

*Nós de Origem, de Processo e de
Saída do IBM SPSS Modeler 17.1*

IBM

Observação

Antes de usar estas informações e o produto por elas suportadas, leia as informações em “Avisos” na página 387.

Informações do produto

Essa edição se aplica à versão 17, liberação 1, modificação 0 do IBM(r) SPSS(r) Modeler e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Prefácio	vii	Importando Dados do Cognos	38
Capítulo 1. Sobre o IBM SPSS Modeler	1	Importando relatórios do Cognos	39
Produtos IBM SPSS Modeler	1	Conexões Cognos	39
IBM SPSS Modeler	1	Seleção de localização do Cognos	40
IBM SPSS Modeler Server	1	Especificando parâmetros para dados ou relatórios	40
IBM SPSS Modeler Administration Console	2	Nó Origem do IBM Cognos TM1	40
IBM SPSS Modeler Batch	2	Importando de dados do IBM Cognos TM1.	41
IBM SPSS Modeler Solution Publisher	2	Nó de Origem do SAS.	42
IBM SPSS Modeler Server Adapters for IBM SPSS Collaboration and Deployment Services	2	Configurando Opções para o Nó de Origem SAS	42
Edições do IBM SPSS Modeler	2	Nó de Origem do Excel	42
Documentação do IBM SPSS Modeler	3	Nó de Origem XML	43
Documentação do SPSS Modeler Professional	3	Selecionando a partir de diversos elementos raiz.	45
Documentação do SPSS Modeler Premium	4	Removendo espaços indesejados dos dados de origem XML	45
Exemplos de Aplicativos	5	Nó Entrada do Usuário	45
Pasta Demos	5	Configurando Opções para o Nó de Entrada do Usuário.	46
Capítulo 2. Nós de Origem.	7	Nó Geração de Simulação	50
Visão geral	7	Configurando Opções para o Nó Geração de Simulação	51
Configurando Armazenamento e Formatação do Campo	9	Clonar Campo	56
Armazenamento de lista e níveis de medição associados	11	Detalhes de Ajuste	57
Caracteres de controle não suportados	12	Especificar parâmetros.	58
Nó de Origem do Analytic Server	12	Distribuições	60
Selecionando uma origem de dados	13	Nó Visualização de Dados	63
Corrigindo as credenciais.	13	Configurando Opções para o Nó Visualização de Dados	64
Nós Suportados	13	Nó de Origem Geoespacial.	64
Nó de Origem da Base de Dados	17	Configurando Opções para o Nó Origem Geoespacial	65
Configurando Opções do Nó de Banco de Dados	18	Nó Visualização Corporativa	65
Incluindo uma conexão com o banco de dados	19	Configurando Opções para o Nó de Visualização Corporativa	66
Especificando valores pré-configurados para uma conexão com o banco de dados.	21	Conexões da Visualização Corporativa	67
Selecionando uma Tabela de Base de Dados	24	Escolhendo o DPD	68
Consultando o Banco de Dados.	25	Escolhendo a Tabela	68
Nó Arquivo Variável	26	Guias do Nó de Origem Comum	68
Configurando Opções para o Nó Arquivo de Variável	27	Configurando Níveis de Medição no Nó de Origem.	68
Importando dados geoespaciais no Nó Arquivo Variável	29	Filtrando Campos do Nó de Origem	69
Nó Arquivo Fixo	30	Capítulo 3. Nós de Operações de Registro	71
Configuração Opções para o Nó Arquivo Fixo..	30	Visão Geral das Operações de Registro	71
Nó do Coleta de dados	32	Selecionar Nó.	72
Opções do Arquivo de Importação do Coleta de dados	32	Nó de Amostra	73
Propriedades de Metadados de Importação do IBM SPSS Data Collection	35	Opções de Nó de Amostra	74
Sequência de Conexão de Banco de Dados	35	Configurações de Cluster e de Estratificação	76
Propriedades Avançadas	36	Tamanhos de Amostra para Camadas.	77
Importando Conjuntos de Múltiplas Respostas	36	Nó Balanceamento	78
Notas de Importação da Coluna do IBM SPSS Data Collection	36	Configurando Opções para o Nó Balanceamento	78
Nó Origem do IBM Cognos BI	37	Nó Agregar	79
Ícones de objeto do Cognos	37	Configurando Opções para o nó Agregado	79
		Agregar configurações de otimização	81

Nó Agregado de RFM	82	Valores dos Dados	137
Configurando Opções para o Nó Agregado de RFM.	82	Definindo Valores Omissos	141
Nó Ordem	83	Verificando Valores de Tipo.	142
Configurações de Otimização de Ordenação ..	84	Configurando o Papel de Campo.	142
Nó Mesclagem	84	Copiando Atributos de Tipo	143
Tipos de Junções	84	Guia Configurações de Formato de Campo ..	143
Especificando Método de Mesclagem e Chaves	86	Filtrando ou Renomeando Campos	145
Selecionando Dados para Junções Parciais . . .	87	Configurando Opções de Filtragem	145
Especificando Condições para uma Mesclagem	88	Derivar Nó	148
Especificando Condições Classificadas para uma	88	Configurando Opções Básicas para o Nó	
Mesclagem	88	Derivar	149
Filtrando Campos com o Nó Mesclagem.	90	Derivando Diversos Campos	149
Configurando a Ordem e a Identificação da		Configurando Opções de Derivar Fórmula ..	150
Entrada.	90	Configurando Opções de Derivar Flag	152
Configurações de Otimização de Mesclagem ..	91	Configurando Opções de Derivar Nominal ..	153
Nó Anexação	92	Configurando Opções de Derivar Estado . . .	153
Configurando Opções de Anexação	92	Configurando Opções de Derivar Contagem ..	154
Nó Distinto	93	Configurando Opções de Derivar Condicional	154
Configurações de Otimização Distintas	95	Recodificando Valores com o Nó Derivar . . .	154
Configurações Compostas Distintas	95	Nó de Preenchimento	155
Nó TS de Fluxo	98	Conversão de Armazenamento Usando o Nó	
Opções do Campo TS de Fluxo.	98	Preenchimento	155
Opções de Modelo do TS de Fluxo	99	Reclassificar Nó	156
Critérios do Modelador Especialista de Séries		Configurando Opções para o Nó Reclassificar	157
Temporais	100	Reclassificando Diversos Campos.	158
Critérios de Suavização Exponencial de Séries		Armazenamento e nível de medição para	
Temporais	100	campos reclassificados	158
Critérios do ARIMA de Séries Temporais . . .	101	Nó Anonimizado	158
Transferir Funções.	102	Configurando opções para o nó anonimizado	159
Manipulando Valores Discrepantes	104	Anonimizando Valores de Campo	160
Opções de Implementação do TS de Fluxo ..	104	Nó de Categorização	160
Nó TCM de Fluxo	105	Configurando Opções para o Nó Categorização	161
Nó TCM de Fluxo - Opções de Séries Temporais	105	Categorias de Largura Fixa	162
Nó TCM de Fluxo - Opções de Observações ..	106	Ladrilhos (contagem igual ou soma).	162
Nó TCM de Fluxo - Opções de Intervalo de		Ranquear Casos	164
Tempo.	107	Desvio Médio/Padrão	164
Nó TCM de Fluxo - Opções de Agregação e de		Categorização ideal	165
Distribuição	107	Visualizando as Categorias Geradas	166
Nó TCM de Fluxo - Opções de Valor Omissos	108	Nó de Análise de RFM	166
Nó TCM de Fluxo - Opções de Dados Gerais	109	Configurações do Nó Análise de RFM	167
Nó TCM de Fluxo - Opções de Construção		Categorização do Nó Análise de RFM	168
Geral	109	Nó Combinação	168
Nó TCM de Fluxo - Opções do Período de		Configurações do Nó Combinação	169
Estimação	109	Nó Partição	170
Nó TCM de Fluxo - Opções de Modelo.	110	Opções do Nó Partição	171
Nó Space-Time-Boxes.	110	Configurar para o Nó Sinalização	172
Definindo densidade de Space-Time-Box ..	112	Configurando Opções para o Nó Configurar	
		para Flag	172
		Nó Reestruturar	173
		Configurando Opções para o Nó Reestruturar	174
		Nó Transpor.	174
		Configurando Opções para o Nó Transpor. ..	174
		Nó Intervalos de Tempo.	176
		Especificando Intervalos de Tempo	177
		Opções de Construção de Intervalo de Tempo	177
		Período de estimação.	178
		Previsões	179
		Intervalos suportados	180
		Nó Histórico	187
		Configurando Opções para o Nó do Histórico	187
		Nó Reordenação de Campo	188
Capítulo 4. Nós de Operações de Campo	115		
Visão Geral de Operações de Campo	115		
Preparação de dados automatizada	117		
Guia Campos	119		
Guia Configurações	119		
Guia Análise	124		
Gerando um Nó Derivar	130		
Nó Tipo	131		
Níveis de Medição	133		
Convertendo Dados Contínuos	136		
O Que É Instanciação?	137		

Configurando Opções de Reordenação de Campo	188
Nó Intervalo de tempo do AS	189
Intervalo de tempo do AS - opções de campo	190
Intervalo de tempo do AS - opções de criação	190
Nó Reprojecção	191
Configurando opções para o nó Reprojetar . .	191
Capítulo 5. Nós de Gráfico.	193
Variáveis de Nós de Gráfico Comuns	193
Estética, Sobreposições, Painéis e Animação . .	194
Utilizando a Guia Saída	195
Utilizando a Guia Anotações	196
Gráficos 3D	196
Nó Elemento do Gráfico.	198
Guia Básica do Gráfico	198
Guia Detalhada do Gráfico	202
Tipos de Visualizações de Gráfico Integradas Disponíveis	204
Criando Visualizações de Mapa	211
Exemplos de Gráficos	212
Guia Aparência do Gráfico	222
Configurando a Localização dos Modelos, Folhas de Estilos e Mapas	223
Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento	224
Convertendo e Distribuindo Arquivos de Forma de Mapa	225
Conceitos Chave para Mapas	226
Usando o Utilitário de Conversão de Mapa . .	226
Distribuindo arquivos de mapas	232
Nó Gráfico	232
Guia Nó do Gráfico	235
Guia Opções de Gráfico	236
Guia Aparência do Gráfico	238
Utilizando um Gráfico de Plotagem	238
Nó Multigráficos	239
Guia Gráfico de Multigráficos	239
Guia Aparência de Multigráficos	241
Utilizando um Gráfico Multigráficos.	241
Nó Gráfico de Tempo	242
Guia Gráfico de Tempo	243
Guia Aparência do Gráfico de Tempo	244
Utilizando um Gráfico de Plotagem de Tempo	244
Nó Distribuição	245
Guia Gráfico de Distribuição	245
Guia Aparência da Distribuição	246
Usando um Nó de Distribuição	246
Nó Histograma	249
Guia Gráfico de Histograma	249
Guia Opções do Histograma	249
Guia Aparência do Histograma	250
Utilizando Histogramas	250
Nó Coleção	251
Guia Gráfico de Coleção.	251
Guia Opções de Coleção.	252
Guia Aparência da Coleção.	252
Utilizando um Gráfico de Coleção	253
Nó da Web	254
Guia Gráfico da Web	255
Guia Opções da Web	256

Guia Aparência da Web	258
Utilizando um Gráfico da Web	258
Nó Avaliação	262
Guia Gráfico de Avaliação	266
Guia Opções de Avaliação	268
Guia Aparência de Avaliação	269
Lendo os Resultados de uma Avaliação de Modelo	269
Usando um Gráfico de Avaliação.	270
Nó de Visualização de Mapa	271
Guia Gráfico de Visualização de Mapa	271
Guia Aparência de Visualização de Mapa . . .	275
Explorando Gráficos	275
Usando Bandas.	276
Utilizando Regiões	279
Usando Elementos Marcados	281
Gerando Nós a partir de Gráficos	282
Editando Visualizações	285
Regras Gerais para Visualizações de Edição . .	286
Editando e Formatando Texto	287
Alterando Cores, Padrões, Traço e Transparência	287
Girando e Alterando a Forma e a Proporção de Aspecto de Elementos de Ponto	288
Alterando o Tamanho dos Elementos Gráficos	288
Especificando Margens e Preenchimento	289
Formatando Números	289
Alterando as Configurações de Eixo e Escala	290
Editando Categorias	291
Alterando os Painéis de Orientação	293
Transformando o Sistema de Coordenadas . .	293
Alterando Estatísticas e Elementos Gráficos . .	294
Alterando a Posição da Legenda	295
Copiando uma Visualização e Dados de Visualização	295
Atalhos de Teclado	296
Incluindo Títulos e Notas de Rodapé	296
Utilizando Folhas de Estilo do Gráfico	296
Imprimindo, Salvando, Copiando e Exportando Gráficos	298

Capítulo 6. Nós de Saída	301
Visão Geral dos Nós de Saída	301
Gerenciando a saída	302
Visualizando a Saída	302
Publicação na Web	303
Visualizando a Saída em um Navegador HTML	304
Exportando Saída	304
Selecionando células e colunas	305
Nó Tabela	305
Guia Configurações do Nó Tabela	305
Guia Formato do Nó da Tabela	306
Guia Saída do Nó Saída	306
Navegador de Tabela.	307
Nó Matriz	308
Guia Configurações do Nó Matriz	308
Guia Aparência do Nó Matriz	309
Navegador de Saída do Nó Matriz	309
Nó Análise	310
Guia Análise do Nó Análise	311
Navegador de Saída da Análise	312
Nó Auditoria de Dados	314

Guia Configurações de Nó Auditoria de Dados	314
Guia de Qualidade de Auditoria de Dados	315
Navegador de Saída de Auditoria de Dados	315
Nó Transformação	321
Guia Opções do nó Transformação	321
Guia Saída do Nó Transformação	322
Visualizador de Saída do Nó Transformação	322
Nó Estatísticas	324
Guia Configurações do Nó Estatísticas	324
Navegador de Saída de Estatísticas	325
Nó Médias	326
Comparando Médias para Grupos de Independentes	326
Comparando Médias Entre Campos Emparelhados	326
Opções do Nó Médias	327
Navegador de Saída do Nó Médias	327
Nó Relatório	328
Guia Modelo do Nó Relatório	329
Navegador de Saída do Nó Relatório	330
Nó Configurar Globais	330
Guia Configurações do Nó Configurar Globais	330
Nó Ajuste de Simulação	331
Ajuste de distribuição	331
Guia Configurações do Nó Ajuste de Simulação	333
Nó Avaliação de Simulação	333
Guia Configurações do Nó de Avaliação de Simulação	334
Saída do Nó Avaliação de Simulação	336
Aplicativos Auxiliares do IBM SPSS Statistics	341
Capítulo 7. Exportar Nós	343
Visão Geral dos Nós de Exportação	343
Nó Exportação do Banco de Dados	344
Guia Exportação do Nó do Banco de Dados	344
Opções de Mesclagem de Exportação de Banco de Dados	345
Opções do Esquema de Exportação de Banco de Dados	346
Opções do Índice de Exportação do Banco de Dados	349
Opções Avançadas de Exportação de Banco de Dados	350
Programação do utilitário de carregamento em massa	352
Nó Exportação de Arquivo Simples	358
Guia Exportação de Arquivo Simples	359
Nó Exportação do IBM SPSS Data Collection	359
Nó Exportação do Analytic Server	360
Nó Exportação do IBM Cognos BI	361
Conexão do Cognos	361
Conexão ODBC	362
Nó Exportação do IBM Cognos TM1	363
Conectando-se a um cubo IBM Cognos TM1 para exportar dados	363

Mapeando dados do IBM Cognos TM1 para exportação	364
Nó Exportação do SAS	365
Guia Exportação do Nó Exportação do SAS	365
Nó Exportação do Excel	365
Guia Exportação do Nó do Excel	365
Nó Exportação XML	366
Gravando dados XML	367
Opções de registros de mapeamento XML	367
Opções de Campos de Mapeamento XML	367
Visualização de mapeamento XML	368

Capítulo 8. IBM SPSS Statistics Nós **369**

Nós do IBM SPSS Statistics - Visão Geral	369
Nó Arquivo de Estatísticas	370
Nó Transformação de Estatísticas	371
Nó Transformação de Estatísticas – Guia Sintaxe	371
Sintaxe Permissível	372
Nó Modelo de Estatísticas	374
Nó do Modelo de Estatísticas - Guia Modelo	374
Nó do Modelo de Estatísticas - Sumarização do Nugget do Modelo	374
Nó Saída de Estatísticas	374
Nó de Saída de Estatísticas – Guia Sintaxe	375
Nó de Saída de Estatísticas – Guia Saída	376
Nó Exportação de Estatísticas	377
Nó Exportação de Estatísticas - Guia Exportação	377
Renomeando ou Filtrando Campos para o IBM SPSS Statistics	378

Capítulo 9. SuperNodes **379**

Visão Geral de SuperNodes	379
Tipos de SuperNodes	379
SuperNodes de Origem	379
SuperNodes de Processo	379
SuperNodes Terminais	380
Criando SuperNodes	380
Aninhando SuperNodes	381
Bloqueando SuperNodes	381
Bloqueando e Desbloqueando um SuperNode	381
Editando um SuperNode Bloqueado	382
Editando SuperNodes	382
Modificando Tipos de SuperNode	382
Anotando e Renomeando SuperNodes	382
Parâmetros de SuperNode	383
SuperNodes e Armazenamento em Cache	385
SuperNodes e Script	385
Salvando e Carregando SuperNodes	386

Avisos **387**

Marcas comerciais	388
-------------------	-----

Índice Remissivo **393**

Prefácio

IBM® SPSS Modeler é o ambiente de trabalho de mineração de dados de força corporativa do IBM Corp.. O SPSS Modeler ajuda as organizações a melhorarem as relações com o cliente e com o cidadão por meio de um entendimento profundo dos dados. As organizações utilizam o insight adquirido do SPSS Modeler para reter clientes rentáveis, identificar oportunidades de venda cruzada, atrair novos clientes, detectar fraude, reduzir o risco e melhorar a entrega de serviço de governo.

A interface visual do SPSS Modeler convida os usuários a aplicarem seus conhecimentos de negócios específicos, levando a modelos preditivos mais poderosos e reduzindo o tempo para a solução. O SPSS Modeler oferece muitas técnicas de modelagem, como previsão, classificação, segmentação e algoritmos de detecção de associação. Quando os modelos são criados, o IBM SPSS Modeler Solution Publisher permite entregá-los aos tomadores de decisão na empresa ou a um banco de dados.

Sobre o IBM Business Analytics

O software IBM Business Analytics fornece informações completas, consistentes e exatas nas quais os tomadores de decisão confiam para melhorar o desempenho de negócios. Um portfólio abrangente de inteligência de negócios, análise preditiva, gerenciamento de desempenho financeiro e estratégia e aplicativos analíticos fornece insight claro, imediato e utilizável sobre o desempenho atual e a capacidade de prever resultados futuros. Combinado com ricas soluções para segmentos de mercados, práticas comprovadas e serviços profissionais, as organizações de qualquer tamanho podem conduzir a maior produtividade, automatizar decisões com confiança e fornecer melhores resultados.

Como parte deste portfólio, o software IBM SPSS Predictive Analytics ajuda as organizações a prever eventos futuros e agir proativamente nesse insight para conduzir os melhores resultados de negócios. Os clientes acadêmicos, comerciais e do governo no mundo todo se baseiam na tecnologia do IBM SPSS como uma vantagem competitiva para atrair, manter e aumentar seus clientes, enquanto reduz fraudes e minimiza riscos. Ao incorporar o software IBM SPSS em suas operações diárias, as organizações se tornam empresas preditivas - capazes de direcionar e automatizar decisões para atenderem às metas de negócios e alcançarem vantagem competitiva mensurável. Para obter informações adicionais ou para entrar em contato com um representante, visite <http://www.ibm.com/spss>.

Suporte técnico

Suporte técnico está disponível para clientes de manutenção. Os clientes podem entrar em contato com o Suporte Técnico para obter assistência com o uso de produtos IBM Corp. ou para obterem ajuda com a instalação para um dos ambientes de hardware suportados. Para obter o Suporte Técnico, consulte o website do IBM Corp. em <http://www.ibm.com/support>. Esteja preparado para se identificar, sua organização e seu contrato de suporte ao solicitar assistência.

Capítulo 1. Sobre o IBM SPSS Modeler

O IBM SPSS Modeler é um conjunto de ferramentas de mineração de dados que permite desenvolver rapidamente modelos preditivos usando o conhecimento de negócios, e implementá-los em operações de negócios para melhorar a tomada de decisão. Projetado em torno do modelo CRISP-DM padrão de mercado, o IBM SPSS Modeler suporta todo o processo de mineração de dados, a partir dos dados para melhores resultados de negócios.

O IBM SPSS Modeler oferece uma variedade de métodos de modelagem a partir do aprendizado de máquina, inteligência artificial e estatísticas. Os métodos disponíveis na paleta Modelagem permitem derivar informações novas a partir dos dados, e desenvolver modelos preditivos. Cada método possui determinadas intensidades e é mais bem adequado para tipos de problemas específicos.

O SPSS Modeler pode ser comprado como um produto independente, ou usado como um cliente na combinação com o SPSS Modeler Server. Várias opções adicionais também estão disponíveis, conforme resumidas nas seções a seguir. Para obter mais informações, consulte <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Produtos IBM SPSS Modeler

A família de produtos IBM SPSS Modeler e o software associado abrangem o seguinte.

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- Adaptadores IBM SPSS Modeler Server para IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler é uma versão funcionalmente completa do produto que você instala e executa em seu computador pessoal. É possível executar o SPSS Modeler no modo local como um produto independente ou usá-lo no modo distribuído com IBM SPSS Modeler Server para melhorar o desempenho em conjuntos de dados grandes.

Com o SPSS Modeler, é possível construir modelos preditivos exatos de maneira rápida e intuitiva, sem programação. Usando a interface visual exclusiva, é possível visualizar facilmente o processo de mineração de dados. Com o suporte da análise avançada integrada ao produto, é possível descobrir tendências e padrões ocultos anteriormente em seus dados. É possível modelar resultados e entender os fatores que os influenciam, permitindo que você aproveite as vantagens das oportunidades de negócios e diminua os riscos.

SPSS Modeler está disponível em duas edições: SPSS Modeler Professional e SPSS Modeler Premium. Consulte o tópico “Edições do IBM SPSS Modeler” na página 2 para obter mais informações.

IBM SPSS Modeler Server

SPSS Modeler usa uma arquitetura de cliente/servidor para distribuir solicitações para operações cheias de recursos para poderosos softwares de servidor, resultando em desempenho mais rápido em conjuntos de dados maiores.

SPSS Modeler Server é um produto licenciado separadamente que é executado de forma contínua no modo de análise distribuído em um host do servidor com uma ou mais instalações do IBM SPSS Modeler. Dessa maneira, o SPSS Modeler Server fornece desempenho superior em conjuntos de dados grandes, pois operações com uso intensivo de memória podem ser executadas no servidor sem fazer download dos dados no computador cliente. IBM SPSS Modeler Server também fornece suporte para otimização de SQL e capacidades de modelagem dentro da base de dados, entregando mais benefícios para o desempenho e a automação.

IBM SPSS Modeler Administration Console

O Modeler Administration Console é um aplicativo gráfico para gerenciar muitas das opções de configuração do SPSS Modeler Server, que também são configuráveis por meio de um arquivo de opções. O aplicativo fornece uma interface com o usuário do console para monitorar e configurar suas instalações do SPSS Modeler Server e está disponível gratuitamente para os atuais clientes do SPSS Modeler Server. O aplicativo pode ser instalado somente em computadores Windows; no entanto, ele pode administrar um servidor instalado em qualquer plataforma suportada.

IBM SPSS Modeler Batch

Embora geralmente a mineração de dados seja um processo interativo, também é possível executar o SPSS Modeler a partir de uma linha de comandos, sem a necessidade de uma interface gráfica com o usuário. Por exemplo, você pode ter tarefas repetidas ou de longa execução que deseja executar sem intervenção do usuário. SPSS Modeler Batch é uma versão especial do produto que fornece suporte para capacidades de análise completa do SPSS Modeler sem acessar a interface com o usuário regular. SPSS Modeler Server é necessário para usar o SPSS Modeler Batch.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher é uma ferramenta que permite criar uma versão do pacote de um fluxo do SPSS Modeler que pode ser executado por um mecanismo de tempo de execução externo ou integrado a um aplicativo externo. Dessa maneira, é possível publicar e implementar fluxos completos do SPSS Modeler para uso em ambientes que não têm o SPSS Modeler instalado. SPSS Modeler Solution Publisher é distribuído como parte do serviço IBM SPSS Collaboration and Deployment Services - Scoring, para o qual uma licença separada é necessária. Com essa licença, você recebe o SPSS Modeler Solution Publisher Runtime, que permite executar os fluxos publicados.

Para obter mais informações sobre SPSS Modeler Solution Publisher, consulte a documentação do IBM SPSS Collaboration and Deployment Services. O IBM SPSS Collaboration and Deployment Services Knowledge Center contém seções chamadas "IBM SPSS Modeler Solution Publisher" e "IBM SPSS Analytics Toolkit."

IBM SPSS Modeler Server Adapters for IBM SPSS Collaboration and Deployment Services

Inúmeros adaptadores para o IBM SPSS Collaboration and Deployment Services estão disponíveis para permitir que o SPSS Modeler e o SPSS Modeler Server interajam com um repositório do IBM SPSS Collaboration and Deployment Services. Dessa forma, um fluxo do SPSS Modeler implementado no repositório pode ser compartilhado por diversos usuários ou acessado a partir do aplicativo thin client IBM SPSS Modeler Advantage. Você instala o adaptador no sistema que hospeda o repositório.

Edições do IBM SPSS Modeler

SPSS Modeler está disponível nas seguintes edições.

SPSS Modeler Professional

SPSS Modeler Professional fornece todas as ferramentas necessárias para você trabalhar com a maioria dos tipos de dados estruturados, como comportamentos e interações controlados em sistemas CRM,

demográficos, comportamento de compra e dados de vendas.

SPSS Modeler Premium

SPSS Modeler Premium é um produto licenciado separadamente que estende o SPSS Modeler Professional para trabalhar com dados especificados, como aqueles usados para análise de entidade ou rede social, e com dados de texto não estruturados. SPSS Modeler Premium abrange os componentes a seguir.

O **IBM SPSS Modeler Entity Analytics** inclui uma dimensão extra à análise preditiva do IBM SPSS Modeler. Considerando que a análise preditiva tenta prever comportamento futuro de dados passados, a análise de entidade foca melhorar a coerência e a consistência dos dados atuais resolvendo conflitos de identidade nos registros em si. Uma identidade pode ser de um indivíduo, de uma organização, de um objeto ou de qualquer outra entidade para a qual possa existir ambiguidade. A resolução de identidade pode ser vital em diversos campos, incluindo gerenciamento de relacionamento com o cliente, detecção de fraude, medidas contra lavagem de dinheiro e segurança nacional e internacional.

IBM SPSS Modeler Social Network Analysis transforma informações sobre relacionamentos em campos que caracterizam o comportamento social dos indivíduos e grupos. Usando dados que descrevem os relacionamentos subjacentes das redes sociais, o IBM SPSS Modeler Social Network Analysis identifica os líderes sociais que influenciam o comportamento dos outros na rede. Além disso, é possível determinar quais pessoas são mais afetadas por outros participantes da rede. Combinando esses resultados com outras medidas, é possível criar perfis abrangentes de indivíduos nos quais basear modelos preditivos. Os modelos que incluem essas informações sociais serão executados melhor do que os modelos que não incluem.

IBM SPSS Modeler Text Analytics usa tecnologias de linguística avançada e processamento de linguagem natural (NLP) para processar rapidamente uma grande variedade de dados de texto não estruturados, extrair e organizar conceitos chave e agrupar esses conceitos em categorias. Categorias e conceitos extraídos podem ser combinados com dados estruturados existentes, como demográficos, e aplicados à modelagem usando o conjunto completo de ferramentas de mineração de dados do IBM SPSS Modeler para gerar decisões melhores e mais focadas.

Documentação do IBM SPSS Modeler

A documentação em formato de ajuda online está disponível no menu Ajuda do SPSS Modeler. Isso inclui documentação para SPSS Modeler, SPSS Modeler Server, bem como o Guia de Aplicativos (também chamado de Tutorial) e outros materiais de apoio.

A documentação completa para cada produto (incluindo instruções de instalação) está disponível em formato PDF sob a pasta *Documentation* em cada produto DVD. Os documentos de instalação também podem ser transferidos por download a partir da web em <http://www.ibm.com/support/docview.wss?uid=swg27046116>.

A documentação em ambos os formatos também estão disponíveis a partir do SPSS Modeler Knowledge Center em http://www-01.ibm.com/support/knowledgecenter/SS3RA7_17.1.0.

Documentação do SPSS Modeler Professional

O conjunto de documentações do SPSS Modeler Professional (excluindo instruções de instalação) é o seguinte.

- **IBM SPSS Modeler User's Guide.** Introdução geral ao uso do SPSS Modeler, incluindo como construir fluxos de dados, manipular valores omissos, construir expressões do CLEM, trabalhar com projetos e relatórios e empacotar fluxos para implementação no IBM SPSS Collaboration and Deployment Services, Predictive Applications ou IBM SPSS Modeler Advantage.

- **IBM SPSS Modeler Source, Process, and Output Nodes.** Descrições de todos os nós usados para ler, processar e emitir dados em diferentes formatos. Efetivamente, isso significa todos os nós além dos de modelagem.
- **IBM SPSS Modeler Modeling Nodes.** Descrições de todos os nós usados para criar modelos de mineração de dados. O IBM SPSS Modeler oferece uma variedade de métodos de modelagem a partir do aprendizado de máquina, inteligência artificial e estatísticas.
- **IBM SPSS Modeler Algorithms Guide.** Descrições das bases matemáticas dos métodos de modelagem usados no IBM SPSS Modeler. Este guia está disponível somente em formato PDF.
- **IBM SPSS Modeler Applications Guide.** Os exemplos neste guia fornecem introduções sintetizadas e direcionadas para técnicas e métodos de modelagem específicos. Uma versão online deste guia também está disponível no menu Ajuda. Consulte o tópico “Exemplos de Aplicativos” na página 5 para obter mais informações.
- **IBM SPSS Modeler Python Scripting and Automation.** Informações sobre como automatizar o sistema por meio de script Python, incluindo as propriedades que podem ser usadas para manipular nós e fluxos.
- **IBM SPSS Modeler Deployment Guide.** Informações sobre como executar fluxos e cenários do IBM SPSS Modeler como passos no processamento de tarefas sob IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler CLEF Developer's Guide.** CLEF fornece a capacidade de integrar programas de terceiros, como rotinas de processamento de dados ou algoritmos de modelagem como nós, no IBM SPSS Modeler.
- **IBM SPSS Modeler In-Database Mining Guide.** Informações sobre como usar o poder do seu banco de dados para melhorar o desempenho e ampliar o intervalo de capacidades analíticas por meio de algoritmos de terceiros.
- **IBM SPSS Modeler Server Administration and Performance Guide.** Informações sobre como configurar e administrar o IBM SPSS Modeler Server.
- **IBM SPSS Modeler Administration Console User Guide.** Informações sobre a instalação e o uso da interface com o usuário do console para monitorar e configurar o IBM SPSS Modeler Server. O console é implementado como um plug-in no aplicativo Deployment Manager.
- **Guia do IBM SPSS Modeler CRISP-DM.** Guia passo a passo para o uso da metodologia CRISP-DM para mineração de dados com SPSS Modeler.
- **IBM SPSS Modeler Batch User's Guide.** Guia completo para o uso do IBM SPSS Modeler no modo em lote, incluindo detalhes da execução do modo em lote e argumentos de linha de comandos. Este guia está disponível somente em formato PDF.

Documentação do SPSS Modeler Premium

O conjunto de documentações do SPSS Modeler Premium (excluindo instruções de instalação) é o seguinte.

- **IBM SPSS Modeler Entity Analytics User Guide.** Informações sobre o uso de análise de entidade com SPSS Modeler, cobrindo instalação e configuração de repositório, nós de análise de entidade e tarefas administrativas.
- **IBM SPSS Modeler Social Network Analysis User Guide.** Um guia para executar análise de rede social com SPSS Modeler, incluindo análise de grupo e análise de difusão.
- **SPSS Modeler Text Analytics User's Guide.** Informações sobre o uso de analítica de texto com SPSS Modeler, cobrindo os nós de mineração de texto, ambiente de trabalho interativo, modelos e outros recursos.

Exemplos de Aplicativos

Enquanto as ferramentas de mineração de dados no SPSS Modeler podem ajudar a resolver uma ampla variedade de negócios e problemas organizacionais, os exemplos de aplicativos fornecem introduções breves e destinadas aos métodos e técnicas de modelagem específicos. Os conjuntos de dados usados aqui são muito menores do que os enormes armazenamentos de dados gerenciados por alguns mineradores de dados, mas os conceitos e métodos envolvidos devem ser escaláveis para aplicativos reais.

É possível acessar os exemplos clicando em **Exemplos de Aplicativos** no menu de Ajuda no SPSS Modeler. Os arquivos de dados e os fluxos de amostra são instalados na pasta *Demos* no diretório de instalação do produto. Consulte o tópico “Pasta Demos” para obter mais informações.

Exemplos de modelagem da base de dados. Consulte os exemplos no *Guia de Mineração dentro do Banco de Dados do IBM SPSS Modeler*.

Exemplos de script. Consulte os exemplos no *Guia de Script e Automação do IBM SPSS Modeler*.

Pasta Demos

Os arquivos de dados e fluxos de amostras usados com exemplos de aplicativo são instalados na pasta *Demos* sob o diretório de instalação do produto. Essa pasta também pode ser acessada a partir do grupo do programa IBM SPSS Modeler no menu Iniciar do Windows ou clicando em *Demos* na lista de diretórios recentes na caixa de diálogo Arquivo Aberto.

Capítulo 2. Nós de Origem

Visão geral

Os nós de origem permitem importar dados armazenados em vários formatos, incluindo arquivos simples, o IBM SPSS Statistics (.sav), SAS, Microsoft Excel e bancos de dados relacionais compatíveis com ODBC. Também é possível gerar dados sintéticos utilizando o nó Entrada do Usuário.

A paleta Origens contém os seguintes nós:



A origem do Analytic Server permite executar um fluxo no Hadoop Distributed File System (HDFS). As informações em uma origem de dados do Analytic Server vêm de uma variedade de locais, como arquivos de texto e bancos de dados. Consulte o tópico “Nó de Origem do Analytic Server” na página 12 para obter mais informações.



O nó Banco de Dados pode ser utilizado para importar dados de uma variedade de outros pacotes utilizando ODBC (Open Database Connectivity), incluindo Microsoft SQL Server, o DB2, o Oracle e outros. Consulte o tópico “Nó de Origem da Base de Dados” na página 17 para obter mais informações.



O nó Arquivo Variável lê arquivos de texto de campo livre, ou seja, arquivos cujos registros contêm um número constante dos campos, mas um número variado de caracteres. Esse nó também é útil para arquivos com texto de cabeçalho de comprimento fixo e certos tipos de anotações. Consulte o tópico “Nó Arquivo Variável” na página 26 para obter mais informações.



O nó Arquivo Fixo importa dados de arquivos de texto de campo fixo, ou seja, arquivos cujos campos não são delimitados, mas iniciam na mesma posição e têm um comprimento fixo. Dados gerados por máquina ou legados são frequentemente armazenados em formato de campo fixo. Consulte o tópico “Nó Arquivo Fixo” na página 30 para obter mais informações.



O nó Arquivo de Estatísticas lê dados do formato de arquivo .sav ou .zsave usado pelo IBM SPSS Statistics, bem como de arquivos de cache salvos em IBM SPSS Modeler que também utilizam o mesmo formato.



O nó do IBM SPSS Data Collection importa dados da pesquisa de opinião a partir de vários formatos usados pelo software de pesquisa de mercado em conformidade com o Modelo de Dados do IBM SPSS Data Collection. O IBM SPSS Data Collection Developer Library deve ser instalado para utilizar esse nó. Consulte o tópico “Nó do Coleta de dados” na página 32 para obter mais informações.



O nó de origem do IBM Cognos BI importa dados a partir de bancos de dados do Cognos BI.



O nó de origem do IBM Cognos TM1 importa dados a partir de bancos de dados do Cognos TM1.



O nó Arquivo SAS importa dados do SAS no IBM SPSS Modeler. Consulte o tópico “Nó de Origem do SAS” na página 42 para obter mais informações.



O nó do Excel importa dados do Microsoft Excel no formato de arquivo .xlsx. Uma origem de dados ODBC não é necessária. Consulte o tópico “Nó de Origem do Excel” na página 42 para obter mais informações.



O nó de origem XML importa dados no formato XML para o fluxo. É possível importar um único arquivo ou todos os arquivos em um diretório. É possível, opcionalmente, especificar um arquivo de esquema a partir do qual a estrutura XML é lida.



O nó Entrada do Usuário fornece uma maneira fácil de criar dados sintéticos, seja desde o início ou alterando dados existentes. Isso é útil, por exemplo, quando desejar criar um conjunto de dados de teste para modelagem. Consulte o tópico “Nó Entrada do Usuário” na página 45 para obter mais informações.



O nó Gerar Simulação fornece uma maneira fácil de gerar dados simulados, seja desde o início utilizando distribuições de estatísticas especificadas pelo usuário ou automaticamente utilizando as distribuições obtidas da execução de um nó Ajuste de Simulação em dados históricos existentes. Isso é útil quando quiser avaliar o resultado de um modelo preditivo quando houver incerteza nas entradas do modelo.



O nó Visualização de Dados pode ser usado para acessar origens de dados definidas em visualizações de dados analíticos do IBM SPSS Collaboration and Deployment Services. Uma visualização de dados analíticos define uma interface padrão para acessar dados e associa diversas origens de dados físicas a essa interface. Consulte o tópico “Nó Visualização de Dados” na página 63 para obter mais informações.



Utilize o nó de origem Geoespacial para exibir dados de mapa ou espaciais na sessão de mineração de dados. Consulte o tópico “Nó de Origem Geoespacial” na página 64 para obter mais informações.



Nota: O nó Visualização Corporativa foi substituído no IBM SPSS Modeler 16.0 pelo nó Visualização de Dados. Para fluxos salvos em liberações anteriores, o nó Visualização Corporativa ainda é suportado. No entanto, ao atualizar ou criar novos fluxos, recomenda-se utilizar o nó Visualização de Dados. O nó Visualização Corporativa cria uma conexão com um IBM SPSS Collaboration and Deployment Services Repository, permitindo ler dados da Visualização Corporativa em um fluxo e empacotar um modelo em um cenário que possa ser acessado a partir do repositório por outros usuários. Consulte o tópico “Nó Visualização Corporativa” na página 65 para obter mais informações.

Para iniciar um fluxo, inclua um nó de origem na tela de fluxo. Em seguida, dê um clique duplo no nó para abrir sua caixa de diálogo. As várias guias na caixa de diálogo permitem ler os dados, visualizar os campos e valores e configurar uma variedade de opções, dentre elas filtros, tipos de dados, papel do campo e verificação de valor omissos.

Configurando Armazenamento e Formatação do Campo

As opções na guia Dados para os nós Arquivo Fixo, Arquivo Variável, Origem XML e Entrada do Usuário permitem especificar o tipo de armazenamento para os campos conforme eles são importados ou criados no IBM SPSS Modeler. Para os nós Arquivo Fixo, Arquivo Variável e Entrada do Usuário, também é possível especificar a formatação de campo e outros metadados.

Para os dados lidos a partir de outras origens, o armazenamento é determinado automaticamente, mas pode ser alterado utilizando uma função de conversão, como `to_integer`, em um nó Preenchimento ou nó Derivar.

Campo Utilize a coluna **Campo** para visualizar e selecionar campos no conjunto de dados atual.

Substituir Marque a caixa de seleção na coluna **Substituir** para ativar as opções nas colunas **Armazenamento** e **Formato de Entrada**.

Armazenamento de Dados

Armazenamento descreve a maneira como os dados são armazenados em um campo. Por exemplo, um campo com valores de 1 e 0 armazena dados de número inteiro. Isso é diferente do nível de medição, que descreve o uso dos dados e não afeta o armazenamento. Por exemplo, talvez você queira configurar o nível de medição para um campo de número inteiro com valores de 1 e 0 para *Flag*. Isto geralmente indica que 1 = *True* e 0 = *False*. Embora o armazenamento deva ser determinado na origem, o nível de medição pode ser alterado utilizando um nó Tipo em qualquer ponto no fluxo. Consulte o tópico “Níveis de Medição” na página 133 para obter mais informações.

Os tipos de armazenamento disponíveis são:

- **Sequência de Caracteres** Usado para campos que contêm dados não numéricos, também chamados de dados alfanuméricos. Uma sequência de caracteres pode incluir qualquer sequência de caracteres, como *fred*, *Classe 2* ou *1234*. Observe que os números em sequências de caracteres não podem ser usados em cálculos.
- **Número Inteiro** Um campo cujos valores são números inteiros.
- **Real** Valores são números que podem incluir decimais (não se limitando a números inteiros). O formato de exibição é especificado na caixa de diálogo Propriedades do Fluxo e pode ser substituído para campos individuais em um nó Tipo (guia Formato).
- **Data** Valores de data especificados em um formato padrão como ano, mês e dia (por exemplo, 2007-09-26). O formato específico é definido na caixa de diálogo Propriedades do Fluxo.
- **Tempo** Tempo medido como uma duração. Por exemplo, uma chamada de serviço que dura 1 hora, 26 minutos e 38 segundos pode ser representada como 01:26:38, dependendo do formato de horário atual, conforme especificado na caixa de diálogo Propriedades do Fluxo.
- **Registro de Data e Hora** Valores que incluem um componente de data e hora, por exemplo, 2007-09-26 09:04:00, mais uma vez dependendo dos formatos de data e hora atuais na caixa de diálogo Propriedades do Fluxo. Observe que os valores de registro de data e hora podem precisar ser incluídos entre aspas duplas para garantir que eles sejam interpretados como um valor único ao invés valores de data e hora separados. (Isto se aplica, por exemplo, ao inserir valores em um nó Entrada do Usuário).
- **Lista** Introduzido no SPSS Modeler versão 17, junto dos novos níveis de medição Geoespacial e de Coleção, um campo de armazenamento Lista contém diversos valores para um único registro. Há versões de lista de todos os outros tipos de armazenamento.

Tabela 1. Ícones de tipo de armazenamento de Lista

Ícone	Tipo de armazenamento
[A]	Lista de sequência de caracteres
[♦]	Lista de número inteiro
[◆]	Lista de números reais
[🕒]	Lista de tempo
[📅]	Lista de data
[🕒📅]	Lista de registro de data e hora
[[]]	Lista com uma profundidade maior que zero

Além disso, para uso com o nível de medição Coleção, há versões de lista dos seguintes níveis de medição.

Tabela 2. Ícones do nível de medição de lista

Ícone	Nível de medição
[✏️]	Lista de contínuo
[🗨️]	Lista de categórico
[🚩]	Lista de flags
[🎯]	Lista de nominal
[📊]	Lista de ordinal

As listas podem ser importadas no SPSS Modeler em um dos três nós de origem (Analytic Server, Geoespacial ou Arquivo de Variável) ou criadas dentro de seus fluxos através do uso dos nós de operação de campo Derivar ou Preenchimento.

Para obter mais informações sobre as Listas e a interação delas com os níveis de medição Coleção e Geoespacial, consulte “Armazenamento de lista e níveis de medição associados” na página 11

Conversões de armazenamento. É possível converter armazenamento de um campo utilizando uma variedade de funções de conversão, como `to_string` e `to_integer`, em um nó Preenchimento. Consulte o tópico “Conversão de Armazenamento Usando o Nó Preenchimento” na página 155 para obter mais informações. Observe que as funções de conversão (e quaisquer outras funções que requerem um tipo específico de entrada, como um valor de data e hora) dependem dos formatos atuais especificados na caixa de diálogo Propriedades do Fluxo. Por exemplo, se desejar converter um campo de sequência de caracteres com valores *Jan 2003*, *Feb 2003*, (e assim por diante), no armazenamento de data, selecione **MON YYYY** como o formato de data padrão para o fluxo. As funções de conversão também estão disponíveis a partir do nó Derivar, para conversão temporária durante um cálculo de derivar. Também é possível utilizar o nó Derivar para executar outras manipulações, como recodificar campos de sequência de caracteres com valores categóricos. Consulte o tópico “Recodificando Valores com o Nó Derivar” na página 154 para obter mais informações.

Leitura de dados combinados. Observe que ao ler campos com armazenamento numérico (número inteiro, real, hora, registro de data e hora ou data), quaisquer valores não numéricos são configurados para nulos ou omissos do sistema. Isso ocorre porque, ao contrário de alguns aplicativos, o IBM SPSS Modeler não permite tipos de armazenamento combinados em um campo. Para evitar isso, quaisquer campos com dados combinados devem ser lidos como sequências de caracteres ao alterar o tipo de armazenamento no nó de origem ou nos aplicativos externos, conforme necessário.

Formato de Entrada de Campo (somente nós Arquivo Fixo, Arquivo Variável e Entrada do Usuário)

Para todos os tipos de armazenamento, exceto Sequência de Caracteres e Número Inteiro, é possível especificar opções de formatação para o campo selecionado utilizando a lista suspensa. Por exemplo, ao mesclar dados de diversos códigos de idiomas, poderá ser necessário especificar um ponto (.) como o separador decimal para um campo, ao passo outro exigirá um separador de vírgula.

As opções de entrada especificadas no nó de origem substituem as opções de formatação especificadas na caixa de diálogo de propriedades do fluxo, entretanto, elas não persistem posteriormente no fluxo. Elas têm o objetivo de analisar a entrada corretamente com base em seu conhecimento dos dados. Os formatos especificados são utilizados como um guia para analisar os dados conforme eles são lidos no IBM SPSS Modeler, não para determinar como eles devem ser formatados após serem lidos no IBM SPSS Modeler. Para especificar a formatação em uma base por campo em outro lugar no fluxo, utilize a guia Formato de um nó Tipo. Consulte o tópico “Guia Configurações de Formato de Campo” na página 143 para obter mais informações.

As opções variam dependendo do tipo de armazenamento. Por exemplo, para o tipo de armazenamento Real, é possível selecionar **Ponto (.)** ou **Vírgula (,)** como o separador decimal. Para campos de registro de data e hora, uma caixa de diálogo separada se abre ao selecionar **Especificar** na lista suspensa. Consulte o tópico “Configurando Opções de Formato de Campo” na página 144 para obter mais informações.

Para todos os tipos de armazenamento, também é possível selecionar **Padrão de fluxo** para utilizar as configurações de padrão de fluxo para importação. As configurações de fluxo são especificadas na caixa de diálogo de propriedades do fluxo.

Opções Adicionais

Várias outras opções podem ser especificadas utilizando a guia Dados:

- Para visualizar as configurações de armazenamento para dados que não estiverem mais conectados por meio do nó atual (dados de treinamento, por exemplo), selecione **Visualizar configurações de campo não utilizado**. É possível limpar os campos anteriores ao clicar em **Limpar**.
- Em qualquer ponto enquanto estiver trabalhando nesta caixa de diálogo, clique em **Atualizar** para recarregar os campos na origem de dados. Isso é útil quando estiver alterando conexões de dados com nó de origem ou quando estiver trabalhando entre as guias na caixa de diálogo.

Armazenamento de lista e níveis de medição associados

Introduzida no SPSS Modeler versão 17 para funcionar com os novos níveis de medição de Geoespacial e Coleção, um campo Armazenamento de lista contém diversos valores para um único registro. As listas são colocadas entre colchetes ([]). Exemplos de listas são: [1,2,4,16] e ["abc", "def"].

As listas podem ser importadas no SPSS Modeler em um dos três nós de origem (Servidor Analítico, Geoespacial ou Arquivo Variável), criados dentro de seus fluxos por meio do uso dos nós de operação de campo Derivar ou Preenchimento, ou geradas pelo nó Mesclagem ao utilizar o método de mesclagem Condição Classificada.

As listas são consideradas como tendo uma profundidade, por exemplo, uma lista simples com itens que estiverem contidos dentro de colchetes simples, no formato [1,3], é registrada no IBM SPSS Modeler com

uma profundidade de zero. Além das listas simples que possuem uma profundidade de zero, é possível utilizar listas aninhadas, em que cada valor dentro de uma lista é uma lista em si.

A profundidade de uma lista aninhada depende do nível de medição associado. Para Sem Tipo não há nenhum limite de profundidade definido, para Coleção a profundidade é zero e, para Geoespacial, a profundidade deve estar entre zero e dois, dependendo do número de itens aninhados.

Para listas de profundidade zero, é possível configurar o nível de medição, como Geoespacial ou Coleção. Esses dois níveis são níveis de medição pai e você configura as informações de subnível de medição na caixa de diálogo Valores. O subnível de medição de uma Coleção determina o nível de medição dos elementos nessa lista. Todos os níveis de medição (exceto sem tipo e geoespacial) estão disponíveis como subníveis para Coleções. O nível de medição Geoespacial possui seis subníveis de Point, LineString, Polygon, MultiPoint, MultiLineString e MultiPolygon; para obter mais informações, consulte “Subníveis de medição Geoespacial” na página 135.

Nota: O nível de medição Coleção pode ser utilizado apenas com listas de profundidade 0, o nível de medição Geoespacial pode ser utilizado apenas com listas de uma profundidade máxima de 2, e o nível de medição Sem Tipo pode ser utilizado com qualquer profundidade da lista.

O exemplo a seguir mostra a diferença entre uma lista de profundidade zero e uma lista aninhada ao utilizar a estrutura dos subníveis de medição Geoespacial Point e LineString:

- O subnível de medição Geoespacial de Point possui uma profundidade de campo de zero:
[1,3] duas coordenadas
[1,3,-1] três coordenadas
- O subnível de medição Geoespacial de LineString possui uma profundidade de campo de um:
[[1,3], [5,0]] duas coordenadas
[[1,3,-1], [5,0,8]] três coordenadas

O campo Point (com uma profundidade de zero) é uma lista normal na qual cada valor é composto de duas ou três coordenadas. O campo de LineString (com uma profundidade de um) é uma lista de pontos, em que cada ponto é composto de uma série adicional de valores da lista.

Para obter mais informações sobre a criação da lista, consulte “Derivando uma lista ou campo geoespacial” na página 152.

Caracteres de controle não suportados

Alguns dos processos no SPSS Modeler não podem manipular dados que incluam vários caracteres de controle. Se seus dados utilizarem estes caracteres, uma mensagem de erro como a seguinte poderá ser exibida:

```
Unsupported control characters found in values of field {0}
```

Os caracteres não suportados são: de 0x0 a 0x3F inclusive, e 0x7F.; no entanto, os caracteres de tabulação (0x9(\t)), nova linha (0xA(\n)) e retorno de linha (0xD(\r)) não causam um problema.

Se for exibida uma mensagem de erro relacionada a caracteres não suportados, em seu fluxo, após o nó Origem, utilize um nó Preenchimento e a expressão CLEM **stripctrlchars** para substituir os caracteres.

Nó de Origem do Analytic Server

A origem do Analytic Server permite executar um fluxo no Hadoop Distributed File System (HDFS). As informações em uma origem de dados do Analytic Server podem vir de uma variedade de locais, incluindo:

- Arquivos de texto no HDFS

- Bancos de dados.
- HCatalog

Geralmente, um fluxo com uma Origem do Analytic Server será executado no HDFS, no entanto, se um fluxo contiver um nó que não seja suportado para execução no HDFS, então o máximo possível do fluxo será "empurrado de volta" ao Analytic Server e, em seguida o SPSS Modeler Server tentará processar o restante do fluxo. Será necessário obter uma subamostra de conjuntos de dados muito grandes, por exemplo, ao colocar um nó Amostra dentro do fluxo.

Origem de dados. Supondo que o administrador do seu SPSS Modeler Server estabeleceu uma conexão, você seleciona uma origem de dados que contém os dados que deseja usar. Uma origem de dados contém os arquivos e os metadados associados a essa origem. Clique em **Selecionar** para exibir uma lista de origens de dados disponíveis. Consulte o tópico "Selecionando uma origem de dados" para obter mais informações.

Se precisar criar uma nova origem de dados ou editar uma existente, clique em **Ativar Editor de Origem de Dados...**

Selecionando uma origem de dados

A tabela Origens de Dados exibe uma lista das origens de dados disponíveis. Selecione a origem que deseja usar e clique em **OK**.

Clique em **Mostrar Proprietário** para exibir o proprietário da origem de dados.

Filtrar por permite filtrar a origem de dados listada em **Palavra-chave**, que verifica os critérios de filtro com relação ao nome da origem de dados e à descrição da origem de dados, ou **Proprietário**. É possível inserir uma combinação de caracteres de sequência de caracteres, numéricos ou curinga descrita abaixo como critérios de filtro. A sequência de caracteres de procura faz distinção entre maiúsculas e minúsculas. Clique em **Atualizar** para atualizar a tabela de Origens de Dados.

- Um sublinhado pode ser utilizado para representar qualquer caractere único na sequência de caracteres de procura.
- % Um sinal de porcentagem pode ser utilizado para representar qualquer sequência de zero ou mais caracteres na sequência de caracteres de procura.

Corrigindo as credenciais

Se suas credenciais para acessar o Analytic Server forem diferentes das credenciais para acessar o SPSS Modeler Server, será necessário inserir as credenciais do Analytic Server ao executar um fluxo no Analytic Server. Se você não souber suas credenciais, entre em contato com o administrador do servidor.

Nós Suportados

Muitos nós do SPSS Modeler são suportados para execução no HDFS, mas pode haver algumas diferenças na execução de certos nós, e alguns não são atualmente suportados. Esse tópico detalha o nível atual de suporte.

Geral

- Alguns caracteres que são normalmente aceitos dentro de um nome de campo de Modelador entre aspas não serão aceitos pelo Analytic Server.
- Para um fluxo do Modeler a ser executado no Analytic Server, ele deve iniciar com um ou mais nós do Analytic Server Source e terminar em um único nó de modelagem ou nó do Analytic Server Export.
- É recomendado que você configure o armazenamento de variáveis respostas contínuas como número real em vez de número inteiro. Modelos de pontuação sempre gravam valores de número real para os arquivos de dados de saída para variáveis respostas contínuas, enquanto o

modelo de dados de saída para as pontuações segue o armazenamento do destino. Assim, se uma variável resposta contínua tiver um armazenamento de número inteiro, haverá uma incompatibilidade entre os valores gravados e o modelo de dados para as pontuações e essa incompatibilidade causará erros quando for tentada a leitura dos dados de pontuação.

Fonte

- Um fluxo que começa com qualquer coisa que não seja um nó de origem do Analytic Server será executado localmente.

Operações de Registro

Todas as Operações de registro são suportadas, com a exceção dos nós TS do fluxo e Caixas de espaço-tempo. Notas adicionais sobre acompanhamento da funcionalidade de nó suportado.

Selecionar

- Suporta o mesmo conjunto de funções suportadas pelo Derivar nó.
- Ao usar o nó Selecionar com a opção de descarte, os campos com valores nulos são descartados no conjunto de resultados. Por exemplo: se os critérios forem descartar linhas nas quais OCCUPATION = "Retired", todas as linhas em que OCCUPATION = "Retired" E OCCUPATION = null serão descartadas. Você deve modificar os critérios de seleção para incluir "not(field = undef)". Por exemplo: atualize os critérios de seleção para ((OCCUPATION = "Retired" e not(OCCUPATION = undef)). O conjunto de resultados conterá linhas em que o campo OCCUPATION é null.

Amostra

- Amostragem de nível de bloqueio não é suportada.
- Métodos de Amostragem complexos não são suportados.

Agregado

- Chaves contíguas não são suportadas. Se você estiver reutilizando um fluxo existente configurado para classificar os dados e, em seguida, usar essa configuração no nó Agregado, altere o fluxo para remover o nó Classificação.
- As estatísticas de ordem (Mediana, primeiro Quartil, terceiro Quartil) são calculadas aproximadamente e suportadas através da guia Otimização.

Classificar

- A guia Otimização não é suportada.

Em um ambiente distribuído, há um número limitado de operações que preservam a ordem de registro estabelecida pelo nó Classificação.

- Uma classificação seguida por um nó Exportação produz uma origem de dados classificados.
- Uma classificação seguida por um nó Amostra com a amostragem **Primeiro** registro retorna os primeiros registros *N*.

Em geral, deve-se colocar um nó Classificar o mais próximo possível das operações que precisam de registros classificados.

Mesclar

- Mesclagem por Ordem não é suportada.
- A guia Otimização não é suportada.
- O Analytic Server não se associa a chaves de sequência vazia; isto é, se uma das chaves pelas quais você estiver mesclando contiver sequências vazias, então quaisquer registros que contiverem a sequência vazia serão descartados da saída mesclada.
- Operações de mesclagem são relativamente lentas. Se você tiver espaço disponível no HDFS, pode ser muito mais rápida mesclar suas origens de dados uma vez e usar a origem mesclada nos seguintes fluxos do que mesclar as origens de dados em cada fluxo.

Transformação de R

A sintaxe R no nó deve consistir em operações de registro em um momento.

Operações de campo

Todas as operações Campo são suportadas, com a exceção dos nós Transpor, Intervalos de Tempo e Histórico. Notas adicionais sobre acompanhamento da funcionalidade de nó suportado.

Preparação de dados automática

- Treinamento de nó não é suportado. Aplicar as transformações em um nó de Preparação de Dados Automática treinado para novos dados é suportado.

Tipo

- A coluna Verificação não é suportada.
- A guia Formato não é suportada.

Derivação

- Todas as funções de Derivação são suportadas, com exceção de funções de sequência.
- Campos de divisão não podem ser derivados no mesmo fluxo que os use como divisões; você precisará criar dois fluxos; um que derive o campo de divisão e um que use o campo como divisões.
- Um campo sinalizador não pode ser usado sozinho em uma comparação; isto é, `if (flagField) then ... endif` causará um erro; a solução alternativa é usar `if (flagField=trueValue) then ... endif`
- Isto é recomendado ao usar o operador `**` para especificar o expoente como um número real, tal como `x**2.0`, em vez de `x**2`, para corresponder resultados no Modelador

Preenchedor

- Suporta o mesmo conjunto de funções suportadas pelo Derivar nó.

Categorização

A funcionalidade a seguir não é suportada.

- Ótima categorização
- Ranqueamentos
- Ladrilhos -> Agrupamento lado a lado: Soma de valores
- Ladrilhos -> Ligações: Manter em atual e designar aleatoriamente
- Ladrilhos -> N customizado: Valores acima de 100 e qualquer valor N em que 100 % N não é igual a zero.

Análise RFM

- A opção Manter em atual para ligações de manipulação não é suportado. Pontuações de recência, frequência e monetárias do RFM nem sempre corresponderão àquelas calculadas pelo Modelador dos mesmos dados. Os intervalos de pontuação serão os mesmos, mas as designações de pontuação (número de categoria) podem diferir em um.

Gráficos

Todos os nós de Gráfico são suportados.

Modelagem

Os nós de Modelagem a seguir são suportados: TCM, Tree-AS, árvore C&R, Busca, CHAID, Linear, Linear-AS, Rede neural, GLE, LSVM, TwoStep-AS, Árvores aleatórias, STP e regras de associação. Notas adicionais sobre a funcionalidade desses nós a seguir.

Linear Ao construir modelos em big data, normalmente você desejará alterar o objetivo para conjuntos de dados muito grandes ou especificar as divisões.

- Treinamento contínuo de modelos PSM existentes não é suportado.

- O objetivo da construção do modelo Padrão apenas será recomendado se os campos de divisão forem definidos de modo que número de registros em cada divisão não seja muito alto, sendo que a definição de "muito alto" depende do poder dos nós individuais no cluster do Hadoop. Por contraste, também é necessário ser cuidadoso para assegurar que as divisões não sejam definidas em partes muito pequenas de maneira que haja muito poucos registros para construir um modelo.
- O objetivo de Boosting não é suportado.
- O objetivo de Bagging não é suportado.
- O objetivo dos conjuntos de dados muito grandes não é recomendado quando há poucos registros; frequentemente ele não construirá um modelo ou construirá um modelo degradado.
- Preparação de Dados Automática não é suportada. Isso pode causar problemas ao tentar construir um modelo sobre dados com muitos valores ausentes; normalmente esse seriam inseridos como parte da preparação de dados automática. Uma solução alternativa seria usar um modelo de árvore ou uma rede neural com a configuração Avançado para inserir valores ausentes selecionados.
- A estatística de previsão não é calculada para modelos de divisão.

Rede Neural

Ao construir modelos em big data, normalmente você desejará alterar o objetivo para conjuntos de dados muito grandes ou especificar as divisões.

- Treinamento contínuo de padrão existente ou modelos PSM não é suportado.
- O objetivo da construção do modelo Padrão apenas será recomendado se os campos de divisão forem definidos de modo que número de registros em cada divisão não seja muito alto, sendo que a definição de "muito alto" depende do poder dos nós individuais no cluster do Hadoop. Por contraste, também é necessário ser cuidadoso para assegurar que as divisões não sejam definidas em partes muito pequenas de maneira que haja muito poucos registros para construir um modelo.
- O objetivo de Boosting não é suportado.
- O objetivo de Bagging não é suportado.
- O objetivo dos conjuntos de dados muito grandes não é recomendado quando há poucos registros; frequentemente ele não construirá um modelo ou construirá um modelo degradado.
- Quando houver muitos valores ausentes nos dados, use a configuração Avançado para inserir valores ausentes.
- A estatística de previsão não é calculada para modelos de divisão.

Árvore C&R, CHAID e Quest

Ao construir modelos em big data, normalmente você desejará alterar o objetivo para conjuntos de dados muito grandes ou especificar as divisões.

- Treinamento contínuo de modelos PSM existentes não é suportado.
- O objetivo da construção do modelo Padrão apenas será recomendado se os campos de divisão forem definidos de modo que número de registros em cada divisão não seja muito alto, sendo que a definição de "muito alto" depende do poder dos nós individuais no cluster do Hadoop. Por contraste, também é necessário ser cuidadoso para assegurar que as divisões não sejam definidas em partes muito pequenas de maneira que haja muito poucos registros para construir um modelo.
- O objetivo de Boosting não é suportado.
- O objetivo de Bagging não é suportado.
- O objetivo dos conjuntos de dados muito grandes não é recomendado quando há poucos registros; frequentemente ele não construirá um modelo ou construirá um modelo degradado.

- Sessões interativas não são suportadas.
- A estatística de previsão não é calculada para modelos de divisão.
- Quando um campo dividido está presente, três modelos construídos localmente no Modeler são um pouco diferentes dos três modelos construídos pelo Analytic Server e, portanto, produzem pontuações diferentes. Os algoritmos em ambos os casos são válidos; os algoritmos usados pelo Analytic Server são simplesmente mais novos. Dado o fato de que três algoritmos tendem a ter muitas regras heurísticas, a diferença entre os dois componentes é normal.

Pontuação do modelo

Todos os modelos suportados para modelagem também são suportados para pontuação. Além disso, nuggets do modelo construído localmente para os nós a seguir são suportados para escoragem: C&RT, Quest, CHAID, Linear e Rede Neural (independentemente se o modelo é padrão, empacotado impulsionado ou para conjuntos de dados muito grandes), Regressão, C5.0, Logística, Genlin, GLMM, Cox, SVM, Rede de Bayes, TwoStep, KNN, Lista de Decisão, Discriminante, Autoaprendizado, Detecção de anomalia, Apriori, Carma, K-Médias, Kohonen, R e Mineração de texto.

- Nenhuma propensão ajustada ou bruta será pontuada. Como uma solução alternativa é possível obter o mesmo efeito calculando manualmente a propensão bruta usando um nó Derivar com as seguintes expressões: `if 'predicted-value' == 'value-of-interest' then 'prob-of-that-value' else 1-'prob-of-that-value'` endif
- Ao pontuar um modelo, o Analytic Server não verifica se todos os campos usados no modelo estão presentes no conjuntos de dados, assim, certifique-se de que isso é verdadeiro antes de executar no Analytic Server

R A sintaxe R na nugget deverá consistir em operações de registro em um momento.

Out Os nós Matriz, Análise, Auditoria de Dados, Transformar, Estatísticas, Médias e Tabela são suportados. Notas adicionais sobre acompanhamento da funcionalidade de nó suportado.

Auditoria de dados

O nó Auditoria de Dados não pode produzir o modo para campos contínuos.

Médias

O nó Médias não pode produzir um erro padrão ou 95% de intervalo de confiança.

Tabela

O nó Tabela é suportado escrevendo uma origem de dados do Analytic Server temporário que contém os resultados de operações de envio de dados. O nó Tabela, em seguida, pagina através do conteúdo dessa origem de dados.

Exportar

Um fluxo pode iniciar com um nó de origem do Analytic Server e terminar com um nó de exportação diferente do nó de exportação do Analytic Server, mas os dados se movimentarão do HDFS para o SPSS Modeler Server, e finalmente para a localização de exportação.

Nó de Origem da Base de Dados

O nó Origem da Base de Dados pode ser utilizado para importar dados de uma variedade de outros pacotes utilizando ODBC (Open Database Connectivity), incluindo Microsoft SQL Server, o DB2, o Oracle e outros.

Para ler ou gravar em um banco de dados, você deverá ter uma origem de dados ODBC instalada e configurada para o banco de dados relevante e com permissões de leitura ou gravação, conforme necessário. O IBM SPSS Data Access Pack inclui um conjunto de drivers ODBC que podem ser usados para esse propósito e esses drivers estão disponíveis no IBM SPSS Data Access Pack DVD ou a partir do site de download. Se você tiver questões sobre como criar ou configurar permissões para origens de dados ODBC, entre em contato com seu administrador de banco de dados.

Drivers ODBC Suportados

Para obter as informações mais recentes sobre quais bancos de dados e drivers ODBC são suportados e testados para uso com o IBM SPSS Modeler 17.1, consulte as matrizes de compatibilidade do produto no site de Suporte corporativo (<http://www.ibm.com/support>).

Onde Instalar os Drivers

Nota: Os drivers ODBC devem ser instalados e configurados em cada computador em que o processamento possa ocorrer.

- Se você estiver executando o IBM SPSS Modeler no modo local (independente), os drivers deverão ser instalados no computador local.
- Se você estiver executando o IBM SPSS Modeler no modo distribuído com relação a um IBM SPSS Modeler Server remoto, os drivers ODBC precisarão ser instalados no computador no qual o IBM SPSS Modeler Server foi instalado. Para o IBM SPSS Modeler Server em sistemas UNIX, consulte também "Configurando Drivers ODBC em Sistemas UNIX" posteriormente nesta seção.
- Se você precisar acessar as mesmas origens de dados a partir do IBM SPSS Modeler e do IBM SPSS Modeler Server, os drivers ODBC deverão ser instalados nos dois computadores.
- Se você estiver executando o IBM SPSS Modeler sobre os Serviços de Terminal, os drivers ODBC precisarão ser instalados no servidor de Serviços de Terminal no qual o IBM SPSS Modeler está instalado.

Acessar dados de um banco de dados

Para acessar dados a partir de um banco de dados, conclua os seguintes passos.

- Instale um driver ODBC e configure uma origem de dados para o banco de dados que deseja utilizar.
- Na caixa de diálogo do nó Banco de Dados, conecte-se a um banco de dados utilizando o modo Tabela ou o modo de SQL Query.
- Selecione uma tabela no banco de dados.
- Utilizando as guias na caixa de diálogo do nó Banco de Dados, é possível alterar os tipos de uso e os campos de dados do filtro.

Mais detalhes sobre os passos anteriores são fornecidos nos tópicos da documentação relacionada.

Nota: Se chamar os Procedimentos Armazenados (SPs) do banco de dados a partir do SPSS Modeler, poderá ser retornado um arquivo de saída único denominado RowsAffected ao invés da saída esperada do SP. Isso ocorre quando o ODBC não retorna informações suficientes para poder determinar o modelo de dados da saída do SP. O SPSS Modeler possui suporte limitado somente para os SPs que retornam a saída e é sugerido que ao invés de utilizar SPs, deve-se extrair a função SELECT a partir do SP e utilizar uma das ações a seguir.

- Crie uma visualização com base na SELECT e escolha a visualização no nó de origem Banco de Dados
- Utilize a SELECT diretamente no nó de origem do Banco de Dados.

Configurando Opções do Nó de Banco de Dados

É possível utilizar as opções na guia Dados da caixa de diálogo do nó de origem Banco de dados para obter acesso a um banco de dados e ler dados na tabela selecionada.

Modo. Selecione **Tabela** para conectar-se a uma tabela utilizando os controles da caixa de diálogo.

Selecione **Query SQL** para consultar o banco de dados selecionado abaixo utilizando SQL. Consulte o tópico "Consultando o Banco de Dados" na página 25 para obter mais informações.

Origem de dados. Para os modos Tabela e SQL Query, é possível inserir um nome no campo Origem de Dados ou selecionar **Incluir nova conexão com o banco de dados** na lista suspensa.

As opções a seguir são utilizadas para conectar-se a um banco de dados e selecionar uma tabela utilizando a caixa de diálogo:

Nome da tabela. Se você souber o nome da tabela que deseja acessar, insira-o no campo Nome da Tabela. Caso contrário, clique no botão **Selecionar** para abrir uma caixa de diálogo listando as tabelas disponíveis.

Tabela de cotação e nomes de coluna. Especifique se deseja que os nomes de colunas e tabelas fiquem entre aspas quando as consultas forem enviadas ao banco de dados (se, por exemplo, incluírem espaços e pontuação).

- A opção **Conforme o necessário** colocará entre aspas os nomes da tabela e os nomes de campo *apenas* se eles incluírem caracteres padrão. Caracteres não padrão incluem caracteres não ASCII, caracteres de espaços e qualquer caractere alfanumérico que não seja um ponto final (.).
- Selecione **Sempre** se quiser que *todos* os nomes de tabelas e campos sejam colocados entre aspas.
- Selecione **Nunca** se você não deseja *nunca* que os nomes de tabelas e colunas sejam postos entre aspas.

Remover espaços à esquerda e à direita. Selecione as opções para descartar espaços à esquerda e à direita nas sequências de caracteres.

Nota. As comparações entre sequências de caracteres que usam e não usam SQL pushback podem gerar diferentes resultados onde existem espaços à direita.

Lendo sequências de caracteres vazias do Oracle. Durante a leitura ou gravação em um banco de dados Oracle, saiba que, ao contrário do IBM SPSS Modeler e da maioria dos outros bancos de dados, o Oracle trata e armazena valores de sequência de caracteres vazios como equivalentes a valores nulos. Isso significa que os mesmos dados extraídos de um banco de dados Oracle podem se comportar diferentemente quando extraídos de um arquivo ou outro banco de dados, e os dados podem retornar resultados diferentes.

Incluindo uma conexão com o banco de dados

Para abrir um banco de dados, primeiro selecione a origem de dados à qual você deseja conectar-se. Na guia Dados, selecione **Incluir nova conexão com o banco de dados** na lista suspensa Origem de Dados.

Isso abre a caixa de diálogo Conexões com o Banco de Dados.

Nota: Para obter uma maneira alternativa para abrir esse diálogo, no menu principal, escolha: **Ferramentas > Bancos de Dados...**

Origens de dados Lista as origens de dados disponíveis. Role para baixo se não for exibido o banco de dados desejado. Após selecionar uma origem de dados e inserir quaisquer senhas, clique em **Conectar-se**. Clique em **Atualizar** para atualizar a lista.

Nome do usuário e senha Se a origem de dados for protegida por senha, insira seu nome de usuário e a senha associada.

Credencial Se uma credencial tiver sido configurada no IBM SPSS Collaboration and Deployment Services, será possível selecionar esta opção para procurá-la no repositório. O nome do usuário e a senha da credencial devem corresponder ao nome do usuário e à senha necessários para acessar o banco de dados.

Conexões Mostra bancos de dados atualmente conectados.

- **Padrão** É possível, opcionalmente, escolher uma conexão como a padrão. Fazer isso faz com que os nós de origem ou de exportação do Banco de Dados tenham essa conexão predefinida como sua origem de dados, embora isso possa ser editado, se desejado.
- **Salvar** Opcionalmente, selecione uma ou mais conexões que deseja exibir novamente em sessões subsequentes.
- **Origem de dados** As sequências de caracteres para os bancos de dados conectados atualmente.
- **Pré-configurado** Indica (com um caractere *) se os valores pré-configurados foram especificados para a conexão com o banco de dados. Para especificar valores pré-configurados, clique nessa coluna na linha correspondente à conexão com o banco de dados e escolha Especificar na lista. Consulte o tópico "Especificando valores pré-configurados para uma conexão com o banco de dados" na página 21 para obter mais informações.

Para remover conexões, selecione uma na lista e clique em **Remover**.

Após concluir suas seleções, clique em **OK**.

Para ler ou gravar em um banco de dados, você deverá ter uma origem de dados ODBC instalada e configurada para o banco de dados relevante e com permissões de leitura ou gravação, conforme necessário. O IBM SPSS Data Access Pack inclui um conjunto de drivers ODBC que podem ser usados para esse propósito e esses drivers estão disponíveis no IBM SPSS Data Access Pack DVD ou a partir do site de download. Se você tiver questões sobre como criar ou configurar permissões para origens de dados ODBC, entre em contato com seu administrador de banco de dados.

Drivers ODBC Suportados

Para obter as informações mais recentes sobre quais bancos de dados e drivers ODBC são suportados e testados para uso com o IBM SPSS Modeler 17.1, consulte as matrizes de compatibilidade do produto no site de Suporte corporativo (<http://www.ibm.com/support>).

Onde Instalar os Drivers

Nota: Os drivers ODBC devem ser instalados e configurados em cada computador em que o processamento possa ocorrer.

- Se você estiver executando o IBM SPSS Modeler no modo local (independente), os drivers deverão ser instalados no computador local.
- Se você estiver executando o IBM SPSS Modeler no modo distribuído com relação a um IBM SPSS Modeler Server remoto, os drivers ODBC precisarão ser instalados no computador no qual o IBM SPSS Modeler Server foi instalado. Para o IBM SPSS Modeler Server em sistemas UNIX, consulte também "Configurando Drivers ODBC em Sistemas UNIX" posteriormente nesta seção.
- Se você precisar acessar as mesmas origens de dados a partir do IBM SPSS Modeler e do IBM SPSS Modeler Server, os drivers ODBC deverão ser instalados nos dois computadores.
- Se você estiver executando o IBM SPSS Modeler sobre os Serviços de Terminal, os drivers ODBC precisarão ser instalados no servidor de Serviços de Terminal no qual o IBM SPSS Modeler está instalado.

Configurando Drivers ODBC em Sistemas UNIX

Por padrão, o DataDirect Driver Manager não está configurado para o IBM SPSS Modeler Server em sistemas UNIX. Para configurar o UNIX para carregar o DataDirect Driver Manager, insira os comandos a seguir:

```
cd <modeler_server_install_directory>/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

Isso remove o link padrão e cria um link para o DataDirect Driver Manager.

Nota: O wrapper do driver UTF16 é necessário para usar os drivers SAP HANA ou IBM DB2 CLI para alguns bancos de dados. Para criar uma ligação para o wrapper do driver UTF16, insira os seguintes comandos:

```
rm -f libspssodbc.so
ln -s libspssodbc_datadirect_utf16.so libspssodbc.so
```

Para configurar o SPSS Modeler Server:

1. Configure o script de inicialização do SPSS Modeler Server `modelersrv.sh` para originar o arquivo de ambiente IBM SPSS Data Access Pack `odbc.sh` incluindo a seguinte linha no `modelersrv.sh`:
. /<pathtoSDAPinstall>/odbc.sh

Em que <pathtoSDAPinstall> é o caminho completo para a instalação do IBM SPSS Data Access Pack.

2. Reinicie o SPSS Modeler Server.

Além disso, para o SAP HANA e IBM DB2 somente, inclua a definição de parâmetro a seguir no DSN em seu arquivo `odbc.ini` para evitar estouros de buffer durante a conexão:

```
DriverUnicodeType=1
```

Nota: O wrapper `libspssodbc_datadirect_utf16.so` também é compatível com os outros drivers ODBC suportados do SPSS Modeler Server.

Especificando valores pré-configurados para uma conexão com o banco de dados

Para alguns bancos de dados, é possível especificar um número de configurações padrão para a conexão com o banco de dados. Todas as configurações se aplicam à exportação do banco de dados.

Os tipos de bancos de dados que suportam esse recurso são os seguintes.

- IBM InfoSphere Warehouse. Consulte o tópico “Configurações para o IBM DB2 InfoSphere Warehouse” para obter mais informações.
- Edições do SQL Server Enterprise e Developer. Consulte o tópico “Configurações para SQL server” na página 22 para obter mais informações.
- Edições do Oracle Enterprise ou Personal. Consulte o tópico “Configurações para Oracle” na página 22 para obter mais informações.
- O IBM Netezza, o IBM DB2 for z/OS e o Teradata se conectam a um banco de dados ou a um esquema de maneira semelhante. Consulte o tópico “Configurações para o IBM Netezza, IBM DB2 for z/OS, IBM DB2 LUW e Teradata” na página 23 para obter mais informações.

Se você estiver conectado a um banco de dados ou a um esquema que não suporta esse recurso, será exibida a mensagem **Nenhuma pré-configuração pode ser definida para essa conexão com o banco de dados**.

Configurações para o IBM DB2 InfoSphere Warehouse

Essas configurações são exibidas para o IBM InfoSphere Warehouse.

Espaço de tabela O espaço de tabela a ser utilizado para exportação. Os administradores de banco de dados podem criar ou configurar espaços de tabela como particionados. É recomendado selecionar um desses espaços de tabela (ao invés de um padrão) a ser usado para exportação do banco de dados.

Usar compactação. Se selecionada, cria tabelas para exportação com compactação (por exemplo, o equivalente a `CREATE TABLE MYTABLE(...) COMPRESS YES;` na SQL).

Não registrar atualizações Se selecionada, evita a criação de log quando criar tabelas e inserir dados (o equivalente a CREATE TABLE MYTABLE(...) NOT LOGGED INITIALLY; na SQL).

Configurações para SQL server

Essas configurações são exibidas para as edições SQL Server Enterprise e Developer.

Usar compactação. Se selecionada, cria tabelas para exportação com compactação.

Compactação para. Escolha o nível de compactação.

- **Linha.** Ativa a compactação de nível de linha (por exemplo, o equivalente a CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW); em SQL).
- **Página.** Ativa a compactação de nível de página (por exemplo, CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE); em SQL).

Configurações para Oracle

Configurações do Oracle - opção Básica

Essas configurações são exibidas para as edições do Oracle Enterprise ou Personal utilizando a opção Básica.

Usar compactação. Se selecionada, cria tabelas para exportação com compactação.

Compactação para. Escolha o nível de compactação.

- **Padrão.** Ativa a compactação padrão (por exemplo, CREATE TABLE MYTABLE(...) COMPRESS; na SQL). Nesse caso, ela possui o mesmo efeito que a opção **Básico**.
- **Básico.** Ativa a compactação básica (por exemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; na SQL).

Configurações do Oracle - opção Avançada

Essas configurações são exibidas para as edições do Oracle Enterprise ou Personal utilizando a opção Avançada.

Usar compactação. Se selecionada, cria tabelas para exportação com compactação.

Compactação para. Escolha o nível de compactação.

- **Padrão.** Ativa a compactação padrão (por exemplo, CREATE TABLE MYTABLE(...) COMPRESS; na SQL). Nesse caso, ela possui o mesmo efeito que a opção **Básico**.
- **Básico.** Ativa a compactação básica (por exemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; na SQL).
- **OLTP.** Ativa a compactação de OLTP (por exemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR OLTP na SQL).
- **Query Baixo/Alto.** (servidores Exadata somente) Ativa compactação columnar híbrida para query (por exemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY LOW; ou CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY HIGH; na SQL). A compactação para a query é útil em ambientes de data warehousing; HIGH fornece uma razão de compactação maior que LOW.
- **Arquivar Baixo/Alto.** (servidores Exadata somente) Ativa compactação columnar híbrida para archive (por exemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE LOW; ou CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE HIGH; na SQL). A compactação para a archive é útil para compactar dados que serão armazenados por longos períodos de tempo; HIGH fornece uma razão de compactação maior que LOW.

Configurações para o IBM Netezza, IBM DB2 for z/OS, IBM DB2 LUW e Teradata

Ao especificar pré-configurações para o IBM Netezza, IBM DB2 for z/OS, IBM DB2 LUW ou Teradata, será solicitado selecionar os seguintes itens:

Usar o banco de dados do adaptador de escoragem no servidor ou **Usar o esquema do adaptador de escoragem no servidor**. Se selecionada, ativa a opção **Banco de dados do adaptador de escoragem no servidor** ou **Esquema do adaptador de escoragem no servidor**.

Banco de dados do adaptador de escoragem no servidor ou **Esquema do adaptador de escoragem no servidor** Na lista suspensa, selecione a conexão que precisar.

Além disso, para Teradata, também é possível configurar os detalhes de banda de query para fornecer metadados adicionais para ajudar com itens como gerenciamento de carga de trabalho, intercalação, identificação e resolução de queries e o rastreamento de uso do banco de dados.

Ortografia da banda de query. Escolha se a banda de query será configurada uma vez durante todo o tempo em que você estiver trabalhando com uma conexão com o banco de dados Teradata (**Para Sessão**) ou se será configurada toda vez que executar um fluxo (**Para Transação**).

Nota: Se configurar a banda de query em um fluxo, a banda será perdida se esse fluxo for copiado para outra máquina. Para evitar isso, é possível utilizar script para executar seu fluxo e utilizar a palavra chave *querybanding* no script para aplicar as configurações que precisar.

Permissões do banco de dados necessárias

Para que os recursos de bando de dados do SPSS Modeler funcionem corretamente, conceda acesso aos seguintes itens a todos os IDs de usuário utilizados:

DB2 LUW

SYSIBM.SYSDUMMY1
SYSIBM.SYSFOREIGNKEYS
SYSIBM.SYSINDEXES
SYSIBM.SYSKEYCOLUSE
SYSIBM.SYSKEYS
SYSIBM.SYSPARMS
SYSIBM.SYSRELS
SYSIBM.SYSROUTINES
SYSIBM.SYSROUTINES_SRC
SYSIBM.SYSSYNONYMS
SYSIBM.SYSTABCONST
SYSIBM.SYSTABCONSTPKC
SYSIBM.SYSTABLES
SYSIBM.SYSTRIGGERS
SYSIBM.SYSVIEWDEP
SYSIBM.SYSVIEWS
SYSCAT.TABLESPACES
SYSCAT.SCHEMATA

DB2/z SYSIBM.SYSDUMMY1
SYSIBM.SYSFOREIGNKEYS
SYSIBM.SYSINDEXES
SYSIBM.SYSKEYCOLUSE
SYSIBM.SYSKEYS
SYSIBM.SYSPARMS
SYSIBM.SYSRELS
SYSIBM.SYSROUTINES
SYSIBM.SYSROUTINES_SRC
SYSIBM.SYSSYNONYMS
SYSIBM.SYSTABCONST
SYSIBM.SYSTABCONSTPKC
SYSIBM.SYSTABLES
SYSIBM.SYSTRIGGERS
SYSIBM.SYSVIEWDEP
SYSIBM.SYSVIEWS
SYSIBM.SYSDUMMYU
SYSIBM.SYSPACKSTMT

Netezza

_V_FUNCTION
_V_DATABASE

Teradata

DBC.Functions
DBC.USERS

Selecionando uma Tabela de Base de Dados

Após conectar-se a uma origem de dados, é possível optar por importar campos a partir de uma tabela ou visualização específica. Na guia Dados da caixa de diálogo Banco de Dados, é possível inserir o nome de uma tabela no campo Nome da tabela ou clicar em **Selecionar** para abrir a caixa de diálogo Selecionar Tabela/Visualização que lista as tabelas e visualizações disponíveis.

Mostrar proprietário da tabela. Selecione se uma origem de dados requer que o proprietário de uma tabela seja especificado antes de poder acessar a tabela. Cancele a seleção dessa opção para origens de dados que não possuem esse requisito.

Nota: os bancos de dados SAS e Oracle geralmente requerem mostrar o proprietário da tabela.

Tabelas/Visualizações. Selecione a tabela ou visualização a ser importada.

Mostrar. Lista as colunas na origem de dados às quais você está atualmente conectado. Clique em uma das opções a seguir para customizar sua visualização das tabelas disponíveis:

- Clique em **Tabelas do Usuário** para visualizar tabelas de banco de dados ordinárias criadas pelos usuários do banco de dados.

- Clique em **Tabelas do Sistema** para visualizar tabelas de banco de dados de propriedade do sistema (por exemplo, tabelas que fornecem informações sobre o banco de dados, como detalhes de índices). Essa opção pode ser utilizada para visualizar as guias utilizadas em bancos de dados do Excel. (Observe que um nó de origem do Excel separado também está disponível. Consulte o tópico “Nó de Origem do Excel” na página 42 para obter mais informações.)
- Clique em **Visualizações** para visualizar tabelas virtuais com base em uma consulta envolvendo uma ou mais tabelas ordinárias.
- Clique em **Sinônimos** para visualizar os sinônimos criados no banco de dados para quaisquer tabelas existentes.

Filtros de Nome/Proprietário. Esses campos permitem filtrar a lista de tabelas exibidas por nome ou proprietário. Por exemplo, digite *SYS* para listar apenas tabelas com esse proprietário. Para procuras curinga, um sublinhado (*_*) pode ser utilizado para representar qualquer caractere único e um sinal de porcentagem (*%*) pode representar qualquer sequência de zero ou mais caracteres.

Configurar Como Padrão. Salva as configurações atuais como o padrão para o usuário atual. Essas configurações serão restauradas no futuro quando um usuário abrir uma nova caixa de diálogo de seletor de tabela *somente para o mesmo nome e login de usuário da origem de dados*.

Consultando o Banco de Dados

Após conectar-se a uma origem de dados, será possível escolher importar os campos utilizando as queries SQL. Na caixa de diálogo principal, selecione **Query SQL** como o modo de conexão. Isso inclui uma janela do editor de consulta na caixa de diálogo. Usando o editor de consulta, é possível criar ou carregar uma mais queries SQL cujo resultado será lido no fluxo de dados.

Se você especificar diversas queries SQL, separe-as com pontos e vírgulas (;) e assegure-se de que não haja diversas instruções SELECT.

Para cancelar e fechar a janela do editor de consulta, selecione **Tabela** como o modo de conexão.

É possível incluir parâmetros de fluxo do SPSS Modeler (um tipo de variável definida pelo usuário) na query SQL. Consulte o tópico “Usando Parâmetros de Fluxo em uma Query SQL” para obter mais informações.

Carregar Query. Clique para abrir o navegador de arquivos, que pode ser utilizado para carregar uma query salva anteriormente.

Salvar Query. Clique para abrir a caixa de diálogo Salvar Query, que pode ser utilizada para salvar a query atual.

Importar Padrão. Clique para importar uma instrução SQL SELECT de exemplo construída automaticamente utilizando a tabela e as colunas selecionadas na caixa de diálogo.

Limpar. Limpe o conteúdo da área de trabalho. Utilize esta opção quando desejar iniciar novamente.

Dividir texto. A opção padrão **Nunca** significa que a query será enviada ao banco de dados como um todo. Como alternativa, é possível escolher **Conforme necessário**, significando que o SPSS Modeler tenta analisar a query e identificar se há instruções SQL que devem ser enviadas para o banco de dados uma após a outra.

Usando Parâmetros de Fluxo em uma Query SQL

Ao gravar uma query SQL para importar campos, é possível incluir parâmetros de fluxo do SPSS Modeler que foram definidos anteriormente. Todos os tipos de parâmetros de fluxo são suportados.

A tabela a seguir mostra como alguns exemplos de parâmetros de fluxo serão interpretados na query SQL.

Tabela 3. Exemplos de parâmetros de fluxo.

Nome do Parâmetro de Fluxo (exemplo)	Armazenamento	Valor de Parâmetro de Fluxo	Interpretado como
PString	Sequência de caracteres	ss	'ss'
PInt	Número inteiro	5	5
PReal	Real	5.5	5.5
PTime	Hora	23:05:01	t{'23:05:01'}
PDate	Date	2011-03-02	d{'2011-03-02'}
PTimeStamp	TimeStamp	2011-03-02 23:05:01	ts{'2011-03-02 23:05:01'}
PColumn	Desconhecido	IntValue	IntValue

Na query SQL, especifique um parâmetro de fluxo da mesma maneira que em uma expressão do CLEM, ou seja, '\$P-<parameter_name>', em que <parameter_name> é o nome que foi definido para o parâmetro de fluxo.

Ao referenciar um campo, o tipo de armazenamento deve ser definido como Desconhecido, e o valor de parâmetro deve estar entre aspas, se necessário. Assim, utilizando os exemplos mostrados na tabela, se você inserir a query SQL:

```
select "IntValue" from Table1 where "IntValue" < '$P-PInt';
```

ela será avaliada como:

```
select "IntValue" from Table1 where "IntValue" < 5;
```

Se tiver que referenciar o campo IntValue por meio do parâmetro PColumn, será necessário especificar a query da seguinte forma para obter o mesmo resultado:

```
select "IntValue" from Table1 where "'$P-PColumn'" < '$P-PInt';
```

Nó Arquivo Variável

É possível usar os nós Arquivo Variável para ler dados de arquivos de texto de campo livre (arquivos cujos registros contêm um número constante de campos, mas um número variado de caracteres), também conhecidos como arquivos de texto delimitado. Esse tipo de nó também é útil para arquivos com texto de cabeçalho de comprimento fixo e certos tipos de anotações. Os registros são lidos um de cada vez e transmitidos pelo fluxo até que o arquivo inteiro seja lido.

Nota sobre leitura de dados geoespaciais

Se o nó contiver dados geoespaciais e for criado como uma exportação a partir de um arquivo simples, deve-se seguir alguns passos extras para configurar os metadados geoespaciais. Para obter informações adicionais, consulte "Importando dados geoespaciais no Nó Arquivo Variável" na página 29.

Notas para leitura de dados de texto delimitado

- Os registros devem ser delimitados por um caractere de nova linha no término de cada linha. O caractere de nova linha não deve ser usado para nenhum outro propósito (como em qualquer nome de campo ou valor). Idealmente, espaços à esquerda e à direita devem ser removidos para economizar espaço, embora isso não seja crítico. Opcionalmente, esses espaços podem ser removidos pelo nó.
- Os campos devem ser delimitados por uma vírgula ou outro caractere que seja utilizado idealmente apenas como um delimitador, o que significa que ele não aparece nos nomes de campos ou nos valores. Se isso não for possível, então todos os campos de texto poderão ser agrupados entre aspas

duplas, se nenhum dos nomes de campos ou valores de texto contiverem aspas duplas. Se nomes de campos ou valores contiverem aspas duplas, então os campos de texto poderão ser agrupados entre aspas simples como uma alternativa, novamente se aspas simples não forem utilizadas em outro lugar dentro de valores. Se nem aspas simples nem aspas duplas puderem ser utilizadas, então os valores de texto precisarão ser corrigidos para remover ou substituir o caractere delimitador ou as aspas simples ou duplas.

- Cada linha, incluindo a linha de cabeçalho, deve conter o mesmo número de campos.
- A primeira linha deve conter os nomes de campo. Se não contiver, desmarque **Ler nomes de campo a partir do arquivo** para fornecer a cada campo um nome genérico, como Field1, Field2, e assim por diante.
- A segunda linha deve conter o primeiro registro de dados. Não deverá haver nenhuma linha em branco ou comentário.
- Os valores numéricos não devem incluir o separador de milhares ou o símbolo de agrupamento--sem o ponto, como em 3.000,00, por exemplo. O indicador decimal (ponto ou ponto final nos EUA ou Reino Unido) deve ser utilizado somente quando apropriado.
- Os valores de data e hora devem estar em um dos formatos que são reconhecidos na caixa de diálogo Opções de Fluxo, como DD/MM/AAAA ou HH:MM:SS. Idealmente, todos os campos de datas e horas no arquivo devem seguir o mesmo formato, e qualquer campo que contiver uma data deve utilizar o mesmo formato para todos os valores dentro desse campo.

Configurando Opções para o Nó Arquivo de Variável

Configure as opções na guia Arquivo da caixa de diálogo do nó Arquivo Variável.

Arquivo Especifique o nome do arquivo. É possível inserir um nome de arquivo ou clicar no botão de reticências (...) para selecionar um arquivo. O caminho de arquivo é mostrado após selecionar um arquivo, e seu conteúdo é exibido com delimitadores no painel abaixo dele.

O texto de amostra que é exibido a partir de sua origem de dados pode ser copiado e colado nos seguintes controles: caracteres de comentário EOL e delimitadores especificados pelo usuário. Utilize Ctrl-C e Ctrl-V para copiar e colar.

Ler nomes de campo a partir do arquivo Seleccionada por padrão, esta opção trata a primeira linha no arquivo de dados como rótulos para a coluna. Se a sua primeira linha não for um cabeçalho, cancele a seleção para fornecer automaticamente a cada campo um nome genérico, como *Field1*, *Field2*, para o número de campos no conjunto de dados.

Especificar o número de campos. Especifique o número de campos em cada registro. O número de campos pode ser detectado automaticamente, desde que os registros sejam terminados em nova linha. Também é possível configurar um número manualmente.

Ignorar caracteres de cabeçalho. Especifique quantos caracteres deseja ignorar no início do primeiro registro.

Caracteres de comentário EOL. Especifique caracteres, como # ou !, para indicar as anotações nos dados. Sempre que um desses caracteres aparecer no arquivo de dados, tudo até, mas não incluindo o caractere de nova linha, será ignorado.

Remover espaços à esquerda e à direita. Selecione opções para descartar espaços à esquerda e à direita nas sequências de caracteres na importação.

Nota. As comparações entre sequências de caracteres que usam e não usam SQL pushback podem gerar diferentes resultados onde existem espaços à direita.

Caracteres inválidos. Selecione **Descartar** para remover caracteres inválidos da origem de dados. Selecione **Substituir por** para substituir os caracteres inválidos por símbolos especificados (um caractere apenas). Os caracteres inválidos são caracteres nulos ou qualquer caractere que não existe no método de codificação especificado.

Codificação. Especifica o método de codificação de texto usado. É possível escolher entre o padrão do sistema, o padrão de fluxo ou UTF-8.

- O padrão do sistema é especificado no Painel de Controle do Windows ou, se estiver em execução no modo distribuído, no computador servidor.
- O padrão de fluxo é especificado na caixa de diálogo Propriedades do Fluxo.

Símbolo decimal Selecione o tipo de separador decimal que é utilizado em sua origem de dados. O **Padrão de fluxo** é o caractere que é selecionado a partir da guia Opções da caixa de diálogo de propriedades do fluxo. Caso contrário, selecione **Ponto (.)** ou **Vírgula (,)** para ler todos os dados nessa caixa de diálogo usando o caractere escolhido como o separador decimal.

Delimitador de linha é o caractere de nova linha Selecione essa opção para utilizar o caractere de nova linha como o delimitador de linha, ao invés de um delimitador de campo. Por exemplo, isso pode ser útil se houver um número ímpar de delimitadores em uma linha que causam a quebra de linha. Observe que a seleção dessa opção significa que não é possível selecionar **Nova linha** na lista Delimitadores.

Nota: Se selecionar esta opção, quaisquer valores em branco no término das linhas de dados serão removidos.

Delimitadores. Usando as caixas de seleção listadas para esse controle, é possível especificar quais caracteres, como a vírgula (,), definem os limites dos campos no arquivo. Você também pode especificar mais de um delimitador, como ", |" para registros que usam vários delimitadores. O delimitador padrão é a vírgula.

Nota: Se a vírgula também estiver definida como o separador decimal, as configurações padrão aqui não funcionarão. Nos casos em que a vírgula for o delimitador de campo e o separador decimal, selecione **Outras** na lista Delimitadores. Em seguida, especifique manualmente uma vírgula no campo de entrada.

Selecione **Permitir diversos delimitadores em branco** para tratar vários caracteres delimitadores em branco adjacentes como um delimitador único. Por exemplo, se um valor de dados for seguido por quatro espaços e, em seguida, outro valor de dados, este grupo seria tratado como dois campos em vez de cinco.

Linhas para varredura de coluna e tipo Especifique quantas linhas e colunas serão varridas em busca de tipos de dados especificados.

Reconhecer datas e horas automaticamente Marque essa caixa de seleção para permitir que o IBM SPSS Modeler tente reconhecer automaticamente as entradas de dados como datas ou horas. Por exemplo, isso significa que uma entrada como 11-07-1965 será identificada como uma data e 02:35:58 será identificada como um horário, no entanto, entradas ambíguas como 07111965 ou 023558 serão mostradas como números inteiros já que não existem delimitadores entre os números.

Nota: Para evitar problemas potenciais de dados ao utilizar arquivos de dados a partir de versões anteriores do IBM SPSS Modeler, esta caixa fica desativada por padrão para informações que são salvas em versões anteriores a 13.

Tratar colchetes como listas Se marcar essa caixa de seleção, os dados incluídos entre colchetes de abertura e de fechamento serão tratados como um valor único, mesmo se o conteúdo incluir caracteres delimitadores, como vírgula e aspas duplas. Por exemplo, isso pode incluir dois ou três dados

geoespaciais dimensionais, em que as coordenadas contidas dentro dos colchetes são processadas como um item de lista único. Para obter mais informações, consulte “Importando dados geoespaciais no Nó Arquivo Variável”

Aspas. Usando as listas suspensas, é possível especificar como as aspas simples e duplas serão tratadas na importação. É possível optar por **Descartar** todas as aspas, **Incluir como texto** pela inclusão das aspas no valor do campo ou **Parear e descartar** para fazer a correspondência de pares de aspas e removê-los. Se uma aspa não tiver correspondência, você receberá uma mensagem de erro. Tanto **Descartar** quanto **Parear e descartar** armazenam o valor de campo (sem aspas) como uma sequência de caracteres.

Nota: Ao utilizar **Emparelhar e descartar**, os espaços são mantidos. Ao utilizar **Descartar**, espaços à direita dentro e fora das aspas são removidos (por exemplo, ' " ab c" , "d ef " , " gh i " ' resultará em 'ab c, d ef, gh i'). Ao utilizar **Incluir como texto**, as aspas são tratadas como caracteres normais, de modo que os espaços à esquerda e à direita sejam removidos naturalmente.

Em qualquer ponto enquanto estiver trabalhando nesta caixa de diálogo, clique em **Atualizar** para recarregar os campos na origem de dados. Isso é útil quando estiver alterando conexões de dados com o nó de origem ou quando estiver trabalhando entre as guias na caixa de diálogo.

Importando dados geoespaciais no Nó Arquivo Variável

Se o nó contiver dados geoespaciais, foi criado como uma exportação a partir de um arquivo simples e é utilizado no mesmo fluxo em que foi criado, esse nó reterá os metadados geoespaciais e nenhum passo de configuração adicional será necessário.

No entanto, se o nó for exportado e usado em um fluxo diferente, os dados da lista geoespacial serão automaticamente convertidos em um formato de sequência de caracteres, portanto, deve-se seguir alguns passos extras para restaurar o tipo de armazenamento da lista e os metadados geoespaciais associados.

Para obter mais informações sobre listas, consulte “Armazenamento de lista e níveis de medição associados” na página 11.

Para obter mais informações sobre os detalhes que podem ser configurados como metadados geoespaciais, consulte “Subníveis de medição Geoespacial” na página 135.

Para configurar os metadados geoespaciais, utilize os passos a seguir.

1. Na guia Arquivo do nó Arquivo Variável, marque a caixa de seleção **Tratar colchetes como listas**. Marcar esta caixa de seleção significa que os dados incluídos entre colchetes de abertura e fechamento são tratados como um valor único, mesmo se o conteúdo incluir caracteres delimitadores, como vírgulas e aspas duplas. A falha ao marcar essa caixa de seleção significa que os dados são lidos como um tipo de armazenamento de sequência de caracteres, quaisquer vírgulas no campo são processadas como delimitadores e sua estrutura de dados é interpretada incorretamente.
2. Se seus dados incluírem aspas simples ou duplas, selecione a opção **Emparelhar e descartar** nos campos **Aspas simples** e **Aspas duplas**, conforme apropriado.
3. Na guia Dados do nó Arquivo Variável, para os campos de dados geoespaciais, marque a caixa de seleção **Substituir** e altere o tipo de **Armazenamento** de uma sequência de caracteres para uma lista.
4. Por padrão, o tipo **Armazenamento** de lista é configurado como uma *Lista de real* e o tipo de armazenamento de valor subjacente do campo da lista é configurado para *Real*. Para alterar o tipo de armazenamento de valor subjacente ou a profundidade, clique em **Especificar...** para exibir a subcaixa de diálogo Armazenamento.
5. Na subcaixa de diálogo Armazenamento, é possível modificar as seguintes configurações:
 - **Armazenamento** Especifique o tipo de armazenamento geral do campo de dados. Por padrão, o tipo de armazenamento é configurado para Lista, no entanto, a lista suspensa contém todos os

outros tipos de armazenamento (String, Integer, Real, Data, Time e Timestamp). Se selecionar qualquer tipo de armazenamento diferente de Lista, as opções de **Valor de armazenamento** e **Profundidade** estarão indisponíveis.

- **Armazenamento de valor** Especifique os tipos de armazenamento dos elementos na lista, ao invés de especificar para o campo como um todo. Ao importar campos geoespaciais, os únicos tipos de armazenamentos relevantes são Real e Integer; a configuração padrão é Real.
- **Profundidade** Especifique a profundidade do campo de lista. A profundidade necessária depende do tipo de campo geoespacial e segue esses critérios:
 - Point – 0
 - LineString – 1
 - Polygon – 1
 - MultiPoint – 1
 - MultiLineString – 2
 - MultiPolygon – 2

Nota: É necessário saber o tipo de campo geoespacial que estiver convertendo de volta em uma lista e a profundidade necessária para esse tipo de campo. Se essas informações forem configuradas incorretamente, não será possível utilizar o campo.

6. Na guia Tipos do nó Arquivo da Variável, para o campo de dados geoespaciais, assegure-se de que a célula **Medição** contenha o nível de medição correto. Para alterar o nível de medição, na célula **Medição**, clique em **Especificar...** para exibir a caixa de diálogo Valores.
7. Na caixa de diálogo Valores, a **Medição**, o **Armazenamento** e a **Profundidade** para a lista são exibidos. Selecione a opção **Especificar valores e rótulos** e, na lista suspensa **Tipo**, selecione o tipo correto para a **Medição**. Dependendo do **Tipo**, poderão ser solicitados mais detalhes, como se os dados representam 2 ou 3 dimensões e qual sistema de coordenadas será utilizado.

Nó Arquivo Fixo

É possível utilizar os nós Arquivo Fixo para importar dados de arquivos de texto de campo fixo (arquivos cujos campos não são delimitados, mas iniciam na mesma posição e possuem um comprimento fixo). Dados gerados por máquina ou legados são frequentemente armazenados em formato de campo fixo. Utilizando a guia Arquivo do nó Arquivo Fixo, é possível especificar facilmente a posição e o comprimento das colunas em seus dados.

Configuração Opções para o Nó Arquivo Fixo

A guia Arquivo do nó Arquivo Fixo permite colocar dados no IBM SPSS Modeler e especificar a posição das colunas e o comprimento de registros. Utilizando a área de janela de visualização de dados no centro da caixa de diálogo, é possível clicar para incluir setas para especificar os pontos de interrupção entre os campos.

Arquivo. Especifique o nome do arquivo. É possível inserir um nome do arquivo ou clicar no botão de reticências (...) para selecionar um arquivo. Após selecionar um arquivo, o caminho de arquivo é mostrado e seu conteúdo é exibido com delimitadores no painel abaixo.

A área de janela de visualização de dados pode ser utilizada para especificar a posição e o comprimento da coluna. A régua de controle na parte superior da janela de visualização ajuda a medir o comprimento das variáveis e a especificar o ponto de interrupção entre elas. É possível especificar linhas de ponto de interrupção clicando na área da régua de controle acima dos campos. Os pontos de interrupção podem ser movidos ao arrastar e podem ser descartados ao arrastá-los para fora da região de visualização de dados.

- Cada linha de ponto de interrupção inclui automaticamente um novo campo na tabela de campos abaixo.

- Posições iniciais indicadas pelas setas são automaticamente incluídas na coluna Início na tabela abaixo.

Orientação por linha. Selecione se deseja ignorar o caractere de nova linha no término de cada registro.

Ignorar linhas de cabeçalho. Especifique quantas linhas deseja ignorar no início do primeiro registro. Isso é útil para ignorar cabeçalhos da coluna.

Comprimento do registro. Especifique o número de caracteres em cada registro.

Campo. Todos os campos que tiverem sido definidos para este arquivo de dados são listados aqui. Há duas maneiras de definir campos:

- Especifique campos interativamente utilizando a área de janela de visualização de dados acima.
- Especifique campos manualmente ao incluir linhas campo vazias na tabela abaixo. Clique no botão à direita da área de janela de campos para incluir novos campos. Em seguida, no campo vazio, insira um nome de campo, uma posição inicial e um comprimento. Essas opções incluem automaticamente setas na área de janela de visualização de dados, que pode ser ajustada facilmente.

Para remover um campo definido anteriormente, selecione o campo na lista e clique no botão Excluir em vermelho.

Iniciar. Especifique a posição do primeiro caractere no campo. Por exemplo, se o segundo campo de um registro começar no décimo-sexto caractere, você insere 16 como o ponto de início.

Comprimento. Especifique quantos caracteres há no maior valor para cada campo. Isso determina o ponto de interrupção para o próximo campo.

Remover espaços à esquerda e à direita. Selecione para descartar espaços à esquerda e à direita nas sequências de caracteres na importação.

Nota. As comparações entre sequências de caracteres que usam e não usam SQL pushback podem gerar diferentes resultados onde existem espaços à direita.

Caracteres inválidos. Selecione **Descartar** para remover caracteres inválidos da entrada de dados. Selecione **Substituir por** para substituir os caracteres inválidos por símbolos especificados (um caractere apenas). Os caracteres inválidos são caracteres nulos (0) ou qualquer caractere que não existe na codificação atual.

Codificação. Especifica o método de codificação de texto usado. É possível escolher entre o padrão do sistema, o padrão de fluxo ou UTF-8.

- O padrão do sistema é especificado no Painel de Controle do Windows ou, se estiver em execução no modo distribuído, no computador servidor.
- O padrão de fluxo é especificado na caixa de diálogo Propriedades do Fluxo.

Símbolo decimal. Selecione o tipo de separador decimal utilizado em sua origem de dados. **Padrão de fluxo** é o caractere selecionado a partir da guia Opções da caixa de diálogo de propriedades do fluxo. Caso contrário, selecione **Ponto (.)** ou **Vírgula (,)** para ler todos os dados nessa caixa de diálogo usando o caractere escolhido como o separador decimal.

Reconhecer datas e horas automaticamente. Marque essa caixa de seleção para permitir que o IBM SPSS Modeler tente reconhecer automaticamente as entradas de dados como datas ou horas. Por exemplo, isso significa que uma entrada como 11-07-1965 será identificada como uma data e 02:35:58 será identificada como uma hora, no entanto, entradas ambíguas como 07111965 ou 023558 serão mostradas como números inteiros já que não existem delimitadores entre os números.

Nota: para evitar problemas potenciais de dados ao utilizar arquivos de dados a partir de versões anteriores do IBM SPSS Modeler, esta caixa fica desativada por padrão para informações salvas em versões anteriores a 13.

Linhas que serão varridas por tipo. Especifique quantas linhas serão varridas por tipos de dados específicos.

Em qualquer ponto enquanto estiver trabalhando nesta caixa de diálogo, clique em **Atualizar** para recarregar os campos na origem de dados. Isso é útil quando alterar as conexões de dados para o nó de origem ou quando trabalhar entre as guias na caixa de diálogo.

Nó do Coleta de dados

Os nós de origem do Coleta de dados importam dados de pesquisa de opinião com base no IBM SPSS Data Collection Survey Reporter Developer Kit utilizado pelo software de pesquisa de mercado a partir do IBM Corp.. Esse formato distingue os **dados do caso** – as respostas reais para as questões reunidas durante uma pesquisa de opinião - dos **metadados** que descrevem como os dados do caso são coletados e organizados. Os metadados consistem em informações como textos da questão, nomes e descrições de variável, definições de variável de respostas diversas, traduções das sequências de caracteres de texto e a definição da estrutura dos dados do caso.

Nota: este nó requer o IBM SPSS Data Collection Survey Reporter Developer Kit, que é distribuído junto dos produtos de software do IBM SPSS Data Collection IBM Corp.. Para obter mais informações, consulte a página da web do IBM SPSS Data Collection em <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>. Além de instalar o Developer Kit, nenhuma configuração adicional será necessária.

Comentários

- Os dados da pesquisa de opinião são lidos a partir do formato VDATA tabular simples ou a partir de origens no formato HDATA hierárquico, se elas incluírem uma origem de metadados (requer o IBM SPSS Data Collection 4.5 ou superior).
- Os tipos são instanciados automaticamente utilizando informações dos metadados.
- Quando os dados da pesquisa de opinião são importados no IBM SPSS Modeler, as questões são renderizadas como campos, com um registro para cada respondente.

Opções do Arquivo de Importação do Coleta de dados

A guia Arquivo no nós do Coleta de dados permite especificar opções para os dados de metadados e do caso que deseja importar.

Configurações de Metadados

Nota: para ver a lista completa de tipos de arquivo de provedor disponíveis, é necessário instalar o IBM SPSS Data Collection Survey Reporter Developer Kit, disponível com o software IBM SPSS Data Collection. Para obter mais informações, consulte a página da web no IBM SPSS Data Collection em <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>

Provedor de Metadados. Os dados de pesquisa de opinião podem ser importados de um número de formatos, conforme suportados pelo software IBM SPSS Data Collection Survey Reporter Developer Kit. Os tipos de provedores disponíveis incluem os seguintes:

- **DataCollectionMDD.** Lê de metadados a partir de um arquivo de definição de questionário (*.mdd*). Este é o formato do Modelo de Dados padrão do IBM SPSS Data Collection.
- **Banco de dados ADO.** Lê dados e metadados do caso a partir de arquivos ADO. Especifique o nome e o local do arquivo *.adoinfo* que contém os metadados. O nome interno desse DSC é *mrADODsc*.
- **In2data Database.** Lê dados do caso e metadados do In2data. O nome interno desse DSC é *mrI2dDsc*.

- **Arquivo de Log de Coleta de Dados.** Lê metadados a partir de um arquivo de log padrão do IBM SPSS Data Collection. Geralmente, os arquivos de log possuem uma extensão de nome do arquivo *.tmp*. No entanto, alguns arquivos de log podem ter outra extensão de nome do arquivo. Se necessário, você pode renomear o arquivo para que tenha uma extensão de arquivo *.tmp*. O nome interno desse DSC é *mrLogDsc*.
- **Arquivo de Definições de Quancept.** Converte metadados em script Quancept. Especifique o nome do arquivo *.qdi* do Quancept. O nome interno desse DSC é *mrQdiDrsDsc*.
- **Quanvert Database.** Lê dados e metadados do caso do Quanvert. Especifique o nome e a localização do arquivo *.qvinf* ou *.pkd*. O nome interno desse DSC é *mrQvDsc*.
- **Banco de Dados de Participação de Coleta de Dados.** Lê tabelas de Tabela de Amostra e de Histórico de um projeto e cria variáveis categóricas derivadas correspondentes às colunas nessas tabelas. O nome interno desse DSC é *mrSampleReportingMDSC*.
- **Arquivo de Estatísticas.** Lê dados e metadados do caso a partir de um arquivo *.sav* do IBM SPSS Statistics. Grava dados do caso em um arquivo *.sav* do IBM SPSS Statistics para análise no IBM SPSS Statistics. Grava metadados a partir de um arquivo *.sav* do IBM SPSS Statistics em um arquivo *.mdd*. O nome interno desse DSC é *mrSavDsc*.
- **Arquivo Surveycraft.** Lê dados e metadados do caso do SurveyCraft. Especifique o nome do arquivo *.vq* do Surveycraft. O nome interno desse DSC é *mrSCDsc*.
- **Arquivo de Script de Coleta de Dados.** Lê a partir de metadados em um arquivo *mrScriptMetadata*. Normalmente, esses arquivos têm uma extensão de nome do arquivo *.mdd* ou *.dms*. O nome interno desse DSC é *mrScriptMDSC*.
- **Arquivo XML Triple-S.** Lê de metadados a partir de um arquivo Triple-S no formato XML. O nome interno desse DSC é *mrTripleSDsc*.

Propriedades de metadados. Opcionalmente, selecione **Propriedades** para especificar a versão da pesquisa de opinião a ser importada, e também o idioma, o contexto e o tipo de rótulo para uso. Consulte o tópico “Propriedades de Metadados de Importação do IBM SPSS Data Collection” na página 35 para obter mais informações.

Configurações de Dados de Caso

Nota: para ver a lista completa de tipos de arquivo de provedor disponíveis, é necessário instalar o IBM SPSS Data Collection Survey Reporter Developer Kit, disponível com o software IBM SPSS Data Collection. Para obter mais informações, consulte a página da web no IBM SPSS Data Collection em <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>

Obter Configurações de Dados do Caso. Ao ler metadados a partir de arquivos *.mdd* somente, clique em **Obter Configurações de Dados do Caso** para determinar quais origens de dados do caso estão associadas aos metadados selecionados, junto das configurações específicas necessárias para acessar uma determinada origem. Esta opção está disponível apenas para arquivos *.mdd*.

Provedor de Dados do Caso. Os tipos de provedor a seguir são suportados:

- **Banco de dados ADO.** Lê dados do caso utilizando a interface ADO da Microsoft. Selecione UDL do OLE-DB para o tipo de dados do caso e especifique uma sequência de caracteres de conexão no campo UDL de Dados do Caso. Consulte o tópico “Sequência de Conexão de Banco de Dados” na página 35 para obter mais informações. O nome interno deste componente é *mrADODsc*.
- **Arquivo de Texto Delimitado (Excel).** Lê dados do caso a partir de um arquivo delimitado por vírgulas (.CSV), assim como pode ser gerado pelo Excel. O nome interno é *mrCsvDsc*.
- **Arquivo de Dados de Coleta de Dados.** Lê dados do caso a partir de um arquivo de Formato de Dados nativo do IBM SPSS Data Collection (IBM SPSS Data Collection 4.5 em diante). O nome interno é *mrDataFileDsc*.
- **In2data Database.** Lê dados e metadados do caso a partir de um arquivo de metadados In2data (*.i2d*). O nome interno é *mrI2dDsc*.

- **Arquivo de Log de Coleta de Dados.** Lê dados do caso a partir de um arquivo de log padrão do IBM SPSS Data Collection. Geralmente, os arquivos de log possuem uma extensão de nome do arquivo *.tmp*. No entanto, alguns arquivos de log podem ter outra extensão de nome do arquivo. Se necessário, você pode renomear o arquivo para que tenha uma extensão de arquivo *.tmp*. O nome interno é *mrLogDsc*.
- **Arquivo de Dados Quantum.** Lê dados do caso de qualquer arquivo ASCII de formato Quantum (*.dat*). O nome interno é *mrPunchDsc*.
- **Arquivo de Dados Quancept.** Lê dados do caso a partir de um arquivo *.drs*, *.drz* ou *.dru* Quancept. O nome interno é *mrQdiDrsDsc*.
- **Banco de dados Quanvert.** Lê dados do caso a partir de um arquivo *qvinfo* ou *.pkd* do Quanvert. O nome interno é *mrQvDsc*.
- **Banco de Dados de Coleta de Dados (MS SQL Server).** Lê dados do caso para um banco de dados relacional do Microsoft SQL Server. Consulte o tópico “Sequência de Conexão de Banco de Dados” na página 35 para obter mais informações. O nome interno é *mrRdbDsc2*.
- **Arquivo de Estatísticas.** Lê dados do caso a partir de um arquivo *.sav* do IBM SPSS Statistics. O nome interno é *mrSavDsc*.
- **Arquivo Surveycraft.** Lê dados do caso a partir de um arquivo *.qdt* do SurveyCraft. Os arquivos *.vq* e *.qdt* devem estar no mesmo diretório, com acesso de leitura e de gravação para ambos os arquivos. Isso não é o modo como eles são criados por padrão ao utilizar SurveyCraft, portanto, um dos arquivos precisa ser movido para importar dados do SurveyCraft. O nome interno é *mrScDsc*.
- **Arquivo de Dados Triple-S.** Lê dados do caso a partir de um arquivo de dados Triple-S, em forma de comprimento fixo ou delimitado por vírgulas. O nome interno é *sr TripleDsc*.
- **XML de Coleta de Dados.** Lê dados do caso a partir de um arquivo de dados XML do IBM SPSS Data Collection. Normalmente, esse formato pode ser usado para transferir dados do caso de uma localização para outra. O nome interno é *mrXmlDsc*.

Tipo de dados do caso. Especifica se os dados do caso são lidos a partir de um arquivo, pasta, UDL do OLE-DB ou DSN ODBC e atualiza as opções da caixa de diálogo de maneira apropriada. As opções válidas dependem do tipo de provedor. Para provedores de banco de dados, é possível especificar opções para a conexão OLE-DB ou ODBC. Consulte o tópico “Sequência de Conexão de Banco de Dados” na página 35 para obter mais informações.

Projeto de dados do caso. Ao ler os dados do caso a partir de um banco de dados do IBM SPSS Data Collection, é possível inserir o nome do projeto. Para todos os outros tipos de dados do caso, essa configuração deverá ser deixada em branco.

Importação Variável

Importar Variáveis do Sistema (Variáveis). Especifica se as variáveis do sistema são importadas, incluindo variáveis que indicam o status da entrevista (em andamento, concluída, data de conclusão, e assim por diante). É possível escolher **Nenhuma**, **Todas** ou **Comum**.

Importar Variáveis de "Códigos". Controla a importação de variáveis que representam os códigos utilizados para respostas "Outras" abertas para variáveis categóricas.

Importar Variáveis "SourceFile". Controla a importação de variáveis que contêm nomes de arquivos de imagens de respostas varridas.

Importar variáveis de múltiplas respostas como. Variáveis de múltiplas respostas podem ser importadas como campos de flag múltiplos (um conjunto de múltiplas dicotomias), que é o método padrão para novos fluxos. Os fluxos criados em liberações do IBM SPSS Modeler anteriores a 12.0 importavam múltiplas respostas em um único campo, com valores separados por vírgulas. O método mais antigo ainda é suportado para permitir que os fluxos existentes sejam executados conforme tal como

anteriormente, no entanto, é recomendado atualizar fluxos mais antigos para usar o novo método. Consulte o tópico “Importando Conjuntos de Múltiplas Respostas” na página 36 para obter mais informações.

Propriedades de Metadados de Importação do IBM SPSS Data Collection

Ao importar dados de pesquisa de opinião do IBM SPSS Data Collection, na caixa de diálogo Propriedades de Metadados, é possível especificar a versão da pesquisa de opinião para importação e também o idioma, o contexto e tipo de rótulo para uso. Observe que apenas um idioma, contexto e tipo de rótulo podem ser importados por vez.

Versão. Cada versão da pesquisa de opinião pode ser considerada como uma captura instantânea dos metadados utilizados para coletar um conjunto específico de dados do caso. Como um questionário sofre mudanças, diversas versões podem ser criadas. É possível importar a versão mais recente, todas as versões ou uma versão específica.

- **Todas as versões.** Selecione essa opção se você desejar usar uma combinação (superconjunto) de todas as versões disponíveis. (Isso é, às vezes, chamado de superversão). Quando há um conflito entre as versões, as versões mais recentes em geral têm precedência sobre as versões mais antigas. Por exemplo, se um rótulo de categoria é diferente em qualquer uma das versões, o texto da versão mais recente será usado.
- **Versão mais recente.** Selecione essa opção se você desejar utilizar a versão mais recente.
- **Especificar versão.** Selecione essa opção se quiser utilizar uma versão específica da pesquisa de opinião.

Escolher todas as versões é útil quando, por exemplo, você deseja exportar dados do caso para mais de uma versão e se houve mudanças nas definições de variável e de categoria que significam que os dados do caso coletados com uma versão não são válidos em outra versão. Selecionar todas as versões para as quais você deseja exportar os dados do caso significa que geralmente é possível exportar os dados do caso coletados com as diferentes versões ao mesmo tempo sem que ocorram erros de validade devido às diferenças entre as versões. No entanto, dependendo das mudanças de versão, alguns erros de validade ainda podem ser encontrados.

Idioma. Questões e o texto associado podem ser armazenados em vários idiomas nos metadados. É possível utilizar o idioma padrão para a pesquisa de opinião ou especificar um idioma específico. Se um item não estiver disponível no idioma especificado, o padrão será utilizado.

Contexto. Selecione o contexto do usuário que deseja usar. O contexto do usuário controla quais textos são exibidos. Por exemplo, selecione **Questão** para exibir os textos de questão ou **Análise** para exibir textos mais curtos adequados para exibição ao analisar os dados.

Tipo de rótulo. Lista os tipos de rótulos que foram definidos. O padrão é **label**, que é utilizado para os textos de questão nas descrições de contexto do usuário e de variável de Questão no contexto do usuário de Análise. Outros tipos de rótulo podem ser definidos para obter instruções, descrições, e assim por diante.

Sequência de Conexão de Banco de Dados

Ao utilizar o nó IBM SPSS Data Collection para importar dados do caso de um banco de dados por meio de um OLE-DB ou ODBC, selecione **Editar** na guia Arquivo para acessar a caixa de diálogo Sequência de Conexões, que permite customizar a sequência de conexões transmitida para o provedor para realizar um ajuste preciso da conexão.

Propriedades Avançadas

Ao utilizar o nó do IBM SPSS Data Collection para importar dados do caso de um banco de dados que requer um login explícito, selecione **Avançado** para fornecer um ID de usuário e senha para acessar a origem de dados.

Importando Conjuntos de Múltiplas Respostas

As variáveis de respostas diversas podem ser importadas a partir do IBM SPSS Data Collection como conjuntos de múltiplas dicotomias, com um campo de flag separado para cada valor possível da variável. Por exemplo, se os respondentes forem solicitados a selecionar em uma lista quais museus eles visitaram, o conjunto incluirá um campo de flag separado para cada museu listado.

Após importar os dados, é possível incluir ou editar conjuntos de múltiplas respostas a partir de qualquer nó que incluir uma guia Filtro. Consulte o tópico “Editando Conjuntos de Múltiplas Respostas” na página 147 para obter mais informações.

Importando Múltiplas Respostas em um Campo Único (para Fluxos de Criados em Liberações Anteriores)

Em liberações mais antigas do IBM SPSS Modeler, ao invés de importar múltiplas respostas conforme descrito acima, elas foram importadas em um campo único, com valores separados por vírgulas. Esse método ainda é usado para suportar fluxos existentes, mas é recomendado que quaisquer fluxos desse tipo sejam atualizados para utilizar o novo método.

Notas de Importação da Coluna do IBM SPSS Data Collection

As colunas a partir dos dados do IBM SPSS Data Collection são lidas no IBM SPSS Modeler conforme sumarizadas na tabela a seguir.

Tabela 4. Sumarização de importação de coluna do IBM SPSS Data Collection

Tipo de Coluna do IBM SPSS Data Collection	IBM SPSS Modeler Armazenamento	Nível de medição
Flag booleano (sim/não)	Sequência de caracteres	Flag (valores 0 e 1)
Catégorico	Sequência de caracteres	Nominal
Registro de data e hora	Registro de data e hora	Contínuo
Duplo (valor de ponto flutuante dentro de um intervalo especificado)	Real	Contínuo
Longo (valor de número inteiro dentro de um intervalo especificado)	Número inteiro	Contínuo
Texto (descrição de texto livre)	Sequência de caracteres	Sem tipo
Nível (indica grades ou loops em uma questão)	Não ocorre no VDATA e não é importado no IBM SPSS Modeler	
Objeto (dados binários como um fac-símile mostrando texto rabiscado ou uma gravação de voz)	Não importado no IBM SPSS Modeler	
Nenhum (tipo desconhecido)	Não importado no IBM SPSS Modeler	
Coluna Respondent.Serial (associa um ID exclusivo a cada respondente)	Número inteiro	Sem tipo

Para evitar possíveis inconsistências entre os rótulos do valor lidos a partir de metadados e os valores reais, todos os valores de metadados são convertidos em minúsculos. Por exemplo, o valor do rótulo *E1720_years* é convertido para *e1720_years*.

Nó Origem do IBM Cognos BI

O nó de origem do IBM Cognos BI permite colocar dados do banco de dados do Cognos BI ou de relatórios de lista única em sua sessão de mineração de dados. Dessa forma, é possível combinar os recursos de inteligência de negócios do Cognos com os recursos de análise preditiva do IBM SPSS Modeler. É possível importar dados relacionais, relacionais modelados dimensionalmente (DMR) e OLAP.

A partir de uma conexão do servidor Cognos, primeiro você seleciona uma localização a partir da qual importar os dados ou relatórios. Uma localização contém um modelo Cognos e todas as pastas, queries, relatórios, visualizações, atalhos, URLs e definições de tarefas associados a esse modelo. Um modelo do Cognos define as regras de negócios, as descrições de dados, os relacionamentos de dados, as dimensões e hierarquias de negócios e outras tarefas administrativas.

Se estiver importando dados, então seleciona os objetos que deseja importar do pacote selecionado. Os objetos que podem ser importados incluem assuntos de query (que representam tabelas de banco de dados) ou itens de query individuais (que representam colunas da tabela). Consulte o tópico “Ícones de objeto do Cognos” para obter mais informações.

Se o pacote tiver filtros definidos, será possível importar um ou mais deles. Se um filtro que importar estiver associado a dados importados, esse filtro será aplicado antes que os dados sejam importados.

Nota: os dados a serem importados devem estar no formato UTF-8.

Se estiver importando um relatório, selecione um pacote ou uma pasta dentro de um pacote, contendo um ou mais relatórios. Em seguida, selecione o relatório individual que deseja importar. *Nota:* apenas relatórios de lista única podem ser importados, listas múltiplas não são suportadas.

Se parâmetros foram definidos, seja para um objeto de dados ou para um relatório, será possível especificar valores para esses parâmetros antes de importar o objeto ou o relatório.

Ícones de objeto do Cognos

Os vários tipos de objetos que você pode importar a partir de um banco de dados do Cognos BI são representados por ícones diferentes, como a tabela a seguir ilustra.

Tabela 5. Ícones de objeto do Cognos.

Ícone	Objeto
	Pacote
	Espaço de Nomes
	Assunto da consulta
	Item de Consulta
	Dimensão da medida
	Medir
	Dimensão
	Nível de Hierarquia
	Nível

Tabela 5. Ícones de objeto do Cognos (continuação).

Ícone	Objeto
	Filtro
	Relatório
	Cálculo independente

Importando Dados do Cognos

Para importar dados de um banco de dados IBM Cognos BI, na guia Dados da caixa de diálogo IBM Cognos BI, assegure-se de que **Modo** esteja configurado para **Dados**.

Conexão. Clique no botão **Editar** para exibir uma caixa de diálogo na qual é possível definir os detalhes de uma nova conexão do Cognos a partir da qual importar dados ou relatórios. Se você já estiver com login efetuado em um servidor Cognos por meio do IBM SPSS Modeler, também será possível editar os detalhes da conexão atual. Consulte o tópico “Conexões Cognos” na página 39 para obter mais informações.

Localização. Quando você tiver estabelecido a conexão do servidor Cognos, clique no botão **Editar** ao lado desse campo para exibir uma lista de pacotes disponíveis a partir da qual importar o conteúdo. Consulte o tópico “Seleção de localização do Cognos” na página 40 para obter mais informações.

Conteúdo. Exibe o nome do pacote selecionado, junto dos namespaces associados ao pacote. Dê um clique duplo em um namespace para exibir os objetos que podem ser importados. Os vários tipos de objetos são denotados por ícones diferentes. Consulte o tópico “Ícones de objeto do Cognos” na página 37 para obter mais informações.

Para escolher um objeto a ser importado, selecione o objeto e clique na parte superior das duas setas para a direita para mover o objeto para a área de janela **Campos para importar**. Selecionar um assunto de query importa todos os seus itens de query. Dar um clique duplo em um assunto de query expande-o para que seja possível escolher um ou mais de seus itens de query individuais. É possível executar diversas seleções com Ctrl-clique (selecionar itens individuais), Shift-clique (selecionar um bloco de itens) e Ctrl-A (selecionar todos os itens).

Para escolher um filtro a ser aplicado (se o pacote possuir filtros definidos), navegue para o filtro na área de janela de Conteúdo, selecione o filtro e clique na mais baixa das duas setas para a direita para mover o filtro para a área **Filtros para aplicar**. É possível executar diversas seleções com Ctrl-clique em (selecionar filtros individuais) e Shift-clique em (selecionar um bloco de filtros).

Campos para importar. Lista os objetos de banco de dados que você escolheu importar no IBM SPSS Modeler para processamento. Se não precisar mais de um objeto específico, selecione-o e clique na seta para a esquerda para retorná-lo à área de janela **Conteúdo**. É possível executar diversas seleções da mesma maneira que para **Conteúdo**.

Filtros para aplicar. Lista os filtros que você escolheu aplicar aos dados antes de eles serem importados. Se não precisar mais de um filtro específico, selecione-o e clique na seta para a esquerda para retorná-lo à área de janela **Conteúdo**. É possível executar diversas seleções da mesma maneira que para **Conteúdo**.

Parâmetros. Se este botão estiver ativado, o objeto selecionado possui parâmetros definidos. É possível utilizar parâmetros para fazer ajustes (por exemplo, executar um cálculo parametrizado) antes de importar os dados. Se os parâmetros estiverem definidos, mas nenhum padrão for fornecido, o botão

exibirá um triângulo de aviso. Clique no botão para exibir os parâmetros e, opcionalmente, editá-los. Se o botão estiver desativado, o relatório não possui parâmetros definidos.

Agregar os dados antes de importar. Marque esta caixa se desejar importar dados agregados ao invés de dados brutos.

Importando relatórios do Cognos

Para importar um relatório predefinido de um banco de dados IBM Cognos BI, na guia Dados da caixa de diálogo IBM Cognos BI, assegure-se de que **Modo** esteja configurado para **Relatório**. *Nota:* apenas relatórios de lista única podem ser importados, listas múltiplas não são suportadas.

Conexão. Clique no botão **Editar** para exibir uma caixa de diálogo na qual é possível definir os detalhes de uma nova conexão do Cognos a partir da qual importar dados ou relatórios. Se você já estiver com login efetuado em um servidor Cognos por meio do IBM SPSS Modeler, também será possível editar os detalhes da conexão atual. Consulte o tópico “Conexões Cognos” para obter mais informações.

Localização. Quando você tiver estabelecido a conexão do servidor Cognos, clique no botão **Editar** ao lado desse campo para exibir uma lista de pacotes disponíveis a partir da qual importar o conteúdo. Consulte o tópico “Seleção de localização do Cognos” na página 40 para obter mais informações.

Conteúdo. Exibe o nome do pacote ou da pasta selecionada que contém os relatórios. Navegue para um relatório específico, selecione-o e clique na seta para a direita para trazer o relatório para o campo **Relatório para importação**.

Relatório para importação. Indica o relatório que você escolheu importar no IBM SPSS Modeler. Se não precisar mais do relatório, selecione-o e clique na seta para a esquerda para retorná-lo para a área de janela **Conteúdo** ou traga um relatório diferente para esse campo.

Parâmetros. Se este botão estiver ativado, o relatório selecionado possui parâmetros definidos. É possível utilizar parâmetros para fazer ajustes antes de importar o relatório (por exemplo, especificar uma data de início e de encerramento para dados do relatório). Se os parâmetros estiverem definidos, mas nenhum padrão for fornecido, o botão exibirá um triângulo de aviso. Clique no botão para exibir os parâmetros e, opcionalmente, editá-los. Se o botão estiver desativado, o relatório não possui parâmetros definidos.

Conexões Cognos

Na caixa de diálogo Cognos Connections, é possível selecionar o servidor Cognos BI a partir do qual importar ou exportar objetos de banco de dados.

URL do servidor Cognos Digite a URL do servidor Cognos BI a partir da qual importar ou exportar. Este é o valor da propriedade do ambiente "URI do Dispatcher Externo" da Configuração do IBM Cognos no servidor Cognos BI. Entre em contato com o administrador do seu sistema de Cognos se você não estiver certo de qual URL usar.

Modo Selecione **Configurar Credenciais** se desejar efetuar login com um namespace, nome do usuário e senha do Cognos específicos (por exemplo, como um administrador). Selecione **Usar conexão anônima** para efetuar login sem credenciais do usuário, caso em que os outros campos não são preenchidos.

Como alternativa, se você tiver uma credencial do IBM Cognos que estiver armazenada no Repositório do IBM SPSS Collaboration and Deployment Services, será possível usar essa credencial ao invés de inserir as informações de nome de usuário e senha ou criar uma conexão anônima. Para usar a credencial existente, selecione **Credenciais Armazenadas** e insira o **Nome da Credencial** ou procure por ela.

O namespace Cognos é modelado por um domínio em IBM SPSS Collaboration and Deployment Services.

ID de Namespace Especifique o provedor de autenticação de segurança do Cognos que é utilizado para efetuar logon no servidor. O provedor de autenticação é usado para definir e manter usuários, grupos e funções e para controlar o processo de autenticação. Observe que este é o ID do Namespace, não o Nome do Namespace (o ID nem sempre é igual ao Nome).

Nome do usuário Insira o nome de usuário do Cognos com o qual efetuar logon no servidor.

Senha Insira a senha associada ao nome de usuário especificado.

Salvar como Padrão Clique nesse botão para armazenar essas configurações como seu padrão para evitar reinseri-las toda vez que abrir o nó.

Seleção de localização do Cognos

A caixa de diálogo Especificar Localização permite selecionar um pacote do Cognos a partir do qual importar os dados ou um pacote ou uma pasta a partir da qual importar os relatórios.

Pastas Públicas. Se estiver importando dados, isso listará os pacotes e as pastas disponíveis a partir do servidor escolhido. Selecione o pacote que deseja utilizar e clique em **OK**. É possível selecionar apenas um pacote por nó de origem do Cognos BI.

Se estiver importando relatórios, isso listará as pastas e os pacotes contendo relatórios que estiverem disponíveis a partir do servidor escolhido. Selecione um pacote ou pasta de relatórios e clique em **OK**. É possível selecionar apenas um pacote ou uma pasta de relatórios por nó de origem do Cognos BI, embora as pastas de relatórios possam conter outras pastas de relatórios, bem como relatórios individuais.

Especificando parâmetros para dados ou relatórios

Se parâmetros tiverem sido definidos no Cognos BI, para um objeto de dados ou um relatório, será possível especificar valores para esses parâmetros antes de importar o objeto ou o relatório. Um exemplo de parâmetros para um relatório seria datas de início e de encerramento do conteúdo do relatório.

Nome. O nome do parâmetro conforme ele é especificado no banco de dados do Cognos BI.

Tipo. Uma descrição do parâmetro.

Valor. O valor a ser designado ao parâmetro. Para inserir ou editar um valor, dê um clique duplo em sua célula na tabela. Os valores não são validados aqui, de modo quaisquer valores inválidos são detectados no tempo de execução.

Remover automaticamente parâmetros inválidos da tabela. Essa opção é selecionada por padrão e removerá todos os parâmetros inválidos encontrados dentro do objeto ou relatório de dados.

Nó Origem do IBM Cognos TM1

O nó de origem do IBM Cognos TM1 permite colocar os dados do Cognos TM1 na sua sessão de mineração de dados. Dessa forma, é possível combinar os recursos de planejamento corporativos do Cognos com os recursos de análise preditiva do IBM SPSS Modeler. É possível importar uma versão simplificada do dados do cubo OLAP multidimensionais.

Nota: O usuário do TM1 precisa das seguintes permissões: privilégio de Gravação de cubos, privilégio de Leitura de processos do SPSS, privilégio de Leitura de dimensões, privilégio de Gravação de elementos de dimensão e privilégio de Gravação de arquivos .blb.

É necessário modificar os dados no TM1 antes que os dados sejam importados.

Nota: Os dados a serem importados devem estar no formato UTF-8.

Em uma conexão do host de administração do IBM Cognos TM1, primeiro selecione um servidor do TM1 a partir do qual importar os dados; um servidor contém um ou mais cubos do TM1. Em seguida, selecione o cubo necessário e, dentro do cubo, selecione as colunas e linhas que deseja importar.

Nota: Antes de poder utilizar os nós Origem ou Exportação do TM1 no SPSS Modeler, deve-se verificar algumas configurações no arquivo `tm1s.cfg`, que é o arquivo de configuração do servidor TM1 no diretório-raiz do servidor TM1.

- `HTTPPortNumber` - configure um número de porta válido; geralmente 1 a 65535.
- `UseSSL` - se configurar para *True*, o HTTPS será usado como o protocolo de transporte. Nesse caso, deve-se importar a certificação do TM1 do JRE SPSS Modeler Server.

Importando de dados do IBM Cognos TM1

Para importar dados de um banco de dados do IBM Cognos TM1, na guia Dados da caixa de diálogo do IBM Cognos TM1, selecione o host de administração do TM1 relevante e detalhes do servidor, do cubo e de dados associados.

Nota: Antes de importar os dados, deve-se executar algum pré-processamento dentro do TM1 para assegurar que os dados estejam em um formato que seja reconhecível para o IBM SPSS Modeler. Isso envolve a filtragem de seus dados utilizando o Editor de Subconjunto para obter a visualização no tamanho e forma corretos para importação.

Observe que valores zero (0) importados do TM1 serão tratados como valores "nulos" (o TM1 não distingue entre valores em branco e zeros). Além disso, observe que os dados não numéricos (ou metadados) a partir de *dimensões regulares* podem ser importados no IBM SPSS Modeler. No entanto, a importação de *medidas* não numéricas não é suportada atualmente.

Host de Administração Digite a URL do host de administração no qual o servidor TM1 ao qual deseja conectar-se está instalado. O host de administração é definido como uma URL única para todos os servidores TM1. Nesta URL, todos os servidores IBM Cognos TM1 instalados e em execução em seu ambiente podem ser descobertos e acessados.

Servidor TM1 Quando tiver estabelecido a conexão com o host de administração, selecione o servidor que contém os dados que deseja importar e clique em **Login**. Se você não tiver se conectado anteriormente a esse servidor, será solicitado a inserir o **Nome do usuário** e a **Senha**; como alternativa, é possível procurar pelos detalhes de login inseridos anteriormente que você salvou como uma **Credencial Armazenada**.

Selecionar uma visualização do cubo do TM1 para importar Exibe o nome dos cubos dentro do servidor TM1 a partir do qual é possível importar dados. Dê um clique duplo em um cubo para exibir os dados de visualização que podem ser importados.

Nota: Apenas os cubos com uma dimensão podem ser importados no IBM SPSS Modeler.

Para escolher dados a serem importados, selecione a visualização e clique na seta para a direita para movê-la para a área **Visualizar para importar**. Se a visualização necessária não aparecer, dê um clique duplo em um cubo para expandir sua lista de visualizações.

Dimensão(ões) de linha. Lista o nome da dimensão da linha nos dados que você escolheu importar. Role pela lista de níveis e selecione aquele que precisar.

Dimensão da coluna Lista o nome da dimensão da coluna nos dados que você escolheu importar. Role pela lista de níveis e selecione aquele que precisar.

Dimensão(ões) de contexto Somente exibição. Mostra as dimensões de contexto que estiverem relacionadas às colunas e linhas selecionadas.

Nó de Origem do SAS

Nota: esse recurso está disponível no SPSS Modeler Professional e no SPSS Modeler Premium.

O nó de origem do SAS permite colocar dados SAS na sua sessão de mineração de dados. É possível importar quatro tipos de arquivos:

- SAS para Windows/OS2 (.sd2)
- SAS para UNIX (.ssd)
- Arquivo de Transporte do SAS (.tpt)
- SAS versão 7/8/9 (.sas7bdat)

Quando os dados são importados, todas as variáveis são mantidas e nenhum tipo de variável é alterado. Todos os casos são selecionados.

Configurando Opções para o Nó de Origem SAS

Importar. Selecione qual tipo de arquivo SAS deseja transportar. É possível escolher **SAS para Windows/OS2 (.sd2)**, **SAS para UNIX (.SSD)**, **Arquivo de Transporte SAS (.tpt)** ou **SAS Versão 7/8/9 (.sas7bdat)**.

Arquivo de importação. Especifique o nome do arquivo. É possível inserir um nome do arquivo ou clicar no botão de reticências (...) para navegar para o local do arquivo.

Membro. Selecione o membro a ser importado a partir do arquivo de transporte SAS selecionado acima. É possível inserir um nome do membro ou clicar em **Selecionar** para procurar todos os membros no arquivo.

Ler formatos de usuário de um arquivo de dados SAS. Selecione para ler formatos do usuário. Os arquivos SAS armazenam dados e formatos de dados (como rótulos de variáveis) em arquivos diferentes. Na maioria das vezes, você desejará para importar também os formatos. Se você tiver um conjunto de dados grande, no entanto, talvez queira cancelar a seleção desta opção para economizar memória.

Arquivo de formato. Se um arquivo de formato for necessário, esta caixa de texto será ativada. É possível inserir um nome do arquivo ou clicar no botão de reticências (...) para navegar para o local do arquivo.

Nomes de variável. Selecione um método para manipular nomes e rótulos de variáveis na importação de um arquivo SAS. Os metadados que você escolher incluir aqui persistirão durante todo o seu trabalho no IBM SPSS Modeler e podem ser exportados novamente para uso no SAS.

- **Ler nomes e rótulos.** Selecione para ler nomes de variáveis e rótulos no IBM SPSS Modeler. Por padrão, esta opção é selecionada e os nomes de variáveis são exibidos no nó Tipo. Os rótulos podem ser exibidos no Construtor de Expressões, nos gráficos, navegadores de modelo e em outros tipos de saída, dependendo das opções especificadas na caixa de diálogo de propriedades do fluxo.
- **Ler rótulos como nomes.** Selecione para ler os rótulos de variável descritivos a partir do arquivo SAS ao invés dos nomes de campo curtos e utilizar esses rótulos como nomes de variáveis no IBM SPSS Modeler.

Nó de Origem do Excel

O nó de origem do Excel permite importar dados do Microsoft Excel no formato de arquivo .xlsx.

Tipo de arquivo. Selecione o tipo de arquivo do Excel que estiver importando.

Importar arquivo. Especifica o nome e a localização do arquivo de planilha para importação.

Usar Intervalo Nomeado. Permite especificar um intervalo nomeado de células conforme definido na planilha do Excel. Clique no botão de reticências (...) para escolher a partir da lista de intervalos disponíveis. Se um intervalo nomeado for utilizado, outras configurações de planilha e de intervalo de dados não serão mais aplicáveis e serão desativadas como resultado.

Escolher planilha. Especifica a planilha a ser importada, seja por índice ou por nome.

- **Por índice.** Especifique o valor de índice para a planilha que você deseja importar, começando com 0 para a primeira planilha, 1 para a segunda planilha e assim por diante.
- **Por nome.** Especifique o nome da planilha que deseja importar. Clique no botão de reticências (...) para escolher na lista de planilhas disponíveis.

Intervalo em planilha. É possível importar dados que começam com a primeira linha preenchida ou com um intervalo de células explícito.

- **O intervalo começa na primeira linha preenchida.** Localiza a primeira célula preenchida e usa como o canto superior esquerdo do intervalo de dados.
- **Intervalo explícito de células.** Permite especificar um intervalo explícito por linha e coluna. Por exemplo, para especificar o intervalo do Excel A1:D5, você pode digitar A1 no primeiro campo e D5 no segundo (ou, como alternativa, R1C1 e R5C4). Todas as linhas no intervalo especificado são retornadas, incluindo as linhas em branco.

Nas linhas em branco. Se mais de uma linha em branco for encontrada, será possível escolher se deseja **Parar a leitura** ou escolher **Retornar linhas em branco** para continuar lendo todos os dados no término da planilha, incluindo linhas em branco.

A primeira linha contém nomes de coluna. Indica que a primeira linha no intervalo especificado deve ser utilizada como nomes de campo (coluna). Se não for selecionada, os nomes de campo serão gerados automaticamente.

Armazenamento de Campo e Nível de Medição

Ao ler os valores a partir do Excel, os campos com armazenamento numérico são lidos com um nível de medição de *Contínuo* por padrão, e os campos de sequência de caracteres são lidos como *Nominal*. É possível alterar manualmente o nível de medição (contínuo versus nominal) na guia Tipo, mas o armazenamento é determinado automaticamente (embora ele possa ser alterado utilizando uma função de conversão, como `to_integer`, em um nó Preenchimento ou nó Derivar, se necessário). Consulte o tópico “Configurando Armazenamento e Formatação do Campo” na página 9 para obter mais informações.

Por padrão, os campos com uma combinação de valores numéricos e de sequência de caracteres são lidos como números, o que significa que quaisquer valores de sequência de caracteres serão configurados para valores nulos (omissos do sistema) no IBM SPSS Modeler. Isso acontece porque, ao contrário do Excel, o IBM SPSS Modeler não permite tipos de armazenamento combinados em um campo. Para evitar isso, é possível configurar manualmente o formato da célula para Texto na planilha do Excel, que faz com que todos os valores (incluindo números) sejam lidos como sequências de caracteres.

Nó de Origem XML

Nota: esse recurso está disponível no SPSS Modeler Professional e no SPSS Modeler Premium.

Utilize o nó de origem XML para importar dados de um arquivo no formato XML em um fluxo do IBM SPSS Modeler. O XML é uma linguagem padrão para troca de dados e, para muitas organizações, é o formato de escolha para esse propósito. Por exemplo, uma agência fiscal do governo pode desejar analisar os dados de declarações de imposto de renda que foram enviadas on-line e que esses dados estejam em formato XML (consulte <http://www.w3.org/standards/xml/>).

A importação de dados XML em um fluxo do IBM SPSS Modeler permite executar uma ampla variedade de funções de análise preditiva na origem. Os dados XML são analisados em um formato tabular em que as colunas correspondem a níveis diferentes de aninhamento dos elementos e atributos XML. Os itens XML são exibidos no formato XPath, (consulte <http://www.w3.org/TR/xpath20/>).

Ler um único arquivo Por padrão, o SPSS Modeler lê um arquivo único, que é especificado no campo **Origem de dados XML**.

Ler todos os arquivos XML em um diretório Escolha esta opção se desejar ler todos os arquivos XML em um diretório específico. Especifique a localização no campo **Diretório** que aparece. Marque a caixa de seleção **Incluir subdiretórios** para ler adicionalmente arquivos XML a partir de todos os subdiretórios do diretório especificado.

Origem de dados XML Digite o caminho e o nome do arquivo completos do arquivo de origem XML que deseja importar, ou utilize o botão Procurar para localizar o arquivo.

Esquema XML (Opcional) Especifique o caminho e o nome do arquivo completos de um arquivo XSD ou DTD a partir do qual ler a estrutura XML, ou utilize o botão Procurar para localizar este arquivo. Se deixar esse campo em branco, a estrutura será lida a partir do arquivo de origem XML. Um arquivo XSD ou DTD pode ter mais de um elemento-raiz. Neste caso, ao alterar o foco para um campo diferente, um diálogo é exibido no qual você escolhe o elemento-raiz que deseja utilizar. Consulte o tópico “Selecionando a partir de diversos elementos raiz.” na página 45 para obter mais informações.

Nota: Os indicadores XSD são ignorados pelo SPSS Modeler

Estrutura XML Uma árvore hierárquica que mostra a estrutura do arquivo de origem XML (ou o esquema, se você especificou um no campo **Esquema XML**). Para definir um limite de registro, selecione um elemento e clique no botão de seta para a direita para copiar o item no campo **Registros**.

Exibir atributos Exibe ou oculta a atributos dos elementos XML no campo **Estrutura XML**.

Registros (expressão XPath) Mostra a sintaxe de XPath para um elemento copiado do campo de estrutura XML. Este elemento é, então, destacado na estrutura XML e define o limite de registro. Um novo registro será criado toda vez que este elemento for encontrado. Se esse campo estiver vazio, o primeiro elemento filho na raiz será usado como o limite de registro.

Ler todos os dados Por padrão, todos os dados no arquivo de origem são lidos no fluxo.

Especifique dados para leitura Escolha esta opção se desejar importar elementos individuais, atributos ou ambos. Escolher essa opção ativa a tabela Campos na qual é possível especificar os dados que deseja importar.

Campos Essa tabela lista os elementos e atributos selecionados para importação, se você tiver selecionado a opção **Especificar dados para leitura**. É possível digitar a sintaxe do XPath de um elemento ou atributo diretamente na coluna XPath, ou selecione um elemento ou atributo na estrutura XML e clicar no botão de seta para a direita para copiar o item na tabela. Para copiar todos os elementos filho e atributos de um elemento, selecione o elemento na estrutura XML e clique no botão de seta para a direita.

- **XPath** A sintaxe XPath dos itens a serem importados.
- **Localização** A localização na estrutura XML dos itens a serem importados. **Caminho fixo** mostra o caminho do item relativo para o elemento destacado na estrutura XML (ou o primeiro elemento filho na raiz, se nenhum elemento estiver destacado). **Qualquer localização** denota um item do nome fornecido em qualquer localização na estrutura XML. **Customizado** será exibido se você digitar uma localização diretamente na coluna XPath.

Selecionando a partir de diversos elementos raiz.

Ao passo que um arquivo XML corretamente formado pode ter somente um elemento raiz único, um arquivo XSD ou DTD pode conter diversas raízes. Se uma das raízes corresponder àquela no arquivo de origem XML, esse elemento raiz será usado, caso contrário, será necessário selecionar uma para ser utilizada.

Escolha a raiz a ser exibida. Selecione o elemento raiz que deseja utilizar. O padrão é o primeiro elemento raiz na estrutura XSD ou DTD.

Removendo espaços indesejados dos dados de origem XML

As quebras de linha nos dados de origem XML podem ser implementadas por uma combinação de caracteres [CR] [LF]. Em alguns casos, essas quebras de linha podem ocorrer no meio de uma sequência de caracteres de texto, por exemplo:

```
<description>An in-depth look at creating applications[CR] [LF]
with XML.</description>
```

Essas quebras de linha podem não estar visíveis quando o arquivo é aberto em alguns aplicativos, por exemplo, um navegador da web. No entanto, quando os dados são lidos no fluxo por meio do nó de origem XML, as quebras de linha são convertidas em uma série de caracteres de espaço.

É possível corrigir isso utilizando um nó Preenchimento para remover esses espaços indesejados:

A seguir há um exemplo de como fazer isso:

1. Anexe um nó Preenchimento ao nó de origem XML.
2. Abra o nó Preenchimento e utilize o seletor de campo para selecionar o campo com os espaços indesejados.
3. Configure **Substituir** para **Baseado na condição** e configure **Condição** para **true**.
4. No campo **Substituir por**, insira `replace(" ", "", @FIELD)` e clique em OK.
5. Anexe um nó Tabela ao nó Preenchimento e execute o fluxo.

Na saída do nó Tabela, o texto agora aparece sem os espaços adicionais.

Nó Entrada do Usuário

O nó Entrada do Usuário fornece uma maneira fácil de criar dados sintéticos, seja desde o início ou alterando dados existentes. Isso é útil, por exemplo, quando desejar criar um conjunto de dados de teste para modelagem.

Criando Dados desde o Início

O nó Entrada do Usuário está disponível a partir da paleta Origens e pode ser incluído diretamente na tela de fluxo.

1. Clique na guia **Origens** na paleta do nó.
2. Arraste e solte ou dê um clique duplo para incluir o nó Entrada do Usuário na tela de fluxo.
3. Dê um clique duplo para abrir sua caixa de diálogo e especificar campos e valores.

Nota: os nós Entrada do Usuário que são selecionados a partir da paleta Origens estarão completamente em branco, sem campos e informações de dados. Isso permite criar dados sintéticos totalmente desde o início.

Gerando Dados a partir de uma Origem de Dados Existente

Também é possível gerar um nó Entrada do Usuário a partir de qualquer nó não terminal no fluxo:

1. Decida em que ponto no fluxo você deseja substituir um nó.
2. Clique com o botão direito no nó que alimentará seus dados no nó Entrada do Usuário e escolha **Gerar Nó de Entrada do Usuário** no menu.
3. O nó Entrada do Usuário é exibido com todos os processos de recebimento de dados anexados a ele, substituindo o nó existente nesse ponto do fluxo de dados. Quando gerado, o nó herda toda a estrutura de dados e as informações do tipo de campo (se disponíveis) a partir dos metadados.

Nota: se os dados não forem executados em todos os nós no fluxo, então os nós não estão totalmente instanciados, significando que os valores de armazenamento e de dados podem não estar disponíveis ao substituir por um nó Entrada do Usuário.

Configurando Opções para o Nó de Entrada do Usuário

A caixa de diálogo de um nó Entrada do Usuário contém várias ferramentas que podem ser usadas para inserir valores e definir a estrutura de dados para dados sintéticos. Para um nó gerado, a tabela na guia Dados contém nomes de campos da origem de dados original. Para um nó incluído a partir da paleta Origens, a tabela está em branco. Utilizando as opções de tabela, é possível executar as seguintes tarefas:

- Incluir novos campos utilizando o botão Incluir um Novo Campo à direita na tabela.
- Renomear campos existentes.
- Especificar armazenamento de dados para cada campo.
- Especificar valores.
- Alterar a ordem dos campos na exibição.

Inserir Dados

Para cada campo, é possível especificar ou inserir valores a partir do conjunto de dados original utilizando o botão de seletor de valor à direita da tabela. Consulte as regras descritas a seguir para obter mais informações sobre como especificar valores. Também é possível escolher deixar o campo em branco - os campos em branco são preenchidos com o nulo do sistema (\$null\$).

Para especificar valores de sequência de caracteres, simplesmente digite-os na coluna Valores, separados por espaços:

```
Fred Ethel Martin
```

Sequências de caracteres que incluem espaços podem ser agrupadas entre aspas duplas:

```
"Bill Smith" "Fred Martin" "Jack Jones"
```

Para campos numéricos, é possível inserir diversos valores da mesma maneira (listados com espaços entre eles):

```
10 12 14 16 18 20
```

Ou é possível especificar a mesma série de números ao configurar seus limites (10, 20) e os passos entre eles (2). Utilizando esse método, você digita:

```
10,20,2
```

Esses dois métodos podem ser combinados ao integrá-los, como:

```
1 5 7 10,20,2 21 23
```

Esta entrada produzirá os seguintes valores:

```
1 5 7 10 12 14 16 18 20 21 23
```

Valores de data e hora podem ser inseridos utilizando o formato padrão atual selecionado na caixa de diálogo Propriedades do Fluxo, por exemplo:

```
11:04:00 11:05:00 11:06:00
```

Para os valores de registro de data e hora, que possuem um componente de data e hora, aspas duplas devem ser utilizadas:

"2007-03-14 11:04:00" "2007-03-14 11:05:00" "2007-03-14 11:06:00"

Para obter detalhes adicionais, consulte os comentários no armazenamento de dados a seguir.

Gerar dados. Permite especificar como os registros são gerados ao executar o fluxo.

- **Todas as combinações.** Gera registros contendo todas as combinações possíveis dos valores de campo, de modo que cada valor de campo apareça em vários registros. Às vezes isso pode gerar mais dados que o desejado, de modo que você geralmente poderá seguir este nó com um nó de amostra.
- **Em ordem.** Gera registros na ordem em que os valores do campo de dados são especificados. Cada valor do campo aparece apenas em um registro. O número total de registros é igual ao maior número de valores para um único campo. Quando os campos têm menos do que o número maior, valores indefinidos (\$null\$) são inseridos.

Mostrar exemplo

Por exemplo, as seguintes entradas geram os registros listados nos dois exemplos de tabela a seguir.

- **Idade.** 30,60,10
- **PA.** LOW
- **Colesterol.** NORMAL HIGH
- **Droga.** (deixar em branco)

Tabela 6. Gerar campo de dados configurado para Todas as combinações.

Idade	BP	Colesterol	Droga
30	LOW	NORMAL	\$null\$
30	LOW	HIGH	\$null\$
40	LOW	NORMAL	\$null\$
40	LOW	HIGH	\$null\$
50	LOW	NORMAL	\$null\$
50	LOW	HIGH	\$null\$
60	LOW	NORMAL	\$null\$
60	LOW	HIGH	\$null\$

Tabela 7. Gerar campo de dados configurado para Na ordem.

Idade	BP	Colesterol	Droga
30	LOW	NORMAL	\$null\$
40	\$null\$	HIGH	\$null\$
50	\$null\$	\$null\$	\$null\$
60	\$null\$	\$null\$	\$null\$

Armazenamento de Dados

Armazenamento descreve a maneira como os dados são armazenados em um campo. Por exemplo, um campo com valores de 1 e 0 armazena dados de número inteiro. Isso é diferente do nível de medição, que descreve o uso dos dados e não afeta o armazenamento. Por exemplo, talvez você queira configurar o nível de medição para um campo de número inteiro com valores de 1 e 0 para *Flag*. Isto geralmente

indica que 1 = *True* e 0 = *False*. Embora o armazenamento deva ser determinado na origem, o nível de medição pode ser alterado utilizando um nó Tipo em qualquer ponto no fluxo. Consulte o tópico “Níveis de Medição” na página 133 para obter mais informações.

Os tipos de armazenamento disponíveis são:

- **Sequência de Caracteres** Usado para campos que contêm dados não numéricos, também chamados de dados alfanuméricos. Uma sequência de caracteres pode incluir qualquer sequência de caracteres, como *fred*, *Classe 2* ou *1234*. Observe que os números em sequências de caracteres não podem ser usados em cálculos.
- **Número Inteiro** Um campo cujos valores são números inteiros.
- **Real** Valores são números que podem incluir decimais (não se limitando a números inteiros). O formato de exibição é especificado na caixa de diálogo Propriedades do Fluxo e pode ser substituído para campos individuais em um nó Tipo (guia Formato).
- **Data** Valores de data especificados em um formato padrão como ano, mês e dia (por exemplo, 2007-09-26). O formato específico é definido na caixa de diálogo Propriedades do Fluxo.
- **Tempo** Tempo medido como uma duração. Por exemplo, uma chamada de serviço que dura 1 hora, 26 minutos e 38 segundos pode ser representada como 01:26:38, dependendo do formato de horário atual, conforme especificado na caixa de diálogo Propriedades do Fluxo.
- **Registro de Data e Hora** Valores que incluem um componente de data e hora, por exemplo, 2007-09-26 09:04:00, mais uma vez dependendo dos formatos de data e hora atuais na caixa de diálogo Propriedades do Fluxo. Observe que os valores de registro de data e hora podem precisar ser incluídos entre aspas duplas para garantir que eles sejam interpretados como um valor único ao invés valores de data e hora separados. (Isto se aplica, por exemplo, ao inserir valores em um nó Entrada do Usuário).
- **Lista** Introduzido no SPSS Modeler versão 17, junto dos novos níveis de medição Geoespacial e de Coleção, um campo de armazenamento Lista contém diversos valores para um único registro. Há versões de lista de todos os outros tipos de armazenamento.

Tabela 8. Ícones de tipo de armazenamento de Lista

Ícone	Tipo de armazenamento
	Lista de sequência de caracteres
	Lista de número inteiro
	Lista de números reais
	Lista de tempo
	Lista de data
	Lista de registro de data e hora
	Lista com uma profundidade maior que zero

Além disso, para uso com o nível de medição Coleção, há versões de lista dos seguintes níveis de medição.

Tabela 9. Ícones do nível de medição de lista

Ícone	Nível de medição
	Lista de contínuo

Tabela 9. Ícones do nível de medição de lista (continuação)

Ícone	Nível de medição
[🏷️]	Lista de categórico
[🚩]	Lista de flags
[🏠]	Lista de nominal
[📊]	Lista de ordinal

As listas podem ser importadas no SPSS Modeler em um dos três nós de origem (Analytic Server, Geoespacial ou Arquivo de Variável) ou criadas dentro de seus fluxos através do uso dos nós de operação de campo Derivar ou Preenchimento.

Para obter mais informações sobre as Listas e a interação delas com os níveis de medição Coleção e Geoespacial, consulte “Armazenamento de lista e níveis de medição associados” na página 11

Conversões de armazenamento. É possível converter armazenamento de um campo utilizando uma variedade de funções de conversão, como `to_string` e `to_integer`, em um nó Preenchimento. Consulte o tópico “Conversão de Armazenamento Usando o Nó Preenchimento” na página 155 para obter mais informações. Observe que as funções de conversão (e quaisquer outras funções que requerem um tipo específico de entrada, como um valor de data e hora) dependem dos formatos atuais especificados na caixa de diálogo Propriedades do Fluxo. Por exemplo, se desejar converter um campo de sequência de caracteres com valores *Jan 2003, Feb 2003*, (e assim por diante), no armazenamento de data, selecione **MON YYYY** como o formato de data padrão para o fluxo. As funções de conversão também estão disponíveis a partir do nó Derivar, para conversão temporária durante um cálculo de derivar. Também é possível utilizar o nó Derivar para executar outras manipulações, como recodificar campos de sequência de caracteres com valores categóricos. Consulte o tópico “Recodificando Valores com o Nó Derivar” na página 154 para obter mais informações.

Leitura de dados combinados. Observe que ao ler campos com armazenamento numérico (número inteiro, real, hora, registro de data e hora ou data), quaisquer valores não numéricos são configurados para nulos ou omissos do sistema. Isso ocorre porque, ao contrário de alguns aplicativos, o IBM SPSS Modeler não permite tipos de armazenamento combinados em um campo. Para evitar isso, quaisquer campos com dados combinados devem ser lidos como sequências de caracteres ao alterar o tipo de armazenamento no nó de origem ou nos aplicativos externos, conforme necessário.

Nota: nós de Entrada do Usuário gerados já podem conter as informações de armazenamento reunidas a partir do nó de origem, se instanciado. Um nó não instanciado não contém informações de tipo de armazenamento ou de uso.

Regras para Especificar Valores

Para campos simbólicos, você deve deixar espaços entre os diversos valores, como:

HIGH MEDIUM LOW

Para campos numéricos, é possível inserir diversos valores da mesma maneira (listados com espaços entre eles):

10 12 14 16 18 20

Ou é possível especificar a mesma série de números ao configurar seus limites (10, 20) e os passos entre eles (2). Utilizando esse método, você digita:

10,20,2

Esses dois métodos podem ser combinados ao integrá-los, como:

1 5 7 10,20,2 21 23

Esta entrada produzirá os seguintes valores:

1 5 7 10 12 14 16 18 20 21 23

Nó Geração de Simulação

O nó Geração de Simulação fornece uma maneira fácil de gerar dados simulados, seja sem dados históricos utilizando distribuições estatísticas especificadas pelo usuário ou automaticamente utilizando as distribuições obtidas da execução de um nó Ajuste de Simulação nos dados históricos existentes. A geração de dados simulados é útil quando desejar avaliar o resultado de um modelo preditivo na presença de incerteza nas entradas de modelo.

Criando dados sem dados históricos

O nó Geração de Simulação está disponível a partir da paleta Origens e pode ser incluído diretamente na tela de fluxo.

1. Clique na guia **Origens** na paleta do nó.
2. Arraste e solte ou dê um clique duplo para incluir o nó Geração de Simulação na tela de fluxo.
3. Dê um clique duplo para abrir sua caixa de diálogo e especificar campos, tipos de armazenamento, distribuições estatísticas e parâmetros de distribuição.

Nota: os nós Geração de Simulação que são selecionados a partir da paleta Origens estarão completamente em branco, sem campos e informações de distribuição. Isso permite plenamente criar dados simulados sem dados históricos.

Gerando dados simulados usando dados históricos existentes

Um nó Geração de Simulação também pode ser criado ao executar um nó terminal Ajuste de Simulação:

1. Clique com o botão direito no nó Ajuste de Simulação e escolha **Executar** no menu.
2. O nó Geração de Simulação aparece na tela de fluxo com uma ligação de atualização para o nó Ajuste de Simulação.
3. Quando gerado, o nó Geração de Simulação herda todos os campos, tipos de armazenamento e informações de distribuição de estatísticas do nó Ajuste de Simulação.

Definindo uma ligação de atualização para um nó de ajuste de simulação

É possível criar uma ligação entre um nó Geração de Simulação e um nó Ajuste de Simulação. Isso será útil se desejar atualizar um ou mais campos com as informações da melhor distribuição de ajuste, determinado pelo ajuste nos dados históricos.

1. Clique com o botão direito no nó Geração de Simulação.
2. No menu, selecione **Definir Ligação de Atualização**. O cursor muda para o cursor de Ligação.
3. Clique em outro nó. Se este nó for um nó Ajuste de Simulação, uma ligação será feita. Se este nó não for um nó Ajuste de Simulação, nenhuma ligação será feita e o cursor voltará para a sua forma normal.

Se os campos no nó Ajuste de Simulação forem diferentes daqueles do nó Geração de Simulação, uma mensagem será exibida informando que há uma diferença.

Quando o nó Ajuste de Simulação é utilizado para atualizar o nó Geração de Simulação vinculada, o resultado depende se os mesmos campos estão presentes nos dois nós e se os campos estão desbloqueados no nó Geração de Simulação. Os resultados da atualização de um nó Ajuste de Simulação são mostrados na tabela a seguir.

Tabela 10. Resultados da atualização de um nó Ajuste de Simulação

Campo no nó Geração de Simulação	Campo na Simulação Nó Ajuste	
	Presente	Omisso
Presente e desbloqueado.	O campo é sobrescrito.	O campo é excluído.
Omisso.	O campo é incluído.	Nenhuma mudança.
Presente e bloqueado.	A distribuição do campo não é sobrescrita. As informações na caixa de diálogo Detalhes do Ajuste e as correlações são atualizadas.	O campo não é sobrescrito. As correlações são configuradas para zero.
A caixa de seleção Não limpar Mín./Máx. ao reajustar é marcada.	O campo é sobrescrito, além dos valores na coluna Mín., Máx.	
A caixa de seleção Não recalcular correlações ao reajustar é marcada.	Se o campo estiver desbloqueado, ele será sobrescrito.	As correlações não são sobrescritas.

Removendo uma ligação de atualização para um nó de ajuste de simulação

É possível remover uma ligação entre um nó Geração de Simulação e um nó Ajuste de Simulação ao executar os seguintes passos:

1. Clique com o botão direito no nó Geração de Simulação.
2. No menu, selecione **Remover Ligação de Atualização**. A ligação é removida.

Configurando Opções para o Nó Geração de Simulação

É possível utilizar as opções na guia Dados da caixa de diálogo do nó Geração de Simulação para fazer o seguinte:

- Visualizar, especificar e editar as informações de distribuição estatística para os campos.
- Visualizar, especificar e editar as correlações entre os campos.
- Especificar o número de iterações e casos para simular.

Selecionar um item. Permite alternar entre as três visualizações do nó Geração de Simulação: Campos Simulados, Correlações e Opções Avançadas.

Visualização Campos Simulados

Se o nó Geração de Simulação tiver sido gerado ou atualizado a partir de um nó de Ajuste de Simulação usando dados históricos, na visualização Campos Simulados, é possível visualizar e editar as informações de distribuição estatísticas para cada campo. As informações a seguir sobre cada campo são copiadas para a guia **Tipos** do nó Geração de Simulação a partir do nó Ajuste de Simulação:

- Nível de medição
- Valores
- Ausente
- Verificar
- Função

Se você não tiver dados históricos, será possível definir campos e especificar suas distribuições ao selecionar um tipo de armazenamento, selecionar um tipo de distribuição e inserir os parâmetros

necessários. Gerar dados desta forma significa que as informações sobre o nível de medição de cada campo não estarão disponíveis até que os dados sejam instanciados, por exemplo, na guia **Tipos** ou em um nó Tipo.

A visualização Campos Simulados contém várias ferramentas que podem ser usadas para executar as tarefas a seguir:

- Incluir e remover campos.
- Alterar a ordem dos campos na exibição.
- Especificar um tipo de armazenamento para cada campo.
- Especifique uma distribuição estatística para cada campo.
- Especifique os valores de parâmetro para a distribuição estatística de cada campo.

Campos Simulados. Essa tabela conterà uma linha vazia se o nó Geração de Simulação foi incluído na tela de fluxo a partir da paleta Origens. Quando essa linha é editada, uma nova linha vazia é incluída na parte inferior da tabela. Se o nó Geração de Simulação tiver sido criado a partir de um nó Ajuste de Simulação, essa tabela conterà uma linha para cada campo dos dados históricos. Linhas extras podem ser incluídas na tabela ao clicar no ícone **Incluir novo campo**.

A tabela Campos Simulados composta das seguintes colunas:

- **Campo.** Contém os nomes dos campos. Os nomes dos campos podem ser editados ao digitar nas células.
- **Armazenamento.** As células nessa coluna contêm uma lista suspensa de tipos de armazenamento. Os tipos de armazenamento disponíveis são **Sequência de Caracteres**, **Número Inteiro**, **Reais Hora**, **Data** e **Registro de Data e Hora**. A opção de tipo de armazenamento determina quais distribuições estão disponíveis na coluna Distribuição. Se o nó Geração de Simulação tiver sido criado a partir de um nó Ajuste de Simulação, o tipo de armazenamento será copiado a partir do nó Ajuste de Simulação.

Nota: Para campos com tipos de armazenamento de data/hora, deve-se especificar os parâmetros de distribuição como números inteiros. Por exemplo, para especificar uma data média de 1º de janeiro de 1970, utilize o número inteiro 0. O número inteiro assinado representa o número de segundos desde (ou antes) a meia-noite de 1º de janeiro de 1970.

- **Status.** Os ícones na Coluna Status indicam o status de ajuste para cada campo.



Nenhuma distribuição foi especificada para o campo ou um ou mais parâmetros de distribuição estão omissos. Para executar a simulação, deve-se especificar uma distribuição para esse campo e inserir valores válidos para os parâmetros.



O campo é configurado para a distribuição de ajuste mais próxima.

Nota: Esse ícone poderá ser exibido sempre se o nó Geração de Simulação for criado a partir de um nó Filtro somente.



A distribuição de ajuste mais próxima foi substituída por uma distribuição alternativa a partir da caixa de subdiálogo Detalhes do Ajuste. Consulte o tópico “Detalhes de Ajuste” na página 57 para obter mais informações.



A distribuição foi especificada ou editada manualmente, e pode incluir um parâmetro especificado em mais de um nível.

- **Bloqueado.** Bloquear um campo simulado, ao marcar a caixa de seleção na coluna com o ícone de bloqueio, exclui o campo da atualização automática por um nó Ajuste de Simulação vinculado. Isso é mais útil ao especificar manualmente uma distribuição e você desejar assegurar que isso não será afetado pelo ajuste de distribuição automática quando um nó Ajuste de Simulação vinculado é executado.
- **Distribuição.** As células nesta coluna contêm uma lista suspensa de distribuições estatísticas. A escolha do tipo de armazenamento determina quais distribuições estão disponíveis nesta coluna para um determinado campo. Consulte o tópico “Distribuições” na página 60 para obter mais informações.

Nota: Não é possível especificar a distribuição Fixa para cada campo. Se desejar que cada campo em seus dados gerados seja fixo, será possível utilizar um nó Entrada do Usuário seguido por um nó Balanceamento.

- **Parâmetros.** Os parâmetros de distribuição associados a cada distribuição ajustada são exibidos nessa coluna. Diversos valores de um parâmetro são separados por vírgula. Especificar diversos valores para um parâmetro gera diversas iterações para a simulação. Consulte o tópico “Iterações” na página 60 para obter mais informações. Se os parâmetros estiverem omissos, isso é refletido no ícone exibido na coluna Status. Para especificar valores para os parâmetros, clique nessa coluna na linha correspondente ao campo de interesse e escolha **Especificar** na lista. Isso abre a subcaixa de diálogo Especificar Parâmetros. Consulte o tópico “Especificar parâmetros” na página 58 para obter mais informações. Essa coluna estará desativada se Empírico for escolhido na coluna Distribuição.
- **Mín., Máx.** Nessa coluna, para algumas distribuições, é possível especificar um valor mínimo, um valor máximo, ou ambos, para os dados simulados. Os dados simulados que forem menores que o valor mínimo e maiores que o valor máximo serão rejeitados, mesmo se forem válidos para a distribuição especificada. Para especificar valores mínimo e máximo, clique nessa coluna na linha correspondente ao campo de interesse e escolha **Especificar** na lista. Isso abre a subcaixa de diálogo Especificar Parâmetros. Consulte o tópico “Especificar parâmetros” na página 58 para obter mais informações. Essa coluna estará desativada se Empírico for escolhido na coluna Distribuição.

Usar Ajuste Mais Próximo. Ativado somente se o nó Geração de Simulação tiver sido criado automaticamente a partir de um nó Ajuste de Simulação usando dados históricos, e uma linha única na tabela Campos Simulados é selecionada. Substitui as informações para o campo na linha selecionada pelas informações da melhor distribuição de ajuste para o campo. Se as informações na linha selecionada tiverem sido editadas, pressionar esse botão reconfigurará as informações de volta para a distribuição de melhor ajuste determinada a partir do nó Ajuste de Simulação.

Detalhes do Ajuste. Ativado somente se o nó Geração de Simulação tiver sido criado automaticamente a partir de um nó Ajuste de Simulação. Isso abre a subcaixa de diálogo Detalhes do Ajuste. Consulte o tópico “Detalhes de Ajuste” na página 57 para obter mais informações.

Várias tarefas úteis podem ser realizadas utilizando os ícones à direita da visualização Campos Simulados. Esses ícones são descritos na tabela a seguir.

Tabela 11. Ícones na visualização Campos Simulados.

Ícone	Dicas de ferramenta	Descrição
	Editar parâmetros de distribuição	Ativado somente quando uma linha única na tabela Campos Simulados estiver selecionada. Abre a subcaixa de diálogo Especificar Parâmetros para a linha selecionada. Consulte o tópico “Especificação de parâmetros” na página 58 para obter mais informações.
	Incluir novo campo	Ativado somente quando uma linha única na tabela Campos Simulados estiver selecionada. Inclui uma nova linha vazia na parte inferior da tabela Campos Simulados.
	Criar várias cópias	Ativado somente quando uma linha única na tabela Campos Simulados estiver selecionada. Isso abre a subcaixa de diálogo Clonar Campo. Consulte o tópico “Clonar Campo” na página 56 para obter mais informações.
	Excluir campo selecionado	Exclui a linha selecionada da tabela Campos Simulados.
	Mover para o topo	Ativado somente se a linha selecionada ainda não for a linha superior da tabela Campos Simulados. Move a linha selecionada para a parte superior da tabela Campos Simulados. Essa ação afeta a ordem dos campos nos dados simulados.
	Mover para cima	Ativado somente se a linha selecionada não for a linha superior da tabela Campos Simulados. Move a linha selecionada para uma posição acima na tabela Campos Simulados. Essa ação afeta a ordem dos campos nos dados simulados.
	Mover para baixo	Ativado somente se a linha selecionada não for a linha inferior da tabela Campos Simulados. Move a linha selecionada para uma posição abaixo na tabela Campos Simulados. Essa ação afeta a ordem dos campos nos dados simulados.
	Mover para o fundo	Ativado somente se a linha selecionada ainda não for a linha inferior da tabela Campos Simulados. Move a linha selecionada para a parte inferior da tabela Campos Simulados. Essa ação afeta a ordem dos campos nos dados simulados.

Não limpar Mín. e Máx. ao reajustar. Quando selecionada, os valores mínimo e máximo não serão sobrescritos quando as distribuições forem atualizadas pela execução de um nó Ajuste de Simulação conectado.

Visualização Correlações

Os campos de entrada para modelos preditivos geralmente são conhecidos como sendo correlacionados, por exemplo, altura e peso. As correlações entre campos que serão simulados devem ser levadas em conta a fim de assegurar que os valores simulados preservem essas correlações.

Se o nó Geração de Simulação tiver sido gerado ou atualizado a partir de um nó Ajuste de Simulação usando dados históricos, na visualização Correlações, é possível visualizar e editar as correlações calculadas entre pares de campos. Se você não tiver dados históricos, será possível especificar as correlações manualmente com base no seu conhecimento de como os campos são correlacionados.

Nota: Antes que quaisquer dados sejam gerados, é verificado automaticamente se a matriz de correlações é semidefinida positiva, e se pode, portanto, ser invertida. Uma matriz poderá ser invertida se suas colunas forem linearmente independentes. Se a matriz de correlações não puder ser invertida, ela será ajustada automaticamente para torná-la invertível.

É possível optar por exibir as correlações em um formato de matriz ou de lista.

Matriz de correlações. Exibe as correlações entre pares de campos em uma matriz. Os nomes de campo são listados, em ordem alfabética, para a esquerda e ao longo da parte superior da matriz. Apenas as células abaixo da diagonal podem ser editadas e um valor entre -1.000 e 1.000, inclusive, deve ser inserido. A célula acima da diagonal é atualizada quando o foco é alterado para fora de sua célula espelhada abaixo da diagonal; em seguida, as duas células exibem o mesmo valor. As células diagonais estão sempre desativadas e sempre têm uma correlação de 1.000. O valor padrão para todas as outras células é 0.000. Um valor 0.000 especifica que não há nenhuma correlação entre o par associado de campos. Apenas os campos contínuos e ordinais são incluídos na matriz. Os campos nominais, categóricos, de flag e campos que são designados à distribuição Fixa não são mostrados na tabela.

Lista de Correlações. Exibe as correlações entre pares de campos em uma tabela. Cada linha da tabela mostra a correlação entre um par de campos. Linhas não podem ser incluídas ou excluídas. As colunas com os títulos Campo 1 e Campo 2 contêm os nomes de campo que não podem ser editados. A coluna Correlação contém as correlações que podem ser editadas; um valor entre -1.000 e 1.000, inclusive, devem ser inseridos. O valor padrão para todas as células é 0.000. Apenas os campos contínuos e ordinais são incluídos na lista. Os campos nominais, categóricos, de flag e campos que são designados à distribuição Fixa não são mostrados na lista.

Reconfigurar correlações. Abre a caixa de diálogo Reconfigurar Correlações. Se os dados históricos estiverem disponíveis, será possível escolher uma das três opções:

- **Ajustado.** Substitui as correlações atuais por aquelas calculadas utilizando os dados históricos.
- **Zeros.** Substitui as correlações atuais por zeros.
- **Cancelar.** Fecha a caixa de diálogo. As correlações são inalteradas.

Se dados históricos não estiverem disponíveis, mas você tiver feito mudanças nas correlações, será possível escolher substituir as correlações atuais por zeros ou cancelar.

Mostrar Como. Selecione **Tabela** para exibir as correlações como uma matriz. Selecione **Lista** para exibir as correlações como uma lista.

Não recalcular correlações ao reajustar. Selecione esta opção se desejar especificar manualmente as correlações e impedir que elas sejam sobrescritas quando ajustar automaticamente as distribuições utilizando um nó Ajuste de Simulação e dados históricos.

Use a tabela de contingência de multiponto ajustada para entradas com distribuição categórica. Por padrão, todos os campos com uma distribuição categórica são incluídos em uma tabela de contingência (ou em uma tabela de contingência multiponto, dependendo do número de campos com uma distribuição categórica). A tabela de contingência é construída, assim como as correlações, quando um nó Ajuste de Simulação é executado. A tabela de contingência não pode ser visualizada. Quando essa opção é selecionada, os campos com uma distribuição categórica são simulados usando as porcentagens reais da tabela de contingência. Ou seja, quaisquer associações entre os campos nominais são recriadas nos novos dados simulados. Quando esta opção estiver desmarcada, os campos com distribuições categóricas são simulados usando as porcentagens esperadas da tabela de contingência. Se modificar um campo, o campo será removido da tabela de contingência.

Visualização Opções Avançadas

Número de casos para simular. Exibe as opções para especificar o número de casos a serem simulados e como quaisquer iterações serão nomeadas.

- **Número máximo de casos.** Isso especifica o número máximo de casos de dados simulados e os valores de destino associados a serem gerados. O valor padrão é 100,00, o valor mínimo é 1000 e o valor máximo é 2.147.483.647.
- **Iterações.** Esse número é calculado automaticamente e não pode ser editado. Uma iteração é criada automaticamente cada vez que um parâmetro de distribuição tiver diversos valores especificados.
- **Total de linhas.** Ativada somente quando o número de iterações for maior que 1. O número é calculado automaticamente utilizando a equação mostrada e não pode ser editado.
- **Criar campo de iteração.** Ativada somente quando o número de iterações é maior que 1. Quando selecionada, o campo **Nome** é ativado. Consulte o tópico “Iterações” na página 60 para obter mais informações.
- **Nome.** Ativada somente quando a caixa de seleção **Criar campo de iteração** estiver selecionada e o número de iterações for maior que 1. Edite o nome do campo de iteração digitando neste campo de texto. Consulte o tópico “Iterações” na página 60 para obter mais informações.

Semente aleatória. Configurar uma semente aleatória permite replicar sua simulação.

- **Replicar resultados.** Quando selecionada, o botão **Gerar** e o campo **Semente aleatória** são ativados.
- **Semente aleatória.** Ativada somente quando a caixa de seleção **Replicar resultados** é marcada. Neste campo, é possível especificar um número inteiro a ser utilizado como a semente aleatória. O valor padrão é 629111597.
- **Gerar.** Ativada somente quando a caixa de seleção **Replicar resultados** é marcada. Cria um pseudonúmero inteiro aleatório entre 1 e 999999999, inclusive, no campo **Semente aleatória**.

Clonar Campo

Na caixa de diálogo Clonar Campo, é possível especificar quantas cópias do campo selecionado serão criadas e como cada cópia será nomeada. É útil ter diversas cópias de campos quando investigar efeitos compostos, por exemplo, taxas de juros ou de crescimento durante um número de períodos de tempo sucessivos.

A barra de título da caixa de diálogo contém o nome do campo selecionado.

Número de cópias a fazer. Contém o número de cópias do campo a serem criadas. Clique nas setas para selecionar o número de cópias a serem criadas. O número mínimo de cópias é 1 e o máximo é 512. O número de cópias é inicialmente configurado para 10.

Caractere(s) de sufixo da cópia. Contém os caracteres que são incluídos no término do nome do campo para cada cópia. Esses caracteres separam o nome do campo do número de cópias. Os caracteres de sufixo podem ser editados ao digitar neste campo. Esse campo pode ser deixado vazio; nesse caso, não haverá caracteres entre o nome do campo e o número de cópias. O caractere padrão é um sublinhado.

Número de cópias inicial. Contém o número de sufixo para a primeira cópia. Clique nas setas para selecionar o número de cópias inicial. O número de cópias inicial mínimo é 1 e o máximo é 1000. O número de cópias inicial padrão é 1.

Passo do número de cópias. Contém o incremento para os números de sufixo. Clique nas setas para selecionar o incremento. O incremento mínimo é 1 e o máximo é 255. O incremento é inicialmente configurado para 1.

Campos. Contém uma visualização dos nomes de campo para as cópias, que será atualizada se quaisquer campos da caixa de diálogo Clonar Campo forem editados. Esse texto é gerado automaticamente e não pode ser editado.

OK. Gera todas as cópias, conforme especificado na caixa de diálogo. As cópias são incluídas na tabela Campos Simulados na caixa de diálogo do nó Geração de Simulação, logo abaixo da linha que contém o campo que foi copiado.

Cancelar. Fecha a caixa de diálogo. Todas as mudanças que tiverem sido feitas são descartadas.

Detalhes de Ajuste

A caixa de diálogo Detalhes do Ajuste estará disponível somente se o nó Geração de Simulação tiver sido criado ou atualizado ao executar um nó Ajuste de Simulação. Ela exibe os resultados do ajuste de distribuição automática para o campo selecionado. As distribuições são ordenadas pela Qualidade do ajuste, com a distribuição de ajuste mais próxima listada primeiro. Nessa caixa de diálogo, é possível executar as tarefas a seguir:

- Examinar as distribuições ajustadas para os dados históricos.
- Selecionar uma das distribuições ajustadas.

Campo. Contém o nome do campo selecionado. Esse texto não pode ser editado.

Tratar como (Medida). Exibe o tipo de medição do campo selecionado. Isso é obtido a partir da tabela Campos Simulados na caixa de diálogo de nó Geração de Simulação. O tipo de medição pode ser alterado ao clicar na seta e selecionar um tipo de medição na lista suspensa. Há três opções: **Contínuo**, **Nominal** e **Ordinal**.

Distribuições. A tabela Distribuições mostra todas as distribuições que são apropriadas para o tipo de medição. As distribuições que tiverem sido ajustadas para os dados históricos são ordenadas pela Qualidade do ajuste, do melhor para o pior ajuste. A Qualidade do ajuste é determinada pela estatística de ajuste escolhida no nó Ajuste de Simulação. As distribuições que não forem ajustadas nos dados históricos são listadas na tabela em ordem alfabética abaixo das distribuições que forem ajustadas.

A tabela Distribuição contém as colunas a seguir:

- **Usar.** O botão de opções selecionado indica qual distribuição está atualmente escolhida para o campo. É possível substituir a distribuição de ajuste mais próxima ao selecionar o botão de opções para a distribuição desejada na coluna Usar. A seleção de um botão de opções na coluna Usar também exibe um gráfico da distribuição sobreposta em um histograma (ou gráfico de barras) dos dados históricos para o campo selecionado. Apenas uma distribuição pode ser selecionada por vez.
- **Distribuição.** Contém o nome da distribuição. Essa coluna não pode ser editada.
- **Estatísticas de Ajuste.** Contém as estatísticas de ajuste calculadas para a distribuição. Essa coluna não pode ser editada. O conteúdo da célula depende do tipo de medição do campo:
 - **Contínuo.** Contém os resultados dos testes de Anderson-Darling e de Kolmogorov-Smirnoff. Os valores p associados aos testes também são mostrados. A estatística de ajuste escolhida como os critérios de Qualidade do ajuste no nó Ajuste de Simulação é mostrada primeiro e utilizada para ordenar as distribuições. As estatísticas de Anderson-Darling são exibidas como $A=aval$ $P=pval$. As

estatísticas de Kolmogorov-Smirnoff são exibidas como $K=kval$ $P=pval$. Se uma estatística não puder ser calculada, um ponto será exibido ao invés de um número.

- **Nominal e Ordinal.** Contém os resultados do teste qui-quadrado. O valor p associado ao teste também é mostrado. As estatísticas são exibidas como $Chi-Sq=val$ $P=pval$. Se a distribuição não foi ajustada, Não ajustada será exibida. Se a distribuição não pode ser matematicamente ajustada, Não pode ser ajustada será exibida.

Nota: a célula está sempre vazia para a distribuição Empírica.

- **Parâmetros.** Contém os parâmetros de distribuição associados a cada distribuição ajustada. Os parâmetros são exibidos como *parameter_name = parameter_value*, com os parâmetros separados por um espaço único. Para a distribuição categórica, os nomes dos parâmetros são as categorias e os valores dos parâmetros são as probabilidades associadas. Se a distribuição não foi ajustada para os dados históricos, a célula estará vazia. Essa coluna não pode ser editada.

Miniatura de histograma. Mostra um gráfico da distribuição selecionada sobreposta em um histograma dos dados históricos do campo selecionado.

Miniatura de distribuição. Mostra uma explicação e a ilustração da distribuição selecionada.

OK. Fecha a caixa de diálogo e atualiza os valores das colunas Medição, Distribuição, Parâmetros e Mín., Máx. de colunas da tabela Campos Simulados para o campo selecionado com as informações da distribuição selecionada. O ícone na coluna Status também é atualizado para refletir se a distribuição selecionada é a distribuição com o ajuste mais próximo dos dados.

Cancelar. Fecha a caixa de diálogo. Todas as mudanças que tiverem sido feitas são descartadas.

Especificar parâmetros

Na caixa de diálogo Especificar Parâmetros, é possível especificar manualmente os valores de parâmetro para a distribuição do campo selecionado. Também é possível escolher uma distribuição diferente para o campo selecionado.

A caixa de diálogo Especificar Parâmetros pode ser aberta de três maneiras:

- Dê um clique duplo em um nome do campo na tabela Campos Simulados na caixa de diálogo do nó Geração de Simulação.
- Clique na coluna Parâmetros ou Mín., Máx. de colunas da tabela Campos Simulados e escolha **Especificar** na lista.
- Na tabela Campos Simulados, selecione uma linha e, em seguida, clique no ícone **Editar parâmetros de distribuição**.

Campo. Contém o nome do campo selecionado. Esse texto não pode ser editado.

Distribuição. Contém a distribuição do campo selecionado. Isso é obtido a partir da tabela Campos Simulados. A distribuição pode ser alterada ao clicar na seta e selecionar uma distribuição na lista suspensa. As distribuições disponíveis dependem do tipo de armazenamento do campo selecionado.

Lados. Essa opção está disponível somente para quando a distribuição dice é selecionada no campo **Distribuição**. Clique nas setas para selecionar o número de lados ou categorias nos quais dividir o campo. O número mínimo de lados é dois e o máximo é 20. O número de lados é inicialmente configurado para 6.

Parâmetros de distribuição. A tabela Parâmetros de distribuição contém uma linha para cada parâmetro da distribuição escolhida.

Nota: A distribuição utiliza um parâmetro de taxa, com um parâmetro de forma $\alpha = k$ e um parâmetro de escala inverso $\beta = 1/\theta$.

A tabela contém duas colunas:

- **Parâmetro.** Contém os nomes dos parâmetros. Essa coluna não pode ser editada.
- **Valor(es).** Contém os valores dos parâmetros. Se o nó Geração de Simulação tiver sido criado ou atualizado a partir de um nó Ajuste de Simulação, as células nessa coluna conterão os valores de parâmetro que foram determinados ao ajustar a distribuição para os dados históricos. Se o nó Geração de Simulação tiver sido incluído na tela de fluxo a partir da paleta de nós Origem, as células nesta coluna estarão vazias. Os valores podem ser editados ao digitar nas células. Consulte o tópico “Distribuições” na página 60 para obter mais informações sobre os parâmetros necessários para cada distribuição e os valores de parâmetro aceitáveis.

Diversos valores para um parâmetro devem ser separados por vírgulas. A especificação de diversos valores para um parâmetro definirá diversas iterações da simulação. É possível especificar diversos valores somente para um parâmetro.

Nota: Para campos com tipos de armazenamento de data/hora, deve-se especificar os parâmetros de distribuição como números inteiros. Por exemplo, para especificar uma data média de 1º de janeiro de 1970, utilize o número inteiro 0.

Nota: Quando a distribuição dice é selecionada, a tabela de Parâmetros de distribuição é um pouco diferente. A tabela contém uma linha para cada lado (ou categoria). A tabela contém uma coluna Valor e uma coluna Probabilidade. A coluna Valor contém um rótulo para cada categoria. Os valores padrão para os rótulos são os números inteiros 1-N, em que N é o número de lados. Os rótulos podem ser editados ao digitar nas células. Qualquer valor pode ser inserido nas células. Se desejar utilizar um valor que não é um número, o tipo de armazenamento do campo de dados deverá ser alterado para sequência de caracteres se o tipo de armazenamento ainda não estiver configurado como sequência de caracteres. A coluna Probabilidade contém a probabilidade de cada categoria. As probabilidades não podem ser editadas e são calculadas como $1/N$.

Visualização. Exibe um gráfico de amostra da distribuição com base nos parâmetros especificados. Se dois ou mais valores tiverem sido especificados para um parâmetro, gráficos de amostra para cada valor do parâmetro serão mostrados. Se dados históricos estiverem disponíveis para o campo selecionado, o gráfico da distribuição será sobreposto em um histograma dos dados históricos.

Configurações Opcionais. Use essas opções para especificar um valor mínimo, um valor máximo ou ambos para os dados simulados. Os dados simulados que forem menores que o valor mínimo e maiores que o valor máximo serão rejeitados, mesmo se forem válidos para a distribuição especificada.

- **Especificar mínimo.** Selecione para ativar o campo **Rejeitar valores abaixo**. A caixa de seleção será desativada se a distribuição Empírica for selecionada.
- **Rejeitar valores abaixo.** Disponível somente se **Especificar Mínimo** for selecionada. Insira um valor mínimo para os dados simulados. Quaisquer valores simulados menores que este valor serão rejeitados.
- **Especificar máximo.** Selecione para ativar o campo **Rejeitar valores acima**. A caixa de seleção será desativada se a distribuição Empírica for selecionada.
- **Rejeitar valores acima.** Disponível somente se **Especificar Máximo** for selecionada. Insira um valor máximo para os dados simulados. Quaisquer valores simulados maiores que este valor serão rejeitados.

OK. Fecha a caixa de diálogo e atualiza os valores das colunas Distribuição, Parâmetros e Mín., Máx. de colunas da tabela Campos Simulados para o campo selecionado. O ícone na coluna Status também é atualizado para refletir a distribuição selecionada.

Cancelar. Fecha a caixa de diálogo. Todas as mudanças que tiverem sido feitas são descartadas.

Iterações

Se tiver especificado mais de um valor para um campo fixo ou um parâmetro de distribuição, um conjunto independente de casos simulados - efetivamente, uma simulação separada - será gerado para cada valor especificado. Isso permite investigar o efeito de variar o campo ou o parâmetro. Cada conjunto de casos simulados é referido como uma *iteração*. Nos dados simulados, as iterações são empilhadas.

Se a caixa de seleção **Criar campo iteração** na visualização Opções Avançadas do diálogo do nó Geração de Simulação for marcada, um campo de iteração será incluído nos dados simulados como um campo nominal com armazenamento numérico. O nome deste campo pode ser editado ao digitar no campo **Nome** na visualização Opções Avançadas. Esse campo contém um rótulo que indica à qual iteração pertence cada caso simulado. O formato dos rótulos depende do tipo de iteração:

- **Iterando um campo fixo.** O rótulo é o nome do campo, seguido por um sinal de igual e pelo valor do campo para essa iteração, que é
field_name = field_value
- **Iterando um parâmetro de distribuição.** O rótulo é o nome do campo, seguido por dois pontos, pelo nome do parâmetro iterado, por um sinal de igual e pelo valor do parâmetro para essa iteração, que é
field_name:parameter_name = parameter_value
- **Iterando um parâmetro de distribuição para uma distribuição categórica ou de intervalo.** O rótulo é o nome do campo, seguido por dois pontos, pelo "Iteration" e pelo número de iteração, que é
field_name: Iteration iteration_number

Distribuições

É possível especificar manualmente a distribuição de probabilidade para qualquer campo ao abrir a caixa de diálogo Especificar Parâmetros para esse campo, selecionar a distribuição desejada na lista

Distribuição e inserir os parâmetros de distribuição na tabela **Parâmetros de Distribuição**. A seguir há algumas notas sobre distribuições específicas:

- **Categórico.** A distribuição categórica descreve um campo de entrada que possui um número fixo de valores numéricos, referidos como categorias. Cada categoria possui uma probabilidade associada, de forma que a soma das probabilidades de todas as categorias seja igual a um.

Nota: Se especificar probabilidades para as categorias que não somam 1, você receberá um aviso.

- **Binomial Negativa - Falhas.** Descreve a distribuição do número de falhas em uma sequência de avaliações antes que um número especificado de sucessos seja observado. O parâmetro *Limite* é o número especificado de sucessos e o parâmetro *Probabilidade* é a probabilidade de sucesso em qualquer avaliação específica.
- **Binomial Negativa - Avaliações.** Descreve a distribuição do número de avaliações que são necessárias antes que um número especificado de sucessos seja observado. O parâmetro *Limite* é o número especificado de sucessos e o parâmetro *Probabilidade* é a probabilidade de sucesso em qualquer avaliação específica.
- **Intervalo.** Essa distribuição consiste em um conjunto de intervalos com uma probabilidade designada a cada intervalo, de modo que a soma das probabilidades de todos os intervalos seja igual a 1. Os valores dentro de um intervalo especificado são obtidos a partir de uma distribuição uniforme definida nesse intervalo. Os intervalos são especificados ao inserir um valor mínimo, um valor máximo e uma probabilidade associada.

Por exemplo, você acredita que o custo de uma matéria-prima tem 40% de chances de cair no intervalo de \$10 – \$15 por unidade, e 60% de chances de cair no intervalo de \$15 – \$20 por unidade. Você modela o custo com uma distribuição Intervalo consistindo em dois intervalos [10 – 15] e [15 – 20], configurando a probabilidade associada ao primeiro intervalo para 0,4 e a probabilidade associada ao segundo intervalo para 0,6. Os intervalos não precisam ser contínuos e podem até mesmo ser sobrepostos. Por exemplo, você pode ter especificado os intervalos \$10 - \$15 e \$20 - \$25 ou \$10 - \$15 e \$13 - \$16.

- **Weibull.** O parâmetro *Localização* é um parâmetro de localização opcional, que especifica onde a origem da distribuição está localizada.

A tabela a seguir mostra as distribuições que estão disponíveis para ajuste de distribuição customizado e os valores aceitáveis para os parâmetros. Algumas dessas distribuições estão disponíveis para ajuste customizado para determinados tipos de armazenamento, mesmo se elas não estiverem ajustadas automaticamente para esses tipos de armazenamento pelo nó Ajuste de Simulação.

Tabela 12. Distribuições disponíveis para ajuste customizado

Distribuição	Tipo de armazenamento suportado para ajuste customizado	Parâmetros	Limites de parâmetro	Notas
Bernoulli	Número inteiro, real, data/hora	Probabilidade	$0 \leq \text{Probabilidade} \leq 1$	
Beta	Número inteiro, real, data/hora	Forma 1 Forma 2 Mínimo Máximo	≥ 0 ≥ 0 $< \text{Máximo}$ $> \text{Mínimo}$	O mínimo e o máximo são opcionais.
Binomial	Número inteiro, real, data/hora	Número de avaliações Probabilidade Mínimo Máximo	$\geq n$, número inteiro $0 \leq \text{Probabilidade} \leq 1$ $< \text{Máximo}$ $> \text{Mínimo}$	O número de avaliações deve ser um inteiro. O mínimo e o máximo são opcionais.
Catagórico	Número inteiro, real, data/hora, sequência de caracteres	Nome da categoria (ou rótulo)	$0 \leq \text{Valor} \leq 1$	Valor é a probabilidade da categoria. Os valores devem somar 1, caso contrário, um aviso é gerado.
Dice	Número inteiro, sequência de caracteres	Lados	$2 \leq \text{Lados} \leq 20$	A probabilidade de cada categoria (lado) é calculada como $1/N$, em que N é o número de lados. As probabilidades não podem ser editadas.
Empírico	Número inteiro, real, data/hora			Não é possível editar a distribuição empírica ou selecioná-la como um tipo. A distribuição empírica está disponível apenas quando houver dados históricos.
Exponencial	Número inteiro, real, data/hora	Escala Mínimo Máximo	> 0 $< \text{Máximo}$ $> \text{Mínimo}$	O mínimo e o máximo são opcionais.

Tabela 12. Distribuições disponíveis para ajuste customizado (continuação)

Distribuição	Tipo de armazenamento suportado para ajuste customizado	Parâmetros	Limites de parâmetro	Notas
Corrigido	Número inteiro, real, data/hora, sequência de caracteres	Valor		Não é possível especificar a distribuição Fixa para cada campo. Se desejar que cada campo em seus dados gerados seja fixo, será possível utilizar um nó Entrada do Usuário seguido por um nó Balanceamento.
Gama	Número inteiro, real, data/hora	Forma Escala Mínimo Máximo	≥ 0 ≥ 0 $< Máximo$ $> Mínimo$	O mínimo e o máximo são opcionais. A distribuição utiliza um parâmetro de taxa, com um parâmetro de forma $\alpha = k$ e um parâmetro de escala inverso $\beta = 1/\theta$.
Lognormal	Número inteiro, real, data/hora	Forma 1 Forma 2 Mínimo Máximo	≥ 0 ≥ 0 $< Máximo$ $> Mínimo$	O mínimo e o máximo são opcionais.
Binomial Negativa - Falhas	Número inteiro, real, data/hora	Limite Probabilidade Mínimo Máximo	≥ 0 $0 \leq Probabilidade \leq 1$ $< Máximo$ $> Mínimo$	O mínimo e o máximo são opcionais.
Binomial Negativa - Avaliações	Número inteiro, real, data/hora	Limite Probabilidade Mínimo Máximo	≥ 0 $0 \leq Probabilidade \leq 1$ $< Máximo$ $> Mínimo$	O mínimo e o máximo são opcionais.
Normal	Número inteiro, real, data/hora	Média Desvio padrão Mínimo Máximo	≥ 0 > 0 $< Máximo$ $> Mínimo$	O mínimo e o máximo são opcionais.
Poisson	Número inteiro, real, data/hora	Média Mínimo Máximo	≥ 0 $< Máximo$ $> Mínimo$	O mínimo e o máximo são opcionais.
Amplitude	Número inteiro, real, data/hora	Begin(X) End(X) Probability(X)	$0 \leq Valor \leq 1$	X é o índice de cada categoria. Os valores de probabilidade devem somar 1.
Triangular	Número inteiro, real, data/hora	Modo Mínimo Máximo	$Mínimo \leq Valor \leq Máximo$ $< Máximo$ $> Mínimo$	

Tabela 12. Distribuições disponíveis para ajuste customizado (continuação)

Distribuição	Tipo de armazenamento suportado para ajuste customizado	Parâmetros	Limites de parâmetro	Notas
Uniforme	Número inteiro, real, data/hora	Mínimo Máximo	< <i>Máximo</i> > <i>Mínimo</i>	
Weibull	Número inteiro, real, data/hora	Taxa Escala Localização Mínimo Máximo	> 0 > 0 ≥ 0 < <i>Máximo</i> > <i>Mínimo</i>	Localização, máximo e mínimo são opcionais.

Nó Visualização de Dados

Utilize o nó Visualização de Dados para incluir dados definidos em uma visualização de dados analíticos do IBM SPSS Collaboration and Deployment Services em seu fluxo. Uma visualização de dados analíticos define uma estrutura para acessar dados que descrevem as entidades usadas em modelos preditivos e regras de negócios. A visualização associa a estrutura de dados com as origens de dados físicos para a análise.

A análise preditiva requer dados organizados em tabelas com cada linha correspondendo a uma entidade para a qual predições são feitas. Cada coluna em uma tabela representa um atributo mensurável da entidade. Alguns atributos podem ser derivados agregando sobre os valores para outro atributo. Por exemplo, as linhas em uma tabela podem representar clientes com as colunas correspondendo ao nome do cliente, sexo, CEP e número de vezes que o cliente fez uma compra acima de \$500 no ano passado. A última coluna é derivada do histórico de pedido do cliente, que é normalmente armazenado em uma ou mais tabelas relacionadas.

O processo de análise preditiva envolve o uso de diferentes conjuntos de dados por todo o ciclo de vida de um modelo. Durante o desenvolvimento inicial de um modelo preditivo, use o histórico de dados que muitas vezes contém resultados conhecidos para o evento que estiver sendo previsto. Para avaliar a efetividade e precisão do modelo, valide um modelo candidato com relação a diferentes dados. Depois de validar o modelo, implemente-o no uso de produção para gerar pontuações para diversas entidades em um processo em lote ou para entidades únicas em um processo em tempo real. Se você combinar o modelo com regras de negócios em um processo de gerenciamento de decisão, você usará dados simulados para validar os resultados da combinação. Entretanto, embora os dados usados difiram nos estágios do processo de desenvolvimento de modelo, cada conjunto de dados deve fornecer o mesmo conjunto de atributos para o modelo. O conjunto de atributos permanece constante; o que muda são os registros de dados que estão sendo analisados.

Uma visualização de dados analíticos consiste nos seguintes componentes que tratam das necessidades especializadas da análise preditiva:

- Um esquema de visualização de dados ou modelo de dados, que define uma interface lógica para acessar dados como um conjunto de atributos organizado em tabelas relacionadas. Os atributos no modelo podem ser derivados de outros atributos.
- Um ou mais planos de acesso a dados que fornecem os atributos de modelo de dados com valores físicos. Você controla os dados disponíveis para o modelo de dados especificando qual plano de acesso a dados permanece ativo para um aplicativo em particular.

Importante: Para utilizar o nó Visualização de Dados, um IBM SPSS Collaboration and Deployment Services Repository deverá primeiramente estar instalado e configurado em seu site. A visualização de dados analíticos referenciada pelo nó geralmente é criada e armazenada no repositório usando o IBM SPSS Collaboration and Deployment Services Deployment Manager.

Configurando Opções para o Nó Visualização de Dados

Use as opções na guia **Dados** da caixa de diálogo do nó Visualização de Dados para especificar as configurações de dados de uma visualização de dados analíticos selecionada a partir do IBM SPSS Collaboration and Deployment Services Repository.

Visualização de dados analíticos. Clique no botão de reticências (...) para selecionar uma visualização de dados analíticos. Se você não estiver atualmente conectado a um servidor de repositório, especifique a URL para o servidor na caixa de diálogo Repositório: Servidor, clique em **OK** e especifique suas credenciais de conexão na caixa de diálogo Repositório: Credenciais. Para obter mais informações sobre a criação de log no repositório e criação de objetos, consulte o Guia do Usuário do IBM SPSS Modeler.

Nome da tabela. Selecione uma tabela no modelo de dados na visualização de dados analíticos. Cada tabela no modelo de dados representa um conceito ou entidade envolvida no processo de análise preditiva. Campos para as tabelas correspondem a atributos das entidades representadas pelas tabelas. Por exemplo, se você está analisando pedidos do cliente, seu modelo de dados pode incluir uma tabela para clientes e outra para pedidos. A tabela de clientes pode ter atributos para o identificador do cliente, idade, sexo, estado civil e país de residência. A tabela de pedidos pode ter atributos para o identificador de pedido, número de itens no pedido, custo total e o identificador para o cliente que fez o pedido. O atributo de identificador de cliente poderia ser usado para associar os clientes na tabela de clientes com seus pedidos na tabela de pedidos.

Plano de acesso a dados. Selecione um plano de acesso a dados na visualização de dados analíticos. Um plano de acesso a dados associa as tabelas de modelo de dados em uma visualização de dados analíticos com origens de dados físicas. Uma visualização de dados analíticos contém diversos planos de acesso a dados. Ao alterar o plano de acesso a dados em uso, altere os dados usados pelo seu fluxo. Por exemplo, se a visualização de dados analíticos contém um plano de acesso de dados para a formação de um modelo e um plano de acesso de dados para testar um modelo, você pode alternar desde os dados de treinamento para os dados de teste alterando o plano de acesso de dados em uso.

Atributos Opcionais. Se um atributo específico não for necessário pelo aplicativo utilizando a visualização de dados analíticos, será possível marcar o atributo como opcional. Diferentemente dos atributos necessários, os atributos opcionais podem incluir valores nulos. Poderá ser necessário ajustar seu aplicativo para incluir manipulação de valores nulos para os atributos opcionais. Por exemplo, ao chamar de uma regra de negócios criada no IBM Operational Decision Manager, o IBM Analytical Decision Management consulta o serviço de regras para determinar quais entradas são necessárias. Se o registro a ser escorado contiver um valor nulo para qualquer um dos campos necessários do serviço de regra, a regra não será chamada e os campos de saída da regra serão preenchidos com valores padrão. Se um campo opcional contiver um valor nulo, a regra será chamada. A regra pode verificar valores nulos para controlar o processamento.

Para especificar atributos como opcionais, clique em **Atributos Opcionais** e selecione os atributos que são opcionais.

Incluir dados XML no campo. Selecione esta opção para criar um campo contendo os dados XML do modelo de objeto executável para cada linha de dados. Essas informações são necessárias se os dados forem utilizados com o IBM Operational Decision Manager. Especifique o nome para esse novo campo.

Nó de Origem Geospacial

Utilize o nó de origem Geospacial para exibir dados de mapa ou espaciais na sessão de mineração de dados. É possível importar dados de uma de duas maneiras:

- Em um arquivo de forma (.shp)
- Ao conectar-se a um servidor ESRI que contém um Sistema de Arquivo Hierárquico que inclui arquivos de mapeamento.

Nota: É possível conectar-se apenas a serviços de mapa público.

Os modelos do Spatio-Temporal Prediction (STP) podem incluir elementos de mapa ou espaciais em suas predições. Para obter mais informações sobre esses modelos, consulte o tópico denominado "Nó de modelagem do Spatio-Temporal Prediction" na seção Modelos de Séries Temporais do guia Nós de Modelagem do Modelador (ModelerModelingNodes.pdf).

Configurando Opções para o Nó Origem Geoespacial

Tipo de origem É possível importar dados a partir de um **Arquivo de Forma** (.shp) ou conectar a um **Serviço de mapa**.

Se estiver utilizando um **Arquivo de forma**, insira o nome do arquivo e o caminho do arquivo ou procure para selecionar o arquivo. O arquivo deve estar no diretório local ou ser acessado a partir de uma unidade mapeada; não é possível acessar o arquivo usando um caminho de convenção de nomenclatura uniforme (UNC).

Nota: Os dados de forma requerem arquivos .shp e .dbf. Os dois arquivos devem ter o mesmo nome e estar na mesma pasta. O arquivo .dbf é importado automaticamente quando selecionar o arquivo .shp. Além disso, pode haver um arquivo .prj que especifica o sistema de coordenadas para os dados de forma.

Se estiver utilizando um **Serviço de Mapa**, insira a URL para o serviço e clique em **Conectar**. Após conectar-se ao serviço, as camadas dentro desse serviço são exibidas na parte inferior da caixa de diálogo em uma estrutura em árvore na área de janela **Mapas disponíveis**; expanda a árvore e selecione a camada desejada.

Nota: É possível conectar-se apenas a serviços de mapa público.

Definição automática dos dados geoespaciais

Por padrão, o SPSS Modeler define automaticamente, quando possível, todos os campos de dados geoespaciais no nó de origem com os metadados corretos. Os metadados podem incluir o nível de medição do campo geoespacial (como Ponto ou Polígono) e o sistema de coordenadas que é utilizado pelos campos, incluindo detalhes como o ponto de origem (por exemplo, latitude 0, longitude 0) e as unidades de medida. Para obter mais informações sobre os níveis de medição, consulte "Subníveis de medição Geoespacial" na página 135.

Os arquivos .shp e .dbf que compõem o arquivo de forma contêm um campo identificador comum que é utilizado como uma chave. Por exemplo, o arquivo .shp pode conter países, com o campo de nome de país utilizado como um identificador, e o arquivo .dbf pode conter informações sobre esses países com o nome do país que também é utilizado como o identificador.

Nota: Se o sistema de coordenadas não for o mesmo que o sistema de coordenadas padrão do SPSS Modeler, poderá ser necessário reprojeter os dados para utilizar o sistema de coordenadas necessário. Para obter informações adicionais, consulte "Nó Reprojeção" na página 191.

Nó Visualização Corporativa

Nota: O nó Visualização Corporativa foi substituído no IBM SPSS Modeler 16.0 pelo nó Visualização de Dados. Para fluxos salvos em liberações anteriores, o nó Visualização Corporativa ainda é suportado. No entanto, ao atualizar ou criar novos fluxos, recomenda-se utilizar o nó Visualização de Dados.

O nó Visualização Corporativa permite criar e manter uma conexão entre uma sessão do IBM SPSS Modeler e uma Visualização Corporativa em um IBM SPSS Collaboration and Deployment Services Repository compartilhado. Isso permite ler dados a partir de uma Visualização Corporativa em um fluxo

do IBM SPSS Modeler e empacotar um modelo do IBM SPSS Modeler em um cenário que possa ser acessado por outros usuários do repositório compartilhado.

Um **cenário** é um arquivo que contém um fluxo do IBM SPSS Modeler com nós, modelos e propriedades adicionais específicas que permitem que ele seja implementado em um IBM SPSS Collaboration and Deployment Services Repository para propósitos de escoragem ou atualização de modelo automática. O uso dos nós Visualização Corporativa com cenários assegura que, em uma situação de multiusuário, todos os usuários trabalhem nos mesmos dados. Uma **conexão** é uma ligação de uma sessão do IBM SPSS Modeler com uma Visualização Corporativa no IBM SPSS Collaboration and Deployment Services Repository.

A **Visualização Corporativa** é o conjunto completo dos dados pertencentes a uma organização, independentemente de onde os dados estiverem localizados fisicamente. Cada conexão consiste em uma seleção específica de uma **Visualização de Aplicativo** (subconjunto da Visualização Corporativa customizado para um aplicativo específico), de uma **Definição de Provedor de Dados** (DPD—liga as tabelas e colunas da Visualização de Aplicativo lógica a uma origem de dados física) e de um **ambiente** (identifica quais colunas específicas devem ser associadas aos segmentos de negócios definidos) únicos. A Visualização Corporativa, as Visualizações de Aplicativo e as definições de DPD são armazenadas com uma versão definida no repositório, embora os dados reais residam em um ou mais bancos de dados ou outras origens externas.

Após uma conexão ter sido estabelecida, especifique uma **tabela** de Visualização de Aplicativo com a qual trabalhar no IBM SPSS Modeler. Em uma Visualização do Aplicativo, uma tabela é uma visualização lógica que consiste em algumas ou todas as colunas de uma ou mais tabelas físicas em um ou mais bancos de dados físicos. Assim, o nó Visualização Corporativa permite que os registros de diversas tabelas de banco de dados sejam vistos como uma única tabela em IBM SPSS Modeler.

requisitos

- Para utilizar o nó Visualização Corporativa, primeiro um IBM SPSS Collaboration and Deployment Services Repository deverá ser instalado e configurado em seu site, com uma Visualização Corporativa, Visualizações de Aplicativos e DPDs já definidos.
- Além disso, o IBM SPSS Collaboration and Deployment Services Enterprise View Driver deve ser instalado em cada computador usado para modificar ou executar o fluxo. Para Windows, simplesmente instale o driver no computador onde o IBM SPSS Modeler ou IBM SPSS Modeler Server está instalado e nenhuma configuração adicional do driver será necessária. No UNIX, uma referência ao script *pev.sh* deve ser incluída no script de inicialização. Entre em contato com o administrador local para obter detalhes sobre como instalar o IBM SPSS Collaboration and Deployment Services Enterprise View Driver.
- Um DPD é definido com relação a uma origem de dados ODBC específica. Para utilizar um DPD a partir do IBM SPSS Modeler, deve-se ter uma origem de dados ODBC definida no host do servidor IBM SPSS Modeler que tenha o mesmo nome e que se conecte ao mesmo armazém de dados, conforme referido no DPD.

Configurando Opções para o Nó de Visualização Corporativa

É possível utilizar as opções na guia Dados da caixa de diálogo Visualização Corporativa:

- Selecionar uma conexão do repositório existente
- Editar uma conexão do repositório existente
- Criar uma nova conexão do repositório
- Selecionar uma tabela de Visualização do Aplicativo

Consulte o *Guia do Administrador do IBM SPSS Collaboration and Deployment Services* para obter detalhes sobre como trabalhar com repositórios.

Conexão. A lista suspensa fornece opções para selecionar uma conexão do repositório existente, editar uma conexão existente ou incluir uma conexão. Se você já estiver com login efetuado em um repositório por meio do IBM SPSS Modeler, escolher a opção **Incluir/Editar uma conexão** exibe a caixa de diálogo Conexões da Visualização Corporativa na qual é possível definir ou editar os detalhes necessários para a conexão atual. Se você não estiver com login efetuado, essa opção exibirá a caixa de diálogo Login do repositório.

Para obter informações sobre criação de log no repositório, consulte o *Guia do Usuário do IBM SPSS Modeler*.

Depois que uma conexão com um repositório for estabelecida, essa conexão permanecerá ativa até que você saia do IBM SPSS Modeler. Uma conexão pode ser compartilhada por outros nós dentro do mesmo fluxo, no entanto, deve-se criar uma nova conexão para cada novo fluxo.

Um login bem-sucedido exibe a caixa de diálogo Conexões da Visualização Corporativa.

Nome da tabela. Esse campo está inicialmente vazio e não poderá ser preenchido até que uma conexão seja criada. Se você souber o nome da tabela de Visualização do Aplicativo que deseja acessar, insira-o no campo Nome da tabela. Caso contrário, clique no botão **Selecionar** para abrir uma caixa de diálogo listando as tabelas de Visualização do Aplicativo disponíveis.

Conexões da Visualização Corporativa

Essa caixa de diálogo permite definir ou editar os detalhes necessários para a conexão do repositório. É possível especificar:

- visualização do Aplicativo e versão
- Ambiente
- Definição de Provedor de Dados (DPD)
- Descrição da conexão

Conexões. Lista as conexões do repositório existentes.

- **Incluir uma nova conexão.** Exibe a caixa de diálogo Recuperar Objeto, na qual é possível procurar e selecionar uma Visualização de Aplicativo a partir do repositório.
- **Copiar a conexão selecionada.** Faz uma cópia de uma conexão selecionada, evitando ter que procurar novamente a mesma Visualização do Aplicativo.
- **Excluir a conexão selecionada.** Exclui a conexão selecionada da lista.

Detalhes da Conexão. Para a conexão selecionada atualmente na área de janela Conexões, exibe a Visualização do Aplicativo, o rótulo da versão, o ambiente, o DPD e o texto descritivo.

- **Visualização do aplicativo.** A lista suspensa exibe a visualização do aplicativo selecionada, se houver. Se as conexões foram feitas com outras Visualizações do aplicativo na sessão atual, elas também aparecerão na lista suspensa. Clique no botão Procurar adjacente para procurar por outras Visualizações do Aplicativo no repositório.
- **Rótulo de versão.** O campo suspenso lista todos os rótulos de versão definidos para a Visualização do Aplicativo especificada. Os rótulos de versão ajudam a identificar versões de objeto do repositório específicas. Por exemplo, pode haver duas versões para uma determinada Visualização de Aplicativo. Ao utilizar rótulos, é possível especificar o rótulo **TEST** para a versão utilizada no ambiente de desenvolvimento e o rótulo **PRODUCTION** para a versão utilizada no ambiente de produção. Selecione um rótulo apropriado.

Nota: Os rótulos não devem incluir caracteres de colchete (" ou ")", caracteres de colchete ("[" ou "]" ou caracteres de colchete duplos (byte duplo japonês), caso contrário, o nome da tabela não será exibido na guia Dados da caixa de diálogo Visualização Corporativa.

- **Ambiente.** O campo suspenso lista todos os ambientes válidos. A configuração do ambiente determina quais DPDs estão disponíveis, assim, a especificação de determinadas colunas deverá estar associada aos segmentos de negócio definidos. Por exemplo, quando **Analítico** é selecionada, apenas as colunas de Visualização do Aplicativo definidas como **Analítico** são retornadas. O ambiente padrão é **Analítico**; também é possível escolher **Operacional**.
- **Provedor de dados** A lista suspensa exibe os nomes de até dez Definições de Provedor de Dados para a Visualização do Aplicativo selecionada. Apenas as DPDs que referenciam a Visualização do Aplicativo selecionada são mostradas. Clique no botão Procurar adjacente para visualizar o nome e o caminho de todas as DPDs relacionadas à Visualização do Aplicativo atual.
- **Descrição.** Texto descritivo sobre a conexão do repositório. Este texto será utilizado para o nome de conexão - clicar em **OK** faz com que o texto apareça na lista suspensa Conexão, na barra de título da caixa de diálogo Visualização Corporativa e como o rótulo do nó Visualização Corporativa na tela.

Escolhendo o DPD

A caixa de diálogo Selecionar Provedor de Dados mostra o nome e o caminho de todos os DPDs que fazem referência à Visualização do Aplicativo atual.

As Visualizações do Aplicativo podem ter diversos DPDs para suportar os diferentes estágios de um projeto. Por exemplo, os dados históricos utilizados para construir um modelo podem vir de um banco de dados, ao passo que os dados operacionais são provenientes de outro banco de dados.

Um DPD é definido com relação a uma origem de dados ODBC específica. Para utilizar um DPD a partir do IBM SPSS Modeler, deve-se ter uma origem de dados ODBC definida no host do servidor IBM SPSS Modeler que tenha o mesmo nome e que se conecte ao mesmo armazém de dados, conforme referido no DPD.

Para escolher um DPD com o qual trabalhar, selecione o nome dele na lista e clique em **OK**.

Escolhendo a Tabela

A caixa de diálogo Selecionar Tabela lista todas as tabelas que forem referenciadas na Visualização do Aplicativo atual. A caixa de diálogo estará vazia se nenhuma conexão tiver sido estabelecida com um IBM SPSS Collaboration and Deployment Services Repository.

Para escolher uma tabela com a qual trabalhar, selecione o nome dela na lista e clique em **OK**.

Guias do Nó de Origem Comum

As opções a seguir podem ser especificadas para todos os nós de origem ao clicar na guia correspondente:

- **Guia Dados.** Utilizada para alterar o tipo de armazenamento padrão.
- **Guia Filtro.** Utilizada para eliminar ou renomear campos de dados. Essa guia oferece a mesma funcionalidade que o nó Filtro. Consulte o tópico “Configurando Opções de Filtragem” na página 145 para obter mais informações.
- **Guia Tipos.** Utilizada para configurar os níveis de medição. Essa guia oferece a mesma funcionalidade que o nó Tipo.
- **Guia Anotações.** Utilizada para todos os nós, essa guia oferece opções para renomear nós, fornecer um ToolTip customizado e armazenar uma anotação longa.

Configurando Níveis de Medição no Nó de Origem

As propriedades do campo podem ser especificadas em um nó de origem ou em um nó Tipo separado. A funcionalidade é semelhante nos dois nós. As seguintes propriedades estão disponíveis:

- **Campo** Dê um clique duplo em qualquer nome de campo para especificar rótulos de valor e de campo para dados no IBM SPSS Modeler. Por exemplo, os metadados do campo importados a partir do IBM

SPSS Statistics podem ser visualizados ou modificados aqui. Da mesma forma, é possível criar novos rótulos para campos e seus valores. Os rótulos que forem especificados aqui são exibidos em todo o IBM SPSS Modeler, dependendo das seleções feitas na caixa de diálogo de propriedades do fluxo.

- **Medição** Este é o nível de medição, utilizado para descrever características dos dados em um determinado campo. Se todos os detalhes de um campo forem conhecidos, ele será chamado de **totalmente instanciado**. Para obter mais informações, consulte “Níveis de Medição” na página 133.

Nota: O nível de medição de um campo é diferente de seu tipo de armazenamento, que indica se os dados são armazenados como sequências de caracteres, números inteiros, números reais, datas, horas, registros de data e hora ou listas.

- **Valores** Esta coluna permite especificar opções para ler os valores de dados a partir do conjunto de dados ou utilizar a opção **Especificar** para especificar os níveis de medição e valores em uma caixa de diálogo separada. Também é possível escolher transmitir campos sem ler seus valores. Para obter mais informações, consulte “Valores dos Dados” na página 137.

Nota: Não será possível corrigir a célula nesta coluna se a entrada **Campo** correspondente contiver uma lista.

- **Omisso** Utilizada para especificar como os valores omissos para o campo serão manipulados. Para obter mais informações, consulte “Definindo Valores Omissos” na página 141.

Nota: Não será possível corrigir a célula nesta coluna se a entrada **Campo** correspondente contiver uma lista.

- **Verificar** Nesta coluna, é possível configurar opções para assegurar que os valores de campo estejam em conformidade com os valores ou intervalos especificados. Para obter mais informações, consulte “Verificando Valores de Tipo” na página 142.

Nota: Não será possível corrigir a célula nesta coluna se a entrada **Campo** correspondente contiver uma lista.

- **Papel** Utilizado para informar aos nós de modelagem se os campos serão **Entrada** (campos preditores) ou **Destino** (campos preditos) para um processo de aprendizado por máquina. **Ambos** e **Nenhum** também são funções disponíveis, junto de **Partição**, que indica um campo usado para registros de partição em amostras separadas para treinamento, teste e validação. O valor **Divisão** especifica que os modelos separados serão construídos para cada valor possível do campo. Para obter mais informações, consulte “Configurando o Papel de Campo” na página 142.

Consulte o tópico “Nó Tipo” na página 131 para obter mais informações.

Quando Instanciar no Nó de Origem

Há duas maneiras pelas quais é possível aprender sobre o armazenamento de dados e os valores de seus campos. Esta **instanciação** pode ocorrer no nó de origem, quando os dados são trazidos pela primeira vez no IBM SPSS Modeler, ou ao inserir um nó Tipo no fluxo de dados.

A instanciação no nó de origem é útil quando:

- O conjunto de dados for pequeno.
- Você pretende derivar os novos campos utilizando o Construtor de Expressões (a instanciação torna os valores de campo disponíveis no Construtor de Expressões).

Geralmente, se o seu conjunto de dados não for muito grande e você não planejar incluir campos posteriormente no fluxo, a instanciação no nó de origem é o método mais conveniente.

Filtrando Campos do Nó de Origem

A guia Filtro em uma caixa de diálogo do nó de origem permite excluir campos a partir de operações de recebimento de dados com base em seu exame inicial dos dados. Isso será útil, por exemplo, se houver campos duplicados nos dados ou se você já estiver familiarizado suficientemente com os dados para

excluir campos irrelevantes. Como alternativa, é possível incluir um nó Filter separado posteriormente no fluxo. A funcionalidade é semelhante nos dois casos. Consulte o tópico “Configurando Opções de Filtragem” na página 145 para obter mais informações.

Capítulo 3. Nós de Operações de Registro

Visão Geral das Operações de Registro

Nós de operações de registro são utilizados para fazer mudanças nos dados no nível de registro. Essas operações são importantes durante as fases de **Entendimento de Dados** e de **Preparação de Dados** da mineração de dados porque elas permitem customizar os dados de acordo com suas necessidades de negócios específicas.

Por exemplo, com base nos resultados da auditoria de dados conduzida utilizando o nó Auditoria de Dados, (paleta Saída), talvez você prefira mesclar os registros dos três últimos meses de compra do cliente. Usando o nó Mesclar, é possível mesclar registros com base nos valores de um campo-chave, como *ID do Cliente*. Ou você pode descobrir que um banco de dados contendo informações sobre ocorrências do website é incontrolável com mais de um milhão de registros. Utilizando um nó Amostra, é possível selecionar um subconjunto de dados para uso na modelagem.

A paleta de Operações de Registro contém os seguintes nós:



O nó Selecionar seleciona ou descartará um subconjunto de registros do fluxo de dados com base em uma condição específica. Por exemplo, é possível selecionar os registros que pertencerem a uma região de vendas específica.



O nó Amostra seleciona um subconjunto de registros. Diversos tipos de amostra são suportados, incluindo amostras estratificadas, em cluster e não aleatórias (estruturadas). A amostragem pode ser útil para melhorar o desempenho e para selecionar grupos de registros ou transações relacionados para análise.



O nó Balanceamento corrige desbalanceamentos em um conjunto de dados, para que ele esteja em conformidade com uma condição especificada. A diretiva de balanceamento ajusta a proporção de registros onde uma condição for true pelo fator especificado.



O nó Agregado substitui uma sequência de registros de entrada por registros de saída resumidos e agregados.



O nó Recency, Frequency, Monetary (RFM) Aggregate permite selecionar os dados transacionais históricos dos clientes, eliminar quaisquer dados não utilizados e combinar todos os dados da transação restantes dos clientes em uma única linha que lista quando eles interagiram pela última vez com você, quantas transações eles realizaram e o valor monetário total dessas transações.



O nó Classificar classifica os registros em ordem crescente ou decrescente com base nos valores de um ou mais campos.



O nó Mesclagem seleciona diversos registros de entrada e cria um registro de saída único contendo alguns ou todos os campos de entrada. Ele é útil para mesclar dados de diferentes origens, como dados do cliente internos e dados demográficos adquiridos.



O nó Anexar concatena conjuntos de registros. Ele é útil para combinar conjuntos de dados com estruturas semelhantes, porém dados diferentes.



O nó Distinto remove registros duplicados seja transmitindo o primeiro registro distinto para o fluxo de dados ou descartando o primeiro registro e transmitindo quaisquer duplicatas para o fluxo de dados.



O nó TS de Fluxo constrói e escora os modelos de séries temporais em uma etapa, sem a necessidade de um nó Intervalos de Tempo.

Muitos dos nós na paleta de Operações de Registro requerem o uso de uma expressão CLEM. Se você estiver familiarizado com o CLEM, será possível digitar uma expressão no campo. No entanto, todos os campos de expressão fornecem um botão que abre o Construtor de Expressões do CLEM, que ajuda a criar essas expressões automaticamente.



Figura 1. Botão do Construtor de Expressões

Selecionar Nó

É possível utilizar os nós Seleção para selecionar ou descartar um subconjunto de registros do fluxo de dados com base em uma condição específica, como BP (pressão arterial) = "HIGH".

Modo. Especifica se os registros que atendem à condição serão incluídos ou excluídos do fluxo de dados.

- **Incluir.** Selecione para incluir registros que atendem à condição de seleção.
- **Descartar.** Selecione para excluir registros que atendem à condição de seleção.

Condição. Exibe a condição de seleção que será utilizada para testar cada registro, que é especificado utilizando uma expressão do CLEM. Insira uma expressão na janela ou utilize o Construtor de Expressões clicando no botão da calculadora (Construtor de Expressões) à direita da janela.

Se escolher descartar registros com base em uma condição, como a seguinte:

```
(var1='value1' and var2='value2')
```

por padrão, o nó Seleção também descarta registros que possuem valores nulos para todos os campos de seleção. Para evitar isso, anexe a seguinte condição à condição original:

```
and not(@NULL(var1) and @NULL(var2))
```

Os nós Seleção também são utilizados para escolher uma proporção de registros. Normalmente, você utilizaria um nó diferente, o nó Amostra, para esta operação. No entanto, se a condição que desejar especificar for mais complexa do que os parâmetros fornecidos, será possível criar sua própria condição utilizando o nó Seleção. Por exemplo, é possível criar uma condição como:

BP = "HIGH" and random(10) <= 4

Isso selecionará cerca de 40% dos registros que mostrarem pressão arterial alta e transmitirá esses registros posteriormente para análise adicional.

Nó de Amostra

É possível utilizar os nós Amostra para selecionar um subconjunto de registros para análise ou para especificar uma proporção de registros a serem descartados. Diversos tipos de amostra são suportados, incluindo amostras estratificadas, em cluster e não aleatórias (estruturadas). A amostragem pode ser utilizada por vários motivos:

- Para melhorar o desempenho ao estimar modelos em um subconjunto dos dados. Os modelos estimados a partir de uma amostra geralmente são tão precisos quanto os derivados do conjunto de dados integral, e poderão ser ainda mais se o desempenho melhorado permitir experimentar diferentes métodos que podem não ter sido tentados de outra forma.
- Para selecionar grupos de registros ou transações relacionados para análise, como selecionar todos os itens em um carrinho de compras on-line (ou de uma cesta de mercado), ou todas as propriedades em uma vizinhança específica.
- Para identificar unidades ou casos para inspeção aleatória no interesse de garantia de qualidade, de prevenção da fraude ou de segurança.

Nota: se você simplesmente deseja particionar dados em amostras de treinamento e de teste para fins de validação, um nó de Partição poderá ser utilizado. Consulte o tópico “Nó Partição” na página 170 para obter mais informações.

Tipos de Amostras

Amostras em cluster. Grupos ou clusters de amostra ao invés de unidades individuais. Por exemplo, suponha que você tenha um arquivo de dados com um registro por estudante. Se você armazenar em cluster por escola e o tamanho da amostra for 50%, então 50% das escolas serão escolhidas e todos os alunos de cada escola selecionada serão selecionados. Os alunos em escolas não selecionadas serão rejeitados. Em média, você espera que aproximadamente 50% dos alunos sejam selecionados, mas como as escolas variam em tamanho, a porcentagem pode não ser exata. Da mesma forma, é possível armazenar em cluster itens do carrinho de compras por ID de transação para assegurar que todos os itens das transações selecionadas sejam mantidos. Para obter um exemplo que armazena em cluster essas propriedades por cidade, consulte o fluxo de amostras *complexsample_property.str*.

Amostras estratificadas. Selecione amostras de modo independente em subgrupos não sobrepostos da população ou em camadas. Por exemplo, é possível assegurar que a amostra de homens e mulheres seja obtida em proporções iguais ou que cada região ou grupo socioeconômico em uma população urbana seja representado. Também é possível especificar um tamanho de amostra diferente para cada camada (por exemplo, se achar que um grupo foi sub-representado nos dados originais). Para obter um exemplo que estratifica essas propriedades de país, consulte o fluxo de amostras *complexsample_property.str*.

Amostragem sistemática ou 1 em n. Quando a seleção aleatória é difícil de obter, a amostra das unidades pode ser obtida sistematicamente (em um intervalo fixo) ou sequencialmente.

Ponderações de amostragem. As ponderações de amostragem são calculadas automaticamente ao desenhar uma amostra complexa e quase correspondem à "frequência" que cada unidade amostrada representa nos dados originais. Portanto, a soma das ponderações sobre a amostra deve estimar o tamanho dos dados originais.

Quadro de Amostragem

Um quadro de amostragem define a fonte potencial de casos a serem incluídos em uma amostra ou estudo. Em alguns casos, pode ser viável identificar cada membro único de uma população e incluir qualquer um deles em uma amostra, por exemplo, quando fizer amostragem de itens que vêm de uma linha de produção. Mais frequentemente, não será possível acessar todos os casos possíveis. Por exemplo, não é possível ter certeza sobre quem votará na eleição até após a eleição ter ocorrido. Neste caso, é possível utilizar o registro eleitoral como seu quadro de amostragem, mesmo se algumas pessoas registradas não votarem, e outras pessoas podem votar, apesar de não terem sido listadas no momento em que você verificou o registro. Quem não estiver no quadro de amostragem não tem perspectiva de ter amostra obtida. Se seu quadro de amostragem é próximo o suficiente por natureza da população que você está tentando avaliar é uma questão que deve ser abordada para cada caso real.

Opções de Nó de Amostra

É possível escolher o método **Simples** ou **Complexo**, conforme apropriado para seus requisitos.

Opções de Amostragem Simples

O método Simples permite selecionar uma porcentagem aleatória de registros, selecionar os registros contíguos ou selecionar cada n -ésimo registro.

Modo. Selecione se deseja transmitir (incluir) ou descartar (excluir) registros para os seguintes modos:

- **Incluir amostra.** Inclui registros selecionados no fluxo de dados e descarta todos os outros. Por exemplo, se você configurar o modo para **Incluir amostra** e configurar a opção **1 em n** para 5, então cada quinto registro será incluído, resultando em um conjunto de dados com quase um quinto do tamanho original. Esse é o modo padrão quando fizer amostragem de dados e o único modo quando utilizar o método complexo.
- **Descartar amostra.** Exclui registros selecionados e inclui todos os outros. Por exemplo, se você configurar o modo para **Descartar amostra** e configurar a opção **1 em n** para 5, então cada quinto registro será descartado. Este modo está disponível apenas com o método simples.

Amostra. Selecione o método de amostragem a partir das seguintes opções:

- **Primeira.** Selecione para usar amostragem de dados contínua. Por exemplo, se o tamanho máximo da amostra for configurado como 10000, então os primeiros 10.000 registros serão selecionados.
- **1 em n.** Selecione para obter amostra de dados ao transmitir ou descartar cada n -ésimo registro. Por exemplo, se n for configurado para 5, então cada quinto registro será selecionado.
- **% Aleatória.** Selecione para obter amostra de uma porcentagem aleatória dos dados. Por exemplo, se configurar a porcentagem para 20, então 20% dos dados serão transmitidos para o fluxo de dados ou descartados, dependendo do modo selecionado. Utilize o campo para especificar uma porcentagem de amostragem. Também é possível especificar um valor semente utilizando o controle **Configurar semente aleatória**.

Usar amostragem de nível de bloco (dentro de banco de dados somente). Esta opção será ativada somente se escolher amostragem de porcentagem aleatória quando executar mineração dentro da base de dados em um banco de dados Oracle ou IBM DB2. Nessas circunstâncias, a amostragem em nível de bloco pode ser mais eficiente.

Nota: Você não obtém um número exato de linhas retornadas cada vez que executar as mesmas configurações de amostra aleatória. Isso ocorre porque cada registro de entrada tem uma probabilidade de $N/100$ de ser incluído na amostra (em que N é a **% Aleatória** que especificar no nó) e as probabilidades são independentes, de modo que os resultados não são exatamente $N\%$.

Tamanho máximo da amostra. Especifica o número máximo de registros a serem incluídos na amostra. Esta opção é redundante e, portanto, desativada quando **Primeira** e **Incluir** forem selecionadas. Observe também que quando utilizada em combinação com a opção **% Aleatória**, essa configuração poderá evitar que determinados registros sejam selecionados. Por exemplo, se você tiver 10 milhões de registros em seu conjunto de dados e selecionar 50% de registros com um tamanho máximo da amostra de três milhões de

registros, então 50% dos primeiros seis milhões de registros serão selecionados e, os quatro milhões de registros restantes não terão nenhuma chance de serem selecionados. Para evitar essa limitação, selecione o método de amostragem **Complexo** e solicite uma amostra aleatória de três milhões de registros sem especificar uma variável de cluster ou de estratificação.

Opções de Amostragem Complexas

As opções de amostra complexas permitem um controle preciso da amostra, incluindo amostras em cluster, estratificadas e ponderadas junto de outras opções.

Cluster e estratificação. Permite especificar campos de ponderação de cluster, de estratificação e de entrada, se necessário. Consulte o tópico “Configurações de Cluster e de Estratificação” na página 76 para obter mais informações.

Tipo de amostra.

- **Aleatório.** Seleciona os clusters ou registros aleatoriamente dentro de cada camada.
- **Sistemático.** Seleciona os registros em um intervalo fixo. Essa opção funciona como o método *1 em n*, exceto que a posição do primeiro registro se altera dependendo de uma semente aleatória. O valor de *n* é determinado automaticamente com base no tamanho ou na proporção da amostra.

Unidades de amostra. É possível selecionar proporções ou contagens como as unidades básicas de amostra.

Tamanho da amostra. É possível especificar o tamanho da amostra de várias maneiras:

- **Fixo.** Permite especificar o tamanho geral da amostra como uma contagem ou proporção.
- **Customizado.** Permite especificar o tamanho de amostra para cada subgrupo ou camada. Essa opção estará disponível somente se um campo de estratificação tiver sido especificado na subcaixa de diálogo Cluster e Estratificação.
- **Variável.** Permite que o usuário selecione um campo que define o tamanho de amostra para cada subgrupo ou camada. Esse campo deve ter o mesmo valor para cada registro dentro de uma camada específica, por exemplo, se a amostra for estratificada por estado, então todos os registros com *county - Surrey* deverão ter o mesmo valor. O campo deve ser numérico e seus valores devem corresponder às unidades de amostra selecionadas. Para proporções, os valores devem ser maiores que 0 e menores que 1; para contagens, o valor mínimo é 1.

Amostra mínima por camada. Especifica um número mínimo de registros (ou um número mínimo de clusters se um campo de cluster estiver especificado).

Máximo de amostra por camada. Especifica o número máximo de registros ou clusters. Se selecionar esta opção sem especificar um campo de cluster ou de estratificação, uma amostra aleatória ou sistemática do tamanho especificado será selecionada.

Configurar semente aleatória. Ao amostrar ou particionar registros com base em uma porcentagem aleatória, esta opção permite duplicar os mesmos resultados em outra sessão. Ao especificar o valor inicial utilizado pelo gerador de número aleatório, é possível assegurar que os mesmos registros sejam designados toda vez que o nó for executado. Insira o valor semente desejado ou clique no botão **Gerar** para gerar automaticamente um valor aleatório. Se essa opção não estiver selecionada, uma amostra diferente será gerada toda vez que o nó for executado.

Nota: ao usar a opção **Configurar semente aleatória** com registros lidos a partir de um banco de dados, um nó Ordenar poderá ser necessário antes da amostragem para assegurar o mesmo resultado sempre que o nó for executado. Isso ocorre porque a semente aleatória depende da ordem de registros, que não é garantido que ela permaneça a mesma em um banco de dados relacional. Consulte o tópico “Nó Ordem” na página 83 para obter mais informações.

Configurações de Cluster e de Estratificação

A caixa de diálogo Cluster e Estratificação permite selecionar os campos de cluster, de estratificação e de ponderação ao obter uma amostra complexa.

Clusters. Especifica um campo categórico utilizado para armazenar registros em cluster. Os registros são amostrados com base na associação de cluster, com alguns clusters incluídos e outros não. Mas se algum registro de um determinado cluster for incluído, todos serão incluídos. Por exemplo, ao analisar associações do produto em carrinhos de compra, é possível armazenar em cluster itens por ID de transação para assegurar que todos os itens das transações selecionadas sejam mantidos. Ao invés de executar amostragem de registros, que destruiria informações sobre quais itens são vendidos em conjunto, é possível obter amostra de transações para assegurar que todos os registros das transações selecionadas sejam preservados.

Estratificar por. Especifica um campo categórico utilizado para estratificar os registros de modo que as amostras sejam selecionadas de modo independente em subgrupos não sobrepostos da população, ou em camadas. Se selecionar uma amostra de 50% estratificada por sexo, por exemplo, então duas amostras de 50% serão obtidas, uma para homens e outra para mulheres. Por exemplo, as camadas podem ser grupos socioeconômicos, categorias de empregos, faixas etárias ou grupos étnicos, permitindo assegurar tamanhos de amostra adequados para os subgrupos de interesse. Se houver três vezes mais mulheres do que homens no conjunto de dados original, essa razão será preservada ao obter amostra separadamente de cada grupo. Diversos campos de estratificação também podem ser especificados (por exemplo, amostragem de linhas de produtos dentro de regiões ou vice-versa).

Nota: se você estratificar por um campo que tenha valores omissos (valores nulos ou omissos do sistema, sequências de caracteres vazias, espaço em branco e valores omissos em branco ou definidos pelo usuário), então não será possível especificar tamanhos de amostra customizados para camadas. Se desejar utilizar tamanhos de amostra customizados quando estratificar por um campo com valores omissos ou em branco, eles deverão ter sido preenchidos anteriormente.

Usar ponderação de entrada. Especifica um campo utilizado para ponderar registros antes da amostragem. Por exemplo, se o campo de ponderação tiver valores que variam de 1 a 5, os registros com ponderação 5 terão cinco vezes mais chances de serem selecionados. Os valores deste campo serão sobrescritos pelas ponderações da saída final geradas pelo nó (consulte o parágrafo a seguir).

Nova ponderação de saída. Especifica o nome do campo no qual as ponderações finais serão gravadas se nenhum campo de ponderação de entrada estiver especificado. (Se um campo de ponderação de entrada for especificado, seus valores serão substituídos pelas ponderações finais, conforme observado acima, e nenhum campo de ponderação de saída separado será criado). Os valores de ponderação de saída indicam o número de registros representados por cada registro de amostra nos dados originais. A soma dos valores de ponderação fornece uma estimativa do tamanho da amostra. Por exemplo, se uma amostra aleatória de 10% for obtida, a ponderação de saída será 10 para todos os registros, indicando que cada registro de amostra representa aproximadamente dez registros nos dados originais. Em uma amostra estratificada ou ponderada, os valores de ponderação de saída podem variar com base na proporção de amostra para cada camada.

Comentários

- A amostragem em cluster será útil se você não puder obter uma lista completa da população que deseja obter amostra, mas puder obter listas completas de determinados grupos ou clusters. Ela também é utilizada quando uma amostra aleatória produz uma lista de sujeitos de teste que seria inviável contatar. Por exemplo, seria mais fácil visitar todos os agricultores de um estado do que uma seleção de agricultores espalhados em cada estado do país.
- É possível especificar campos de cluster e de estratificação para obter amostra de clusters de modo independente dentro de cada camada. Por exemplo, é possível obter amostra de valores de propriedade estratificados por estado e amostra de cluster por cidade dentro de cada estado. Isso

assegurar que uma amostra independente de cidades seja obtida a partir de cada estado. Algumas cidades serão incluídas e outras não, mas para cada cidade que for incluída, todas as propriedades dessa cidade são incluídas.

- Para selecionar uma amostra aleatória de unidades a partir de cada cluster, é possível unir dois nós de Amostra. Por exemplo, primeiro é possível obter amostra de cidades estratificadas por estado, conforme descrito acima. Em seguida, anexar um segundo nó de Amostra e selecionar *town* como um campo de estratificação, permitindo obter amostra de uma proporção de registros a partir de cada cidade.
- Nos casos em que uma combinação de campos é necessária para identificar exclusivamente os clusters, um novo campo pode ser gerado utilizando um nó Derivar. Por exemplo, se diversas lojas utilizarem o mesmo sistema de numeração para transações, será possível derivar um novo campo que concatena os IDs de loja e da transação.

Tamanhos de Amostra para Camadas

Ao desenhar uma amostra estratificada, a opção padrão é obter amostra da mesma proporção de registros ou clusters a partir de cada camada. Se um grupo exceder outro por um fator de 3, por exemplo, você geralmente deseja preservar a mesma proporção na amostra. Se esse não for o caso, no entanto, será possível especificar o tamanho da amostra separadamente para cada camada.

A caixa de diálogo Tamanhos de Amostra para Camadas lista cada valor do campo de estratificação, permitindo substituir o padrão por essa camada. Se diversos campos de estratificação forem selecionados, cada combinação possível de valores será listada, permitindo especificar o tamanho de cada grupo étnico dentro de cada cidade, por exemplo, ou de cada cidade dentro de cada estado. Os tamanhos são especificados como proporções ou contagens, conforme determinado pela configuração atual no nó Amostra.

Para Especificar Tamanhos de Amostra para Camadas

1. No nó Amostra, selecione **Complexo** e também um ou mais campos de estratificação. Consulte o tópico “Configurações de Cluster e de Estratificação” na página 76 para obter mais informações.
2. Selecione **Customizado** e depois **Especificar Tamanhos**.
3. Na caixa de diálogo Tamanhos da Amostra para Camadas, clique no botão **Ler Valores** na parte inferior esquerda para preencher a exibição. Poderá ser necessário instanciar valores em uma origem ou nó Tipo de envio de dados. Consulte o tópico “O Que É Instanciação?” na página 137 para obter mais informações.
4. Clique em qualquer linha para substituir o tamanho padrão para essa camada.

Notas sobre Tamanho da Amostra

Tamanhos de amostra customizados poderão ser úteis se camadas diferentes tiverem variâncias diferentes, por exemplo, para tornar os tamanhos de amostra proporcionais ao desvio padrão. (Se os casos dentro da camada forem mais variados, será necessário tirar mais amostras deles para obter uma amostra representativa). Ou, se uma camada for pequena, talvez você queira utilizar uma proporção de amostra maior para assegurar que um número mínimo de observações seja incluído.

Nota: se você estratificar por um campo que tenha valores omissos (valores nulos ou omissos do sistema, sequências de caracteres vazias, espaço em branco e valores omissos em branco ou definidos pelo usuário), então não será possível especificar tamanhos de amostra customizados para camadas. Se desejar utilizar tamanhos de amostra customizados quando estratificar por um campo com valores omissos ou em branco, eles deverão ter sido preenchidos anteriormente.

Nó Balanceamento

É possível utilizar nós Balanceamento para corrigir desbalanceamentos em conjuntos de dados de modo que estejam de acordo com os critérios de teste especificados. Por exemplo, suponha que um conjunto de dados possua apenas dois valores, *baixo* ou *alto*, e que 90% dos casos são *baixos* ao passo que apenas 10% dos casos são *altos*. Muitas técnicas de modelagem possuem problemas com tais dados parciais porque eles tendem a aprender apenas o resultado *baixo* e a ignorar o resultado *alto*, o que é muito raro. Se os dados forem bem balanceados com números aproximadamente iguais de resultados *baixos* e *altos*, os modelos terão uma chance melhor de localizar padrões que distinguem os dois grupos. Nesse caso, um nó Balanceamento é útil para criar uma diretiva de balanceamento que reduz casos com um resultado *baixo*.

O balanceamento é executado ao duplicar e, em seguida, descartar registros com base nas condições que você especificar. Os registros para os quais não há nenhuma condição são sempre passados. Como esse processo funciona ao duplicar e/ou descartar registros, a sequência original de seus dados é perdida em operações posteriores. Assegure-se de derivar quaisquer valores relacionados à sequência antes de incluir um nó Balanceamento no fluxo de dados.

Nota: os nós Balanceamentos podem ser gerados automaticamente a partir de gráficos de distribuição e histogramas. Por exemplo, é possível balancear seus dados para mostrar proporções iguais em todas as categorias de um campo categórico, conforme mostrado em um gráfico de distribuição.

Exemplo. Ao construir um fluxo RFM para identificar clientes recentes que tenham respondido positivamente a campanhas de marketing anteriores, o departamento de marketing de uma empresa de vendas utiliza um nó Balanceamento para balancear as diferenças entre as respostas verdadeiras e falsas nos dados.

Configurando Opções para o Nó Balanceamento

Registrar diretivas de balanceamento. Lista as diretivas balanceamento atuais. Cada diretiva inclui um fator e uma condição que instruem o software a "aumentar a proporção de registros por um fator especificado quando a condição for verdadeira". Um fator inferior a 1,0 significa que a proporção de registros indicados será diminuída. Por exemplo, se desejar diminuir o número de registros em que a droga Y é a droga de tratamento, será possível criar uma diretiva de balanceamento com um fator de 0,7 e uma condição Drug = "drugY". Esta diretiva significa que o número de registros em que a droga Y é a droga de tratamento será reduzido para 70% para todas as operações de recebimento de dados.

Nota: os fatores de balanceamento para redução podem ser especificados para quatro casas decimais. Fatores configurados abaixo de 0,0001 resultarão em um erro, já que os resultados não são calculados corretamente.

- **Crie condições** ao clicar no botão à direita do campo de texto. Isso insere uma linha vazia para inserir novas condições. Para criar uma expressão do CLEM para a condição, clique no botão Construtor de Expressões.
- **Exclua diretivas** usando o botão excluir vermelho.
- **Ordene as diretivas** utilizando os botões de seta para cima e para baixo.

Somente dados de treinamento de balanceamento. Se um campo de partição estiver presente no fluxo, essa opção balanceará os dados somente na partição de treinamento. Em particular, isso poderá ser útil se estiver gerando escores de propensão ajustada, que requer um teste desbalanceado ou uma partição de validação. Se nenhum campo de partição estiver presente no fluxo (ou se diversos campos de partição forem especificados), então essa opção será ignorada e todos os dados serão balanceados.

Nó Agregar

Agregação é uma tarefa de preparação de dados frequentemente usada para reduzir o tamanho de um conjunto de dados. Antes de continuar com a agregação, deve-se levar tempo para limpar os dados, concentrando especialmente nos valores ausentes. Depois de ter agregado, informações potencialmente úteis sobre valores ausentes podem ser perdidas.

É possível usar um nó Agregar para substituir uma sequência de registros de entrada com resumo, registros de saída agregada. Por exemplo, é possível ter um conjunto de registros de vendas de entrada, como os mostrados na tabela a seguir.

Tabela 13. Exemplo de entrada de registro de vendas

Idade	Sexo	Região	Ramificação	Sales
23	M	S	8	4
45	M	S	16	4
37	M	S	8	5
30	M	S	5	7
44	M	N	4	9
25	M	N	2	11
29	F	S	16	6
41	F	N	4	8
23	F	N	6	2
45	F	N	4	5
33	F	N	6	10

É possível agregar esses registros com *Sexo* e *Região* como campos-chave. Então, escolha agregar *Idade* com o modo **Média** e *Vendas* com o modo **Soma**. Selecione **Incluir contagem de registro no campo** na caixa de diálogo do nó Agregar e sua saída agregada seria conforme mostrado na tabela a seguir.

Tabela 14. Exemplo de registro agregado

Idade (média)	Sexo	Região	Vendas (soma)	Contagem de Registro
35,5	F	N	25	4
29	F	S	6	1
34,5	M	N	20	2
33,75	M	S	20	4

Nesta contagem você aprende, por exemplo, que a idade média das quatro equipes de vendas femininas na região Norte é de 35,5, e o total da soma das suas vendas foi 25 unidades.

Nota: Os campos como *Ramificação* serão automaticamente descartados quando nenhum modo de agregação for especificado.

Configurando Opções para o nó Agregado

No nó Agregado, você especificar o seguinte.

- Um ou mais campos-chave a serem utilizados como categorias para a agregação
- Um ou mais campos agregados para os quais calcular os valores agregados
- Um ou mais modos de agregação (tipos de agregação) para saída para cada campo agregado

Também é possível especificar os modos de agregação padrão a serem utilizados para os campos recém-incluídos e utilizar expressões (semelhantes às fórmulas) para categorizar a agregação.

Observe que para desempenho incluído, as operações de agregações podem se beneficiar da ativação do processamento paralelo.

Campos-chave. Lista campos que podem ser utilizados como categorias para agregação. Campos contínuos (numéricos) e categóricos podem ser utilizados como chaves. Se você escolher mais de um campo-chave, os valores serão combinados para produzir um valor da chave para agregar registros. Um registro agregado será gerado para cada campo-chave exclusivo. Por exemplo, se *Sex* e *Region* forem os campos-chave, cada combinação exclusiva de *M* e *F* com regiões *N* e *S* (quatro combinações exclusivas) terá um registro agregado. Para incluir um campo-chave, utilize o botão Seletor de Campo à direita da janela.

O restante da caixa de diálogo é dividido em duas áreas principais – **Agregados Básicos** e **Expressões Agregadas**.

Agregados Básicos

Campos agregados. Lista os campos para os quais valores serão agregados, bem como os modos de agregação selecionados. Para incluir campos nesta lista, utilize o botão Seletor de Campo à direita. Os modos de agregação a seguir estão disponíveis.

Nota: Alguns modos não são aplicáveis a campos não numéricos (por exemplo, **Soma** para um campo de data/hora). Os modos que não puderem ser utilizados com um campo agregado selecionado são desativados.

- **Soma.** Selecione para retornar valores somados para cada combinação de campo-chave. A soma é o total dos valores, em todos os casos com valores não omissos.
- **Média.** Selecione para retornar os valores médios para cada combinação de campo-chave. A média é uma medida da tendência central e é a média aritmética (a soma dividida pelo número de casos).
- **Mín.** Selecione para retornar valores mínimos para cada combinação de campo-chave.
- **Máx.** Selecione para retornar valores máximos para cada combinação de campo-chave.
- **SDev.** Selecione para retornar o desvio padrão para cada combinação de campo-chave. O desvio padrão é uma medida da dispersão ao redor da média e é igual à raiz quadrada da medição de variância.
- **Mediana.** Selecione para retornar os valores de mediana para cada combinação de campo-chave. A mediana é uma medida da tendência central que não é sensível a valores extremos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos). Também conhecida como o 50^o percentil ou 2^o quartil.
- **Contagem.** Selecione para retornar a contagem de valores não nulos para cada combinação de campo-chave.
- **Variância.** Selecione para retornar os valores de variância para cada combinação de campo-chave. A variância é uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividida por um a menos do número de casos.
- **1^o Quartil.** Selecione para retornar os valores do 1^o quartil (25^o percentil) para cada combinação de campo-chave.
- **3^o Quartil.** Selecione para retornar os valores do 3^o quartil (75^o percentil) para cada combinação de campo-chave.

Nota: Ao executar um fluxo contendo um nó Agregado, os valores retornados para os 1^o e 3^o Quartis quando efetuar SQL pushback para um banco de dados Oracle podem ser diferentes daqueles retornados no modo nativo.

Modo padrão. Especifique o modo de agregação padrão a ser utilizado para os campos incluídos recentemente. Se você usar frequentemente a mesma agregação, selecione um ou mais modos aqui e utilize o botão Aplicar a Todos à direita para aplicar os modos selecionados a todos os campos listados acima.

Extensão do novo nome de campo. Selecione para incluir um sufixo ou um prefixo, como "1" ou "novo", para duplicar campos agregados. Por exemplo, o resultado de uma agregação de valores mínimos no campo *Age* produzirá um nome de campo chamado *Age_Min_1* se tiver selecionado a opção de sufixo e especificado "1" como a extensão. *Nota:* as extensões de agregação como *_Min* ou *Max_* são automaticamente incluídas no novo campo, indicando o tipo de agregação executado. Selecione **Sufixo** ou **Prefixo** para indicar seu estilo de extensão preferencial.

Incluir contagem de registro no campo. Selecione para incluir um campo extra em cada registro de saída chamado *Record_Count*, por padrão. Esse campo indica quantos registros de entrada foram agregados para formar cada registro agregado. Crie um nome customizado para esse campo ao digitar no campo de edição.

Nota: Valores nulos do sistema são excluídos quando os agregados são calculados, mas são incluídos na contagem de registro. Valores em branco, por outro lado, são incluídos na agregação e também na contagem de registros. Para excluir valores em branco, é possível utilizar um nó Preenchimento para substituir valores em branco por valores nulos. Também é possível remover espaços em branco utilizando um nó Seleção.

Expressões Agregadas

As expressões são semelhantes às fórmulas que são criadas a partir de valores, nomes de campos, operadores e funções. Ao contrário das funções que funcionam em um registro único por vez, as expressões agregadas operam em um grupo, conjunto ou coleção de registros.

Nota: Expressões agregadas poderão ser criadas somente se o fluxo incluir uma conexão com o banco de dados (por meio de um nó de origem de Banco de Dados).

Novas expressões são criadas como campos derivados; para criar uma expressão, utilize as funções *Agregados de Banco de Dados* que estão disponíveis a partir do Construtor de Expressões.

Para obter mais informações sobre o Construtor de Expressões, consulte o Guia do Usuário do IBM SPSS Modeler (*ModelerUsersGuide.pdf*).

Observe que há uma conexão entre os **Campos-Chave** e quaisquer expressões agregadas que forem criadas porque as expressões agregadas são agrupadas pelo campo-chave.

As expressões agregadas válidas são aquelas que são avaliadas para agregar resultados; a seguir há alguns exemplos de expressões agregadas válidas e as regras que as controlam:

- É possível usar funções escalares para combinar diversas funções de agregação para produzir um resultado de agregação único. Por exemplo:

$\max(C01) - \min(C01)$

- Uma função de agregação pode operar sobre o resultado de diversas funções escalares. Por exemplo:

$\text{sum}(C01 * C01)$

Agregar configurações de otimização

Na guia Otimização, especifique o seguinte.

As chaves são contínuas. Selecione essa opção se você souber que todos os registros com os mesmos valores da chave são agrupados na entrada (por exemplo, se a entrada for classificada nos campos de chave). Fazer isso poderá melhorar o desempenho.

Permitir aproximação para Mediana e Quartis. As estatísticas de Ordem (Mediana, 1º Quartil e 3º Quartil) não são atualmente suportadas quando processar dados no Analytic Server. Se estiver utilizando o Analytic Server, será possível marcar essa caixa de seleção para utilizar um valor aproximado para essas estatísticas ao invés de ser calculado ao categorizar os dados e, em seguida, calcular uma estimativa para a estatística com base na distribuição entre as categorias. Por padrão, essa opção fica desmarcada.

Número de categorias. Disponível apenas se marcar a caixa de seleção **Permitir aproximação para Mediana e Quartis**. Selecione o número de categorias a serem utilizadas ao estimar a estatística; o número de categorias afeta a % **máxima de erro**. Por padrão, o número de categorias é 1000, que corresponde a um erro máximo de 0,1 por cento do intervalo.

Nó Agregado de RFM

O nó Agregado de Recency, Frequency, Monetary (RFM) permite selecionar os dados transacionais históricos dos clientes, eliminar quaisquer dados não utilizados e combinar todos os dados da transação restantes desses clientes em uma única linha, usando o ID de cliente exclusivo deles como uma chave, que lista quando eles interagiram pela última vez com você (recência), quantas transações eles realizaram (frequência) e o valor total dessas transações (monetário).

Antes de continuar com qualquer agregação, deve-se dedicar um tempo para limpar os dados, concentrando especialmente em quaisquer valores omissos.

Após ter identificado e transformado os dados utilizando o nó Agregado de RFM, é possível utilizar um nó Análise de RFM para realizar análise adicional. Consulte o tópico “Nó de Análise de RFM” na página 166 para obter mais informações.

Observe que assim que o arquivo de dados for executado por meio do nó Agregado de RFM, ele não terá nenhum valor de destino; portanto, antes de poder utilizá-lo como entradas para análise preditiva adicional com quaisquer nós de modelagem, como C5.0 ou CHAID, será necessário mesclá-lo com outros dados do cliente (por exemplo, ao corresponder os IDs de cliente). Consulte o tópico “Nó Mesclagem” na página 84 para obter mais informações.

Os nós Agregado de RFM e Análise de RFM no IBM SPSS Modeler são configurados para utilizar a categorização independente, ou seja, eles classificam e categorizam dados em cada medida de recência, de frequência e monetária, sem considerar seus valores ou as outras duas medidas.

Configurando Opções para o Nó Agregado de RFM

A guia Configurações do nó Agregado de RFM contém os seguintes campos.

Calcular recência relativa a. Especifique a data a partir da qual a recência das transações será calculada. Ela pode ser uma **Data fixa** que você inserir ou a **Data de hoje**, conforme configurado pelo sistema. A **Data de hoje** é inserida por padrão e é atualizada automaticamente quando o nó for executado.

IDs são contíguos. Se seus dados estiverem pré-ordenados de modo que todos os registros com o mesmo ID apareçam juntos no fluxo de dados, selecione esta opção para acelerar o processamento. Se seus dados não estiverem pré-ordenados (ou se você não tiver certeza), deixe essa opção desmarcada e o nó ordenará os dados automaticamente.

ID. Selecione o campo a ser utilizado para identificar o cliente e suas transações. Para exibir os campos a partir dos quais é possível selecionar, utilize o botão Seletor de Campo à direita.

Data. Selecione o campo de data a ser utilizado para calcular a recência. Para exibir os campos a partir dos quais é possível selecionar, utilize o botão Seletor de Campo à direita.

Observe que isto requer um campo com um armazenamento de data ou de registro data e hora, no formato apropriado para uso como entrada. Por exemplo, se você tiver um campo de sequência de caracteres com valores como *Jan 2007*, *Fev 2007*, e assim por diante, será possível convertê-lo em um campo de data utilizando um nó Preenchimento e a função `to_date()`. Consulte o tópico “Conversão de Armazenamento Usando o Nó Preenchimento” na página 155 para obter mais informações.

Valor. Selecione o campo a ser utilizado para calcular o valor monetário total de transações do cliente. Para exibir os campos a partir dos quais é possível selecionar, utilize o botão Seletor de Campo à direita. *Nota:* este deve ser um valor numérico.

Extensão do novo nome de campo. Selecione para anexar um sufixo ou um prefixo, como "12_month", à recência recém-gerada, à frequência e ao campo monetário. Selecione **Sufixo** ou **Prefixo** para indicar seu estilo de extensão preferencial. Por exemplo, isso pode ser útil ao examinar vários períodos de tempo.

Descartar registros com valor abaixo de. Se necessário, é possível especificar um valor mínimo abaixo do qual quaisquer detalhes da transação não serão utilizados ao calcular os totais de RFM. As unidades de valor se relacionam ao campo **Valor** selecionado.

Incluir somente transações recentes. Se estiver analisando um banco de dados grande, será possível especificar que apenas os últimos registros são utilizados. É possível escolher utilizar os dados registrados após uma determinada data ou dentro de um período recente:

- **Data de transação após.** Especifique a data da transação após a qual os registros serão incluídos em sua análise.
- **Transação dentro da última.** Especifique o número e o tipo de períodos (dias, semanas, meses ou anos) de volta a partir da data **Calcular recência relativa a** após a qual os registros serão incluídos em sua análise.

Salvar a data da segunda transação mais recente. Selecione essa caixa se desejar saber a data da segunda transação mais recente de cada cliente. Além disso, em seguida, é possível selecionar também a caixa **Salvar a data da terceira transação mais recente.** Por exemplo, isso pode ajudar a identificar clientes que possam ter realizado transações em algum tempo atrás considerável, mas apenas uma transação recente.

Nó Ordem

É possível utilizar os nós Ordem para classificar os registros em ordem crescente ou decrescente com base nos valores de um ou mais campos. Por exemplo, os nós Ordem são frequentemente utilizados para visualizar e selecionar registros com os valores de dados mais comuns. Geralmente, primeiro deve-se agregar os dados utilizando o nó Agregado e, em seguida, utilizar o nó Ordem para classificar os dados agregados em ordem decrescente de contagens de registro. Exibir esses resultados em uma tabela permite explorar os dados e tomar decisões, como selecionar os registros dos 10 melhores clientes.

A guia Configurações do nó Ordem contém os seguintes campos.

Ordenar por. Todos os campos selecionados para serem utilizados como chaves de ordenação são exibidos em uma tabela. Um campo-chave funciona melhor para ordenação quando ele é numérico.

- **Inclua campos** nessa lista usando o botão Seletor de Campo à direita.
- **Selecione uma ordem** ao clicando na seta **Crescente** ou **Decrescente** na coluna *Ordem* da tabela.
- **Exclua campos** usando o botão excluir vermelho.
- **Ordene as diretivas** utilizando os botões de seta para cima e para baixo.

Padrão de ordenação. Selecione **Crescente** ou **Decrescente** para uso como o padrão de ordenação quando novos campos são incluídos acima.

Nota: O nó Ordem não será aplicado se houver um nó Distinto abaixo no fluxo de modelo. Para obter informações sobre o nó Distinto, consulte “Nó Distinto” na página 93.

Configurações de Otimização de Ordenação

Se estiver trabalhando com dados que você já sabe que estão ordenados por alguns campos-chave, é possível especificar quais campos já estão ordenados, para que o sistema possa ordenar o restante dos dados de modo mais eficiente. Por exemplo, você deseja ordenar por *Idade* (decrésciente) e *Droga* (crescente), mas sabe que seus dados são já estão ordenados por *Idade* (decrésciente).

Os dados são pré-ordenados. Especifica se os dados já estão ordenados por um ou mais campos.

Especificar ordenação existente. Especifique os campos que já estiverem ordenados. Utilizando a caixa de diálogo Selecionar Campos, inclua campos na lista. Na coluna *Ordem*, especifique se cada campo é classificado em ordem crescente ou decrésciente. Se estiver especificando diversos campos, assegure-se de listá-los na ordenação correta. Use as setas para a direita da lista para organizar os campos na ordem correta. Se você errar ao especificar a ordenação existente correta, um erro aparecerá ao executar o fluxo, exibindo o número de registro no qual a ordenação está inconsistente com o que você especificou.

Nota: a velocidade da ordenação poderá ser beneficiada com a ativação do processamento paralelo.

Nó Mesclagem

A função de um nó Mesclagem é tomar diversos registros de entrada e criar um registro de saída único contendo todos ou alguns dos campos de entrada. Esta é uma operação útil quando desejar mesclar dados de diferentes origens, como dados do cliente interno e dados demográficos de compras. É possível mesclar dados das seguintes formas.

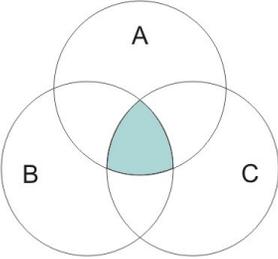
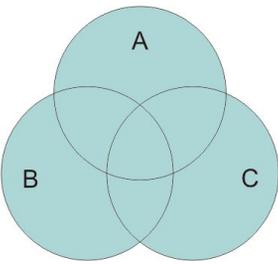
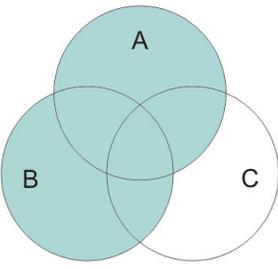
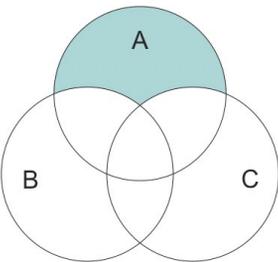
- Mesclar por **Ordem** concatena os registros correspondentes de todas as origens na ordem de entrada até que a menor origem de dados seja esgotada. Se usar esta opção, é importante que seus dados tenham sido classificados usando o nó Ordem.
- Mesclar utilizando um campo **Chave**, como *ID do Cliente*, para especificar como corresponder registros de uma origem de dados com registros de outra(s). Vários tipos de junções são possíveis, incluindo junção interna, junção externa completa, junção externa parcial e antijunção. Consulte o tópico “Tipos de Junções” para obter mais informações.
- Mesclar por **Condição** significa que é possível especificar uma condição a ser satisfeita para que a mesclagem ocorra. É possível especificar a condição diretamente no nó ou construir a condição utilizando o Construtor de Expressões.
- Mesclar por **Condição Classificada** é uma junção externa à esquerda na qual você especifica uma condição a ser satisfeita para que a mesclagem ocorra e também uma expressão de ranqueamento que ordena de baixo para cima. Mais frequentemente utilizada para mesclar dados geoespaciais, é possível especificar a condição diretamente no nó ou construir a condição utilizando o Construtor de Expressões.

Tipos de Junções

Ao utilizar um campo-chave para mesclagem de dados, é útil gastar algum tempo pensando sobre quais registros serão excluídos e quais serão incluídos. Há uma variedade de junções, que são discutidas em detalhes a seguir.

Os dois tipos básicos de junções são referidos como junções internas e externas. Esses métodos são frequentemente utilizados para mesclar tabelas a partir de conjuntos de dados relacionados com base em valores comuns de um campo-chave, como *ID do Cliente*. As junções internas permitem uma mesclagem limpa e um conjunto de dados de saída que inclui apenas os registros completos. As junções externas também incluem registros completos dos dados mesclados, mas elas também permitem incluir dados exclusivos de uma ou mais tabelas de entrada.

Os tipos de junções permitidos são descritos em mais detalhes a seguir.

	<p>Uma junção interna inclui apenas registros em que um valor para o campo-chave é comum para todas as tabelas de entrada. Ou seja, os registros não correspondidos não serão incluídos no conjunto de dados de saída.</p>
	<p>Uma junção externa integral inclui todos os registros, correspondentes e não correspondentes, a partir das tabelas de entrada. Junções externas esquerdas e direitas são referidas como junções externas parciais e são descritas a seguir.</p>
	<p>Uma junção externa parcial inclui todos os registros correspondidos utilizando o campo-chave, bem como registros não correspondidos das tabelas especificadas. (Ou, em outras palavras, todos os registros de algumas tabelas e apenas os registros correspondentes de outras). Tabelas (como A e B mostradas aqui) podem ser selecionadas para inclusão na junção externa utilizando o botão Selecionar na guia Mesclar. As junções parciais também são chamadas de junções externas direitas ou esquerdas quando apenas duas tabelas estiverem sendo mescladas. Como o IBM SPSS Modeler permite a mesclagem de mais de duas tabelas, isso é referido como uma junção externa parcial.</p>
	<p>Uma antijunção inclui apenas os registros não correspondidos para a primeira tabela de entrada (Tabela A mostrada aqui). Esse tipo de junção é o oposto de uma junção interna e não inclui registros completos no conjunto de dados de saída.</p>

Por exemplo, se você tiver informações sobre fazendas em um conjunto de dados e solicitações de seguro relacionadas à fazenda em outro, é possível corresponder os registros da primeira origem com a segunda origem utilizando as opções de Mesclagem.

Para determinar se um cliente em sua fazenda de amostra registrou uma solicitação de seguro, utilize a opção de junção interna para retornar uma lista mostrando onde todos os IDs correspondem a partir de duas amostras.

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0...
2	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0...
3	id620	name620	north	880.000	74.000	6.000	426988.000	rapeseed	arable_de	118885.0...

Figura 2. Saída de amostra para uma mesclagem de junção interna

O uso da opção de junção externa integral retorna registros correspondentes e não correspondentes partir das tabelas de entrada. O valor omissivo do sistema (\$null\$) será utilizado para quaisquer valores

incompletos.

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalu
1	id601	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	decomm...	74703.1C
2	id602	name602	north	1780.000	42.000	9.000	734118.000	maize	\$null\$	\$nul
3	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0
4	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0
5	id606	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	arable_de	122135.0

Figura 3. Saída de amostra para uma mesclagem de junção externa integral

Uma junção externa parcial inclui todos os registros correspondidos utilizando o campo-chave, bem como registros não correspondidos das tabelas especificadas. A tabela exibe todos os registros correspondidos a partir do campo de ID, bem como os registros correspondidos a partir do primeiro conjunto de dados.

	id	claimtype	claimvalue	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	\$null\$	\$null\$	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id604	decomm...	281082.0...	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes
3	id605	decomm...	122006.0...	name605	north	1700.000	46.000	8.000	621148.000	wheat
4	id607	\$null\$	\$null\$	name607	southeast	1820.000	29.000	6.000	211605.000	maize
5	id608	\$null\$	\$null\$	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
6	id609	\$null\$	\$null\$	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
7	id615	\$null\$	\$null\$	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
8	id618	\$null\$	\$null\$	name618	southeast	1180.000	98.000	3.000	368646.000	maize

Figura 4. Saída de amostra para uma mesclagem de junção externa parcial

Se estiver utilizando a opção antijunção, a tabela retornará apenas os registros não correspondidos para a primeira tabela de entrada.

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id607	name607	southeast	1820.000	29.000	6.000	211605.000	maize
3	id608	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
4	id609	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
5	id615	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
6	id618	name618	southeast	1180.000	98.000	3.000	368646.000	maize
7	id619	name619	north	840.000	64.000	8.000	457552.000	potatoes

Figura 5. Saída de amostra para uma mesclagem de antijunção

Especificando Método de Mesclagem e Chaves

A guia Mesclagem do nó Mesclagem contém os seguintes campos.

Método de Mesclagem Selecione o método a ser utilizado para mesclar os registros. Selecionar **Chaves** ou **Condição** ativa a metade inferior da caixa de diálogo.

- **Ordenar** Mescla registros em ordem de modo que o n -ésimo registro de cada entrada seja mesclado para produzir o n -ésimo registro de saída. Quando qualquer registro é executado fora de um registro de entrada correspondente, nenhum outro registro de saída é produzido. Isso significa que o número de registros que são criados é o número de registros no conjunto de dados menor.
- **Chaves** Utiliza um campo-chave, como *ID Transação*, para mesclar os registros com o mesmo valor no campo-chave. Esse é equivalente a uma "junção equivalente" de um banco de dados. Se um valor de chave ocorrer mais de uma vez, todas as combinações possíveis serão retornadas. Por exemplo, se os registros com o mesmo valor do campo-chave *A* contiverem valores diferentes *B*, *C* e *D* em outros campos, os campos mesclados produzirão um registro separado para cada combinação de *A* com o valor *B*, *A* com o valor *C* e *A* com o valor *D*.

Nota: valores nulos não são considerados idênticos no método mesclagem por chave e não serão juntados.

- **Condição** Utilize esta opção para especificar uma condição para a mesclagem. Para obter mais informações, consulte “Especificando Condições para uma Mesclagem” na página 88.
- **Condição classificada** Utilize esta opção para especificar se cada pareamento de linha nos conjuntos de dados primários e todos os secundários deve ser mesclada; utilize a expressão de ranqueamento para classificar diversas correspondências em qualquer ordem de baixo para cima. Para obter mais informações, consulte “Especificando Condições Classificadas para uma Mesclagem” na página 88.

Chaves possíveis Lista apenas os campos com nomes de campo que corresponderem exatamente em todas as origens de dados de entrada. Selecione um campo nesta lista e utilize o botão de seta para incluí-lo como um campo-chave utilizado para mesclagem de registros. Mais de um campo-chave pode ser utilizado. É possível renomear campos de entrada não correspondentes utilizando um nó Filtro ou a guia Filtro de um nó de origem.

Chaves para mesclagem Lista todos os campos utilizados para mesclar registros de todas as origens de dados de entrada com base nos valores dos campos-chave. Para remover uma chave da lista, selecione uma e utilize o botão de seta para retorná-la para a lista de Chaves Possíveis. Quando mais de um campo-chave é selecionado, a opção abaixo é ativada.

Combinar campos-chave duplicados Quando mais de um campo-chave é selecionado acima, esta opção assegura que haja somente um campo de saída com esse nome. Esta opção é ativada por padrão, exceto no caso em que os fluxos tiverem sido importados de versões anteriores do IBM SPSS Modeler. Quando essa opção é desativada, os campos-chave duplicados devem ser renomeados ou excluídos utilizando a guia Filtrar na caixa de diálogo do nó Mesclagem.

Incluir somente registros correspondentes (junção interna) Selecione para mesclar apenas registros completos.

Incluir registros correspondentes e não correspondentes (junção externa completa) Selecione para executar uma "junção externa completa". Isso significa que se os valores para o campo-chave não estiverem presentes em todas as tabelas de entrada, os registros incompletos ainda serão retidos. O valor indefinido (\$null\$) é incluído no campo-chave e também no registro de saída.

Incluir registros correspondentes e não correspondentes selecionados (junção externa parcial) Selecione para executar uma "junção externa parcial" das tabelas que você selecionar em uma subcaixa de diálogo. Clique em **Selecionar** para especificar as tabelas para as quais registros incompletos serão retidos na mesclagem.

Incluir registros no primeiro conjunto de dados que não corresponder a nenhum outro (antijunção) Selecione para executar um tipo de "antijunção", em que apenas os registros não correspondentes do primeiro conjunto de dados são transmitidos posteriormente. É possível especificar a ordem dos conjuntos de dados de entrada utilizando as setas na guia Entradas. Este tipo de junção não inclui registros completos no conjunto de dados de saída. Para obter mais informações, consulte “Tipos de Junções” na página 84.

Selecionando Dados para Junções Parciais

Para junção externa parcial, deve-se selecionar uma ou mais tabelas para as quais os registros incompletos serão retidos. Por exemplo, talvez você queira reter todos os registros a partir de uma tabela Cliente enquanto mantém apenas os registros correspondidos da tabela de Empréstimos Hipotecários.

Coluna Junção Externa. Na coluna *Junção Externa*, selecione os conjuntos de dados a serem incluídos em sua totalidade. Para uma junção parcial, registros sobrepostos serão retidos, bem como registros incompletos para conjuntos de dados selecionados aqui. Consulte o tópico “Tipos de Junções” na página 84 para obter mais informações.

Especificando Condições para uma Mesclagem

Ao configurar o método de mesclagem para **Condição**, é possível especificar uma ou mais condições a serem atendidas para que a mesclagem ocorra.

É possível inserir as condições diretamente no campo Condição ou construí-las com a ajuda do Construtor de Expressões clicando no símbolo de calculadora à direita do campo.

Incluir tags para duplicar os nomes dos campos e evitar conflitos de mesclagem Se dois ou mais dos conjuntos de dados a serem mesclados contiverem os mesmos nomes de campo, marque esta caixa de seleção para incluir uma tag de prefixo diferente no início dos cabeçalhos de coluna do campo. Por exemplo, se houver dois campos chamados *Name*, o resultado da mesclagem deverá conter *1_Name* e *2_Name*. Se a tag for renomeada na origem de dados, o novo nome será utilizado ao invés da tag de prefixo numerada. Se você não marcar esta caixa de seleção e houver nomes duplicados nos dados, um aviso será exibido à direita da caixa de seleção.

Especificando Condições Classificadas para uma Mesclagem

Uma mesclagem de Condição Classificada pode ser considerada como uma mesclagem de junção externa à esquerda por condição, em que o lado esquerdo da mesclagem é o conjunto de dados primário no qual cada registro é um evento. Por exemplo, em um modelo que é utilizado para localizar padrões nos dados criminais, cada registro no conjunto de dados primário seria um crime e suas informações associadas (localização, tipo, e assim por diante). Neste exemplo, o lado direito pode conter os conjuntos de dados geoespaciais relevantes.

A mesclagem utiliza uma condição de mesclagem e uma expressão de ranqueamento. A condição de mesclagem pode utilizar uma função geoespacial como *within* ou *close_to*. Durante a mesclagem, todos os conjuntos de dados do lado direito são incluídos no conjunto de dados do lado esquerdo, mas diversas correspondências resultam em um campo de lista. Por exemplo:

- Lado esquerdo: dados criminais
- Lado direito: conjunto de dados de países e conjunto de dados de estradas
- Condições de mesclagem: dados criminais *within* países e *close_to* estradas, junto de uma definição do que é contado como *close_to*.

Neste exemplo, se um crime ocorreu em uma distância *close_to* necessária de três estradas (e o número de correspondências a serem retornadas é configurado para pelo menos três), então todas as três estradas serão retornadas como um item da lista.

Ao configurar o método de mesclagem para **Condição classificada**, é possível especificar uma ou mais condições a serem atendidas para que a mesclagem ocorra.

Conjunto de dados primário Selecione o conjunto de dados primário para a mesclagem; os campos de todos os outros conjuntos de dados são incluídos no conjunto de dados selecionado. Isso pode ser considerado como o lado esquerdo de uma mesclagem de junção externa.

Ao selecionar um conjunto de dados primário, todos os outros conjuntos de dados de entrada que estiverem conectados ao nó Mesclagem serão automaticamente listados na tabela **Mesclagens**.

Incluir tags para duplicar os nomes dos campos e evitar conflitos de mesclagem Se dois ou mais dos conjuntos de dados a serem mesclados contiverem os mesmos nomes de campo, marque esta caixa de seleção para incluir uma tag de prefixo diferente no início dos cabeçalhos de coluna do campo. Por exemplo, se houver dois campos que forem chamados *Name*, o resultado da mesclagem deverá conter *1_Name* e *2_Name*. Se a tag for renomeada na origem de dados, o novo nome será utilizado ao invés da tag de prefixo numerada. Se você não marcar esta caixa de seleção e houver nomes duplicados nos dados, um aviso será exibido à direita da caixa de seleção.

Mesclagens

Conjunto de Dados

Mostra o nome dos conjuntos de dados secundários que estiverem conectados como entradas ao nó Mesclagem. Por padrão, quando houver mais de um conjunto de dados secundário, eles serão listados na ordem em que eles foram conectados ao nó Mesclagem.

Condição de mesclagem

Insira as condições exclusivas para a mesclagem de cada um dos conjuntos de dados na tabela com o conjunto de dados primário. É possível digitar as condições diretamente na célula ou construí-las com a ajuda do Construtor de Expressões clicando no símbolo de calculadora à direita da célula. Por exemplo, é possível utilizar predicados geoespaciais para criar uma condição de mesclagem que coloca os dados criminais de um conjunto de dados dentro dos dados do estado de outro conjunto de dados. A condição de mesclagem padrão depende do nível de medição geoespacial, conforme mostrado na lista a seguir.

- Point, LineString, MultiPoint, MultiLineString - a condição padrão de *close_to*.
- Polygon, MultiPolygon - a condição padrão de *within*.

Para obter mais informações sobre esses níveis, consulte “Subníveis de medição Geoespacial” na página 135.

Se um conjunto de dados contiver diversos campos geoespaciais de tipos diferentes, a condição padrão que é utilizada depende do primeiro nível de medição que for localizado nos dados, na seguinte ordem decrescente.

- Point
- LineString
- Polygon

Nota: Os padrões estão disponíveis somente quando houver um campo de dados geoespaciais no banco de dados secundário.

Expressão de classificação em posto

Especifique uma expressão para classificar a mesclagem dos conjuntos de dados; essa expressão é utilizada para classificar diversas correspondências em uma ordem que seja baseada nos critérios de ranqueamento. É possível digitar as condições diretamente na célula ou construí-las com a ajuda do Construtor de Expressões clicando no símbolo de calculadora à direita da célula.

As expressões de ranqueamento padrão de distância e de área são fornecidas no Construtor de Expressões e ambas são classificadas da baixa para alta, significando, por exemplo, que a correspondência superior para distância é o valor menor. Um exemplo de ranqueamento por distância é quando o conjunto de dados primário contém os crimes e suas localizações associadas e cada outro conjunto de dados contém objetos com localizações; neste caso, a distância entre os crimes e os objetos pode ser utilizada como critério de ranqueamento. A expressão de ranqueamento padrão depende do nível de medição geoespacial, conforme mostrado na lista a seguir.

- Point, LineString, MultiPoint, MultiLineString - a expressão padrão é *distance*.
- Polygon, MultiPolygon - a expressão padrão é *area*.

Nota: Os padrões estão disponíveis somente quando houver um campo de dados geoespaciais no banco de dados secundário.

Número de correspondências

Especifique o número de correspondências que são retornadas, com base nas expressões de condição e de ranqueamento. O número padrão de correspondência depende do nível de medição geoespacial no conjunto de dados secundário, conforme mostrado na lista abaixo, no entanto, é possível dar um clique duplo na célula para inserir seu próprio valor, até um máximo de 100.

- Point, LineString, MultiPoint, MultiLineString - valor padrão de 3.
- Polygon, MultiPolygon - valor padrão de 1.
- Conjunto de dados que não contém nenhum campo geoespacial - valor padrão de 1.

Como um exemplo, se você configurar uma mesclagem que é baseada em uma **Condição de Mesclagem** de *close_to* e uma **Expressão de Ranqueamento** de *distance*, as três principais correspondências (mais próximas) dos conjuntos de dados secundários para cada registro no conjunto de dados primário são retornadas como os valores no campo de lista resultante.

Filtrando Campos com o Nó Mesclagem

Os nós Mesclagem incluem uma maneira conveniente de filtrar ou renomear campos duplicados como resultado da mesclagem de diversas origens de dados. Clique na guia **Filtrar** na caixa de diálogo para selecionar opções de filtragem.

As opções apresentadas aqui são quase idênticas àquelas do nó Filtro. No entanto, há opções adicionais não discutidas aqui que estão disponíveis no menu Filtro. Consulte o tópico “Filtrando ou Renomeando Campos” na página 145 para obter mais informações.

Campo. Exibe os campos de entrada a partir de origens de dados atualmente conectadas.

Tag. Lista o nome (ou número) da tag associado à ligação de origem de dados. Clique na guia **Entradas** para alterar ligações ativas para esse nó Mesclagem.

Nó de origem. Exibe o nó de origem cujos dados estão sendo mesclados.

Nó conectado. Exibe o nome do nó para o nó que está conectado ao nó Mesclagem. Frequentemente, a mineração de dados complexos requer várias operações de mesclagem ou anexação que podem incluir o mesmo nó de origem. O nome do nó conectado fornece uma maneira de diferenciá-los.

Filtrar. Exibe as conexões atuais entre o campo de entrada e saída. Conexões ativas mostram uma seta sólida. Conexões com um X vermelho indicam campos filtrados.

Campo. Lista os campos de saída após mesclagem ou anexação. Campos duplicados são exibidos em vermelho. Clique no campo Filtro acima para desativar campos duplicados.

Visualizar campos atuais. Selecione para visualizar informações sobre os campos selecionados a serem utilizados como campos-chave.

Visualizar as configurações de campos não utilizados. Selecione para visualizar informações sobre os campos que não estiverem atualmente em uso.

Configurando a Ordem e a Identificação da Entrada

Utilizando a guia Entradas nas caixas de diálogo dos nós Mesclagem e Anexação, é possível especificar a ordem das origens de dados de entrada e fazer quaisquer mudanças no nome da tag para cada origem.

Tags e ordem de conjuntos de dados de entrada. Selecione para mesclar ou anexar somente registros completos.

- **Tag.** Lista os nomes de tag atuais para cada origem de dados de entrada. Os nomes de tags, ou **tags**, são uma maneira de identificar exclusivamente os links de dados para a operação de mesclagem ou anexação. Por exemplo, é como se a água de diversos canos fosse combinada em um ponto e fluísse por um único cano. Os dados do IBM SPSS Modeler fluem da mesma forma e o ponto de mesclagem é geralmente uma interação complexa entre as várias origens de dados. As tags fornecem uma maneira de gerenciar as entradas ("canos") para um nó Mesclagem ou Anexação para que, se o nó for salvo ou desconectado, os links permaneçam e sejam facilmente identificados.

Ao conectar origens de dados adicionais a um nó Mesclagem ou Anexação, as tags padrão são criadas automaticamente utilizando números para representar a ordem na qual você conectou os nós. Essa ordem não está relacionada à ordem dos campos nos conjuntos de dados de entrada ou de saída. É possível alterar a tag padrão inserindo um novo nome na coluna *Tags*.

- **Nó de Origem.** Exibe o nó de origem cujos dados estão sendo combinados.
- **Nó Conectado.** Exibe o nome do nó para o nó que está conectado ao nó Mesclagem ou Anexação. Frequentemente, a mineração de dados complexos requer várias operações de mesclagem que podem incluir o mesmo nó de origem. O nome do nó conectado fornece uma maneira de diferenciá-los.
- **Campos.** Lista o número de campos em cada origem de dados.

Visualizar tags atuais. Selecione para visualizar tags que estão sendo ativamente utilizadas pelo nó Mesclagem ou Anexação. Em outras palavras, as tags atuais identificam links para o nó que possui dados fluindo por ele. Utilizando a metáfora de cano, as tags atuais representam os canos com um fluxo de água existente.

Visualizar as configurações de tag não utilizada. Selecione para visualizar as tags ou ligações que foram anteriormente utilizadas para conexão com o nó Mesclagem ou Anexação, mas que não estão atualmente conectadas a uma origem de dados. Isso é semelhante a canos vazios que ainda estão intactos em um sistema de tubulação. É possível escolher conectar esses "canos" a uma nova origem ou removê-los. Para remover tags não utilizadas do nó, clique em **Limpar**. Isso limpa todas as tags não utilizadas de uma vez.

Configurações de Otimização de Mesclagem

O sistema fornece duas opções que podem ajudá-lo a mesclar dados com mais eficiência em determinadas situações. Essas opções permitem otimizar a mesclagem quando um conjunto de dados de entrada é significativamente maior que outros conjuntos de dados ou quando seus dados já estiverem ordenados por todos ou por alguns campos-chave que você estiver usando para mesclagem.

Nota: As otimizações desta guia se aplicam apenas à execução do nó nativo do IBM SPSS Modeler, ou seja, quando o nó Mesclagem não realiza SQL pushback. As configurações de otimização não têm efeito sobre a geração de SQL.

Um conjunto de dados de entrada é relativamente grande. Selecione para indicar que um dos conjuntos de dados de entrada é muito maior que os outros. O sistema armazenará em cache os menores conjuntos de dados na memória e, em seguida, executará mesclagem ao processar o conjunto de dados maior sem armazená-lo em cache ou ordená-lo. Geralmente você utilizará este tipo de junção com dados projetados utilizando um esquema de estrela ou um design semelhante, em que há uma grande tabela central de dados compartilhados (por exemplo, em dados transacionais). Se você selecionar essa opção, clique em **Selecionar** para especificar o conjunto de dados grande. Observe que é possível selecionar apenas *um* conjunto de dados grande. A tabela a seguir resume quais junções podem ser otimizadas utilizando este método.

Tabela 15. Sumarização das otimizações de junção.

Tipo de Junção	Pode ser otimizado para um conjunto de dados de entrada grande?
Interno	Sim
Parcial	Sim, se não houver registros incompletos no conjunto de dados grande.
Completa	Não
Antijunção	Sim, se o conjunto de dados grande for a primeira entrada.

Todas as entradas já estão ordenadas por um ou mais campos-chave. Selecione para indicar que os dados de entrada já estão ordenados por um ou mais campos-chave que você está usando para mesclagem. Assegure-se de que *todos* os seus conjuntos de dados de entrada sejam ordenados.

Especificar ordenação existente. Especifique os campos que já estiverem ordenados. Utilizando a caixa de diálogo Selecionar Campos, inclua campos na lista. É possível selecionar somente a partir dos campos-chave que estiverem sendo utilizados para a mesclagem (especificados na guia Mesclar). Na coluna *Ordem*, especifique se cada campo é classificado em ordem crescente ou decrescente. Se estiver especificando diversos campos, assegure-se de listá-los na ordenação correta. Use as setas para a direita da lista para organizar os campos na ordem correta. Se você errar ao especificar a ordenação existente correta, um erro aparecerá ao executar o fluxo, exibindo o número de registro no qual a ordenação está inconsistente com o que você especificou.

Dependendo da distinção entre maiúsculas e minúsculas do método de ordenação utilizado pelo banco de dados, a otimização poderá não funcionar corretamente quando uma ou mais entradas são ordenadas pelo banco de dados. Por exemplo, se você tiver duas entradas em que uma faz distinção entre maiúsculas e minúsculas e a outra não faz, os resultados da ordenação poderão ser diferentes. A otimização de mesclagem faz com que os registros sejam processados utilizando sua ordenação. Como resultado, se as entradas forem ordenadas utilizando métodos de ordenação diferentes, o nó Mesclagem relatará um erro e exibirá o número do registro no qual a ordenação está inconsistente. Quando todas as entradas forem de uma origem ou ordenadas utilizando ordenações mutuamente inclusivas, os registros poderão ser mesclados com sucesso.

Nota: a velocidade da mesclagem poderá ser beneficiada com a ativação do processamento paralelo.

Nó Anexação

É possível utilizar os nós Anexação para concatenar conjuntos de registros. Diferentemente dos nós Mesclagem, que juntam registros a partir de diferentes origens, os nós Anexação leem e transmitem posteriormente todos os registros de uma origem até que não haja mais nenhum deles. Em seguida, os registros da próxima origem são lidos utilizando a mesma estrutura de dados (número de registros, número de campos, e assim por diante) como a primeira entrada, ou entrada primária. Quando a origem primária tiver mais campos que a outra origem de entrada, a sequência de caracteres nula do sistema (\$null\$) será utilizada para quaisquer valores incompletos.

Os nós Anexação são úteis para combinar conjuntos de dados com estruturas semelhantes, mas dados diferentes. Por exemplo, você pode ter dados de transação armazenados em arquivos diferentes para períodos de tempo diferentes, como um arquivo de dados de vendas para março e um arquivo de dados de vendas separado para abril. Supondo que eles tenham a mesma estrutura (os mesmos campos na mesma ordem), o nó Anexação os juntará em um arquivo grande que, em seguida, poderá ser analisado.

Nota: para anexar arquivos, os níveis de medição de campo devem ser semelhantes. Por exemplo, um campo *Nominal* não pode ser anexado a um campo cujo nível de medição é *Contínuo*.

Configurando Opções de Anexação

Corresponder campos por. Selecione um método a ser utilizado quando corresponder campos para anexação.

- **Posição.** Selecione para anexar conjuntos de dados com base na posição dos campos na origem de dados principal. Ao utilizar esse método, os dados devem ser ordenados para assegurar a anexação adequada.
- **Nome.** Selecione para anexar conjuntos de dados com base no nome dos campos nos conjuntos de dados de entrada. Selecione também **Corresponder maiúsculas e minúsculas** para ativar a distinção entre maiúsculas e minúsculas ao corresponder nomes de campo.

Campo de saída. Lista os nós de origem que estiverem conectados ao nó Anexação. O primeiro nó na lista é a origem de entrada primária. É possível ordenar os campos na exibição clicando no título da coluna. Na realidade, essa ordenação não reordena os campos no conjunto de dados.

Incluir campos a partir de. Selecione **Somente conjunto de dados principal** para produzir campos de saída com base nos campos no conjunto de dados principal. O conjunto de dados principal é a primeira entrada, especificada na guia Entradas. Selecione **Todos os conjuntos de dados** para produzir campos de saída para todos os campos em todos os conjuntos de dados, não importa se houver um campo correspondente em todos os conjuntos de dados de entrada.

Identificar registros ao incluir um conjunto de dados de origem no campo. Selecione para incluir um campo adicional no arquivo de saída cujos valores indicam o conjunto de dados de origem para cada registro. Especifique o nome no campo de texto. O nome do campo padrão é *Entrada*.

Nó Distinto

Os registros duplicados em um conjunto de dados devem ser removidos antes de a mineração de dados poder iniciar. Por exemplo, em um banco de dados de marketing, as pessoas podem aparecer diversas vezes com informações de endereço ou de empresa diferentes. É possível utilizar o nó Distinto para localizar ou remover registros duplicados em seus dados ou criar um único registro composto a partir de um grupo de registros duplicados.

Para utilizar o nó Distinto, deve-se primeiro definir um conjunto de campos-chave que determinam quando dois registros são considerados duplicatas.

Se você não selecionar todos os campos como campos-chave, então dois registros "duplicados" poderão não ser realmente idênticos pois eles ainda poderão diferir nos valores dos campos restantes. Neste caso, também é possível definir uma ordenação que é aplicada dentro de cada grupo de registros duplicados. Esta ordenação fornece um controle preciso sobre qual registro é tratado como o primeiro dentro de um grupo. Caso contrário, todas as duplicatas serão consideradas como permutáveis e qualquer registro poderá ser selecionado. Como a ordem de chegada dos registros não é levada em conta, não é necessário o uso do nó Ordem de envio de dados (consulte "Ordenando registros dentro do nó distinto" abaixo).

Modo. Especifique se deseja criar um registro composto ou incluir ou excluir (descartar) o primeiro registro.

- **Criar um registro composto para cada grupo.** Fornece uma maneira de agregar campos não numéricos. Selecionar essa opção torna a guia Composto disponível na qual você especifica como criar os registros compostos. Consulte "Configurações Compostas Distintas" na página 95 para obter mais informações.
- **Incluir somente o primeiro registro em cada grupo.** Seleciona o primeiro registro de cada grupo de registros duplicados e descarta o resto. O *primeiro* registro é determinado pela ordenação definida a seguir, e não pela ordem de chegada dos registros.
- **Descartar somente o primeiro registro de cada grupo.** Descarta o primeiro registro de cada grupo de registros duplicados e seleciona o restante. O *primeiro* registro é determinado pela ordenação definida a seguir, e não pela ordem de chegada dos registros. Esta opção é útil para *localizar* duplicatas em seus dados para que elas possam ser examinadas posteriormente no fluxo.

Campos-chave para agrupamento. Lista um ou mais campos utilizados para determinar se os registros são idênticos. É possível:

- Incluir campos nessa lista usando o botão de seletor de campo à direita.
- Excluir campos da lista utilizando o botão X vermelho (remover).

Nos grupos, ordenar registros por. Lista os campos utilizados para determinar como os registros são ordenados dentro de cada grupo de duplicatas, e se eles são classificados em ordem crescente ou decrescente. É possível:

- Incluir campos nessa lista usando o botão de seletor de campo à direita.
- Excluir campos da lista utilizando o botão X vermelho (remover).
- Mover os campos utilizando os botões para cima ou para baixo, se estiver ordenando mais de um campo.

Uma ordenação deverá ser especificada se você tiver escolhido incluir ou excluir o primeiro registro em cada grupo, de modo que é importante saber qual registro é tratado como o primeiro.

Você também pode desejar especificar uma ordenação se tiver escolhido criar um registro composto para determinadas opções na guia Composto. Consulte “Configurações Compostas Distintas” na página 95 para obter mais informações.

Padrão de ordenação. Especifique se, por padrão, os registros são classificados em ordem **Crescente** ou **Decrescente** dos valores de chave de classificação.

Ordenando registros no nó Distinto

Se a ordem de registros dentro de um grupo de duplicatas for importante, então deve-se especificar a ordem utilizando a opção **Nos grupos, ordenar registros por** no nó Distinto. Não confie em um nó Ordem de envio de dados. Lembre-se de que a ordem de chegada dos registros não é levado em conta -- apenas a ordem especificada no nó.

Se nenhum campo de classificação for especificado (ou se especificar campos de classificação insuficientes), então os registros dentro de cada grupo de duplicatas estarão não ordenados (ou ordenados incompletamente) e os resultados poderão ser imprevisíveis.

Por exemplo, suponha que temos um conjunto muito grande de registros de log referentes a um número de máquinas. O log contém dados como os seguintes:

Tabela 16. Dados do log da máquina

Registro de data e hora	Máquina	Temperatura
17h00min22s	Máquina A	31
13h11min30s	Máquina B	26
16h49min59s	Máquina A	30
18h06min30s	Máquina X	32
16h17min33s	Máquina A	29
19h59min04s	Máquina C	35
19h20min55s	Máquina Y	34
15h36min14s	Máquina X	28
12h30min41s	Máquina Y	25
14h45min49s	Máquina C	27
19h42min00s	Máquina B	34
20h51min09s	Máquina Y	36
19h07min23s	Máquina X	33

Para reduzir o número de registros para o registro mais recente de cada máquina, utilize **Machine** como o campo-chave e use **Timestamp** como o campo de classificação (em ordem decrescente). A ordem de entrada não afeta o resultado, porque a seleção de ordenação especifica quais das muitas linhas de uma determinada Máquina devem ser retornadas, e a saída dos dados final seria a seguinte.

Tabela 17. Dados do log da máquina ordenados

Registro de data e hora	Máquina	Temperatura
17h00min22s	Máquina A	31
19h42min00s	Máquina B	34
19h59min04s	Máquina C	35

Tabela 17. Dados do log da máquina ordenados (continuação)

Registro de data e hora	Máquina	Temperatura
19h07min23s	Máquina X	33
20h51min09s	Máquina Y	36

Configurações de Otimização Distintas

Se os dados nos quais você estiver trabalhando tiverem somente um número pequeno de registros ou se eles já tiverem sido ordenados, será possível otimizar a maneira de manipular esses dados para que o IBM SPSS Modeler possa processá-los com mais eficiência.

Nota: se você selecionar **Dados de entrada possuem um número baixo de chaves distintas** ou se utilizar a geração de SQL para o nó, qualquer linha dentro do valor de chave distinto poderá ser retornada; para controlar qual linha será retornada dentro de uma chave distinta, é necessário especificar a ordenação utilizando os campos **Nos grupos, ordenar registros por** na guia Configurações. As opções de otimização não afetarão a saída dos resultados pelo nó Distinto desde que uma ordenação tenha sido especificada na guia Configurações.

O conjunto de dados de entrada possui um número baixo de chaves distintas. Selecione esta opção se você tiver um número pequeno de registros ou um número pequeno de valores exclusivos de um ou mais campos-chave, ou ambos. Fazer isso poderá melhorar o desempenho.

O conjunto de dados de entrada já está ordenado por campos de agrupamento e campos de ordenação na guia Configurações. Selecione essa opção somente se seus dados já estiverem ordenados por todos os campos listados em **Nos grupos, ordenar registros por** na guia Configurações, e se a ordenação crescente ou decrescente dos dados for a mesma. Fazer isso poderá melhorar o desempenho.

Desativar geração de SQL. Selecione esta opção para desativar a geração de SQL para o nó.

Configurações Compostas Distintas

Se os dados nos quais você estiver trabalhando tiverem diversos registros, por exemplo, para a mesma pessoa, será possível otimizar a maneira como os dados são manipulados ao criar um registro composto ou agregado único para processar. Se você tiver o IBM SPSS Modeler Entity Analytics instalado, essa opção também poderá ser usada para combinar ou comprimir os registros duplicados que são a saída a partir do SPSS Entity Analytics.

Nota: Esta guia estará disponível apenas quando selecionar **Criar um registro composto para cada grupo** na guia Configurações.

Como exemplo, suponha que o SPSS Entity Analytics marque três registros como sendo a mesma entidade, conforme mostrado na tabela a seguir.

Tabela 18. Exemplo de diversos registros para a mesma entidade.

\$EA-ID	Nome	Idade	Banco	Nível de Escolaridade	Total da Dívida
0003	Bob Jones	27	K	Médio	27000
0003	Robert Jones	35	N	Superior	42000
0003	Robbie Jones	27	D	PhD	7000

Nossa intenção é agregar esses três em um único registro que, em seguida, utilizaremos mais tarde. É possível utilizar o nó Agregado para incluir o total da dívida e calcular uma média de idade; no entanto,

não é possível calcular a média de detalhes como nomes, bancos, e assim por diante. Se não especificarmos quais detalhes iremos utilizar para criar um registro composto, poderemos derivar um único registro.

Na nossa tabela podemos criar um registro composto ao escolher os seguintes detalhes.

- Para **Nome**, utilize o Primeiro registro
- Para **Idade**, selecione a Mais alta
- Para o **Banco**, concatene todos os valores sem delimitador
- Para **Nível de Escolaridade**, selecione o Primeiro localizado na lista (Faculdade de PhD)
- Para **Dívida**, selecione o Total

Ao combinar (ou agregar) esses detalhes, terminamos com um único registro composto que contém os seguintes detalhes.

- Nome: Bob Jones
- Idade: 35
- Banco: KND
- Nível de Escolaridade: PhD
- Dívida: 76000

Isso fornece um melhor perfil de Bob Jones, uma pessoa de 35 anos de idade formada em PhD, que possui pelo menos três contas bancárias conhecidas e uma dívida total grande.

Configurando opções para a guia Composta

Campo. Esta coluna mostra todos os campos, exceto os campos-chave no modelo de dados, em sua ordenação natural; se o nó não estiver conectado, nenhum campo será mostrado. Para classificar as linhas em ordem alfabética por nome do campo, clique no cabeçalho da coluna. É possível selecionar mais de uma linha usando Shift-clique ou Ctrl-clique. Além disso, se você clicar com o botão direito em um campo, um menu será exibido a partir do qual é possível escolher selecionar todas as linhas, classificar as linhas em ordem crescente ou decrescente de nome do campo ou de valor, selecionar campos por tipo de medida ou de armazenamento ou selecionar um valor para incluir automaticamente a mesma entrada **Preencher com valores baseados em** para cada linha selecionada.

Preencher com valores baseados em. Selecione o tipo de valor a ser utilizado para o registro composto para o **Campo**. As opções disponíveis dependem do tipo de campo.

- Para campos de intervalo numérico, é possível escolher a partir de:
 - Primeiro registro no grupo
 - Último registro no grupo
 - Total
 - Média
 - Mínimo
 - Máximo
 - Custom
- Para campos de data e hora, é possível escolher a partir de:
 - Primeiro registro no grupo
 - Último registro no grupo
 - Mais Antigo
 - Mais Recente
 - Custom
- Para campos de sequência de caracteres e sem tipo, é possível escolher a partir de:

- Primeiro registro no grupo
- Último registro no grupo
- Alfanumérico primeiro
- Alfanumérico por último
- Custom

Em cada caso, é possível utilizar a opção **Customizado** para exercer mais controle sobre qual valor será usado para preencher o registro composto. Consulte “Distinto Composto - Guia Customizado” para obter mais informações.

Incluir contagem de registro no campo. Selecione esta opção para incluir um campo extra em cada registro de saída, denominado Record_Count por padrão. Esse campo indica quantos registros de entrada foram agregados para formar cada registro agregado. Para criar um nome customizado para este campo, digite sua entrada no campo de edição.

Distinto Composto - Guia Customizado

A caixa de diálogo Preenchimento Customizado fornece mais controle sobre qual valor é utilizado para concluir o novo registro composto. Observe que primeiro é necessário instanciar seus dados usando essa opção se customizar apenas uma única linha de Campo na guia Composto.

Nota: Esta caixa de diálogo está disponível apenas quando selecionar o valor Customizado na coluna **Preencher com valores baseados em** na guia Composto.

Dependendo do tipo de campo, é possível escolher a partir de uma das seguintes opções.

- **Selecionar por frequência.** Escolha um valor com base na frequência com a qual ele ocorre no registro de dados.

Nota: Não disponível para Campos com um tipo de Contínuo, Sem Tipo ou Data/Hora.

- **Usar.** Selecione a partir do Mais ou Menos Frequente.
- **Empates.** Se houver dois ou mais registros com a mesma frequência de ocorrência, especifique como selecionar o registro necessário. É possível escolher a partir de uma das quatro opções: Usar Primeiro, Usar Último, Usar Mais Baixo ou Usar Mais Alto.
- **Inclui valor (T/F).** Selecione esta opção para converter um campo para um flag que identifica se qualquer um dos registros em um grupo tem um valor especificado. Em seguida, é possível selecionar o **Valor** na lista de valores para o campo selecionado.

Nota: Não disponível se você selecionar mais de uma linha de Campo na guia Composto

- **Primeira correspondência na lista.** Selecione esta opção para priorizar qual valor fornecer ao registro composto. Em seguida, é possível selecionar um dos **Itens** na lista de valores para o campo selecionado.

Nota: Não disponível se você selecionar mais de uma linha de Campo na guia Composto

- **Concatenar valores.** Selecione esta opção para reter todos os valores em um grupo ao concatená-los em uma sequência de caracteres. Deve-se especificar um delimitador a ser utilizado entre cada valor.

Nota: Esta será a única opção disponível se você selecionar uma ou mais linhas de Campo com um tipo de Contínuo, Sem Tipo ou Data/Hora.

- **Usar delimitador.** É possível optar por utilizar um **Espaço** ou uma **Vírgula** como um valor de delimitador na sequência de caracteres concatenada. Como alternativa, no campo **Outro**, é possível inserir seu próprio caractere de valor de delimitador.

Nota: Disponível apenas se você selecionar a opção **Concatenar valores**.

Nó TS de Fluxo

O nó Séries Temporais de Fluxo (TS) pode ser utilizado para construir e escorar os modelos de séries temporais em um passo, e pode ser utilizado com ou sem um nó Intervalos de Tempo de envio de dados. Um modelo de série temporal separado é criado para cada campo de destino, no entanto, nuggets do modelo não são incluídos na paleta de modelos gerados e as informações de modelo não podem ser procuradas.

Nem sempre é necessário usar o nó Intervalos de Tempo com o nó TS de Fluxo, desde que os pontos a seguir sejam satisfeitos:

- Os dados estão espaçados uniformemente.
- Os dados têm rótulos de séries temporais.
- Nenhum registro adicional precisa ser incluído.
- Os dados possuem uma *periodicidade* padrão, por exemplo, cinco dias por semana ou oito horas por dia.

Os métodos para modelar dados de séries temporais requerem um intervalo uniforme entre cada medição, com quaisquer valores omissos indicados por linhas vazias. Se seus dados ainda não atenderem a esse requisito, um nó Intervalos de Tempo de envio de dados poderá transformar valores conforme necessário. Consulte o tópico “Nó Intervalos de Tempo” na página 176 para obter mais informações.

Outros pontos a serem observados com relação aos nós Séries Temporais são:

- Os campos devem ser numéricos.
- Os campos de data não podem ser utilizados como entradas.
- As partições são ignoradas.

O nó TS de Fluxo estima modelos de suavização exponencial, Média Móvel Integrada Autorregressiva (ARIMA) univariada e ARIMA multivariada (ou função de transferência) para séries temporais e produz previsões com base nos dados de séries temporais. Também está disponível um Modelador Especialista, que tenta identificar e estimar automaticamente o modelo ARIMA ou de suavização exponencial de melhor ajuste para uma ou mais variáveis de destino. Para obter mais informações sobre os modelos de séries temporais em geral, consulte o tópico *Modelos de Séries Temporais* no documento *Nós de Modelagem do IBM SPSS Modeler 17.1*.

O nó TS de Fluxo é suportado para uso em um ambiente de implementação de fluxo, por meio do IBM SPSS Modeler Solution Publisher, utilizando o IBM SPSS Collaboration and Deployment Services Scoring Service ou o IBM InfoSphere Warehouse.

Opções do Campo TS de Fluxo

A guia Campos é onde você especifica os campos a serem utilizados na construção do modelo. Antes de poder construir um modelo, é necessário especificar quais campos você deseja utilizar como destinos e como entradas. Geralmente, o nó TS de Fluxo utiliza as informações do campo a partir de um nó Tipo de envio de dados. Se você estiver utilizando um nó Tipo para selecionar campos de entrada e de destino, não será necessário alterar nada nesta guia.

Usar configurações do nó de tipo. Essa opção instrui o nó a utilizar as informações de campo a partir de um nó Tipo de envio de dados. Esse é o padrão.

Usar configurações customizadas. Essa opção instrui o nó a utilizar as informações de campo especificadas aqui ao invés das informações fornecidas em qualquer nó ou nós Tipo de envio de dados. Após selecionar esta opção, especifique os campos abaixo. Observe que os campos armazenados como datas não são aceitos como campos de destino ou de entrada.

- **Destinos.** Selecione um ou mais campos de destino. Isso é semelhante a configurar o papel do campo para *Destino* em um nó Tipo. Os campos de destino para um modelo de série temporal devem ter um nível de medição de *Contínuo*. Um modelo separado é criado para cada campo de destino. Um campo de destino considera todos os campos de *Entrada* especificados, exceto si mesmo, como entradas possíveis. Assim, o mesmo campo poderá ser incluído em ambas as listas; esse campo será utilizado como uma entrada possível para todos os modelos, exceto aquele no qual ele é um destino.
- **Entradas.** Selecione um ou mais campos de entrada. Isso é semelhante a configurar o papel do campo para *Entrada* em um nó Tipo. Os campos de entrada para um modelo de série temporal devem ser numéricos.

Opções de Modelo do TS de Fluxo

Método. É possível escolher Modelador Especialista, Suavização Exponencial ou ARIMA. Selecione **Crítérios** para especificar opções para o método selecionado.

- **Modelador Especialista.** Escolha esta opção para utilizar o Modelador Especialista, que localiza automaticamente o modelo de melhor ajuste para cada série.
- **Suavização Exponencial.** Utilize esta opção para especificar um modelo de suavização exponencial customizado.
- **ARIMA.** Utilize esta opção para especificar um modelo ARIMA customizado.

Derivar campos para limites de confiança superiores e inferiores. Selecione esta opção para gerar campos para os intervalos de confiança para as previsões do modelo e autocorrelações residuais. Quando selecionada, o controle **Largura do limite de confiança (%)** é ativado.

Largura do limite de confiança (%). Especifique os intervalos de confiança que são calculados. É possível especificar quaisquer valores positivos menores que 100. Por padrão, um intervalo de confiança de 95% é utilizado.

Opções de Intervalos de Tempo

Nesta seção da caixa de diálogo, é possível escolher utilizar as especificações para as estimativas e previsões de um nó Intervalos de Tempo de envio de dados ou especificar essas configurações no nó TS do Fluxo.

Usar configurações a partir do nó Intervalos de Tempo. Selecione essa opção para usar informações sobre especificações para estimativas e previsões que são feitas em um nó Intervalos de Tempo de envio de dados.

A primeira linha das informações indica se quaisquer registros são excluídos do modelo ou utilizados como validação.

A segunda linha fornece informações sobre quaisquer períodos de previsão especificados no nó Intervalos de Tempo.

Se a primeira linha lê **Nenhum intervalo de tempo definido**, o que indica que nenhum nó Intervalos de Tempo está conectado. Esta situação causará um erro ao tentar executar o fluxo. Se você selecionar **Usar configurações do nó Intervalos de Tempo**, deve-se incluir um nó Intervalos de Tempo acima do nó TS do Fluxo.

Especificar configurações. Selecione esta opção para especificar as configurações para as estimativas e previsões se não houver nenhum nó Intervalos de Tempo acima do nó de TS do Fluxo.

Estimar até. Selecione **Última observação válida** para estimar o modelo utilizando todos os registros até a última observação válida. Os registros que são utilizados para previsão iniciam a partir da primeira observação que não for válida. As observações *Válidas* são aquelas em que todos os campos de destino

possuem valores que não são nulos. Selecione **Contar de volta a partir do último registro** para reservar um número de registros mais recentes para previsão. Os registros mais recentes podem ter valores para os campos de destino, embora isso não seja necessário. Se **Contar de volta a partir do último registro** for selecionada, o controle **Offset** será ativado, permitindo especificar o número de registros a serem utilizados somente para previsão. O número padrão de registros é quatro, e o máximo é 99.

Nota: Os registros que devem ser utilizados para previsão devem ter valores para todos os campos de entrada.

Intervalo de tempo. Especifique o intervalo de tempo usado para estimação e previsão. As opções a seguir estão disponíveis: Anos, Trimestres, Meses, Semanas, Dias, Horas, Minutos e Segundos.

Critérios do Modelador Especialista de Séries Temporais

Tipo de modelo. As opções a seguir estão disponíveis:

- **Todos os modelos.** O Modelador Especialista considera os modelos de suavização ARIMA e exponencial.
- **Apenas modelos de suavização exponencial.** O Modelador Especialista considera apenas os modelos de suavização exponencial.
- **Apenas modelos ARIMA.** O Modelador Especialista considera apenas os modelos ARIMA.

Modelador Especialista considera modelos sazonais. Essa opção será ativada apenas se uma periodicidade tiver sido definida para o conjunto de dados ativo. Quando essa opção é selecionada, o Modelador Especialista considera modelos sazonais e não sazonais. Se essa opção não estiver selecionada, o Modelador Especialista considerará apenas modelos não sazonais.

Eventos e Intervenções. Permite designar determinados campos de entrada como campos de evento ou de intervenção. Fazer isso identifica um campo como contendo dados de séries temporais afetados por eventos (situações recorrentes previsíveis, por exemplo, promoções de vendas) ou intervenções (incidentes únicos, por exemplo, indisponibilidade de energia ou greve de funcionário). O Modelador Especialista não considera funções de transferência arbitrárias para entradas identificadas como campos de evento ou de intervenção.

Os campos de entrada devem ter um nível de medição de *Flag*, *Nominal* ou *Ordinal* e ser numéricos (por exemplo, 1/0, não True/False, para um campo de flag), antes de serem incluídos nessa lista.

Valores discrepantes

Detectar valores discrepantes automaticamente. Por padrão, a detecção automática de valores discrepantes não é executada. Selecione esta opção para executar detecção automática de valores discrepantes e, em seguida, selecione os tipos de valores discrepantes desejados. Consulte o tópico “Manipulando Valores Discrepantes” na página 104 para obter mais informações.

Critérios de Suavização Exponencial de Séries Temporais

Tipo de modelo. Os modelos de suavização exponencial são classificados como ¹. Os modelos sazonais estarão disponíveis apenas se a periodicidade for sazonal. As periodicidades sazonais são: períodos cíclicos, anos, trimestres, meses, dias por semana, horas por dia, minutos por dia, e segundos por dia.

- **Simples.** Esse modelo é apropriado para uma série na qual não houver nenhuma tendência ou sazonalidade. O único parâmetro de suavização relevante é nível. A suavização exponencial simples é mais semelhante a um ARIMA com zero ordem de autorregressão, uma ordem de diferenciação, uma ordem de média de movimentação e nenhuma constante.

1. Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1-28.sazonais ou não sazonais

- **Tendência linear de Holt.** Esse modelo é apropriado para uma série na qual houver uma tendência linear e nenhuma sazonalidade. Seus parâmetros de suavização relevantes são nível e tendência e, nesse modelo, eles não são restritos pelos valores uns dos outros. O modelo de Holt é mais geral do que o modelo de Brown, mas poderá demorar mais para calcular estimativas para grandes séries. A suavização exponencial de Holt é mais semelhante a um ARIMA com zero ordem de autorregressão, duas ordens de diferenciação e duas ordens de média móvel.
- **Tendência linear de Brown.** Esse modelo é apropriado para uma série na qual houver uma tendência linear e nenhuma sazonalidade. Seus parâmetros de suavização relevantes são nível e tendência, no entanto, nesse modelo, eles são considerados iguais. O modelo de Brown é, portanto, um caso especial do modelo de Holt. A suavização exponencial de Brown é mais semelhante a um ARIMA com zero ordem de autorregressão, duas ordens de diferenciação e duas ordens de média móvel, com o coeficiente para a segunda ordem de média móvel igual à metade do coeficiente da primeira ordem ao quadrado.
- **Tendência amortecida.** Esse modelo é apropriado para uma série na qual houver uma tendência linear que está desaparecendo e nenhuma sazonalidade. Seus parâmetros de suavização relevantes são nível, tendência e tendência de amortecimento. A suavização exponencial amortecida é mais semelhante a um ARIMA com uma ordem de autorregressão, uma ordem da diferenciação e duas ordens de média móvel.
- **Sazonal simples.** Esse modelo é apropriado para uma série sem nenhuma tendência e com um efeito sazonal que seja constante ao longo do tempo. Seus parâmetros de suavização relevantes são nível e season. A suavização exponencial sazonal é mais semelhante a um ARIMA com zero ordem de autorregressão, uma ordem de diferenciação, uma ordem da diferenciação sazonal e as ordens 1, p e $p+1$ de média móvel, em que p é o número de períodos em um intervalo sazonal. Para dados mensais, $p=12$.
- **Aditiva de Winter.** Esse modelo é apropriado para uma série na qual houver uma tendência linear e um efeito sazonal que seja constante ao longo do tempo. Seus parâmetros de suavização relevantes são nível, tendência e season. A suavização exponencial aditiva de Winter é muito semelhante a um ARIMA com zero ordem de autorregressão, uma ordem de diferenciação; uma ordem de diferenciação sazonal e $p+1$ ordens de média móvel, em que p é o número de períodos em um intervalo sazonal. Para dados mensais, $p=12$.
- **Multiplicativa de Winter.** Esse modelo é apropriado para uma série na qual houver uma tendência linear e um efeito sazonal que altera com a magnitude da série. Seus parâmetros de suavização relevantes são nível, tendência e season. A suavização exponencial multiplicativa de Winter não é semelhante a nenhum modelo ARIMA.

Transformação de Resposta. É possível especificar uma transformação para ser executada em cada variável dependente antes de ser modelada. Consulte o tópico *Transformações de Séries* no documento *Nós de Modelagem do IBM SPSS Modeler 17.1* para obter mais informações.

- **Nenhum.** Nenhuma transformação é executada.
- **Raiz quadrada.** A transformação raiz quadrada é executada.
- **Logarítmica natural.** A transformação logarítmica natural é executada.

Critérios do ARIMA de Séries Temporais

O nó Séries Temporais permite construir modelos ARIMA sazonais ou não sazonais customizados, também conhecidos como modelos Box-Jenkins, com ou sem um conjunto fixo de variáveis de entrada (preditoras) ². É possível definir funções de transferência para qualquer uma ou todas variáveis de entrada e especificar detecção automática de valores discrepantes ou de um conjunto explícito de valores discrepantes.

2. Box, G. E. P., G. M. Jenkins, e G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3ª ed. Englewood Cliffs, N.J.: Prentice Hall.

Todas as variáveis de entrada especificadas são explicitamente incluídas no modelo. Isso está em contraste com o uso do Modelador Especialista, em que as variáveis de entrada serão incluídas apenas se elas possuírem um relacionamento estatisticamente significativo com a variável de destino.

Modelo

A guia Modelo permite especificar a estrutura de um modelo ARIMA customizado.

Ordens ARIMA. Insira valores para os vários componentes ARIMA de seu modelo nas células correspondentes da grade de Estrutura. Todos os valores devem ser números inteiros não negativos. Para componentes autorregressivos e de média móvel, o valor representa a ordem máxima. Todas as ordens inferiores positivas serão incluídas no modelo. Por exemplo, se você especificar 2, o modelo incluirá as ordens 2 e 1. As células na coluna Sazonal serão ativadas apenas se uma periodicidade tiver sido definida para o conjunto de dados ativo.

- **Autorregressiva (p).** O número de ordens autorregressivas no modelo. As ordens autorregressivas especificam quais valores anteriores da série são utilizados para prever valores atuais. Por exemplo, uma ordem autorregressiva de 2 especifica que o valor de dois períodos de tempo da série no passado será utilizado para prever o valor atual.
- **Diferença (d).** Especifica a ordem de diferenciação aplicada à série antes de estimar os modelos. A diferenciação é necessária quando tendências estiverem presentes (as séries com tendências normalmente são não estacionárias e a modelagem ARIMA assume estacionariedade) e é utilizada para remover seus efeitos. A ordem da diferenciação corresponde ao grau de tendência das séries -- a diferenciação de primeira ordem considera tendências lineares, a diferenciação de segunda ordem considera tendências quadráticas, e assim por diante.
- **Média Móvel (q).** O número de ordens de média móvel no modelo. As ordens de média móvel especificam como os desvios da média de série para valores anteriores são utilizados para prever valores atuais. Por exemplo, as ordens de média móvel de 1 e 2 especificam que os desvios do valor médio das séries de cada um dos dois últimos períodos de tempo são considerados ao prever valores atuais da série.

Ordens Sazonais. Os componentes autorregressivo, média móvel e de diferenciação sazonais desempenham os mesmos papéis que seus correspondentes não sazonais. Para ordens sazonais, no entanto, os valores atuais da série são afetados pelos valores anteriores da série separados por um ou mais períodos sazonais. Por exemplo, para dados mensais (período de sazonal de 12), uma ordem sazonal de 1 significa que o valor de série atual é afetado pelo valor de série 12 períodos anteriores ao período atual. Em seguida, uma ordem sazonal de 1, para os dados mensais, será o mesmo que especificar uma ordem não sazonal de 12.

Transformação de Resposta. É possível especificar uma transformação para ser executada em cada variável de destino antes de ser modelada. Consulte o tópico *Transformação de Série* no documento *Modelando Nós do IBM SPSS Modeler 17.1* para obter mais informações.

- **Nenhum.** Nenhuma transformação é executada.
- **Raiz quadrada.** A transformação raiz quadrada é executada.
- **Logarítmica natural.** A transformação logarítmica natural é executada.

Incluir constante no modelo. A inclusão de uma constante é padrão, a menos que você tenha certeza de que o valor de série médio geral é 0. Excluir a constante é recomendado quando a diferenciação é aplicada.

Transferir Funções

A guia Funções de Transferência permite definir funções de transferência para qualquer um ou todos os campos de entrada. As funções de transferência permitem especificar a maneira pela qual os valores passados destes campos são utilizados para prever valores futuros da série de resposta.

A guia será exibida apenas se os campos de entrada (com o papel configurado como *Entrada*) forem especificados, seja no nó Tipo ou na guia Campos do nó TS de Fluxo (selecione **Usar configurações customizadas – Entradas**).

A lista superior mostra todos os campos de entrada. As informações restantes nesta caixa de diálogo são específicas para o campo de entrada selecionado na lista.

Ordens de Função de Transferência. Insira valores para os vários componentes da função de transferência nas células correspondentes da grade Estrutura. Todos os valores devem ser números inteiros não negativos. Para componentes de numerador e denominador, o valor representa a ordem máxima. Todas as ordens inferiores positivas serão incluídas no modelo. Além disso, a ordem 0 é sempre incluída para componentes de numerador. Por exemplo, se você especificar 2 para o numerador, o modelo incluirá as ordens 2, 1 e 0. Se você especificar 3 para o denominador, o modelo incluirá as ordens 3, 2, e 1. As células na coluna Sazonal serão ativadas apenas se uma periodicidade tiver sido definida para o conjunto de dados ativo.

Numerador. A ordem de numerador da função de transferência especifica quais valores anteriores da série independente (preditora) selecionada são utilizados para prever valores atuais da série dependente. Por exemplo, uma ordem de numerador de 1 especifica que o valor de um período de série independente no passado -- assim como o valor atual da série independente -- é utilizado para prever o valor atual de cada série dependente.

Denominador. A ordem de denominador da função de transferência especifica como os desvios da média de série, para os valores anteriores da série independente (preditora) selecionada, são utilizados para prever os valores atuais da série dependente. Por exemplo, uma ordem de denominador de 1 especifica que os desvios do valor médio de um período de série independente no passado são considerados ao prever o valor atual de cada série dependente.

Diferença. Especifica a ordem de diferenciação aplicada à série independente (preditora) selecionada antes de estimar os modelos. A diferenciação é necessária quando tendências estiverem presentes e é utilizada para remover seu efeito.

Ordens Sazonais. Os componentes de numerador, denominador e diferenciação sazonais desempenham os mesmos papéis que seus correspondentes não sazonais. Para ordens sazonais, no entanto, os valores atuais da série são afetados pelos valores anteriores da série separados por um ou mais períodos sazonais. Por exemplo, para dados mensais (período de sazonal de 12), uma ordem sazonal de 1 significa que o valor de série atual é afetado pelo valor de série 12 períodos anteriores ao período atual. Em seguida, uma ordem sazonal de 1, para os dados mensais, será o mesmo que especificar uma ordem não sazonal de 12.

Atraso. Configurar um atraso faz com que a influência do campo de entrada seja atrasada pelo número de intervalos especificados. Por exemplo, se o atraso for configurado como 5, o valor do campo de entrada no tempo t não afetará as previsões até que decorram cinco períodos ($t + 5$).

Transformação. A especificação de uma função de transferência para um conjunto de variáveis independentes também inclui uma transformação opcional a ser executada nessas variáveis.

- **Nenhum.** Nenhuma transformação é executada.
- **Raiz quadrada.** A transformação raiz quadrada é executada.
- **Logarítmica natural.** A transformação logarítmica natural é executada.

Manipulando Valores Discrepantes

A guia Valores Discrepantes fornece um número de opções para manipular valores discrepantes nos dados³.

Não detectar valores discrepantes ou modelá-los. Por padrão, os valores discrepantes não são detectados nem modelados. Selecione esta opção para desativar qualquer detecção ou modelagem de valores discrepantes.

Detectar valores discrepantes automaticamente. Selecione esta opção para executar detecção automática de valores discrepantes, e selecione um ou mais dos tipos de valores discrepantes mostrados.

Tipo de Valores Discrepantes para Detectar. Selecione um ou mais tipos de valores discrepantes que deseja detectar. Os tipos suportados são:

- Aditivo (padrão)
- Mudança de nível (padrão)
- Inovador
- Transiente
- Aditivo sazonal
- Tendência local
- Curva de nível aditiva

Consulte o tópico *Valores Discrepantes* no documento *Nós de Modelagem do IBM SPSS Modeler 17.1* para obter mais informações.

Opções de Implementação do TS de Fluxo

Se estiver usando um nó TS de Fluxo em um ambiente de implementação de fluxo, por meio do IBM SPSS Modeler Solution Publisher, por padrão, sempre que um modelo de série temporal for usado para escorar registros, o modelo é reconstruído e novas previsões são feitas. É possível alterar esse comportamento padrão utilizando as opções na guia Implementação para especificar quando os modelos de série temporal são reconstruídos.

Nota: As configurações na guia Implementação são ignoradas para uso normal do nó TS de Fluxo no IBM SPSS Modeler.

Forçar reconstrução completa quando necessário (pode alterar o tipo de modelo). Selecione esta opção se desejar que os modelos de série temporal sejam reconstruídos apenas se houver um número especificado de novos registros. A reconstrução do modelo para cada terceiro ou quarto novo registro ao invés de para cada novo registro pode melhorar o desempenho. Selecionar essa opção ativa os controles do **Número mínimo de novos registros para que a reconstrução seja necessária**.

Número mínimo de novos registros para que a reconstrução seja necessária. É possível escolher quando os modelos de série temporal são reconstruídos com base no número de novos registros. Selecione uma das seguintes opções:

- Se **Usar porcentagem de registros (%)** for selecionada, os modelos de série temporal serão reconstruídos quando uma porcentagem especificada do número total de registros for de novos registros. A porcentagem padrão é 10%.
- Se **Usar número absoluto de registros** for selecionada, o modelo de série temporal será reconstruído quando um número especificado de registros for de novos registros. O número mínimo de novos registros que podem ser especificados para o modelo ser reconstruído é um e o máximo é 100. O número padrão de registros é quatro.

3. Pena, D., G. C. Tiao, e R. S. Tsay, eds. 2001. *A course in time series analysis*. Nova York: John Wiley and Sons.

Campo para identificar novos registros. Selecione o campo que é observado para detectar quando há novos registros. O campo que é escolhido deve ser aquele que ordena os registros na série temporal. Apenas campos contínuos podem ser escolhidos.

Nó TCM de Fluxo

O nó TCM de Fluxo pode ser usado para construir e escorar modelos casuais temporais em um passo.

Para obter mais informações sobre a modelagem causal temporal, consulte o tópico Modelos Causais Temporais na seção Modelos de Séries Temporais do guia Nós de Modelagem do SPSS Modeler.

Nó TCM de Fluxo - Opções de Séries Temporais

Na guia Campos, utilize as configurações de **Séries Temporais** para especificar as séries a serem incluídas no sistema de modelo.

Selecione a opção para a estrutura de dados que se aplica aos seus dados. Para dados multidimensionais, clique em **Selecionar Dimensões** para especificar os campos de dimensão. A ordem especificada dos campos dimensão define a ordem na qual eles aparecem em todos os diálogos e saída subsequentes. Utilize os botões de seta para cima e para baixo na subdiálogo Selecionar Dimensões para reordenar os campos de dimensão.

Para dados baseados em coluna, o termo *série* tem o mesmo significado que o termo *campo*. Para dados multidimensionais, os campos que contiverem séries temporais são referidos como campos de *métrica*. Uma série temporal para dados multidimensionais é definida por um campo de métrica e por um valor para cada um dos campos da dimensão. As considerações a seguir se aplicam a ambos dados dimensionais e baseados em coluna.

- As séries que forem especificadas como entradas candidatas ou como destino e entrada são consideradas para inclusão no modelo para cada destino. O modelo para cada destino sempre inclui valores em lag do destino em si.
- As séries que forem especificadas como entradas forçadas são sempre incluídas no modelo para cada destino.
- Pelo menos uma série deve ser especificada como um destino ou como destino e entrada.
- Quando **Usar papéis predefinidos** é selecionada, os campos que tiverem um papel de Entrada são configurados como entradas candidatas. Nenhum papel predefinido é mapeado para uma entrada forçada.

Dados multidimensionais

Para dados multidimensionais, especifique os campos de métrica e papéis associados em uma grade, em que cada linha na grade especifica uma única métrica e papel. Por padrão, o sistema de modelo inclui séries de todas as combinações dos campos de dimensão para cada linha na grade. Por exemplo, se houver dimensões para *region* e *brand*, então, por padrão, especificar a métrica *sales* como uma resposta significa que há uma série de respostas de vendas separadas para cada combinação de *region* e *brand*.

Para cada linha na grade, é possível customizar o conjunto de valores para qualquer um dos campos de dimensão clicando no botão de reticências para uma dimensão. Esta ação abre o subdiálogo Selecionar Valores de Dimensão. Também é possível incluir, excluir ou copiar linhas da grade.

A coluna **Contagem de Séries** exibe o número de conjuntos de valores de dimensão que estiverem atualmente especificados para a métrica associada. O valor exibido pode ser maior que o número real de séries (uma série por conjunto). Essa condição ocorre quando algumas das combinações especificadas de valores de dimensão não correspondem à série contida pela métrica associada.

Nó TCM de fluxo - selecionar valores da dimensão

Para dados multidimensionais, é possível customizar a análise ao especificar quais valores de dimensão se aplicam a um campo de métrica específico com um determinado papel. Por exemplo, se *sales* for um campo de métrica e *channel* for uma dimensão com valores 'retail' e 'web', será possível especificar que as vendas da 'web' representam uma entrada e as vendas 'retail' representam um destino.

Todos os valores

Especifica que todos os valores do campo de dimensão atual são incluídos. Esta opção é a padrão.

Selecionar valores para incluir ou excluir

Utilize esta opção para especificar o conjunto de valores para o campo de dimensão atual. Quando **Incluir** é selecionado para o **Modo**, apenas os valores que forem especificados na lista **Valores Selecionados** são incluídos. Quando **Excluir** é selecionado para o **Modo**, todos os valores diferentes dos valores que são especificados na lista **Valores Selecionados** são incluídos.

É possível filtrar o conjunto de valores a partir do qual escolher. Os valores que atenderem à condição de filtro aparecem na guia **Correspondido** e os valores que não atenderem à condição de filtro aparecem na guia **Não Correspondido** da lista **Valores Desmarcados**. A guia **Tudo** lista todos os valores desmarcados, independentemente de qualquer condição do filtro.

- É possível utilizar asteriscos (*) para indicar caracteres curingas quando especificar um filtro.
- Para limpar o filtro atual, especifique um valor vazio para o termo de procura no diálogo Filtrar Valores Exibidos.

Nó TCM de Fluxo - Opções de Observações

Na guia Campos, utilize as configurações de **Observações** para especificar os campos que definem as observações.

Observações que são definidas pelas datas/horas

É possível especificar que as observações sejam definidas por um campo de data, de hora ou de registro de data e hora. Além do campo que define as observações, selecione o intervalo de tempo apropriado que descreve as observações. Dependendo do intervalo de tempo especificado, também é possível especificar outras configurações, como o intervalo entre as observações (incremento) ou o número de dias por semana. As considerações a seguir se aplicam ao intervalo de tempo:

- Utilize o valor **Irregular** quando as observações estiverem espaçadas irregularmente no tempo, como o horário no qual uma ordem de vendas é processada. Quando **Irregular** é selecionada, deve-se especificar o intervalo de tempo que é usado para a análise, a partir das configurações de **Intervalo de Tempo** na guia Especificações de Dados.
- Quando as observações representam uma data e hora e o intervalo de tempo é horas, minutos ou segundos, então utilize **Horas por dia**, **Minutos por dia** ou **Segundos por dia**. Quando as observações representam um tempo (duração) sem referência a uma data e o intervalo de tempo é horas, minutos ou segundos, então utilize **Horas (não periódico)**, **Minutos (não periódico)** ou **Segundos (não periódico)**.
- Com base no intervalo de tempo selecionado, o procedimento poderá detectar observações omissas. Detectar observações omissas é necessário porque o procedimento supõe que todas as observações estão igualmente espaçadas no tempo e que não há observações omissas. Por exemplo, se o intervalo de tempo for Dias e a data 2014-10-27 for seguida por 2014-10-29, então há uma observação omissa para 2014-10-28. Os valores são imputados para quaisquer observações omissas. Configurações para manipular valores omissos podem ser especificadas a partir da guia Especificações de Dados.
- O intervalo de tempo especificado permite que o procedimento detecte diversas observações no mesmo intervalo de tempo que precisam ser agregadas e alinhe essas observações em um limite de intervalo, como o primeiro do mês, para assegurar que as observações sejam igualmente espaçadas. Por exemplo, se o intervalo de tempo for Meses, então diversas datas no

mesmo mês serão agregadas. Esse tipo de agregação é referido como *agrupamento*. Por padrão, as observações são sumarizadas quando agrupadas. É possível especificar um método diferente para agrupamento, como a média das observações, a partir das configurações de **Agregação e Distribuição** na guia Especificações de Dados.

- Para alguns intervalos de tempo, configurações adicionais podem definir quebras nos intervalos normais igualmente espaçados. Por exemplo, se o intervalo de tempo for Dias, mas apenas dias da semana forem válidos, será possível especificar que há cinco dias em uma semana, com a semana iniciando na segunda-feira.

Observações que são definidas por períodos ou períodos cíclicos

Observações podem ser definidas por um ou mais campos de número inteiro que representam períodos ou ciclos repetitivos de períodos, até um número arbitrário de níveis de ciclo. Com esta estrutura, é possível descrever uma série de observações que não se enquadram em um dos intervalos de tempo padrão. Por exemplo, um ano fiscal com apenas 10 meses pode ser descrito com um campo de ciclo que representa anos e com um campo de período que representa meses, em que o comprimento de um ciclo é 10.

Os campos que especificam períodos cíclicos definem uma hierarquia de níveis periódicos, em que o nível mais baixo é definido pelo campo **Período**. O próximo nível mais alto é especificado por um campo de ciclo cujo nível é 1, seguido por um campo de ciclo cujo nível é 2, e assim por diante. Os valores de campo de cada nível, exceto o mais alto, devem ser periódicos com relação ao próximo nível mais alto. Os valores do nível mais alto não podem ser periódicos. Por exemplo, no caso do ano fiscal de 10 meses, os meses são periódicos dentro dos anos e os anos não são periódicos.

- O comprimento de um ciclo em um determinado nível é a periodicidade do próximo nível mais baixo. Para o exemplo do ano fiscal, há apenas um nível de ciclo e o comprimento do ciclo é 10, já que o próximo nível mais baixo representa meses e há 10 meses no ano fiscal especificado.
- Especificar o valor inicial para qualquer campo periódico que não iniciar a partir de 1. Essa configuração é necessária para detectar valores omissos. Por exemplo, se um campo periódico iniciar a partir de 2, mas o valor inicial estiver especificado como 1, então o procedimento irá supor que há um valor omissos para o primeiro período em cada ciclo desse campo.

Nó TCM de Fluxo - Opções de Intervalo de Tempo

O intervalo de tempo que é utilizado para a análise pode diferir do intervalo de tempo das observações. Por exemplo, se o intervalo de tempo das observações for Dias, será possível escolher Meses para o intervalo de tempo para análise. Os dados são então agregados de dados diários para dados mensais antes de o modelo ser construído. Também é possível optar por distribuir os dados de um intervalo de tempo maior para um intervalo de tempo menor. Por exemplo, se as observações forem trimestrais, será possível distribuir os dados de trimestrais para mensais.

As opções disponíveis para o intervalo de tempo no qual a análise é feita dependem de como as observações são definidas e do intervalo de tempo dessas observações. Em particular, quando as observações são definidas por períodos cíclicos, então apenas a agregação é suportada. Nesse caso, o intervalo de tempo da análise deve ser maior ou igual ao intervalo de tempo das observações.

O intervalo de tempo para a análise é especificado nas configurações de **Intervalo de Tempo** na guia Especificações de Dados. O método pelo qual os dados são agregados ou distribuídos é especificado nas configurações de **Agregação e Distribuição** na guia Especificações de Dados.

Nó TCM de Fluxo - Opções de Agregação e de Distribuição

Funções de agregação

Quando o intervalo de tempo que é utilizado para a análise é maior que o intervalo de tempo das observações, os dados de entrada são agregados. Por exemplo, a agregação é feita quando o

intervalo de tempo das observações for Dias e o intervalo de tempo para análise for Meses. As funções de agregação a seguir estão disponíveis: mean, sum, mode, min ou max.

Funções de distribuição

Quando o intervalo de tempo que é utilizado para a análise é menor que o intervalo de tempo das observações, os dados de entrada são distribuídos. Por exemplo, a distribuição é feita quando o intervalo de tempo das observações for Trimestres e o intervalo de tempo para análise for Meses. As funções de distribuição a seguir estão disponíveis: mean ou sum.

Funções de agrupamento

O agrupamento é aplicado quando as observações são definidas por datas/horas e diversas observações ocorrem no mesmo intervalo de tempo. Por exemplo, se o intervalo de tempo das observações for Meses, então diversas datas no mesmo mês serão agrupadas e associadas ao mês em que elas ocorrem. As funções de agrupamento a seguir estão disponíveis: mean, sum, mode, min ou max. O agrupamento é feito sempre quando as observações são definidas por datas/horas e o intervalo de tempo das observações é especificado como Irregular.

Nota: Embora o agrupamento seja uma forma de agregação, ele é feito antes de qualquer manipulação de valores omissos, ao passo que a agregação formal é feita após qualquer manipulação de valores omissos. Quando o intervalo de tempo das observações é especificado como Irregular, a agregação é feita apenas com a função de agrupamento.

Agregar observações entre dias com as do dia anterior

Especifica se as observações com tempos que ultrapassam um limite de um dia são agregados aos valores para o dia anterior. Por exemplo, para observações por hora com um dia de oito horas que começa às 20h, essa configuração especifica se as observações entre 0h e 4h são incluídas nos resultados agregados para o dia anterior. Essa configuração se aplicará apenas se o intervalo de tempo das observações for Horas por dia, Minutos por dia ou Segundos por dia e o intervalo de tempo para análise for Dias.

Configurações customizadas para campos especificados

É possível especificar as funções de agregação, distribuição e agrupamento em uma base de campo por campo. Essas configurações substituem as configurações padrão para as funções de agregação, distribuição e agrupamento.

Nó TCM de Fluxo - Opções de Valor Omisso

Os valores omissos nos dados de entrada são substituídos por um valor imputado. Os métodos de substituição a seguir estão disponíveis:

Interpolação linear

Substitui valores omissos utilizando um interpolação linear. O último valor válido antes do valor omissos e o primeiro valor válido após o valor omissos são utilizados para a interpolação. Se a primeira ou a última observação na série tiver um valor omissos, então os dois valores não omissos mais próximos no início ou no término da série serão utilizados.

Média de série

Substitui valores omissos pela média para toda a série.

Média de pontos próximos

Substitui valores omissos pela média de valores circundantes válidos. O span de pontos próximos é o número de valores válidos antes e após o valor omissos que são utilizados para calcular a média.

Mediana de pontos próximos

Substitui valores omissos pela mediana de valores válidos circundantes. O span de pontos próximos é o número de valores válidos antes e após o valor omissos que são utilizados para calcular a mediana.

Tendência linear

Esta opção utiliza todas as observações não omissas na série para ajustar um modelo de regressão linear simples, que é então utilizado para imputar os valores omissos.

Outras configurações:

Porcentagem máxima de valores omissos (%)

Especifica a porcentagem máxima de valores omissos que são permitidos para qualquer série. As séries com mais valores omissos que o máximo especificado são excluídas da análise.

Nó TCM de Fluxo - Opções de Dados Gerais

Número máximo de valores distintos por campo de dimensão

Esta configuração se aplica a dados multidimensionais e especifica o número máximo de valores distintos que são permitidos para qualquer campo de dimensão. Por padrão, este limite é configurado como 10000, mas poderá ser aumentado para um número arbitrariamente grande.

Nó TCM de Fluxo - Opções de Construção Geral

Largura de intervalo de confiança (%)

Esta configuração controla os intervalos de confiança para ambas as previsões e parâmetros do modelo. É possível especificar quaisquer valores positivos menores que 100. Por padrão, um intervalo de confiança de 95% é utilizado.

Número máximo de entradas por cada resposta

Essa configuração especifica o número máximo de entradas que são permitidas no modelo para cada resposta. É possível especificar um número inteiro no intervalo de 1 a 20. O modelo para cada resposta sempre inclui valores de lag de si mesmo, portanto, configurar esse valor para 1 especifica que a única entrada é a própria resposta.

Tolerância do modelo

Essa configuração controla o processo iterativo que é utilizado para determinar o melhor conjunto de entradas para cada resposta. É possível especificar qualquer valor que seja maior que zero. O padrão é 0,001.

Limite de valor discrepante (%)

Uma observação será sinalizada como um valor discrepante se a probabilidade de que o modelo seja um valor discrepante, conforme calculado a partir do mesmo, exceder esse limite. É possível especificar um valor no intervalo de 50 a 100.

Número de Lags para Cada Entrada

Esta configuração especifica o número de termos de lag para cada entrada no modelo para cada resposta. Por padrão, o número de termos de lag é determinado automaticamente a partir do intervalo de tempo que é utilizado para a análise. Por exemplo, se o intervalo de tempo for meses (com um incremento de um mês), então o número de lags será 12. Opcionalmente, é possível especificar explicitamente o número de lags. O valor especificado deve ser um número inteiro no intervalo de 1 a 20.

Continuar a estimação usando modelos existentes

Se você já gerou um modelo causal temporal, selecione esta opção para reutilizar as configurações de critérios que forem especificadas para esse modelo, ao invés de construir um novo modelo.

Dessa forma, é possível economizar tempo ao reestimar e produzir uma nova previsão que seja baseada nas mesmas configurações de modelo como antes, porém utilizando dados mais recentes.

Nó TCM de Fluxo - Opções do Período de Estimação

Por padrão, o período de estimação começa no momento da primeira observação e termina no momento da observação mais recente em todas as séries.

Por horários de início e de encerramento

É possível especificar o início e o término do período de estimação ou especificar apenas o início ou o término. Se você omitir o início ou o término do período de estimação, o valor padrão será utilizado.

- Se as observações forem definidas por um campo de data/hora, então insira valores para o início e o término no mesmo formato que é utilizado para o campo de data/hora.
- Para as observações que forem definidas por períodos cíclicos, especifique um valor para cada um dos campos de períodos cíclicos. Cada campo é exibido em uma coluna separada.

Por intervalos de tempo mais recentes ou mais antigos

Define o período de estimação como um número especificado de intervalos de tempo que começam no primeiro intervalo de tempo ou terminam no último intervalo de tempo nos dados, com um offset opcional. Neste contexto, o intervalo de tempo refere-se ao intervalo de tempo da análise. Por exemplo, suponha que as observações sejam mensais, mas o intervalo de tempo da análise seja trimestral. Especificar **Mais recente** e um valor de 24 para o **Número de intervalos de tempo** significa os últimos 24 trimestres.

Opcionalmente, é possível excluir um número especificado de intervalos de tempo. Por exemplo, especificar os últimos 24 intervalos de tempo e 1 para o número para excluir significa que o período de estimação consiste em 24 intervalos que precedem o último.

Nó TCM de Fluxo - Opções de Modelo

Nome do modelo

É possível especificar um nome customizado para o modelo ou aceitar o nome gerado automaticamente, que é *TCM*.

Previsão

A opção **Estender registros para o futuro** configura o número de intervalos de tempo para prever além do término do período de estimação. O intervalo de tempo nesse caso é o intervalo de tempo da análise, que é especificado na guia Especificações de Dados. Quando previsões são solicitadas, modelos autorregressivos são construídos automaticamente para quaisquer séries de entrada que não sejam também respostas. Esses modelos são, então, utilizados para gerar valores para essas séries de entrada no período de previsão. Não há limite máximo para essa configuração.

Nó Space-Time-Boxes

Os Space-Time-Boxes (STB) são uma extensão dos locais espaciais com Geohash. Mais especificamente, um STB é uma sequência de caracteres alfanuméricos que representa uma região formada regularmente de espaço e tempo.

Por exemplo, o STB **dr5ru7|2013-01-01 00:00:00|2013-01-01 00:15:00** é composto das três partes a seguir:

- O geohash **dr5ru7**
- O registro de data e hora de início **2013-01-01 00:00:00**
- O registro de data e hora de término **2013-01-01 00:15:00**

Como um exemplo, é possível utilizar as informações de espaço e tempo para melhorar a confiança de que duas entidades sejam as mesmas porque elas estão praticamente no mesmo local ao mesmo tempo. Como alternativa, é possível aprimorar a precisão da identificação de relacionamento ao mostrar que duas entidades estão relacionadas devido à sua proximidade no espaço e no tempo.

É possível escolher o modo **Registros Individuais** ou **Hangouts**, conforme apropriado para seus requisitos. Esses dois modos requerem os mesmos detalhes básicos, conforme a seguir:

Campo de latitude. Selecione o campo que identifica a latitude (no sistema de coordenadas WGS84).

Campo de longitude. Selecione o campo que identifica a longitude (no sistema de coordenadas WGS84).

Campo do registro de data e hora. Selecione o campo que identifica a hora ou data.

Opções de Registro Individuais

Utilize este modo para incluir um campo adicional a um registro para identificar sua localização em um determinado momento.

Derivar. Selecione uma ou mais densidades de espaço e tempo a partir das quais derivar o novo campo. Consulte “Definindo densidade de Space-Time-Box” na página 112 para obter mais informações.

Extensão do nome do campo. Digite a extensão que deseja incluir em um ou mais nomes de campos novos. É possível escolher incluir essa extensão como um **Sufixo** ou **Prefixo**.

Opções de Hangout

Um hangout pode ser considerado como uma localização e/ou tempo em que uma entidade é localizada continuamente ou repetidamente. Por exemplo, isso pode ser utilizado para identificar um veículo que faz execuções de transporte regulares e identificar quaisquer desvios da norma.

O detector de hangout monitora o movimento das entidades e sinaliza condições em que uma entidade é observada como “pendente” na área. O detector de hangout designa automaticamente cada hangout sinalizado para um ou mais STBs e utiliza rastreamento de entidade e de evento na memória para detectar hangouts com eficiência otimizada.

Densidade de STB. Selecione a densidade de espaço e tempo a partir da qual derivar o novo campo. Por exemplo, um valor de **STB_GH4_10MINS** corresponde a uma caixa geohash de quatro caracteres com um tamanho aproximado de 20 km por 20 km e uma janela de tempo de 10 minutos. Consulte “Definindo densidade de Space-Time-Box” na página 112 para obter mais informações.

Campo ID da entidade. Selecione a entidade a ser utilizada como o identificador de hangout. Esse campo de ID identifica o evento.

Número mínimo de eventos. Um evento é uma linha nos dados. Selecione o número mínimo de ocorrências de um evento para a entidade a serem consideradas como pendentes. Um hangout também deverá qualificar-se com base no campo **Tempo de permanência é pelo menos** a seguir.

Tempo de permanência é pelo menos. Especifique a duração mínima na qual a entidade deve persistir na mesma localização. Isso pode ajudar a excluir, por exemplo, um carro esperando o semáforo abrir, de ser considerado como pendente. Um hangout também deverá qualificar-se com base no campo **Número mínimo de eventos** anterior.

A seguir há mais detalhes sobre o que se qualifica como um hangout:

Deixar que e_1, \dots, e_n denote todos os eventos ordenados por tempo que são recebidos de um determinado ID de entidade durante um período de tempo (t_1, t_n) . Esses eventos se qualificarão como um hangout se:

- $n \geq$ número mínimo de eventos
- $t_n - t_1 \geq$ tempo mínimo de permanência
- Todos os eventos e_1, \dots, e_n ocorrerem no mesmo STB

Permitir que hangouts ultrapassem os limites de STB. Se essa opção for selecionada, a definição de um hangout será menos rigorosa e poderá incluir, por exemplo, uma entidade que sair em mais de um Space-Time-Box. Por exemplo, se seus STBs forem definidos como horas inteiras, selecionar essa opção

reconheceria uma entidade que estiver saindo para uma hora como válida, mesmo se a hora for consistida dos 30 minutos antes da meia-noite e dos 30 minutos após a meia-noite. Se essa opção não for selecionada, 100% do tempo hangout deverá estar dentro de um único Space-Time-Box.

Proporção mín. de eventos na caixa de tempo qualificativa (%). Disponível apenas se **Permitir que hangouts ultrapassem os limites de STB.** estiver selecionada. Utilize isso para controlar o grau em que um hangout relatado em um STB pode de fato se sobrepor ao outro. Selecione a proporção mínima de eventos que devem ocorrer dentro de um único STB para identificar um hangout. Se estiver configurado para 25%, e a proporção de eventos for 26%, isso se qualificará como sendo um hangout.

Por exemplo, suponha que você configure o detector de hangout para requerer pelo menos dois eventos (número mínimo de eventos = 2) e um tempo de ajuda instantânea adjacente de pelo menos 2 minutos em uma caixa de espaço geohash de 4 bytes e uma caixa de tempo 10 minutos (STB_NAME = STB_GH4_10MINS). Quando um hangout for detectado, digamos que a entidade flutue na mesma caixa de espaço geohash de 4 bytes enquanto os três eventos qualificativos ocorrem dentro de um span de tempo de 10 minutos entre 16h57 e 17h07 às 16h58, 17h01 e 17h03. O valor percentual da caixa de tempo qualificativa especifica os STBs que são elegíveis para o hangout, conforme a seguir:

- **100%.** O hangout é relatado na caixa de tempo de 17h - 17h10 e não está na caixa de tempo 16h50 - 17h (os eventos das 17h01 e 17h03 atendem a todas as condições que são necessárias para qualificação de um hangout e 100% desses eventos ocorreram na caixa de tempo 17h - 17h10).
- **50%.** Os hangouts nas duas caixas de tempo são relatados (os eventos das 17h01 e 17h03 atendem a todas as condições que são necessárias para qualificação de um hangout e 50% desses eventos ocorreram na caixa de tempo 16h50 - 17h e pelo menos 50% desses eventos ocorreram na caixa de tempo 17h - 17h10).
- **0%.** Hangouts nas duas caixas de tempo são relatados.

Quando 0% é especificado, os relatórios de hangout incluem os STBs representando cada caixa de tempo que é tocada pela duração qualificativa. A duração qualificativa precisa ser menor ou igual à duração correspondente da caixa de tempo no STB. Em outras palavras, nunca deverá haver uma configuração em que um STB de 10 minutos é configurado em conjunto com uma duração qualificativa de 20 minutos.

Um hangout é relatado assim que as condições qualificativas forem atendidas, e não é relatado mais de uma vez por STB. Suponha que três eventos elegíveis para um hangout e um total de 10 eventos ocorrem dentro de uma duração qualificativa dentro do mesmo STB. Nesse caso, o hangout é relatado quando o terceiro evento qualificativo ocorre. Nenhum dos sete eventos adicionais aciona um relatório de hangout.

Nota:

- Os dados de evento em memória do detector de hangout não são compartilhados entre os processos. Portanto, uma entidade específica possui afinidade com um determinado nó detector de hangout. Ou seja, os dados de pedido recebidos de uma entidade devem sempre ser transmitidos de modo consistente para o nó detector de hangout que rastreia essa entidade, que é normalmente o mesmo nó em toda a execução.
- Os dados de evento em memória do detector de hangout são voláteis. Sempre que o detector de hangout for encerrado e reiniciado, quaisquer hangouts de trabalho em andamento serão perdidos. Isso significa que parar e reiniciar o processo pode fazer com que o sistema deixe de relatar hangouts reais. Uma solução potencial envolve a reprodução de alguns dos dados de pedido históricos (por exemplo, voltar 48 horas e reproduzir os registros de pedido que são aplicáveis a qualquer nó que foi reiniciado).
- O detector de hangout deve ser alimentado por dados em ordem sequencial temporal.

Definindo densidade de Space-Time-Box

Escolha o tamanho (densidade) de seu Space-Time-Boxes (STB) ao especificar a área física e o tempo decorrido a ser incluído em cada.

Geodensidade. Selecione o tamanho da área a ser incluída em cada STB.

Intervalo de tempo. Selecione o número de horas a serem incluídas em cada STB.

Nome do campo. Prefixado com STB, este é preenchido automaticamente com base em suas seleções nos dois campos anteriores.

Capítulo 4. Nós de Operações de Campo

Visão Geral de Operações de Campo

Após uma exploração de dados inicial, você provavelmente precisará selecionar, limpar ou construir os dados em preparação para análise. A paleta Operações de Campo contém muitos nós úteis para essa transformação e preparação.

Por exemplo, utilizando um nó Derivar, você pode criar um atributo que não estiver atualmente representado nos dados. Ou pode utilizar um nó Categorização para recodificar os valores do campo automaticamente para análises destinadas. Você provavelmente utilizará um nó Tipo com frequência, que permite designar um nível de medição, valores e um papel de modelagem para cada campo no conjunto de dados. Suas operações são úteis para manipular valores omissos e modelagem de recebimento de dados.

A paleta Operações de Campo contém os seguintes nós:



O nó Automated Data Preparation (ADP) pode analisar seus dados e identificar correções, selecionar campos que são problemáticos ou que provavelmente não serão úteis, derivar novos atributos quando apropriado e aprimorar o desempenho por meio de técnicas de triagem e de amostragem inteligentes. É possível utilizar o nó de forma totalmente automatizada, permitindo que o nó escolha e aplique correções, ou é possível visualizar as mudanças antes que elas sejam feitas e aceitá-las, rejeitá-las ou modificá-las conforme desejado.



O nó Tipo especifica metadados e propriedades de campo. Por exemplo, é possível especificar um nível de medição (contínua, nominal, ordinal ou sinalizador) para cada campo, configurar opções para manipular valores omissos e nulos do sistema, configurar a função de um campo para fins de modelagem, especificar rótulos de campo e de valor e especificar valores para um campo.



O nó Filtro filtra (descarta) os campos, renomeia-os e mapeia-os de um nó de origem para outro.



O nó Derivar modifica valores de dados ou cria novos campos a partir de um ou mais campos existentes. Ele cria campos do tipo fórmula, sinalizador, nominal, estado, contagem e condicional.



O nó Combinação combina dois ou mais nuggets do modelo para obter previsões mais precisas do que pode ser obtido a partir de qualquer modelo individual.



O nó de Preenchimento substitui valores do campo altera o armazenamento. É possível optar por substituir valores baseados em uma condição de CLEM, como @BLANK(@FIELD). Como alternativa, é possível optar por substituir todos os espaços em branco ou valores nulos por um valor específico. Um nó de Preenchimento é geralmente utilizado em conjunto com um nó Tipo para substituir valores omissos.



O nó Anonimizar transforma a maneira com que os nomes e valores de campo são representados no recebimento de dados, ocultando, assim, os dados originais. Isto poderá ser útil se desejar permitir que outros usuários construam modelos utilizando dados sensíveis, como nomes de cliente ou outros detalhes.



O nó Reclassificar transforma um conjunto de valores categóricos em outro. A reclassificação é útil para reduzir as categorias ou para reagrupar dados para análise.



O nó Categorização cria automaticamente novos campos nominais (conjunto) com base nos valores de um ou mais campos existentes contínuos (intervalo numérico). Por exemplo, é possível transformar um campo de receita contínuo em um novo campo categórico contendo grupos de receitas como desvios da média. Depois de criar categorias para o novo campo, é possível gerar um nó Derivar com base nos pontos de corte.



O nó Recency, Frequency, Monetary (RFM) Analysis permite determinar de modo quantitativo quais podem ser seus melhores clientes ao examinar quando foi a última vez que eles compraram de você (recência), com que frequência eles compraram (frequência) e quanto eles gastaram em todas as transações (monetário).



O nó Partição gera um campo de partição que divide os dados em subconjuntos separados para o treinamento, teste e estágios de validação de construção de modelo.



O nó Configurar para Sinalizador deriva diversos campos de sinalização com base nos valores categóricos definidos para um ou mais campos nominais.



O nó Reestruturar converte um campo nominal ou sinalizador em um grupo de campos que podem ser preenchidos com os valores do outro campo. Por exemplo, dado um campo chamado *tipo de pagamento* com valores de *crédito*, *dinheiro* e *débito*, três novos campos seriam criados (*crédito*, *dinheiro* e *débito*), cada um dos quais podendo conter o valor do pagamento real feito.



O nó Transpor troca os dados em linhas e colunas para que os registros se tornem campos e os campos se tornem registros.



O nó Intervalos de Tempo especifica os intervalos e cria rótulos (se necessário) para modelar dados de séries temporais. Se os valores não forem uniformemente espaçados, o nó poderá preencher ou agregar valores conforme necessário para gerar um intervalo uniforme entre os registros.



O nó Intervalos de Tempo original não é compatível com o Analytic Server (AS). O nó Intervalos de Tempo do AS (novo no SPSS Modeler release 17.0) contém um subconjunto das funções do nó Intervalos de Tempo existente que pode ser usado com o Analytic Server.

Utilize o nó Intervalos de Tempo do AS para especificar intervalos e derivar um novo campo de tempo para estimativa ou previsão. Uma variedade completa de intervalos de tempo é suportada, desde segundos até anos.



O nó Histórico cria novos campos contendo dados de campos em registros anteriores. Os nós Históricos são mais frequentemente utilizados para dados sequenciais, como dados de séries temporais. Antes de utilizar um nó Histórico, você pode querer classificar os dados utilizando um nó Classificar.



O nó Reordenar Campo define a ordem natural utilizada para exibir campos de recebimento de dados. Esta ordem afeta a exibição de campos em uma variedade de locais, como tabelas, listas e o Seletor de Campo. Esta operação é útil ao trabalhar com conjuntos de dados grandes para tornar os campos de interesse mais visíveis.



No SPSS Modeler, itens como as funções espaciais do Construtor de Expressões, o nó Spatio-Temporal Prediction (STP), e o nó Visualização de Mapa usam o sistema de coordenadas projetado. Utilize o nó Reprojeter para alterar o sistema de coordenadas de quaisquer dados que importar que utilizam um sistema de coordenadas geográficas.

Vários desses nós podem ser gerados diretamente a partir do relatório de auditoria criado por um nó Auditoria de Dados. Consulte o tópico “Gerando Outros Nós para Preparação de Dados” na página 320 para obter mais informações.

Preparação de dados automatizada

A preparação de dados para análise é um dos passos mais importantes em qualquer projeto e, tradicionalmente, um dos que mais consomem tempo. A Preparação de Dados Automatizada (ADP) manipula a tarefa para você, analisando seus dados e identificando correções, verificando campos que são problemáticos ou que podem não ser úteis, derivando novos atributos quando apropriado e melhorando o desempenho por meio de técnicas de triagem inteligentes. É possível utilizar o algoritmo de forma totalmente **automática**, permitindo que ele escolha e aplique as correções, ou é possível utilizá-lo de forma **interativa**, visualizando as mudanças antes que elas sejam feitas e aceitando-as ou rejeitando-as conforme desejado.

Utilizar a ADP permite deixar seus dados prontos para construção de modelo de modo rápido e fácil, sem precisar de conhecimento prévio dos conceitos estatísticos envolvidos. Os modelos tendem a criar e escorar mais rapidamente e, além disso, o uso da ADP melhora a robustez dos processos de modelagem automatizados, como atualização e champion / challenger de modelo.

Nota: quando a ADP prepara um campo para análise, ela cria um novo campo contendo os ajustes ou as transformações, ao invés de substituir os valores e propriedades existentes do campo antigo. O campo antigo não é utilizado na análise adicional e seu papel é configurado para Nenhum.

Exemplo. Uma empresa de seguros com recursos limitados para investigar solicitações de seguro de um proprietário de residência deseja construir um modelo para sinalizar solicitações suspeitas potencialmente fraudulentas. Antes de construir o modelo, eles aprontarão os dados para modelagem utilizando preparação de dados automatizada. Como eles desejam ser capazes de revisar as transformações propostas antes que as transformações sejam aplicadas, eles utilizaram preparação de dados automatizada no modo interativo.

Um grupo de indústria automotiva mantém o controle das vendas de uma variedade de veículos pessoais. Em um esforço para serem capazes de identificar modelos com alto e baixo desempenho, eles desejam estabelecer um relacionamento entre as vendas de veículos e as características do veículo. Eles utilizarão a preparação de dados automatizada para preparar os dados para análise e os modelos de construção utilizando os dados "antes" e "após" a preparação para ver como os resultados diferem.

Qual é o seu objetivo? A preparação de dados automatizada recomenda passos de preparação de dados que afetarão a velocidade com que outros algoritmos podem construir modelos e melhorar o poder preditivo desses modelos. Isso pode incluir a transformação, construção e seleção de variáveis. A resposta também pode ser transformada. É possível especificar as prioridades de construção de modelo nas quais o processo de preparação de dados deve se concentrar.

- **Balanciar velocidade e precisão.** Esta opção prepara os dados para dar prioridade igual para a velocidade com a qual os dados são processados pelos algoritmos de construção de modelo e também para a precisão das previsões.
- **Otimizar para velocidade.** Esta opção prepara os dados para dar prioridade à velocidade com a qual os dados são processados pelos algoritmos de construção de modelo. Selecione essa opção quando estiver trabalhando com conjuntos de dados muito grandes ou procurando por uma resposta rápida.
- **Otimizar para precisão.** Esta opção prepara os dados para dar prioridade à precisão das previsões produzidas pelos algoritmos de construção de modelo.
- **Análise customizada.** Selecione essa opção quando desejar alterar manualmente o algoritmo na guia Configurações. Observe que essa configuração será selecionada automaticamente se mudanças subsequentes forem feitas nas opções na guia Configurações que forem incompatíveis com um dos outros objetivos.

Treinando o Nó

O nó ADP é implementado como um nó do processo e funciona de maneira semelhante ao nó Tipo; **treinar** o nó ADP corresponde a instanciar o nó Tipo. Quando a análise tiver sido executada, as transformações especificadas serão aplicadas aos dados sem análise adicional, contanto que o modelo de dados anterior não seja alterado. Assim como os nós Tipo e Filtro, se o nó ADP for desconectado, ele lembrará o modelo de dados e as transformações de modo que ele não precisará ser retreinado se reconectado, o que permite treiná-lo em um subconjunto de dados típicos e, em seguida, copiá-lo ou implementá-lo para uso nos dados ativos com a frequência necessária.

Usando a Barra de Ferramentas

A barra de ferramentas permite executar e atualizar a exibição da análise de dados e gerar nós que possam ser utilizados em conjunto com os dados originais.

- **Gerar** A partir deste menu, é possível gerar um Filtro ou um nó Derivar. Observe que esse menu está disponível somente quando houver uma análise mostrada na guia Análise.

O nó Filtro remove campos de entrada transformados. Se você configurar o nó ADP para deixar os campos de entrada originais no conjunto de dados, isso restaurará o conjunto de entradas original permitindo interpretar o campo de escore em termos das entradas. Por exemplo, isso poderá ser útil se desejar produzir um gráfico do campo de escore com relação a várias entradas.

O nó Derivar pode restaurar o conjunto de dados original e as unidades de destino. Será possível gerar um nó Derivar somente quando o nó ADP contiver uma análise que escala novamente um destino do intervalo (ou seja, a nova escala Box-Cox é selecionada no painel Preparar Entradas e Destino). Não será possível gerar um nó Derivar se o destino não for um intervalo ou se a nova escala Box-Cox não estiver selecionada. Consulte o tópico “Gerando um Nó Derivar” na página 130 para obter mais informações.

- **Exibir** Contém opções que controlam o que é mostrado na guia Análise. Isso inclui os controles de edição do gráfico e as seleções de exibição para o painel principal e visualizações vinculadas.
- **Visualização** Exibe uma amostra das transformações que serão aplicadas aos dados de entrada.

- **Analisar Dados** Inicia uma análise utilizando as configurações atuais e exibe os resultados na guia Análise.
- **Limpar Análise** Exclui a análise existente (disponível apenas quando uma análise atual existir).

Status do Nó

O status do nó ADP na tela do IBM SPSS Modeler é indicada por uma seta ou por uma marcação no ícone que mostra se a análise ocorreu ou não.

Guia Campos

Antes de poder construir um modelo, é necessário especificar quais campos você deseja utilizar como destinos e como entradas. Com algumas exceções, todos os nós de modelagem utilizarão as informações de campo a partir de um nó Tipo de envio de dados. Se você estiver utilizando um nó Tipo para selecionar campos de entrada e de destino, não será necessário alterar nada nesta guia.

Usar configurações do nó de tipo. Essa opção instrui o nó a utilizar as informações de campo a partir de um nó Tipo de envio de dados. Esse é o padrão.

Usar configurações customizadas. Essa opção instrui o nó a utilizar as informações de campo especificadas aqui ao invés das informações fornecidas em qualquer nó ou nós Tipo de envio de dados. Após selecionar esta opção, especifique os campos abaixo conforme necessário.

Resposta. Para modelos que requerem um ou mais campos de destino, selecione os campos de destino. Isso é semelhante a configurar o papel do campo para *Destino* em um nó Tipo.

Entradas. Selecione um ou mais campos de entrada. Isso é semelhante a configurar o papel do campo para *Entrada* em um nó Tipo.

Guia Configurações

A guia Configurações é composta por vários grupos diferentes de configurações que podem ser modificadas para realizar um ajuste preciso de como o algoritmo processa seus dados. Se você fizer mudanças nas configurações padrão que forem incompatíveis com os outros objetivos, a guia Objetivo será atualizada automaticamente para selecionar a opção **Customizar análise**.

Configurações de Campo

Usar campo de frequência. Esta opção permite selecionar um campo como uma ponderação de frequência. Use esta opção se os registros em seus dados de treinamento representarem mais de uma unidade cada; por exemplo, se estiver usando dados agregados. Os valores do campo devem ser o número de unidades representadas por cada registro.

Usar campo de ponderação. Esta opção permite selecionar um campo como uma ponderação de caso. As ponderações de caso são utilizadas para considerar as diferenças na variância nos níveis do campo de saída.

Como manipular campos que são excluídos da modelagem. Especifique o que acontece com os campos excluídos; é possível escolher filtrá-los dos dados ou simplesmente configurar o *Papel* deles para **Nenhum**.

Nota: esta ação também será aplicada ao destino, se ele estiver transformado. Por exemplo, se a nova versão derivada do destino for utilizada como o campo de **Destino**, o destino original será filtrado ou configurado para **Nenhuma**.

Se os campos de entrada não corresponderem à análise existente. Especifique o que acontece se um ou mais campos de entrada obrigatórios estiverem omissos do conjunto de dados de entrada quando executar um nó ADP treinado.

- **Parar a execução e manter a análise existente.** Isso para o processo de execução, preserva as informações de análise atuais e exibe um erro.
- **Limpar as análises existentes e analisar os novos dados.** Isso limpa a análise existente, analisa os dados de entrada e aplica as transformações recomendadas nesses dados.

Preparar datas e horas

Muitos algoritmos de modelagem são incapazes de manipular diretamente data e hora em detalhes; essas configurações permitem derivar novos dados de duração que podem ser utilizados como entradas de modelo a partir de datas e horas em seus dados existentes. Os campos que contêm datas e horas devem ser predefinidos com tipos de armazenamento de data ou hora. Os campos de data e hora originais não serão recomendados como entradas de modelo após a preparação de dados automatizada.

Preparar datas e horas para modelagem. Cancelar a seleção desta opção desativa todos os outros controles de Preparar Datas e Horas enquanto mantém as seleções.

Calcular o tempo decorrido até a data de referência. Isso produz o número de anos/meses/dias desde uma data de referência para cada variável que contém datas.

- **Data de Referência.** Especifique a data a partir da qual a duração será calculada com relação às informações de data nos dados de entrada. A seleção de **Data de hoje** significa que a data atual do sistema é sempre utilizada quando o ADP é executado. Para utilizar uma data específica, selecione **Data fixa** e insira a data necessária. A data atual é inserida automaticamente no campo **Data fixa** quando o nó é criado pela primeira vez.
- **Unidades para Duração de Data.** Especifique se o ADP deve decidir automaticamente a unidade de duração de data ou selecione a partir de **Unidades fixas** de Anos, Meses ou Dias.

Calcular o tempo decorrido até o tempo de referência. Isso produz o número de horas/minutos/segundos desde um tempo de referência para cada variável que contém os tempos.

- **Tempo de Referência.** Especifique o tempo a partir do qual a duração será calculada com relação às informações de tempo nos dados de entrada. A seleção de **Tempo atual** significa que o tempo atual do sistema é sempre utilizado quando o ADP é executado. Para utilizar um tempo específico, selecione **Tempo fixo** e insira os detalhes necessários. O tempo atual é inserido automaticamente no campo **Tempo fixo** quando o nó é criado pela primeira vez.
- **Unidades para Duração de Tempo.** Especifique se o ADP deve decidir automaticamente a unidade de duração de tempo ou selecione a partir de **Unidades fixas** de Horas, Minutos ou Segundos.

Extrair Elementos de Tempo Cíclicos. Utilize essas configurações para dividir um campo de data ou hora único em um ou mais campos. Por exemplo, se você selecionar todas as três caixas de seleção de data, o campo de data de entrada "1954-05-23" será dividido em três campos: 1954, 5 e 23, com cada um utilizando o sufixo definido no painel **Nomes de Campos**, e o campo de data original é ignorado.

- **Extrair de datas.** Para quaisquer entradas de data, especifique se você deseja extrair anos, meses, dias ou qualquer combinação.
- **Extrair de tempos.** Para quaisquer entradas de tempo, especifique se você deseja extrair horas, minutos, segundos ou qualquer combinação.

Excluir Campos

Dados de qualidade ruim podem afetar a precisão de suas previsões, portanto, é possível especificar o nível de qualidade aceitável para variáveis de entrada. Todos os campos que forem constantes ou que tiverem 100% de valores omissos são automaticamente excluídos.

Excluir campos de entrada de qualidade baixa. Cancelar a seleção desta opção desativa todos os outros controles de Excluir Campos enquanto mantém as seleções.

Excluir campos com valores omissos em excesso. Os campos com mais que a porcentagem especificada de valores omissos são removidos da análise adicional. Especifique um valor maior ou igual a 0, que é equivalente a cancelar a seleção dessa opção, e menor ou igual a 100, embora campos com todos valores omissos sejam excluídos automaticamente. O padrão é 50.

Excluir campos nominais com categorias exclusivas em excesso. Os campos nominais que tiverem mais que o número especificado de categorias são removidos da análise adicional. Especifique um número inteiro positivo. O padrão é 100. Isso é útil para remover automaticamente da modelagem os campos que contêm informações exclusivas de registro, como o ID, o endereço ou o nome.

Excluir campos categóricos com valores em excesso em uma única categoria. Campos ordinais e nominais com uma categoria que contém mais que a porcentagem especificada de registros são removidos da análise adicional. Especifique um valor maior ou igual a 0, equivalente a cancelar a seleção dessa opção, e menor ou igual a 100, embora campos constantes sejam automaticamente excluídos. O padrão é 95.

Preparando Entradas e Destinos

Como nenhum dado está sempre em um estado perfeito para processamento, você pode querer ajustar algumas das configurações antes de executar uma análise. Por exemplo, isso pode incluir a remoção de valores discrepantes, especificação de como manipular valores omissos ou ajuste do tipo.

Nota: se alterar os valores neste painel, a guia **Objetivos** será atualizada automaticamente para selecionar a opção **Análise customizada**.

Preparar os campos de entrada e de destino para modelagem. Ativa ou desativa todos os campos no painel.

Ajustar Tipo e Melhorar Qualidade dos Dados. Para entradas e o destino, é possível especificar diversas transformações de dados separadamente; isso ocorre porque você pode não querer alterar os valores do destino. Por exemplo, uma predição de receitas em dólares é mais significativa do que uma predição medida em log(dólares). Além disso, se o destino tiver valores omissos, não haverá nenhum ganho preditivo para preenchimento de valores omissos, ao passo que o preenchimento de valores omissos em entradas poderá permitir alguns algoritmos processem informações que do contrário seriam perdidas.

Configurações adicionais para essas transformações, como valor de corte de valor discrepante, são comuns para o destino e as entradas.

É possível selecionar as configurações a seguir para entradas e/ou destinos:

- **Ajustar o tipo de campos numéricos.** Selecione esta opção para determinar se os campos numéricos com um nível de medição de *Ordinal* podem ser convertidos em *Contínuo*, ou vice-versa. É possível especificar os valores limite mínimo e máximo para controlar a conversão.
- **Reordenar campos nominais.** Selecione esta opção para ordenar campos nominais (conjunto) da categoria menor para a maior.
- **Substituir valores discrepantes em campos contínuos.** Especifique se é necessário substituir valores discrepantes; utilize isto em conjunto com as opções **Método para substituir valores discrepantes** abaixo.
- **Campos contínuos: substituir valores omissos pela média.** Selecione esta opção para substituir valores omissos de variáveis contínuas (intervalo).
- **Campos nominais: substituir valores omissos pelo modo.** Selecione esta opção para substituir valores omissos de variáveis nominais (conjunto).
- **Campos ordinais: substituir valores omissos pela mediana.** Selecione esta opção para substituir valores omissos de variáveis ordinais (conjunto ordenado).

Número máximo de valores para campos ordinais. Especifique o limite para redefinir campos ordinais (conjunto ordenado) como contínuos (intervalo). O padrão é 10, portanto, se um campo ordinal tiver mais de 10 categorias, ele será redefinido como contínuo (intervalo).

Número mínimo de valores para campos contínuos. Especifique o limite para redefinir campos de escala ou contínuos (intervalo) como ordinais (conjunto ordenado). O padrão é 5, portanto, se um campo contínuo tiver menos de 5 valores, ele será redefinido como ordinal (conjunto ordenado).

Valor de corte de valor discrepante. Especifique o critério de valor de corte de valor discrepante, medido em desvios padrão, o padrão é 3.

Método para substituir valores discrepantes. Selecione se os valores discrepantes devem ser substituídos por um corte (coerção) com o valor de corte ou excluídos e configurados como valores omissos. Quaisquer valores discrepantes configurados como valores omissos seguem as configurações de manipulação de valor omissos selecionadas acima.

Colocar todos os campos de entrada contínuos em uma escala comum. Para normalizar os campos de entrada contínuos, marque esta caixa de seleção e escolha o método de normalização. O padrão é **Transformação de score z**, em que é possível especificar a **Média final** que possui um padrão de 0, e o **Desvio padrão final**, que tem um padrão de 1. Como alternativa, é possível escolher utilizar **Transformação mín./máx.** e especificar os valores mínimo e máximo, cujos valores padrão são 0 e 100, respectivamente.

Este campo é especialmente útil quando selecionar **Executar construção de variável** no painel Construir e Selecionar Variáveis.

Escalar novamente um destino contínuo com uma transformação Box-Cox. Para normalizar um campo de destino contínuo (escala ou intervalo), marque essa caixa de seleção. A transformação Box-Cox tem valores padrão de 0 para a **Média final** e 1 para o **Desvio padrão final**.

Nota: se escolher normalizar o destino, a dimensão do destino será transformada. Nesse caso, poderá ser necessário gerar um nó Derivar para aplicar uma transformação inversa para transformar as unidades novamente em um formato reconhecível para processamento adicional. Consulte o tópico “Gerando um Nó Derivar” na página 130 para obter mais informações.

Construção e Seleção de Variável

Para melhorar o poder preditivo de seus dados, é possível transformar os campos de entrada ou construir novos campos com base nos existentes.

Nota: se alterar os valores neste painel, a guia **Objetivos** será atualizada automaticamente para selecionar a opção **Análise customizada**.

Transformar, construir e selecionar campos de entrada para melhorar o poder preditivo. Ativa ou desativa todos os campos no painel.

Mesclar categorias de dispersão para maximizar associação com a resposta. Selecione esta opção para tornar um modelo mais simples ao reduzir o número de variáveis a serem processadas em associação com a resposta. Se necessário, altere o valor da probabilidade a partir do padrão de 0,05.

Observe que, se todas as categorias forem mescladas em uma só, as versões originais e derivadas do campo serão excluídas porque elas não possuirão nenhum valor como um preditor.

Quando não houver nenhuma resposta, mesclar as categorias de dispersão com base nas contagens. Se estiver lidando com dados que não possuem nenhuma resposta, será possível escolher mesclar categorias

de dispersão de variáveis ordinais (conjunto ordenado) e/ou nominais (conjunto). Especifique a porcentagem mínima de casos ou registros nos dados que identificam as categorias a serem mescladas; o padrão é 10.

As categorias são mescladas utilizando as seguintes regras:

- A mesclagem não é executada em campos binários.
- A mesclagem parará se houver somente duas categorias durante a mesclagem.
- A mesclagem também parará se não houver nenhuma categoria original, nem qualquer categoria criada durante a mesclagem ou houver menos que a porcentagem mínima especificada de casos.

Categorizar campos contínuos enquanto preserva o poder preditivo. Se você tiver dados que incluem uma variável resposta categórica, será possível categorizar entradas contínuas com associações fortes para melhorar o desempenho do processamento. Se necessário, altere o valor da probabilidade para subconjuntos homogêneos a partir do padrão de 0,05.

Se a operação de categorização resultar em uma categoria única para um campo específico, as versões originais e categorizadas do campo serão excluídas porque elas não possuirão nenhum valor como um preditor.

Nota: a categorização no ADP difere de categorização ideal utilizada em outras partes do IBM SPSS Modeler. A categorização ideal utiliza informações de entropia para converter uma variável contínua em uma variável categórica, o que requer ordenar e armazenar todos os dados na memória. O ADP utiliza subconjuntos homogêneos para categorizar uma variável contínua, o que significa que a categorização do ADP não precisa ordenar dados nem armazena todos os dados na memória. O uso do método de subconjunto homogêneo para categorizar uma variável contínua significa que o número de categorias após a categorização é sempre menor ou igual ao número de categorias da resposta.

Executar seleção de variável. Selecione esta opção para remover variáveis com um coeficiente de correlação baixo. Se necessário, altere o valor da probabilidade a partir do padrão de 0,05.

Esta opção se aplica somente às variáveis de entrada contínuas em que a resposta seja contínua, e às variáveis de entrada categóricas.

Executar construção de variável. Selecione esta opção para derivar novas variáveis a partir de uma combinação de diversas variáveis existentes (que são, então, descartadas da modelagem).

Esta opção se aplica somente às variáveis de entrada contínuas em que a resposta seja contínua, ou onde não houver nenhuma resposta.

Nomes de campo

Para identificar facilmente recursos novos e transformados, o ADP cria e aplica os novos nomes, prefixos ou sufixos de base. É possível corrigir esses nomes para serem mais relevantes para suas próprias necessidades e dados. Se desejar especificar outros rótulos, isso precisará ser feito em um nó Tipo de recebimento de dados.

Campos Transformados e Construídos. Especifique as extensões de nome a serem aplicadas aos campos de destino e de entrada transformados.

Observe que no nó do ADP, configurar campos de sequência de caracteres para conter nada poderá causar um erro dependendo de como você escolheu manipular campos não utilizados. Se **Como manipular campos que foram excluídos da modelagem** for configurada para **Filtrar campos não utilizados** no painel Configurações de Campo da guia Configurações, as extensões de nome para entradas e o destino poderão ser configuradas para nada. Os campos originais são filtrados e os campos transformados salvos sobre eles; nesse caso, os novos campos transformados terão o mesmo nome que os originais.

No entanto, se escolher **Configurar a direção de campos não utilizados para 'Nenhum'**, então extensões de nome vazias ou nulas para o destino e entradas causarão um erro porque você está tentando criar nomes de campo duplicados.

Além disso, especifique o nome do prefixo a ser aplicado a todos os recursos que forem construídos por meio das configurações de Seleção e Construção. O novo nome é criado ao anexar um sufixo numérico nesse nome raiz de prefixo. O formato do número depende de quantos novos recursos são derivados, por exemplo:

- 1 a 9 recursos construídos serão denominados: feature1 a feature9.
- 10 a 99 recursos construídos serão denominados: feature01 a feature99.
- 100 a 999 recursos construídos serão denominados: feature001 a feature999, e assim por diante.

Isso assegura que os recursos construídos sejam classificados em uma ordem sensível, não importa quantos houver.

Durações Calculadas a partir de Datas e Horas. Especifique as extensões de nome a serem aplicadas às durações calculadas a partir das datas e horas.

Elementos Cíclicos Extraídos de Datas e Horas. Especifique as extensões de nome a serem aplicadas aos elementos cíclicos extraídos das datas e horas.

Guia Análise

1. Quando estiver satisfeito com as configurações do ADP, incluindo quaisquer mudanças feitas nas guias Objetivo, Campos e Configurações, clique em **Analisar Dados**; o algoritmo aplica as configurações nas entradas de dados e exibe os resultados na guia Análise.

A guia Análise contém saída tabular e gráfica que sumariza o processamento de seus dados e exibe recomendações sobre como os dados podem ser modificados ou melhorados para escoragem. Em seguida, é possível revisar e aceitar ou rejeitar as recomendações.

A guia Análise é formada por dois painéis, a visualização principal à esquerda e a visualização vinculada ou auxiliar à direita. Há três visualizações principais:

- Sumarização de Processamento de Campo (o padrão). Consulte o tópico “Sumarização de Processamento de Campo” na página 125 para obter mais informações.
- Campos. Consulte o tópico “Campos” na página 125 para obter mais informações.
- Sumarização da Ação. Consulte o tópico “Sumarização da Ação” na página 126 para obter mais informações.

Há quatro visualizações vinculadas/auxiliares:

- Poder Preditivo (o padrão). Consulte o tópico “Poder Preditivo” na página 127 para obter mais informações.
- Tabela de Campos. Consulte o tópico “Tabela de Campos” na página 127 para obter mais informações.
- Detalhes do Campo. Consulte o tópico “Detalhes do Campo” na página 127 para obter mais informações.
- Detalhes da Ação. Consulte o tópico “Detalhes da Ação” na página 128 para obter mais informações.

Ligações entre as visualizações

Dentro da visualização principal, o texto sublinhado nas tabelas controla a exibição na visualização vinculada. Clicar no texto permite obter detalhes sobre um determinado campo, conjunto de campos ou passo de processamento. A ligação selecionada pela última vez é mostrada em uma cor mais escura, o que ajuda a identificar a conexão entre o conteúdo dos dois painéis de visualização.

Reconfigurando as visualizações

Para exibir novamente as recomendações de Análise originais e abandonar quaisquer mudanças feitas nas visualizações Análise, clique em **Reconfigurar** na parte inferior do painel de visualização principal.

Sumarização de Processamento de Campo

A tabela Sumarização de Processamento de Campo fornece uma captura instantânea do impacto geral projetado de processamento, incluindo mudanças no estado das variáveis e o número de variáveis construídas.

Observe que como nenhum modelo é realmente construído, não há uma medida ou um gráfico da mudança no poder preditivo geral antes e após a preparação de dados, no entanto, é possível exibir gráficos do poder preditivo de preditores individuais recomendados.

A tabela exibe as seguintes informações:

- O número de campos de destino.
- O número de preditores originais (entrada).
- Os preditores recomendados para uso em análise e modelagem. Isso inclui o número total de campos recomendados, o número de campos originais e não transformados recomendados, o número de campos transformados recomendados (exceto versões intermediárias de qualquer campo, campos derivados de preditores de data/hora e preditores construídos), o número de campos recomendados que são derivados dos campos de data/hora e o número de preditores construídos recomendados.
- O número de preditores de entrada não recomendados para uso em qualquer forma, seja em sua forma original, como um campo derivado ou como entrada para um preditor construído.

Se qualquer uma das informações **Campos** estiver sublinhada, clique para exibir mais detalhes em uma visualização vinculada. Detalhes do **Destino**, **Variáveis de entrada** e **Variáveis de entrada não utilizadas** são mostrados na visualização vinculada da Tabela de Campos. Veja o tópico “Tabela de Campos” na página 127 para obter mais informações. **As variáveis recomendadas para uso na análise** são exibidas na visualização vinculada Poder Preditivo Consulte o tópico “Poder Preditivo” na página 127 para obter mais informações.

Campos

A visualização principal Campos exibe os campos processados e se o ADP recomenda o uso deles em modelos de recebimento de dados. É possível substituir a recomendação para qualquer campo, por exemplo, para excluir variáveis construídas ou incluir variáveis que o ADP recomenda excluir. Se um campo tiver sido transformado, é possível decidir se deseja aceitar a transformação sugerida ou usar a versão original.

A visualização Campos consiste em duas tabelas, uma para o destino e outra para preditores que foram processados ou criados.

Tabela de destino

A tabela **Destino** será mostrada apenas se um destino estiver definido nos dados.

A tabela contém duas colunas:

- **Nome.** Este é o nome ou o rótulo do campo de destino e o nome original é sempre utilizado, mesmo se o campo tiver sido transformado.
- **Nível de Medição.** Isso exibe o ícone que representa o nível de medição; passe o mouse sobre o ícone para exibir um rótulo (contínuo, ordinal, nominal, e assim por diante) que descreve os dados.

Se o destino foi transformado, a coluna **Nível de Medição** refletirá a versão transformada final. *Nota:* não é possível desligar transformações para o destino.

Tabela Preditores

A tabela **Preditores** é sempre mostrada. Cada linha da tabela representa um campo. Por padrão, as linhas são classificadas em ordem decrescente de poder preditivo.

Para variáveis ordinárias, o nome original é sempre utilizado como o nome da linha. As versões originais e derivadas de campos de data/hora aparecem na tabela (em linhas separadas), que também inclui preditores construídos.

Observe que as versões transformadas de campos mostrados na tabela sempre representam as versões finais.

Por padrão, apenas campos recomendáveis são mostrados na tabela Preditores. Para exibir os campos restantes, selecione a caixa **Incluir campos não recomendados na tabela** acima da tabela e, em seguida, esses campos são exibidos na parte inferior da tabela.

A tabela contém as seguintes colunas:

- **Versão para Utilizar.** Isso exibe uma lista suspensa que controla se um campo será utilizado posteriormente e se as transformações sugeridas serão utilizadas. Por padrão, a lista suspensa reflete as recomendações.

Para preditores ordinários que foram transformados, a lista suspensa possui três opções: **Transformado**, **Original** e **Não utilizar**.

Para preditores ordinários não transformados, as opções são: **Original** e **Não utilizar**.

Para campos de data/hora derivados e preditores construídos, as opções são: **Transformado** e **Não utilizar**.

Para campos de data originais, a lista suspensa é desativada e configurada para **Não utilizar**.

Nota: para preditores com as versões originais e transformadas, a mudança entre as versões **Original** e **Transformado** atualiza automaticamente as configurações de **Nível de Medição** e de **Poder Preditivo** para essas variáveis.

- **Nome.** Cada nome de campo é uma ligação. Clique em um nome para exibir mais informações sobre o campo na visualização vinculada. Consulte o tópico “Detalhes do Campo” na página 127 para obter mais informações.
- **Nível de Medição.** Isso exibe o ícone que representa o tipo de dados; passe o mouse sobre o ícone para exibir um rótulo (contínuo, ordinal, nominal, e assim por diante) que descreve os dados.
- **Poder Preditivo.** O poder preditivo é exibido apenas para campos recomendados pelo ADP. Essa coluna não será exibida se não houver nenhum destino definido. O poder preditivo varia de 0 a 1, com valores maiores indicando "melhores" preditores. Em geral, o poder preditivo é útil para comparar preditores dentro de uma análise do ADP, mas os valores de poder preditivo não devem ser comparados nas análises.

Sumarização da Ação

Para cada ação executada pela preparação de dados automatizada, os preditores de entrada são transformados e/ou filtrados e os campos que sobreviverem a uma ação serão utilizados na próxima. Os campos que sobreviverem até o último passo são, então, recomendados para uso na modelagem, ao passo que entradas para preditores transformados e construídos são filtradas.

A Sumarização da Ação é uma tabela simples que lista as ações de processamento executadas pelo ADP. Quando qualquer **Ação** for sublinhada, clique para exibir mais detalhes em uma visualização vinculada sobre as ações executadas. Consulte o tópico “Detalhes da Ação” na página 128 para obter mais informações.

Nota: apenas as versões transformadas originais e finais de cada campo são mostradas, não quaisquer versões intermediárias que foram utilizadas durante a análise.

Poder Preditivo

Exibido por padrão quando a análise é executada pela primeira vez, ou quando selecionar **Preditores recomendados para uso em análise** na visualização principal de Sumarização de Processamento de Campo, o gráfico exibe o poder preditivo dos preditores recomendados. Os campos são ordenados por poder preditivo, com o campo com o valor mais alto aparecendo na parte superior.

Para versões transformadas de preditores ordinários, o nome do campo reflete sua opção de sufixo no painel Nomes de Campo da guia Configurações, por exemplo: *_transformed*.

Os ícones do nível de medição são exibidos após os nomes de campo individuais.

O poder preditivo de cada preditor recomendado é calculado a partir de uma regressão linear ou modelo naïve Bayes, dependendo se o destino for contínuo ou categórico.

Tabela de Campos

Exibida quando clicar em **Destino**, **Preditores** ou **Preditores não utilizados** na visualização principal de Sumarização de Processamento de Campo, a visualização Tabela de Campos exibe uma tabela simples listando as variáveis relevantes.

A tabela contém duas colunas:

- **Nome.** O nome do preditor.

Para destinos, o nome ou o rótulo original do campo é utilizado, mesmo se o destino tiver sido transformado.

Para versões transformadas de preditores ordinários, o nome reflete sua opção de sufixo no painel Nomes de Campo da guia Configurações, por exemplo: *_transformed*.

Para campos derivados de datas e horas, o nome da versão transformada final é utilizado, por exemplo: *bdate_years*.

Para preditores construídos, o nome do preditor construído é utilizado, por exemplo: *Predictor1*.

- **Nível de medição.** Isso exibe o ícone que representa o tipo de dados.

Para o Destino, o **Nível de medição** sempre reflete a versão transformada (se o destino foi transformado), por exemplo, alterado de ordinal (conjunto ordenado) para contínuo (intervalo, escala), ou vice-versa.

Detalhes do Campo

Exibida quando clicar em qualquer **Nome** na visualização principal Campos, a visualização Detalhes do Campo contém a distribuição, os valores omissos e gráficos de poder preditivo (se aplicável) para o campo selecionado. Além disso, o histórico de processamento para o campo e o nome do campo transformado também são mostrados (se aplicável).

Para cada conjunto de gráficos, duas versões são mostradas lado a lado para comparar o campo com e sem as transformações aplicadas; se uma versão transformada do campo não existir, um gráfico será mostrado somente para a versão original. Para campos de data ou hora derivados e preditores construídos, os gráficos são mostrados somente para o novo preditor.

Nota: se um campo for excluído devido a categorias em excesso, somente o histórico de processamento será mostrado.

Gráfico de Distribuição

A distribuição de campo contínuo é mostrada como um histograma, com uma curva normal sobreposta e uma linha de referência vertical para o valor médio, e os campos categóricos são exibidos como um gráfico de barras.

Os histogramas são rotulados para mostrar o desvio e a assimetria padrão, no entanto, a assimetria não será exibida se o número de valores for 2 ou menos ou se a variância do campo original for menor que 10-20.

Passa o mouse sobre o gráfico para exibir a média para histogramas ou a contagem e a porcentagem do número total de registros para categorias nos gráficos de barras.

Gráfico de Valor Omisso

Os gráficos de pizza comparam a porcentagem dos valores omitidos com e sem as transformações aplicadas e os rótulos do gráfico mostram a porcentagem.

Se o ADP executou a manipulação de valor omisso, o gráfico de pizza de pós-transformação também incluirá o valor de substituição como um rótulo -- ou seja, o valor utilizado no lugar dos valores omitidos.

Passa o mouse sobre o gráfico para exibir a contagem de valores omitidos e porcentagem do número total de registros.

Gráfico de Poder Preditivo

Para campos recomendados, os gráficos de barras exibem o poder preditivo antes e após a transformação. Se o destino tiver sido transformado, o poder preditivo calculado será referente ao destino transformado.

Nota: os gráficos de poder preditivo não serão mostrados se nenhum destino estiver definido ou se o destino for clicado no painel de visualização principal.

Passa o mouse sobre o gráfico para exibir o valor do poder preditivo.

Tabela de Históricos de Processamento

A tabela mostra como a versão transformada de um campo foi derivada. As ações executadas pelo ADP são listadas na ordem em que foram realizadas, no entanto, para determinados passos, diversas ações podem ter sido executadas para um campo específico.

Nota: esta tabela não é mostrada para os campos que não foram transformados.

As informações na tabela são divididas em duas ou três colunas:

- **Ação.** O nome da ação. Por exemplo, Preditores Contínuos. Consulte o tópico “Detalhes da Ação” para obter mais informações.
- **Detalhes.** A lista de processamentos executados. Por exemplo, Transformação para unidades padrão.
- **Função.** Mostrada apenas para preditores construídos, isto exibe a combinação linear de campos de entrada, por exemplo, $.06*age + 1,21*height$.

Detalhes da Ação

Exibida ao selecionar qualquer **Ação** sublinhada na visualização principal de Sumarização da Ação, a visualização vinculada Detalhes da Ação exibe informações comuns e específicas da ação para cada passo de processamento que for concluído, com os detalhes específicos da ação exibidos primeiro.

Para cada ação, a descrição é utilizada como o título na parte superior da visualização vinculada. Os detalhes específicos da ação são exibidos abaixo do título e podem incluir detalhes do número de preditores derivados, campos reformulados, transformações de destino, categorias mescladas ou reordenadas e preditores construídos ou excluídos.

Conforme cada ação é processada, o número de preditores usados no processamento pode alterar, por exemplo, conforme preditores são excluídos ou mesclados.

Nota: se uma ação foi desativada ou nenhum destino foi especificado, uma mensagem de erro será exibida no lugar dos detalhes da ação quando a ação é clicada na visualização principal Sumarização da Ação.

Há nove ações possíveis, no entanto, nem todas elas são necessariamente ativas para cada análise.

Tabela de Campos de Texto

A tabela exibe o número de:

- Valores em branco à esquerda aparados.
- Preditores excluídos da análise.

Tabela de Preditores de Data e Hora

A tabela exibe o número de:

- Durações derivadas dos preditores de data e hora.
- Elementos de data e hora.
- Preditores de data e hora derivados, no total.

A data ou hora de referência é exibida como uma nota de rodapé se quaisquer durações de data forem calculadas.

Tabela de Verificação do Preditor

A tabela exibe o número dos seguintes preditores excluídos do processamento:

- Constantes.
- Preditores com valores omissos em excesso.
- Preditores com casos em excesso em uma única categoria.
- Campos nominais (conjuntos) com excesso de categorias.
- Preditores verificados, no total.

Verificar Tabela de Nível de Medição

A tabela exibe os números de campos reformulados, divididos no seguinte:

- Campos ordinais (conjuntos ordenados) reformulados como campos contínuos.
- Campos contínuos reformulados como campos ordinais.
- Número total de reformulados.

Se nenhum campo de entrada (destino ou preditor) era contínuo ou ordinal, isso é mostrado como uma nota de rodapé.

Tabela de Valores Discrepantes

A tabela exibe contagens de quantos valores discrepantes foram manipulados.

- O número de campos contínuos para os quais os valores discrepantes foram localizados e aparados, ou o número de campos contínuos para os quais os valores discrepantes foram localizados e configurados como omissos, dependendo de suas configurações no painel Preparar Entradas e Destino na guia Configurações.
- O número de campos contínuos excluídos porque eram constantes, após a manipulação de valor discrepante.

Uma nota de rodapé mostra o valor de corte do valor discrepante, ao passo que outra nota de rodapé será mostrada se nenhum campo de entrada (destino ou preditor) era contínuo.

Tabela de Valores Omissos

A tabela exibe os números de campos que tiveram valores omissos substituídos, divididos em:

- Destino. Esta linha não será mostrada se nenhum destino for especificado.
- Preditores. Este é dividido ainda mais no número de nominais (conjuntos), ordinais (conjunto ordenado) e contínuos.
- O número total de valores omissos substituídos.

Tabela de Destino

A tabela exibe se o destino foi transformado, mostrado como:

- Transformação Box-Cox para normalidade. Este é dividido ainda mais em colunas que mostram os critérios especificados (desvio médio e padrão) e Lambda.
- Categorias de destino reordenadas para melhorar a estabilidade.

Tabela de Preditores Categóricos

A tabela exibe o número de preditores categóricos:

- Cujas categorias foram reordenadas da mais baixa para a mais alta para melhorar a estabilidade.
- Cujas categorias foram mescladas para maximizar a associação com o destino.
- Cujas categorias foram mescladas para manipular categorias de dispersão.
- Excluídos devido à baixa associação com o destino.
- Excluídos porque eram constantes após a mesclagem.

Uma nota de rodapé será mostrada se não houver nenhum preditor categórico.

Tabela de Preditores Contínuos

Há duas tabelas. A primeira exibe um dos seguintes números de transformações:

- Valores do preditor transformados para unidades padrão. Além disso, isso mostra o número de preditores transformados, a média especificada e o desvio padrão.
- Valores do preditor mapeados para um intervalo comum. Além disso, isso mostra o número de preditores transformados utilizando uma transformação min-max, bem como os valores mínimo e máximo especificados.
- Valores do preditor categorizados e o número de preditores categorizados.

A segunda tabela exibe os detalhes da construção de espaço do preditor, mostrados como o número de preditores:

- Construídos.
- Excluídos devido a uma baixa associação com o destino.
- Excluídos porque eram constantes após a categorização.
- Excluídos porque eram constantes após a construção.

Uma nota de rodapé será mostrada se nenhum preditor contínuo era de entrada.

Gerando um Nó Derivar

Ao gerar um nó Derivar, ele aplica a transformação de destino inversa no campo de escore. Por padrão, o nó insere o nome do campo de escore que seria produzido por um modelador automático (como

Classificador Automático ou Numeração Automática) ou pelo nó Combinação. Se um destino (intervalo) de escala tiver sido transformado, o campo de escore será mostrado em unidades transformadas, por exemplo, log(\$\$) ao invés de \$. Para interpretar e utilizar os resultados, deve-se converter o valor predito de volta para a escala original.

Nota: será possível gerar um nó Derivar somente quando o nó ADP contiver uma análise que escala novamente um destino do intervalo (ou seja, a nova escala Box-Cox é selecionada no painel Preparar Entradas e Destino). Não será possível gerar um nó Derivar se o destino não for um intervalo ou se a nova escala Box-Cox não estiver selecionada.

O nó Derivar é criado no modo Múltiplo e utiliza @FIELD na expressão para que seja possível incluir o destino transformado, se necessário. Por exemplo, utilizando os seguintes detalhes:

- Nome do campo de destino: response
- Nome do campo de destino transformado: response_transformed
- Nome do campo de escore: \$XR-response_transformed

O nó Derivar criaria um novo campo: \$XR-response_transformed_inverse.

Nota: se você não estiver utilizando um modelador automático ou o nó Combinação, será necessário editar o nó Derivar para transformar o campo de escore correto para seu modelo.

Variáveis resposta contínuas normalizadas

Por padrão, se marcar a caixa de seleção **Escalar novamente uma variável resposta contínua com uma transformação Box-Cox** no painel Preparar Entradas e Destino, isso transformará o destino e criará um novo campo que será o destino para sua construção de modelo. Por exemplo, se seu destino original era *response*, o novo destino será *response_transformed*; modelos posteriores do nó ADP selecionarão esse novo destino automaticamente.

No entanto, isso pode causar problemas, dependendo do destino original. Por exemplo, se o destino era *Age*, os valores do novo destino não serão *Years*, mas uma versão transformada de *Years*. Isso significa que não é possível examinar e interpretar os escores porque eles não estão em unidades reconhecíveis. Nesse caso, é possível aplicar uma transformação inversa que converterá as unidades transformadas novamente no seu formato original. Para fazer isso:

1. Após clicar em **Analisar Dados** para executar a análise ADP, selecione *Nó Derivar* no menu *Gerar*.
2. Coloque o nó de Derivar após o nugget na tela do modelo.

O nó Derivar restaurará o campo de escore para as dimensões originais de forma que a predição esteja nos valores de *Years* originais.

Por padrão, o nó Derivar transforma o campo de escore gerado por um modelador automático ou por um modelo combinado. Se estiver construindo um modelo individual, será necessário editar o nó Derivar para derivar a partir do campo de escore real. Se desejar avaliar seu modelo, deve-se incluir o destino transformado no campo **Derivar De** no nó Derivar. Isso aplica a mesma transformação inversa ao destino e qualquer nó Avaliação ou Análise posterior usará os dados transformados corretamente, desde que você altere esses nós para utilizar nomes de campos ao invés de metadados.

Se também desejar restaurar o nome original, será possível utilizar um nó Filtro para remover o campo de destino original se ele ainda estiver lá e renomear os campos de destino e de escore.

Nó Tipo

As propriedades do campo podem ser especificadas em um nó de origem ou em um nó Tipo separado. A funcionalidade é semelhante nos dois nós. As seguintes propriedades estão disponíveis:

- **Campo** Dê um clique duplo em qualquer nome de campo para especificar rótulos de valor e de campo para dados no IBM SPSS Modeler. Por exemplo, os metadados do campo importados a partir do IBM SPSS Statistics podem ser visualizados ou modificados aqui. Da mesma forma, é possível criar novos rótulos para campos e seus valores. Os rótulos que forem especificados aqui são exibidos em todo o IBM SPSS Modeler, dependendo das seleções feitas na caixa de diálogo de propriedades do fluxo.
- **Medição** Este é o nível de medição, utilizado para descrever características dos dados em um determinado campo. Se todos os detalhes de um campo forem conhecidos, ele será chamado de **totalmente instanciado**. Para obter mais informações, consulte “Níveis de Medição” na página 133.

Nota: O nível de medição de um campo é diferente de seu tipo de armazenamento, que indica se os dados são armazenados como sequências de caracteres, números inteiros, números reais, datas, horas, registros de data e hora ou listas.

- **Valores** Esta coluna permite especificar opções para ler os valores de dados a partir do conjunto de dados ou utilizar a opção **Especificar** para especificar os níveis de medição e valores em uma caixa de diálogo separada. Também é possível escolher transmitir campos sem ler seus valores. Para obter mais informações, consulte “Valores dos Dados” na página 137.

Nota: Não será possível corrigir a célula nesta coluna se a entrada **Campo** correspondente contiver uma lista.

- **Omisso** Utilizada para especificar como os valores omissos para o campo serão manipulados. Para obter mais informações, consulte “Definindo Valores Omissos” na página 141.

Nota: Não será possível corrigir a célula nesta coluna se a entrada **Campo** correspondente contiver uma lista.

- **Verificar** Nesta coluna, é possível configurar opções para assegurar que os valores de campo estejam em conformidade com os valores ou intervalos especificados. Para obter mais informações, consulte “Verificando Valores de Tipo” na página 142.

Nota: Não será possível corrigir a célula nesta coluna se a entrada **Campo** correspondente contiver uma lista.

- **Papel** Utilizado para informar aos nós de modelagem se os campos serão **Entrada** (campos preditores) ou **Destino** (campos preditos) para um processo de aprendizado por máquina. **Ambos** e **Nenhum** também são funções disponíveis, junto de **Partição**, que indica um campo usado para registros de partição em amostras separadas para treinamento, teste e validação. O valor **Divisão** especifica que os modelos separados serão construídos para cada valor possível do campo. Para obter mais informações, consulte “Configurando o Papel de Campo” na página 142.

Várias outras opções podem ser especificadas utilizando a janela do nó Tipo:

- Utilizando o botão do menu de ferramentas, é possível escolher **Ignorar Campos Únicos** quando um nó Tipo tiver sido instanciado (através suas especificações, valores de leitura ou executando o fluxo). Ignorar campos únicos ignora automaticamente campos com apenas um valor.
- Utilizando o botão do menu de ferramentas, é possível escolher **Ignorar Conjuntos Grandes** quando um nó Tipo tiver sido instanciado. Ignorar conjuntos grandes ignora automaticamente conjuntos com um número grande de membros.
- Utilizando o botão do menu de ferramentas, é possível escolher **Converter Números Inteiros Contínuos em Ordinais Para** quando um nó Tipo tiver sido instanciado. Consulte o tópico “Convertendo Dados Contínuos” na página 136 para obter mais informações.
- Utilizando o botão do menu de ferramentas, é possível gerar um nó Filtro para descartar campos selecionados.
- Utilizando os botões de comutação óculos de sol, é possível configurar o padrão para todos os campos para Ler ou Transmitir. A guia Tipos no nó de origem transmite campos por padrão, ao passo que o próprio nó Tipo lê valores por padrão.

- Utilizando o botão **Limpar Valores**, é possível limpar as mudanças nos valores de campo feitas nesse nó (valores não herdados) e ler novamente valores a partir das operações de envio de dados. Esta opção é útil para reconfigurar as mudanças que podem ter sido feitas para campos específicos anteriormente.
- Usando o botão **Limpar Todos os Valores**, é possível reconfigurar valores para **todos** os campos lidos no nó. Esta opção configura efetivamente a coluna **Valores** para **Ler** para todos os campos. Essa opção é útil para reconfigurar valores para todos os campos e ler novamente valores e tipos de operações de envio de dados.
- Utilizando o menu de contexto, é possível optar por **Copiar** os atributos de um campo para outro. Consulte o tópico “Copiando Atributos de Tipo” na página 143 para obter mais informações.
- Utilizando a opção **Visualizar configurações de campos não utilizados**, é possível visualizar as configurações de tipo para campos que não estiverem mais presentes nos dados ou que se conectaram uma vez a esse nó Tipo. Isso é útil ao reutilizar um nó Tipo para conjuntos de dados que foram alterados.

Níveis de Medição

O nível de medição (anteriormente conhecido como "conhecido" ou "tipo de uso") descreve o uso dos campos de dados no IBM SPSS Modeler. O nível de medição pode ser especificado na guia Tipos de uma origem ou nó Tipo. Por exemplo, talvez você queira configurar o nível de medição para um campo de número inteiro com valores de 1 e 0 para *Flag*. Isto geralmente indica que 1 = *True* e 0 = *False*.

Armazenamento versus medição. Observe que o nível de medição de um campo é diferente de seu tipo de armazenamento, que indica se os dados são armazenados como uma sequência de caracteres, número inteiro, número real, data, hora ou registro de data e hora. Embora os tipos de dados possam ser modificados em qualquer ponto em um fluxo utilizando um nó Tipo, o armazenamento deve ser determinado na origem durante a leitura dos dados no IBM SPSS Modeler (embora ele possa ser subsequentemente alterado utilizando uma função de conversão). Consulte o tópico “Configurando Armazenamento e Formatação do Campo” na página 9 para obter mais informações.

Alguns nós de modelagem indicam os tipos de nível de medição permitidos para seus campos de entrada e de destino por meio dos ícones em suas guias Campos.

Ícones do nível de medição

Tabela 19. Ícones do nível de medição

Ícone	Nível de medição
	Padrão
	Contínuo
	Categórico
	Sinalizador
	Nominal
	Ordinal
	Sem tipo
	Coleção

Tabela 19. Ícones do nível de medição (continuação)

Ícone	Nível de medição
	Geo-espacial

Os níveis de medida a seguir estão disponíveis:

- **Padrão** Dados cujo tipo de armazenamento e valores são desconhecidos (por exemplo, porque não foram lidos ainda) estão disponíveis como **<Padrão>**.
- **Contínuo** Usado para descrever valores numéricos, como um intervalo de 0-100 ou 0,75-1,25. Um valor contínuo pode ser um número inteiro, número real ou data/hora.
- **Catagórico** Usado para valores da sequência de caracteres quando um número exato de valores distintos é desconhecido. Este é um tipo de dados **não instanciado**, o que significa que todas as informações possíveis sobre o armazenamento e o uso dos dados ainda não são conhecidas. Uma vez que os dados são lidos, o nível de medição será *Flag*, *Nominal* ou *Sem tipo*, dependendo do número máximo de membros para campos nominais especificados na caixa de diálogo Propriedades do fluxo.
- **Flag** Usado para dados com dois valores distintos que indicam a presença ou ausência de um traço, como true e false, Yes e No ou 0 e 1. Os valores usados podem variar, mas um sempre deve ser designado como o valor "true", e o outro como o valor "false". Os dados podem ser representados como texto, número inteiro, número real, data, hora ou registro de data e hora.
- **Nominal** Usado para descrever dados com múltiplos valores distintos, cada um deles tratados como um membro de um conjunto como small/medium/large. Os dados nominais podem ter qualquer armazenamento—numérico, sequência de caracteres ou data/hora. Observe que configurar o nível de medição como *Nominal* não altera automaticamente os valores do armazenamento da sequência de caracteres.
- **Ordinal** Usado para descrever dados com múltiplos valores distintos que têm uma ordem inerente. Por exemplo, categorias de salário ou ranqueamentos de satisfação podem ser digitados como dados ordinais. A ordem é definida pela ordem de ordenação natural dos elementos de dados. Por exemplo, 1, 3, 5 é o padrão de ordenação para um conjunto de números inteiros, enquanto HIGH, LOW, NORMAL (ordem ascendente alfabética) é a ordem para um conjunto de configurações. O nível de medição ordinal permite que você defina um conjunto de dados categóricos como dados ordinais para as finalidades de visualização, construção de modelo e exportação para outros aplicativos (como IBM SPSS Statistics) que reconhecem dados ordinais como um tipo distinto. Você pode usar um campo ordinal em qualquer lugar em que um campo nominal pode ser usado. Além disso, campos de qualquer tipo de armazenamento (real, número inteiro, sequência de caracteres, data, hora e assim por diante) pode ser definido como ordinal.
- **Sem tipo** Usado para dados que não se adequam a qualquer um dos tipos acima, para campos com um valor único ou para dados nominais em que o conjunto tem mais membros do que o máximo definido. Também é útil para casos em que o nível de medição seria de outra forma configurado com muitos membros (como um número de conta). Quando você seleciona **Sem tipo** para um campo, a função é automaticamente configurada como **Nenhuma**, com **ID de registro** como a única alternativa. O tamanho máximo padrão para conjuntos é de 250 valores exclusivos. Esse número pode ser ajustado ou desativado na guia Opções da caixa de diálogo Propriedades do fluxo que pode ser acessada a partir do menu Ferramentas.
- **Coleta** Usado para identificar dados não geo-espaciais que são gravados em uma lista. Uma coleta é efetivamente um campo de lista de zero de profundidade, em que os elementos nessa lista têm um dos outros níveis de medição.
Para obter mais informações sobre listas, consulte o tópico Armazenamento de lista e níveis de medição associados na seção Nós de origem do SPSS Modeler Source, Process, and Output Nodes Guide.
- **Geo-espacial** Usado com o tipo de armazenamento Lista para identificar dados geo-espaciais. As listas podem ser uma Lista de número inteiro ou uma Lista de campos reais com uma espessura de lista que esteja entre zero e dois, inclusive.

Para obter mais informações, consulte o tópico Geospatial Measurement Sublevels na seção Nó Tipo do SPSS Modeler Source, Process, and Output Nodes Guide.

É possível especificar manualmente os níveis de medição ou permitir que o software leia os dados e determine o nível de medição com base nos valores que ele lê.

Como alternativa, quando você tiver vários campos de dados contínuos que devem ser tratados como dados categóricos, será possível escolher uma opção para convertê-los. Consulte o tópico “Convertendo Dados Contínuos” na página 136 para obter mais informações.

Para Uso de Digitação Automática

1. Em um nó Tipo ou na guia Tipos de um nó de origem, configure a coluna *Valores* para <Read> para os campos desejados. Isso tornará os metadados disponíveis para todos os nós de recebimento de dados. É possível configurar rapidamente todos os campos para <Read> ou <Pass> usando botões de óculos de sol na caixa de diálogo.
2. Clique em **Ler Valores** para ler valores da origem de dados imediatamente.

Para Configurar Manualmente o Nível de Medição de um Campo

1. Selecione um campo na tabela.
2. Na lista suspensa na coluna *Medição*, selecione um nível de medição para o campo.
3. Como alternativa, é possível utilizar Ctrl-A ou Ctrl-clique para selecionar diversos campos antes de utilizar a lista suspensa para selecionar um nível de medição.

Subníveis de medição Geoespacial

O nível de medição Geoespacial, que é utilizado com o tipo de armazenamento de Lista, possui seis subníveis que são utilizados para identificar diferentes tipos de dados geoespaciais.

- **Point** – Identifica uma localização específica; por exemplo, o centro de uma cidade.
- **Polygon** – Uma série de pontos que identifica o único limite de uma região e sua localização; por exemplo, um estado.
- **LineString** – Também conhecida como uma Polilinha ou apenas uma Linha, uma LineString é uma série de pontos que identifica a rota de uma linha. Por exemplo, uma LineString pode ser um item fixo, como rodovia, rio ou via férrea, ou a trilha de algo que se move, como uma rota de voo de um avião ou a navegação de um navio.
- **MultiPoint** – Utilizado quando cada linha em seus dados contém diversos pontos por região. Por exemplo, se cada linha representar uma rua da cidade, os diversos pontos para cada rua poderão ser utilizados para identificar cada poste de iluminação dessa rua.
- **MultiPolygon** – Utilizado quando cada linha em seus dados contém vários polígonos. Por exemplo, se cada linha representar o contorno de um país, os EUA poderão ser registrados como vários polígonos para identificar áreas diferentes, como o continente principal, o Alasca e o Havaí.
- **MultiLineString** – Utilizado quando cada linha em seus dados contém várias linhas. Como as linhas não podem se ramificar, é possível utilizar uma MultiLineString para identificar um grupo de linhas. Por exemplo, dados como hidrovias ou a malha ferroviária em cada país.

Esses subníveis de medição são utilizados com o tipo de armazenamento Lista. Para obter informações adicionais, consulte “Armazenamento de lista e níveis de medição associados” na página 11.

Restrições

Deve-se ter em mente algumas restrições quando utilizar dados geoespaciais.

- O sistema de coordenadas pode afetar o formato dos dados. Por exemplo, um sistema de coordenadas projetadas utiliza os valores de coordenadas x, y e (quando necessário) z, ao passo que um sistema de coordenadas Geográficas utiliza os valores de coordenadas longitude, latitude e (quando necessário) um valor para altitude ou profundidade.

Para obter mais informações sobre sistemas de coordenadas, consulte o tópico Configurando Opções Geoespaciais para Fluxos na seção Trabalhando com Fluxos do Guia do Usuário do SPSS Modeler.

- Uma LineString não pode cruzar a si mesma.
- Um Polygon não possui fechamento automático; para cada Polygon, assegure-se de que o primeiro e o segundo pontos sejam definidos como os mesmos.
- A direção dos dados em um MultiPolygon é importante; o sentido horário indica uma forma sólida e o sentido anti-horário indica um buraco. Por exemplo, se você registrar uma área de um país que contém lagos, a borda da área do continente principal poderá ser registrada no sentido horário e o formato de cada lago no sentido anti-horário.
- Um Polygon não pode intersectar com si mesmo. Um exemplo dessa interseção seria se você tentasse representar o limite do polígono como uma linha contínua no formato da figura 8.
- MultiPolygons não podem sobrepor um ao outro.
- Para campos Geoespaciais, os únicos tipos de armazenamento relevantes são **Real** e **Número Inteiro** (a configuração padrão é **Real**).

Ícones de subnível de medição Geoespacial

Tabela 20. Ícones de subnível de medição Geoespacial

Ícone	Nível de medição
	Point
	Polygon
	LineString
	MultiPoint
	MultiPolygon
	MultiLineString

Convertendo Dados Contínuos

O tratamento de dados categóricos como contínuos pode ter um sério efeito na qualidade de um modelo, especialmente se ele for o campo de destino; por exemplo, produzir um modelo de regressão ao invés de um modelo binário. Para evitar isso, é possível converter intervalos de números inteiros em tipos categóricos, como *Ordinal* ou *senalização*.

1. A partir do botão de menu Operações e Gerar (com o símbolo da ferramenta), selecione **Converter Números Inteiros Contínuos em Ordinais**. O diálogo de valores de conversão é exibido.
2. Especifique o tamanho do intervalo que será convertido automaticamente; isto se aplica a qualquer intervalo até e incluindo o tamanho que você inserir.
3. Clique em **OK**. Os intervalos afetados são convertidos em um *senalização* ou *Ordinal* e exibidos na guia Tipos do nó Tipo.

Resultados da Conversão

- Quando um campo *Contínuo* com armazenamento de número inteiro é alterado para *Ordinal*, os valores inferior e superior são expandidos para incluir todos os valores de número inteiro do valor mais baixo para o valor mais alto. Por exemplo, se o intervalo for 1, 5, o conjunto de valores será 1, 2, 3, 4, 5.
- Se o campo *Contínuo* alterar para *senalização*, os valores inferior e superior se tornarão os valores falso e verdadeiro para o campo de sinalização.

O Que É Instanciação?

Instanciação é o processo de leitura ou de especificação de informações, como tipo de armazenamento e valores para um campo de dados. Para otimizar os recursos do sistema, a instanciação é um processo direcionado ao usuário, em que você informa o software para ler valores ao especificar as opções na guia Tipos em um nó de origem ou executar dados por meio de um nó Tipo.

- Dados com tipos desconhecidos também são referidos como **não instanciados**. Dados cujo tipo de armazenamento e valores são desconhecidos são exibidos na coluna *Medição* da guia Tipos como **<Default>**.
- Quando você tem algumas informações sobre um armazenamento de campo, como sequência de caracteres ou numérico, os dados são chamados **parcialmente instanciados**. **Catégorico** ou **Contínuo** são níveis de medição parcialmente instanciados. Por exemplo, **Catégorico** especifica que o campo é simbólico, mas não se sabe se ele é nominal, ordinal ou de flag.
- Quando todos os detalhes sobre um tipo são conhecidos, incluindo os valores, um nível de medição **totalmente instanciado**, como nominal, ordinal, de sinalização ou contínuo, é exibido nesta coluna.
Nota: o tipo *contínuo* é utilizado para campos de dados parcialmente instanciados e totalmente instanciados. Os dados Contínuos podem ser números inteiros ou números reais.

Durante a execução de um fluxo de dados com um nó Tipo, tipos não instanciados se tornam imediatamente parcialmente instanciados com base nos valores de dados iniciais. Quando todos os dados tiverem passado pelo nó, todos os dados se tornarão totalmente instanciados, a menos que os valores tenham sido configurados para **<Pass>**. Se a execução for interrompida, os dados permanecerão parcialmente instanciados. Após a guia Tipos ter sido instanciada, os valores de um campo serão estáticos naquele ponto no fluxo. Isso significa que quaisquer mudanças posteriores não afetarão os valores de um campo específico, mesmo se executar novamente o fluxo. Para alterar ou atualizar os valores com base em novos dados ou manipulações incluídas, é necessário editá-los na própria guia Tipos ou configurar o valor de um campo para **<Leitura>** ou **<Leitura +>**.

Quando Instanciar

Geralmente, se o seu conjunto de dados não for muito grande e você não planejar incluir campos posteriormente no fluxo, a instanciação no nó de origem é o método mais conveniente. No entanto, a instanciação em um nó Tipo separado é útil quando:

- O conjunto de dados for grande e o fluxo filtrar um subconjunto antes do nó Tipo.
- Os dados tiverem sido filtrados no fluxo.
- Os dados tiverem sido mesclados ou anexados no fluxo.
- Novos campos de dados são derivados durante o processamento.

Valores dos Dados

Utilizando a coluna *Valores* da guia Tipos, é possível ler valores automaticamente a partir dos dados ou especificar níveis de medição e valores em uma caixa de diálogo separada.

As opções disponíveis na lista suspensa Valores fornecem instruções para digitação automática, conforme mostrado na tabela a seguir.

Tabela 21. Instruções para digitação automática

Opção	Função
<Leitura>	Os dados serão lidos quando o nó for executado.
<Read+>	Os dados serão lidos e anexados aos dados atuais (se houver algum).
<Pass >	Nenhum dado é lido.
<Atual>	Mantém valores de dados atuais.

Tabela 21. Instruções para digitação automática (continuação)

Opção	Função
Especificar...	Uma caixa de diálogo separada é ativada para especificar opções de valores e de nível de medição.

Executar um Nó Tipo ou clicar em **Ler Valores** irá digitar ou ler valores automaticamente a partir de sua origem de dados com base em sua seleção. Esses valores também podem ser especificados manualmente utilizando a opção Especificar ou dando um clique duplo em uma célula na coluna *Campo*.

Após ter feito mudanças nos campos no nó Tipo, será possível reconfigurar as informações de valor utilizando os seguintes botões na barra de ferramentas da caixa de diálogo:

- Utilizando o botão **Limpar Valores**, é possível limpar as mudanças nos valores de campo feitas nesse nó (valores não herdados) e ler novamente valores a partir das operações de envio de dados. Esta opção é útil para reconfigurar as mudanças que podem ter sido feitas para campos específicos anteriormente.
- Usando o botão **Limpar Todos os Valores**, é possível reconfigurar valores para **todos** os campos lidos no nó. Esta opção configura efetivamente a coluna *Valores* para **Ler** para todos os campos. Essa opção é útil para reconfigurar valores para todos os campos e ler novamente valores e níveis de medição de operações de envio de dados.

Utilizando a Caixa de Diálogo de Valores

Clicar na coluna **Valores** ou **Omisso** da guia Tipos exibe uma lista suspensa de valores predefinidos. Escolher a opção **Especificar...** nesta lista abre uma caixa de diálogo separada na qual é possível configurar opções para leitura, especificação, rotulagem e manipulação de valores para o campo selecionado.

Muitos dos controles são comuns a todos os tipos de dados. Esses controles comuns são discutidos aqui.

Medição Exibe o nível de medição selecionado atualmente. É possível alterar a configuração para refletir a maneira com que você pretende utilizar os dados. Por exemplo, se um campo chamado *day_of_week* contiver números que representam dias individuais, talvez você queira alterar isto para dados nominais para criar um nó de distribuição que examina cada categoria individualmente.

Armazenamento Exibe o tipo de armazenamento, se conhecido. Os tipos de armazenamento não são afetados pelo nível de medição que você escolher. Para alterar o tipo de armazenamento, é possível utilizar a guia Dados nos nós de origem Arquivo Fixo e Arquivo de Variável ou uma função de conversão em um nó Preenchimento.

Campo de Modelo Para campos gerados como resultado da escoragem de um nugget do modelo, detalhes do campo de modelo também podem ser visualizados. Esses detalhes incluem o nome do campo de destino e também o papel do campo na modelagem (se um valor predito, probabilidade, propensão, e assim por diante).

Valores Selecione um método para determinar valores para o campo selecionado. As seleções que forem feitas aqui substituem quaisquer seleções feitas anteriormente a partir da coluna **Valores** da caixa de diálogo do nó Tipo. As opções para leitura de valores incluem o seguinte:

- **Ler a partir dos dados** Selecione para ler valores quando o nó é executado. Esta opção é a mesma que **<Read>**.
- **Transmitir** Selecione para não ler dados para o campo atual. Esta opção é a mesma que **<Pass>**.
- **Especificar valores e rótulos** As opções aqui são utilizadas para especificar valores e rótulos para o campo selecionado. Utilizada com a verificação de valor, use essa opção para especificar valores que são baseados no conhecimento do campo atual. Esta opção ativa controles exclusivos para cada tipo de campo. As opções para valores e rótulos são cobertas individualmente nos tópicos subsequentes.

Nota: Não é possível especificar valores ou rótulos para um campo cujo nível de medição é Sem Tipo ou <Default>.

- **Estender valores a partir dos dados** Selecione para anexar os dados atuais aos valores que você inserir aqui. Por exemplo, se `field_1` possuir um intervalo de (0,10), e você inserir uma série de valores de (8,16), o intervalo será estendido ao incluir o 16, sem remover o mínimo original. O novo intervalo será (0,16). Escolher esta opção configura automaticamente a opção de digitação automática para <Read+>.

Comprimento máx. da lista Disponível somente para os dados com um nível de medição de Geoespacial ou Coleção. Configure o comprimento máximo da lista ao especificar o número de elementos que a lista pode conter.

Verificar valores Selecione um método de forçar os valores de acordo com os valores contínuos, de flag ou nominais especificados. Esta opção corresponde à coluna **Verificar** na caixa de diálogo do nó Tipo, e as configurações feitas aqui substituem aquelas na caixa de diálogo. Utilizada com a opção **Especificar valores e rótulos**, a verificação de valor permite que os valores nos dados estejam em conformidade com os valores esperados. Por exemplo, se você especificar valores como 1, 0 e, em seguida, utilizar a opção **Descartar**, será possível descartar todos os registros com valores diferentes de 1 ou 0.

Definir em brancos Selecione para ativar os seguintes controles que você utiliza para declarar valores omissos ou em branco em seus dados.

- **Valores omissos** Utilize esta tabela para definir valores específicos (como 99 ou 0) como em branco. O valor deve ser apropriado para o tipo de armazenamento do campo.
- **Intervalo** Utilizado para especificar um intervalo de valores omissos, por exemplo, idades de 1 a 17 ou maiores que 65. Se um valor limite for deixado em branco, o intervalo será ilimitado, por exemplo, se um limite inferior de 100 for especificado sem um limite superior, então todos os valores maiores ou iguais a 100 serão definidos como omissos. Os valores limite são inclusivos, por exemplo, um intervalo com um limite inferior de 5 e um limite superior de 10 inclui 5 e 10 na definição do intervalo. Um intervalo de valores omissos pode ser definido para qualquer tipo de armazenamento, incluindo data/hora e sequência de caracteres (caso em que a ordenação alfabética é usada para determinar se um valor está dentro do intervalo).
- **Espaço Nulo/Em Branco** Também é possível especificar nulos do sistema (exibido nos dados como \$null\$) e espaço em branco (valores de sequência de caracteres sem caracteres visíveis) como em branco.

Nota: O nó Tipo também trata sequências de caracteres vazias como espaço em branco para propósitos de análise, embora elas sejam armazenadas de forma diferente internamente e possam ser manipuladas de forma diferente em determinados casos.

Nota: Para codificar em brancos como \$null\$ ou indefinidos, utilize o nó Preenchimento.

Descrição Utilize esta caixa de texto para especificar um rótulo de campo. Esses rótulos aparecem em vários locais, como em gráficos, tabelas, saída e navegadores de modelo, dependendo das seleções feitas na caixa de diálogo Propriedades do Fluxo.

Especificando Valores e Rótulos para Dados Contínuos

O nível de medição *Contínuo* é utilizado para campos numéricos. Há três tipos de armazenamento para dados contínuos:

- Real
- Número inteiro
- Data/hora

A mesma caixa de diálogo é utilizada para editar todos os campos contínuos; o tipo de armazenamento é exibido apenas para referência.

Especificando Valores

Os controles a seguir são exclusivos para campos contínuos e são usados para especificar um intervalo de valores:

Inferior. Especifique um limite inferior para o intervalo de valores.

Superior. Especifique um limite superior para o intervalo de valores.

Especificando Rótulos

É possível especificar rótulos para qualquer valor de um campo de intervalo. Clique no botão **Rótulos** para abrir uma caixa de diálogo separada para especificar rótulos de valores.

Subcaixa de Diálogo de Valores e Rótulos: Clicar em **Rótulos** na caixa de diálogo Valores de um campo de intervalo abre uma nova caixa de diálogo na qual é possível especificar rótulos para qualquer valor no intervalo.

É possível utilizar as colunas *Valores* e *Rótulos* nesta tabela para definir pares de valores e rótulos. Os pares definidos atualmente são mostrados aqui. É possível incluir novos pares de rótulos ao clicar em uma célula vazia e inserir um valor e seu rótulo correspondente. *Nota:* incluir pares de valor/valor-rótulo nessa tabela não fará com que nenhum novo valor seja incluído no campo. Ao invés disso, ele simplesmente cria metadados para o valor do campo.

Os rótulos que forem especificados no nó Tipo são exibidos em muitos locais (como ToolTips, rótulos de saída, e assim por diante), dependendo das seleções feitas na caixa de diálogo de propriedades do fluxo.

Especificando Valores e Rótulos para Dados Nominiais e Ordinais

Os níveis de medição Nominal (conjunto) e ordinal (conjunto ordenado) indicam que os valores de dados são utilizados discretamente como um membro do conjunto. Os tipos de armazenamento para um conjunto podem ser sequência de caracteres, número inteiro, número real ou data/hora.

Os controles a seguir são exclusivos para campos nominiais e ordinais e são utilizados para especificar os valores e rótulos:

Valores. A coluna *Valores* na tabela permite especificar os valores com base em seu conhecimento do campo atual. Utilizando esta tabela, é possível inserir valores esperados para o campo e verificar a conformidade do conjunto de dados com estes valores utilizando a lista suspensa Verificar Valores. Usando os botões de seta e exclusão, é possível modificar valores existentes, bem como reordenar ou excluir valores.

Rótulos. A coluna *Rótulos* permite especificar rótulos para cada valor no conjunto. Esses rótulos aparecem em várias localizações, como em gráficos, tabelas, saída e navegadores de modelo, dependendo das seleções feitas na caixa de diálogo de propriedades do fluxo.

Especificando Valores para um Flag

Os campos de flag são utilizados para exibir dados que tiverem dois valores distintos. Os tipos de armazenamento para flags podem ser sequência de caracteres, número inteiro, número real ou data/hora.

True. Especifique um valor de flag para o campo quando a condição for atendida.

False. Especifique um valor de flag para o campo quando a condição não for atendida.

Rótulos. Especifique rótulos para cada valor no campo de sinalização. Esses rótulos aparecem em várias localizações, como em gráficos, tabelas, saída e navegadores de modelo, dependendo das seleções feitas na caixa de diálogo de propriedades do fluxo.

Especificando Valores para Dados de Coleção

Os campos de coleção são utilizados para exibir dados não geoespaciais que estiverem em uma lista.

O único item que pode ser configurado para o nível **Medição** da Coleção é **Medida de lista**. Por padrão, esta medida é configurada para Sem Tipo, mas é possível selecionar outro valor para configurar o nível de medição dos elementos dentro da lista. É possível escolher uma das seguintes opções:

- Sem tipo
- Contínuo
- Nominal
- Ordinal
- Sinalizador

Especificando Valores para Dados Geoespaciais

Os campos geoespaciais são utilizados para exibir dados geoespaciais que estiverem em uma lista.

Para o nível de **Medição** Geoespacial, é possível configurar as seguintes opções para configurar o nível de medição dos elementos na lista:

Tipo Selecione o subnível de medição do campo geoespacial. Os subníveis disponíveis são determinados pela profundidade do campo de lista, e os padrões são: Point (profundidade zero), LineString (profundidade de um) e Polygon (profundidade de um).

Para obter mais informações sobre subníveis, consulte “Subníveis de medição Geoespacial” na página 135.

Para obter mais informações sobre as profundidades de lista, consulte “Armazenamento de lista e níveis de medição associados” na página 11.

Sistema de coordenadas Esta opção estará disponível somente se você alterou o nível de medição para geoespacial a partir de um nível não geoespacial. Para aplicar um sistema de coordenadas aos seus dados geoespaciais, marque essa caixa de seleção. Por padrão, o sistema de coordenadas configurado na área de janela **Ferramentas > Propriedades do Fluxo > Opções > Geoespacial** é mostrado. Para utilizar um sistema de coordenadas diferente, clique no botão **Alterar** para exibir a caixa de diálogo Selecionar Sistema de Coordenadas e escolha o sistema que precisar.

Para obter mais informações sobre sistemas de coordenadas, consulte o tópico Configurando Opções Geoespaciais para Fluxos na seção Trabalhando com Fluxos do Guia do Usuário do SPSS Modeler.

Definindo Valores Omissos

A coluna **Omisso** da guia Tipos indica se uma manipulação de valor omissos foi definida para um campo. As configurações possíveis são:

Ligado (*). Indica que a manipulação de valores omissos está definida para este campo. Isso pode ser feito por meio de um nó Preenchimento de recebimento de dados ou por meio de uma especificação explícita utilizando a opção Especificar (veja abaixo).

Desligado. O campo não tem nenhuma manipulação de valor omissos definida.

Especificar. Escolha essa opção para exibir um diálogo no qual é possível declarar que valores explícitos sejam considerados como valores omissos para esse campo.

Verificando Valores de Tipo

Ativar a opção Verificar para cada campo examina todos os valores nesse campo para determinar se eles estão de acordo com as configurações de tipo atuais ou os valores que você especificou na caixa de diálogo Especificar Valores. Isso é útil para limpar conjuntos de dados e reduzir o tamanho de um conjunto de dados em uma única operação.

A configuração da coluna *Verificar* na caixa de diálogo do Tipo determina o que acontece quando um valor fora dos limites do tipo é descoberto. Para alterar as configurações de Verificar para um campo, utilize a lista suspensa para esse campo na coluna *Verificar*. Para definir as configurações de Verificar para todos os campos, clique na coluna *Campo* e pressione Ctrl-A. Em seguida, use a lista suspensa para qualquer campo na coluna *Verificar*.

As configurações de Verificar a seguir estão disponíveis:

Nenhum. Os valores serão transmitidos sem verificação. Essa é a configuração padrão.

Anular. Altere valores fora dos limites para o sistema nulo (\$null\$).

Forçar. Campos cujos níveis de medição estiverem totalmente instanciados serão verificados em busca de valores que estiverem fora dos intervalos especificados. Valores não especificados serão convertidos em um valor legal para esse nível de medição utilizando as regras a seguir:

- Para flags, qualquer valor diferente dos valores true e false é convertido para o valor false.
- Para conjuntos (nominal ou ordinal), qualquer valor desconhecido é convertido para o primeiro membro de valores do conjunto.
- Números maiores que o limite superior de um intervalo são substituídos pelo limite superior.
- Números menores que o limite inferior de um intervalo são substituídos pelo limite inferior.
- Valores nulos em um intervalo recebem o valor de ponto médio para esse intervalo.

Descartar. Quando valores ilegais são localizados, o registro inteiro é descartado.

Avisar. O número de itens ilegais é contado e relatado na caixa de diálogo de propriedades do fluxo quando todos os dados tiverem sido lidos.

Interromper. O primeiro valor ilegal encontrado termina a execução do fluxo. O erro é relatado na caixa de diálogo de propriedades do fluxo.

Configurando o Papel de Campo

O papel de um campo especifica como ele é utilizado na construção de modelo, por exemplo, se um campo for uma entrada ou um destino (aquilo que está sendo predito).

Nota: os papéis de ID Partição, Frequência e Registro podem, cada um, ser aplicados a um único campo somente.

Os papéis a seguir estão disponíveis:

Entrada. O campo será utilizado como uma entrada para aprendizado de máquina (um campo preditor).

Resposta. O campo será utilizado como uma saída ou destino para aprendizado de máquina (um dos campos que o modelo tentará prever).

Ambos. O campo será utilizado como uma entrada e uma saída pelo nó a priori. Todos os outros nós de modelagem ignorarão o campo.

Nenhum. O campo será ignorado pelo aprendizado de máquina. Campos cujo nível de medição foi configurado para **Sem tipo** são configurados automaticamente para **Nenhum** na coluna *Papel*.

Partição. Indica um campo utilizado para particionar os dados em amostras separadas para propósitos de treinamento, de teste e de validação (opcional). O campo deve ser um tipo de conjunto instanciado com dois ou três valores possíveis (conforme definido na caixa de diálogo Valores Campo). O primeiro valor representa a amostra de treinamento, o segundo representa a amostra de teste e o terceiro (se presente) representa a amostra de validação. Quaisquer valores adicionais são ignorados, e campos de flag não podem ser utilizados. Observe que para utilizar a partição em uma análise, o particionamento deve ser ativado na guia Opções de Modelo no nó de construção ou de análise de modelo apropriado. Os registros com valores nulos para o campo de partição são excluídos da análise quando o particionamento é ativado. Se diversos campos de partição tiverem sido definidos no fluxo, um campo de partição único deverá ser especificado na guia Campos em cada nó de modelagem aplicável. Se um campo adequado ainda não existir em seus dados, será possível criar um utilizando um nó Partição ou um nó Derivar. Consulte o tópico “Nó Partição” na página 170 para obter mais informações.

Divisão. (Campos nominais, ordinais e de flag somente) Especifica que um modelo deve ser construído para cada valor possível do campo.

Frequência. (Campos numéricos somente) Configurar esse papel permite que o valor do campo seja utilizado como um fator de ponderação de frequência para o registro. Esse recurso é suportado pelos modelos Árvore C&R, CHAID, QUEST e Linear, e todos os outros nós ignoraram esse papel. A ponderação de frequência é ativada por meio da opção **Utilizar ponderação de frequência** na guia Campos dos nós de modelagem que suportam o recurso.

ID do registro. O campo será utilizado como o identificador de registro exclusivo. Esse recurso é ignorado pela maioria dos nós, no entanto, ele é suportado pelos modelos Lineares e é necessário para os nós de mineração dentro da base de dados do IBM Netezza.

Copiando Atributos de Tipo

É possível copiar facilmente os atributos de um tipo, como valores, opções de verificação e valores omissos de um campo para outro:

1. Clique com o botão direito no campo cujos atributos você deseja copiar.
2. No menu de contexto, escolha **Copiar**.
3. Clique com o botão direito em um ou mais campos cujos atributos deseja alterar.
4. No menu de contexto, escolha **Colar Especial**. *Nota:* é possível selecionar diversos campos utilizando o método Ctrl-clique ou utilizando a opção **Selecionar Campos** no menu de contexto.

Uma nova caixa de diálogo é aberta, na qual é possível selecionar os atributos especificados que deseja colar. Se estiver colando em diversos campos, as opções que forem selecionadas aqui serão aplicadas a todos os campos de destino.

Colar os seguintes atributos. Selecione na lista abaixo para colar os atributos de um campo para outro.

- **Tipo.** Selecione para colar o nível de medição.
- **Valores.** Selecione para colar os valores do campo.
- **Omisso.** Selecione para colar configurações de valores omissos.
- **Verificar.** Selecione para colar opções de verificação de valor.
- **Papel.** Selecione para colar o papel de um campo.

Guia Configurações de Formato de Campo

A guia Formato nos nós de Tabela e Tipo mostra uma lista de campos atuais ou não utilizados e as opções de formatação para cada campo. A seguir há uma descrição de cada coluna na tabela de formatação de campo:

Campo. Isso mostra o nome do campo selecionado.

Formato. Dê um clique duplo em uma célula nessa coluna para especificar formatação para campos em uma base individual utilizando a caixa de diálogo que é aberta. Consulte o tópico “Configurando Opções de Formato de Campo” para obter mais informações. A formatação especificada aqui substitui a formatação especificada nas propriedades do fluxo geral.

Nota: os nós Exportação do Statistics e Saída do Statistics exportam arquivos *.sav* que incluem a formatação por campo em seus metadados. Se um formato por campo não suportado pelo formato de arquivo *.sav* do IBM SPSS Statistics for especificado, o nó utilizará o formato padrão do IBM SPSS Statistics.

Justificar. Utilize esta coluna para especificar como os valores devem ser justificados dentro da coluna da tabela. A configuração padrão é **Automático**, que justifica à esquerda valores simbólicos e à direita valores numéricos. É possível substituir o padrão ao selecionar **Esquerda**, **Direita** ou **Centro**.

Largura da Coluna. Por padrão, as larguras das colunas são calculadas automaticamente com base nos valores do campo. Para substituir o cálculo de largura automático, clique em uma célula da tabela e utilize a lista suspensa para selecionar uma nova largura. Para inserir uma largura customizada não listada aqui, abra a caixa do subdiálogo Formatos de Campo ao dar um clique duplo em uma célula da tabela na coluna Campo ou Formato. Como alternativa, é possível clicar com o botão direito em uma célula e escolher **Configurar Formato**.

Visualizar Campos Atuais. Por padrão, a caixa de diálogo mostra a lista de campos ativos no momento. Para visualizar a lista de campos não utilizados, selecione **Visualizar configurações de campos não utilizados**.

Menu de Contexto. O menu de contexto para esta guia fornece várias opções de seleção e de atualização de configuração. Clique com o botão direito em uma coluna para exibir este menu.

- **Selecionar Tudo.** Seleciona todos os campos.
- **Selecionar Nenhum.** Limpa a seleção.
- **Selecionar Campos.** Seleciona campos com base em características de tipo ou de armazenamento. As opções são **Selecionar Categóricos**, **Selecionar Contínuos** (numéricos), **Selecionar Sem Tipo**, **Selecionar Sequências de Caracteres**, **Selecionar Números** ou **Selecionar Data/Hora**. Consulte o tópico “Níveis de Medição” na página 133 para obter mais informações.
- **Configurar Formato.** Abre uma subcaixa de diálogo para especificar opções de data, hora e decimais em uma base por campo.
- **Configurar Justificação.** Configura a justificação para o campo selecionado. As opções são **Automático**, **Centro**, **Esquerda** ou **Direita**.
- **Configurar Largura da Coluna.** Configura a largura do campo para campos selecionados. Especifique **Automático** para ler a largura dos dados. Ou é possível configurar a largura do campo para 5, 10, 20, 30, 50, 100 ou 200.

Configurando Opções de Formato de Campo

A formatação de campo é especificada em uma subcaixa de diálogo disponível a partir da guia Formato nos nós Tipo e Tabela. Se você tiver selecionado mais de um campo antes de abrir esta caixa de diálogo, então as configurações do primeiro campo na seleção serão utilizadas para todos. Clicar em **OK** após fazer as especificações aqui aplicará essas configurações a todos os campos selecionados na guia Formato.

As opções a seguir estão disponíveis em uma base por campo. Muitas dessas configurações também podem ser especificadas na caixa de diálogo de propriedades do fluxo. Todas as configurações feitas no nível de campo substituem o padrão especificado para o fluxo.

Formato de data. Selecione um formato de data para ser usado para campos de armazenamento de data ou quando sequências de caracteres forem interpretadas como datas pelas funções de data do CLEM.

Formato de hora. Selecione um formato de hora para ser usado para campos de armazenamento de hora ou quando sequências de caracteres forem interpretadas como horas pelas funções de hora do CLEM.

Formato de exibição numérico. É possível escolher entre os formatos de exibição padrão (####.###), científico (#.###E+##) ou moeda (\$###.##).

Símbolo decimal. Selecione uma vírgula (,) ou um ponto (.) como um separador decimal.

Símbolo de Agrupamento. Para formatos de exibição de número, selecione o símbolo usado para agrupar valores (por exemplo, a vírgula em 3.000,00). As opções incluem nenhum, ponto final, vírgula, espaço e definido por código de idioma (caso em que o padrão para o código de idioma atual é usado).

Casas decimais (padrão, científica, moeda, exportação). Para formatos de exibição numéricos, especifica o número de casas decimais a serem usadas durante a exibição de números reais. Essa opção é especificada separadamente para cada formato de exibição.

Justificar. Especifica como os valores devem ser justificados dentro da coluna. A configuração padrão é **Automático**, que justifica à esquerda valores simbólicos e à direita valores numéricos. É possível substituir o padrão ao selecionar esquerda, direita ou centro.

Largura da coluna. Por padrão, as larguras das colunas são calculadas automaticamente com base nos valores do campo. É possível especificar uma largura customizada em intervalos de cinco usando as setas à direita da caixa de listagem.

Filtrando ou Renomeando Campos

É possível renomear ou excluir campos em qualquer ponto em um fluxo. Por exemplo, como pesquisador médico, talvez você não esteja preocupado com o nível de potássio (dados de nível de campo) de pacientes (dados de nível de registro); portanto, é possível filtrar o campo *K* (potássio). Isso pode ser feito utilizando um nó Filtro separado ou utilizando a guia Filtro em um nó de origem ou de saída. A funcionalidade é a mesma, independentemente de qual nó ele estiver sendo acessado.

- Nos nós de origem, como Arquivo de Variável, Fixo Arquivo, Arquivo do Statistics e XML, é possível renomear ou filtrar os campos conforme os dados são lidos no IBM SPSS Modeler.
- Utilizando um nó Filtro, é possível renomear ou filtrar os campos em qualquer ponto no fluxo.
- Nos nós Exportação do Statistics, Transformação do Statistics, Modelo do Statistics e Saída do Statistics, é possível filtrar ou renomear campos para conformidade com os padrões de nomenclatura do IBM SPSS Statistics. Consulte o tópico “Renomeando ou Filtrando Campos para o IBM SPSS Statistics” na página 378 para obter mais informações.
- É possível utilizar a guia Filtro em qualquer um dos nós acima para definir ou editar conjuntos de múltiplas respostas. Consulte o tópico “Editando Conjuntos de Múltiplas Respostas” na página 147 para obter mais informações.
- Por fim, é possível utilizar um nó Filtro para mapear os campos de um nó de origem para outro.

Configurando Opções de Filtragem

A tabela utilizada na guia Filtro mostra o nome de cada campo conforme ele entra e sai do nó. É possível utilizar as opções nesta tabela para renomear ou filtrar os campos que forem duplicados ou desnecessários para operações de recebimento de dados.

- **Campo.** Exibe os campos de entrada a partir de origens de dados atualmente conectadas.
- **Filtrar.** Exibe o status do filtro de todos os campos de entrada. Os campos filtrados incluem um X vermelho nesta coluna, indicando que este campo não será transmitido posteriormente. Clique na coluna *Filtrar* para um campo selecionado para ativar ou desativar a filtragem. Também é possível selecionar as opções para diversos campos simultaneamente utilizando o método de seleção Shift-clique.

- **Campo.** Exibe os campos conforme eles saem do nó Filtro. Os nomes duplicados são exibidos em vermelho. É possível editar os nomes de campo ao clicar nesta coluna e inserir um novo nome. Também é possível remover os campos clicando na coluna *Filtrar* para desativar os campos duplicados.

Todas as colunas na tabela podem ser classificadas clicando no cabeçalho da coluna.

Visualizar campos atuais. Selecione para visualizar os campos para conjuntos de dados ativamente conectados ao nó Filtro. Esta opção é selecionada por padrão e é o método mais comum de usar os nós Filtro.

Visualizar as configurações de campos não utilizados. Selecione para visualizar os campos para conjuntos de dados que se conectaram uma vez, mas que não estão mais conectados ao nó Filtro. Esta opção é útil quando copiar nós Filtro de um fluxo para outro ou quando salvar e recarregar nós Filtro.

Menu do Botão Filtrar

Clique no botão Filtrar no canto superior esquerdo da caixa de diálogo para acessar um menu que fornece um número de atalhos e outras opções.

É possível escolher:

- Remover todos os campos.
- Incluir todos os campos.
- Alternar todos os campos.
- Remover duplicatas. *Nota:* selecionar essa opção remove todas as ocorrências do nome duplicado, incluindo a primeira.
- Renomear campos e conjuntos de múltiplas respostas para conformidade com outros aplicativos. Consulte o tópico “Renomeando ou Filtrando Campos para o IBM SPSS Statistics” na página 378 para obter mais informações.
- Truncar nomes de campo.
- Anonimar campo e nomes de conjunto de múltiplas respostas.
- Usar nomes de campo de entrada.
- Editar Conjuntos de Múltiplas Respostas. Consulte o tópico “Editando Conjuntos de Múltiplas Respostas” na página 147 para obter mais informações.
- Configure o estado de filtro padrão.

Também é possível utilizar os botões de comutação de seta na parte superior da caixa de diálogo para especificar se você deseja incluir ou descartar campos por padrão. Isso é útil para grandes conjuntos de dados em que apenas alguns campos devem ser incluídos posteriormente. Por exemplo, é possível selecionar apenas os campos que deseja manter e especificar que todos os outros devem ser descartados (ao invés de selecionar individualmente todos os campos para descarte).

Truncando Nomes de Campo

No menu do botão Filtrar (canto superior esquerdo da guia Filtro), é possível escolher truncar nomes de campo.

Comprimento máximo. Especifique um número de caracteres para limitar o comprimento de nomes de campo.

Número de dígitos. Se os nomes de campo, quando truncados, não foram mais exclusivos, eles serão truncados ainda mais e diferenciados ao incluir dígitos no nome. É possível especificar o número de dígitos usado. Utilize os botões de seta para ajustar o número.

Por exemplo, a tabela a seguir ilustra como os nomes de campo em um conjunto de dados médico são truncados utilizando as configurações padrão (maximum length=8 e number of digits=2).

Tabela 22. Truncamento de nome de campo

Nomes de campo	Nomes de Campo Truncados
Entrada de Paciente 1	Patien01
Entrada de Paciente 2	Patien02
Frequência Cardíaca	HeartRat
BP	BP

Anonimizando Nomes de Campo

É possível anonimizar nomes de campo a partir de qualquer nó que incluir uma guia Filtro ao clicar no menu do botão Filtro no canto superior esquerdo e escolher **Anonimar Nomes de Campo**. Os nomes de campos anonimizados consistem em um prefixo de sequência de caracteres mais um valor numérico exclusivo.

Anonimar valores de. Escolha **Campos selecionados somente** para anonimizar apenas os nomes de campos já selecionados na guia Filtro. O padrão é **Todos os campos**, que anonimiza todos os nomes de campo.

Prefixo de nomes de campo. O prefixo padrão para nomes de campo anonimizados é **anon_** ou escolha **Customizado** e digite seu próprio prefixo se desejar um diferente.

Anonimar conjuntos de múltiplas respostas. Anonimiza os nomes de conjuntos de múltiplas respostas da mesma maneira que os campos. Consulte o tópico “Editando Conjuntos de Múltiplas Respostas” para obter mais informações.

Para restaurar os nomes de campo originais, escolha **Usar Nomes de Campo de Entrada** a partir do menu do botão de filtro.

Editando Conjuntos de Múltiplas Respostas

É possível incluir ou editar conjuntos de múltiplas respostas a partir de qualquer nó que incluir uma guia Filtro ao clicar no canto superior esquerdo e escolher **Editar Conjuntos de Múltiplas Respostas**.

Os conjuntos de múltiplas respostas são utilizados para registrar os dados que possam ter mais de um valor para cada caso, por exemplo, quando perguntar aos respondentes de uma pesquisa de opinião quais museus eles visitaram e quais revistas eles leem. Os conjuntos de múltiplas respostas podem ser importados no IBM SPSS Modeler usando um nó de origem do Coleta de dados ou um nó de origem Arquivo de Estatísticas e podem ser definidos no IBM SPSS Modeler utilizando um nó Filtro.

Clique em **Novo** para criar um novo conjunto de múltiplas respostas ou clique em **Editar** para modificar um conjunto existente.

Nome e rótulo. Especifica o nome e uma descrição para o conjunto.

Tipo. Questões de múltiplas respostas podem ser manipuladas de duas maneiras:

- **Conjunto de múltiplas dicotomias.** Um campo de flag separado é criado para cada resposta possível, assim, se houver 10 revistas, haverá 10 campos de flag, cada um dos quais podendo ter valores, como 0 ou 1 para *true* ou *false*. O valor contado permite especificar qual valor será contado como verdadeiro. Este método é útil quando desejar permitir que os respondentes escolham todas as opções que se aplicam.
- **Conjunto de categorias múltiplas.** Um campo nominal é criado para cada resposta até o número máximo de respostas de um determinado respondente. Cada campo nominal possui valores que representam as respostas possíveis, como 1 para o *Time*, 2 para *Newsweek* e 3 para *PC Week*. Esse método é mais útil quando desejar limitar o número de respostas, por exemplo, ao pedir que os respondentes escolham as três revistas que eles leem com mais frequência.

Campos no conjunto. Utilize os ícones à direita para incluir ou remover campos.

Comentários

- Todos os campos incluídos em um conjunto de múltiplas respostas devem ter o mesmo armazenamento.
- Os conjuntos são distintos dos campos que eles incluem. Por exemplo, excluir um conjunto não exclui os campos que ele inclui, somente as ligações entre esses campos. O conjunto ainda está visível acima do ponto de exclusão, mas não é visível abaixo.
- Se os campos forem renomeados utilizando um nó Filtro (diretamente na guia ou escolhendo Renomear para as opções IBM SPSS Statistics, **Truncar** ou **Anonimar** no menu Filtro), quaisquer referências a esses campos utilizados em conjuntos de múltiplas respostas também serão atualizadas. No entanto, todos os campos em um conjunto de múltiplas respostas que forem descartados pelo nó Filtro não serão removidos do conjunto de múltiplas respostas. Esses campos, embora não estejam mais visíveis no fluxo, ainda são referenciados pelo conjunto de múltiplas respostas, o que pode ser uma consideração ao exportar, por exemplo.

Derivar Nó

Um dos recursos mais poderosos no IBM SPSS Modeler é a capacidade de modificar valores de dados e de derivar novos campos a partir de dados existentes. Durante os projetos de mineração de dados mais longos, é comum executar várias derivações, como extrair um ID de cliente a partir de uma sequência de caracteres dos dados do log da web ou criar um valor de fidelidade do cliente com base nos dados da transação e demográficos. Todas essas transformações podem ser executadas utilizando uma variedade de nós de operações de campo.

Vários nós fornecem a capacidade para derivar novos campos:



O nó Derivar modifica valores de dados ou cria novos campos a partir de um ou mais campos existentes. Ele cria campos do tipo fórmula, sinalizador, nominal, estado, contagem e condicional.



O nó Reclassificar transforma um conjunto de valores categóricos em outro. A reclassificação é útil para reduzir as categorias ou para reagrupar dados para análise.



O nó Categorização cria automaticamente novos campos nominais (conjunto) com base nos valores de um ou mais campos existente contínuos (intervalo numérico). Por exemplo, é possível transformar um campo de receita contínuo em um novo campo categórico contendo grupos de receitas como desvios da média. Depois de criar categorias para o novo campo, é possível gerar um nó Derivar com base nos pontos de corte.



O nó Configurar para Sinalizador deriva diversos campos de sinalização com base nos valores categóricos definidos para um ou mais campos nominais.



O nó Reestruturar converte um campo nominal ou sinalizador em um grupo de campos que podem ser preenchidos com os valores do outro campo. Por exemplo, dado um campo chamado *tipo de pagamento* com valores de *crédito*, *dinheiro* e *débito*, três novos campos seriam criados (*crédito*, *dinheiro* e *débito*), cada um dos quais podendo conter o valor do pagamento real feito.



O nó Histórico cria novos campos contendo dados de campos em registros anteriores. Os nós Históricos são mais frequentemente utilizados para dados sequenciais, como dados de séries temporais. Antes de utilizar um nó Histórico, você pode querer classificar os dados utilizando um nó Classificar.

Utilizando o Nó Derivar

Utilizando o nó Derivar, é possível criar seis tipos de novos campos a partir de um ou mais campos existentes:

- **Fórmula.** O novo campo é o resultado de uma expressão arbitrária do CLEM.
- **Flag.** O novo campo é um flag, representando uma condição especificada.
- **Nominal.** O novo campo é nominal, o que significa que seus membros são um grupo de valores especificados.
- **Estado.** O novo campo é um de dois estados. A alternância entre esses estados é acionada por uma condição especificada.
- **Contagem.** O novo campo baseia-se no número de vezes em que uma condição foi verdadeira.
- **Condicional.** O novo campo é o valor de uma das duas expressões, dependendo do valor de uma condição.

Cada um desses nós contém um conjunto de opções especiais na caixa de diálogo do nó Derivar. Essas opções são discutidas nos tópicos subsequentes.

Configurando Opções Básicas para o Nó Derivar

Na parte superior da caixa de diálogo dos nós Derivar há um número de opções para selecionar o tipo do nó Derivar que forem necessárias.

Modo. Selecione **Único** ou **Múltiplos**, dependendo se deseja derivar diversos campos. Quando **Múltiplos** é selecionada, a caixa de diálogo se altera para incluir opções para diversos campos Derivar.

Derivar campo. Para nós Derivar simples, especifique o nome do campo que deseja derivar e inclua em cada registro. O nome padrão é *DeriveN*, em que *N* é o número de nós Derivar que você criou até aqui durante a sessão atual.

Derivar como. Selecione um tipo de nó Derivar, como **Fórmula** ou **Nominal**, na lista suspensa. Para cada tipo, um novo campo é criado com base nas condições que forem especificadas na caixa de diálogo específica do tipo.

Selecionar uma opção na lista suspensa incluirá um novo conjunto de controles na caixa de diálogo principal de acordo com as propriedades de cada tipo de nó Derivar.

Tipo de campo. Selecione um nível de medição, como **contínuo**, **categorico** ou **flag**, para o nó recém-derivado. Essa opção é comum a todas as formas de nós Derivar.

Nota: derivar novos campos geralmente requer o uso de funções especiais ou expressões matemáticas. Para ajudá-lo a criar essas expressões, um Construtor de Expressões está disponível na caixa de diálogo de todos os tipos de nós Derivar e fornece verificação de regra e também uma lista completa de expressões do CLEM.

Derivando Diversos Campos

Configurar o modo para **Múltiplos** dentro de um nó Derivar fornece a capacidade de derivar diversos campos com base na mesma condição dentro do mesmo nó. Esse recurso economiza tempo quando quiser fazer transformações idênticas em vários campos em seu conjunto de dados. Por exemplo, se deseja construir um modelo de regressão que preveja o salário atual com base no salário inicial e

experiência anterior, poderá ser benéfico aplicar uma transformação logarítmica a todas as três variáveis defasadas. Ao invés de incluir um novo nó Derivar para cada transformação, é possível aplicar a mesma função a todos os campos de uma só vez. Simplesmente selecione todos os campos dos quais é possível derivar um novo campo e, em seguida, digite a expressão de derivação utilizando a função @FIELD dentro dos parênteses de campo.

Nota: a função @FIELD é uma ferramenta importante para derivação de diversos campos ao mesmo tempo. Ela permite consultar o conteúdo do campo ou dos campos atuais sem especificar o nome do campo exato. Por exemplo, uma expressão do CLEM utilizada para aplicar uma transformação logarítmica a diversos campos é $\log(@FIELD)$.

As opções a seguir estão incluídas na caixa de diálogo quando selecionar o modo **Múltiplo**:

Derivar de. Utilize o Seletor de Campo para selecionar campos a partir dos quais derivar os novos campos. Um campo de saída será gerado para cada campo selecionado. *Nota:* os campos selecionados não precisam ter o mesmo tipo de armazenamento, no entanto, a operação Derivar falhará se a condição não for válida para *todos* os campos.

Extensão do nome do campo. Digite a extensão que deseja incluir em um ou mais nomes de campos novos. Por exemplo, para um novo campo contendo o log de *Current Salary*, será possível incluir a extensão *log_* no nome do campo, produzindo *log_Current Salary*. Utilize os botões de opções para escolher se deseja incluir a extensão como um prefixo (no início) ou como um sufixo (no término) do nome do campo. O nome padrão é *DeriveN*, em que *N* é o número de nós Derivar que você criou até aqui durante a sessão atual.

Como no nó Derivar no modo único, agora é necessário criar uma expressão a ser utilizada para derivar um novo campo. Dependendo do tipo de operação Derivar selecionado, há várias opções para criar uma condição. Essas opções são discutidas nos tópicos subsequentes. Para criar uma expressão, é possível simplesmente digitar em um ou mais campos de fórmula ou utilizar o Construtor de Expressões clicando no botão da calculadora. Lembre-se de utilizar a função @FIELD ao referenciar manipulações em diversos campos.

Selecionando Diversos Campos

Para todos os nós que executam operações em diversos campos de entrada, como Derivar (modo múltiplo), Agregado, Ordem, Múltigráficos e Gráfico de Tempo, é possível selecionar facilmente diversos campos utilizando a caixa de diálogo Selecionar Campos.

Ordenar por. É possível ordenar campos disponíveis para visualização ao selecionar uma das opções a seguir:

- **Natural.** Visualize a ordem dos campos conforme eles foram transmitidos para o fluxo de dados no nó atual.
- **Nome.** Use a ordem alfabética para ordenar campos para visualização.
- **Tipo.** Visualize campos ordenados pelo seu nível de medição. Essa opção é útil quando selecionar campos com um nível de medição específico.

Selecione os campos da lista um por vez ou utilize os métodos Shift-clique e Ctrl-clique para selecionar vários campos. Também é possível utilizar os botões abaixo da lista para selecionar grupos de campos com base em seu nível de medição, ou para selecionar ou cancelar seleção de todos os campos na tabela.

Configurando Opções de Derivar Fórmula

Os nós Derivar Fórmula criam um novo campo para cada registro em um conjunto de dados com base nos resultados de uma expressão do CLEM. Esta expressão não pode ser condicional. Para derivar valores que são baseados em uma expressão condicional, utilize o tipo flag ou condicional do nó Derivar.

Fórmula Especifique uma fórmula ao usar a linguagem do CLEM para derivar um valor para o novo campo.

Nota: Como o SPSS Modeler não consegue saber qual nível de medição deve ser utilizado para um campo de lista derivada, para os níveis de medição Coleção e Geoespacial, é possível clicar em **Especificar...** para abrir a caixa de diálogo Valores e configurar o nível de medição necessário. Para obter informações adicionais, consulte “Configurando valores de lista derivados”.

Para campos Geoespaciais, os únicos tipos de armazenamento relevantes são **Real** e **Número Inteiro** (a configuração padrão é **Real**).

Configurando valores de lista derivados

A caixa de diálogo Valores é exibida quando você seleciona **Especificar...** a partir da lista suspensa **Tipo de campo** da Fórmula do nó Derivar. Nessa caixa de diálogo, configure os valores de nível de submedição a serem utilizados para os níveis de medição **Tipo de campo** de fórmula para Coleção ou Geoespacial.

Medição Selecione **Coleção** ou **Geoespacial**. Se selecionar qualquer outro nível de medição, a caixa de diálogo exibirá uma mensagem de que não há valores editáveis.

Coleção

O único item que pode ser configurado para o nível **Medição** da Coleção é **Medida de lista**. Por padrão, esta medida é configurada para Sem Tipo, mas é possível selecionar outro valor para configurar o nível de medição dos elementos dentro da lista. É possível escolher uma das seguintes opções:

- Sem tipo
- Categórico
- Contínuo
- Nominal
- Ordinal
- Sinalizador

Geo-espacial

Para o nível de **Medição** Geoespacial, é possível selecionar as seguintes opções para configurar o nível de medição dos elementos na lista:

Tipo Selecione o subnível de medição do campo geoespacial. Os subníveis disponíveis são determinados pela profundidade do campo de lista, e os padrões são os seguintes:

- Point (profundidade zero)
- LineString (profundidade de um)
- Polygon (profundidade de um)
- MultiPoint (profundidade de um)
- MultiLineString (profundidade de dois)
- MultiPolygon (profundidade de dois)

Para obter mais informações sobre subníveis, consulte o tópico Subníveis de Medição Geoespacial na seção Nó de Tipo do Guia de Nós Origem, Processo e Saída do SPSS Modeler.

Para obter mais informações sobre profundidades de lista, consulte o tópico Armazenamento de Lista e Níveis de Medição Associados na seção Nó de Origem do Guia Nós de Origem, Processo e Saída do SPSS Modeler.

Sistema de coordenadas Esta opção estará disponível somente se você alterou o nível de medição para geoespacial a partir de um nível não geoespacial. Para aplicar um sistema de coordenadas aos seus dados geoespaciais, marque essa caixa de seleção. Por padrão, o sistema de coordenadas configurado na área de janela **Ferramentas > Propriedades do Fluxo > Opções > Geoespacial** é mostrado. Para utilizar um sistema de coordenadas diferente, clique no botão **Alterar** para exibir a caixa de diálogo Selecionar Sistema de Coordenadas e escolha o sistema que corresponda aos seus dados.

Para obter mais informações sobre sistemas de coordenadas, consulte o tópico Configurando Opções Geoespaciais para Fluxos na seção Trabalhando com Fluxos do Guia do Usuário do SPSS Modeler.

Derivando uma lista ou campo geoespacial

Há ocasiões em que os dados que devem ser registrados como um item da lista são importados no SPSS Modeler com os atributos errados. Por exemplo, como campos geoespaciais separados, como uma coordenada x e uma coordenada y, ou uma longitude e latitude, como linhas individuais em um arquivo .csv. Nessa situação, deve-se combinar os campos individuais em um campo de lista única utilizando o nó Derivar.

Nota: Deve-se saber qual é o campo x (ou longitude) campo e qual é o campo y (ou campo de latitude) quando combinar dados geoespaciais. Deve-se combinar os seus dados para que o campo de lista resultante possua a ordem de elementos: [x, y] ou [Longitude, Latitude], que são os formatos padrão para coordenadas geoespaciais.

Os passos a seguir mostram um exemplo simples de derivar um campo de lista.

1. No seu fluxo, anexe um nó Derivar ao nó Origem.
2. Na guia Configurações do nó Derivar, selecione **Fórmula** a partir da lista **Derivar como**.
3. Em **Tipo de Campo**, selecione **Coleção** para uma lista geoespacial ou **Geoespacial**. Por padrão, o SPSS Modeler utiliza uma abordagem de "melhor palpite" para configurar o detalhe de lista correta; é possível selecionar **Especificar...** para abrir a caixa de diálogo Valores. Nesta caixa de diálogo, é possível inserir informações adicionais sobre os dados em sua lista; por exemplo, para uma lista geoespacial, é possível alterar o nível de medição.
4. Na área de janela **Fórmula**, insira a fórmula para combinar os dados no formato de lista correto. Como alternativa, clique no botão de calculadora para abrir o Construtor de Expressões.

Um exemplo simples de uma fórmula para derivar uma lista é [x, y], em que x e y são campos separados na origem de dados. O novo campo derivado que é criado é uma lista na qual o valor para cada registro é os valores x e y concatenados para esse registro.

Nota: Os campos que são combinados em uma lista dessa maneira devem ter o mesmo tipo de armazenamento.

Para obter mais informações sobre listas e profundidades de listas, consulte "Armazenamento de lista e níveis de medição associados" na página 11.

Configurando Opções de Derivar Flag

Os nós Derivar Flag são usados para indicar uma condição específica, como pressão arterial alta ou inatividade de conta do cliente. Um campo de flag é criado para cada registro e, quando a condição verdadeira é atendida, o valor de flag para verdadeiro é incluído no campo.

Valor True. Especifique um valor para incluir no campo de flag para registros que corresponderem à condição especificada abaixo. O padrão é T.

Valor False. Especifique um valor para incluir no campo de flag para registros que *não* corresponderem à condição especificada abaixo. O padrão é F.

Verdadeiro quando. Especifique uma condição do CLEM para avaliar determinados valores de cada registro e forneça ao registro um valor verdadeiro ou um valor falso (definido acima). Observe que o valor verdadeiro será fornecido aos registros no caso de valores numéricos não falsos.

Nota: para retornar uma sequência de caracteres vazia, deve-se digitar as aspas de abertura e de fechamento sem nada entre elas, como "". Sequências de caracteres vazias geralmente são utilizadas, por exemplo, como o valor false para permitir que os valores true sobressaiam mais claramente em uma tabela. Da mesma forma, as aspas devem ser utilizadas se desejar um valor de sequência de caracteres que, de outra forma, seria tratado como um número

Exemplo

Em liberações do IBM SPSS Modeler anteriores a 12.0, várias respostas foram importadas para um único campo, com valores separados por vírgulas. Por exemplo:

```
museum_of_design,institute_of_textiles_and_fashion  
museum_of_design  
archeological_museum  
$null$  
national_art_gallery,national_museum_of_science,other
```

Para preparar esses dados para análise, é possível utilizar a função `hassubstring` para gerar um campo de flag separado para cada resposta com uma expressão, como:

```
hassubstring(museums,"museum_of_design")
```

Configurando Opções de Derivar Nominal

Os nós Derivar Nominal são utilizados para executar um conjunto de condições do CLEM para determinar qual condição cada registro satisfaz. Conforme uma condição é atendida para cada registro, um valor (indicando qual conjunto de condições foi atendido) será incluído no novo campo derivado.

Valor padrão. Especifique um valor a ser utilizado no novo campo, se nenhuma das condições for atendida.

Configurar campo para. Especifique um valor a ser inserido no novo campo quando uma condição específica for atendida. Cada valor na lista tem uma condição associada que você especifica na coluna adjacente.

Se essa condição for verdadeira. Especifique uma condição para cada membro no campo de configuração para listar. Utilize o Construtor de Expressões para selecionar a partir de funções e campos disponíveis. É possível utilizar os botões de seta e de excluir para reordenar ou remover condições.

Uma condição funciona ao testar os valores de um campo específico no conjunto de dados. Conforme cada condição é testada, os valores especificados acima serão designados para o novo campo para indicar qual condição foi atendida, se houver. Se nenhuma das condições for atendida, o valor padrão será utilizado.

Configurando Opções de Derivar Estado

Os nós Derivar Estado são um tanto semelhantes aos nós Derivar Flag. Um nó Flag configura valores de acordo com o cumprimento de uma condição *única* para o registro atual, ao passo que um nó Derivar Estado pode alterar os valores de um campo dependendo de como ele cumpre *duas condições independentes*. Isso significa que o valor será alterado (ligado ou desligado) conforme cada condição for atendida.

Estado inicial. Selecione se deseja dar a cada registro do novo campo o valor **Ligado** ou **Desligado** inicialmente. Observe que esse valor pode ser alterado conforme cada condição for atendida.

Valor "On". Especifique o valor para o novo campo quando a condição Ligado for atendida.

Alternar "On" quando. Especifique uma condição do CLEM que alterará o estado para Ligado quando a condição for verdadeira. Clique no botão de calculadora para abrir o Construtor de Expressões.

Valor "Off". Especifique o valor para o novo campo quando a condição Desligado for atendida.

Alternar "Off" quando. Especifique uma condição do CLEM que alterará o estado para Desligado quando a condição for falsa. Clique no botão de calculadora para abrir o Construtor de Expressões.

Nota: para especificar uma sequência de caracteres vazia, deve-se digitar as aspas de abertura e de fechamento sem nada entre elas, como "". Da mesma forma, as aspas devem ser utilizadas se desejar um valor de sequência de caracteres que, de outra forma, seria tratado como um número.

Configurando Opções de Derivar Contagem

Um nó Derivar Contagem é utilizado para aplicar uma série de condições aos valores de um campo numérico no conjunto de dados. Conforme cada condição é atendida, o valor do campo de contagem derivado é aumentado em um incremento configurado. Este tipo de nó Derivar é útil para dados de séries temporais.

Valor inicial. Configura um valor usado na execução para o novo campo. O valor inicial deve ser uma constante numérica. Utilize os botões de seta para aumentar ou diminuir o valor.

Incrementar quando. Especifique a condição do CLEM que, quando atendida, alterará o valor derivado com base no número especificado em Incrementar em. Clique no botão de calculadora para abrir o Construtor de Expressões.

Incremento em. Configure o valor utilizado para incrementar a contagem. É possível utilizar uma constante numérica ou o resultado de uma expressão do CLEM.

Reconfigurar quando. Especifique uma condição que, quando atendida, reconfigura o valor derivado para o valor inicial. Clique no botão de calculadora para abrir o Construtor de Expressões.

Configurando Opções de Derivar Condicional

Os nós Derivar Condicional utilizam uma série de instruções If-Then-Else para derivar o valor do novo campo.

If. Especifique uma condição do CLEM que será avaliada para cada registro após a execução. Se a condição for verdadeira (ou não falsa, no caso de números), o novo campo receberá o valor especificado abaixo pela expressão Then. Clique no botão de calculadora para abrir o Construtor de Expressões.

Then. Especifique um valor ou uma expressão do CLEM para o novo campo quando a instrução If acima for verdadeira (ou não falsa). Clique no botão de calculadora para abrir o Construtor de Expressões.

Else. Especifique um valor ou uma expressão do CLEM para o novo campo quando a instrução If acima for falsa. Clique no botão de calculadora para abrir o Construtor de Expressões.

Recodificando Valores com o Nó Derivar

Os nós Derivar também podem ser utilizados para recodificar valores, por exemplo, ao converter um campo de sequências de caracteres com valores categóricos em um campo nominal numérico (conjunto).

1. Para Derivar Como, selecione o tipo de campo (Nominal, Flag, etc.), conforme apropriado.
2. Especifique as condições para recodificar valores. Por exemplo, é possível configurar o valor como 1 se Drug='drugA', 2 se Drug='drugB', e assim por diante.

Nó de Preenchimento

Os nós Preenchimento são utilizados para substituir valores do campo e alterar o armazenamento. É possível escolher substituir valores com base em uma condição especificada do CLEM, como @BLANK(FIELD). Como alternativa, é possível optar por substituir todos os espaços em branco ou valores nulos por um valor específico. Os nós Preenchimento são geralmente usados em conjunto com o nó Tipo para substituir valores omissos. Por exemplo, é possível preencher em brancos com o valor médio de um campo ao especificar uma expressão como @GLOBAL_MEAN. Esta expressão preencherá todos os brancos com o valor médio, conforme calculado por um nó Configurar Globais.

Preencher nos campos. Utilizando o Seletor de Campo (botão à direita do campo de texto), selecione os campos a partir do conjunto de dados cujos valores serão examinados e substituídos. O comportamento padrão é substituir valores dependendo das expressões Condição e Substituir por especificadas abaixo. Também é possível selecionar um método alternativo de substituição utilizando as opções Substituir abaixo.

Nota: ao selecionar diversos campos para substituir por um valor definido pelo usuário, é importante que os tipos de campo sejam semelhantes (todos numéricos ou todos simbólicos).

Substituir. Selecione para substituir os valores de um ou mais campos selecionados utilizando um dos seguintes métodos:

- **Baseado na condição.** Esta opção ativa o campo Condição e o Construtor de Expressões para criar uma expressão utilizada como uma condição para substituição pelo valor especificado.
- **Sempre.** Substitui todos os valores do campo selecionado. Por exemplo, é possível utilizar esta opção para converter o armazenamento de receita em uma sequência de caracteres utilizando a seguinte expressão do CLEM: (to_string(income)).
- **Valores em branco.** Substitui todos os valores em branco especificados pelo usuário no campo selecionado. A condição padrão @BLANK(@FIELD) é utilizada para selecionar brancos. *Nota:* é possível definir espaços em branco utilizando a guia Tipos do nó de origem ou com um nó Tipo.
- **Valores nulos.** Substitui todos os valores nulos do sistema no campo selecionado. A condição padrão @NULL(@FIELD) é utilizada para selecionar nulos.
- **Valores em branco e nulos.** Substitui valores em branco e nulos do sistema no campo selecionado. Essa opção é útil quando você não tiver certeza se nulos foram definidos como valores omissos ou não.

Condição . Essa opção estará disponível quando você tiver selecionado a opção **Baseado na condição**. Utilize esta caixa de texto para especificar uma expressão do CLEM para avaliar os campos selecionados. Clique no botão de calculadora para abrir o Construtor de Expressões.

Substituir por. Especifique uma expressão do CLEM para dar um novo valor para os campos selecionados. Também é possível substituir o valor por um valor nulo ao digitar undef na caixa de texto. Clique no botão de calculadora para abrir o Construtor de Expressões.

Nota: quando um ou mais campos selecionados são sequências de caracteres, deve-se substituí-los por um valor de sequência de caracteres. Utilizar o padrão 0 ou outro valor numérico como o valor de substituição para campos de sequência de caracteres resultará em um erro.

Conversão de Armazenamento Usando o Nó Preenchimento

Utilizando a condição Substituir de um nó Preenchimento, é possível converter facilmente o armazenamento de campo para um único ou diversos campos. Por exemplo, utilizando a função de conversão to_integer, é possível converter *income* de uma sequência de caracteres em um número inteiro utilizando a seguinte expressão do CLEM: to_integer(income).

É possível visualizar as funções de conversão disponíveis e criar automaticamente uma expressão do CLEM usando o Construtor de Expressões. Na lista suspensa de Funções, selecione **Conversão** para visualizar uma lista de funções de conversão de armazenamento. As seguintes funções de conversão estão disponíveis:

- `to_integer(ITEM)`
- `to_real(ITEM)`
- `to_number(ITEM)`
- `to_string(ITEM)`
- `to_time(ITEM)`
- `to_timestamp(ITEM)`
- `to_date(ITEM)`
- `to_datetime(ITEM)`

Convertendo valores de data e hora. Observe que as funções de conversão (e quaisquer outras funções que requerem um tipo específico de entrada, como um valor de data e hora) dependem dos formatos atuais especificados na caixa de diálogo de opções de fluxo. Por exemplo, se desejar converter um campo de sequência de caracteres com valores *Jan 2003, Feb 2003, etc.*, no armazenamento de data, selecione **MON YYYY** como o formato de data padrão para o fluxo.

As funções de conversão também estão disponíveis a partir do nó Derivar, para conversão temporária durante um cálculo de derivar. Também é possível utilizar o nó Derivar para executar outras manipulações, como recodificar campos de sequência de caracteres com valores categóricos. Consulte o tópico “Recodificando Valores com o Nó Derivar” na página 154 para obter mais informações.

Reclassificar Nó

O nó Reclassificar permite transformação de um conjunto de valores categóricos em outro. A reclassificação é útil para reduzir as categorias ou para reagrupar dados para análise. Por exemplo, é possível reclassificar os valores para *Produto* em três grupos, como *Utensílios, Cama, Mesa e Banho e Aparelhos Domésticos*. Geralmente, essa operação é executada diretamente a partir de um nó Distribuição ao agrupar valores e gerar um nó Reclassificar. Consulte o tópico “Usando um Nó de Distribuição” na página 246 para obter mais informações.

A reclassificação pode ser executada para um ou mais campos simbólicos. Também é possível escolher substituir os novos valores do campo existente ou gere um novo campo.

Quando utilizar um nó Reclassificar

Antes de utilizar um nó Reclassificar, considere se outro nó Operações de Campo é mais apropriado para a tarefa em mãos:

- Para transformar intervalos numéricos em conjuntos utilizando um método automático, como ranqueamentos ou percentis, deve-se utilizar um nó Categorização. Consulte o tópico “Nó de Categorização” na página 160 para obter mais informações.
- Para classificar os intervalos numéricos em conjuntos manualmente, deve-se utilizar um nó Derivar. Por exemplo, se desejar reduzir os valores de salário em categorias de faixa salarial específicas, deve-se utilizar um nó Derivar para definir cada categoria manualmente.
- Para criar um ou mais campos de flag com base nos valores de um campo categórico, como *Mortgage_type*, deve-se utilizar um nó Configurar para Flag.
- Para converter um campo categórico em armazenamento numérico, é possível utilizar um nó Derivar. Por exemplo, é possível converter os valores *Não* e *Sim* para 0 e 1, respectivamente. Consulte o tópico “Recodificando Valores com o Nó Derivar” na página 154 para obter mais informações.

Configurando Opções para o Nó Reclassificar

Há três passos para uso do nó Reclassificar:

1. Primeiro, selecione se deseja reclassificar diversos campos ou um único campo.
2. Em seguida, escolha se deseja recodificar no campo existente ou criar um novo campo.
3. Em seguida, utilize as opções dinâmicas na caixa de diálogo do nó Reclassificar para mapear conjuntos conforme desejado.

Modo. Selecione **Único** para reclassificar as categorias para um campo. Selecione **Diversos** para ativar as opções que permitem a transformação de mais de um campo por vez.

Reclassificar em. Selecione **Novo campo** para manter o campo nominal original e derivar um campo adicional contendo os valores reclassificados. Selecione **Campo existente** para sobrescrever os valores no campo original pelas novas classificações. Isso é essencialmente uma operação de "preenchimento".

Após especificar o modo e as opções de substituição, deve-se selecionar o campo de transformação e especificar os novos valores de classificação utilizando as opções dinâmicas na metade inferior da caixa de diálogo. Essas opções variam dependendo do modo selecionado acima.

Reclassificar campo(s). Utilize o botão Seletor de Campo à direita para selecionar um (modo Único) ou mais (modo Diversos) campos categóricos.

Novo nome de campo. Especifique um nome para o novo campo nominal contendo valores recodificados. Esta opção está disponível apenas no modo Único quando **Novo campo** é selecionado acima. Quando **Campo existente** é selecionado, o nome de campo original é retido. Ao trabalhar no modo Múltiplo, esta opção é substituída pelos controles para especificar uma extensão incluída em cada novo campo. Consulte o tópico "Reclassificando Diversos Campos" na página 158 para obter mais informações.

Reclassificar valores. Essa tabela permite um mapeamento claro dos valores de conjunto antigos para os valores que forem especificados aqui.

- **Valor original.** Essa coluna lista os valores existentes em um ou mais campos de seleção.
 - **Novo valor.** Use essa coluna para digitar novos valores de categoria ou selecionar um na lista suspensa. Ao gerar automaticamente um nó Reclassificar utilizando os valores de um gráfico Distribuição, esses valores são incluídos na lista suspensa. Isso permite mapear rapidamente valores existentes para um conjunto de valores conhecido. Por exemplo, as organizações de assistência médica às vezes agrupam diagnósticos de forma diferente com base na rede ou no código do idioma. Após uma fusão ou aquisição, todas as partes precisarão reclassificar dados novos ou mesmo existentes em um modo consistente. Ao invés de digitar manualmente cada valor de destino a partir de uma lista longa, é possível ler a lista principal de valores no IBM SPSS Modeler, executar um gráfico de Distribuição para o campo *Diagnóstico* e gerar um nó Reclassificar (valores) para esse campo diretamente do gráfico. Esse processo tornará todos os valores de Diagnóstico de destino disponíveis a partir da lista suspensa de Novos Valores.
4. Clique em **Obter** para ler os valores originais para um ou mais campos selecionados acima.
 5. Clique em **Copiar** para colar valores originais na coluna *Novo valor* para os campos que ainda não foram mapeados. Os valores originais não mapeados são incluídos na lista suspensa.
 6. Clique em **Limpar novo** para apagar todas as especificações na coluna *Novo valor*. *Nota:* esta opção não apaga os valores da lista suspensa.
 7. Clique em **Automático** para gerar automaticamente números inteiros consecutivos para cada um dos valores originais. Somente valores de número inteiro (sem valores reais, como 1,5, 2,5, e assim por diante) podem ser gerados.

Por exemplo, é possível gerar automaticamente números de ID do produto consecutivos para nomes de produtos ou números de cursos para ofertas de classes universitárias. Esta funcionalidade corresponde à transformação Recodificação Automática para os conjuntos no IBM SPSS Statistics.

Para uso de valores não especificados. Essa opção é utilizada para preencher valores não especificados no novo campo. É possível escolher manter o valor original ao selecionar **Valor original** ou especificar um valor padrão.

Reclassificando Diversos Campos

Para mapear valores de categoria para mais de um campo por vez, configure o modo para **Múltiplos**. Isso permite novas configurações na caixa de diálogo Reclassificar, que são descritas abaixo.

Reclassificar campos. Utilize o botão Seletor de Campo à direita para selecionar os campos que deseja transformar. Utilizando o Seletor de Campo, é possível selecionar todos os campos de uma vez ou campos de um tipo semelhante, como nominal ou de flag.

Extensão do nome do campo. Quando recodificar diversos campos simultaneamente, é mais eficiente especificar uma extensão comum incluída em todos os novos campos ao invés de nomes de campo individuais. Especifique uma extensão como `_recode` e selecione se deseja anexar ou pré-anexar essa extensão aos nomes de campo originais.

Armazenamento e nível de medição para campos reclassificados

O nó Reclassificar sempre cria um campo nominal a partir da operação de recodificação. Em alguns casos, isso pode alterar o nível de medição do campo quando usar o modo de reclassificação **Campo existente**.

O armazenamento do novo campo (como os dados são *armazenados* ao invés de como eles são *utilizados*) é calculado com base nas opções da guia Configurações a seguir:

- Se valores não especificados forem configurados para usar um valor padrão, o tipo de armazenamento será determinado ao examinar os novos valores e também o valor padrão e ao determinar o armazenamento apropriado. Por exemplo, se todos os valores puderem ser analisados como números inteiros, o campo terá o tipo de armazenamento de número inteiro.
- Se valores não especificados forem configurados para utilizar os valores originais, o tipo de armazenamento será baseado no armazenamento do campo original. Se todos os valores puderem ser analisados como o armazenamento do campo original, então esse armazenamento será preservado, caso contrário, o armazenamento será determinado ao localizar o tipo de armazenamento mais apropriado que inclua valores antigos e novos. Por exemplo, reclassificar um conjunto de números inteiros { 1, 2, 3, 4, 5 } com a reclassificação 4 => 0, 5 => 0 gera um novo conjunto de números inteiros { 1, 2, 3, 0 }, ao passo que a reclassificação 4 => "Over 3", 5 => "Over 3" gera o conjunto de sequências de caracteres { "1", "2", "3", "Over 3" }.

Nota: se o tipo original era não instanciado, o novo tipo também será não instanciado.

Nó Anonimizado

O nó Anonimizado permite disfarçar nomes de campos e/ou valores de campo ao trabalhar com dados que devem ser incluídos em um modelo abaixo do nó. Dessa forma, o modelo gerado pode ser livremente distribuído (por exemplo, para Suporte Técnico) sem nenhum risco de que usuários não autorizados possam visualizar dados confidenciais, como registros de funcionários ou registros médicos de pacientes.

Dependendo de onde o nó Anonimizado for colocado no fluxo, poderá ser necessário fazer mudanças nos outros nós. Por exemplo, se inserir um nó Anonimizado acima de um nó Seleção, os critérios de seleção no nó Seleção precisarão ser alterados se eles estiverem atuando em valores que agora se tornam anonimizados.

O método a ser utilizado para anonimizar depende de vários fatores. Para nomes de campo e todos os valores dos campos, exceto os níveis de medição Contínuo, os dados são substituídos por uma sequência de caracteres no formato:

prefix_Sn

em que *prefix_* é uma sequência de caracteres especificada pelo usuário ou a sequência de caracteres padrão *anon_* e *n* é um valor de número inteiro que inicia em 0 e é incrementado para cada valor exclusivo (por exemplo, *anon_S0*, *anon_S1*, e assim por diante).

Valores do campo de tipo Contínuo devem ser transformados porque os intervalos numéricos lidam com valores de número inteiro ou reais ao invés de sequências de caracteres. Assim, eles podem ser anonimizados somente ao transformar o intervalo em um intervalo diferente, disfarçando, portanto, os dados originais. A transformação de um valor *x* no intervalo é executada da seguinte maneira:

$$A*(x + B)$$

em que:

A é um fator de escala, que deve ser maior que 0.

B é um offset de conversão a ser incluído nos valores.

Exemplo

No caso de um campo *AGE* no qual o fator de escala *A* é configurado como 7 e o offset de conversão *B* é configurado como 3, os valores para *AGE* são transformados em:

$$7*(AGE + 3)$$

Configurando opções para o nó anonimizado

Aqui é possível escolher quais campos devem ter seus valores mascarados adicionalmente ao recebimento de dados.

Observe que os campos de dados devem ser instanciados no envio de dados a partir do nó anonimizado antes que as operações anonimizadas possa ser executadas. É possível instanciar os dados clicando no botão **Ler Valores** em um nó Tipo ou na guia Tipos de um nó de origem.

Campo. Lista os campos no conjunto de dados atual. Se qualquer nome de campo já tiver sido anonimizado, os nomes anonimizados serão mostrados aqui.

Medição. O nível de medição do campo.

Anonimizar Valores. Selecione um ou mais campos, clique nessa coluna e escolha **Sim** para anonimizar o valor do campo usando o prefixo padrão **anon_**; escolha **Especifique** para exibir uma caixa de diálogo na qual é possível digitar seu próprio prefixo ou, no caso de valores de campo do tipo *Continuous*, especificar se a transformação de valores de campo deve usar valores aleatórios ou especificados pelo usuário. Observe que os tipos de arquivos *Continuous* e tipos não *Continuous* não podem ser especificados na mesma operação; deve-se fazer isso separadamente para cada tipo de campo.

Visualizar campos atuais. Selecione para visualizar arquivos para conjuntos de dados ativamente conectados ao nó anonimizado. Essa opção é selecionada por padrão.

Visualizar as configurações de campos não utilizados. Selecione para visualizar campos para conjuntos de dados que já foram conectados ao nó, mas que não estão mais conectados. Esta opção é útil ao copiar nós de um fluxo para outro ou ao salvar e recarregar nós.

Especificando Como Valores de Campo Serão Anonimizados

A caixa de diálogo Substituir Valores permite escolher se deve utilizar o prefixo padrão para valores de campo anonimizados ou utilizar um prefixo customizado. Clicar em **OK** nesta caixa de diálogo altera a configuração de Valores Anônimos na guia Configurações para **Sim** para o campo ou campos selecionados.

Prefixo de valores de campo. O prefixo padrão para valores de campo anonimizados é **anon_**; escolha **Customizado** e insira seu próprio prefixo se desejar um diferente.

A caixa de diálogo Transformar Valores é exibida apenas para campos do tipo Contínuo e permite especificar se a transformação de valores de campo é utilizar valores de campo aleatórios ou valores especificados pelo usuário.

Aleatório. Escolha esta opção para utilizar valores aleatórios para a transformação. **Configurar semente aleatória** é selecionada por padrão; especifique um valor no campo **Semente** ou utilize o valor padrão.

Fixo. Escolha esta opção para especificar seus próprios valores para a transformação.

- **Escalar por.** O número pelo qual valores de campo serão multiplicados na transformação. O valor mínimo é 1 e o máximo é normalmente 10, mas isso pode ser reduzido para evitar estouro.
- **Converter por.** O número que será incluído nos valores do campo na transformação. O valor mínimo é 0 e o máximo é normalmente 1000, mas isso pode ser reduzido para evitar estouro.

Anonimizando Valores de Campo

Os campos que foram selecionados para anonimização na guia Configurações têm seus valores anonimizados:

- Quando executar o fluxo que contém o nó Anonimizar
- Ao visualizar os valores

Para visualizar os valores, clique no botão **Anonimizar Valores** na guia Valores Anonimizados. Em seguida, selecione um nome de campo da lista suspensa.

Se o nível de medição for Contínuo, a exibição mostrará:

- Os valores mínimo e máximo do intervalo original
- A equação utilizada para transformar os valores

Se o nível de medição for algo diferente de Contínuo, a tela exibirá os valores originais e anonimizados para esse campo.

Se a exibição aparecer com um plano de fundo amarelo, isso indica que a configuração para o campo selecionado foi alterada desde a última vez em que os valores foram anonimizados ou que as mudanças foram feitas nos dados acima do nó Anonimizar, de modo que os valores anonimizados podem não estar mais corretos. O conjunto atual de valores é exibido; clique no botão **Anonimizar Valores** novamente para gerar um novo conjunto de valores de acordo com a configuração atual.

Anonimizar Valores. Cria valores anonimizados para o campo selecionado e os exibe na tabela. Se estiver utilizando a semente aleatória para um campo do tipo Contínuo, clicar neste botão repetidamente criará um conjunto diferente de valores todas as vezes.

Limpar Valores. Limpa os valores originais e anonimizados da tabela.

Nó de Categorização

O nó Categorização permite criar automaticamente novos campos nominais com base nos valores de um ou mais campos existentes contínuos (intervalo numérico). Por exemplo, é possível transformar um campo de receita contínuo em um novo campo categórico contendo grupos de receitas de largura igual ou como desvios da média. Como alternativa, é possível selecionar um campo "supervisor" categórico para preservar a intensidade da associação original entre os dois campos.

A categorização pode ser útil por diversos motivos, incluindo:

- **Requisitos de algoritmo.** Determinados algoritmos, como Naive Bayes e Regressão Logística, requerem entradas categóricas.
- **Desempenho.** Algoritmos como logística multinomial poderão executar melhor se o número de valores distintos de campos de entrada for reduzido. Por exemplo, use o valor de mediana ou médio para cada categoria ao invés de usar os valores originais.
- **Privacidade de Dados.** Informações pessoais sensíveis, como salários, podem ser relatadas em intervalos ao invés de em figuras de salário reais a fim de proteger a privacidade.

Um número de métodos de categorização está disponível. Depois de criar categorias para o novo campo, é possível gerar um nó Derivar com base nos pontos de corte.

Quando utilizar um nó Categorização

Antes de utilizar um nó Categorização, considere se outra técnica é mais apropriada para a tarefa em mãos:

- Para especificar manualmente os pontos de corte para categorias, como intervalos de salário predefinidos específicos, utilize um nó Derivar. Consulte o tópico “Derivar Nó” na página 148 para obter mais informações.
- Para criar novas categorias para os conjuntos existentes, utilize um nó Reclassificar. Consulte o tópico “Reclassificar Nó” na página 156 para obter mais informações.

Manipulação de valor omissos

O nó Categorização manipula os valores omissos das seguintes maneiras:

- **Em brancos especificados pelo usuário.** Valores omissos especificados como em branco são incluídos durante a transformação. Por exemplo, se você designou -99 para indicar um valor em branco utilizando o nó Tipo, esse valor será incluído no processo de categorização. Para ignorar em brancos durante a categorização, deve-se utilizar um nó Preenchimento para substituir os valores em branco pelo valor nulo do sistema.
- **Valores omissos do sistema (\$null\$).** Os valores nulos são ignorados durante a transformação de categorização e permanecem nulos após a transformação.

A guia Configurações fornece opções para técnicas disponíveis. A guia Visualização exibe pontos de corte estabelecidos para dados anteriormente executados através do nó.

Configurando Opções para o Nó Categorização

Utilizando o nó Categorização, é possível gerar automaticamente categorias utilizando as seguintes técnicas:

- Categorização de largura fixa
- Ladrilhos (contagem ou soma igual)
- Médio e o desvio padrão
- Ranqueamentos
- Otimizado relativo a um campo categórico "supervisor"

A parte inferior da caixa de diálogo altera dinamicamente dependendo do método de categorização que selecionar.

Campos de categoria. Transformações pendentes de campos contínuos (intervalo numérico) são exibidas aqui. O nó Categorização permite categorizar diversos campos simultaneamente. Incluir ou remover os campos utilizando os botões à direita.

Método de categorização. Selecione o método utilizado para determinar os pontos de corte para as novas categorias de campo. Os tópicos subseqüentes descrevem as opções disponíveis em cada caso.

Limites da categoria. Especifique como os limites de categoria devem ser criados.

- **Sempre recalcular.** Os pontos de corte e alocações de categoria são sempre recalculados quando o nó for executado.
- **Ler a partir da guia Valores de Categoria, se disponível.** Os pontos de corte e alocações de categoria são calculados somente conforme necessário (por exemplo, quando novos dados são incluídos).

Os tópicos a seguir discutem as opções para os métodos disponíveis de categorização.

Categorias de Largura Fixa

Ao escolher **Largura fixa** como o método de categorização, um novo conjunto de opções é exibido na caixa de diálogo.

Nome da extensão. Especifique uma extensão a ser utilizada para um ou mais campos gerados. *_BIN* é a extensão padrão. Também é possível especificar se a extensão é incluída no início (**Prefixo**) ou no término (**Sufixo**) do nome do campo. Por exemplo, é possível gerar um novo campo chamado *income_BIN*.

Largura da categoria. Especifique um valor (número inteiro ou real) usado para calcular a "largura" da categoria. Por exemplo, é possível utilizar o valor padrão 10 para categorizar o campo *Age*. Como *Age* possui um intervalo de 18 a 65, as categorias geradas seriam mostradas na tabela a seguir.

Tabela 23. Categorias para Idade com intervalo de 18 a 65

Categoria 1	Categoria 2	Categoria 3	Categoria 4	Categoria 5	Categoria 6
>=13 a <23	>=23 a <33	>=33 a <43	>=43 a <53	>=53 a <63	>=63 a <73

O início dos intervalos de categoria é calculado utilizando o menor valor varrido menos a metade da largura da categoria (conforme especificado). Por exemplo, nas categorias mostradas acima, 13 é utilizado para iniciar os intervalos de acordo com o seguinte cálculo: $18 [\text{menor valor de dados}] - 5 [0,5 \times (\text{largura da categoria de } 10)] = 13$.

Nº de categorias. Use essa opção para especificar um número inteiro utilizado para determinar o número de categorias de largura fixa para o(s) novo(s) campo(s).

Após ter executado o nó Categorização em um fluxo, será possível visualizar os limites de categoria gerados clicando no guia **Visualização** na caixa de diálogo do nó Categorização. Consulte o tópico "Visualizando as Categorias Geradas" na página 166 para obter mais informações.

Ladrilhos (contagem igual ou soma)

O método de criação de bin lado a lado cria campos nominais que podem ser usados para dividir registros digitalizados em grupos de percentil (ou percentis, decis, etc), de forma que cada grupo contém o mesmo número de registros, ou a soma dos valores em cada grupo é igual. Os registros são classificados em ordem crescente, com base no valor do campo de bin especificado, de modo que aos registros com os valores mais baixos para a variável bin selecionada é atribuída uma classificação de 1, o próximo conjunto de registros são classificados em 2, e assim por diante. Os valores do limite para cada bin são gerados automaticamente com base nos dados e métodos de criação de ladrilhos usados.

Extensão do nome do ladrilho. Especifique uma extensão usada para o(s) campo(s) gerado(s) usando ladrilhos-p padrão. A extensão padrão é *_TILE* e *N*, em que *N* é o número do ladrilho. Também é possível especificar se a extensão é incluída no início (**Prefixo**) ou no término (**Sufixo**) do nome do campo. Por exemplo, seria possível gerar um novo campo, chamado *income_BIN4*.

Extensão de ladrilho customizada. Especifique uma extensão usada para um intervalo de ladrilho customizado. O padrão é *_TILEN*. Observe que *N* neste caso *não* será substituído pelo número customizado.

Os ladrilhos-p disponíveis são:

- **Quartil.** Gere 4 bins, cada um contendo 25% dos casos.
- **Quintil.** Gere 5 bins, cada um contendo 20% dos casos.
- **Decil.** Gere 10 bins, cada um contendo 10% dos casos.
- **Vingtile.** Gere 20 bins, cada um contendo 5% dos casos.
- **Percentil.** Gere 100 bins, cada um contendo 1% dos casos.
- **N customizado.** Selecione para especificar o número de bins. Por exemplo, um valor de 3 poderia produzir 3 categorias em faixas (2 pontos de corte), cada uma contendo 33,3% dos casos.

Observe que se houver menos valores discretos nos dados que o número de ladrilhos especificado, todos os ladrilhos não serão usados. Em tais casos, é provável que a nova distribuição reflita a distribuição original dos seus dados.

Método de criação de ladrilhos. Especifica o método usado para designar registros para bins.

- **Contagem de registro.** Busca designar um número igual de registros para cada bin.
- **Soma de valores.** Busca designar registros para bins de modo que a soma dos valores em cada bin seja igual. Ao direcionar esforços de vendas, por exemplo, esse método pode ser usado para designar possíveis clientes para grupos decis com base no valor por registro, com o valor de possíveis clientes mais alto no bin superior. Por exemplo, uma empresa farmacêutica pode classificar os médicos em grupos decis, com base no número de prescrições que eles escrevem. Enquanto cada decil deveria conter aproximadamente o mesmo número de scripts, o número de indivíduos que contribuem com esses scripts não seria o mesmo, com os indivíduos que escrevem a maioria dos scripts concentrados em decil 10. Observe que essa abordagem supõe que todos os valores sejam maiores que zero, e pode produzir resultados inesperados se este não for o caso.

Empates. Uma condição de empate resulta quando os valores em cada lado de um ponto de corte são idênticos. Por exemplo, se você estiver designando decis e mais de 10% de registros tiverem o mesmo valor para o campo de bin, todos eles não caberão no mesmo bin sem forçar o limite de uma maneira ou de outra. Os empates podem ser movidos para cima, para o próximo bin, ou mantidos no atual, mas devem ser resolvidos para que todos os registros com valores idênticos caiam no mesmo local, mesmo se isto faça com que alguns bins tenham mais registros do que o esperado. Os limites de bins subsequentes também podem ser ajustados como resultado, fazendo com que os valores sejam designados de forma diferente para o mesmo conjunto de números com base no método usado para resolver empates.

- **Incluir no próximo.** Selecione para mover os valores de empate para a próxima categoria.
- **Manter no atual.** Mantém os valores de empate na categoria atual (inferior). Este método pode resultar em um total menor de categorias sendo criado.
- **Designar aleatoriamente.** Selecione para alocar os valores de empate aleatoriamente em um local. Isso tenta manter o número de registros em cada bin em uma quantidade igual.

Exemplo: Criação de ladrilho por contagem de registro

A tabela a seguir ilustra como os valores de campo simplificados são classificados no momento da criação de ladrilhos por contagem de registro. Observe que os resultados variam, dependendo da opção de empate selecionada.

Tabela 24. Exemplo de criação de ladrilho por contagem de registro.

Valores	Incluir para Próxima	Manter na Atual
10	1	1
13	2	1
15	3	2
15	3	2

Tabela 24. Exemplo de criação de ladrilho por contagem de registro (continuação).

Valores	Incluir para Próxima	Manter na Atual
20	4	3

O número de itens por bin é calculado como:

número total de valor / número de ladrilhos

No exemplo simplificado acima, o número desejado de itens por bin é 1,25 (5 valores / 4 quartis). O valor 13 (sendo número de valor 2) estende o limite de contagem desejado de 1,25 e, portanto, é tratado de forma diferente, dependendo da opção de empates selecionada. No modo **Incluir para o Próximo**, ele é incluído no bin 2. No modo **Manter no Atual**, ele é deixado no bin 1, empurrando para fora o intervalo de valores para o bin 4 que dos valores de dados existentes. Como resultado, apenas três bins são criados e os limites para cada bin são ajustados de acordo, conforme mostrado na tabela a seguir.

Tabela 25. Resultado do exemplo de criação de bin.

Categoria	Inferior	Superior
1	>=10	<15
2	>=15	<20
3	>=20	<=20

Nota: A velocidade da criação de bin por ladrilho pode se beneficiar da ativação do processamento paralelo.

Ranquear Casos

Ao escolher **Ranqueamento** como o método de categorização, um novo conjunto de opções é exibido na caixa de diálogo.

O ranqueamento cria novos campos contendo classificações, ranqueamentos fracionais e valores de percentil para campos numéricos, dependendo das opções especificadas abaixo.

Ordem de ranqueamento. Selecione **Crescente** (o valor mais baixo é marcado como 1) ou **Decrescente** (o valor mais alto é marcado como 1).

Posição. Selecione para classificar os casos em ordem crescente ou decrescente, conforme especificado acima. O intervalo de valores no novo campo será 1- N , em que N é o número de valores discretos no campo original. Valores empatados recebem a média de seus ranqueamentos.

Ranqueamento fracional. Selecione para classificar os casos, em que o valor do novo campo é igual à classificação dividida pela soma das ponderações dos casos não omissos. As classificações fracionárias entram no intervalo de 0 a 1.

Porcentagem de ranqueamento fracional. Cada classificação é dividida pelo número de registros com valores válidos e multiplicada por 100. A porcentagem das classificações fracionárias entra no intervalo de 1 a 100.

Extensão. Para todas as opções de ranqueamento, é possível criar extensões customizadas e especificar se a extensão é incluída no início (**Prefixo**) ou no término (**Sufixo**) do nome do campo. Por exemplo, é possível gerar um novo campo chamado *income_P_RANK*.

Desvio Médio/Padrão

Ao escolher **Desvio médio/padrão** como o método de categorização, um novo conjunto de opções é exibido na caixa de diálogo.

Este método gera um ou mais novos campos com categorias em banda com base nos valores de desvio médio e padrão da distribuição de um ou mais campos especificados. Selecione o número de desvios a serem utilizados abaixo.

Nome da extensão. Especifique uma extensão a ser utilizada para um ou mais campos gerados. *_SDBIN* é a extensão padrão. Também é possível especificar se a extensão é incluída no início (**Prefixo**) ou no término (**Sufixo**) do nome do campo. Por exemplo, é possível gerar um novo campo chamado *income_SDBIN*.

- **+/- 1 desvio padrão.** Selecione para gerar três categorias.
- **+/- 2 desvios padrão.** Selecione para gerar cinco categorias.
- **+/- 3 desvios padrão.** Selecione para gerar sete categorias.

Por exemplo, selecionar +/-1 desvio padrão resulta em três categorias conforme calculado e mostrado na tabela a seguir.

Tabela 26. Exemplo de categoria de desvio padrão.

Categoria 1	Categoria 2	Categoria 3
$x < (\text{Mean} - \text{Std. Dev})$	$(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$	$x > (\text{Mean} + \text{Std. Dev})$

Em uma distribuição normal, 68% dos casos caem dentro de um desvio padrão da média, 95% dentro de dois desvios padrão e 99% dentro de três desvios padrão. Observe, no entanto, que a criação de categorias em banda com base em desvios padrão pode resultar em algumas categorias sendo definidas fora do intervalo de dados reais e até mesmo fora do intervalo de valores de dados possíveis (por exemplo, um intervalo de salário negativo).

Categorização ideal

Se o campo que deseja categorizar estiver fortemente associado a outro campo categórico, será possível selecionar o campo categórico como um campo "supervisor" para criar as categorias de maneira a preservar a intensidade da associação original entre os dois campos.

Por exemplo, suponha que você tenha utilizado a análise do cluster para agrupar estados com base em taxas de delinquência para empréstimos de habitação, com as taxas mais altas no primeiro cluster. Neste caso, você pode escolher *Porcentagem de vencimento* e *Porcentagem de execução hipotecária* como os campos Categoria e o campo de associação de cluster gerados pelo modelo como o campo supervisor.

Nome de extensão Especifique uma extensão a ser utilizada para um ou mais campos gerados e se ele deve ser incluído no início (**Prefixo**) ou no término (**Sufixo**) do nome do campo. Por exemplo, é possível gerar um novo campo chamado *pastdue_OPTIMAL* e outro chamado *infoclosure_OPTIMAL*.

Campo supervisor Um campo categórico utilizado para construir as categorias.

Pré-categorizar campos para melhorar o desempenho com conjuntos de dados grandes Indica se pré-processamento deve ser utilizado para aperfeiçoar a categorização ideal. Isso agrupa valores de escala em um número grande de categorias utilizando um método simples de categorização não supervisionada, representa valores dentro de cada categorização pela média e ajusta a ponderação de caso adequadamente antes de continuar com a categorização supervisionada. Em termos práticos, este método troca um grau de precisão para velocidade e é recomendado para grandes conjuntos de dados. Também é possível especificar o número máximo de categorias em que qualquer variável deve terminar após o pré-processamento quando essa opção for utilizada.

Mesclar categorias que possuem contagens de caso relativamente pequenas com um vizinho maior. Se ativada, indica que uma categoria será mesclada se a razão de seu tamanho (número de casos) com o de uma categoria vizinha é menor que o limite especificado; observe que os limites maiores podem resultar em mais mesclagens.

Configurações de Ponto de Corte

A caixa de diálogo Configurações de Ponto de Corte permite especificar opções avançadas para o algoritmo de categorização ideal. Essas opções informam ao algoritmo como calcular as categorizações usando o campo de destino.

Terminais de categoria. É possível especificar se os terminais inferiores ou superiores devem ser inclusivos (inferior $\leq x$) ou exclusivos (inferior $< x$).

Primeira e última categorias. Para a primeira e última categorias, é possível especificar se essas categorias devem ser ilimitadas (estendendo para o infinito positivo ou negativo) ou limitadas pelos pontos de dados mais baixos ou mais altos.

Visualizando as Categorias Geradas

A guia Valores de Categorização no nó Categorização permite visualizar os limites para as categorias geradas. Utilizando o menu Gerar, também é possível gerar um nó Derivar que possa ser utilizado para aplicar esses limites de um conjunto de dados para outro.

Campo Categorizado. Utilize a lista suspensa para selecionar um campo para visualização. Os nomes de campo mostrados utilizam o nome de campo original para maior clareza.

Ladrilho. Utilize a lista suspensa para selecionar um ladrilho, como 10 ou 100, para visualização. Essa opção está disponível somente quando categorias tiverem sido geradas utilizando o método de ladrilho (contagem ou soma iguais).

Limites da categoria. Os valores limite são mostrados aqui para cada categoria gerada, junto do número de registros que caem em cada categoria. Para o método de categorização ideal somente, o número de registros em cada categoria é mostrado como uma porcentagem do total. Observe que os limites não são aplicáveis quando o método de categorização de ranqueamento é utilizado.

Ler Valores. Lê valores categorizados a partir do conjunto de dados. Observe que os limites também são sobrescritos quando novos dados são executados por meio do fluxo.

Gerando um Nó Derivar

É possível utilizar o menu Gerar para criar um nó Derivar com base nos limites atuais. Isso é útil para aplicar limites de categoria estabelecidos de um conjunto de dados para outro. Além disso, quando esses pontos de divisão são conhecidos, uma operação Derivar é mais eficiente (significado mais rápida) do que uma operação de Categorização ao trabalhar com grandes conjuntos de dados.

Nó de Análise de RFM

O nó Recency, Frequency, Monetary (RFM) Analysis permite determinar de modo quantitativo quais podem ser seus melhores clientes ao examinar quando foi a última vez que eles compraram de você (recência), com que frequência eles compraram (frequência) e quanto eles gastaram em todas as transações (monetário).

O raciocínio por trás da análise de RFM é que os clientes que compram um produto ou um serviço uma vez têm maior probabilidade de comprar novamente. Os dados do cliente categorizados são separados em um número de categorias, com os critérios de categorização ajustados conforme necessário. Em cada uma das categorias, os clientes são designados a um escore; esses escores são, então, combinados para

fornecer um escore de RFM geral. Esse escore é uma representação da associação do cliente nas categorias criadas para cada um dos parâmetros de RFM. Esses dados customizados podem ser suficientes para suas necessidades, por exemplo, ao identificar os clientes mais frequentes e clientes de alto valor; como alternativa, eles podem ser transmitidos em um fluxo para modelagem e análise adicionais.

Observe, no entanto, que embora a capacidade de analisar e classificar os escores de RFM seja uma ferramenta útil, deve-se estar ciente de determinados fatores ao utilizá-la. Pode haver uma tentação de destinar clientes com os ranqueamentos mais altos, no entanto, solicitações em excesso desses clientes podem levar a ressentimento e uma queda real nos negócios de repetição. Também vale lembrar que os clientes com escores baixos não devem ser negligenciados, mas, ao invés disso, podem ser cultivados para tornarem-se melhores clientes. Por outro lado, escores altos em si não refletem necessariamente uma boa perspectiva de vendas, dependendo do mercado. Por exemplo, um cliente na categoria 5 para Recência, significando que ele comprou muito recentemente, pode, na realidade, não ser o melhor cliente-alvo para alguém que vende produtos de vida mais longa e caros, como carros ou televisores.

Nota: dependendo de como os dados são armazenados, poderá ser necessário preceder o nó de Análise de RFM com um nó Agregado de RFM para transformar os dados em um formato utilizável. Por exemplo, os dados de entrada devem estar no formato do cliente, com uma linha por cliente; se os dados dos clientes estiverem no formato transacional, utilize um nó Agregado de RFM acima para derivar os campos de recência, de frequência e monetário. Consulte o tópico “Nó Agregado de RFM” na página 82 para obter mais informações.

Os nós Agregado de RFM e Análise de RFM no IBM SPSS Modeler são configurados para utilizar a categorização independente, ou seja, eles classificam e categorizam dados em cada medida de recência, de frequência e monetária, sem considerar seus valores ou as outras duas medidas.

Configurações do Nó Análise de RFM

Recência. Utilizando o Seletor de Campo (botão à direita da caixa de texto), selecione o campo de recência. Isso pode ser uma data, um registro de data e hora ou número simples. Observe que quando uma data ou um registro de data e hora representa a data da transação mais recente, o valor mais alto é considerado o mais recente e, quando um número é especificado, ele representa o tempo decorrido desde a transação mais recente e o valor mais baixo é considerado o mais recente.

Nota: se o nó Análise de RFM for precedido no fluxo por um nó Agregado de RFM, os campos Recência, Frequência e Monetário gerados pelo nó Agregado de RFM deverão ser selecionados como entradas no nó Análise de RFM.

Frequência. Utilizando o Seletor de Campo, selecione o campo de frequência a ser utilizado.

Monetário. Utilizando o Seletor de campo, selecione o campo monetário a ser utilizado.

Número de categorias. Para cada um dos três tipos de saída, selecione o número de categorias a serem criadas. O padrão é 5.

Nota: o número mínimo de categorias é 2 e o máximo é 9.

Ponderação. Por padrão, a importância mais alta ao calcular escores é fornecida aos dados de recência, seguida por frequência e, em seguida, por monetário. Se necessário, é possível corrigir a ponderação que afeta um ou vários desses para alterar qual recebe a importância mais alta.

O escore RFM é calculado da seguinte forma: (Escore de recência x Ponderação de recência) + (Escore de frequência x Ponderação de frequência) + (Escore de monetário x Ponderação de monetário).

Empates. Especifique como escores idênticos (empatados) devem ser categorizados. As opções são:

- **Incluir no próximo.** Selecione para mover os valores de empate para a próxima categoria.

- **Mantener no atual.** Mantém os valores de empate na categoria atual (inferior). Este método pode resultar em um total menor de categorias sendo criado. (Esse é o valor padrão.)

Limites da categoria. Especifique se os escores de RFM e alocações de categorias são sempre recalculados quando o nó é executado ou se são calculados apenas conforme necessário (por exemplo, quando novos dados tiverem sido incluídos). Se você selecionar **Ler a partir da guia Valores de Categoria, se disponível**, será possível editar os pontos de corte superior e inferior para as diferentes categorias na guia Valores de Categoria.

Quando executado, o nó Análise de RFM categoriza os campos de recência bruta, frequência e monetário e inclui os novos campos a seguir no conjunto de dados:

- Escore de recência. Um ranqueamento (valor de categoria) para Recência
- Escore de frequência. Um ranqueamento (valor de categoria) para Frequência
- Escore de monetário. Um ranqueamento (valor de categoria) para Monetário
- Escore de RFM. A soma ponderada dos escores de recência, frequência e monetário.

Incluir valores discrepantes nas categorias de término. Se você marcar essa caixa de seleção, os registros que estiverem abaixo da categoria inferior serão incluídos na categoria inferior, e os registros acima da categoria mais alta serão incluídos na categoria mais alta, caso contrário, eles receberão um valor nulo. Essa caixa estará disponível apenas se selecionar **Ler a partir da guia Valores de Categoria, se disponível**.

Categorização do Nó Análise de RFM

A guia Valores de Categorização permite visualizar e, em alguns casos, corrigir os limites para as categorias geradas.

Nota: será possível corrigir os valores nesta guia somente se selecionar **Ler a partir da guia Valores de Categoria, se disponível** na guia Configurações.

Campo Categorizado. Utilize a lista suspensa para selecionar um campo e dividi-lo em categorias. Os valores disponíveis são aqueles selecionados na guia Configurações.

Tabela de valores de categoria. Os valores limite são mostrados aqui para cada categoria gerada. Se selecionar **Ler a partir da guia Valores de Categoria, se disponível** na guia Configurações, será possível alterar os pontos de corte superior e inferior de cada categoria ao dar um clique duplo na célula relevante.

Ler Valores. Lê valores categorizados a partir do conjunto de dados e preenche a tabela de valores de categoria. Observe que se você selecionar **Sempre recalcular** na guia Configurações, os limites de categoria serão sobrescritos quando novos dados forem executados por meio do fluxo.

Nó Combinação

O nó Combinação combina dois ou mais nuggets do modelo para obter predições mais precisas do que pode ser obtido a partir de quaisquer modelos individuais. Ao combinar predições a partir de diversos modelos, as limitações em modelos individuais podem ser evitadas, resultando em uma precisão geral maior. Os modelos combinados desta maneira geralmente executam pelo menos tão bem quanto o melhor dos modelos individuais e muitas vezes melhor.

Essa combinação de nós ocorre automaticamente nos nós de modelagem Classificador Automático, Numeração Automática e Cluster Automático automatizados.

Após utilizar um nó Combinação, será possível utilizar um nó Análise ou um nó Avaliação para comparar a precisão dos resultados combinados com cada um dos modelos de entrada. Para isso, assegure-se de que a opção **Filtrar campos gerados por modelos combinados** não esteja selecionada na guia Configurações no nó Combinação.

Campos de Saída

Cada nó Combinação gera um campo contendo os resultados combinados. O nome baseia-se no campo de destino especificado e é prefixado com \$XF_, \$XS_ ou \$XR_, dependendo do nível de medição do campo--flag, nominal (conjunto) ou contínuo (intervalo), respectivamente. Por exemplo, se o destino for um campo de flag denominado *resposta*, o campo de saída será *\$XF_response*.

Campos de confiança ou de propensão. Para campos de flag e nominais, campos de confiança ou de propensão adicionais são criados com base no método de combinação, conforme detalhado na tabela a seguir.

Tabela 27. Criação do campo no método de combinação.

Método de combinação	Nome do campo
Votação Votação de confiança ponderada Bruto-propensão-votação ponderada Ajustado-propensão-votação ponderada A confiança mais alta vence	\$XFC_<field>
Média de propensão bruta	\$XFRP_<field>
Média de propensão bruta ajustada	\$XFAP_<field>

Configurações do Nó Combinação

Campo de destino para combinação. Selecione um campo único que seja utilizado como o destino por dois ou mais modelos de envio de dados. Os modelos de envio de dados podem utilizar variáveis resposta de flag, nominais e contínuas, mas pelo menos dois dos modelos devem compartilhar o mesmo destino para combinar os escores.

Filtrar campos gerados por modelos combinados. Remove da saída todos os campos adicionais gerados pelos modelos individuais que são alimentados no nó Combinação. Marque esta caixa de seleção se você estiver interessado apenas no escore combinado de todos os modelos de entrada. Assegure-se de que esta opção esteja desmarcada se, por exemplo, desejar utilizar um nó Análise ou nó Avaliação para comparar a precisão do escore combinado com a precisão de cada um dos modelos de entrada individuais.

As configurações disponíveis dependem do nível de medição do campo selecionado como o destino.

Variáveis Resposta Contínuas

Para uma variável resposta contínua, os escores terão a média calculada. Este é o único método disponível para combinar escores.

Ao calcular a média dos escores ou das estimativas, o nó Combinação utiliza um cálculo de erro padrão para trabalhar a diferença entre os valores medidos ou estimados e os valores reais e para mostrar quão próximas essas estimativas corresponderam. Os cálculos de erro padrão são gerados por padrão para novos modelos, entretanto, é possível desmarcar a caixa de seleção para modelos existentes, por exemplo, se eles tiverem que ser gerados novamente.

Respostas categóricas

Para variáveis resposta categóricas, um número de métodos é suportado, incluindo **votação**, que funciona ao calcular o número de vezes que cada valor predito possível é escolhido e selecionar o valor com o maior total. Por exemplo, se três dos cinco modelos preverem *sim* e os outros dois preverem *não*, então *sim* ganhará por 3 votos a 2. Como alternativa, os votos podem ser **ponderados** com base no valor de confiança ou de propensão para cada predição. Em seguida, as ponderações são somadas, e o valor com o maior total será selecionado novamente. A confiança para a predição final é a soma das ponderações do valor vencedor dividida pelo número de modelos incluídos na combinação.

Todos os campos categóricos. Para os campos de flag e nominais, os seguintes métodos são suportados:

- Votação
- Votação de confiança ponderada
- A confiança mais alta vence

Campos de flag somente. Somente para os campos de flag, um número de métodos com base na propensão também está disponível:

- Votação de propensão ponderada bruta
- Votação de propensão ponderada ajustada
- Média de propensão bruta
- Propensão ajustada média

Empates de votação. Para métodos de votação, é possível especificar como os empates são resolvidos.

- **Seleção aleatória.** Um dos valores empatados é escolhido aleatoriamente.
- **Confiança mais alta.** O valor empatado que foi predito com a confiança mais alta vence. Observe que essa não é necessariamente a mesma confiança mais alta de todos os valores preditos.
- **Propensão bruta ou ajustada (somente campos de flag).** O valor empatado que foi predito com a propensão absoluta maior, em que a propensão absoluta é calculada como:

$$\frac{\text{abs}(0.5 - \text{propensity}) * 2}{2}$$

Ou, no caso de propensão ajustada:

$$\text{abs}(0.5 - \text{adjusted propensity}) * 2$$

Nó Partição

Os nós Partição são utilizados para gerar um campo de partição que divide os dados em subconjuntos ou amostras separados para os estágios de treinamento, de teste e de validação da construção de modelo. Ao utilizar uma amostra para gerar o modelo e uma amostra separada para testá-lo, é possível obter uma boa indicação do quão bem o modelo será generalizado para conjuntos de dados maiores que forem semelhantes aos dados atuais.

O nó Partição gera um campo nominal com o papel configurado como **Partição**. Como alternativa, se um campo apropriado já existir em seus dados, ele poderá ser designado como uma partição utilizando um nó Tipo. Nesse caso, nenhum nó Partição separado é necessário. Qualquer campo nominal instanciado com dois ou três valores pode ser utilizado como uma partição, mas os campos de flag não podem ser utilizados. Consulte o tópico “Configurando o Papel de Campo” na página 142 para obter mais informações.

Diversos campos de partição podem ser definidos em um fluxo, e se for o caso, um campo de partição único deverá ser selecionado na guia Campos em cada nó de modelagem que utilizar particionamento. (Se apenas uma partição estiver presente, ela será utilizada automaticamente sempre que o particionamento estiver ativado).

Ativar particionamento. Para utilizar a partição em uma análise, o particionamento deve ser ativado na guia Opções de Modelo no nó de construção ou de análise de modelo apropriado. Desmarcar esta opção permite desativar o particionamento sem remover o campo.

Para criar um campo de partição com base em algum outro critério, como um intervalo de data ou localização, também é possível utilizar um nó Derivar. Consulte o tópico “Derivar Nó” na página 148 para obter mais informações.

Exemplo. Ao construir um fluxo de RFM para identificar clientes recentes que tenham respondido positivamente as campanhas de marketing anteriores, o departamento de marketing de uma empresa de vendas utiliza um nó Partição para dividir os dados em partições de treinamento e de teste.

Opções do Nó Partição

Campo de partição. Especifica o nome do campo criado pelo nó.

Partições. É possível particionar os dados em duas amostras (treino e teste) ou três (treino, teste e validação).

- **Treino e teste.** Particiona os dados em duas amostras, permitindo treinar o modelo com uma amostra e testar com outra.
- **Treino, teste e validação.** Particiona os dados em três amostras, permitindo treinar o modelo com uma amostra, testar e refinar o modelo utilizando uma segunda amostra e validar seus resultados com uma terceira amostra. Embora esse processo reduza o tamanho de cada partição apropriadamente, ele poderá ser mais adequado quando estiver trabalhando com um conjunto de dados muito grande.

Tamanho da partição. Especifica o tamanho relativo de cada partição. Se a soma dos tamanhos de partição for menor que 100%, então os registros não incluídos em uma partição serão descartados. Por exemplo, se um usuário tiver 10 milhões de registros e especificou tamanhos de partição de 5% de treinamento e 10% de teste, após executar o nó, deverá haver aproximadamente 500.000 registros de treinamento e um milhão de registros de teste, com o restante sendo descartado.

Valores. Especifica os valores utilizados para representar cada amostra de partição nos dados.

- **Usar valores definidos pelo sistema ("1," "2" e "3").** Utiliza um número inteiro para representar cada partição, por exemplo, todos os registros que caírem na amostra de treinamento possuem um valor igual a 1 para o campo de partição. Isso assegura que os dados sejam móveis entre os códigos de idioma e que, se o campo de partição for reinstanciado em outro lugar (por exemplo, ler os dados de volta a partir de um banco de dados), a ordenação será preservada (de modo que 1 ainda representará a partição de treinamento). No entanto, os valores requerem alguma interpretação.
- **Anexar rótulos aos valores definidos pelo sistema.** Combina o número inteiro com um rótulo, por exemplo, os registros da partição de treinamento possuem um valor de 1_Training. Isso permite que alguém que examina os dados identifique qual é cada valor e preserve a ordenação. No entanto, os valores são específicos para um determinado código de idioma.
- **Usar rótulos como valores.** Utiliza o rótulo sem nenhum número inteiro, por exemplo, **Treinamento**. Isso permite especificar os valores ao editar os rótulos. No entanto, ele torna os dados específicos do código de idioma, e a reinstanciação de uma coluna de partição colocará os valores em sua ordenação natural, que pode não corresponder à sua ordem "semântica".

Semente. Disponível apenas quando **Designação de partição repetível** está selecionada. Ao amostrar ou particionar registros com base em uma porcentagem aleatória, esta opção permite duplicar os mesmos resultados em outra sessão. Ao especificar o valor inicial utilizado pelo gerador de número aleatório, é possível assegurar que os mesmos registros sejam designados toda vez que o nó for executado. Insira o valor semente desejado ou clique no botão **Gerar** para gerar automaticamente um valor aleatório. Se essa opção não estiver selecionada, uma amostra diferente será gerada toda vez que o nó for executado.

Nota: Ao usar a opção **Semente** com registros lidos a partir de um banco de dados, um nó Ordem poderá ser necessário antes da amostragem para assegurar o mesmo resultado sempre que o nó for executado. Isso ocorre porque a semente aleatória depende da ordem de registros, que não é garantido que ela permaneça a mesma em um banco de dados relacional. Consulte o tópico “Nó Ordem” na página 83 para obter mais informações.

Usar campo único para designar partições. Disponível apenas quando **Designação de partição repetível** está selecionada. (Para bancos de dados de Camada 1 somente) Marque esta caixa para utilizar o SQL pushback para designar registros às partições. Na lista suspensa, escolha um campo com valores exclusivos (como um campo de ID) para assegurar que os registros sejam designados de forma aleatória, porém repetível.

As camadas do banco de dados são explicadas na descrição do nó Origem da base de dados. Consulte o tópico “Nó de Origem da Base de Dados” na página 17 para obter mais informações.

Gerando nós de seleção

Utilizando o menu Gerar no nó Partição, é possível gerar automaticamente um nó Seleção para cada partição. Por exemplo, é possível selecionar todos os registros na partição de treinamento para obter uma avaliação ou análise adicional utilizando apenas a essa partição.

Configurar para o Nó Sinalização

O nó Configurar para Flag é usado para derivar campos de flag com base nos valores categóricos definidos para um ou mais campos nominais. Por exemplo, seu conjunto de dados pode conter um campo nominal, *BP* (pressão arterial), com os valores *Alto*, *Normal* e *Baixo*. Para facilitar a manipulação de dados, é possível criar um campo de sinalização para pressão alta, que indica se o paciente tem pressão alta ou não.

Configurando Opções para o Nó Configurar para Flag

Configurar campos. Lista todos os campos de dados com um nível de medição de *Nominal* (conjunto). Selecione um na lista para exibir os valores no conjunto. É possível escolher a partir desses valores para criar um campo de sinalização. Observe que os dados devem ser totalmente instanciados utilizando uma origem de envio de dados ou nó Tipo antes de poder ver os campos nominais disponíveis (e seus valores). Consulte o tópico “Nó Tipo” na página 131 para obter mais informações.

Extensão do nome do campo. Selecione para ativar controles para especificar uma extensão que será incluída como um sufixo ou prefixo para o novo campo de sinalização. Por padrão, os novos nomes de campo são criados automaticamente ao combinar o nome de campo original com o valor do campo em um rótulo, como *Fieldname_fieldvalue*.

Valores de conjunto disponíveis. Os valores no conjunto selecionado acima são exibidos aqui. Selecione um ou mais valores para os quais deseja gerar flags. Por exemplo, se os valores em um campo chamado *blood_pressure* forem *Alto*, *Médio* e *Baixo*, será possível selecionar *Alto* e incluí-lo na lista à direita. Isso criará um campo com um flag para registros com um valor indicando a pressão sanguínea alta.

Criar campos de flag. Os campos de flag recém-criados são listados aqui. É possível especificar opções para nomear o novo campo utilizando os controles da extensão do nome do campo.

Valor True. Especifique o valor real utilizado pelo nó quando configurar um flag. Por padrão, esse valor é T.

Valor False. Especifique o valor false utilizado pelo nó quando configurar um flag. Por padrão, esse valor é F.

Agregar chaves. Selecione para agrupar registros com base nos campos-chave especificados abaixo. Quando **Agregar chaves** é selecionada, todos os campos de flag em um grupo serão "ativados" se *qualquer* registro foi configurado como true. Utilize o Seletor de Campo para especificar quais campos-chave serão utilizados para agregar registros.

Nó Reestruturar

O nó Reestruturar pode ser utilizado para gerar diversos campos com base nos valores de um campo nominal ou de flag. Os campos recém-gerados podem conter valores de outros flags de campo ou numéricos (0 e 1). A funcionalidade deste nó é semelhante à funcionalidade do nó Configurar para Flag. No entanto, ela oferece mais flexibilidade. Ela permite criar campos de qualquer tipo (incluindo flags numéricos), utilizando os valores de outro campo. Em seguida, é possível executar agregação ou outras manipulações com outros nós de recebimento de dados. (O nó Configurar para Flag permite agregar campos em um passo, que poderá ser conveniente se você estiver criando campos de flag).

Por exemplo, o conjunto de dados a seguir contém um campo nominal, *Account*, com valores *Savings* e *Draft*. O saldo de abertura e o saldo atual são registrados para cada conta, e alguns clientes possuem diversos tipos de contas. Suponhamos que você queira saber se cada cliente possui um tipo de conta específico e, em caso afirmativo, quanto dinheiro cada tipo de conta possui. É possível utilizar o nó Reestruturar para gerar um campo para cada um dos valores de *Conta*, e você seleciona *Current_Balance* como o valor. Cada novo campo é preenchido com o saldo atual para o registro especificado.

Tabela 28. Dados de amostra antes da reestruturação.

CustID	Account	Open_Bal	Current_Bal
12701	Rascunho	1000	1005.32
12702	Economias	100	144,51
12703	Economias	300	321.20
12703	Economias	150	204.51
12703	Rascunho	1200	586.32

Tabela 29. Dados de amostra após a reestruturação.

CustID	Account	Open_Bal	Current_Bal	Account_Draft_Current_Bal	Account_Savings_Current_Bal
12701	Rascunho	1000	1005.32	1005.32	\$null\$
12702	Economias	100	144,51	\$null\$	144,51
12703	Economias	300	321.20	\$null\$	321.20
12703	Economias	150	204.51	\$null\$	204.51
12703	Rascunho	1200	586.32	586.32	\$null\$

Utilizando o Nó Reestruturar com o Nó Agregado

Em muitos casos, talvez você queira emparelhar o nó Reestruturar com um nó Agregado. No exemplo anterior, um cliente (com o ID 12703) possui três contas. É possível utilizar um nó Agregado para calcular o saldo total para cada tipo de conta. O campo-chave é *CustID*, e os campos agregados são os novos campos reestruturados, *Account_Draft_Current_Bal* e *Account_Savings_Current_Bal*. A tabela a seguir mostra os resultados.

Tabela 30. Dados de amostra após a reestruturação e agregação.

CustID	Record_Count	Account_Draft_Current_Bal_Sum	Account_Savings_Current_Bal_Sum
12701	1	1005.32	\$null\$
12702	1	\$null\$	144,51
12703	3	586.32	525.71

Configurando Opções para o Nó Reestruturar

Campos disponíveis. Lista todos os campos de dados com um nível de medição de *Nominal* (configurado) ou de *sinalização*. Selecione um na lista para exibir os valores no conjunto ou no flag e, em seguida, escolha entre esses valores para criar os campos reestruturados. Observe que os dados devem ser totalmente instanciados utilizando uma origem de envio de dados ou nó Tipo antes de poder ver os campos disponíveis (e seus valores). Consulte o tópico “Nó Tipo” na página 131 para obter mais informações.

Valores disponíveis. Os valores no conjunto selecionado acima são exibidos aqui. Selecione um ou mais valores para os quais deseja gerar campos reestruturados. Por exemplo, se os valores em um campo chamado *Pressão Arterial* forem *Alto*, *Médio* e *Baixo*, será possível selecionar *Alto* e incluí-lo na lista à direita. Isso criará um campo com um valor especificado (consulte abaixo) para registros com um valor de *Alto*.

Criar campos reestruturados. Os campos reestruturados recém-criados são listados aqui. Por padrão, os novos nomes de campo são criados automaticamente ao combinar o nome de campo original com o valor do campo em um rótulo, como *Fieldname_fieldvalue*.

Incluir nome do campo. Cancele a seleção para remover o nome de campo original como um prefixo dos novos nomes de campo.

Usar valores de outros campos. Especifique um ou mais campos cujos valores serão utilizados para preencher os campos reestruturados. Utilize o Seletor de Campo para selecionar um ou mais campos. Para cada campo escolhido, um novo campo é criado. O nome do campo de valor é anexado ao nome do campo reestruturado, por exemplo, *BP_High_Age* ou *BP_Low_Age*. Cada campo herda o tipo do campo valor original.

Criar flags de valor numérico. Selecione para preencher os novos campos com flags de valor numérico (0 para false e 1 para true) ao invés de utilizar um valor de outro campo.

Nó Transpor

Por padrão, as colunas são campos e linhas são registros ou observações. Se necessário, é possível utilizar um nó Transpor para trocar os dados nas linhas e colunas para que os campos se tornem registros e os registros se tornem campos. Por exemplo, se você tiver dados de séries temporais em que cada série é uma linha ao invés de uma coluna, será possível transpor os dados antes da análise.

Configurando Opções para o Nó Transpor

Novos nomes de campo

Novos nomes de campos podem ser gerados automaticamente com base em um prefixo especificado ou lidos a partir de um campo existente nos dados.

Usar prefixo. Esta opção gera novos nomes de campo automaticamente com base no prefixo especificado (*Field1*, *Field2*, e assim por diante). É possível customizar o prefixo conforme necessário. Com essa opção, deve-se especificar o número de campos a serem criados, independentemente do número de linhas nos

dados originais. Por exemplo, se **Número de novos campos** for configurado como 100, todos os dados além das primeiras 100 linhas serão descartados. Se houver menos de 100 linhas nos dados originais, alguns campos serão nulos. (É possível aumentar o número de campos conforme necessário, no entanto, o propósito dessa configuração é evitar a transposição de um milhão de registros em um milhão de campos, o que produziria um resultado não gerenciável).

Por exemplo, suponha que você tenha dados com séries em linhas e um campo separado (coluna) para cada mês. É possível transpor isso para que cada série esteja em um campo separado, com uma linha para cada mês.

Ler a partir do campo. Lê nomes de campo a partir de um campo existente. Com essa opção, o número de novos campos é determinado pelos dados, até o máximo especificado. Cada valor do campo selecionado se torna um novo campo nos dados de saída. O campo selecionado pode ter qualquer tipo de armazenamento (número inteiro, sequência de caracteres, data, e assim por diante), mas para evitar nomes de campo duplicados, cada valor do campo selecionado deve ser exclusivo (em outras palavras, o número de valores deve corresponder ao número de linhas). Se nomes de campo duplicados forem encontrados, um aviso será exibido.

- **Ler Valores.** Se o campo selecionado não tiver sido instanciado, selecione esta opção para preencher a lista de novos nomes de campo. Se o campo já tiver sido instanciado, então esse passo não será necessário.
- **número máximo de valores para ler.** Ao ler nomes de campos a partir dos dados, um limite superior é especificado para evitar a criação de um número excessivamente grande de campos. (Conforme observado acima, a transposição de um milhão de registros em um milhão de campos produziria um resultado não gerenciável).

Por exemplo, se a primeira coluna em seus dados especificar o nome para cada série, será possível utilizar esses valores como nomes de campos nos dados transpostos.

Transpor. Por padrão, somente campos contínuos (intervalo numérico) são transpostos (armazenamento de número inteiro ou real). Opcionalmente, é possível escolher um subconjunto de campos numéricos ou de campos de sequências de caracteres de transposição. No entanto, todos os campos transpostos devem ser do mesmo tipo de armazenamento, numérico ou de sequência de caracteres, mas não ambos, já que a combinação dos campos de entrada geraria valores combinados dentro de cada coluna de saída, o que viola a regra de que todos os valores de um campo devem ter o mesmo armazenamento. Outros tipos de armazenamento (data, hora, registro de data e hora) não podem ser transpostos.

- **Tudo numérico.** Transpõe todos os campos numéricos (armazenamento de número inteiro e real). O número de linhas na saída corresponde ao número de campos numéricos nos dados originais.
- **Tudo sequência de caracteres.** Transpõe todos os campos de sequência de caracteres.
- **Customizado.** Permite selecionar um subconjunto de campos numéricos. O número de linhas na saída corresponde ao número de campos selecionados. *Nota:* esta opção está disponível apenas para campos numéricos.

Nome do ID de linha. Especifica o nome do campo de ID linha criado pelo nó. Os valores deste campo são determinados pelos nomes dos campos nos dados originais.

Dica: ao transpor dados de série temporal de linhas para colunas, se seus dados originais incluírem uma linha, como data, mês ou ano, que rotula o período para cada medição, assegure-se de ler estes rótulos no IBM SPSS Modeler como nomes de campos (conforme demonstrado nos exemplos acima, que mostram o mês ou a data como nomes de campos nos dados originais, respectivamente), ao invés de incluir o rótulo na primeira linha de dados. Isso evitará combinação de rótulos e valores em cada coluna (o que forçaria os números serem lidos como sequências de caracteres, já que os tipos de armazenamento não podem ser combinados dentro de uma coluna).

Nó Intervalos de Tempo

O nó Intervalos de Tempo permite especificar intervalos e gerar rótulos para dados de séries temporais a serem utilizados em um nó de modelagem de Séries Temporais, em um nó TS de Fluxo ou em um nó Gráfico de Tempo para estimativa ou previsão. Uma variedade completa de intervalos de tempo é suportada, desde segundos até anos. Por exemplo, se você tiver uma série de medições diárias com início em 3 de janeiro de 2005, será possível rotular registros nessa data, com a segunda linha sendo 4 de janeiro, e assim por diante. Também é possível especificar a periodicidade – por exemplo, cinco dias por semana ou oito horas por dia.

Além disso, é possível especificar o intervalo de registros a serem utilizados para estimativa. É possível escolher se deseja excluir os primeiros registros na série e se validações devem ser especificadas. Isso permite testar o modelo ao reter os registros mais recentes nos dados de série temporal para comparar seus valores conhecidos com os valores estimados para esses períodos.

Também é possível especificar quantos períodos de tempo no futuro você deseja prever, e é possível especificar valores futuros para uso na previsão pelos nós de modelagem Séries Temporais ou pelos nós TS de Fluxo de recebimento de dados.

O nó Intervalos de Tempo gera um campo *TimeLabel* em um formato apropriado para o intervalo e o período especificados e também um campo *TimeIndex* que designa um número inteiro exclusivo para cada registro. Um número de campos adicionais também pode ser gerado, dependendo do intervalo ou do período selecionado (como o minuto ou o segundo dentro do qual uma medição cai).

É possível preencher ou agregar valores, conforme necessário, para assegurar que as medições sejam igualmente espaçadas. Os métodos para modelar dados de séries temporais requerem um intervalo uniforme entre cada medição, com quaisquer valores omissos indicados por linhas vazias. Se seus dados ainda não atenderem a esse requisito, o nó poderá transformá-los para isso.

Comentários

- Intervalos periódicos podem não corresponder ao tempo real. Por exemplo, uma série com base em uma semana de trabalho de cinco dias padrão trata a diferença entre sexta-feira e segunda-feira como um único dia.
- O nó Intervalos de Tempo supõe que cada série está em um campo ou coluna, com uma linha para cada medição. Se necessário, é possível transpor seus dados para atender a esse requisito. Consulte o tópico “Nó Transpor” na página 174 para obter mais informações.
- Para séries que não estiverem igualmente espaçadas, é possível especificar um campo que identifica a data ou a hora de cada medição. Observe que isto requer um campo de data, de hora ou de registros de data e hora, no formato apropriado para uso como entrada. Se necessário, é possível converter um campo existente (como um campo de rótulo de sequência de caracteres) nesse formato utilizando um nó Preenchimento. Consulte o tópico “Conversão de Armazenamento Usando o Nó Preenchimento” na página 155 para obter mais informações.
- Ao visualizar detalhes para o rótulo e campos de índice gerados, geralmente é útil ativar a exibição de rótulos de valor. Por exemplo, ao visualizar uma tabela com valores gerados para os dados mensais, é possível clicar no ícone de rótulos de valor na barra de ferramentas para ver *Janeiro*, *Fevereiro*, *Março*, e assim por diante, ao invés de 1, 2, 3, e assim por diante.



Figura 6. Ícone de rótulos de valor

Especificando Intervalos de Tempo

A guia Intervalos permite especificar o intervalo e a periodicidade para construir ou rotular as séries. As configurações específicas dependem do intervalo selecionado. Por exemplo, se escolher **Horas por dia**, será possível especificar o número de dias por semana, o dia em que cada semana começa, o número de horas em cada dia e a hora em que cada dia começa. Consulte o tópico “Intervalos suportados” na página 180 para obter mais informações.

Rotulando ou Construindo as Séries

É possível rotular registros consecutivamente ou construir a série com base em um campo de data, de registro de data e hora ou de hora especificado.

- **Iniciar rotulagem a partir do primeiro registro.** Especifique a data e/ou a hora de início para rotular registros consecutivos. Se estiver rotulando em horas por dia, por exemplo, especifique a data e a hora quando a série inicia, com um registro único para cada hora a partir disso. Além de incluir rótulos, esse método não altera os dados originais. Ao invés disso, ele supõe que os registros já estejam igualmente espaçados, com um intervalo uniforme entre cada medição. Quaisquer medições omissas devem ser indicadas por linhas vazias nos dados.
- **Construir a partir dos dados.** Para séries que não estiverem igualmente espaçadas, é possível especificar um campo que identifica a data ou a hora de cada medição. Observe que isto requer um campo de data, de hora ou de registros de data e hora, no formato apropriado para uso como entrada. Por exemplo, se você tiver um campo de sequência de caracteres com valores como *Jan 2000*, *Feb 2000*, e assim por diante, será possível convertê-lo em um campo de data utilizando um nó Preenchimento. Consulte o tópico “Conversão de Armazenamento Usando o Nó Preenchimento” na página 155 para obter mais informações. A opção **Construir a partir dos dados** também transforma os dados para corresponder ao intervalo especificado pelo preenchimento ou agregação de registros conforme necessário, por exemplo, ao “acumular” semanas em meses ou substituir os registros omissos por valores em branco ou extrapolados. É possível especificar as funções utilizadas para preencher ou agregar registros na guia Construção. Consulte o tópico “Opções de Construção de Intervalo de Tempo” para obter mais informações.

Extensão do novo nome de campo. Permite especificar um prefixo ou um sufixo que é aplicado a todos os campos gerados pelo nó. Por exemplo, utilizando o prefixo *\$TI_* padrão, os campos criados pelo nó serão denominados *\$TI_TimeIndex*, *\$TI_TimeLabel*, e assim por diante.

Formato de data. Especifica o formato para o campo *TimeLabel* criado pelo nó, conforme aplicável para o intervalo atual. A disponibilidade dessa opção depende da seleção atual.

Formato de hora. Especifica o formato para o campo *TimeLabel* criado pelo nó, conforme aplicável para o intervalo atual. A disponibilidade dessa opção depende da seleção atual.

Opções de Construção de Intervalo de Tempo

A guia Construção no nó Intervalos de Tempo permite especificar opções para agregar e preencher campos para corresponder ao intervalo especificado. Essas configurações se aplicam apenas quando a opção **Construir a partir de dados** estiver selecionada na guia Intervalos. Por exemplo, se você tiver uma combinação de dados semanais e mensais, será possível agregar ou “acumular” os valores semanais para atingir um intervalo mensal uniforme. Como alternativa, é possível configurar o intervalo para semanalmente e preencher as séries ao inserir valores em branco para quaisquer semanas que estiverem faltando ou ao extrapolar valores omissos utilizando uma função de preenchimento especificada.

Ao preencher ou agregar dados, quaisquer campos de data ou de registro de data e hora existentes serão efetivamente substituídos pelos campos *TimeLabel* e *TimeIndex* gerados e, em seguida, eliminados da saída. Os campos sem tipo também são eliminados. Os campos que medem o tempo como uma duração são preservados, por exemplo, um campo que mede a duração de uma chamada de serviço ao invés da hora em que a chamada iniciou, contanto que eles estejam armazenados internamente como campos de tempo

ao invés de registro de data e hora. Consulte o tópico “Configurando Armazenamento e Formatação do Campo” na página 9 para obter mais informações. Outros campos são agregados com base nas opções especificadas na guia Construção.

- **Usar campos e funções padrão.** Especifica que todos os campos devem ser agregados ou preenchidos conforme necessário, com exceção dos campos de data, de registro de data e hora e sem tipo, conforme observado acima. A função padrão é aplicada com base no nível de medição, por exemplo, campos contínuos são agregados utilizando a média, ao passo que campos nominais utilizam o modo. É possível alterar o padrão para um ou mais níveis de medição na parte inferior da caixa de diálogo.
- **Especificar campos e funções.** Permite especificar os campos a serem preenchidos ou agregados e a função utilizada para cada. Quaisquer campos não selecionados são eliminados da saída. Utilize os ícones no lado direito para incluir ou remover campos da tabela ou clique na célula na coluna apropriada para alterar a função de agregação ou de preenchimento utilizada para esse campo para substituir o padrão. Os campos sem tipo são excluídos da lista e não podem ser incluídos na tabela.

Padrões. Especifica as funções de agregação e de preenchimento utilizadas por padrão para diferentes tipos de campos. Esses padrões são aplicados quando **Usar padrão** é selecionada e também são aplicados como o padrão inicial para quaisquer novos campos incluídos na tabela. (A mudança nos Padrões não altera nenhuma das configurações existentes na tabela, mas se aplica a todos os campos incluídos subsequentemente).

Função de agregação. As funções de agregação a seguir estão disponíveis:

- **Contínuo.** As funções disponíveis para campos contínuos incluem **Média, Soma, Modo, Mín. e Máx..**
- **Nominal.** As opções incluem **Modo, Primeiro e Último.** Primeiro significa o primeiro valor não nulo (classificado por data) no grupo de agregação e último significa o último valor não nulo no grupo.
- **Flag.** As opções incluem **True se algum for verdadeiro, Modo, Primeiro e Último.**

Função de preenchimento. As seguintes funções de preenchimento estão disponíveis:

- **Contínuo.** As opções incluem **Em Branco e Média de pontos mais recentes**, o que significa a média dos três valores não nulos mais recentes antes do período de tempo que será criado. Se não houver três valores, o novo valor estará em branco. Os valores recentes incluem apenas valores reais e um valor preenchido criado anteriormente não é considerado na procura de um valor não nulo.
- **Nominal. Em Branco e Valor mais recente.** Mais Recente refere-se ao valor não nulo mais recente antes do período de tempo que será criado. Mais uma vez, apenas valores reais são considerados na procura de um valor recentes.
- **Flag.** As opções incluem **Em Branco, True e False.**

Número máximo de registros no conjunto de dados resultante. Especifica um limite superior para o número de registros criados, o que poderá do contrário se tornar muito grande, especialmente quando o intervalo de tempo for configurado para segundos (seja deliberadamente ou de outra forma). Por exemplo, uma série com apenas dois valores (Jan. 1, 2000 e Jan. 1, 2001) geraria 31.536.000 registros se preenchida para segundos (60 segundos x 60 minutos x 24 horas x 365 dias). O sistema parará o processamento e exibirá um aviso se o máximo especificado for excedido.

Campo de Contagem

Ao agregar ou preencher valores, um novo campo *Contagem* é criado indicando o número de registros envolvidos na determinação do novo registro. Se quatro valores semanais forem agregados em um único mês, por exemplo, a contagem seria 4. Para um registro preenchido, a contagem é 0. O nome do campo é *Contagem*, mais o prefixo ou sufixo especificado na guia Intervalo.

Período de estimação

A guia Estimação do nó Intervalos de Tempo permite especificar o intervalo de registros usados na estimação do modelo, bem como quaisquer validações. Essas configurações podem ser substituídas nos

nós de modelagem recebimento de dados conforme necessário, mas especificá-las aqui pode ser mais conveniente do que especificá-las para cada nó individualmente.

Iniciar Estimativa. É possível iniciar o período de estimação no início dos dados ou excluir valores mais antigos que podem ser de uso limitado na predição. Dependendo dos dados, você pode descobrir que a redução do período de estimação pode acelerar o desempenho (e reduzir a quantidade de tempo gasto na preparação de dados) sem nenhuma perda significativa de precisão da previsão.

Terminar Estimativa. É possível estimar o modelo usando todos os registros até o término dos dados ou "validar" os registros mais recentes para avaliar o modelo. No último caso, você está efetivamente "prevendo" valores que são já conhecidos, permitindo comparar valores observados e preditos para avaliar a eficácia do modelo.

Previsões

A guia Previsão do nó Intervalos de Tempo permite especificar o número de registros que você deseja prever e especificar valores futuros para uso na previsão feita pelos nós de modelagem Séries Temporais ou nós TS de Fluxo de recebimento de dados. Essas configurações podem ser substituídas nos nós de recebimento de dados conforme necessário, mas especificá-las aqui pode ser mais conveniente do que especificá-las para cada nó individualmente.

Estender registros no futuro. Especifica o número de registros para prever além do período de estimação. Observe que esses registros podem ou não ser "previsões" dependendo do número de validações especificadas na guia Estimativa.

Campo indicador de futuro. Rótulo do campo gerado que indica se um registro contém dados de previsão. O valor padrão para o rótulo é *\$TI_Future*.

Valores Futuros para Uso na Previsão. Para cada registro que deseja prever (exceto validações), se você estiver utilizando campos preditores (com o papel configurado para *Entrada*), deve-se especificar valores estimados para o período de previsão para cada preditor. É possível especificar valores manualmente ou escolher a partir de uma lista.

- **Campo.** Clique no botão seletor de campo e escolha quaisquer campos que possam ser utilizados como preditores. Observe que os campos selecionados aqui podem ou não ser utilizados na modelagem; para realmente utilizar um campo como um preditor, ele deverá ser selecionado em um nó de modelagem de recebimento de dados. Esta caixa de diálogo simplesmente fornece um local conveniente para especificar valores futuros para que eles possam ser compartilhados por diversos nós de modelagem de recebimento de dados sem especificá-los separadamente em cada nó. Além disso, observe que a lista de campos disponíveis pode ser restringida pelas seleções na guia Construção. Por exemplo, se **Especificar campos e funções** for selecionada na guia Construção, quaisquer campos não agregados ou preenchidos serão eliminados do fluxo e não poderão ser utilizados na modelagem.

Nota: se valores futuros forem especificados para um campo que não estiver mais disponível no fluxo (porque ele foi eliminado ou devido a seleções atualizadas feitas na guia Construção), o campo será mostrado em vermelho na guia Previsão.

- **Valores.** Para cada campo, é possível escolher a partir de uma lista de funções ou clicar em **Especificar** para inserir valores manualmente ou escolher a partir de uma lista de valores predefinidos. Se os campos preditores estiverem relacionados a itens que estiverem sob o seu controle, ou que de outra forma são conhecidos antecipadamente, deve-se inserir valores manualmente. Por exemplo, se você estiver prevendo a receita do próximo mês para um hotel com base no número de reservas de quarto, é possível especificar o número de reservas que você realmente tem para esse período. Por outro lado, se um campo preditor estiver relacionado a algo fora de seu controle, como preço de ações, será possível utilizar uma função como o valor mais recente ou a média de pontos recentes.

As funções disponíveis dependem do nível de medição do campo.

Tabela 31. Funções disponíveis para níveis de medição.

Nível de medição	Funções
Campo Contínuo ou Nominal	Em branco Média de pontos recentes Valor mais recente Especificar
campo de sinalização	Em branco Valor mais recente Verdadeiro Falso Especificar

Média de pontos recentes—Calcula o valor futuro da média dos últimos três pontos de dados.

Valor mais recente—Configura o valor futuro para aquele do ponto de dados mais recente.

True/False—Configura o valor futuro de um campo de sinalização para True ou False, conforme especificado.

Especificar—Abre uma caixa de diálogo para especificar valores futuros manualmente ou escolhê-los a partir de uma lista predefinida.

Valores futuros

Aqui é possível especificar valores futuros para uso na previsão por meio de nós de modelagem Séries Temporais ou nós TS de Fluxo de recebimento de dados. Essas configurações podem ser substituídas nos nós de recebimento de dados conforme necessário, mas especificá-las aqui pode ser mais conveniente do que especificá-las para cada nó individualmente.

É possível inserir valores manualmente ou clicar no botão do seletor no lado direito da caixa de diálogo para escolher a partir de uma lista de valores definidos para o campo atual.

O número de valores futuros que pode ser especificado corresponde ao número de registros pelos quais você está estendendo a série temporal no futuro.

Intervalos suportados

O nó Intervalos de Tempo suporta uma variedade completa de intervalos desde segundos até anos, bem como períodos cíclicos (por exemplo, sazonais) e não-cíclicos. Especifique o intervalo no campo Intervalo de Tempo na guia Intervalos.

Pontos

Selecione **Períodos** para rotular uma série não cíclica existente que não corresponder a nenhum dos outros intervalos especificados. A série já deverá estar na ordem correta, com um intervalo uniforme entre cada medição. A opção **Construir a partir dos dados** não está disponível quando esse intervalo é selecionado.

Saída da amostra

Os registros são rotulados incrementalmente com base no valor de início especificado (*Período 1, Período 2*, e assim por diante). Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 32. Exemplos de criação de novo campo

\$TI_TimeIndex (Integer)	\$TI_TimeLabel (String)	\$TI_Period (Integer)
1	Período 1	1
2	Período 2	2
3	Período 3	3
4	Período 4	4

Tabela 32. Exemplos de criação de novo campo (continuação)

\$TI_TimeIndex (Integer)	\$TI_TimeLabel (String)	\$TI_Period (Integer)
5	Período 5	5

Períodos cíclicos

Selecione **Períodos Cíclicos** para rotular uma série existente com um ciclo de repetição que não se ajustar em nenhum dos intervalos padrão. Por exemplo, essa opção poderá ser utilizada se você tiver apenas 10 meses em seu ano fiscal. A série já deverá estar na ordem correta, com um intervalo uniforme entre cada medição. (A opção **Construir a partir dos dados** não está disponível quando esse intervalo é selecionado).

Saída da amostra

Os registros são rotulados incrementalmente com base no ciclo e no ponto de início especificados (*Ciclo 1, Período 1, Ciclo 1, Período 2*, e assim por diante). Por exemplo, com o número de períodos por ciclo configurado para 3, os novos campos são criados conforme mostrado na tabela a seguir.

Tabela 33. Exemplos de criação de novo campo.

\$TI_TimeIndex (Integer)	\$TI_TimeLabel (String)	\$TI_Cycle (Número Inteiro)	\$TI_Period (Integer)
1	Ciclo 1, Período 1	1	1
2	Ciclo 1, Período 2	1	2
3	Ciclo 1, Período 3	1	3
4	Ciclo 2, Período 1	2	1
5	Ciclo 2, Período 2	2	2

Anos

Para anos, é possível especificar o ano de início para rotular registros consecutivos ou selecionar **Construir a partir dos dados** para especificar um campo de registro de data e hora ou de data que identifica o ano para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 34. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Sequência de Caracteres)	\$TI-Year (Integer)
1	2000	2000
2	2001	2001
3	2002	2002
4	2003	2003
5	2004	2004

Trimestres

Para uma série trimestral, é possível especificar o mês em que o ano fiscal começa. Também é possível especificar o trimestre e o ano de início (por exemplo, Q1 2000) para rotular registros consecutivos ou selecionar **Construir a partir de dados** para escolher um campo de registro de data e hora ou de data que identifica o trimestre e o ano para cada registro.

Saída da amostra

Para um ano fiscal que inicia em janeiro, novos campos seriam criados e preenchidos conforme mostrado na tabela a seguir.

Tabela 35. Exemplos de criação de novo campo

\$TI-TimeIndex (Número Inteiro)	\$TI-TimeLabel (String)	\$TI-Year (Integer)	\$TI-Quarter (Número inteiro com rótulos)
1	Q1 2000	2000	1 (Q1)
2	Q2 2000	2000	2 (Q2)
3	Q3 2000	2000	3 (Q3)
4	Q4 2000	2000	4 (Q4)
5	Q1 2001	2001	1 (Q1)

Se o ano iniciar em um mês diferente de janeiro, novos campos serão mostrados na tabela a seguir (supondo um ano fiscal iniciando em julho). Para visualizar os rótulos que identificam os meses para cada trimestre, ative a exibição dos rótulos de valor ao clicar no ícone da barra de ferramentas.



Figura 7. Ícone de rótulos de valor

Tabela 36. Exemplos de criação de novo campo

\$TI-TimeIndex (Número Inteiro)	\$TI-TimeLabel (String)	\$TI-Year (Integer)	\$TI-Quarter (Número inteiro com rótulos)
1	Q1 2000/2001	1	1 (Q1 Jul-Sep)
2	Q2 2000/2001	1	2 (Q2 Oct-Dec)
3	Q3 2000/2001	1	3 (Q3 Jan-Mar)
4	Q4 2000/2001	1	4 (Q4 Apr-Jun)
5	Q1 2001/2002	2	1 (Q1 Jul-Sep)

Meses

É possível selecionar o ano e o mês de início para rotular registros consecutivos ou selecionar **Construir a partir dos dados** para escolher um campo de registro de data e hora ou de data que indica o mês para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 37. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Data)	\$TI-Year (Integer)	\$TI-Months (Número Inteiro com rótulos)
1	Jan 2000	2000	1 (Janeiro)
2	Feb 2000	2000	2 (Fevereiro)
3	Mar 2000	2000	3 (Março)
4	Apr 2000	2000	4 (Abril)
5	May 2000	2000	5 (Maio)

Semanas (Não Periódicas)

Para uma série semanal, é possível selecionar o dia da semana em que o ciclo começa.

Observe que as semanas podem ser somente não periódicas devido a meses, trimestres e até mesmo anos diferentes que não têm necessariamente o mesmo número de semanas. Entretanto, dados com registro de data e hora podem ser facilmente agregados ou preenchidos para um nível semanal para modelos não periódicos.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 38. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Data)	\$TI-Week (Número Inteiro)
1	1999-12-27	1
2	2000-01-03	2
3	2000-01-10	3
4	2000-01-17	4
5	2000-01-24	5

O campo *\$TI-TimeLabel* para uma semana mostra o primeiro dia dessa semana. Na tabela anterior, o usuário começa a rotular a partir de 1º de janeiro de 2000. No entanto, a semana inicia na segunda-feira, e 1º de janeiro de 2000 é um sábado. Assim, a semana que inclui 1º de janeiro começa em 27 de dezembro de 1999 e é o rótulo do primeiro ponto.

O formato de Data determina as sequências de caracteres produzidas para o campo *\$TI-TimeLabel*.

Dias por Semana

Para as medições diárias que caem em um ciclo semanal, é possível especificar o número de dias por semana e o dia em que cada semana começa. É possível especificar a data de início para rotular registros consecutivos ou selecionar **Construir a partir dos dados** para escolher um campo de registro de data e hora ou de data que indica a data de cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 39. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Data)	\$TI-Week (Número Inteiro)	\$TI-Day (Número Inteiro com rótulos)
1	Jan 5 2005	1	3 (Quarta-feira)
2	Jan 6 2005	1	4 (Quinta-feira)
3	Jan 7 2005	1	5 (Sexta-feira)
4	Jan 10 2005	2	1 (Segunda-feira)
5	Jan 11 2005	2	2 (Terça-feira)

Nota: a semana sempre inicia em 1 para o primeiro período de tempo e não percorre com base no calendário. Assim, a semana 52 é seguida pela semana 53, 54, e assim por diante. A semana não reflete a semana do ano, apenas o número de incrementos semanais na série.

Dias (Não Periódicos)

Escolha dias não periódicos se você tiver medições diárias que não se ajustam em um ciclo semanal regular. É possível especificar a data de início para rotular registros consecutivos ou selecionar **Construir a partir dos dados** para escolher um campo de registro de data e hora ou de data que indica a data de cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 40. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Data)
1	Jan 5 2005
2	Jan 6 2005
3	Jan 7 2005
4	Jan 8 2005
5	Jan 9 2005

Horas por Dia

Para medições por hora que se ajustam em um ciclo diário, é possível especificar o número de dias por semana, o número de horas no dia (como um dia útil de oito horas), o dia em que a semana começa e o horário em que cada dia começa. As horas pode ser especificadas em minutos com base em um relógio de 24 horas (por exemplo, 14h05 = 2:05 p.m.)

É possível especificar a data e hora de início para rotular registros consecutivos ou selecionar **Construir a partir de dados** para escolher um campo de registro de data e hora que identifica a data e hora para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 41. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Registro de Data e Hora)	\$TI-Day (Número Inteiro com rótulos)	\$TI-Hour (Número Inteiro com rótulos)
1	Jan 5 2005 8:00	3 (Quarta-feira)	8 (8:00)
2	Jan 5 2005 9:00	3 (Quarta-feira)	9 (9:00)
3	Jan 5 2005 10:00	3 (Quarta-feira)	10 (10:00)
4	Jan 5 2005 11:00	3 (Quarta-feira)	11 (11:00)
5	Jan 5 2005 12:00	3 (Quarta-feira)	12 (12:00)

Horas (Não Periódicas)

Escolha esta opção se tiver medições por hora que não se ajustam em um ciclo diário regular. É possível especificar o horário de início para rotular registros consecutivos ou selecionar **Construir a partir dos dados** para escolher um campo de registro de data e hora ou de hora que indica o horário para cada registro.

As horas são baseadas em um relógio de 24 horas (13h00 = 1:00 pm) e não quebra (hora 25 após a hora 24).

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 42. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Sequência de Caracteres)	\$TI-Hour (Número Inteiro com rótulos)
1	8:00	8 (8:00)
2	9:00	9 (9:00)
3	10:00	10 (10:00)
4	11:00	11 (11:00)
5	12:00	12 (12:00)

Minutos por Dia

Para as medições por minuto que caem em um ciclo diário, é possível especificar o número de dias por semana, o dia em que a semana começa, o número de horas no dia e o horário em que o dia começa. As horas são especificadas com base em um relógio de 24 horas e podem ser especificadas para minuto e segundo utilizando dois pontos (por exemplo, 2:05:17 p.m. = 14h05min17s). Também é possível especificar o número de minutos para incrementar (a cada minuto, a cada dois minutos, e assim por diante, em que o incremento deve ser um valor múltiplo de 60).

É possível especificar a data e hora de início para rotular registros consecutivos ou selecionar **Construir a partir de dados** para escolher um campo de registro de data e hora que identifica a data e hora para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 43. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Registro de Data e Hora)	\$TI-Minute
1	2005-01-05 08:00:00	0
2	2005-01-05 08:01:00	1
3	2005-01-05 08:02:00	2
4	2005-01-05 08:03:00	3
5	2005-01-05 08:04:00	4

Minutos (Não Periódicos)

Escolha esta opção se tiver medições por minuto que não se ajustam em um ciclo diário regular. É possível especificar o número de minutos para incrementar (a cada minuto, a cada dois minutos, e assim por diante, em que o valor especificado deve ser um número múltiplo de 60).

É possível especificar o horário de início para rotular registros consecutivos ou selecionar **Construir a partir dos dados** para escolher um campo de registro de data e hora ou de hora que identifica o horário para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 44. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Sequência de Caracteres)	\$TI-Minute
1	8:00	0
2	8:01	1
3	8:02	2
4	8:03	3
5	8:04	4

- A sequência de caracteres *TimeLabel* é criada utilizando dois pontos entre a hora e o minuto. A hora não é agrupada-hora 25 após a hora 24.
- Os minutos incrementam pelo valor especificado na caixa de diálogo. Por exemplo, se o incremento for 2, o *TimeLabel* será 8:00, 8:02 e assim por diante, e os minutos serão 0, 2, e assim por diante.

Segundos por Dia

Para os intervalos de segundos que caem em um ciclo diário, é possível especificar o número de dias por semana, o dia em que a semana começa, o número de horas no dia e o horário em que o dia começa. As horas são especificadas com base em um relógio de 24 horas e podem ser especificadas para minuto e segundo utilizando dois pontos (por exemplo, 2:05:17 p.m. = 14h05min17s). Também é possível especificar o número de segundos para incrementar (a cada segundo, a cada dois segundos, e assim por diante, em que o valor especificado deve ser um número múltiplo de 60).

É possível especificar a data e hora para iniciar a rotulagem de registros consecutivos ou selecionar **Construir a partir de dados** para escolher um campo de registro de data e hora que identifica a data e hora para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 45. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Registro de Data e Hora)	\$TI-Minute	\$TI-Second
1	2005-01-05 08:00:00	0	0
2	2005-01-05 08:00:01	0	1
3	2005-01-05 08:00:02	0	2
4	2005-01-05 08:00:03	0	3
5	2005-01-05 08:00:04	0	4

Segundos (Não Periódicos)

Escolha esta opção se tiver medições feitas por segundo que não se ajustam em um ciclo diário regular. É possível especificar o número de segundos para incrementar (a cada segundo, a cada dois segundos, e assim por diante, em que o valor especificado deve ser um número múltiplo de 60).

Especifique o horário de início da rotulagem de registros consecutivos ou selecione **Construir a partir dos dados** para escolher um campo de registro de data e hora ou de hora que identifica o horário para cada registro.

Saída da amostra

Novos campos são criados conforme mostrado na tabela a seguir.

Tabela 46. Exemplos de criação de novo campo

\$TI-TimeIndex (Integer)	\$TI-TimeLabel (Sequência de Caracteres)	\$TI-Minute	\$TI-Second
1	8:00:00	0	0
2	8:00:01	0	1
3	8:00:02	0	2
4	8:00:03	0	3
5	8:00:04	0	4

- A sequência de caracteres *TimeLabel* é criada utilizando dois pontos entre a hora e o minuto e entre o minuto e o segundo. A hora não é agrupada-a hora após 24 será a hora 25.
- Os segundos são incrementados por qualquer número que for especificado como o incremento. Se o incremento for 2, o *TimeLabel* será 8:00:00, 8:00:02, e assim por diante, e os segundos serão 0, 2, e assim por diante.

Nó Histórico

Os nós Históricos são mais frequentemente utilizados para dados sequenciais, como dados de séries temporais. Eles são utilizados para criar novos campos contendo dados de campos em registros anteriores. Ao utilizar um nó Histórico, talvez você queira ter dados pré-ordenados por um determinado campo. É possível utilizar um nó Ordem para fazer isso.

Configurando Opções para o Nó do Histórico

Campos selecionados. Utilizando o Seletor de Campo (botão à direita da caixa de texto), selecione os campos para os quais você deseja um histórico. Cada campo selecionado é utilizado para criar novos campos para todos os registros no conjunto de dados.

Offset. Especifique o registro mais recente anterior ao registro atual a partir do qual você deseja extrair os valores de campo de histórico. Por exemplo, se o Offset for configurado para 3, conforme cada registro passa pelo nó, os valores de campo para o terceiro registro anterior serão incluídos no registro atual. Utilize as configurações de Span para especificar a partir de que ponto anterior os registros serão extraídos. Utilize as setas para ajustar o valor de offset.

Span. Especifique a partir de quantos registros anteriores você deseja extrair os valores. Por exemplo, se o Offset for configurado como 3 e o Span for configurado como 5, cada registro que passar pelo nó terá cinco campos incluídos nele para cada campo especificado na lista Campos Selecionados. Isso significa que quando o nó estiver processando o registro 10, os campos serão incluídos do registro 7 ao registro 3. Utilize as setas para ajustar o valor do span.

Onde o histórico está indisponível. Selecione uma das opções a seguir para manipular registros que não possuem valores de histórico. Isso geralmente se refere aos vários primeiros registros na parte superior do conjunto de dados para os quais não há registros anteriores para uso como um histórico.

- **Descartar registros.** Selecione para descartar registros em que nenhum valor histórico está disponível para o campo selecionado.
- **Deixar histórico indefinido.** Selecione para manter registros em que nenhum valor histórico está disponível. O campo de histórico será preenchido com um valor indefinido, exibido como `$null$`.
- **Preencher valores com.** Especifique um valor ou uma sequência de caracteres a ser utilizada para registros em que nenhum valor histórico está disponível. O valor de substituição padrão é `undef`, o sistema nulo. Os valores nulos são exibidos utilizando a sequência de caracteres `$null$`.

Ao selecionar um valor de substituição, lembre-se das seguintes regras para que uma execução apropriada ocorra:

- Campos selecionados devem ser do mesmo tipo de armazenamento.
- Se todos os campos selecionados tiverem armazenamento numérico, o valor de substituição deverá ser analisado como um número inteiro.
- Se todos os campos selecionados tiverem armazenamento real, o valor de substituição deverá ser analisado como um real.
- Se todos os campos selecionados tiverem armazenamento simbólico, o valor de substituição deverá ser analisado como uma sequência de caracteres.
- Se todos os campos selecionados tiverem armazenamento de data/hora, o valor de substituição deverá ser analisado como um campo de data/hora.

Se alguma das condições acima não for atendida, você receberá um erro quando executar o nó Histórico.

Nó Reordenação de Campo

O nó Reordenação de Campo permite definir a ordem natural utilizada para exibir campos de recebimento de dados. Esta ordem afeta a exibição de campos em uma variedade de locais, como tabelas, listas e o Seletor de Campo. Esta operação é útil, por exemplo, ao trabalhar com conjuntos de dados grandes para tornar os campos de interesse mais visíveis.

Configurando Opções de Reordenação de Campo

Há duas maneiras de reordenar campos: ordenação customizada e ordenação automática

Ordenação Customizada

Selecione **Ordem Customizada** para ativar uma tabela de nomes de campos e de tipos na qual é possível visualizar todos os campos e utilizar os botões de seta para criar uma ordem customizada.

Para reordenar campos:

1. Selecione um campo na tabela. Utilize o método Ctrl-clique para selecionar diversos campos.
2. Utilize os botões de seta simples para mover um ou mais campos uma linha para cima ou para baixo.
3. Utilize os botões de seta de linha para mover um ou mais campos para a parte inferior ou superior da lista.
4. Especifique a ordem de campos não incluídos aqui ao mover para cima ou para baixo a linha divisória, indicada como [outros campos].

Mais informações sobre [outros campos]

Outros campos. O propósito da linha divisória [outros campos] é dividir a tabela em duas partes.

- Os campos que aparecerem acima da linha divisória serão ordenados (conforme eles aparecem na tabela) na parte superior de todas as ordens naturais usadas para exibir os campos abaixo desse nó.
- Os campos que aparecerem abaixo da linha divisória serão ordenados (conforme eles aparecem na tabela) na parte inferior de todas as ordens naturais usadas para exibir os campos abaixo desse nó.

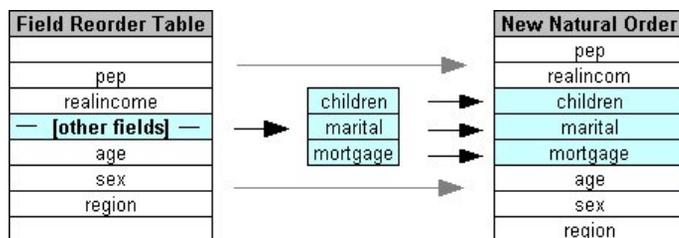


Figura 8. Diagrama ilustrando como “outros campos” são incorporados na nova ordem de campo

- Todos os demais campos que não aparecerem na tabela de reordenação de campo aparecerão entre esses campos “superiores” e “inferiores”, conforme indicado pela posição da linha divisória.

As opções de ordenação customizadas adicionais incluem:

- Classifique os campos em ordem ascendente ou descendente ao clicar nas setas acima de cada cabeçalho da coluna (**Tipo**, **Nome** e **Armazenamento**). Ao ordenar por coluna, os campos não especificados aqui (indicados pela linha [outros campos]) são ordenados por último em sua ordem natural.
- Clique em **Limpar Não Utilizados** para excluir todos os campos não utilizados do nó Reordenação de Campo. Os campos não utilizados são exibidos na tabela com fonte vermelha. Isso indica que o campo foi excluído em operações de envio de dados.
- Especifique a ordenação para quaisquer novos campos (exibidos com um ícone de raio para indicar um campo novo ou não especificado). Ao clicar em **OK** ou **Aplicar**, o ícone desaparece.

Note: se os campos forem incluídos acima após uma ordem customizada ter sido aplicada, os novos campos serão anexados na parte inferior da lista customizada.

Ordenação Automática

Selecione **Ordenação Automática** para especificar um parâmetro para ordenação. As opções da caixa de diálogo se alteram dinamicamente para fornecer opções para ordenação automática.

Ordenar Por. Selecione uma das três maneiras de ordenar campos lidos no nó Reordenação. Os botões de seta indicam se a ordem será crescente ou decrescente. Selecione um para fazer uma mudança.

- Nome
- Tipo
- Armazenamento

Os campos incluídos antes do nó Reordenação de Campo após a ordenação automática ter sido aplicada serão colocados automaticamente em sua posição correta com base no tipo de ordem selecionado.

Nó Intervalo de tempo do AS

O nó original Intervalos de Tempo no SPSS Modeler não é compatível com o Analytic Server (AS). O nó Intervalos de Tempo do AS (novo na liberação SPSS Modeler 17.0) contém um subconjunto das funções do nó Intervalos de Tempo que podem ser usados com o Analytic Server.

Use o nó Intervalos de Tempo do AS para especificar intervalos e derivar um novo campo de tempo para estimar ou prever. Uma variedade completa de intervalos de tempo é suportada, desde segundos até anos.

Use o nó para derivar um novo campo de tempo; o novo campo terá o mesmo tipo de armazenamento que o campo de tempo de entrada que você escolheu. O nó gera os seguintes itens:

- O campo especificado na guia Campos como o **Campo de Tempo**, juntamente com o prefixo/sufixo escolhido. Por padrão, o prefixo é \$TI_.
- Os campos especificados na guia Campos, como os **Campos de dimensão**.
- Os campos especificados na guia Campos, como os **Campos para agregar**.

Vários campos extras também podem ser gerados, dependendo do intervalo ou período selecionado (como o minuto ou segundo dentro do qual uma medida está).

Intervalo de tempo do AS - opções de campo

Use a guia Campos no nó Intervalos de Tempo do AS para selecionar os dados a partir dos quais o novo intervalo de tempo é derivado.

Fields Exibe todos os campos de entrada para o nó com seus ícones de tipo de medição. Todos os campos de tempo têm o tipo de medida 'contínuo'. Selecione o campo a ser usado como a entrada.

Campo de Tempo Mostra o campo de entrada a partir do qual o novo intervalo de tempo é derivado; apenas um único campo contínuo é permitido. O campo é usado pelo nó Intervalos de Tempo como a chave de agregação para converter o intervalo. O novo campo possui o mesmo tipo de armazenamento que o campo de tempo de entrada escolhido. Se você selecionar um campo de número inteiro, ele será considerado como um índice de tempo.

Campos de dimensões Opcionalmente, é possível incluir campos aqui para criar séries temporais individuais baseadas nos valores do campo. Como um exemplo, com dados geoespaciais, é possível usar um campo de ponto como uma dimensão. Neste exemplo, a saída de dados do nó Intervalos de Tempo do AS é classificado nas séries temporais para cada valor de ponto no campo de ponto.

Campos para agregar Selecione os campos a serem agregadas como parte de alterar o período do campo de tempo. Apenas os campos que você seleciona aqui estão disponíveis na guia Compilação para a guia **Configurações customizadas para campos especificados**. Qualquer campo não incluído é filtrado dos dados que saem do nó. Isso significa que todos os campos restantes na lista **Campos** são filtrados dos dados.

Intervalo de tempo do AS - opções de criação

Use a guia Compilação para especificar opções para alterar o intervalo de tempo e como os campos nos dados são agregados, com base em seu tipo de medição.

Ao agregar dados, quaisquer campos de data, hora ou registro de data e hora existentes são substituídos pelos campos gerados e são eliminados da saída. Outros campos são agregados com base nas opções especificadas nessa guia.

Intervalo de tempo Selecione o intervalo e a periodicidade para construção das séries. Para obter mais informações, consulte "Intervalos suportados" na página 180.

Configurações padrão Selecione a agregação padrão a ser aplicada aos dados de diferentes tipos. O padrão é aplicado com base no nível de medição; por exemplo, campos contínuos são agregados pela soma, enquanto os campos nominais usam o modo. É possível configurar o padrão para 3 diferentes níveis de medição:

- **Contínuo** As funções disponíveis para os campos contínuos incluem **Soma, Média, Mín, Máx, Mediana, 1º Quartil e 3º Quartil**.
- **Nominal** As opções incluem **Modo, Mín e Máx**.
- **sinalização** As opções são **Verdadeiro se houver algum verdadeiro** ou **Falso se houver algum falso**.

Configurações customizadas para campos especificados É possível especificar exceções para as configurações padrão de agregação para campos individuais. Use os ícones à direita para incluir ou

remover campos da tabela, ou clique na célula na coluna apropriada para alterar a função de agregação para esse campo. Os campos sem tipo são excluídos da lista e não podem ser incluídos na tabela.

Nova extensão do nome do campo Especifique o **Prefixo** ou **Sufixo** aplicado a todos os campos gerados pelo nó.

Nó Reprojecção

Com dados geoespaciais ou de mapa, duas das maneiras mais comuns que são utilizadas para identificar as coordenadas são os sistemas de coordenadas projetadas e de coordenadas geográficas. No IBM SPSS Modeler, itens como as funções espaciais do Construtor de Expressões, o Nó Spatio-Temporal Prediction (STP) e o Nó Visualização de Mapa usam o sistema de coordenadas projetadas, portanto, todos os dados que você importar que estiverem registrados com um sistema de coordenadas geográficas deverão ser reprojitados. Sempre que possível, os campos geoespaciais (quaisquer campos com um nível de medição geoespacial) são automaticamente reprojitados quando utilizados (não quando eles são importados). Quando todos os campos não puderem ser reprojitados automaticamente, use o nó Reprojecção para alterar o sistema de coordenadas. Reprojetar dessa forma significa que é possível corrigir situações em que um erro ocorre devido ao uso de um sistema de coordenadas incorreto.

Situações de exemplo em que poderá ser necessário reprojitar para alterar o sistema de coordenadas são mostradas na lista a seguir:

- *Anexar* Se tentar anexar dois conjuntos de dados com diferentes sistemas de coordenadas para um campo geoespacial, o SPSS Modeler exibirá a seguinte mensagem de erro: Os sistemas de coordenadas de <Field1> e <Field2> não são compatíveis. Reprojete um ou os dois campos para o mesmo sistema de coordenadas.
<Field1> e <Field2> são os nomes dos campos geoespaciais que causaram o erro.
- *Expressão if/else* Se utilizar uma expressão que contém uma instrução if/else com campos geoespaciais ou tipos de retorno em ambas as partes da expressão, mas com diferentes sistemas de coordenadas, o SPSS Modeler exibe a seguinte mensagem de erro: A expressão condicional contém sistemas de coordenadas incompatíveis: <arg1> e <arg2>.
<arg1> e <arg2> são os outros argumentos then ou else que retornam um tipo geoespacial com diferentes sistemas de coordenadas.
- *Construindo uma lista de campos geoespaciais* Para criar um campo de lista que consiste em vários campos geoespaciais, todos os argumentos do campo geoespacial que são fornecidos para a expressão da lista devem estar no mesmo sistema de coordenadas. Se não estiverem, então a seguinte mensagem de erro será exibida: Os sistemas de coordenadas <Field1> e <Field2> não são compatíveis. Reprojete um ou os dois campos para o mesmo sistema de coordenadas.

Para obter mais informações sobre sistemas de coordenadas, consulte o tópico Configurando Opções Geoespaciais para Fluxos na seção Trabalhando com Fluxos do Guia do Usuário do SPSS Modeler.

Configurando opções para o nó Reprojetar Campos

Campos geoespaciais

Por padrão, essa lista está vazia. É possível mover campos geoespaciais para essa lista a partir da lista **Campos a serem reprojitados** para assegurar que esses campos não sejam reprojitados.

Campos a serem reprojitados

Por padrão, essa lista contém todos os campos geoespaciais que forem a entrada para este nó. Todos os campos nesta lista são reprojitados para o sistema de coordenadas que você configurar na área **Sistema de Coordenadas**.

Sistema de Coordenadas

Padrão de fluxo

Selecione esta opção para utilizar o sistema de coordenadas padrão.

Especificar

Se selecionar esta opção, será possível usar o botão **Alterar** para exibir a caixa de diálogo Selecionar Sistema de Coordenadas e escolher o sistema de coordenadas a ser utilizado para reprojeção.

Para obter mais informações sobre sistemas de coordenadas, consulte o tópico Configurando Opções Geoespaciais para Fluxos na seção Trabalhando com Fluxos do Guia do Usuário do SPSS Modeler.

Capítulo 5. Nós de Gráfico

Variáveis de Nós de Gráfico Comuns

Várias fases do processo de mineração de dados utilizam gráficos e diagramas para explorar os dados introduzidos no IBM SPSS Modeler. Por exemplo, é possível conectar um nó Gráfico ou Distribuição a uma origem de dados para obter um insight dos tipos de dados e das distribuições. Em seguida, é possível executar manipulações de registro e de campo para preparar os dados para as operações de modelagem de recebimento de dados. Outro uso comum de gráficos é verificar a distribuição e os relacionamentos entre campos derivados recentemente.

A paleta Gráficos contém os seguintes nós:



O nó Elemento do Gráfico oferece muitos tipos diferentes de gráficos em um único nó. Utilizando esse nó, é possível escolher os campos de dados que deseja explorar e, em seguida, selecionar um gráfico a partir dos disponíveis para os dados selecionados. O nó filtra automaticamente todos os tipos de gráficos que não funcionariam com as opções de campo.



O nó Gráfico mostra o relacionamento entre os campos numéricos. É possível criar um gráfico utilizando pontos (gráfico de dispersão) ou linhas.



O nó Distribuição mostra a ocorrência de valores simbólicos (categóricos), como tipo ou gênero da hipoteca. O nó Distribuição pode ser usado geralmente para mostrar desbalanceamentos nos dados, que poderão então ser corrigidos utilizando um nó Balanceamento antes de criar um modelo.



O nó Histograma mostra a ocorrência de valores para campos numéricos. Ele é normalmente utilizado para explorar os dados antes de manipulações e construções de modelo. Semelhante ao nó Distribuição, o nó Histograma revela frequentemente desequilíbrios nos dados.



O nó de Coleção mostra a distribuição de valores para um campo numérico com relação aos valores de outro campo. (Ele cria gráficos semelhantes a histogramas). Ele é útil para ilustrar uma variável ou campo cujos valores se alteram ao longo do tempo. Usando um gráfico 3D, também é possível incluir um eixo simbólico exibindo distribuições por categoria.



O nó Multigráficos cria uma representação que exhibe diversos campos Y em um único campo X. Os campos Y são representados como linhas coloridas, em que cada linha é equivalente a um nó Gráfico com Estilo configurado para **Linha** e o Modo X configurado para **Classificar**. Os multigráficos são úteis quando deseja explorar a flutuação de diversas variáveis ao longo do tempo.



O nó Web ilustra a intensidade do relacionamento entre os valores de dois ou mais campos simbólicos (categóricos). O gráfico utiliza linhas de várias larguras para indicar a intensidade da conexão. O nó Web pode ser usado, por exemplo, para explorar o relacionamento entre a compra de um conjunto de itens em um site de e-commerce.



O nó Gráfico de Tempo exibe um ou mais conjuntos de dados de séries temporais. Geralmente, um nó Intervalos de Tempo deverá ser usado primeiro para criar um campo *TimeLabel* que será utilizado para rotular o eixo *x*.



O nó Avaliação ajuda a avaliar e comparar modelos preditivos. O gráfico de avaliação mostra quão bem os modelos preveem resultados específicos. Ele classifica os registros com base no valor previsto e na confiança da previsão. Ele divide os registros em grupos de tamanhos iguais (**quantis**) e, em seguida, representa o valor do critério de negócios para cada quantil do mais alto para o mais baixo. Diversos modelos são mostrados como linhas separadas na representação.



O nó Visualização de Mapa pode aceitar diversas conexões de entrada e exibir dados geoespaciais em um mapa como uma série de camadas. Cada camada é um campo geoespacial único, por exemplo, a camada de base pode ser um mapa de um país e, acima dele, você pode ter uma camada para estradas, uma camada para rios e uma camada para as cidades.

Quando tiver incluído um nó de gráfico em um fluxo, será possível dar um clique duplo no nó para abrir uma caixa de diálogo para especificar opções. A maioria dos gráficos contém um número de opções exclusivas apresentadas em uma ou mais guias. Há também várias opções de guias comuns para todos os gráficos. As seções a seguir contêm mais informações sobre estas opções comuns.

Quando tiver configurado as opções para um nó de gráfico, será possível executá-lo a partir da caixa de diálogo ou como parte de um fluxo. Na janela do gráfico gerado, será possível gerar os nós Derivar (Conjunto e Sinalização) e Seleção com base em uma seleção ou região de dados, efetivamente "subconfigurando" os dados. Por exemplo, é possível utilizar esse recurso poderoso para identificar e excluir valores discrepantes.

Estética, Sobreposições, Painéis e Animação

Sobreposições e Estética

A estética (e sobreposições) inclui dimensionalidade em uma visualização. O efeito de uma estética (agrupamento, armazenamento em cluster ou empilhamento) depende do tipo de visualização, do tipo de campo (variável) e do tipo de gráfico do elemento e estatística. Por exemplo, um campo categórico para cor pode ser utilizado para pontos de grupo em um gráfico de dispersão ou para criar as pilhas em um gráfico de barras empilhadas. Um intervalo numérico contínuo para cor também pode ser utilizado para indicar os valores do intervalo para cada ponto em um gráfico de dispersão.

Deve-se experimentar as estéticas e as sobreposições para localizar aquela que atenda as suas necessidades. As descrições a seguir podem ajudar a selecionar a correta.

Nota: Nem todas as estéticas ou sobreposições estão disponíveis para todos os tipos de visualização.

- **Cor.** Quando a cor é definida por um campo categórico, ele divide a visualização com base nas categorias individuais, uma cor para cada categoria. Quando a cor é um intervalo numérico contínuo, ele varia a cor com base no valor do campo intervalo. Se o elemento gráfico (por exemplo, uma barra ou caixa) representar mais de um registro/caso e um campo de intervalo for utilizado para cor, a cor variará com base na *média* do campo de intervalo.
- **Forma.** A forma é definida por um campo categórico que divide a visualização em elementos de formas diferentes, uma para cada categoria.
- **Transparência.** Quando a transparência é definida por um campo categórico, ele divide a visualização com base nas categorias individuais, um nível de transparência para cada categoria. Quando a transparência é um intervalo numérico contínuo, ele varia a transparência com base no valor do campo

intervalo. Se o elemento gráfico (por exemplo, uma barra ou caixa) representar mais de um registro/caso e um campo de intervalo for utilizado para a transparência, a cor varia com base na *média* do campo de intervalo. No maior valor, os elementos gráficos são totalmente transparentes. E no menor valor, eles são totalmente opacos.

- **Rótulo de Dados.** Os rótulos de dados são definidos por qualquer tipo de campo cujos valores são usados para criar rótulos que são anexados aos elementos gráficos.
- **Tamanho.** Quando o tamanho é definido por um campo categórico, ele divide a visualização com base nas categorias individuais, um tamanho para cada categoria. Quando o tamanho é um intervalo numérico contínuo, ele varia o tamanho com base no valor do campo intervalo. Se o elemento gráfico (por exemplo, uma barra ou caixa) representar mais de um registro/caso e um campo de intervalo for utilizado para tamanho, o tamanho variará com base na *média* do campo de intervalo.

Painel e Animação

Painel. Painel, também conhecido como faceta, cria uma tabela de gráficos. Um gráfico é gerado para cada categoria nos campos de painel e todos os painéis aparecem simultaneamente. Painel é útil para verificar se a visualização está sujeita às condições dos campos de painel. Por exemplo, é possível agrupar em painel um histograma por sexo para determinar se as distribuições de frequência são iguais entre homens e mulheres. Ou seja, é possível verificar se o salário está sujeito a diferenças por sexo. Selecione um campo categórico para painel.

Animação. A animação é semelhante a painel em que diversos gráficos são criados a partir dos valores do campo de animação, mas esses gráficos não são mostrados juntos. Ao invés disso, use os controles no modo Explorar para animar a saída e inverter em uma sequência de gráficos individuais. Além disso, ao contrário de painel, a animação não requer um campo categórico. É possível especificar um campo contínuo cujos valores são divididos em intervalos automaticamente. É possível variar o tamanho do intervalo com os controles de animação em modo de exploração. Nem todas as visualizações oferecem animação.

Utilizando a Guia Saída

Para todos os tipos de gráficos, é possível especificar as opções a seguir para o nome do arquivo e exibição de gráficos gerados.

Nota: os gráficos de nó de distribuição possuem configurações adicionais.

Nome de saída. Especifica o nome do gráfico produzido quando o nó é executado. **Automático** escolhe um nome com base no nó que gera a saída. Como opção, é possível selecionar **Customizado** para especificar um nome diferente.

Saída para tela. Selecione para gerar e exibir o gráfico em uma nova janela.

Saída para arquivo. Selecione para salvar a saída como um arquivo.

- **Gráfico de saída.** Selecione para produzir a saída em um formato de gráfico. Disponível somente em nós de Distribuição.
- **Tabela de saída.** Selecione para produzir a saída em um formato de tabela. Disponível somente em nós de Distribuição.
- **Nome do arquivo.** Especifique um nome utilizado para o gráfico ou tabela gerado. Utilize o botão de reticências (...) para especificar um arquivo e localização específicos.
- **Tipo de arquivo.** Especifique o tipo de arquivo na lista suspensa. Para todos os nós do gráfico, exceto o nó Distribuição com uma opção **Tabela de Saída**, os tipos de arquivo de gráfico disponíveis são os seguintes.
 - Bitmap (.*bmp*)
 - PNG (.*png*)

- Objeto de saída (.cou)
- JPEG (.jpg)
- HTML (.html)
- Documento ViZml (.xml) para uso em outros aplicativos IBM SPSS Statistics.

Para a opção **Tabela de Saída** no nó de Distribuição, os tipos de arquivos disponíveis são os seguintes.

- Dados delimitados por guia (.tab)
- Dados delimitados por vírgula (.csv)
- HTML (.html)
- Objeto de saída (.cou)

Paginar saída. Quando salvar a saída como HTML, esta opção é ativada para permitir o controle do tamanho de cada página HTML. (Aplica-se somente ao nó Distribuição).

Linhas por página. Quando **Paginar saída** é selecionada, esta opção é ativada para permitir determinar o comprimento de cada página HTML. A configuração padrão é 400 linhas. (Aplica-se somente ao nó Distribuição).

Utilizando a Guia Anotações

Utilizada para todos os nós, essa guia oferece opções para renomear nós, fornecer um ToolTip customizado e armazenar uma anotação longa.

Gráficos 3D

Gráficos e gráficos de coleção no IBM SPSS Modeler possuem a capacidade de exibir informações sobre um terceiro eixo. Isso fornece flexibilidade adicional na visualização de dados para selecionar subconjuntos ou derivar novos campos para modelagem.

Ao criar um gráfico 3D, é possível clicar nele e arrastar o mouse para girá-lo e visualizá-lo a partir de qualquer ângulo.

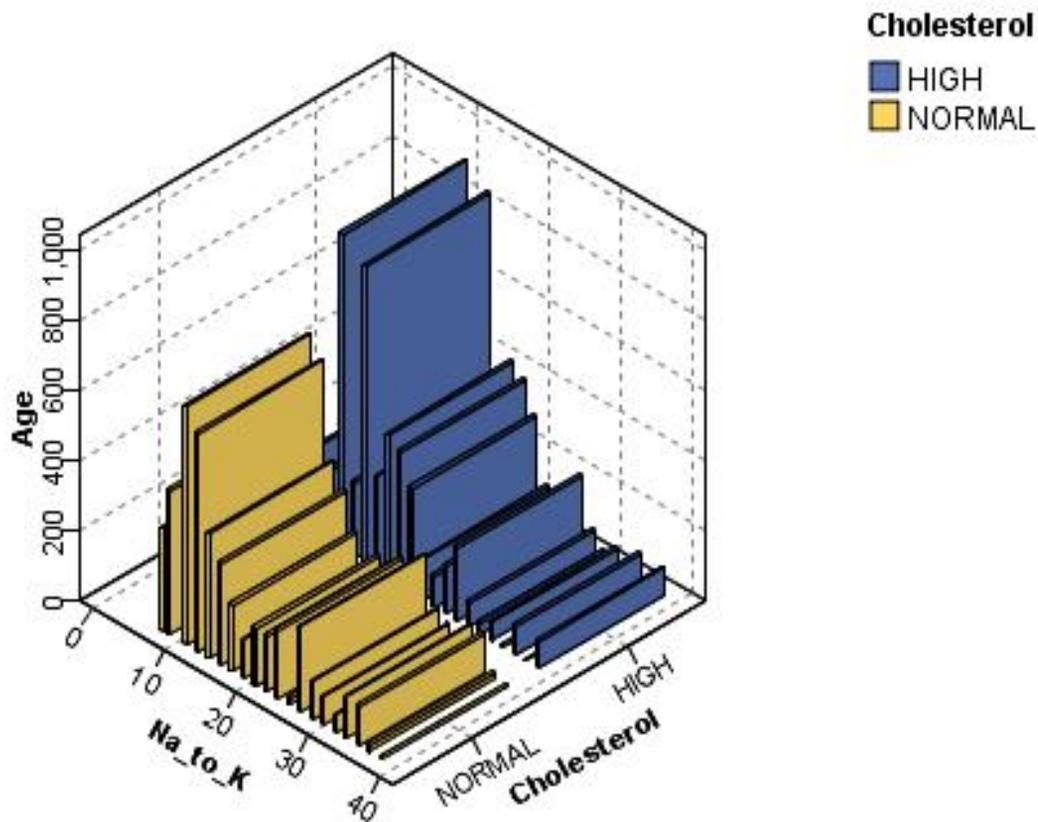


Figura 9. Gráfico de coleção com os eixos x, y e z

Há duas maneiras de criar gráficos 3D no IBM SPSS Modeler: representando informações em um terceiro eixo (gráficos 3D reais) e exibindo gráficos com efeitos 3D. Os dois métodos estão disponíveis para gráficos e coleções.

Para Representar Informações em Terceiro Eixo

1. Na caixa de diálogo do nó de gráfico, clique na guia **Plot**.
2. Clique no botão 3D para ativar opções para o eixo z.
3. Utilize o botão Seletor de Campo para selecionar um campo para o eixo z. Em alguns casos, somente campos simbólicos são permitidos aqui. O Seletor de Campo exibirá os campos apropriados.

Para Incluir Efeitos 3D em um Gráfico

1. Após ter criado um gráfico, clique na guia **Gráfico** na janela de saída.
2. Clique no botão 3D para alternar a visualização para um gráfico tridimensional.

Nó Elemento do Gráfico

O nó Elemento do Gráfico permite escolher a partir de muitas saídas de gráficos diferentes (gráficos de barras, gráficos de pizza, histogramas, gráficos de dispersão, heatmaps, etc.) em um único nó. Ao começar, na primeira guia, escolhendo os campos de dados que deseja explorar, o nó apresenta uma opção de tipos de gráfico que funcionam para seus dados. O nó filtra automaticamente todos os tipos de gráficos que não funcionariam com as opções de campo. É possível definir opções de gráfico detalhadas ou mais avançadas na guia Detalhado.

Nota: o nó Elemento do Gráfico deve ser conectado a um fluxo com dados em outro para editar o nó ou selecionar tipos de gráfico.

Há dois botões que permitem controlar quais modelos de visualização (e folhas de estilo e mapas) estão disponíveis:

Gerenciar. Gerencie modelos de visualização, folhas de estilo e mapas em seu computador. É possível importar, exportar, renomear e excluir modelos de visualização, folhas de estilo e mapas em sua máquina local. Consulte o tópico “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224 para obter mais informações.

Localização. Altere a localização na qual os modelos de visualização, folhas de estilo e mapas são armazenados. A localização atual é listada à direita do botão. Consulte o tópico “Configurando a Localização dos Modelos, Folhas de Estilos e Mapas” na página 223 para obter mais informações.

Guia Básica do Gráfico

Se você não tiver certeza de qual tipo de visualização representaria melhor seus dados, utilize a guia Básico. Ao selecionar seus dados, será exibido um subconjunto de tipos de visualização que são apropriados para os dados. Para obter exemplos, consulte “Exemplos de Gráficos” na página 212.

1. Selecione um ou mais campos (variáveis) na lista. Utilize Ctrl+Clique para selecionar diversos campos.
Observe que o nível de medição do campo determina o tipo de visualizações que está disponível. É possível alterar o nível de medição ao clicar com o botão direito no campo na lista e escolher uma opção. Para obter mais informações sobre os tipos de nível de medição disponíveis, consulte “Tipos de Campo (Variável)” na página 200.
2. Selecione um tipo de visualização. Para obter descrições dos tipos disponíveis, consulte “Tipos de Visualizações de Gráfico Integradas Disponíveis” na página 204.
3. Para determinadas visualizações, é possível escolher uma estatística de sumarização. Diferentes subconjuntos de estatísticas estão disponíveis dependendo se a estatística for baseada em contagem ou calculada a partir de um campo contínuo. As estatísticas disponíveis também dependem do próprio modelo. Uma lista completa de estatísticas que podem estar disponíveis está a seguir no próximo passo.
4. Se desejar definir mais opções, como estética opcional e campos de painel, clique em **Detalhado**. Consulte o tópico “Guia Detalhada do Gráfico” na página 202 para obter mais informações.

Estatísticas de Sumarização Calculadas a partir de um Campo Contínuo

- *Média.* Uma medida de tendência central. A média aritmética, a soma dividida pelo número de casos.
- *Median.* O valor acima e abaixo do qual metade dos casos cai, o 50º percentil. Se houver um número par de casos, a mediana será a média dos dois casos intermediários quando forem classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).
- *Modo.* O valor que ocorre mais frequentemente. Se vários valores compartilharem a maior frequência da ocorrência, cada um deles será um modo.

- *Mínimo*. O menor valor de uma variável numérica.
- *Máximo*. O maior valor de uma variável numérica.
- *Intervalo*. A diferença entre os valores mínimo e máximo.
- *Intervalo Méd.*. O meio do intervalo, ou seja, o valor cuja diferença do mínimo é igual a sua diferença do máximo.
- *Sum*. A soma ou o total dos valores, em todos os casos com valores não omissos.
- *Soma Acumulativa*. A soma acumulativa dos valores. Cada elemento gráfico mostra a soma de um subgrupo mais a soma total de todos os grupos anteriores.
- *Percentual de Soma*. A porcentagem dentro de cada subgrupo com base em um campo somado, comparado com a soma em todos os grupos.
- *Percentual de Soma Acumulativa*. A porcentagem acumulativa dentro de cada subgrupo com base em um campo somado, comparado com a soma em todos os grupos. Cada elemento gráfico mostra a porcentagem de um subgrupo mais a porcentagem total de todos os grupos anteriores.
- *Variância*. Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.
- *Desvio padrão*. Uma medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.
- *Erro padrão*. Uma medida do quanto o valor de uma estatística de teste varia de amostra para amostra. Ela é o desvio padrão da distribuição de amostragem para uma estatística. Por exemplo, o erro padrão da média é o desvio padrão das médias da amostra.
- *Curtose*. Uma medida da extensão até a qual as observações se agrupam em torno de um ponto central. Para a uma distribuição normal, o valor da estatística de curtose é zero. A curtose positiva indica que, com relação a uma distribuição normal, as observações são agrupadas mais ao centro da distribuição e têm rastros mais finos até os valores extremos da distribuição, no ponto em que os rastros da distribuição leptocúrtica são mais espessos com relação a uma distribuição normal. Já a curtose negativa indica que, com relação a uma distribuição normal, as observações são menos agrupadas e possuem rodapés mais espessos até os valores extremos da distribuição, no ponto em que os rastros da distribuição platykúrtic são mais finos com relação a uma distribuição normal.
- *Assimetria*. Uma medida da assimetria de uma distribuição. A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

As seguintes estatísticas da região podem resultar em mais de um elemento gráfico por subgrupo. Ao utilizar os elementos gráficos de intervalo, de área ou de borda, uma estatística de região resulta em um elemento gráfico mostrando o intervalo. Todos os outros elementos gráficos resultam em dois elementos separados, um mostrando o início do intervalo e o outro mostrando o término do intervalo.

- **Região: Intervalo**. O intervalo de valores entre os valores mínimo e máximo.
- **Região: Intervalo de Confiança de 95% da Média**. Um intervalo de valores que tem uma chance de 95% de incluir a média da população.
- **Região: Intervalo de Confiança de 95% de Individual**. Um intervalo de valores que tem uma chance de 95% de incluir o valor predito fornecido ao caso individual.
- **Região: 1 Desvio Padrão acima/abaixo da Média**. Um intervalo de valores entre 1 *desvio padrão* acima e abaixo da *média*.
- **Região: 1 Erro Padrão acima/abaixo da Média**. Um intervalo de valores entre 1 *erro padrão* acima e abaixo da *média*.

- **Contagem.** O número de linhas/casos.
- **Contagem acumulativa.** O número acumulativo de linhas/casos. Cada elemento gráfico mostra a contagem de um subgrupo mais a contagem total de todos os grupos anteriores.
- **Percentual de Contagem.** A porcentagem de linhas/casos em cada subgrupo em comparação com o número total de linhas/casos.
- **Percentual Acumulativo de Contagem.** A porcentagem acumulativa de linhas/casos em cada subgrupo em comparação com o número total de linhas/casos. Cada elemento gráfico mostra a porcentagem de um subgrupo mais a porcentagem total de todos os grupos anteriores.

Tipos de Campo (Variável)

Ícones aparecem ao lado dos campos na lista de campos para indicar o tipo de campo e o tipo de dados. Os ícones também identificam conjuntos de múltiplas respostas.

Tabela 47. Ícones do nível de medição.

Nível de medição	Numérico	Sequência de caracteres	Data	Hora
Contínuo		n/d		
Conjunto classificado				
Set				

Tabela 48. Ícones de conjunto de múltiplas respostas.

Tipo de conjunto de múltiplas respostas	Ícone
Conjunto de múltiplas respostas, diversas categorias	
Conjunto de múltiplas respostas, diversas dicotomias	

Nível de medição

Um nível de medição de campo é importante ao criar uma visualização. A seguir há uma descrição dos níveis de medição. É possível alterar temporariamente o nível de medição ao clicar com o botão direito em um campo na lista e escolher uma opção. Na maioria dos casos, é necessário considerar apenas as duas classificações maiores dos campos, categóricos e contínuos:

Categórico. Dados com um número limitado de valores ou categorias distintas (por exemplo, sexo ou religião). Os campos categóricos podem ser campos de sequência de caracteres (alfanuméricos) ou numéricos que usam códigos numéricos para representar categorias (por exemplo, 0 = *male* e 1 = *female*). Também referidos como dados qualitativos. Conjuntos, conjuntos ordenados e flags são todos campos categóricos.

- **Conjunto.** Um campo/variável cujos valores representam categorias sem nenhum ranqueamento intrínseco (por exemplo, o departamento da empresa no qual um funcionário trabalha). Exemplos de variáveis nominais incluem região, CEP e filiação religiosa. Também conhecido como uma variável nominal.
- **Conjunto Ordenado.** Um campo/variável cujos valores representarem categorias com algum ranqueamento intrínseco (por exemplo, níveis de satisfação de serviço de muito insatisfeito para muito

satisfeito). Exemplos de conjuntos ordenados incluem escores de atitude que representam o grau de satisfação ou de confiança e os escores de classificação de preferência. Também conhecido como uma variável ordinal.

- *sinalização*. Um campo/variável com dois valores distintos, como Sim e Não ou 1 e 2. Também conhecido como uma variável dicotômica ou binária.

Contínuo. Dados medidos em um intervalo ou razão de escala, em que os valores de dados indicam a ordem dos valores e a distância entre os valores. Por exemplo, um salário de \$72.195 é maior que um salário de \$52.398, e a distância entre os dois valores é de \$19.797. Também conhecido como dados quantitativos, de escala ou de intervalo numérico.

Os campos categóricos definem categorias na visualização, geralmente para desenhar elementos gráficos separados ou para agrupar elementos gráficos. Campos contínuos geralmente são resumidos dentro de categorias e de campos categóricos. Por exemplo, uma visualização padrão de renda para categorias como sexo exibe a renda média para homens e a renda média para mulheres. Os valores brutos para os campos contínuos também podem ser representados, como em um gráfico de dispersão. Por exemplo, um gráfico de dispersão pode mostrar o salário atual e o salário inicial para cada caso. Um campo categórico pode ser utilizado para agrupar os casos por sexo.

Tipos de Dados

O nível de medição não é a única propriedade de um campo que determina seu tipo. Um campo também é armazenado como um tipo de dados específico. Os tipos de dados possíveis são sequências de caracteres (dados não numéricos, como letras), valores numéricos (números reais) e datas. Ao contrário do nível de medição, o tipo de dados de um campo não pode ser alterado temporariamente. Deve-se alterar o modo como os dados são armazenados no conjunto de dados original.

Conjuntos de múltiplas respostas

Alguns arquivos de dados suportam um tipo especial de "campo" chamado **conjunto de múltiplas respostas**. Na realidade, os conjuntos de múltiplas respostas não são "campos" no sentido normal. Os conjuntos de múltiplas respostas utilizam vários campos para registrar respostas para questões em que o respondente pode dar mais de uma resposta. Os conjuntos de múltiplas respostas são tratados como campos categóricos, e a maioria das ações que pode ser executada com campos categóricos também pode ser executada com os conjuntos de múltiplas respostas.

Os conjuntos de múltiplas respostas podem ser conjuntos de múltiplas dicotomias ou conjunto de categorias múltiplas.

Conjuntos de múltiplas dicotomias. Um conjunto de múltiplas dicotomias geralmente consiste em diversos campos dicotômicos: campos com apenas dois valores possíveis de um sim/não, presente/ausente, natureza verificada/não verificada. Embora os campos possam não ser estritamente dicotômicos, todos os campos no conjunto são codificados da mesma maneira.

Por exemplo, uma pesquisa de opinião fornece cinco possíveis respostas à questão: "Quais das seguintes fontes você confia para notícias?" O respondente pode indicar diversas opções ao selecionar uma caixa para cada opção. As cinco respostas se tornam cinco campos no arquivo de dados, codificados como 0 para *Não* (não selecionado) e 1 para *Sim* (selecionado).

Conjuntos de categorias múltiplas. Um conjunto de categorias múltiplas consiste em diversos campos, todos codificados da mesma maneira, geralmente com muitas categorias de respostas possíveis. Por exemplo, um item da pesquisa de opinião diz, "Indique até três nacionalidades que melhor descrevam sua herança étnica". Pode haver centenas de respostas possíveis, mas para propósitos de codificação, a lista é limitada a 40 nacionalidades mais comuns, com o restante relegado para uma categoria "outra". No arquivo de dados, as três opções se tornam três campos, cada um com 41 categorias (40 nacionalidades codificadas e uma categoria "outra").

Guia Detalhada do Gráfico

Utilize a guia Detalhada quando souber qual tipo de visualização deseja criar ou quando deseja incluir estética opcional, como painéis e/ou animação, em uma visualização. Para obter exemplos, consulte “Exemplos de Gráficos” na página 212.

1. Se você selecionou um tipo de visualização na guia Básico, ele será exibido. Caso contrário, escolha um na lista suspensa. Para obter informações sobre os tipos de visualização, consulte “Tipos de Visualizações de Gráfico Integradas Disponíveis” na página 204.
2. Logo à direita da imagem miniatura da visualização há controles para especificar os campos (variáveis) necessários para o tipo de visualização. Todos esses campos devem ser especificados.
3. Para determinadas visualizações, é possível selecionar uma estatística de sumarização. Em alguns casos (como com gráficos de barras), é possível utilizar uma dessas opções de sumarização para a estética de transparência. Para obter descrições das estatísticas de sumarização, consulte “Guia Básica do Gráfico” na página 198.
4. É possível selecionar uma ou mais das estéticas opcionais. Elas podem incluir dimensionalidade ao permitir incluir outros campos na visualização. Por exemplo, é possível utilizar um campo para variar o tamanho dos pontos de um gráfico de dispersão. Para obter mais informações sobre estética opcional, consulte “Estética, Sobreposições, Painéis e Animação” na página 194. Observe que a estética de transparência não é suportada em script.
5. Se estiver criando uma visualização de mapa, o grupo **Arquivos de Mapeamento** mostrará o arquivo ou arquivos de mapeamento que serão utilizados. Se houver um arquivo de mapeamento padrão, esse arquivo será exibido. Para alterar o arquivo de mapeamento, clique em **Selecionar um Arquivo de Mapeamento** para exibir a caixa de diálogo Selecionar Mapas. Também é possível especificar o arquivo de mapeamento padrão nessa caixa de diálogo. Consulte o tópico “Selecionando Arquivos de Mapeamento para Visualizações de Mapa” para obter mais informações.
6. É possível selecionar uma ou mais das opções de painel ou de animação. Para obter mais informações sobre as opções de painel e de animação, consulte “Estética, Sobreposições, Painéis e Animação” na página 194.

Selecionando Arquivos de Mapeamento para Visualizações de Mapa

Se você selecionar um modelo de visualização de mapa, será necessário um arquivo de mapeamento que define as informações geográficas para desenhar o mapa. Se houver um arquivo de mapeamento padrão, ele será utilizado para a visualização do mapa. Para escolher um arquivo de mapeamento diferente, clique em **Selecionar um Arquivo de Mapeamento** na guia Detalhada para exibir a caixa de diálogo Selecionar Mapas.

A caixa de diálogo Selecionar Mapas permite escolher um arquivo de mapeamento primário e um arquivo de mapeamento de referência. Os arquivos de mapeamento definem as informações geográficas para desenhar o mapa. Seu aplicativo é instalado com um conjunto de arquivos de mapeamento padrão. Se houver outros arquivos de forma ESRI que deseja utilizar, primeiro é necessário converter os arquivos de forma em arquivos SMZ. Consulte o tópico “Convertendo e Distribuindo Arquivos de Forma de Mapa” na página 225 para obter mais informações. Após converter o mapa, clique em **Gerenciar...** na caixa de diálogo Seletor de Modelo para importar o mapa no sistema Gerenciar para que ele fique disponível na caixa de diálogo Selecionar Mapas.

A seguir estão alguns pontos a serem considerados ao especificar arquivos de mapeamento:

- Todos os modelos de mapa requerem pelo menos um arquivo de mapeamento.
- O arquivo de mapeamento normalmente vincula um atributo de chave de mapa a uma chave de dados.
- Se o modelo não requerer uma chave de mapa que é vinculada a uma chave de dados, ele requererá um arquivo de mapa de referência e campos que especificam coordenadas (como longitude e latitude) para desenhar os elementos no mapa de referência.

- Os modelos de mapa de sobreposição requerem dois mapas: um arquivo de mapeamento primário e um arquivo de mapeamento de referência. O mapa de referência é desenhado primeiro para que ele fique atrás do arquivo de mapeamento principal.

Para obter informações sobre a terminologia de mapa como atributos e variáveis, consulte o “Conceitos Chave para Mapas” na página 226.

Arquivo de Mapeamento. É possível selecionar qualquer arquivo de mapeamento que estiver no sistema Gerenciar. Esses incluem arquivos de mapeamentos pré-instalados e arquivos de mapeamento que forem importados. Para obter mais informações sobre como gerenciar arquivos de mapeamento, consulte “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224.

Chave de Mapa. Especifique o atributo que deseja utilizar como a chave que vincula o arquivo de mapeamento à chave de dados.

Salvar este arquivo de mapeamento e as configurações como o padrão. Marque esta caixa de seleção se quiser utilizar o arquivo de mapeamento selecionado como o padrão. Se você tiver especificado um arquivo de mapeamento padrão, não será necessário especificar um arquivo de mapeamento toda vez que criar uma visualização de mapa.

Chave de Dados. Esse controle lista o mesmo valor que aparece na guia Seletor de Modelo Detalhado. Ele é fornecido aqui para conveniência se precisar alterar a chave devido ao arquivo de mapeamento específico que escolher.

Exibir todas as variáveis de mapa na visualização. Quando essa opção é selecionada, todas as variáveis no mapa são renderizadas na visualização, mesmo se não houver nenhum valor de chave de dados correspondente. Se desejar ver apenas as variáveis para as quais você tiver dados, desmarque essa opção. As variáveis identificadas pelas chaves de mapa mostradas na lista **Chaves de Mapa Não Correspondidas** não serão renderizadas na visualização.

Comparar Valores de Mapa e de Dados. A chave de mapa e a chave de dados vinculam-se para criar a visualização do mapa. A chave de mapa e a chave de dados devem desenharem a partir do mesmo domínio (por exemplo, países e regiões). Clique em **Comparar** para testar se os valores de chave de dados e de chave de mapa correspondem. O ícone exibido informa o estado da comparação. Esses ícones são descritos abaixo. Se uma comparação tiver sido feita e se houver valores de chave de dados sem quaisquer valores de chave de mapa correspondentes, os valores de chave de dados aparecerão na lista **Chaves de Dados Não correspondidas**. Na lista **Chaves de Mapa Não Correspondidas**, também é possível ver quais valores de chave de mapa não têm valores de chave de dados correspondentes. Se **Exibir todas as variáveis de mapa na visualização** não for marcada, as variáveis identificadas por esses valores de chave de mapa não serão renderizadas.

Tabela 49. Ícones de Comparação.

Ícone	Descrição
	Nenhuma comparação foi executada. Este é o estado padrão antes de clicar em Comparar . Deve-se continuar com cuidado, já que você não sabe se os valores de chave de dados e de chave de mapa correspondem.
	Uma comparação foi executada e os valores de chave de dados e de chave de mapa correspondem completamente. Para cada valor na chave de dados, há uma variável correspondente identificada pela chave de mapa.

Tabela 49. Ícones de Comparação (continuação).

Ícone	Descrição
	Uma comparação foi executada, e alguns valores de chave de dados e de chave de mapa não correspondem. Para alguns valores de chave de dados, não há nenhuma variável correspondente identificada pela chave de mapa. Deve-se continuar com cuidado. Se continuar, a visualização do mapa não incluirá todos os valores de dados.
	Uma comparação foi executada, e nenhum valor de chave de dados e de chave de mapa corresponde. Deve-se escolher uma chave de dados ou uma chave de mapa diferente porque nenhum mapa será renderizado se você continuar.

Tipos de Visualizações de Gráfico Integradas Disponíveis

É possível criar vários tipos diferentes de visualização. Todos os tipos integrados a seguir estão disponíveis nas guias Básico e Detalhado. Algumas das descrições para os modelos (especialmente os modelos de mapa) identificam os campos (variáveis) especificados na guia Detalhado utilizando **Texto especial**.

Tabela 50. Tipos de gráfico disponíveis.

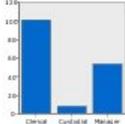
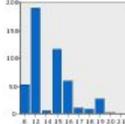
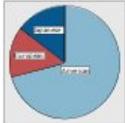
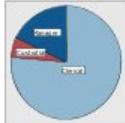
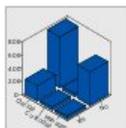
Ícone de gráfico	Descrição	Ícone de gráfico	Descrição
	<p>Bar</p> <p>Calcula uma estatística de sumarização para um campo numérico contínuo e exibe os resultados de cada categoria de um campo categórico como barras.</p> <p><i>Requer:</i> Um campo categórico e um campo contínuo.</p>		<p>Contagem de barras</p> <p>Exibe a proporção de linhas/casos em cada categoria de um campo categórico como barras. Também é possível utilizar o nó do gráfico Distribuição para produzir este gráfico. Esse nó oferece algumas opções adicionais. Consulte o tópico “Nó Distribuição” na página 245 para obter mais informações.</p> <p><i>Requer:</i> um campo categórico único.</p>
	<p>Pizza</p> <p>Calcula a soma de um campo numérico contínuo e exibe a proporção dessa soma distribuída em cada categoria de um campo categórico como fatias de uma pizza.</p> <p><i>Requer:</i> Um campo categórico e um campo contínuo.</p>		<p>Setor de contagens</p> <p>Exibe a proporção de linhas/casos em cada categoria de um campo categórico como fatias de uma pizza.</p> <p><i>Requer:</i> um campo categórico único.</p>

Tabela 50. Tipos de gráfico disponíveis (continuação).

Ícone de gráfico



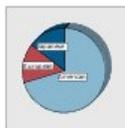
Descrição

Barra 3D

Calcula uma estatística de sumarização para um campo numérico contínuo e exibe os resultados para a passagem das categorias entre dois campos categóricos.

Requer: Um par de campos categóricos e um campo contínuo.

Ícone de gráfico

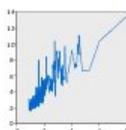


Descrição

Gráfico de setores em 3D

Isso é o mesmo que Pizza, mas com um efeito 3D adicional.

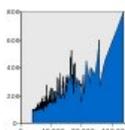
Requer: um campo categórico e um campo contínuo.



Linha

Calcula uma estatística de sumarização para um campo para cada valor de outro campo e desenha uma linha conectando os valores. Também é possível utilizar o nó gráfico Plotagem para produzir um gráfico de plotagem de linha. Esse nó oferece algumas opções adicionais. Consulte o tópico “Nó Gráfico” na página 232 para obter mais informações.

Requer: um par de campos de qualquer tipo.



Área

Calcula uma estatística de sumarização para um campo para cada valor de outro campo e desenha uma área conectando os valores. A diferença entre uma linha e a área é mínima em que uma área se parece com uma linha com o espaço colorido abaixo dela. No entanto, se utilizar uma estética de cor, isto resulta em uma divisão simples da linha e um empilhamento da área.

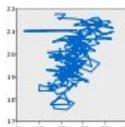
Requer: um par de campos de qualquer tipo.



Área 3D

Exibe os valores de um campo representados com relação aos valores de outro e divididos por um campo categórico. Um elemento de área é desenhado para cada categoria.

Requer: um campo categórico e um par de campos de qualquer tipo.



Caminho

Exibe os valores de um campo representados com relação aos valores de outro campo, com uma linha conectando os valores na ordem em que eles aparecem no conjunto de dados original. A ordem é a principal diferença entre um caminho e uma linha.

Requer: um par de campos de qualquer tipo.

Tabela 50. Tipos de gráfico disponíveis (continuação).

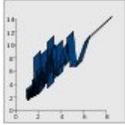
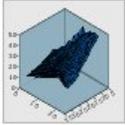
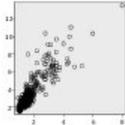
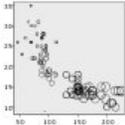
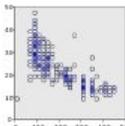
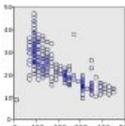
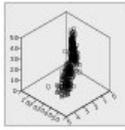
Ícone de gráfico	Descrição	Ícone de gráfico	Descrição
	<p>Faixa de Opções</p> <p>Calcula uma estatística de sumarização para um campo para cada valor de outro campo e desenha uma faixa de opções conectando os valores. Uma faixa de opções é essencialmente uma linha com efeitos em 3D. Não é um gráfico 3D real.</p> <p><i>Requer:</i> um par de campos de qualquer tipo.</p>		<p>Superfície</p> <p>Exibe os valores de três campos representados com relação aos valores de um outro, com uma superfície conectando os valores.</p> <p><i>Requer:</i> três campos de qualquer tipo.</p>
	<p>Gráfico de dispersão</p> <p>Exibe os valores de um campo representados com relação aos valores de outro. Este gráfico pode destacar o relacionamento entre os campos (se houver um). Também é possível utilizar o nó do gráfico Plotagem para produzir um gráfico de dispersão. Esse nó oferece algumas opções adicionais. Consulte o tópico “Nó Gráfico” na página 232 para obter mais informações.</p> <p><i>Requer:</i> um par de campos de qualquer tipo.</p>		<p>Gráfico de Bolhas</p> <p>Assim como um gráfico de dispersão básico, exibe os valores de um campo representados com relação aos valores de outro. A diferença é que os valores de um terceiro campo são utilizados para variar o tamanho dos pontos individuais.</p> <p><i>Requer:</i> três campos de qualquer tipo.</p>
	<p>Gráfico de dispersão categórico</p> <p>Assim como um gráfico de dispersão básico, exibe os valores de um campo representados com relação aos valores de outro. A diferença é que valores semelhantes são categorizados em grupos e que a estética de cor e de tamanho é usada para indicar o número de casos em cada categoria.</p> <p><i>Requer:</i> um par de campos contínuos.</p>		<p>Gráfico de dispersão categórico</p> <p>Consulte a descrição de Gráfico de Dispersão Categorizado. A diferença é a forma das categorias subjacentes, que se parecem com hexágonos e não com círculos. O gráfico de dispersão hexa categorizado resultante é semelhante ao gráfico de dispersão categorizado. No entanto, o número de valores em cada categoria diferirá entre os gráficos devido à forma das categorias subjacentes.</p> <p><i>Requer:</i> um par de campos contínuos.</p>

Tabela 50. Tipos de gráfico disponíveis (continuação).

Ícone de gráfico



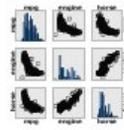
Descrição

Gráfico de dispersão 3D

Exibe os valores de três campos representados uns aos outros. Este gráfico pode destacar o relacionamento entre os campos (se houver um). Também é possível utilizar o nó do gráfico Plotagem para produzir um gráfico de dispersão 3D. Esse nó oferece algumas opções adicionais. Consulte o tópico “Nó Gráfico” na página 232 para obter mais informações.

Requer: três campos de qualquer tipo.

Ícone de gráfico

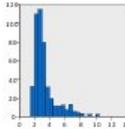


Descrição

Matriz de gráfico de dispersão (SPLOM)

Exibe os valores de um campo representados com relação aos valores de cada outro campo. Um SPLOM é como uma tabela de gráficos de dispersões. O SPLOM também inclui um histograma de cada campo.

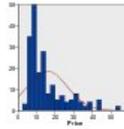
Requer: dois ou mais campos contínuos.



Histograma

Exibe a distribuição de frequência de um campo. Um histograma pode ajudar a determinar o tipo de distribuição e ver se a distribuição é defasada. Também é possível utilizar o nó do gráfico Histograma para produzir este gráfico. Esse nó oferece algumas opções adicionais. Consulte o tópico “Guia Gráfico de Histograma” na página 249 para obter mais informações.

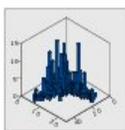
Requer: um campo único de qualquer tipo.



Histograma com distribuição normal

Exibe a distribuição de frequência de um campo contínuo com uma curva sobreposta da distribuição normal.

Requer: um campo contínuo único.



Histograma em 3D

Exibe a distribuição de frequência de um par de campos contínuos.

Requer: um par de campos contínuos.



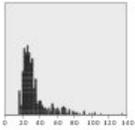
Densidade em 3D

Exibe a distribuição de frequência de um par de campos contínuos. Isso é semelhante a um histograma 3D, com a única diferença de que uma superfície é utilizada ao invés de barras para mostrar a distribuição.

Requer: um par de campos contínuos.

Tabela 50. Tipos de gráfico disponíveis (continuação).

Ícone de gráfico



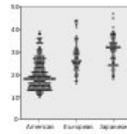
Descrição

Gráfico de pontos

Exibe os casos/linhas individuais e os empilha nos pontos de dados distintos no eixo x. Este gráfico é semelhante a um histograma em que ele mostra a distribuição dos dados, mas exibe cada caso/linha ao invés de uma contagem agregada de uma categoria específica (intervalo de valores).

Requer: um campo único de qualquer tipo.

Ícone de gráfico



Descrição

Gráfico de pontos em 2D

Exibe os casos/linhas individuais e os empilha nos pontos de dados distintos no eixo y para cada categoria de um campo categórico.

Requer: Um campo categórico e um campo contínuo.

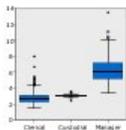
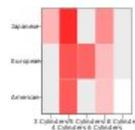


Diagrama em caixa

Calcula as cinco estatísticas (mínimo, primeiro quartil, mediana, terceiro quartil e máximo) para um campo contínuo para cada categoria de um campo categórico. Os resultados são exibidos como elementos boxplot/esquema. Os boxplots podem ajudar a ver como a distribuição de dados contínuos varia entre as categorias.

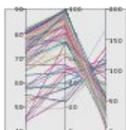
Requer: Um campo categórico e um campo contínuo.



Heat map

Calcula a média para um campo contínuo para a passagem das categorias entre dois campos categóricos.

Requer: Um par de campos categóricos e um campo contínuo.



Paralelo

Cria eixos paralelos para cada campo e desenha uma linha através do valor do campo para cada linha/caso nos dados.

Requer: dois ou mais campos contínuos.



Coropleta de contagens

Calcula a contagem para cada categoria de um campo categórico (**Chave de Dados**) e desenha um mapa que utiliza saturação de cor para representar as contagens nas variáveis de mapa que correspondem às categorias.

Requer: um campo categórico. Um arquivo de mapeamento cuja chave corresponde às categorias de **Chave de Dados**.

Tabela 50. Tipos de gráfico disponíveis (continuação).

Ícone de gráfico	Descrição	Ícone de gráfico	Descrição
	<p>Coropleta de Médias/Mediana/Soma</p> <p>Calcula a média, a mediana ou a soma de um campo contínuo (Cor) para cada categoria de um campo categórico (Chave de Dados) e desenha um mapa que utiliza saturação de cor para representar as estatísticas calculadas nas variáveis de mapa que correspondem às categorias.</p> <p><i>Requer:</i> Um campo categórico e um campo contínuo. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>		<p>Coropleta de valores</p> <p>Desenha um mapa que usa cor para representar valores de um campo categórico (Cor) para as variáveis de mapa que correspondem aos valores definidos por outro campo categórico (Chave de Dados). Se houver diversos valores categóricos do campo Cor para cada variável, o valor modal será utilizado.</p> <p><i>Requer:</i> um par de campos categóricos. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>
	<p>Coordenadas em uma coropleta de contagens</p> <p>Isso é semelhante à Coropleta de Contagens, exceto que há dois campos contínuos adicionais (Longitude e Latitude) que identificam coordenadas para desenhar pontos no mapa coroplético.</p> <p><i>Requer:</i> um campo categórico e um par de campos contínuos. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>		<p>Coordenadas em uma Coropleta de Médias/Mediana/Soma</p> <p>Isso é semelhante à Coropleta de Médias/Mediana/Soma, exceto que há dois campos contínuos adicionais (Longitude e Latitude) que identificam coordenadas para desenhar pontos no mapa coroplético.</p> <p><i>Requer:</i> um campo categórico e três campos contínuos. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>
	<p>Coordenadas em uma coropleta de valores</p> <p>Isso é semelhante à Coropleta de Valores, exceto que há dois campos contínuos adicionais (Longitude e Latitude) que identificam coordenadas para desenhar pontos no mapa coroplético.</p> <p><i>Requer:</i> um par de campos categóricos e um par de campos contínuos. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>		<p>Barras de contagens em um mapa</p> <p>Calcula a proporção de linhas/casos em cada categoria de um campo categórico (Categorias) para cada mapa de variável (Chave de Dados) e desenha um mapa e os gráficos de barras no centro de cada variável de mapa.</p> <p><i>Requer:</i> um par de campos categóricos. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>

Tabela 50. Tipos de gráfico disponíveis (continuação).

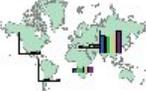
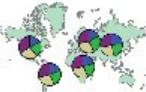
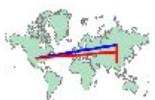
Ícone de gráfico	Descrição	Ícone de gráfico	Descrição
	<p>Barras em um mapa</p> <p>Calcula uma estatística de sumarização para um campo contínuo (Valores) e exibe os resultados para cada categoria de um campo categórico (Categorias) para cada variável de mapa (Chave de Dados) como gráficos de barras posicionados no centro de cada variável de mapa.</p> <p><i>Requer:</i> Um par de campos categóricos e um campo contínuo. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>		<p>Setores de contagens de um mapa</p> <p>Exibe a proporção de linhas/casos em cada categoria de um campo categórico (Categorias) para cada mapa de variável (Chave de Dados) e desenha um mapa e as proporções como fatias de um gráfico de pizza no centro de cada variável de mapa.</p> <p><i>Requer:</i> um par de campos categóricos. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>
	<p>Sector em um mapa</p> <p>Calcula a soma de um campo contínuo (Valores) em cada categoria de um campo categórico (Categorias) para cada variável de mapa (Chave de Dados) e desenha um mapa e as somas como fatias de um gráfico de pizza no centro de cada variável de mapa.</p> <p><i>Requer:</i> Um par de campos categóricos e um campo contínuo. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>		<p>Gráfico de linha em um mapa</p> <p>Calcula uma estatística de sumarização para um campo contínuo (Y) para cada valor de outro campo (X) para cada variável de mapa (Chave de Dados) e desenha um mapa e gráficos de linha conectando os valores no centro de cada variável de mapa.</p> <p><i>Requer:</i> um campo categórico e um par de campos de qualquer tipo. Um arquivo de mapeamento cuja chave corresponde às categorias de Chave de Dados.</p>
	<p>Coordenadas em um mapa de referência</p> <p>Desenha um mapa e pontos utilizando campos contínuos (Longitude e Latitude) que identificam as coordenadas para os pontos.</p> <p><i>Requer:</i> um par de campos de intervalo. Um arquivo de mapeamento.</p>		<p>Setas em um mapa de referência</p> <p>Desenha um mapa e setas utilizando campos contínuos que identificam os pontos de início (Início Longo e Lat. de Início) e os terminais (Término Longo e Lat. de Término) para cada seta. Cada registro/caso nos dados resulta em uma seta no mapa.</p> <p><i>Requer:</i> quatro campos contínuos. Um arquivo de mapeamento.</p>

Tabela 50. Tipos de gráfico disponíveis (continuação).

Ícone de gráfico	Descrição	Ícone de gráfico	Descrição
	<p>Mapa de Sobreposição de Ponto</p> <p>Desenha um mapa de referência e sobrepõe outro mapa de ponto sobre ele com variáveis de ponto coloridas por um campo categórico (Cor).</p> <p><i>Requer:</i> um par de campos categóricos. Um arquivo de mapeamento de ponto cuja chave corresponde às categorias Chave de Dados. Um arquivo de mapeamento de referência.</p>		<p>Mapa poligonal de sobreposição</p> <p>Desenha um mapa de referência e sobrepõe outro mapa de polígono sobre ele com variáveis de polígono coloridas por um campo categórico (Cor).</p> <p><i>Requer:</i> um par de campos categóricos. Um arquivo de mapeamento de polígono cuja chave corresponde às categorias Chave de Dados. Um arquivo de mapeamento de referência.</p>
	<p>Mapa linear de sobreposição</p> <p>Desenha um mapa de referência e sobrepõe outro mapa de linha sobre ele com variáveis de linha coloridas por um campo categórico (Cor).</p> <p><i>Requer:</i> um par de campos categóricos. Um arquivo de mapeamento de linha cuja chave corresponde às categorias Chave de Dados. Um arquivo de mapeamento de referência.</p>		

Criando Visualizações de Mapa

Para muitas visualizações, é necessário fazer apenas duas escolhas: campos (variáveis) de interesse e um modelo para visualizar esses campos. Nenhuma opção ou ação adicional é necessária. As visualizações de mapa requerem pelo menos um passo adicional: selecionar um arquivo de mapeamento que define as informações geográficas para a visualização do mapa.

Os passos básicos para criar um mapa simples são os seguintes:

1. Selecione os campos de interesse na guia Básico. Para obter informações sobre o tipo e o número de campos necessários para diferentes visualizações de mapa, consulte “Tipos de Visualizações de Gráfico Integradas Disponíveis” na página 204.
2. Selecione um modelo de mapa.
3. Clique na guia Detalhado.
4. Verifique se a **Chave de Dados** e outras listas suspensas necessárias estão configuradas para os campos corretos.
5. No grupo Arquivos de Mapeamento, clique em **Selecionar um Arquivo de Mapeamento**.
6. Use a caixa de diálogo Selecionar Mapas para escolher o arquivo de mapeamento e a chave de mapa. Os valores da chave de mapa devem corresponder aos valores para o campo especificado pela **Chave de Dados**. É possível utilizar o botão **Comparar** para comparar esses valores. Se você selecionar um modelo de mapa de sobreposição, também será necessário escolher um mapa de referência. O mapa de referência não possui chave para os dados. Ele é utilizado como o plano de fundo para o mapa principal. Para obter mais informações sobre a caixa de diálogo Selecionar Mapas, consulte “Selecionando Arquivos de Mapeamento para Visualizações de Mapa” na página 202.

7. Clique em **OK** para fechar a caixa de diálogo Selecionar Mapas.
8. No Seletor de Modelo de Gráfico, clique em **Executar** para criar a visualização do mapa.

Exemplos de Gráficos

Essa seção inclui vários exemplos diferentes para demonstrar as opções disponíveis. Os exemplos também fornecem informações para interpretar as visualizações resultantes.

Esses exemplos utilizam o fluxo denominado *graphboard.str*, que faz referência aos arquivos de dados denominados *employee_data.sav*, *customer_subset.sav* e *worldsales.sav*. Esses arquivos estão disponíveis na pasta *Demos* de qualquer instalação do IBM SPSS Modeler Client. Isso pode ser acessado a partir do grupo do programa IBM SPSS Modeler no menu Iniciar do Windows. O arquivo *graphboard.str* está na pasta *streams*.

É recomendado ler os exemplos na ordem apresentada. Os exemplos subsequentes são construídos sobre os anteriores.

Exemplo: Gráfico de Barras com uma Estatísticas de Sumarização

Vamos criar um gráfico de barras que sumariza um campo numérico contínuo/variável para cada conjunto/variável categórica. Especificamente, criaremos um gráfico de barras que mostra a média salarial para homens e mulheres.

Este e vários dos seguintes exemplos usam *Dados do funcionário*, que é um conjunto de dados hipotético contendo informações sobre os funcionários de uma empresa.

1. Inclua um nó de origem Arquivo de Estatísticas que aponte para *employee_data.sav*.
2. Incluir um Elemento do gráfico e abri-lo para edição.
3. Na guia Básico, selecione *Sexo* e *Salário Atual*. (Use Ctrl+Click para selecionar diversos campos/variáveis).
4. Selecione **Barra**.
5. Na lista suspensa Sumarização, selecione **Média**.
6. Clique em **Executar**.
7. Na exibição resultante, clique no botão da barra de ferramentas "Exibir rótulos de campo e de valor" (o segundo do grupo de dois no centro da barra de ferramentas).

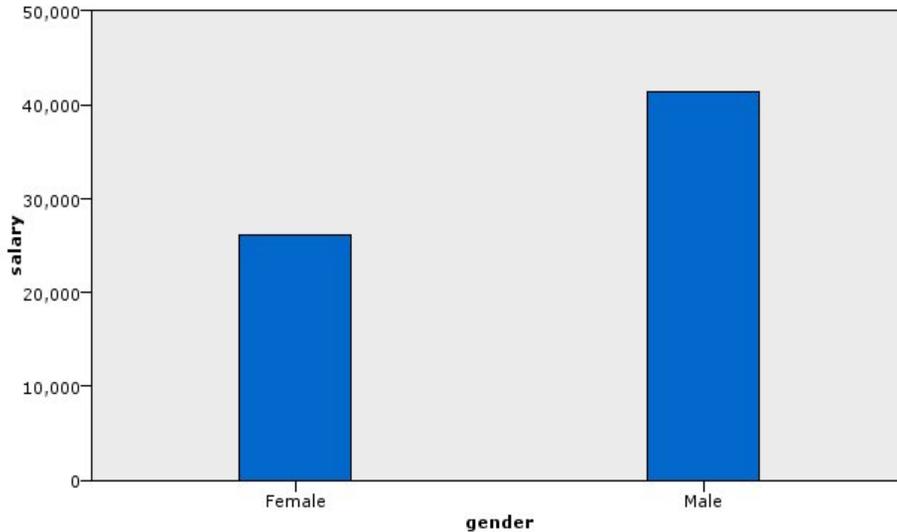


Figura 10. Gráfico de barras com uma estatísticas de sumarização

É possível observar o seguinte:

- Com base na altura das barras, é evidente que a média salarial para homens é maior que a média salarial para mulheres.

Exemplo: Gráfico de Barras Empilhadas com uma Estatística de Sumarização

Vamos criar agora um gráfico de barras empilhadas para ver se a diferença na média salarial entre homens e mulheres está sujeita ao tipo de emprego. Em média, as mulheres ganham mais que os homens em determinados tipos de empregos.

Nota: Este exemplo utiliza *Dados de funcionário*.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Categoria de Emprego* e *Salário Atual*. (Use Ctrl+Click para selecionar diversos campos/variáveis).
3. Selecione **Barra**.
4. Na lista Sumarização, selecione **Média**.
5. Clique na guia Detalhado. Observe que suas seleções na guia anterior são refletidas aqui.
6. No grupo Estética Opcional, escolha *sexo* na lista suspensa Cor.
7. Clique em **Executar**.

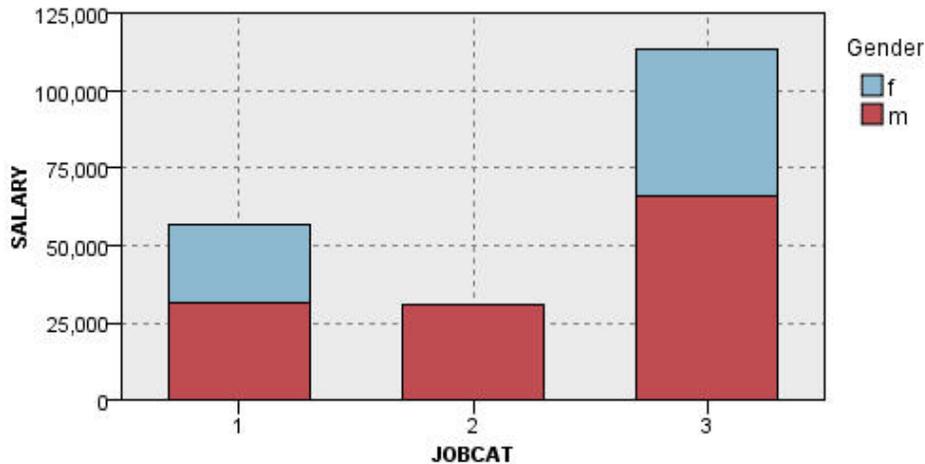


Figura 11. Gráfico de barras empilhadas

É possível observar o seguinte:

- A diferença nas médias salariais para cada tipo de emprego não parece ser tão grande quanto era no gráfico de barras que comparou as médias salariais para todos os homens e mulheres. Talvez haja um número variado de homens e mulheres em cada grupo. É possível verificar isso ao criar um gráfico de barras de contagens.
- Independentemente do tipo de emprego, a média salarial para homens é sempre maior que a média salarial para mulheres.

Exemplo: Histograma Agrupado em Painéis

Criaremos um histograma que será agrupado em painel por sexo para que possamos comparar as distribuições de frequência de salário entre homens e mulheres. A distribuição de frequência mostra quantos casos/linhas estão dentro faixas salariais específicas. O histograma agrupado em painéis pode ajudar a analisar ainda mais a diferença de salários entre os sexos.

Nota: este exemplo utiliza *Dados de funcionário*.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Salário Atual*.
3. Selecione **Histograma**.
4. Clique na guia Detalhado.
5. No grupo Painéis e Animação, escolha *sexo* na lista suspensa No Painel
6. Clique em **Executar**.

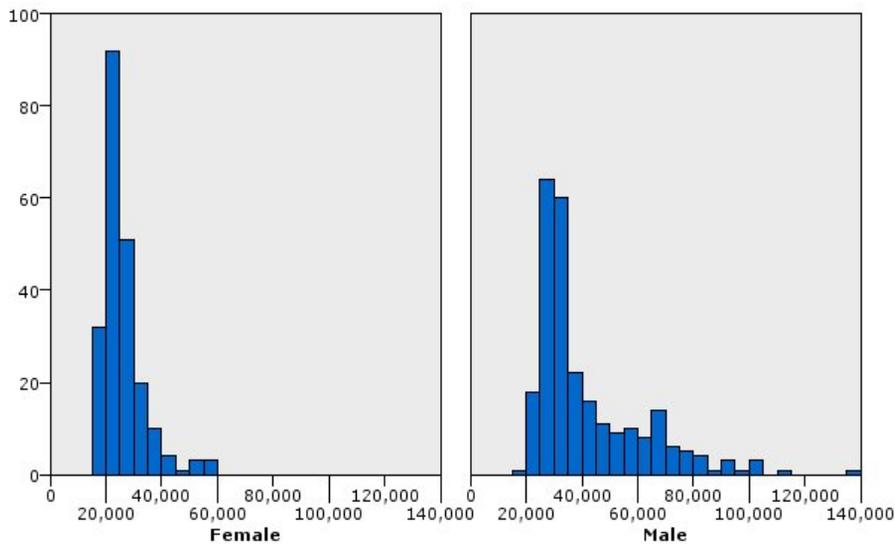


Figura 12. Histograma Agrupado em Painéis

É possível observar o seguinte:

- Nenhuma distribuição de frequência é uma distribuição normal. Ou seja, os histogramas não se parecem com curvas em sino, o que ocorreria se os dados fossem normalmente distribuídos.
- As barras mais altas estão no lado esquerdo de cada gráfico. Portanto, tanto homens quanto mulheres ganham salários menores ao invés de maiores.
- As distribuições de frequência de salário entre homens e mulheres não são iguais. Observe o formato dos histogramas. Há mais homens que ganham salários maiores do que mulheres que ganham salários maiores.

Exemplo: gráfico de pontos agrupados em painéis

Assim como um histograma, um gráfico de pontos mostra a distribuição de um amplitude numérica contínua. E ao contrário de um histograma, que mostra contagens de intervalos categorizados de dados, um gráfico de pontos mostra cada linha/caso nos dados. Portanto, um gráfico de pontos fornece maior granularidade em comparação com o histograma. Na realidade, o uso de um gráfico de pontos pode ser o ponto de início preferencial quando analisar as distribuições de frequência.

Nota: este exemplo utiliza *Dados de funcionário*.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Salário Atual*.
3. Selecione **Gráfico de Pontos**.
4. Clique na guia Detalhado.
5. No grupo Painéis e Animação, escolha *sexo* na lista suspensa No Painel
6. Clique em **Executar**.
7. Maximize a janela de saída resultante para visualizar o gráfico mais claramente.

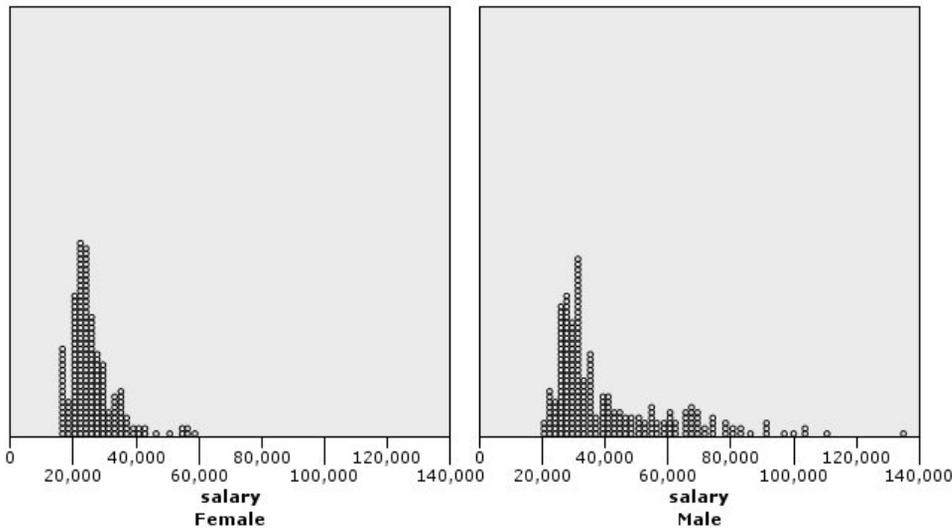


Figura 13. Gráfico de pontos agrupado em painéis

Comparado com o histograma (consulte “Exemplo: Histograma Agrupado em Painéis” na página 214), é possível observar o seguinte:

- O pico de 20.000 que aparece no histograma para mulheres é menos dramático no gráfico de pontos. Há muitos casos/linhas concentrados ao redor desse valor, mas a maioria desses valores está mais próxima de 25.000. Este nível de granularidade não é evidente no histograma.
- Embora o histograma para homens sugira que a média salarial deles diminui gradativamente após 40.000, o gráfico de pontos mostra que a distribuição é razoavelmente uniforme após esse valor, até 80.000. Em qualquer valor salarial nesse intervalo, há três ou mais homens que ganham esse salário específico.

Exemplo: Boxplot

Um boxplot é outra visualização útil para ver como os dados são distribuídos. Um boxplot contém várias medidas estatísticas que iremos explorar após criar a visualização.

Nota: este exemplo utiliza *Dados de funcionário*.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Sexo* e *Salário Atual*. (Use Ctrl+Click para selecionar diversos campos/variáveis).
3. Selecione **Boxplot**.
4. Clique em **Executar**.

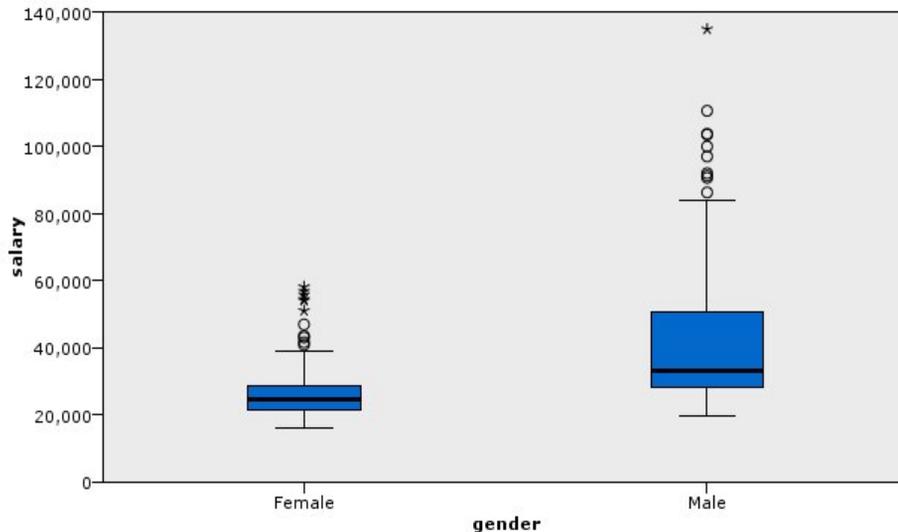


Figura 14. Diagrama em caixa

Vamos explorar as diferentes partes do boxplot:

- A linha escura no meio das caixas é a mediana de *salário*. Metade dos casos/linhas tem um valor maior que a mediana e metade tem um valor menor. Assim como a média, a mediana é uma medida da tendência central. Ao contrário da média, a mediana é menos influenciada pelos casos/linhas com valores extremos. Neste exemplo, a mediana é menor que a média (em comparação com “Exemplo: Gráfico de Barras com uma Estatísticas de Sumarização” na página 212). A diferença entre a média e a mediana indica que há alguns casos/linhas com valores extremos que elevam a média. Ou seja, alguns funcionários têm salários maiores.
- A parte inferior da caixa indica o 25º percentil. Vinte e cinco por cento dos casos/linhas possuem valores abaixo do 25º percentil. A parte superior da caixa representa o 75º percentil. Vinte e cinco por cento dos casos/linhas possuem valores acima do 75º percentil. Isso significa que 50% dos casos/linhas estão dentro da caixa. A caixa é muito mais curta para mulheres do que para homens. Esta é uma dica de que *salário* varia menos para mulheres do que para homens. As partes superior e inferior da caixa são geralmente chamadas de **juntas**.
- As barras T que se estendem a partir das caixas são chamadas de **fences internos** ou **whiskers**. Essas barras se estendem até uma vez e meia a altura da caixa ou, se nenhum caso/linha possuir um valor nesse intervalo, até os valores mínimo ou máximo. Se os dados forem distribuídos normalmente, espera-se que aproximadamente 95% dos dados estejam entre os fences internos. Neste exemplo, os fences internos estendem menos para mulheres do que para os homens, o que é outra indicação de que *salário* varia menos para mulheres do que para homens.
- Os pontos são **valores discrepantes**. Estes são definidos como valores que não caem nos fences internos. Valores discrepantes são valores extremos. Os asteriscos ou estrelas são **valores discrepantes extremos**. Esses representam casos/linhas que possuem valores mais de três vezes a altura das caixas. Há diversos valores discrepantes para mulheres e homens. Lembre-se de que a média é maior que a mediana. A média maior é causada por esses valores discrepantes.

Exemplo: Gráfico de Pizza

Agora usaremos um conjunto de dados diferente para explorar alguns outros tipos de visualização. O conjunto de dados é *customer_subset*, que é um arquivo de dados hipotético que contém informações sobre clientes.

Vamos primeiro criar um gráfico de pizza para verificar a proporção de clientes em diferentes regiões geográficas.

1. Inclua um nó de origem Arquivo de Estatísticas que aponte para *customer_subset.sav*.

2. Incluir um Elemento do gráfico e abri-lo para edição.
3. Na guia Básico, selecione *Indicador geográfico* .
4. Selecione **Setor de Contagens**.
5. Clique em **Executar**.

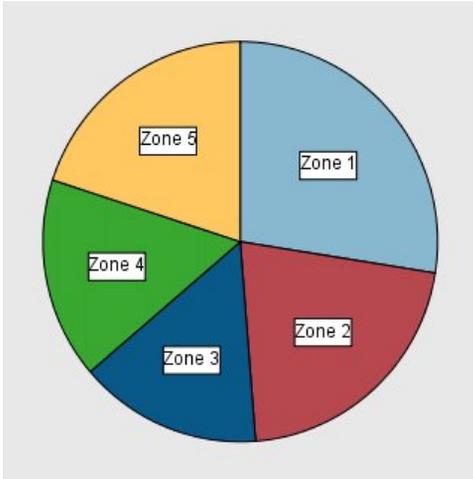


Figura 15. Gráfico de pizza

É possível observar o seguinte:

- A Zona 1 possui mais clientes do que em cada uma das outras zonas.
- Os clientes são igualmente distribuídos entre as outras zonas.

Exemplo: Heat Map

Vamos criar agora um heat map categórico para verificar a renda média dos clientes em diferentes regiões geográficas e faixas etárias.

Nota: este exemplo utiliza *customer_subset*.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Indicador geográfico*, *Categoria de idade* e *Renda familiar em milhares* , nessa ordem. (Use Ctrl+Clique para selecionar diversos campos/variáveis).
3. Selecione **Heat Map**.
4. Clique em **Executar**.
5. Na janela de saída resultante, clique no botão da barra de ferramentas "Exibir rótulos de campo e de valor" (o botão da direita dos dois localizados no centro da barra de ferramentas).

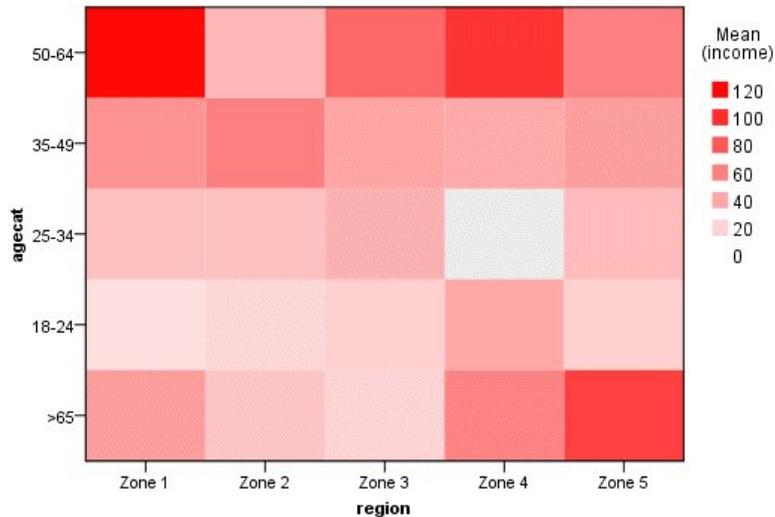


Figura 16. Heat map categórico

É possível observar o seguinte:

- Um heat map é uma tabela que utiliza cores ao invés de números para representar os valores para as células. Um vermelho claro intenso indica o maior valor, ao passo que cinza indica um valor baixo. O valor de cada célula é a média do campo/variável contínuo para cada par de categorias.
- Exceto na Zona 2 e na Zona 5, o grupo de clientes cuja idade está entre 50 e 64 anos possui uma renda familiar média maior que os outros grupos.
- Não há clientes entre as idades de 25 e 34 anos na zona 4.

Exemplo: Matriz de Gráfico de Dispersão (SPLOM)

Criaremos uma matriz de gráfico de dispersão de diversas variáveis diferentes para que possamos determinar se existem quaisquer relacionamentos entre as variáveis no conjunto de dados.

Nota: este exemplo utiliza `customer_subset`.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Idade em anos*, *Renda familiar em milhares* e *Dívida de cartão de crédito em milhares*. (Use Ctrl+Click para selecionar diversos campos/variáveis).
3. Selecione **SPLOM**.
4. Clique em **Executar**.
5. Maximize a janela de saída para visualizar a matriz mais claramente.

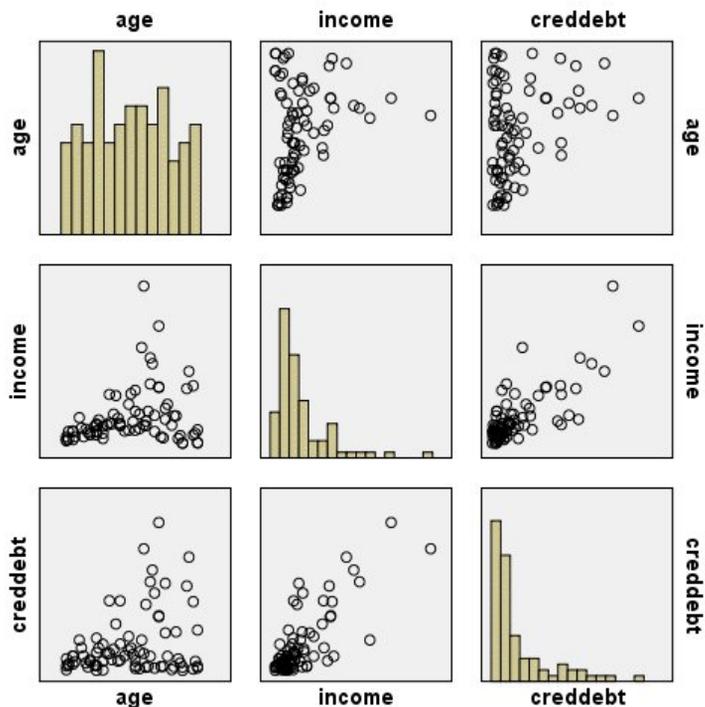


Figura 17. Matriz de gráfico de dispersão (SPLOM)

É possível observar o seguinte:

- Os histogramas exibidos na diagonal mostram a distribuição de cada variável no SPLOM. O histograma para *age* aparece na célula superior esquerda, o histograma para *income* aparece na célula central e o histograma para *creddebt* aparece na célula inferior direita. Nenhuma das variáveis parece normalmente distribuída. Ou seja, nenhum dos histogramas lembra uma curva de sino. Além disso, observe que os histogramas para *income* e *creddebt* são positivamente defasados.
- Não parece haver qualquer relacionamento entre *age* e as outras variáveis.
- Há um relacionamento linear entre *income* e *creddebt*. Ou seja, *creddebt* aumenta à medida que *income* aumenta. Talvez você queira criar gráficos de dispersão individuais dessas variáveis e de outras variáveis relacionadas para explorar os relacionamentos ainda mais.

Exemplo: Coropletas (Mapa de Cores) de Somas

Agora criaremos uma visualização de mapa. Em seguida, no exemplo subsequente, criaremos uma variação dessa visualização. O conjunto de dados é *worldsales*, que é um arquivo de dados hipotéticos que contém a receita de vendas por continente e produto.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Continente* e *Receita*. (Use Ctrl+Click para selecionar diversos campos/variáveis).
3. Selecione **Coropleta de Somas**.
4. Clique na guia Detalhado.
5. No grupo Estética Opcional, escolha *Continente* na lista suspensa Rótulo de Dados.
6. No grupo Arquivos de Mapeamento, clique em **Selecionar um Arquivo de Mapeamento**.
7. Na caixa de diálogo Selecionar Mapas, verifique se **Mapa** está configurado para *Continentes* e se **Chave de Mapa** está configurada para *CONTINENT*.
8. Nos grupos Comparar Mapa e Valores de Dados, clique em **Comparar** para assegurar que as chaves de mapa correspondam às chaves de dados. Nesse exemplo, todos os valores de chave de dados possuem chaves e variáveis de mapa correspondentes. Também podemos ver que não há dados para a Oceania.

9. Na caixa de diálogo Seleccionar Mapas, clique em **OK**.
10. Clique em **Executar**.



Figura 18. Coropleta de Somas

Nessa visualização de mapa, podemos ver facilmente que a receita é maior na América do Norte e menor na América do Sul e na África. Cada continente está rotulado porque utilizamos *Continente* para a estética de rótulo de dados.

Exemplo: Gráficos de Barras em um Mapa

Este exemplo mostra como as receitas são divididas por produto em cada continente.

Nota: este exemplo utiliza *worldsales*.

1. Incluir um Elemento do gráfico e abri-lo para edição.
2. Na guia Básico, selecione *Continente*, *Produto* e *Receita*. (Use Ctrl+Click para selecionar diversos campos/variáveis).
3. Selecione **Barras em um Mapa**.
4. Clique na guia Detalhado.
Ao utilizar mais de um campo de um tipo específico, é importante verificar se cada campo é designado ao slot correto.
5. Na lista suspensa Categorias, escolha *Produto*.
6. Na lista suspensa Valores, escolha *Receita*.
7. Na lista suspensa Chave de Dados, escolha *Continente*.
8. Na lista suspensa Sumarização, escolha *Soma*.
9. No grupo Arquivos de Mapeamento, clique em **Selecionar um Arquivo de Mapeamento**.
10. Na caixa de diálogo Seleccionar Mapas, verifique se **Mapa** está configurado para *Continentes* e se **Chave de Mapa** está configurada para *CONTINENT*.

11. Nos grupos Comparar Mapa e Valores de Dados, clique em **Comparar** para assegurar que as chaves de mapa correspondam às chaves de dados. Nesse exemplo, todos os valores de chave de dados possuem chaves e variáveis de mapa correspondentes. Também podemos ver que não há dados para a Oceania.
12. Na caixa de diálogo Seleccionar Mapas, clique em **OK**.
13. Clique em **Executar**.
14. Maximize a janela de saída resultante para visualizar a exibição mais claramente.

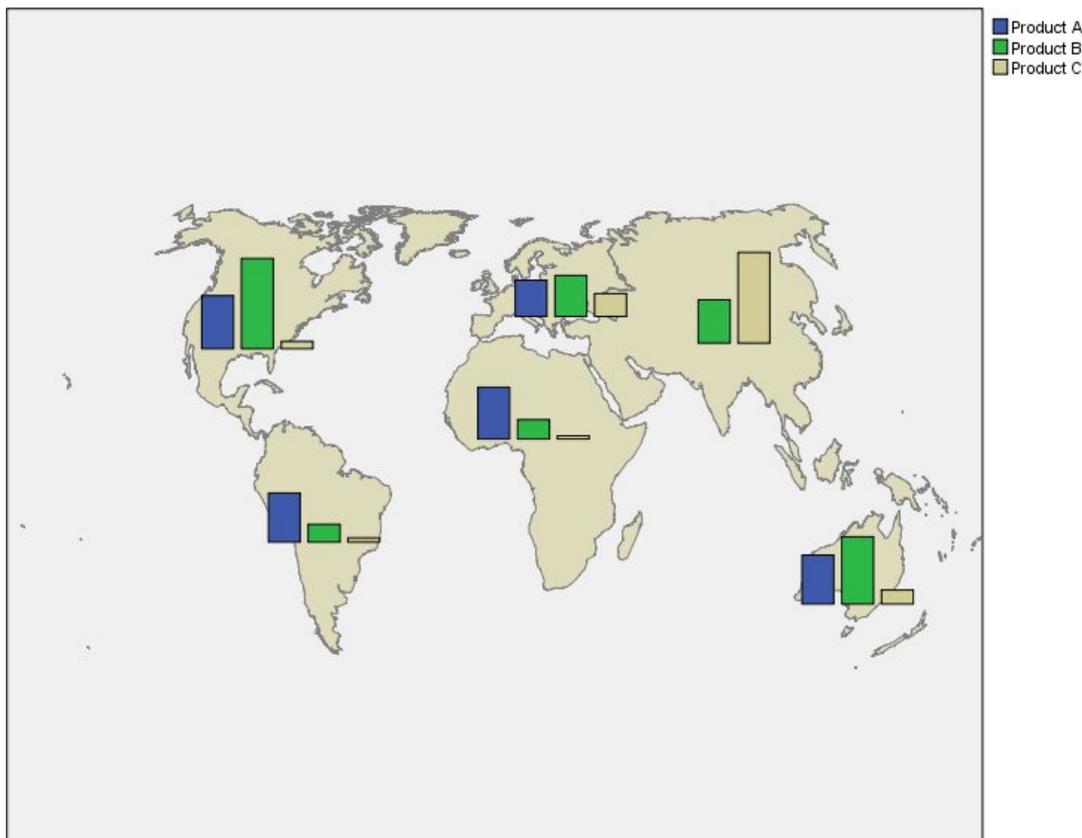


Figura 19. Gráficos de Barras em um Mapa

É possível observar o seguinte:

- A distribuição de receita total dos produtos é muito semelhante na América do Sul e na África.
- O *Produto C* gera a menor receita em qualquer lugar, exceto na Ásia.
- Há uma receita mínima ou nenhuma receita do *Produto A* na Ásia.

Guia Aparência do Gráfico

É possível especificar opções de aparência antes da criação do gráfico.

Opções de Aparência Geral

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Amostragem. Especifique um método para conjuntos de dados maiores. É possível especificar um tamanho máximo de conjunto de dados ou utilizar o número padrão de registros. O desempenho é aprimorado para grandes conjuntos de dados ao selecionar a opção **Amostra**. Como alternativa, é possível escolher criar um gráfico de todos os pontos de dados ao selecionar **Usar todos os dados**, no entanto, deve-se observar que isso pode diminuir drasticamente o desempenho do software.

Opções de Aparência de Folha de Estilo

Há dois botões que permitem controlar quais modelos de visualização (e folhas de estilo e mapas) estão disponíveis:

Gerenciar. Gerencie modelos de visualização, folhas de estilo e mapas em seu computador. É possível importar, exportar, renomear e excluir modelos de visualização, folhas de estilo e mapas em sua máquina local. Consulte o tópico “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224 para obter mais informações.

Localização. Altere a localização na qual os modelos de visualização, folhas de estilo e mapas são armazenados. A localização atual é listada à direita do botão. Consulte o tópico “Configurando a Localização dos Modelos, Folhas de Estilos e Mapas” para obter mais informações.

O exemplo a seguir mostra onde as opções de aparência são colocadas em um gráfico. (*Nota:* nem todos os gráficos utilizam todas essas opções).

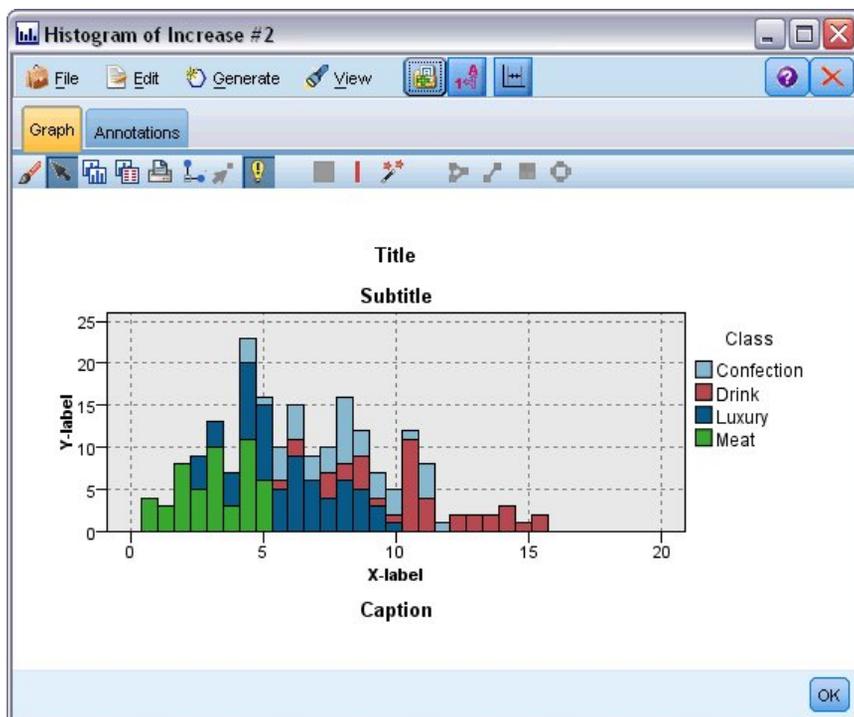


Figura 20. Posição das várias opções de aparência do gráfico

Configurando a Localização dos Modelos, Folhas de Estilos e Mapas

Os modelos de visualização, as folhas de estilo de visualização e os arquivos de mapeamento são armazenados em uma pasta local específica ou no IBM SPSS Collaboration and Deployment Services Repository. Ao selecionar modelos, folhas de estilo e mapas, somente aqueles integrados nesta localização são exibidos. Ao manter todos os modelos, folhas de estilo e arquivos de mapeamento em um local, os aplicativos do IBM SPSS poderão acessá-los facilmente. Para obter informações sobre como incluir

modelos, folhas de estilo e arquivos de mapeamento adicionais nesta localização, consulte “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento”.

Como Configurar a Localização de Modelos, de Folhas de Estilo e de Arquivos de Mapeamento

1. Em uma caixa de diálogo de modelo ou de folha de estilo, clique em **Localização...** para exibir a caixa de diálogo Modelos, Folhas de Estilo e Mapas.
2. Selecione uma opção para a localização padrão para modelos, folhas de estilo e arquivos de mapeamento:

Máquina Local. Os modelos, folhas de estilo e arquivos de mapeamento estão localizados em uma pasta específica em seu computador local. No Windows XP, essa pasta é *C:\Documents and Settings\\Application Data\SPSSInc Graphboard *. A pasta não pode ser alterada.

IBM SPSS Collaboration and Deployment Services Repository. Os modelos, folhas de estilo e arquivos de mapeamento estão localizados em uma pasta especificada pelo usuário no IBM SPSS Collaboration and Deployment Services Repository. Para identificar a pasta específica, clique em **Pasta**. Para obter mais informações, consulte “Utilizando o IBM SPSS Collaboration and Deployment Services Repository como a Localização de Modelo, Folha de Estilo e do Arquivo de Mapa”.

3. Clique em **OK**.

Utilizando o IBM SPSS Collaboration and Deployment Services Repository como a Localização de Modelo, Folha de Estilo e do Arquivo de Mapa

Os modelos e folhas de estilo de visualização podem ser armazenados no IBM SPSS Collaboration and Deployment Services Repository. Essa localização é uma pasta específica no IBM SPSS Collaboration and Deployment Services Repository. Se isto for configurado como a localização padrão, quaisquer modelos, folhas de estilo e arquivos de mapeamento nessa localização estarão disponíveis para seleção.

Como Configurar uma pasta no IBM SPSS Collaboration and Deployment Services Repository como a Localização para Modelos, Folhas de Estilo e Arquivos de Mapeamento

1. Em uma caixa de diálogo com um botão Localização, clique em **Localização....**
2. Selecione IBM SPSS Collaboration and Deployment Services Repository.
3. Clique na **Pasta**.
Nota: se você ainda não estiver conectado ao IBM SPSS Collaboration and Deployment Services Repository, será solicitado a fornecer informações de conexão.
4. Na caixa de diálogo Selecionar Pasta, selecione a pasta na qual os modelos, as folhas de estilo e os arquivos de mapeamento são armazenados.
5. Opcionalmente, é possível selecionar um rótulo em **Recuperar Rótulo**. Apenas modelos, folhas de estilo e arquivos de mapeamento com esse rótulo serão exibidos.
6. Se estiver procurando por uma pasta que contém um determinado modelo ou folha de estilo, você poderá querer procurar pelo modelo, pela folha de estilo ou pelo arquivo de mapeamento na guia Procurar. A caixa de diálogo Selecionar Pasta seleciona automaticamente a pasta na qual o modelo, a folha de estilo ou o arquivo de mapeamento encontrado está localizado.
7. Clique em **Selecionar Pasta**.

Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento

É possível gerenciar os modelos, as folhas de estilo e os arquivos de mapeamento localmente em seu computador usando a caixa de diálogo Gerenciar Modelos, Folhas de Estilo e Mapas. Esta caixa de diálogo permite importar, exportar, renomear e excluir modelos de visualização, folhas de estilo e arquivos de mapeamento localmente em seu computador.

Clique em **Gerenciar ...** em uma das caixas de diálogo na qual você seleciona modelos, folhas de estilo ou mapas.

Caixa de Diálogo Gerenciar Modelos, Folhas de Estilos e Mapas

A guia Modelo lista todos os modelos locais. A guia Folha de Estilo lista todas as folhas de estilo locais, além de exibir visualizações de exemplo com dados de amostra. É possível selecionar uma das folhas de estilo para aplicar seus estilos às visualizações de exemplo. Consulte o tópico “Aplicando Folhas de Estilo” na página 297 para obter mais informações. A guia Mapa lista todos os arquivos de mapa locais. Esta guia também exibe as chaves de mapa, incluindo valores de amostra, um comentário, se um foi fornecido ao criar o mapa, e uma visualização do mapa.

Os seguintes botões operam em qualquer guia que estiver ativada no momento.

Importar. Importe um modelo de visualização, uma folha de estilo ou um arquivo de mapeamento a partir do sistema de arquivos. Importar um modelo, uma folha de estilo ou um arquivo de mapeamento torna-o disponível para o aplicativo IBM SPSS. Se outro usuário enviou a um modelo, uma folha de estilo ou um arquivo de mapeamento, você importará o arquivo antes de utilizá-lo em seu aplicativo.

Exportar. Exporte um modelo de visualização, uma folha de estilo ou um arquivo de mapeamento para o sistema de arquivos. Exporte um modelo, uma folha de estilo ou um arquivo de mapeamento quando desejar enviá-lo para outro usuário.

Renomear. Renomeie o modelo de visualização, a folha de estilo ou um arquivo de mapeamento selecionado. Não é possível alterar um nome para um que já esteja sendo utilizado.

Exportar Chave de Mapa. Exporte as chaves de mapa como um arquivo de valores separados por vírgula (CSV). Esse botão é ativado somente na guia Mapa.

Excluir. Exclua um ou mais modelos de visualização, folhas de estilo ou arquivos de mapeamento selecionados. É possível selecionar diversos modelos, folhas de estilo ou arquivos de mapeamento com Ctrl-clique. Não há nenhuma ação de desfazer para a exclusão, portanto, continue com cuidado.

Convertendo e Distribuindo Arquivos de Forma de Mapa

O Seletor de Modelo de Gráfico permite criar visualizações de mapa a partir da combinação de um modelo de visualização e de um arquivo do SMZ. Os arquivos SMZ são semelhantes aos arquivos de forma do ESRI (formato de arquivo SHP) pelo fato de conterem as informações geográficas para desenhar um mapa (por exemplo, fronteiras de país), mas eles são otimizados para visualizações do mapa. O Seletor de Modelo de Gráfico é pré-instalado com um número selecionado de arquivos SMZ. Se você tiver um arquivo de forma ESRI existente que deseja utilizar para visualizações de mapa, primeiro é necessário converter o arquivo de forma em um arquivo SMZ utilizando o Utilitário de Conversão de Mapa. O Utilitário de Conversão de Mapa suporta arquivos de forma ESRI de ponto, polilinha ou polígono (tipos de forma 1, 3 e 5) contendo uma camada única.

Além de converter arquivos de forma ESRI, o Utilitário de Conversão de Mapas permite modificar o nível de detalhes do mapa, alterar rótulos de variável, mesclar variáveis e mover variáveis, dentre muitas outras mudanças opcionais. Também é possível utilizar o Utilitário de Conversão de Mapa para modificar um arquivo SMZ existente (incluindo aqueles pré-instalados).

editando arquivos do SMZ pré-instalados

1. Exporte o arquivo SMZ a partir de Gerenciar sistema. Consulte o tópico “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224 para obter mais informações.
2. Use o Utilitário de Conversão de Mapa para abrir e editar o arquivo SMZ exportado. É recomendável salvar o arquivo com um nome diferente. Consulte o tópico “Usando o Utilitário de Conversão de Mapa” na página 226 para obter mais informações.
3. Importe o arquivo SMZ modificado em Gerenciar sistema. Consulte o tópico “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224 para obter mais informações.

Recursos Adicionais para Arquivos de Mapeamento

Dados geoespaciais no formato de arquivo SHP, que podem ser utilizados para suportar suas necessidades de mapeamento, estão disponíveis a partir de muitas origens privadas e públicas. Verifique com websites do governo local se buscar dados livres. Muitos dos modelos neste produto são baseados em dados disponíveis publicamente obtidos a partir de GeoCommons () e do U.S. Census Bureau (<http://www.census.gov>).

AVISO IMPORTANTE: As informações relacionadas a produtos não IBM foram obtidas junto aos fornecedores destes produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou estes produtos e não pode confirmar a precisão do desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não IBM. Perguntas sobre os recursos de produtos não IBM devem ser encaminhadas aos fornecedores desses produtos. As referências nestas informações a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a estes websites. Os materiais nesses websites não fazem parte dos materiais para este programa IBM, a menos que indicado em um arquivo Avisos que acompanha este programa IBM, e a utilização dos materiais dos sites é de seu próprio risco.

Conceitos Chave para Mapas

Entender alguns conceitos chave relacionados aos arquivos de forma ajudará a utilizar o Utilitário de Conversão de Mapa de modo eficiente.

Um **arquivo de forma** fornece as informações geográficas para desenhar um mapa. Há três tipos de arquivos de forma que o Utilitário de Conversão de Mapa suporta:

- **Ponto.** O arquivo de forma identifica as localizações dos pontos, como cidades.
- **Polilinha.** O arquivo de forma identifica os caminhos e suas localizações, como rios.
- **Polígono.** O arquivo de forma identifica um regiões delimitadas e suas localizações, como países.

Normalmente você utilizará um arquivo de forma de polígono. Mapas Coropléticos são criados a partir de arquivos de forma de polígono. Os mapas coropléticos utilizam cor para representar um valor dentro de polígonos individuais (regiões). Arquivos de forma de ponto e polilinha geralmente são sobrepostos em um arquivo de forma de polígono. Um exemplo é um arquivo de forma de ponto das cidades dos EUA sobreposto em um arquivo de forma de polígono dos estados desse mesmo país.

Um arquivo de forma é composto de **variáveis**. As variáveis são as entidades geográficas individuais. Por exemplo, as variáveis podem ser países, estados, cidades, e assim por diante. O arquivo de forma também contém dados sobre as variáveis. Esses dados são armazenados em **atributos**. Os atributos são semelhantes aos campos ou variáveis em um arquivo de dados. Há pelo menos um atributo que é a **chave de mapa** para a variável. A chave de mapa pode ser um rótulo, como um nome de país ou de estado. A chave de mapa é o que você ligará para uma variável/campo em um arquivo de dados para criar uma visualização de mapa.

Observe que será possível reter apenas um ou mais atributos-chave no arquivo SMZ. O Utilitário de Conversão de Mapa não suporta salvar atributos adicionais. Isso significa que será necessário criar diversos arquivos SMZ se desejar agregar em níveis diferentes. Por exemplo, se desejar agregar estados e regiões dos EUA, será necessário separar arquivos SMZ: um que possua uma chave que identifica estados e outro que possua uma chave que identifica regiões.

Usando o Utilitário de Conversão de Mapa

Como iniciar o Utilitário de Conversão de Mapa

A partir dos menus, escolha:

Ferramentas > Utilitário de Conversão de Mapa

Há quatro telas principais (passos) no Utilitário de Conversão de Mapa. Um dos passos também envolve os subpassos para obter um controle mais detalhado sobre a edição do arquivo de mapeamento.

Passo 1 - Escolher Arquivos de Destino e de Origem

Primeiro é necessário selecionar um arquivo de mapeamento de origem e um destino para o arquivo de mapeamento convertido. Serão necessários os arquivos *.shp* e *.dbf* para o arquivo de forma.

Selecione um arquivo *.shp* (ESRI) ou *.smz* para conversão. Navegue até um arquivo de mapeamento existente em seu computador. Esse é o arquivo que será convertido e salvo como um arquivo SMZ. O arquivo *.dbf* para o arquivo de forma *deve* ser armazenado na mesma localização com um nome de arquivo de base que corresponda ao arquivo *.shp*. O arquivo *.dbf* é necessário porque contém as informações sobre o atributo do arquivo *.shp*.

Configurar um destino e nome do arquivo para o arquivo de mapeamento convertido. Insira um caminho e nome do arquivo para o arquivo SMZ que será criado a partir da origem do mapa original.

- **Importar no Seletor de Modelo.** Além de salvar um arquivo no sistema de arquivos, é possível, opcionalmente, incluir o mapa na lista Gerenciar do Seletor de Modelo. Se escolher essa opção, o mapa estará disponível automaticamente no Seletor de Modelo para os produtos do IBM SPSS instalados em seu computador. Se você não importar no Seletor de Modelo agora, será necessário importá-lo manualmente mais tarde. Para obter mais informações sobre como importar mapas no sistema de Gerenciamento do Seletor de Modelo, consulte “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224.

Passo 2 - Escolher a chave do mapa

Agora você escolherá quais chaves de mapa serão incluídas no arquivo SMZ. Em seguida, é possível alterar algumas opções que afetarão a renderização do mapa. Os passos subsequentes no Utilitário de Conversão de Mapa incluem uma visualização do mapa. As opções de renderização que você escolher serão utilizadas para gerar a visualização do mapa.

Escolher chave de mapa primária. Selecione o atributo que é a chave primária para identificar e rotular variáveis no mapa. Por exemplo, chave primária de um mapa-múndi pode ser o atributo que identifica os nomes de países. A chave primária também vincula seus dados às variáveis de mapa, portanto, assegure-se de que os valores (rótulos) do atributo que você escolher correspondam aos valores em seus dados. Rótulos de exemplo são exibidos quando você escolhe um atributo. Se precisar alterar esses rótulos, você será capaz de fazer isso em um passo posterior.

Escolher chaves adicionais a serem incluídas. Além da chave de mapa primária, verifique quaisquer outros atributos-chave que deseja incluir no arquivo SMZ gerado. Por exemplo, alguns atributos podem conter rótulos traduzidos. Se esperar dados codificados em outros idiomas, você poderá querer preservar esses atributos. Observe que é possível escolher apenas as chaves adicionais que representam as mesmas variáveis que a chave primária. Por exemplo, se a chave primária for os nomes completos de estados dos EUA, será possível selecionar apenas as chaves alternativas que representam os estados desse mesmo país, como abreviações de estado.

Suavizar automaticamente o mapa. Os arquivos de forma com polígonos geralmente contêm muitos pontos de dados e muitos detalhes de visualizações de mapa estatísticas. O excesso de detalhes pode causar distração e impactar negativamente o desempenho. É possível reduzir o nível de detalhe e generalizar o mapa com suavização. O mapa terá uma aparência mais nítida e, conseqüentemente, renderizará mais rápido. Quando o mapa é suavizado automaticamente, o ângulo máximo é 15 graus e a porcentagem a ser mantida é 99. Para obter informações sobre essas configurações, consulte “Suavizar o mapa” na página 228. Observe que você tem a oportunidade de aplicar suavização adicional posteriormente em outro passo.

Remover limites entre polígonos em contato na mesma variável. Algumas variáveis podem incluir subvariáveis que possuem limites internos para as variáveis principais de interesse. Por exemplo, um

mapa-múndi de continentes pode conter limites internos para os países contidos em cada continente. Se escolher essa opção, os limites internos não aparecerão no mapa. No mapa-múndi de continentes de exemplo, escolher essa opção remove os limites de países enquanto retém os limites de continente.

Passo 3 - Editar o mapa

Agora que você especificou opções básicas para o mapa, é possível editar as mais específicas. Essas modificações são opcionais. Esse passo do Utilitário de Conversão de Mapa guiará você nas tarefas associadas e exibirá uma visualização do mapa para que seja possível verificar suas mudanças. Algumas tarefas podem não estar disponíveis, dependendo do tipo de arquivo de forma (ponto, polilinha ou polígono) e sistema de coordenadas.

Cada tarefa possui os seguintes controles comuns no lado esquerdo do Utilitário de Conversão de Mapa.

Mostrar os rótulos no mapa. Por padrão, rótulos de variável não são mostrados na visualização. É possível escolher exibir os rótulos. Embora os rótulos possam ajudar a identificar as variáveis, eles podem interferir na seleção direta no mapa de visualização. Ative esta opção quando necessário, por exemplo, quando estiver editando os rótulos de variável.

Cor da visualização de mapa. Por padrão, o mapa de visualização exibe áreas com uma cor sólida. Todas as variáveis têm a mesma cor. É possível escolher ter um sortimento de cores designadas às variáveis de mapa individuais. Esta opção pode ajudar a distinguir diferentes variáveis no mapa. Isso será útil especialmente quando mesclar variáveis e desejar ver como as novas variáveis são representadas na visualização.

Cada tarefa também tem o seguinte controle comum no lado direito do Utilitário de Conversão de Mapa.

Desfazer. Clique em **Desfazer** para reverter para o último estado. É possível desfazer um máximo de 100 mudanças.

Suavizar o mapa: Os arquivos de forma com polígonos geralmente contêm muitos pontos de dados e muitos detalhes de visualizações de mapa estatísticas. O excesso de detalhes pode causar distração e impactar negativamente o desempenho. É possível reduzir o nível de detalhe e generalizar o mapa com suavização. O mapa terá uma aparência mais nítida e, conseqüentemente, renderizará mais rápido. Essa opção não está disponível para os mapas de ponto e polilinha.

Ângulo máximo O ângulo máximo, que deve ter um valor entre 1 e 20, especifica a tolerância para conjuntos de suavização de pontos que forem quase lineares. Um valor maior permite maior tolerância para a suavização linear e eliminará subseqüentemente mais pontos, resultando em um mapa mais generalizado. Para aplicar suavização linear, o Utilitário de Conversão de Mapa verifica o ângulo interno formado por cada conjunto de três pontos no mapa. Se 180 menos o ângulo for menor que o valor especificado, o Utilitário de Conversão de Mapa eliminará o ponto médio. Em outras palavras, o Utilitário de Conversão de Mapa verifica se a linha formada pelos três pontos é quase reta. Se for, o Utilitário de Conversão de Mapa tratará a linha como uma linha reta entre os terminais e descartará o ponto médio.

Porcentagem para manter. A porcentagem a ser mantida, que deve ter um valor entre 90 e 100, determina a quantia de área terrestre para manter quando o mapa for suavizado. Esta opção afeta apenas as variáveis que possuem diversos polígonos, como no caso se uma variável incluísse ilhas. Se a área total de uma variável menos um polígono for maior que a porcentagem especificada da área original, o Utilitário de Conversão de Mapa eliminará o polígono do mapa. O Utilitário de Conversão de Mapa nunca removerá todos os polígonos para a variável. Ou seja, sempre haverá pelo menos um polígono para a variável, independentemente da quantia de suavização aplicada.

Depois de escolher um ângulo e porcentagem máximos a serem mantidos, clique em **Aplicar**. A visualização será atualizada com as mudanças de suavização. Se precisar suavizar o mapa novamente,

repeita até obter o nível de suavidade desejada. Observe que há um limite para suavização. Se você suavizar repetidamente, chegará a um ponto em que nenhuma suavização adicional poderá ser aplicada ao mapa.

Editar os rótulos de variável: É possível editar os rótulos de variável conforme necessário (talvez para corresponder aos dados esperados) e também reposicionar os rótulos no mapa. Mesmo se você não achar necessário alterar os rótulos, deve-se revisá-los antes de criar visualizações no mapa. Como os rótulos não são mostrados por padrão na visualização, você também pode desejar selecionar **Mostrar os rótulos no mapa** para exibi-los.

Chaves. Selecione a chave que contém os rótulos de variável que deseja revisar e/ou editar.

Variáveis. Essa lista exibe os rótulos de variável contidos na chave selecionada. Para editar o rótulo, dê um clique duplo nele na lista. Se os rótulos forem mostrados no mapa, também é possível dar um clique duplo nos rótulos de variável diretamente na visualização do mapa. Se desejar comparar os rótulos com um arquivo de dados real, clique em **Comparar**.

X/Y. Essas caixas de texto listam o ponto central atual do rótulo da variável selecionada no mapa. As unidades são exibidas nas coordenadas do mapa. Elas podem ser locais, coordenadas cartesianas (por exemplo, o Sistema de Coordenadas de Plano de Estado) ou coordenadas geográficas (em que X é longitude e a latitude é Y). Insira as coordenadas para a nova posição do rótulo. Se os rótulos forem mostrados, também será possível clicar e arrastar um rótulo no mapa para movê-lo. As caixas de texto serão atualizadas com a nova posição.

Comparar. Se você tiver um arquivo de dados que contém valores de dados que são destinados a corresponder os rótulos de variáveis para uma chave específica, clique em **Comparar** para exibir a caixa de diálogo Comparar com uma Origem de Dados Externa. Nessa caixa de diálogo, é possível abrir o arquivo de dados e comparar seus valores diretamente com aqueles nos rótulos de variável da chave do mapa.

Caixa de diálogo Comparar com uma Origem de Dados Externa: A caixa de diálogo Comparar com uma Origem de Dados Externa permite abrir um arquivo de valores separados por tabulação (com uma extensão *.txt*), um arquivo de valores separados por vírgula (com uma extensão *.csv*) ou um arquivo de dados formatado para o IBM SPSS Statistics (com uma extensão *.sav*). Quando o arquivo é aberto, é possível selecionar um campo no arquivo de dados para comparar com os rótulos de variável em uma chave de mapa específica. Em seguida, é possível corrigir quaisquer discrepâncias no arquivo de mapeamento.

Campos no arquivo de dados. Escolha o campo cujos valores você deseja comparar com os rótulos de variável. Se a primeira linha no arquivo *.txt* ou *.csv* contiver rótulos descritivos para cada campo, marque **Usar primeira linha como rótulos da coluna**. Caso contrário, cada campo será identificado por sua posição no arquivo de dados (por exemplo, "Coluna 1", "Coluna 2", e assim por diante).

Chave para comparar. Escolha a chave de mapa cujos rótulos de variável você deseja comparar com os valores de campo do arquivo de dados.

Comparar. Clique quando estiver pronto para comparar os valores.

Resultados da Comparação. Por padrão, a tabela Resultados da Comparação lista apenas os valores do campo não correspondidos no arquivo de dados. O aplicativo tenta localizar um rótulo de variável relacionado, geralmente ao verificar espaços inseridos ou omissos. Clique na lista suspensa no *Rótulo do Mapa* para corresponder o rótulo de variável no arquivo de mapeamento com o valor do campo exibido. Se não houver nenhum rótulo de variável correspondente em seu arquivo de mapeamento, escolha *Deixar Não Correspondido*. Se desejar ver todos os valores do campo, mesmo aqueles que já corresponderem a um rótulo de variável, desmarque **Exibir apenas casos não correspondidos**. Isso poderá ser feito se desejar substituir uma ou mais correspondências.

É possível utilizar cada variável apenas uma vez para correspondê-la a um valor do campo. Se desejar corresponder diversas variáveis com um valor de campo único, será possível mesclar as variáveis e, em seguida, corresponder a nova variável mesclada com o valor do campo. Para obter mais informações sobre como mesclar variáveis, consulte “Mesclar variáveis”.

Mesclar variáveis: A mesclagem de variáveis é útil para criar regiões maiores em um mapa. Por exemplo, se estiver convertendo um mapa de estados, será possível mesclar os estados (as variáveis neste exemplo) para regiões Norte, Sul, Leste e Oeste maiores.

Chaves. Selecione a chave de mapa que contém os rótulos de variável que ajudarão a identificar as variáveis que deseja mesclar.

Variáveis. Clique na primeira variável que deseja mesclar. Dê um ctrl-clique nas outras variáveis que deseja mesclar. Observe que as variáveis também serão selecionadas na visualização de mapa. É possível clicar e dar um Ctrl-clique nas variáveis diretamente na visualização do mapa, além de selecioná-las na lista.

Após selecionar as variáveis que deseja mesclar, clique em **Mesclar** para exibir a caixa de diálogo Nome e Variável Mesclada, na qual será possível aplicar um rótulo na nova variável. Você pode querer selecionar **Colorir a visualização de mapa** após a mesclagem de variável para assegurar que os resultados sejam os esperados.

Após mesclar as variáveis, você também pode querer mover o rótulo para a nova variável. Isso pode ser feito na tarefa *Editar os rótulos da variável*. Consulte o tópico “Editar os rótulos de variável” na página 229 para obter mais informações.

Caixa de diálogo Nomear a Variável Mesclada: A caixa de diálogo Nomear a Variável Mesclada permite designar rótulos para a nova variável mesclada.

A tabela Rótulos exibe informações de cada chave no arquivo de mapeamento e permite designar um rótulo para cada chave.

Novo Rótulo. Insira um novo rótulo para a variável mesclada a ser designada para a chave de mapa específica.

Chave. A chave do mapa para a qual você está designando o novo rótulo.

Rótulos Antigos. Os rótulos para as variáveis que serão mescladas na nova variável.

Remover limites entre polígonos em contato. Marque esta opção para remover os limites das variáveis que foram mescladas. Por exemplo, se você mesclou estados em regiões geográficas, essa opção removerá os limites em torno dos estados individuais.

Mover variáveis: É possível mover variáveis no mapa. Isso pode ser útil quando desejar colocar variáveis juntas, como um continente e ilhas periféricas.

Chaves. Selecione a chave de mapa que contém os rótulos de variável que ajudarão a identificar as variáveis que você deseja mover.

Variáveis. Clique na variável que deseja mover. Observe que a variável será selecionada na visualização de mapa. Também é possível clicar na variável diretamente na visualização do mapa.

X/Y. Essas caixas de texto listam o ponto central atual da variável selecionada no mapa. As unidades são exibidas nas coordenadas do mapa. Elas podem ser locais, coordenadas cartesianas (por exemplo, o Sistema de Coordenadas de Plano de Estado) ou coordenadas geográficas (em que X é longitude e a

latitude é Y). Insira as coordenadas para a nova posição da variável. Também é possível clicar e arrastar uma variável no mapa para movê-la. As caixas de texto serão atualizadas com a nova posição.

Excluir variáveis: É possível excluir variáveis indesejadas do mapa. Isso pode ser útil quando deseja remover alguma desordem ao excluir variáveis que não forem de interesse na visualização de mapa.

Chaves. Selecione a chave de mapa que contém os rótulos de variável que ajudarão a identificar as variáveis que você deseja excluir.

Variáveis. Clique na variável que deseja excluir. Se deseja excluir diversas variáveis simultaneamente, dê um Ctrl-clique nas variáveis adicionais. Observe que as variáveis também serão selecionadas na visualização de mapa. É possível clicar e dar um Ctrl-clique nas variáveis diretamente na visualização do mapa, além de selecioná-las na lista.

Excluir elementos individuais: Além de excluir variáveis inteiras, é possível excluir alguns dos elementos individuais que formam as variáveis, como lagos e pequenas ilhas. Essa opção não está disponível para mapas de ponto.

Elementos. Clique os elementos que deseja excluir. Se deseja excluir diversos elementos simultaneamente, dê um Ctrl-clique nos elementos adicionais. Observe que os elementos também serão selecionados na visualização de mapa. É possível clicar e dar um Ctrl-clique nos elementos diretamente na visualização do mapa, além de selecioná-los na lista. Como a lista de nomes de elemento não é descritiva (cada elemento é designado a um número dentro da variável), deve-se verificar a seleção na visualização de mapa para assegurar que você selecionou os elementos desejados.

Configurar a projeção:

A projeção de mapa especifica a maneira com que a Terra tridimensional é representada em duas dimensões. Todas as projeções causam distorções. No entanto, algumas projeções são mais adequadas dependendo se você estiver visualizando um mapa global ou um mapa mais local. Além disso, algumas projeções preservam a forma das variáveis originais. As projeções que preservam a forma são projeções conformais. Essa opção está disponível apenas para os mapas com coordenadas geográficas (longitude e latitude).

Diferentemente de outras opções no Utilitário de Conversão de Mapa, a projeção pode ser alterada após uma visualização do mapa ser criada.

Projeção. Selecione uma projeção de mapa. Se estiver criando um mapa global ou hemisférico, utilize as projeções *Local*, *Mercator* ou *Winkel Tripel*. Para áreas menores, utilize as projeções *Local*, *Cônica Conformal de Lambert* ou *Mercator Transversal*. Todas as projeções utilizam o elipsoide WGS83 para o datum.

- A projeção **Local** é sempre utilizada quando o mapa é criado com um sistema de coordenadas local, como o Sistema de Coordenadas Planas do Estado. Esses sistemas de coordenadas são definidos por coordenadas Cartesianas ao invés de coordenadas geográficas (longitude e latitude). Na projeção Local, linhas horizontais e verticais são igualmente espaçadas em um sistema de coordenadas Cartesianas. A projeção Local não é conformal.
- A projeção **Mercator** é uma projeção conformal de mapas globais. As linhas horizontais e verticais são retas e sempre perpendiculares entre si. Observe que a projeção Mercator se estende ao infinito à medida que se aproxima dos Polos Norte e Sul, portanto, ela não poderá ser utilizada se o seu mapa incluir o Polo Norte ou Sul. A distorção é maior quando o mapa se aproxima desses limites.
- A projeção **Winkel Tripel** é uma projeção não conformal de mapas globais. Embora não seja conformal, ela fornece um bom equilíbrio entre forma e tamanho. Exceto para o Equador e o Meridiano Central, todas as linhas são curvadas. Se o mapa global incluir o Polo Norte ou Sul, essa será uma boa opção de projeção.

- Como o nome sugere, a projeção **Cônica Conformal de Lambert** é uma projeção conformal utilizada para mapas de regiões continentais ou terrestres menores que se estendem na direção Leste e Oeste em comparação com a direção Norte e o Sul.
- O **Mercator Transversal** é outra projeção conformal para mapas de regiões continentais ou terrestres menores. Utilize esta projeção para as regiões terrestres que se estendem na direção Norte e Sul em comparação com a direção Leste e Oeste.

Passo 4 - Concluir

Neste ponto, é possível incluir um comentário para descrever o arquivo de mapeamento e também criar um arquivo de dados de amostra a partir das chaves de mapa.

Chaves de mapa. Se houver diversas chaves no arquivo de mapeamento, selecione uma chave de mapa cujos rótulos de variável você deseja exibir na visualização. Se criar um arquivo de dados a partir do mapa, esses rótulos serão utilizados para os valores de dados.

Comentário. Insira um comentário que descreva o mapa ou que forneça informações adicionais que podem ser relevantes para seus usuários, como as origens para os arquivos de forma originais. O comentário aparecerá no sistema de Gerenciamento do Seletor de Modelo de Gráfico.

Criar um conjunto de dados a partir dos rótulos de variáveis. Selecione essa opção se desejar criar um arquivo de dados a partir dos rótulos de variável exibidos. Ao clicar em **Procurar...**, será possível especificar uma localização e um nome de arquivo. Se incluir uma extensão *.txt*, o arquivo será salvo como um arquivo de valores separados por tabulação. Se incluir uma extensão *.csv*, o arquivo será salvo como um arquivo de valores separados por vírgula. Se incluir uma extensão *.sav*, o arquivo será salvo no formato do IBM SPSS Statistics. SAV é o padrão quando nenhuma extensão é especificada.

Distribuindo arquivos de mapas

No primeiro passo do Utilitário de Conversão de Mapa, você escolheu uma localização na qual deseja salvar o arquivo SMZ convertido. Você também pode ter escolhido incluir o mapa no sistema Gerenciar para o Seletor de Modelo de Gráfico. Se você escolheu salvar no sistema Gerenciar, o mapa estará disponível em qualquer produto do IBM SPSS que executar no mesmo computador.

Para distribuir o mapa para outros usuários, será necessário enviar o SMZ para eles. Esses usuários podem, então, utilizar o sistema Gerenciar para importar o mapa. É possível simplesmente enviar o arquivo cuja localização você especificou no passo 1. Se desejar enviar um arquivo que está no sistema Gerenciar, primeiro é necessário exportá-lo:

1. No Seletor de Modelo, clique em **Gerenciar...**
2. Clique na guia Mapa.
3. Selecione o mapa que deseja distribuir.
4. Clique em **Exportar...** e escolha uma localização na qual deseja salvar o arquivo.

Agora é possível enviar o arquivo de mapeamento físico para outros usuários. Os usuários precisarão reverter esse processo e importar o mapa no sistema Gerenciar.

Nó Gráfico

Os nós do gráfico mostram o relacionamento entre campos numéricos. É possível criar um gráfico utilizando pontos (também conhecido como um gráfico de dispersão), ou é possível utilizar linhas. É possível criar três tipos de gráficos de linha ao especificar um Modo X na caixa de diálogo.

Modo X = Ordenar

Configurar o Modo X para **Ordenar** faz com que os dados sejam ordenados pelos valores do campo representados no eixo x. Isso produz uma linha única que vai da esquerda para a direita no gráfico. Usar

um campo nominal como uma sobreposição produz diversas linhas de diferentes matizes que vão da esquerda para a direita no gráfico.

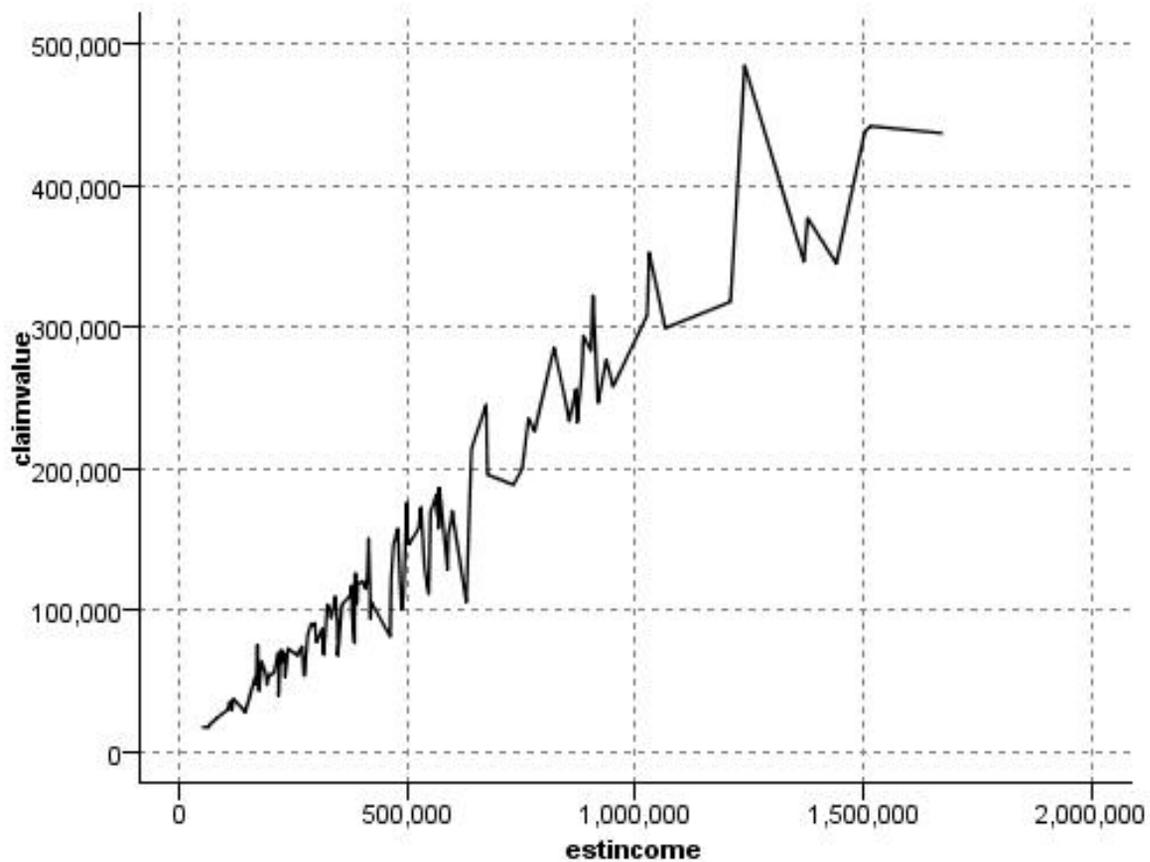


Figura 21. Gráfico de linha com o Modo X configurado para Ordenar

Modo X = Sobreposição

Configurar o Modo X para **Sobreposição** cria diversos gráficos de linhas no mesmo gráfico. Os dados não são ordenados para um gráfico de sobreposição; desde que os valores no eixo x aumentem, os dados serão representados em uma única linha. Se os valores diminuïrem, uma nova linha inicia. Por exemplo, conforme x move de 0 a 100, os valores de y serão representados em uma única linha. Quando x fica abaixo de 100, uma nova linha é representada além da primeira. O gráfico concluído pode ter diversos gráficos úteis para comparar várias séries de valores y . Esse tipo de gráfico é útil para dados com um componente de tempo periódico, como a demanda de energia elétrica em períodos sucessivos de 24 horas.

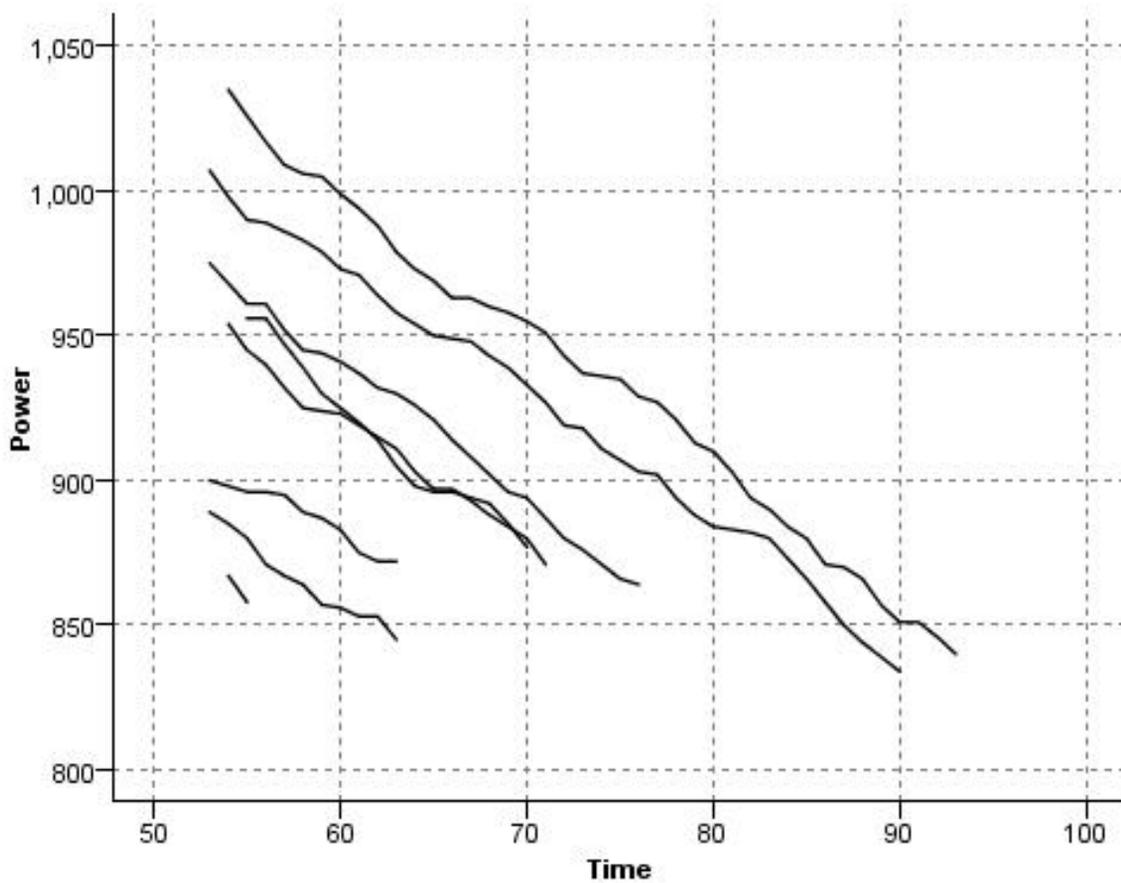


Figura 22. Gráfico de linha com o Modo X configurado para Sobreposição

Modo X = Conforme Lido

Configurar o Modo X para **Conforme Lido** representa os valores x e y conforme eles são lidos a partir da origem de dados. Essa opção é útil para dados com um componente de séries temporais em que você se interessa pelas tendências ou padrões que dependem da ordem dos dados. Poderá ser necessário ordenar os dados antes de criar esse tipo de gráfico. Também pode ser útil comparar dois gráficos semelhantes com o Modo X configurado para **Ordenar** e **Conforme Lido** para determinar o quanto um padrão depende da ordenação.

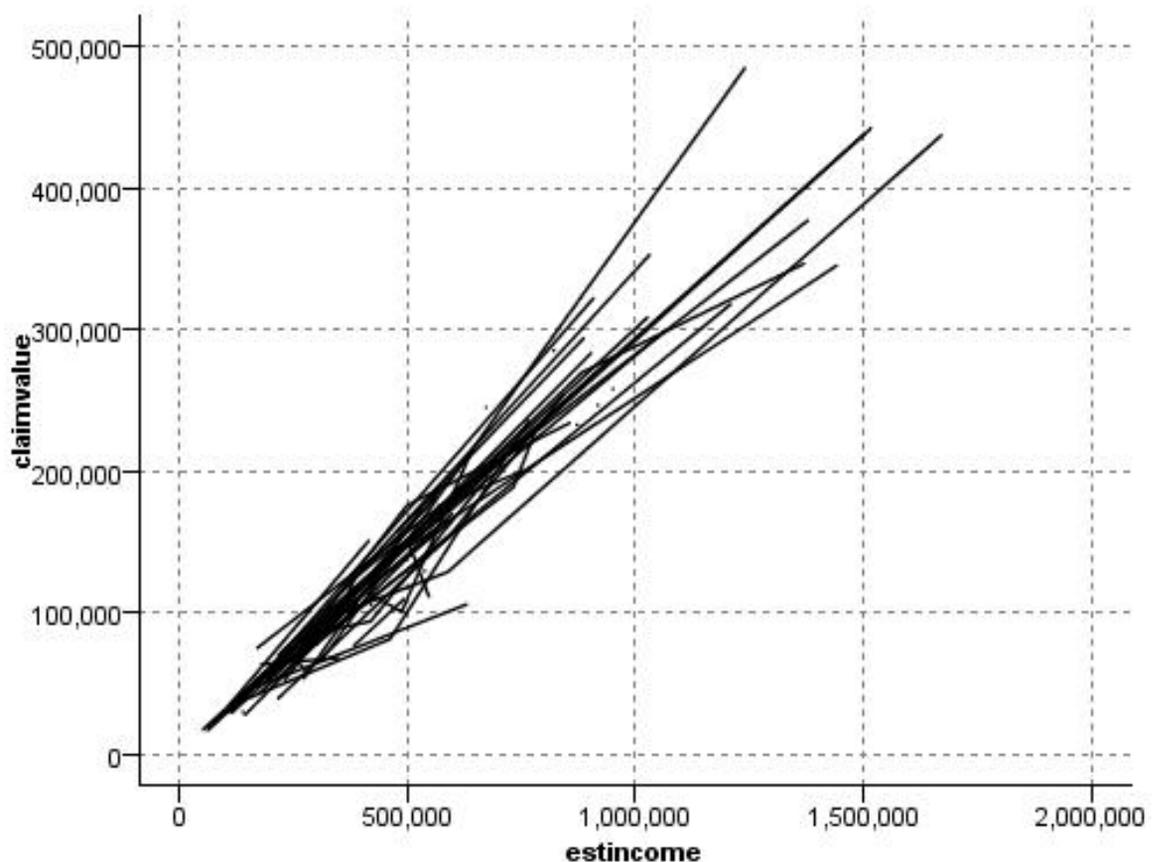


Figura 23. Gráfico de linha mostrado anteriormente como Ordenar, executado novamente com o Modo X configurado como Conforme Lido

Também é possível utilizar o elemento do Gráfico para produzir gráficos de dispersão e gráficos de linha. No entanto, há mais opções a serem escolhidas a partir desse nó. Consulte o tópico “Tipos de Visualizações de Gráfico Integradas Disponíveis” na página 204 para obter mais informações.

Guia Nó do Gráfico

Os gráficos mostram valores de um campo Y com relação a valores de um campo X. Muitas vezes, esses campos correspondem a uma variável dependente e a uma variável independente, respectivamente.

Campo X. Na lista, selecione o campo a ser exibido no eixo *x* horizontal.

Campo Y. Na lista, selecione o campo a ser exibido no eixo *y* vertical.

Campo Z. Ao clicar no botão do gráfico 3D, é possível, em seguida, escolher um campo na lista para exibir no eixo *z*.

Sobreposição. Há várias maneiras de ilustrar categorias para os valores de dados. Por exemplo, é possível usar *maincrop* como uma sobreposição de cores para indicar os valores de *estincome* e de *claimvalue* para a agricultura principal cultivada pelos requerentes da solicitação. Consulte o tópico “Estética, Sobreposições, Painéis e Animação” na página 194 para obter mais informações.

Tipo de sobreposição. Especifica se uma função de sobreposição ou um suavizador é exibido. As funções de suavizador e de sobreposição são sempre calculadas como uma função de y .

- **Nenhum.** Nenhuma sobreposição é exibida.
- **Suavizador.** Exibe uma linha de ajuste suavizada calculada usando uma regressão por quadrados mínimos (LOESS) iterativa, robusta e localmente ponderada. Este método calcula efetivamente uma série de regressões, cada uma focando uma pequena área no gráfico. Isso produz uma série de linhas de regressão "locais" que são, então, juntadas para criar uma curva suave.

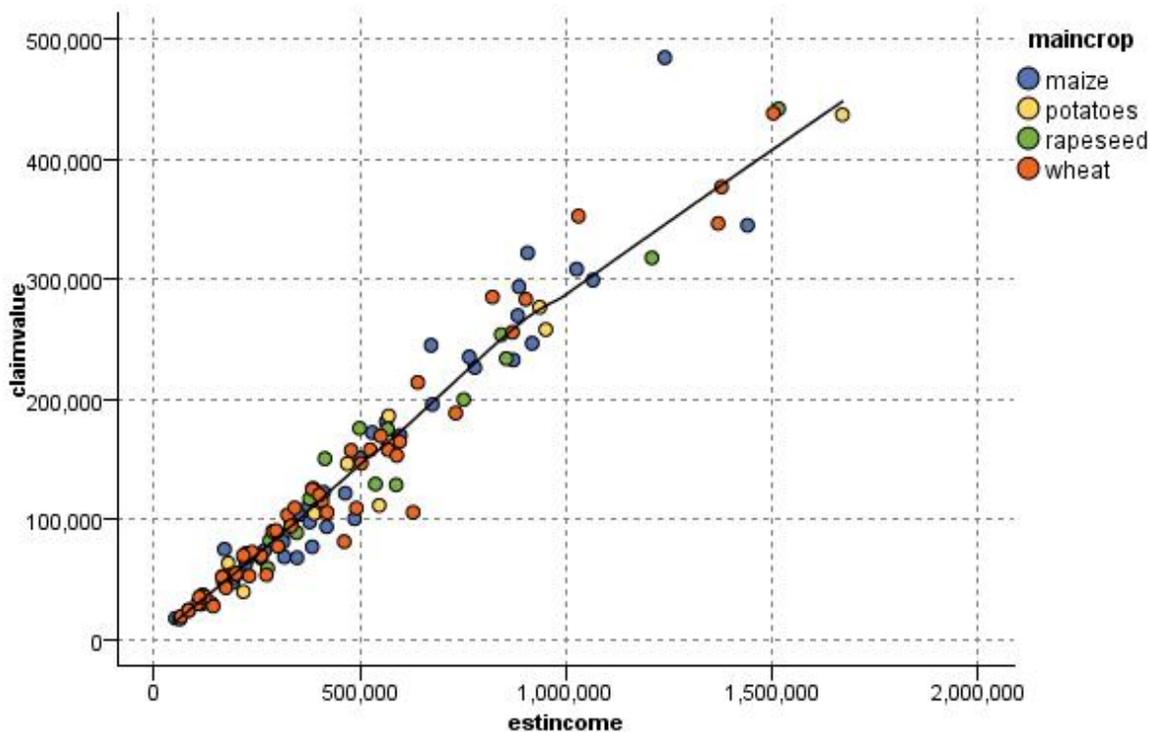


Figura 24. Gráfico com uma sobreposição de suavizador LOESS

- **Função.** Selecione para especificar uma função conhecida para comparar com os valores reais. Por exemplo, para comparar os valores reais versus valores preditos, é possível criar um gráfico da função $y = x$ como uma sobreposição. Especifique uma função para $y =$ na caixa de texto. A função padrão é $y = x$, mas é possível especificar qualquer tipo de função, como uma função quadrática ou uma expressão arbitrária, em termos de x .

Nota: as funções de sobreposição não estão disponíveis para um gráfico de painel ou de animação.

Após de ter configurado as opções para um gráfico, será possível executar o gráfico diretamente a partir da caixa de diálogo ao clicar em **Executar**. No entanto, talvez você queira utilizar a guia Opções para especificações adicionais, como categorização, Modo de X e estilo.

Guia Opções de Gráfico

Estilo. Selecione **Ponto** ou **Linha** para o estilo de gráfico. A seleção de **Linha** ativa o controle **Modo X**. A seleção de **Ponto** irá utilizar um sinal de mais (+) como a forma de ponto padrão. Quando o gráfico é criado, é possível alterar a forma do ponto e alterar o tamanho.

Modo X Para gráficos de linha, deve-se escolher um Modo X para definir o estilo do gráfico de linha. Selecione **Ordem**, **Sobreposição** ou **Como Ler**. Para **Sobreposição** ou **Como Leitura**, deve-se especificar um tamanho de conjunto de dados máximo utilizado para tirar amostra dos n primeiros registros. Caso contrário, o padrão de 2.000 registros será usado.

Intervalo X automático. Selecione para utilizar o intervalo inteiro de valores nos dados ao longo desse eixo. Cancele a seleção para utilizar um subconjunto explícito de valores com base em seus valores **Mín.** e **Máx** especificados. Insira valores ou utilize as setas. Valores automáticos são selecionados por padrão para permitir uma construção rápida do gráfico.

Intervalo Y automático. Selecione para utilizar o intervalo inteiro de valores nos dados ao longo desse eixo. Cancele a seleção para utilizar um subconjunto explícito de valores com base em seus valores **Mín.** e **Máx** especificados. Insira valores ou utilize as setas. Valores automáticos são selecionados por padrão para permitir uma construção rápida do gráfico.

Intervalo Z automático. Apenas quando um gráfico 3D for especificado na guia Gráfico. Selecione para utilizar o intervalo inteiro de valores nos dados ao longo desse eixo. Cancele a seleção para utilizar um subconjunto explícito de valores com base em seus valores **Mín.** e **Máx** especificados. Insira valores ou utilize as setas. Valores automáticos são selecionados por padrão para permitir uma construção rápida do gráfico.

Jitter. Também conhecido como **agitação**, o jitter é útil para gráficos de ponto de um conjunto de dados no qual muitos valores são repetidos. Para ver uma distribuição mais clara de valores, é possível utilizar o jitter para distribuir os pontos aleatoriamente em torno do valor real.

Nota para usuários de versões anteriores do IBM SPSS Modeler: o valor do jitter utilizado em um gráfico utiliza uma métrica diferente nesta liberação do IBM SPSS Modeler. Em versões anteriores, o valor era um número real e agora é uma proporção do tamanho do quadro. Isso significa que os valores de agitação em fluxos antigos são susceptíveis de serem muito grandes. Para esta liberação, quaisquer valores de agitação diferentes de zero serão convertidos para o valor 0,2.

Número máximo de registros para representar. Especifique um método para representar grandes conjuntos de dados. É possível especificar um tamanho máximo de conjunto de dados ou utilizar o padrão de 2.000 registros. O desempenho é aprimorado para grandes conjuntos de dados quando selecionar as opções **Categoria** ou **Amostra**. Como alternativa, é possível escolher criar um gráfico de todos os pontos de dados ao selecionar **Usar todos os dados**, no entanto, deve-se observar que isso pode diminuir drasticamente o desempenho do software.

Nota: quando o modo X é configurado para **Sobreposição** ou **Como Leitura**, essas opções são desativadas e somente os primeiros n registros são utilizados.

- **Categorizar.** Selecione para ativar a categorização quando o conjunto de dados contém mais do que o número de registros especificado. A categorização divide o gráfico em grades finas antes de realmente representar e conta o número de pontos que apareceriam em cada uma das células da grade. No gráfico final, um ponto é representado por célula no centroide da categoria (média de todos os locais de ponto na categoria). O tamanho dos símbolos representados indica o número de pontos nessa região (a menos que você tenha utilizado o tamanho como uma sobreposição). Utilizar o centroide e o tamanho para representar o número de pontos torna o gráfico categorizado uma maneira superior para representar grandes conjuntos de dados porque ele impede a representação em excesso em regiões densas (massas de cor indiferenciadas) e reduz artefatos de símbolo (padrões de densidade artificiais). Os artefatos de símbolo ocorrem quando determinados símbolos (principalmente o símbolo de mais [+]) colidem de uma maneira que produz áreas densas não presentes nos dados brutos.
- **Amostra.** Selecione para obter uma amostra aleatória dos dados de acordo com o número de registros inserido no campo de texto. O padrão é 2.000.

Guia Aparência do Gráfico

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Rótulo X. Aceite o rótulo do eixo x (horizontal) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Y. Aceite o rótulo do eixo y (vertical) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Z. Disponível apenas para gráficos 3D, aceite o rótulo do eixo z gerado automaticamente ou selecione **Customizado** para especificar um rótulo customizado.

Exibir linha de grade. Seleccionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

Utilizando um Gráfico de Plotagem

Os gráficos e multigráficos são essencialmente gráficos de X em função Y . Por exemplo, se você estiver explorando potenciais fraudes em requisições de concessões agrícolas, talvez você queira representar a receita reclamada na requisição com relação à receita estimada por uma rede neural. O uso de uma sobreposição, como tipo de cultura, ilustra se há um relacionamento entre as reclamações (valor ou número) e o tipo de cultura.

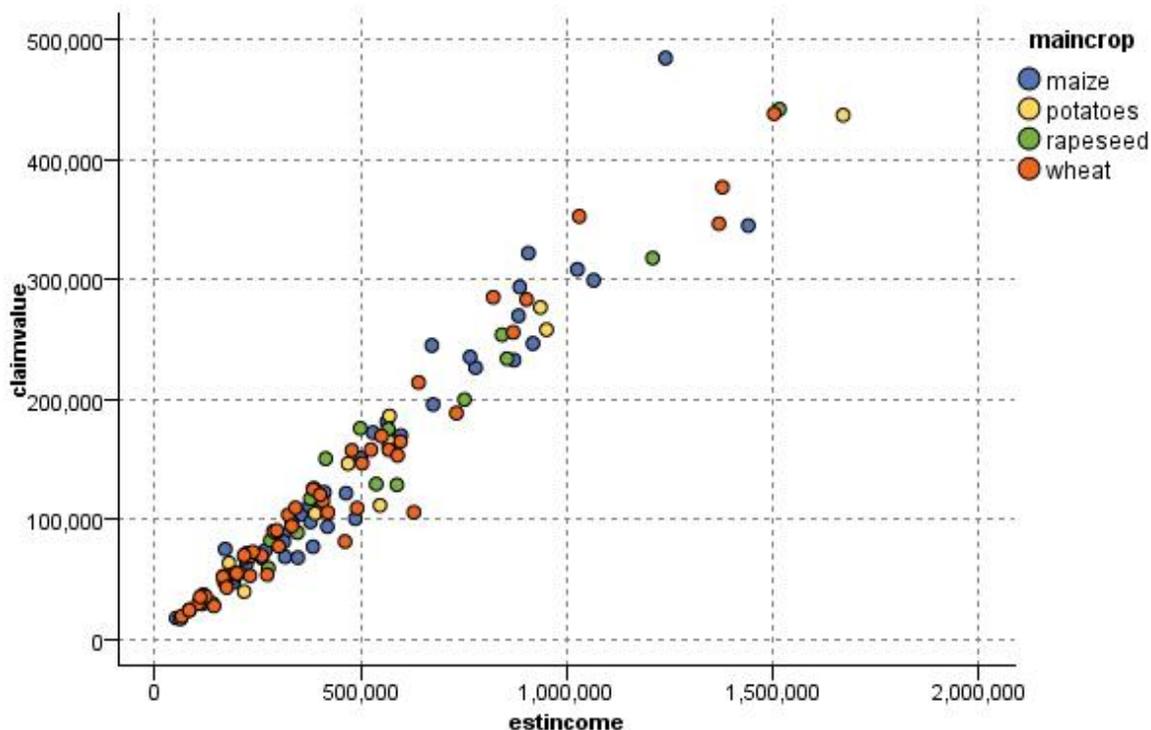


Figura 25. Gráfico do relacionamento entre a receita estimada e o valor da reclamação com o tipo de cultura principal como uma sobreposição

Como gráficos, multigráficos e gráficos de avaliação são bidimensionais exibindo Y em função de X, é fácil interagir com eles ao definir regiões, marcar elementos ou até mesmo desenhar bandas. Também é possível gerar nós para os dados representados por essas regiões, bandas ou elementos. Consulte o tópico “Explorando Gráficos” na página 275 para obter mais informações.

Nó Multigráficos

Um multigráficos é um tipo especial de gráfico que exhibe diversos campos Y em um único campo X. Os campos Y são representados como linhas coloridas, em que cada linha é equivalente a um nó Gráfico com Estilo configurado para **Linha** e o Modo X configurado para **Classificar**. Os multigráficos são úteis quando você tiver dados de sequência de tempo e desejar explorar a flutuação de diversas variáveis ao longo do tempo.

Guia Gráfico de Multigráficos

Campo X. Na lista, selecione o campo a ser exibido no eixo x horizontal.

Campos Y. Selecione um ou mais campos na lista para exibir ao longo do intervalo de valores do campo X. Utilize o botão Seletor de Campo para selecionar diversos campos. Clique no botão Excluir para remover campos da lista.

Sobreposição. Há várias maneiras de ilustrar categorias para os valores de dados. Por exemplo, é possível utilizar uma sobreposição de animação para exibir diversos gráficos para cada valor nos dados. Isso é útil para conjuntos que contêm mais de 10 categorias. Quando utilizada para conjuntos com mais de 15 categorias, é possível observar uma redução no desempenho. Consulte o tópico “Estética, Sobreposições, Painéis e Animação” na página 194 para obter mais informações.

Normalizar. Selecione para escalar todos os valores Y para o intervalo de 0 a 1 para exibir no gráfico. A normalização ajuda a explorar o relacionamento entre as linhas que podem de outra forma estar obscuras devido a diferenças no intervalo de valores para cada série, e é recomendada quando representar várias linhas no mesmo gráfico ou quando comparar os gráficos em painéis lado a lado. (A normalização não é necessária quando todos os valores de dados estiverem dentro de um intervalo semelhante).

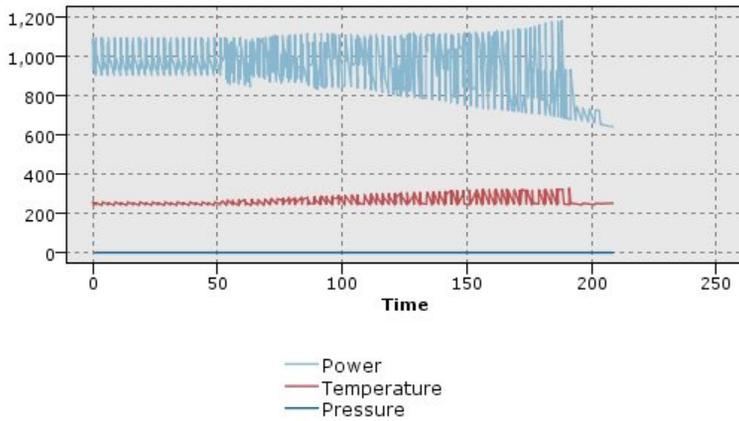


Figura 26. Multigráficos padrão mostrando oscilações de uma usina de energia ao longo do tempo (note que sem a normalização, o gráfico Pressão é impossível de se ver)

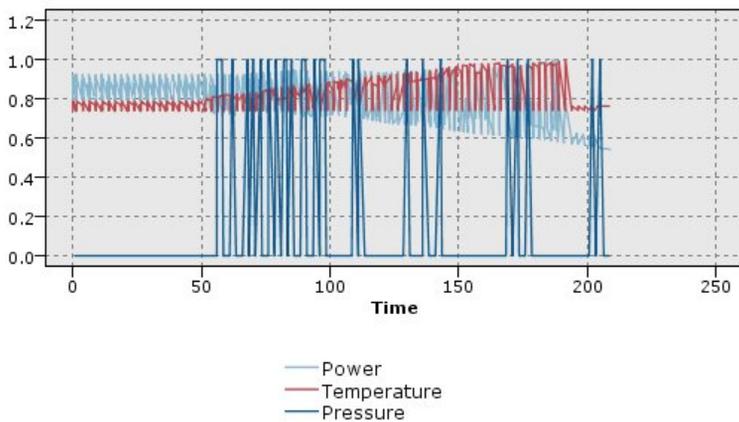


Figura 27. Multigráficos normalizados mostrando um gráfico para Pressão

Função de sobreposição. Selecione para especificar uma função conhecida para comparar com os valores reais. Por exemplo, para comparar os valores reais versus valores preditos, é possível criar um gráfico da função $y = x$ como uma sobreposição. Especifique uma função para $y =$ na caixa de texto. A função padrão é $y = x$, mas é possível especificar qualquer tipo de função, como uma função quadrática ou uma expressão arbitrária, em termos de x .

Nota: as funções de sobreposição não estão disponíveis para um gráfico de painel ou de animação.

Quando o número de registros for maior que. Especifique um método para representar grandes conjuntos de dados. É possível especificar um tamanho máximo de conjunto de dados ou utilizar o padrão de 2.000 pontos. O desempenho é aprimorado para grandes conjuntos de dados quando

selecionar as opções **Categoria** ou **Amostra**. Como alternativa, é possível escolher criar um gráfico de todos os pontos de dados ao selecionar **Usar todos os dados**, no entanto, deve-se observar que isso pode diminuir drasticamente o desempenho do software.

Nota: quando o modo X é configurado para **Sobreposição** ou **Como Leitura**, essas opções são desativadas e somente os primeiros n registros são utilizados.

- **Categorizar.** Selecione para ativar a categorização quando o conjunto de dados contém mais do que o número de registros especificado. A categorização divide o gráfico em grades finas antes de realmente representar e conta o número de conexões que apareceriam em cada uma das células da grade. No gráfico final, uma conexão é utilizada por célula no centroide da categoria (média de todos os pontos de conexão na categoria).
- **Amostra.** Selecione para obter uma amostra aleatória dos dados de acordo com o número de registros especificado.

Guia Aparência de Multigráficos

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Rótulo X. Aceite o rótulo do eixo x (horizontal) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Y. Aceite o rótulo do eixo y (vertical) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Exibir linha de grade. Selecionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

Utilizando um Gráfico Multigráficos

Os gráficos e multigráficos são essencialmente gráficos de X em função Y . Por exemplo, se você estiver explorando potenciais fraudes em requisições de concessões agrícolas, talvez você queira representar a receita reclamada na requisição com relação à receita estimada por uma rede neural. O uso de uma sobreposição, como tipo de cultura, ilustra se há um relacionamento entre as reclamações (valor ou número) e o tipo de cultura.

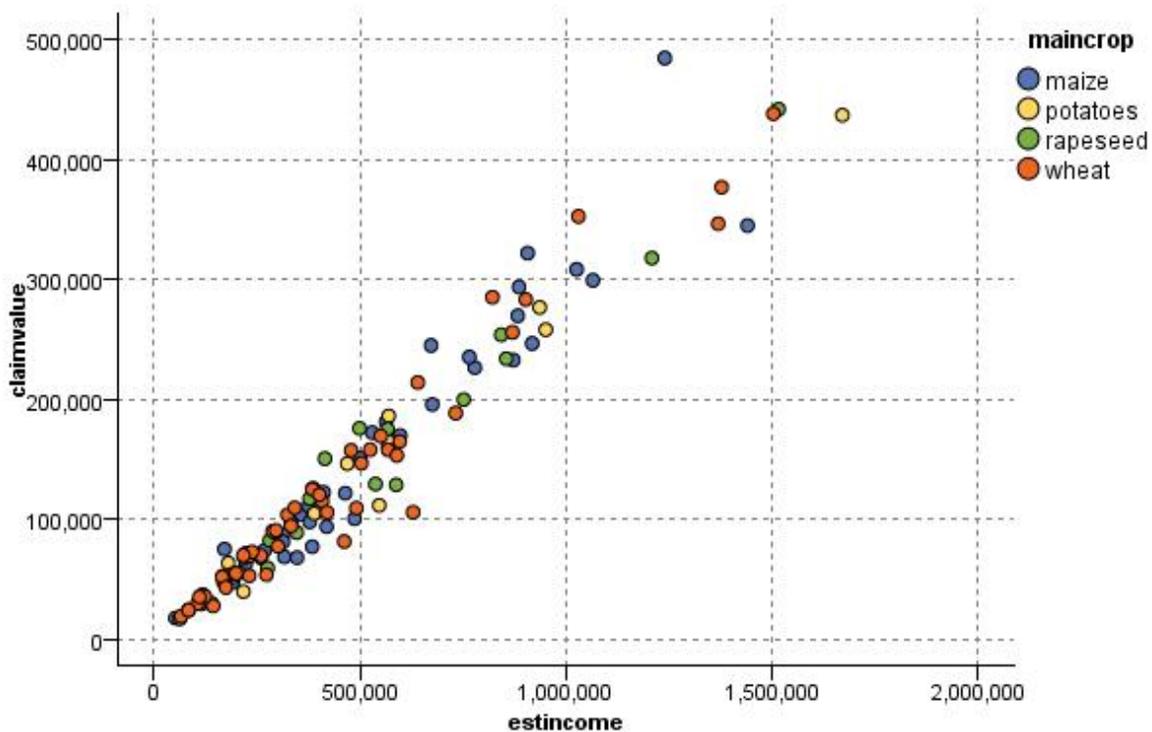


Figura 28. Gráfico do relacionamento entre a receita estimada e o valor da reclamação com o tipo de cultura principal como uma sobreposição

Como gráficos, multigráficos e gráficos de avaliação são bidimensionais exibindo Y em função de X, é fácil interagir com eles ao definir regiões, marcar elementos ou até mesmo desenhar bandas. Também é possível gerar nós para os dados representados por essas regiões, bandas ou elementos. Consulte o tópico “Explorando Gráficos” na página 275 para obter mais informações.

Nó Gráfico de Tempo

Os nós Gráfico de Tempo permitem visualizar uma ou mais séries temporais representadas ao longo do tempo. As séries que você representar devem conter valores numéricos e supõe-se que elas ocorram em um intervalo de tempo no qual os períodos são uniformes. Um nó Intervalos de Tempo geralmente é utilizado antes de um nó Gráfico de Tempo para criar um campo *TimeLabel*, que é utilizado por padrão para rotular o eixo x nos gráficos. Consulte o tópico “Nó Intervalos de Tempo” na página 176 para obter mais informações.

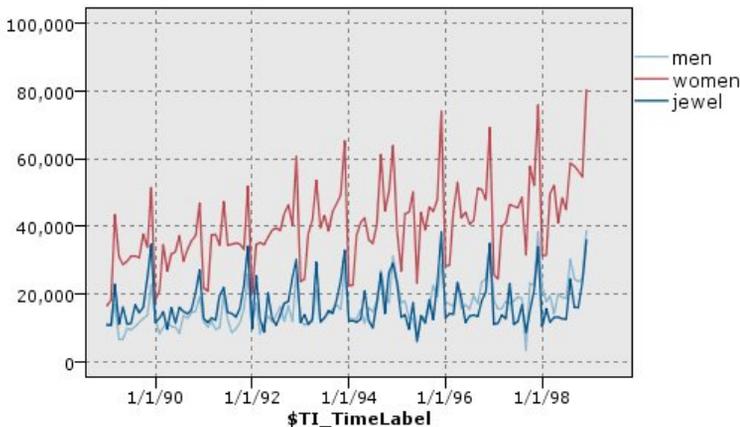


Figura 29. Representando vendas de roupas e jóias masculinas e femininas ao longo do tempo

Criando Intervenções e Eventos

É possível criar campos de Evento e de Intervenção a partir do gráfico de tempo ao gerar um nó Derivar (flag ou nominal) a partir dos menus de contexto. Por exemplo, é possível criar um campo de evento no caso de uma greve ferroviária, em que o estado da unidade será True se o evento ocorreu e False se não ocorreu. Para um campo de Intervenção, no caso de um aumento de preços, por exemplo, é possível usar um nó derivar contagem para identificar a data do aumento, com 0 para o preço antigo e 1 para o novo preço. Consulte o tópico “Derivar Nó” na página 148 para obter mais informações.

Guia Gráfico de Tempo

Gráfico. Fornece uma opção de como representar dados de série temporal.

- **Séries selecionadas.** Representa os valores para a série de tempo selecionada. Se selecionar essa opção quando representar os intervalos de confiança, desmarque a caixa de seleção **Normalizar**.
- **Modelos de Séries Temporais selecionados.** Utilizado em conjunto com um modelo de Série Temporal, esta opção representa todos os campos relacionados (valores reais e preditos, bem como os intervalos de confiança) para uma ou mais séries de tempo selecionadas. Esta opção desativa algumas outras opções na caixa de diálogo. Essa é a opção preferencial se representar os intervalos de confiança.

Série. Selecione um ou mais campos com dados de série temporal que deseja representar. Os dados devem ser numéricos.

Rótulo do eixo X. Escolha o rótulo padrão ou um campo único a ser utilizado como o rótulo para o eixo x nas representações. Se escolher Padrão, o sistema usará o campo TimeLabel criado em um nó Intervalos de Tempo anterior ou números inteiros sequenciais se não houver nenhum nó de Intervalos de Tempo. Consulte o tópico “Nó Intervalos de Tempo” na página 176 para obter mais informações.

Exibir séries em painéis separados. Especifica se cada série é exibida em um painel separado. Como alternativa, se você não escolher exibir em painéis, todas as séries temporais serão representadas no mesmo gráfico e suavizadores não estarão disponíveis. Ao representar todas as séries temporais no mesmo gráfico, cada série será representada por uma cor diferente.

Normalizar. Selecione para escalar todos os valores Y para o intervalo de 0 a 1 para exibir no gráfico. A normalização ajuda a explorar o relacionamento entre as linhas que podem de outra forma estar obscuras devido a diferenças no intervalo de valores para cada série, e é recomendada quando representar várias linhas no mesmo gráfico ou quando comparar os gráficos em painéis lado a lado. (A normalização não é necessária quando todos os valores de dados estiverem dentro de um intervalo semelhante).

Exibição. Selecione um ou mais elementos para exibir em seu gráfico. É possível escolher entre linhas, pontos e (LOESS) suavizadores. Os suavizadores estarão disponíveis apenas se as séries forem exibidas em painéis separados. Por padrão, o elemento de linha é selecionado. Assegure-se de selecionar pelo menos um elemento de gráfico antes de executar o nó de gráfico, caso contrário, o sistema retornará um erro informando que nada foi selecionado para representar.

Limitar registros. Selecione esta opção se desejar limitar o número de registros representados. Especifique o número de registros, lidos a partir do início de seu arquivo de dados, que será representado na opção **Número máximo de registros para representar**. Por padrão, esse número é configurado para 2.000. Se desejar representar os últimos n registros em seu arquivo de dados, será possível usar um nó Ordem antes desse nó para organizar os registros em ordem decrescente de tempo.

Guia Aparência do Gráfico de Tempo

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Rótulo X. Aceite o rótulo do eixo x (horizontal) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Y. Aceite o rótulo do eixo y (vertical) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Exibir linha de grade. Selecionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

Layout. Para gráficos de tempo somente, é possível especificar se os valores de tempo são representados ao longo do eixo horizontal ou vertical.

Utilizando um Gráfico de Plotagem de Tempo

Após criar um gráfico de plotagem de tempo, há várias opções para ajustar a exibição do gráfico e gerar nós para análise adicional. Consulte o tópico “Explorando Gráficos” na página 275 para obter mais informações.

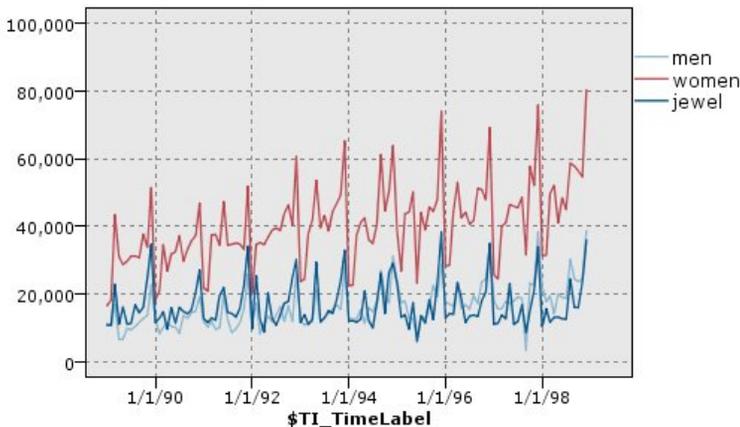


Figura 30. Representando vendas de roupas e jóias masculinas e femininas ao longo do tempo

Após ter criado um gráfico de tempo, definido bandas e analisado os resultados, será possível utilizar as opções no menu Gerar e no menu de contexto para criar os nós Seleção ou Derivar. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

Nó Distribuição

Um gráfico ou uma tabela de distribuição mostra a ocorrência de valores simbólicos (não numéricos), tais como tipo de hipoteca ou sexo, em um conjunto de dados. Um uso típico do nó de Distribuição é mostrar os desbalanceamentos nos dados que podem ser corrigidos utilizando um nó Balanceamento antes de criar um modelo. É possível gerar automaticamente um nó Balanceamento utilizando o menu Gerar na janela de gráfico ou de tabela de distribuição.

Também é possível utilizar o nó Gráfico para produzir gráficos de barras de contagens. No entanto, há mais opções a serem escolhidas a partir desse nó. Consulte o tópico “Tipos de Visualizações de Gráfico Integradas Disponíveis” na página 204 para obter mais informações.

Nota: para mostrar a ocorrência de valores numéricos, deve-se utilizar um nó Histograma.

Guia Gráfico de Distribuição

Gráfico. Selecione o tipo de distribuição. Selecione **Campos selecionados** para mostrar a distribuição do campo selecionado. Selecione **Todos os flags (valores verdadeiros)** para mostrar a distribuição de valores verdadeiros para os campos de flag no conjunto de dados.

Campo. Selecione um campo nominal ou de flag para o qual mostrar a distribuição de valores. Somente os campos que não tiverem sido configurados explicitamente como numéricos aparecem na lista.

Sobreposição. Selecione um campo nominal ou de flag para uso como uma sobreposição de cores, ilustrando a distribuição de seus valores dentro de cada valor do campo especificado. Por exemplo, é possível utilizar uma resposta de campanha de marketing (*pep*) como uma sobreposição para número de filhos (*children*) para ilustrar responsividade por tamanho de família. Consulte o tópico “Estética, Sobreposições, Painéis e Animação” na página 194 para obter mais informações.

Normalizar por cor. Selecione as barras de escala de modo que todas as barras ocupem toda a largura do gráfico. Os valores de sobreposição iguais à proporção de cada barra facilitam as comparações entre as categorias.

Ordenar. Selecione o método utilizado para exibir valores no gráfico de distribuição. Selecione **Alfabética** para usar ordem alfabética ou **Por contagem** para listar valores em ordem decrescente de ocorrência.

Escala proporcional. Selecione para escalar a distribuição de valores de forma que o valor com a maior contagem preencha a largura total do gráfico. Todas as outras barras são escaladas com relação a este valor. Cancelar a seleção desta opção escala as barras de acordo com as contagens totais de cada valor.

Guia Aparência da Distribuição

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Rótulo X. Aceite o rótulo do eixo x (horizontal) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Y. Aceite o rótulo do eixo y (vertical) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Exibir linha de grade. Selecionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

Usando um Nó de Distribuição

Os nós de distribuição são utilizados para mostrar a distribuição de valores simbólicos em um conjunto de dados. Eles são usados frequentemente antes dos nós de manipulação para explorar os dados e corrigir quaisquer desbalanceamentos. Por exemplo, se instâncias de respondentes sem filhos ocorrerem com uma frequência muito maior que outros tipos de respondentes, talvez você queira reduzir essas instâncias de modo que uma regra mais útil possa ser gerada em operações de mineração de dados posteriores. Um nó Distribuição ajuda a examinar e a tomar decisões sobre tais desbalanceamentos.

O nó Distribuição é incomum pelo fato de produzir um gráfico e uma tabela para analisar seus dados.

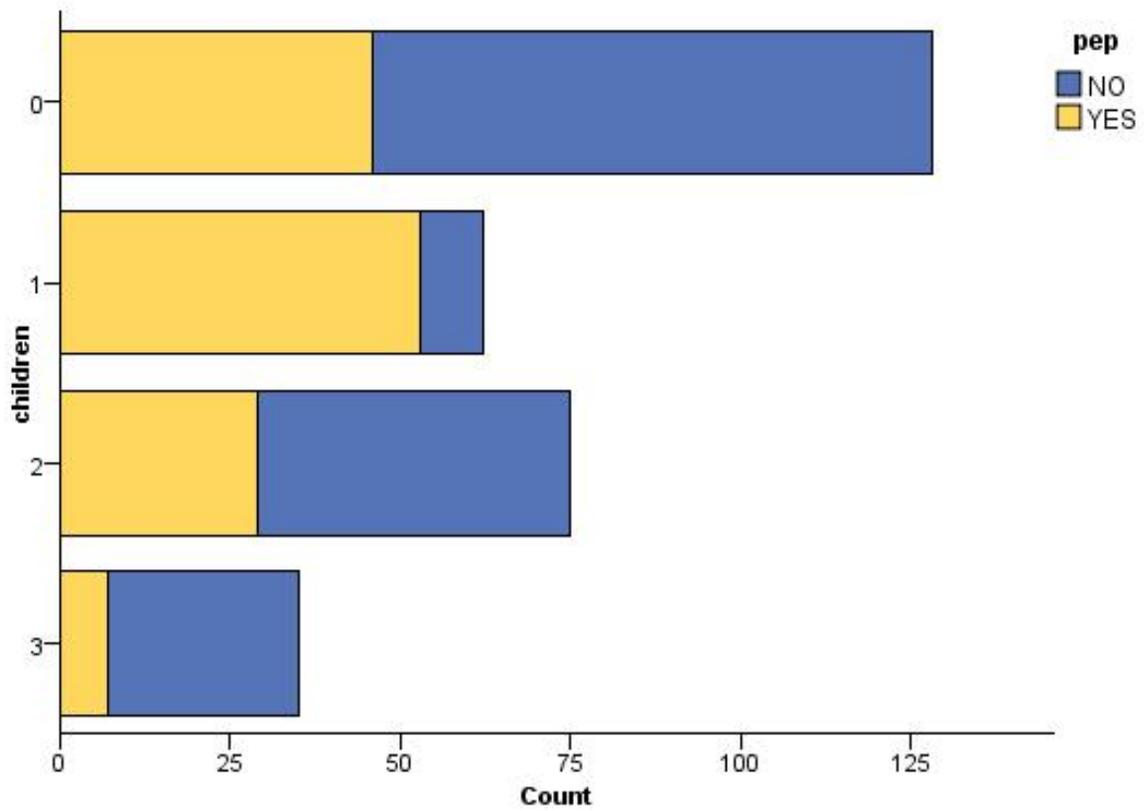


Figura 31. Gráfico de distribuição mostrando o número de pessoas com ou sem filhos que responderam a uma campanha de marketing

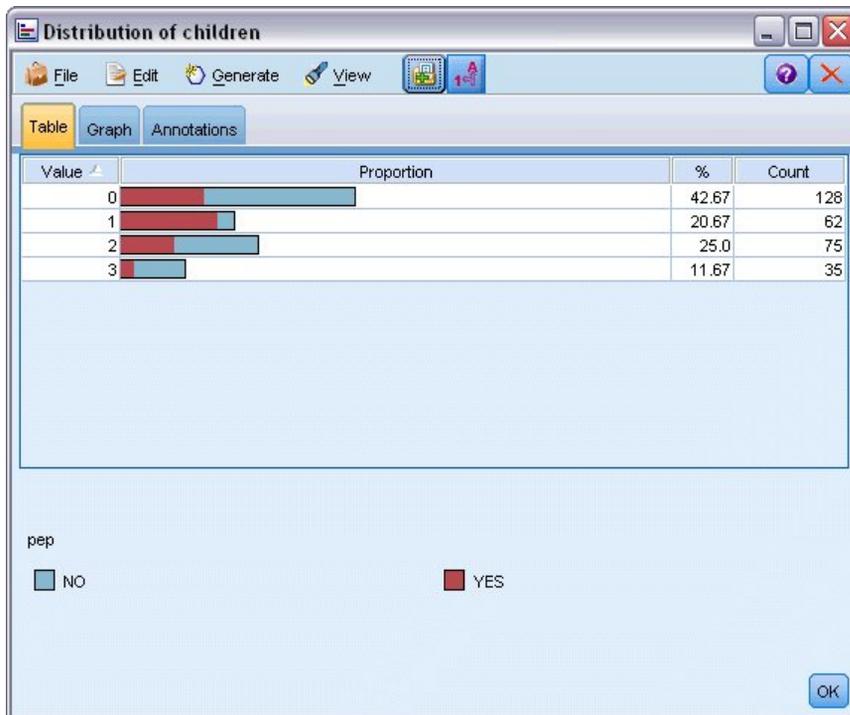


Figura 32. Tabela de distribuição mostrando a proporção de pessoas com ou sem filhos que responderam a uma campanha de marketing

Após ter criado uma tabela e um gráfico de distribuição e examinado os resultados, será possível utilizar as opções de menus para agrupar valores, copiar valores e gerar um número de nós para a preparação de dados. Além disso, é possível copiar ou exportar as informações do gráfico e da tabela para uso em outros aplicativos, como o MS Word ou MS PowerPoint. Consulte o tópico “Imprimindo, Salvando, Copiando e Exportando Gráficos” na página 298 para obter mais informações.

Para Selecionar e Copiar Valores a partir de uma Tabela de Distribuição

1. Clique e mantenha pressionado o botão do mouse e arraste-o sobre as linhas para selecionar um conjunto de valores. Também é possível utilizar o menu Editar para **Selecionar Todos os Valores**.
2. No menu Editar, escolha **Copiar Tabela** ou **Copiar Tabela (incluindo nomes de campo)**.
3. Cole na área de transferência ou no aplicativo desejado.

Nota: as barras não são copiadas diretamente. Ao invés disso, os valores de tabela são copiados. Isso significa que valores sobrepostos não serão exibidos na tabela copiada.

Para Agrupar Valores a partir de uma Tabela de Distribuição

1. Selecione os valores para agrupamento utilizando o método Ctrl+clique.
2. No menu Editar, escolha **Agrupar**.

Nota: quando agrupar e desagrupar valores, o gráfico na guia Gráfico é redesenhado automaticamente para mostrar as mudanças.

Também é possível:

- Desagrupar valores ao selecionar o nome do grupo na lista de distribuição e escolher **Desagrupar** no menu Editar.
- Editar grupos ao selecionar o nome do grupo na lista de distribuição e escolher **Editar grupo** no menu Editar. Isso abre uma caixa de diálogo na qual os valores podem ser mudados para e do grupo.

Opções de Menu Gerar

É possível utilizar as opções no menu Gerar para selecionar um subconjunto de dados, derivar um campo de flag, reagrupar os valores, reclassifique valores ou balancear os dados a partir de um gráfico ou tabela. Essas operações geram um nó de preparação de dados e o coloca na tela de fluxo. Para utilizar o nó gerado, conecte-o a um fluxo existente. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

Nó Histograma

Os nós Histogramas mostram a ocorrência de valores para campos numéricos. Eles são normalmente utilizados para explorar os dados antes de manipulações e construções de modelo. Semelhantes ao nó Distribuição, os nós Histogramas geralmente são usados para revelar desbalanceamentos nos dados. Embora o nó do Gráfico também possa ser usado para produzir um histograma, há mais opções para escolher a partir desse nó. Consulte o tópico “Tipos de Visualizações de Gráfico Integradas Disponíveis” na página 204 para obter mais informações.

Nota: para mostrar a ocorrência de valores para campos simbólicos, deve-se utilizar um nó de Distribuição.

Guia Gráfico de Histograma

Campo. Selecione um campo numérico para o qual mostrar a distribuição de valores. Apenas os campos que não tiverem sido definidos explicitamente como simbólicos (categóricos) serão listados.

Sobreposição. Selecione um campo simbólico para mostrar as categorias de valores para o campo especificado. Selecionar um campo de sobreposição converte o histograma em um gráfico empilhado com cores utilizadas para representar diferentes categorias do campo de sobreposição. Utilizando o nó Histograma, há três tipos de sobreposições: cor, painel e animação. Consulte o tópico “Estética, Sobreposições, Painéis e Animação” na página 194 para obter mais informações.

Guia Opções do Histograma

Intervalo X automático. Selecione para utilizar o intervalo inteiro de valores nos dados ao longo desse eixo. Cancele a seleção para utilizar um subconjunto explícito de valores com base em seus valores **Mín.** e **Máx** especificados. Insira valores ou utilize as setas. Valores automáticos são selecionados por padrão para permitir uma construção rápida do gráfico.

Categorias. Selecione **Por número** ou **Por largura**.

- Selecione **Por número** para exibir um número fixo de barras cuja largura depende do intervalo e do número de categorias especificadas. Indique o número de categorias a serem utilizadas no gráfico na opção **Nº de categorias**. Utilize as setas para ajustar o número.
- Selecione **Por largura** para criar um gráfico com barras de largura fixa. O número de categorias depende da largura especificada e do intervalo de valores. Indica a largura das barras na opção **Largura da categoria**.

Normalizar por cor. Selecione para ajustar todas as barras para a mesma altura, exibindo valores sobrepostos como uma porcentagem do total de casos em cada barra.

Exibir curva normal. Selecione para incluir uma curva normal no gráfico mostrando o média e a variância dos dados.

Separar bandas para cada cor. Selecione para exibir cada valor sobreposto como uma banda separada no gráfico.

Guia Aparência do Histograma

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Rótulo X. Aceite o rótulo do eixo x (horizontal) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Y. Aceite o rótulo do eixo y (vertical) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Exibir linha de grade. Selecionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

Utilizando Histogramas

Os histogramas mostram a distribuição de valores em um campo numérico cujos valores variam ao longo do eixo x . Os histogramas operam de forma semelhante aos gráficos de coleções. As coleções mostram a distribuição de valores para um campo numérico *com relação aos valores de outro*, ao invés da ocorrência de valores de um campo único.

Após ter criado um gráfico, é possível examinar os resultados e definir bandas para dividir valores ao longo do eixo x ou definir regiões. Também é possível marcar elementos dentro do gráfico. Consulte o tópico “Explorando Gráficos” na página 275 para obter mais informações.

É possível utilizar as opções no menu Gerar para criar os nós Balanceamento, Seleção ou Derivar utilizando os dados no gráfico ou, mais especificamente, dentro de bandas, regiões ou elementos marcados. Este tipo de gráfico é frequentemente utilizado antes dos nós de manipulação para explorar os dados e corrigir quaisquer desbalanceamentos ao gerar um nó Balanceamento a partir do gráfico para uso no fluxo. Também é possível gerar um nó Derivar Flag para incluir um campo mostrando em qual banda cada registro cai ou um nó Seleção para selecionar todos os registros dentro de um conjunto ou intervalo de valores específico. Essas operações ajudam a focar em um subconjunto específico de dados para exploração adicional. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

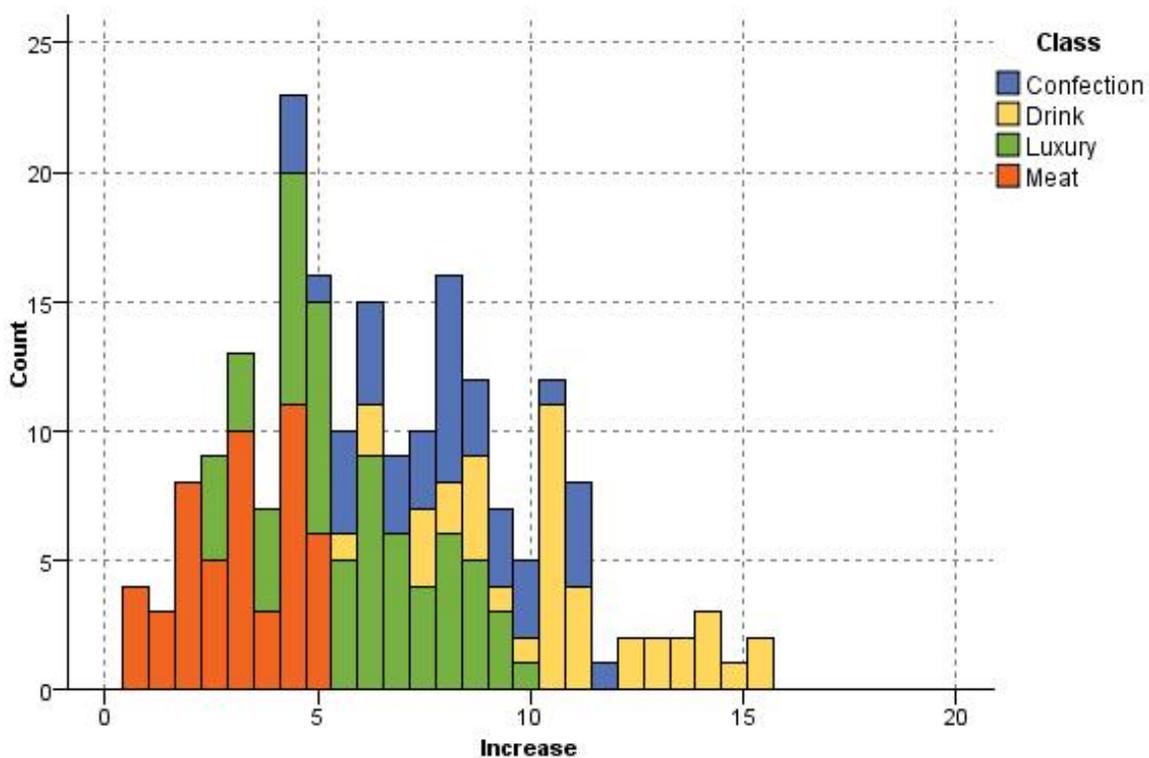


Figura 33. Histograma mostrando a distribuição do aumento de compras por categoria devido à promoção

Nó Coleção

As coleções são semelhantes aos histogramas, exceto que as coleções mostram a distribuição de valores para um campo numérico com relação aos valores de outro, ao invés da ocorrência de valores para um campo único. Uma coleção é útil para ilustrar uma variável ou um campo cujos valores se alteram ao longo do tempo. Usando um gráfico 3D, também é possível incluir um eixo simbólico exibindo distribuições por categoria. Duas Coleções dimensionais são mostradas como gráficos de barras empilhadas, com sobreposições quando utilizadas. Consulte o tópico “Estética, Sobreposições, Painéis e Animação” na página 194 para obter mais informações.

Guia Gráfico de Coleção

Coletar. Selecione um campo cujos valores serão coletados e exibidos sobre o intervalo de valores para o campo especificado em **Sobre**. Somente os campos que não foram definidos como simbólicos são listados.

Sobre. Selecione um campo cujos valores serão utilizados para exibir o campo especificado em **Coletar**.

Por. Ativada ao criar um gráfico 3D, esta opção permite selecionar um campo nominal ou de flag usado para exibir o campo de coleta por categorias.

Operação. Selecione o que cada barra no gráfico de coleção representa. As opções incluem **Soma**, **Média**, **Máx.**, **Mín.** e **Desvio Padrão**.

Sobreposição. Selecione um campo simbólico para mostrar as categorias de valores para o campo selecionado. Selecionar um campo de sobreposição converte a coleção e cria diversas barras de cores variadas para cada categoria. Este nó tem três tipos de sobreposições: cor, painel e animação. Consulte o tópico “Estética, Sobreposições, Painéis e Animação” na página 194 para obter mais informações.

Guia Opções de Coleção

Intervalo X automático. Selecione para utilizar o intervalo inteiro de valores nos dados ao longo desse eixo. Cancele a seleção para utilizar um subconjunto explícito de valores com base em seus valores **Mín.** e **Máx** especificados. Insira valores ou utilize as setas. Valores automáticos são selecionados por padrão para permitir uma construção rápida do gráfico.

Categorias. Selecione **Por número** ou **Por largura**.

- Selecione **Por número** para exibir um número fixo de barras cuja largura depende do intervalo e do número de categorias especificadas. Indique o número de categorias a serem utilizadas no gráfico na opção **Nº de categorias**. Utilize as setas para ajustar o número.
- Selecione **Por largura** para criar um gráfico com barras de largura fixa. O número de categorias depende da largura especificada e do intervalo de valores. Indica a largura das barras na opção **Largura da categoria**.

Guia Aparência da Coleção

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Sobre rótulo. Aceite o rótulo gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo de coleção. Aceite o rótulo gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Por rótulo. Aceite o rótulo gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Exibir linha de grade. Selecionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

O exemplo a seguir mostra onde as opções de aparência são colocadas em uma versão 3D do gráfico.

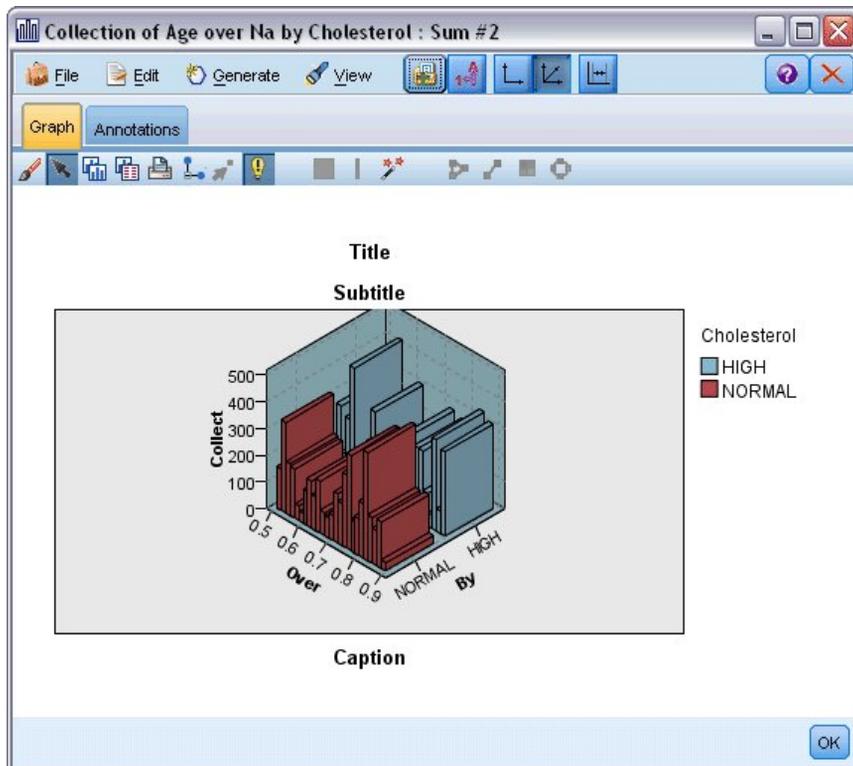


Figura 34. Posição das opções de aparência do gráfico no gráfico Coleção em 3D

Utilizando um Gráfico de Coleção

As coleções mostram a distribuição de valores para um campo numérico *com relação aos valores de outro*, ao invés da ocorrência de valores de um campo único. Os histogramas operam de forma semelhante aos gráficos de coleções. Os histogramas mostram a distribuição de valores em um campo numérico cujos valores variam ao longo do eixo x .

Após ter criado um gráfico, é possível examinar os resultados e definir bandas para dividir valores ao longo do eixo x ou definir regiões. Também é possível marcar elementos dentro do gráfico. Consulte o tópico “Explorando Gráficos” na página 275 para obter mais informações.

É possível utilizar as opções no menu Gerar para criar os nós Balanceamento, Seleção ou Derivar utilizando os dados no gráfico ou, mais especificamente, dentro de bandas, regiões ou elementos marcados. Este tipo de gráfico é frequentemente utilizado antes dos nós de manipulação para explorar os dados e corrigir quaisquer desbalanceamentos ao gerar um nó Balanceamento a partir do gráfico para uso no fluxo. Também é possível gerar um nó Derivar Flag para incluir um campo mostrando em qual banda cada registro cai ou um nó Seleção para selecionar todos os registros dentro de um conjunto ou intervalo de valores específico. Essas operações ajudam a focar em um subconjunto específico de dados para exploração adicional. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

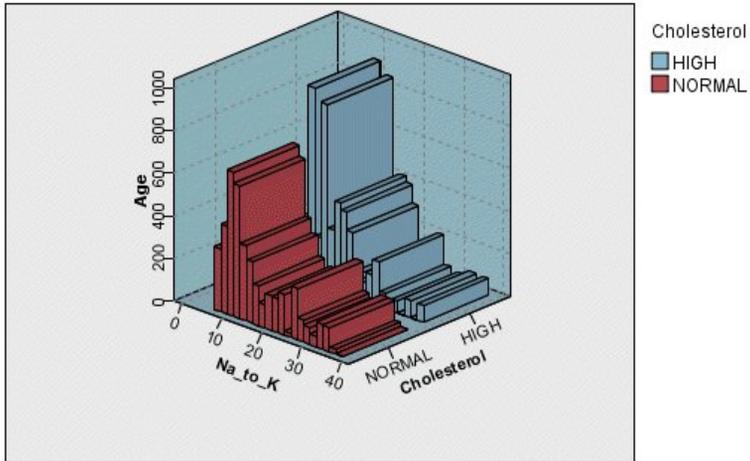


Figura 35. Gráfico de coleção 3D mostrando a soma de Na_to_K de Age para os níveis de colesterol elevado e normal

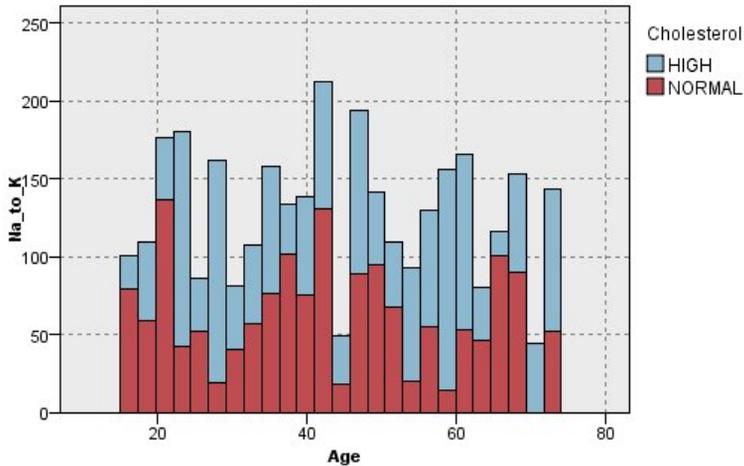


Figura 36. Gráfico de coleção sem o eixo z exibido, mas com Colesterol como sobreposição de cor

Nó da Web

Os nós da web mostram a intensidade dos relacionamentos entre valores de dois ou mais campos simbólicos. O gráfico exibe conexões utilizando tipos variados de linhas para indicar a intensidade da conexão. É possível utilizar um nó Web, por exemplo, para explorar o relacionamento entre a compra de vários itens em um site de e-commerce ou em uma loja de varejo tradicional.

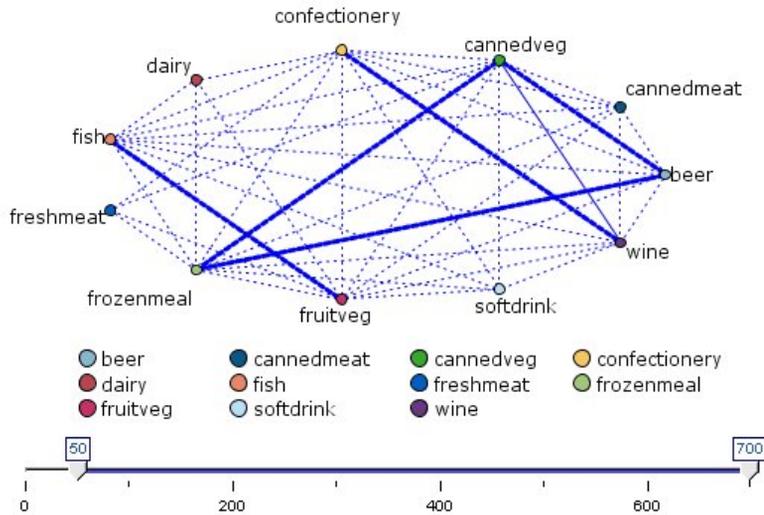


Figura 37. Gráfico da web mostrando os relacionamentos entre a compra de itens da loja

Webs Direcionadas

Os nós Web Direcionada são semelhantes aos nós Web por mostrarem a intensidade dos relacionamentos entre os campos simbólicos. No entanto, gráficos da web direcionada mostram conexões somente de um ou mais campos De para um campo Para único. As conexões são unidirecionais por terem somente uma direção.

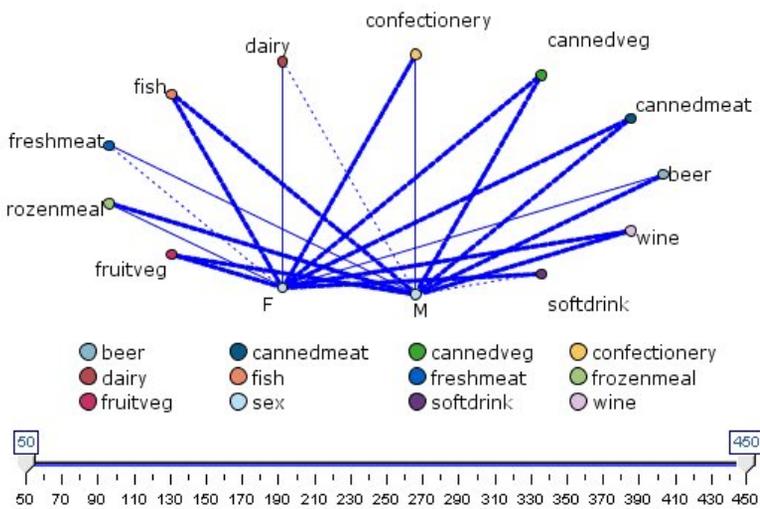


Figura 38. Gráfico da web direcionada mostrando o relacionamento entre a compra de itens da loja e o sexo

Assim como os nós Web, o gráfico exibe conexões utilizando tipos variados de linhas para indicar a intensidade da conexão. É possível utilizar um nó Web Direcionada, por exemplo, para explorar os relacionamentos entre sexo e uma tendência para determinados itens de compra.

Guia Gráfico da Web

Web. Selecione para criar um gráfico da web ilustrando a intensidade dos relacionamentos entre todos os campos especificados.

Web direcionada. Selecione para criar um gráfico da web direcional ilustrando a intensidade dos relacionamentos entre diversos campos e os valores de um campo, como sexo ou religião. Quando essa opção é selecionada, um campo Para é ativado e o controle Campos abaixo é renomeado como Campos De para maior clareza.

Campo Para (Web direcionada somente). Selecione um campo de sinalização ou nominal usado para uma web direcionada. Somente os campos que não tiverem sido configurados explicitamente como numéricos são listados.

Campos/Campos De. Selecione campos para criar um gráfico da web. Somente os campos que não tiverem sido configurados explicitamente como numéricos são listados. Utilize o botão Seletor de Campo para selecionar diversos campos ou selecione campos por tipo.

Nota: para uma web direcionada, esse controle é utilizado para selecionar campos De.

Mostrar flags verdadeiros somente. Selecione para exibir somente flags verdadeiros para um campo de sinalização. Essa opção simplifica a exibição da web e é geralmente utilizada para dados quando a ocorrência de valores positivos for de especial importância.

Valores de linha são. Selecione um tipo de limite na lista suspensa.

- **Absoluto** configura os limites com base no número de registros que possuem cada par de valores.
- **Porcentagens gerais** mostra o número absoluto de casos representados pelo link como uma proporção de todas as ocorrências de cada par de valores representados no gráfico da web.
- **Porcentagens de campo/valor menores e Porcentagens de campo/valor maiores** indicam qual campo/valor utilizar para avaliar as porcentagens. Por exemplo, suponha que 100 registros possuam o valor *drugY* para o campo *Drug* e apenas 10 têm o valor *LOW* para o campo *BP*. Se sete registros tiverem os valores *drugY* e *LOW*, esta porcentagem será 70% ou 7%, dependendo de qual campo estiver referenciando, menor (*BP*) ou maior (*Drug*).

Nota: para gráficos da web direcionados, as terceira e quarta opções acima não estão disponíveis. Ao invés disso, é possível selecionar **Porcentagem do campo/valor "Para"** e **Porcentagem do campo/valor "De"**.

Ligações fortes são mais pesadas. Selecionado por padrão, este é o modo padrão de visualizar ligações entre os campos.

Ligações fracas são mais pesadas. Selecione para reverter o significado das ligações exibidas em linhas em negrito. Essa opção é frequentemente utilizada para detecção de fraude ou exame de valores discrepantes.

Guia Opções da Web

A guia Opções para os nós da Web contém um número de opções adicionais para customizar o gráfico de saída.

Número de Ligações. As opções a seguir são utilizadas para controlar o número de ligações exibidas no gráfico de saída. Algumas dessas opções, como **Ligações fracas acima** e **Ligações fortes acima**, também estão disponíveis na janela do gráfico de saída. Também é possível utilizar uma régua de controle no gráfico final para ajustar o número de ligações exibidas.

- **Número máximo de ligações para exibir.** Especifique um número indicando o número máximo de ligações para mostrar no gráfico de saída. Utilize as setas para ajustar o valor.
- **Mostrar somente ligações acima.** Especifique um número indicando o valor mínimo para o qual mostrar uma conexão na web. Utilize as setas para ajustar o valor.

- **Mostrar todas as ligações.** Especifique para exibir todas as ligações, independentemente dos valores mínimos e máximos. A seleção dessa opção pode aumentar o tempo de processamento se houver um grande número de campos.

Descartar se houver pouquíssimos registros. Selecione para ignorar as conexões que forem suportadas por pouquíssimos registros. Configure o limite para essa opção ao inserir um número em **Mín. de registros/linhas**.

Descartar se houver registros em excesso. Selecione para ignorar conexões fortemente suportadas. Insira um número em **linhas/registros máx.**

Ligações fracas abaixo. Especifique um número indicando o limite para conexões fracas (linhas pontilhadas) e conexões regulares (linhas normais). Todas as conexões abaixo deste valor são consideradas fracas.

Ligações fortes acima. Especifique um limite para conexões fortes (linhas intensas) e conexões regulares (linhas normais). Todas as conexões acima desse valor são consideradas fortes.

Tamanho da Ligação. Especifique opções para controlar o tamanho das ligações:

- **Tamanho da ligação varia continuamente.** Selecione para exibir um intervalo de tamanhos de ligação que refletem a variação com relação às intensidades baseadas nos valores de dados reais.
- **Tamanho da ligação mostra as categorias forte/normal/fraca.** Selecione para exibir três intensidades de conexões--forte, normal e fraca. Os pontos de corte para essas categorias podem ser especificados acima e também no gráfico final.

Exibição da Web. Selecione um tipo de exibição da web:

- **Layout de círculo.** Selecione para utilizar a exibição da web padrão.
- **Layout de rede.** Selecione para utilizar um algoritmo para agrupar as ligações mais fortes. Isso é destinado a destacar ligações fortes utilizando diferenciação espacial e também linhas ponderadas.
- **Layout Direcionado.** Selecione para criar uma exibição da web direcionada que utiliza a seleção **Para Campo** a partir da guia Gráfico como o foco para a direção.
- **Layout da Grade.** Selecione para criar uma exibição da web que esteja organizada em um padrão de grade regularmente espaçado.

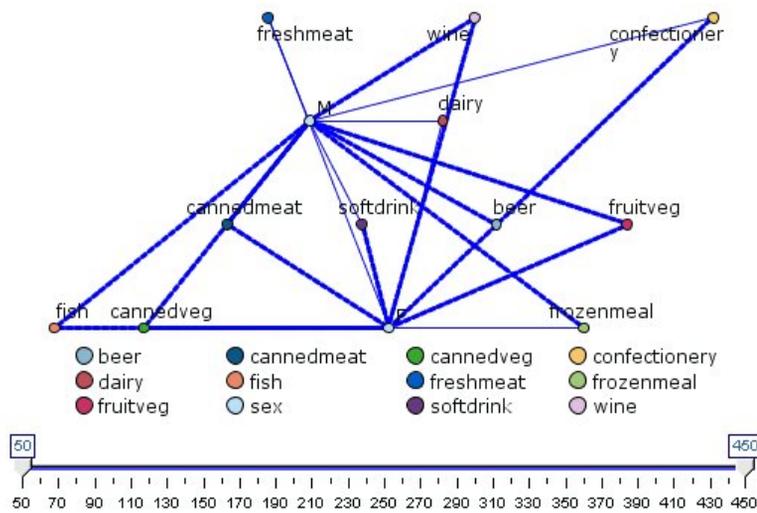


Figura 39. Gráficos da web mostrando conexões fortes de frozenmeal e cannedveg para outros itens de uma mercearia

Guia Aparência da Web

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Mostrar legenda. É possível especificar se a legenda é exibida. Para gráficos com um número grande de campos, ocultar a legenda pode melhorar a aparência do gráfico.

Usar rótulos como nós. É possível incluir o texto do rótulo dentro de cada nó ao invés de exibir rótulos adjacentes. Para gráficos com um número pequeno de campos, isso pode resultar em um gráfico mais legível.

Relationship between gender and grocery purchases

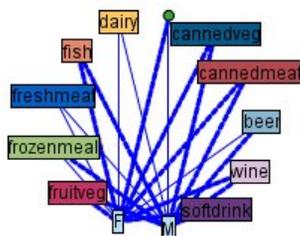


Figura 40. Gráfico da Web mostrando os rótulos como nós

Utilizando um Gráfico da Web

Nós da web são utilizados para mostrar a intensidade dos relacionamentos entre valores de dois ou mais campos simbólicos. As conexões são exibidas em um gráfico com tipos variados de linhas para indicar as conexões de aumento de intensidade. É possível utilizar um nó da web, por exemplo, para explorar o relacionamento entre os níveis de colesterol, pressão arterial e a droga que foi eficaz no tratamento de doenças do paciente.

- Conexões fortes são mostradas com uma linha pesada. Isso indica que os dois valores estão fortemente relacionados e devem ser explorados ainda mais.
- Conexões médias são mostradas com uma linha de ponderação normal.
- Conexões fracas são mostradas com uma linha pontilhada.
- Se nenhuma linha for mostrada entre dois valores, isso significa que os dois valores nunca ocorrem no mesmo registro ou que esta combinação ocorre em um número de registros abaixo do limite especificado na caixa de diálogo de nó da web.

Após criar um nó da web, há várias opções para ajustar a exibição do gráfico e gerar nós para análise adicional.

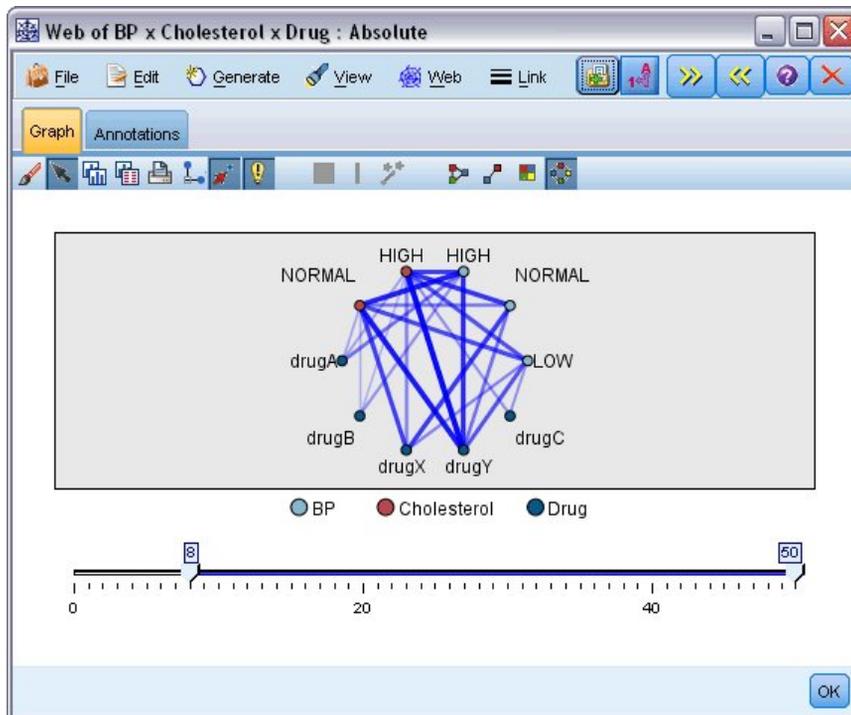


Figura 41. Gráfico da web indicando um número de relacionamentos fortes, como pressão arterial normal com DrugX e colesterol alto com DrugY

Para os nós da Web e nós da Web Direcionada, é possível:

- Alterar o layout da exibição da web.
- Ocultar pontos para simplificar a exibição.
- Alterar os limites de controle de estilos de linha.
- Destacar linhas entre valores para indicar um relacionamento "selecionado".
- Gerar um nó Seleção para um ou mais registros "selecionados" ou um nó Derivar Flag associado a um ou mais relacionamentos na web.

Para Ajustar Pontos

- **Mover** pontos ao clicar em um ponto e arrastá-lo para a nova localização. A web será redesenhada para refletir a nova localização.
- **Ocultar** pontos ao clicar com o botão direito em um ponto na web e escolher **Ocultar** ou em **Ocultar e Replanejar** no menu de contexto. **Ocultar** apenas oculta o ponto selecionado e todas as linhas associadas a ele. **Ocultar e Replanejar** redesenha a web, ajustando para quaisquer mudanças que tiverem sido feitas. Quaisquer movimentações manuais são desfeitas.
- **Mostrar** mostra todos os pontos ocultos ao escolher **Revelar Tudo** ou **Revelar Tudo e Replanejar** no menu da web na janela de gráfico. A seleção de **Revelar tudo e replanejar** redesenha a web, ajustando para incluir todos os pontos anteriormente ocultos e suas conexões.

Para Seleção, ou "Destacar", Linhas

As linhas selecionadas são destacadas em vermelho.

1. Para selecionar uma única linha, clique com o botão esquerdo na linha.
2. Para selecionar diversas linhas, execute um dos seguintes procedimentos:
 - Utilizando o cursor, desenhe um círculo em torno dos pontos cujas linhas deseja selecionar.
 - Mantenha pressionada a tecla Ctrl e clique com o botão esquerdo nas linhas individuais que deseja selecionar.

É possível cancelar a seleção de todas as linhas selecionadas clicando no plano de fundo do gráfico, ou escolhendo **Limpar Seleção** a partir do menu da web na janela do gráfico.

Para Visualizar a Web Usando um Layout Diferente

No menu da web, escolha **Layout de Círculo**, **Layout de Rede**, **Layout Direcionado** ou **Layout de Grade** para alterar o layout do gráfico.

Para Ativar e Desativar a Régua de Controle de Ligações:

No menu Visualizar, escolha **Régua de Controle de Ligações**.

Para Selecionar ou Sinalizar Registros para um Relacionamento Único

1. Clique com o botão direito na linha que representa o relacionamento de interesse.
2. No menu de contexto, escolha **Gerar Nó Seleção Para Ligação** ou **Gerar Nó Derivar Para Ligação**.

Um nó Seleção ou Derivar é incluído automaticamente na tela de fluxo com as opções e condições apropriadas especificadas:

- O nó Seleção seleciona todos os registros no relacionamento fornecido.
- O nó Derivar gera um flag indicando se o relacionamento selecionado se aplica a registros no conjunto de dados inteiro. O campo de sinalização é nomeado ao unir os dois valores no relacionamento com um sublinhado, como *LOW_drugC* ou *drugC_LOW*.

Para Selecionar ou Sinalizar Registros para um Grupo de Relacionamentos

1. Selecione uma ou mais linhas na exibição da web que representam relacionamentos de interesse.
 2. No menu Gerar na janela do gráfico, escolha **Nó Seleção ("And")**, **Nó Seleção ("Or")**, **Nó Derivar ("And")** ou **Nó Derivar ("Or")**.
- Os nós "Or" fornecem a disjunção das condições. Isso significa que o nó será aplicado aos registros que possuir algum dos relacionamentos selecionados.
 - Os nós "E" fornecem o conjunto de condições. Isso significa que o nó será aplicado apenas aos registros que possuem todos os relacionamentos selecionados. Ocorrerá um erro se qualquer um dos relacionamentos selecionados for mutuamente exclusivo.

Após ter concluído sua seleção, um nó Seleção ou um nó Derivar é incluído automaticamente na tela de fluxo com as opções e condições especificadas apropriadas.

Ajustando Limites da Web

Após criar um gráfico da web, é possível ajustar os limites que controlam os estilos de linha utilizando a régua de controle da barra de ferramentas para alterar a linha visível mínima. Também é possível visualizar opções de limite adicionais clicando no botão de seta dupla amarela na barra de ferramentas para expandir a janela do gráfico da web. Em seguida, clique na guia **Controles** para visualizar opções adicionais.

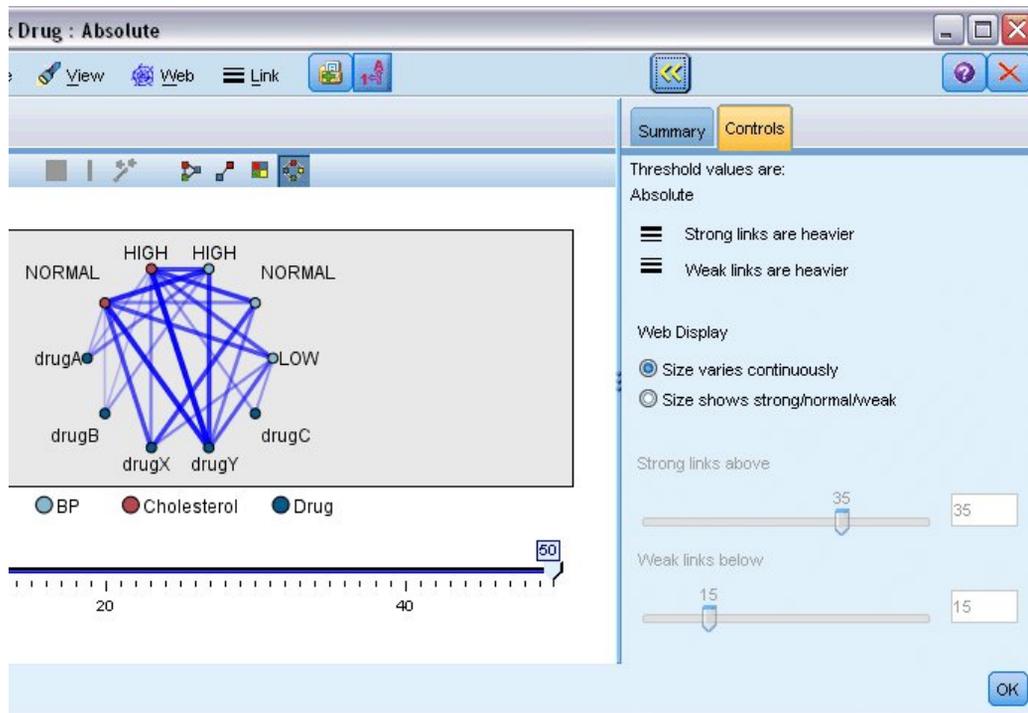


Figura 42. Janela expandida apresentando opções de exibição e de limite

Valores limite são. Mostra o tipo de limite selecionado durante a criação na caixa de diálogo do nó da Web.

Ligações fortes são mais pesadas. Selecionada por padrão, este é o modo padrão de visualizar as ligações entre os campos.

Ligações fracas são mais pesadas. Selecione para reverter o significado das ligações exibidas em linhas em negrito. Essa opção é frequentemente utilizada para detecção de fraude ou exame de valores discrepantes.

Exibição da Web. Especifique opções para controlar o tamanho das ligações no gráfico de saída:

- **Tamanho varia continuamente.** Selecione para exibir um intervalo de tamanhos de ligação que refletem a variação com relação às intensidades baseadas nos valores de dados reais.
- **Tamanho mostra forte/normal/fraca.** Selecione para exibir três intensidades de conexões--forte, normal e fraca. Os pontos de corte para essas categorias podem ser especificados acima e também no gráfico final.

Ligações fortes acima. Especifique um limite para conexões fortes (linhas intensas) e conexões regulares (linhas normais). Todas as conexões acima desse valor são consideradas fortes. Utilize a régua de controle para ajustar o valor ou insira um número no campo.

Ligações fracas abaixo. Especifique um número indicando o limite para conexões fracas (linhas pontilhadas) e conexões regulares (linhas normais). Todas as conexões abaixo deste valor são consideradas fracas. Utilize a régua de controle para ajustar o valor ou insira um número no campo.

Após ajustar os limites para uma web, é possível replanejar ou redesenhar a exibição da web com os novos valores limite por meio do menu da web localizado na barra de ferramentas do gráfico da web. Após localizar configurações que revelam os padrões mais significativos, é possível atualizar as configurações originais no nó da web (também chamado de nó da Web Pai) escolhendo **Atualizar Nó Pai** a partir do menu da Web na janela do gráfico.

Criando uma Sumarização da Web

É possível criar um documento de sumarização da web que lista ligações fortes, médias e fracas ao clicar no botão de seta dupla amarela na barra de ferramentas para expandir a janela de gráfico da web. Em seguida, clique na guia **Sumarização** para visualizar tabelas para cada tipo de ligação. As tabelas podem ser expandidas e reduzidas utilizando os botões de comutação para cada uma.

Para imprimir a sumarização, escolha o seguinte no menu na janela do gráfico da web:

Arquivo > Imprimir Sumarização

Nó Avaliação

O nó Avaliação oferece uma maneira fácil de avaliar e de comparar modelos preditivos para escolher o melhor modelo para seu aplicativo. Os gráficos de avaliação mostram o desempenho dos modelos na predição de resultados específicos. Eles funcionam ao ordenar os registros com base no valor predito e na confiança da predição, dividindo os registros em grupos de tamanhos iguais (**quantis**) e, em seguida, representando o valor do critério de negócios para cada quantil, do mais alto para o mais baixo. Diversos modelos são mostrados como linhas separadas na representação.

Os resultados são manipulados pela definição de um valor específico ou de um intervalo de valores como uma **ocorrência**. As ocorrências geralmente indicam sucesso de algum tipo (como uma venda para um cliente) ou um evento de interesse (como um diagnóstico médico específico). É possível definir critérios de ocorrência na guia Opções da caixa de diálogo ou utilizar os critérios de ocorrência padrão conforme a seguir:

- Os campos de saída de **Flag** são diretos, em que as ocorrências correspondem aos valores *true*.
- Para campos de saída **Nominais**, o primeiro valor no conjunto define uma ocorrência.
- Para campos de saída **Contínuos**, as ocorrências igualam valores maiores que o ponto central do intervalo do campo.

Há seis tipos de gráficos de avaliação, cada um deles enfatizando um critério de avaliação diferente.

Gráficos de Ganhos

Os ganhos são definidos como a proporção do total de ocorrências que ocorre em cada quantil. Os ganhos são calculados como (número de ocorrências no quantil / número total de ocorrências) ? 100%.

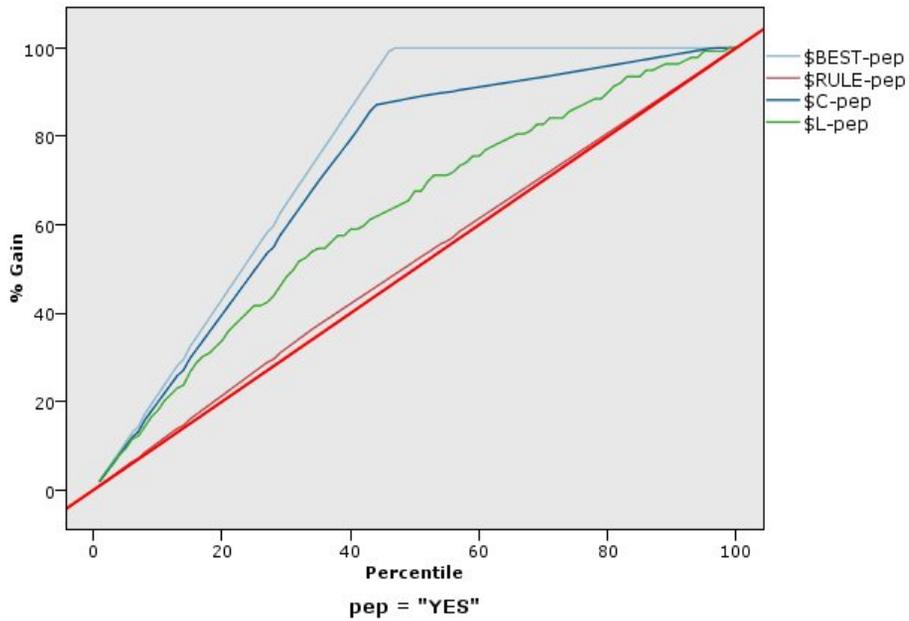


Figura 43. Gráfico de ganhos (acumulativos) com a linha de base, a melhor linha e a regra de negócios exibidas

Gráficos de Elevação

A elevação compara a porcentagem de registros em cada quantil que forem ocorrências com a porcentagem geral de ocorrências nos dados de treinamento. Ela é calculada como (ocorrências no quantil / registros no quantil) / (total de concorrências / total de registros).

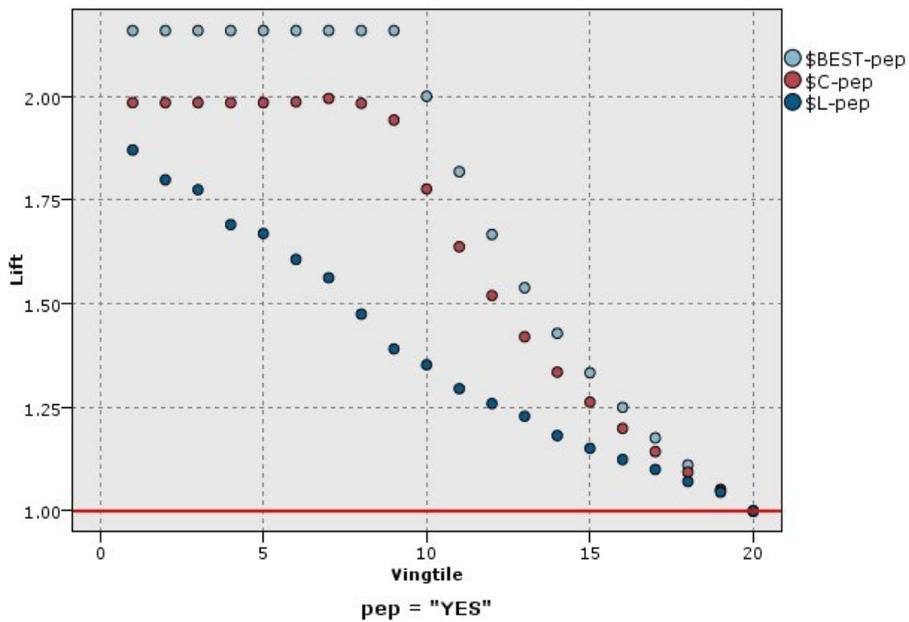


Figura 44. Gráfico de elevação (acumulativo) usando pontos e a melhor linha

Gráficos de Resposta

A resposta é simplesmente a porcentagem de registros no quantil que são ocorrências. A resposta é calculada como (ocorrências no quantil / registros no quantil) * 100%.

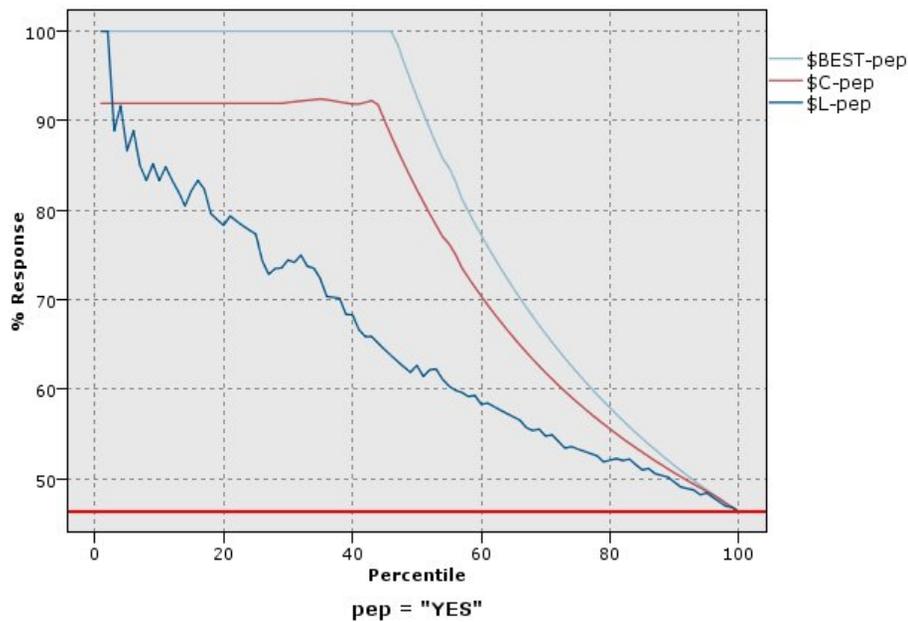


Figura 45. Gráfico de resposta (acumulativo) com a melhor linha

Gráficos de Lucros

O lucro é igual à **receita** de cada registro menos o **custo** do registro. Os lucros para um quantil são simplesmente a soma dos lucros de todos os registros no quantil. Assume-se que as receitas se aplicam apenas a ocorrências, mas os custos se aplicam a todos os registros. Os lucros e os custos podem ser corrigidos ou podem ser definidos por campos nos dados. Os lucros são calculados como (soma de receitas dos registros no quantil - soma dos custos dos registros no quantil).

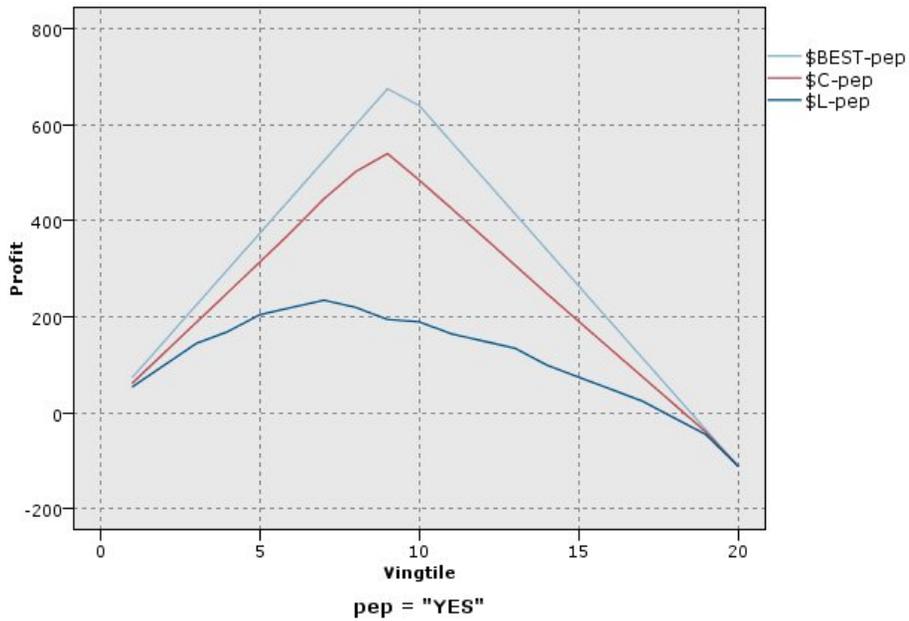


Figura 46. Gráfico de lucro (acumulativo) com a melhor linha

Gráficos ROI

O ROI (retorno sobre investimento) é similar ao lucro no que concerne à definição de receitas e despesas. O ROI compara os lucros com os custos do quantil. O ROI é calculado como (lucros do quantil / custos do quantil) ? 100%.

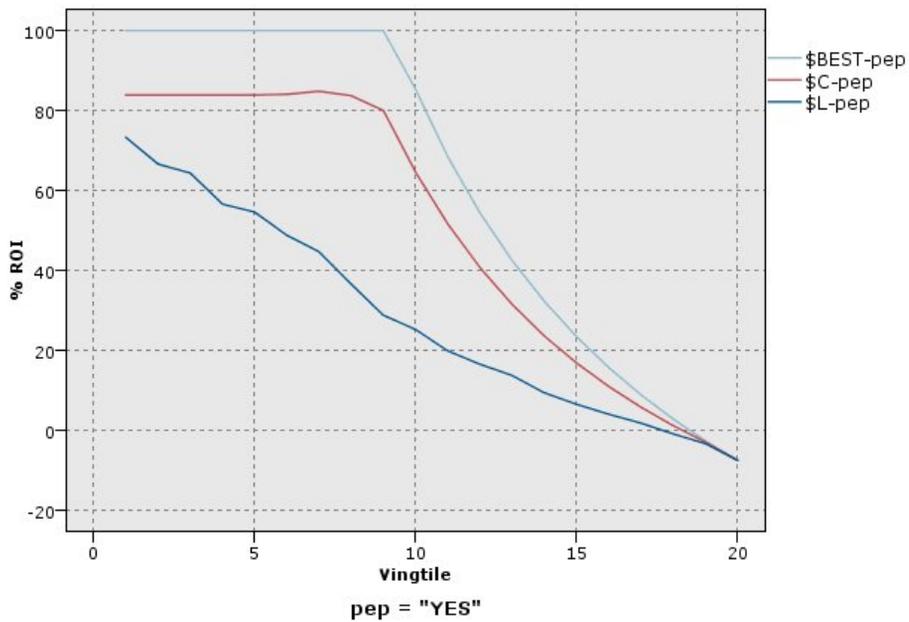


Figura 47. Gráfico ROI (acumulativo) com a melhor linha

Gráficos ROC

O ROC (característica de operação do receptor) pode ser utilizado somente com classificadores binários. O ROC pode ser utilizado para visualizar, organizar e selecionar classificadores com base em seu desempenho. Um gráfico ROC representa a taxa de verdadeiros positivos (ou sensibilidade) com relação à taxa de falsos positivos do classificador. Um gráfico ROC representa os trade-offs relativos entre benefícios (verdadeiros positivos) e custos (falsos positivos). Um verdadeiro positivo é uma instância que é uma ocorrência e é classificada como uma ocorrência. Portanto, a taxa de verdadeiros positivos é calculada como o número de verdadeiros positivos / número de instâncias que forem realmente ocorrências. Um falso positivo é uma instância que é uma perda e é classificada como uma ocorrência. Portanto, a taxa de falsos positivos é calculada como o número de falsos positivos / número de instâncias que forem realmente perdas.

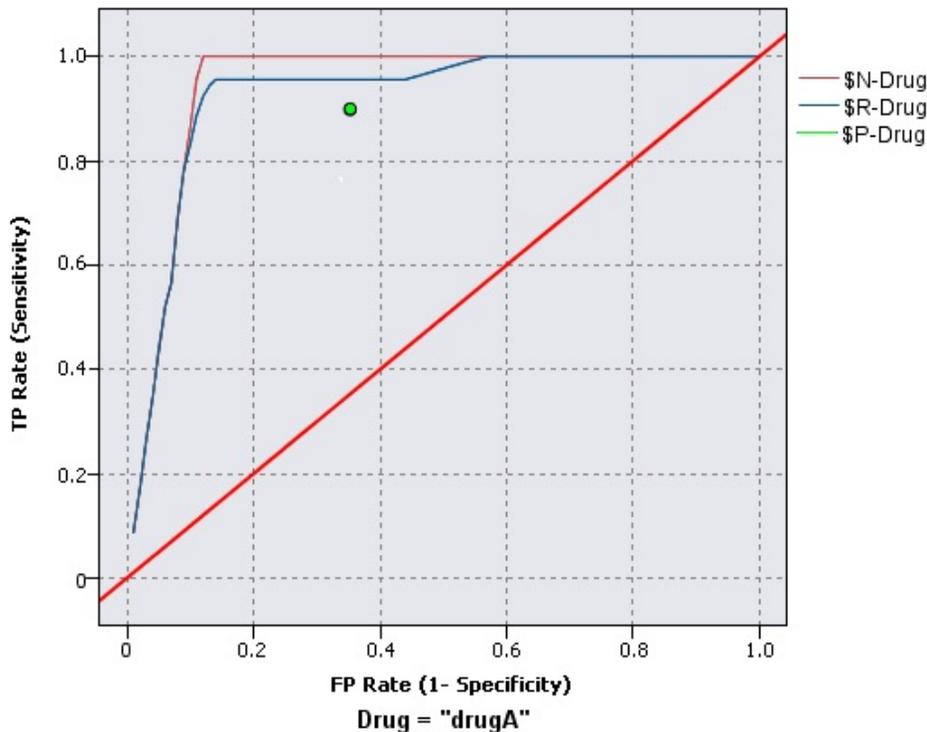


Figura 48. Gráfico ROC com a melhor linha

Os gráficos de avaliação também podem ser acumulativos, de modo que cada ponto é igual ao valor do quantil correspondente mais todos os quantis maiores. Os gráficos acumulativos geralmente conduzem melhor o desempenho geral dos modelos, ao passo que os gráficos não acumulativos geralmente se destacam ao indicar áreas de problema específicas para os modelos.

Guia Gráfico de Avaliação

Tipo de gráfico. Selecione um dos tipos a seguir: **Ganhos**, **Resposta**, **Elevação**, **Lucro**, **ROI** (retorno sobre o investimento) ou **ROC** (característica de operação do receptor).

Gráfico acumulativo. Selecione para criar um gráfico acumulativo. Os valores nos gráficos acumulativos são representados para cada quantil, mais todos os quantis maiores. (O **Gráfico acumulativo** não está disponível para gráficos ROC).

Incluir linha de base. Selecione para incluir uma linha de base no gráfico, indicando uma distribuição perfeitamente aleatória das ocorrências em que a confiança se torna irrelevante. (**Incluir linha de base** não está disponível para gráficos de Lucro e de ROI).

Incluir melhor linha. Selecione para incluir uma melhor linha no gráfico, indicando confiança perfeita (em que ocorrências = 100% dos casos). (**Incluir melhor linha** não está disponível para gráficos ROC).

Usar critérios de lucro para todos os tipos de gráficos. Selecione para utilizar os critérios de lucro (custo, receita e ponderação) quando calcular as medidas de avaliação, ao invés da contagem normal de ocorrências. Para modelos com determinados destinos numéricos, como um modelo que prevê a receita obtida de um cliente em resposta a uma oferta, o valor do campo de destino fornece uma melhor medida do desempenho do modelo do que a contagem de ocorrências. Selecionar essa opção ativa os campos **Custos, Receita e Ponderação** para os gráficos Ganhos, Resposta e Elevação. Para utilizar os critérios de lucro para esses três tipos de gráfico, é recomendado configurar **Receita** para que seja o campo de destino, **Custo** para 0,0 para que o lucro seja igual à receita e especificar uma condição de ocorrência definida pelo usuário de "true", para que todos os registros sejam contados como ocorrências. (**Usar critérios de lucro para todos os tipos de gráficos** não está disponível para gráficos ROC).

Localizar campos preditos/preditores utilizando. Selecione **Metadados do campo de saída do modelo** para procurar os campos preditos no gráfico utilizando seus metadados ou selecione **Formato de nome do campo** para procurá-los por nome.

Campos de escore do gráfico. Marque esta caixa de seleção para ativar o seletor de campo de escore. Em seguida, selecione um ou mais campos de intervalo, contínuos ou de escore; ou seja, campos que não são estritamente modelos preditivos, mas que podem ser úteis para classificar registros em termos de propensão para que sejam uma ocorrência. O nó Avaliação pode comparar qualquer combinação de um ou mais campos de escore com um ou mais modelos preditivos. Um exemplo típico pode ser comparar vários campos RFM com seu melhor modelo preditivo.

Resposta. Selecione o campo de destino utilizando o seletor de campo. Escolha qualquer campo de flag ou nominal instanciado com dois ou mais valores.

Nota: este campo de destino é aplicável apenas para campos de escore (modelos preditivos definem seus próprios destinos) e será ignorado se um critério de ocorrência customizado for configurado na guia Opções.

Dividir por partição. Se um campo de partição for utilizado para dividir registros em amostras de treinamento, de teste e de validação, selecione esta opção para exibir um gráfico de avaliação separado para cada partição. Consulte o tópico "Nó Partição" na página 170 para obter mais informações.

Nota: ao dividir por partição, os registros com valores nulos no campo partição são excluídos da avaliação. Isso nunca será um problema se um nó Partição for utilizado, uma vez que os nós Partição não geram valores nulos.

Gráfico. Selecione na lista suspensa o tamanho de quantis para representar no gráfico. As opções incluem **Quartis, Quintis, Decis, Vintis, Percentis e 1000-tiles**. (**Gráfico** não está disponível para gráficos ROC).

Estilo. Selecione **Linha** ou **Ponto**.

Para todos os tipos de gráficos, exceto gráficos ROC, controles adicionais permitem especificar custos, receitas e ponderações.

- **Custos.** Especifique o custo associado a cada registro. É possível selecionar custos **Fixo** ou **Variável**. Para custos fixos, especifique o valor do custo. Para custos variáveis, clique no botão Seletor de Campo para selecionar um campo como o campo de custo. (**Custos** não está disponível para gráficos ROC).
- **Renda.** Especifique a renda associada a cada registro que representa uma ocorrência. É possível selecionar custos **Fixo** ou **Variável**. Para rendas fixas, especifique o valor da renda. Para renda variável, clique no botão Seletor de Campo para selecionar um campo como o campo de renda. (**Renda** não está disponível para gráficos ROC).

- **Ponderação.** Se os registros em seus dados representarem mais de uma unidade, será possível usar ponderações de frequência para ajustar os resultados. Especifique a ponderação associada a cada registro usando as ponderações **Fixa** ou **Variável**. Para ponderações fixas, especifique o valor da ponderação (o número de unidades por registro). Para ponderações variáveis, clique no botão Seletor de Campo para selecionar um campo como o campo de ponderação. (**Ponderação** não está disponível para gráficos ROC).

Guia Opções de Avaliação

A guia Opções para gráficos de avaliação fornece flexibilidade na definição de ocorrências, critérios de escoragem e regras de negócios exibidos no gráfico. Também é possível configurar opções para exportar os resultados da avaliação do modelo.

Ocorrência definida pelo usuário. Selecione para especificar uma condição customizada utilizada para indicar uma ocorrência. Esta opção é útil para definir o resultado de interesse, ao invés de deduzi-lo a partir do tipo de campo de destino e da ordem de valores.

- **Condição.** Quando **Ocorrência definida pelo usuário** é selecionada acima, deve-se especificar uma expressão do CLEM para uma condição de ocorrência. Por exemplo, @TARGET = "YES" é uma condição válida que indica que um valor de *Sim* para o campo de destino será contado como uma ocorrência na avaliação. A condição especificada será utilizada para todos os campos de destino. Para criar uma condição, digite no campo ou utilize o Construtor de Expressões para gerar uma expressão de condição. Se os dados forem instanciados, será possível inserir valores diretamente do Construtor de Expressões.

Score definido pelo usuário. Selecione para especificar uma condição usada para escorar casos antes de designá-los aos quantis. O score padrão é calculado a partir do valor predito e da confiança. Utilize o campo Expressão para criar uma expressão de escoragem customizada.

- **Expressão.** Especifique uma expressão do CLEM usada para escoragem. Por exemplo, se uma saída numérica no intervalo 0-1 for ordenada para que os valores inferiores sejam melhores que os superiores, será possível definir uma ocorrência como @TARGET < 0,5 e o score associado como 1 - @PREDICTED. A expressão de score deve resultar em um valor numérico. Para criar uma condição, digite no campo ou utilize o Construtor de Expressões para gerar uma expressão de condição.

Incluir regra de negócios. Selecione para especificar uma condição de regra que reflete os critérios de interesse. Por exemplo, talvez você queira exibir uma regra para todos os casos em que mortgage = "Y" and income >= 33000. As regras de negócios são desenhadas no gráfico e rotuladas na chave como *Regra*. (**Incluir regra de negócios** não é suportada para os gráficos ROC).

- **Condição.** Especifique uma expressão do CLEM utilizada para definir uma regra de negócios no gráfico de saída. Simplesmente digite no campo ou utilize o Construtor de Expressões para gerar uma expressão de condição. Se os dados forem instanciados, será possível inserir valores diretamente do Construtor de Expressões.

Exportar resultados no arquivo. Selecione para exportar os resultados da avaliação do modelo em um arquivo de texto delimitado. É possível ler este arquivo para executar análises especializadas nos valores calculados. Configure as seguintes opções para exportação:

- **Nome do arquivo.** Insira o nome do arquivo para o arquivo de saída. Utilize o botão reticências (...) para navegar até a pasta desejada.
- **Delimitador.** Insira um caractere, como uma vírgula ou um espaço, para uso como o delimitador de campo.

Incluir nomes de campo. Selecione essa opção para incluir nomes de campo como a primeira linha do arquivo de saída.

Nova linha depois de cada registro. Selecione essa opção para iniciar cada registro em uma nova linha.

Guia Aparência de Avaliação

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Texto. Aceite o rótulo de texto gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo X. Aceite o rótulo do eixo x (horizontal) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Rótulo Y. Aceite o rótulo do eixo y (vertical) gerado automaticamente ou selecione **Customizado** para especificar um rótulo.

Exibir linha de grade. Selecionada por padrão, esta opção exibe uma linha de grade atrás do plot ou do gráfico que permite determinar mais facilmente os pontos de corte da região e da banda. As linhas de grade são exibidas sempre em branco, a menos que o segundo plano do gráfico seja branco; nesse caso, elas são exibidas em cinza.

Lendo os Resultados de uma Avaliação de Modelo

A interpretação de um gráfico de avaliação depende, até certo ponto, do tipo de gráfico, mas existem algumas características comuns a todos os gráficos de avaliação. Para gráficos acumulativos, linhas mais altas indicam melhores modelos, especialmente no lado esquerdo do gráfico. Em muitos casos, ao comparar diversos modelos em que as linhas se cruzam, um modelo é maior em uma parte do gráfico e outro é maior em uma parte diferente do gráfico. Nesse caso, é necessário considerar qual parte da amostra você desejará (que define um ponto no eixo x) quando decidir qual modelo escolher.

A maioria dos gráficos não acumulativos será muito semelhante. Para bons modelos, gráficos não acumulativos devem ser altos em direção ao lado esquerdo do gráfico e baixos em direção ao lado direito do gráfico. (Se um gráfico não acumulativo mostrar um padrão do tipo dente de serra, será possível suavizá-lo ao reduzir o número de quantis para representação e executar novamente o gráfico). As quedas acentuadas no lado esquerdo do gráfico ou picos no lado direito podem indicar áreas em que o modelo está prevendo mal. Uma linha simples que atravessa todo o gráfico indica um modelo que essencialmente não fornece nenhuma informação.

Gráficos de ganhos. Os gráficos de ganhos acumulativos iniciam em 0% e terminam em 100% à medida que você vai da esquerda para a direita. Para um bom modelo, o gráfico de ganhos se elevará acentuadamente até 100% e, em seguida, se estabilizará. Um modelo que não fornece nenhuma informação seguirá a diagonal da parte inferior esquerda até a parte superior direita (mostrado no gráfico se **Incluir linha de base** estiver selecionada).

Gráficos de elevação. Os gráficos de elevação acumulativos tendem a iniciar acima de 1,0 e descer gradativamente até atingirem 1,0 à medida que você vai da esquerda para a direita. A borda direita do gráfico representa o conjunto de dados inteiro, portanto, a razão das ocorrências nos quantis acumulativos com essas ocorrências é 1,0. Para um bom modelo, a elevação deve começar bem acima de 1,0 à esquerda, permanecer em um patamar alto à medida que você move para a direita e, em seguida, cair acentuadamente para 1,0 no lado direito do gráfico. Para um modelo que não fornece nenhuma informação, a linha flutuará ao redor de 1,0 no gráfico inteiro. (Se **Incluir linha de base** for selecionada, uma linha horizontal em 1,0 será mostrada no gráfico para referência).

Gráficos de resposta. Gráficos de resposta acumulativa tendem a ser muito semelhantes aos gráficos de elevação, exceto para o ajuste de escala. Os gráficos de resposta geralmente iniciam próximo a 100% e descem gradativamente até atingirem a taxa de resposta geral (total de ocorrências / total de registros) na

borda direita do gráfico. Para um bom modelo, a linha deve iniciar próximo ou em 100% à esquerda, permanecer em um patamar alto conforme você o acompanha para a direita e, em seguida, cair acentuadamente até a taxa de resposta geral no lado direito do gráfico. Para um modelo que não fornece nenhuma informação, a linha flutuará ao redor da taxa de resposta geral no gráfico inteiro. (Se **Incluir linha de base** estiver selecionada, uma linha horizontal na taxa de resposta geral será mostrada no gráfico para referência).

Gráficos de lucro. Os gráficos de lucros acumulativos mostram a soma dos lucros à medida que você aumenta o tamanho da amostra selecionada, da esquerda para a direita. Os gráficos de lucro geralmente iniciam quase em 0, aumentam gradativamente à medida que você move para a direita até atingirem um pico ou um patamar no meio e, em seguida, diminuem em direção à borda direita do gráfico. Para um bom modelo, os lucros mostrarão um pico bem-definido em algum lugar no meio do gráfico. Para um modelo que não fornece nenhuma informação, a linha será relativamente reta e poderá aumentar, diminuir ou nivelar, dependendo da estrutura de custo/renda aplicada.

Gráficos ROI. Os gráficos ROI (retorno sobre investimento) acumulativo tendem a ser semelhantes aos gráficos de resposta e gráficos de elevação, exceto para o ajuste de escala. Os gráficos ROI iniciam geralmente acima de 0% e descem gradativamente até atingirem o ROI geral para o conjunto de dados inteiro (que pode ser negativo). Para um bom modelo, a linha deve começar bem acima de 0%, permanecer em um patamar alto conforme você move para a direita e, em seguida, cair acentuadamente para o ROI geral no lado direito do gráfico. Para um modelo que não fornece nenhuma informação, a linha deve flutuar em torno do valor ROI geral.

Gráficos ROC. As curvas ROC geralmente têm a forma de um gráfico de ganhos acumulativos. A curva inicia na coordenada (0,0) e termina na coordenada (1,1) à medida que você vai da esquerda para a direita. Um gráfico que se eleva acentuadamente para a coordenada (0,1) e, em seguida, se estabiliza é uma indicação de um bom classificador. Um modelo que classifica instâncias aleatoriamente como ocorrências ou perdas seguirá a diagonal da parte inferior esquerda até a parte superior direita (mostrado no gráfico se **Incluir linha de base** estiver selecionada). Se nenhum campo de confiança for fornecido para um modelo, o modelo será representado como um ponto único. O classificador com o limite de classificação ideal está localizado próximo à coordenada (0,1) ou no canto superior esquerdo do gráfico. Esta localização representa um número alto de instâncias que estiverem classificadas corretamente como ocorrências e um número baixo de instâncias que estiverem classificadas incorretamente como ocorrências. Os pontos acima da linha diagonal representam bons resultados de classificação. Os pontos abaixo da linha diagonal representam resultados de classificação ruins que são piores do que se as instâncias fossem classificadas aleatoriamente.

Usando um Gráfico de Avaliação

O uso do mouse para explorar um gráfico de avaliação é semelhante ao uso de um histograma ou de um gráfico de coleção. O eixo x representa os escores de modelo entre os quantis especificados, como vintis ou decis.

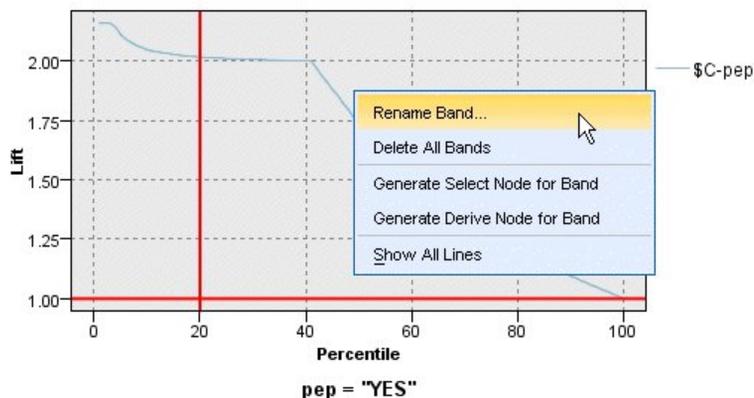


Figura 49. Trabalhando com um gráfico de avaliação

É possível particionar o eixo x em bandas tal como faria para um histograma usando o ícone de divisor para exibir opções para dividir automaticamente o eixo em bandas iguais. Consulte o tópico “Explorando Gráficos” na página 275 para obter mais informações. É possível editar manualmente os limites de bandas ao selecionar **Bandas de Gráfico** no menu Editar.

Após ter criado um gráfico de avaliação, definido bandas e examinado os resultados, será possível utilizar as opções no menu Gerar e no menu de contexto para criar automaticamente os nós com base nas seleções no gráfico. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

Ao gerar nós a partir de um gráfico de avaliação, será solicitado a selecionar um modelo único a partir de todos os modelos disponíveis no gráfico.

Selecione um modelo e clique em **OK** para gerar o novo nó na tela de fluxo.

Nó de Visualização de Mapa

O nó Visualização de Mapa pode aceitar diversas conexões de entrada e exibir dados geoespaciais em um mapa como uma série de camadas. Cada camada é um campo geoespacial único, por exemplo, a camada de base pode ser um mapa de um país e, acima dele, você pode ter uma camada para estradas, uma camada para rios e uma camada para as cidades.

Embora a maioria dos conjuntos de dados geoespaciais normalmente contém um campo geoespacial único, quando houver diversos campos geoespaciais em uma única entrada, é possível escolher quais campos deseja exibir. Dois campos da mesma conexão de entrada não podem ser exibidos ao mesmo tempo, entretanto, é possível copiar e colar a conexão de entrada e exibir um campo diferente um do outro.

Guia Gráfico de Visualização de Mapa Camadas

Esta tabela exibe informações sobre as entradas para o nó do mapa. A ordem das camadas determina a ordem na qual as camadas são exibidas na visualização do mapa e na saída visual quando o nó é executado. A linha superior na tabela é a camada 'top' e a linha inferior é a camada 'base'; em outras palavras, cada camada é mostrada no mapa na frente da camada diretamente abaixo dela na tabela.

Nota: Quando uma camada na tabela contiver um campo geoespacial tridimensional, somente os eixos x e y são representados. O eixo z é ignorado.

Nome Nomes são criados automaticamente para cada camada e são formados utilizando o seguinte formato: tag[source node:connected node]. Por padrão, a tag é mostrada como um número, com 1 representando a primeira entrada que é conectada, 2 para a segunda de entrada, e assim por diante. Se necessário, pressione o botão **Editar Camada** para alterar a tag na caixa de diálogo Alterar Opções de Camada do Mapa. Por exemplo, é possível alterar a tag para que seja "estradas" ou "cidades" para refletir a entrada de dados.

Tipo Mostra o ícone de tipo de medição do campo geoespacial que é selecionado como a camada. Se os dados de entrada contiverem diversos campos com um tipo de medição geoespacial, a seleção padrão usará a ordenação as seguir:

1. Point
2. Linestring
3. Polygon
4. MultiPoint
5. MultiLinestring
6. MultiPolygon

Nota: Se houver dois campos com o mesmo tipo de medição, o primeiro campo (em ordem alfabética por nome) será selecionado por padrão.

Símbolo

Nota: Esta coluna é preenchida somente para os campos Point e MultiPoint.

Mostra o símbolo que é utilizado para os campos Point e MultiPoint. Se necessário, pressione o botão **Editar Camada** para alterar o símbolo na caixa de diálogo Alterar Opções de Camada do Mapa.

Cor Mostra a cor que é selecionada para representar a camada no mapa. Se necessário, pressione o botão **Editar Camada** para alterar a cor na caixa de diálogo Alterar Opções de Camada do Mapa. A cor é aplicada a itens diferentes dependendo do tipo de medição.

- Para Points ou MultiPoints, a cor é aplicada no símbolo da camada.
- Para Linestrings ou Polygons, a cor é aplicada na forma inteira. Os polígonos sempre têm um contorno em preto, e a cor que é mostrada na coluna é a cor que é usada para preencher a forma.

Visualização

Esta área de janela mostra uma visualização da seleção atual de entradas na tabela **Camadas**. A visualização leva em conta a ordem das camadas, do símbolo, da cor e de quaisquer outras configurações de exibição que estiverem associadas às camadas e, quando possível, atualiza a exibição sempre que as configurações alterarem. Se você alterar detalhes em outro lugar no seu fluxo, como nos campos geoespaciais para uso como camadas, ou se corrigir detalhes, como funções de agregação associadas, poderá ser necessário clicar no botão **Atualizar Dados** para atualizar a visualização.

Utilize a **Visualização** para definir suas configurações de exibição antes de executar seu fluxo. Para proteger contra atrasos de tempo que possam ser causados pelo uso de um conjunto de dados grande, a visualização obtém uma amostra de cada camada e cria uma exibição dos 100 primeiros registros.

Alterar camadas de mapa

É possível utilizar a caixa de diálogo Alterar Opções de Camada de Mapa para corrigir vários detalhes de qualquer camada que é mostrada na guia **Gráfico** do nó Visualização de Mapa.

Detalhes da entrada

Tag Por padrão, a tag é um número, que pode ser substituído por uma tag mais significativa para ajudar a identificar a camada no mapa. Por exemplo, a tag pode ser o nome da entrada de dados, como "Cidades".

Camada de campo

Se você tiver mais de um campo geoespacial em seus dados de entrada, utilize essa opção para selecionar o campo que deseja exibir como uma camada no mapa.

Por padrão, as camadas que podem ser escolhidas estão na ordenação a seguir.

- Point
- Linestring
- Polygon
- MultiPoint
- MultiLinestring
- MultiPolygon

Configurações de exibição

Usar categorização hexa

Nota: Esta opção afeta apenas campos de ponto e multiponto.

A categorização hexa (hexagonal) combina pontos imediatos (com base em suas coordenadas x e y) em um único ponto para exibir no mapa. O ponto único é mostrado como um hexágono, mas é renderizado efetivamente como um polígono.

Como o hexágono é renderizado como um polígono, quaisquer campos de ponto com categorização hexa ativada são tratados como polígonos. Isso significa que se escolher **Ordenar por tipo** na caixa de diálogo do nó do mapa, quaisquer camadas de ponto que tiverem a categorização hexa aplicada serão renderizadas acima das camadas de polígono, mas abaixo das camadas linestring e de ponto.

Se utilizar a categorização hexa para um campo multiponto, o campo será primeiro convertido em um campo de ponto ao categorizar os valores de multiponto para calcular o ponto central. Os pontos centrais são utilizados para calcular as categorias hexas.

Agregação

Nota: Esta coluna está disponível somente quando marcar a caixa de seleção **Usar categorização hexa** e também selecionar uma **Sobreposição**.

Se selecionar um campo **Sobreposição** para uma camada de pontos que usa categorização hexa, todos os valores nesse campo deverão ser agregados para todos os pontos dentro do hexágono. Especifique uma função de agregação para todos os campos de sobreposição que desejar aplicar ao mapa. As funções de agregação disponíveis dependem do tipo de medição.

- As funções de agregação para um tipo de medição Contínuo, com armazenamento Real ou Número Inteiro:
 - Soma
 - Média
 - Min
 - Max
 - Mediana
 - Primeiro Quartil
 - Terceiro Quartil

- As funções de agregação para um tipo de medição Contínuo, com armazenamento Hora, Data ou Registro de Data e Hora:
 - Média
 - Min
 - Max
- Funções de agregação para tipos de medição Nominal ou Categórico:
 - Modo
 - Min
 - Max
- Funções de agregação para um tipo de medição Flag:
 - True, se alguma for verdadeira
 - False, se alguma for false

Cor

Utilize esta opção para escolher uma cor padrão a ser aplicada a todas as variáveis do campo geoespacial ou a um campo de sobreposição, que colore as variáveis com base nos valores de outro campo nos dados.

Se selecionar **Padrão**, será possível escolher uma cor a partir da paleta de cores que é mostrada na área de janela **Ordem de Cor de Categoria do Gráfico** na guia Exibir da caixa de diálogo Opções do Usuário.

Se selecionar **Sobreposição**, será possível escolher qualquer campo a partir da origem de dados que contém o campo geoespacial que foi selecionado como o **Campo de camada**.

- Para campos de sobreposição nominais ou categóricos, a paleta de cores a partir da qual é possível escolher é a mesma que aquela mostrada para as opções de cores **Padrão**.
- Para campos de sobreposição contínuos e ordinais, uma segunda lista suspensa é exibida, a partir da qual você seleciona uma cor. Quando você seleciona uma cor, a sobreposição é aplicada ao variar a saturação dessa cor de acordo com os valores no campo contínuo ou ordinal. O valor mais alto utiliza a cor escolhida na lista suspensa e os valores mais baixos são mostrados por saturações equivalentemente inferiores.

Símbolo

Nota: Ativada somente para os tipos de medição ponto e multiponto.

Utilize esta opção para escolher se deseja ter um símbolo **Padrão**, que é aplicado a todos os registros de seu campo geoespacial, ou um símbolo de **Sobreposição**, que altera o ícone de símbolo para os pontos com base nos valores de outro campo em seus dados.

Se selecionar **Padrão**, será possível escolher um dos símbolos padrão a partir de uma lista suspensa para representar os dados de ponto no mapa.

Se selecionar **Sobreposição**, será possível escolher qualquer campo nominal, ordinal ou categórico a partir da origem de dados que contém o campo geoespacial que foi selecionado como o **Campo de camada**. Para cada valor no campo de sobreposição, um símbolo diferente é exibido no mapa.

Por exemplo, seus dados podem conter um campo de ponto que representa os locais de lojas e a sobreposição pode ser um campo de tipo de loja. Neste exemplo, todas as lojas de alimentos podem ser identificadas no mapa por um símbolo de cruz e todas as lojas de eletrônicos por um símbolo quadrado.

Tamanho

Nota: Ativada somente para os tipos de medição ponto, multiponto, linestring e multilinestring.

Utilize esta opção para escolher se deseja ter um tamanho **Padrão**, que é aplicado a todos os registros de seu campo geoespacial, ou um tamanho de **Sobreposição**, que altera o tamanho do ícone de símbolo ou a espessura da linha com base nos valores de outro campo em seus dados.

Se selecionar **Padrão**, será possível escolher um valor de largura de pixel. As opções disponíveis são 1, 2, 3, 4, 5, 10, 20 ou 30.

Se selecionar **Sobreposição**, será possível escolher qualquer campo a partir da origem de dados que contém o campo geoespacial que foi selecionado como o **Campo de camada**. A espessura da linha ou do ponto varia dependendo do valor do campo escolhido.

Transparência

Utilize esta opção para escolher se deseja ter uma transparência **Padrão**, que é aplicada a todos os registros de seu campo geoespacial, ou uma transparência de **Sobreposição**, que altera a transparência do símbolo, da linha ou do polígono com base nos valores de outro campo em seus dados.

Se você selecionar **Padrão**, será possível escolher a partir de uma seleção de níveis de transparência que começam a partir de 0% (opaca) e aumenta em incrementos de 10% até 100% (transparente).

Se selecionar **Sobreposição**, será possível escolher qualquer campo a partir da origem de dados que contém o campo geoespacial que foi selecionado como o **Campo de camada**. Um nível diferente de transparência é exibido no mapa para cada valor no campo de sobreposição. A transparência é aplicada à cor escolhida a partir da lista suspensa de cores para o ponto, linha ou polígono.

Rótulo dos dados

Nota: Esta opção não estará disponível se marcar a caixa de seleção **Usar categorização hexa**.

Utilize esta opção para selecionar um campo para uso como rótulos de dados no mapa. Por exemplo, se aplicado a uma camada de polígono, o rótulo de dados poderá ser o campo de nome, contendo o nome de cada polígono. Se selecionar o campo de nome, esses nomes serão exibidos no mapa.

Guia Aparência de Visualização de Mapa

É possível especificar opções de aparência antes da criação do gráfico.

Título. Insira o texto a ser utilizado para o título do gráfico.

Subtítulo. Insira o texto a ser utilizado para o subtítulo do gráfico.

Legenda. Insira o texto a ser utilizado para a legenda do gráfico.

Explorando Gráficos

Ao passo que o modo de Edição permite editar o layout e a aparência do gráfico, o modo Explorar permite explorar analiticamente os dados e os valores representados pelo gráfico. O principal objetivo da exploração é analisar os dados e, em seguida, identificar valores utilizando bandas, regiões e marcação para gerar nós Seleção, Derivar ou Balanceamento. Para selecionar este modo, escolha **Visualizar > Modo Explorar** nos menus (ou clique no ícone da barra de ferramentas).

Ao passo que alguns gráficos podem usar todas as ferramentas de exploração, outros aceitam apenas um. O modo de Exploração inclui:

- Definir e editar bandas, que são utilizadas para dividir os valores ao longo de um eixo x de escala. Consulte o tópico “Usando Bandas” na página 276 para obter mais informações.

- Definir e editar regiões, que são utilizadas para identificar um grupo de valores dentro da área retangular. Consulte o tópico “Utilizando Regiões” na página 279 para obter mais informações.
- Marcar e desmarcar elementos para selecionar os valores que podem ser utilizados para gerar um nó Seleção ou Derivar. Consulte o tópico “Usando Elementos Marcados” na página 281 para obter mais informações.
- Gerar nós utilizando os valores identificados pelas bandas, regiões, elementos marcados e ligações da web para uso em seu fluxo. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

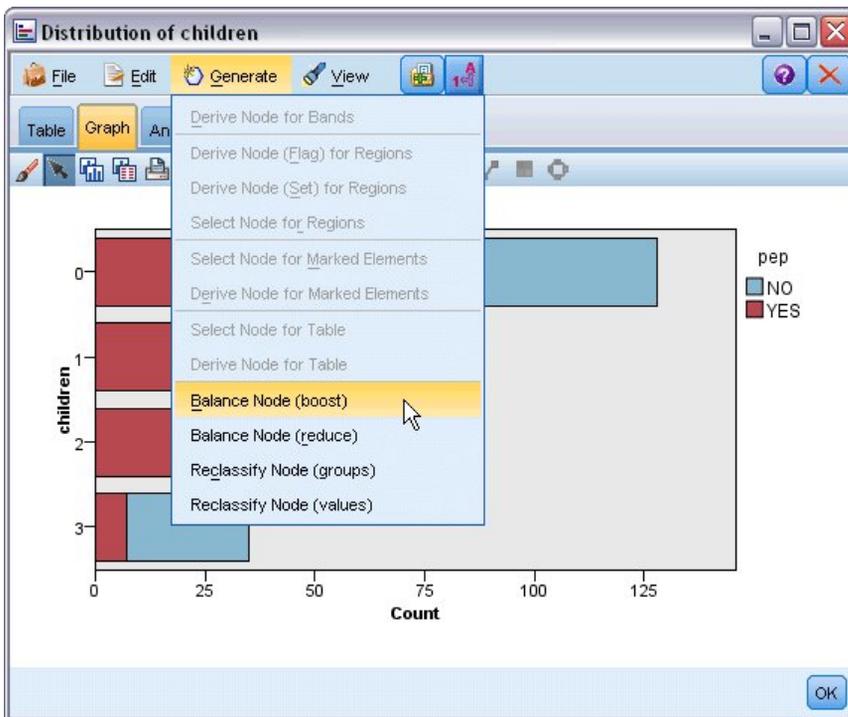


Figura 50. Gráfico com o menu gerar mostrando

Usando Bandas

Em qualquer gráfico com um campo em escala no eixo x , é possível desenhar linhas de banda verticais para dividir o intervalo de valores no eixo x . Se um gráfico tiver diversos painéis, uma linha de banda desenhada em um painel também será representada nos outros painéis.

Nem todos os gráficos aceitam bandas. Alguns desses gráficos que podem ter bandas incluem: histogramas, gráficos de barras e distribuições, gráficos (linha, dispersão, tempo, etc.), coleções e gráficos de avaliação. Nos gráficos com painel, as bandas aparecem em todos os painéis. Em alguns casos em um SPLOM, uma linha de banda horizontal é exibida pelo fato de o eixo no qual a banda de campo/variável foi desenhada ter sido invertido.

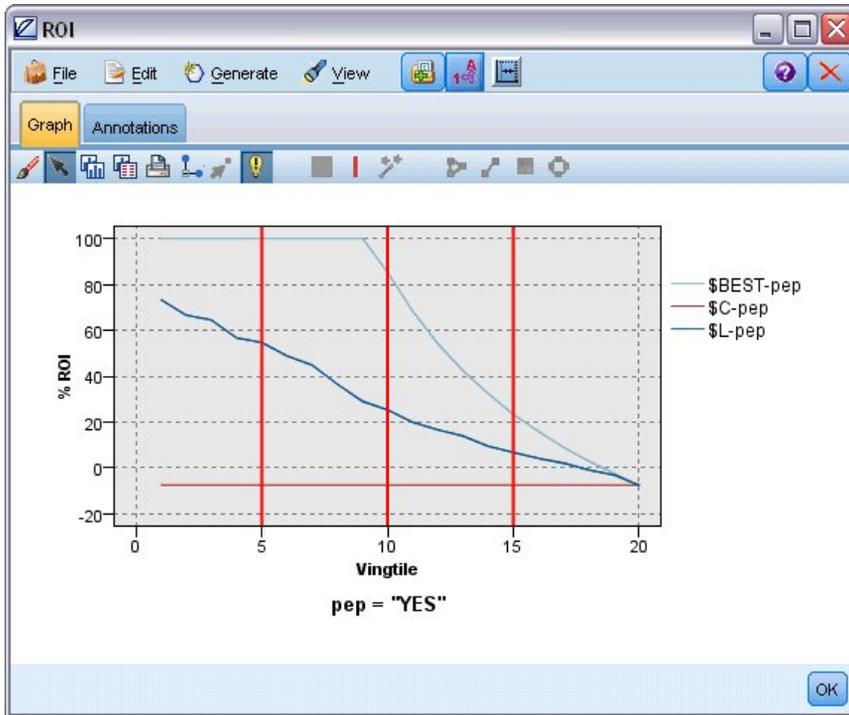


Figura 51. Gráfico com três bandas

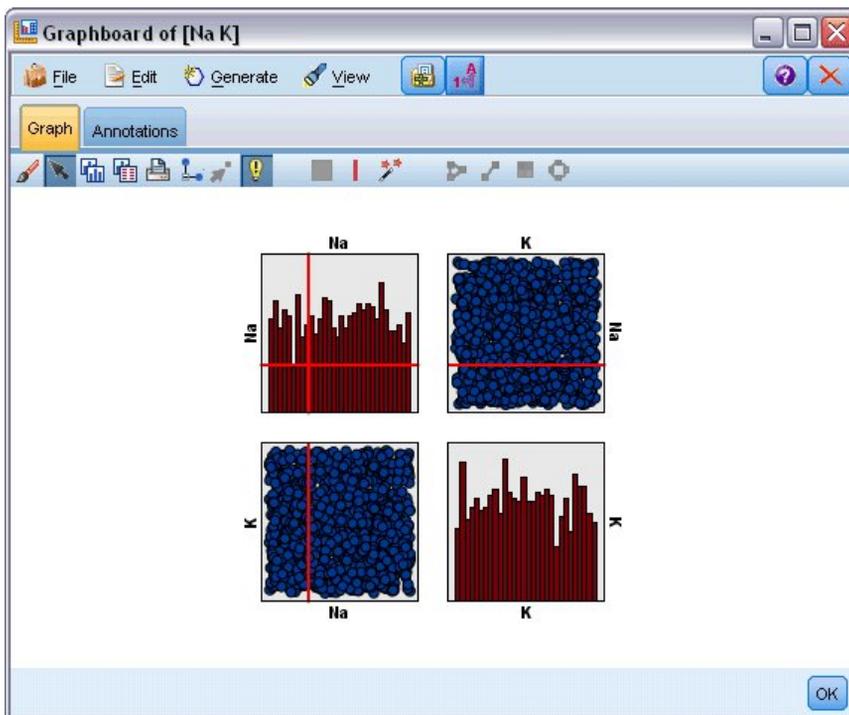


Figura 52. SPLOM com bandas

Definindo Bandas

Em um gráfico sem bandas, incluir uma linha de banda divide o gráfico em duas bandas. O valor da linha de banda representa o ponto inicial, também referido como o limite inferior, da segunda banda ao

ler o gráfico da esquerda para a direita. Da mesma forma, em um gráfico com duas bandas, incluir uma linha de banda divide uma dessas bandas em duas, resultando em três bandas. Por padrão, as bandas são nomeadas *bandN*, em que *N* é igual ao número de bandas da esquerda para a direita no eixo *x*.

Após definir uma banda, será possível arrastar e soltar essa banda para reposicioná-la no eixo *x*. É possível ver mais atalhos ao clicar com o botão direito em uma banda para tarefas como renomear, excluir ou gerar nós para essa banda específica.

Para definir bandas:

1. Verifique se você está no modo Explorar. Nos menus, escolha **Visualizar > Modo Explorar**.
2. Na barra de ferramentas do modo Explorar, clique no botão Desenhar Banda.



Figura 53. Botão da barra de ferramentas Desenhar Bandas

3. Em um gráfico que aceita bandas, clique no ponto do valor do eixo *x* no qual você deseja definir uma linha de banda.

Nota: como alternativa, clique no ícone de barra de ferramentas **Dividir Gráfico em Bandas**, insira o número de bandas iguais que deseja e clique em **Dividir**.



Figura 54. Ícone divisor usado para expandir a barra de ferramentas com opções para dividir em bandas

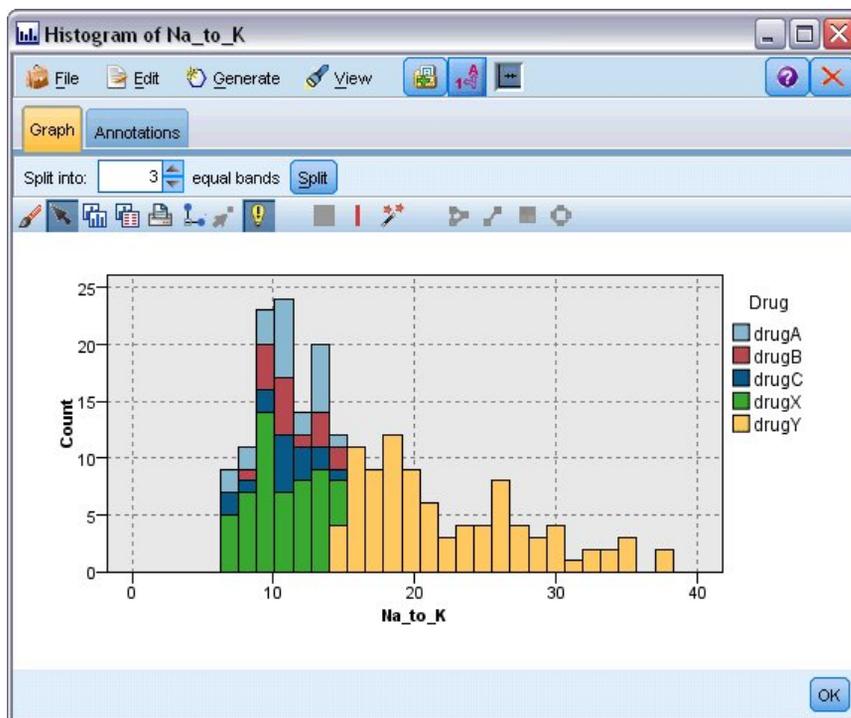


Figura 55. Criando barra de ferramentas de bandas iguais com bandas ativadas

Editando, Renomeando e Excluindo Bandas

É possível editar as propriedades de bandas existentes na caixa de diálogo Editar Bandas de Gráfico ou por meio dos menus de contexto no próprio gráfico.

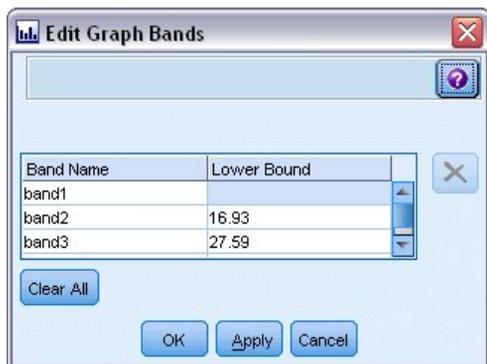


Figura 56. Caixa de diálogo Editar Bandas de Gráfico

Para editar bandas:

1. Verifique se você está no modo Explorar. Nos menus, escolha **Visualizar > Modo Explorar**.
2. Na barra de ferramentas do modo Explorar, clique no botão Desenhar Banda.
3. Nos menus, escolha **Editar > Bandas de Gráfico**. A caixa de diálogo Editar Bandas de Gráfico é aberta.
4. Se houver diversos campos no gráfico (como com gráficos SPLOM), será possível selecionar o campo desejado na lista suspensa.
5. Inclua uma nova banda digitando um nome e um limite inferior. Pressione a tecla Enter para iniciar uma nova linha.
6. Edite o limite de uma banda ao ajustar o valor de **Limite Inferior**.
7. Renomeie uma banda ao inserir um novo nome da banda.
8. Exclua uma banda ao selecionar a linha na tabela e clicar no botão Excluir.
9. Clique em **OK** para aplicar suas mudanças e fechar a caixa de diálogo.

Nota: como alternativa, é possível excluir e renomear bandas diretamente no gráfico ao clicar com o botão direito na linha de banda e escolher a opção desejada nos menus de contexto.

Utilizando Regiões

Em qualquer gráfico com dois eixos de escala (ou intervalo), é possível desenhar regiões para agrupar valores dentro de uma área retangular desenhada chamada região. Uma **região** é uma área do gráfico descrita pelos seus valores X e Y mínimo e máximo. Se um gráfico tiver diversos painéis, uma região desenhada em um painel também será representada nos outros painéis.

Nem todos os gráficos aceitam regiões. Alguns desses gráficos que aceitam regiões incluem: gráficos (de linha, de dispersão, bolha, tempo, etc.), SPLOM e coleções. Estas regiões são desenhadas no espaço X,Y e, portanto, não podem ser definidas em gráficos 1D, 3D ou animados. Nos gráficos com painel, as regiões aparecem em todos os painéis. Com uma matriz de gráfico de dispersão (SPLOM), uma região correspondente aparecerá nos gráficos superiores correspondentes, mas não nos gráficos de diagonais já que eles mostram apenas um campo em escala.

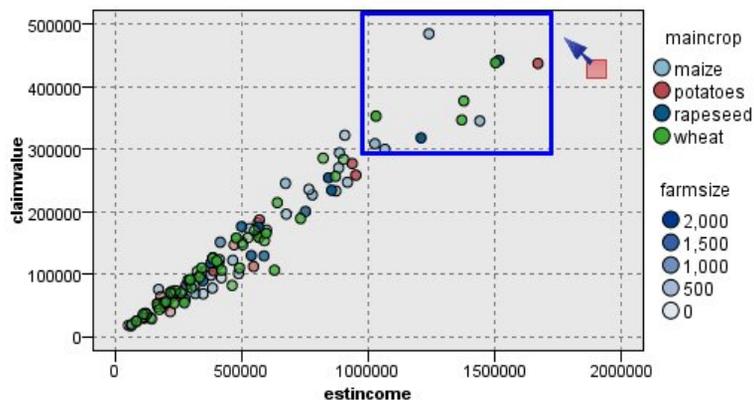


Figura 57. Definindo uma região de altos valores de solicitação

Definindo Regiões

Sempre que você define uma região, você está criando um agrupamento de valores. Por padrão, cada nova região é chamada *Region<N>*, em que *N* corresponde ao número de regiões já criadas.

Após definir uma região, é possível clicar com o botão direito na linha da região para obter alguns atalhos básicos. No entanto, é possível ver muitos outros atalhos ao clicar com o botão direito dentro da região (não na linha) para tarefas como renomear, excluir ou gerar nós Seleção e Derivar para essa região específica.

É possível selecionar subconjuntos de registros com base na inclusão deles em uma determinada região ou em uma de várias regiões. Também é possível incorporar as informações de região para um registro ao produzir um nó Derivar para sinalizar registros com base na inclusão deles em uma região. Consulte o tópico “Gerando Nós a partir de Gráficos” na página 282 para obter mais informações.

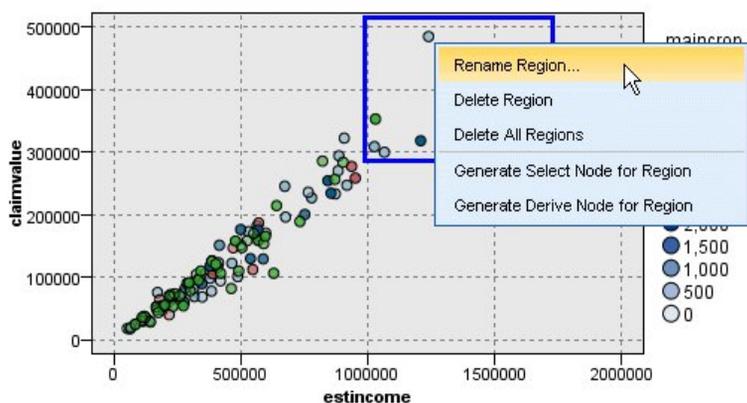


Figura 58. Explorando a região de altos valores de solicitação

Para definir regiões:

1. Verifique se você está no modo Explorar. Nos menus, escolha **Visualizar > Modo Explorar**.
2. Na barra de ferramentas do modo Explorar, clique no botão **Desenhar Região**.



Figura 59. Botão da barra de ferramentas Desenhar Região

3. Em um gráfico que aceita as regiões, clique e arraste o mouse para desenhar a região retangular.

Editando, Renomeando e Excluindo Regiões

É possível editar as propriedades de regiões existentes na caixa de diálogo Editar Regiões de Gráfico ou por meio dos menus de contexto no próprio gráfico.

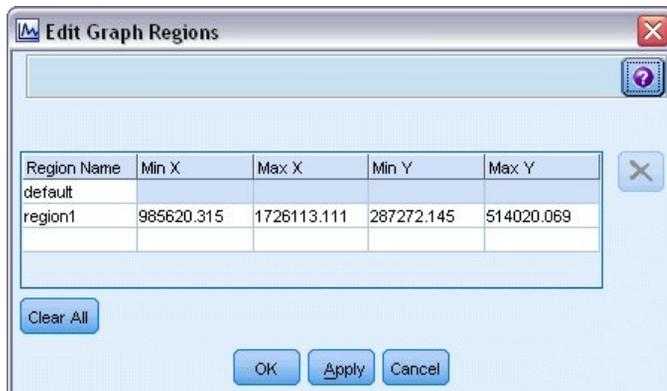


Figura 60. Especificando propriedades para as regiões definidas

Para editar regiões:

1. Verifique se você está no modo Explorar. Nos menus, escolha **Visualizar > Modo Explorar**.
2. Na barra de ferramentas do modo Explorar, clique no botão Desenhar Região.
3. Nos menus, escolha **Editar > Regiões de Gráfico**. A caixa de diálogo Editar Regiões de Gráfico é aberta.
4. Se você tiver diversos campos no gráfico (por exemplo, gráficos SPLOM), deve-se definir o campo para a região nas colunas *FieldA* e *FieldB*.
5. Inclua uma nova região em uma nova linha ao digitar um nome, selecionar nomes de campo (se aplicável) e definir os limites máximo e mínimo para cada campo. Pressione a tecla Enter para iniciar uma nova linha.
6. Edite os limites de região existente ao ajustar os valores **Mín.** e **Máx.** para *A* e *B*.
7. Renomeie uma região ao alterar o nome da região na tabela.
8. Exclua uma região ao selecionar a linha na tabela e clicar no botão Excluir.
9. Clique em **OK** para aplicar suas mudanças e fechar a caixa de diálogo.

Nota: como alternativa, é possível excluir e renomear regiões diretamente no gráfico ao clicar com o botão direito na linha de região e escolher a opção desejada nos menus de contexto.

Usando Elementos Marcados

É possível marcar elementos, como barras, fatias e pontos, em qualquer gráfico. As linhas, áreas e as superfícies não podem ser marcadas em gráficos diferentes de gráfico de tempo, multigráficos e gráficos de avaliação já que as linhas referenciam campos nesses casos. Sempre que você marca um elemento, você está essencialmente destacando todos os dados representados por esse elemento. Em qualquer gráfico em que o mesmo caso é representado em mais de um local (como SPLOM), a marcação é o mesmo que pintura com pincel. É possível marcar elementos em gráficos e até mesmo dentro de bandas e regiões. Sempre que marcar um elemento e, em seguida, voltar para o modo de Edição, a marcação ainda

permanecerá visível.

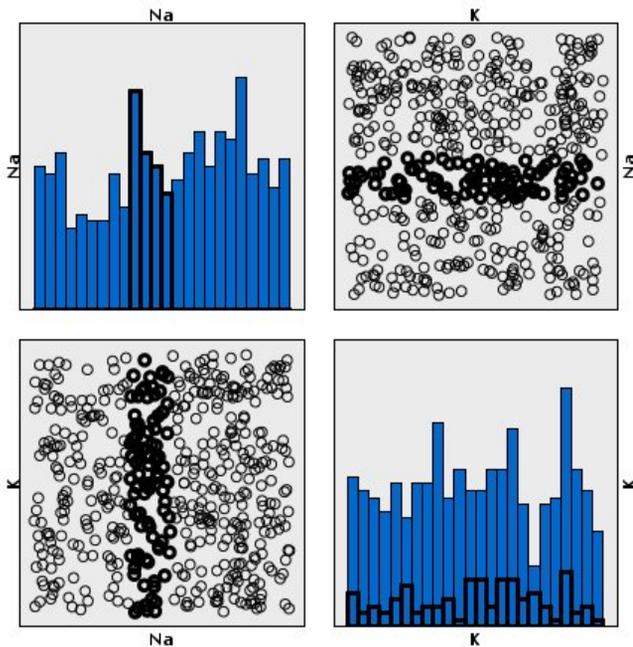


Figura 61. Marcando elementos em um SPLOM

É possível marcar e desmarcar os elementos ao clicar nos elementos no gráfico. Ao clicar em um elemento pela primeira vez para marcá-lo, o elemento aparecerá com uma cor de borda espessa para indicar que ele foi marcado. Se clicar no elemento novamente, a borda desaparecerá e o elemento não estará mais marcado. Para marcar diversos elementos, é possível manter a tecla Ctrl pressionada enquanto clica nos elementos ou arrastar o mouse em torno de cada um dos elementos que deseja marcar utilizando a "varinha mágica". Lembre-se de que, se clicar em outra área ou elemento sem manter a tecla Ctrl pressionada, todos os elementos marcados anteriormente serão desmarcados.

É possível gerar nós Seleção e Derivar a partir dos elementos marcados no gráfico. Consulte o tópico "Gerando Nós a partir de Gráficos" para obter mais informações.

Para marcar os elementos:

1. Verifique se você está no modo Explorar. Nos menus, escolha **Visualizar > Modo Explorar**.
2. Na barra de ferramentas do modo Explorar, clique no botão Marcar Elementos.
3. Clique no elemento que precisar ou clique e arraste o mouse para desenhar uma linha em torno da região que contém diversos elementos.

Gerando Nós a partir de Gráficos

Um dos recursos mais poderosos oferecidos pelos gráficos do IBM SPSS Modeler é a capacidade de gerar nós a partir de um gráfico ou de uma seleção dentro do gráfico. Por exemplo, no gráfico de plotagem de tempo, é possível gerar nós Derivar e Seleção com base em uma seleção ou região de dados, "subconfigurando" efetivamente os dados. Por exemplo, é possível utilizar esse recurso poderoso para identificar e excluir valores discrepantes.

Sempre que for possível desenhar uma banda, também é possível gerar um nó Derivar. Em gráficos com dois eixos da escala, é possível gerar os nós Derivar ou Seleção a partir das regiões desenhadas no gráfico. Em gráficos com elementos marcados, é possível gerar nós Derivar, nós Seleção e, em alguns casos, nós Filtro a partir desses elementos. A geração do nó de balanceamento é ativada para qualquer

gráfico que mostrar uma distribuição de contagens.

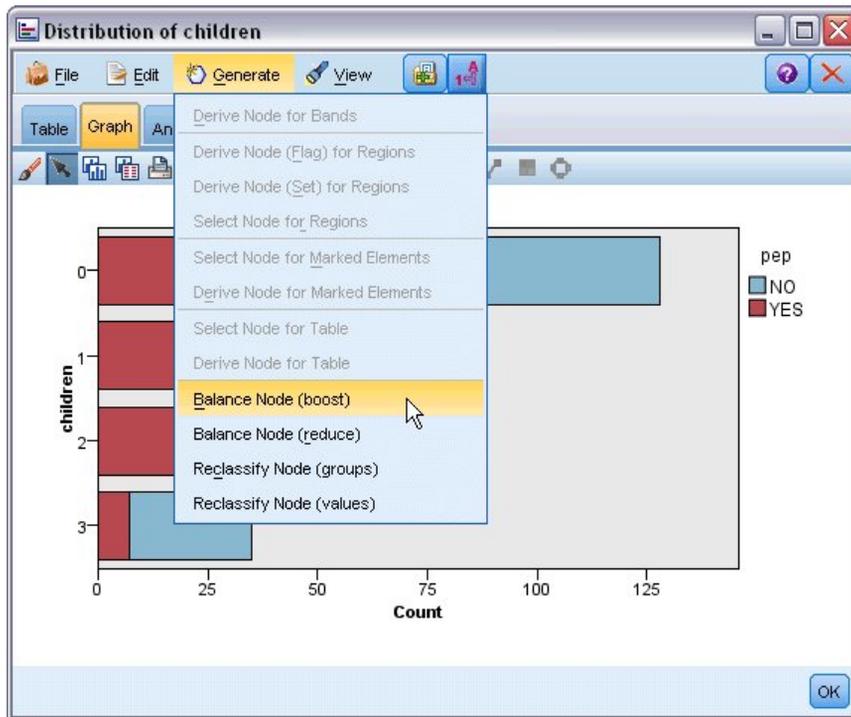


Figura 62. Gráfico com o menu gerar mostrando

Sempre que gerar um nó, ele é colocado na tela de fluxo diretamente para que seja possível conectá-lo a um fluxo existente. Os nós a seguir podem ser gerados a partir de gráficos: Seleção, Derivar, Balanceamento, Filtro e Reclassificar.

Nós Seleção

Os nós Seleção podem ser gerados para testar a inclusão dos registros dentro de uma região, a exclusão de todos os registros que estiverem fora da região ou a reversão para o processamento de recebimento de dados.

- **Para bandas.** É possível gerar um nó Seleção que inclua ou exclua os registros dentro dessa banda. **Selecionar nó para Bandas somente** está disponível apenas por meio dos menus contextuais, já que é necessário selecionar a banda que será usada no nó Seleção.
- **Para regiões.** É possível gerar um nó Seleção que inclua ou exclua os registros dentro de uma região.
- **Para elementos marcados.** É possível gerar nós Seleção para capturar os registros correspondentes aos elementos marcados ou às ligações de gráfico da web.

Nós Derivar

Os nós Derivar podem ser gerados a partir de regiões, bandas e elementos marcados. Todos os gráficos podem produzir nós Derivar. No caso de gráficos de avaliação, uma caixa de diálogo para selecionar o modelo se abre. No caso de gráficos da web, o **Nó Derivar ("And")** e o **Nó Derivar ("Or")** são possíveis.

- **Para bandas.** É possível gerar um nó Derivar que produz uma categoria para cada intervalo marcado no eixo, utilizando os nomes de banda listados na caixa de diálogo Editar Bandas como nomes de categoria.
- **Para regiões.** É possível gerar um nó Derivar (**Derivar como flag**) que cria um campo de sinalização chamado *in_region* com os flags configurados para *T* para os registros dentro de qualquer região e *F* para os registros fora de todas as regiões. Também é possível gerar um nó Derivar (**Derivar como**

conjunto) que produz um conjunto com um valor para cada região com um novo campo chamado *região* para cada registro, que assume como seu valor o nome da região na qual os registros caem. Os registros fora de todas as regiões recebem o nome da região padrão. Nomes de valor se tornam os nomes de região listados na caixa de diálogo Editar regiões.

- **Para elementos marcados.** É possível gerar um nó Derivar que calcula um flag que é *True* para todos os elementos marcados e *False* para todos os outros registros.

Nós Balanceamento

Os nós de balanceamento podem ser gerados para corrigir desbalanceamentos nos dados, como redução da frequência de valores comuns (use a opção de menu **Nó Balanceamento (reduzir)**) ou boosting da ocorrência de valores infrequentes (use a opção de menu **Nó Balanceamento (boost)**). A geração do nó de balanceamento é ativada para qualquer gráfico que mostrar uma distribuição de contagens, como Histograma, Ponto, Coleção, Barra de Contagens, Setor de Contagens e Multigráficos.

Nós Filtro

Os nós Filtro podem ser gerados para renomear ou filtrar campos com base nas linhas ou nos nós marcados no gráfico. No caso de gráficos de avaliação, a melhor linha de ajuste não gera um nó de filtro.

Nós Reclassificar

Os nós Reclassificar podem ser gerados para recodificar valores. Essa opção é utilizada para gráficos de distribuição. É possível gerar um nó Reclassificar para **grupos** para recodificar valores específicos de um campo exibido, dependendo de sua inclusão em um grupo (selecione grupos utilizando Ctrl+clique na guia **Tabelas**). Também é possível gerar um nó de reclassificação para **valores** para recodificar dados em um conjunto de inúmeros valores existente, como reclassificar dados em um conjunto de valores padrão a fim de mesclar dados financeiros a partir de várias empresas para análise.

Nota: se os valores forem predefinidos, será possível lê-los no IBM SPSS Modeler como um arquivo simples e utilizar uma distribuição para exibir todos os valores. Em seguida, gere um nó Reclassificar (valores) para esse campo diretamente do gráfico. Fazer isso colocará todos os valores de destino na coluna *Novos valores* do nó Reclassificar (lista suspensa).

Ao configurar opções para o nó Reclassificar, a tabela permite um mapeamento claro de valores de conjunto antigos para novos valores que você especificar:

- **Valor original.** Essa coluna lista os valores existentes para um ou mais campos de seleção.
- **Novo valor.** Use essa coluna para digitar novos valores de categoria ou selecione um na lista suspensa. Ao gerar automaticamente um nó Reclassificar utilizando os valores de um gráfico Distribuição, esses valores são incluídos na lista suspensa. Isso permite mapear rapidamente valores existentes para um conjunto de valores conhecido. Por exemplo, as organizações de assistência médica às vezes agrupam diagnósticos de forma diferente com base na rede ou no código do idioma. Após uma fusão ou aquisição, todas as partes precisarão reclassificar dados novos ou mesmo existentes em um modo consistente. Ao invés de digitar manualmente cada valor de destino em uma lista longa, é possível ler a lista principal de valores no IBM SPSS Modeler, executar um gráfico de Distribuição para o campo *Diagnóstico* e gerar um nó Reclassificar (valores) para esse campo diretamente do gráfico. Esse processo tornará todos os valores de Diagnóstico de destino disponíveis a partir da lista suspensa de Novos Valores.

Para obter mais informações sobre o Nó Reclassificar, consulte “Configurando Opções para o Nó Reclassificar” na página 157.

Gerando Nós a partir de Gráficos

É possível utilizar o menu Gerar na janela de saída do gráfico para gerar os nós. O nó gerado será colocado na tela de fluxo. Para utilizar o nó gerado, conecte-o a um fluxo existente.

Para gerar um nó a partir de um gráfico:

1. Verifique se você está no modo Explorar. Nos menus, escolha **Visualizar > Modo Explorar**.
2. Na barra de ferramentas do modo Explorar, clique no botão Região.
3. Defina bandas, regiões ou quaisquer elementos marcados necessários para gerar o nó.
4. No menu Gerar, escolha o tipo de nó que deseja produzir. Somente aqueles que forem possíveis serão ativados.

Nota: como alternativa, também é possível gerar nós diretamente a partir do gráfico ao clicar com o botão direito e escolher a opção de geração desejada nos menus de contexto.

Editando Visualizações

Ao passo que o modo Explorar permite explorar analiticamente os dados e os valores representados pela visualização, o modo de Edição permite alterar o layout e a aparência da visualização. Por exemplo, é possível alterar as fontes e as cores para corresponder à guia de estilo de sua organização. Para selecionar este modo, escolha **Visualizar > Modo de Edição** nos menus (ou clique no ícone da barra de ferramentas).

No modo de Edição, existem várias barras de ferramentas que afetam os diferentes aspectos do layout da visualização. Se houver algumas que você não utiliza, será possível ocultá-las para aumentar a quantidade de espaço na caixa de diálogo na qual o gráfico é exibido. Para selecionar ou cancelar a seleção das barras de ferramentas, clique no nome da barra de ferramentas relevante no menu Visualizar.

Nota: para incluir detalhes adicionais em suas visualizações, é possível aplicar rótulos de título, de nota de rodapé e de eixo. Consulte o tópico “Incluindo Títulos e Notas de Rodapé” na página 296 para obter mais informações.

Há várias opções para editar uma visualização no **modo de Edição**. É possível:

- Editar texto e formatá-lo.
- Alterar a cor de preenchimento, a transparência e o padrão de quadros e de elementos gráficos.
- Alterar a cor e os traços de bordas e linhas.
- Girar e alterar a forma e a proporção de aspecto de elementos de ponto.
- Alterar o tamanho dos elementos gráficos (como barras e pontos).
- Ajustar o espaço em torno dos itens ao utilizar margens e preenchimento.
- Especificar a formatação para números.
- Alterar as configurações de eixo e de escala.
- Ordenar, excluir e reduzir categorias em um eixo categórico.
- Configurar a orientação dos painéis.
- Aplicar transformações em um sistema de coordenadas.
- Alterar estatísticas, tipos de elementos gráficos e modificadores de colisão.
- Alterar a posição da legenda.
- Aplicar folhas de estilo de visualização.

Os tópicos a seguir descrevem como executar essas várias tarefas. Também é recomendado ler as regras gerais para editar gráficos.

Como Alternar para ao Modo de Edição

A partir dos menus, escolha:

Visualizar > Modo de Edição

Regras Gerais para Visualizações de Edição

Modo de Edição

Todas as edições são feitas no modo de Edição. Para ativar o modo de Edição, a partir dos menus, escolha:

Visualizar > Modo de Edição

Seleção

As opções disponíveis para edição dependem da seleção. Opções de barra de ferramentas e da paleta de propriedades diferentes são ativadas dependendo do que estiver selecionado. Somente os itens ativados se aplicam à seleção atual. Por exemplo, se um eixo for selecionado, as guias Escala, Marcações Principais e Marcações Secundárias estarão disponíveis na paleta de propriedades.

A seguir há algumas dicas para selecionar itens na visualização:

- Clique em um item para selecioná-lo.
- Selecione um elemento gráfico (como pontos em um gráfico de dispersão ou barras em um gráfico de barras) com um único clique. Após a seleção inicial, clique novamente para restringir a seleção para grupos de elementos gráficos ou para um único elemento gráfico.
- Pressione Esc para cancelar a seleção de tudo.

Paletas

Quando um item é selecionado na visualização, as várias paletas são atualizadas para refletir a seleção. As paletas contêm controles para fazer edições na seleção. As paletas podem ser barras de ferramentas ou um painel com diversos controles e guias. As paletas podem ser ocultadas, para assegurar que a paleta necessária seja exibida para fazer as edições. Verifique o menu Visualizar para as paletas que estiverem atualmente exibidas.

É possível reposicionar as paletas ao clicar e arrastar o espaço vazio em uma paleta de barra de ferramentas ou o lado esquerdo das outras paletas. O feedback visual permite saber onde é possível acoplar a paleta. Para paletas que não forem de barra de ferramentas, também é possível clicar no botão Fechar para ocultar a paleta e no botão Desacoplar para exibir a paleta em uma janela separada. Clique no botão Ajuda para exibir a ajuda para a paleta específica.

Configurações Automáticas

Algumas configurações fornecem uma opção **-auto-**. Isso indica que valores automáticos são aplicados. Quais configurações automáticas serão utilizadas depende dos valores específicos de visualização e de dados. É possível inserir um valor para substituir a configuração automática. Se desejar restaurar a configuração automática, exclua o valor atual e pressione Enter. A configuração exibirá **-auto-** novamente.

Removendo/Ocultando os Itens

É possível remover/ocultar vários itens na visualização. Por exemplo, é possível ocultar a legenda ou o rótulo do eixo. Para excluir um item, selecione-o e pressione Excluir. Se o item não permitir a exclusão, nada acontecerá. Se você acidentalmente excluir um item, pressione Ctrl+Z para desfazer a exclusão.

Estado

Algumas barras de ferramentas refletem o estado da seleção atual, outras não. A paleta de propriedades sempre reflete o estado. Se uma barra de ferramentas *não* refletir o estado, isso será mencionado no tópico que descreve a barra de ferramentas.

Editando e Formatando Texto

É possível editar o texto em vigor e alterar a formatação de um bloco de texto inteiro. Observe que não é possível editar o texto que estiver vinculado diretamente aos valores de dados. Por exemplo, não é possível editar um rótulo de marcação porque o conteúdo do rótulo é derivado dos dados subjacentes. No entanto, é possível formatar qualquer texto na visualização.

Como Editar o Texto Em Vigor

1. Dê um clique duplo no bloco de texto. Esta ação seleciona todo o texto. Todas as barras de ferramentas ficam desativadas neste momento, já que não é possível alterar nenhuma outra parte da visualização durante a edição de texto.
2. Digite para substituir o texto existente. Também é possível clicar no texto novamente para exibir um cursor. Posicione o cursor onde desejar e digite o texto adicional.

Como Formatar o Texto

1. Selecione o quadro que contém o texto. Não dê um clique duplo no texto.
2. Formate o texto utilizando a barra de ferramentas de fonte. Se a barra de ferramentas não estiver ativada, assegure-se de que somente o *quadro* que contém o texto seja selecionado. Se o texto em si for selecionado, a barra de ferramentas será desativada.

É possível alterar a fonte:

- Cor
- Família (por exemplo, Arial ou Verdana)
- Tamanho (a unidade é pt, a menos que você indique uma unidade diferente, como pc)
- Ponderação
- Alinhamento relativo ao quadro de texto

A formatação se aplica a todo o texto em um quadro. Não é possível alterar a formatação de letras ou de palavras individuais em qualquer bloco de texto específico.

Alterando Cores, Padrões, Traço e Transparência

Muitos itens diferentes em uma visualização possuem um preenchimento e borda. O exemplo mais óbvio é uma barra em um gráfico de barras. A cor das barras é a cor de preenchimento. Elas também podem ter uma borda sólida preta ao redor delas.

Há outros itens menos óbvios na visualização que possuem cores de preenchimento. Se a cor de preenchimento for transparente, você poderá não saber que existe um preenchimento. Por exemplo, considere o texto em um rótulo do eixo. Ele aparece como se esse texto fosse um texto "flutuante", mas na realidade aparece em um quadro que possui uma cor de preenchimento transparente. É possível ver o quadro selecionando o rótulo do eixo.

Qualquer quadro na visualização pode ter um estilo de preenchimento e de borda, incluindo o quadro ao redor da visualização inteira. Além disso, qualquer preenchimento possui um nível de opacidade/transparência associado que pode ser ajustado.

Como Alterar as Cores, Padrões, Traço e Transparência

1. Selecione o item que deseja formatar. Por exemplo, selecione as barras em um gráfico de barras ou um quadro contendo texto. Se a visualização for dividida por uma variável ou campo categórico, também é possível selecionar o grupo que corresponde a uma categoria individual. Isso permite

alterar a estética padrão designada para esse grupo. Por exemplo, é possível alterar a cor de um dos grupos de empilhamento em um gráfico de barras empilhadas.

2. Para alterar a cor de preenchimento, a cor da borda ou o padrão de preenchimento, utilize a barra de ferramentas de cores.

Nota: essa barra de ferramentas não reflete o estado da seleção atual.

Para alterar uma cor ou preenchimento, é possível clicar no botão para selecionar a opção exibida ou clicar na seta suspensa para escolher outra opção. Para cores, observe que há uma cor semelhante a branca com uma linha diagonal vermelha atravessando-a. Essa é a cor transparente. É possível utilizar isso, por exemplo, para ocultar as bordas nas barras em um histograma.

- O primeiro botão controla a cor de preenchimento. Se a cor estiver associada a um campo contínuo ou ordinal, esse botão alterará a cor de preenchimento para a cor associada ao valor mais alto nos dados. É possível usar a guia Cor na paleta de propriedades para alterar a cor associada ao valor mais baixo e aos dados omissos. A cor dos elementos alterará de forma incremental da cor Baixa para a cor Alta conforme os valores dos dados subjacentes aumentam.
 - O segundo botão controla a cor da borda.
 - O terceiro botão controla o padrão de preenchimento. O padrão de preenchimento utiliza a cor da borda. Portanto, o padrão de preenchimento será visível apenas se houver uma cor de borda visível.
 - O quarto controle é uma régua de controle e uma caixa de texto que controlam a opacidade da cor de preenchimento e do padrão. Uma porcentagem inferior significa menos opacidade e mais transparência. 100% é totalmente opaco (nenhuma transparência).
3. Para alterar o traço de uma borda ou linha, utilize a barra de ferramentas de linha.

Nota: essa barra de ferramentas não reflete o estado da seleção atual.

Assim como a outra barra de ferramentas, é possível clicar no botão para selecionar a opção exibida ou clicar na seta suspensa para escolher outra opção.

Girando e Alterando a Forma e a Proporção de Aspecto de Elementos de Ponto

É possível girar elementos de ponto, designar uma forma predefinida diferente ou alterar a proporção de aspecto (a razão da largura com a altura).

Como Modificar Elementos de Ponto

1. Selecione os elementos de ponto. Não é possível girar ou alterar a forma ou a proporção de aspecto de elementos de ponto individuais.
2. Utilize a barra de ferramentas de símbolo para modificar os pontos.
 - O primeiro botão permite alterar a forma dos pontos. Clique na seta suspensa e selecione um formato predefinido.
 - O segundo botão permite girar os pontos para uma posição de bússola específica. Clique na seta suspensa e, em seguida, arraste o ponteiro para a posição desejada.
 - O terceiro botão permite alterar a proporção de aspecto. Clique na seta suspensa e, em seguida, clique e arraste o retângulo que aparece. A forma do retângulo representa a proporção de aspecto.

Alterando o Tamanho dos Elementos Gráficos

É possível alterar o tamanho dos elementos gráficos na visualização. Estes incluem barras, linhas e pontos, entre outros. Se o elemento gráfico for dimensionado por uma variável ou campo, o tamanho especificado será o tamanho *mínimo*.

Como Alterar o Tamanho dos Elementos Gráficos

1. Selecione os elementos gráficos que deseja redimensionar.

- Utilize a régua de controle ou insira um tamanho específico para a opção disponível na barra de ferramentas do símbolo. A unidade é pixels, a menos que você indique uma unidade diferente (consulte a seguir uma lista completa de abreviações de unidade). Também é possível especificar uma porcentagem (como 30%), que significa que um elemento gráfico utiliza a porcentagem especificada do espaço disponível. O espaço disponível depende do tipo de elemento gráfico e da visualização específica.

Tabela 51. Abreviações de unidade válidas

Abreviação	Unidade
cm	centímetro
pol.	polegada
mm	milímetro
pc	paica
pt	ponto
px	pixel

Especificando Margens e Preenchimento

Se houver muito ou pouco espaçamento ao redor ou dentro de um quadro na visualização, será possível alterar suas configurações de margem e de preenchimento. A **margem** é a quantidade de espaço entre o quadro e outros itens em torno dela. O **preenchimento** é a quantidade de espaço entre a borda do quadro e o conteúdo do quadro.

Como Especificar Margens e Preenchimento

- Selecione o quadro para o qual deseja especificar margens e preenchimento. Isso pode ser um quadro de texto, o quadro ao redor da legenda ou até mesmo o quadro de dados exibindo os elementos gráficos (como barras e pontos).
- Utilize a guia Margens na paleta de propriedades para especificar as configurações. Todos os tamanhos estão em pixels, a menos que você indique uma unidade diferente (como centímetro ou polegada).

Formatando Números

É possível especificar o formato de números em rótulos de marcação ou em rótulos de valores de dados em um eixo contínuo que exibem um número. Por exemplo, é possível especificar que os números exibidos nos rótulos de marcação são mostrados em milhares.

Como Especificar Formatos Numéricos

- Selecione os rótulos de marcação ou os rótulos de valores de dados de eixo contínuo se eles contiverem números.
- Clique na guia **Formato** na paleta de propriedades.
- Selecione as opções de formatação de número que desejar:
 - Prefixo.** Um caractere para exibir no início do número. Por exemplo, insira um símbolo de dólar (\$) se os números forem salários em dólares americanos.
 - Sufixo.** Um caractere para exibir no término do número. Por exemplo, insira um sinal de porcentagem (%) se os números forem porcentagens.
 - Mín. de Dígitos de Número Inteiro.** O número mínimo de dígitos a ser exibido na parte de número inteiro de uma representação decimal. Se o valor real não contiver o número mínimo de dígitos, a parte de número inteiro do valor será preenchida com zeros.

Máx. de Dígitos de Número Inteiro. O número máximo de dígitos a ser exibido na parte de número inteiro de uma representação decimal. Se o valor real exceder o número máximo de dígitos, a parte de número inteiro do valor será substituída por asteriscos.

Mín. de Dígitos Decimais. O número mínimo de dígitos a ser exibido na parte decimal de uma representação decimal ou científica. Se o valor real não contiver o número mínimo de dígitos, a parte decimal do valor será preenchida com zeros.

Máx. de Dígitos Decimais. O número máximo de dígitos a ser exibido na parte decimal de uma representação decimal ou científica. Se o valor real exceder o número máximo de dígitos, o decimal será arredondado para o número apropriado de dígitos.

Científico. Especifica se os números serão exibidos em notação científica. A notação científica é útil para números muito grandes ou muito pequenos. **-auto-** permite que o aplicativo determine quando a notação científica é apropriada.

Ajuste de escala. Um fator de escala, que é um número pelo qual o valor original será dividido. Utilize um fator de escala quando os números forem grandes, mas não desejar que o rótulo se estenda demais para acomodar o número. Se você alterar o formato numérico dos rótulos de marcação, assegure-se de editar o título do eixo para indicar como o número deverá ser interpretado. Por exemplo, suponha que o eixo de escala exiba salários e que os rótulos sejam 30.000, 50.000 e 70.000. É possível inserir um fator de escala de 1000 para exibir 30, 50 e 70. Em seguida, deve-se editar o título do eixo de escala para incluir o texto em milhares.

Parênteses para -ve. Especifica se parênteses devem ser exibidos ao redor de valores negativos.

Agrupamento. Especifica se um caractere deve ser exibido entre grupos de dígitos. O código do idioma atual do computador determina qual caractere é utilizado para agrupamento de dígitos.

Alterando as Configurações de Eixo e Escala

Há várias opções para modificar os eixos e escalas.

Como alterar as configurações de eixo e de escala

1. Selecione qualquer parte do eixo (por exemplo, o rótulo do eixo ou os rótulos de marcação).
2. Use as guias Escala, Marcações Principais e Marcações Secundárias na paleta de propriedades para alterar as configurações de eixo e de escala.

Guia Escala

Nota: a guia Escala não aparece para gráficos nos quais os dados são pré-agregados (por exemplo, histogramas).

Tipo. Especifica se a escala é linear ou transformada. As transformações de escala ajudam a entender os dados ou fazer suposições necessárias para inferência estatística. Nos gráficos de dispersão, é possível utilizar uma escala transformada se o relacionamento entre as variáveis ou campos independentes e dependentes for linear. As transformações de escala também podem ser utilizadas para fazer um histograma defasado mais simétrico para que ele se pareça com uma distribuição normal. Observe que você está transformando apenas a escala na qual os dados são exibidos e não está transformando os dados reais.

- **linear.** Especifica uma escala linear não transformada.
- **log.** Especifica uma escala transformada de log de base 10. Para acomodar valores zero e negativos, essa transformação utiliza uma versão modificada da função de log. Essa função "log seguro" é definida como $\text{sign}(x) * \log(1 + \text{abs}(x))$. De modo que $\text{safeLog}(-99)$ é igual a:
 $\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$
- **potência.** Especifica uma escala transformada da potência usando um expoente de 0,5. Para acomodar valores negativos, essa transformação utiliza uma versão modificada da função de potência. Essa função "potência segura" é definida como $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$. De modo que $\text{safePower}(-100)$ é igual a:

$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$

Min/Max/Baixo Ideal/Alto Ideal. Especifica o intervalo para a escala. A seleção de **Baixo Ideal** e **Alto Ideal** permite que o aplicativo selecione uma escala apropriada com base nos dados. O mínimo e máximo são "ideais" porque eles são geralmente valores inteiros maiores ou menores que os valores máximo e mínimo dos dados. Por exemplo, se os dados estiverem em um intervalo de 4 a 92, um baixo e alto ideais para a escala poderão ser 0 e 100 ao invés do mínimo e máximo reais dos dados. Cuidado para não configurar um intervalo que seja muito pequeno e ocultar itens importantes. Além disso, observe que não será possível configurar um mínimo e máximo explícitos se a opção **Incluir zero** for selecionada.

Margem Baixa/Margem Alta. Cria margens na extremidade mais baixa e/ou alta do eixo. A margem aparece perpendicular ao eixo selecionado. A unidade é pixels, a menos que você indique uma unidade diferente (como centímetro ou polegada). Por exemplo, se você configurar a **Margem Alta** para 5 para o eixo vertical, uma margem horizontal de 5 pixels correrá ao longo da parte superior do quadro de dados.

Reverter. Especifica se a escala é revertida.

Incluir zero. Indica que a escala deve incluir 0. Essa opção é normalmente utilizada para gráficos de barras para assegurar que as barras comecem em 0, ao invés de um valor próximo à altura da barra menor. Se essa opção for selecionada, **Mín.** e **Máx.** serão desativados porque não será possível configurar um mínimo e máximo customizados para o intervalo da escala.

Guias Marcações Principais/Marcações Secundárias

Marcações ou **marcas de seleção** são as linhas que aparecem em um eixo. Elas indicam valores em intervalos ou categorias específicos. **Marcações principais** são as marcações com rótulos. Essas também são mais longas que outras marcações. **Marcações secundárias** são marcações que aparecem entre as marcações principais. Algumas opções são específicas para o tipo de marcação, mas a maioria das opções está disponível para as marcações principais e secundárias.

Mostrar marcações. Especifica se as marcações principais e secundárias são exibidas em um gráfico.

Mostrar linhas de grade. Especifica se as linhas de grade são exibidas nas marcações principais ou secundárias. As **linhas de grade** são linhas que cruzam um gráfico inteiro de eixo a eixo.

Posição. Especifica a posição das marcações com relação ao eixo.

Comprimento. Especifica o comprimento das marcações. A unidade é pixels, a menos que você indique uma unidade diferente (como centímetro ou polegada).

Base. *Aplica-se apenas às marcações principais.* Especifica o valor no qual a primeira marcação principal aparece.

Delta. *Aplica-se apenas às marcações principais.* Especifica a diferença entre as marcações principais. Ou seja, as marcações principais aparecerão em cada n -ésimo valor, em que n é o valor delta.

Divisões. *Aplica-se apenas às marcações secundárias.* Especifica o número de divisões de marcação secundária entre as marcações principais. O número de marcações secundárias é um a menos do número de divisões. Por exemplo, suponha que haja marcações principais em 0 e 100. Se você inserir 2 como o número de divisões de marcações secundárias, haverá *uma* marcação secundária em 50, dividindo o intervalo 0 a 100 e criando *duas* divisões.

Editando Categorias

É possível editar as categorias em um eixo categórico de várias maneiras:

- Alterar a ordenação para exibir as categorias.

- Excluir categorias específicas.
- Incluir uma categoria que não aparece no conjunto de dados.
- Reduzir/combinaar categorias pequenas em uma categoria.

Como Alterar a Ordenação de Categorias

1. Selecione um eixo categórico. A paleta Categorias exibe as categorias no eixo.

Nota: se a paleta não estiver visível, assegure-se de ter ativado-a. No menu Visualizar no IBM SPSS Modeler, escolha **Categorias**.

2. Na paleta Categorias, selecione uma opção de ordenação na lista suspensa:

Customizado. Classifique as categorias com base na ordem em que elas aparecem na paleta. Utilize os botões de seta para mover as categorias para a parte superior da lista, para cima, para baixo e para a parte inferior da lista.

Dados. Classifique as categorias com base na ordem em que elas ocorrem no conjunto de dados.

Nome. Classifique as categorias em ordem alfabética, utilizando os nomes conforme exibidos na paleta. Isso pode ser o valor ou o rótulo, dependendo se o botão da barra de ferramentas para exibir valores e rótulos estiver selecionado.

Valor. Classifique categorias pelo valor de dados subjacente, utilizando os valores exibidos entre parênteses na paleta. Somente origens de dados com metadados (como arquivos de dados do IBM SPSS Statistics) suportam esta opção.

Estatística. Ordene as categorias com base nas estatísticas calculadas para cada categoria. Os exemplos de estatísticas incluem contagens, porcentagens e médias. Essa opção estará disponível somente se uma estatística for utilizada no gráfico.

Como Incluir uma Categoria

Por padrão, apenas as categorias que aparecem no conjunto de dados estão disponíveis. É possível incluir uma categoria na visualização, se necessário.

1. Selecione um eixo categórico. A paleta Categorias exibe as categorias no eixo.

Nota: se a paleta não estiver visível, assegure-se de ter ativado-a. No menu Visualizar no IBM SPSS Modeler, escolha **Categorias**.

2. Na paleta Categorias, clique no botão Incluir categoria:



Figura 63. Botão Incluir categoria

3. Na caixa de diálogo Incluir uma nova categoria, insira um nome para a categoria.
4. Clique em **OK**.

Como Excluir Categorias Específicas

1. Selecione um eixo categórico. A paleta Categorias exibe as categorias no eixo.

Nota: se a paleta não estiver visível, assegure-se de ter ativado-a. No menu Visualizar no IBM SPSS Modeler, escolha **Categorias**.

2. Na paleta Categorias, selecione um nome de categoria na lista Incluir e, em seguida, clique no botão X. Para mover a categoria de volta, selecione o nome dela na lista Excluído e, em seguida, clique na seta para a direita da lista.

Como Reduzir/Combinaar Categorias Pequenas

É possível combinar as categorias que forem muito pequenas para que não haja necessidade de exibi-las separadamente. Por exemplo, se você tiver um gráfico de pizza com muitas categorias, considere reduzir as categorias com uma porcentagem menor que 10. A redução está disponível apenas para estatísticas que forem aditivas. Por exemplo, não é possível incluir médias juntas porque as médias não são aditivas. Portanto, a combinação ou a redução de categorias utilizando uma média não está disponível.

1. Selecione um eixo categórico. A paleta Categorias exibe as categorias no eixo.

Nota: se a paleta não estiver visível, assegure-se de ter ativado-a. No menu Visualizar no IBM SPSS Modeler, escolha **Categorias**.

2. Na paleta Categorias, selecione **Reduzir** e especifique uma porcentagem. Todas as categorias cuja porcentagem do total for menor que o número especificado são combinadas em uma categoria. A porcentagem é baseada na estatística mostrada no gráfico. A redução está disponível apenas para estatísticas baseadas em contagem e de adição (soma).

Alterando os Painéis de Orientação

Se estiver utilizando os painéis em sua visualização, será possível alterar sua orientação.

Como Alterar a Orientação dos Painéis

1. Selecione qualquer parte da visualização.
2. Clique na guia **Painéis** na paleta de propriedades.
3. Selecione uma opção a partir de **Layout**:

Tabela. Organiza os painéis como uma tabela, em que há uma linha ou coluna designada a cada valor individual.

Transposto. Organiza os painéis como uma tabela, mas também troca as linhas e as colunas originais. Esta opção não é a mesma que transpõem o próprio gráfico. Observe que o eixo x e o eixo y são inalterados quando selecionar essa opção.

Lista. Organiza os painéis como uma lista, em que cada célula representa uma combinação de valores. As colunas e linhas não são mais designadas a valores individuais. Esta opção permite que os painéis sejam agrupados, se necessário.

Transformando o Sistema de Coordenadas

Muitas visualizações são exibidas em um sistema de coordenadas simples e retangulares. É possível transformar o sistema de coordenadas, conforme necessário. Por exemplo, é possível aplicar uma transformação polar no sistema de coordenadas, incluir efeitos de aplicação de sombra oblíquos e transpor os eixos. Também é possível desfazer qualquer uma dessas transformações, se elas já estiverem aplicadas à visualização atual. Por exemplo, um gráfico de pizza é desenhado em um sistema de coordenadas polares. É possível desfazer a transformação polar e exibir o gráfico de pizza como uma única barra empilhada em um sistema de coordenadas retangulares.

Como Transformar o Sistema de Coordenadas

1. Selecione o sistema de coordenadas que deseja transformar. Selecione o sistema de coordenadas ao selecionar o quadro ao redor do gráfico individual.
2. Clique na guia **Coordenadas** na paleta de propriedades.
3. Selecione as transformações que deseja aplicar ao sistema de coordenadas. Também é possível cancelar a seleção de uma transformação para desfazê-la.

Transposta. A mudança da orientação do eixo é chamada de **Transposição**. Ela é semelhante ao trocar os eixos vertical e horizontal em uma visualização 2D.

Polar. Uma transformação polar produz os elementos gráficos em um ângulo e distância específicos a partir do centro do gráfico. Um gráfico de pizza é uma visualização 1D com uma transformação polar que desenha as barras individuais em ângulos específicos. Um gráfico de radar é uma visualização 2D

com uma transformação polar que desenha elementos gráficos em ângulos e distâncias específicos a partir do centro do gráfico. Uma visualização 3D deve também incluir uma dimensão de profundidade adicional.

Oblíqua. Uma transformação oblíqua inclui um efeito 3D nos elementos gráficos. Essa transformação inclui a profundidade nos elementos gráficos, mas a profundidade é meramente decorativa. Ela não é influenciada pelos valores de dados específicos.

Mesma Razão. Aplicar a mesma razão específica que a mesma distância em cada escala representa a mesma diferença em valores de dados. Por exemplo, 2 cm em ambas as escalas representam uma diferença de 1000.

% de Inserção de pré-transformação. Se os eixos forem cortados após a transformação, você poderá querer incluir as inserções no gráfico antes de aplicar a transformação. As inserções reduzem as dimensões numa certa porcentagem antes de quaisquer transformações serem aplicadas ao sistema de coordenadas. Você tem o controle sobre as dimensões *x* inferior, *x* superior, *y* inferior e *y* superior, nessa ordem.

% de Inserção de pós-transformação. Se desejar alterar a proporção de aspecto do gráfico, será possível incluir as inserções no gráfico após aplicar a transformação. As inserções reduzem as dimensões numa certa porcentagem após quaisquer transformações serem aplicadas ao sistema de coordenadas. Essas inserções também poderão ser aplicadas mesmo se nenhuma transformação for aplicada no gráfico. Você tem o controle sobre as dimensões *x* inferior, *x* superior, *y* inferior e *y* superior, nessa ordem.

Alterando Estatísticas e Elementos Gráficos

É possível converter um elemento gráfico em outro tipo, alterar a estatística utilizada para extrair o elemento gráfico ou especificar o modificador de colisão que determina o que acontece quando os elementos gráficos se sobrepõem.

Como Converter um Elemento de Gráfico

1. Selecione o elemento gráfico que deseja converter.
2. Clique na guia **Elemento** na paleta de propriedades.
3. Selecione um novo tipo de elemento gráfico na lista Tipo.

Tabela 52. Tipos de elementos gráficos

Tipo de Elemento Gráfico	Descrição
Ponto	Um marcador que identifica um ponto de dados específico. Um elemento de ponto é utilizado em gráficos de dispersão e em outras visualizações relacionadas.
Intervalo	Uma forma retangular desenhada em um valor de dados específico, preenchendo o espaço entre uma origem e outro valor de dados. Um elemento de intervalo é utilizado em gráficos de barras e em histogramas.
Linha	Uma linha que conecta valores de dados.
Caminho	Uma linha que conecta os valores de dados na ordem em que eles aparecem no conjunto de dados.
Área	Uma linha que conecta os elementos de dados com a área entre a linha e uma origem preenchida.
Polígono	Uma forma de múltiplos lados que engloba uma região de dados. Um elemento de polígono pode ser utilizado em um gráfico de dispersão categorizado ou em um mapa.
Esquema	Um elemento consiste em uma caixa com whiskers e marcadores indicando valores discrepantes. Um elemento de esquema é utilizado para boxplots.

Como Alterar a Estatística

1. Selecione o elemento gráfico cuja estatística deseja alterar.
2. Clique na guia **Elemento** na paleta de propriedades.

Como Especificar o Modificador de Colisão

O modificador de colisão determina o que acontece quando elementos gráficos se sobrepõem.

1. Selecione o elemento gráfico para o qual deseja especificar o modificador de colisão.
2. Clique na guia **Elemento** na paleta de propriedades.
3. Na lista suspensa do Modificador, selecione um modificador de colisão. **-auto-** permite que o aplicativo determine qual modificador de colisão é apropriado para o tipo de elemento gráfico e estatística.

Sobreposição. Desenhe elementos gráficos um sobre o outro quando eles tiverem o mesmo valor.

Pilha. Elementos gráficos de pilha que normalmente são sobrepostos quando têm os mesmos valores de dados.

Dodge. Move elementos gráficos ao lado de outros elementos gráficos que aparecem com o mesmo valor, ao invés de sobrepô-los. Os elementos gráficos são organizados simetricamente. Ou seja, os elementos gráficos são movidos para os lados opostos de uma posição central. O dodging é muito semelhante ao armazenamento em cluster.

Pile. Move elementos gráficos ao lado de outros elementos gráficos que aparecem com o mesmo valor, ao invés de sobrepô-los. Os elementos gráficos são organizados assimetricamente. Ou seja, os elementos gráficos são organizados um sobre o outro, com o elemento gráfico na parte inferior posicionado em um valor específico na escala.

Jitter (normal). Reposiciona aleatoriamente os elementos gráficos com o mesmo valor de dados utilizando uma distribuição normal.

Jitter (uniforme). Reposiciona aleatoriamente os elementos gráficos com o mesmo valor de dados utilizando uma distribuição uniforme.

Alterando a Posição da Legenda

Se um gráfico incluir uma legenda, a legenda será normalmente exibida à direita de um gráfico. É possível alterar esta posição se necessário.

Como Alterar a Posição da Legenda

1. Selecione a legenda.
2. Clique na guia **Legenda** na paleta de propriedades.
3. Selecione uma posição.

Copiando uma Visualização e Dados de Visualização

A paleta Geral inclui botões para copiar a visualização e seus dados.



Figura 64. Botão Copiar visualização

Copiando a visualização. Essa ação copia a visualização para a área de transferência como uma imagem. Diversos formatos de imagem estão disponíveis. Ao colar a imagem em outro aplicativo, é possível escolher uma opção "colar especial" para selecionar um dos formatos de imagem disponíveis para colar.



Figura 65. Botão Copiar dados de visualização

Copiando os dados de visualização. Essa ação copia os dados subjacentes que são utilizados para desenhar a visualização. Os dados são copiados para a área de transferência como texto simples ou texto formatado em HTML. Ao colar os dados em outro aplicativo, é possível escolher a opção "colar especial" para escolher um desses formatos para colar.

Atalhos de Teclado

Tabela 53. Atalhos de teclado

Tecla de Atalho	Função
Ctrl+Espaço	Alterna entre os modos Explorar e Edição
Excluir	Exclui um item de visualização
Ctrl+Z	Desfazer
Ctrl+Y	Refazer
F2	Exibe a estrutura de tópicos para selecionar itens no gráfico

Incluindo Títulos e Notas de Rodapé

Para todos os tipos de gráficos, é possível incluir rótulos de título, de nota de rodapé ou de eixo exclusivos para ajudar a identificar o que é mostrado no gráfico.

Incluindo Títulos nos Gráficos

1. Nos menus, escolha **Editar > Incluir Título do Gráfico**. Uma caixa de texto contendo <TITLE> é exibida acima do gráfico.
2. Verifique se você está no modo de Edição. Nos menus, escolha **Visualizar > Modo de Edição**.
3. Dê um clique duplo no texto <TITLE>.
4. Digite o título necessário e pressione Retornar.

Incluindo Notas de Rodapé nos Gráficos

1. Nos menus, escolha **Editar > Incluir Nota de Rodapé do Gráfico**. Uma caixa de texto contendo <FOOTNOTE> é exibida abaixo do gráfico.
2. Verifique se você está no modo de Edição. Nos menus, escolha **Visualizar > Modo de Edição**.
3. Dê um clique duplo no texto <FOOTNOTE>.
4. Digite o título necessário e pressione Retornar.

Utilizando Folhas de Estilo do Gráfico

O gráfico básico exibe informações, como cores, fontes, símbolos e espessura de linha, que são controladas por uma folha de estilo. Há uma folha de estilo padrão fornecida com o IBM SPSS Modeler, no entanto, é possível fazer mudanças nela se necessário. Por exemplo, você pode ter um esquema de cores corporativas para apresentações que deseja utilizar em seus gráficos. Consulte o tópico "Editando Visualizações" na página 285 para obter mais informações.

Nos nós de gráfico, é possível utilizar o modo de Edição para fazer mudanças de estilo na aparência de um gráfico. Em seguida, é possível utilizar o menu **Editar > Estilos** para salvar as mudanças como uma

folha de estilo a serem aplicadas a todos os gráficos que forem gerados posteriormente a partir do nó de gráfico atual, ou como uma nova folha de estilo padrão para todos os gráficos que forem produzidos usando o IBM SPSS Modeler.

Há cinco opções de folha de estilo disponíveis a partir da opção **Estilos** no menu Editar:

- **Alternar Folha de Estilo.** Exibe uma lista de folhas de estilo diferentes armazenadas que podem ser selecionadas para alterar a aparência de seus gráficos. Consulte o tópico “Aplicando Folhas de Estilo” para obter mais informações.
- **Armazenar Estilos no Nó.** Isso armazena modificações nos estilos do gráfico selecionado para que elas sejam aplicadas a quaisquer gráficos futuros criados a partir do mesmo nó de gráfico no fluxo atual.
- **Armazenar Estilos como Padrão.** Isso armazena modificações nos estilos do gráfico selecionado para que elas sejam aplicadas a todos os gráficos futuros criados a partir de qualquer nó de gráfico em qualquer fluxo. Após selecionar essa opção, será possível utilizar **Aplicar Estilos Padrão** para alterar quaisquer outros gráficos existentes para utilizar os mesmos estilos.
- **Aplicar Estilos Padrão.** Isso altera os estilos do gráfico selecionado para aqueles que estão atualmente salvos como os estilos padrão.
- **Aplicar Estilos Originais.** Isso altera os estilos de um gráfico de volta para aqueles fornecidos como o padrão original.

Aplicando Folhas de Estilo

É possível aplicar uma folha de estilo de visualização que especifica as propriedades de estilo da visualização. Por exemplo, a folha de estilo pode definir fontes, traços e cores, entre outras opções. Até certo ponto, as folhas de estilo fornecem um atalho para as edições que tiverem que ser executadas manualmente. Observe, no entanto, que uma folha de estilo é limitada a mudanças de *estilo*. Outras mudanças, como a posição da legenda ou o intervalo de escala, não são armazenadas na folha de estilo.

Como Aplicar uma Folha de Estilo

1. A partir dos menus, escolha:
Editar > Estilos > Alternar Folha de Estilo
2. Utilize a caixa de diálogo Alternar Folha de Estilo para selecionar uma folha de estilo.
3. Clique em **Aplicar** para aplicar a folha de estilo à visualização sem fechar o diálogo. Clique em **OK** para aplicar a folha de estilo e fechar a caixa de diálogo.

Caixa de Diálogo Alternar/Selecionar Folha de Estilo

A tabela na parte superior da caixa de diálogo lista todas as folhas de estilo de visualização que estiverem atualmente disponíveis. Algumas folhas de estilo são pré-instaladas, ao passo que outras podem ter sido criadas no IBM SPSS Visualization Designer (um produto separado).

A parte inferior da caixa de diálogo mostra visualizações de exemplo com dados de amostra. Selecione uma das folhas de estilo para aplicar seus estilos às visualizações de exemplo. Esses exemplos podem ajudá-lo a determinar como a folha de estilo afetará sua visualização real.

A caixa de diálogo também oferece as seguintes opções.

Estilos Existentes. Por padrão, uma folha de estilo pode sobrescrever todos os estilos na visualização. É possível alterar esse comportamento.

- **Sobrescrever todos os estilos.** Ao aplicar a folha de estilo, sobrescreve todos os estilos na visualização, incluindo os estilos modificados na visualização durante a sessão de edição atual.
- **Preservar estilos modificados.** Ao aplicar a folha de estilo, sobrescreve apenas os estilos que *não* foram modificados na visualização durante a sessão de edição atual. Os estilos que foram modificados durante a sessão de edição atual são preservados.

Gerenciar. Gerencie modelos de visualização , folhas de estilo e mapas em seu computador. É possível importar, exportar, renomear e excluir modelos de visualização , folhas de estilo e mapas em sua máquina local. Consulte o tópico “Gerenciando Modelos, Folhas de Estilo e Arquivos de Mapeamento” na página 224 para obter mais informações.

Localização. Altere a localização na qual os modelos de visualização , folhas de estilo e mapas são armazenados. A localização atual é listada à direita do botão. Consulte o tópico “Configurando a Localização dos Modelos, Folhas de Estilos e Mapas” na página 223 para obter mais informações.

Imprimindo, Salvando, Copiando e Exportando Gráficos

Cada gráfico possui um número de opções que permitem salvar ou imprimir o gráfico ou exportá-lo para outro formato. A maioria dessas opções está disponíveis a partir do menu Arquivo. Além disso, no menu Editar, é possível escolher copiar o gráfico ou os dados dentro dele para utilizar em outro aplicativo.

Imprimindo

Para imprimir o gráfico, utilize o item de menu ou o botão **Imprimir**. Antes de imprimir, é possível utilizar **Configuração da Página** e **Visualização de Impressão** para configurar as opções de impressão e visualizar a saída.

Salvando Gráficos

Para salvar o gráfico em um arquivo de saída do IBM SPSS Modeler (*.cou), escolha **Arquivo > Salvar** ou **Arquivo > Salvar Como** a partir dos menus.

ou

Para salvar o gráfico no repositório, escolha **Arquivo > Armazenar Saída** a partir dos menus.

Copiando Gráficos

Para copiar o gráfico para uso em outro aplicativo, como o MS Word ou MS PowerPoint, escolha **Edit > Copiar Gráfico** a partir dos menus.

Copiando Dados

Para copiar os dados para uso em outro aplicativo, como o MS Excel ou MS Word, escolha **Edit > Copiar Dados** a partir dos menus. Por padrão, os dados serão formatados como HTML. Utilize **Colar Especial** no outro aplicativo para visualizar outras opções de formatação ao colar.

Exportando Gráficos

A opção **Exportar Gráfico** permite exportar o gráfico em um dos seguintes formatos: Bitmap (.bmp), JPEG (.jpg), PNG (.png), HTML (.html), ou documento ViZml (.xml) para uso em outros aplicativos do IBM SPSS Statistics.

Para exportar gráficos, escolha **Arquivo > Exportar Gráfico** a partir dos menus e, em seguida, escolha o formato.

Exportando Tabelas

A opção **Exportar Tabela** permite exportar a tabela em um dos seguintes formatos: delimitado por guia (.guia), delimitado por vírgula (.csv) ou HTML (.html).

Para exportar tabelas, escolha **Arquivo > Exportar Tabela** a partir dos menus e, em seguida, escolha o formato.

Capítulo 6. Nós de Saída

Visão Geral dos Nós de Saída

Os nós de saída fornecem os meios para obter informações sobre seus dados e modelos. Eles também fornecem um mecanismo para exportar dados em vários formatos para estabelecer interface com as outras ferramentas de software.

Os nós de saída a seguir estão disponíveis:



O nó Tabela exibe os dados em formato de tabela, que também podem ser gravados em um arquivo. Isso é útil a qualquer momento em que precisar inspecionar seu valores de dados ou exportá-los em um formato facilmente legível.



O nó Matriz cria uma tabela que mostra relacionamentos entre campos. Ele é usado mais normalmente para mostrar o relacionamento entre dois campos simbólicos, como também pode mostrar relacionamentos entre campos sinalizadores ou campos numéricos.



O nó Análise avalia a capacidade de modelos preditivos de gerar previsões exatas. Os nós de análise executam várias comparações entre os valores preditos e os valores reais para um ou mais nuggets do modelo. Eles também podem comparar modelos preditivos entre si.



O nó Auditoria de Dados fornece uma primeira visão abrangente dos dados, incluindo estatísticas de resumo, histogramas e distribuição para cada campo, bem como informações sobre valores discrepantes, valores omissos e valores extremos. Os resultados são exibidos em uma matriz de fácil leitura e podem ser classificados e utilizados para gerar gráficos de tamanho completo e nós de preparação de dados.



O nó Transformar permite selecionar e visualizar os resultados das transformações antes de aplicá-las aos campos selecionados.



O nó Estatísticas fornece informações de resumo básicas sobre campos numéricos. Ela calcula as estatísticas de resumo para campos individuais e correlações entre os campos.



O nó Média compara a média entre grupos independentes ou entre pares de campos relacionados para testar se há uma diferença significativa. Por exemplo, é possível comparar receitas médias antes e depois de realizar uma promoção ou comparar receitas de clientes que não receberam a promoção com aqueles que receberam.



O nó Relatório cria relatórios formatados contendo texto fixo, bem como dados e outras expressões derivadas dos dados. Especifique o formato do relatório utilizando os modelos de texto para definir as construções de saída de texto e de dados fixos. É possível fornecer formatação de texto customizada utilizando tags HTML no modelo e configurando as opções na guia Saída. É possível incluir valores de dados e de outra saída condicional utilizando expressões do CLEM no modelo.



O nó Configurar Globais varre os dados e calcula os valores de resumo que podem ser utilizados em expressões do CLEM. Por exemplo, é possível utilizar esse nó para calcular as estatísticas para um campo chamado *age* e, em seguida, utilizar a média geral de *age* em expressões do CLEM ao inserir a função @GLOBAL_MEAN(*age*).



O nó Ajuste de Simulação examina a distribuição estatística dos dados em cada campo e gera (ou atualiza) um nó Gerar Simulação com a melhor distribuição de ajuste designada a cada campo. Em seguida, o nó Gerar Simulação poderá ser utilizado para gerar dados simulados.



O nó Avaliação de Simulação avalia um campo de destino previsto especificado e apresenta informações de distribuição e de correlação sobre o campo de destino.

Gerenciando a saída

O gerenciador de Saída mostra os diagramas, gráficos e tabelas gerados durante uma sessão do IBM SPSS Modeler. É possível reabrir uma saída sempre dando um clique duplo nela no gerenciador—não é necessário executar novamente o fluxo ou o nó correspondente.

Para Visualizar o Gerenciador de Saída

Abra o menu Visualizar e escolha **Gerenciadores**. Clique na guia **Saídas**.

No Gerenciador de Saída, é possível:

- Exibir objetos de saída existentes, como histogramas, gráficos de avaliação e tabelas.
- Renomear objetos de saída.
- Salvar objetos de saída no disco ou no IBM SPSS Collaboration and Deployment Services Repository (se disponível).
- Incluir arquivos de saída no projeto atual.
- Excluir objetos de saída não salvos da sessão atual.
- Abrir objetos de saída salvos ou recuperá-los a partir do IBM SPSS Collaboration and Deployment Services Repository (se disponível).

Para acessar essas opções, clique com o botão direito em qualquer lugar na guia Saídas.

Visualizando a Saída

A saída na tela é exibida como uma janela do navegador de saída. A janela do navegador de saída possui seu próprio conjunto de menus que permitem imprimir ou salvar a saída ou exportá-la em outro formato. Observe que opções específicas podem variar dependendo do tipo de saída.

Imprimindo, salvando e exportando dados. Mais informações estão disponíveis, conforme a seguir:

- Para imprimir a saída, utilize a opção de menu ou botão **Imprimir**. Antes de imprimir, é possível utilizar **Configuração da Página** e **Visualização de Impressão** para configurar as opções de impressão e visualizar a saída.
- Para salvar a saída em um arquivo de saída do IBM SPSS Modeler (.cou), escolha **Salvar** ou **Salvar Como** no menu Arquivo.
- Para salvar a saída em outro formato, como texto ou HTML, escolha **Exportar** no menu Arquivo. Consulte o tópico “Exportando Saída” na página 304 para obter mais informações.
Observe que esses formatos poderão ser selecionados somente se a saída contiver dados que possam ser exportados sensivelmente dessa forma. Por exemplo, o conteúdo de uma árvore de decisão pode ser exportado como texto, mas o conteúdo de um modelo K-médias não faria sentido como texto.
- Para salvar a saída em um repositório compartilhado para que outros usuários possam visualizá-la utilizando o IBM SPSS Collaboration and Deployment Services Deployment Portal, escolha **Publicar na Web** no menu Arquivo. Observe que essa opção requer uma licença separada para o IBM SPSS Collaboration and Deployment Services.

Selecionando células e colunas. O menu Editar contém várias opções para selecionar, cancelar a seleção e copiar as células e colunas, conforme apropriado para o tipo de saída atual. Consulte o tópico “Selecionando células e colunas” na página 305 para obter mais informações.

Gerando novos nós. O menu Gerar permite gerar novos nós com base no conteúdo do navegador de saída. As opções variam dependendo do tipo de saída e dos itens na saída que estiverem atualmente selecionadas. Para obter detalhes sobre as opções de geração do nó para um tipo específico de saída, consulte a documentação para essa saída.

Publicação na Web

O recurso Publicação na Web permite publicar determinados tipos de saída de fluxo em um IBM SPSS Collaboration and Deployment Services Repository central compartilhado que forma a base do IBM SPSS Collaboration and Deployment Services. Se utilizar essa opção, outros usuários que precisarem visualizar essa saída poderão fazer isso utilizando acesso à Internet e uma conta do IBM SPSS Collaboration and Deployment Services--eles não precisam ter o IBM SPSS Modeler instalado.

A tabela a seguir lista os nós do IBM SPSS Modeler que suportam o recurso de Publicação na Web. A saída desses nós é armazenada no IBM SPSS Collaboration and Deployment Services Repository em formato de objeto de saída (.cou) e pode ser visualizada diretamente no IBM SPSS Collaboration and Deployment Services Deployment Portal.

Outros tipos de saída poderão ser visualizados apenas se o aplicativo relevante (por exemplo, IBM SPSS Modeler, para objetos de fluxo) estiver instalado na máquina do usuário.

Tabela 54. Nós que suportam Publicação na Web.

Tipo do nó	Nó
Gráficos	todos
Out	Tabela
	Matriz
	Auditoria de dados
	Transformar
	Médias
	Análise
	Statistics
	Relatório (HTML)
IBM SPSS Statistics	Saída de estatísticas

Publicando a saída na web

Para publicar a saída na web:

1. Em um fluxo do IBM SPSS Modeler, execute um dos nós listados na tabela. Fazer isso cria um objeto de saída (por exemplo, uma tabela, matriz ou objeto de relatório) em uma nova janela.

2. Na janela de objeto de saída, escolha:

Arquivo > Publicar na Web

Nota: se desejar apenas exportar arquivos HTML simples para uso com um navegador da web padrão, escolha **Exportar** no menu Arquivo e selecione **HTML**.

3. Conecte-se ao IBM SPSS Collaboration and Deployment Services Repository.

Ao conectar-se com sucesso, o diálogo Repositório: Armazenar será exibido, oferecendo um número de opções de armazenamento.

4. Após escolher as opções de armazenamento desejadas, clique em **Armazenar**.

Visualizando saída publicada pela web

Deve-se ter uma conta do IBM SPSS Collaboration and Deployment Services configurada para usar esse recurso. Se você tiver o aplicativo relevante instalado para o tipo de objeto que deseja visualizar (por exemplo, IBM SPSS Modeler ou IBM SPSS Statistics), a saída será exibida no próprio aplicativo ao invés de ser exibida no navegador.

Para visualizar a saída publicada na web:

1. Aponte o navegador para `http://<repos_host>:<repos_port>/peb`

em que *repos_host* e *repos_port* são os nomes do host e o número da porta para o host do IBM SPSS Collaboration and Deployment Services.

2. Insira os detalhes de login de sua conta do IBM SPSS Collaboration and Deployment Services.

3. Clique em **Repositório de Conteúdo**.

4. Navegue para ou procure pelo objeto que deseja visualizar.

5. Clique no nome do objeto. Para alguns tipos de objetos, como gráficos, poderá haver um atraso enquanto o objeto é renderizado no navegador.

Visualizando a Saída em um Navegador HTML

Na guia Avançado dos nuggets do modelo Linear, Logística e PCA/Fator, é possível visualizar as informações exibidas em um navegador separado, como o Internet Explorer. As informações são geradas como HTML, permitindo salvá-la e reutilizá-lo em outro local, como em uma intranet corporativa ou em um site da Internet.

Para exibir as informações em um navegador, clique no botão Ativar, situado abaixo do ícone do modelo na parte superior esquerda da guia Avançado do nugget do modelo.

Exportando Saída

Na janela do navegador de saída, é possível escolher exportar a saída em outro formato, como texto ou HTML. Os formatos de exportação variam de acordo com o tipo de saída, mas, em geral, são semelhantes às opções de tipo de arquivo disponíveis se você selecionar **Salvar no arquivo** no nó usado para gerar a saída.

Nota: Esses formatos poderão ser selecionados somente se a saída contiver dados que possam ser exportados sensivelmente dessa forma. Por exemplo, o conteúdo de uma árvore de decisão pode ser exportado como texto, mas o conteúdo de um modelo K-médias não faria sentido como texto.

Para Exportar a Saída

1. No navegador de saída, abra o menu Arquivo e escolha **Exportar**. Em seguida, escolha o tipo de arquivo que deseja criar:

- **Delimitado por Tabulações (*.tab)**. Esta opção gera um arquivo de texto formatado contendo os valores de dados. Este estilo é sempre útil para gerar uma representação de texto simples das informações que podem ser importadas em outros aplicativos. Esta opção está disponível para os nós Tabela, Matriz e Médias.
- **Delimitado por Vírgulas (*.dat)**. Esta opção gera um arquivo de texto delimitado por vírgulas contendo os valores de dados. Esse estilo é muitas vezes útil como uma maneira rápida de gerar um arquivo de dados que possa ser importado em planilhas ou em outros aplicativos de análise de dados. Esta opção está disponível para os nós Tabela, Matriz e Médias.
- **Delimitado por Tabulações Transpostas (*.tab)**. Esta opção é idêntica à opção Delimitado por Tabulações, porém os dados são transpostos para que as linhas representem os campos e as colunas representem os registros.
- **Delimitado por Vírgulas Transpostas (*.dat)**. Esta opção é idêntica à opção Delimitado por Vírgulas, porém os dados são transpostos para que as linhas representem os campos e as colunas representem os registros.
- **HTML (*.html)**. Esta opção grava a saída formatada em HTML em um ou mais arquivos.

Selecionando células e colunas

Diversos nós, incluindo o nó Tabela, o nó Matriz e o nó Médias, geram resultado tabular. Essas tabelas de saída podem ser visualizadas e manipuladas de maneiras semelhantes, incluindo selecionar células, copiar toda ou parte da tabela para a Área de Transferência, gerar novos nós com base na seleção atual e salvar e imprimir a tabela.

Selecionando células. Para selecionar uma célula, clique nela. Para selecionar um intervalo retangular de células, clique em um canto do intervalo desejado, arraste o mouse para o outro canto do intervalo e solte o botão do mouse. Para selecionar uma coluna inteira, clique no título da coluna. Para selecionar diversas colunas, utilize Shift + clique ou Ctrl + clique nos títulos da coluna.

Ao fazer uma nova seleção, a seleção antiga é limpa. Ao manter a tecla Ctrl pressionada enquanto seleciona, é possível incluir a nova seleção em qualquer seleção existente ao invés de limpar a seleção antiga. É possível usar esse método para selecionar diversas regiões não contínuas na tabela. O menu Editar também contém as opções **Selecionar Tudo** e **Limpar Seleção**.

Reordenando colunas. Os navegadores de saída do nó Tabela e do nó Médias permitem mover colunas na tabela clicando em um título da coluna e arrastando-a para a localização desejada. É possível mover somente uma coluna por vez.

Nó Tabela

O nó Tabela cria uma tabela que lista os valores em seus dados. Todos os campos e todos os valores no fluxo são incluídos, o que facilita inspecionar seu valores de dados ou exportá-los em um formato facilmente legível. Opcionalmente, é possível destacar registros que atendam a uma determinada condição.

Nota: A menos que você esteja trabalhando com conjuntos de dados pequenos, é recomendado selecionar um subconjunto dos dados para transmiti-lo para o nó Tabela. O nó Tabela não pode ser exibido corretamente quando o número de registros supera um tamanho que possa estar contido na estrutura de exibição (por exemplo, 100 milhões de linhas).

Guia Configurações do Nó Tabela

Destacar registros onde. É possível destacar registros na tabela ao inserir uma expressão do CLEM que é verdadeira para os registros a serem destacados. Esta opção é ativada apenas quando **Gerar para tela** estiver selecionada.

Guia Formato do Nó da Tabela

A guia Formato contém opções usadas para especificar a formatação em uma base por campo. Essa guia é compartilhada com o nó de Tipo. Consulte o tópico “Guia Configurações de Formato de Campo” na página 143 para obter mais informações.

Guia Saída do Nó Saída

Para nós que geram a saída em estilo de tabela, a guia Saída permite especificar o formato e a localização dos resultados.

Nome de saída. Especifica o nome da saída produzido quando o nó é executado. **Automático** escolhe um nome com base no nó que gera a saída. Opcionalmente, é possível selecionar **Customizado** para especificar um nome diferente.

Saída para tela (o padrão). Cria um objeto de saída para visualizar on-line. O objeto de saída aparecerá na guia Saídas da janela do gerenciador quando o nó de saída for executado.

Saída para arquivo. Salva a saída em um arquivo quando o nó é executado. Se escolher essa opção, insira um nome de arquivo (ou navegue para um diretório e especifique um nome de arquivo utilizando o botão Seletor de Arquivo) e selecione um tipo de arquivo. Observe que alguns tipos de arquivo podem estar indisponíveis para determinados tipos de saída.

Os dados são gerados no formato de codificação padrão do sistema, que é especificado no Painel de Controle do Windows ou, se estiver executando no modo distribuído, no computador servidor.

- **Dados (delimitados por tabulação) (*.tab).** Esta opção gera um arquivo de texto formatado contendo os valores de dados. Este estilo é muitas vezes útil para gerar uma representação em texto simples das informações que pode ser importada em outros aplicativos. Esta opção está disponível para os nós Tabela, Matriz e Médias.
- **Dados (delimitado por vírgula) (*.dat).** Esta opção gera um arquivo de texto delimitado por vírgulas contendo os valores de dados. Esse estilo é muitas vezes útil como uma maneira rápida de gerar um arquivo de dados que possa ser importado em planilhas ou em outros aplicativos de análise de dados. Esta opção está disponível para os nós Tabela, Matriz e Médias.
- **HTML (*.html).** Esta opção grava a saída formatada em HTML em um ou mais arquivos. Para saída tabular (a partir dos nós Tabela, Matriz ou Médias), um conjunto de arquivo HTML possui um painel de conteúdo listando os nomes de campo e os dados em uma tabela HTML. A tabela poderá ser dividida em vários arquivos HTML se o número de linhas na tabela exceder a especificação de **Linhas por página**. Nesse caso, o painel de conteúdo contém ligações para todas as páginas da tabela e fornece um meio de navegar na tabela. Para saída não tabular, um arquivo HTML único é criado contendo os resultados do nó.

Nota: se a saída HTML contiver formatação apenas para a primeira página, selecione **Paginar saída** e ajuste a especificação **Linhas por página** para incluir toda a saída em uma única página. Ou, se o modelo de saída para nós como o nó Relatório contiver tags HTML customizadas, assegure-se de ter especificado **Customizado** como o tipo de formato.

- **Arquivo de Texto (*.txt).** Esta opção gera um arquivo de texto contendo a saída. Esse estilo é geralmente útil para gerar a saída que pode ser importada em outros aplicativos, como processadores de texto ou software de apresentação. Esta opção não está disponível para alguns nós.
- **Objeto de saída (*.cou).** Os objetos de saída salvos nesse formato podem ser abertos e visualizados no IBM SPSS Modeler, incluídos nos projetos e publicados e controlados usando o IBM SPSS Collaboration and Deployment Services Repository.

Visualização de saída. Para o nó Médias, é possível especificar se a saída simples ou avançada é exibida por padrão. Observe que também é possível alternar entre essas visualizações ao procurar a saída gerada. Consulte o tópico “Navegador de Saída do Nó Médias” na página 327 para obter mais informações.

Formato. Para o nó Relatório, é possível escolher se a saída é formatada automaticamente ou formatada utilizando HTML incluído no modelo. Selecione **Customizado** para permitir formatação HTML no modelo.

Título. Para o nó Relatório, é possível especificar o texto do título opcional que aparecerá na parte superior da saída de relatório.

Destacar texto inserido. Para o nó Relatório, selecione esta opção para destacar o texto gerado pelas expressões do CLEM no modelo de Relatório. Consulte o tópico “Guia Modelo do Nó Relatório” na página 329 para obter mais informações. Esta opção não é recomendada ao utilizar a formatação **Customizado**.

Linhas por página. Para o nó Relatório, especifique um número de linhas a serem incluídas em cada página durante a formatação **Automática** do relatório de saída.

Transpor dados. Esta opção transpõe os dados antes da exportação, para que as linhas representem os campos e as colunas representem os registros.

Nota: para tabelas grandes, as opções acima podem ser um pouco ineficientes, especialmente quando trabalhar com um servidor remoto. Nesses casos, o uso de um nó de saída Arquivo fornece um desempenho muito melhor. Consulte o tópico “Nó Exportação de Arquivo Simples” na página 358 para obter mais informações.

Navegador de Tabela

O navegador de tabela exibe dados tabulares e permite executar operações padrão, incluindo selecionar e copiar células, reordenar colunas e salvar e imprimir a tabela. Consulte o tópico “Selecionando células e colunas” na página 305 para obter mais informações. Essas são as mesmas operações que podem ser realizadas quando visualizar os dados em um nó.

Exportando dados da tabela. É possível exportar dados do navegador de tabela ao escolher:

Arquivo > Exportar

Consulte o tópico “Exportando Saída” na página 304 para obter mais informações.

Os dados são exportados no formato de codificação padrão do sistema, que é especificado no Painel de Controle do Windows ou, se estiver executando no modo distribuído, no computador servidor.

Procurando na tabela. O botão de procura (com o ícone de binóculos) na barra de ferramentas principal ativa a barra de ferramentas de procura, permitindo procurar por valores específicos na tabela. É possível procurar para frente ou para trás na tabela, especificar uma procura com distinção entre maiúsculas e minúsculas (o botão **Aa**) e interromper uma procura em andamento com o botão de procura de interrupção.

Gerando novos nós. O menu Gerar contém operações de geração do nó.

- **Nó Seleção ("Registros").** Gera um nó Seleção que seleciona os registros para os quais qualquer célula na tabela é selecionada.
- **Selecionar ("And").** Gera um nó Seleção que seleciona registros que contenham *todos* os valores selecionados na tabela.
- **Selecionar ("Or").** Gera um nó Seleção que seleciona registros que contenham *qualquer* um dos valores selecionados na tabela.
- **Derivar ("Registros").** Gera um nó Derivar para criar um novo campo de sinalização. O campo de sinalização contém *T* para registros para os quais qualquer célula na tabela é selecionada e *F* para os registros restantes.

- **Derivar ("And").** Gera um nó Derivar para criar um novo campo de sinalização. O campo de sinalização contém *T* para registros contendo *todos* os valores selecionados na tabela e *F* para os registros restantes.
- **Derivar ("Or").** Gera um nó Derivar para criar um novo campo de sinalização. O campo de sinalização contém *T* para registros contendo *qualquer* um dos valores selecionados na tabela e *F* para os registros restantes.

Nó Matriz

O nó Matriz permite criar uma tabela que mostra os relacionamentos entre os campos. É mais comumente utilizado para mostrar o relacionamento entre dois campos categóricos (flag, nominal ou ordinal), mas também pode ser utilizado para mostrar relacionamentos entre campos contínuos (intervalo numérico).

Guia Configurações do Nó Matriz

A guia Configurações permite especificar opções para a estrutura da matriz.

Campos. Selecione um tipo de seleção de campo a partir de uma das opções a seguir:

- **Selecionado.** Esta opção permite selecionar um campo categórico para as linhas e outro para as colunas da matriz. As linhas e colunas da matriz são definidas pela lista de valores para o campo categórico selecionado. As células da matriz contêm as estatísticas de sumarização selecionadas abaixo.
- **Todos os flags (valores verdadeiros).** Esta opção solicita uma matriz com uma linha e uma coluna para cada campo de sinalização nos dados. As células da matriz contêm as contagens de positivos duplos para cada combinação de flag. Em outras palavras, para uma linha correspondente a *comprou pão* e uma coluna correspondente a *comprou queijo*, a célula na interseção dessa linha e coluna contém o número de registros para os quais *comprou pão* e *comprou queijo* são verdadeiros.
- **Todos numéricos.** Esta opção solicita uma matriz com uma linha e uma coluna para cada campo numérico. As células da matriz representam a soma dos produtos cruzados para o par de campos correspondente. Em outras palavras, para cada célula na matriz, os valores do campo de linha e do campo de coluna são multiplicados para cada registro e, em seguida, somados entre os registros.

Incluir valores omissos. Especifica se valores omissos do usuário (em branco) e omissos do sistema (nulos) são incluídos na saída da linha e da coluna. Por exemplo, se o valor *N/A* foi definido como omissos do usuário para o campo de coluna selecionado, uma coluna separada com o rótulo *N/A* será incluída na tabela (supondo que esse valor realmente ocorra nos dados) assim como qualquer outra categoria. Se esta opção for desmarcada, a coluna *N/A* será excluída, independentemente de quantas vezes ela ocorre.

Nota: a opção para incluir valores omissos se aplica apenas quando os campos selecionados tiverem tabulação cruzada. Valores em branco são mapeados para \$null\$ e são excluídos da agregação para o campo de função quando o modo for **Selecionado** e o conteúdo for configurado para **Função**, e para todos os campos numéricos quando o modo for configurado como **Todos Numéricos**.

Conteúdo da célula. Se você tiver escolhido **Selecionado** para os campos acima, será possível especificar a estatística a ser utilizada nas células da matriz. Selecione uma estatística baseada em contagem ou selecione um campo de sobreposição para sumarizar os valores de um campo numérico com base nos valores dos campos de linha e de coluna.

- **Tabulações cruzadas.** Os valores da célula são contagens e/ou porcentagens de quantos registros possuem a combinação correspondente de valores. É possível especificar as sumarizações de tabulação cruzada que desejar utilizando as opções na guia Aparência. O valor qui-quadrado global também é exibido junto do significado. Consulte o tópico “Navegador de Saída do Nó Matriz” na página 309 para obter mais informações.
- **Função.** Se você selecionar uma função de sumarização, os valores da célula serão uma função dos valores do campo de sobreposição selecionado para os casos que possuem os valores de linha ou de

coluna apropriados. Por exemplo, se o campo de linha for *Região*, o campo de coluna for *Produto* e o campo de sobreposição for *Receita*, então a célula na linha *Nordeste* e a coluna *Widget* conterão a soma (ou a média, mínima ou máxima) da receita dos widgets vendidos na região nordeste. A função de sumarização padrão é **Média**. É possível selecionar outra função para sumarizar o campo de função. As opções incluem **Média**, **Soma**, **SDev** (desvio padrão), **Máx** (máximo) e **Mín** (mínimo).

Guia Aparência do Nó Matriz

A guia Aparência permite controlar as opções de ordenação e de destaque para a matriz, bem como as estatísticas apresentadas para matrizes de tabulação cruzada.

Linhas e colunas. Controla a ordenação dos títulos de linha e de coluna na matriz. O padrão é **Não ordenado**. Selecione **Crescente** ou **Decrescente** para ordenar os títulos da linha e da coluna na direção especificada.

Sobreposição. Permite destacar valores extremos na matriz. Os valores são destacados com base nas contagens de célula (para matrizes de tabulação cruzada) ou em valores calculados (para matrizes de função).

- **Destacar superior.** É possível solicitar que os valores mais altos na matriz sejam destacados (em vermelho). Especifique o número de valores a serem destacados.
- **Destacar inferior.** Também é possível solicitar que os valores mais baixos na matriz sejam destacados (em verde). Especifique o número de valores a serem destacados.

Nota: para as duas opções de destaque, os empates podem fazer com que mais valores que o solicitado sejam destacados. Por exemplo, se você tiver uma matriz com seis zeros entre as células e solicitar **Destacar 5 valores da parte inferior**, todos os seis zeros serão destacados.

Conteúdo da célula de tabulação cruzada. Para tabulações cruzadas, é possível especificar as estatísticas básicas contidas na matriz para matrizes de tabulação cruzada. Essas opções não estão disponíveis quando a opção **Tudo Numérico** ou **Função** é selecionada na guia Configurações.

- **Contagens.** As células incluem o número de registros com o valor da linha que possuem o valor da coluna correspondente. Este é o único conteúdo da célula padrão.
- **Valores esperados.** O valor esperado para o número de registros na célula, supondo que não exista nenhum relacionamento entre as linhas e colunas. Os valores esperados baseiam-se na fórmula a seguir:

$$p(\text{row value}) * p(\text{column value}) * \text{total number of records}$$

- **Residuais.** A diferença entre os valores observados e esperados.
- **Porcentagem de linha.** A porcentagem de todos os registros com o valor da linha que possuem o valor da coluna correspondente. As porcentagens somam 100 dentro das linhas.
- **Porcentagem de coluna.** A porcentagem de todos os registros com o valor da coluna que possuem o valor da linha correspondente. As porcentagens somam 100 dentro das colunas.
- **Porcentagem do total.** A porcentagem de todos os registros que possuem a combinação de valor de coluna e valor de linha. As porcentagens somam 100 ao longo de toda a matriz.
- **Incluir totais de linha e de coluna.** Inclui uma linha e uma coluna na matriz para totais de coluna e linha.
- **Aplicar Configurações.** (somente Navegador de Saída) Permite você para fazer alterações na aparência da saída da Matriz de nó sem precisar fechar e reabrir o navegador de saída. Faça as mudanças nesta guia do Navegador de saída, clique neste botão e, em seguida, selecione a guia Matriz para ver o efeito das mudanças.

Navegador de Saída do Nó Matriz

O navegador de matriz exibe dados de tabulação cruzada e permite executar operações na matriz, incluindo selecionar células, copiar a matriz para a Área de Transferência total ou parcialmente, gerar

novos nós com base na seleção da matriz e salvar e imprimir a matriz. O navegador da matriz também pode ser utilizado para exibir a saída a partir de determinados modelos, como modelos Naive Bayes da Oracle.

Os menus Arquivo e Editar fornecem opções comuns para imprimir, salvar e exportar a saída, bem como selecionar e copiar os dados. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.

Qui-quadrado. Para uma tabulação cruzada de dois campos categóricos, o qui-quadrado de Pearson global também é exibido abaixo da tabela. Esse teste indica a probabilidade de que dois campos não estejam relacionados, com base na diferença entre as contagens observadas e as contagens esperadas, se nenhum relacionamento existir. Por exemplo, se não houver nenhum relacionamento entre a satisfação do cliente e a localização da loja, você espera taxas de satisfação semelhantes em todas as lojas. No entanto, se os clientes em determinadas lojas relatarem consistentemente taxas mais altas que em outras, você pode suspeitar que isso não foi uma coincidência. Quanto maior a diferença, menor a probabilidade de que isso tenha sido resultado apenas de um erro de amostragem de tentativa.

- O teste qui-quadrado indica a probabilidade de que os dois campos não estejam relacionados, caso em que quaisquer diferenças entre as frequências observadas e esperadas são o resultado somente de uma tentativa. Se esta probabilidade for muito pequena, geralmente menor que 5%, então o relacionamento entre os dois campos é considerado significativo.
- Se houver apenas uma coluna ou linha (um chi-quadrado unidirecional), os graus de liberdade serão o número de células menos um. Para qui-quadrado bidirecional, os graus de liberdade são o número de linhas menos um vezes o número de colunas menos um.
- Tenha cuidado ao interpretar a estatística qui-quadrado se qualquer uma das frequências de célula esperadas for menor que cinco.
- O teste qui-quadrado está disponível somente para uma tabulação cruzada de dois campos. (Quando **Todos flags** ou **Todos os números** estiver selecionada na guia Configurações, este teste não será exibido).

Menu Gerar. O menu Gerar contém operações de geração do nó. Estas operações estão disponíveis apenas para matrizes de tabulação cruzada, e pelo menos uma célula deverá estar selecionada na matriz.

- **Nó Seleção.** Gera um nó Seleção que seleciona os registros que correspondem a qualquer célula selecionada na matriz.
- **Nó Derivar (Flag).** Gera um nó Derivar para criar um novo campo de sinalização. O campo de sinalização contém T para registros que correspondem a qualquer célula selecionada na matriz e F para os registros restantes.
- **Nó Derivar (Conjunto).** Gera um nó Derivar para criar um novo campo nominal. O campo nominal contém uma categoria para cada conjunto contínuo de células selecionadas na matriz.

Nó Análise

O nó Análise permite avaliar a capacidade de um modelo para gerar predições exatas. Os nós de análise executam várias comparações entre os valores preditos e os valores reais (seu campo de destino) para um ou mais nuggets do modelo. Os nós Análise também podem ser utilizados para comparar modelos preditivos com outros modelos preditivos.

Ao executar um nó Análise, uma sumarização dos resultados da análise é automaticamente incluída na seção Análise na guia Sumarização de cada nugget do modelo no fluxo executado. Os resultados da análise detalhada aparecem na guia Saídas da janela do gerenciador ou podem ser gravadas diretamente em um arquivo.

Nota: como os nós Análise comparam valores preditos com valores reais, eles são úteis apenas com modelos supervisionados (aqueles que requerem um campo de destino). Para modelos não supervisionados, como algoritmos de armazenamento em cluster, não há nenhum resultado real disponível para uso como base para comparação.

Guia Análise do Nó Análise

A guia Análise permite especificar os detalhes da análise.

Matriz de coincidência (para variáveis resposta simbólicas ou categóricas). Mostra o padrão de correspondências entre cada campo gerado (predito) e seu campo de destino para variáveis resposta categóricas (flag, nominal ou ordinal). Uma tabela é exibida com as linhas definidas pelos valores reais e as colunas definidas pelos valores preditos, com o número de registros que possuem esse padrão em cada célula. Isso é útil para identificar erros sistemáticos de predição. Se houver mais de um campo gerado relacionado ao mesmo campo de saída, mas produzido por modelos diferentes, os casos em que esses campos concordam e discordam são contados e os totais exibidos. Para os casos em que eles concordarem, outro conjunto de estatísticas corretas/erradas é exibido.

Avaliação de desempenho. Mostra estatísticas de avaliação de desempenho para modelos com saídas categóricas. Essa estatística, relatada para cada categoria de um ou mais campos de saída, é uma medida do conteúdo médio de informações (em bits) do modelo para prever registros pertencentes a essa categoria. Ela leva em consideração a dificuldade do problema de classificação, de modo que as predições precisas para categorias raras obtenham um índice de avaliação de desempenho mais alto do que as predições precisas para categorias comuns. Se o modelo não fizer nada além de supor uma categoria, o índice de avaliação de desempenho para essa categoria será 0.

Métricas de avaliação (AUC e Gini, somente classificadores binários). Para classificadores binários, esta opção relata as métricas de AUC (área abaixo da curva) e de avaliação de coeficiente de Gini. Essas duas métricas de avaliação são calculadas juntas para cada modelo binário. Os valores das métricas são relatados em uma tabela no navegador de saída da análise.

A métrica de avaliação AUC é calculada como a área abaixo de uma curva ROC (característica de operação do receptor) e é uma representação escalar do desempenho esperado de um classificador. A AUC está sempre entre 0 e 1, com um número maior representando um classificador melhor. Uma curva ROC diagonal entre as coordenadas (0,0) e (1,1) representa um classificador aleatório e possui uma AUC de 0,5. Portanto, um classificador realista não terá uma AUC menor que 0,5.

A métrica de avaliação de coeficiente de Gini às vezes é utilizada como uma métrica de avaliação alternativa para a AUC e as duas medidas estão estreitamente relacionadas. O coeficiente de Gini é calculado como o dobro da área entre a curva ROC e a diagonal, ou como $Gini = 2AUC - 1$. O coeficiente de Gini está sempre entre 0 e 1, com um número maior representando um classificador melhor. O coeficiente de Gini é negativo no evento improvável de que a curva ROC esteja abaixo da diagonal.

Figuras de confiança (se disponível). Para modelos que geram um campo de confiança, esta opção relata estatísticas dos valores de confiança e o relacionamento deles com as predições. Há duas configurações para esta opção:

- **Limite para.** Relata o nível de confiança acima do qual a precisão será a porcentagem especificada.
- **Melhorar precisão.** Relata o nível de confiança acima do qual a precisão será melhorada pelo fator especificado. Por exemplo, se a precisão geral for de 90% e esta opção estiver configurada como 2,0, o valor relatado será a confiança necessária para uma precisão de 95%.

Localizar campos preditos/preditores utilizando. Determina como campos preditos são correspondidos com o campo de destino original.

- **Metadados de campo de saída de modelo.** Corresponde campos preditos com o destino com base nas informações do campo de modelo, permitindo uma correspondência mesmo se um campo predito tiver sido renomeado. Informações de campo de modelo também podem ser acessadas por qualquer campo

predito a partir da caixa de diálogo Valores utilizando um nó Tipo. Consulte o tópico “Utilizando a Caixa de Diálogo de Valores” na página 138 para obter mais informações.

- **Formato do nome do campo.** Corresponde campos com base na convenção de nomenclatura. Por exemplo, valores preditos gerados por um nugget do modelo C5.0 para um destino denominado *response* devem estar em um campo denominado *\$C-response*.

Separar por partição. Se um campo de partição for utilizado para dividir registros em amostras de treinamento, teste e de validação, selecione esta opção para exibir os resultados separadamente para cada partição. Consulte o tópico “Nó Partição” na página 170 para obter mais informações.

Nota: ao separar por partição, os registros com valores nulos no campo de partição serão excluídos da análise. Isso nunca será um problema se um nó Partição for utilizado, uma vez que os nós Partição não geram valores nulos.

Análise definida pelo usuário. É possível especificar seu próprio cálculo de análise a ser utilizado na avaliação de seu(s) modelo(s). Utilize expressões do CLEM para especificar o que deve ser calculado para cada registro e como combinar os escores de nível de registro em um escore geral. Utilize funções **@TARGET** e **@PREDICTED** para referenciar o valor de destino (saída real) e o valor predito, respectivamente.

- **If.** Especifique uma expressão condicional se precisar utilizar cálculos diferentes, dependendo de alguma condição.
- **Then.** Especifique o cálculo se a condição If for true.
- **Else.** Especifique o cálculo se a condição If for false.
- **Usar.** Selecione uma estatística para calcular um escore geral dos escores individuais.

Dividir análise por campos. Mostra os campos categóricos disponíveis para dividir a análise. Além da análise geral, uma análise separada será relatada para cada categoria de cada campo de divisão.

Navegador de Saída da Análise

O navegador de saída da análise exibe os resultados da execução do nó Análise. As opções de salvamento, de exportação e de impressão usuais estão disponíveis no menu Arquivo. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.

Ao procurar a saída da Análise pela primeira vez, os resultados são expandidos. Para ocultar os resultados após visualizá-los, utilize o controle de expansão à esquerda do item para reduzir os resultados específicos que desejar ocultar ou clique no botão **Reduzir Tudo** para reduzir todos os resultados. Para ver os resultados novamente após reduzi-los, utilize o controle de expansão à esquerda do item para mostrar os resultados ou clique no botão **Expandir Tudo** para mostrar todos os resultados.

Resultados para o campo de saída. A saída da Análise contém uma seção para cada campo de saída para o qual houver um campo de predição correspondente criado por um modelo gerado.

Comparando. Na seção de campo de saída há uma subseção para cada campo de predição associado a esse campo de saída. Para campos de saída categóricos, o nível superior desta seção contém uma tabela que mostra o número e a porcentagem das predições corretas e incorretas e o número total de registros no fluxo. Para campos de saída numéricos, esta seção mostra as seguintes informações:

- **Erro Mínimo.** Mostra o erro mínimo (a diferença entre os valores observados e preditos).
- **Erro Máximo.** Mostra o erro máximo.
- **Erro Médio.** Mostra a média de erros em todos os registros. Isso indica se há um **viés** sistemático (uma tendência mais forte de superestimar do que subestimar, ou vice-versa) no modelo.
- **Erro Médio Absoluto.** Mostra a média dos valores absolutos dos erros em todos os registros. Indica a magnitude média do erro, independente da direção.
- **Desvio Padrão.** Mostra o desvio padrão dos erros.

- **Correlação Linear.** Mostra a correlação linear entre os valores preditos e reais. Essa estatística varia entre -1,0 e 1,0. Valores próximos de +1,0 indicam uma forte associação positiva, de modo que os valores preditos altos são associados aos valores reais altos e valores preditos baixos são associados a valores reais baixos. Valores próximos de -1,0 indicam uma forte associação negativa, de modo que valores preditos altos são associados a valores reais baixos, e vice-versa. Valores próximos de 0,0 indicam uma associação fraca, de modo que os valores preditos são mais ou menos independentes dos valores reais. *Nota:* uma entrada em branco aqui indica que a correlação linear não pode ser calculada nesse caso porque os valores reais ou preditos são constantes.
- **Ocorrências.** Mostra o número de registros usados na análise.

Matriz de Coincidência. Para campos de saída categóricos, se você solicitou uma matriz de coincidência nas opções de análise, uma subseção aparecerá aqui contendo a matriz. As linhas representam os valores reais observados, e as colunas representam valores preditos. A célula na tabela indica o número de registros para cada combinação de valores preditos e reais.

Avaliação de Desempenho. Para os campos de saída categóricos, se você solicitou as estatísticas de avaliação de desempenho nas opções de análise, os resultados da avaliação de desempenho aparecerão aqui. Cada categoria de saída é listada com sua estatística de avaliação de desempenho.

Relatório de Valores de Confiança. Para os campos de saída categóricos, se você solicitou valores de confiança nas opções de análise, os valores aparecerão aqui. As seguintes estatísticas são relatadas para valores de confiança do modelo:

- **Range.** Mostra o intervalo (valores menor e maior) de valores de confiança para registros nos dados em fluxo.
- **Média de Corretos.** Mostra a confiança média para registros que são classificados corretamente.
- **Média de Incorretos.** Mostra a confiança média para registros que são classificados incorretamente.
- **Sempre Corretas Acima.** Mostra o limite de confiança acima do qual as predições são sempre corretas e mostra a porcentagem de casos que atendem a esse critério.
- **Sempre Incorretas Abaixo.** Mostra o limite de confiança abaixo do qual as predições são sempre incorretas e mostra a porcentagem de casos que atendem a esse critério.
- **X% de Precisão Acima.** Mostra o nível de confiança no qual precisão é X%. O X é aproximadamente o valor especificado para **Limite para** nas opções de Análise. Para alguns modelos e conjuntos de dados, não é possível escolher um valor de confiança que forneça o limite exato especificado nas opções (geralmente devido a clusters de casos semelhantes com o mesmo valor de confiança perto do limite). O limite relatado é o valor mais próximo do critério de precisão especificado que pode ser obtido com um limite de valor de confiança único.
- **X de Dobra Correta Acima.** Mostra o valor de confiança no qual a precisão é X vezes melhor do que é para o conjunto de dados geral. O X é o valor especificado para **Melhorar precisão** nas opções de Análise.

Concordância entre. Se dois ou mais modelos gerados que preveem o mesmo campo de saída forem incluídos no fluxo, será possível ver as estatísticas sobre a **concordância** entre as predições geradas pelos modelos. Isto inclui o número e a porcentagem de registros para os quais as predições concordam (para campos de saída categóricos) ou estatísticas de sumarização do erro (para campos de saída contínuos). Para campos categóricos, ele inclui uma análise das predições em comparação com os valores reais para o subconjunto de registros no qual os modelos concordam (gera o mesmo valor predito).

Métricas de Avaliação. Para classificadores binários, se você solicitou as métricas de avaliação nas opções de análise, os valores das métricas de avaliação de coeficiente AUC e Gini serão mostrados em uma tabela nesta seção. A tabela possui uma linha para cada modelo de classificador binário. A tabela de métricas de avaliação é mostrada para cada campo de saída, ao invés de para cada modelo.

Nó Auditoria de Dados

O nó Auditoria de Dados fornece uma visão inicial abrangente dos dados que você colocar no IBM SPSS Modeler, apresentada em uma matriz de fácil leitura que pode ser ordenada e utilizada para gerar gráficos de tamanho integral e uma variedade de nós de preparação de dados.

- A guia Auditoria exibe um relatório que fornece estatísticas de sumarização, histogramas e gráficos de distribuição que podem ser úteis ao obter um entendimento preliminar dos dados. O relatório também exibe o ícone de armazenamento antes do nome do campo.
- A guia Qualidade no relatório de auditoria exibe informações sobre valores discrepantes, valores extremos e valores omissos e oferece ferramentas para manipular esses valores.

Utilizando o Nó Auditoria de Dados

O nó Auditoria de Dados pode ser conectado diretamente a um nó de origem ou abaixo de um nó Tipo instanciado. Também é possível gerar um número de nós de preparação de dados com base nos resultados. Por exemplo, é possível gerar um nó de Filtro que exclua campos com valores omissos em excesso para que seja útil na modelagem e gerar um SuperNode que impute valores omissos para qualquer ou para todos os campos restantes. É daí que vem o verdadeiro poder da auditoria, permitindo não só avaliar o estado atual de seus dados, mas também executar uma ação com base na avaliação.

Verificação ou amostragem de dados. Como uma auditoria inicial é particularmente eficiente quando lidar com “big data”, um nó Amostra pode ser utilizado para reduzir o tempo de processamento durante a exploração inicial ao selecionar apenas um subconjunto de registros. O nó Auditoria de Dados também pode ser utilizado em combinação com nós como Seleção de Variável e Detecção de Anomalias nos estágios de exploração de análise.

Guia Configurações de Nó Auditoria de Dados

A guia Configurações permite especificar os parâmetros básicos para a auditoria.

Padrão. É possível simplesmente conectar o nó ao seu fluxo e clicar em **Executar** para gerar um relatório de auditoria para todos os campos com base nas configurações padrão, conforme a seguir:

- Se não houver nenhuma configuração do nó Tipo, todos os campos serão incluídos no relatório.
- Se houver configurações de Tipo (independentemente se elas estiverem instanciadas ou não), todos os campos de *Entrada*, de *Destino* e *Ambos* serão incluídos na exibição. Se houver um campo de *Destino* único, utilize-o como o campo de Sobreposição. Se houver mais de um campo de *Destino* especificado, nenhuma sobreposição padrão foi especificada.

Usar campos customizados. Selecione esta opção para selecionar campos manualmente. Utilize o botão seletor de campo à direita para selecionar campos individualmente ou por tipo.

Campo de sobreposição. O campo de sobreposição é utilizado para desenhar gráficos miniaturas mostrados no relatório de auditoria. No caso de um campo contínuo (intervalo numérico), as estatísticas bivariadas (covariância e correlação) também são calculadas. Se um campo de *Destino* único estiver presente com base nas configurações do nó Tipo, ele será utilizado como o campo de sobreposição padrão conforme descrito acima. Como alternativa, é possível selecionar **Utilizar campos customizados** para especificar uma sobreposição.

Exibição. Permite especificar se os gráficos são disponíveis na saída e escolher as estatísticas exibidas por padrão.

- **Gráficos.** Exibe um gráfico para cada campo selecionado, seja um gráfico de distribuição (barras), um histograma ou um gráfico de dispersão, conforme apropriado para os dados. Os gráficos são exibidos como miniaturas no relatório inicial, mas gráficos e nós de gráfico em tamanho integral também podem ser gerados. Consulte o tópico “Navegador de Saída de Auditoria de Dados” na página 315 para obter mais informações.

- **Estatísticas Básicas/Avançadas.** Especifica o nível de estatísticas exibido na saída por padrão. Embora essa configuração determine a exibição inicial, todas as estatísticas estão disponíveis na saída, independentemente dessa configuração. Consulte o tópico “Exibir Estatísticas” na página 316 para obter mais informações.

Mediana e modo. Calcula a mediana e o modo para todos os campos no relatório. Observe que com grandes conjuntos de dados, essas estatísticas podem aumentar o tempo de processamento, já que elas demoram mais tempo do que outras para calcular. No caso apenas da mediana, o valor relatado pode basear-se em uma amostra de 2000 registros (ao invés de no conjunto de dados inteiro) em alguns casos. Esta amostragem é feita em uma base por campo nos casos em que os limites de memória seriam de outra forma excedidos. Quando a amostragem estiver em vigor, os resultados serão rotulados como tal na saída (*Mediana de Amostra* ao invés de apenas *Mediana*). Todas as estatísticas diferentes de mediana são sempre calculadas utilizando o conjunto de dados integral.

Campos vazios ou sem tipo. Quando utilizados com dados instanciados, campos sem tipos não são incluídos no relatório de auditoria. Para incluir campos sem tipo (incluindo campos vazios), selecione **Limpar Todos os Valores** em qualquer nó Tipo de envio de dados. Isso assegura que os dados não sejam instanciados, fazendo com que todos os campos sejam incluídos no relatório. Por exemplo, isso pode ser útil se desejar obter uma lista completa de todos os campos ou gerar um nó Filtro que excluirá aqueles que estiverem vazios. Consulte o tópico “Filtrando Campos com Dados Omissos” na página 319 para obter mais informações.

Guia de Qualidade de Auditoria de Dados

A guia Qualidade no nó Auditoria de Dados fornece opções para a manipulação de valores omissos, valores discrepantes e valores extremos.

Valores ausentes

- **Contagem de registros com valores válidos.** Selecione esta opção para mostrar o número de registros com valores válidos para cada campo avaliado. Observe que valores nulos (indefinidos), valores em branco, espaços em branco e sequências de caracteres vazias são sempre tratados como valores inválidos.
- **Decompor contagens de registros com valores inválidos.** Selecione esta opção para mostrar o número de registros com cada tipo de valor inválido para cada campo.

Valores Discrepantes e Valores Extremos

Método de detecção de valores discrepantes e valores extremos. Dois métodos são suportados:

Desvio padrão da média. Detecta valores discrepantes e extremos com base no número de desvios padrão da média. Por exemplo, se você tiver um campo com uma média de 100 e um desvio padrão de 10, será possível especificar 3,0 para indicar que qualquer valor abaixo de 70 ou acima de 130 deve ser tratado como um valor discrepante.

Amplitude interquartil. Detecta valores discrepantes e extremos com base na amplitude interquartil, que é o intervalo dentro do qual os dois quartis centrais caem (entre o 25^o e o 75^o percentis). Por exemplo, com base na configuração padrão de 1,5, o limite inferior deve ser os valores discrepantes $Q1 - 1,5 * IQR$ e o limite superior deve ser $Q3 + 1,5 * IQR$. Observe que a utilização desta opção pode diminuir o desempenho em grandes conjuntos de dados.

Navegador de Saída de Auditoria de Dados

O navegador Auditoria de Dados é uma ferramenta poderosa para obter uma visão geral de seus dados. A guia de Auditoria exibe gráficos em miniatura, ícones de armazenamento e estatísticas para todos os campos, ao passo que a guia Qualidade exibe informações sobre valores discrepantes, valores extremos e valores omissos. Com base nas estatísticas iniciais de gráficos e de sumarização, você pode decidir

recodificar um campo numérico, derivar um novo campo ou reclassificar os valores de um campo nominal. Ou você pode querer explorar ainda mais utilizando uma visualização mais sofisticada. Isso pode ser feito a partir do navegador de relatório de auditoria utilizando o menu Gerar para criar qualquer número de nós que podem ser utilizados para transformar ou visualizar os dados.

- Ordene as colunas ao clicar no cabeçalho da coluna ou reordene as colunas utilizando arrastar e soltar. A maioria das operações de saída padrão também é suportada. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.
- Visualize valores e intervalos para os campos ao dar um clique duplo em um campo nas colunas Medição ou Exclusivo.
- Utilize a barra de ferramentas ou o menu Editar para mostrar ou ocultar rótulos de valor ou para escolher as estatísticas que deseja exibir. Consulte o tópico “Exibir Estatísticas” para obter mais informações.
- Verifique os ícones de armazenamento à esquerda dos nomes de campo. Armazenamento descreve a maneira como os dados são armazenados em um campo. Por exemplo, um campo com valores de 1 e 0 armazena dados de número inteiro. Isso é diferente do nível de medição, que descreve o uso dos dados e não afeta o armazenamento. Consulte o tópico “Configurando Armazenamento e Formatação do Campo” na página 9 para obter mais informações.

Visualizando e gerando gráficos

Se nenhuma sobreposição estiver selecionada, a guia Auditoria exibirá gráficos de barra (para campos nominais ou de flag) ou histogramas (campos contínuos).

Para uma sobreposição de campo nominal ou de flag, os gráficos são coloridos pelos valores da sobreposição.

Para uma sobreposição de campo contínuo, gráficos de dispersão bidimensionais são gerados ao invés de barras e histogramas unidimensionais. Nesse caso, o eixo x é mapeado para o campo de sobreposição, permitindo ver a mesma escala em todos os eixos x conforme você lê para baixo na tabela.

- Para campos de Flag ou Nominais, coloque o cursor do mouse sobre uma barra para exibir o valor ou o rótulo subjacente em uma Dica de Ferramenta.
- Para campos de Flag ou Nominais, utilize a barra de ferramentas para alternar a orientação dos gráficos miniaturas de horizontal para vertical.
- Para gerar um gráfico em tamanho integral a partir de qualquer miniatura, clique duas vezes na miniatura ou selecione uma miniatura e escolha **Saída do Gráfico** no menu Gerar. *Nota:* se um gráfico miniatura foi baseado em dados de amostra, o gráfico gerado conterá todos os casos se o fluxo de dados original ainda estiver aberto.

Um gráfico poderá ser gerado somente se o nó Auditoria de Dados que criou a saída estiver conectado ao fluxo.

- Para gerar um nó de gráfico correspondente, selecione um ou mais campos na guia de Auditoria e escolha **Nó de Gráfico** no menu Gerar. O nó resultante é incluído na tela de fluxo e poderá ser utilizado para recriar o gráfico sempre que o fluxo for executado.
- Se um conjunto de sobreposição possuir mais de 100 valores, um aviso será emitido e a sobreposição não será incluída.

Exibir Estatísticas

A caixa de diálogo Exibir Estatísticas permite escolher as estatísticas exibidas na guia Auditoria. As configurações iniciais são especificadas no nó Auditoria de Dados. Consulte o tópico “Guia Configurações de Nó Auditoria de Dados” na página 314 para obter mais informações.

Mínimo. O menor valor de uma variável numérica.

Máximo. O maior valor de uma variável numérica.

Sum. A soma ou o total dos valores, em todos os casos com valores não omissos.

Intervalo. A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.

Média. Uma medida de tendência central. A média aritmética, a soma dividida pelo número de casos.

Erro padrão de média. Uma medida do quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

desvio padrão . Uma medida de dispersão em torno da média, igual à raiz quadrada da variância. O desvio padrão é medido nas mesmas unidades que a variável original.

Variância. Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.

Assimetria. Uma medida da assimetria de uma distribuição. A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

Erro Padrão de Assimetria. A razão de assimetria com seu erro padrão pode ser utilizada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for inferior a -2 ou superior a +2). Um valor positivo grande para assimetria indica um rodapé direito longo, e um valor negativo extremo indica um rodapé esquerdo longo.

Curtose. Uma medida da extensão até a qual as observações se agrupam em torno de um ponto central. Para a uma distribuição normal, o valor da estatística de curtose é zero. A curtose positiva indica que, com relação a uma distribuição normal, as observações são agrupadas mais ao centro da distribuição e têm rastros mais finos até os valores extremos da distribuição, no ponto em que os rastros da distribuição leptocúrtica são mais espessos com relação a uma distribuição normal. Já a curtose negativa indica que, com relação a uma distribuição normal, as observações são menos agrupadas e possuem rodapés mais espessos até os valores extremos da distribuição, no ponto em que os rastros da distribuição platykúrtica são mais finos com relação a uma distribuição normal.

Erro Padrão de Curtose. A razão de curtose com seu erro padrão pode ser utilizada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for inferior a -2 ou superior a +2). Um valor positivo grande para curtose indica que os rodapés da distribuição são maiores do que aqueles de uma distribuição normal, e um valor negativo para curtose indica rodapés mais curtos (como aqueles de uma distribuição uniforme em forma de caixa).

unique. Avalia todos os efeitos simultaneamente, ajuste cada efeito a todos os outros efeitos de qualquer tipo.

Válidos. Casos válidos que não possuem o valor omissão do sistema, nem um valor definido como omissão de usuário. Observe que valores nulos (indefinidos), valores em branco, espaços em branco e sequências de caracteres vazias são sempre tratados como valores inválidos.

Median. O valor acima e abaixo do qual metade dos casos cai, o 50º percentil. Se houver um número par de casos, a mediana será a média dos dois casos intermediários quando forem classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).

Modo. O valor que ocorre mais frequentemente. Se vários valores compartilharem a maior frequência da ocorrência, cada um deles será um modo.

Observe que a mediana e o modo são suprimidos por padrão para melhorar o desempenho, mas podem ser selecionados na guia Configurações no nó Auditoria de Dados. Consulte o tópico “Guia Configurações de Nó Auditoria de Dados” na página 314 para obter mais informações.

Estatísticas para Sobreposições

Se um campo de sobreposição contínuo (intervalo numérico) estiver em uso, as estatísticas a seguir também estarão disponíveis:

Covariância. Uma medida de associação não padronizada entre duas variáveis, igual ao desvio de produto vetorial dividido por N-1.

Guia Qualidade do navegador de Auditoria de Dados

A guia Qualidade no navegador de Auditoria de Dados exibe os resultados da análise de qualidade de dados e permite especificar o tratamento para valores discrepantes, extremos e valores omissos.

Imputando valores omissos: O relatório de auditoria lista a porcentagem de registros completos para cada campo, junto do número de valores válidos, nulos e em branco. É possível escolher imputar os valores omissos para campos específicos conforme apropriado e, em seguida, gerar um SuperNode para aplicar essas transformações.

1. Na coluna **Imputar Omissos**, especifique o tipo de valores que deseja imputar, se houver. É possível escolher imputar em brancos, nulos, ambos ou especificar uma condição ou expressão customizada que seleciona os valores para imputar.

Há vários tipos de valores omissos reconhecidos pelo IBM SPSS Modeler:

- **Valores omissos do sistema ou nulos.** Esses são valores de não sequência de caracteres que foram deixados em branco no arquivo de origem ou banco de dados e não foram definidos especificamente como "omissos" em um nó Tipo ou de origem. Valores omissos do sistema são exibidos como **\$null\$**. Observe que sequências de caracteres vazias não são consideradas nulas no IBM SPSS Modeler, embora elas possam ser tratadas como nulas por certos bancos de dados.
- **Sequências de caracteres vazias e espaço em branco.** Valores de sequência de caracteres vazia e espaços em branco (sequências de caracteres sem caracteres visíveis) são tratados como sendo diferentes de valores nulos. Sequências de caracteres vazias são tratadas como sendo equivalentes a espaço em branco para a maioria dos propósitos. Por exemplo, se você selecionar a opção para tratar um espaço em branco como brancos em um nó Tipo ou de origem, essa configuração se aplicará também às sequências de caracteres vazias.
- **Valores omissos definidos pelo usuário ou em branco.** São valores como desconhecido, 99 ou -1 que são definidos explicitamente em um nó de origem ou nó de Tipo como omissos. Opcionalmente, também é possível escolher tratar nulos e espaços em branco como brancos, o que permite que eles sejam sinalizados para tratamento especial e sejam excluídos da maioria dos cálculos. Por exemplo, é possível usar a função @BLANK para tratar esses valores, com outros tipos de valores omissos, como brancos.

2. Na coluna **Método**, especifique o método que deseja utilizar.

Os métodos a seguir estão disponíveis para imputar valores omissos:

Fixo. Substitui um valor fixo (a média do campo, o ponto médio do intervalo ou a constante especificada).

Aleatório. Substitui um valor aleatório com base em uma distribuição uniforme ou normal.

Expressão. Permite especificar uma expressão customizada. Por exemplo, você poderia substituir valores por uma variável global criada pelo nó Configurar Globais.

Algoritmo. Substitui um valor predito por um modelo baseado no algoritmo C&RT. Para cada campo imputado usando esse método, haverá um modelo C&RT separado, com um nó de Preenchimento que substitui brancos e nulos pelo valor predito pelo modelo. Um nó de Filtro é usado para remover os campos de predição gerados pelo modelo.

3. Para gerar um SuperNode de Valores Omissos, nos menus escolha:

Gerar > SuperNode de Valores Omissos

A caixa de diálogo SuperNode de Valores Omissos é exibida.

4. Selecione **Todos os campos** ou **Somente campos selecionados** e especifique um tamanho de amostra, se desejar. (A amostra especificada é uma porcentagem; por padrão, 10% de todos os registros são amostrados).
5. Clique em **OK** para incluir o SuperNode gerado na tela de fluxo.
6. Anexe o SuperNode ao fluxo para aplicar as transformações.

No SuperNode, uma combinação de nugget do modelo, Preenchimento e Filtro é usada conforme apropriado. Para entender como ele funciona, é possível editar o SuperNode e clicar em **Aumentar Zoom**, e também é possível incluir, editar ou remover nós específicos dentro do SuperNode para realizar um ajuste preciso do comportamento.

Manipulando Valores Discrepantes e Valores Extremos: O relatório de auditoria lista o número de valores discrepantes e os valores extremos são listados para cada campo com base nas opções de detecção especificadas no nó de Auditoria de Dados. Consulte o tópico “Guia de Qualidade de Auditoria de Dados” na página 315 para obter mais informações. É possível escolher forçar, descartar ou anular esses valores para campos específicos conforme apropriado e, em seguida, gerar um SuperNode para aplicar as transformações.

1. Na coluna **Ação**, especifique a manipulação de valores discrepantes e extremos para campos específicos, conforme desejado.

As seguintes ações estão disponíveis para a manipulação de valores discrepantes e extremos:

- **Forçar.** Substitui valores discrepantes e valores extremos pelo valor mais próximo que não seria considerado extremo. Por exemplo, se um valor discrepante for definido para ser algo acima ou abaixo de três desvios padrão, então todos os valores discrepantes serão substituído pelo valor mais alto ou mais baixo dentro desse intervalo.
- **Descartar.** Descarta os registros com valores discrepantes ou extremos para o campo especificado.
- **Anular.** Substitui valores discrepantes e extremos pelo valor nulo ou omissos do sistema.
- **Forçar valores discrepantes / descartar extremos.** Descarta somente valores extremos.
- **Forçar valores discrepantes / anular extremos.** Anula somente valores extremos.

2. Para gerar o SuperNode, a partir dos menus escolha:

Gerar > SuperNode de Valor Discrepante e Extremo

A caixa de diálogo SuperNode de Valor Discrepante é exibida.

3. Selecione **Todos os campos** ou **Somente campos selecionados** e, em seguida, clique em **OK** para incluir o SuperNode gerado na tela de fluxo.
4. Anexe o SuperNode ao fluxo para aplicar as transformações.

Opcionalmente, é possível editar os SuperNodes e aumentar o zoom para navegar ou fazer mudanças. No SuperNode, os valores são descartados, forçados ou anulados utilizando uma série de nós Seleção e/ou Preenchimento conforme apropriado.

Filtrando Campos com Dados Omissos: No navegador de Auditoria de Dados, é possível criar um nó Filtro novo com base nos resultados da análise de Qualidade utilizando a caixa de diálogo Gerar Filtro a partir da Qualidade.

Modo. Selecione a operação **Incluir** ou **Excluir** desejada para campos especificados.

- **Campos selecionados.** O nó Filtro será incluirá/excluirá os campos selecionados na guia Qualidade. Por exemplo, é possível ordenar a tabela na coluna % **Completo**, utilizar Shift-clique para selecionar os campos menos completos e, em seguida, gerar um nó Filtro que exclui esses campos.
- **Campos com porcentagem de qualidade maior que.** O nó Filtro incluirá/excluirá campos em que a porcentagem de registros completos é maior que o limite especificado. O limite padrão é 50%.

Filtrando Campos Vazios ou Sem Tipo

Observe que, após os valores de dados terem sido instanciados, campos sem tipo ou vazios serão excluídos dos resultados de auditoria e da maior parte da outra saída no IBM SPSS Modeler. Esses campos são ignorados para propósitos de modelagem, que podem inchar ou bagunçar os dados. Se isso ocorrer, será possível usar o navegador de Auditoria de Dados a partir do qual gerar um nó Filtro que remove esses campos do fluxo.

1. Para assegurar que todos os campos sejam incluídos na auditoria, incluindo campos vazios ou sem tipos, clique em **Limpar Todos os Valores** na origem ou nó Tipo de envio de dados ou configure valores para <Pass> para todos os campos.
2. No navegador Auditoria de Dados, ordene a coluna % **Completo**, selecione os campos que tiverem zero valores válidos (ou algum outro limite) e utilize o menu Gerar para produzir um nó Filtro que pode ser incluído no fluxo.

Selecionando Registros com Dados Omissos: No navegador Auditoria de Dados, é possível criar um novo nó Seleção com base nos resultados da análise de qualidade.

1. No navegador de Auditoria de Dados, escolha a guia Qualidade.
2. No menu, escolha:

Gerar > Nó Seleção de Valores Omissos

A caixa de diálogo Gerar Nó Seleção é exibida.

Selecionar quando o registro for. Especifique se os registros devem ser mantidos quando forem **Válidos** ou **Inválidos**.

Procurar valores inválidos em. Especifique onde verificar valores inválidos.

- **Todos os campos.** O nó Seleção verificará todos os campos em busca de valores inválidos.
- **Campos selecionados na tabela.** O nó Seleção verificará apenas os campos selecionados atualmente na tabela de saída Qualidade.
- **Campos com porcentagem de qualidade maior que.** O nó Seleção verificará os campos em que a porcentagem de registros completos é maior que o limite especificado. O limite padrão é 50%.

Considerar um registro inválido se um valor inválido for encontrado em. Especifique a condição para identificar um registro como inválido.

- **Qualquer um dos campos acima.** O nó Seleção considerará um registro inválido se *qualquer* um dos campos especificados acima contiver um valor inválido para esse registro.
- **Todos os campos acima.** O nó Seleção irá considerar um registro inválido apenas se *todos* os campos especificados acima contiverem valores inválidos para esse registro.

Gerando Outros Nós para Preparação de Dados

Uma variedade de nós utilizados na preparação de dados pode ser gerada diretamente a partir do navegador de Auditoria de Dados, incluindo os nós Reclassificação, Categorização e Derivar. Por exemplo:

- É possível derivar um novo campo com base nos valores de *claimvalue* e *farmincome* ao selecionar ambos no relatório de auditoria e escolher **Derivar** a partir do menu Gerar. O novo nó é incluído na tela de fluxo.

- Da mesma forma, é possível determinar, com base nos resultados da auditoria, que a recodificação *farmincome* em categorias baseadas em percentil fornece uma análise mais focada. Para gerar um nó Categorização, selecione a linha do campo na exibição e escolha **Categorização** a partir do menu Gerar.

Quando um nó é gerado e incluído na tela de fluxo, deve-se anexá-lo ao fluxo e abrir o nó para especificar opções para um ou mais arquivos selecionados.

Nó Transformação

A normalização dos campos de entrada é um passo importante antes de utilizar técnicas de escoragem tradicionais, como regressão, regressão logística e análise discriminante. Essas técnicas realizam suposições sobre as distribuições normais de dados que podem não ser verdadeiras para muitos arquivos de dados brutos. Uma abordagem para lidar com dados do mundo real é aplicar transformações que movem um elemento de dados brutos em direção a uma distribuição mais normal. Além disso, os campos normalizados podem facilmente ser comparados entre si, por exemplo, rendimento e idade estão em escalas totalmente diferentes em um arquivo de dados brutos, mas quando normalizadas, o impacto relativo de cada um pode ser facilmente interpretado.

O Nó Transformação fornece um visualizador de saída que permite executar uma avaliação visual rápida da melhor transformação para utilizar. É possível ver rapidamente se as variáveis são normalmente distribuídas e, se necessário, escolher a transformação desejada e aplicá-la. É possível selecionar diversos campos e executar uma transformação por campo.

Após selecionar as transformações preferenciais para os campos, será possível gerar nós Derivar ou Preenchimento que desempenham as transformações e anexar esses nós para o fluxo. O nó Derivar cria novos campos, ao passo que o nó Preenchimento transforma os existentes. Consulte o tópico “Gerando Gráficos” na página 323 para obter mais informações.

Guia Campos do Nó Transformação

Na guia Campos, é possível especificar quais campos dos dados deseja utilizar para visualizar possíveis transformações e aplicá-las. Apenas campos numéricos podem ser transformados. Clique no botão seletor de campo e selecione um ou mais campos numéricos a partir da lista exibida.

Guia Opções do nó Transformação

A guia Opções permite especificar o tipo de transformações que você deseja incluir. É possível escolher incluir todas as transformações disponíveis ou selecionar transformações individualmente.

No último caso, também é possível inserir um número para compensar os dados para as transformações inversas e logarítmicas. Fazer isso é útil em situações em que uma grande proporção de zeros nos dados causa um viés na média e nos resultados do desvio padrão.

Por exemplo, suponha que você tenha um campo denominado *BALANCE* que possui alguns valores zero e deseja utilizar a transformação inversa nele. Para evitar um viés indesejável, selecione **Inverso (1/x)** e insira 1 no campo **Utilizar um offset de dados**. (Observe que esse offset não está relacionado àquele executado pela função de sequência @OFFSET no IBM SPSS Modeler).

Todas as fórmulas. Indica que todas as transformações disponíveis devem ser calculadas e mostradas na saída.

Selecione as fórmulas. Permite selecionar as diferentes transformações a serem calculadas e mostradas na saída.

- **Inverso (1/x).** Indica que a transformação inversa deve ser exibida na saída.
- **Log (log n).** Indica que a transformação \log_n deve ser exibida na saída.
- **Log (log 10).** Indica que a transformação \log_{10} deve ser exibida na saída.

- **Exponencial.** Indica que a transformação exponencial (e^x) deve ser exibida na saída.
- **Raiz Quadrada.** Indica que a transformação raiz quadrada deve ser exibida na saída.

Guia Saída do Nó Transformação

A guia Saída permite especificar o formato e a localização da saída. É possível escolher exibir os resultados na tela ou enviá-los para um dos tipos de arquivo padrão. Consulte o tópico “Guia Saída do Nó Saída” na página 306 para obter mais informações.

Visualizador de Saída do Nó Transformação

O visualizador de saída permite ver os resultados da execução do Nó Transformação. O visualizador é uma ferramenta poderosa que exibe diversas transformações por campo em visualizações em miniatura da transformação, permitindo comparar os campos rapidamente. É possível utilizar as opções em seu menu Arquivo para salvar, exportar ou imprimir a saída. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.

Para cada transformação (diferente de Transformação Seleccionada), uma legenda é exibida abaixo no formato:

Mean (Standard deviation)

Gerando Nós para a Transformação

O visualizador de saída fornece um ponto de início útil para sua preparação de dados. Por exemplo, talvez você queira normalizar o campo *AGE* para poder utilizar uma técnica de escoragem (como regressão logística ou análise discriminante) que assume uma distribuição normal. Com base nas estatísticas iniciais de gráficos e de sumarização, você pode decidir transformar o campo *AGE* de acordo com uma distribuição específica (por exemplo, log). Após selecionar a distribuição preferida, é possível, em seguida, gerar um nó derivar com uma transformação padronizada a ser usada para escoragem.

É possível gerar os seguintes nós de operações de campo a partir do visualizador de saída:

- Derivar
- Preenchedor

Um nó Derivar cria novos campos com as transformações desejadas, ao passo que o nó Preenchimento transforma os campos existentes. Os nós são colocados na tela na forma de um SuperNode.

Se selecionar a mesma transformação para campos diferentes, um nó Derivar ou Preenchimento conterá as fórmulas para esse tipo de transformação para todos os campos aos quais a transformação se aplica. Por exemplo, suponha que você tenha selecionado os campos e as transformações mostrados na tabela a seguir para gerar um nó Derivar.

Tabela 55. Exemplo de geração do nó Derivar.

Campo	Transformação.
<i>AGE</i>	Distribuição Atual
<i>INCOME</i>	Log
<i>OPEN_BAL</i>	Inverso
<i>BALANCE</i>	Inverso

Os nós a seguir estão contidos no SuperNode:

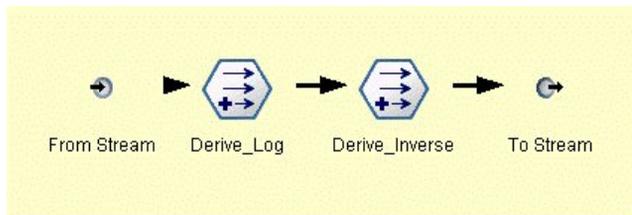


Figura 66. SuperNode na tela

Neste exemplo, o nó *Derive_Log* possui a fórmula de log para o campo *INCOME* e o nó *Derive_Inverse* possui as fórmulas inversas para os campos *OPEN_BAL* e *BALANCE*.

Para Gerar um Nó

1. Para cada campo no visualizador de saída, selecione a transformação desejada.
2. No menu Gerar, escolha **Nó Derivar** ou **Nó Preenchimento**, conforme desejado.

Fazer isso exibe a caixa de diálogo Gerar Nó Derivar ou Gerar Nó Preenchimento, conforme apropriado.

Escolha **Transformação não padronizada** ou **Transformação padronizada (score z)**, conforme desejado. A segunda opção aplica um score z na transformação; os scores z representam valores como uma função de distância da média da variável em desvios padrão. Por exemplo, se aplicar a transformação logarítmica no campo *AGE* e escolher uma transformação padronizada, a equação final para o nó gerado será:

$$(\log(\text{AGE}) - \text{Mean}) / \text{SD}$$

Quando um nó for gerado e aparecer na tela de fluxo:

1. Anexe-o ao fluxo.
2. Para um SuperNode, opcionalmente, dê um clique duplo no nó para visualizar seu conteúdo.
3. Opcionalmente, dê um clique duplo em um nó Derivar ou Preenchimento para modificar as opções para o campo ou campos selecionados.

Gerando Gráficos: É possível gerar a saída do histograma em tamanho integral a partir de um histograma miniatura no visualizador de saída.

Para Gerar um Gráfico

1. Dê um clique duplo em um gráfico miniatura no visualizador de saída.

ou

Selecione um gráfico miniatura no visualizador de saída.

2. No menu Gerar, escolha **Saída do gráfico**.

Isso exibe o histograma com uma curva de distribuição normal sobreposta. Isso permite comparar quão próximo cada transformação disponível corresponde a uma distribuição normal.

Nota: será possível gerar um gráfico somente se o nó Transformação que criou a saída estiver conectado ao fluxo.

Outras Operações: No visualizador de saída, também é possível:

- Ordenar a grade da saída pela coluna Campo.
- Exportar a saída para um arquivo HTML. Consulte o tópico “Exportando Saída” na página 304 para obter mais informações.

Nó Estatísticas

O nó Estatísticas fornece informações de sumarização sobre campos numéricos. É possível obter as estatísticas de sumarização para campos e correlações individuais entre os campos.

Guia Configurações do Nó Estatísticas

Examinar. Selecione um ou mais campos para os quais deseja estatísticas de sumarização individuais. Você pode selecionar vários campos.

Estatísticas. Selecione as estatísticas para o relatório. As opções disponíveis incluem **Contagem, Média, Soma, Mín., Máx., Intervalo, Variância, Desv. Pad., Erro de Média Pad., Mediana e Modo.**

Correlacionar. Selecione um ou mais campos que deseja correlacionar. Você pode selecionar vários campos. Quando campos de correlação são selecionados, a correlação entre cada campo Examinar e um ou mais campos de correlação é listada na saída.

Configurações de Correlação. É possível especificar opções para exibir a intensidade das correlações na saída.

Configurações de correlação

O IBM SPSS Modeler pode caracterizar correlações com rótulos descritivos para ajudar a destacar os relacionamentos importantes. A **correlação** mede a intensidade do relacionamento entre os dois campos contínuos (intervalo numérico). Ela aceita valores entre -1,0 e 1,0. Valores próximos de +1,0 indicam uma associação positiva forte para que valores altos em um campo sejam associados a valores altos no outro e valores baixos sejam associados com valores baixos. Valores próximos de -1,0 indicam uma associação negativa forte para que valores altos de um campo sejam associados a valores baixos no outro, e vice-versa. E valores próximos de 0,0 indicam uma associação fraca, de modo que os valores para os dois campos sejam mais ou menos independentes.

Utilizando a caixa de diálogo Configurações de Correlação, é possível controlar a exibição dos rótulos de correlação, alterar os limites que definem as categorias e alterar os rótulos utilizados para cada intervalo. Como a maneira de caracterizar os valores de correlação depende enormemente do domínio do problema, você poderá querer customizar os intervalos e os rótulos para ajustar à sua situação específica.

Mostrar rótulos de intensidade de correlação na saída. Essa opção é selecionada por padrão. Cancele a seleção dessa opção para omitir os rótulos descritivos da saída.

Intensidade de Correlação. Há duas opções para definir e rotular a intensidade das correlações:

- **Definir a intensidade de correlação por importância (1-p).** Rotula correlações com base na importância, definida como 1 menos a significância ou 1 menos a probabilidade de que a diferença nas médias possa ser explicada por uma única tentativa. Quanto mais próximo esse valor estiver de 1, maior a chance de que os dois campos sejam *não* independentes, ou seja, que algum relacionamento existe entre eles. Rotular correlações com base na importância é geralmente recomendado sobre o valor absoluto porque ele considera a variabilidade dos dados, por exemplo, um coeficiente de 0,6 pode ser altamente significativo em um conjunto de dados e nada significativo em outro. Por padrão, os valores de importância entre 0,0 e 0,9 são rotulados como *Fraco*, aqueles entre 0,9 e 0,95 são rotulados como *Médio* e aqueles entre 0,95 e 1,0 são rotulados como *Forte*.
- **Definir intensidade de correlação por valor absoluto.** Rotula correlações com base no valor absoluto do coeficiente de correlação de Pearson, que varia entre -1 e 1, conforme descrito acima. Quanto mais próximo o valor absoluto desta medida estiver de 1, mais forte será a correlação. Por padrão, as correlações padrão entre 0,0 e 0,3333 são rotuladas como *Fraco*, aquelas entre 0,3333 e 0,6666 são rotuladas como *Médio* e aquelas entre 0,6666 e 1,0 são rotuladas como *Forte*. Observe, no entanto, que a significância de qualquer valor fornecido é difícil de generalizar de um conjunto de dados para outro; por essa razão, definir correlações com base na probabilidade ao invés de no valor absoluto é recomendado na maioria dos casos.

Navegador de Saída de Estatísticas

O navegador de saída do nó Estatísticas exibe os resultados da análise estatística e permite executar operações, incluindo selecionar campos, gerar novos nós com base na seleção e salvar e imprimir os resultados. As opções de salvamento, de exportação e de impressão usuais estão disponíveis no menu Arquivo, e as opções de edição usuais estão disponíveis no menu Editar. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.

Ao procurar a saída de Estatísticas pela primeira vez, os resultados são expandidos. Para ocultar os resultados após visualizá-los, utilize o controle de expansão à esquerda do item para reduzir os resultados específicos que desejar ocultar ou clique no botão **Reduzir Tudo** para reduzir todos os resultados. Para ver os resultados novamente após reduzi-los, utilize o controle de expansão à esquerda do item para mostrar os resultados ou clique no botão **Expandir Tudo** para mostrar todos os resultados.

A saída contém uma seção para cada campo *Examinar*, contendo uma tabela das estatísticas solicitadas.

- **Contagem.** O número de registros com valores válidos para o campo.
- **Média.** O valor médio (média) para o campo em todos os registros.
- **Soma.** A soma de valores para o campo em todos os registros.
- **Mín.** O valor mínimo para o campo.
- **Máx.** O valor máximo para o campo.
- **Intervalo.** A diferença entre os valores mínimo e máximo.
- **Variância.** Uma medida da variabilidade dos valores de um campo. Ela é calculada ao tomar a diferença entre cada valor e a média geral, elevá-la ao quadrado, somar todos os valores e dividir pelo número de registros.
- **Desvio Padrão.** Outra medida de variabilidade nos valores de um campo, calculada como a raiz quadrada da variância.
- **Erro Padrão da Média** Uma medida da incerteza na estimativa da média do campo, se supor que a média será aplicada aos novos dados.
- **Mediana.** O valor "médio" para o campo; ou seja, o valor que divide a metade superior dos dados da metade inferior dos mesmos (com base nos valores do campo).
- **Modo.** O valor único mais comum nos dados.

Correlações. Se campos de correlação forem especificados, a saída também conterá uma seção listando a correlação de Pearson entre o campo Examinar e cada campo de correlação, além de rótulos descritivos opcionais para os valores de correlação. Consulte o tópico “Configurações de correlação” na página 324 para obter mais informações.

Menu Gerar. O menu Gerar contém operações de geração do nó.

- **Filtro.** Gera um nó Filtro para filtrar campos que não estiverem correlacionados, ou que estiverem fracamente correlacionados com outros campos.

Gerando um Nó Filtro a partir de Estatísticas

O nó Filtro gerado a partir de um navegador de saída de Estatísticas filtrará os campos com base em suas correlações com outros campos. Ele funciona ao classificar as correlações em ordem de valor absoluto, usando as maiores correlações (de acordo com o critério configurado na caixa de diálogo Gerar Filtro a partir de Estatísticas), e ao criar um filtro que transmite todos os campos que aparecem em qualquer uma dessas correlações grandes.

Modo. Decida como selecionar as correlações. **Incluir** faz com que os campos que aparecem nas correlações especificadas sejam retidos. **Excluir** faz com que os campos sejam filtrados.

Incluir/Excluir campos que aparecem em. Define o critério para seleção de correlações.

- **Número máximo de correlações.** Seleciona o número especificado de correlações e inclui/exclui campos que aparecem em qualquer uma dessas correlações.
- **Maior porcentagem de correlações (%).** Seleciona a porcentagem especificada ($n\%$) de correlações e inclui/exclui campos que aparecem em qualquer uma dessas correlações.
- **Correlações maiores que.** Seleciona correlações maiores em valor absoluto do que o limite especificado.

Nó Médias

O nó Média compara a média entre grupos independentes ou entre pares de campos relacionados para testar se há uma diferença significativa. Por exemplo, é possível comparar receitas médias antes e depois de realizar uma promoção ou comparar receitas de clientes que não receberam a promoção com aqueles que receberam.

É possível comparar médias de duas formas diferentes, dependendo de seus dados:

- **Entre grupos em um campo.** Para comparar grupos independentes, selecione um campo de teste e um campo de agrupamento. Por exemplo, é possível excluir uma amostra de clientes de "validação" quando enviar uma promoção e comparar as receitas médias do grupo de validação com todos os outros grupos. Nesse caso, é possível especificar um campo de teste único que indica a receita de cada cliente, com um campo de sinalização ou nominal indicando se eles receberam a oferta. As amostras são independentes no sentido de que cada registro é designado a um grupo ou outro, e não é possível vincular um membro específico de um grupo a um membro específico de outro. Também é possível especificar um campo nominal com mais de dois valores para comparar as médias de diversos grupos. Quando executado, o nó calcula um teste ANOVA unidirecional nos campos selecionados. Nos casos em que houver apenas dois grupos de campo, os resultados do ANOVA unidirecional são essencialmente os mesmos que um teste t de amostras independentes. Consulte o tópico "Comparando Médias para Grupos de Independentes" para obter mais informações.
- **Entre pares de campos.** Ao comparar médias para dois campos relacionados, os grupos deverão ser emparelhados de alguma forma para que os resultados sejam significativos. Por exemplo, é possível comparar as receitas média do mesmo grupo de clientes antes e depois de executar uma promoção ou comparar taxas de uso de um serviço entre pares de marido-esposa para ver se eles são diferentes. Cada registro contém duas medidas separadas, porém relacionadas, que podem ser comparadas de forma significativa. Quando executado, o nó calcula um teste t de amostras emparelhadas em cada par de campos selecionado. Consulte o tópico "Comparando Médias Entre Campos Emparelhados" para obter mais informações.

Comparando Médias para Grupos de Independentes

Selecione **Entre grupos em um campo** no nó Médias para comparar a média para dois ou mais grupos independentes.

Campos de agrupamento. Selecione um campo de sinalização ou nominal numérico com dois ou mais valores distintos que divide os registros nos grupos que deseja comparar, como aqueles que receberam uma oferta com aqueles que não receberam. Independentemente do número de campos de teste, somente um campo de agrupamento pode ser selecionado.

Campos de teste. Selecione um ou mais campos numéricos que contenham as medidas que deseja testar. Um teste separado será conduzido para cada campo que selecionar. Por exemplo, é possível testar o impacto de uma determinada promoção quanto ao uso, receita e perda de clientes.

Comparando Médias Entre Campos Emparelhados

Selecione **Entre pares de campos** no nó Médias para comparar médias entre campos separados. O campos devem estar relacionados de alguma forma para que os resultados sejam significativos, como receitas antes e após uma promoção. Diversos pares de campos também podem ser selecionados.

Campo um. Selecionar um campo numérico que contenha a primeira das medidas que deseja comparar. Em um estudo antes e depois, este seria o campo "antes".

Campo dois. Selecione o segundo campo que deseja comparar.

Incluir. Inclui o par selecionado na lista de pares de campos de Teste.

Repita as seleções de campos conforme necessário para incluir diversos pares na lista.

Configurações de correlação. Permite especificar opções para rotular a intensidade das correlações. Consulte o tópico "Configurações de correlação" na página 324 para obter mais informações.

Opções do Nó Médias

A guia Opções permite configurar os valores p de limite utilizados para rotular resultados como importante, marginal ou insignificante. Também é possível editar o rótulo para cada ranqueamento. A importância é medida em uma escala de porcentagem e pode ser amplamente definida como 1 menos a probabilidade de obter um resultado (como a diferença nas médias entre dois campos) tão extremo quanto ou mais extremo que o resultado observado em apenas uma tentativa. Por exemplo, um valor p maior que 0,95 indica menos de 5% de chance de que o resultado possa ser explicado somente por uma tentativa.

Rótulos de importância. É possível editar os rótulos utilizados para rotular cada par ou grupo de campos na saída. Os rótulos padrão são *importante*, *marginal* e *insignificante*.

Valores de corte. Especifica o limite para cada ranqueamento. Geralmente valores p maiores que 0,95 classificam como importantes, ao passo que valores inferiores a 0,9 seriam insignificantes, e esses limites podem ser ajustados conforme necessário.

Nota: medidas de importância estão disponíveis em um número de nós. Os cálculos específicos dependem do nó e do tipo de campos de destino e de entrada utilizados, e os valores ainda podem ser comparados, já que tudo é medido em uma escala de porcentagem.

Navegador de Saída do Nó Médias

O navegador de saída Médias exibe dados de tabulação cruzada e permite executar operações padrão, incluindo selecionar e copiar a tabela uma linha por vez, ordenar por qualquer coluna e salvar e imprimir a tabela. Consulte o tópico "Visualizando a Saída" na página 302 para obter mais informações.

As informações específicas na tabela dependem do tipo de comparação (grupos dentro de um campo ou campos separados).

Ordenar por. Permite ordenar a saída por uma coluna específica. Clique na seta para cima ou para baixo para alterar a direção da ordenação. Como alternativa, é possível clicar em qualquer título de coluna para ordenar por essa coluna. (Para alterar a direção da ordenação dentro da coluna, clique novamente).

Visualização. É possível escolher **Simple**s ou **Avançado** para controlar o nível de detalhe na exibição. A visualização avançada inclui todas as informações da visualização simples, mas com detalhes adicionais fornecidos.

Saída de Médias Comparando Grupos em um Campo

Ao comparar grupos dentro de um campo, o nome do campo de agrupamento é exibido acima da tabela de saída, e as médias e estatísticas relacionadas são relatadas separadamente para cada grupo. A tabela inclui uma linha separada para cada campo de teste.

As colunas a seguir são exibidas:

- **Campo.** Lista os nomes dos campos de teste selecionados.

- **Médias por grupo.** Exibe a média para cada categoria do campo de agrupamento. Por exemplo, é possível comparar quem recebeu uma oferta especial (*Nova Promoção*) com aqueles que não receberam (*Padrão*). Na visualização avançada, o desvio padrão, o erro padrão e a contagem também são exibidos.
- **Importância.** Exibe o valor e o rótulo da importância. Consulte o tópico “Opções do Nó Médias” na página 327 para obter mais informações.

Saída Avançada

Na visualização avançada, as seguintes colunas adicionais são exibidas.

- **Teste F.** Esse teste baseia-se na razão da variância entre os grupos e a variância dentro de cada grupo. Se as médias forem as mesmas para todos os grupos, você espera que a razão de F esteja próxima de 1, já que ambas são estimativas da mesma variância da população. Quanto maior essa razão, maior será a variação entre os grupos e maior será a chance de existir uma diferença significativa.
- **df.** Exibe os graus de liberdade.

Saída de Médias Comparando Pares de Campos

Ao comparar campos separados, a tabela de saída inclui uma linha para cada par de campo selecionado.

- **Campo Um/Dois.** Exibe os nomes do primeiro e do segundo campo em cada par. Na visualização avançada, o desvio padrão, o erro padrão e a contagem também são exibidos.
- **Média Um/Dois.** Exibe a média para cada campo, respectivamente.
- **Correlação.** Mede a intensidade do relacionamento entre dois campos contínuos (intervalo numérico). Valores próximos de +1,0 indicam uma forte associação positiva, e valores próximos de -1,0 indicam uma forte associação negativa. Consulte o tópico “Configurações de correlação” na página 324 para obter mais informações.
- **Diferença Média.** Exibe a diferença entre as duas médias de campo.
- **Importância.** Exibe o valor e o rótulo da importância. Consulte o tópico “Opções do Nó Médias” na página 327 para obter mais informações.

Saída Avançada

A saída avançada inclui as seguintes colunas:

Intervalo de Confiança de 95%. Os limites inferior e superior do intervalo dentro do qual a média real deverá cair em 95% de todas as amostras possíveis deste tamanho a partir dessa população.

Teste T. A estatística t é obtida ao dividir a diferença média pelo seu erro padrão. Quanto maior o valor absoluto desta estatística, maior a probabilidade de que as médias não sejam as mesmas.

df. Exibe os graus de liberdade para a estatística.

Nó Relatório

O nó Relatório permite criar relatórios formatados contendo texto fixo, bem como dados e outras expressões derivadas dos dados. Especifique o formato do relatório ao utilizar os modelos de texto para definir as construções de texto fixo e de saída de dados. É possível fornecer formatação de texto customizada utilizando tags HTML no modelo e configurando as opções na guia Saída. Valores de dados e outra saída condicional são incluídos no relatório utilizando as expressões do CLEM no modelo.

Alternativas para o Nó Relatório

O nó Relatório é mais geralmente utilizado para listar a saída de registros ou de casos a partir de um fluxo, como todos os registros que atendem a uma determinada condição. Neste sentido, ele pode ser considerado como uma alternativa menos estruturada para o nó Tabela.

- Se desejar um relatório que liste as informações do campo ou tudo mais que estiver definido no fluxo ao invés dos dados em si (como definições de campo especificadas em um nó Tipo), um script poderá ser utilizado.
- Para gerar um relatório que inclua diversos objetos de saída (como uma coleção de modelos, tabelas e gráficos gerados por um ou mais fluxos) e que possa ser gerado em diversos formatos (incluindo texto, HTML e Microsoft Word/Office), um projeto do IBM SPSS Modeler poderá ser utilizado.
- Para produzir uma lista de nomes de campo sem usar script, é possível usar um nó Tabela precedido por um nó Amostra que descarta todos os registros. Isso produz uma tabela sem linhas, que poderá ser transposta ou exportada para produzir uma lista de nomes de campo em uma única coluna. (Selecione **Transpor dados** na guia Saída no nó Tabela para fazer isso).

Guia Modelo do Nó Relatório

Criando um modelo. Para definir o conteúdo do relatório, crie um modelo na guia Modelo do nó Relatório. O modelo consiste em linhas de texto, cada uma especificando algo sobre o conteúdo do relatório, e algumas linhas de tag especial utilizadas para indicar o escopo das linhas de conteúdo. Dentro de cada linha de conteúdo, as expressões do CLEM entre colchetes ([]) são avaliadas antes que a linha seja enviada para o relatório. Há três escopos possíveis para uma linha no modelo:

Fixo. As linhas que não forem marcadas de outra forma são consideradas fixas. As linhas fixas são copiadas para o relatório apenas uma vez, depois que quaisquer expressões que elas contiverem forem avaliadas. Por exemplo, a linha

```
This is my report, printed on [@TODAY]
```

copiará uma única linha para o relatório, contendo o texto e a data atual.

Global (iterar ALL). As linhas contidas entre as tags especiais #ALL e # são copiadas para o relatório uma vez para cada registro de dados de entrada. As expressões do CLEM (entre colchetes) são avaliadas com base no registro atual de cada linha de saída. Por exemplo, as linhas

```
#ALL
For record [@INDEX], the value of AGE is [AGE]
#
```

incluem uma linha para cada registro indicando o número de registro e a idade.

Para gerar uma lista de todos os registros:

```
#ALL
[Age] [Sex] [Cholesterol] [BP]
#
```

Condiciona (iterar WHERE). As linhas contidas entre as tags especiais #WHERE <condition> e # são copiadas para o relatório uma vez para cada registro quando a condição especificada for verdadeira. A condição é uma expressão CLEM. (Na condição WHERE, os colchetes são opcionais). Por exemplo, as linhas

```
#WHERE [SEX = 'M']
Male at record no. [@INDEX] has age [AGE].
#
```

gravam uma linha no arquivo para cada registro com um valor de M para o sexo. O relatório completo conterá as linhas fixas, globais e condicionais definidas ao aplicar o modelo nos dados de entrada.

É possível especificar opções para exibir ou salvar os resultados utilizando a guia Saída, comum para vários tipos de nós de saída. Consulte o tópico “Guia Saída do Nó Saída” na página 306 para obter mais informações.

Gerando Dados em Formato HTML ou XML

É possível incluir tags HTML ou XML diretamente no modelo para gravar relatórios em um desses formatos. Por exemplo, o modelo a seguir produz uma tabela HTML.

Este relatório será gravado em HTML.
Somente registros nos quais Age estiver acima de 60 são incluídos.

```
<HTML>
<TABLE border="2">
<TR>
<TD>Age</TD>
<TD>BP</TD>
<TD>Cholesterol</TD>
<TD>Drug</TD>
</TR>

#WHERE Age > 60
<TR>
<TD>[Age]</TD>
<TD>[BP]</TD>
<TD>[Cholesterol]</TD>
<TD>[Drug]</TD>
</TR>
#
</TABLE>
</HTML>
```

Navegador de Saída do Nó Relatório

O navegador de relatório mostra o conteúdo do relatório gerado. As opções de salvamento, de exportação e de impressão usuais estão disponíveis no menu Arquivo, e as opções de edição usuais estão disponíveis no menu Editar. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.

Nó Configurar Globais

O nó Configurar Globais varre os dados e calcula os valores de resumo que podem ser utilizados em expressões do CLEM. Por exemplo, é possível utilizar um nó Configurar Globais para calcular estatísticas de um campo chamado *idade* e, em seguida, utilizar a média geral de *idade* nas expressões do CLEM ao inserir a função @GLOBAL_MEAN(*idade*).

Guia Configurações do Nó Configurar Globais

Globais a serem criados. Selecione um ou mais campos para os quais deseja que os globais estejam disponíveis. Você pode selecionar vários campos. Para cada campo, especifique as estatísticas a serem calculadas, assegurando-se de que as estatísticas desejadas sejam selecionadas nas colunas ao lado do nome do campo.

- **MEAN.** O valor médio (média) para o campo em todos os registros.
- **SUM.** A soma de valores para o campo em todos os registros.
- **MIN.** O valor mínimo para o campo.
- **MAX.** O valor máximo para o campo.
- **SDEV.** O desvio padrão, que é uma medida da variabilidade dos valores de um campo e é calculada como a raiz quadrada da variância.

Operação(ões) padrão. As opções selecionadas aqui serão utilizadas quando novos campos forem incluídos na lista Globais acima. Para alterar a configuração padrão das estatísticas, selecione ou cancele a seleção das estatísticas, conforme apropriado. Também é possível utilizar o botão **Aplicar** para aplicar as operações padrão a todos os campos na lista.

Nota: Algumas operações não são aplicáveis a campos não numéricos (por exemplo, Soma para um campo de data/hora). As operações que não puderem ser utilizadas com um campo selecionado são desativadas

Limpar todos os globais antes de executar. Selecione esta opção para remover todos os valores globais antes de calcular novos valores. Se essa opção não estiver selecionada, os valores calculados recentemente substituem valores mais antigos, e os globais que não forem recalculados permanecem disponíveis também.

Exibir visualização de globais criados após a execução. Se selecionar essa opção, a guia Globais da caixa de diálogo de propriedades do fluxo aparecerá após a execução para exibir os valores globais calculados.

Nó Ajuste de Simulação

O nó Ajuste de Simulação ajusta um conjunto de distribuições estatísticas candidatas para cada campo nos dados. O ajuste de cada distribuição para um campo é avaliado utilizando um critério de Qualidade do ajuste. Quando um nó Ajuste de Simulação é executado, um nó Geração de Simulação é construído (ou um nó existente é atualizado). Cada campo é designado a sua melhor distribuição de ajuste. Em seguida, o nó Geração de Simulação poderá ser utilizado para gerar dados simulados para cada campo.

Embora o nó Ajuste de Simulação seja um nó terminal, ele não inclui um modelo na paleta de modelos gerados, inclui uma saída ou um gráfico na guia saídas ou exporta dados.

Nota: Se os dados históricos forem esparsos (ou seja, existem muitos valores ausentes), poderá ser difícil para o componente de ajuste localizar valores válidos suficientes para ajustar as distribuições nos dados. Nos casos em que os dados são esparsos, é necessário remover os campos esparsos se eles não forem necessários ou imputar os valores omissos antes do ajuste. Utilizando as opções na guia **Qualidade** do nó de Auditoria de Dados, é possível visualizar o número de registros completos, identificar quais campos são esparsos e selecionar um método de imputação. Se houver um número insuficiente de registros para ajuste de distribuição, é possível utilizar um nó Balanceamento para aumentar o número de registros.

Usando um nó Ajuste de Simulação para criar automaticamente um nó Geração de Simulação

Quando o nó Ajuste de Simulação é executado pela primeira vez, um nó Geração de Simulação é criado com uma ligação de atualização para o nó Ajuste de Simulação. Se o nó Ajuste de Simulação for executado novamente, um novo nó Geração de Simulação será criado apenas se a ligação de atualização tiver sido removida. Um nó Ajuste de Simulação também pode ser utilizado para atualizar um nó Geração de Simulação conectado. O resultado depende se os mesmos campos estiverem presentes nos dois nós e se os campos estiverem desbloqueados no nó Geração de Simulação. Consulte o tópico “Nó Geração de Simulação” na página 50 para obter mais informações.

Um nó Ajuste de Simulação pode ter apenas uma ligação de atualização com um nó Geração de Simulação. Para definir uma ligação de atualização para um nó Geração de Simulação, siga esses passos:

1. Clique com o botão direito no nó Ajuste de Simulação.
2. No menu, selecione **Definir Ligação de Atualização**.
3. Clique no nó Geração de Simulação para o qual deseja definir uma ligação de atualização.

Para remover uma ligação de atualização entre um nó Ajuste de Simulação e um nó de Geração de Simulação, clique com o botão direito na ligação de atualização e selecione **Remover Ligação**.

Ajuste de distribuição

Uma distribuição estatística é a frequência teórica da ocorrência de valores que uma variável pode assumir. No nó Ajuste de Simulação, um conjunto de distribuições estatísticas teóricas é comparado com cada campo de dados. As distribuições que estiverem disponíveis para ajuste são descritas no tópico “Distribuições” na página 60. Os parâmetros da distribuição teórica são ajustados para oferecer o melhor ajuste para os dados de acordo com uma medida de qualidade do ajuste, que é o critério do Anderson-Darling ou o critério do Kolmogorov-Smirnov. Os resultados do ajuste de distribuição pelo nó Ajuste de Simulação mostram quais distribuições foram ajustadas, as melhores estimativas dos parâmetros para cada distribuição e quão bem cada distribuição ajusta os dados. Durante o ajuste de distribuição, as correlações entre campos com tipos de armazenamento numérico e as contingências entre os campos com uma distribuição categórica também são calculadas. Os resultados do ajuste de distribuição são usados para criar um nó Geração de Simulação.

Antes que quaisquer distribuições sejam ajustadas em seus dados, os primeiros 1000 registros são examinados para ver se há valores omissos. Se houver valores omissos em excesso, o ajuste da distribuição não será possível. Se isso ocorrer, deve-se decidir se uma das opções a seguir é apropriada:

- Usar um nó de envio de dados para remover registros com valores omissos.
- Usar um nó de envio de dados para imputar valores para valores omissos.

O ajuste de distribuição não exclui valores omissos de usuário. Se seus dados tiverem valores omissos de usuário e você desejar que esses valores sejam excluídos do ajuste de distribuição, então deve-se configurar esses valores para omissos do sistema.

O papel de um campo não é levado em conta quando as distribuições são ajustadas. Por exemplo, campos com o papel **Destino** são tratados da mesma forma que os campos com papéis de **Entrada**, **Nenhum**, **Ambos**, **Partição**, **Divisão**, **Frequência** e **ID**.

Os campos são tratados de maneira diferente durante o ajuste de distribuição de acordo com seu tipo de armazenamento e com o nível de medição. O tratamento dos campos durante o ajuste de distribuição é descrito na tabela a seguir.

Tabela 56. Ajuste de distribuição de acordo com o tipo de armazenamento e com o nível de medição dos campos

Tipo de armazenamento	Nível de medição					
	Contínuo	Catagórico	Sinalizador	Nominal	Ordinal	Sem tipo
Sequência de caracteres	Impossível	Catagórico, distribuições dice e fixas são ajustadas				O campo é ignorado e não transmitido para o nó Geração de Simulação.
Número inteiro	Todas as distribuições são ajustadas. As correlações e contingências são calculadas.	A distribuição catagórica é ajustada. As correlações não são calculadas.			Distribuições binomiais, binomiais negativas e de Poisson são ajustadas, e as correlações calculadas.	
Real						
Hora						
Data						
Registro de data e hora						
Desconhecido	O tipo de armazenamento apropriado é determinado a partir dos dados.					

Os campos com nível de medição ordinal são tratados como campos contínuos e são incluídos na tabela de correlações no nó Geração de Simulação. Se desejar que uma distribuição diferente de binomial, binomial negativa ou de Poisson seja ajustada em um campo ordinal, deve-se alterar o nível de medição do campo para contínuo. Se você tiver definido anteriormente um rótulo para cada valor de um campo ordinal e, em seguida, alterar o nível de medição para contínuo, os rótulos serão perdidos.

Os campos que tiverem valores únicos são tratados da mesma maneira que os campos com diversos valores durante o ajuste de distribuição. Os campos com o tipo de armazenamento de hora, data ou registro de data e hora são tratados como numéricos.

Ajustando Distribuições para Campos de Divisão

Se seus dados contiverem um campo de divisão e você desejar que o ajuste de distribuição seja executado separadamente para cada divisão, os dados deverão ser transformados utilizando um nó Reestruturação de envio de dados. Utilizando o nó Reestruturação, gere um novo campo para cada valor do campo de divisão. Esses dados reestruturados podem, em seguida, ser utilizados para ajuste de distribuição no Ajuste de Simulação.

Guia Configurações do Nó Ajuste de Simulação

Nome do nó de origem. É possível gerar o nome do nó Geração de Simulação gerado (ou atualizado) automaticamente ao selecionar **Automático**. O nome gerado automaticamente será o nome especificado no nó Ajuste de Simulação se um nome customizado tiver sido especificado (ou Sim Gen, se nenhum nome customizado tiver sido especificado no nó Ajuste de Simulação). Selecione **Customizado** para especificar um nome customizado no campo de texto adjacente. A menos que o campo de texto seja editado, o nome customizado padrão é Sim Gen.

Opções de Ajuste Com estas opções, é possível especificar como as distribuições são ajustadas para os campos e como o ajuste das distribuições é avaliado.

- **Número de casos para amostra.** Isso especifica o número de casos a serem utilizados quando ajustar distribuições para os campos no conjunto de dados. Selecione **Todos os casos** para ajustar as distribuições para todos os registros nos dados. Se o conjunto de dados for muito grande, você poderá querer considerar a limitação do número de casos que são utilizados para ajuste de distribuição. Selecione **Limitar aos N primeiros casos** para utilizar apenas os N primeiros casos. Clique nas setas para especificar o número de casos a serem utilizados. Como alternativa, é possível utilizar um nó de envio de dados para tirar uma amostra aleatória dos registros para ajuste de distribuição.
- **Critérios de Qualidade do ajuste (campos contínuos somente).** Para campos contínuos, selecione o teste do Anderson-Darling ou o teste da Qualidade do ajuste do Kolmogorov-Smirnoff para classificar as distribuições quando ajustá-las para os campos. O teste do Anderson-Darling é selecionado por padrão e recomendado especialmente quando desejar assegurar o melhor ajuste possível nas regiões de rodapé. As duas estatísticas são calculadas para cada distribuição candidata, mas somente a estatística selecionada é utilizada para ordenar as distribuições e determinar a melhor distribuição de ajuste.
- **Categorias (distribuição empírica somente).** Para campos contínuos, a distribuição Empírica é a função de distribuição acumulativa dos dados históricos. Ela é a probabilidade de cada valor ou de um intervalo de valores e é derivada diretamente dos dados. É possível especificar o número de categorias que são utilizadas para calcular a distribuição Empírica para campos contínuos ao clicar nas setas. O padrão é 100 e o máximo é 1000.
- **Campo de ponderação (opcional).** Se o conjunto de dados contiver um campo de ponderação, clique no ícone seletor de campo e selecione o campo de ponderação na lista. O campo de ponderação é então excluído do processo de ajuste de distribuição. A lista mostra todos os campos no conjunto de dados que tiverem o nível de medição contínuo. É possível selecionar apenas um campo de ponderação.

Nó Avaliação de Simulação

O nó Avaliação de Simulação é um nó terminal que avalia um campo especificado, fornece uma distribuição do campo e produz gráficos de distribuições e de correlações. Este nó é utilizado principalmente para avaliar campos contínuos. Portanto, ele complementa o gráfico de avaliação, que é gerado por um nó Avaliação e é útil para avaliar campos discretos. Outra diferença é que o nó Avaliação de Simulação avalia uma predição única em várias iterações, ao passo que o nó Avaliação avalia diversas predições, cada uma com uma iteração única. As iterações são geradas quando mais de um valor for

especificado para um parâmetro de distribuição no nó Geração de Simulação. Consulte o tópico “Iterações” na página 60 para obter mais informações.

O nó Avaliação de Simulação é projetado para ser utilizado com dados que foram obtidos dos nós Ajuste de Simulação e Geração de Simulação. O nó pode, no entanto, ser utilizado com qualquer outro nó. Qualquer número de passos de processamento pode ser colocado entre o nó de Geração de Simulação e o nó Avaliação de Simulação.

Importante: O nó Avaliação de Simulação requer no mínimo 1000 registros com valores válidos para o campo de destino.

Guia Configurações do Nó de Avaliação de Simulação

Na guia Configurações do nó Avaliação de Simulação, é possível especificar o papel de cada campo em seu conjunto de dados e customizar a saída que é gerada pela simulação.

Selecionar um item. Permite alternar entre as três visualizações do nó Avaliação de Simulação: Campos, Funções de Densidade e Saídas.

Visualização Campos

Campo de destino. Esse é um campo obrigatório. Clique na seta para selecionar o campo de destino de seu conjunto de dados a partir da lista suspensa. O campo selecionado pode ter um nível de medição contínuo, nominal ou ordinal, mas não pode ter data ou um nível de medição não especificado.

Campo de iteração (opcional). Se os dados tiverem um campo iteração que indica à qual iteração pertence cada registro em seus dados, ele deverá ser selecionado aqui. Isso significa que cada iteração será avaliada separadamente. Apenas os campos com níveis de medição contínuos, ordinais ou nominais podem ser selecionados.

Dados de entrada já ordenados pela iteração. Ativada somente se um campo de iteração estiver especificado no campo **Campo de iteração (opcional)**. Selecione esta opção se tiver certeza de que seus dados de entrada já estão ordenados pelo campo de iteração especificado em **Campo de iteração (opcional)**.

Número máximo de iterações para representar. Ativada somente se um campo de iteração estiver especificado no campo **Campo de iteração (opcional)**. Clique nas setas para especificar o número de iterações para representar. A especificação deste número evita a tentativa de representar iterações em excesso em um único gráfico, o que tornaria o gráfico difícil de interpretar. O nível mais baixo para o qual o número máximo de iterações pode ser configurado é 2 e o nível mais alto é 50. O número máximo de iterações para representar é inicialmente configurado para 10.

Campos de entrada para tornado de correlação. O gráfico tornado de correlação é um gráfico de barras que exhibe os coeficientes de correlação entre a resposta especificada e cada uma das entradas especificadas. Clique no ícone seletor de campo para selecionar os campos de entrada a serem incluídos no gráfico tornado, a partir de uma lista de entradas simuladas disponíveis. Apenas os campos de entrada com níveis de medição contínuos e ordinais podem ser selecionados. Campos nominais, sem tipo e de entrada de data não estão disponíveis na lista e não podem ser selecionados.

Visualização Funções de Densidade

Com as opções nessa visualização, é possível customizar a saída para as funções de densidade de probabilidade e funções de distribuição acumulativas para variáveis resposta contínuas, bem como gráficos de barras de valores preditos para variáveis resposta categóricas.

Funções de densidade. As funções de densidade são as médias primárias de análise do conjunto de resultados de sua simulação.

- **Função de densidade de probabilidade (PDF).** Selecione esta opção para produzir uma função de densidade de probabilidade para o campo de destino. A função de densidade de probabilidade exibe a distribuição dos valores de resposta. É possível utilizar a função de densidade de probabilidade para determinar a probabilidade de que a resposta esteja dentro de uma região específica. Para variáveis resposta categóricas (respostas com um nível de medição de nominal ou ordinal) um gráfico de barras é gerado exibindo a porcentagem de casos que caem em cada categoria da resposta.
- **Função de distribuição acumulativa (CDF).** Selecione esta opção para produzir uma função de distribuição acumulativa para o campo de destino. A função de distribuição acumulativa exibe a probabilidade de que o valor da resposta seja menor ou igual a um valor especificado. Ela está disponível apenas para variáveis resposta contínuas.

Linhas de referência (contínuas). Essas opções serão ativadas se a **Função de densidade de probabilidade (PDF)**, **Função de distribuição acumulativa (CDF)** ou ambas forem selecionadas. Com essas opções, é possível incluir várias linhas de referência verticais fixas nas funções de densidade de probabilidade e funções de distribuição acumulativa.

- **Média.** Selecione esta opção para incluir uma linha de referência no valor médio do campo de destino.
- **Mediana.** Selecione esta opção para incluir uma linha de referência no valor de mediana do campo de destino.
- **Desvios Padrão.** Selecione esta opção para incluir linhas de referência a mais e a menos com relação a um número especificado de desvios padrão da média do campo de destino. Selecionar essa opção ativa o campo **Número** adjacente. Clique nas setas para especificar o número de desvios padrão. O número mínimo de desvios padrão é 1 e o máximo é 10. O número de desvios padrão é inicialmente configurado para 3.
- **Percentis.** Selecione esta opção para incluir linhas de referência em dois valores de percentil da distribuição do campo de destino. Selecionar essa opção ativa os campos de texto **Inferior** e **Superior**. Por exemplo, inserir um valor de 90 no campo de texto **Superior** inclui uma linha de referência no 90º percentil da resposta, que é o valor abaixo do qual 90% das observações caem. Da mesma forma, um valor de 10 no campo de texto **Inferior** representa o décimo percentil da resposta, que é o valor abaixo do qual 10% das observações caem.
- **Linhas de referência customizadas.** Selecione esta opção para incluir linhas de referência em valores especificados ao longo do eixo horizontal. Selecionar essa opção ativa a tabela **Valores** adjacente. Sempre que um número válido for inserido na tabela **Valores**, uma nova linha vazia é anexada na parte inferior da tabela. Um número *válido* é um número dentro do intervalo de valores do campo de destino

Nota: Quando diversas funções de densidade ou funções de distribuição (a partir de diversas iterações) são exibidas em um gráfico único, as linhas de referência (diferentes de linhas customizadas) são aplicadas separadamente à cada função.

Variável resposta categórica (PDF somente). Essas opções são ativadas apenas se a **Função de densidade de probabilidade (PDF)** for selecionada.

- **Valores de categoria para relatório.** Para modelos com campos de destino categóricos, o resultado do modelo é um conjunto de probabilidades preditas, uma para cada categoria, de que o valor de resposta caia em cada categoria. A categoria com a probabilidade mais alta é considerada a categoria predita e utilizada na geração do gráfico de barras para a função de densidade de probabilidade. Selecione **Categoria predita** para gerar o gráfico de barras. Selecione **Probabilidades preditas** para gerar histogramas da distribuição de probabilidades preditas para cada uma das categorias do campo de destino. Também é possível selecionar **Ambos** para gerar os dois tipos de gráficos.
- **Agrupamento para análise de sensibilidade.** Simulações que incluem iterações de análise de sensibilidade geram um campo de destino independente (ou campo de destino predito de um modelo) para cada iteração que for definida pela análise. Há uma iteração para cada valor do parâmetro de distribuição que está sendo variado. Quando as iterações estiverem presentes, o gráfico de barras da

categoria predita para um campo de destino categórico é exibido como um gráfico de barras em cluster que inclui os resultados de todas as iterações. Selecione o **Agrupar categorias** ou **Agrupar iterações**.

Visualização Saídas

Valores de percentil de distribuições de resposta. Com essas opções, é possível escolher criar uma tabela de valores de percentil das distribuições de resposta e especificar os percentis a serem exibidos.

Criar uma tabela de valores de percentil. Para campos de destino contínuos, selecione esta opção para obter uma tabela de percentis especificados das distribuições de resposta. Escolha uma das seguintes opções para especificar os percentis:

- **Quartis.** Os quartis são os 25^o, 50^o e 75^o percentis da distribuição do campo de destino. As observações são divididas em quatro grupos de tamanhos iguais.
- **Intervalos.** Se desejar um número igual de grupos diferente de quatro, selecione **Intervalos**. Selecionar essa opção ativa o campo **Número** adjacente. Clique nas setas para especificar o número de intervalos. O número mínimo de intervalos é 2 e o máximo é 100. O número de intervalos é inicialmente configurado para 10.
- **Percentis customizados.** Selecione **Percentis customizados** para especificar percentis individuais, por exemplo, o 99^o percentil. Selecionar essa opção ativa a tabela **Valores** adjacente. Sempre que um número válido, entre 1 e 100, for inserido na tabela **Valores**, uma nova linha vazia é anexada na parte inferior da tabela.

Saída do Nó Avaliação de Simulação

Quando o nó Avaliação de Simulação é executado, a saída é incluída no gerenciador de Saída. O navegador de saída de Avaliação de Simulação exibe os resultados da execução do nó Avaliação de Simulação. As opções de salvamento, de exportação e de impressão usuais estão disponíveis no menu **Arquivo**, e as opções de edição usuais estão disponíveis no menu **Editar**. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações. O menu **Exibir** é ativado apenas se um dos gráficos for selecionado. Ele não é ativado para a tabela de distribuição ou para as saídas de informações. No menu **Visualização**, é possível selecionar **Modo de Edição** para alterar o layout e a aparência do gráfico ou **Modo Explorar** para explorar os dados e os valores representados pelo gráfico. O modo Estático corrige as linhas de referência do gráfico (e régua de controle) em suas posições atuais para que elas não possam ser movidas. O modo Estático é o único modo no qual é possível copiar, exportar ou imprimir o gráfico com suas linhas de referência. Para selecionar este modo, clique em **Modo Estático** no menu **Visualizar**.

A janela do navegador de saída de Avaliação de Simulação consiste de dois painéis. No lado esquerdo da janela, há um painel de navegação que exibe representações em miniatura dos gráficos que foram gerados quando o nó Avaliação de Simulação foi executado. Quando uma miniatura é selecionada, a saída do gráfico é exibida no painel à direita da janela.

Painel de Navegação

O painel de navegação do navegador de saída contém as miniaturas dos gráficos que são gerados a partir de uma simulação. As miniaturas que são mostradas no painel de navegação dependem do nível de medição do campo de destino e das opções que estiverem selecionadas na caixa de diálogo do nó Avaliação de Simulação. Descrições das miniaturas são fornecidas na tabela a seguir.

Tabela 57. Miniaturas do painel de navegação

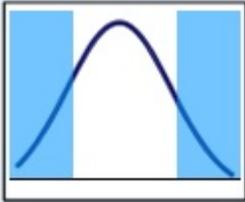
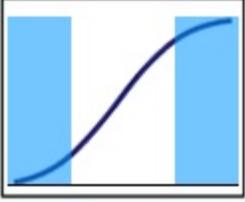
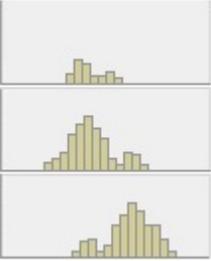
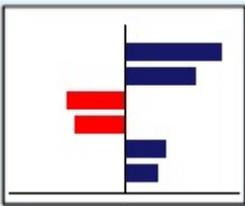
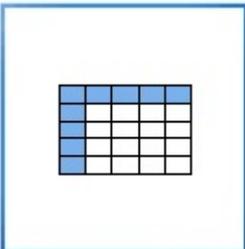
Miniatura	Descrição	Comentários
	<p>Função de densidade de probabilidade</p>	<p>Esta miniatura será mostrada apenas se o nível de medição do campo de destino for contínuo e se Função de Densidade de Probabilidade (PDF) estiver selecionada na visualização Funções de Densidade da caixa de diálogo do nó Avaliação de Simulação.</p> <p>Se o nível de medição do campo de destino for categórico, essa miniatura não será mostrada.</p>
	<p>Função de distribuição acumulativa</p>	<p>Esta miniatura será mostrada apenas se o nível de medição do campo de destino for contínuo e se Função de Distribuição Acumulativa (CDF) estiver selecionada na visualização Funções de Densidade da caixa de diálogo do nó Avaliação de Simulação.</p> <p>Se o nível de medição do campo de destino for categórico, essa miniatura não será mostrada.</p>
	<p>Valores de categoria predita</p>	<p>Esta miniatura será mostrada apenas se o nível de medição do campo de destino for categórico, se Função de Densidade de Probabilidade (PDF) estiver selecionada na visualização Funções de Densidade da caixa de diálogo do nó Avaliação de Simulação e se Categoria predita ou Ambas estiver selecionada na área Valores de categoria para relatório.</p> <p>Se o nível de medição do campo de destino for contínuo, essa miniatura não será mostrada.</p>
	<p>Probabilidades de categoria predita</p>	<p>Esta miniatura será mostrada apenas se o nível de medição do campo de destino for categórico, se Função de Densidade de Probabilidade (PDF) estiver selecionada na visualização Funções de Densidade da caixa de diálogo do nó Avaliação de Simulação e se Probabilidades preditas ou Ambas estiver selecionada na área Valores de categoria para relatório.</p> <p>Se o nível de medição do campo de destino for contínuo, essa miniatura não será mostrada.</p>

Tabela 57. Miniaturas do painel de navegação (continuação)

Miniatura	Descrição	Comentários
	Gráficos de Tornado	Esta miniatura será mostrada apenas se um ou mais campos de entrada forem selecionados no campo Campos de entrada para tornado de correlação na visualização Campos da caixa de diálogo do nó Avaliação de Simulação.
	Tabela de distribuição	Esta miniatura será mostrada apenas se o nível de medição do campo de destino for contínuo e se Criar uma tabela de valores de percentil estiver selecionada na visualização Saídas da caixa de diálogo do nó Avaliação de Simulação. O menu Visualizar é desativado para este gráfico. Se o nível de medição do campo de destino for categórico, essa miniatura não será mostrada.
	Informações	Essa miniatura é sempre mostrada. O menu Visualizar é desativado para esta saída.

Saída de Gráfico

Os tipos de gráficos de saída que estão disponíveis dependem do nível de medição do campo de destino, se um campo de iteração é usado e das opções que estiverem selecionadas na caixa de diálogo do nó Avaliação de Simulação. Diversos gráficos que são gerados a partir de uma simulação possuem recursos interativos que podem ser utilizados para customizar a exibição. Recursos interativos estão disponíveis ao clicar em **Opções de Gráfico**. Todos os gráficos de simulação são visualizações de gráfico.

Gráficos de função de densidade de probabilidade para variáveis resposta contínuas. Esse gráfico mostra a probabilidade e a frequência, com a escala de probabilidade no eixo vertical esquerdo e a escala de frequência no eixo vertical direito. O gráfico possui duas linhas de referência verticais deslizantes que dividem o gráfico em regiões separadas. A tabela abaixo do gráfico exibe o percentual da distribuição que estiver em cada uma das regiões. Se diversas funções de densidade forem exibidas no mesmo gráfico (por causa de iterações), a tabela terá uma linha separada para as probabilidades que estiverem associadas a cada função de densidade, uma coluna extra que contém o nome da iteração e uma cor que é associada a cada função de densidade. As iterações são listadas na tabela em ordem alfabética, de acordo com o rótulo de iteração. Se nenhum rótulo de iteração estiver disponível, o valor da iteração será usado. A tabela não pode ser editada.

Cada uma das linhas de referência possui uma régua de controle (triângulo invertido) que pode ser utilizada para mover facilmente a linha. Cada régua de controle possui um rótulo que indica sua posição atual. Por padrão, as régua de controle são posicionadas nos 5º e 95º percentis da distribuição. Se houver diversas iterações, as régua de controle serão posicionadas nos 5º e 95º percentis da primeira iteração que estiver listada na tabela. Não é possível mover as linhas de modo que se cruzem uma a outra.

Um número de recursos adicionais está disponível clicando em **Opções de Gráfico**. Em particular, é possível configurar explicitamente as posições das réguas de controle, incluir linhas de referência fixas e alterar a visualização do gráfico de uma curva contínua para um histograma. Consulte o tópico “Opções de gráfico” na página 340 para obter mais informações. Clique com o botão direito no gráfico para copiar ou exportar o gráfico.

Gráficos de função de distribuição acumulativa para variáveis resposta contínuas. Esse gráfico tem as mesmas duas linhas de referência verticais móveis e uma tabela associada que são descritas para o gráfico de função de densidade de probabilidade. Os controles da régua de controle e a tabela se comportam da mesma maneira que a função de densidade de probabilidade quando houver diversas iterações. As mesmas cores que são utilizadas para identificar qual função de densidade pertence a cada iteração são utilizadas para as funções de distribuição.

Esse gráfico também fornece acesso à caixa de diálogo Opções de Gráfico, que permite configurar explicitamente as posições das réguas de controle, incluir linhas de referência fixas e especificar se a função de distribuição acumulativa é exibida como uma função de aumento (o padrão) ou uma função de diminuição. Consulte o tópico “Opções de gráfico” na página 340 para obter mais informações. Clique com o botão direito no gráfico para copiar, exportar ou editar o gráfico. Selecionar **Editar** abre o gráfico em uma janela do editor de gráfico flutuante.

Gráfico de valores de categoria preditos para variáveis resposta categóricas. Para campos de variável resposta categórica, um gráfico de barras exibe os valores preditos. Os valores preditos são exibidos como o percentual do campo de destino que é predito cair em cada categoria. Para campos de variável resposta categórica com iterações de análise de sensibilidade, os resultados para a categoria de resposta predita são exibidos como um gráfico de barras agrupadas que inclui os resultados para todas as iterações. O gráfico é armazenado em cluster por categoria ou por iteração, dependendo de qual opção foi selecionada na área **Agrupamento para análise de sensibilidade** na visualização Funções de Densidade da caixa de diálogo do nó Avaliação de Simulação. Clique com o botão direito no gráfico para copiar, exportar ou editar o gráfico. Selecionar **Editar** abre o gráfico em uma janela do editor de gráfico flutuante.

Gráfico de probabilidades de categoria predita para variáveis resposta categóricas. Para campos de variável resposta categórica, um histograma exibe a distribuição de probabilidades preditas para cada uma das categorias da resposta. Para campos de variável resposta categórica com iterações de análise de sensibilidade, os histogramas são exibidos por categoria ou por iteração, dependendo de qual opção foi selecionada na área **Agrupamento para análise de sensibilidade** na visualização Funções de Densidade da caixa de diálogo do nó Avaliação de Simulação. Se os histogramas forem agrupados por categoria, uma lista suspensa contendo os rótulos da iteração permitirá escolher qual iteração será exibida. Também é possível selecionar a iteração a ser exibida ao clicar com o botão direito no gráfico e selecionar a iteração no submenu **Iteração**. Se os histogramas forem agrupados por iteração, uma lista suspensa contendo os nomes de categoria permitirá escolher qual categoria será exibida. Também é possível selecionar a categoria a ser exibida ao clicar com o botão direito no gráfico e selecionar a categoria no submenu **Categoria**.

Esse gráfico está disponível apenas para um subconjunto de modelos e a opção para gerar todas as probabilidades do grupo deverá ser selecionada no nugget do modelo. Por exemplo, no nugget do modelo de Logística, deve-se selecionar **Anexar todas as probabilidades**. Os nuggets do modelo a seguir suportam essa opção:

- Logística, SVM, Bayes, Rede Neural e KNN
- Modelos de mineração dentro da base de dados do DB2/ISW para regressão logística, árvores de decisão e Naive Bayes

Por padrão, a opção para gerar todas as probabilidades de grupo não está selecionada nestes nuggets do modelo.

Gráficos Tornados. O gráfico tornado é um gráfico de barras que mostra a sensibilidade do campo de destino para cada uma das entradas especificadas. A sensibilidade é medida pela correlação do destino

com cada entrada. O título do gráfico contém o nome do campo de destino. Cada barra no gráfico representa a correlação entre o campo de destino e um campo de entrada. As entradas simuladas que são incluídas no gráfico são as entradas que são selecionadas no campo **Campos de entrada para tornado de correlação** na visualização Campos da caixa de diálogo do nó Avaliação de Simulação. Cada barra é rotulada com o valor da correlação. As barras são ordenadas pelo valor absoluto das correlações, do maior para o menor valor. Se houver iterações, um gráfico separado é gerado para cada iteração. Cada gráfico possui um subtítulo, que contém o nome da iteração.

Tabela de distribuição. Essa tabela contém o valor do campo de destino, abaixo do qual a porcentagem especificada das observações cai. A tabela contém uma linha para cada valor de percentil que é especificado na visualização Saídas da caixa de diálogo do nó Avaliação de Simulação. Os valores de percentil podem ser quartis, um número diferente de percentis igualmente espaçados ou percentis especificados individualmente. A tabela de distribuição contém uma coluna para cada iteração.

Informações. Essa seção fornece uma sumarização geral dos campos e registros que são utilizados na avaliação. Ela também mostra os campos de entrada e as contagens de registro, que são divididos para cada iteração.

Opções de gráfico

Na caixa de diálogo Opções de Gráfico, é possível customizar a exibição dos gráficos de funções de densidade de probabilidade e de funções de distribuição acumulativa ativados que são gerados a partir de uma simulação.

Visualização. A lista suspensa **Visualização** se aplica apenas ao gráfico de função de densidade de probabilidade. É possível usá-la para alternar a visualização do gráfico de uma curva contínua para um histograma. Esse recurso é desativado quando diversas funções de densidade (a partir de várias iterações) são exibidas no mesmo gráfico. Quando houver diversas funções de densidade, as funções densidade podem ser visualizadas apenas como curvas contínuas.

Ordem. A lista suspensa **Ordem** se aplica apenas ao gráfico de função de distribuição acumulativa. Ela especifica se a função de distribuição acumulativa é exibida como uma função ascendente (o padrão) ou uma função decrescente. Quando exibido como uma função decrescente, o valor da função em um determinado ponto no eixo horizontal é a probabilidade de que o campo de destino estar à direita desse ponto.

Posições da régua de controle. O campo de texto **Superior** contém a posição atual da linha de referência da régua de controle direita. O campo de texto **Inferior** contém a posição atual da linha de referência da régua de controle esquerda. É possível configurar explicitamente as posições das régua de controle ao inserir valores nos campos de texto **Superior** e **Inferior**. O valor no campo de texto **Inferior** deve ser estritamente menor que o valor no campo de texto **Superior**. É possível remover a linha de referência à esquerda ao selecionar **-Infinito**, configurando efetivamente a posição para infinito negativo. Essa ação desativa o campo de texto **Inferior**. É possível remover a linha de referência à direita ao selecionar **Infinito**, configurando efetivamente sua posição para o infinito. Essa ação desativa o campo de texto **Superior**. Não é possível remover as duas linhas de referência; selecionar **-Infinito** desativa a caixa de seleção **Infinito**, e vice-versa.

Linhas de referência. É possível incluir várias linhas de referência verticais fixas nas funções de densidade de probabilidade e funções de distribuição acumulativa.

- **Média.** É possível incluir uma linha de referência na média do campo de destino.
- **Mediana.** É possível incluir uma linha de referência na mediana do campo de destino.
- **Desvios Padrão.** É possível incluir linhas de referência a mais e a menos com relação a um número especificado de desvios padrão da média do campo de destino. É possível inserir o número de desvios padrão a serem utilizados no campo de texto adjacente. O número mínimo de desvios padrão é 1 e o máximo é 10. O número de desvios padrão é inicialmente configurado para 3.

- **Percentis.** É possível incluir linhas de referência em um ou dois valores de percentis da distribuição para o campo de destino ao inserir valores nos campos de texto **Inferior** e **Superior**. Por exemplo, um valor de 95 no campo de texto **Superior** representa o 95º percentil, que é o valor abaixo do qual 95% das observações caem. Da mesma forma, um valor de 5 no campo de texto **Inferior** representa o décimo percentil, que é o valor abaixo do qual 5% das observações caem. Para o campo de texto **Inferior**, o valor mínimo de percentil é 0 e o máximo é 49. Para o campo de texto **Superior**, o valor de percentil mínimo é 50 e o máximo é 100.
- **Posições customizadas.** É possível incluir linhas de referência com valores especificados ao longo do eixo horizontal. É possível remover linhas de referência customizadas ao excluir a entrada da grade.

Ao clicar em **OK**, as réguas de controle, os rótulos acima das réguas de controle, as linhas de referência e a tabela abaixo do gráfico são atualizados para refletir as opções que estão selecionadas na caixa de diálogo Opções de Gráfico. Clique em **Cancelar** para fechar a caixa de diálogo sem fazer nenhuma mudança. As linhas de referência podem ser removidas ao cancelar a seleção da opção associada no diálogo Opções de Gráfico e clicar em **OK**.

Nota: Quando diversas funções de densidade ou funções de distribuição são exibidas em um gráfico único (devido aos resultados das iterações de análise de sensibilidade), as linhas de referência (diferentes de linhas customizadas) são aplicadas separadamente à cada função. Apenas as linhas de referência para a primeira iteração são mostradas. Os rótulos da linha de referência incluem o rótulo de iteração. O rótulo de iteração é derivado do envio de dados, geralmente a partir de um nó de Geração de Simulação. Se nenhum rótulo de iteração estiver disponível, o valor da iteração será usado. As opções **Média**, **Mediana**, **Desvios Padrão** e **Percentis** são desativadas para funções de distribuição acumulativa com diversas iterações.

Aplicativos Auxiliares do IBM SPSS Statistics

Se uma versão compatível do IBM SPSS Statistics estiver instalada e licenciada em seu computador, será possível configurar o IBM SPSS Modeler para processar dados com a funcionalidade do IBM SPSS Statistics utilizando os nós Transformação do Statistics, Modelo do Statistics, Saída do Statistics ou Exportação do Statistics.

Para obter informações sobre a compatibilidade do produto com a versão atual do IBM SPSS Modeler, consulte o site de Suporte Corporativo em <http://www.ibm.com/support>.

Para configurar o IBM SPSS Modeler para trabalhar com o IBM SPSS Statistics e outros aplicativos, escolha:

Ferramentas > Opções > Aplicativos Auxiliares

IBM SPSS Statistics Interativo. Insira o caminho e o nome completos do comando (por exemplo, *C:\Program Files\IBM\SPSS\Statistics\<nn>\stats.exe*) a ser utilizado ao ativar o IBM SPSS Statistics diretamente em um arquivo de dados produzido pelo nó Exportação de Estatísticas. Consulte o tópico “Nó Exportação de Estatísticas” na página 377 para obter mais informações.

Conexão. Se o IBM SPSS Statistics Server estiver localizado no mesmo host que o IBM SPSS Modeler Server, será possível ativar uma conexão entre os dois aplicativos, que aumenta a eficiência e deixa os dados no servidor durante a análise. Selecione **Servidor** para ativar a opção **Porta** abaixo. A configuração padrão é **Local**.

Porta. Especifique a porta do servidor para o IBM SPSS Statistics Server.

Utilitário de Localização do IBM SPSS Statistics. Para permitir que o IBM SPSS Modeler use os nós Transformação de Estatísticas, Modelo de Estatísticas e Saída de Estatísticas, você deverá ter uma cópia do IBM SPSS Statistics instalada e licenciada no computador no qual o fluxo é executado.

- Se estiver executando o IBM SPSS Modeler no modo local (independente), a cópia licenciada do IBM SPSS Statistics deverá estar no computador local. Clique neste botão para especificar a localização da instalação do local do IBM SPSS Statistics que deseja utilizar para o licenciamento.
- Além disso, se executar no modo distribuído com relação a um IBM SPSS Modeler Server remoto, também é necessário executar um utilitário no host do IBM SPSS Modeler Server para criar o arquivo *statistics.ini*, que indicará para o IBM SPSS Statistics o caminho de instalação do IBM SPSS Modeler Server. Para fazer isso, no prompt de comandos, mude para o diretório IBM SPSS Modeler Server *bin* e, para Windows, execute:

```
statisticsutility -location=<IBM SPSS Statistics_installation_path>/
```

Como alternativa, para UNIX, execute:

```
./statisticsutility -location=<IBM SPSS Statistics_installation_path>/bin
```

Se você não tiver uma cópia licenciada do IBM SPSS Statistics em sua máquina local, ainda será possível executar o nó Arquivo de Estatísticas com relação a um servidor IBM SPSS Statistics, porém as tentativas de executar outros nós do IBM SPSS Statistics exibirão uma mensagem de erro.

Comentários

Se você tiver problema ao executar os nós de procedimento do IBM SPSS Statistics, considere as seguintes dicas:

- Se nomes de campo usados no IBM SPSS Modeler tiverem mais de oito caracteres (para versões anteriores do IBM SPSS Statistics 12.0), mais de 64 caracteres (para o IBM SPSS Statistics 12.0 e versões subsequentes) ou contiver caracteres inválidos, será necessário renomeá-los ou truncá-los antes de sua leitura no IBM SPSS Statistics. Consulte o tópico “Renomeando ou Filtrando Campos para o IBM SPSS Statistics” na página 378 para obter mais informações.
- Se o IBM SPSS Statistics foi instalado após o IBM SPSS Modeler, poderá ser necessário especificar a localização do IBM SPSS Statistics, conforme explicado acima.

Capítulo 7. Exportar Nós

Visão Geral dos Nós de Exportação

Os nós de exportação fornecem um mecanismo para exportar dados em vários formatos para estabelecer interface com as outras ferramentas de software.

Os nós de exportação a seguir estão disponíveis:



O nó Exportação de Banco de Dados grava dados em uma origem de dados relacionais compatível com ODBC. Para gravar em uma origem de dados ODBC, a origem de dados deverá existir e você deverá ter permissão de gravação para ela.



O nó Flat File Export gera dados para um arquivo de texto delimitado. Ele é útil para exportar dados que podem ser lidos por outro software de análise ou de planilha.



O nó Exportação de Estatísticas gera dados no formato IBM SPSS Statistics *.sav* ou *.zsav*. Os arquivos *.sav* ou *.zsav* podem ser lidos pelo IBM SPSS Statistics Base e por outros produtos. Este também é o formato utilizado para arquivos em cache no IBM SPSS Modeler.



O nó de exportação do IBM SPSS Data Collection gera dados no formato utilizado pelo software de pesquisa de mercado do IBM SPSS Data Collection. O IBM SPSS Data Collection Data Library deve ser instalado para usar este nó.



O nó de exportação do IBM Cognos BI exporta dados em um formato que pode ser lido por bancos de dados do Cognos BI.



O nó de exportação do IBM Cognos TM1 exporta dados em um formato que pode ser lido por bancos de dados do Cognos TM1.



O nó de exportação SAS gera dados em formato do SAS a serem lidos no SAS ou em um pacote de software compatível com o SAS. Três os formatos de arquivo SAS estão disponíveis: SAS para Windows/OS2, SAS para UNIX ou SAS Versão 7/8.



O nó de exportação do Excel gera dados no formato de arquivo *.xlsx* do Microsoft Excel. Opcionalmente, você pode optar por ativar o Excel automaticamente e abrir o arquivo exportado quando o nó for executado.



O nó de exportação XML gera dados para um arquivo no formato XML. Opcionalmente, é possível criar um nó de origem XML para ler os dados exportados de volta no fluxo.

Nó Exportação do Banco de Dados

É possível utilizar os nós Banco de Dados para gravar dados em origens de dados relacionais compatíveis com o ODBC, que são explicados na descrição do nó de origem do Banco de Dados. Consulte o tópico “Nó de Origem da Base de Dados” na página 17 para obter mais informações.

Utilize os passos gerais a seguir para gravar dados em um banco de dados:

1. Instale um driver ODBC e configure uma origem de dados para o banco de dados que deseja utilizar.
2. Na guia Exportar do nó Banco de Dados, especifique a origem de dados e a tabela na qual deseja gravar. É possível criar uma nova tabela ou inserir dados em uma existente.
3. Especifique opções adicionais conforme necessário.

Esses passos são descritos em mais detalhes nos vários tópicos a seguir.

Guia Exportação do Nó do Banco de Dados

Nota: Alguns dos bancos de dados para os quais é possível exportar podem não suportar nomes de colunas em tabelas que tiverem mais de 30 caracteres de comprimento. Se uma mensagem de erro for exibida, que menciona que a sua tabela possui um nome de coluna incorreto, reduza o tamanho do nome para menos de 30 caracteres.

Origem de dados. Mostra a origem de dados selecionada. Insira o nome ou selecione-o na lista suspensa. Se o banco de dados desejado não for exibido na lista, selecione **Incluir nova conexão com o banco de dados** e localize seu banco de dados a partir da caixa de diálogo Conexões com o Banco de Dados. Consulte o tópico “Incluindo uma conexão com o banco de dados” na página 19 para obter mais informações.

Nome da tabela. Insira o nome da tabela para a qual deseja enviar os dados. Se selecionar a opção **Inserir na tabela**, será possível selecionar uma tabela existente no banco de dados clicando no botão **Selecionar**.

Criar tabela. Selecione esta opção para criar uma nova tabela de banco de dados ou para sobrescrever uma tabela de banco de dados existente.

Inserir na tabela. Selecione esta opção para inserir os dados como novas linhas em uma tabela de base de dados existente.

Mesclar tabela. (Quando disponível) Selecione essa opção para atualizar colunas de banco de dados selecionadas com valores de campos de dados de origem correspondentes. Selecionar essa opção ativa o botão **Mesclar**, que exibe um diálogo a partir do qual é possível mapear os campos de dados de origem para colunas de banco de dados.

Descartar tabela existente. Selecione esta opção para excluir qualquer tabela existente com o mesmo nome ao criar uma nova tabela.

Excluir linhas existentes. Selecione esta opção para excluir linhas existentes da tabela antes de exportar, ao inserir em uma tabela.

Nota: se qualquer uma das duas opções acima forem selecionadas, você receberá uma mensagem **Aviso de sobrescrição** quando executar o nó. Para suprimir os avisos, cancele a seleção de **Avisar quando um nó sobrescreve uma tabela de base de dados** na guia Notificações da caixa de diálogo Opções do Usuário.

Tamanho da sequência de caracteres padrão. Os campos que forem marcados como sem tipo em um nó Tipo de envio de dados são gravados no banco de dados como campos de sequência de caracteres. Especifique o tamanho das sequências de caracteres a serem utilizadas para campos sem tipo.

Clique em **Esquema** para abrir uma caixa de diálogo na qual é possível configurar várias opções de exportação (para bancos de dados que suportam esse recurso), configure tipos de dados SQL para seus campos e especifique a chave primária para propósitos de indexação do banco de dados. Consulte o tópico “Opções do Esquema de Exportação de Banco de Dados” na página 346 para obter mais informações.

Clique em **Índices** para especificar opções para indexar a tabela exportada a fim de melhorar o desempenho do banco de dados. Consulte o tópico “Opções do Índice de Exportação do Banco de Dados” na página 349 para obter mais informações.

Clique em **Avançado** para especificar as opções de carregamento em massa e de confirmação do banco de dados. Consulte o tópico “Opções Avançadas de Exportação de Banco de Dados” na página 350 para obter mais informações.

Colocar nomes de tabelas e colunas entre aspas. Selecione opções utilizadas ao enviar uma instrução CREATE TABLE para o banco de dados. As tabelas ou colunas com espaços ou caracteres não padrão devem ser colocadas entre aspas.

- **Conforme necessário.** Selecione para permitir que o IBM SPSS Modeler determine automaticamente quando aspas são necessárias em uma base individual.
- **Sempre.** Selecione para sempre colocar nomes de tabelas e de colunas entre aspas.
- **Nunca.** Selecione para desativar o uso de aspas.

Gerar um nó de importação para esses dados. Selecione para gerar um nó de origem da Base de Dados para os dados, conforme exportados para a origem e tabela de dados especificadas. Após a execução, esse nó é incluído na tela de fluxo.

Opções de Mesclagem de Exportação de Banco de Dados

Este diálogo permite mapear campos a partir dos dados de origem para colunas na tabela de banco de dados de destino. Quando um campo de dados origem é mapeado para uma coluna de banco de dados, o valor da coluna é substituído pelo valor dos dados de origem quando o fluxo é executado. Campos de origem não mapeados são deixados inalterados no banco de dados.

Mapear Campos. Aqui é onde você especifica o mapeamento entre os campos de dados de origem e as colunas de banco de dados. Os campos de dados de origem com o mesmo nome que as colunas no banco de dados são mapeados automaticamente.

- **Mapear.** Mapeia um campo de dados de origem selecionado na lista do campo à esquerda do botão para uma coluna de banco de dados selecionada na lista à direita. É possível mapear mais de um campo por vez, mas o número de entradas selecionadas nas duas listas deve ser igual.
- **Remover mapeamento.** Remove o mapeamento de uma ou mais colunas de banco de dados selecionadas. Este botão é ativado quando selecionar um campo ou uma coluna de banco de dados na tabela à direita do diálogo.
- **Incluir.** Inclui um ou mais campos de dados de origem selecionados na lista de campos à esquerda do botão para a lista à direita pronta para o mapeamento. Este botão é ativado quando selecionar um campo na lista à esquerda e nenhum campo com esse nome existir na lista à direita. Clicar nesse botão

mapeia o campo selecionado para uma nova coluna de banco de dados com o mesmo nome. A palavra <NEW> é exibida após o nome da coluna do banco de dados para indicar que esse é um novo campo.

Mesclar Linhas. É possível utilizar um campo de chave, como *ID de Transação*, para mesclar os registros com o mesmo valor no campo-chave. Esse é equivalente a uma "junção equivalente" de um banco de dados. Os valores de chave devem ser aqueles de chaves primárias, ou seja, devem ser exclusivos e não podem conter valores nulos.

- **Chaves possíveis.** Lista todos os campos localizados em todas as origens de dados de entrada. Selecione um ou mais campos nessa lista e utilize o botão de seta para incluí-los como campos-chave para mesclar registros. Qualquer campo de mapa com uma coluna de banco de dados mapeada correspondente está disponível como uma chave, com a exceção de que os campos incluídos como novas colunas de banco de dados (mostrados com <NEW> após o nome) não estão disponíveis.
- **Chaves para mesclagem.** Lista todos os campos utilizados para mesclar registros de todas as origens de dados de entrada com base nos valores dos campos-chave. Para remover uma chave da lista, selecione uma e utilize o botão de seta para retorná-la para a lista de Chaves Possíveis. Quando mais de um campo-chave é selecionado, a opção abaixo é ativada.
- **Incluir somente os registros que existirem no banco de dados.** Executa uma junção parcial; se o registro estiver no banco de dados e no fluxo, os campos mapeados serão atualizados.
- **Incluir registros no banco de dados.** Executa uma junção externa; todos os registros no fluxo serão mesclados (se o mesmo registro existir no banco de dados) ou incluídos (se o registro ainda não existir no banco de dados).

Para mapear um campo de dados de origem para uma nova coluna de banco de dados

1. Clique no nome do campo de origem na lista à esquerda, em **Mapear Campos**.
2. Clique no botão **Incluir** para concluir o mapeamento.

Para mapear um campo de dados de origem para uma coluna de banco de dados existente

1. Clique no nome do campo de origem na lista à esquerda, em **Mapear Campos**.
2. Clique no nome da coluna em **Coluna de Banco de Dados** à direita.
3. Clique no botão **Mapear** para concluir o mapeamento.

Para remover um mapeamento

1. Na lista à direita, no Campo, clique no nome do campo para o qual deseja remover o mapeamento.
2. Clique no botão **Remover mapeamento**.

Para cancelar a seleção de um campo em qualquer uma das listas

Mantenha pressionada a tecla CTRL e clique no nome do campo.

Opções do Esquema de Exportação de Banco de Dados

Na caixa de diálogo Esquema de exportação de banco de dados, é possível configurar opções para exportação do banco de dados (para bancos de dados que suportam essas opções), configurar tipos de dados SQL para seus campos, especificar quais campos são chaves primárias e customizar a instrução CREATE TABLE gerada na exportação.

A caixa de diálogo possui várias partes:

- A seção na parte superior (se exibida) contém opções para exportar para um banco de dados que suporta essas opções. Esta seção não será exibida se você não estiver conectado a tal banco de dados.
- O campo de texto no centro exibe o modelo utilizado para gerar o comando CREATE TABLE, que, por padrão, segue o formato:

```
CREATE TABLE <table-name> <(table columns)>
```

- A tabela na parte inferior permite especificar o tipo de dados SQL para cada campo e também indicar quais campos são as chaves primárias, conforme descrito abaixo. A caixa de diálogo gera automaticamente os valores dos parâmetros <table-name> e <(table columns)> com base nas especificações na tabela.

Configurando opções de exportação do banco de dados

Se essa seção for exibida, será possível especificar um número de definições para exportação no banco de dados. Os tipos de bancos de dados que suportam esse recurso são os seguintes.

- IBM InfoSphere Warehouse. Consulte o tópico “Opções para o IBM DB2 InfoSphere Warehouse” na página 348 para obter mais informações.
- Edições do SQL Server Enterprise e Developer. Consulte o tópico “Opções para o SQL server” na página 348 para obter mais informações.
- Edições do Oracle Enterprise ou Personal. Consulte o tópico “Opções para Oracle” na página 348 para obter mais informações.

Customizando instruções CREATE TABLE

Utilizando a parte do campo de texto desta caixa de diálogo, é possível incluir opções extras específicas do banco de dados na instrução CREATE TABLE.

1. Marque a caixa de seleção **Customizar o comando CREATE TABLE** para ativar a janela de texto.
2. Inclua quaisquer opções específicas do banco de dados na instrução. Assegure-se de reter os parâmetros de texto <table-name> e <(table-columns)> porque eles são substituídos pelo nome da tabela e por definições de coluna reais pelo IBM SPSS Modeler.

Configurando tipos de dados SQL

Por padrão, o IBM SPSS Modeler permite que o servidor de base de dados designe os tipos de dados SQL automaticamente. Para substituir o tipo automático para um campo, localize a linha correspondente ao campo e selecione o tipo desejado na lista suspensa na coluna *Tipo* da tabela do esquema. É possível utilizar Shift-clique para selecionar mais de uma linha.

Para tipos que utilizam um argumento de comprimento, de precisão ou de escala (BINARY, VARBINARY, CHAR, VARCHAR, NUMERIC e NUMBER), deve-se especificar um comprimento ao invés de permitir que o servidor de base de dados designe um comprimento automático. Por exemplo, especificar um valor sensível, como VARCHAR(25), para comprimento assegura que o tipo de armazenamento no IBM SPSS Modeler seja sobrescrito, caso seja sua intenção. Para substituir a designação automática, selecione **Especificar** na lista suspensa Tipo e substitua a definição de tipo pela instrução de definição de tipo SQL desejada.

A maneira mais fácil de fazer isso é primeiramente selecionar o tipo que estiver mais próximo da definição de tipo desejada e, em seguida, selecionar **Especificar** para editar essa definição. Por exemplo, para configurar o tipo de dados SQL para VARCHAR(25), primeiro configure o tipo para **VARCHAR(lenght)** a partir da lista suspensa Tipo e, em seguida, selecione **Especificar** e substitua o comprimento do texto pelo valor 25.

Chaves principais

Se uma ou mais colunas na tabela exportada tiver que ter um valor exclusivo ou uma combinação de valores para cada linha, será possível indicar isso ao marcar a caixa de seleção **Chave Primária** para cada campo que se aplica. A maioria dos bancos de dados não permite que a tabela seja modificada de modo que invalide uma restrição de chave primária e criará automaticamente um índice sobre a chave primária para ajudar a aplicar essa restrição. (Opcionalmente, é possível criar índices para outros campos na caixa de diálogo Índices. Consulte o tópico “Opções do Índice de Exportação do Banco de Dados” na página 349 para obter mais informações.)

Opções para o IBM DB2 InfoSphere Warehouse

Espaço de tabela. O espaço de tabela a ser utilizado para exportação. Os administradores de banco de dados podem criar ou configurar espaços de tabela como particionados. É recomendado selecionar um desses espaços de tabela (ao invés de um padrão) a ser usado para exportação do banco de dados.

Particionar os dados por campo. Especifica o campo de entrada a ser utilizado para particionamento.

Usar compactação. Se selecionada, cria tabelas para exportação com compactação (por exemplo, o equivalente a `CREATE TABLE MYTABLE(...) COMPRESS YES;` na SQL).

Opções para o SQL server

Usar compactação. Se selecionada, cria tabelas para exportação com compactação.

Compactação para. Escolha o nível de compactação.

- **Linha.** Ativa a compactação de nível de linha (por exemplo, o equivalente a `CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW);` em SQL).
- **Página.** Ativa a compactação de nível de página (por exemplo, `CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE);` em SQL).

Opções para Oracle

Configurações do Oracle - opção Básica

Usar compactação. Se selecionada, cria tabelas para exportação com compactação.

Compactação para. Escolha o nível de compactação.

- **Padrão.** Ativa a compactação padrão (por exemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` na SQL). Nesse caso, ela possui o mesmo efeito que a opção **Básico**.
- **Básico.** Ativa a compactação básica (por exemplo, `CREATE TABLE MYTABLE(...) COMPRESS BASIC;` na SQL).

Configurações do Oracle - opção Avançada

Usar compactação. Se selecionada, cria tabelas para exportação com compactação.

Compactação para. Escolha o nível de compactação.

- **Padrão.** Ativa a compactação padrão (por exemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` na SQL). Nesse caso, ela possui o mesmo efeito que a opção **Básico**.
- **Básico.** Ativa a compactação básica (por exemplo, `CREATE TABLE MYTABLE(...) COMPRESS BASIC;` na SQL).
- **OLTP.** Ativa a compactação de OLTP (por exemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR OLTP` na SQL).
- **Query Baixo/Alto.** (servidores Exadata somente) Ativa compactação columnar híbrida para query (por exemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY LOW;` ou `CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY HIGH;` na SQL). A compactação para a query é útil em ambientes de data warehousing; HIGH fornece uma razão de compactação maior que LOW.
- **Arquivar Baixo/Alto.** (servidores Exadata somente) Ativa compactação columnar híbrida para archive (por exemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE LOW;` ou `CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE HIGH;` na SQL). A compactação para a archive é útil para compactar dados que serão armazenados por longos períodos de tempo; HIGH fornece uma razão de compactação maior que LOW.

Opções do Índice de Exportação do Banco de Dados

A caixa de diálogo Índices permite criar índices em tabelas de banco de dados exportadas a partir do IBM SPSS Modeler. É possível especificar os conjuntos de campos que deseja incluir e customizar o comando CREATE INDEX, conforme necessário.

A caixa de diálogo tem duas partes:

- O campo de texto na parte superior exibe um modelo que pode ser utilizado para gerar um ou mais comandos CREATE INDEX que, por padrão, seguem o formato:

```
CREATE INDEX <index-name> ON <table-name>
```

- A tabela na parte inferior da caixa de diálogo permite incluir especificações para cada índice que deseja criar. Para cada índice, especifique o nome do índice e os campos ou colunas para incluir. A caixa de diálogo gera automaticamente os valores dos parâmetros <index-name> e <table-name> de acordo.

Por exemplo, a SQL gerada para um único índice nos campos *empid* e *deptid* pode ser semelhante a esta:

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID)
```

É possível incluir diversas linhas para criar diversos índices. Um comando CREATE INDEX separado é gerado para cada linha.

Customizando o comando CREATE INDEX

Opcionalmente, é possível customizar o comando CREATE INDEX para todos os índices ou somente de um índice específico. Isso fornece a flexibilidade para acomodar requisitos ou opções de banco de dados específicos e de aplicar customizações a todos os índices ou apenas aqueles específicos, conforme necessário.

- Selecione **Customizar o comando CREATE INDEX** na parte superior da caixa de diálogo para modificar o modelo utilizado para todos os índices incluídos subsequentemente. Observe que as mudanças não serão aplicadas automaticamente aos índices que já tiverem sido incluídos na tabela.
- Selecione uma ou mais linhas na tabela e, em seguida, clique em **Atualizar índices selecionados** na parte superior da caixa de diálogo para aplicar as customizações atuais a todas as linhas selecionadas.
- Marque a caixa de seleção **Customizar** em cada linha para modificar o modelo de comando somente para esse índice.

Observe que os valores dos parâmetros <index-name> e <table-name> são gerados automaticamente pela caixa de diálogo com base nas especificações da tabela e não podem ser editados diretamente.

BITMAP KEYWORD. Se estiver utilizando um banco de dados Oracle, será possível customizar o modelo para criar um índice de bitmap ao invés de um índice padrão, conforme a seguir:

```
CREATE BITMAP INDEX <index-name> ON <table-name>
```

As indexações de bitmap podem ser úteis para indexar colunas com um pequeno número de valores distintos. A SQL resultante pode ser semelhante a esta:

```
CREATE BITMAP INDEX MYTABLE_IDX1 ON MYTABLE(COLOR)
```

Palavra-chave UNIQUE. A maioria dos bancos de dados suporta a palavra-chave UNIQUE no comando CREATE INDEX. Isso impõe uma restrição de exclusividade semelhante a uma restrição de chave primária na tabela subjacente.

```
CREATE UNIQUE INDEX <index-name> ON <table-name>
```

Observe que essa especificação não é necessária para campos realmente designados como chaves primárias. A maioria dos bancos de dados criará automaticamente um índice para quaisquer campos especificados como campos de chave primária no comando CREATE TABLE, portanto, não é necessário criar explicitamente os índices nesses campos. Consulte o tópico “Opções do Esquema de Exportação de Banco de Dados” na página 346 para obter mais informações.

Palavra-chave FILLFACTOR. Alguns parâmetros físicos para o índice podem ter um ajuste preciso. Por exemplo, o SQL Server permite que o usuário obtenha um trade-off do tamanho do índice (após a criação inicial) com relação aos custos de manutenção conforme mudanças futuras forem feitas na tabela.

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID) WITH FILLFACTOR=20
```

Outros Comentários

- Se um índice já existir com o nome especificado, a criação do índice falhará. Todas as falhas serão inicialmente tratadas como avisos, permitindo que índices subsequentes sejam criados e, em seguida, relatados novamente como um erro no log de mensagens após todos os índices terem sido tentados.
- Para obter melhor desempenho, os índices devem ser criados após os dados terem sido carregados na tabela. Os índices devem conter pelo menos uma coluna.
- Antes de executar o nó, é possível visualizar a SQL gerada no log de mensagens.
- Para tabelas temporárias gravadas no banco de dados (ou seja, quando o armazenamento em cache do nó estiver ativado) as opções para especificar chaves primárias e índices não estão disponíveis. No entanto, o sistema pode criar índices na tabela temporária conforme apropriado, dependendo de como os dados são utilizados nos nós de recebimento de dados. Por exemplo, se os dados em cache forem subsequentemente unidos por uma coluna *DEPT*, faria sentido indexar a tabela em cache nesta coluna.

Otimização de Índices e da Consulta

Em alguns sistemas de gerenciamento de banco de dados, quando uma tabela de banco de dados tiver sido criada, carregada e indexada, um passo adicional será necessário antes que o otimizador possa utilizar os índices para acelerar a execução da consulta na nova tabela. Por exemplo, no Oracle, o otimizador de consulta baseado em custo requer que uma tabela seja analisada antes que seus índices possam ser utilizados na otimização da consulta. O arquivo de propriedades ODBC interno para Oracle (não visível pelo usuário) contém uma opção para que isso aconteça, conforme a seguir:

```
# Defines SQL to be executed after a table and any associated indexes  
# have been created and populated  
table_analysis_sql, 'ANALYZE TABLE <table-name> COMPUTE STATISTICS'
```

Este passo é executado sempre que uma tabela for criada no Oracle (independentemente se as chaves primárias ou os índices estiverem definidos). Se necessário, o arquivo de propriedades ODBC para bancos de dados adicionais pode ser customizado de maneira semelhante – entre em contato com o Suporte para obter assistência.

Opções Avançadas de Exportação de Banco de Dados

Ao clicar no botão Avançado na caixa de diálogo do nó de exportação de Banco de Dados, uma nova caixa de diálogo é exibida na qual é possível especificar detalhes técnicos para exportar resultados para um banco de dados.

Usar confirmação de lote. Selecione para desativar confirmações linha por linha ao banco de dados.

Tamanho do lote. Especifica o número de registros a serem enviados para o banco de dados antes de confirmar na memória. Diminuir esse número fornece maior integridade de dados ao custo de velocidades de transferência mais lentas. Você pode querer fazer um ajuste preciso desse número para obter um desempenho ideal com seu banco de dados.

Opções do InfoSphere Warehouse. Exibidas apenas se estiver conectado a um banco de dados do InfoSphere Warehouse (IBM DB2 9.7 ou posterior). **Não registrar atualizações** permite evitar o registro de eventos ao criar tabelas e inserir dados.

Usar o carregamento em massa. Especifica um método para carregamento de dados em massa para o banco de dados diretamente do IBM SPSS Modeler. Algumas experiências podem ser necessárias para selecionar quais opções de carregamento em massa são apropriadas para um cenário específico.

- **Via ODBC.** Selecione para utilizar a API do ODBC para executar inserções de diversas linhas com maior eficiência do que a exportação normal para o banco de dados. Escolha a partir de ligação entre linhas ou entre colunas nas opções abaixo.
- **Via carregador externo.** Selecione para utilizar um programa de utilitário de carregamento em massa customizado específico para seu banco de dados. Selecionar essa opção ativa uma variedade de opções abaixo.

Opções Avançadas do ODBC. Essas opções estão disponíveis apenas quando **Via ODBC** é selecionada. Observe que essa funcionalidade pode não ser suportada por todos os drivers ODBC.

- **Entre linhas.** Selecione a ligação entre linhas para utilizar a chamada `SQLBulkOperations` para carregar dados no banco de dados. A ligação entre linhas geralmente melhora a velocidade em comparação com o uso de inserções parametrizadas que inserem dados basicamente registro por registro.
- **Entre colunas.** Selecione para utilizar ligação entre colunas para carregar dados no banco de dados. A ligação entre colunas melhora o desempenho ao ligar cada coluna de banco de dados (em uma instrução `INSERT` parametrizada) a uma matriz de N valores. Executar a instrução `INSERT` uma vez faz com que N linhas sejam inseridas no banco de dados. Este método pode aumentar significativamente o desempenho.

Opções do Carregador Externo. Quando **Via carregador externo** é especificada, diversas opções são exibidas para exportar o conjunto de dados em um arquivo e especificar e executar um programa carregador customizado para carregar os dados desse arquivo no banco de dados. O IBM SPSS Modeler pode estabelecer interface com carregadores externos para muitos sistemas de banco de dados populares. Vários scripts foram incluídos com o software e estão disponíveis junto de uma documentação técnica no subdiretório *scripts*. Observe que para utilizar esta funcionalidade, o Python 2.7 deve ser instalado na mesma máquina que o IBM SPSS Modeler ou o IBM SPSS Modeler Server, e o parâmetro `python_exe_path` deve ser configurado no arquivo *options.cfg*. Consulte o tópico “Programação do utilitário de carregamento em massa” na página 352 para obter mais informações.

- **Usar delimitador.** Especifica qual caractere delimitador deverá ser usado no arquivo exportado. Selecione **Tab** para delimitar com tabulação e **Espaço** para delimitar com espaços. Selecione **Outro** para especificar outro caractere, como uma vírgula (,).
- **Especificar arquivo de dados.** Selecionar para inserir o caminho a ser utilizado para o arquivo de dados gravado durante o carregamento em massa. Por padrão, um arquivo temporário é criado no diretório `temp` no servidor.
- **Especificar programa carregador.** Selecione para especificar um programa de carregamento em massa. Por padrão, o software procura no subdiretório *scripts* da instalação do IBM SPSS Modeler um script Python a ser executado para um banco de dados específico. Vários scripts foram incluídos com o software e estão disponíveis junto de uma documentação técnica no subdiretório *scripts*.
- **Gerar log.** Selecione para gerar um arquivo de log no diretório especificado. O arquivo de log contém informações de erro e será útil se a operação de carregamento em massa falhar.
- **Verificar tamanho da tabela.** Selecione para executar a verificação de tabela que assegura que o aumento do tamanho da tabela corresponda ao número de linhas exportadas a partir do IBM SPSS Modeler.
- **Opções extras do carregador.** Especifica os argumentos adicionais para o programa carregador. Utilize aspas duplas para argumentos que contêm espaços.

As aspas duplas são incluídas nos argumentos opcionais ao escapá-las com uma barra invertida. Por exemplo, a opção especificada como `-comment "This is a \"comment\""` inclui o flag `-comment` e o próprio comentário renderizado como `This is a "comment"`.

Uma barra invertida pode ser incluída ao escapar com outra barra invertida. Por exemplo, a opção especificada como `-specialdir "C:\\Test Scripts\\"` inclui o flag `-specialdir` e o diretório renderizado como `C:\Test Scripts\`.

Programação do utilitário de carregamento em massa

O nó de exportação do banco de dados inclui opções para carregamento em massa na caixa de diálogo Opções Avançadas. Os programas de utilitários de carregamento em massa podem ser usados para carregar dados de um arquivo de texto em um banco de dados.

A opção **Usar carregamento em massa - via carregador externo** configura o IBM SPSS Modeler a fazer três coisas:

- Criar qualquer tabela de banco de dados necessária.
- Exportar os dados para um arquivo de texto.
- Chamar um programa de utilitário de carregamento em massa para carregar os dados deste arquivo na tabela de banco de dados.

Normalmente, o programa de utilitário de carregamento em massa não é o utilitário de carregamento de banco de dados em si (por exemplo, utilitário `sqlldr` do Oracle), mas um pequeno script ou programa que forma os argumentos corretos, cria qualquer arquivo auxiliar específico do banco de dados (como um arquivo de controle) e, em seguida, chama o utilitário de carregamento de banco de dados. As informações nas seções a seguir ajudarão a editar um utilitário de carregamento em massa existente.

Como alternativa, é possível gravar seu próprio programa para carregamento em massa. Consulte o tópico “Desenvolvendo programas de utilitário de carregamento em massa” na página 356 para obter mais informações. Observe que isso não é coberto sob um contrato de Suporte Técnico padrão e deve-se entrar em contato com um representante de Serviços IBM para obter assistência.

Scripts para carregamento em massa

O IBM SPSS Modeler é fornecido com uma série de programas de utilitário de carregamento em massa para bancos de dados diferentes que são implementados utilizando os scripts Python. Ao executar um fluxo contendo um nó de exportação de banco de dados com a opção **Via carregador externo** selecionada, o IBM SPSS Modeler cria a tabela de banco de dados (se necessário) via ODBC, exporta os dados para um arquivo temporário no host que está executando o IBM SPSS Modeler Server e, em seguida, chama o script de carregamento em massa. Este script, por sua vez, executa utilitários fornecidos pelo fornecedor DBMS para fazer upload de dados dos arquivos temporários para o banco de dados.

Nota: A instalação do IBM SPSS Modeler não inclui um intérprete de tempo de execução Python, de modo que uma instalação separada do Python é necessária. Consulte o tópico “Opções Avançadas de Exportação de Banco de Dados” na página 350 para obter mais informações.

Os scripts são fornecidos (na pasta `\scripts` do diretório de instalação do IBM SPSS Modeler) para os bancos de dados listados na tabela a seguir.

Tabela 58. Scripts do utilitário de carregamento em massa.

Banco de Dados	Nome do script	Informações adicionais
IBM DB2	<code>db2_loader.py</code>	Consulte o tópico “Carregamento de dados em massa para bancos de dados do IBM DB2” na página 353 para obter mais informações.

Tabela 58. Scripts do utilitário de carregamento em massa (continuação).

Banco de Dados	Nome do script	Informações adicionais
IBM Netezza	<i>netezza_loader.py</i>	Consulte o tópico “Carregamento de dados em massa para bancos de dados do IBM Netezza” para obter mais informações.
Oracle	<i>oracle_loader.py</i>	Consulte o tópico “Carregamento de dados em massa para bancos de dados Oracle” na página 354 para obter mais informações.
Servidor SQL	<i>mssql_loader.py</i>	Consulte o tópico “Carregamento de dados em massa para bancos de dados SQL Server” na página 355 para obter mais informações.
Teradata	<i>teradata_loader.py</i>	Consulte o tópico “Carregamento de dados em massa para bancos de dados Teradata” na página 355 para obter mais informações.

Carregamento de dados em massa para bancos de dados do IBM DB2

Os pontos a seguir podem ajudá-lo a configurar um carregamento em massa do IBM SPSS Modeler para um banco de dados IBM DB2 usando a opção do Carregador Externo na caixa de diálogo Opções Avançadas de Exportação do BD.

Assegure-se de que o utilitário do processador de linha de comandos (CLP) do DB2 esteja instalado.

O script *db2_loader.py* chama o comando DB2 LOAD. Assegure-se de que o processador de linha de comandos (*db2* no UNIX, *db2cmd* no Windows) esteja instalado no servidor no qual o *db2_loader.py* deverá ser executado (normalmente, o host que executa o IBM SPSS Modeler Server).

Verifique se o nome do alias do banco de dados local é o mesmo que o nome do banco de dados real.

O alias do banco de dados DB2 local é o nome usado pelo software cliente do DB2 para referenciar um banco de dados em uma instância do DB2 local ou remota. Se o alias do banco de dados local for diferente do nome do banco de dados remoto, forneça a opção do carregador extra:

```
-alias <local_database_alias>
```

Por exemplo, o banco de dados remoto é denominado STARS no host GALAXY, mas o alias do banco de dados local do DB2 no host que executa o IBM SPSS Modeler Server é STARS_GALAXY. Use a opção do carregador extra

```
-alias STARS_GALAXY
```

Codificação de dados de caractere não ASCII

Se estiver carregando dados em massa que não estiverem no formato ASCII, assegure-se de que a variável de página de códigos na seção de configuração do *db2_loader.py* esteja configurada corretamente em seu sistema.

Sequências de caracteres em branco

As sequências de caracteres em branco são exportadas para o banco de dados como valores NULL.

Carregamento de dados em massa para bancos de dados do IBM Netezza

Os pontos a seguir podem ajudá-lo a configurar um carregamento em massa do IBM SPSS Modeler para um banco de dados IBM Netezza usando a opção do Carregador Externo na caixa de diálogo Opções Avançadas de Exportação do BD.

Assegure-se de que o utilitário *nzload* do Netezza esteja instalado

O script *netezza_loader.py* chama o utilitário *nzload* do Netezza. Assegure-se de que o *nzload* esteja instalado e configurado corretamente no servidor em que o *netezza_loader.py* deverá ser executado.

Exportando dados não ASCII

Se a sua exportação contiver dados que não estão em formato ASCII, poderá ser necessário incluir `-encoding UTF8` no campo **Opções extras do carregador** na caixa de diálogo Opções Avançadas de Exportação do DB. Isto deve assegurar que os dados não ASCII sejam corretamente transferidos por upload.

Dados em formato de Data, Hora e Registro de Data e Hora

Nas propriedades do fluxo, configure o formato de data para **DD-MM-AAAA** e o formato de hora para **HH:MM:SS**.

Sequências de caracteres em branco

As sequências de caracteres em branco são exportadas para o banco de dados como valores NULL.

Ordem diferente de colunas na tabela de fluxo e de destino ao inserir dados em uma tabela existente

Se a ordem de colunas no fluxo for diferente daquela da tabela de destino, os valores de dados serão inseridos nas colunas erradas. Utilize um nó Reordenação de Campo para assegurar que a ordem das colunas no fluxo corresponda à ordem na tabela de destino. Consulte o tópico “Nó Reordenação de Campo” na página 188 para obter mais informações.

Rastreamento progresso *nzload*

Ao executar o IBM SPSS Modeler no modo local, inclua `-sts` no campo **Opções extras do carregador** na caixa de diálogo Opções Avançadas de Exportação do BD para ver as mensagens de status a cada 10000 linhas na janela de comandos aberta pelo utilitário *nzload*.

Carregamento de dados em massa para bancos de dados Oracle

Os pontos a seguir poderão ajudar a configurar um carregamento em massa a partir do IBM SPSS Modeler para um banco de dados Oracle utilizando a opção Carregador Externo na caixa de diálogo Opções Avançadas de Exportação do BD

Assegure-se de que o utilitário *sqlldr* do Oracle esteja instalado

O script *oracle_loader.py* chama o utilitário *sqlldr* do Oracle. Observe que o *sqlldr* não é incluído automaticamente no Oracle Client. Assegure-se de que *sqlldr* esteja instalado no servidor em que o *oracle_loader.py* deverá ser executado.

Especifique o SID ou o nome do serviço do banco de dados

Se você estiver exportando dados para um servidor Oracle não local ou se o servidor Oracle local possuir diversos bancos de dados, será necessário especificar o seguinte no campo **Opções de carregador extras** na caixa de diálogo Opções Avançadas de Exportação do BD para transmitir o SID ou o nome do serviço:

```
-database <SID>
```

Editando a seção de configuração em *oracle_loader.py*

Em sistemas UNIX (e opcionalmente, no Windows), edite a seção de configuração no início do script *oracle_loader.py*. Aqui, os valores para as variáveis de ambiente `ORACLE_SID`, `NLS_LANG`, `TNS_ADMIN` e `ORACLE_HOME` podem ser especificados, se apropriado, bem como o caminho completo do utilitário *sqlldr*.

Dados em formato de Data, Hora e Registro de Data e Hora

Nas propriedades do fluxo, normalmente deve-se configurar o formato de data para **AAAA-MM-DD** e o formato de hora para **HH:MM:SS**.

Se precisar utilizar formatos de data e hora diferentes do especificado anteriormente, consulte sua documentação Oracle e edite o arquivo de script *oracle_loader.py*.

Codificação de dados de caractere não ASCII

Se você estiver carregando dados em massa que não estiverem no formato ASCII, assegure-se de que a variável de ambiente `NLS_LANG` esteja corretamente configurada em seu sistema. Este é lido pelo utilitário do carregador *sqlldr* do Oracle. Por exemplo, o valor correto para `NLS_LANG` para Shift-JIS no Windows é `Japanese_Japan.JA16SJIS`. Para obter mais detalhes sobre `NLS_LANG`, verifique sua documentação do Oracle.

Sequências de caracteres em branco

As sequências de caracteres em branco são exportadas para o banco de dados como valores NULL.

Carregamento de dados em massa para bancos de dados SQL Server

Os pontos a seguir poderão ajudar a configurar um carregamento em massa a partir do IBM SPSS Modeler para um banco de dados SQL Server utilizando a opção Carregador Externo na caixa de diálogo Opções Avançadas de Exportação do BD.

Assegure-se de que o utilitário *bcp.exe* do SQL Server esteja instalado

O script *mssql_loader.py* chama o utilitário *bcp.exe* do SQL Server. Assegure-se de que *bcp.exe* esteja instalado no servidor em que o *mssql_loader.py* deverá ser executado.

Utilizar espaços como delimitador não funciona

Evite escolher espaço como o delimitador na caixa de diálogo Opções Avançadas de Exportação do BD.

Verifique a opção de tamanho de tabela recomendada

Recomenda-se ativar a opção **Verificar tamanho de tabela** na caixa de diálogo Opções Avançadas de Exportação do BD. As falhas no processo de carregamento em massa nem sempre são detectadas, e a ativação desta opção executa uma verificação extra de que o número correto de linhas foi carregado.

Sequências de caracteres em branco

As sequências de caracteres em branco são exportadas para o banco de dados como valores NULL.

Especifique a instância nomeada completa do SQL server

Pode haver ocasiões em que o SPSS Modeler não pode acessar o SQL Server devido a um nome do host não completo e exibe o erro a seguir:

Erro encontrado durante a execução do carregador em massa externo. O arquivo de log pode fornecer mais detalhes.

Para corrigir esse erro, inclua a sequência de caracteres a seguir, incluindo as aspas duplas, no campo **Opções de carregador extras**:

```
"-S mhreboot.spss.com\SQLSERVER"
```

Carregamento de dados em massa para bancos de dados Teradata

Os pontos a seguir poderão ajudar a configurar um carregamento em massa a partir do IBM SPSS Modeler para um banco de dados Teradata utilizando a opção Carregador Externo na caixa de diálogo Opções Avançadas de Exportação do BD.

Assegure-se de que o utilitário *fastload* do Teradata esteja instalado

O script *teradata_loader.py* chama o utilitário *fastload* do Teradata. Assegure-se de que o *fastload* esteja instalado e configurado corretamente no servidor em que o *teradata_loader.py* deverá ser executado.

Os dados podem ser carregados em massa somente para tabelas vazias

Somente tabelas vazias podem ser utilizadas como destinos para um carregamento em massa. Se uma tabela de destino contiver quaisquer dados anteriores ao carregamento em massa, a operação falhará.

Dados em formato de Data, Hora e Registro de Data e Hora

Nas propriedades do fluxo, configure o formato de data para **AAAA-MM-DD** e o formato de hora para **HH:MM:SS**.

Sequências de caracteres em branco

As sequências de caracteres em branco são exportadas para o banco de dados como valores NULL.

ID do processo do Teradata (tdpid)

Por padrão, o *fastload* exporta dados para o sistema do Teradata com `tdpid=dbc`. Geralmente, haverá uma entrada no arquivo HOSTS que associa `dbccop1` ao endereço IP do servidor Teradata. Para utilizar um servidor diferente, especifique o seguinte no campo **Opções de carregador extras** na caixa de diálogo Opções Avançadas de Exportação do BD para transmitir o `tdpid` deste servidor:

```
-tdpid <id>
```

Espaços nos nomes de tabelas e colunas

Se nomes de tabela ou de colunas contiverem espaços, a operação de carregamento em massa falhará. Se possível, renomeie os nomes de tabela ou de coluna para remover espaços.

Desenvolvendo programas de utilitário de carregamento em massa

Este tópico explica como desenvolver um programa de utilitário de carregamento em massa que pode ser executado a partir do IBM SPSS Modeler para carregar dados de um arquivo de texto em um banco de dados. Observe que isso não é coberto sob um contrato de Suporte Técnico padrão e deve-se entrar em contato com um representante de Serviços IBM para obter assistência.

Utilizando Python para construir programas de utilitário de carregamento em massa

Por padrão, o IBM SPSS Modeler procura por um programa de utilitário de carregamento em massa padrão com base no tipo de banco de dados. Consulte Tabela 58 na página 352.

É possível utilizar o script *test_loader.py* para auxiliar no desenvolvimento de programas de utilitário de carregamento em lote. Consulte o tópico “Testando programas de utilitário de carregamento em massa” na página 358 para obter mais informações.

Objetos transmitidos para o programa de utilitário de carregamento em massa

O IBM SPSS Modeler grava dois arquivos que são transmitidos para o programa de utilitário de carregamento em massa.

- **Arquivo de dados.** Esse contém os dados a serem carregados, em formato de texto.
- **Arquivo de esquema.** Este é um arquivo XML que descreve os nomes e os tipos das colunas e fornece informações sobre como o arquivo de dados é formatado (por exemplo, qual caractere é utilizado como o delimitador entre campos).

Além disso, o IBM SPSS Modeler transmite outras informações, como nome da tabela, nome de usuário e senha como argumentos ao chamar o programa de carregamento em massa.

Nota: para sinalizar conclusão com sucesso no IBM SPSS Modeler, o programa de utilitário de carregamento em massa deve excluir o arquivo de esquema.

Argumentos transmitidos para o programa de utilitário de carregamento em massa

Os argumentos transmitidos para o programa são listados na tabela a seguir.

Tabela 59. Argumentos transmitidos para o utilitário de carregamento em massa.

Argumento	Descrição
schemafile	Caminho do arquivo de esquema.
data file	Caminho do arquivo de dados.
servername	Nome do servidor DBMS; pode estar em branco.
databasename	Nome do banco de dados no servidor DBMS; pode estar em branco.
username	Nome do usuário para efetuar login no banco de dados.
password	Senha para efetuar login no banco de dados.
tablename	Nome da tabela para carregar.
ownername	Nome do proprietário da tabela (também conhecido como nome de esquema).
logfile	Nome do arquivo de log (se estiver em branco, nenhum arquivo de log é criado).
rowcount	Número de linhas no conjunto de dados.

Quaisquer opções especificadas no campo **Opções extras do carregador** na caixa de diálogo Opções Avançadas de Exportação do BD são transmitidas para o programa de utilitário de carregamento em massa após esses argumentos padrão.

Formato do arquivo de dados

Os dados são gravados no arquivo de dados em formato de texto, em que cada campo é separado por um caractere delimitador que é especificado na caixa de diálogo Opções Avançadas de Exportação do BD. A seguir há um exemplo de como um arquivo de dados delimitado por guia pode aparecer.

```
48 F HIGH NORMAL 0.692623 0.055369 drugA
15 M NORMAL HIGH 0.678247 0.040851 drugY
37 M HIGH NORMAL 0.538192 0.069780 drugA
35 F HIGH HIGH 0.635680 0.068481 drugA
```

O arquivo é gravado na codificação local utilizada pelo IBM SPSS Modeler Server (ou pelo IBM SPSS Modeler se não estiver conectado ao IBM SPSS Modeler Server). Alguma formatação é controlada pelas configurações de fluxo do IBM SPSS Modeler.

Formato do arquivo de esquema

O arquivo de esquema é um arquivo XML que descreve o arquivo de dados. A seguir há um exemplo para acompanhar o arquivo de dados anterior.

```
<?xml version="1.0" encoding="UTF-8"?>
<DBSCHEMA version="1.0">
<table delimiter="\t" commit_every="10000" date_format="YYYY-MM-DD" time_format="HH:MM:SS"
append_existing="false" delete_datafile="false">
<column name="Age" encoded_name="416765" type="integer"/>
<column name="Sex" encoded_name="536578" type="char" size="1"/>
<column name="BP" encoded_name="4250" type="char" size="6"/>
<column name="Cholesterol" encoded_name="43686F6C65737465726F6C" type="char" size="6"/>
<column name="Na" encoded_name="4E61" type="real"/>
<column name="K" encoded_name="4B" type="real"/>
<column name="Drug" encoded_name="44727567" type="char" size="5"/>
</table>
</DBSCHEMA>
```

As duas tabelas a seguir listam os atributos dos elementos <table> e <column> do arquivo de esquema.

Tabela 60. Atributos do elemento <table>.

Atributo	Descrição
delimiter	O caractere delimitador de campo (TAB é representado como \t).
commit_every	O intervalo de tamanho do lote (como na caixa de diálogo Opções Avançadas de Exportação do BD).
date_format	O formato utilizado para representar datas.
time_format	O formato utilizado para representar tempos.
append_existing	true se a tabela a ser carregada já contiver dados; caso contrário, false.
delete_datafile	true se o programa de carregamento em massa tiver que excluir o arquivo de dados na conclusão do carregamento.

Tabela 61. Atributos do elemento <column>.

Atributo	Descrição
name	O nome da coluna.
encoded_name	O nome da coluna convertido na mesma codificação que o arquivo de dados e a saída como uma série de números hexadecimais de dois dígitos.
type	O tipo de dados da coluna: um de integer, true, char, time, date e datetime.
size	Para o tipo de dados char, a largura máxima da coluna em caracteres.

Testando programas de utilitário de carregamento em massa

É possível testar o carregamento em massa usando um script de teste *test_loader.py* incluído na pasta *\scripts* do diretório de instalação do IBM SPSS Modeler. Isso é útil ao tentar desenvolver, depurar ou solucionar problemas de programas ou scripts de carregamento em massa para uso com o IBM SPSS Modeler.

Para utilizar o script de teste, prossiga conforme a seguir.

1. Execute o script *test_loader.py* para copiar os arquivos de esquema e de dados para os arquivos *schema.xml* e *data.txt* e criar um arquivo em lote do Windows (*test.bat*).
2. Edite o arquivo *test.bat* para selecionar o programa ou script de utilitário de carregamento em massa para teste.
3. Execute *test.bat* a partir de um shell de comando para testar o programa ou script de carregamento em massa escolhido.

Nota: na realidade, a execução de *test.bat* não carrega dados no banco de dados.

Nó Exportação de Arquivo Simples

O nó de exportação Arquivo Simples permite gravar dados em um arquivo de texto delimitado. Ele é útil para exportar dados que podem ser lidos por outro software de análise ou de planilha.

Se seus dados contiverem informações geoespaciais, será possível exportá-los como um arquivo simples e, se um nó de origem Arquivo de Variável for gerado para uso dentro do mesmo fluxo, todo o armazenamento, medição e metadados geoespaciais serão retidos no novo nó de origem. No entanto, se você exportar os dados e, em seguida, importá-los em um fluxo diferente, deve-se executar alguns passos extras para configurar os metadados geoespaciais no novo nó de origem. Para obter mais informações, consulte o tópico “Nó Arquivo Variável” na página 26.

Nota: Não é possível gravar arquivos no formato de cache antigo, já que o IBM SPSS Modeler não mais utiliza esse formato para arquivos de cache. Os arquivos de cache do IBM SPSS Modeler agora são salvos

em formato *.sav* do IBM SPSS Statistics, que pode ser gravado utilizando um nó de exportação Statistics. Para obter mais informações, consulte o tópico “Nó Exportação de Estatísticas” na página 377.

Guia Exportação de Arquivo Simples

Exportar arquivo. Especifica o nome do arquivo. Insira um nome de arquivo ou clique no botão seletor de arquivo para navegar para a localização do arquivo.

Modo de gravação Se **Sobrescrever** for selecionado, quaisquer dados existentes no arquivo especificado serão sobrescritos. Se **Anexar** for selecionado, a saída será incluída no final do arquivo existente, preservando os dados que contém.

- **Incluir nomes de campo.** Se essa opção for selecionada, os nomes de campos serão gravados em uma nova linha no arquivo de saída. Essa opção está disponível apenas para o modo de gravação **Sobrescrever**.

Nova linha depois de cada registro. Se essa opção for selecionada, cada registro será gravado em uma nova linha no arquivo de saída.

Separador de campo. Especifica o caractere a ser inserido entre os valores do campo no arquivo de texto gerado. As opções são **Vírgula**, **Tabulação**, **Espaço** e **Outras**. Se você selecionar **Outras**, digite o caractere delimitador desejado na caixa de texto.

Aspas de símbolos Especifica o tipo de aspas a ser usado para valores de campos simbólicos. As opções são **Nenhum** (os valores não estão entre aspas), **Simples (')**, **Duplas (")** e **Outras**. Se você selecionar **Outras**, digite o caractere de aspas desejado na caixa de texto.

Codificação. Especifica o método de codificação de texto usado. É possível escolher entre o padrão do sistema, o padrão de fluxo ou UTF-8.

- O padrão do sistema é especificado no Painel de Controle do Windows ou, se estiver em execução no modo distribuído, no computador servidor.
- O padrão de fluxo é especificado na caixa de diálogo Propriedades do Fluxo.

Símbolo decimal. Especifica como os decimais devem ser representados nos dados.

- **Fluxo padrão.** O separador decimal definido pelo fluxo atual da configuração padrão será usado. Normalmente, este será o separador decimal definido pelas configurações do código do idioma do computador.
- **Ponto final (.)**. O caractere de ponto final será usado como o separador decimal.
- **Vírgula (,)**. O caractere vírgula será usado como o separador decimal.

Gerar um nó de importação para esses dados. Selecione essa opção para gerar automaticamente um nó de origem do Arquivo de Variável que lerá o arquivo de dados exportado. Consulte o tópico “Nó Arquivo Variável” na página 26 para obter mais informações.

Nó Exportação do IBM SPSS Data Collection

O nó de exportação do IBM SPSS Data Collection salva os dados no formato utilizado pelo software de pesquisa de mercado do IBM SPSS Data Collection com base no Modelo de Dados do IBM SPSS Data Collection. Este formato distingue os dados do caso - as respostas reais para as questões reunidas durante uma pesquisa de opinião - dos metadados que descrevem como os dados do caso são coletados e organizados. Os metadados consistem em informações como textos da questão, nomes e descrições de variáveis, conjuntos de múltiplas respostas, traduções dos vários textos e as definições da estrutura dos dados do caso. Consulte o tópico “Nó do Coleta de dados” na página 32 para obter mais informações.

Nota: este nó requer o IBM SPSS Data Collection Data Model versão 4.0 ou superior, que é distribuído com o software do IBM SPSS Data Collection. Para obter mais informações, consulte a página da web no

IBM SPSS Data Collection em <http://www.ibm.com/software/analytics/spss/products/data-collection/>. Além de instalar o Data Model, nenhuma configuração adicional será necessária.

Arquivo de metadados. Especifica o nome do arquivo de definição de questionário (.mdd) no qual os metadados exportados serão salvos. Um questionário padrão é criado com base nas informações de tipo de campo. Por exemplo, um campo nominal (conjunto) pode ser representado como uma questão única com a descrição de campo utilizada como o texto da questão e uma caixa de seleção separada para cada valor definido.

Mesclar metadados. Especifica se os metadados sobrescreverão as versões existentes ou se serão mesclados com os metadados existentes. Se a opção de mesclagem for selecionada, uma nova versão será criada toda vez que o fluxo for executado. Isso permite rastrear versões de um questionário conforme ele é submetido a mudanças. Cada versão pode ser considerada como uma captura instantânea dos metadados utilizados para coletar um conjunto específico de dados do caso.

Ativar variáveis do sistema. Especifica se as variáveis do sistema são incluídas no arquivo .mdd exportado. Essas incluem variáveis como *Respondent.Serial*, *Respondent.Origin* e *DataCollection.StartTime*.

Configurações de dados do caso. Especifica o arquivo de dados (.sav) do IBM SPSS Statistics no qual os dados do caso são exportados. Observe que todas as restrições de nomes de variáveis e de valores se aplicam aqui, portanto, por exemplo, você pode precisar alternar para a guia Filtro e utilizar a opção "Renomear para o IBM SPSS Statistics" no menu de opções de Filtro para corrigir caracteres inválidos nos nomes de campos.

Gerar um nó de importação para esses dados. Selecione esta opção para gerar automaticamente um nó de origem do IBM SPSS Data Collection que lerá o arquivo de dados exportado.

Conjuntos de múltiplas respostas. Quaisquer conjuntos de múltiplas respostas definidos no fluxo serão preservados automaticamente quando o arquivo for exportado. É possível visualizar e editar conjuntos de múltiplas respostas a partir de qualquer nó com uma guia Filtro. Consulte o tópico "Editando Conjuntos de Múltiplas Respostas" na página 147 para obter mais informações.

Nó Exportação do Analytic Server

O nó Exportação do Analytic Server permite gravar dados de sua análise em uma origem de dados existente do Analytic Server. Isso pode ser, por exemplo, arquivos de texto no Hadoop Distributed File System (HDFS) ou um banco de dados.

Geralmente, um fluxo com um nó Exportação do Analytic Server também inicia com nós de Origem de Dados do Analytic Server e é enviado para o Analytic Server e executado no HDFS. Como alternativa, um fluxo com origens de dados "local" pode terminar com um nó Exportação do Analytic Server para fazer o upload de conjuntos de dados relativamente pequenos (não mais que 100.000 registros) para uso com o Analytic Server.

Origem de dados. Selecione uma origem de dados que contém os dados que deseja usar. Uma origem de dados contém os arquivos e os metadados associados a essa origem. Clique em **Selecionar** para exibir uma lista de origens de dados disponíveis. Consulte o tópico "Selecionando uma origem de dados" na página 13 para obter mais informações.

Se precisar criar uma nova origem de dados ou editar uma existente, clique em **Ativar Editor de Origem de Dados...**

Modo. Selecione **Anexar** para incluir na origem de dados existente, ou **Sobrescrever** para substituir o conteúdo da origem de dados.

Gerar um nó Importação para estes dados. Selecione para gerar um nó de origem para os dados, conforme exportados para a origem de dados especificada. Esse nó é incluído na tela de fluxo.

Nó Exportação do IBM Cognos BI

O nó Exportação do IBM Cognos BI permite exportar dados de um fluxo do IBM SPSS Modeler para o Cognos BI, em formato UTF-8. Dessa forma, o Cognos BI pode fazer uso de dados transformados ou escorados a partir do IBM SPSS Modeler. Por exemplo, é possível utilizar o Cognos BI Report Studio para criar um relatório com base nos dados exportados, incluindo os valores de previsões e de confiança. O relatório poderá, então, ser salvo no servidor Cognos BI e distribuído aos usuários do Cognos BI.

Nota: Você pode exportar apenas os dados relacionais, não dados de OLAP.

Para exportar dados para o Cognos BI, é necessário especificar o seguinte:

- Conexão Cognos - a conexão com o servidor Cognos BI
- Conexão ODBC – a conexão com o servidor de dados Cognos que o servidor Cognos BI utiliza

Na conexão do Cognos, especifique uma origem de dados do Cognos a ser utilizada. Essa origem de dados deve usar o mesmo login que a origem de dados ODBC.

Exporte os dados do fluxo real no servidor de dados e os metadados do pacote no servidor Cognos BI.

Assim como com qualquer outro nó de exportação, também é possível utilizar a guia Publicar da caixa de diálogo do nó para publicar o fluxo para implementação utilizando o IBM SPSS Modeler Solution Publisher.

Conexão do Cognos

Este é o local onde você especifica a conexão com o servidor Cognos BI que deseja utilizar para a exportação. O procedimento envolve exportar os metadados para um novo pacote no servidor Cognos BI, enquanto o fluxo de dados é exportado para o servidor de dados Cognos.

Conexão. Clique no botão **Editar** para exibir uma caixa de diálogo na qual é possível definir a URL e outros detalhes do servidor Cognos BI para o qual deseja exportar os dados. Se você já estiver com login efetuado em um servidor Cognos BI por meio do IBM SPSS Modeler, será possível também editar os detalhes da conexão atual. Consulte o tópico “Conexões Cognos” na página 39 para obter mais informações.

Origem de dados. O nome da origem de dados Cognos (geralmente um banco de dados) para o qual você está exportando os dados. A lista suspensa mostra todas as origens de dados do Cognos que podem ser acessadas a partir da conexão atual. Clique no botão **Atualizar** para atualizar a lista.

Pasta. O caminho e o nome da pasta no servidor Cognos BI na qual o pacote de exportação deve ser criado.

Nome do pacote. O nome do pacote na pasta especificada que deve conter os metadados exportados. Este deve ser um novo pacote com um assunto de query único; não é possível exportá-lo para um pacote existente.

Modo. Especifica como você deseja executar a exportação:

- **Publicar pacote agora.** (padrão) Executa a operação de exportação assim que clicar em **Executar**.
- **Exportar script de ação.** Cria um script XML que pode ser executados mais tarde (por exemplo, utilizando o Framework Manager) para executar a exportação. Digite o caminho e o nome do arquivo para o script no campo **Arquivo** ou utilize o botão **Editar** para especificar o nome e o local do arquivo de script.

Gerar um nó de importação para esses dados. Selecione para gerar um nó de origem para os dados, conforme exportados para a origem de dados e para a tabela especificadas. Ao clicar em **Executar**, esse nó é incluído na tela de fluxo.

Conexão ODBC

Aqui você especifica a conexão com o servidor de dados Cognos (ou seja, o banco de dados) para o qual os dados em fluxo devem ser exportados.

Nota: assegure-se de que a origem de dados que você especificar aqui aponte para a mesma origem especificada no painel **Conexões do Cognos**. Assegure-se também de que a origem de dados de conexão do Cognos utilize o mesmo login que a origem de dados ODBC.

Origem de dados. Mostra a origem de dados selecionada. Insira o nome ou selecione-o na lista suspensa. Se o banco de dados desejado não for exibido na lista, selecione **Incluir nova conexão com o banco de dados** e localize seu banco de dados a partir da caixa de diálogo Conexões com o Banco de Dados. Consulte o tópico “Incluindo uma conexão com o banco de dados” na página 19 para obter mais informações.

Nome da tabela. Insira o nome da tabela para a qual deseja enviar os dados. Se selecionar a opção **Inserir na tabela**, será possível selecionar uma tabela existente no banco de dados clicando no botão **Selecionar**.

Criar tabela. Selecione esta opção para criar uma nova tabela de banco de dados ou para sobrescrever uma tabela de banco de dados existente.

Inserir na tabela. Selecione esta opção para inserir os dados como novas linhas em uma tabela de base de dados existente.

Mesclar tabela. (Quando disponível) Selecione essa opção para atualizar colunas de banco de dados selecionadas com valores de campos de dados de origem correspondentes. Selecionar essa opção ativa o botão **Mesclar**, que exibe um diálogo a partir do qual é possível mapear os campos de dados de origem para colunas de banco de dados.

Descartar tabela existente. Selecione esta opção para excluir qualquer tabela existente com o mesmo nome ao criar uma nova tabela.

Excluir linhas existentes. Selecione esta opção para excluir linhas existentes da tabela antes de exportar, ao inserir em uma tabela.

Nota: se qualquer uma das duas opções acima forem selecionadas, você receberá uma mensagem **Aviso de sobrescrição** quando executar o nó. Para suprimir os avisos, cancele a seleção de **Avisar quando um nó sobrescreve uma tabela de base de dados** na guia Notificações da caixa de diálogo Opções do Usuário.

Tamanho da sequência de caracteres padrão. Os campos que forem marcados como sem tipo em um nó Tipo de envio de dados são gravados no banco de dados como campos de sequência de caracteres. Especifique o tamanho das sequências de caracteres a serem utilizadas para campos sem tipo.

Clique em **Esquema** para abrir uma caixa de diálogo na qual é possível configurar várias opções de exportação (para bancos de dados que suportam esse recurso), configure tipos de dados SQL para seus campos e especifique a chave primária para propósitos de indexação do banco de dados. Consulte o tópico “Opções do Esquema de Exportação de Banco de Dados” na página 346 para obter mais informações.

Clique em **Índices** para especificar opções para indexar a tabela exportada a fim de melhorar o desempenho do banco de dados. Consulte o tópico “Opções do Índice de Exportação do Banco de Dados” na página 349 para obter mais informações.

Clique em **Avançado** para especificar as opções de carregamento em massa e de confirmação do banco de dados. Consulte o tópico “Opções Avançadas de Exportação de Banco de Dados” na página 350 para obter mais informações.

Colocar nomes de tabelas e colunas entre aspas. Selecione opções utilizadas ao enviar uma instrução CREATE TABLE para o banco de dados. As tabelas ou colunas com espaços ou caracteres não padrão devem ser colocadas entre aspas.

- **Conforme necessário.** Selecione para permitir que o IBM SPSS Modeler determine automaticamente quando aspas são necessárias em uma base individual.
- **Sempre.** Selecione para sempre colocar nomes de tabelas e de colunas entre aspas.
- **Nunca.** Selecione para desativar o uso de aspas.

Gerar um nó de importação para esses dados. Selecione para gerar um nó de origem para os dados, conforme exportados para a origem de dados e para a tabela especificadas. Ao clicar em **Executar**, esse nó é incluído na tela de fluxo.

Nó Exportação do IBM Cognos TM1

O nó Exportação do IBM Cognos BI permite exportar dados a partir de um fluxo do IBM SPSS Modeler para o Cognos TM1. Dessa forma, o Cognos BI pode fazer uso de dados transformados ou escorados a partir do IBM SPSS Modeler.

Nota: É possível exportar somente medidas que não forem de dados de dimensão de contexto; como alternativa, é possível incluir novos elementos no cubo.

Para exportar dados para o Cognos BI, é necessário especificar o seguinte:

- A conexão com o servidor Cognos TM1.
- O Cubo para o qual os dados serão exportados.
- O mapeamento dos nomes de dados SPSS para as dimensões e medidas do TM1 equivalentes.

Nota: O usuário do TM1 precisa das seguintes permissões: privilégio de Gravação de cubos, privilégio de Leitura de processos do SPSS, privilégio de Leitura de dimensões, privilégio de Gravação de elementos de dimensão e privilégio de Gravação de arquivos .blb.

Assim como com qualquer outro nó de exportação, também é possível utilizar a guia Publicar da caixa de diálogo do nó para publicar o fluxo para implementação utilizando o IBM SPSS Modeler Solution Publisher.

Nota: Antes de poder utilizar os nós Origem ou Exportação do TM1 no SPSS Modeler, deve-se verificar algumas configurações no arquivo tm1s.cfg, que é o arquivo de configuração do servidor TM1 no diretório-raiz do servidor TM1.

- HTTPPortNumber - configure um número de porta válido; geralmente 1 a 65535.
- UseSSL - se configurar para *True*, o HTTPS será usado como o protocolo de transporte. Nesse caso, deve-se importar a certificação do TM1 do JRE SPSS Modeler Server.

Conectando-se a um cubo IBM Cognos TM1 para exportar dados

O primeiro passo na exportação de dados em um banco de dados IBM Cognos TM1 é selecionar o host de administração do TM1 relevante, além do servidor e do cubo associados, na guia **Conexão** da caixa de diálogo IBM Cognos TM1.

Nota: Somente valores "nulos" reais serão descartados quando exportar dados no TM1. Valores zero (0) serão exportados como um valor válido. Observe também que somente os campos com um tipo de armazenamento de *sequência de caracteres* podem ser mapeados para dimensões na guia Mapeamento. Antes de exportar para o TM1, deve-se utilizar o cliente do IBM SPSS Modeler para converter tipos de dados não sequência de caracteres em sequências de caracteres.

Host de Administração Digite a URL do host de administração no qual o servidor TM1 ao qual deseja conectar-se está instalado. O host de administração é definido como uma URL única para todos os servidores TM1. Nesta URL, todos os servidores IBM Cognos TM1 instalados e em execução em seu ambiente podem ser descobertos e acessados.

Servidor TM1 Quando tiver estabelecido a conexão com o host de administração, selecione o servidor que contém os dados que deseja importar e clique em **Login**. Se você não tiver se conectado anteriormente a esse servidor, será solicitado a inserir o **Nome do usuário** e a **Senha**; como alternativa, é possível procurar pelos detalhes de login inseridos anteriormente que você salvou como uma **Credencial Armazenada**.

Selecionar um cubo do TM1 para exportar Exibe o nome dos cubos dentro do servidor do TM1 no qual é possível exportar dados.

Para escolher os dados a serem exportados, selecione o cubo e clique na seta para a direita para mover o cubo para o campo **Exportar no cubo**. Quando tiver selecionado o cubo, utilize a guia Mapeamento para mapear as dimensões e medidas do TM1 para os campos SPSS relevantes ou para um valor fixo (operação *Selecionar*).

Mapeando dados do IBM Cognos TM1 para exportação

Após selecionar o host de administração TM1 e o servidor e cubo TM1 associados, use a guia Mapeamento na caixa de diálogo IBM Cognos TM1 Export para mapear as dimensões e medidas TM1 para os campos SPSS ou configure as dimensões TM1 em um valor fixo.

Nota: Somente os campos com um tipo de armazenamento de *sequência de caracteres* podem ser mapeados para dimensões. Antes de exportar para o TM1, deve-se utilizar o cliente do IBM SPSS Modeler para converter tipos de dados não sequência de caracteres em sequências de caracteres.

Campos Lista os nomes de campos de dados a partir do arquivo de dados SPSS que estiverem disponíveis para exportação.

Dimensões do TM1 Mostra o cubo do TM1 selecionado na guia Conexões, junto de suas dimensões regulares, da dimensão de medida e dos elementos da dimensão da medida selecionada. Selecione o nome da dimensão ou da medida do TM1 para mapear para o campo de dados do SPSS.

As opções a seguir estão disponíveis no guia Mapeamento.

Selecionar uma dimensão de medida Na lista de dimensões do cubo selecionado, selecione uma para ser a dimensão de medida.

Ao selecionar uma dimensão, exceto a dimensão de medida, e clicar em **Selecionar**, um diálogo será exibido mostrando os elementos da folha da dimensão selecionada. É possível selecionar apenas os elementos da folha. Os elementos selecionados são rotulados com um **S**.

Mapear Mapeia o campo de dados do SPSS selecionado para a dimensão ou medida do TM1 selecionada (uma dimensão regular ou uma medida ou elemento específico da dimensão de medida). Campos mapeados são rotulados com um **M**.

Remover mapeamento Remove o mapeamento do campo de dados do SPSS selecionado da dimensão ou medida do TM1 selecionada. Observe que é possível remover mapeamento somente de um mapeamento único por vez. O campo de dados SPSS de mapeamento removido é movido de volta para a coluna esquerda.

Criar Novo Cria uma nova medida na dimensão de medida do TM1. Um diálogo é exibido no qual você insere o novo **Nome de medida do TM1**. Essa opção está disponível somente para dimensões de medida, não para dimensões regulares.

Para obter mais informações sobre o TM1, consulte a documentação do IBM Cognos TM1 em http://www-01.ibm.com/support/knowledgecenter/SS9RXT_10.2.2/com.ibm.swg.ba.cognos.ctm1.doc/welcome.html.

Nó Exportação do SAS

Nota: esse recurso está disponível no SPSS Modeler Professional e no SPSS Modeler Premium.

O nó de exportação SAS permite gravar dados em formato do SAS para serem lidos no SAS ou em um pacote de software compatível com o SAS. É possível exportar em três formatos de arquivo SAS: SAS para Windows/OS2, SAS para UNIX ou SAS Versão 7/8.

Guia Exportação do Nó Exportação do SAS

Exportar arquivo. Especifique o nome do arquivo. Insira um nome de arquivo ou clique no botão seletor de arquivo para navegar para a localização do arquivo.

Exportação. Especifique o formato do arquivo de exportação. As opções são **SAS para Windows/OS2**, **SAS para UNIX** ou **SAS Versão 7/8**.

Exportar nomes de campo. Selecione as opções para exportar nomes e rótulos de campos a partir do IBM SPSS Modeler para uso com o SAS.

- **Nomes e rótulos de variável.** Selecione para exportar nomes e rótulos de campo do IBM SPSS Modeler. Os nomes são exportados como nomes de variáveis do SAS, enquanto os rótulos são exportadas como rótulos de variáveis do SAS.
- **Nomes como rótulos de variável.** Selecione para utilizar os nomes de campo do IBM SPSS Modeler como rótulos de variável no SAS. O IBM SPSS Modeler permite caracteres nos nomes de campos que são inválidos em nomes de variáveis do SAS. Para evitar a possibilidade de criar nomes de SAS inválidos, selecione **Nomes e rótulos da variável**.

Gerar um nó de importação para esses dados. Selecione esta opção para gerar automaticamente um nó de origem do SAS que lerá o arquivo de dados exportado. Consulte o tópico “Nó de Origem do SAS” na página 42 para obter mais informações.

Nó Exportação do Excel

O nó de exportação do Excel gera dados no formato .xlsx do Microsoft Excel. Opcionalmente, é possível escolher ativar o Excel automaticamente e abrir o arquivo exportado quando o nó for executado.

Guia Exportação do Nó do Excel

Nome do arquivo. Insira um nome de arquivo ou clique no botão seletor de arquivo para navegar para a localização do arquivo. O nome do arquivo padrão é *excelexp.xlsx*.

Tipo de arquivo. O formato de arquivo .xlsx do Excel é suportado.

Criar novo arquivo. Cria um novo arquivo do Excel.

Inserir no arquivo existente. O conteúdo é substituído a partir da célula designada pelo campo **Iniciar na célula**. Outras células na planilha são deixadas com seu conteúdo original.

Incluir nomes de campo. Especifica se nomes de campos devem ser incluídos na primeira linha da planilha.

Iniciar na célula. O local da célula utilizado para o primeiro registro de exportação (ou para o primeiro nome de campo se **Incluir nomes de campo** estiver selecionada). Os dados são preenchidos à direita e para baixo a partir desta célula inicial.

Escolher planilha. Especifica a planilha à qual deseja exportar os dados. É possível identificar a planilha por índice ou por nome:

- **Por índice.** Se estiver criando um novo arquivo, especifique um número de 0 a 9 para identificar a planilha para a qual deseja exportar, iniciando com 0 para a primeira planilha, 1 para a segunda planilha, e assim por diante. Será possível utilizar valores de 10 ou superior apenas se uma planilha já existir nesta posição.
- **Por nome.** Se estiver criando um novo arquivo, especifique o nome utilizado para a planilha. Se estiver inserindo em um arquivo existente, os dados serão inseridos nesta planilha se ela existir, caso contrário, uma nova planilha com esse nome será criada.

Ativar o Excel. Especifica se o Excel é ativado automaticamente no arquivo exportado quando o nó é executado. Observe que ao executar no modo distribuído com relação ao IBM SPSS Modeler Server, a saída é salva no sistema de arquivos do servidor, e o Excel será ativado no Cliente com uma cópia do arquivo exportado.

Gerar um nó de importação para esses dados. Selecione esta opção para gerar automaticamente um nó de origem do Excel que lerá o arquivo de dados exportado. Consulte o tópico "Nó de Origem do Excel" na página 42 para obter mais informações.

Nó Exportação XML

O nó Exportação XML permite gerar dados no formato XML, utilizando a codificação UTF-8. Opcionalmente, é possível criar um nó de origem XML para ler os dados exportados de volta no fluxo.

Arquivo de exportação XML. O caminho e o nome do arquivo completos do arquivo XML para o qual deseja exportar os dados.

Usar esquema XML. Marque esta caixa de seleção se desejar utilizar um esquema ou um DTD para controlar a estrutura dos dados exportados. Fazer isso ativa o botão **Mapear**, descrito a seguir.

Se não utilizar um esquema ou um DTD, a estrutura padrão a seguir será utilizada para os dados exportados:

```
<records>
  <record>
    <fieldname1>value</fieldname1>
    <fieldname2>value</fieldname2>
    :
    <fieldnameN>value</fieldnameN>
  </record>
  <record>
    :
    :
  </record>
  :
  :
</records>
```

Os espaços em um nome de campo são substituídos por sublinhados; por exemplo, "My Field" se torna <My_Field>.

Mapear. Se tiver escolhido utilizar um esquema XML, esse botão abrirá um diálogo no qual é possível especificar qual parte da estrutura XML deve ser utilizada para iniciar cada novo registro. Consulte o tópico “Opções de registros de mapeamento XML” para obter mais informações.

Campos mapeados. Indica o número de campos que foram mapeados.

Gerar um nó de importação para esses dados. Selecione esta opção para gerar automaticamente um nó de origem XML que lerá o arquivo de dados exportado novamente no fluxo. Consulte o tópico “Nó de Origem XML” na página 43 para obter mais informações.

Gravando dados XML

Quando um elemento XML é especificado, o valor do campo é colocado dentro da identificação de elemento:

```
<element>value</element>
```

Quando um atributo é mapeado, o valor do campo é colocado como o valor para o atributo:

```
<element attribute="value">
```

Se um campo for mapeado para um elemento acima do elemento <records>, o campo será gravado apenas uma vez e será uma constante para todos os registros. O valor para este elemento virá do primeiro registro.

Se um valor nulo tiver que ser gravado, isso será feito ao especificar um conteúdo vazio. Para elementos, isto é:

```
<element></element>
```

Para atributos, é:

```
<element attribute="">
```

Opções de registros de mapeamento XML

A guia Registros permite especificar qual parte da estrutura XML será usada para iniciar cada novo registro. Para mapear corretamente em um esquema, é necessário especificar o delimitador de registro.

Estrutura XML. Uma árvore hierárquica mostrando a estrutura do esquema XML especificado na tela anterior.

Registros (expressão XPath). Para configurar o delimitador de registro, selecione um elemento na estrutura XML e clique no botão de seta para a direita. Sempre que este elemento for encontrado nos dados de origem, um novo registro será criado no arquivo de saída.

Nota: se você selecionar o elemento raiz na estrutura XML, apenas um único registro poderá ser gravado e todos os outros registros serão ignorados.

Opções de Campos de Mapeamento XML

A guia Campos é utilizada para mapear os campos no conjunto de dados para elementos ou atributos na estrutura XML quando um arquivo de esquema é utilizado.

Nomes de campos que correspondem a um nome de elemento ou de atributo são mapeados automaticamente contanto que o nome do elemento ou do atributo seja exclusivo. Portanto, se houver um elemento e um atributo com o nome field1, não haverá mapeamento automático. Se houver apenas um item na estrutura chamado field1, um campo com esse nome no fluxo será mapeado automaticamente.

Campos. A lista de campos no modelo. Selecione um ou mais campos como a parte de origem do mapeamento. É possível utilizar os botões na parte inferior da lista para selecionar todos os campos, ou todos os campos com um nível de medição específica.

Estrutura XML. Selecione um elemento na estrutura XML como o destino do mapa. Para criar o mapeamento, clique em Mapear. O mapeamento é exibido. O número de campos que foram mapeados dessa maneira é exibido abaixo dessa lista.

Para remover um mapeamento, selecione o item na lista de estrutura XML e clique em **Remover mapeamento**.

Exibir atributos. Exibe ou oculta os atributos, se houver, dos elementos XML na estrutura XML.

Visualização de mapeamento XML

Na guia Visualizar, clique em **Atualizar** para abrir uma visualização do XML que será gravado.

Se o mapeamento estiver incorreto, retorne para a guia Registros ou Campos para corrigir os erros e clique em **Atualizar** novamente para ver o resultado.

Capítulo 8. IBM SPSS Statistics Nós

Nós do IBM SPSS Statistics - Visão Geral

Para complementar o IBM SPSS Modeler e suas habilidades de mineração de dados, o IBM SPSS Statistics fornece a capacidade de realizar análise estatística e gerenciamento de dados adicionais.

Se você tiver uma cópia compatível e licenciada do IBM SPSS Statistics instalada, será possível conectar-se a ela a partir do IBM SPSS Modeler e executar manipulação e análise de dados complexas e de diversos passos que, do contrário, não seriam suportadas pelo IBM SPSS Modeler. Para o usuário avançado, há também a opção para modificar adicionalmente a análise utilizando sintaxe de comando. Consulte as Notas sobre a Liberação para obter informações sobre compatibilidade de versão.

Se disponíveis, os nós do IBM SPSS Statistics são mostrados em uma parte dedicada da paleta de nós.

Nota: é recomendado instanciar seus dados em um nó Tipo antes de utilizar os nós Transformação, Modelo ou Saída do IBM SPSS Statistics. Isso também é um requisito quando utilizar o comando de sintaxe AUTORECODE.

A paleta do IBM SPSS Statistics contém os seguintes nós:



O nó Arquivo de Estatísticas lê dados do formato de arquivo *.sav* ou *.zsav* usado pelo IBM SPSS Statistics, bem como de arquivos de cache salvos em IBM SPSS Modeler que também utilizam o mesmo formato.



O nó Transformação de Estatísticas executa uma seleção de comandos de sintaxe do IBM SPSS Statistics com relação às origens de dados no IBM SPSS Modeler. Esse nó requer uma cópia licenciada do IBM SPSS Statistics.



O nó Modelo de Estatísticas permite analisar e trabalhar com seus dados executando os procedimentos do IBM SPSS Statistics que produzem o PMML. Esse nó requer uma cópia licenciada do IBM SPSS Statistics.



O nó Saída de Estatísticas permite chamar um procedimento do IBM SPSS Statistics para analisar seus dados do IBM SPSS Modeler. Uma ampla variedade de procedimentos de analítica do IBM SPSS Statistics está disponível. Esse nó requer uma cópia licenciada do IBM SPSS Statistics.



O nó Exportação de Estatísticas gera dados no formato IBM SPSS Statistics *.sav* ou *.zsav*. Os arquivos *.sav* ou *.zsav* podem ser lidos pelo IBM SPSS Statistics Base e por outros produtos. Este também é o formato utilizado para arquivos em cache no IBM SPSS Modeler.

Nota: se a sua cópia do SPSS Statistics estiver licenciada para um único usuário e um fluxo for executado com duas ou mais ramificações, cada uma das quais contendo um nó do SPSS Statistics, você poderá receber um erro de licenciamento. Isso ocorre quando a sessão do SPSS Statistics de uma ramificação não

tiver terminado antes da sessão de outra ramificação tentar iniciar. Se possível, projete novamente o fluxo de modo que diversas ramificações com nós do SPSS Statistics não sejam executadas em paralelo.

Nó Arquivo de Estatísticas

É possível utilizar o nó Arquivo de Estatísticas para ler dados diretamente a partir de um arquivo (.sav ou .zsav) do IBM SPSS Statistics salvo. Este formato agora é utilizado para substituir o arquivo de cache a partir de versões anteriores do IBM SPSS Modeler. Se desejar importar um arquivo de cache salvo, deve-se utilizar o nó Arquivo do IBM SPSS Statistics.

Importar arquivo. Especifique o nome do arquivo. É possível inserir um nome do arquivo ou clicar no botão de reticências (...) para selecionar um arquivo. O caminho do arquivo é mostrado após ter selecionado um arquivo.

O arquivo possui senha criptografada. Selecione esta caixa se você souber que o arquivo está protegido por senha; será solicitado a inserir a **Senha**. Se o arquivo for protegido por senha, e você não inserir a senha, uma mensagem de aviso será exibida se tentar alterar para outra guia, atualizar os dados, visualizar o conteúdo do nó ou tentar executar um fluxo contendo o nó.

Nota: Os arquivos protegidos por senha podem ser abertos somente pelo IBM SPSS Modeler versão 16 ou superior.

nomes de variáveis. Selecione um método de manipulação de nomes de variáveis e de rótulos na importação a partir de um arquivo .sav ou .zsav do IBM SPSS Statistics. Os metadados que você escolher incluir aqui persistirão durante todo o seu trabalho no IBM SPSS Modeler e podem ser exportados novamente para uso no IBM SPSS Statistics.

- **Ler nomes e rótulos.** Selecione para ler os nomes de variáveis e rótulos no IBM SPSS Modeler. Por padrão, esta opção é selecionada e os nomes de variáveis são exibidos no nó Tipo. Os rótulos podem ser exibidos nos gráficos, navegadores de modelo e em outros tipos de saída, dependendo das opções especificadas na caixa de diálogo de propriedades do fluxo. Por padrão, a exibição dos rótulos na saída é desativada.
- **Ler rótulos como nomes.** Selecione para ler os rótulos de variável descritivos a partir do arquivo .sav ou .zsav do IBM SPSS Statistics ao invés dos nomes de campo curtos e utilizar esses rótulos como nomes de variáveis no IBM SPSS Modeler.

Valores. Selecione um método de manipulação de valores e de rótulos na importação a partir de um arquivo .sav ou .zsav do IBM SPSS Statistics. Os metadados que você escolher incluir aqui persistirão durante todo o seu trabalho no IBM SPSS Modeler e podem ser exportados novamente para uso no IBM SPSS Statistics.

- **Ler dados e rótulos.** Selecione para ler os valores reais e os rótulos de valor no IBM SPSS Modeler. Por padrão, esta opção é selecionada e os valores em si são exibidos no nó Tipo. Os rótulos de valor podem ser exibidos no Construtor de Expressões, nos gráficos, navegadores de modelo e em outros tipos de saída, dependendo das opções especificadas na caixa de diálogo de propriedades do fluxo.
- **Ler rótulos como dados.** Selecione se deseja usar rótulos de valor do arquivo .sav ou .zsav em vez dos códigos numérico ou simbólico usados para representar os valores. Por exemplo, selecionar essa opção para dados com um campo de gênero cujos valores de 1 e 2 representam de fato *masculino* e *feminino*, respectivamente, converterá o campo em uma sequência de caracteres e importará *masculino* e *feminino* como valores reais.

É importante considerar valores omissos em seus dados do IBM SPSS Statistics antes de selecionar essa opção. Por exemplo, se um campo numérico usar rótulos apenas para valores omissos (0 = *Sem resposta*, -99 = *Desconhecido*), então, selecionar a opção acima importará apenas os rótulos de valores *Sem resposta* e *Desconhecido* e converterá o campo em uma sequência de caracteres. Nesses casos, deve-se importar os próprios valores e configurar valores omissos em um nó Tipo.

Usar informações de formato de campo para determinar o armazenamento. Se essa caixa estiver desmarcada, os valores de campo que são formatados no arquivo *.sav* como números inteiros (ou seja, campos especificados como *Fn.0* na Visualização de Variável no IBM SPSS Statistics) são importados usando o armazenamento de número inteiro. Todos os outros valores de campo, exceto sequências de caracteres, são importados como números reais.

Se esta caixa estiver selecionada (padrão), todos os valores de campos, exceto sequências de caracteres, serão importados como números reais, independentemente se estiverem formatados no arquivo *.sav* como números inteiros ou não.

Conjuntos de múltiplas respostas. Os conjuntos de múltiplas respostas definidos no arquivo IBM SPSS Statistics serão preservados automaticamente quando o arquivo for importado. É possível visualizar e editar conjuntos de múltiplas respostas a partir de qualquer nó com uma guia Filtro. Consulte o tópico “Editando Conjuntos de Múltiplas Respostas” na página 147 para obter mais informações.

Nó Transformação de Estatísticas

O nó Transformação de Estatísticas permite concluir transformações de dados usando a sintaxe do comando IBM SPSS Statistics. Isso permite concluir uma série de transformações não suportadas pelo IBM SPSS Modeler e também a automação de transformações complexas de diversos passos, incluindo a criação de um número de campos a partir de um único nó. Ele é semelhante ao nó Saída de Estatísticas, exceto que os dados são retornados para o IBM SPSS Modeler para análise adicional, ao passo que, no nó Saída, os dados são retornados como os objetos de saída solicitados, como gráficos ou tabelas.

Deve-se ter uma versão compatível do IBM SPSS Statistics instalada e licenciada em seu computador para utilizar esse nó. Consulte o tópico “Aplicativos Auxiliares do IBM SPSS Statistics” na página 341 para obter mais informações. Consulte as Notas sobre a Liberação para obter informações de compatibilidade.

Se necessário, é possível utilizar a guia Filtro para filtrar ou renomear campos para eles estejam em conformidade com os padrões de nomenclatura do IBM SPSS Statistics. Consulte o tópico “Renomeando ou Filtrando Campos para o IBM SPSS Statistics” na página 378 para obter mais informações.

Referência da Sintaxe. Para obter detalhes sobre procedimentos específicos do IBM SPSS Statistics, consulte o guia de *Referência de Sintaxe de Comando do IBM SPSS Statistics*, incluindo sua cópia do software do IBM SPSS Statistics. Para visualizar o guia a partir da guia Sintaxe, escolha a opção **Editor de sintaxe** e clique no botão Ativar Ajuda de Sintaxe do IBM SPSS Statistics.

Nota: nem toda a sintaxe do IBM SPSS Statistics é suportada por este nó. Consulte o tópico “Sintaxe Permissível” na página 372 para obter mais informações.

Nó Transformação de Estatísticas – Guia Sintaxe

Opção do diálogo do IBM SPSS Statistics

Se você não estiver familiarizado com a sintaxe do IBM SPSS Statistics para um procedimento, a maneira mais simples de criar sintaxe no IBM SPSS Modeler é escolher a opção **Diálogo do IBM SPSS Statistics**, selecionar a caixa de diálogo para o procedimento, preencher a caixa de diálogo e clicar em OK. Fazer isso coloca a sintaxe na guia Sintaxe do nó do IBM SPSS Statistics que está usando com o IBM SPSS Modeler. Em seguida, é possível executar o fluxo para obter a saída a partir do procedimento.

Opção do editor de Sintaxe do IBM SPSS Statistics

Verificar. Após inserir seus comandos de sintaxe na parte superior da caixa de diálogo, utilize este botão para validar suas entradas. Qualquer sintaxe incorreta é identificada na parte inferior da caixa de diálogo.

Para assegurar que o processo de verificação não demore muito tempo, ao validar a sintaxe, ela será verificada com relação a uma amostra representativa de seus dados para assegurar que as entradas sejam válidas ao invés de verificar com relação ao conjunto de dados inteiro.

Sintaxe Permissível

Se você tiver muita sintaxe anterior a partir do IBM SPSS Statistics ou se estiver familiarizado com os recursos de preparação de dados do IBM SPSS Statistics, será possível utilizar o nó Transformação de Estatísticas para executar muitas das suas transformações existentes. Como uma diretriz, o nó permite transformar dados de maneiras previsíveis--por exemplo, ao executar comandos em loop ou alterar, incluir, classificar, filtrar ou selecionar os dados.

Os exemplos dos comandos que podem ser executados são:

- Calcular números aleatórios de acordo com uma distribuição binomial:

```
COMPUTE newvar = RV.BINOM(10000,0.1)
```

- Recodificar uma variável em uma nova variável:

```
RECODE Age (Lowest thru 30=1) (30 thru 50=2) (50 thru Highest=3) INTO AgeRecoded
```

- Substituir valores omissos:

```
RMV Age_1=SMEAN(Age)
```

A sintaxe do IBM SPSS Statistics que é suportada pelo nó Transformação de Estatísticas é listada abaixo.

Nome do comando

ADD VALUE LABELS

APPLY DICTIONARY

AUTORECODE

BREAK

CD

CLEAR MODEL PROGRAMS

CLEAR TIME PROGRAM

CLEAR TRANSFORMATIONS

COMPUTE

COUNT

CREATE

DATE

DEFINE-!ENDDFINE

DELETE VARIABLES

DO IF

DO REPEAT

ELSE

ELSE IF

END CASE

END FILE

END IF

END INPUT PROGRAM

END LOOP

END REPEAT

Nome do comando

EXECUTE
FILE HANDLE
FILE LABEL
FILE TYPE-END FILE TYPE
FILTER
FORMATS
IF
INCLUDE
INPUT PROGRAM-END INPUT PROGRAM
INSERT
LEAVE
LOOP-END LOOP
MATRIX-END MATRIX
MISSING VALUES
N OF CASES
NUMERIC
PERMISSIONS
PRESERVE
RANK
RECODE
RENAME VARIABLES
RESTORE
RMV
SAMPLE
SELECT IF
SET
SORT CASES
STRING
SUBTITLE
TEMPORARY
TITLE
UPDATE
V2C
VALIDATEDATA
VALUE LABELS
VARIABLE ATTRIBUTE
VARSTOCASES
VECTOR

Nó Modelo de Estatísticas

O nó Modelo de Estatísticas permite analisar e trabalhar com seus dados executando os procedimentos do IBM SPSS Statistics que produzem o PMML. Em seguida, os nuggets do modelo que você criar podem ser utilizados de forma usual nos fluxos do IBM SPSS Modeler para escoragem, e assim por diante.

Deve-se ter uma versão compatível do IBM SPSS Statistics instalada e licenciada em seu computador para utilizar esse nó. Consulte o tópico “Aplicativos Auxiliares do IBM SPSS Statistics” na página 341 para obter mais informações. Consulte as Notas sobre a Liberação para obter informações de compatibilidade.

Os procedimentos analíticos do IBM SPSS Statistics que estiverem disponíveis dependem do tipo de licença que você possui.

Nó do Modelo de Estatísticas - Guia Modelo

Nome do modelo É possível gerar o nome do modelo automaticamente com base no campo de destino ou ID (ou tipo de modelo em casos em que nenhum campo não é especificado) ou especifique um nome customizado.

Selecionar um diálogo. Clique para exibir uma lista de procedimentos do IBM SPSS Statistics disponíveis que podem ser selecionados e executados. A lista mostra somente os procedimentos que produzem PMML e para os quais você estiver licenciado, e não inclui procedimentos gravados pelo usuário.

1. Clique no procedimento necessário; o diálogo relevante do IBM SPSS Statistics é exibido.
2. No diálogo IBM SPSS Statistics, insira detalhes para o procedimento.
3. Clique em **OK** para retornar para o nó Modelo de Estatísticas; a sintaxe do IBM SPSS Statistics é exibida na guia Modelo.
4. Para retornar para o diálogo do IBM SPSS Statistics a qualquer momento, por exemplo, para modificar sua query, clique no botão de exibição do diálogo do IBM SPSS Statistics à direita do botão de seleção de procedimento.

Nó do Modelo de Estatísticas - Sumarização do Nugget do Modelo

Ao executar o nó Modelo de Estatísticas, ele executa o procedimento do IBM SPSS Statistics associado e cria um nugget do modelo que pode ser utilizado nos fluxos do IBM SPSS Modeler para escoragem.

A guia Sumarização do nugget do modelo exibe informações sobre os campos, sobre as configurações de construção e sobre o processo de estimação do modelo. Os resultados são exibidos em uma visualização em árvore que pode ser expandida ou reduzida ao clicar nos itens específicos.

O botão **Visualizar Modelo** exibe os resultados em uma forma modificada do Visualizador de Saída do IBM SPSS Statistics. Para obter mais informações sobre esse visualizador, consulte a documentação do IBM SPSS Statistics.

As opções de exportação e de impressão usuais estão disponíveis no menu Arquivo. Consulte o tópico “Visualizando a Saída” na página 302 para obter mais informações.

Nó Saída de Estatísticas

O nó Saída de Estatísticas permite chamar um procedimento do IBM SPSS Statistics para analisar seus dados do IBM SPSS Modeler. É possível visualizar os resultados em uma janela do navegador ou salvar os resultados no formato de arquivo de saída do IBM SPSS Statistics. Uma ampla variedade de procedimentos analíticos do IBM SPSS Statistics está acessível a partir do IBM SPSS Modeler.

Deve-se ter uma versão compatível do IBM SPSS Statistics instalada e licenciada em seu computador para utilizar esse nó. Consulte o tópico “Aplicativos Auxiliares do IBM SPSS Statistics” na página 341 para obter mais informações. Consulte as Notas sobre a Liberação para obter informações de compatibilidade.

Se necessário, é possível utilizar a guia Filtro para filtrar ou renomear campos para eles estejam em conformidade com os padrões de nomenclatura do IBM SPSS Statistics. Consulte o tópico “Renomeando ou Filtrando Campos para o IBM SPSS Statistics” na página 378 para obter mais informações.

Referência da Sintaxe. Para obter detalhes sobre procedimentos específicos do IBM SPSS Statistics, consulte o guia de *Referência de Sintaxe de Comando do IBM SPSS Statistics*, incluindo sua cópia do software do IBM SPSS Statistics. Para visualizar o guia a partir da guia Sintaxe, escolha a opção **Editor de sintaxe** e clique no botão Ativar Ajuda de Sintaxe do IBM SPSS Statistics.

Nó de Saída de Estatísticas – Guia Sintaxe

Utilize esta guia para criar a sintaxe para o procedimento do IBM SPSS Statistics que você deseja utilizar para analisar seus dados. A sintaxe é composta de duas partes: uma **instrução** e as **opções** associadas. A instrução especifica a análise ou a operação a ser executada e os campos a serem utilizados. As opções especificam todo o resto, incluindo quais estatísticas serão exibidas, os campos derivados para salvar, e assim por diante.

Opção do diálogo do IBM SPSS Statistics

Se você não estiver familiarizado com a sintaxe do IBM SPSS Statistics para um procedimento, a maneira mais simples de criar sintaxe no IBM SPSS Modeler é escolher a opção **Diálogo do IBM SPSS Statistics**, selecionar a caixa de diálogo para o procedimento, preencher a caixa de diálogo e clicar em OK. Fazer isso coloca a sintaxe na guia Sintaxe do nó do IBM SPSS Statistics que está usando com o IBM SPSS Modeler. Em seguida, é possível executar o fluxo para obter a saída a partir do procedimento.

É possível, opcionalmente, gerar um nó de origem do Arquivo de Estatísticas para importar os dados resultantes. Isso é útil, por exemplo, se um procedimento gravar campos, como escores, no conjunto de dados ativo além da exibição da saída.

Para criar a sintaxe:

1. Clique no botão **Selecionar um diálogo**.
2. Escolha uma das seguintes opções:
 - **Analisar.** Lista o conteúdo do menu Análise do IBM SPSS Statistics; escolha o procedimento que deseja utilizar.
 - **Outro.** Se mostrada, lista os diálogos criados pelo Construtor de Diálogo Personalizado no IBM SPSS Statistics, bem como quaisquer outros diálogos do IBM SPSS Statistics que não aparecem no menu Analisar e para os quais você tiver uma licença. Se não houver nenhum diálogo aplicável, essa opção não será mostrada.

Nota: os diálogos de Preparação de Dados Automática não são mostrados.

Se você tiver um diálogo customizado do IBM SPSS Statistics que cria novos campos, esses campos não poderão ser utilizados no IBM SPSS Modeler porque o nó de Saída de Estatísticas é um nó terminal.

Opcionalmente, marque a caixa **Gerar um nó de importação para os dados resultantes** para criar um nó de origem Arquivo de Estatísticas que pode ser utilizado para importar os dados resultantes em outro fluxo. O nó é colocado na tela, com os dados contidos no arquivo *.sav* especificado no campo **Arquivo** (a localização padrão é o diretório de instalação do IBM SPSS Modeler).

Opção do editor de sintaxe

Para salvar a sintaxe que foi criada para um procedimento frequentemente utilizado:

1. Clique no botão Opções de Arquivo (o primeiro na barra de ferramentas).
2. Selecione **Salvar** ou **Salvar Como** no menu.
3. Salve o arquivo como um arquivo *.sps*.

Para utilizar os arquivos de sintaxe criados anteriormente, substituindo o conteúdo atual, se houver, do editor de Sintaxe:

1. Clique no botão Opções de Arquivo (o primeiro na barra de ferramentas).
2. Escolha **Abrir** no menu.
3. Selecione um arquivo *.sps* para colar o conteúdo na guia Sintaxe do nó de Saída.

Para inserir a sintaxe salva anteriormente sem substituir o conteúdo atual:

1. Clique no botão Opções de Arquivo (o primeiro na barra de ferramentas).
2. Escolha **Inserir** no menu
3. Selecione um arquivo *.smallerpktsrecv* para colar o conteúdo no nó de Saída no ponto especificado pelo cursor.

Opcionalmente, marque a caixa **Gerar um nó de importação para os dados resultantes** para criar um nó de origem Arquivo de Estatísticas que pode ser utilizado para importar os dados resultantes em outro fluxo. O nó é colocado na tela, com os dados contidos no arquivo *.sav* especificado no campo **Arquivo** (a localização padrão é o diretório de instalação do IBM SPSS Modeler).

Quando clicar em **Executar**, os resultados serão mostrados no Visualizador de Saída do IBM SPSS Statistics. Para obter mais informações sobre o visualizador, consulte a documentação do IBM SPSS Statistics.

Nó de Saída de Estatísticas – Guia Saída

A guia Saída permite especificar o formato e a localização da saída. É possível escolher exibir os resultados na tela ou enviá-los para um dos tipos de arquivo disponíveis.

Nome de saída. Especifica o nome da saída produzido quando o nó é executado. **Automático** escolhe um nome com base no nó que gera a saída. Opcionalmente, é possível selecionar **Customizado** para especificar um nome diferente.

Saída para tela (o padrão). Cria um objeto de saída para visualizar on-line. O objeto de saída aparecerá na guia Saídas da janela do gerenciador quando o nó de saída for executado.

Saída para arquivo. Salva a saída em um arquivo quando executar o nó. Se escolher essa opção, insira um nome de arquivo no campo **Nome do arquivo** (ou navegue para um diretório e especifique um nome de arquivo utilizando o botão Seletor de Arquivo) e selecione um tipo de arquivo.

Tipo de arquivo. Escolha o tipo de arquivo para o qual deseja enviar a saída.

- **Documento HTML (*.html).** Grava a saída no formato HTML.
- **Arquivo do Visualizador do IBM SPSS Statistics (*.spv).** Grava a saída em um formato que possa ser lido pelo Visualizador de Saída do IBM SPSS Statistics.
- **Arquivo de Relatórios da Web do IBM SPSS Statistics (*.spw).** Grava a saída no formato de Relatório da Web do IBM SPSS Statistics, que pode ser publicado em um repositório do IBM SPSS Collaboration and Deployment Services e posteriormente visualizado em um navegador da web. Consulte o tópico “Publicação na Web” na página 303 para obter mais informações.

Nota: se selecionar **Saída para tela**, a diretiva OMS VIEWER=N0 do IBM SPSS Statistics não terá efeito e, além disso, as APIs de script (módulo *Básico* e *Python SpssClient*) não estarão disponíveis no IBM SPSS Modeler.

Nó Exportação de Estatísticas

O nó Exportação de Estatísticas permite exportar dados no formato *.sav* do IBM SPSS Statistics. Os arquivos *.sav* do IBM SPSS Statistics podem ser lidos pelo IBM SPSS Statistics Base e por outros módulos. Este também é o formato utilizado para arquivos de cache do IBM SPSS Modeler.

Mapear nomes de campo do IBM SPSS Modeler para nomes de variável do IBM SPSS Statistics pode, às vezes, causar erros porque os nomes de variável do IBM SPSS Statistics são limitados a 64 caracteres e não podem incluir determinados caracteres, como espaços, símbolo de dólar (\$) e hífen (-). Há duas maneiras de ajustar para estas restrições:

- É possível renomear campos de acordo com os requisitos de nome de variável do IBM SPSS Statistics ao clicar na guia Filtro. Consulte o tópico “Renomeando ou Filtrando Campos para o IBM SPSS Statistics” na página 378 para obter mais informações.
- Escolha exportar os nomes de campos e rótulos a partir do IBM SPSS Modeler.

Nota: o IBM SPSS Modeler grava arquivos *.sav* em formato UTF-8 Unicode. O IBM SPSS Statistics suporta apenas arquivos em formato UTF-8 Unicode a partir da liberação 16.0. Para evitar a possibilidade de distorção de dados, os arquivos *.sav* salvos com a codificação Unicode não devem ser usados em liberações do IBM SPSS Statistics anteriores a 16.0. Para obter mais informações, consulte a ajuda do IBM SPSS Statistics.

Conjuntos de múltiplas respostas. Quaisquer conjuntos de múltiplas respostas definidos no fluxo serão preservados automaticamente quando o arquivo for exportado. É possível visualizar e editar conjuntos de múltiplas respostas a partir de qualquer nó com uma guia Filtro. Consulte o tópico “Editando Conjuntos de Múltiplas Respostas” na página 147 para obter mais informações.

Nó Exportação de Estatísticas - Guia Exportação

Exportar arquivo. Especifica o nome do arquivo. Insira um nome de arquivo ou clique no botão seletor de arquivo para navegar para a localização do arquivo.

Tipo de arquivo. Selecione se o arquivo deve ser salvo no formato normal *.sav* ou compactado *.zsav*.

Criptografar arquivo com senha. Selecione essa caixa para proteger o arquivo com uma senha; será solicitado a inserir e confirmar a **Senha** em uma caixa de diálogo separada.

Nota: Os arquivos protegidos por senha podem ser abertos somente pelo IBM SPSS Modeler versão 16 ou superior ou pelo IBM SPSS Statistics versão 21 ou superior.

Exportar nomes de campo. Especifica um método de manipular nomes e rótulos de variáveis na exportação do IBM SPSS Modeler em um arquivo *.sav* ou *.zsav* do IBM SPSS Statistics.

- **Nomes e rótulos de variável.** Selecione para exportar nomes e rótulos de campo do IBM SPSS Modeler. Os nomes são exportados como nomes de variável do IBM SPSS Statistics, enquanto os rótulos são exportados como rótulos de variáveis do IBM SPSS Statistics.
- **Nomes como rótulos de variável.** Selecione para utilizar os nomes de campo do IBM SPSS Modeler como rótulos de variável no IBM SPSS Statistics. O IBM SPSS Modeler permite caracteres nos nomes de campos que são inválidos nos nomes de variável do IBM SPSS Statistics. Para evitar criação de nomes possivelmente inválidos do IBM SPSS Statistics, selecione **Nomes como rótulos da variável** ou utilize a guia Filtro para ajustar os nomes dos campos.

Ativar Aplicativo. Se o IBM SPSS Statistics foi instalado em seu computador, será possível selecionar essa opção para chamar o aplicativo diretamente no arquivo de dados salvo. As opções para ativar o aplicativo devem ser especificadas na caixa de diálogo Aplicativos Auxiliares. Consulte o tópico

“Aplicativos Auxiliares do IBM SPSS Statistics” na página 341 para obter mais informações. Para simplesmente criar um arquivo *.sav* ou *.zsav* do IBM SPSS Statistics sem abrir um programa externo, desmarque essa opção.

Gerar um nó de importação para esses dados. Selecione essa opção para gerar automaticamente um nó de origem do Arquivo de Estatísticas que lerá o arquivo de dados exportado. Consulte o tópico “Nó Arquivo de Estatísticas” na página 370 para obter mais informações.

Renomeando ou Filtrando Campos para o IBM SPSS Statistics

Antes de exportar ou de implementar dados do IBM SPSS Modeler em aplicativos externos, como IBM SPSS Statistics, poderá ser necessário renomear ou ajustar os nomes de campos. As caixas de diálogo Transformação de Estatísticas, Saída de Estatísticas e Exportação de Estatísticas contêm uma guia Filtro para facilitar esse processo.

Uma descrição básica de funcionalidade da guia Filtro é discutida em outro local. Consulte o tópico “Configurando Opções de Filtragem” na página 145 para obter mais informações. Este tópico fornece dicas para leitura de dados no IBM SPSS Statistics.

Para ajustar os nomes de campo de acordo com as convenções de nomenclatura do IBM SPSS Statistics:

1. Na guia Filtro, clique no botão da barra de ferramentas Menu de Opções de Filtro (o primeiro na barra de ferramentas).
2. Selecione Renomear Para IBM SPSS Statistics.
3. No diálogo Renomear Para IBM SPSS Statistics, é possível escolher substituir caracteres inválidos nos nomes de arquivos por um caractere **Hash (#)** ou por um **Sublinhado (_)**.

Renomear conjuntos de múltiplas respostas. Selecione esta opção se desejar ajustar os nomes de conjuntos de múltiplas respostas, que podem ser importados no IBM SPSS Modeler usando um nó de origem Arquivo de Estatísticas. Eles são usados para registrar dados que podem ter mais de um valor para cada caso, como em respostas de pesquisa de opinião.

Capítulo 9. SuperNodes

Visão Geral de SuperNodes

Uma das razões pelas quais a interface de programação visual do IBM SPSS Modeler é muito fácil de aprender é que cada nó possui uma função claramente definida. No entanto, para processamento complexo, uma longa sequência de nós poderá ser necessária. Eventualmente, isso pode bagunçar a tela de fluxo e dificultar o acompanhamento dos diagramas de fluxo. Há duas maneiras para evitar a desordem de um fluxo longo e complexo:

- É possível dividir uma sequência de processamento em vários fluxos que alimentam um no outro. O primeiro fluxo, por exemplo, cria um arquivo de dados que o segundo fluxo utiliza como entrada. O segundo fluxo cria um arquivo que o terceiro utiliza como entrada, e assim por diante. É possível gerenciar esses diversos fluxos ao salvá-los em um **projeto**. Um projeto fornece organização para diversos fluxos e suas saídas. No entanto, um arquivo de projeto contém somente uma referência aos objetos que ele contém e ainda haverá diversos arquivos de fluxo para gerenciar.
- É possível criar um **SuperNode** como uma alternativa mais aperfeiçoada quando trabalhar com processos de fluxo complexos.

Os SuperNodes agrupam diversos nós em um único nó ao encapsular seções de um fluxo de dados. Isso fornece inúmeros benefícios para o minerador de dados:

- Os fluxos são mais limpos e mais gerenciáveis.
- Os nós podem ser combinados em um SuperNode específico de negócios.
- Os SuperNodes podem ser exportados em bibliotecas para reutilização em diversos projetos de mineração de dados.

Tipos de SuperNodes

Os SuperNodes são representados no fluxo de dados por um ícone de estrela. O ícone é sombreado para representar o tipo de SuperNode e a direção na qual o fluxo deve fluir para ou a partir dele.

Há três tipos de SuperNodes:

- SuperNodes de Origem
- SuperNodes de Processo
- SuperNodes de Terminal

SuperNodes de Origem

Os SuperNodes de origem contêm uma origem de dados tal como um nó de origem normal e podem ser utilizados em qualquer lugar em que um nó de origem normal possa ser utilizado. O lado esquerdo de um SuperNode de origem é sombreado para indicar que ele está "fechado" à esquerda e que os dados devem fluir de modo que sejam recebidos *em* um SuperNode.

Os SuperNodes de origem possuem apenas um ponto de conexão à direita, mostrando que os dados saem do SuperNode e vão para o fluxo.

SuperNodes de Processo

Os SuperNodes de Processo contêm apenas nós de processo e têm o sombreado removido para mostrar que os dados podem fluir na *entrada* e na *saída* deste tipo de SuperNode.

Os SuperNodes de Processo possuem pontos de conexão nos lados esquerdo e direito, mostrando que os dados entram no SuperNode e saem para retornar ao fluxo. Embora os SuperNodes possam conter

fragmentos de fluxo adicionais e até mesmo fluxos extras, os dois pontos de conexão devem fluir por um caminho único conectando os pontos *Do Fluxo* e *Para Fluxo*.

Nota: às vezes os SuperNodes de processo também são referidos como *SuperNodes de Manutenção*.

SuperNodes Terminais

Os SuperNodes Terminais contêm um ou mais nós terminais (gráfico, tabela, e assim por diante) e podem ser utilizados da mesma maneira que um nó terminal. Um SuperNode terminal é sombreado no lado direito para indicar que ele está "fechado" à direita e que os dados podem fluir apenas *para* um SuperNode terminal.

Os SuperNodes Terminais possuem apenas um ponto de conexão à esquerda, mostrando que os dados entram no SuperNode a partir do fluxo e terminam dentro do SuperNode.

Os SuperNodes Terminais também podem conter scripts que são utilizados para especificar a ordem de execução de todos os nós terminais dentro do SuperNode. Consulte o tópico "SuperNodes e Script" na página 385 para obter mais informações.

Criando SuperNodes

A criação de um SuperNode "reduz" o fluxo de dados ao encapsular vários nós em um nó. Após criar ou carregar um fluxo na tela, existem várias maneiras de criar um SuperNode.

Seleção múltipla

A maneira mais simples de criar um SuperNode é selecionar todos os nós que desejar encapsular:

1. Utilize o mouse para selecionar diversos nós na tela de fluxo. Também é possível utilizar Shift-clique para selecionar um fluxo ou uma seção de um fluxo. *Nota:* os nós que forem selecionados devem ser de um fluxo contínuo ou bifurcado. Não é possível selecionar nós que não forem adjacentes ou que não estiverem conectados de alguma forma.
2. Em seguida, utilizando um dos três métodos a seguir, encapsule os nós selecionados:
 - Clique no ícone SuperNode (em forma de uma estrela) na barra de ferramentas.
 - Clique com o botão direito no SuperNode e, no menu de contexto, escolha:
Criar SuperNode > Da Seleção
 - No menu SuperNode, escolha:
Criar SuperNode > Da Seleção

Todas essas três opções encapsulam os nós em um SuperNode sombreado para refletir seu tipo - origem, processo ou terminal - com base em seu conteúdo.

Seleção Única

Também é possível criar um SuperNode ao selecionar um nó único e utilizar as opções de menu para determinar o início e o término do SuperNode ou encapsular tudo junto do nó selecionado.

1. Clique no nó que determina o início de encapsulamento.
2. No menu SuperNode, escolha:
Criar SuperNode > A Partir Daqui

Os SuperNodes também podem ser criados mais interativamente ao selecionar o início e o término da seção de fluxo para encapsular nós:

1. Clique no primeiro ou no último nó que deseja incluir no SuperNode.
2. No menu SuperNode, escolha:

Criar SuperNode > Selecionar..

3. Como alternativa, é possível utilizar as opções de menu de contexto ao clicar com o botão direito no nó desejado.
4. O cursor se transforma em um ícone do SuperNode, indicando que outro ponto deve ser selecionado no fluxo. Mova para cima ou para baixo até a "outra extremidade" do fragmento do SuperNode e clique em um nó. Esta ação substituirá todos os nós que estiverem no meio pelo ícone de estrela do SuperNode.

Nota: os nós que forem selecionados devem ser de um fluxo contínuo ou bifurcado. Não é possível selecionar nós que não forem adjacentes ou que não estiverem conectados de alguma forma.

Aninhando SuperNodes

Os SuperNodes podem ser aninhados dentro de outros SuperNodes. As mesmas regras para cada tipo de SuperNode (origem, processo e terminal) se aplicam aos SuperNodes aninhados. Por exemplo, um SuperNode de processo com aninhamento deve ter um fluxo de dados contínuo através de todos os SuperNodes aninhados para que ele permaneça um processo SuperNode. Se um dos SuperNodes aninhados for terminal, então os dados não fluirão mais pela hierarquia.

Embora os SuperNodes de terminal e de origem possam conter outros tipos de SuperNodes aninhados, as mesmas regras básicas para criação de SuperNodes se aplicam.

Bloqueando SuperNodes

Após ter criado um SuperNode, será possível bloqueá-lo com uma senha para evitar que ele seja alterado. Por exemplo, isso poderá ser feito se estiver criando fluxos, ou partes de fluxos, como modelos de valor fixo para uso por outras pessoas em sua organização que tiverem pouca experiência com a configuração de investigações do IBM SPSS Modeler.

Quando um SuperNode é bloqueado, os usuários ainda podem inserir valores na guia Parâmetros para todos os parâmetros que tiverem sido definidos, e um SuperNode bloqueado poderá ser executado sem inserir a senha.

Nota: o bloqueio e o desbloqueio não podem ser executados utilizando scripts.

Bloqueando e Desbloqueando um SuperNode

Aviso: senhas perdidas não podem ser recuperadas.

É possível bloquear ou desbloquear um SuperNode a partir de qualquer uma das três guias.

1. Clique em **Bloquear Nó**.
2. Insira e confirme a senha.
3. Clique em **OK**.

Um SuperNode protegido por senha é identificado na tela de fluxo por um símbolo de cadeado pequeno no canto superior esquerdo do ícone SuperNode.

Desbloquear um SuperNode

1. Para remover permanentemente a proteção de senha, clique em **Desbloquear Nó**; será solicitada a senha.
2. Insira a senha e clique em **OK**; o SuperNode não é mais protegido por senha e o símbolo de cadeado não é mais exibido ao lado do ícone no fluxo.

Editando um SuperNode Bloqueado

Se tentar definir parâmetros ou aumentar o zoom para exibir um SuperNode bloqueado, será solicitado a inserir a senha.

Insira a senha e clique em **OK**.

Agora é possível editar as definições de parâmetro e aumentar ou diminuir o zoom sempre que precisar até fechar o fluxo no qual o SuperNode se encontra.

Observe que isso não remove a proteção de senha, somente permite acesso para trabalhar com o SuperNode. Consulte o tópico “Bloqueando e Desbloqueando um SuperNode” na página 381 para obter mais informações.

Editando SuperNodes

Após criar um SuperNode, ele poderá ser examinado mais de perto ao aumentar o zoom nele; se o SuperNode estiver bloqueado, será solicitado a inserir a senha. Consulte o tópico “Editando um SuperNode Bloqueado” para obter mais informações.

Para visualizar o conteúdo de um SuperNode, é possível utilizar o ícone de aumentar o zoom a partir da barra de ferramentas do IBM SPSS Modeler ou o seguinte método:

1. Clique com o botão direito em um SuperNode.
2. No menu de contexto, escolha **Aumentar Zoom**.

O conteúdo do SuperNode selecionado será exibido em um ambiente do IBM SPSS Modeler um pouco diferente, com conectores mostrando o fluxo de dados por meio do fluxo ou do fragmento de fluxo. Neste nível na tela de fluxo, há várias tarefas que podem ser executadas:

- Modificar o tipo de SuperNode-origem, processo ou terminal.
- Criar parâmetros ou editar os valores de um parâmetro. Os parâmetros são utilizados em script e em expressões do CLEM.
- Especificar as opções de armazenamento em cache para o SuperNode e seus subnós.
- Criar ou modificar um script de SuperNode (SuperNodes de terminal somente).

Modificando Tipos de SuperNode

Em algumas circunstâncias, é útil alterar o tipo de um SuperNode. Essa opção está disponível apenas quando estiver com zoom aumentado em um SuperNode e se aplica somente ao SuperNode nesse nível. Os três tipos de SuperNodes são explicados na tabela a seguir.

Tabela 62. Tipos de SuperNode.

Tipo de SuperNode	Descrição
SuperNode de Origem	Uma conexão saindo
SuperNode de Processo	Duas conexões: uma chegando e outra saindo
SuperNode de Terminal	Uma conexão chegando

Para Alterar o Tipo de um SuperNode

1. Assegure-se de estar com zoom aumentado no SuperNode.
2. No menu SuperNode, escolha **Tipo de SuperNode** e, em seguida, escolha o tipo.

Anotando e Renomeando SuperNodes

É possível renomear um SuperNode conforme ele aparece no fluxo, bem como escrever anotações utilizadas em um projeto ou relatório. Para acessar essas propriedades:

- Clique com o botão direito em um SuperNode (zoom diminuído) e escolha **Renomear e Anotar**.
- Como alternativa, no menu SuperNode, escolha **Renomear e Anotar**. Essa opção está disponível nos modos de zoom aumentado e zoom diminuído.

Em ambos os casos, uma caixa de diálogo é aberta com a guia Anotações selecionada. Utilize as opções aqui para customizar o nome exibido na tela de fluxo e fornecer documentação referente às operações de SuperNode.

Utilizando Comentários com SuperNodes

Se criar um SuperNode a partir de um nó ou nugget comentado, deve-se incluir o comentário na seleção para criar o SuperNode se desejar que o comentário apareça no SuperNode. Se omitir o comentário da seleção, o comentário permanecerá no fluxo quando o SuperNode for criado.

Ao expandir um SuperNode que inclui comentários, os comentários são reestabelecidos ao ponto em que estavam antes de o SuperNode ter sido criado.

Ao expandir um SuperNode que inclui objetos comentados, mas os comentários não foram incluídos no SuperNode, os objetos são reestabelecidos ao ponto em que estavam, mas os comentários não são reanexados.

Parâmetros de SuperNode

No IBM SPSS Modeler, é possível configurar variáveis definidas pelo usuário, como *Minvalue*, cujos valores podem ser especificados quando usados em script ou em expressões do CLEM. Essas variáveis são chamadas **parâmetros**. É possível configurar parâmetros para fluxos, sessões e SuperNodes. Quaisquer parâmetros configurados para um SuperNode estão disponíveis ao construir expressões do CLEM nesse SuperNode ou em quaisquer nós aninhados. Parâmetros configurados para SuperNodes aninhados não estão disponíveis para seu SuperNode pai.

Há dois passos para criar e configurar parâmetros para SuperNodes:

1. Defina parâmetros para o SuperNode.
2. Em seguida, especifique o valor para cada parâmetro do SuperNode.

Esses parâmetros podem, então, ser utilizados em expressões do CLEM para quaisquer nós encapsulados.

Definindo Parâmetros de SuperNode

Os parâmetros para um SuperNode podem ser definidos nos modos zoom diminuído e zoom aumentado. Os parâmetros definidos se aplicam a todos os nós encapsulados. Para definir os parâmetros de um SuperNode, primeiro é necessário acessar a guia Parâmetros da caixa de diálogo SuperNode. Utilize um dos seguintes métodos para abrir a caixa de diálogo:

- Dê um clique duplo em um SuperNode no fluxo.
- No menu SuperNode, escolha **Configurar Parâmetros**.
- Como alternativa, quando tiver o zoom aumentado em um SuperNode, escolha **Configurar Parâmetros** no menu de contexto.

Após ter aberto a caixa de diálogo, a guia Parâmetros será exibida com quaisquer parâmetros definidos anteriormente.

Para Definir um Novo Parâmetro

Clique no botão **Definir Parâmetros** para abrir a caixa de diálogo.

Nome. Nomes de parâmetros são listados aqui. É possível criar um novo parâmetro inserindo um nome nesse campo. Por exemplo, para criar um parâmetro para temperatura mínima, você poderia digitar

minvalue. Não inclua o prefixo \$P- que denota um parâmetro nas expressões do CLEM. Esse nome também é usado para exibição no construtor de expressões do CLEM.

Nome longo. Lista o nome descritivo para cada parâmetro criado.

Armazenamento. Selecione um tipo de armazenamento da lista. Armazenamento indica como valores de dados são armazenados no parâmetro. Por exemplo, quando você trabalha com valores contendo zeros à esquerda que deseja preservar (como 008), é necessário selecionar **Sequência de Caracteres** como o tipo de armazenamento. Caso contrário, os zeros serão removidos do valor. Os tipos de armazenamento disponíveis são sequência de caracteres, número inteiro, real, hora, data e registro de data e hora. Para parâmetros de data, observe que os valores devem ser especificados usando uma notação padrão ISO, conforme mostrado no parágrafo seguinte.

Valor. Lista o valor atual de cada parâmetro. Ajuste o parâmetro conforme necessário. Observe que para parâmetros de data, os valores devem ser especificados na notação padrão ISO (ou seja, YYYY-MM-DD). Datas especificadas em outros formatos não são aceitas.

Tipo (opcional). Se você pretende implementar o fluxo em um aplicativo externo, selecione um nível de medição da lista. Caso contrário, é recomendável deixar a coluna *Tipo* no estado em que se encontra. Se desejar especificar restrições de valor para o parâmetro, como limites mínimo e máximo para um intervalo numérico, selecione **Especificar** na lista.

Observe que as opções nome longo, armazenamento e tipo podem ser configuradas para parâmetros somente por meio da interface com o usuário. Essas opções não podem ser configuradas usando scripts.

Clique nas setas à direita para mover o parâmetro selecionado mais para cima ou para baixo na lista de parâmetros disponíveis. Use o botão Excluir (marcado com um X) para remover o parâmetro selecionado.

Configurando Valores para Parâmetros do SuperNode

Após ter definido os parâmetros para um SuperNode, será possível especificar valores utilizando os parâmetros em uma expressão ou em um script do CLEM.

Para Especificar Parâmetros de um SuperNode

1. Dê um clique duplo no ícone do SuperNode para abrir a caixa de diálogo SuperNode.
2. Como alternativa, no menu SuperNode, escolha **Configurar Parâmetros**.
3. Clique na guia **Parâmetros**. *Nota:* os campos nessa caixa de diálogo são os campos definidos ao clicar no botão **Definir Parâmetros** nesta guia.
4. Insira um valor na caixa de texto para cada parâmetro que você criou. Por exemplo, é possível configurar o valor *minvalue* para um limite específico de interesse. Em seguida, este parâmetro pode ser utilizado em várias operações, como selecionar registros acima ou abaixo deste limite para exploração adicional.

Usando Parâmetros do SuperNode para Acessar Propriedades do Nó

Os parâmetros do SuperNode também podem ser utilizados para definir propriedades do nó (também conhecidas como **parâmetros de slot**) para nós encapsulados. Por exemplo, suponha que você deseja especificar que um SuperNode treine uma Rede Neural encapsulada por um determinado período de tempo utilizando uma amostra aleatória dos dados disponíveis. Utilizando parâmetros, é possível especificar valores para o período de tempo e amostra de porcentagem.

Suponha que seu SuperNode de exemplo contenha um nó Amostra chamado *Amostra* e um nó Rede Neural chamado *Treinar*. É possível utilizar as caixas de diálogo do nó para especificar a configuração de **Amostra** do nó Amostra como % **Aleatória** e a configuração **Parar em** do nó Rede Neural para **Tempo**. Após essas opções serem especificadas, será possível acessar as propriedades do nó com os parâmetros e especificar determinados valores para o SuperNode. Na caixa de diálogo SuperNode, clique em **Definir Parâmetros** e crie os parâmetros mostrados na seguinte tabela.

Tabela 63. Parâmetros a serem criados

Parâmetro	Valor	Nome longo
Train.time	5	Tempo para treinar (minutos)
Sample.random	10	Porcentagem de amostra aleatória

Nota: os nomes de parâmetros, como *Sample.random*, utilizam a sintaxe correta para referenciar as propriedades do nó, em que *Sample* representa o nome do nó e *random* é uma propriedade do nó.

Após ter definido esses parâmetros, será possível modificar facilmente os valores para as duas propriedades do nó Amostra e Rede Neural sem reabrir cada caixa de diálogo. Ao invés disso, basta selecionar **Configurar Parâmetros** no menu SuperNode para acessar a guia Parâmetros da caixa de diálogo SuperNode, na qual é possível especificar novos valores para % **Aleatória** e **Tempo**. Isto é útil particularmente quando explorar os dados durante diversas iterações de construção de modelo.

SuperNodes e Armazenamento em Cache

A partir de um SuperNode, todos os nós, exceto os nós terminais, podem ser armazenados em cache. O armazenamento em cache é controlado ao clicar com o botão direito em um nó e escolher uma das várias opções no menu de contexto do Cache. Essa opção de menu está disponível tanto fora de um SuperNode quanto para os nós encapsulados em um SuperNode.

Há várias diretrizes para caches do SuperNode:

- Se algum dos nós encapsulados em um SuperNode tiver o armazenamento em cache ativado, o SuperNode também terá.
- Desativar o cache em um SuperNode desativa o cache para *todos* os nós encapsulados.
- Na realidade, ativar o armazenamento em cache em um SuperNode ativa o cache no último subnó armazenável em cache. Em outras palavras, se o último subnó for um nó Seleção, o cache será ativado para esse nó Seleção. Se o último subnó for um nó terminal (que não permite armazenamento em cache), o próximo nó acima que suportar armazenamento em cache será ativado.
- Após ter configurado caches para os subnós de um SuperNode, quaisquer atividades anteriores do nó em cache, como incluir ou editar nós, esvaziarão os caches.

SuperNodes e Script

É possível utilizar a linguagem de script do IBM SPSS Modeler para escrever programas simples que manipulam e executam o conteúdo de um SuperNode de terminal. Por exemplo, talvez você queira especificar a ordem de execução para um fluxo complexo. Como um exemplo, se um SuperNode contiver um nó Configurar Globais que precisa ser executado antes de um nó Gráfico, será possível criar um script que executa o nó Configurar Globais primeiro. Em seguida, os valores calculados por este nó, como o desvio médio ou padrão, poderão ser utilizados quando o nó Gráfico for executado.

A guia Script da caixa de diálogo SuperNode está disponível apenas para os SuperNodes de terminal.

Para Abrir a Caixa de Diálogo Script para um SuperNode de Terminal

- Clique com o botão direito na tela do SuperNode e escolha **Script do SuperNode**.
- Como alternativa, nos modos zoom aumentado e zoom diminuído, é possível escolher **Script do SuperNode** a partir do menu SuperNode.

Nota: os scripts de SuperNode são executados apenas com o fluxo e o SuperNode quando tiver selecionado **Executar este script** na caixa de diálogo.

Opções específicas para script e seu uso no IBM SPSS Modeler são discutidas no *Guia de Script e Automação*, que está disponível no IBM SPSS Modeler DVD.

Salvando e Carregando SuperNodes

Uma das vantagens dos SuperNodes é que eles podem ser salvos e reutilizados em outros fluxos. Ao salvar e carregar SuperNodes, observe que eles usam uma extensão *.slb*.

Para Salvar um SuperNode

1. Aumentar Zoom no SuperNode.
2. No menu SuperNode, escolha **Salvar SuperNode**.
3. Especifique um nome de arquivo e um diretório na caixa de diálogo.
4. Selecione se deseja incluir o SuperNode salvo no projeto atual.
5. Clique em **Salvar**.

Para Carregar um SuperNode

1. No menu Inserir na janela do IBM SPSS Modeler, escolha **SuperNode**.
2. Selecione um arquivo de SuperNode (*.slb*) no diretório atual ou procure por um diferente.
3. Clique em **Carregar**.

Nota: os SuperNodes importados possuem os valores padrão para todos os seus parâmetros. Para alterar os parâmetros, dê um clique duplo em um SuperNode na tela de fluxo.

Avisos

Estas informações foram desenvolvidas para produtos e serviços oferecidos no mundo todo.

A IBM pode não oferecer os produtos, serviços ou recursos discutidos neste documento em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser utilizados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM pode ser usado em substituição a este produto, programa ou serviço. Entretanto a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes pendentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
Botafogo
Rio de Janeiro, RJ
CEP 22290-240

Para pedidos de licença relacionados a informações de DBCS (Conjunto de Caracteres de Byte Duplo), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

O parágrafo a seguir não se aplica ao Reino Unido nem a nenhum país em que tais disposições não estejam de acordo com a legislação local: A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA" SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS A ELAS NÃO SE LIMITANDO, AS GARANTIAS IMPLÍCITAS DE NÃO INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em certas transações; portanto, essa disposição pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas mudanças periódicas nas informações aqui contidas; tais mudanças serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

As referências nestas informações à documentação ou a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a esses documentos ou websites. Os materiais contidos nesses websites não fazem parte dos materiais desse produto IBM e o uso desses websites é de inteira responsabilidade do Cliente.

A IBM pode usar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Os licenciados deste programa que desejarem obter informações sobre este assunto com o propósito de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) o uso mútuo das informações trocadas, deverão entrar em contato com:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
Botafogo
Rio de Janeiro, RJ
CEP 22290-240

Tais informações podem estar disponíveis, sujeitas aos termos e condições apropriados, incluindo, em alguns casos, o pagamento de uma taxa.

O programa licenciado descrito nesta publicação e todo o material licenciado disponível são fornecidos pela IBM sob os termos do Contrato com o Cliente IBM, do Contrato Internacional de Licença do Programa IBM ou de qualquer outro contrato equivalente.

Todos os dados de desempenho aqui contidos foram determinados em um ambiente controlado. Portanto, os resultados obtidos em outros ambientes operacionais podem variar significativamente. Algumas medidas podem ter sido tomadas em sistemas em nível de desenvolvimento e não há garantia de que estas medidas serão iguais em sistemas geralmente disponíveis. Além disso, algumas medidas podem ter sido estimadas por extrapolação. Os resultados reais podem variar. Os usuários deste documento devem verificar os dados aplicáveis para seu ambiente específico.

As informações relativas a produtos não IBM foram obtidas junto aos fornecedores dos respectivos produtos, a partir de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou estes produtos e não pode confirmar a precisão de seu desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não IBM. As dúvidas sobre os recursos de produtos não IBM devem ser encaminhadas diretamente aos seus fornecedores.

Todas as declarações relacionadas aos objetivos e intenções futuras da IBM estão sujeitas a mudanças ou cancelamento sem aviso prévio e representam apenas metas e objetivos.

Estas informações contêm exemplos de dados e relatórios usados nas operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de indivíduos, empresas, marcas e produtos. Todos estes nomes são fictícios e qualquer semelhança com nomes e endereços usados por uma empresa real é mera coincidência.

Se estas informações estiverem sendo exibidas em formato eletrônico, as fotografias e ilustrações coloridas podem não aparecer.

Marcas comerciais

IBM, o logotipo IBM e ibm.com são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em vários países no mundo todo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas comerciais da IBM está disponível na web em "Copyright and trademark information", em www.ibm.com/legal/copytrade.shtml.

Intel, o logotipo Intel, Intel Inside, o logotipo Intel Inside, Intel Centrino, o logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas comerciais ou marcas registradas da Intel Corporation ou suas subsidiárias nos Estados Unidos e em outros países.

Linux é uma marca registrada de Linus Torvalds nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou em outros países.

UNIX é uma marca registrada do The Open Group nos Estados Unidos e/ou em outros países.

Java e todas as marcas comerciais e logotipos baseados em Java são marcas comerciais ou marcas registradas da Oracle e/ou de seus afiliados.

Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresa.

Glossário

C

Covariância. Uma medida de associação não padronizada entre duas variáveis, igual ao desvio de produto vetorial dividido por $N-1$.

K

Curtose. Uma medida da extensão até a qual as observações se agrupam em torno de um ponto central. Para a uma distribuição normal, o valor da estatística de curtose é zero. A curtose positiva indica que, com relação a uma distribuição normal, as observações são agrupadas mais ao centro da distribuição e têm rastros mais finos até os valores extremos da distribuição, no ponto em que os rastros da distribuição leptocúrtica são mais espessos com relação a uma distribuição normal. Já a curtose negativa indica que, com relação a uma distribuição normal, as observações são menos agrupadas e possuem rodapés mais espessos até os valores extremos da distribuição, no ponto em que os rastros da distribuição platykúrtica são mais finos com relação a uma distribuição normal.

M

Máximo. O maior valor de uma variável numérica.

Média. Uma medida de tendência central. A média aritmética, a soma dividida pelo número de casos.

Median. O valor acima e abaixo do qual metade dos casos cai, o 50º percentil. Se houver um número par de casos, a mediana será a média dos dois casos intermediários quando forem classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).

Mínimo. O menor valor de uma variável numérica.

Modo. O valor que ocorre mais frequentemente. Se vários valores compartilharem a maior frequência da ocorrência, cada um deles será um modo.

R

Intervalo. A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.

S

Assimetria. Uma medida da assimetria de uma distribuição. A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

desvio padrão . Uma medida de dispersão em torno da média, igual à raiz quadrada da variância. O desvio padrão é medido nas mesmas unidades que a variável original.

Desvio padrão. Uma medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.

Erro padrão. Uma medida do quanto o valor de uma estatística de teste varia de amostra para amostra. Ela é o desvio padrão da distribuição de amostragem para uma estatística. Por exemplo, o erro padrão da média é o desvio padrão das médias da amostra.

Erro Padrão de Curtose. A razão de curtose com seu erro padrão pode ser utilizada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for inferior a -2 ou superior a +2). Um valor positivo grande para curtose indica que os rodapés da distribuição são maiores do que aqueles de uma distribuição normal, e um valor negativo para curtose indica rodapés mais curtos (como aqueles de uma distribuição uniforme em forma de caixa).

Erro padrão de média. Uma medida do quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

Erro Padrão de Assimetria. A razão de assimetria com seu erro padrão pode ser utilizada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for inferior a -2 ou superior a +2). Um valor positivo grande para assimetria indica um rodapé direito longo, e um valor negativo extremo indica um rodapé esquerdo longo.

Sum. A soma ou o total dos valores, em todos os casos com valores não omissos.

U

unique. Avalia todos os efeitos simultaneamente, ajuste cada efeito a todos os outros efeitos de qualquer tipo.

V

Válidos. Casos válidos que não possuem o valor omissos do sistema, nem um valor definido como omissos de usuário.

Variância. Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.

Índice Remissivo

A

abrindo
 objetos de saída 302
acumulando dados de séries
 temporais 177
agregando dados de série temporal 177, 190
agregando registros 172
agrupando valores 246
amostragem 1 em n 74
amostragem de dados contínuos 74
amostras de teste
 dados de particionamento 170, 171
amostras de treinamento
 balanceamento 78
 dados de particionamento 170, 171
amostras de validação
 dados de particionamento 170, 171
amostras em cluster 73, 74, 76
amostras estratificadas 73, 74, 76, 77
amostras não aleatórias 73, 74
amostras ponderadas 76
amostras sistemáticas 73, 74
animação em gráficos 194
ano fiscal
 nó Intervalos de Tempo 181
anonimizando nomes de campos 147
ANOVA
 nó Média 326
ANOVA unidirecional
 nó Média 326
antijunção 84
aplicando zoom 382
aplicativos auxiliares 341
aproximação de mediana 81
aproximação de quartil 81
armazenamento 138
 convertendo 154, 155
armazenamento de campo
 convertendo 154
arquivo de dados
 employee_data.sav 371
arquivos .dbf 64
arquivos .sav 370
arquivos .sd2 (SAS) 42
arquivos .shp 64, 65
arquivos .slb 386
arquivos .ssd (SAS) 42
arquivos .tpt (SAS) 42
arquivos .zsav 370
Arquivos DAT
 exportação 365
 exportando 304
 salvamento 306
Arquivos de dados do IBM SPSS
 Statistics
 importando dados de pesquisa de
 opinião 32
arquivos de forma de mapas
 conceitos 226
arquivos de forma de mapas (*continuação*)
 editando mapas do SMZ
 pré-instalados 225
 tipos 226
 utilizando no Seletor de Modelo de
 Gráfico 225
arquivos de formas 225
arquivos de formato 42
arquivos de mapeamento
 excluindo 224
 exportando 224
 importando 224
 localização 223
 renomeando 224
 seleccionando no Seletor de Modelo de
 Gráfico 202
arquivos de saída
 salvamento 306
arquivos de texto 26
 exportação 365
arquivos de transporte
 nó de origem SAS 42
arquivos delimitados por vírgula
 exportação 365
 exportando 304
 salvamento 306
arquivos do ESRI 225
arquivos do Excel
 exportação 365
arquivos do SMZ
 criar 225
 editando arquivos do SMZ
 pré-instalados 225
 excluindo 224
 exportando 224
 importando 224
 pré-instalados 225
 renomeando 224
 visão geral 225
arquivos simples 26
arquivos XLSX
 exportação 365
aspas
 importando arquivos de texto 27
 para exportação do banco de
 dados 344
atributos
 em mapas 226
atributos de tipo 143
atributos do campo 143
auditoria
 auditoria de dados iniciais 314
 nó de Auditoria de Dados 314
autorregressão
 modelos ARIMA 101
avaliação de modelo 262
avaliando modelos 310

B

banco de dados
 carregamento em massa 350, 352
banco de dados Quanvert
 importação 32
bancos de dados ADO
 importação 32
bancos de dados grandes 71
 executando uma auditoria de
 dados 314
Bancos de dados In2data
 importação 32
banda de query
 Teradata 23
bandas em gráficos 276
bins decis 162
bins percentis 162
bins quartis 162
bins quintis 162
bins vingtiles 162
bloqueando SuperNodes 381
boxplot 204
 exemplo 216

C

cache
 SuperNodes 385
calcular durações
 preparação de dados
 automatizada 120
cálculo de duração
 preparação de dados
 automatizada 120
camadas em mapas geoespaciais 271
Campo de contagem
 nó Intervalos de Tempo 177
 preenchendo ou agregando séries
 temporais 177
campo TimeIndex
 nó Intervalos de Tempo 177
campo TimeLabel
 nó Intervalos de Tempo 177
campos
 derivando diversos campos 149
 rótulos de campo e de valor 138
 transpondo 174
campos-chave 79, 172
campos de chave primária
 Nó de exportação do banco de
 dados 346
campos de partição 142, 170, 171
campos de rótulo
 rotulando registros de saída 142
caracteres de comentário
 em arquivos de variável 27
caracteres de controle 12
caracteres de controle hexadecimais 12
caracteres de controle não suportados 12
caracteres EOL 27

- carregamento em massa 350, 352
 - casas decimais
 - formatos de exibição 144
 - casos de ranqueamento 164
 - categorização ideal 165
 - categorização supervisionada 165
 - cenário 65
 - chaves contínuas 81
 - classificando
 - campos 188
 - campos pré-ordenados 84
 - registros 83
 - cluster 285, 294
 - codificação simulada 172
 - Cognos, consulte IBM Cognos BI 39
 - comando CREATE INDEX 349
 - combinando dados 92
 - a partir de diversos arquivos 84
 - comentários
 - utilizando com SuperNodes 382
 - como visualizar
 - saída HTML no navegador 304
 - concatenando registros 92
 - condições
 - classificado 88
 - especificando para uma mesclagem 88
 - especificando uma série 154
 - condições classificadas
 - especificando para uma mesclagem 88
 - conexões
 - para o repositório do IBM SPSS Collaboration and Deployment Services 65
 - conexões com o banco de dados
 - definindo 19
 - valores pré-configurados 21
 - configurações automáticas 286
 - configurar semente aleatória
 - registros de amostragem 171
 - conjunto de dados principal 92
 - conjuntos
 - convertendo em flags 172, 173
 - transformando 157, 158
 - conjuntos de categorias múltiplas 147
 - conjuntos de múltiplas dicotomias 147
 - conjuntos de múltiplas respostas
 - conjuntos de categorias múltiplas 147
 - conjuntos de múltiplas dicotomias 147
 - definindo 147
 - excluindo 147
 - nas visualizações 200
 - nó de origem do IBM SPSS Data Collection 32, 36
 - nó de origem do IBM SPSS Statistics 370
 - Construtor de Expressões 71
 - consultas
 - Nó de origem da base de dados 17, 18
 - consultas SQL
 - Nó de origem da base de dados 17, 18, 25
 - contagens
 - nó Categorização 162
 - saída de estatísticas 325
 - contagens iguais
 - nó Categorização 162
 - convertendo conjuntos em flags 173
 - convertendo conjuntos em sinalizações 172
 - converter níveis de medição 136
 - coordenadas polares 293
 - copiando atributos de tipo 143
 - copiando visualizações 295
 - coroplético
 - exemplo 220
 - correlações 324
 - nó Média 328
 - probabilidade 324
 - rótulos descritivos 324
 - saída de estatísticas 325
 - significância 324
 - valor absoluto 324
 - correlações de Pearson
 - nó Média 328
 - saída de estatísticas 325
 - criar
 - novos campos 148, 149
 - CRISP-DM
 - entendimento de dados 7
 - custos
 - gráficos de avaliação 266
- D**
- dados
 - agregando 79
 - armazenamento 155
 - tipo de armazenamento 138
 - dados anuais
 - nó Intervalos de Tempo 181
 - dados atrasados 187
 - dados categóricos 136
 - dados contínuos 136, 139
 - dados CSV
 - importação 32
 - dados de amostragem 77
 - dados de particionamento 170, 171
 - gráficos de avaliação 266
 - nó Análise 311
 - dados de pesquisa de mercado
 - importação 32, 36
 - nó de origem do IBM SPSS Data Collection 32, 36
 - dados de pesquisa de opinião
 - importação 32, 36
 - nó de origem do IBM SPSS Data Collection 32
 - dados de pesquisa de opinião do IBM SPSS Data Collection
 - importação 32
 - dados de resumo 79
 - dados de série temporal
 - agregando 176, 177, 189
 - construindo a partir de dados 177
 - definindo 176, 177, 178
 - identificando 176, 177, 178
 - intervalos 177
 - período de estimação 178
 - dados de série temporal (*continuação*)
 - preenchimento 176, 177
 - validações 178
 - dados de texto de campo fixo 30
 - dados de texto de campo livre 26
 - dados de texto delimitado 26
 - dados desbalanceados 78
 - dados do caso
 - nó de origem do IBM SPSS Data Collection 32
 - dados geo-espaciais 26
 - combinando 88
 - derivando 152
 - em arquivos de variável 29
 - exportando 358
 - importando 27
 - listar em arquivos de variável 29
 - mesclagem da condição classificada 88
 - Restrições 135
 - dados geoespaciais em mapas 271
 - dados imparciais 78
 - dados mensais
 - nó Intervalos de Tempo 182
 - dados nominais 140
 - dados ordinais 140
 - dados parciais 78
 - dados Quancept
 - importação 32
 - dados Quantum
 - importação 32
 - dados semanais
 - nó Intervalos de Tempo 183
 - dados simulados
 - nó Geração de Simulação 50
 - dados sintéticos
 - nó Entrada do Usuário 45
 - dados Surveycraft
 - importação 32
 - dados trimestrais
 - nó Intervalos de Tempo 181
 - dados Triple-S
 - importação 32
 - data
 - anonimizando 158
 - armazenamento 155
 - auditoria 314
 - caracteres de controle não suportados 12
 - entendendo 71
 - explorar 314
 - preparação 71
 - data/hora 133
 - datas
 - configurando formatos 144
 - de campos 350
 - definição de densidade de Space-Time-Boxes 112
 - Definição de Provedor de Dados 65
 - delimitadores 27
 - densidade
 - 3-D 204
 - densidade 3D 204
 - derivando diversos 149
 - desbloqueando SuperNodes 381
 - descartando
 - campos 145

- desempenho
 - classificando 84
 - combinando 91
 - dados de amostragem 73
- designando tipos de dados 115
- desvio médio/padrão
 - utilizado para campos de categorização 164
- desvio padrão
 - nó Configurar Globais 330
 - nó de Categorização 164
 - saída de estatísticas 325
- desvio padrão para agregação 79
- digitação automática 133, 137
- diminuindo dados 72, 73
- direção dos campos 142
- disfarçando dados para utilização em um modelo 158
- distribuição 249
- diversas entradas 84
- diversos campos
 - selecionando 150
- documentação 3
- documentos MDD
 - importação 32
- dodge 285, 294
- DPD 65
- duplicado
 - campos 84
 - fields 145
 - registros 93

E

- editando gráficos
 - tamanho dos elementos gráficos 288
- editando visualizações 285
 - categorias 291
 - combinando categorias 291
 - configurações automáticas 286
 - cores e padrões 287
 - eixos 290
 - escalas 290
 - excluindo categorias 291
 - forma de ponto 288
 - formatos numéricos 289
 - incluindo efeitos em 3D. 293
 - margens 289
 - ordenado categorias 291
 - painéis 293
 - posição da legenda 295
 - preenchimento 289
 - proporção de aspecto de ponto 288
 - reduzindo categorias 291
 - regras 286
 - rotação de ponto 288
 - seleção 286
 - text 287
 - traço 287
 - transformando o sistema de coordenadas 293
 - transparência 287
 - transpor 293
- editor de consulta
 - Nó de origem da base de dados 25
- editor de query
 - Nó de origem da base de dados 25

- elementos de marcação 279, 281
- elementos de saída 338
- elementos de tempo cíclicos
 - preparação de dados automatizada 120
- elementos gráficos
 - alteração 294
 - convertendo 294
 - modificadores de colisão 294
- em branco
 - em tabelas de Matriz 308
- empates
 - nó Categorização 162
- empilhar 285, 294
- encapsulando nós 380
- erro padrão da média
 - saída de estatísticas 325
- escoragem
 - opções do gráfico de avaliação 268
- escores de propensão
 - balanceamento de dados 78
- escores de propensão ajustada
 - balanceamento de dados 78
- esquema
 - Nó de exportação do banco de dados 346
- estatística de F
 - nó Média 327
- estatísticas
 - editando em visualizações 294
 - nó de Auditoria de Dados 314
 - nó Matriz 308
- estatísticas de avaliação de desempenho 311
- estatísticas de sumarização
 - nó de Auditoria de Dados 314
- eventos
 - criar 242
- Excel
 - ativando a partir do IBM SPSS Modeler 365
- excluindo
 - arquivos de mapeamento 224
 - folhas de estilo de visualização 224
 - modelos de visualização 224
 - objetos de saída 302
- exclusão de campo não utilizado
 - preparação de dados automatizada 119
- execução
 - especificando a ordem de 385
- exemplos
 - Guia de Aplicativos 3
 - visão geral 5
- exemplos de aplicativos 3
- explorando dados
 - nó de Auditoria de Dados 314
- explorando gráficos 275
 - bandas de gráfico 276
 - elementos de marcação 281
 - regiões 279
 - varinha de mágica 281
- exportação do Analytic Server 360
- exportando
 - arquivos de mapeamento 224
 - dados do IBM Cognos TM1 363
 - folhas de estilo de visualização 224

- exportando (*continuação*)
 - modelos de visualização 224
 - saída 304
 - SuperNodes 386
- exportando dados
 - Arquivos DAT 365
 - em um banco de dados 344
 - formato de arquivo simples 358
 - formato SAS 365
 - formato XML 366
 - geoespacial 358
 - nó do IBM Cognos BI Export 39, 361, 362
 - nó Exportação do IBM Cognos TM1 363
 - para Excel 365
 - para IBM SPSS Statistics 377
 - text 365
- exportar casas decimais 144
- expressões do CLEM 71
- extensão
 - campo derivado 149

F

- fatores de balanceamento 78
- fatores de escala 78
- fields
 - anonimizando dados 158
 - reordenando 188
 - selecionando diversos 150
- filtrando campos 90, 145
 - para o IBM SPSS Statistics 378
- folhas de estilo
 - excluindo 224
 - exportando 224
 - importando 224
 - renomeando 224
- folhas de estilo de visualização
 - aplicando 297
 - excluindo 224
 - exportando 224
 - importando 224
 - localização 223
 - renomeando 224
- forçando valores 142
- formato de armazenamento da lista 11
- formato de exibição científico 144
- formato de exibição de moeda 144
- formato HDATA
 - nó de origem do IBM SPSS Data Collection 32
- formato VDATA
 - nó de origem do IBM SPSS Data Collection 32
- formatos
 - data 9
- formatos de armazenamento 9
- formatos de exibição
 - casas decimais 144
 - científico 144
 - moeda 144
 - números 144
 - símbolo de agrupamento 144
- formatos de exibição de número 144
- formatos de hora 144
- formatos de saída 306

fórmula de derivação de campo 150
 frequências
 nó Categorização 162
 função de Máx.
 agregação de séries temporais 177, 190
 função de Média
 agregação de séries temporais 177, 190
 função de Mín.
 agregação de séries temporais 177, 190
 função de Modo
 agregação de séries temporais 177, 190
 função de Primeiro
 agregação de séries temporais 177
 função de Soma
 agregação de séries temporais 177, 190
 função de Último
 agregação de séries temporais 177
 função Em Branco
 preenchendo séries temporais 177
 função hassubstring 152
 Função mais recente
 preenchendo séries temporais 177
 funções
 especificando para campos 142
 funções de modelagem
 especificando para campos 142
 funções de transferência 102
 atraso 102
 ordens de denominador 102
 ordens de diferença 102
 ordens do numerador 102
 ordens sazonais 102

G

Geo-espacial
 configurando opções de importação 65
 gerando flags 172, 174
 gerando nós a partir de gráficos 282
 nós Balanceamento 282
 nós Derivar 282
 nós Filtro 282
 nós Reclassificar 282
 Selecionar nós 282
 gerenciador de saída 302
 gerenciadores
 guia saídas 302
 girando gráficos 3D 196
 gráfico
 tipos de gráfico 204
 gráfico de áreas 204
 3-D 204
 gráfico de áreas 3D
 descrição 204
 gráfico de barras 204
 3-D 204
 de contagens 204
 em um mapa 204
 exemplo 212, 213
 gráfico de barras 3D 204

gráfico de barras empilhadas
 exemplo 213
 gráfico de bolha 204
 gráfico de caminho 204
 gráfico de coordenadas paralelas 204
 gráfico de dispersão 204
 3-D 204
 categorizado 204
 hexa categorizado 204
 gráfico de dispersão 3D 204
 gráfico de dispersão categorizado 204
 categorias hexa 204
 gráfico de dispersão hexa
 categorizado 204
 gráfico de faixa de opções 204
 gráfico de linha 204
 em um mapa 204
 gráfico de máximos, mínimos e encerramentos 204
 gráfico de máximos e mínimos 204
 gráfico de pizza 204
 3-D 204
 em um mapa 204
 exemplo 217
 utilizando contagens 204
 gráfico de pizza 3D 204
 gráfico de pontos 204
 2D 204
 exemplo 215
 gráfico de pontos 2D 204
 gráfico de superfície 204
 gráficos
 3-D 196
 a partir do Elemento do Gráfico 198
 categorias 276
 coleções 251
 copiar 298
 distribuições 245
 esquema de cores padrão 296
 excluindo regiões 279
 explorando 275
 exportando 298
 folha de estilo 296
 gerando a partir de uma auditoria de dados 320
 gerando nós 282
 girando imagem 3D 196
 gráficos 232
 gráficos de avaliação 262
 guia anotações 196
 guias Saída 195
 histogramas 249
 imprimindo 298
 multigráficos 239
 nota de rodapé 296
 regiões 279
 rótulos do eixo 296
 salvamento 298
 salvando layouts editados 296
 salvando mudanças no layout 296
 salvando saída 306
 séries temporais 242
 tamanho dos elementos gráficos 288
 título 296
 visualização do mapa 271
 webs 254
 gráficos 3D 196

gráficos de associação 254
 gráficos de dispersão 232, 239
 gráficos de elevação 262, 269
 gráficos de ganhos 262, 269
 gráficos de linha 232, 239
 gráficos de lucro 262, 269
 gráficos de ponto 232, 239
 gráficos de resposta 262, 269
 graus de liberdade
 nó Matriz 309
 nó Média 327, 328
 guia sintaxe
 nó de Saída de Estatísticas 375

H

heat map 204
 exemplo 218
 histograma 204
 3-D 204
 exemplo 214
 histograma 3D 204
 HTML
 salvando saída 306

I

IBM SPSS Modeler 1
 documentação 3
 IBM SPSS Modeler Server 1
 IBM SPSS Statistics
 ativando a partir do IBM SPSS Modeler 341, 374, 377
 localização de licença 341
 nomes de campo válidos 378
 ícones, IBM Cognos BI 37
 idioma
 nó de origem do IBM SPSS Data Collection 35
 importação
 dados do IBM Cognos BI 38
 relatórios do IBM Cognos BI 39
 importância
 comparando médias 327
 nó Média 327, 328
 importando
 arquivos de mapeamento 224
 dados do IBM Cognos TM1 41
 folhas de estilo de visualização 224
 modelos de visualização 224
 SuperNodes 386
 incluindo
 registros 79
 incrementos de minuto
 nó Intervalos de Tempo 185
 incrementos de segundo
 nó Intervalos de Tempo 186
 indexação de tabelas de banco de dados 349
 índices BITMAP
 tabelas de banco de dados 349
 instanciação 133, 137
 nó de origem 69
 instruções if-then-else 154
 integração
 modelos ARIMA 101

- intervalo
 - saída de estatísticas 325
- intervalos 133
 - dados de série temporal 176, 189
 - valores omissos 138
- intervalos de células
 - arquivos do Excel 42
- intervalos de confiança
 - nó Média 327, 328
- intervalos de número inteiro 139
- intervalos reais 139
- intervenções
 - criar 242

J

- jitter 285, 294
- jittering 236
- junção externa 84
- junção interna 84
- junções 84, 86
 - externa parcial 87
- junções parciais 84, 87

L

- ladrilhos
 - nó Categorização 162
- layout de rede para gráficos da web 256
- Layout direcionado para gráficos da web 256
- legenda
 - posição 295
- ligação entre colunas 350
- ligação entre linhas 350
- ligações
 - nó Web 256
- limites
 - visualizando limites de categoria 166
- LineString 135
- linha de base
 - opções do gráfico de avaliação 266
- linhas em branco
 - arquivos do Excel 42
- lista 11, 133
 - comprimento máximo 138
 - depth 11
 - derivando 152
 - níveis de medição geoespaciais 135
- listar em arquivos de variável 29
- listar tipo de armazenamento 29
- listas
 - tipo de dado de coleção 141
 - tipo de dados geoespaciais 141

M

- manipulação em branco 138
 - nó de Categorização 161
 - preenchendo valores 155
- manipulando valores omissos 115
- mapa
 - com gráficos de barras 204
 - com gráficos de linha 204
 - com gráficos de pizza 204
 - com pontos 204

- mapa (*continuação*)
 - com setas 204
 - cor 204
 - sobreposição 204
- mapa coroplético 204
- mapa de coordenadas 204
- mapa de cores 204
 - exemplo 220
- mapa de fluxo 204
- mapa de sobreposição 204
- mapas
 - convertendo arquivos de forma do ESRI 225
 - diminuindo 227, 228
 - distribuindo 232
 - excluindo elementos individuais 231
 - excluindo variáveis 231
 - mesclando variáveis 230
 - movendo variáveis 230
 - projeção 231
 - rótulos de variáveis 229
 - suavizando 227, 228
- mapeamento
 - dados para exportação para o IBM Cognos TM1 364
- mapeando campos 345
- matriz de coincidência
 - nó Análise 311
- matriz de gráfico de dispersão
 - exemplo 219, 221
- matriz de gráfico de dispersão (SPLOM) 204
- máximo
 - nó Configurar Globais 330
 - saída de estatísticas 325
- média
 - nó Configurar Globais 330
 - nó de Categorização 164
 - saída de estatísticas 325
- Média de função mais recente
 - preenchendo séries temporais 177
- média móvel
 - modelos ARIMA 101
- mediana
 - saída de estatísticas 325
- médias
 - comparando 326, 327
- medições diárias
 - nó Intervalos de Tempo 183, 184
- medições por hora
 - nó Intervalos de Tempo 184
- melhor linha
 - opções do gráfico de avaliação 266
- membro (importação do SAS)
 - definição 42
- metadados 138
- Metadados
 - nó de origem do IBM SPSS Data Collection 32
- método de chave 84
- mínimo
 - nó Configurar Globais 330
 - saída de estatísticas 325
- modelador especialista
 - critérios nos modelos de série temporal 100

- modelo de processo CRISP-DM
 - preparação de dados 115
- modelos
 - anonimizando dados para 158
 - ARIMA 101
 - excluindo 224
 - exportando 224
 - importando 224
 - nó Relatório 329
 - renomeando 224
- modelos ARIMA
 - constante 101
 - critérios nos modelos de série temporal 101
 - funções de transferência 102
 - ordens autorregressivas 101
 - ordens de diferenciação 101
 - ordens de média móvel 101
 - ordens sazonais 101
 - valores discrepantes 104
- modelos causais temporais 108
 - nó TCM de fluxo 105
- modelos de série temporal
 - critérios ARIMA 101
 - critérios de suavização
 - exponencial 100
 - funções de transferência 102
 - periodicidade 102
 - transformação de séries 102
 - valores discrepantes 104
- modelos de séries temporais
 - critérios do modelador
 - especialista 100
- modelos de visualização
 - excluindo 224
 - exportando 224
 - importando 224
 - localização 223
 - renomeando 224
- modelos do IBM SPSS Statistics 374
 - detalhes do nugget avançados 374
 - nugget do modelo 374
 - opções de modelo 374
 - sobre 374
- modificadores de colisão 294
- modificando valores de dados 148
- modo
 - saída de estatísticas 325
- MultiLineString 135
- MultiPoint 135
- MultiPolygon 135

N

- navegação 336
- navegador de análise
 - interpretando 312
- navegador de auditoria de dados
 - gerando gráficos 320
 - gerando nós 320
 - Menu Arquivo 315
 - Menu Editar 315
- navegador de estatísticas
 - gerando nós de filtro 325
 - interpretando 325
 - menu Gerar 325

- navegador de matriz
 - menu Gerar 309
- navegador de qualidade
 - gerando nós de filtro 319
 - gerando nós de seleção 320
- navegador de relatório 330
- navegador de tabela
 - menu Gerar 307
 - procurando 307
 - reordenando colunas 305, 307
 - selecionando células 305, 307
- níveis de medição geoespaciais 11, 133, 135
- nível de medição
 - alterando em visualizações 198
 - coleção 11, 141, 151
 - definidos 133
 - geoespacial 11, 135, 141, 151
 - nas visualizações 200
 - restrições em dados geoespaciais 135
- nível de medição de coleção 141, 151
- nível de medição geoespacial 141, 151
- nó Agregado
 - aproximação para mediana 81
 - aproximação para quartil 81
 - configurações de otimização 81
 - configurando opções 79
 - performance 79
 - processamento paralelo 79
- nó Agregado de RFM
 - opções de configuração 82
- nó Agregar
 - visão geral 79
- nó Ajuste de simulação
 - ajuste de distribuição 331
 - guia configurações 333
- nó Ajuste de Simulação 331
 - ajuste de distribuição 331
 - configurações de saída 333
- nó Amostra
 - amostras aleatórias 73, 74
 - amostras em cluster 73, 74, 76
 - amostras estratificadas 73, 74, 76, 77
 - amostras não aleatórias 73, 74
 - amostras ponderadas 76
 - amostras sistemáticas 73, 74
 - quadro de amostragem 73
 - Tamanhos de Amostra para Camadas 77
- nó Análise 310
 - guia análise 311
 - guia saída 306
- nó Análise de RFM
 - configurações 167
 - valores de categorização 168
 - visão geral 166
- nó Anexação
 - campos de tag 90
 - correspondência de campo 92
 - opções de configuração 92
 - Visão Geral do NIS 92
- nó Anonimizado
 - criando valores anonimizados 160
 - opções de configuração 159
 - Visão Geral do NIS 158
- nó Arquivo de Estatísticas 370
- nó Arquivo Fixo
 - configurando opções 30
 - reconhecimento automático de data 30
 - Visão Geral do NIS 30
- nó Arquivo Variável 26
 - importando dados geoespaciais 29
 - metadados geoespaciais 29
 - opções de configuração 27
 - reconhecimento automático de data 27
- nó Avaliação
 - guia aparência 269, 275
- nó Avaliação de Simulação 333, 336, 338, 340
 - configurações de saída 334
- nó Balanceamento
 - configurando opções 78
 - gerando a partir de gráficos 282
 - Visão Geral do NIS 78
- nó Coleção 251
 - guia aparência 252
 - guia de opções 251, 252
 - usando o gráfico 253
- nó Combinação
 - campos de saída 168
 - combinando escores 168
- nó Configurar Globais 330
 - guia configurações 330
- nó Configurar para Sinalização 172
- nó Configurar para Sinalizador 172
- nó da Visualização Corporativa 65
- nó de Auditoria de Dados 314
 - guia configurações 314
 - guia saída 306
- nó de Avaliação 262
 - condição de ocorrência 268
 - expressão de escore 268
 - guia de opções 268
 - guia gráfico 266
 - lendo resultados 269
 - regra de negócios 268
 - usando o gráfico 270
- Nó de avaliação de Simulação
 - guia configurações 334
- nó de Categorização
 - categorias de desvio médio/padrão 164
 - categorias de largura fixa 162
 - configurando opções 161
 - ideal 165
 - pré-visualizando categorias 166
 - ranqueamentos 164
 - Visão Geral do NIS 160
- nó de criação de bin
 - contagens iguais 162
 - somas iguais 162
- nó de exportação Arquivo Simples 358
 - guia de exportação 359
- Nó de exportação do banco de dados 344
 - esquema 346
 - fonte de dados 344
 - guia de exportação 344
 - mapeando campos de dados de origem para as colunas do banco de dados 345
- Nó de exportação do banco de dados (*continuação*)
 - nome da tabela 344
 - opções de mesclagem 345
 - tabelas de indexação 349
- nó de exportação do Excel 365
- nó de exportação do IBM Cognos TM1
 - exportando dados 363
 - mapeando dados de exportação 364
- nó de exportação do IBM SPSS Data Collection 359
- nó de exportação do ODBC. Consulte o nó de exportação da base de dados 344
- nó de exportação SAS 365
- nó de exportação XML 366
- nó de importação do Excel
 - gerando a partir da saída 365
- nó de Mesclagem
 - configurações de otimização 91
 - opções de configuração 86, 88
- nó de Ordem
 - configurações de otimização 84
 - visão geral 83
- Nó de origem da base de dados 17
 - consultas SQL 18
 - editor de consulta 25
 - editor de query 25
 - selecionando tabelas e visualizações 24
- nó de origem do Excel 42
- nó de origem do IBM Cognos BI 37, 39, 40
 - importando dados 38
 - importando relatórios 39
- Nó de origem do IBM Cognos BI ícones 37
- nó de origem do IBM Cognos TM1 40
 - importando dados 41
- nó de origem do IBM SPSS Data Collection 32, 36
 - arquivos de log 32
 - arquivos de metadados 32
 - configurações de conexão com o banco de dados 35, 36
 - conjuntos de múltiplas respostas 36
 - idioma 35
 - tipos de rótulos 35
- nó de origem Geoespacial
 - arquivos .dbf 64, 65
 - arquivos .shp 64, 65
 - Serviço de mapa 64, 65
- nó de origem SAS
 - arquivos .sd2 (SAS) 42
 - arquivos .ssd (SAS) 42
 - arquivos .tpt (SAS) 42
 - arquivos de transporte 42
- nó de origem XML 43
- nó de Saída de Estatísticas 374
 - guia Sintaxe 375
- nó de visualização de Mapa 271
 - alterar opções de camada 272
 - guia gráfico 271
- nó Derivar
 - condicional 154
 - configurando opções 149
 - contagem 154

- nó Derivar (*continuação*)
 - convertendo armazenamento de campo 154
 - derivando diversos 149
 - derivando um campo de lista 152
 - derivando um campo geoespacial 152
 - estado 153
 - fórmula 150
 - gerando a partir da preparação de dados automatizada 130
 - gerando a partir de categorias 160
 - gerando a partir de gráficos 282
 - gerando a partir de ligações de gráfico da web 258
 - gerando a partir de um nó
 - Categorização 166
 - nominal 153
 - recodificando valores 154
 - sinalizador 152
 - valor da fórmula 151
 - valores de coleção 151
 - valores geoespaciais 151
 - Visão Geral do NIS 148
- nó Distinto
 - configurações compostas 95, 97
 - configurações de otimização 95
 - ordenando registros 93
 - Visão Geral do NIS 93
- nó Distribuição 245
 - guia aparência 246
 - guia gráfico 245
 - usando o gráfico 246
 - utilizando a tabela 246
- nó do arquivo de cache 370
- nó do IBM Cognos BI Export 39, 361, 362
- nó Elemento do Gráfico 198
 - guia aparência 222
- nó Entrada do Usuário
 - opções de configuração 46
 - visão geral 45
- nó Estatísticas 324
 - correlações 324
 - estatísticas 324
 - guia configurações 324
 - guia saída 306
 - rótulos de correlação 324
- nó Exportação de Estatísticas 377
 - guia Exportar 377
- nó Exportação do IBM Cognos TM1 363
- nó Filtro
 - configurando opções 145
 - conjuntos de múltiplas respostas 147
 - Visão Geral do NIS 145
- nó Geração de simulação
 - opções de configuração 51
- nó Geração de Simulação
 - visão geral 50
- nó Gráfico 232
 - guia aparência 238
 - guia de opções 236
 - guia gráfico 235
 - usando o gráfico 238
- nó Gráfico de Tempo 242
 - guia aparência 244
 - guia gráfico 243
- nó Gráfico de Tempo (*continuação*)
 - usando o gráfico 244
- nó Histograma 249
 - guia aparência 250
 - guia gráfico 249
 - usando o gráfico 250
- nó Histórico 187
 - visão geral 187
- nó Intervalos de Tempo 177, 178
 - Analytic Server 189
 - visão geral 176
- nó Intervalos de Tempo do AS 190
 - visão geral 189
- Nó Intervalos de tempo do servidor analítico
 - visão geral 189
- nó Matriz 308
 - destaque 309
 - guia aparência 309
 - guia configurações 308
 - guia saída 306
 - navegador de saída 309
 - ordenando linhas e colunas 309
 - porcentagens de coluna 309
 - porcentagens de linha 309
 - tabulação cruzada 309
- nó Média 326
 - campos emparelhados 326
 - grupos independentes 326
 - guia saída 306
 - importância 327
 - navegador de saída 327
- nó Mesclagem 84
 - campos de tag 90
 - filtrando campos 90
 - visão geral 84
- nó Multigráficos 239
 - guia aparência 241
 - guia gráfico 239
 - usando o gráfico 241
- nó origem do Microsoft Excel 42
- nó Partição 170, 171
- nó Preenchimento
 - Visão Geral do NIS 155
- nó Preparação Automática de Dados 117
- nó Reclassificar 157, 158
 - gerando a partir de uma distribuição 246
 - visão geral 156
 - Visão Geral do NIS 160
- nó Reestruturar 173, 174
 - com o nó Agregado 173
- nó Relatório 328
 - guia modelo 329
 - guia saída 306
- nó Reordenação de Campo 188
 - opções de configuração 188
 - ordenação automática 188
 - ordenação customizada 188
- nó Reprojeção 191
- nó Reprojeter 191
- nó RFM Agregado
 - visão geral 82
- nó Saída do IBM SPSS Statistics
 - guia Saída 376
- nó Seleção
 - gerando a partir de gráficos 282
- nó Seleção (*continuação*)
 - visão geral 72
- nó Selecionar
 - gerando a partir de ligações de gráfico da web 258
- nó Space-Time-Boxes
 - definindo densidade 112
 - visão geral 110
- nó Tabela 305
 - configurações de saída 305
 - guia configurações 305
 - guia saída 306
- nó TCM de fluxo 105, 106, 107, 109, 110
- nó Tipo
 - configurando o papel de modelagem 142
 - copiando tipos 143
 - dados contínuos 139
 - dados nominais 140
 - dados ordinais 140
 - limpando valores 68
 - manipulação em branco 138
 - opções de configuração 133, 135, 136
 - tipo de campo de sinalização 140
 - tipo de dado de coleção 141
 - tipo de dados geoespaciais 141
 - visão geral 131
- nó Transformação 321
- nó Transformação de Estatísticas 371
 - guia Sintaxe 371
 - opções de configuração 371
 - sintaxe permissível 372
- nó Transpor 174
 - campos de sequência de caracteres 174
 - Campos numéricos 174
 - nomes de campo 174
- nó TS de Fluxo
 - opções de campo 98
 - opções de implementação 104
 - opções de modelo 99
 - visão geral 98
- Nó Visualização de Dados 63
 - opções de configuração 64
- nó Web 254
 - ajustando limites 260
 - ajustando pontos 258
 - alterar layout 258
 - definindo ligações 256
 - guia aparência 258
 - guia de opções 256
 - guia gráfico 255
 - régua de controle 258
 - régua de controle de ligações 258
 - sumarização da web 262
 - usando o gráfico 258
- nomes de campo 146
 - anonimizando 147
 - exportação de dados 344, 359, 365, 377
- nomes de variáveis
 - exportação de dados 344, 359, 365, 377
- normalizar valores
 - nós de gráfico 239, 243
- normalizar variável resposta
 - contínua 121, 130

- nós de exportação 343
 - exportação do Analytic Server 360
- nós de gráfico 193
 - animação 194
 - Avaliação 262
 - Coleção 251
 - Distribuição 245
 - Gráfico 232
 - Graphboard 198
 - Histograma 249
 - Multigráficos 239
 - painéis 194
 - Plotagem de tempo 242
 - sobreposições 194
 - visualização do mapa 271
 - Web 254
- nós de operações de campo 115
 - gerando a partir de uma auditoria de dados 320
- nós de operações de registro 71
 - nó Intervalos de Tempo 176
 - nó Intervalos de Tempo do AS 189
- nós de origem
 - nó Arquivo de Estatísticas 370
 - nó Arquivo Fixo 30
 - nó Arquivo Variável 26
 - nó da Visualização Corporativa 65
 - Nó de origem da base de dados 17
 - nó de origem do Excel 42
 - nó de origem do IBM Cognos BI 37, 39, 40
 - nó de origem do IBM Cognos TM1 40
 - nó de origem Geoespacial 64
 - nó de origem SAS 42
 - nó de origem XML 43
 - nó Entrada do Usuário 45, 46
 - nó Geração de Simulação 50, 51
 - origem do Analytic Server 12
 - tipos de instanciação 69
 - visão geral 7
- nós de saída 301, 305, 308, 310, 314, 324, 328, 330, 331, 333, 334, 336, 338, 340, 374
 - guia saída 306
 - publicação na web 303
- Nós do IBM SPSS Statistics 369
- nulos 138
 - em tabelas de Matriz 308

O

- ocorrências
 - opções do gráfico de avaliação 268
- ODBC
 - carregamento em massa por meio de 350, 352
 - conexão com o nó Exportação do IBM Cognos BI 362
 - Nó de origem da base de dados 17
- opções
 - IBM SPSS Statistics 341
- opções de camada em mapas 272
- Opções de gráfico 340
- opções de mesclagem, exportação do banco de dados 345
- opções de modelo
 - nó Modelo de Estatísticas 374

- Oracle 17
- ordem crescente 83
- ordem da coluna
 - navegador de tabela 305, 307
- ordem de execução
 - especificando 385
- ordem de mesclagem 84
- ordem decrescente 83
- ordem dos dados de entrada 90
- ordem natural
 - alterando 188
- ordenação
 - campos pré-ordenados 95
 - nó Distinto 93
- ordenando dados 83, 188
- ordens sazonais
 - modelos ARIMA 101
- origem do Analytic Server 12
- origens de dados
 - conexões com o banco de dados 19

P

- painel 194
- palavra-chave FILLFACTOR
 - indexação de tabelas de banco de dados 349
- palavra-chave UNIQUE
 - indexação de tabelas de banco de dados 349
- paletas
 - em movimento 286
 - exibindo 286
 - ocultando 286
- parâmetros
 - configurando para SuperNodes 383
 - no IBM Cognos BI 40
 - propriedades do nó 384
 - SuperNodes 383, 384
- parâmetros de fluxo 25
- Parâmetros de SuperNode 383, 384
- performance
 - nó Agregado 79
 - nós Categorização 166
 - nós Derivar 166
- periodicidade
 - dados de série temporal 176, 189
 - modelador de série temporal 102
- período de estimação 178
- períodos
 - nó Intervalos de Tempo 180
- períodos cíclicos
 - nó Intervalos de Tempo 181
- pesos
 - gráficos de avaliação 266
- planilhas
 - importando a partir do Excel 42
- planos de acesso a dados 64
- Point 135
- Polygon 135
- pontos de corte
 - nó de Categorização 160
- preenchendo dados de série temporal 177
- preparação de dados automática
 - construção 122

- preparação de dados automática
 - (continuação)
 - normalizar variável resposta
 - contínua 121
 - preparação de destino 121
 - preparação de entrada 121
 - preparando destinos 121
 - preparando entradas 121
 - seleção de variável 122
- preparação de dados automatizada
 - análise de campo 125
 - campos 119
 - campos de nome 123
 - configurações de campo 119
 - detalhes da ação 128
 - detalhes do campo 127
 - excluir campos 120
 - exclusão de campo não utilizado 119
 - exclusão de campos não utilizados 119
 - geração de nó derivar 130
 - ligações entre as visualizações 124
 - normalizar variável resposta
 - contínua 130
 - objetivos 117
 - poder preditivo 127
 - preparar as datas e horas 120
 - reconfigurar visualizações 124
 - sumarização da ação 126
 - sumarização de processamento de campo 125
 - tabela de campos 127
 - visualização do modelo 124
- preparação de dados geoespaciais
 - nó Reprojeter 191
- Primeiro quartil
 - agregação de séries temporais 190
- processamento paralelo
 - classificando 84
 - combinando 91
 - nó Agregado 79
- procurando
 - navegador de tabela 307
- profundidade da lista 11
- propriedades
 - nó 384
- propriedades do nó 384
- publicação na web 303
- Python
 - scripts de carregamento em massa 350, 352

Q

- quadro de amostragem 73
- qualidade de dados
 - navegador de Auditoria de Dados 318
- queries SQL
 - Nó de origem da base de dados 25
- qui-quadrado
 - nó Matriz 309
- Qui-quadrado de Pearson
 - nó Matriz 309

R

- ranqueamentos fracionais 164
- recência
 - configurando data relativa 82
- recodificação automática 156, 157
- recodificar 156, 157, 160
- reconhecimento automático de data 27, 30
- reconhecimento de data 27, 30
- reestruturando dados 173
- regiões em gráficos 279
- registro
 - comprimento 30
 - contagens 79
 - rótulos 142
- registros
 - combinando 84
 - transpondo 174
- registros compostos 95
 - configurações customizadas 97
- registros exclusivos 93
- registros incompletos 86
- regra de negócios
 - opções do gráfico de avaliação 268
- regressão por quadrados mínimos
 - localmente ponderada
 - nó Gráfico 235
- relatório de qualidade
 - navegador de Auditoria de Dados 318
- relatórios
 - salvando saída 306
- renda
 - gráficos de avaliação 266
- renomeando
 - arquivos de mapeamento 224
 - campos para exportação 378
 - folhas de estilo de visualização 224
 - modelos de visualização 224
- renomeando objetos de saída 302
- Repositório do IBM SPSS Collaboration and Deployment Services
 - conectando-se com 65
 - utilizando como localização para modelos de visualização, folhas de estilo e mapas 224
- representando associações 254
- reprojeção de dados do mapa 191
- reprojeção de dados geoespaciais 191
- resíduos
 - nó Matriz 309
- restrições em dados geoespaciais 135
- resultado tabular
 - reordenando colunas 305
 - selecione células 305
- ROI
 - gráficos 262, 269
- rótulos 140
 - especificando 138, 139, 140
 - exportando 365, 377
 - importando 42, 370
- rótulos de valor
 - nó Arquivo de Estatísticas 370
- rótulos de variáveis
 - nó Arquivo de Estatísticas 370
 - nó Exportação de Estatísticas 377

S

- saída 336, 338
 - exportando 304
 - gerando novos nós a partir de 302
 - HTML 304
 - imprimindo 302
 - salvamento 302
- saída de gráfico 338
- saída de impressão 302
- saída de matriz
 - salvando como texto 306
- saída HTML
 - nó Relatório 329
 - visualizar no navegador 304
- saída XML
 - nó Relatório 329
- salvamento
 - objetos de saída 302, 306
 - saída 302
- SAS
 - configurando opções de importação 42
- script
 - SuperNodes 385
- selecionando linhas (casos) 72
- selecionando valores 276, 279, 281
- semente aleatória
 - registros de amostragem 171
- séries temporais 187
- serviço de mapa
 - nó de origem Geoespacial 64, 65
- servidor ESRI 64
- significância
 - intensidade de correlação 324
- símbolo de agrupamento
 - formatos de exibição de número 144
- símbolo decimal 27
 - formatos de exibição de número 144
 - nó de exportação Arquivo Simplex 359
- sintaxe do XPath 43
- sistema de coordenadas geográficas 191
- sistema de coordenadas projetadas 191
- sistemas de coordenadas
 - transformando 293
- sobreposição de gráfico de cores 194
- sobreposição de gráfico de forma 194
- sobreposição de gráfico de painel 194
- sobreposição de gráfico de tamanho 194
- sobreposições para gráficos 194
- sobrescrevendo tabelas da base de dados 344
- soma
 - nó Configurar Globais 330
 - saída de estatísticas 325
- SPLOM 204
 - exemplo 219, 221
- suavização exponencial
 - critérios nos modelos de série temporal 100
- suavizador
 - nó Gráfico 235
- suavizador LOESS
 - nó Gráfico 235
- suavizador lowess, Consulte suavizador LOESS
 - nó Gráfico 235

- substituindo valores de campos 155
- SuperNodes 379
 - aninhando 381
 - aumentando o zoom 382
 - bloqueando 381
 - carregando 386
 - configurando parâmetros 383
 - criando caches para 385
 - criar 380
 - desbloqueando 381
 - editando 382
 - proteção de senha 381, 382
 - salvamento 386
 - script 385
 - SuperNodes de origem 379
 - SuperNodes de processo 379
 - SuperNodes terminais 380
 - tipos de 379
 - utilizando com comentários 382

T

- tabelas
 - salvando como texto 306
 - salvando saída 306
 - unindo 84
- tabulação cruzada
 - nó Matriz 308, 309
- tags 84, 90
- tamanho de confirmação 350
- Teradata
 - banda de query 23
- Terceiro quartil
 - agregação de séries temporais 190
- teste t
 - amostras emparelhadas 326
 - amostras independentes 326
 - nó Média 326, 328
- text
- data 30
- texto
 - data 26
 - delimitado 26
 - timestamp 133
- tipo 9
- tipo de coleção 141
- tipo de configuração 133
- tipo de flag 133, 140
- tipo de geoespaciais 141
- tipo de uso 9, 133
- tipos de armazenamento
 - lista 29
- tipos de campo
 - nas visualizações 200
- tipos de dados 30, 115, 133
 - instanciação 137
- tipos de gráfico
 - gráfico 204
- tipos de rótulos
 - nó de origem do IBM SPSS Data Collection 35
- tipos de variável
 - nas visualizações 200
- tipos de verificação 142
- transformação de diferenciação
 - modelos ARIMA 101

- transformação de diferenciação sazonal
 - modelos ARIMA 101
- transformação logarítmica
 - modelador de série temporal 102
- transformação logarítmica natural
 - modelador de série temporal 102
- transformação raiz quadrada
 - modelador de série temporal 102
- transformações
 - reclassificar 156, 160
 - recodificar 156, 160
- transparência em gráficos 194
- transpondo dados 174
- transpondo de dados 174
- True se qualquer função for verdadeira
 - agregação de séries temporais 177, 190
- truncando nomes de campo 145, 146

U

- unindo conjuntos de dados 92
- Utilitário de conversão de mapa 225, 226

V

- validações
 - modelagem de séries temporais 178
- valor da chave para agregação 79
- valor de contagem para agregação 79
- valor de mediana para agregação 79, 81
- valor de quartil para agregação 79, 81
- valor de variância para agregação 79
- valor máximo para agregação 79
- valor médio para agregação 79
- valor médio para registros 79
- valor mínimo para agregação 79
- valor p
 - importância 327
- valor semente
 - amostragem e registros 171
- valores
 - especificando 138
 - lendo 137
 - rótulos de campo e de valor 138
- valores ausentes 115
- valores de coleção para derivar
 - fórmula 151
- valores de limpeza 68
- valores discrepantes
 - em modelos de série temporal 104
 - modelos ARIMA 104
- valores discrepantes aditivos
 - modelador de série temporal 104
- valores discrepantes aditivos sazonais
 - modelador de série temporal 104
- valores discrepantes de mudança de nível
 - modelador de série temporal 104
- valores discrepantes de tendência local
 - modelador de série temporal 104
- valores discrepantes inovadores
 - modelador de série temporal 104
- valores discrepantes temporários
 - modelador de série temporal 104

- valores em branco
 - em tabelas de Matriz 308
- valores esperados
 - nó Matriz 309
- valores falsos 140
- valores geoespaciais para derivar
 - fórmula 151
- valores globais 330
- valores indefinidos 86
- valores nulos
 - em tabelas de Matriz 308
- valores omissos 138, 141
 - Em nós Agregar 79
 - em tabelas de Matriz 308
- valores omissos de usuário
 - em tabelas de Matriz 308
- valores omissos do sistema
 - em tabelas de Matriz 308
- valores para derivar fórmula 151
- valores pré-configurados, conexão com o
 - banco de dados 21
- valores somados 79
- valores verdadeiros 140
- variância
 - saída de estatísticas 325
- variáveis
 - em mapas 226
- variáveis de código
 - nó de origem do IBM SPSS Data Collection 32
- variáveis de SourceFile
 - nó de origem do IBM SPSS Data Collection 32
- Variáveis do sistema
 - nó de origem do IBM SPSS Data Collection 32
- varinha mágica em gráficos 281
- visualização
 - gráficos e diagramas 193
- visualização de dados analíticos 64
- visualização do mapa
 - exemplo 220
- visualização do modelo
 - na preparação de dados automatizada 124
- visualizações
 - categorias 291
 - copiar 295
 - cores e padrões 287
 - editando 285
 - eixos 290
 - escalas 290
 - forma de ponto 288
 - formatos numéricos 289
 - margens 289
 - modo de edição 285
 - painéis 291, 293
 - posição da legenda 295
 - preenchimento 289
 - proporção de aspecto de ponto 288
 - rotação de ponto 288
 - text 287
 - traço 287
 - transformando o sistema de coordenadas 293
 - transparência 287
 - transpor 291, 293

- visualizações de mapa
 - criar 211



Impresso no Brasil