

IBM SPSS Modeler 14.2 ユー ザー ガイド



注： サポートされている情報および製品をご利用いただく前に、「注意事項」（ p. 300 ）の一般情報をお読みください。

本マニュアルには、SPSS Inc., an IBM Company が所有する情報が含まれています。これらの情報は使用許諾契約書に基づいて提供され、著作権法によって保護されています。本文書に記載されている情報には、製品の保証は含まれていません。また本マニュアルに記載されている文は製品の保証を規定しないものとします。

IBM または SPSS に情報を送信すると、あなたに対する義務を負うことなく、適切とする方法でその情報を使用または配布する非独占的権利と IBM および SPSS 付与するものとします。

© Copyright IBM Corporation 1994, 2011..

はじめに

IBM® SPSS® Modeler は、IBM Corp. が開発した企業強化用のデータ マイニング ワークベンチです。SPSS Modeler を使用すると、企業はデータを詳しく調べることで顧客および一般市民とのリレーションシップを強化することができます。企業は、SPSS Modeler を使って得られた情報に基づいて利益をもたらす顧客を獲得し、抱き合わせ販売の機会を見つけ、新規顧客を引き付け、不正を発見し、リスクを減少させ、政府機関へのサービスの提供を改善することができます。

SPSS Modeler の視覚的インターフェイスを使用すると、特定ビジネスの専門知識を適用し、より強力な予測モデルを実現し、解決までの時間を短縮します。SPSS Modeler では、予測、分類、セグメント化、および関連性検出アルゴリズムなど、さまざまなモデル作成手法を提供しています。モデルを作成した後は、IBM® SPSS® Modeler Solution Publisher により、企業全体の意思決定者やデータベースにモデルを配布することが可能になります。

IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者がビジネス パフォーマンスを向上させるために信頼する完全で、一貫した正確な情報を提供します。ビジネス インテリジェンス、予測分析、財務実績および戦略管理、および分析アプリケーションの包括的なポートフォリオを利用することによって、現在の実績を明確、迅速に理解し、将来の結果を予測することができます。豊富な業界のソリューション、実績ある実例、専門サービスと組み合わせ、さまざまな規模の組織が、高い生産性を実現、意思決定を自信を持って自動化し、より良い決定をもたらします。

このポートフォリオの一部として、IBM SPSS Predictive Analytics ソフトウェアを使用する組織は、将来のイベントを予測し、その洞察に基づいて積極的に行動し、より優れた業績を実現することができます。全世界の企業、政府、学術分野のお客様が IBM SPSS の技術を活用し、不正行為を減少させ、リスクを軽減させながら、顧客の獲得、保持、成長において、競争優位を高めることができます。IBM SPSS ソフトウェアを日々の業務に取り入れることによって、組織は業務目標を達成し、大きな競争的優位を獲得することができるよう、意思決定を方向付け、自動化することができるようになります。お問い合わせは、<http://www.ibm.com/spss> を参照してください。

テクニカル サポート

お客様はテクニカル サポートをご利用いただけます。IBM Corp. 製品の使用方法、または対応するハードウェア環境へのインストールについてサポートが必要な場合は、テクニカル サポートにご連絡ください。テクニカ

ル サポートの詳細は、IBM Corp. Web ページ <http://www.ibm.com/support> を参照してください。ご本人、組織、サポートの同意を確認できるものをご用意ください。

内容

1	IBM SPSS Modeler について	1
	IBM SPSS Modeler Server	1
	IBM SPSS Modeler のオプション	2
	IBM SPSS Text Analytics	2
	IBM SPSS Modeler ドキュメント	3
	アプリケーションの例	4
	Demos フォルダ	5
2	新機能	6
	IBM SPSS Modeler 14.2 の新機能および変更された機能	6
	IBM SPSS Modeler 14.2 の新しいノード	7
3	IBM SPSS Modeler の概要	8
	はじめに	8
	IBM SPSS Modeler の起動	8
	コマンドラインからの起動	9
	IBM SPSS Modeler Server への接続	10
	一時ディレクトリの変更	14
	複数の IBM SPSS Modeler セッションの起動	15
	IBM SPSS Modeler インターフェイスについて	15
	IBM SPSS Modeler ストリーム領域	16
	ノードパレット	16
	IBM SPSS Modeler マネージャ	18
	IBM SPSS Modeler プロジェクト	20
	IBM SPSS Modeler ツールバー	21
	ツールバーのカスタマイズ	22
	IBM SPSS Modeler ウィンドウのカスタマイズ	23
	IBM SPSS Modeler でのマウスの使用方法	24
	ショートカットキーの使用	25
	印刷	26
	IBM SPSS Modeler 自動化	27

4 データマイニングについて 28

データマイニングの概要	28
データへの評価	30
データマイニングの戦略的方法	31
CRISP-DM プロセスモデル	32
モデルの種類	34
データマイニングの例	42

5 ストリームの構築 43

ストリーム構築の概要	43
データストリームの作成	43
ノードの作業	44
ストリームの作業	57
ストリームの説明	75
ストリームの実行	77
モデルの処理	78
ノードおよびストリームへのコメントおよび注釈の追加	79
データストリームの保存	90
ファイルのロード	94
データストリームのマッピング	95
ヒントとショートカット	101

6 欠損値の処理 106

欠損値の概要	106
欠損値の処理	108
欠損値を含むレコードの処理	108
欠損値を含むフィールドの処理	109
欠損値の代入または置換	110
欠損値用 CLEM 関数	110

7 CLEM 式の作成

113

CLEM について	113
CLEM の例	116
値とデータ型	118
式と条件	119
ストリーム、セッション、およびスーパーノード パラメータ	120
文字列の処理	121
空白および欠損値の処理	122
数値の処理	123
時間と日付の処理	123
複数フィールドの要約	124
複数レスポンス データの処理	126
Clem 式ビルダー	127
Clem 式ビルダーへのアクセス	128
CLEM 式の作成	128
関数の選択	129
フィールド、パラメータ、およびグローバル変数の選択	130
値の表示または選択	131
CLEM 式の検査	132
検索と置換	133

8 CLEM 言語に関するリファレンス

136

CLEM リファレンス概要	136
CLEMデータ型	136
整数	137
実数	137
文字	138
文字列	138
リスト	138
Fields	138
日付(D)	139
Time	140
CLEM演算子	141
関数のリファレンス	143
関数の表記方法について	144
情報関数	145

変換関数	146
比較関数	147
論理関数	149
数値関数	150
三角関数	151
確率関数	152
ビット単位の整数演算	152
乱数関数	154
文字列関数	154
SoundEx 関数	160
日付および時刻の関数	160
シーケンス関数	165
グローバル関数	171
空白値とヌル値処理関数	172
特殊フィールド	173

9 リポジトリでの IBM SPSS Modeler の使用 175

IBM SPSS Collaboration and Deployment Services Repository について	175
IBM SPSS Collaboration and Deployment Services Repository オブジェクトの保存と展開	177
IBM SPSS Collaboration and Deployment Services Repository への接続	179
IBM SPSS Collaboration and Deployment Services Repository の資格情報の入力	180
IBM SPSS Collaboration and Deployment Services Repository コンテンツを参照	181
IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの保存	182
オブジェクトの特性の設定	182
ストリームの保存	188
プロジェクトの格納	189
ノードの格納	189
出力オブジェクトの格納	190
モデルおよびモデル パレットの格納	190
IBM SPSS Collaboration and Deployment Services Repository からのオブジェクトの取得	191
取得するオブジェクトの選択	193
オブジェクトのバージョンの選択	194
IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの検索	195
IBM SPSS Collaboration and Deployment Services Repository オブジェクトの変更	198
フォルダの編集、名前の変更、削除	198

IBM SPSS Collaboration and Deployment Services Repository オブジェクトのロックおよびロック解除	199
IBM SPSS Collaboration and Deployment Services Repository オブジェクトの削除	199
IBM SPSS Collaboration and Deployment Services Repository オブジェクトのプロパティの管理	200
フォルダのプロパティの表示	201
オブジェクトのプロパティの表示および編集	203
オブジェクトのバージョン ラベルの管理	205
ストリームの展開	206
ストリーム展開のオプション	208
スコアリング枝	212

10 外部アプリケーションへのエクスポート 219

外部アプリケーションへのエクスポートについて	219
IBM SPSS Modeler Advantage でストリームを開く	220
Predictive Applications 4.x ウィザード	220
Predictive Applications ウィザードを使用する前に	221
2 値の予測を傾向スコアとしてエクスポート	222
ステップ 1 :Predictive Applications ウィザードの概要	223
ステップ 2 :ターミナル ノードの選択	224
ステップ 3 :UCV ノードの選択	224
ステップ 4 :パッケージの指定	227
ステップ 5 :パッケージの生成	227
ステップ 6 :要約	230
PMML としてのモデルのインポートおよびエクスポート	230
PMML をサポートするモデルの種類	233

11 プロジェクトとレポート 236

プロジェクトの概要	236
CRISP-DM ビュー	237
クラス ビュー	238
プロジェクトの作成	239
新規プロジェクトの作成	240
プロジェクトへの追加	240
プロジェクト IBM SPSS Collaboration and Deployment Services Repository への転送	242
プロジェクトのプロパティの設定	243

プロジェクトの注釈	245
オブジェクトのプロパティ	247
プロジェクトを閉じる	248
レポートの生成	248
生成されたレポートの保存とエクスポート	252

12 IBM SPSS Modeler のカスタマイズ 255

IBM SPSS Modeler オプションのカスタマイズ	255
IBM SPSS Modeler オプションの設定	255
システム オプション	255
デフォルト ディレクトリ の設定	257
[ユーザー オプション] ダイアログ ボックス の設定	257
ユーザー 情報の設定	267
ノード パレット のカスタマイズ	267
パレット マネージャ のカスタマイズ	267
パレット タブ の表示 の変更	273
CEMI ノード の管理	274

13 ストリームおよびノードのパフォーマンスの考察 275

ノードの順序	275
ノードのキャッシュ	276
パフォーマンス: プロセス ノード	278
パフォーマンス: モデル作成 ノード	280
パフォーマンス: CLEM 式	281

付録

A IBM SPSS Modeler のアクセス機能 282

IBM SPSS Modeler のアクセス機能の概要	282
アクセス機能サポートの種類	282
視力の弱いユーザーのためのアクセシビリティ	282
視覚障害ユーザーのためのアクセシビリティ	283

キーボードのアクセス機能	284
画面読み上げソフトウェアの使用	294
使用上のヒント	296
他のソフトウェアとの相性	297
JAWS および Java	297
IBM SPSS Modeler でのグラフの使用	297
B Unicode のサポート	298
IBM SPSS Modeler 内での Unicode のサポート	298
C 注意事項	300
索引	304

IBM SPSS Modeler について

IBM® SPSS® Modeler は、ビジネスの専門知識を活用して予測モデルを迅速に作成したり、また作成したモデルをビジネス オペレーションに展開して意志決定を改善できるようにする、一連のデータ マイニング ツールです。SPSS Modeler は業界標準の CRISP-DM モデルをベースに設計されたものであり、データ マイニング プロセス全体をサポートして、データに基づいてより良いビジネスの成果を達成できるようにします。

SPSS Modeler ではさまざまなモデル作成方法を提供しています。[モデル作成] パレットを利用して、データから新しい情報を引き出したり、予測モデルを作成することができます。各手法によって、利点や適した問題の種類が異なります。

SPSS Modeler は、スタンドアロン製品として購入または SPSS Modeler Server と組み合わせて使用することができます。後のセクションで説明されているとおり、多くの追加オプションも使用することができます。詳細は、<http://www.ibm.com/software/analytics/spss/products/modeler/> を参照してください。

IBM SPSS Modeler Server

SPSS Modeler は、クライアント/サーバー アーキテクチャを使用し、リソース主体の操作が必要な要求を、強力なサーバー ソフトウェアへ分散されるようになりました。その結果、規模が比較的大きいデータ セットを処理するパフォーマンスを実現しました。ここに挙げた以外にも、ほかの製品やアップデートも利用できる可能性があります。詳細は、<http://www.ibm.com/software/analytics/spss/products/modeler/> を参照してください。

SPSS Modeler: SPSS Modeler はこの製品のすべての機能を搭載したバージョンであり、ユーザーのデスクトップ コンピュータにインストールし、そのコンピュータで実行します。スタンドアロン製品としてローカル モードで実行するか、大規模なデータ セットを使用する場合にパフォーマンスを向上させるために IBM® SPSS® Modeler Server と組み合わせて実行することができます。

SPSS Modeler Server: SPSS Modeler Server は、1 つまたは複数の IBM® SPSS® Modeler のインストールと同時に分散分析モードで継続的に実行し、大規模なデータセットを使用する際にパフォーマンスが大幅に向上しますが、それは、データをクライアント コンピュータへダウンロードする

ことなく、メモリー主体の操作をサーバー上で実行できるからです。また、SPSS Modeler Server は SQL 最適化のサポート、データベース内モデル作成機能を提供し、パフォーマンスおよび自動化にさらなるメリットをもたらします。分析を実行するには、少なくとも 1 つの SPSS Modeler をインストールしておく必要があります。

IBM SPSS Modeler のオプション

次のコンポーネントおよび機能を個別に購入し、ライセンス供与を受け SPSS Modeler と合わせて使用できます。追加の製品や更新が利用できる可能性があることに注意してください。 詳細は、<http://www.ibm.com/software/analytics/spss/products/modeler/> を参照してください。

- SPSS Modeler Server へのアクセスにより、大規模なデータセット上のスケーラビリティおよびパフォーマンスを向上させ、SQL 最適化のサポート、およびインデータベース モデリング能力を提供します。
- SPSS Modeler Solution Publisher は、SPSS Modeler 環境の外側でのリアルタイムまたは自動スコアリングで使用します。 詳細は、 [2 章 IBM SPSS Modeler Solution Publisher in IBM SPSS Modeler 14.2 Solution Publisher](#) を参照してください。
- アダプタを使用して IBM SPSS Collaboration and Deployment Services またはシンククライアント アプリケーションの IBM SPSS Modeler Advantage に展開します。 詳細は、 [9 章 p.177 IBM SPSS Collaboration and Deployment Services Repository オブジェクトの保存と展開](#) を参照してください。

IBM SPSS Text Analytics

IBM® SPSS® Text Analytics は、SPSS Modeler の完全に統合されたアドオンで、高度な言語テクノロジーと自然言語処理（NLP: Natural Language Processing）を使用して、さまざまな無構造テキスト データを高速で処理し、重要な概念を抽出および整理してカテゴリに分類します。抽出されたコンセプトとカテゴリを、人口統計のような既存の構造化データと組み合わせ、IBM® SPSS® Modeler の豊富なデータ マイニング ツールを適用する方法で、焦点を絞ったより良い決定を下すことができます。

- テキスト マイニング ノードは、テキスト リンクとクラスタの高度な洞察を実行できるインタラクティブ ワークベンチと同様にコンセプトおよびカテゴリ モデル作成を提供、独自のカテゴリを作成し。言語リソースのテンプレートを調整します。

- ブログやその他の Web ベースのソースなど、多くのインポート形式がサポートされています。
- CRM (Customer Relationship Management、顧客関係管理) やゲノム研究のような特定の分野用のカスタム テンプレート、ライブラリ、辞書も付属しています。

注：このコンポーネントを利用するには、別途、ライセンスが必要です。詳細は、<http://www.ibm.com/software/analytics/spss/products/modeler/>を参照してください。

IBM SPSS Modeler ドキュメント

オンライン ヘルプ形式の完全なドキュメントは、SPSS Modeler の [ヘルプ] メニューから使用できます。SPSS Modeler、SPSS Modeler Server、および SPSS Modeler Solution Publisher のアプリケーション ガイドやその他サポート資料が含まれています。

各製品の PDF 形式の完全なドキュメントは、各製品 DVD の ¥Documentation フォルダにもあります。

- **IBM SPSS Modeler ユーザー ガイド:**SPSS Modeler の使用方法への全体的な入門で、データ ストリームの構築方法、欠損地の処理方法、CLEM 式の処理方法、プロジェクトおよびレポートの処理方法、IBM SPSS Collaboration and Deployment Services、予測アプリケーション製品、または IBM SPSS Modeler Advantage へ展開するストリームのパッケージ化方法が含まれています。
- **IBM SPSS Modeler 入力ノード、プロセス ノード、出力ノード:** さまざまな形式のデータを読み込み、処理し、出力するために使用するすべてのノードの説明があります。これは、モデル作成ノード以外のすべてのノードについての説明です。
- **IBM SPSS Modeler モデル作成ノード:** データ マイニング モデルの作成に使用するすべてのノードの説明。IBM® SPSS® Modeler には、マシン学習、人工知能、および統計に基づいたさまざまなモデル作成手法が用意されています。詳細は、3 章 **モデル作成ノードの概要 in IBM SPSS Modeler 14.2 Modeling Nodes** を参照してください。
- **IBM SPSS Modeler アルゴリズム ガイド:**SPSS Modeler で使用されている手法の数学的な基礎の説明があります。
- **IBM SPSS Modeler アプリケーション ガイド:** 本ガイドの例では、特定のモデル作成手法および技術に関する簡単で、目的に沿った説明を行います。本ガイドのオンライン バージョンは、[ヘルプ] メニューからも利用できます。詳細は、p.4 **アプリケーションの例** を参照してください。
- **IBM SPSS Modeler スクリプトとオートメーション:** スクリプトの実行によるシステムのオートメーションの情報で、ノードおよびストリームを操作するために使用することができるプロパティが含まれています。

- **IBM SPSS Modeler 展開ガイド:**SPSS Modeler のストリームやシナリオを IBM® SPSS® Collaboration and Deployment Services Deployment Manager のジョブを処理するステップとしての実行についての情報。
- **IBM SPSS Modeler CLEF 開発者ガイド:**CLEF では、SPSS Modeler のノードとしてデータ処理ルーチンやモデル作成アルゴリズムなどのサードパーティ製のプログラムを統合します。
- **IBM SPSS Modeler データベース内 マイニング ガイド:** ユーザーのデータベースを最大限に活用して、パフォーマンスを改善する方法と、サードパーティ製のアプローチを使用して分析可能な範囲を拡大する方法についての情報があります。
- **IBM SPSS Modeler Server およびパフォーマンス ガイド:**IBM® SPSS® Modeler Server の設定と管理の方法について説明します。
- **IBM SPSS Modeler 管理コンソール ユーザー ガイド:**SPSS Modeler Server を監視して設定するためのコンソール ユーザー インターフェイスのインストールおよび使用に関する情報。コンソールは、Deployment Manager アプリケーションへのプラグインとして実装されます。
- **IBM SPSS Modeler Solution Publisherガイド:** SPSS Modeler Solution Publisher はアドオン コンポーネントです。組織はこれを使用すると、標準的な SPSS Modeler 環境の外部へストリームを公開できます。
- **IBM SPSS Modeler CRISP-DM Guide.** CRISP-DM 手法を使用した SPSS Modeler によるデータ マイニングの段階的なガイドです。

アプリケーションの例

SPSS Modeler のデータ マイニング ツールは、多様なビジネスおよび組織の問題解決を支援しますが、アプリケーションの例では、特定のモデル作成手法および技術に関する簡単で、目的に沿った説明を行います。ここで使用されるデータセットは、データ マイニング作業によって管理された巨大なデータ ストアよりも非常に小さいですが、関係するコンセプトや方法は実際のアプリケーションに対して大規模です。

SPSS Modeler の [ヘルプ] メニューから [アプリケーションの例] を選択すると、例にアクセスすることができます。データ ファイルとサンプル ストリームは、製品のインストール ディレクトリの Demos フォルダにインストールされています。詳細は、[p.5 Demos フォルダ](#) を参照してください。

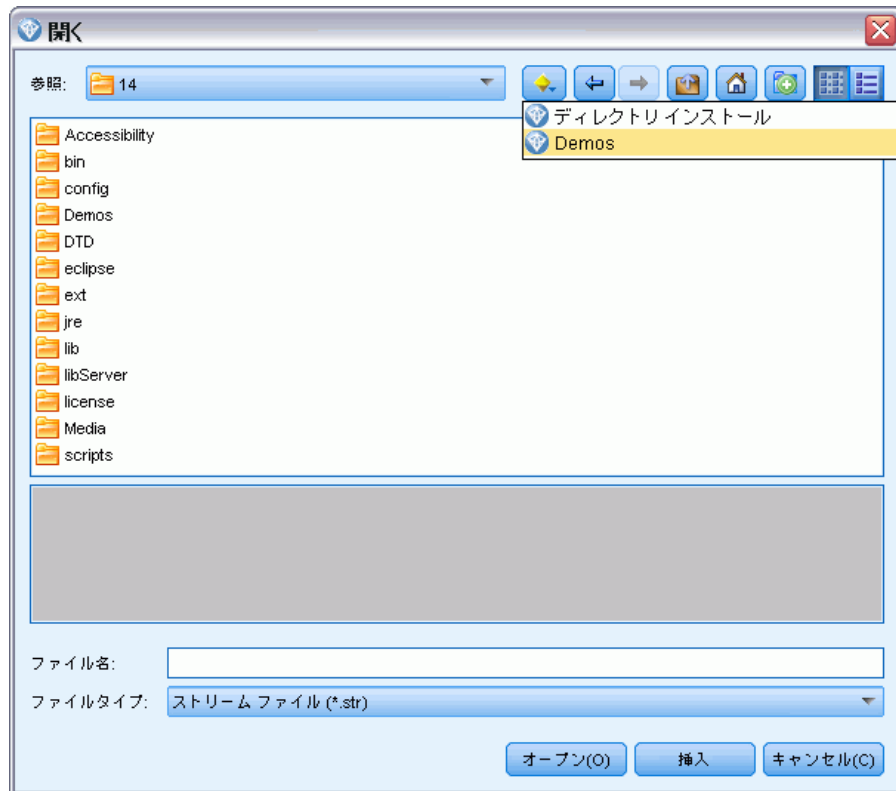
データベース モデル作成の例: 例は、『IBM SPSS Modeler データベース内マイニング ガイド』を参照してください。

スクリプトの例: 例は、『IBM SPSS Modeler スクリプトとオートメーション ガイド』を参照してください。

Demos フォルダ

アプリケーションの例で使用されるデータ ファイルとサンプル ストリームは、製品のインストール ディレクトリの Demos フォルダにインストールされています。このフォルダには、Windows [スタート] メニューの **IBM SPSS Modeler 14.2** プログラム グループから、または [ファイルを開く] ダイアログ ボックスの最近使ったディレクトリの一覧から [Demos] を選択してアクセスすることもできます。

図 1-1
最近使用されたディレクトリの一覧から Demos フォルダを選択



新機能

IBM SPSS Modeler 14.2 の新機能および変更された機能

IBM® SPSS® Modeler 14.2 リリースにより、次の機能が追加されます。

IBM Netezza Analytics のデータベース内マイニングのサポート: このリリースでは、IBM Netezza Analytics テクノロジーのデータベース内マイニングをサポートしています。詳細は、[6 章 IBM SPSS Modeler](#) そして [IBM Netezza Analytics in IBM SPSS Modeler 14.2 データベース内マイニング ガイド](#) を参照してください。

サポートは、Netezza ディジジョン ツリーおよび Netezza K-Means の新しいデータベース モデリング ノードの形式で提供されます。詳細は、[p. 7 IBM SPSS Modeler 14.2 の新しいノード](#) を参照してください。

Netezza データベースの階層 1 サポート: Netezza データベースは現在階層 1 レベルでサポートされています。つまり、可能なすべての SQL プッシュバックを、データベース独自の SQL 最適化によって使用できます。詳細は、[2 章 データへのアクセス in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド](#) を参照してください。

サンプル ノードの SQL 生成に対するサポート拡張: 単純なサンプリングを使用する場合の SQL 生成のサポートは、Windows および UNIX で実行する DB2 データベース、IBM Netezza および Teradata データベース向けに拡張されました。詳細は、[6 章 SQL 生成をサポートしているノード in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド](#) を参照してください。

Cognos 10 のサポート: IBM Cognos BI 入力ノードおよびエクスポート ノードでは、Cognos BI バージョン 8.4 のほか Cognos BI バージョン 10.1 をサポートしています。バージョン 8.4 (デフォルト) からバージョン 10.1 に切り替えるには、サポート サイト (<http://www.ibm.com/support>) で入手可能な該当の Cognos 10.1 ユーティリティ (Windows バッチ ファイルまたは UNIX シェル スクリプト) を実行します。今後バージョン 8.4 に戻す場合、同じサイトから入手可能な該当の Cognos 8.4 ユーティリティを実行する必要があります。

IBM System z のサーバー側のサポート: SPSS Modeler は、Red Hat Enterprise Linux、Red Hat Enterprise Linux Advanced Platform、または SuSE Linux Enterprise Server を実行する IBM System z システムのサーバー側サポー

トを追加しました。詳細は、IBM® SPSS® Modeler Server DVD の『IBM SPSS Modeler Server 14.2 for UNIX Installation Instructions』を参照してください。

IBM SPSS Modeler 14.2 の新しいノード



Netezza ディシジョン ツリー ノードは、将来の観測を予測または分類するために使用できるディシジョン ツリーを作成します。この方法は再帰的なデータ区分を使用して学習レコードを複数のサブグループに分割し、各ステップで不純性を最小限に抑えます。ツリーのノードが「純粹」であると考えられるのは、ノード中にあるケースの 100% が、対象フィールドのある特定のカテゴリ（クラス ラベル）に分類される場合です。詳細は、6 章 [Netezza ディシジョン ツリー in IBM SPSS Modeler 14.2 データベース内マイニング ガイド](#) を参照してください。



Netezza K-Means ノードは、クラスタ分析を実行し、データのメンバーを類似したレコードのグループに分割できます。アルゴリズムは、同じ基本プロセスの反復を数回繰り返し、学習インスタンスは最も近いクラスタに割り当てられます。すべてのクラスタ中心は、特定のクラスタに割り当てられたインスタンスの平均属性値のベクトルとして再計算されます。詳細は、6 章 [Netezza K-Means in IBM SPSS Modeler 14.2 データベース内マイニング ガイド](#) を参照してください。

IBM SPSS Modeler の概要

はじめに

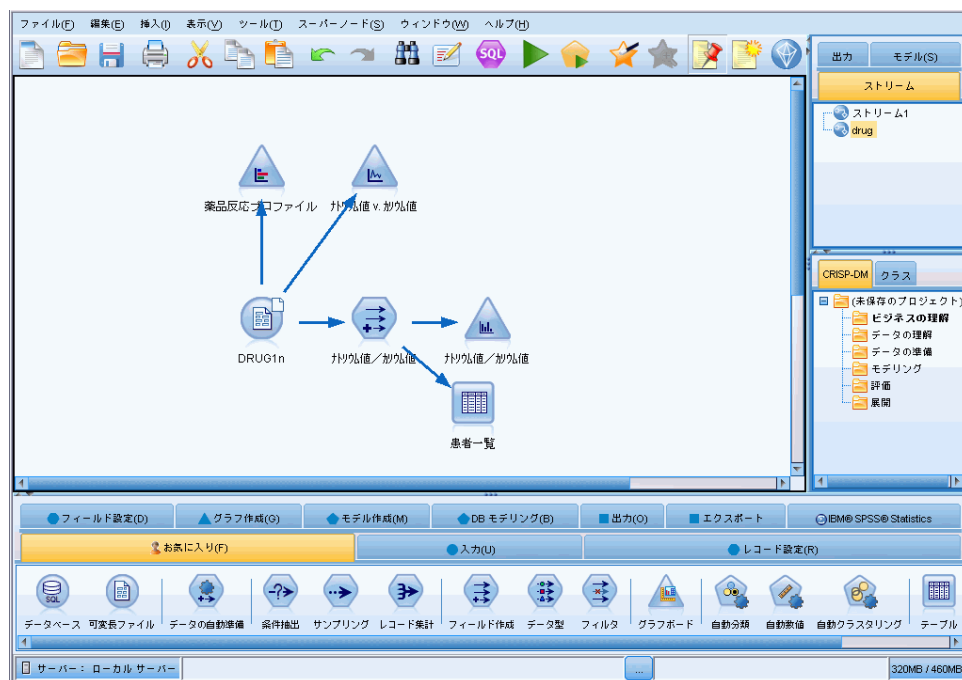
IBM® SPSS® Modeler はデータマイニングアプリケーションで、大きなデータセット中の有益なリレーションシップを見つけだすための、戦略的なアプローチ手段を提供しています。従来の統計的な手法とは対照的に、作業開始時に何を見つけだそうとしているのかがわからなくてもかまいません。データの探索、さまざまなモデルの適合、およびさまざまなデータの関連の調査などの作業を行いながら、有益な情報を見つけだしていくことができます。

IBM SPSS Modeler の起動

アプリケーションを起動するには、以下のメニューをクリックします。
[スタート] > [すべてのプログラム] > IBM SPSS Modeler14.2 > IBM SPSS Modeler14.2

数秒後にメイン ウィンドウが表示されます。

図 3-1
IBM SPSS Modeler のメイン アプリケーション ウィンドウ



コマンドラインからの起動

オペレーティング システムのコマンド ラインを使用し、次のようにして IBM® SPSS® Modeler を起動できます。

- ▶ IBM® SPSS® Modeler がインストールされているコンピュータで、DOS つまりコマンド プロンプト ウィンドウを開きます。
- ▶ SPSS Modeler インターフェイスをインタラクティブ モードで起動するには、**modelerclient** コマンドを入力し、続いてたとえば次のような適切な引数を入力します。

```
modelerclient -stream report.str -execute
```

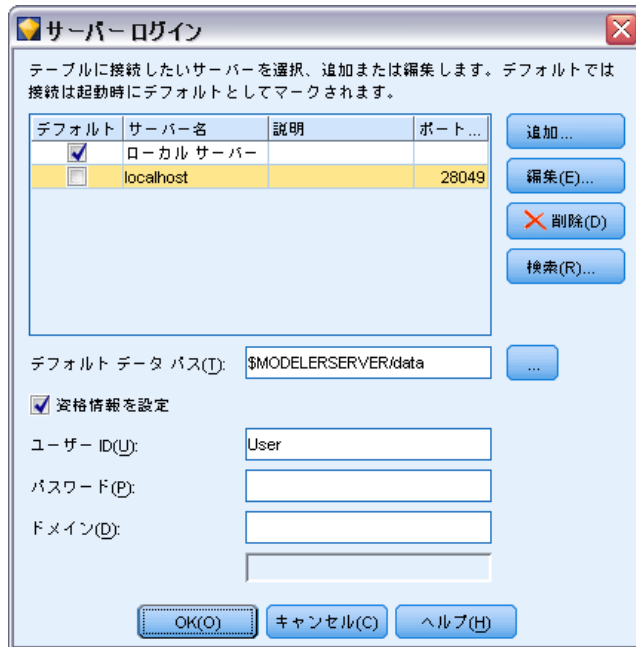
使用可能な引数 (フラグ) により、サーバーへの接続、ストリームのロード、スクリプトの実行、または必要に応じて他のパラメータの指定を行うことができます。

IBM SPSS Modeler Server への接続

IBM® SPSS® Modeler は、スタンドアロンのアプリケーションとして、または、IBM® SPSS® Modeler Server に直接または IBM® SPSS® Collaboration and Deployment Services から Coordinator of Processes プラグインを介して SPSS Modeler Server またはサーバー クラスタに接続されたクライアントとして、実行できます。現在の接続ステータスは、SPSS Modeler ウィンドウの左下に表示されます。

サーバーに接続する場合は、接続するサーバー名を手動で入力するか、以前定義した名前を選択できます。ただし、IBM SPSS Collaboration and Deployment Services を使用する場合、[サーバーへのログイン] ダイアログ ボックスからサーバーまたはサーバー クラスタのリストを使用して検索することができます。ネットワーク上で実行する Statistics サービスを介して参照する機能は、Coordinator of Processes で使用できます。詳細は、[D 付録 サーバー クラスタでの負荷バランシング in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド](#) を参照してください。

図 3-2
[サーバー ログイン] ダイアログ ボックス



サーバーに接続するには

- ▶ [ツール] メニューの [サーバーへのログイン] をクリックします。[サーバーへのログイン] ダイアログ ボックスが開きます。または、SPSS Modeler ウィンドウの接続ステータス領域をダブルクリックします。

- ▶ ダイアログ ボックスで、ローカル サーバーのコンピュータに接続するオプションを指定するか、テーブルから接続を選択します。
 - [追加] または [編集] をクリックして、接続を追加または編集します。詳細は、[p.12 IBM SPSS Modeler Server 接続の追加および編集](#) を参照してください。
 - [検索] をクリックして、SPSS COP のサーバーまたはサーバー クラスターにアクセスします。詳細は、[p.13 IBM SPSS Collaboration and Deployment Services のサーバーの検索](#) を参照してください。

サーバー テーブル： このテーブルには、定義されたサーバー接続のセットが含まれています。このテーブルには、デフォルト接続、サーバー名、説明、ポート番号が表示されています。既存の接続を選択または検索、あるいは新しい接続を手動で追加することができます。特定のサーバーをデフォルト接続として設定するには、接続のテーブルの [デフォルト] 列のチェック ボックスを選択します。

デフォルト データ パス： サーバー コンピュータ上のデータへのパスを指定します。[...] ボタンをクリックして、目的の場所を指定することもできます。

資格情報の設定： このボックスのチェックを解除すると、**シングル サインオン**機能を有効にし、ローカル コンピュータのユーザー名とパスワードの詳細を入力してサーバーにログインします。この でシングル サインオンを使用できない場合、またはこのボックスをチェックしてシングル サインオンを無効にした場合（たとえば、管理者アカウントにログインした場合）、資格情報を入力するための次のフィールドが表示されます。

ユーザー ID： サーバーにログインするユーザー名を入力します。

パスワード： 指定したユーザー名に対応するパスワードを入力します。

ドメイン： サーバーにログインする際に使用するドメイン名を指定します。サーバー コンピュータが クライアント コンピュータとは異なる Windows ドメインにある場合のみ、ドメイン名が必要です。

- ▶ [OK] をクリックして、接続を完了します。

サーバーとの接続を切断するには

- ▶ [ツール] メニューの [サーバーへのログイン] をクリックします。[サーバーへのログイン] ダイアログ ボックスが開きます。または、SPSS Modeler ウィンドウの接続ステータス領域をダブル クリックします。
- ▶ ダイアログ ボックスで、[ローカル サーバー] を選択し、[OK] をクリックします。

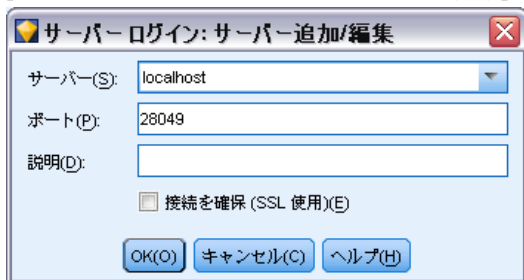
IBM SPSS Modeler Server 接続の追加および編集

[サーバーへのログイン] ダイアログ ボックスでサーバー接続を手動で編集または追加することができます。[追加] をクリックすると、サーバー接続の詳細を入力できる空の [サーバーの追加/編集] ダイアログ ボックスにアクセスすることができます。[サーバーへのログイン] ダイアログ ボックスで既存の接続を選択して [編集] をクリックすると、[サーバーの追加/編集] ダイアログ ボックスが開いて接続の詳細が表示され、その接続を変更することができます。

注： IBM® SPSS® Collaboration and Deployment Services から追加されたサーバー接続は、名前、ポート、およびそのほかの詳細が IBM SPSS Collaboration and Deployment Services で定義されているため、編集することができません。

図 3-3

[サーバーへのログイン: サーバーの追加/編集] ダイアログ ボックス



サーバー接続を追加するには

- ▶ [ツール] メニューの [サーバーへのログイン] をクリックします。[サーバーへのログイン] ダイアログ ボックスが開きます。
- ▶ ダイアログ ボックスで、[追加] をクリックします。[サーバーへのログイン: サーバーの追加/編集] ダイアログ ボックスが表示されます。
- ▶ サーバー接続の詳細を入力して [OK] をクリックします。接続が保存され、[サーバーへのログイン] ダイアログ ボックスに戻ります。
 - **サーバー:** 利用できるサーバーを指定するか、またはリストから選択します。サーバー コンピュータは、英数字の名前（たとえば、myserver）、または、サーバー コンピュータに割り当てられた IP アドレス（たとえば、202.123.456.78）で識別できます。
 - **ポート:** サーバーが待機しているポート番号を入力します。デフォルトのポート番号がうまく動作しない場合は、システム管理者に問い合わせして正しいポート番号を取得してください。

- **説明：** サーバー接続の説明をオプションで入力します。
- **接続を確保 (SSL 使用)：** SSL (Secure Sockets Layer) 接続を使用するかどうかを指定します。SSL は、ネットワークを介してセキュアなデータ送信を行うために一般的に使用されているプロトコルです。この機能を使用するには、IBM® SPSS® Modeler Server をホストするサーバー側で SSL を有効にする必要があります。必要な場合、詳細を各サイトの管理者に問い合わせてください。

サーバー接続を編集するには

- ▶ [ツール] メニューの [サーバーへのログイン] をクリックします。[サーバーへのログイン] ダイアログ ボックスが開きます。
- ▶ ダイアログ ボックスで、編集する接続を選択し、[編集] をクリックします。[サーバーへのログイン: サーバーの追加/編集] ダイアログ ボックスが表示されます。
- ▶ サーバー接続の詳細を変更して [OK] をクリックします。変更が保存され、[サーバーへのログイン] ダイアログ ボックスに戻ります。

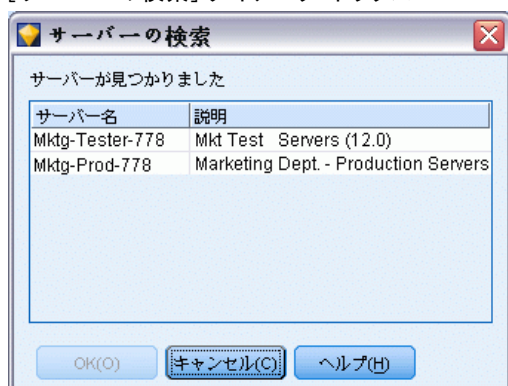
IBM SPSS Collaboration and Deployment Services のサーバーの検索

サーバー接続を手動で入力する代わりに、IBM® SPSS® Collaboration and Deployment Services で使用できる Coordinator of Processes (COP) を介してネットワークで使用可能なサーバーまたはサーバー クラスタを選択できます。サーバー クラスタは、Coordinator of Processes が処理要求に応答するのに最適なサーバーを決定するサーバーのグループです。 [詳細は、D 付録 サーバー クラスタでの負荷バランシング in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド](#) を参照してください。

[サーバーへのログイン] ダイアログ ボックスでサーバーを手動で追加できますが、使用できるサーバーを検索して、正しいサーバー名やポート番号を知らなくてもサーバーに接続することができます。この情報は、自動的に提供されます。ただし、ユーザー名、ドメインおよびパスワードなどの、正しいログオン情報が必要です。

注：Coordinator of Processes 機能へアクセスしていない場合、接続するサーバー名を手動で入力したり、以前定義した名前を選択することができます。 [詳細は、p.12 IBM SPSS Modeler Server 接続の追加および編集](#) を参照してください。

図 3-4
[サーバーの検索] ダイアログ ボックス



サーバーおよびクラスタを検索するには

- ▶ [ツール] メニューの [サーバーへのログイン] をクリックします。[サーバーへのログイン] ダイアログ ボックスが開きます。
- ▶ ダイアログ ボックスで [検索] をクリックすると、[サーバーの検索] ダイアログ ボックスが表示されます。Coordinator of Processes をブラウズしようとする際に IBM SPSS Collaboration and Deployment Services にログオンしていない場合、ログオンを指示するメッセージが表示されます。
[詳細は、9 章 p.179 IBM SPSS Collaboration and Deployment Services Repository への接続 を参照してください。](#)
- ▶ リストからサーバーまたはサーバー クラスタを選択します。
- ▶ [OK] をクリックしてダイアログ ボックスが閉じられ、選択した接続が [サーバーへのログイン] ダイアログ ボックスのテーブルに追加されます。

一時ディレクトリの変更

IBM® SPSS® Modeler Server が行う処理や操作の中には、一時ファイルを作成する必要があるものもあります。IBM® SPSS® Modeler のデフォルトでは、システムの一時ファイル用ディレクトリに一時ファイルが作成されます。一時ディレクトリの場所を変更するには、次の手順に従ってください。

- ▶ 新規ディレクトリ spss およびそのサブディレクトリ servertemp を作成します。
- ▶ SPSS Modeler のインストール ディレクトリ中の /config にある、options.cfg を編集します。次のファイルの temp_directory パラメータを編集して読み込みます。temp_directory, "C:/spss/servertemp"

- ▶ この作業を行った後は、SPSS Modeler Server サーバー サービスを再起動する必要があります。再起動するには、Windows コントロール パネルの [管理ツール] にある [サービス] タブをクリックしてください。サービスを停止した後、再び開始すると変更内容が有効になります。また、マシンを再起動しても、サービスが再開されます。

これで、新しいディレクトリに一時ファイルが作成されるようになります。

注： この作業を行う際に、スラッシュの種類を間違えて指定することのないように注意してください。SPSS Modeler のように、普通のスラッシュを使用しています。

複数の IBM SPSS Modeler セッションの起動

複数の IBM® SPSS® Modeler を一度に起動する必要がある場合、IBM® SPSS® Modeler および Windows の設定を変更する必要があります。たとえば、2 つの個別のサーバー ライセンスを持ち、同じクライアント マシンから 2 つの異なるサーバーに対して 2 つのストリームを実行する場合に、変更を行う必要があります。

複数の SPSS Modeler セッションを有効化する手順は、次のとおりです。

- ▶ 次のメニューをクリックします。
[スタート] > [すべてのプログラム] > IBM SPSS Modeler14.2
- ▶ IBM SPSS Modeler14.2 のショートカット（アイコンを表示）で、右クリックして [プロパティ] を選択します。
- ▶ [対象] テキスト ボックスで、文字列の終わりに `-noshare` を追加します。
- ▶ Windows の Explorer で、次の項目を選択します。
ツール > フォルダ オプション...
- ▶ [ファイル タイプ] タブで、[SPSS Modeler ストリーム] オプションを選択し、[詳細] をクリックします。
- ▶ [ファイル タイプの編集] ダイアログ ボックスで、[SPSS Modeler で開く] を選択し、[編集] をクリックします。
- ▶ [アクションの実行に使用するアプリケーション] テキスト ボックスで、`-stream` 引数の前に `-noshare` を追加します。

IBM SPSS Modeler インターフェイスについて

IBM® SPSS® Modeler の使いやすいインターフェイスでは、データ マイニング処理の各ポイントで、特定ビジネスの専門知識が必要となります。予測、分類、セグメント化、関連性検出などのモデリング アルゴリズム

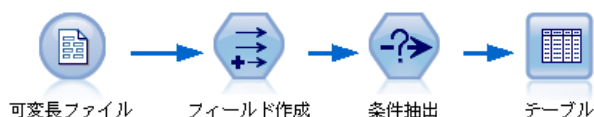
によって、強力かつ正確なモデルが作成されます。モデルの結果は簡単に展開でき、データベース、IBM® SPSS® Statistics、およびさまざまなアプリケーションで使用することができます。

SPSS Modeler での作業は、3 ステップのデータの処理で行われます。

- まず、データを SPSS Modeler に読み込みます。
- そして一連の操作によってデータを実行します。
- 最後に、データを宛て先に送信します。

各操作により、入力から各レコード単位でデータが流れて処理されて、最後に終点（モデルまたはデータ出力タイプ）に到達するため、この操作の流れは**データ ストリーム**と呼ばれます。

図 3-5
単純なストリーム



IBM SPSS Modeler ストリーム領域

ストリーム領域は、IBM® SPSS® Modeler ウィンドウ内で最も広い領域を占めています。ここで、データ ストリームの作成と操作を行います。

ストリームは、業務に関連するデータ操作のダイアグラムをインターフェイスのメイン領域に描画することによって作成します。各操作はノードと呼ばれるアイコンで表されます。ノードは、データの流れと操作を表すストリーム中に配置され、相互に接続されます。

SPSS Modeler では、同じストリーム領域内で、または別のストリーム領域を開いて、複数のストリームに関する作業を同時に行えます。セッションの間、ストリームは SPSS Modeler ウィンドウの右上にあるストリーム マネージャに格納されます。

ノード パレット

IBM® SPSS® Modeler のほとんどのデータおよびモデル作成ツールは、ストリーム領域下のウィンドウの下部にある【ノード パレット】にあります。

たとえば、[レコード設定] パレット タブには、選択、結合、追加など、データ レコードの操作を実行するために使用できるノードが含まれています。

ストリーム領域にノードを追加するには、ノード パレットのアイコンをダブル クリックするか、またはアイコンをストリーム領域にドラッグ アンド ドロップします。次にそれらのノードを接続し、データの流を表すストリームを作成していきます。

図 3-6
ノード パレットの [レコード設定] タブ



それぞれのパレット タブには、次のようなストリーム操作の各段階で使われる関連ノード群が用意されています。

- **入力**：ノードによって SPSS Modeler にデータが入力されます。
- **レコード設定**：抽出、結合、および追加などの、データ レコードの操作を実行するノードです。
- **フィールド設定**：フィルタリング、新規フィールドの作成、およびフィールドの測定レベルの判断などの、データ フィールドに対する操作に用いられます。
- **グラフ**：モデル作成の前後に、データをグラフィカルに表示するために用いられます。グラフには、散布図、ヒストグラム、Web グラフ ノード、および評価グラフなどがあります。
- **モデリング**。ニューラル ネットワーク、ディシジョン ツリー、クラスタリング アルゴリズム、およびデータのシーケンス化などの、SPSS Modeler で利用できる優れたモデル作成 アルゴリズムを表すノードです。
- **データベース モデリング**：Microsoft SQL Server、IBM DB2、および Oracle データベースで使用できるモデル作成アルゴリズムを使用します。
- **出力**：SPSS Modeler で表示できるさまざまなデータ、グラフ、モデルの結果の出力を作成します。
- **エクスポート**：IBM® SPSS® Data Collection または Excel などの外部アプリケーションで表示できるさまざまな出力を作成します。
- **SPSS Statistics**：IBM® SPSS® Statistics プロシージャを実行するほか、SPSS Statistics との間でデータのインポートまたはエクスポートを行います。

SPSS Modeler を使用していく過程で、このパレットの内容を自分に合わせてカスタマイズしていくことができます。詳細は、12 章 p.267 ノードパレットのカスタマイズを参照してください。

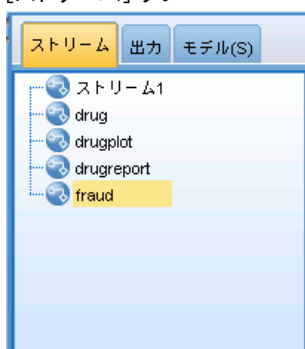
ノード パレットの下部にあるレポート パネルには、データ ストリームにデータを読み込む場合など、各種操作の進行状況に関するフィードバック情報が表示されます。ステータス パネルもノード パレットの下にあります。このウィンドウには、アプリケーションの現在の処理状況や、ユーザーへのフィードバックが必要な場合の指示などが表示されます。

IBM SPSS Modeler マネージャ

ウィンドウの右上がマネージャ パネルです。3 つのタブがあり、ストリーム、出力、モデルの管理を行います。

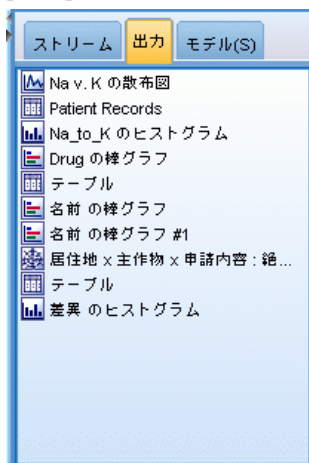
[ストリーム] タブでは、セッション中に作成されたストリームを開いたり、名前を変更したり、保存したり、削除することができます。

図 3-7
[ストリーム] タブ



[出力] タブには、IBM® SPSS® Modeler のストリーム操作で作成されたグラフおよびテーブルなどのさまざまなファイルが表示されます。ここに記載されているテーブル、グラフ、およびレポートを表示したり、名前を変更したり、または閉じることができます。

図 3-8
[出力] タブ



[モデル] タブは、最も強力なマネージャ タブです。このタブにはすべてのモデル ナゲットが表示されます。それらは現在のセッションの SPSS Modeler で生成されるモデルです。これらのモデルは、[モデル] タブから直接参照したり、領域中のストリームに追加することができます。

図 3-9
モデル ナゲットを含む [モデル] タブ

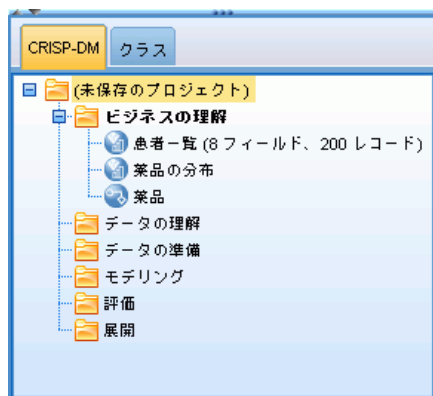


IBM SPSS Modeler プロジェクト

ウィンドウの右下には、データマイニングプロジェクト（データマイニングタスクに関連するファイルのグループ）の作成と管理に使用するプロジェクトパネルがあります。IBM® SPSS® Modeler で作成したプロジェクトを表示するには、クラスビューを使用する方法と、CRISP-DMビューを使用する方法があります。

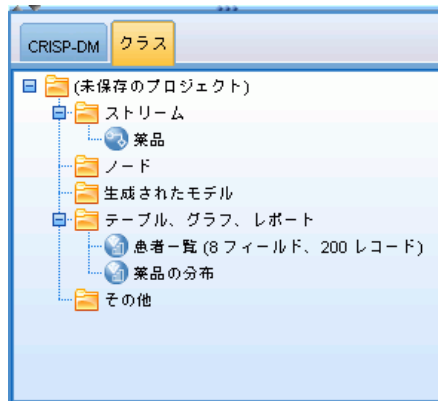
[CRISP-DM] タブでは、世界的に知られている方法論で実績のある CRISP-DM (Cross-Industry Standard Process for Data Mining) に基づいてプロジェクトを編成することができます。データマイニングを熟知している方でも、初めてデータマイニングを行う方でも、CRISP-DM ツールを使用することにより、より円滑にプロジェクトを編成し、最良の結果を得ることができます。

図 3-10
CRISP-DM ビュー



[クラス] タブでは、SPSS Modeler で行った作業内容を、作成したオブジェクトの種類に応じて編成することができます。このビューは、データ、ストリーム、モデルなどを確認、整理するために役立ちます。










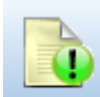

図 3-11
クラス ビュー



IBM SPSS Modeler ツールバー

IBM® SPSS® Modeler ウィンドウの上部には、アイコンが配置されたツールバーがあります。このツールバーは、役に立つさまざまな機能を提供しています。ツールバー ボタンとその機能を次に示します。

	新規ストリームの作成		ストリームを開く
	ストリームの保存		現在のストリームの印刷
	切り取ってクリップボードに移動		クリップボードにコピー
	貼り付けの選択		最後の操作を元に戻す
	やり直し (R)		ノードの検索
	ストリームのプロパティを編集		SQL 生成をプレビュー

	ストリームを実行		選択したストリームを実行
	ストリームの中止（ストリームの実行中にだけ利用可能）		選択したノードをスーパーノードにカプセル化
	ズーム イン（スーパーノード専用）		ズーム アウト（スーパーノード専用）
	ストリーム中にマークアップがありません		コメントの挿入
	ストリーム マークアップの非表示（ある場合）		非表示のストリーム マークアップを表示
	IBM® SPSS® Modeler Advantage でストリームを開く		

ストリーム マークアップはストリームのコメント、モデル リンク、およびスコアリング枝の表示で構成されています。

ストリーム コメントの詳細は、「[ノードおよびストリームへのコメントおよび注釈の追加](#)」（p. 79）を参照してください。

スコアリング枝表示の詳細は、「[スコアリング枝](#)」（p. 212）を参照してください。

モデルのリンクについては、『IBM SPSS モデル作成ノード ガイド』を参照してください。

ツールバーのカスタマイズ

ツールバーは、次のようにさまざまな観点から変更できます。

- 表示するかどうか
- アイコンのツールヒントを使用できるかどうか
- 大きいアイコンまたは小さいアイコンのどちらを使用するか

ツールバー表示をオンまたはオフにするには

- ▶ メイン メニューで次の各項目をクリックします。

表示 > ツールバー > 表示

ツールヒントまたはアイコンのサイズ設定を変更するには

- ▶ メイン メニューで次の各項目をクリックします。

表示 > ツールバー > カスタマイズ

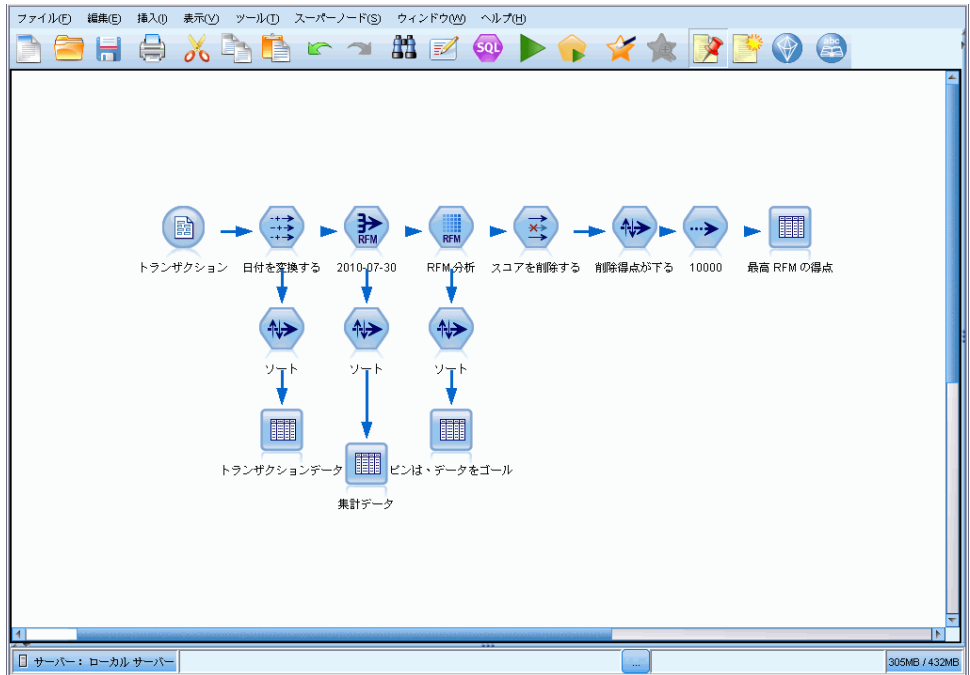
必要に応じて、[ツールヒントを表示] または [大きいボタン] をクリックします。

IBM SPSS Modeler ウィンドウのカスタマイズ

IBM® SPSS® Modeler インターフェイスのさまざまな構成要素の間にあるディバイダを使って、そのサイズを変更したり、ツールを閉じることができます。たとえば、大きいストリームを作成して作業を行う場合などに、各ディバイダにある小さい矢印をクリックしてノード パレット、マネージャ パネル、およびプロジェクト パネルを閉じることができます。これらのウィンドウを閉じることによってストリーム領域が広がるため、大きいストリームや複数のストリームで作業を行うために十分な領域を確保することができます。

または、[表示] メニューで、[ノード パレット]、[マネージャ] または [プロジェクト] をクリックして、これらの項目の表示をオンまたはオフにします。

図 3-12
最大化されたストリーム キャンバス



ノード パレット、マネージャ パネル、およびプロジェクト パネルを閉じる代わりに、SPSS Modeler ウィンドウの横と下にある青いスクロールバーを動かして、ストリーム領域をスクロールすることもできます。

画面マークアップの表示を制御できます。画面マークアップはストリームのコメント、モデル リンク、およびスコアリング枝の表示で構成されています。この表示をオンまたはオフにするには、次をクリックします。

表示 > ストリーム マークアップ

IBM SPSS Modeler でのマウスの使用方法

IBM® SPSS® Modeler でよく使われるマウス操作を次に示します。

- **シングルクリック:** マウスの右または左ボタンを使用して、メニューからオプションを選択したり、ポップアップ メニューを開いたり、さまざまな標準コントロールやオプションにアクセスすることができます。ノードをクリックしたままマウスを動かしてドラッグすれば、ノードを移動できます。

- **ダブルクリック**：マウスの左ボタンをダブルクリックすることにより、ストリーム領域にノードを配置したり、既存のノードを編集します。
- **中央ボタンのクリック**：マウスの中央ボタンをクリックしてカーソルをドラッグすることにより、ストリーム領域中のノードを接続します。マウスの中央ボタンをダブルクリックすると、ノードの接続が解除されます。中央ボタンがないマウスを使用している場合は、代わりに Alt キーを押しながらクリックしたり、ドラッグしてください。

ショートカット キーの使用

大部分の IBM® SPSS® Modeler ビジュアル プログラミング操作には、対応するショートカット キーが用意されています。たとえば、ノードをクリックしてキーボードの Del キーを押すと、ノードを削除することができます。同様に、Ctrl キーを押したまま S キーを押すと、ストリームを素早く保存できます。このようなコントロール コマンドは、Ctrl+S のように、Ctrl と他のキー名で示されています。

Ctrl+X (切り取り) のように、標準の Windows 操作で使われているショートカット キーも数多くあります。SPSS Modeler では、後述するアプリケーション独自のショートカット キーのほかに、これらの標準のショートカット キーを利用することもできます。

注：一部、SPSS Modeler で使われていた古いショートカット キーが Windows 標準のショートカット キーと重複している場合があります。これらの古いショートカット キーを使用する場合は、Alt キーも一緒に押しながら使用してください。たとえば、Ctrl+Alt+C キーを使用して、キャッシュをオンとオフに切り換えることができます。

テーブル 3-1
サポートしているショートカット キー

ショートカット キー	関数
Ctrl+A	すべてを選択する
Ctrl+X	切り取り
Ctrl+N	新規ストリーム
Ctrl+O	ストリームを開く
Ctrl+P	プリント
Ctrl+C	コピー
Ctrl+V	貼り付け
Ctrl + Z	元に戻す
Ctrl+Q	選択したノードの下流にあるすべてのノードを選択
Ctrl+W	下流のすべてのノードの選択を解除 (Ctrl+Q と切り替わる)

ショートカットキー	関数
Ctrl+E	選択したノードから実行
Ctrl+S	現在のストリームを保存
Alt+矢印キー	選択したストリーム領域上のノードを矢印の方向に移動
Shift+F10	選択したノードのポップアップメニューを表示

テーブル 3-2
古いホットキーに対応するショートカットキー

ショートカットキー	関数
Ctrl+Alt+D	ノードの複製
Ctrl+Alt+L	ノードのロード
Ctrl+Alt+R	ノード名の変更
Ctrl+Alt+U	ユーザー入力ノードの生成
Ctrl+Alt+C	キャッシュのオン/オフの切り替え
Ctrl+Alt+F	キャッシュの取消
Ctrl+Alt+X	スーパーノードの展開
Ctrl+Alt+Z	ズーム イン/ズーム アウト
Del	ノードまたは接続の削除

印刷

IBM® SPSS® Modeler では、次のオブジェクトを印刷できます。

- ストリームのダイアグラム
- グラフ作成
- [テーブル]
- レポート (レポート ノードおよびプロジェクト レポートから)
- スクリプト ([ストリームのプロパティ]、[スタンドアロン スクリプト]、または [スーパーノード スクリプト] ダイアログ ボックスから)
- モデル (モデル ブラウザ、現在フォーカスのあるダイアログ ボックスのタブ、ツリー ビューア)
- 注釈 (出力の [注釈] タブを使って)

オブジェクトを印刷するには、次のようにします。

- プレビューを行わないでオブジェクトを印刷するには、ツールバーの [印刷] ボタンをクリックします。

- 印刷前に印刷の設定を行うには、[ファイル] メニューの [ページ設定] を選択します。
- 印刷前にプレビューを行う場合は、[ファイル] メニューの [印刷プレビュー] を選択します。
- 標準の印刷ダイアログ ボックスに、選択されているプリンタのオプションを表示して、さまざまなオプションを設定するには、[ファイル] メニューの [印刷] を選択します。

IBM SPSS Modeler 自動化

高度なデータ マイニング作業は複雑で長期間になることもあるため、IBM® SPSS® Modeler にはさまざまな種類のコーディングや自動化のサポート機能が用意されています。

- **Control Language for Expression Manipulation (CLEM)** は、SPSS Modeler ストリーム中を流れるデータの分析と操作を行うための強力な言語です。CLEM を使用すれば、経費と収入データから利益を算出するような簡単な操作から、Web ログ データを有益な情報を含む一連のフィールドやレコードに変換するような複雑な操作まで、さまざまなストリーム操作を行うことができます。詳細は、[7 章 p.113 CLEM について](#) を参照してください。
- **スクリプト**は、ユーザー インターフェースのプロセスを自動化する強力なツールです。スクリプトは、マウスやキーボードを使用して実行するのと同じような操作を実行できます。CLEM のサブセットを使用して、ノードにオプションを設定し、フィールドを作成することができます。また、出力を指定して生成されたモデルを操作することができます。詳細は、[2 章 スクリプトの概要 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド](#) を参照してください。

データマイニングについて

データマイニングの概要

データマイニングでは、さまざまな手法を使用して、データ本体に関する貴重な情報を探し出します。データマイニングでは、意思決定支援、予測、推定などの領域で使われるさまざまな手法を利用して情報を取り出します。データ量は膨大ですが、利用価値の低いものやそのままの形では直接役に立たないものがあります。価値があるのは、データに隠されている貴重な情報です。

データマイニングは、データに関するユーザー（またはユーザー側の専門家）の知識と、高度で有効な分析手法が結びついたときに大きな成果を上げます。分析手法とは、データの基本的な関係と特性を◆◆コンピュータを使って識別する手法です。データマイニングの過程では、履歴データからモデルが作成されます。このモデルは、工程の予測やパターン検出などさまざまな目的で使用されます。これは、**マシン学習**または**モデル作成**と呼ばれる、モデルを構築するための手法です。

モデル作成手法

IBM® SPSS® Modeler には、数多くのマシン学習およびモデル作成手法が含まれており、それらは、解くことができる問題の種類に応じておおまかに分類できます。

- 予測モデルリング手法には、ディシジョン ツリー、ニューラル ネットワーク、および統計モデルが含まれます。
- クラスタリング モデルは、類似したレコードのグループを識別し、そのグループに従ってレコードにラベルを付けます。クラスタリング手法には、Kohonen、K-Means、および TwoStep が含まれます。
- アソシエーション ルールは、特定の結果（特定の製品の購入）と条件セット（複数の他の製品の購入）を関連付けます。
- スクリーニング モデルを使用することにより、データ スクリーニングを行って、モデル作成で関心の対象となる可能性が最高に高いフィールドとレコードを探し出したり、既知のパターンに適合しない可能性がある外れ値を識別したりできます。ここで使用できる方法には、機能選択と異常性検出が含まれます。

データの操作と発見

SPSS Modeler はまた、さまざまな機能を使って自分の専門知識をデータに適用できます。

- **データの操作**：既存のデータから新しいデータ項目を構成したり、データを意味のあるサブセットに分割します。さまざまなソースからのデータを結合したり、フィルタリングすることができます。
- **参照と視覚化**：データ検査ノードを使ってデータのさまざまな側面を表示し、グラフや統計量などを含めた初期検査を行います。視覚化されたインタラクティブ グラフィックを利用したり、グラフィックをエクスポートしてプロジェクト レポートに入れることができます。
- **統計量**：データ内の変数から予測された関係を確認します。SPSS Modeler では、IBM® SPSS® Statistics からの統計量を使用することもできます。
- **仮説の検定**：データの動作をモデル化し、これらのモデルを検査します。

通常、これらの機能は、データ中の有効な一連の属性を識別するために用いられます。次にこれらの属性を、基本的なルールと関係を識別するモデル作成手法に適用します。

典型的なアプリケーション

データマイニング テクニックの典型的なアプリケーションには、次の例が含まれます。

ダイレクトメール：どの人口統計的なグループが、最も高い応答率を持っているのかを識別する。この情報は、将来のメール送付の際に、応答を最大にするために使用できます。

信用情報のスコアリング：個人の信用履歴を、ローン審査に使用する。

人事：過去の雇用業務を理解して、雇用プロセスを効率化するディシジョン ルールを作成する。

医療分野の研究：医学的な証拠に基づいて適当な処置を提案するディシジョン ルールを作成する。

市場分析：地域、価格、および顧客の特長などのどの変数が売り上げに関連しているかを識別する。

品質管理：製品の製造行程からのデータを分析して、製品の欠陥を識別する変数を決定する。

政策研究：最も重要な変数を選択するディシジョン ルールを使用した政策決定で、調査データを使用する。

健康管理：ユーザー調査および臨床データを組み合わせて、健康に寄与している変数を見出す。

用語

属性、フィールド、および変数という用語は、検討中の全ケースに共通する一つのデータ項目を示します。特定のケースを示す属性値の集合は、レコード、例、またはケースといいます。

データへの評価

データ マイニングでは、使用するデータが一定の基準を満たしていない限り、成果を期待できません。次の項で、データの側面をいくつか示し、検討すべきアプリケーションを示します。

データが使用できることを確認してください。

これは漠然としているように思われますが、データを使用できても、簡単に利用できるフォーマットにはなっていない場合があることに注意してください。IBM® SPSS® Modeler では、データをデータベース (ODBC を介して) またはファイルからインポートできます。ただし、直接アクセスできないマシン上で、データが別のフォーマットで保持されている場合があります。そのようなデータは、ダウンロードするか、適切な形式でダンプしないと、使用できません。データが異なるデータベースやソースに散在しており、それらを同時に取得する必要がある場合もあります。データは、オンライン上にあるとは限りません。データが紙にしか記されていない場合は、データ マイニングを始める前に、データを入力する必要があります。

データが関連する属性をカバーしているかどうかの確認

データ マイニングの目的は関連する属性を識別することなので、この確認などは奇妙に思われるかもしれませんが、使用できるデータを調べ、記録されていない関連因子を見つけることは、非常に役に立ちます。たとえば、アイスクリームの販売数を予測するとき、小売店や過去の販売記録については大量の情報があったとしても、大きな影響を与えると思われる天候や気温については情報がないという場合があります。属性が欠けていても、データ マイニングで有効な結果を出すことはできますが、結果として出された予測の精度が落ちる場合があります。

このような状況を評価するために、データの総合的な検査を行うことができます。まず始めに、データ ソースにデータ検査ノードを接続して実行し、レポートを生成、調査することも検討してください。詳細は、[6 章 データ検査ノード in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

ノイズを含んだデータの注意

データには、誤ったデータやさまざまに判断できる主観的なデータが含まれていることがよくあります。これらは一括してノイズと呼ばれます。データ内のノイズが正常な場合もあります。基本ルールがある場合もありますが、すべてのケースにあてはまるとは限りません。

通常、データにノイズが多くなると、それだけ正確な結果を得ることが難しくなります。しかし、SPSS Modeler のマシン学習手法はノイズデータを処理することができ、約 50% 近くのノイズを含むデータセットに対しても有効に使用できます。

データが十分にあることを確認してください。

データマイニングでは、データセットのサイズは必ずしも重要ではありません。データセットの標本性、可能な結果に到達するためのデータセットの範囲、および変数の組み合わせの方がはるかに重要です。

通常、考慮する属性が多くなると、有効な標本範囲を満たすのに必要なレコード数も多くなります。

データが標本として有効で、共通する基本ルールがある場合は、数千（数百でも可）レコードのデータサンプルでも、百万単位のデータサンプルの場合と同様に、正確な結果をすばやく得ることができます。

データのエキスパートの探索

多くの場合、ユーザーは自分のデータを処理するので、データの内容と意味はよく理解しています。しかし、別の部門や顧客のデータを処理する場合は、データについてよく知っている専門家の力を借りた方がはるかに効率が上がります。専門家は関連する属性の識別方法を指導したり、真に貴重な情報の集まりと、データセットの異常などの理由により生じた無益な情報を区別し、データマイニングの結果を解釈するための支援を行います。

データマイニングの戦略的方法

日常の企業努力と同様に、データマイニングは、計画的かつ体系的な方法で行えば、より大きい効果をあげることができます。IBM® SPSS® Modeler などの最新鋭のデータマイニングツールを使用したときでも、データマイニングの多くの作業で、熟練したビジネスアナリストが注意深く処理を監視する必要があります。よりよい計画を策定するためには、次の事項を確認してください。

- 解決すべき重要な問題は何か？
- どのようなデータソースを使用できるか、データのどの部分が現在の問題に関連しているか？

- データ マイニングを開始する前に、どのような事前処理とデータの事前処理が必要か？
- どのようなデータ マイニング手法を使用できるか？
- データ マイニング分析の結果をどのように評価するか？
- データ マイニングから最大限の情報を得るにはどうするか？

通常のデータ マイニング プロセスは、すぐに複雑になっていく傾向があります。複雑なビジネス問題、複数のデータ ソース、データ ソースのさまざまなデータ品質、データ マイニング手法の配列、データ マイニングの結果を測定するさまざまな方法など、追跡する項目は多岐にわたります。

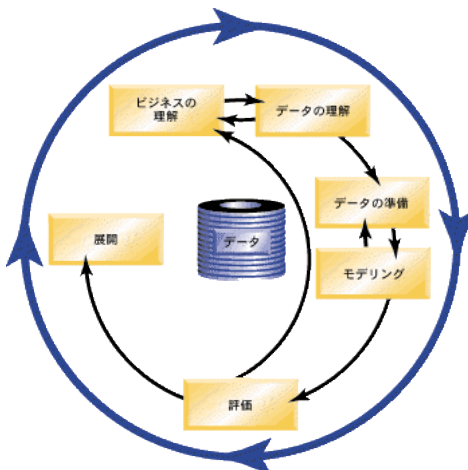
追跡を継続するには、データ マイニングのために明確に定義されたプロセス モデルがあると役に立ちます。プロセス モデルを使用して、この項で前述した質問に回答し、重要なポイントが処理されていることを確認することができます。プロセス モデルは、複雑なデータを処理しているときにユーザーが迷子にならないように、データ マイニングのロード マップとして機能します。

SPSS Modeler で使用を推奨しているデータ マイニング プロセス モデルは、CRISP-DM (Cross-Industry Standard Process for Data Mining) です。名前からわかるように、このモデルは広範囲な産業やビジネス問題に適用できる一般モデルとして設計されています。

CRISP-DM プロセス モデル

一般 CRISP-DM プロセス モデルには 6 つのフェーズがあり、データ マイニングの主要な課題を解決します。6 つのフェーズは、データ マイニングをさらに大きなビジネスの実践に組み込むために設計された、循環するプロセス内に一緒に適合されます。

図 4-1
CRISP-DM プロセス モデル



次に 6 つのフェーズを説明します。

- **ビジネスの理解**：データマイニングで最も重要なフェーズです。このフェーズには、ビジネス対象の決定、状況の評価、データマイニングの目標の決定、プロジェクト計画の作成があります。
- **データの理解**：データは、データマイニングの「生の素材」を提供します。このフェーズは、必要なデータリソースとそのリソースの特性を理解するためのフェーズです。このフェーズには、初期データの収集、データの説明、データの検査、データ品質の妥当性検証があります。[出力] パレットにあるデータ検査ノードは、データの理解に必要不可欠なツールです。
- **データの準備**：データリソースのカatalogを作成したら、そのデータをデータマイニング用に準備する必要があります。準備作業には、データの選択、クリーニング、構成、構築、統合、およびフォーマットなどがあります。
- **モデリング**。これはデータマイニングの最重要部分で、複雑な分析手法を使用してデータから情報を取得します。このフェーズには、モデル化手法の選択、テスト設計の生成、モデルの構築、評価などがあります。
- **評価**：いったんモデルを選択したら、ビジネス目標に到達するために、データマイニングの結果がどのように役に立つかを評価することができます。このフェーズには、結果の評価、データマイニングプロセスのレビュー、次のステップの決定があります。
- **展開**：これまでの努力が報われるとき、実りを収穫するときです。このフェーズでは、新しい知識を日常のビジネス処理に統合して、それまでのビジネス上の問題を解決します。このフェーズには、展開、モニタリング、メンテナンス、最終レポートの作成、プロジェクトのレビューがあります。

このプロセスモデルには、重要なポイントがいくつかあります。まず、通常は前の段落で概要を示した順序で処理を進めますが、多くの場所で各フェーズは非線型的に相互に影響し合います。たとえば、通常、データの準備はモデル作成の前に行われます。しかし、モデル作成フェーズで出された決定と収集された情報によっては、データの準備フェーズの再考が必要になる場合があります。データの準備フェーズを変更すると、次は新しいモデル作成を検討する必要があります。このように、2 つのフェーズは、適切に解決されるまで相互にフィードバックします。同様に、評価フェーズで誤った問いに答えようとしていたことがわかり、元のビジネスの理解から再評価しなければならない場合もあります。この時点でビジネスの理解を訂正しよりよい目標を設定してから、残りの処理を進めることができます。

2 番目の主要な点は、データマイニングの反復特性です。ユーザーの仕事が、単にデータマイニングプロジェクトを計画し実行して、自分のデータを片付けてから家に帰るだけ、ということはほとんどありません。データマイニングを使用して顧客の要望を解決するということが当面の

目標です。データマイニングの1サイクルから知識を得ると、多くの場合は顧客の要望を判別してそれを満たすための新しい問い、新しい課題、新しい機会が生じます。これらの新しい問い、課題、機会は、データを再度マイニングすることで解決できます。このマイニングと新しい機会の識別プロセスは、ビジネスについての考え方の一部になり、ビジネス戦略全体の基礎になります。

ここでは、CRISP-DM プロセスモデルについて簡単に概要だけ説明します。モデルの詳細は、次のリソースを参照してください。

- 『CRISP-DM ガイド』は、インストール ディスクの ¥Documentation フォルダの他のドキュメントと共にアクセスできます。
- [スタート] メニューから、または IBM® SPSS® Modeler 内で [ヘルプ] メニューから [CRISP-DM のヘルプ] をクリックして利用可能な、CRISP-DM ヘルプ システム。

モデルの種類

IBM® SPSS® Modeler ではさまざまなモデル作成方法を提供しています。[モデル作成] パレットを利用して、データから新しい情報を引き出したリ、予測モデルを作成することができます。各手法によって、利点や適した問題の種類が異なります。

SPSS Modeler アプリケーション ガイドでは、これらの手法の多くの例が、モデル作成プロセスの概要とともに提供されています。このガイドは、オンライン チュートリアル、および PDF 形式で使用できます。[詳細は、1 章 p.4 アプリケーションの例 を参照してください。](#)

モデル作成方法は、次の 3 つのカテゴリに分けられます。

- Classification
- アソシエーション
- セグメント化

分類モデル

分類モデルでは、1 つまたは複数の入力フィールドの値を使用し、1 つまたは複数の出力、または対象フィールドの値を予測します。これらの手法の例として、ディ氏ジョン ツリー (C&R ツリー、QUEST、CHAID および C5.0 アルゴリズム)、回帰 (線型、ロジスティック、一般化線型、Cox 回帰アルゴリズム)、ニューラル ネットワーク、サポート ベクターマシン、Bayesian ネットワークがあります。

組織は分類モデルを活用して、既知の結果に基づく予測に役立てることができます。たとえば、顧客が購入するか立ち去るか、または特定の取引が既知の詐欺パターンに当てはまるかどうか、です。モデル作成手法

には、マシン学習、ルール算出、サブグループ識別、統計的手法、および多重モデル生成が含まれます。

分類ノード



自動分類ノードは、2種類の結果 (yes/no、churn/don't churn など) を生じる多くの異なるモデルを作成および比較し、与えられた分析への最善のアプローチを選ぶことができるようになります。多くのモデル作成アルゴリズムに対応し、希望する方法、各特定のオプション、そして結果を比較するための基準を選択することができます。このノードで、指定されたオプションに基づいてモデルのセットが生成され、指定された基準に基づいて最善の候補がランク付けされます。詳細は、[5章 自動分類ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



自動数値ノードでは、多くのさまざまな方法を使用し、連続する数値範囲の結果を求めてモデルを推定し比較します。このノードは、自動分類ノードと同じ方法で動作し、1回のモデル作成のパスで、複数の組み合わせのオプションを使用し試すアルゴリズムを選択することができます。使用できるアルゴリズムには、ニューラルネットワーク、C&R Tree、CHAID、線型回帰、一般化線型回帰、サポートベクトルマシン (SVM) が含まれています。モデルは、相関、相対エラー、または使用された変数の数に基づいて比較できます。詳細は、[5章 自動数値ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



C&R Tree (分類と回帰ツリー) ノードは、ディビジョン ツリーを生成し、将来の観測値を予測または分類できるようにします。この方法は再帰的なデータ区分を使用して学習レコードを複数のセグメントに分割し、各ステップで不純性を最小限に抑えます。ツリーのノードが「純粹」であると考えられるのは、ノード中にあるケースの 100% が、対象フィールドのある特定のカテゴリに分類される場合です。対象フィールドおよび入力フィールドは、数値範囲またはカテゴリ (名義型、順序型、フラグ) が使用できます。すべての分岐は 2 分割です (2 つのサブグループのみ)。詳細は、[6章 C&R ツリー ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



QUEST ノードには、ディビジョン ツリーの構築用に 2 分岐の方法が用意されています。これは、大規模な C&R ツリー分析が必要とする処理時間を短縮すると同時に、より多くの分割を可能にする入力値が優先される分類ツリー内の傾向を低減するように設計されています。入力フィールドは、数値範囲 (連続型) にできませんが、対象変数はカテゴリでなければなりません。すべての分割は 2 分岐です。詳細は、[6章 QUEST ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



CHAID ノードはディシジョン ツリーを生成し、カイ二乗統計値を使用して最適な分割を識別します。C&RT Tree および QUEST ノードと異なり、CHAID は、非 2 分岐ツリーを生成できます。これは、ある分岐が 3 個以上の枝葉を持つことを意味します。対象フィールドおよび入力フィールドは、数値範囲（連続型）またはカテゴリとなります。Exhaustive CHAID は CHAID の修正版で、可能性のある分割すべてを調べることで、よりよい結果を得られますが、計算時間も長くなります。詳細は、6 章 [CHAID ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



C5.0 ノードは、ディシジョン ツリーとルール セットのどちらかを構築します。このモデルは、各レベルで最大の情報の対応をもたらすフィールドに基づいてサンプルを分割します。対象フィールドは、カテゴリでなければなりません。複数の分割を 2 つ以上のサブグループに分割できます。詳細は、6 章 [C5.0 ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



ディシジョン リスト ノードは、母集団に関連する与えられた 2 値の結果の高いもしくは低い尤度を示すサブグループまたはセグメントを識別します。たとえば、離れる可能性の少ないもしくはキャンペーンに好意的に答える可能性のある顧客を探することができます。顧客区分を追加し、結果を比較するために他のモデルを並べて表示することによって、ビジネスに関する知識をモデルに導入することができます。ディシジョン リスト モデルは、ルールのリストから構成され、各ルールには条件と結果が含まれます。ルールは順番に適用され、一致する最初のルールで、結果が決まります。詳細は、9 章 [ディシジョン リスト in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



線型回帰モデルは、対象と 1 つまたは複数の予測値との線型の関係に基づいて連続型対象を予測します。詳細は、10 章 [線型モデル in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



線型回帰は、データを要約する一般的な統計手法であり、予測された出力値と実際の出力値の違いを最小限にする直線または面を当てはめることにより予測を行います。



因子分析ノードには、データの複雑性を整理する強力なデータ分解手法が 2 種類あります。主成分分析 (PCA) : 入力フィールドの線型結合が検出されます。成分が互いに直交する (直角に交わる) 場合に、フィールドのセット全体の分散を把握するのに役立ちます。因子分析 : 一連の観測フィールド内の相関パターンを説明する基本因子が識別されます。どちらの手法でも、元のフィールド セットの情報を効果的に要約する少数の派生フィールドの検出が目標です。 [詳細は、10 章 因子分析ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



フィールド選択ノードで、(欠損値の割合などの) 諸基準に基づいて入力フィールドをスクリーニングして除去にかけ、指定した目標に相対的な残りの入力フィールドの重要度をランク付けします。たとえば、数百の潜在的入力フィールドを含むデータセットがあるとして、患者予後のモデリングにはどれが役に立つのでしょうか? [詳細は、4 章 フィールド選択ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



判別分析によって、ロジスティック回帰より厳密な仮説を立てることができますが、これらの仮説が一致した場合、ロジスティック回帰分析に対する様々な代替あるいは補足になります。 [詳細は、10 章 判別分析ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



ロジスティック回帰は、入力フィールドの値に基づいてレコードを分類する統計手法です。線型回帰と似ていますが、数値範囲ではなくカテゴリ対象フィールドを使用します。 [詳細は、10 章 ロジスティック ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



一般化線型モデルは、指定したリンク関数によって従属変数が因子および共変量と線型関係になるよう、一般線型モデルを拡張したものです。さらにこのモデルでは、非正規分布の従属変数を使用することができます。線型回帰、ロジスティック回帰、カウント データに関するログ線型モデル、そして区間打ち切り生存モデルなど、統計モデルの機能が数多く含まれています。 [詳細は、10 章 GenLin ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



Cox 回帰ノードを使用すると、打ち切りレコードの存在下でイベントまでの時間のデータの生存モデルを構築します。モデルは、対象のイベントが入力変数の指定の値で指定の時間 (t) に発生する確率を予測する生存関数を作成します。 [詳細は、10 章 Cox ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



サポート ベクター マシン (SVM) ノードを使用すると、オーバーフィットすることなく、データを 2 つのグループのいずれかに分類することができます。SVM は、非常に多数の入力フィールドを含むデータセットなど、広範なデータセットを処理することができます。詳細は、[15 章 SVM ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



ベイジ ネットワーク ノードを使用すると、観測された情報および記録された情報を実際の知識を組み合わせることによって確率モデルを作成し、発生 の 尤度を確立できます。ノードは主に分類に使用される Tree Augmented Naïve Bayes (TAN) および Markov Blanket ネットワークに焦点を当てています。詳細は、[7 章 Bayesian Network ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



SLRM (自己学習応答モデル) ノードを使用するとモデルを構築でき、単一または少数の新しいケースを使用して全データを使用するモデルの保持をすることなく、モデルの再見積もりを行うことができます。詳細は、[14 章 SLRM ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



時系列ノードは、時系列から指数平滑法、1 変量の自己回帰型統合移動平均法 (ARIMA)、および多変量 ARIMA (または転送関数) モデルを推測し、将来のパフォーマンスの予測を作成します。時系列ノードは、時間区分ノードによって常に先行される必要があります。詳細は、[13 章 時系列モデル作成ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



k が整数である場合、 k 最近隣 (KNN) ノードは、新しいケースを、予測領域の新しいケースに最も近い k 個のオブジェクトのカテゴリまたは値と関連付けます。類似したケースはお互いに近く、類似していないケースはお互いに離れています。詳細は、[16 章 KNN ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。

アソシエーション モデル

アソシエーション モデルでは、イベント、購入、属性など、1 つまたは複数のエンティティが 1 つまたは複数のその他のエンティティと関連するデータ内のパターンを検出します。モデルは、これらの関係性を定義するルール セットを構築します。データ内のフィールドは、入力および対象のいずれのフィールドとしても機能します。これらのアソシエーションは手動で検出できますが、アソシエーション ルール アルゴリズムはより迅速に検出が可能で、より複雑なパターンも検証できます。Apriori および Carma モデルが、アソシエーション アルゴリズムの使用例です。他にアソ

シエーション モデルの 1 つとしてシーケンス検出モデルがあり、時間構造データのシーケンス パターンを検索します。

アソシエーション モデルは、たとえば、商品 X を交友した顧客は Y と Z も購入するなど、複数の結果を予測する場合にもっと役立ちます。アソシエーション モデルは、特定の結論（何かを購入する意思決定など）と、一連の条件とを関連付けます。アソシエーション ルールのアルゴリズムは、どのような属性の間にも関連を成立させることができるという点で、より一般的なディシジョン ツリーのアルゴリズム (C5.0 や C&R Trees など) より勝っています。ディシジョン ツリーのアルゴリズムは、一つの結果にいたるルールを構築するのに対し、アソシエーション ルールのアルゴリズムは、それぞれが異なる結果にいたる多数のルールを見つけようとします。

アソシエーション ノード



Apriori ノードで、データからルール セットを抽出し、情報内容が最も充実したルールを引き出します。Apriori には、5 種類のルール選択方法があり、高度なインデックス作成方法を使用して、大きなデータセットが効率的に処理されます。大きな問題の場合は、一般に、Apriori の方が高速に学習できます。保持できるルール数に特に制限はありません。また、最大 32 の前提条件を持つルールを処理できます。Apriori では、入力フィールドと出力フィールドのすべてがカテゴリであることが必要ですが、この種類のデータに合わせて最適化されているので、よりよいパフォーマンスを実現します。詳細は、[12 章 Apriori ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



CARMA モデルは、入力または対象フィールドを指定しなくても、データからルールのセットを抽出します。Apriori とは対照的に、CARMA ノードは、前提条件サポートではなく、ルール サポート（前提条件と結果の両方のサポート）の構築の設定ができます。これは、生成されたルールをさまざまなアプリケーションで活用できることを意味します。たとえば、この休暇シーズンに販売促進する項目を結果とする、商品またはサービス（前提条件）のリストを調べることができます。詳細は、[12 章 CARMA ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



シーケンス ノードで、シーケンシャルな、または時間経過が伴うデータ内のアソシエーション ルールを検出します。予測可能な順序で起こる傾向にあるアイテム セットのリストを、シーケンスと呼びます。たとえば、顧客がひげそりとアフター シェービング クリームを購入した場合、その顧客は次の購入時にシェービング クリームを購入する可能性があります。シーケンス ノードは CARMA アソシエーション ルール アルゴリズムに基づいているため、効率的な 2 段階通過法でシーケンスが検出されます。詳細は、[12 章 シーケンス ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。

セグメント化モデル

セグメント化モデルでは、データを入力フィールドの類似したパターンを持つレコードのセグメント、またはクラスタに分割します。入力フィールドにのみ関心があるため、セグメント モデルには出力フィールドまたは対象フィールドの概念はありません。セグメント モデルの例として、Kohonen ネットワーク、K-Means クラスタリング、TwoStep クラスタリングおよび異常値検出があります。

「クラスタリング モデル」とも呼ばれるセグメント化モデルは、特定の結果が不明である場合に適しています（たとえば、詐欺の新しいパターンを識別する場合や、既存の顧客ベースから関心の対象となるグループを識別する場合です）。クラスタリング モデルは、類似したレコードのグループを識別し、そのグループに従ってレコードにラベルを付けます。この作業は、各グループとそれぞれの特性に関する事前の知識を活用せずに実行されます。これは、クラスタリング モデルと他のモデル作成手法との違いであり、クラスタリング モデルには、モデルが予測する定義済みの出力フィールドや対象フィールドはありません。これらのモデルには、正、誤という回答はありません。モデルの価値は、データのグループ構成を把握し、それらのグループについて役に立つ説明を提供できるかどうかで決まります。クラスタリング モデルは、クラスタやセグメントを作成するためによく利用されます。このクラスタやセグメントは、後の分析で入力として使用されます（たとえば、潜在的な顧客を、等質のサブグループに分類する方法です）。

セグメント化ノード



自動クラスタ ノードは、同様の特性を持つレコードのグループを識別するクラスタリング モデルを推定し、比較します。ノードは他の自動化モデル作成ノードと同じように動作し、複数の組み合わせのオプションを単一のモデル作成の実行で検証できます。モデルは、クラスタ モデルの有用性をフィルタリングおよびランク付けする基本的な指標を使用して比較し、特定のフィールドの重要度に基づいて指標を提供します。 [詳細は、5 章 自動クラスタ ノード in IBM SPSS Modeler 14.2 Modeling Nodes を参照してください。](#)



K-Means ノードで、データ セットが異なるグループ（つまりクラスタ）へ、クラスタリングされます。この方法で、固定数のクラスタを定義し、クラスタにレコードを繰り返し割り当てて、これ以上調整してもモデルが改善されなくなるまで、クラスタの中心を調整します。K-means では、結果を予測するのではなく、「非監視学習」として知られるプロセスが使用されます。 [詳細は、11 章 K-Means ノード in IBM SPSS Modeler 14.2 Modeling Nodes を参照してください。](#)



Kohonen ノードは、ニューラル ネットワークの一種であり、データ セットをクラスタ化して異なるグループを形成する目的で使用できます。ネットワークの学習が完了すると、類似のレコードは出力マップで互い近くに表示され、違いの大きいレコードほど離れたところに表示されます。強度の高いユニットを識別するために生成されたモデル内で、各ユニットが獲得した観察の数値を調べることができます。これは、適切なクラスタ数についてのヒントになる場合があります。詳細は、[11 章 Kohonen ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



TwoStep ノードで、2 段階のクラスタ化手法が使用されます。最初のステップでは、データを 1 度通過させて、未処理の入力データを管理可能な一連のサブクラスタに圧縮します。2 番目のステップでは、階層クラスタ化手法を使用して、サブクラスタをより大きなクラスタに結合させていきます。TwoStep には、学習データに最適なクラスタ数を自動的に推定するという利点があります。また、フィールド タイプの混在や大規模データ セットも効率よく処理できます。詳細は、[11 章 TwoStep クラスタ ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。



異常値検出ノードで、「正常な」データのパターンに合致しない異常ケースや外れ値を識別します。このノードで、外れ値が既知のパターンに当てはまらなかったり、何を探しているのかはつきりしなかったりする場合でも、外れ値を識別できます。詳細は、[4 章 異常値検出ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。

データベース内マイニング モデル

SPSS Modeler は、Oracle Data Miner、IBM DB2 InfoSphere Warehouse、および Microsoft Analysis Services などのデータベース ベンダーから入手できる、データ マイニングとモデル作成ツールとの統合をサポートしています。データベース内および SPSS Modeler アプリケーション内のモデル、スコアおよびストア モデルすべての作成が可能です。詳細は、製品 DVD に収録されている、『SPSS Modeler データベース内マイニング ガイド』を参照してください。

IBM SPSS Statistics モデル

コンピュータに IBM® SPSS® Statistics をインストールしライセンスが付与されている場合、SPSS Modeler 内から特定の SPSS Statistics ルーチンにアクセスおよび実行して、モデルを作成およびスコアリングできます。詳細は、[8 章 IBM SPSS Statistics ノードの概要 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

詳細情報

モデル作成アルゴリズムについての詳細なマニュアルも利用できます。詳細は、製品 DVD に収録されている、『SPSS Modeler アルゴリズム ガイド』を参照してください。

データマイニングの例

データマイニングについて学習する現実的で最良の方法は、最初に例を使うことです。『IBM® SPSS® Modeler アプリケーション ガイド』には多数のアプリケーション例があり、特定のモデル作成手法および技術に関して、目的別に簡単に説明しています。詳細は、[1 章 p.4 アプリケーションの例](#) を参照してください。

ストリームの構築

ストリーム構築の概要

IBM® SPSS® Modeler を使用したデータマイニングでは、ストリームと呼ばれる、一連のノードを通過するデータの流を処理する過程に注目します。この一連のノードは、データに関して行う処理や操作を表し、ノード間のリンクはデータの流（データフロー）の方向を示します。通常は、データストリームを使って SPSS Modeler にデータを読み込み、そのデータを一連の処理や操作を経由して、テーブルやビューアなどに出力します。

たとえば、データソースを開いて新しいフィールドを追加し、そのフィールドの値に基づいてレコードを選択して、結果をテーブルに表示する場合を考えてみましょう。この場合、データストリームは次の4つのノードで構成されます。



可変長ノード：データをデータソースから読み込むために使用します。



フィールド作成ノード：データセットに新しく計算されたフィールドを追加するために使用します。



条件抽出ノード：データストリームから除外するレコードの選択基準を設定するために使用します。



テーブルノード：操作結果を画面に表示するために使用します。

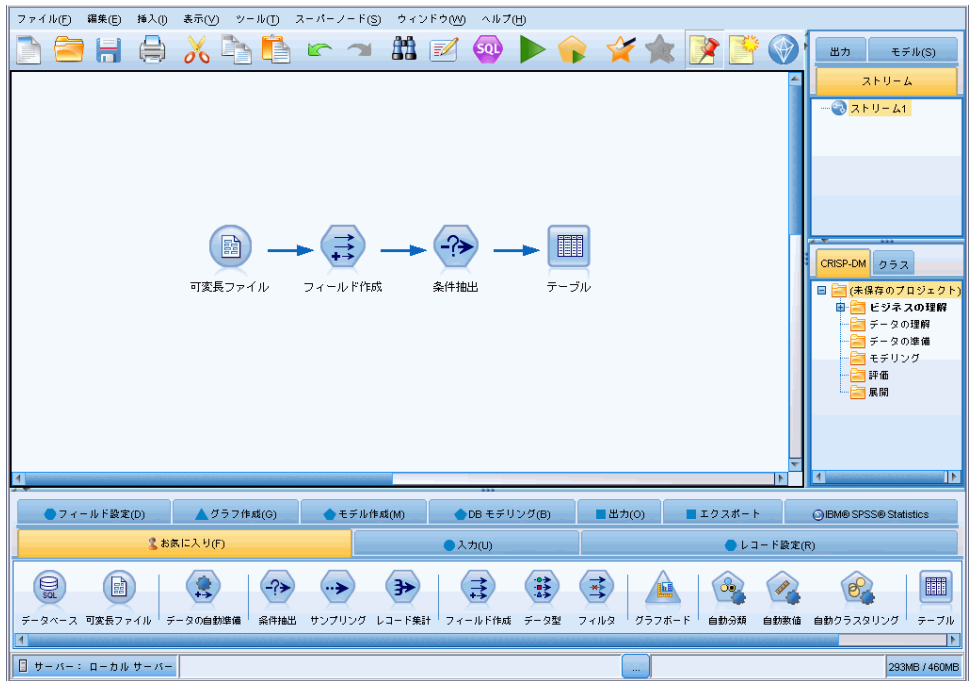
データストリームの作成

IBM® SPSS® Modeler 特有の優れたインターフェイスにより、データストリームのダイアグラムを使って、データマイニング作業を視覚的に行うことができます。基本的には、データストリームは次の手順で作成していきます。

- ストリーム領域にノードを追加する。
- ノードを接続してストリームを形成する。

- 任意のノードまたはストリーム オプションを指定する。
- ストリームを実行する。

図 5-1
ストリーム領域上の完成したストリーム



このセクションでは、ノードを使ったより複雑なデータ ストリームの作成方法を詳細に説明していきます。また、ノードやストリームのオプションや設定方法も取り上げていきます。SPSS Modeler に付属のデモ データ（プログラム インストール ディレクトリの Demos フォルダ）を使ったストリーム構築の例は、「アプリケーションの例」（p.4）を参照してください。

ノードの作業

ノードは、IBM® SPSS® Modeler でデータを探索するために用いられます。作業領域内のさまざまなノードが、それぞれさまざまなオブジェクトや処理を表しています。SPSS Modeler ウィンドウの下部にあるパレットには、ストリームの構築に使われるすべてのノードが用意されています。

スーパーノードには、いくつかの種類があります。**入力ノード**はデータをストリームに投入し、ノード パレットの [入力] タブにあります。**プロセス ノード**は、各データ レコードおよびフィールドで操作を行い、パレットの [レコード設定] タブおよび [フィールド設定] タブに表示されます。**出力ノード**は、さまざまなデータの出力、グラフ、モデル結果を生成し、ノード パレットの [グラフ] タブ、[出力] タブ、[エクスポート]

タブに表示されます。**モデル作成ノード**は統計アルゴリズムを使用してモデル ナゲットを作成し、ノード パレットの [モデル作成] タブ、(有効化されている場合は) [DB モデリング] タブに表示されます。詳細は、3 章 p. 16 **ノード パレット** を参照してください。

ノード間を接続してストリームを形成し、それを実行することによって、関係を視覚化して、結論を導き出します。ストリームはスクリプトに似ています。ストリームは保存して、さまざまなデータ ファイルで再利用できます。

ストリーム データを処理する実行可能なノードは、**ターミナル ノード**と呼ばれます。モデル作成ノードまたは出力ノードがストリームまたはストリーム枝の最後に表示される場合、ターミナル ノードとなります。高度なノードをターミナル ノードに接続できません。

注： ノード パレットはユーザー設定できます。詳細は、12 章 p. 267 **ノード パレットのカスタマイズ** を参照してください。

ストリームへのノードの追加

ノード パレットからストリームにノードを追加するには、次の複数の方法があります。

- パレットのノードをダブル クリックする。注： ノードをダブル クリックすると、現在のストリームとノードが自動的に接続されます。詳細は、p. 46 **ストリーム内のノードの接続** を参照してください。
- パレットからストリーム領域にノードをドラッグ アンド ドロップする。
- パレット上のノードをクリックして、次にストリーム領域をクリックする。
- IBM® SPSS® Modeler の [挿入] メニューから適切なオプションを選択する。

ノードをストリーム領域に追加したら、そのノードをダブル クリックすると、ノードに関するダイアログ ボックスが表示されます。使用できるオプションは、追加するノードの種類によって異なります。ダイアログ ボックス内の特定のコントロールについての情報を表示するには、[ヘルプ] ボタンをクリックします。

ノードの除去

データ ストリームからノードを除去するには、ノードをクリックして Del キーを押すか、または、右クリックしてメニューから [削除] を選択します。

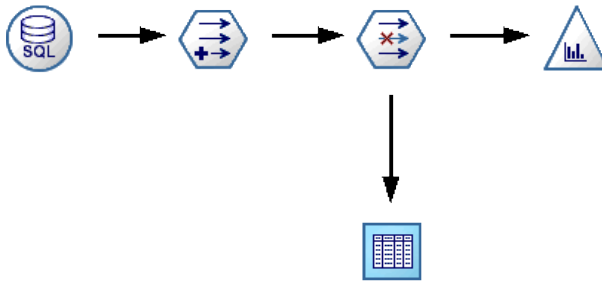
ストリーム内のノードの接続

ストリーム領域に追加した各ノードは、接続しないとデータ ストリームにはなりません。ノード間の接続は、ある操作から別の操作へとデータの流れる方向を表しています。ノード間を接続してデータ ストリームを作成するには、さまざまな方法があります。ダブル クリック、マウスの中央ボタンの使用、または手作業

ダブル クリックでノードを追加、接続するには

データ ストリームを形成するもっとも簡単な方法は、パレット上のノードをダブル クリックすることです。この操作を行うと、ストリーム領域上で選択されているノードに、ダブル クリックしたノードが自動的に接続されます。たとえば、ストリーム領域にデータベース ノードがある場合、このノードを選択した後に、フィールド作成ノードなどのパレット上のノードをダブル クリックします。これによって、既存のデータベース ノードにフィールド作成ノードが自動的に接続されます。この手順を、ヒストグラムやテーブル ノードなどのターミナル ノードに到達するまで繰り返すことができます。ターミナル ノードでは、任意のノードが最後の非ターミナル ノードに接続されます。

図 5-2
パレット上のノードをダブル クリックして作成されたストリーム



マウスの中央ボタンを使ってノードを接続するには

ストリーム領域で、あるノードをマウスの中央ボタンを使ってクリックした後、別のノードにドラッグします(マウスの中央ボタンがない場合は、Alt キーを押しながらクリック、ドラッグしてください)。

図 5-3
マウスの中央ボタンを使ったノードの接続



ノードを手作業で接続するには

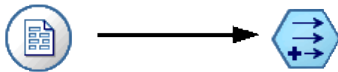
マウスの中央ボタンがなく、手作業でノードを接続したい場合は、ノードのポップアップメニューを使って、ストリーム領域上の他のノードに接続することができます。

- ▶ 接続を開始するノードを右クリックします。右クリックすると、ノードメニューが表示されます。
- ▶ メニューで [接続] をクリックします。
- ▶ 始点ノードとカーソルの両方に接続アイコンが表示されます。領域上の目的のノードをクリックすると、2つのノードが接続されます。

図 5-4
ポップアップメニューの [接続] オプションを使ったノードの接続



図 5-5
接続済みノード



ノードを接続する際には、注意しなければいけない事柄があります。次のような接続を作成しようとすると、エラーメッセージが表示されます。

- 入力ノードへの接続
- ターミナル ノードからの接続
- ノードの接続数が最大入力接続数を超える場合
- すでに接続されている 2 つのノードの接続
- データがすでに通過したノードに戻る場合（循環性）

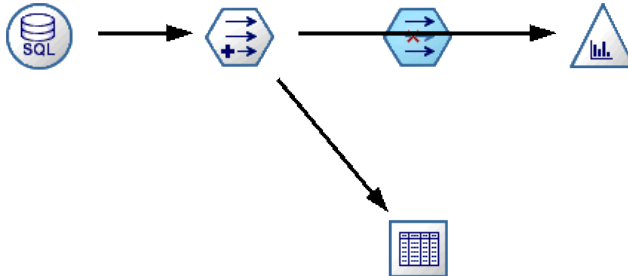
ストリーム内でのノードの迂回

データ ストリーム内でノードを迂回する場合、そのノードのすべての入力接続と出力接続が、そのノードの入力ノードから出力ノードに直接接続されます。ノードに入力接続または出力接続がない場合、ノードの接続はすべて、経路は変更されずに破棄されます。

たとえば、新規フィールドを作成し、フィールドのフィルタリングを行った後、結果をヒストグラムとテーブルに表示するストリームを考えてみましょう。フィールドをフィルタリングする前にもデータのグラフとテーブルを参照したい場合、ストリームに新たにヒストグラム ノードとテーブル ノードを追加するか、またはフィルタ ノードを迂回することがで

きます。フィルタ ノードを迂回すると、グラフとテーブルはフィールド作成ノードから直接接続されます。フィルタ ノードの接続は解除されます。

図 5-6
前に接続していたフィルタ ノードの迂回



ノードを迂回するには

- ▶ ストリーム領域で、迂回するノードをマウスの中央ボタンでダブルクリックします。または、Alt キーを押しながらダブルクリックすることもできます。

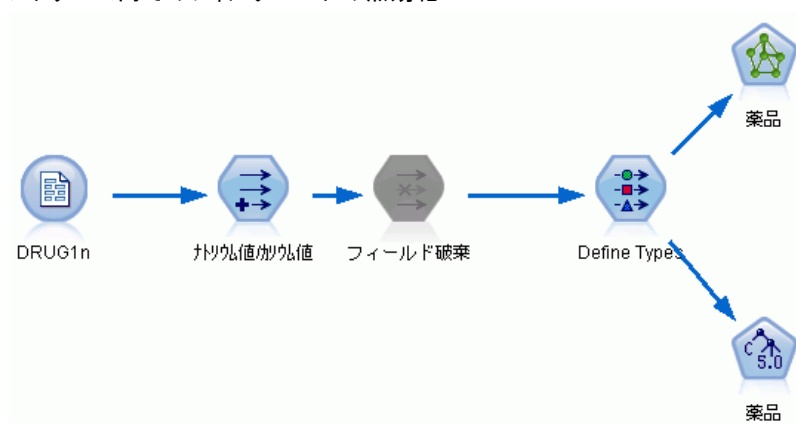
注： この操作を元に戻すには、[編集] メニューの [元に戻す] をクリックするか、または Ctrl+Z キーを押してください。

ストリーム内でのノードの無効化

ストリーム内の 1 回の入力によるプロセス ノードを無効化できます。その結果として、ストリームの実行時にノードが無視されます。これにより、ノードを削除またはバイパスする必要がなくなり、残りのノードを接続したままにできます。ノード設定を開いて編集することもできますが、ノードを再度有効化した後で変更が有効となります。

たとえば、複数のフィールドを除外するストリームがあり、減らされたデータ セットによりモデルを作成するとします。フィールドを除外せずに同じモデル作成したい場合、モデルの結果が改善されるかどうかを確認するためにフィルタ ノードを無効化することができます。フィルタ ノードを無効化すると、モデル作成ノードへの接続は、フィールド作成ノードからデータ型ノードに直接通過します。

図 5-7
ストリーム内でのフィルタ ノードの無効化



ノードを無効化するには

- ▶ ストリーム領域で、無効にするノードを右クリックします。
- ▶ ポップアップ メニューの [ノードの無効化] をクリックします。

または、[編集] メニューの [ノード] > [ノードの無効化] を選択することもできます。ストリームにノードを戻す場合、同じようにして [ノードの有効化] を選択します。

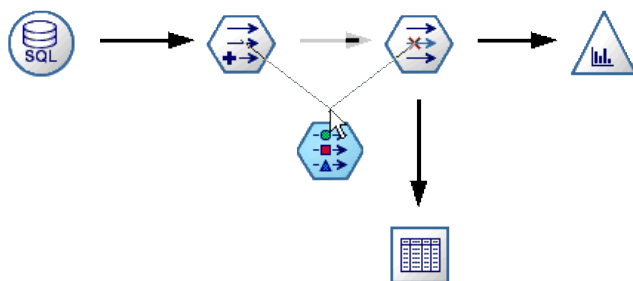
注： この操作を元に戻すには、[編集] メニューの [元に戻す] をクリックするか、または Ctrl+Z キーを押してください。

この操作を元に戻すには、[編集] メニューの [元に戻す] をクリックするか、または Ctrl+Z キーを押してください。

既存の接続へのノードの追加

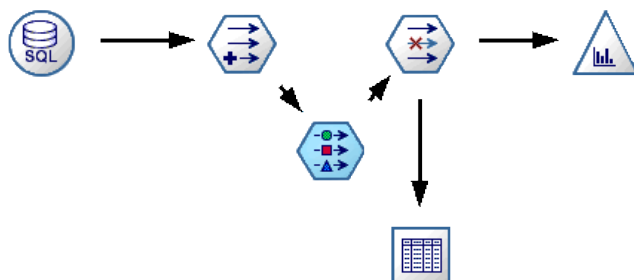
2 つのノードを接続する矢印をドラッグすることにより、接続された 2 つのノードの間に新しいノードを追加できます。

図 5-8
接続された 2 つのノード間への新規ノードの追加



- ▶ マウスの中央ボタンを使用して、接続の矢印をクリックし、ノードの挿入先までドラッグします。中央ボタンがない場合は、Alt キーを押しながらクリック アンド ドラッグすることもできます。

図 5-9
新規ストリーム



- ▶ ボタンを押したまま挿入するノードまで接続をドラッグし、マウス ボタンを離します。

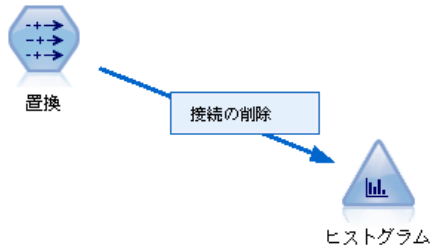
注： ノードを迂回すれば、ノードから新しい接続を削除して、元の接続を復元できます。

ノード間の接続の削除

2 つのノード間の接続を削除する手順は、次のとおりです。

- ▶ 接続の矢印を右クリックします。
- ▶ メニューで [接続の削除] をクリックします。

図 5-10
ストリーム内のノード間の接続の削除



ノードとのすべての接続を削除するには、次のいずれかを実行します。

- ノードを選択して、F3 を押します。
- ノードを選択して、メインメニューから次の各項目をクリックします。
編集 > ノード > [接続の解除(S)]

ノードのオプションの設定

ノードの作成と接続が完了したら、オプションを設定してノードをカスタマイズできます。ノードを右クリックして、次のいずれかのオプションを選択します。

図 5-11
ノードのポップアップメニュー



- 選択したノードのダイアログ ボックスを開くには、[編集] をクリックします。
- あるノードから別のノードに手作業で接続するには、[接続] をクリックします。
- 選択したノードに対するすべてのリンクを削除するには、[接続の解除] をクリックします。
- [編集] ダイアログ ボックスの [注釈] タブを表示するには、[名前の変更と注釈] をクリックします。
- [新規コメント] をクリックして、ノードに関連するコメントを追加します。詳細は、[p. 79 ノードおよびストリームへのコメントおよび注釈の追加](#) を参照してください。
- [ノードの無効化] をクリックして、処理時にノードを「非表示」にします。処理する場合にノードを再度表示するには、[ノードの有効化] をクリックします。詳細は、[p. 48 ストリーム内でのノードの無効化](#) を参照してください。

- ストリーム領域からノードを削除するには、[切り取り] または [削除] をクリックします。注： [切り取り] をクリックした場合、そのノードを貼り付けることができます。ただし、[削除] を選択した場合は、貼り付けることはできません。
- 接続を持たないノードのコピーを作成するには、[コピー] をクリックします。コピーしたノードは、既存のストリームに追加することも、新しいストリームに追加することもできます。
- 以前に保存したノードを開いて、そのオプション設定を現在選択しているノードに取り込むには、[ノードのロード] をクリックします。注： この操作を行うには、ノードの種類が同じでなければなりません。
- [ノードの取得] をクリックして、接続した IBM® SPSS® Collaboration and Deployment Services Repository からノードを取得します。 [詳細は、5 章 IBM SPSS Collaboration and Deployment Services Repository 内のオブジェクトへのアクセス in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。](#)
- ノードの詳細をファイルに保存するには、[ノードの保存] をクリックします。保存したノード定義は、同じ種類のノードにしかロードできません。
- [ノードを格納] をクリックして、接続した IBM SPSS Collaboration and Deployment Services Repository の選択したノードを格納します。 [詳細は、5 章 IBM SPSS Collaboration and Deployment Services Repository 内のオブジェクトへのアクセス in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。](#)
- 選択したノードのキャッシュに関するオプションを設定するには、[キャッシュ] をクリックしてメニューを表示します。
- データの新規ソースへのマッピングや必須フィールドの指定などのオプションを設定するには、[データ マッピング] をクリックしてメニューを表示します。
- 現在のストリームにスーパーノードを作成するには、[スーパーノードの作成] をクリックしてメニューを表示します。 [詳細は、9 章 スーパーノードの作成 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。](#)
- 選択したノードを置き換えるには、[ユーザー入力ノードの生成] をクリックします。このノードは、元のノードと同じフィールドを持ちます。 [詳細は、2 章 ユーザー入力ノード in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。](#)
- 選択したノードの下流にあるすべてのターミナル ノードを実行するには、[ここから実行] をクリックします。

ノードのキャッシュ オプション

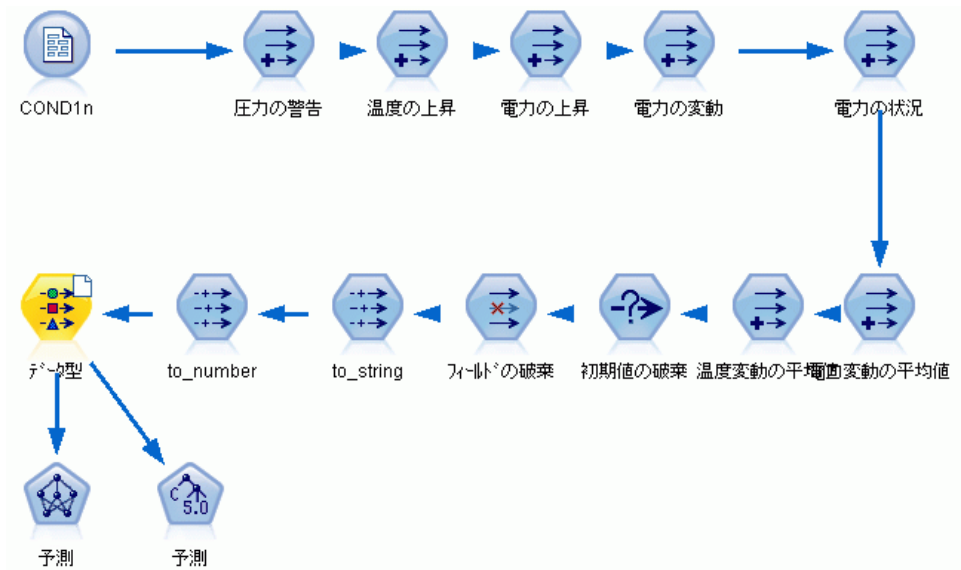
ストリームの実行を最適化するために、ターミナル ノード以外の任意のノードに**キャッシュ**を設定することができます。ノードにキャッシュを設定すると、次のデータ ストリームの実行時にそのノードを通過するデータがキャッシュされます。データがキャッシュされたら、以降のデータはデータ ソースの代わりに、キャッシュ（ディスクの一時ディレクトリに格納）から読み込まれます。

キャッシュは、ソート、結合、集計など、時間のかかる操作に最も有用です。たとえば、データベースから販売データを読み込む入力ノードと、販売地域ごとに集計するレコード集計ノードがあるとします。この場合、データ セット全体ではなく集計済みデータをキャッシュしたいので、入力ノードではなくレコード集計ノードにキャッシュを設定します。

注： 入力ノードでキャッシュする場合、IBM® SPSS® Modeler に読み込まれるように元のデータのコピーを保存するため、多くの環境でパフォーマンスは改善されません。

キャッシュを有効にしたノードの右上には、小さい文書アイコンが表示されます。データがキャッシュされると、この文書アイコンが緑に変わります。

図 5-12
新規作成フィールドを格納するデータ型ノードのキャッシュ



キャッシュを有効にするには

- ▶ ストリーム領域でノードを右クリックし、メニューの [キャッシュ] をクリックします。
- ▶ キャッシュ設定用のサブメニューから、[使用する] を選択します。

- ▶ キャッシュを解除するには、ノードを右クリックして、キャッシュ設定用のサブメニューから [使用しない] を選択します。

データベース内でのノードのキャッシュ

データベース内で実行されるストリームの場合、ファイル システムでなくデータベース内の一時テーブルへ、データを中流でキャッシュできます。この機能を SQL 最適化と組み合わせると、パフォーマンスが著しく向上する可能性があります。たとえば、データ マイニング ビューを作成するために複数のテーブルを併合するストリームからの出力をキャッシュし、必要に応じて再使用できます。すべての下流ノードで自動的に SQL を生成するようにすると、パフォーマンスがさらに向上する可能性があります。

256 文字を超える文字列でキャッシュするデータベースを使用する場合、キャッシュ ノードの上流に、フィールド値が読み込まれるデータ型ノードがあることを確認してください。それにより、一時テーブルの該当する列を適切な幅に設定し、文字列を調整します。

データベースのキャッシングを利用するには、SQL 最適化とデータベースキャッシングの両方を有効にする必要があります。Client の設定は、Server の最適化設定によって無効になるので注意してください。詳細は、[12 章 p. 263 最適化オプションの設定](#) を参照してください。

データベースのキャッシングが有効にされると、任意の非ターミナルノードを右クリックするだけでその場所のデータがキャッシュされ、次にストリームが実行されるときに自動的に、データベース内にキャッシュが直接作成されます。データベースキャッシングまたは SQL 最適化が有効にされていないと、キャッシュは、代わりにファイル システムへ書き出されます。

注： 次のデータベースは、キャッシュのために一時テーブルをサポートします。DB2、Netezza、Oracle、SQL Server および Teradata。その他のデータベースでは、データベースキャッシングに通常のテーブルが使用されます。SQL コードを特定のデータベース向けにカスタマイズできます。サポートに連絡してください。

キャッシュを取り消すには

ノードに表示されている文書アイコンが白い場合、キャッシュが空であることを表しています。キャッシュがいっぱいになると、文書アイコンは濃い緑色になります。キャッシュの内容を置き換える場合は、最初にキャッシュを取り消してからデータ ストリームを再実行して、キャッシュにデータを格納しなおす必要があります。

- ▶ ストリーム領域でノードを右クリックし、メニューの [キャッシュ] をクリックします。

- ▶ キャッシュ設定用のサブメニューから、[取り消し] を選択します。

キャッシュを保存するには

キャッシュの内容は、IBM® SPSS® Statistics データ ファイル (*.sav) として保存できます。その後、ファイルをキャッシュとして再ロードしたり、キャッシュ ファイルをデータ ソースとして使用するノードを設定することができます。さらに、別のプロジェクトで保存したキャッシュをロードすることもできます。

- ▶ ストリーム領域でノードを右クリックし、メニューの [キャッシュ] をクリックします。
- ▶ キャッシュ設定用のサブメニューから、[キャッシュの保存] を選択します。
- ▶ [キャッシュの保存] ダイアログ ボックスで、キャッシュ ファイルを保存する場所を指定します。
- ▶ [ファイル名] テキスト ボックスに名前を入力します。
- ▶ [ファイルの種類] リストで [*sav] が選択されていることを確認し、[保存] をクリックします。

キャッシュをロードするには

キャッシュ ファイルをノードから削除する前に保存した場合、それを再ロードすることができます。

- ▶ ストリーム領域でノードを右クリックし、メニューの [キャッシュ] をクリックします。
- ▶ キャッシュ設定用のサブメニューから、[キャッシュのロード] を選択します。
- ▶ [キャッシュのロード] ダイアログ ボックスで、目的のキャッシュ ファイルを選択し、[ロード] をクリックします。

ノードのデータをプレビューする

ストリーム構築時に予測していたようにデータが変更されているか確認するために、各重要なステップごとにテーブル ノードを使用してデータを実行する場合があります。このような手間を省くために、ノード適用時に作成されるデータのサンプルを表示する各ノードからプレビューを生成し、各ノードの構築時間を短縮できます。

モデル ナゲットのノード上流の場合、プレビューでは入力フィールドを表示します。モデルナゲット、またはナゲットのノードの下流（ターミナル ノードを除く）の場合、プレビューでは入力フィールドと生成されたフィールドが表示されます。

表示される行のデフォルト数は 10 行です。ただし、ストリームのプロパティで変更できます。詳細は、[p. 58 ストリームのオプションの設定](#) を参照してください。

図 5-13
モデル ナゲットからのデータプレビュー

	year_built	volume_interior	volume_other	lot_size	taxable_value	\$XR-taxable_value	\$XRE-taxable_value
1	979	166	11	100	90500	105523.184	1526.604
2	988	603	73	497	420000	355497.802	14929.477
3	987	303	75	91	152500	147713.631	1733.849
4	926	228	12	55	92000	69313.952	8116.212
5	988	666	145	441	390000	367551.491	16353.591
6	970	563	60	800	435000	395173.708	4759.398
7	902	355	41	225	130000	109865.220	3694.157
8	972	468	78	625	380500	334714.268	6367.557
9	986	315	14	107	158000	145149.106	1263.839
10	975	282	27	141	134000	138747.302	1892.730

[ノードの生成] メニューから、複数の種類のノードを作成できます。詳細は、[6 章 テーブル ブラウザ in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

ノードをロックする

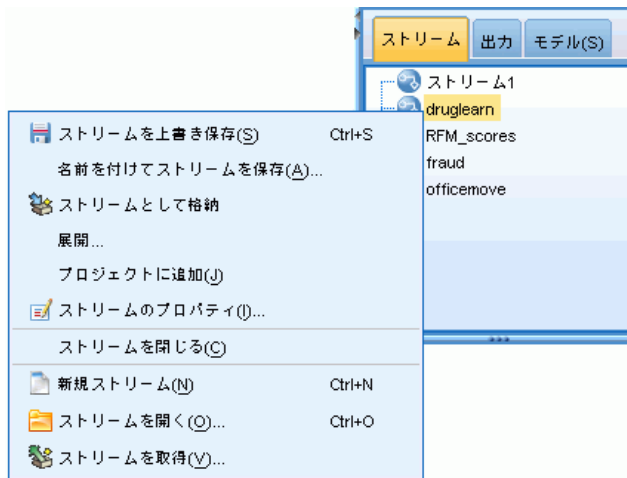
その他のユーザーがストリームの 1 つまたは複数のノード設定を修正できないようにするために、スーパーノードという特別な種類のノードにノードをカプセル化して、パスワード保護を適用してノードをロックすることができます。詳細は、[9 章 スーパーノードのロック in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

ストリームの作業

ストリーム領域上の入力ノード、プロセス ノード、およびターミナルノードを接続したら、データ ストリームの作成は完了です。ノードの集合となるストリームは、保存したり、注釈を付けたり、プロジェクトに追加することができます。また、最適化、日付と時刻の設定、パラメータ、およびスクリプトなどのさまざまなオプションを設定することができます。これらのプロパティの詳細は後述します。

IBM® SPSS® Modeler では、複数のデータ ストリームを同時に使用、変更することができます。メイン ウィンドウの右側には、現在表示しているストリーム、出力、およびモデルを把握するために役立つマネージャ (管理) パネルが用意されています。マネージャ パネルが表示されない場合、[表示] メニューの [マネージャ] をクリックし、[ストリーム] タブをクリックします。

図 5-14
マネージャ パネルの [ストリーム] タブとポップアップ メニュー オプション



このタブから、次のような作業を行うことができます。

- ストリームへのアクセス。
- ストリームの保存。
- 現在のプロジェクトへのストリームの保存。
- ストリームを閉じる。
- 新規ストリームを開く。
- IBM SPSS Collaboration and Deployment Services リポジトリでストリームの保存と検索を行う (ユーザーのサイトで利用可能な場合)。詳細は、[9 章 p.175 IBM SPSS Collaboration and Deployment Services Repository](#) についてを参照してください。

これらの設定を利用するには、[ストリーム] タブでストリームを右クリックしてください。

ストリームのオプションの設定

現在のストリームに対して、さまざまなオプションを指定できます。その大半は、CLEM 式に適用されます。

ストリーム オプションを設定するには

- ▶ [ファイル] メニューから、[ストリームのプロパティ] を選択します（またはマネージャ パネルの [ストリーム] タブからストリームを選択し、ポップアップメニューを右クリックして [ストリームのプロパティ] をクリックします）。
- ▶ [オプション] タブをクリックします。

また、[ツール] メニューから次の各項目を選択することもできます。
ストリームのプロパティ > オプション

図 5-15
[ストリームのプロパティ] ダイアログ ボックスの [オプション] タブ



計算フォーマット: CLEM 式で使用する三角関数の角度単位として、[ラジアン] または [次数] を選択します。

日付/時刻のインポートフォーマット: 日付/時刻フィールドの日付/時刻ストレージ（記憶域）を使用するか、または日付/時刻を文字列変数としてインポートするかを選択します。

日付のフォーマット：日付ストレージ（記憶域）フィールドが使用する、または文字列が CLEM 日付関数によって日付として解釈された場合に使用する日付のフォーマットを選択します。

時間のフォーマット：時間ストレージ（記憶域）フィールドが使用する、または文字列が CLEM 時間関数によって時間として解釈された場合に使用する時間のフォーマットを選択します。

日/分をロールオーバー：時間のフォーマットで、マイナスの時差を、前日の時間、または当日の時間を表すものとして解釈するかどうかを選択します。

数字の表示フォーマット：標準 (####.###)、科学的 (#.###E+##)、または通貨表記フォーマット (\$###.##) から選択できます。

小数点以下の表示（標準、科学的、通貨）：数値の表示フォーマットで、実数を表示または印刷するとき使用する、小数点以下の桁数を指定します。このオプションは、各表示フォーマットごとに別々に指定します。

小数点記号：桁区切り記号として、カンマ (,)、またはピリオド (.) を選択します。

グループ化記号：数字の表示フォーマットで、値をグループ化するのに使用する記号を選択します（例：3,000.00 のカンマ）。オプションには、なし、ピリオド、カンマ、スペース、およびロケール定義（現在のロケールがデフォルトとして使用されている場合）。

基準日付（1 月 1 日）：CLEM 日付関数で単一日付を扱う場合に、使用する基準の日付（常に 1 月 1 日）を選択します。

次世紀までのロールオーバーの閾値：2 桁で年号を表す場合に、前の世紀の年号かどうかを判断するための、閾値とする年を指定します。たとえば、閾値の年として 1930 を指定すると 05/11/02 が 2002 年であると想定します。同じ設定で 30 以降の日付に 20 世紀を使用するため、05/11/73 は 1973 年であると想定されます。

文字コード：テキスト文字コードについてストリームのデフォルトの方法を指定します（注：可変長ファイル ソース ノードおよびファイル エクスポート ノードのみに適用；他のノードではこの設定を使用しません。多くのデータ ファイルには埋め込み文字コード情報があります）。システム デフォルトまたは UTF-8 のいずれかを選択できます。システム デフォルトは、Windows のコントロール パネル、または分散モードで動作している場合はサーバー コンピュータから指定できます。詳細は、[B 付録 p.298 IBM SPSS Modeler 内での Unicode のサポート](#) を参照してください。

データプレビューで表示する行の最大数：データのプレビューがノードに必要な場合に表示する行の最大数。詳細は、[p.56 ノードのデータをプレビューする](#) を参照してください。

名義からフィールドの最大メンバー数:名義型（セット型）フィールドの最大メンバー数を指定する場合に選択します。以降のフィールドのデータ型は、**データ型不明**になります。このオプションは、特に大きい名義型フィールドで作業を行うような場合に役立ちます。注：測定レベルが**【データ型不明】**に設◆◆されると、役割が自動的に**【なし】**に設定されます。つまり、このフィールドをモデリングには利用できません。

ニューラル ネットワーク、Kohonen、および K-Means モデルのセット サイズを制限：ニューラル ネットワーク、Kohonen ネットワーク、および K-Means モデル作成に使われる名義型フィールドの、最大メンバー数を指定する場合に選択します。デフォルトは 20 です。以降のフィールドは無視され、フィールドに問題があることを示す警告が発せられます。

ルールセットの評価：ルールセット モデルの評価方法を決定します。デフォルトでは、ルールセットは**【評決】**を使用して個別のルールからの予測を組み合わせ、最終的な予測を決定します。ルールセットが該当する最初のルールを使用するように設定するには、**【最初のヒット】**を選択します。[詳細は、6 章 ルール セットのモデル ナゲット in IBM SPSS Modeler 14.2 Modeling Nodes を参照してください。](#) このオプションは、ディシジョン リスト モデルには適用されません。このモデルでは、アルゴリズムに定義されたように、最初のヒットが常に使用されます。

実行時に入力ノードをリフレッシュ:現在のストリームの実行時に、すべての入力ノードを自動的にリフレッシュする場合に指定します。この操作は、現在のストリーム上にあるすべての入力ノード（ユーザー入力ノードを除く）が自動的にリフレッシュされることを除いては、入力ノードで**【リフレッシュ】** ボタンをクリックするのと同じ働きをします。

注：このオプションを選択すると、データが変更されていない場合でも、下流にあるノードのキャッシュが消去されます。キャッシュの消去はストリームの実行ごとに一度だけ行われます。つまり、1 回の実行において下流のキャッシュを一時ストレージ（記憶域）として使用することができます。たとえば、複雑なフィールド作成操作の後に中流にキャッシュを作成し、このフィールド作成ノードの下流に複数のグラフやレポートを追加した場合を考えてみましょう。ストリームの実行時に、フィールド作成ノードのキャッシュは消去され、その後最初のグラフまたはレポートがキャッシュに格納されます。以降のターミナル ノードは、そのデータをフィールド作成ノードのキャッシュから読み込みます。

出力中のフィールドと値ラベルを表示する:テーブル中の フィールド、値ラベル、グラフ、およびその他の出力を表示します。ラベルが存在しない場合、フィールド名とデータ値が代わりに表示されます。デフォルトでは、ラベルはオフになっていますが、IBM® SPSS® Modeler の他の場所で必要になった場合には、個々のラベルごとにオンとオフを切り替えることができます。ツールバーから利用できるトグル ボタンを使って、出力ウィンドウにラベルを表示させることもできます。

図 5-16
フィールドラベルと値ラベルの切り替えに使われるツールバー アイコン



デフォルトとして保存: 指定されたオプションは、現在のストリームに対してだけ適用されます。このボタンをクリックすると、これらのオプションをすべてのストリームのデフォルトとして設定します。

ストリーム レイアウトのオプションの設定

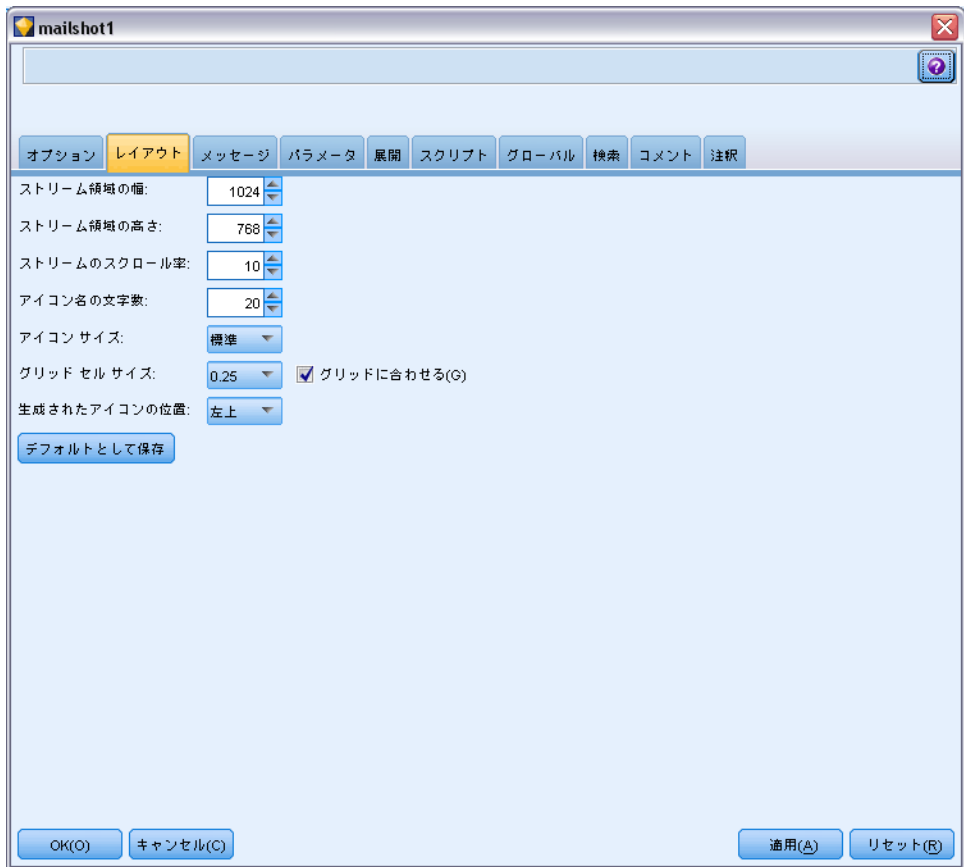
[ストリームのプロパティ] ダイアログ ボックスの [レイアウト] タブでは、ストリーム領域の表示や使用に関するさまざまなオプションを指定することができます。

レイアウト オプションを設定するには

- ▶ [ファイル] メニューから、[ストリームのプロパティ] を選択します（またはマネージャ パネルの [ストリーム] タブからストリームを選択し、ポップアップ メニューを右クリックして [ストリームのプロパティ] をクリックします）。
- ▶ [レイアウト] タブをクリックします。

また、[ツール] メニューから次の各項目を選択することもできます。
ストリームのプロパティ > レイアウト

図 5-17
[ストリームのプロパティ] ダイアログ ボックスの [レイアウト] タブ



ストリーム領域の幅 :ストリーム領域の幅をピクセル単位で指定します。

ストリーム領域の高さ :ストリーム領域の高さをピクセル単位で指定します。

ストリームのスクロール率 : ストリーム領域のスクロール率を指定して、ノードが領域上のある場所から別の場所へドラッグされる場合に、ストリーム領域ウィンドウのスクロールの速度を制御します。大きい数字を指定すると、スクロール率が高速になります。

アイコン名の文字数 :ストリーム領域上のノード名の文字数制限を指定します。

アイコン サイズ :ストリーム領域に大きいノード アイコン (標準) を表示するか、または小さいノード アイコンを表示するかを選択します。

グリッド セル サイズ : リストからグリッド セル サイズを選択します。ここに指定した数値は、ストリーム領域上のノードを、目に見えないグリッドを使って配置するために用いられます。デフォルトは 0.25 です。

グリッドに合わせる: アイコンを目に見えないグリッド パターンに沿って並べる場合に選択します (デフォルト)。

生成されたアイコンの位置: モデル ナゲットから生成するノードのアイコンを配置するキャンバスの場所を選択します。デフォルトは左上です。

デフォルトとして保存: 指定されたオプションは、現在のストリームに対してだけ適用されます。このボタンをクリックすると、これらのオプションをすべてのストリームのデフォルトとして設定します。

ストリーム操作メッセージの表示

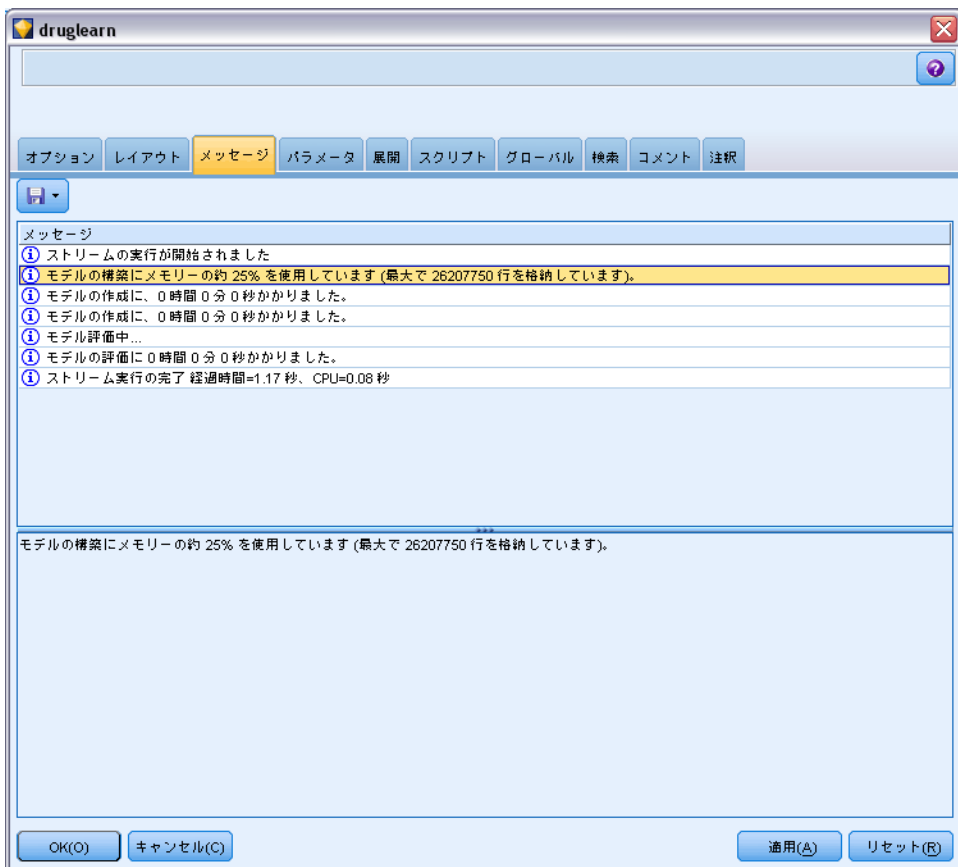
ストリームの実行、最適化、およびモデル構築および評価の時間などのストリーム操作に関するメッセージは、[ストリームのプロパティ] ダイアログ ボックスの [メッセージ] タブから簡単に参照することができます。このテーブルには、エラー メッセージも報告されます。

ストリーム メッセージを表示するには

- ▶ [ファイル] メニューから、[ストリームのプロパティ] を選択します (またはマネージャ パネルの [ストリーム] タブからストリームを選択し、ポップアップ メニューを右クリックして [ストリームのプロパティ] をクリックします)。
- ▶ [メッセージ] タブをクリックします。

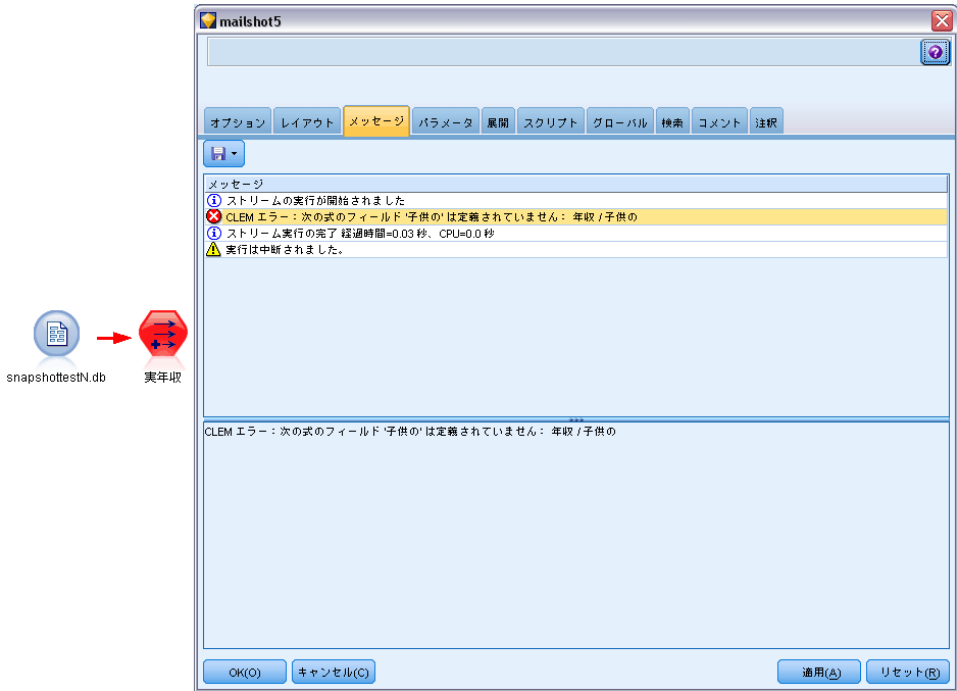
また、[ツール] メニューから次の各項目を選択することもできます。
ストリームのプロパティ > メッセージ

図 5-18
[ストリームのプロパティ] ダイアログ ボックスの [メッセージ] タブ



ここには、ストリーム操作に関するメッセージだけでなく、エラー メッセージも報告されます。ストリームの実行がエラーにより中断された場合、このダイアログ ボックスの [メッセージ] タブにそのエラー メッセージが表示されます。また、ストリーム領域上でエラーの発生したノードは赤で強調表示されます。

図 5-19
ストリームの実行でのエラーの報告



[ユーザー オプション] ダイアログ ボックスで、SQL 最適化とログ記録オプションが有効な場合、生成された SQL についての情報も表示されます。詳細は、12 章 p. 263 最適化オプションの設定を参照してください。

ここに表示されたメッセージを保存するには、ドロップダウン リスト ([メッセージ] タブのすぐ下、左側) から、[メッセージの保存] をクリックします。また、あるストリームのメッセージをすべて消去するには、リストから [すべてのメッセージを消去] を選択します。

ストリームとセッション パラメータの設定

パラメータは、CLEM 式とスクリプトで使用するために定義できます。実際のところ、パラメータはユーザー定義の変数であり、保存されて、現在のストリーム、セッション、またはスーパーノードで継続します。さらに、スクリプトを使用する場合と同様に、ユーザー インターフェイスからもアクセスできます。たとえば、ストリームを保存すると、そのストリームに設定されているパラメータも保存されます。(これは、ローカル スクリプト変数と異なる点です。ローカル スクリプト変数は、宣言されたスクリプト内でのみ使用できます。)通常パラメータは、スクリプト中でパラメータ値を指定する CLEM 式の一部として使用されます。

パラメータの有効範囲は、それがどこで設定されたかによって異なります。

- ストリーム パラメータは、ストリーム スクリプト内またはストリーム プロパティのダイアログ ボックス内で設定でき、ストリーム内のすべてのノードで使用できます。Clem 式ビルダーの [パラメータ] リストに表示されます。
- セッション パラメータは、スタンドアロン スクリプト内または [セッション パラメータ] ダイアログ ボックス内で設定できます。セッション パラメータは、現在のセッションのすべてのストリーム ([マネージャ] ウィンドウの [ストリーム] タブに表示されているすべてのストリーム) で利用できます。

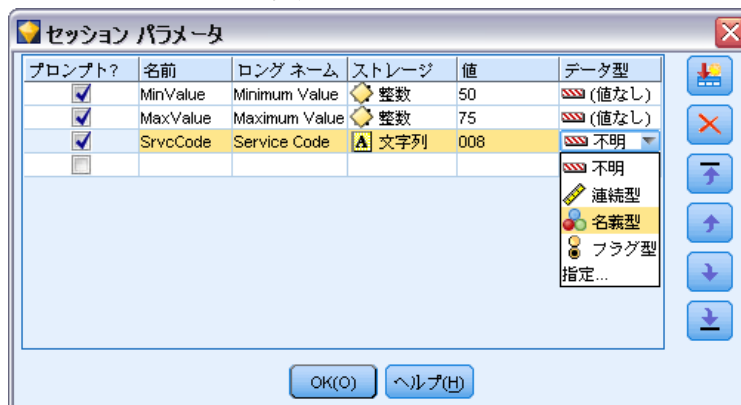
パラメータは、スーパーノード用にも設定できます。この場合、スーパーノード内にカプセル化されたノードでだけ表示できます。詳細は、9 章 [スーパーノードのパラメータの定義 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

スクリプトでのパラメータの設定と使用方法の詳細は、3 章の「[ストリーム、セッション、およびスーパーノード パラメータ](#)」を参照してください。

ユーザー インターフェイスからストリームとセッションのパラメータを設定するには

- ▶ ストリーム パラメータを設定するには、次のメニュー項目を選択してください。
ツール > ストリームのプロパティ > パラメータ
- ▶ セッション パラメータを設定するには、[ツール] メニューの [セッション パラメータの設定] を選択します。

図 5-20
セッションのパラメータの設定



プロンプト?: ランタイムにこのパラメータに値を入力するよう、プロンプトを表示したい場合は、このボックスをオンにします。

名前： ここにはパラメータ名が表示されます。新しくパラメータを作成するには、このフィールドに名前を入力します。たとえば、最低気温を表すパラメータを作成する場合に、`minvalue` と入力することができます。CLEM 式内でパラメータを示す接頭辞の `$P-` を付けないようにしてください。ここで指定した名前は、CLEM 式ビルダーにも表示されます。

ロング ネーム： 作成したパラメータを説明する名前が表示されます。

ストレージ： リストからストレージ タイプを選択します。ストレージで、データ値がパラメータ内にどのように格納されるかを示します。たとえば、「008」のように先頭に 0 がある値を扱う場合に、その 0 を保持する必要があるならば、ストレージ タイプとして **[文字列]** を選択する必要があります。選択しないと、値から 0 が削除されます。ストレージ タイプとしては、文字列、整数、実数、時間、日付、またはタイムスタンプを利用できます。日付のパラメータには、次の段落で示す ISO 規格の表記を使用して値を指定する必要があります。

値： 各パラメータの現在の値が表示されます。必要に応じてパラメータを調整してください。日付のパラメータには、ISO 規格の表記（つまり、YYYY-MM-DD）を使用して値を指定する必要があります。他のフォーマットで指定された日付は受け入れられません。

データ型 (オプション)： ストリームを外部アプリケーションに展開する場合は、使用する測定レベルをリストから選択します。それ以外の場合は、データ型の欄はそのままにしておくことをお勧めします。数値範囲の上限および下限など、パラメータに値の制約を指定したい場合、リストから **[指定]** を選択します。

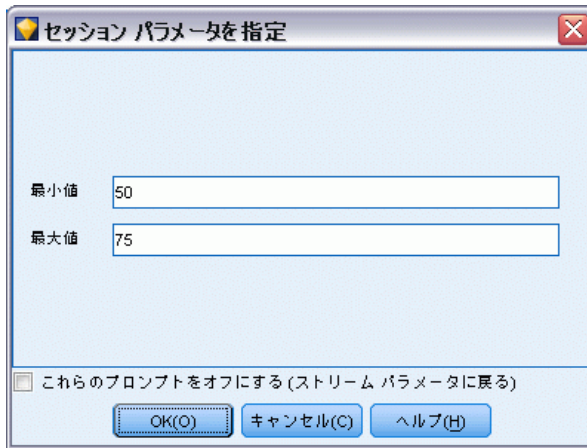
ロング ネーム、ストレージ、およびデータ型のオプションは、ユーザーインターフェイスを通じてだけ、パラメータに設定できます。これらのオプションは、スクリプトを使用して設定できません。

右にある矢印をクリックして、選択したパラメータを使用可能なパラメータのリストの上または下に移動することができます。選択したパラメータを削除するには、削除ボタン (X マーク) を使用します。

パラメータ値のライタイム プロンプトの指定

さまざまな機会、同じパラメータに異なる値を入力する必要があるストリームがある場合、1 つまたは複数のストリームまたはセッション パラメータ値にランタイム プロンプトを指定できます。

図 5-21
パラメータ値のライタイム プロンプト



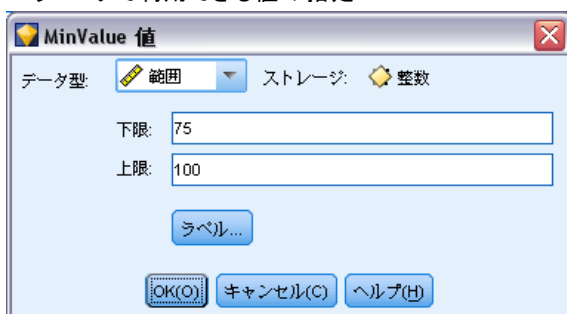
パラメータ: (オプション) パラメータの値を入力、またはデフォルト値がある場合はデフォルト値のままにします。

これらのプロンプトの無効化 ストリーム実行時にこれらのプロンプトを表示しない場合は、このボックスをオンにします。パラメーターが定義されているストリーム プロパティまたはセッション プロパティのダイアログ ボックスの [プロンプト?] チェック ボックスをオンにすると、プロンプトを再度表示することができます。 [詳細は、 p.66 ストリームとセッション パラメータの設定 を参照してください。](#)

パラメータ タイプの値の制約の指定

データ モデリング ストリームを読み込む外部アプリケーションにストリームを展開する際に、パラメータの値の制約をを行うことができます。このダイアログ ボックスでは、ストリームを実行する外部ユーザーが利用できる値を指定することができます。データ型によって、ダイアログ ボックスで指定できる値の範囲は動的に変わります。ここに表示されるオプションは、データ型ノードからの値に対して利用できるオプションと同じものです。

図 5-22
パラメータで利用できる値の指定



タイプ：現在選択されている測定レベルを表示します。IBM® SPSS® Modeler でのパラメータの利用目的に応じて、値を変更することができます。

ストレージ：わかっている場合に、ストレージ タイプを表示します。ストレージ タイプは、SPSS Modeler における作業で選択する測定レベル（連続型、名義型、フラグ型）には影響されません。ストレージ タイプは、[パラメータ] タブで変更することができます。

このダイアログ ボックスの下半分は、[データ型] で選択した測定レベルによって異なります。

連続型測定レベル

下限：パラメータ値の下限を指定します。

上限：パラメータ値の上限を指定します。

ラベル：集計範囲フィールドの任意値のラベルを指定できます。ラベル ボタンをクリックして、値ラベルを指定する個別のダイアログ ボックスを開きます。

名義型測定レベル

値：名義型フィールドとして使用されるパラメータの値を指定することができます。この値は、SPSS Modeler ストリーム中では強制されませんが、外部展開アプリケーションのドロップダウン リストで使用されます。矢印および削除ボタンを使って、既存の値を変更したり、値の並び替えや削除などの作業を行えます。

フラグ型測定レベル

真 (true)：条件を満たす場合のパラメータのフラグ値を指定します。

偽 (false)：条件を満たさない場合のパラメータのフラグ値を指定します。

ラベル: フラグ型フィールドの任意の値のラベルを指定できます。

ストリーム展開のオプション

[ストリームのプロパティ] ダイアログ ボックスの [展開] タブで、IBM® SPSS® Collaboration and Deployment Services 内のシナリオとして、モデル リフレッシュ、自動ジョブ スケジュール、あるいは IBM® SPSS® Decision Management または Predictive Applications 5.x による高度な使用のオプションを指定できます。すべてのストリームは、展開のために指定されたスコアリング枝が必要です。追加の要件およびオプションは展開の種類によって異なります。詳細は、9 章 p.177 IBM SPSS Collaboration and Deployment Services Repository オブジェクトの保存と展開 を参照してください。

ストリームのグローバル値の表示

[ストリームのプロパティ] ダイアログ ボックスの [グローバル] タブでは、現在のストリームに設定されているグローバル値を表示することができます。グローバル値は、グローバル ノードを使って作成され、選択したフィールドの平均、合計、または標準偏差などの統計量を判断するために用いられます。詳細は、6 章 グローバル ノード in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。

グローバル ノードが実行されると、ストリーム操作内でこれらの値をさまざまに利用することができます。詳細は、8 章 p.171 グローバル関数 を参照してください。

ストリームのグローバル値を表示するには

- ▶ [ファイル] メニューから、[ストリームのプロパティ] を選択します（またはマネージャ パネルの [ストリーム] タブからストリームを選択し、ポップアップ メニューを右クリックして [ストリームのプロパティ] をクリックします）。
- ▶ [グローバル] タブをクリックします。

また、[ツール] メニューから次の各項目を選択することもできます。

ストリームのプロパティ > グローバル

図 5-23
ストリームで利用できるグローバル値の表示



利用できるグローバル値: このテーブルには、利用できるグローバル値が表示されています。ここでグローバル値を編集することはできません。ただし、テーブルの右にある [すべての値の消去] ボタンを使って、ストリームのグローバル値をすべて消去することができます。

ストリーム内のノード検索

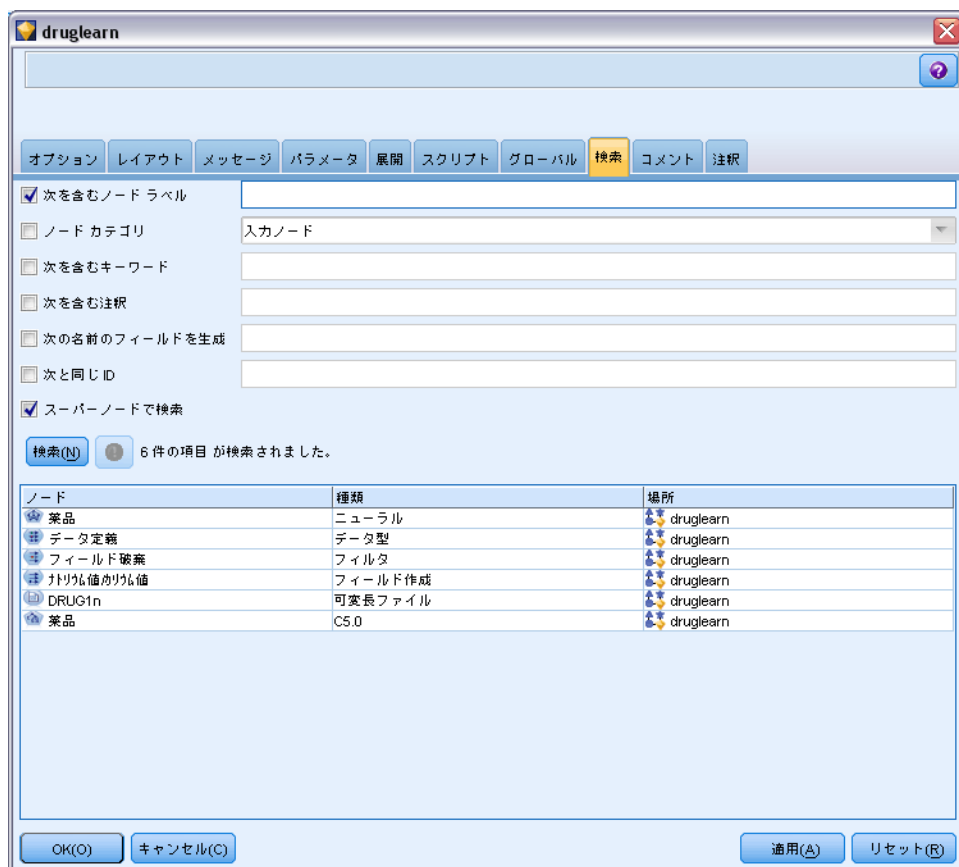
ノード名、カテゴリ、識別子などの検索基準を指定してストリーム内のノードを検索できます。この機能は、多くのノードを含む複雑なストリームに役立ちます。

ストリーム内のノードを検索するには

- ▶ [ファイル] メニューから、[ストリームのプロパティ] を選択します（またはマネージャ パネルの [ストリーム] タブからストリームを選択し、ポップアップメニューを右クリックして [ストリームのプロパティ] をクリックします）。
- ▶ [検索] タブをクリックします。

また、[ツール] メニューから次の各項目を選択することもできます。
ストリームのプロパティ > 検索

図 5-24
ストリーム内のノード検索



その他のオプションを除外してノード ID によって検索する（[次の ID に一致] フィールドを使用）という場合を除き、複数のオプションを指定して検索を絞り込みます。

以下のノードラベルを含む： このボックスをチェックして、ノードラベルのすべてまたは一部を入力して特定のノードを検索します。検索では大文字と小文字の区別を行わず、複数の単語をテキストの単一部分として扱います。

ノードカテゴリ： このボックスをチェックして、特定の種類のノードを検索するリストのカテゴリをクリックします。[プロセスノード] は、ノードパレットの [レコード設定] タブまたは [フィールド設定] タブのノードを指し、[モデルノードの適用] は、モデルナゲットを指します。

次のキーワードを含む： このボックスをチェックして、1 つまたは複数のキーワードを入力して、ノードダイアログボックスの [注釈] タブの [キーワード] フィールド入力されたテキストを含むノードを検索しま

す。入力したキーワードは完全に一致する必要があります。代替を検索するセミコロンで区切った複数のキーワードを使用すると（たとえば、`proton;neutron` と入力）、これらのキーワードをいずれかを持つすべてのノードを検索します。詳細は、[p. 88 \[注釈\]](#) を参照してください。

次の注釈を含む： このボックスをチェックして、1 つまたは複数のキーワードを入力して、ノード ダイアログ ボックスの [注釈] タブのメイン テキスト エリアにこのテキストを含むノードを検索します。検索では大文字と小文字の区別を行わず、複数の単語をテキストの単一部分として扱います。詳細は、[p. 88 \[注釈\]](#) を参照してください。

生成するフィールド名： このボックスをチェックして、生成するフィールドの名前を入力します（例：`$C-Drug`）。このオプションを使用して、特定のフィールド名を生成するモデル作成ノードを検索できます。フィールド名は 1 つだけ入力してください。完全に一致する必要があります。

次の ID と一致： このボックスをチェックしてノード ID を入力し、この ID を持つ特定のノードを検索します（このオプションを選択すると、前述のオプションはすべて無効になります。）。Node ID はノード作成時にシステムによって割り当てられ、ID を使用してスクリプトまたは自動化の目的でノードを参照できます。ノード ID は 1 つだけ入力してください。完全に一致する必要があります。詳細は、[p. 88 \[注釈\]](#) を参照してください。

スーパーノードで検索： このボックスはデフォルトで選択されています。そのため、スーパーノード内外のノードに検索が実行されます。ストリームの最上位で、スーパーノード外のノードのみを検索する場合は、このボックスの選択を解除します。

検索： 使用するすべてのオプションを指定し、このボタンをクリックすると検索が開始されます。

指定されたオプションに一致するノードは、ダイアログ ボックスの下部に表示されます。リスト内のノードを選択すると、ストリーム領域内で強調表示されます。

ストリーム名の変更

[ストリームのプロパティ] ダイアログ ボックスの [注釈] タブでは、ストリームを説明する注釈を追加したり、ストリームの名前を作成することができます。これらのオプションは、プロジェクト パネルに追加するストリームのレポートを生成する際に役に立ちます。詳細は、[p. 88 \[注釈\]](#) を参照してください。

ストリームの説明

作成する各ストリームについて、IBM® SPSS® Modeler は、ストリームの内容に関する情報を含むストリームの説明を作成します。ストリームが何を行っているかを確認したいが SPSS Modeler をインストールしていない場合、たとえば IBM® SPSS® Collaboration and Deployment Services を使用してストリームにアクセスする場合に役に立ちます。

図 5-25
ストリームの説明のセクションを開く

The screenshot shows the IBM SPSS Modeler interface with a table of stream metadata and a list of content sections.

IBM SPSS Modeler	
ストリームドキュメント	newschancart_commented
作成日:	2006/01/27 1:17
作成者:	llanders
最終保存日:	2010/05/28 20:51
最終保存者:	djones

内容

- [説明 & コメント](#)
- [スコアリング情報](#)
- [モデル作成情報](#)

説明 & コメント

コメント	
This stream predicts whether customers are likely to subscribe to our interactive news service, based on characteristics of customers who have already subscribed	
This is the source data about our existing customers	
This is where we tell the stream what we are trying to predict	

ストリームの説明が、さまざまなセクションで構成されている HTML 文書の形式で表示されます。

ストリームの一般情報

ここには、ストリーム名、いつストリームが作成されたか、最後に保存されたのはいつかについての詳細が表示されます。

説明とコメント

このセクションには次が表示されます。

- ストリームの注釈（「[\[注釈\]](#)」（p. 88）を参照）

- 特定のノードに接続していないコメント
- ストリームのモデル作成枝およびスコアリング枝のノードに接続するコメント

スコアリング情報

ここでは、ストリームのスコアリング枝に関連するさまざまな見出しの情報が表示されます。

- **コメント**：スコアリング枝のノードにのみリンクするコメントが表示されます。
- **入力**：ストレージタイプ (string、integer、real など) とともに、入力フィールドが表示されます。
- **出力**：ストレージタイプとともに、モデル作成ノードで生成された追加フィールドなど、出力フィールドが表示されます。
- **パラメータ**：ストリームのスコアリング枝に関連するパラメータを表示し、モデルをスコアリングするごとに表示または編集できます。[ストリームのプロパティ] ダイアログで [展開] タブ の [スコアリングパラメータ] ボタンをクリックすると、これらのパラメータが識別されます。
- **モデルノード**：モデル名と種類を表示します (ニューラル ネットワーク、C&R Tree など)。[ストリームのプロパティ] ダイアログ ボックスで [展開] タブ の [モデルノード] フィールドに選択されたモデルナゲットです。
- **モデルの詳細**：前の見出しで識別されたモデル ナゲットの詳細を表示します。可能な場合、モデルの予測値の重要度および評価グラフが表示されます。

モデル作成情報

ストリームのモデル作成枝に関連する情報が表示されます。

- **コメント**：モデル作成枝のノードに接続するコメントまたは注釈が表示されます。
- **入力**：モデル作成枝の役割とともに入力フィールドが表示されます (たとえば入力、出力、分割など、フィールド方向の値の形式)。
- **パラメータ**：ストリームのモデル作成枝に関連するパラメータを表示し、モデルを更新するごとに表示または編集できます。[ストリームのプロパティ] ダイアログ ボックスで [展開] タブ の [モデル構築パラメータ] ボタンをクリックすると、これらのパラメータが識別されます。
- **モデル作成ノード**：モデルを生成または更新するために使用されるモデル作成ノードの名前と種類を表示します。

ストリームの説明のプレビュー

[ストリームのプロパティ] ダイアログ ボックスでオプションをクリックして、Web ブラウザでストリームの説明の内容を表示することができます。説明の内容は、ダイアログ ボックスの [展開] タブで指定するオプションによって異なります。詳細は、[9 章 p.208 ストリーム展開のオプション](#) を参照してください。

ストリームの説明を表示するには

- ▶ IBM® SPSS® Modeler のメイン メニューで次の各項目をクリックします。
ツール > ストリームのプロパティ > Deployment
- ▶ 展開タイプ、指定のスコアリング ノードおよびスコアリング パラメータを設定します。
- ▶ 展開タイプが [モデルのリフレッシュ] の場合、オプションで次を選択できます。
 - モデル作成ノードとモデル構築パラメータ
 - ストリームのスコアリング枝のモデル ナゲット
- ▶ [ストリームの説明をプレビュー] ボタンをクリックします。

ストリームの説明のエクスポート

ストリームの説明の内容を HTML ファイルにエクスポートできます。

ストリームの説明をエクスポートするには

- ▶ メイン メニューで次の各項目をクリックします。
File > ストリームの説明のエクスポート
- ▶ HTML ファイルの名前を入力して、[保存] をクリックします。

ストリームの実行

ストリームに必要なオプションを設定して、ノード間を接続したら、ストリーム内のノードにデータを流すことによって、ストリームを実行できます。IBM® SPSS® Modeler でストリームを実行するには、さまざまな方法があります。以下を行うことができます。

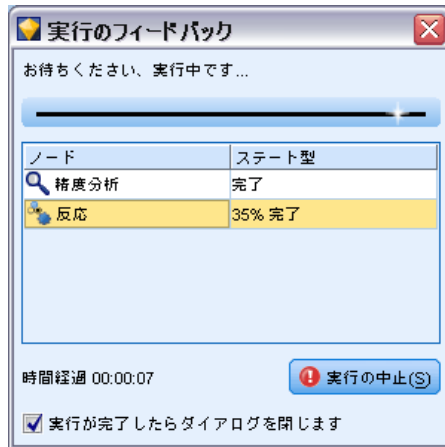
- [ツール] メニューから [実行] を選択します。
- ツールバーの [実行...] ボタンをクリックします。これらのボタンを利用すれば、ストリーム全体を実行することも、単に選択したターミナルノードだけを実行することもできます。詳細は、[3 章 p.21 IBM SPSS Modeler ツールバー](#) を参照してください。

- ターミナル ノードを右クリックし、ポップアップ メニューから [実行] をクリックして、1 つのデータ ストリームを実行します。
- 非ターミナル ノードを右クリックし、ポップアップ メニューから [ここから実行] をクリックして、1 つのデータ ストリームの一部を実行します。これにより、選択したノードの後のこれらの操作のみを実行できます。

実行中のストリームを中止するには、ツールバーにある赤い中止ボタンをクリックするか、または[ツール] メニューの [実行の中止] をクリックします。

何らかのストリームの実行が 3 秒以上かかる場合は、[実行のフィードバック] ダイアログ ボックスが表示され、進行状況を示します。

図 5-26
[実行のフィードバック] ダイアログ ボックス



一部のノードでは、ストリーム実行に関して追加情報を提供するためにさらなる表示を行うことがあります。これらの表示は、ダイアログ ボックスの対応する行を選択することで表示されます。最初の行は自動的に選択されます。

モデルの処理

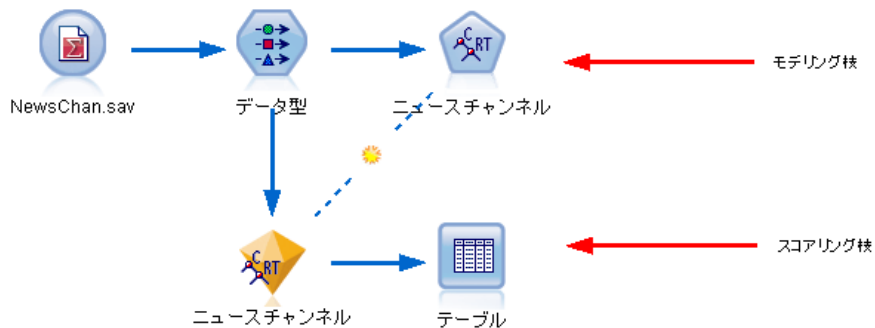
ストリームにモデル作成ノード (ノード パレットの [モデル作成] タブまたは [DB モデリング] タブのモデル作成ノード) が含まれている場合、ストリーム実行時に**モデル ナゲット**が作成されます。モデル ナゲットは、**モデル**のコンテナです。つまり、入力データに対して予測を生成できる一連のルール、式または方程式で、予測分析の中心となるものです。

図 5-27
モデル ナゲット



モデル作成ノードを正常に実行すると、対応するモデル ナゲットがストリーム領域に置かれ、金色のダイヤモンドの形のアイコンで表示されます（名前は「ナゲット」）。ナゲットを開いて内容を参照し、モデルに関する詳細を表示できます。予測を表示するには、1 つまたは複数のターミナル ノードを接続および実行、出力は読み取り可能な形式で予測を表示します。

図 5-28
ストリームのモデル作成枝およびスコアリング枝



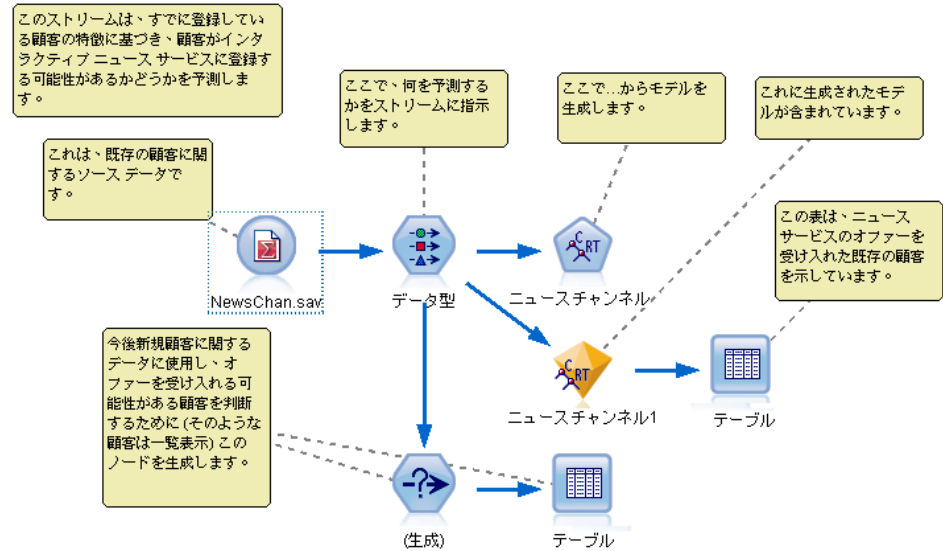
一般的なモデル作成ストリームは、2 つの枝で構成されています。**モデル作成枝**には、モデル作成ノード、そしてそれに先行する入力ノードおよび処理ノードが含まれています。**スコアリング枝**は、モデル作成ノードを実行すると作成され、予測の表示に使用するモデル ナゲットおよびターミナル ノードが含まれています。

詳細については、『IBM® SPSS® Modeler モデル作成ノード ガイド』を参照してください。

ノードおよびストリームへのコメントおよび注釈の追加

ストリームについて、組織内の他の部門に説明が必要な場合があります。説明を行うために、ストリーム、ノード、モデル ナゲットに説明のコメントを追加することができます。

図 5-29
コメントを追加したストリーム



他のユーザーがコメントを画面上で表示したり、コメントを含むストリームの画像を印刷することができます。

ストリームまたはスーパーノードのすべてのコメントを一覧表示し、リスト内のコメントの順序を変更、コメントテキストを編集、コメントの前景または背景の色を変更できます。詳細は、[p. 86 ストリームのコメントの一覧表示](#)を参照してください。

[ストリームのプロパティ] ダイアログ ボックス、ノードのダイアログ ボックス、またはモデル ナゲットのウィンドウの [注釈] タブを使用して、説明をテキスト注釈の形式でストリーム、ノード、ナゲットに追加することもできます。これらの説明は、[注釈] タブが開いているときのみ表示できますが、ストリームの注釈を画面上のコメントとして表示することもできます。詳細は、[p. 88 \[注釈\]](#)を参照してください。


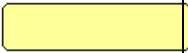

コメント

コメントはテキスト ボックスの形式で、任意の文字数のテキストを入力でき、必要に応じてコメントを追加できます。コメントは独立（ストリームオブジェクトに接続していない）していたり、ストリーム内のノードまたはモデル ナゲットに接続している場合があります。独立しているコメントは、通常ストリーム全体の目的を説明するために使用され、接続しているコメントは、接続しているノードまたはナゲットを説明します。ノードおよびナゲットは複数のコメントを接続し、ストリームは独立したコメントをいくつも持つことができます。

注：ストリームの注釈を画面上のコメントとして表示することもできますが、これらをノードまたはナゲットに接続することはできません。詳細は、p. 88 注釈のコメントへの変換を参照してください。

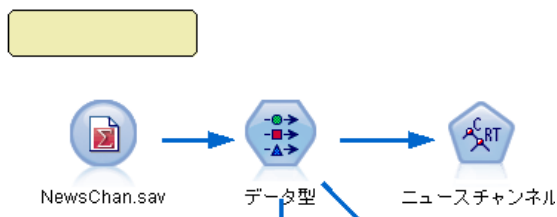
テキスト ボックスの外観は、次の表が示すように、コメント（または画面上のコメントとして表示される注釈）の現在のモードを示すよう変更されます。

テーブル 5-1
コメントおよび注釈のテキスト ボックスのモード

[コメント] テキスト ボックス	[注釈] テキスト ボックス	モード	指示内容	取得方法
		編集	コメントが開き、編集できます。	新しいコメントまたは注釈を作成するか、既存のコメントをダブルクリックします。
		最後の選択	コメントを移動、サイズ変更または削除できます。	編集後、ストリームの背景をクリックするか、既存のコメントまたは注釈をシングルクリックします。
		表示	編集が完了しています。	編集後、別のノード、コメントまたは注釈をクリックします。

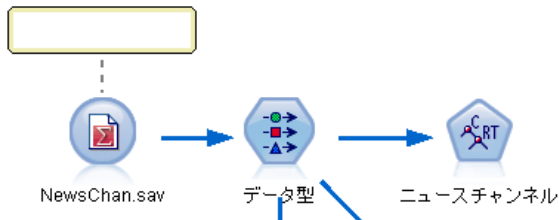
新しい独立したコメントを作成すると、コメントは最初ストリーム領域の左上の隅に表示されます。

図 5-30
新しい独立したコメント



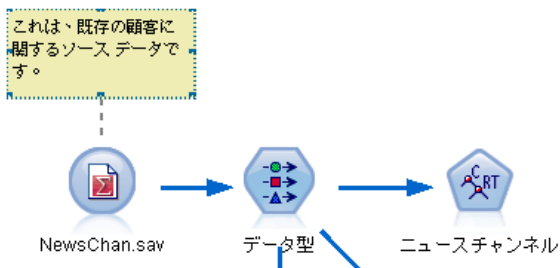
コメントをノードまたはナゲットに接続している場合、コメントは最初接続されているストリーム オブジェクトの上に表示されます。

図 5-31
ノードに接続された新しいコメント



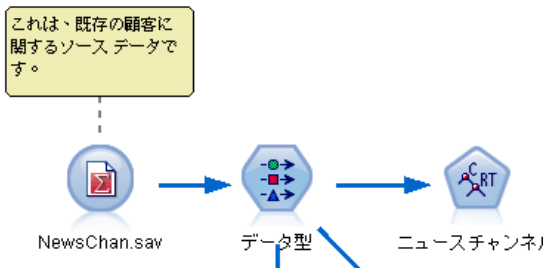
テキスト ボックスは白色の場合が、テキストが入力できることを示します。テキストを入力した場合、テキスト ボックスの外側をクリックします。コメントの背景が黄色に変化し、テキスト入力が完了したことを示します。コメントが選択されたままになっており、移動、サイズ変更、または削除できます。

図 5-32
編集モードのコメント



もう一度クリックすると、境界線が実線に変わり、編集が完了していることを示します。

図 5-33
完了したコメント



コメントをダブルクリックすると、テキスト ボックスの背景色が白に変わって編集モードとなり、コメント テキストを編集できます。

スーパーノードにコメントを追加することもできます。詳細は、9章 [スーパーノードの注釈付けと名前の変更 in IBM SPSS Modeler 14.2 入力ノード、プロセスノード、出力ノード](#) を参照してください。

コメントに関する操作

コメントに多くの操作を行うことができます。以下を行うことができます。

- 独立したコメントの追加
- コメントのノードまたはナゲットへの接続
- コメントの編集
- コメントのサイズ変更
- コメントの移動
- コメントの接続解除
- コメントの削除
- ストリームのすべてのコメントの表示または非表示

独立したコメントを追加するには

- ▶ ストリームに何も選択されていないことを確認します。
- ▶ 次のいずれかを実行します。
 - メインメニューで次の各項目をクリックします。
[挿入(O)] > 新規コメント
 - ストリームの背景を右クリックし、ポップアップメニューで [新規コメント] をクリックします。
 - ツールバーにある [新規コメント] ボタンをクリックします。
- ▶ コメントテキストを入力（またはクリップボードからテキストを貼り付け）ます。
- ▶ ストリームのノードをクリックして、コメントを保存します。

コメントをノードまたはナゲットに接続するには

- ▶ ストリーム領域で 1 つまたは複数のノードまたはナゲットを選択します。
- ▶ 次のいずれかを実行します。
 - メインメニューで次の各項目をクリックします。
[挿入(O)] > 新規コメント
 - ストリームの背景を右クリックし、ポップアップメニューで [新規コメント] をクリックします。
 - ツールバーにある [新規コメント] ボタンをクリックします。

- ▶ コメント テキストを入力します。
- ▶ ストリームの別のノードをクリックして、コメントを保存します。
または、次を行うことができます。
- ▶ 独立したコメントの挿入（前項を参照）
- ▶ 次のいずれかを実行します。
 - コメントを選択して F2 キーを押し、ノードまたはナゲットを選択します。
 - ノードまたはナゲットを選択して F2 キーを押し、コメントを選択します。
 - (3 ボタンのマウスのみ) マウス ポインタをコメントに移動し、中央のボタンを押したままマウス ポインタをノードまたはナゲットに移動して、マウス ボタンを離します。

コメントを追加のノードまたはナゲットに接続するには

コメントがすでにノードまたはナゲットに接続されている、または現在ストリーム レベルにあり、追加のノードまたはナゲットにコメントを接続する場合は、次のいずれかを実行します。

- コメントを選択して F2 キーを押し、ノードまたはナゲットを選択します。
- ノードまたはナゲットを選択して F2 キーを押し、コメントを選択します。
- (3 ボタンのマウスのみ) マウス ポインタをコメントに移動し、中央のボタンを押したままマウス ポインタをノードまたはナゲットに移動して、マウス ボタンを離します。

既存のコメントを編集するには

- ▶ 次のいずれかを実行します。
 - コメント テキスト ボックスをダブルクリックします。
 - テキスト ボックスを選択して Enter キーを押します。
 - テキスト ボックスを右クリックして、メニューを表示し、[編集] をクリックします。
- ▶ コメント テキストを編集します。編集時、Ctrl + C キーを使用してテキストをコピーするなど、標準的な Windows ショートカット キーを使用できます。編集中にそのほかのオプションが、コメントのポップアップメニューに表示されます。
- ▶ テキスト ボックスの外部を 1 回クリックするとサイズ変更のコントロールが表示され、もう 1 回クリックするとコメントが完了します。

コメント テキスト ボックスをサイズ変更するには

- ▶ コメントを選択してサイズ変更のコントロールを表示します。
- ▶ コントロールをクリック アンド ドラッグしてボックスのサイズを変更します。
- ▶ テキスト ボックスの外部をクリックして、変更を保存します。

既存のコメントを移動するには

接続したオブジェクト（あれば）を移動せずにコメントを移動する場合、次のいずれかを実行します。

- コメントにマウス ポインタを移動し、左のマウス ボタンを押したままコメントを新しい位置にドラッグします。
- コメントを選択し、Alt キーを押したまま矢印キーを使用してコメントを移動します。

コメントを接続したノードまたはナゲットとともに移動する場合は、次のようにします。

- ▶ 移動するすべてのオブジェクトを選択します。
- ▶ 次のいずれかを実行します。
 - オブジェクトのいずれかにマウス ポインタを移動し、左のマウス ボタンを押したままオブジェクトを新しい位置にドラッグします。
 - オブジェクトのいずれかを選択し、Alt キーを押したまま矢印キーを使用してオブジェクトを移動します。

コメントとノードまたはナゲットの接続を解除するには

- ▶ 接続を解除するコメントを選択します。
- ▶ 次のいずれかを実行します。
 - F3 キーを押します。
 - 選択したコメントを右クリックして、メニューの [接続の解除] をクリックします。

コメントを削除するには

- ▶ 削除するコメントを選択します。
- ▶ 次のいずれかを実行します。
 - Del キーを押します。
 - 選択したコメントを右クリックして、メニューの [削除] をクリックします。

コメントがノードまたはナゲットに接続している場合、接続も同時に削除されます。

コメントが本来、独立したコメントに変換されたストリームまたはスーパーノードの注釈である場合、コメントは領域から削除されますが、テキストはストリームまたはスーパーノードの [注釈] タブに保持されます。

ストリームのコメントの表示または非表示

- ▶ 次のいずれかを実行します。
 - メイン メニューで次の各項目をクリックします。
表示 > コメント
 - ツールバーにある [コメントの表示/非表示] ボタンをクリックします。

ストリームのコメントの一覧表示

特定のストリームまたはスーパーノードに作成されたすべてのコメントのリストを表示できます。

このリスト上で、次のことができます。

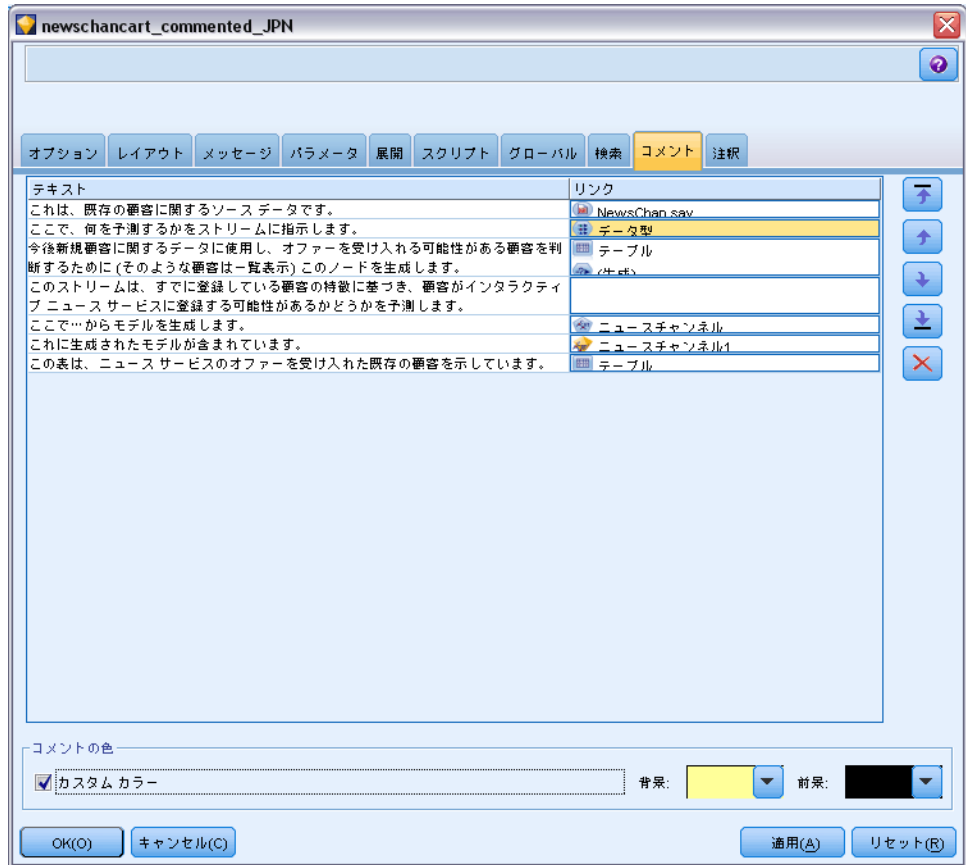
- コメントの順序の変更
- コメント テキストを編集
- コメントの前景色または背景色の変更

コメントの一覧表示

ストリームに作成されたコメントを一覧表示するには、次のいずれかを実行します。

- メイン メニューで次の各項目をクリックします。
ツール > ストリームのプロパティ > コメント
- マネージャ パネルでストリームを右クリックし、[ストリームのプロパティ] を選択して [コメント] をクリックします。
- ストリーム領域上でストリームの背景を右クリックし、[ストリームのプロパティ] を選択して [コメント] をクリックします。

図 5-34
ストリームのコメントを一覧表示



テキスト： コメントのテキスト。テキストをダブルクリックして、フィールドを編集可能なテキスト ボックスに変更します。

リンク。 コメントが接続されるノードの名前。このフィールドが空の場合、コメントがストリームに適用されます。

配置ボタン： リスト内の選択したコメントを上下に移動させます。

コメントの色： コメントの前景色または背景色を変更するには、コメントを選択し、[カスタム カラー] チェック ボックスを選択し、[背景] または [前景] リスト (または両方) から色を選択します。[適用] をクリックし、ストリームの背景をクリックして、変更の内容を表示します。[OK] をクリックして、変更を保存します。

注釈のコメントへの変換

ストリームまたはスーパーノードに作成された注釈は、コメントに変換できます。

ストリームの場合、注釈はストリーム領域の独立したコメントに変換されます（つまり、どのノードにも接続されません）。

スーパーノードの注釈をコメントに変換する場合、コメントはストリーム領域のスーパーノードに接続されませんが、スーパーノードにズームインすると確認することができます。

ストリームの注釈をコメントに変換するには

- ▶ [ツール] メニューから [ストリームのプロパティ] を選択します。（代わりに、マネージャ パネル内でストリームを右クリックして、[ストリームのプロパティ] を選択することもできます。）
- ▶ [注釈] タブをクリックします。
- ▶ [注釈をコメントとして表示] チェック ボックスを選択します。
- ▶ [OK] をクリックします。

スーパーノードの注釈をコメントに変換するには

- ▶ 領域の [スーパーノード] アイコンをダブルクリックします。
- ▶ [注釈] タブをクリックします。
- ▶ [注釈をコメントとして表示] チェック ボックスを選択します。
- ▶ [OK] をクリックします。

[注釈]

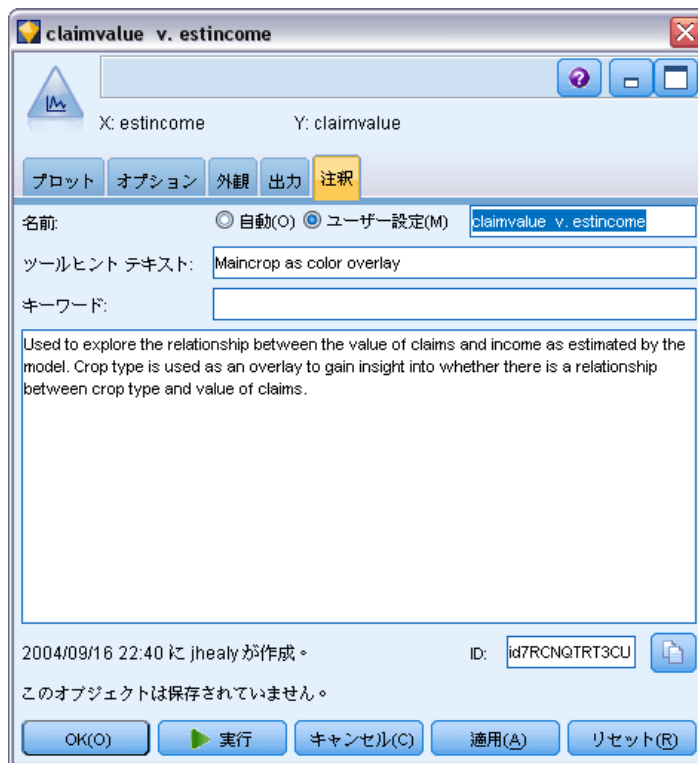
ノード、ストリーム、およびモデルには、さまざまな方法で注釈を付けることができます。詳細な注釈を追加したり、新しい名前を指定することができます。これらのオプションは、プロジェクト パネルに追加するストリームのレポートを生成する際に役に立ちます。ノードおよびモデル ナゲットには、ストリーム領域にある、よく似たノードと区別するのに役立つように、ツールヒント テキストを追加することもできます。

注釈の追加

ノードまたはモデル ナゲットを編集すると、ダイアログ ボックスが表示されます。このダイアログ ボックスの [注釈] タブでは、注釈に関するさまざまなオプションを設定できます。このタブは、次の方法で表示することもできます。

- ▶ ノードまたはナゲットに注釈を付けるには、ストリーム領域にあるノードまたはナゲットを右クリックし、[名前の変更と注釈] をクリックします。[編集] ダイアログ ボックスの [注釈] タブが表示されます。
- ▶ ストリームに注釈を付けるには、[ツール] メニューから [ストリームのプロパティ] をクリックします。(代わりに、マネージャ パネル内でストリームを右クリックして、[ストリームのプロパティ] を選択することもできます。)[注釈] タブをクリックします。

図 5-35
[注釈] タブの設定



名前: ストリーム領域に表示される、ノードの自動的に生成された名前を変更する、または固有の名前を指定する場合は、[ユーザー設定] を選択します。

ツールヒント テキスト: (ノードおよびモデル ナゲットに対してのみ) ストリーム領域のツールヒントとして使用されるテキストを入力します。特に、似たようなノードが多数あるような作業に役立ちます。

キーワード: プロジェクト レポート内で使用したりストリーム内のノードを検索する場合に、リポジトリに保存されているオブジェクトの検索または追跡時に使用したりするキーワードを指定します (「IBM SPSS Collaboration and Deployment Services Repository について」 (p. 175) を参照)。複数のキーワードをセミコロンで区切って指定できます income;

crop type; claim value 各キーワードの前後にある空白文字は削除されます。たとえば、income ; crop type は、income;crop type と同じ結果になります(ただし、キーワード内にある空白文字は削除されません。たとえば、スペースが 1 個の crop type とスペースが 2 個の crop type は同一ではありません。)

メイン テキスト エリアには、ノードの操作やノード作成時の決定事項などの操作に関する注釈を記入します。たとえば、ストリームを共有、再利用する場合、「空白が多いフィールドをフィルタ ノードを使って破棄する。」のような情報を記載しておけば、大変役に立ちます。ノードに注釈を付けると、ノードに情報が保存されます。これらの注釈を、プロジェクト パネルを使って作成したプロジェクト レポートに表示するように設定することもできます。 [詳細は、11 章 p.236 プロジェクトの概要 を参照してください。](#)

注釈をコメントとして表示： (ストリームおよびスーパーノードの注釈についてのみ) このボックスをチェックすると、注釈をストリーム領域上に表示される独立したコメントに変換します。 [詳細は、p.79 ノードおよびストリームへのコメントおよび注釈の追加 を参照してください。](#)

ID： スクリプトまたはオートメーションの目的でノードを参照するために使用できる一意の ID を表示します。この値はノード作成時に自動的に生成され、変更することはありません。「0」の文字との混乱を避けるため、ノード ID にはゼロを使用しません。必要に応じて、右側にあるコピー ボタンを使用して ID をスクリプトなどの場所にコピーして貼り付けます。 [詳細は、3 章 ノードの参照 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。](#)

データ ストリームの保存

ストリームを作成したら、今後再利用するためにストリームを保存できます。

ストリームを保存するには

- ▶ [ファイル] メニューの [ストリームを上書き保存] または [名前を付けて保存 (ストリーム)] をクリックします。
- ▶ [保存] ダイアログ ボックスで、ストリーム ファイルを保存するフォルダを指定します。
- ▶ [ファイル名] テキスト ボックスにストリームの名前を入力します。
- ▶ 保存したストリームを現在のプロジェクトに追加する場合は、[プロジェクトに追加] を選択します。

[保存] をクリックすると、指定したディレクトリにストリームが、ファイル拡張子 *.str で保存されます。

自動バックアップファイル： 毎回、ストリームが保存されると、そのファイルの以前に保存されたバージョンは、自動的にバックアップとして保存されます。バックアップ ファイルはのファイル名には、ハイフンが追加されず（例：mystream.str-）。バックアップ バージョンを復元するには、ハイフンを削除してそのファイルをオープンするだけです。

ステートの保存

ストリームのほかに、**ステート**を保存することもできます。ステートには、現在表示されているストリーム ダイアグラムとモデル ナゲットが含まれています（マネージャ パネルの [モデル] タブに記載）。

ステートを保存するには

- ▶ [ファイル] メニューで次の各項目をクリックします。
State > [ステートの保存] または [名前を付けてステートを保存]
- ▶ [保存] ダイアログ ボックスで、ステートを保存するフォルダを指定します。

[保存] をクリックすると、指定したディレクトリにステートが、ファイル拡張子 *.cst で保存されます。

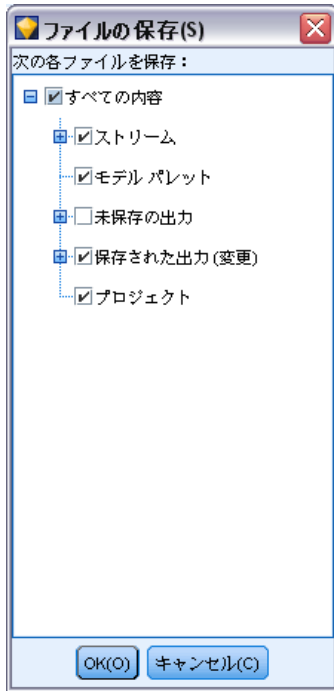
ノードの保存

ストリーム領域でノードを右クリックし、ポップアップ メニューから [ノードの保存] をクリックして、個々のノードを保存することもできます。ファイルの拡張子は、*.nod を使用します。

複数のストリーム オブジェクトの保存

ストリーム、プロジェクト、またはモデル ナゲットなどの複数のオブジェクトが保存されていない状態で IBM® SPSS® Modeler を終了する場合、ソフトウェアを終了する前にそれらのオブジェクトを保存するよう指示するメッセージが表示されます。これらのオブジェクトを保存することを選択した場合、各オブジェクトを保存するためのオプションがダイアログ ボックスに表示されます。

図 5-36
複数のオブジェクトの保存



- ▶ 保存するオブジェクトに対応するチェックボックスを選択してください。
- ▶ 各オブジェクトを指定した場所に保存するには、[OK] をクリックします。
各オブジェクトに対応する標準の[保存] ダイアログボックスが表示されます。保存が完了すると、アプリケーションが終了します。

出力の保存

IBM® SPSS® Modeler 出力ノードから生成されたテーブル、グラフ、およびレポートは、出力オブジェクト (*.cou) フォーマットで保存できます。

- ▶ 保存したい出力を表示するときは、出力ウィンドウのメニューで、次をクリックします。
File > [保存(S)]
- ▶ 出力ファイルの名前と場所を指定します。
- ▶ 必要に応じて、現在のプロジェクトにファイルを追加するには、[保存] ダイアログボックスにある [ファイルをプロジェクトに追加] を選択します。 [詳細は、11 章 p.236 プロジェクトの概要 を参照してください。](#)

また、[マネージャー] パネルにリストされている任意の出力オブジェクトを右クリックして、ポップアップメニューから [保存] を選択することもできます。

暗号化および復号化の情報

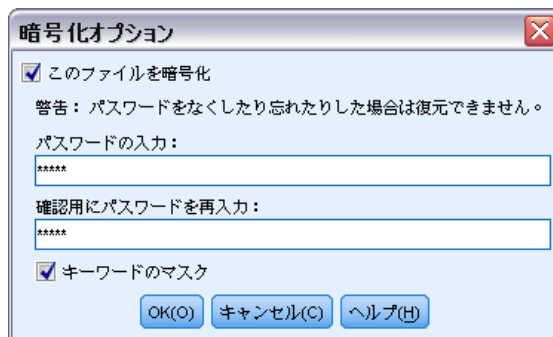
ストリーム、ノード、プロジェクト、出力ファイル、またはモデル ナゲットを保存するときは、無許可の使用を防止するために暗号化することができます。こうするには、保存するときに追加のオプションを選択して、保存するアイテムにパスワードを追加します。この暗号化は、保存するあらゆるアイテムに設定して、セキュリティを追加できます。これは、IBM® SPSS® Modeler と IBM® SPSS® Modeler Server の間でファイルを受け渡す場合に使用する SSL 暗号とは異なります。

暗号化されたアイテムを開くときは、パスワードを入力するように促すプロンプトが表示されます。正しいパスワードを入力すると、アイテムが自動的に復号化され、通常通りに開きます。

項目を暗号化するには

- ▶ [保存] ダイアログ ボックスで、暗号化するアイテムについて、[オプション] をクリックします。[暗号化オプション] ダイアログ ボックスが開きます。

図 5-37
ファイルを保存するときの暗号化オプション



- ▶ [このファイルを暗号化] を選択します。
- ▶ オプションで、さらにセキュリティを高めるために、[キーワードのマスク] を選択します。こうすることで、入力したものがすべてドットの列として表示されます。
- ▶ パスワードを入力します。警告 : パスワードを忘れた場合は、ファイルまたはモデルは開くことができなくなります。
- ▶ [キーワードのマスク] を選択した場合は、パスワードを再入力して、正しくパスワードを入力したことを確認します。

- ▶ [OK] をクリックして、[保存] ダイアログ ボックスに戻ります。

注：暗号化保護されたアイテムのコピーを保存する場合は、[暗号化オプション] ダイアログ ボックスで設定を変更しない限り、新しいアイテムは元のパスワードを使用して自動的に暗号化フォーマットで保存されます。

ファイルのロード

IBM® SPSS® Modeler で保存された、さまざまなオブジェクトを再ロードすることができます。

- ストリーム (.str)
- ステート (.cst)
- モデル (.gm)
- モデル パレット (.gen)
- ノード (.nod)
- 出力 (.cou)
- プロジェクト (.cpj)

新しくファイルを開く

ストリームは、[ファイル] メニューから直接ロードすることができます。

- ▶ [ファイル] メニューの [ストリームを開く] をクリックします。

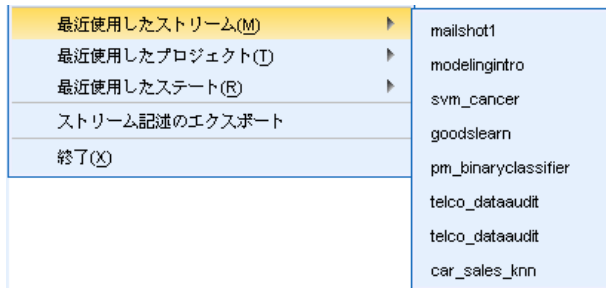
その他の種類のファイルは、[ファイル] メニューのサブメニューを使って開くことができます。たとえば、モデルをロードするには、[ファイル] メニューから次の各項目を選択します。

Models > モデルを開く、またはモデル パレットのロード

最近使用したファイルを開く

最近使用したファイルを素早くロードするには、[ファイル] メニューの下の部分にあるオプションを使用します。

図 5-38
[ファイル] メニューから最近使用したオプションを開く



[最近使用したストリーム]、[最近使用したプロジェクト]、または [最近使用したステート] を選択して、最近使用したファイルのリストを開きます。

データ ストリームのマッピング

マッピング ツールを使って、新しいデータ ソースを既存のストリームに接続できます。マッピング ツールは接続を設定するだけでなく、既存のストリームのフィールドを新しいソースのフィールドで置換する方法を指定する場合にも役立ちます。新しいデータ ソース用にデータ ストリーム全体を作成しななくても、単に既存のストリームに接続するだけで済みます。

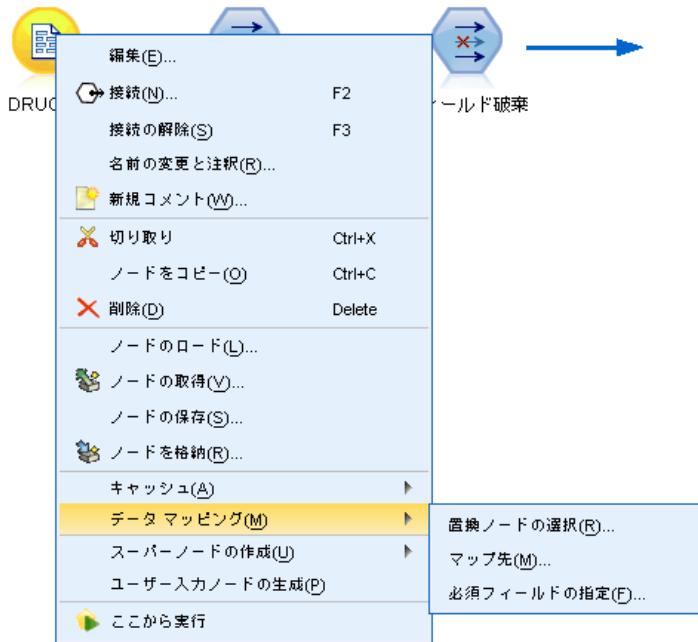
データ マッピング ツールを利用すれば、2 つのストリーム フラグメントを結合し、すべての (必須) フィールド名を正しく一致させることができます。基本的に、データのマッピングを行うと、名前を変更して適切なフィールドを一致させる新しいフィルタ ノードが単純に作成されます。

データのマッピングを行うには、次の 2 種類の方法があります。

置換ノードの選択： この方法は、置換されるノードを選択することから始まります。まず置換するノードを右クリックします。ポップアップ メニューの [データ マッピング] > [置換ノードの選択] オプションを使用して、置換するノードを選択します。

マップ先： この方法は、ストリームに導入するノードから始まります。まず導入するノードを右クリックします。ポップアップ メニューの [データ マッピング] > [マップ] オプションを使用して、結合するノードを選択します。この方法は、ターミナル ノードにマッピングする場合などに適しています。
注：レコード結合ノードまたはレコード追加ノードにマップすることはできません。この場合は、普段と同じように、ストリームをレコード結合ノードに接続してください。

図 5-39
データ マッピング オプションの選択



データ マッピングは、ストリーム構築に密に統合されています。既に接続があるノードに接続する場合、接続を置き換える、またはそのノードにマッピングするオプションが提供されます。

データのテンプレートへのマッピング

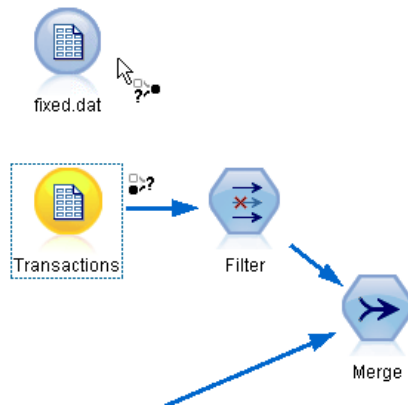
テンプレート ストリームのデータ ソースを、IBM® SPSS® Modeler に独自のデータを取り込む新しい入力ノードで置換するには、ポップアップメニューの [データ マッピング] の [置換ノードの選択] オプションを使用する必要があります。このオプションは、レコード結合ノード、レコード集計ノード、およびターミナル ノードを除く、他のすべてのノードで利用することができます。データ マッピング ツールを使ってこの操作を行うことにより、既存のストリーム操作と新規データ ソースのフィールドが正しく一致することが保証されます。データ マッピング作業の概要を次に示します。

ステップ 1: 元の入力ノードの必須フィールドを指定します。 ストリーム操作が適切に行われるためには、必須フィールドを指定する必要があります。 [詳細は、 p. 99 必須フィールドの指定 を参照してください。](#)

ステップ 2: 新しいデータ ソースをストリーム領域に追加します。 入力ノードを使って、新しい置換データを取り込みます。

ステップ 3: テンプレート入力ノードを置換します。 テンプレート入力ノードのポップアップメニューの [データ マッピング] オプションを使用して、[置換ノードの選択] をクリックし、置換データの入力ノードを選択します。

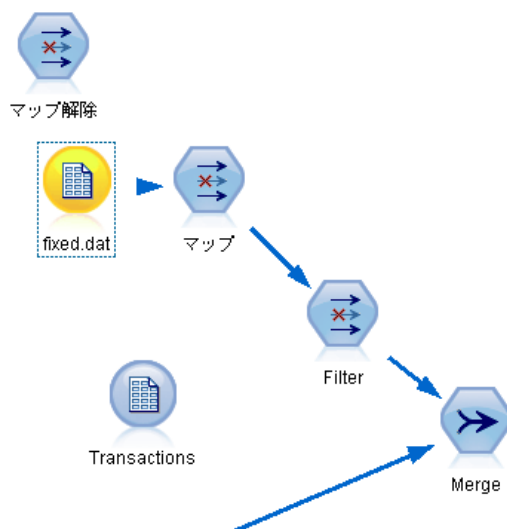
図 5-40
置換入力ノードの選択



ステップ 4: マップされたフィールドを確認します。 表示されているダイアログ ボックスで、置換データ ソースからストリームへ、フィールドが正しくマッピングされているかどうかを確認します。マップされていない必須フィールドは赤色で表示されています。これらのフィールドはストリーム操作で使用されるので、下流の操作が正しく行われるようにするためには、新しいデータ ソース中の適切なフィールドと置き換える必要があります。 [詳細は、p. 100 マップされたフィールドの調査](#) を参照してください。

ダイアログ ボックスですべての必須フィールドが正しくマップされていることを確認したら、マップと呼ばれるフィルタ ノードを使い、古いデータ ソースの接続を解除して、新しいデータ ソースをストリームに接続します。このフィルタ ノードは、ストリーム中のフィールドの実際のマッピングを指示します。ストリーム領域には、マップ解除フィルタ ノードも含まれています。マップ解除フィルタ ノードをストリームに追加して、フィールド名のマッピングを元に戻すことができます。これにより、フィールドのマッピングが解除されます。ただし、フィールドを再選択してオーバーレイするには、下流のターミナル ノードを編集する必要があります。このことに注意してください。

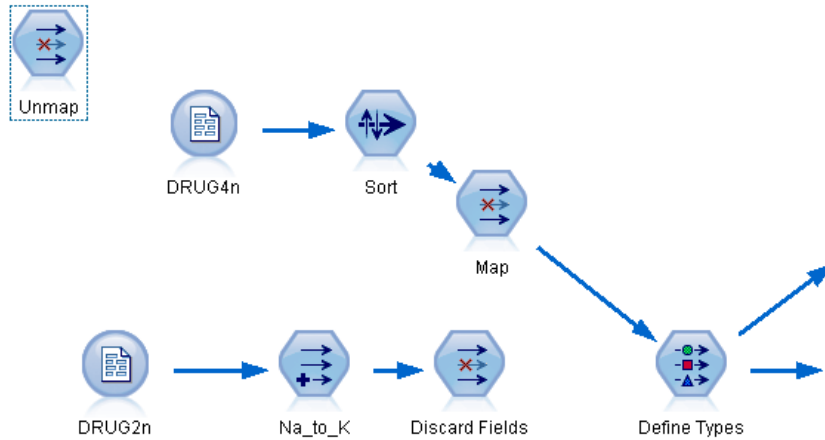
図 5-41
ストリームに正しくマップされた新しいデータソース



ストリーム間のマッピング

ノードの接続と同様に、この方法のデータ マッピングでは、あらかじめ必須フィールドを設定しておく必要はありません。この方法では、データマッピングのポップアップメニューから【マップ先】を使って、あるストリームを別のストリームに単に接続します。このデータ マッピング手法は、ターミナル ノードへのマッピングや、ストリーム間のコピーと貼り付けに役立ちます。注：【マップ先】オプションを使って、レコード結合ノード、レコード追加ノード、およびすべての種類の入力ノードにマップすることはできません。

図 5-42
あるストリームのソートノードを別のストリームのデータ型ノードにマッピング



ストリーム間でデータをマップするには

- ▶ 新規ストリームへの接続に使用するノードを右クリックします。
- ▶ メニューで次の各項目をクリックします。
[データ マッピング(M)] > マップ先
- ▶ カーソルを使って、対象ストリーム上の目的のノードを選択します。
- ▶ 表示されたダイアログ ボックスで、フィールドが正しく対応しているかどうかを確認して、[OK] をクリックします。

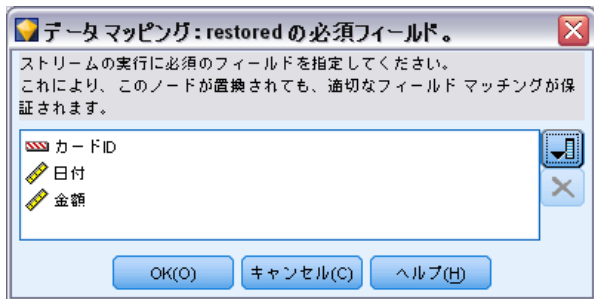
必須フィールドの指定

既存のストリームにマッピングする場合、通常はストリームの作成者により必須フィールドが指定されています。これらの必須フィールドは、特定のフィールドが下流の操作で使用されるかどうかを示します。たとえば、既存のストリームで、Churn というフィールドを使用するモデルを作成している場合を考えてみましょう。フィールド Churn がないとモデルを作成できないため、このストリームでは Churn は必須フィールドになります。同様に、フィールド作成ノードなどの操作ノードで使われるフィールドは、新規フィールドを作成するために必要なフィールドになります。このようなフィールドを明示的に必須フィールドと設定すれば、新しい入力ノード中の適切なフィールドが確実にマップされます。必須フィールドが関連付けられていないと、エラー メッセージが表示されます。ある操作ノードまたは出力ノードを不要と判断したら、それらのノードをストリームから削除し、該当するフィールドを [必須フィールド] リストから削除します。

必須フィールドを設定するには

- ▶ 置換するテンプレート ストリームの入力ノードを右クリックします。
- ▶ メニューで次の各項目をクリックします。
[データ マッピング(M)] > 必須フィールドの指定

図 5-43
必須フィールドの指定

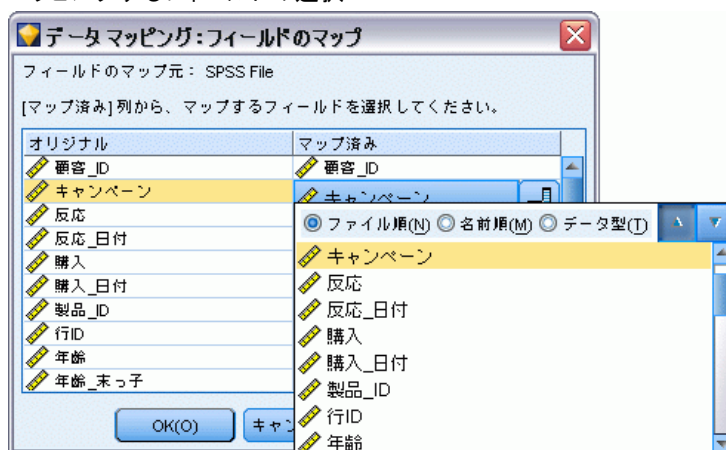


- ▶ フィールド ピッカーを使用して、一覧にフィールドを追加したり除去することができます。フィールド ピッカーを開くには、フィールド リストの右にあるアイコンをクリックしてください。

マップされたフィールドの調査

あるデータ ストリームまたはデータ ソースから別のデータ ストリームまたはデータ ソースにマップする場所を選択したら、マッピングするフィールドを選択する、またはシステム デフォルトのマッピングが正しいかどうかを確認するダイアログ ボックスが表示されます。ストリームまたはデータ ソースの必須フィールドが設定されているけれども、それが一致していない場合、それらのフィールドは赤で表示されます。データ ソースのマッピングされていないフィールドは、変更されずにフィルタ ノードを通過します。ただし、必須でないフィールドもマッピングできることに注意してください。

図 5-44
マッピングするフィールドの選択



オリジナル： テンプレートまたは既存のストリーム中のすべてのフィールド、つまり下流に流れるすべてのフィールドが表示されます。新しいデータソースのフィールドは、これらのフィールドにマップされます。

マップ済み： テンプレートフィールドにマッピングするために選択したフィールドの一覧を表示します。これらのフィールドは、ストリーム操作で使われるオリジナルのフィールドと一致させるために、フィールド名を変更する必要がある場合があります。テーブルセルをクリックすると、リストに利用可能なフィールドが表示されます。

マップするフィールドがわからない場合は、マップする前に入力データを調べることをお勧めします。たとえば、入力ノードの [データ型] タブを使って、入力データの概要を確認することができます。

ヒントとショートカット

ここで説明しているヒントやショートカットを覚えておくと、作業をより簡単かつ円滑に行うことができます。

- **ストリームを素早く構築するにはダブルクリックを使用する：** パレット上のノードを単にダブルクリックすれば、現在のストリーム上にそのノードが追加、接続されます。
- **キーの組み合わせを使って下流のノードを選択する：** Ctrl+Q キーと Ctrl+W キーを使用して、下流のすべてのノード選択を切り替えます。
- **ショートカットキーを使ってノードの接続、接続解除を行う：** ストリーム領域上のノードを選択している場合、F2 キーを押した後、Tab キーを使って接続先ノードを選択し、次に Shift キーを押しながらスペースバーを押すことにより、元のノードと接続先ノードを接続することが

できます。選択したノードのすべての入力接続および出力接続を解除するには、F3 キーを押します。

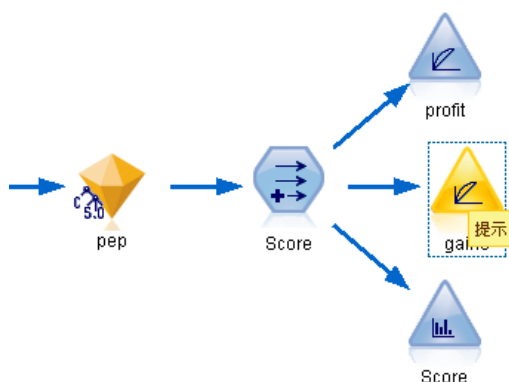
- **[ノード パレット] タブによく使うノードを設定する：** [ツール] メニューの [パレット] をクリックすると、ダイアログ ボックスが表示されます。ここから、[ノード パレット] タブにノードを追加したり、削除することができます。

図 5-45
パレット マネージャ



- **ノードの名前の変更と追加に関するヒント：** 各ノードのダイアログ ボックスには [注釈] タブが用意されています。このタブでは、ノードのストリーム領域上での独自の名前を設定したり、ストリームを構築するために役立つツールヒントを追加することができます。また、進捗状況の追跡を行う目的で、または処理の説明、必要な作業、達成すべきビジネス上の目的などを記録するために、長い注釈を記入できる欄も用意されています。

図 5-46
ツールヒントとユーザーが設定したノード名



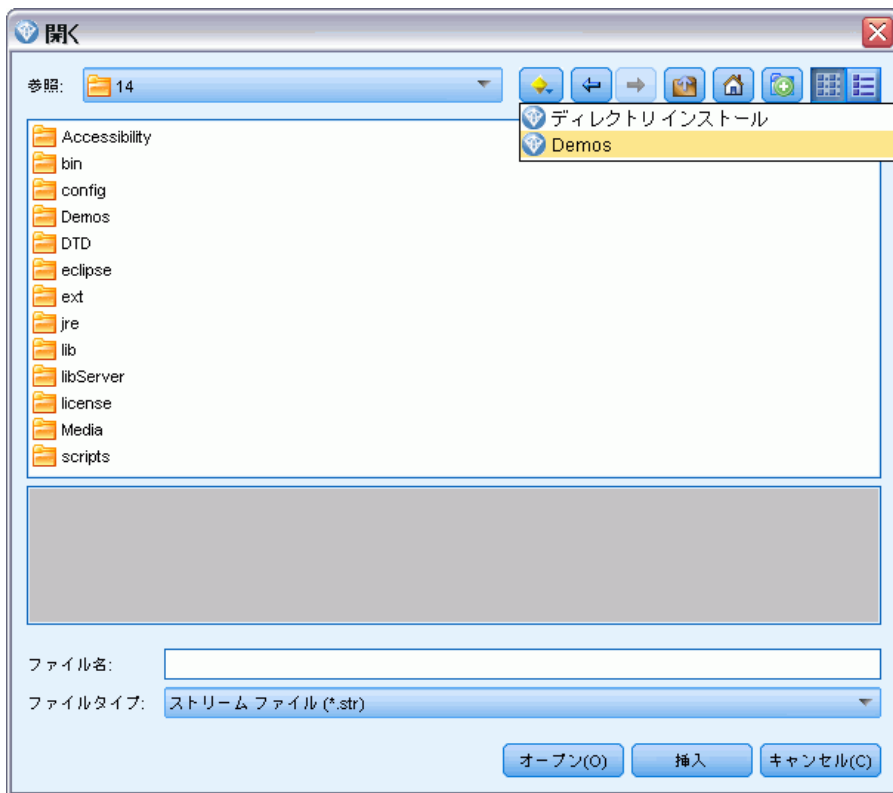
- CLEM 式に値を自動的に入力する**：フィールド作成ノードやフィルタ ノードなどのさまざまなダイアログ ボックスから利用できる式ビルダーを使って、CLEM 式にフィールド値を自動的に挿入することができます。式ビルダーの値ボタンをクリックして、既存のフィールド値から選択してください。

図 5-47
[値] ボタン



- ファイルを素早く参照する**：[開く] ダイアログでファイルの参照時に、[ファイル] リスト（黄色いダイヤモンドの形をしたボタンをクリック）を使って、前に使用したディレクトリや、IBM® SPSS® Modeler のデフォルト ディレクトリに素早くアクセスすることができます。進むボタンおよび戻るボタンを使って、過去に利用したディレクトリを参照することができます。

図 5-48
最近使用されたディレクトリの一覧から Demos フォルダを選択



- **不要な出力ウィンドウを最小化する:** 出力ウィンドウの右上角にある赤い X ボタンを使えば、出力ウィンドウを素早く閉じて削除することができます。こうすることにより、マネージャ パネルの [出力] タブに、注目する結果や興味のある結果だけを保持することができます。

このソフトウェアでは、さまざまなキーボード ショートカットを利用することができます。詳細は、[A 付録 p.284 キーボードのアクセス機能](#) を参照してください。

ご存じですか

- マウスをドラッグして、ストリーム領域上の複数のノード グループを選択することができます。
- あるストリームから別のストリームにノードをコピーして貼り付けることができます。

- ダイアログ ボックスや出力ウィンドウからヘルプを参照することができます。
- CRISP-DM (Cross Industry Standard Process for Data Mining) に関するヘルプを参照できます。([ヘルプ] メニューの [CRISP-DM ヘルプ] をクリックします。)

欠損値の処理

欠損値の概要

データマイニングのデータの前処理フェーズでは、データの欠損値の置換が必要な場合がしばしばあります。欠損値とは、データセット中の不明な値、収集されていない値、または誤って入力された値です。通常、そのような値はフィールドに対して無効になります。たとえば、[性別] フィールドの値は M または F でなくてはなりません。このフィールドに Y や Z などの値があった場合は、値は無効で空白として解釈しても問題ありません。同様に、[年齢] フィールドに負の値がある場合も、意味がないため空白として扱われます。このような明白に誤った値が故意に入力されたり、質問に対して回答したくないことを示すために、フィールドを空欄のまま放置することはよくあります。時には、年齢を入力していないなどの未回答が、特定の結果を予測するための因子であるかどうかを判断するために、空白の意味を詳しく調べる必要があることもあります。

一部のモデル作成手法では、欠損値がより効果的に処理されます。たとえば、C5.0、および Apriori は、データ型ノードで明示的に「欠損値」と宣言された値を効果的に処理しています。他のモデル作成手法では、欠損値を処理できないため、学習に時間がかかったり、モデルの精度が低下してしまいます。

IBM® SPSS® Modeler の欠損値には、次の 2 種類があります。

- **ヌル値またはシステム欠損値**：これらの値は、データベースまたはソースファイルに空白のまま残された文字列以外の値であり、入力ノードまたはデータ型ノードで特に「欠損値」として定義されていません。システム欠損値は \$null\$ 値として表示されます。空の文字列が特定のデータベースでヌルとして処理される場合でも、SPSS Modeler では空の文字列をヌルとは見なさないことに注意してください。
- **空文字列と空白文字**：空文字の値と空白文字（表示されない文字による文字列）をヌル値の重複レコードとして処理します。空白の文字列は、ほとんどの目的に対してホワイトスペースとして扱われます。たとえば、オプションを選択してソースまたはデータ型ノードで空白文字を空白値として扱う場合、この設定は空白の文字列も同様に適用します。
- **空白値またはユーザー定義の欠損値**：これには、unknown、99、または -1 などの値で、入力ノードまたはデータ型ノード中では明確に欠損値として定義されています。オプションでヌルと空白文字を「空白」として処理することもできます。そうすることによって、特別な処理のためにフラグを付けたり、ほとんどの計算から除外することができるようになります。たとえば、@BLANK 関数を使用して、これらの値を他の欠損値と

共に空白値として処理することができます。詳細は、4章 [値] ダイアログ ボックスの使用 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。

図 6-1
連続型変数の欠損値の指定

The screenshot shows the 'age 値' dialog box with the following settings:

- データ型: 範囲 (Range)
- ストレージ: 整数 (Integer)
- 値: 値とラベルを指定 (Specify values and labels)
- 下限: 18
- 上限: 67
- 値の検査: なし (None)
- 空白を定義 (Define missing):
- 欠損値 (Missing value): (empty field)
- 範囲 (Range): -1 ~ 17
- ヌル (Null):
- 空白文字 (Blank text):

混在データの読み込み: 数値ストレージ（整数、実数、時間、タイムスタンプ、日付のいずれか）を持つフィールドで読み込みを行う場合、数値以外の値が「ヌル」または「システムによる欠損」に設定されることに注意してください。これは、他のアプリケーションと異なり、フィールド内にストレージの混在を許さないためです。これを回避するためには、必要に応じて入力ノードまたは外部アプリケーション内の記憶域タイプを変更し、混在データを含むフィールドを文字列として読み込む必要があります。詳細は、2章 フィールドのストレージと形式の設定 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。

Oracle からの空の文字列の読み取り: Oracle データベースとの間の読み書き時には、SPSS Modeler やその他のほとんどのデータベースと異なり、空の文字列がヌル値と同様に取り扱われ、格納されることに注意してください。つまり、Oracle データベースから抽出されたデータは、同じデータがファイルやその他のデータベースから抽出された場合とは異なって動作し、また異なる結果が返ることがあります。

欠損値の処理

業務やドメインの知識の観点から、欠損値をどのように取り扱うかを決定する必要があります。多くの場合、学習時間を短縮し、精度を向上させるには、データセットから空白を削除する必要があります。その一方で、空白値から、新しいビジネス チャンスを発見したり、新たな洞察を加えられることもあります。最適テクニックを選択するときに、次に示すデータの側面を検討する必要があります。

- データ セットのサイズ
- 空白を含むフィールドの数
- 失われた情報の量

一般的に、アプローチは次の 2 通りあります。

- 欠損値を含むフィールドまたはレコードを除外する。
- 様々な方法で欠損値を代入、置換または強制する。

これら 2 つのアプローチは、データ検査ノードを使用し、大半を自動化することができます。たとえば、フィルタ ノードを生成し、モデル作成で有用な欠損値をあまりに多く含むフィールドを除外することができます。また、スーパー ノードを作成して維持するいずれかまたはすべてのフィールドに対して欠損値を代入することができます。このため検査によって実際にデータの現在の状態を評価するだけでなく、評価に基づいて行動を起こすことができます。詳細は、[8 章 分析用のデータの準備 \(データ検査\) in IBM SPSS Modeler 14.2 アプリケーション ガイド](#) を参照してください。

欠損値を含むレコードの処理

欠損値の大部分が少数のレコードに集中している場合、それらのレコードだけを除外することができます。たとえば、銀行にはローン顧客に関する詳細で完全なレコードがあるのが普通です。しかし、行員用のローンの承認では空白に関する制限が緩和されており、行員ローン用に収集されたデータにいくつかの空白のフィールドが含まれていることがあります。このような場合のこれらの欠損値の処理には、2 つのオプションがあります。

- 条件抽出ノードを使用して、スタッフ レコードを削除できます。
- データ セットが大きい場合、空白を含むレコードを破棄することができます。詳細は、[6 章 欠損データを含むレコードの選択 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

欠損値を含むフィールドの処理

欠損値の大部分が少数のフィールドに集中している場合は、レコードレベルではなくフィールドレベルで欠損値に対処することができます。この方法では、欠損値の処理方法を決める前に、特定のフィールドの相対重要度を調べることがもできます。フィールドがモデル作成において重要ではない場合、欠損値の過多に関わらず、そのフィールドを保持する価値はないと思われます。

たとえば、マーケットリサーチ会社がデータの収集に 50 の質問から成る一般的なアンケートを使用する場合を考えてみましょう。そのうちの 2 つの質問は年齢と政治信条に関するもので、これは多くの人が提供をためらう情報です。この場合、年齢と政治信条には多くの欠損値が存在することになります。

フィールドの尺度

どの方法を使用するかを決める際には、欠損値のあるフィールドの尺度も考慮する必要があります。

数値型フィールド:連続型などの数値型フィールドの場合は、数値以外の値を除去してからモデルを作成する必要があります。数値型フィールドに空白が含まれていると、多くのモデルは機能しません。

カテゴリフィールド:数値型やフラグ型などのカテゴリフィールドの場合は、欠損値の変更は必須ではありませんが、変更するとモデルの精度が向上します。たとえば、フィールド [性別] を使用するモデルは、Y や Z のような意味のない値があっても機能しますが、M や F 以外の値をすべて削除すればモデルの精度が向上します。

スクリーニングまたは削除フィールド

あまりに多くの欠損値を含むフィールドのスクリーニングには、次のようなオプションがあります。

- データ検査ノードを使用し、品質に基づいてフィールドをフィルタリングすることができます。詳細は、[6 章 欠損データを含むフィールドのフィルタリング in IBM SPSS Modeler 14.2 入力ノード、プロセスノード、出力ノード](#) を参照してください。
- フィールド選択ノードを使用して、指定した欠損値のパーセンテージ以上のフィールドをスクリーニングし、指定した目標値を基準にした重要度に基づいてフィールドをランク付けします。詳細は、[4 章 フィールド選択ノード in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。
- フィールドを削除する代わりに、データ型ノードを使ってフィールドの役割をなしに設定することもできます。これにより、フィールドはデータセット内に維持されますが、モデル作成プロセスからは除外されます。

欠損値の代入または置換

このようにわずかな欠損値が存在しているような場合は、値を挿入して空白と置換すると効果的な場合があります。この処理はデータ検査レポートから行うことができ、それにより必要に応じて特定のフィールドに対しオプションを指定し、様々な方法を使用して値を代入するスーパー ノードを作成することができます。これは最もフレキシブルな方法で、単一ノード内にある多くのフィールドに対し処理を指定することができます。

次の方法で、欠損値の代入ができます。

固定： 固定値で置き換えます（指定のフィールド計測、範囲の中間または一定数）。

無作為： 正常または均一分布に基づいたランダム値で置き換えます。

式： ユーザー設定の式を指定することができます。たとえば、値をグローバル値設定ノードで作成されたグローバル変数と置き換えることができます。

アルゴリズム： C&RT アルゴリズムの基づいたモデルによって予測された値で置き換えます。この方法で代入された各フィールドに対し、空白値やヌル値をモデルで予測された値と置き換える置換ノードとともに、個別の C&RT モデルが作成されます。フィルタ ノードを使用して、モデルが生成した予測値を削除します。

詳細は、[6 章 値の代入 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

代わりに、特定のフィールドに対して強制するには、データ型ノードを使用して、フィールドのデータ型が正しい値だけを利用するために、空白値の置換が必要なフィールドの [検査] 列を [強制] に設定します。 [詳細は、4 章 データ型ノード in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

欠損値用 CLEM 関数

欠損値の処理には、さまざまな関数を利用できます。条件抽出ノードや置換ノードで欠損値を破棄したり、値を代入するためによく用いられる関数を次に示します。

- `count_nulls(LIST)`
- `@BLANK(FIELD)`
- `@NULL(FIELD)`
- `undef`

@FIELD 関数とともに @ 関数を使用して、1 つまたは複数のフィールドに空白やヌル値が存在するかどうかを判断することができます。空白またはヌル値があるフィールドに単にフラグを設定することも、それを他の演算子を使って適切な値に置換することもできます。

次に示すように、フィールドのリスト全般にわたってヌルをカウントできます。

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

フィールドのリストを入力として受け容れる関数を使用すると、特別な関数 @FIELDS_BETWEEN と @FIELDS_MATCHING を以下の例のように使用できます。

```
count_nulls(@FIELDS_MATCHING('card*'))
```

図 6-2
置換ノードを使って選択したフィールドの空白値を 0 に置換



undef 関数を使って、フィールドを \$null\$ として表示されるシステム欠損値で埋めることができます。たとえば、任意の数値を置換するには、次のような条件ステートメントを使用することができます。

```
if not(年齢 > 17) or not(年齢 < 66) then undef else 年齢 endif
```

このステートメントでは、範囲内にはない値を、`$null$` として表示されるシステム欠損値で置き換えます。`not()` 関数を使うと、負の数値も含めた他のすべての数値を指定することもできます。詳細は、[8 章 p.172 空白値とヌル値処理関数](#) を参照してください。

レコードの破棄に関する注意

条件抽出ノードを使ってレコードを破棄する場合、シンタックスでは 3 値論理を用いて、選択したステートメントのヌル値を自動的に含めることに注意してください。選択した式からヌル値（システム欠損値）を除外するには、式で `and not` を使って明示的に指定しなければなりません。たとえば、処方薬の種類が薬品 C であるすべてのレコードを選択し、含めるには、次のステートメントを使用します。

```
薬品 = '薬品C' and not(@NULL(薬品))
```

前のバージョンでは、このような状況の場合ヌル値を除外していました。

CLEM 式の作成

CLEM について

Control Language for Expression Manipulation (CLEM) は、IBM® SPSS® Modeler ストリームで利用するデータの分析と処理を行うための強力な言語です。CLEM を使用すれば、経費と収入データから利益を算出するような簡単な操作から、Web ログ データを有益な情報を含む一連のフィールドやレコードに変換するような複雑な操作まで、さまざまなストリーム操作を行うことができます。

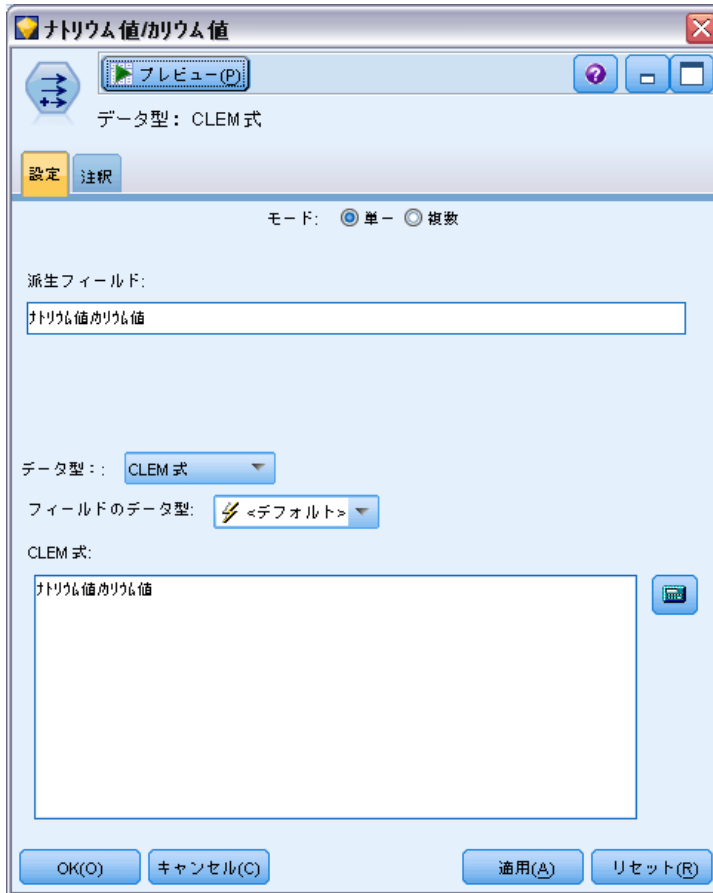
CLEM は、SPSS Modeler で次の目的のために使用されます。

- レコード フィールドに対する条件の比較と評価。
- 新規フィールドの値の作成。
- 既存のフィールドの新しい値の作成。
- レコードのシーケンスの推測。
- レコードからレポートへのデータ挿入。

スクリプト :CLEM 言語のサブセットは、また、ユーザー インターフェイスでスクリプトを使用する場合に使用することができます。これによって、同じデータ操作の大部分を自動的に行うことができます。 [詳細は、2 章 スクリプトの概要 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。](#)

CLEM 式は SPSS Modeler でのデータ準備に欠くことができず、レコードとフィールド設定（選択、バランス、置換）からプロットと出力（分析、レポート、テーブル）まで、広範囲にわたるノードに使用できます。たとえば、CLEM をフィールド作成ノードに使用して、比率などの、式に準拠する新規フィールドを作成できます。

図 7-1
式に基づいて新規フィールドを作成するフィールド作成ノード



CLEM 式は、また、グローバル検索と置き換え操作にも使用できます。たとえば、@NULL(@FIELD) 式を置換ノードに使用して、整数値 0 でシステム欠損値を置き換えることができます（空白とも呼ばれるユーザー欠損値を置き換える場合は、@BLANK 関数を使います）。

図 7-2
システム欠損値を 0 に置き換える置換ノード



また、より複雑な CLEM 式を作成することも可能です。たとえば、ルールの条件セットに基づいて、新規フィールドを作成することもできます。

図 7-3
あるフィールドの値をその前のフィールドと比較する条件式



CLEM の例

正しい指定形式や、CLEM で利用できる式の種類を理解するために、次の式の例を参照してください。

簡単な式

フィールド After および Before の値に基づいて新しいフィールドを作成する簡単な式の例を次に示します。

$(\text{After} - \text{Before}) / \text{Before} * 100.0$

フィールドの値を参照するには、フィールド名を引用符で囲まないことに注意してください。

同様に、次の式は単にフィールド salary の各値の対数を返します。

```
log(salary)
```

複雑な式

より複雑で長い式を使用することもできます。次の式は、2 つのフィールド (\$KX-Kohonen および \$KY-Kohonen) の値が、指定した範囲内の場合に true (真) の値を返します。ここでは、フィールド名に特殊文字が含まれているため、フィールド名を単一引用符で囲んでいることに注意してください。

```
(' $KX-Kohonen ' >= -0.2635771036148072 and ' $KX-Kohonen ' <= 0.3146203637123107
and ' $KY-Kohonen ' >= -0.18975617885589602 and
' $KY-Kohonen ' <= 0.17674794197082522) -> T
```

文字列関数など多くの関数では、さまざまなパラメータを正しい形式で指定する必要があります。次の例では、商品が有機栽培 (organic)、遺伝子組み換え食品 (genetically modified)、または従来農法 (conventional) かを判断するために、フィールド produce_ID の先頭文字を取得するには、関数 **subscrs** を以下のように使用します。式の結果は、->`result` により記述されます。

```
subscrs(1,produce_ID) -> `c`
```

次の式も同様です。

```
stripchar('3','123') -> `12`
```

文字は常に単一逆引用符で囲むことに注意してください。

式中での関数の組み合わせ

CLEM 式は複数の関数の組み合わせで成り立っていることもあります。関数 **subscr** および **lowertoupper** を組み合わせて、produce_ID の先頭文字を取得し、それを大文字に変換する式の例を次に示します。

```
lowertoupper(subscr(1,produce_ID)) -> `C`
```

この式は、次のようにさらに短い形式で記述することもできます。

```
lowertoupper(produce_ID(1)) -> `C`
```

よく使われる関数の組み合わせの例を、もう一つ次に示します。

```
locchar_back(`n`, (length(web_page)), web_page)
```

この式は、フィールド web_page 中の値を、値の最後の文字から先頭方向に向かって、文字 `n` を検索していきます。この式で値の長さに 7 のような特定の数字を使用すると、その値 (7 文字) 未満の値に対して式が不正

になってしまいます。そこで、この式では `length` 関数と組み合わせることによって、動的に現在の値の長さを計算して使用します。

特殊関数

さまざまな特殊関数（先頭に `@` 記号がある）を利用することができます。よく使われる特殊関数を次に示します。

`@BLANK('referrer ID')->T`

特殊関数は、組み合わせて使用され、同時に複数のフィールドに対して、フィールドが空の場合にフラグを設定するためによく用いられる方法です。

`@BLANK(@FIELD)->T`

CLEM マニュアルには、ほかにもさまざまな例が記載されています。 [詳細は、8 章 p.136 CLEM リファレンス概要 を参照してください。](#)

値とデータ型

CLEM 式は、値、フィールド名、演算子、および関数から構成される式と同じようなものです。有効で最も単純な CLEM 式は、1 個の値またはフィールド名になります。次に、有効な値の例を示します。

3

1.79

'banana'

次に、有効なフィールド名の例を示します。

Product_ID

'\$P-NextField'

「商品」は Market Basket データ セットからのフィールド名で、「'\$P-NextField'」はパラメータ名になります。式の値は、指定されたフィールドの値になります。通常、フィールド名は文字から始まります。数字や下線 (`_`) を記述することもできます。フィールド名を引用符で囲めば、この規則に従わないフィールド名も使用できます。CLEM の値は次のいずれかになります。

- 文字列: "c1"、"Type 2"、"a piece of free text" など
- 整数: 12、0、-189 など
- 実数: 12.34、0.0、-0.0045 など
- 日付/時刻フィールド: 05/12/2002、12/05/2002、12/05/02 など

次の要素も使用できます。

- 文字コード：`a` or 3 など
- アイテムのリスト：[1 2 3]、['Type 1' 'Type 2'] など

通常は文字コードやリストがフィールドの値になることはありません。これらの要素は、CLEM 関数の引数として使われます。

引用規則

本ソフトウェアでは、CLEM 式で使われるフィールド、値、パラメータ、および文字列などを柔軟に指定することができますが、次の規則にしたがって式を作成することをお勧めします。

- **文字列**：文字列を指定する場合は、常に二重引用符を使用してください（"Type 2" または "value"）。単一引用符を使用することもできますが、引用符で囲まれたフィールドと誤解される危険性があります。
- **文字**：` のような単一の逆向き引用符を常に使用します。たとえば、関数 `stripchar('d','drugA')` に文字 `d` があるとします。これに関する唯一の例外は、文字列中の特定の文字を参照するために整数を使用する場合です。たとえば、関数 `lowertoupper("drugA"(5))` → "A" に文字 `5` があるとします。注：標準の U.K. および U.S. キーボードでは、逆引用符文字（アクサン グラフ、Unicode 0060）のキーは、Esc キーの下にあります。
- **フィールド**—通常、CLEM 式でフィールドを使用する場合、引用符は付けません（`subscr(2,arrayID)` → CHAR）。スペースや他の特殊文字を入れる必要があるような場合には、単一引用符を使用することもあります（'Order Number'）。データ セット中に単一引用符で囲まれているのに未定義のフィールドがあると、それは文字列として読み込まれてしまいます。
- **パラメータ**—常に単一引用符で囲んでください（'\$P-threshold'）。

式と条件

CLEM 式が返す結果の例を次に示します（新しい値を作成する場合に使用）。

```
Weight * 2.2
Age + 1
sqrt(Signal-Echo)
```

または、次のように、真 (true) あるいは偽 (false) を評価することもできます（条件に基づいて抽出する場合に使用）。

```
Drug = "drugA"
Age < 16
not(PowerFlux) and Power > 2000
```

CLEM 式の中では、次のように演算子と関数を自由に組み合わせることができます。

```
sqrt(abs(Signal)) * max(T1, T2) + Baseline
```

括弧と演算子の優先順位によって、式が計算される順序が決まります。この例では、計算は次の順序で行われます。

- まず、**abs(Signal)** が計算され、その結果に対して **sqrt** が計算されます。
- **max(T1, T2)** が計算されます。
- 2 つの結果が乗算されます。x の優先度は + より高くなります。
- 最後に、上の結果に **Baseline** が加算されます。

優先順位を降順（最初に計算されるものから最後に計算されるものの順）に記載すると、次のようになります。

- 関数の引数
- 関数の呼び出し
- xx
- x / mod div rem
- +-
- ><>= <= /== == =/=

優先順位を変更する場合、または計算の順序がよくわからない場合は、次のように括弧を使用して、計算の順序を明確に指定することができます。

```
sqrt(abs(Signal)) * (max(T1, T2) + Baseline)
```

ストリーム、セッション、およびスーパーノード パラメータ

パラメータは、CLEM 式とスクリプトで使用するために定義できます。実際のところ、パラメータはユーザー定義の変数であり、保存されて、現在のストリーム、セッション、またはスーパーノードで継続します。さらに、スクリプトを使用する場合と同様に、ユーザー インターフェイスからもアクセスできます。たとえば、ストリームを保存すると、そのストリームに設定されているパラメータも保存されます。（これは、ローカル スクリプト変数と異なる点です。ローカル スクリプト変数は、宣言されたスクリプト内でのみ使用できます。）通常パラメータは、スクリプト中でパラメータ値を指定する CLEM 式の一部として使用されます。

パラメータの有効範囲は、それがどこで設定されたかによって異なります。

- ストリーム パラメータは、ストリーム スクリプト内またはストリーム プロパティのダイアログ ボックス内で設定でき、ストリーム内のすべてのノードで使用できます。Clem 式ビルダーの [パラメータ] リストに表示されます。
- セッション パラメータは、スタンドアロン スクリプト内または [セッション パラメータ] ダイアログ ボックス内で設定できます。セッション パラメータは、現在のセッションのすべてのストリーム ([マネージャ] ウィンドウの [ストリーム] タブに表示されているすべてのストリーム) で利用できます。

パラメータは、スーパーノード用にも設定できます。この場合、スーパーノード内にカプセル化されたノードでだけ表示できます。詳細は、9 章 [スーパーノードのパラメータの定義 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

CLEM 式でのパラメータの使用

CLEM 式で、パラメータは `$P-pname` の形式で表されます。ここで、`pname` はパラメータの名前です。CLEM 式でパラメータを使用する場合は、'`$P-scale`' のように単一引用符で囲む必要があります。

利用できるパラメータは、Clem 式ビルダーを使って簡単に参照することができます。現在のパラメータを表示するには

- ▶ CLEM 式を利用できる任意のダイアログ ボックスで、Clem 式ビルダー ボタンをクリックします。
- ▶ [フィールド] リストで、[パラメータ] を選択します。

CLEM 式に挿入するパラメータをリストから選択することができます。詳細は、[p. 130 フィールド、パラメータ、およびグローバル変数の選択](#) を参照してください。

文字列の処理

文字列に対しては、次のような操作を行うことができます。

- 文字列の大文字または小文字への変換、`uppertolower(CHAR)`。
- 文字列変数から `'ID_'` または `'$'` など指定の文字の削除 `-stripchar(CHAR,STRING)`。
- 文字列変数の長さ (文字数) の判断 `-length(STRING)`。
- 文字列値の文字順のチェック `-alphabefore(STRING1, STRING2)`。
- 先頭または末尾の空白文字を値から削除 `-trim(STRING)`、`trim_start(STRING)`、または `trimend(STRING)`。

- 指定した文字列から最初または最後のn 文字を抽出—**startstring(LENGTH, STRING) or endstring(LENGTH, STRING)**。たとえば、製品名と4桁のIDコードを結合する (ACME CAMERA-D109) フィールド名 item があるとします。4桁のコードのみを含む新規フィールドを作成するには、フィールド作成コードで次の式を指定します。

`endstring(4, item)`

- 特定のパターン的一致—**STRING matches PATTERN**。たとえば、職名に「market」が付いている人を選択するには、条件抽出ノードで次のように指定します。

`job_title matches "*"market"`

- 文字列内のサブ文字列のすべてのインスタンスの置換—**replace(SUBSTRING, NEWSUBSTRING, STRING)**。垂直線 (|) など、サポートされていない文字のすべてのインスタンスを、テキストマイニングを行う前にセミコロンと置き換えるには、置換ノードの **replace** 関数を使用します。[対象フィールド] で、サポートされていない文字のあるすべてのフィールドを選択します。[置換] 条件の場合、[常に] を選択し、[置換値] で次の条件を指定します。

`replace(';', '|', @FIELD)`

- 特定のサブ文字列の有無に基づいた、フラグ型フィールドの作成。たとえば、それぞれの回答に次のような式を使用して個別のフラグ型フィールドを生成するために、フィールド生成ノードの文字列関数を使用できます。

`hassubstring(museums, "museum_of_design")`

詳細は、8章 p.154 文字列関数 を参照してください。

空白および欠損値の処理

空白または欠損値を置き換えることは、データマイナーに一般的なデータ準備タスクです。CLEMでは、空白処理を自動化するさまざまなツールを用意しています。空白の処理には、置換ノードがよく使用されますが、次の関数はCLEM式を利用できる任意のノードで使用することができます。

- **@BLANK(FIELD)** : Ageなどの特定のフィールドの値が空のレコードを判断するために用いられます。
- **@NULL(FIELD)** : 指定されたフィールドの値がシステム欠損値のレコードを判断するために用いられます。IBM® SPSS® Modelerでは、システム欠損値は `$null$` 値として表示されます。

図 7-4
システム欠損値を 0 に置き換える置換ノード



詳細は、 8 章 p.172 空白値とヌル値処理関数 を参照してください。

数値の処理

IBM® SPSS® Modeler では、次のような数値に対する標準の操作を利用することができます。

- 指定した角度のサインの計算- $\sin(\text{NUM})$ 。
- 数値フィールドの自然対数の算出- $\log(\text{NUM})$ 。
- 2 つの数値の合計の算出- $\text{NUM1} + \text{NUM2}$ 。

詳細は、 8 章 p.150 数値関数 を参照してください。

時間と日付の処理

時間および日付の形式は、データ ソースやロケールによってさまざまです。使用する日付と時間の書式はストリームごとに異なり、[ストリームのプロパティ] ダイアログ ボックスで設定されます。日付/時間フィールドの処理に一般的に使われる関数の例を次に示します。

経過時間の算出

基準日からの経過時間は、次のような関数ファミリーを使用すれば簡単に算出することができます。この関数は、基準日から日付文字列 **DATE** で指定された日付までの月数を、実数で返します。これは、1 か月を 30.0 日と仮定した近似値になります。

`date_in_months(Date)`

日付/時間値の比較

日付/時間フィールドの値は、次のような関数を使ってレコード間で比較することができます。この関数は、日付文字列 **DATE1** で指定された日付が、日付文字列 **DATE2** で指定された日付よりも前の場合に、true (真) の値を返します。それ以外の場合、この関数は 0 の値を返します。

`date_before(Date1, Date2)`

差分の計算

次のような関数を使用すれば、2 つの時間や 2 つの日付の差分を算出することもできます。

`date_weeks_difference(Date1, Date2)`

この関数は、日付文字列 **DATE1** が表す日付から日付文字列 **DATE2** が表す日付までの時間を、実数の週数で返します。ここでは、1 週間を 7.0 日と仮定しています。**DATE2** が **DATE1** よりも前の場合、この関数は負の数値を返します。

今日の日付

関数 **@TODAY** を使用すれば、データ セットに現在の日付を追加することができます。現在の日付は、[ストリームのプロパティ] ダイアログ ボックスで指定されている日付の形式を使って、指定されたフィールドまたは新しいフィールドに文字列として追加されます。詳細は、8 章 p.160 [日付および時刻の関数](#) を参照してください。

複数フィールドの要約

CLEM 言語には、複数フィールドにわたって要約統計量を返すさまざまな関数が含まれます。これらの関数は調査データの分析に特に役立つことがあります、その場合、1 つの質問に対する複数のレスポンスを複数フィールドに保存することもできます。詳細は、p.126 [複数レスポンス データの処理](#) を参照してください。

比較関数

複数フィールドにわたって、`min_n` 関数と `max_n` 関数を使って値を比較できます。たとえば、次のような場合です。

```
max_n(['card1fee' 'card2fee' 'card3fee' 'card4fee'])
```

また、さまざまな個数関数を使って特定基準を満たす値のカウントを取得することができ、たとえ、値が複数フィールドに保存されている場合でもそれは可能です。たとえば、5 年を越えて保持されてきたカードの数をカウントするには次のようにします。

```
count_greater_than(5, ['cardtenure' 'card2tenure' 'card3tenure'])
```

同じフィールド セットにわたってヌル値をカウントするには次のようにします。

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

この例でカウントされているのは、カードを保持する人たちの数ではなく、保持されているカードの数であることに注意してください。 [詳細は、8 章 p. 147 比較関数 を参照してください。](#)

数値関数

複数フィールドにわたって、`sum_n` 関数、`mean_n` 関数、`sdev_n` 関数を使って統計を取得できます。たとえば、次のようになります。

```
sum_n(['card1bal' 'card2bal' 'card3bal'])
```

```
mean_n(['card1bal' 'card2bal' 'card3bal'])
```

[詳細は、8 章 p. 150 数値関数 を参照してください。](#)

フィールドのリストを生成

フィールドのリストを入力として解釈する関数を使用する場合、`@FIELDS_BETWEEN(start, end)` および `@FIELDS_MATCHING(pattern)` の特殊関数を入力として使用できます。たとえば、フィールドの順序が上記の `sum_n` の例に示すようなものと仮定すると、以下は同等になります。

```
sum_n(@FIELDS_BETWEEN(card1bal, card3bal))
```

また、「card」で始まるすべてのフィールドにわたってヌル値の数をカウントするには、次のようにします。

```
count_nulls(@FIELDS_MATCHING('card*'))
```

[詳細は、8 章 p. 173 特殊フィールド を参照してください。](#)

複数レスポンス データの処理

さまざまな比較関数を使用して、次のような複数レスポンス データを分析することができます。

- value_at
- first_index / last_index
- first_non_null / last_non_null
- first_non_null_index / last_non_null_index
- min_index / max_index

たとえば、特定の購買を決定した 1 番目、2 番目、3 番目に重要な理由（たとえば、価格、人に勧められた、レビュー、地元の提供業者など）を尋ねる複数回答の質問があるとします。この場合、次のように最初に表示されたフィールドのインデックスを作成して価格の重要度を決定します。

```
first_index("price", [Reason1 Reason2 Reason3])
```

同様に、顧客に質問して購買の尤度の順に 3 つの車をランク付けし、次のように 3 つの個別のフィールドにコード化します。

customer id	car1	car2	car3
101	1	3	2
102	3	2	1
103	2	3	1

この場合、`min_index` 関数を使用して、最も好きな車（ランク #1、または最も低いランク）のフィールドのインデックスを決定します。

```
min_index(['car1' 'car2' 'car3'])
```

詳細は、[8 章 p. 147 比較関数](#) を参照してください。

複数レスポンス セットの参照

特別な `@MULTI_RESPONSE_SET` 関数を使用して、複数レスポンス セットのすべてのフィールドを参照することができます。たとえば、上記の 3 つの car フィールドが `car_rankings` という名前の複数レスポンス セットにある場合、次の関数が同じ結果を返します。

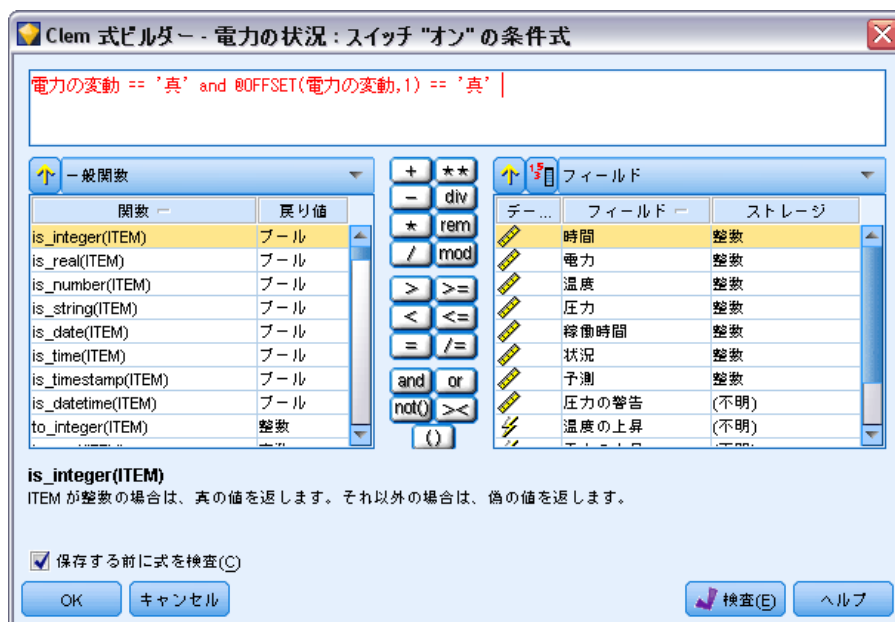
```
max_index(@MULTI_RESPONSE_SET("car_rankings"))
```

詳細は、[4 章 複数レスポンス セット編集 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

Clem 式ビルダー

CLEM 式を手動で入力したり式ビルダーを使用したりすることができ、それによって、現在ストリームのデータ フィールドに加えて、CLEM の関数と演算子のリストがすべて表示されます。このため、フィールド名と関数名を正確に覚えていなくても、すばやく CLEM 式を作成できます。加えて、式ビルダーのコントロールによってフィールドや値に適切な引用符が自動的に追加され、正しい式をシンタックス的に作成することが簡単になります。

図 7-5
[Clem 式ビルダー] ダイアログ ボックス



注： スクリプトまたはパラメータの設定では、Clem 式ビルダーを利用できません。

Clem 式ビルダーへのアクセス

Clem 式ビルダーは、条件抽出、バランス、フィールド作成、置換、散布図、線グラフ、精度分析、レポート、およびテーブル ノードの式などの CLEM 式を使うノードのすべてで使用できます。Clem 式ビルダーを表示するには、CLEM 式フィールドの右にある計算機ボタンをクリックします。

図 7-6
ノードと Clem 式ビルダー ボタン



CLEM 式の作成

Clem 式ビルダーには、フィールド、関数、および演算子の完全なリストがあるだけでなく、データがインスタンス化されている場合はデータ値へのアクセス手段も提供されています。

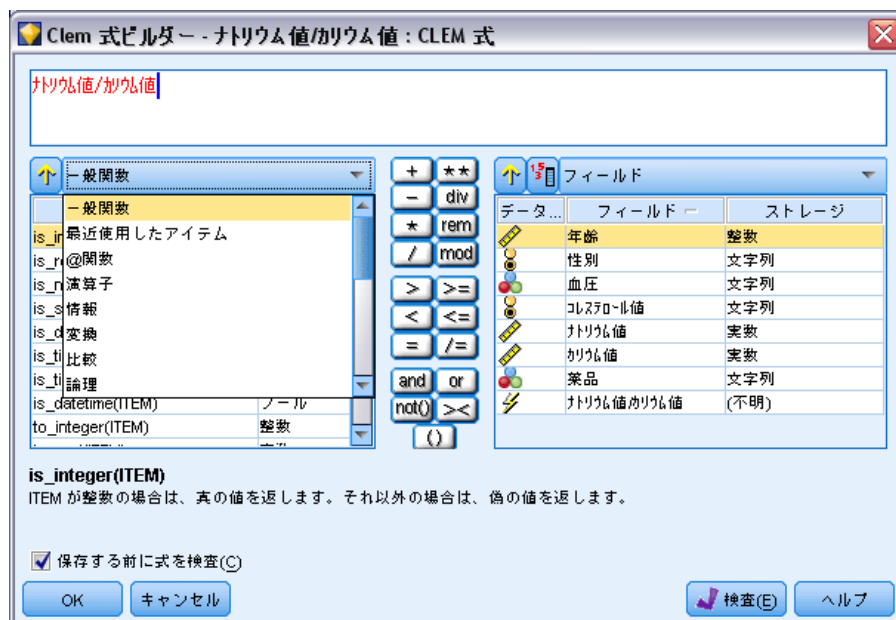
Clem 式ビルダーを使用して CLEM 式を作成する手順

- ▶ Clem 式フィールドとフィールドのリストを参照しながら、ウィンドウに入力します。
- or
- ▶ スクロール リストから適切なフィールドと関数を選択します。
 - ▶ 黄色の矢印ボタンをダブル クリックまたはクリックして、CLEM 式フィールドにフィールドや関数を追加します。
 - ▶ オペランドを式に挿入するには、ダイアログ ボックスの中央にあるオペランド ボタンを使用します。

関数の選択

関数リストには、使用可能なすべての CLEM 関数と演算子が表示されています。リストをスクロールして関数を選択します。または、すばやく目的のものを探すには、ドロップダウン リストを使って関数または演算子のサブセットを表示します。利用できる関数は、探しやすいようにカテゴリにグループ化されています。

図 7-7
[関数] ドロップダウン リスト



特に役に立つ 2 種類のカテゴリがあります。

- **すべての関数:** 利用できるすべての CLEM 関数が表示されます。
- **最近使用した演算子:** 現在のセッション内で使われた CLEM 関数が表示されます。

関数のグループを選択して、ダブルクリックすると CLEM 式フィールド中のカーソルの位置に関数が挿入されます。

フィールド、パラメータ、およびグローバル変数の選択

[フィールド] リストには、現在データ ストリームで利用できるすべてのフィールドが表示されます。リストをスクロールして、目的のフィールドを選択してください。フィールドを上の式に追加するには、ダブルクリックするか、または黄色の矢印キーを使用します。

図 7-8
Clem 式ビルダー : [フィールド] リスト



詳細は、[p. 120 ストリーム、セッション、およびスーパーノード パラメータ](#) を参照してください。

フィールドの他に、次の項目から選択することもできます。

複数レスポンス セット: 詳細については、『IBM SPSS Modeler 入力、プロセス、出力ノードガイド』を参照してください。

最近使用した演算子: 現在のセッション内で使われたフィールド、マルチアンサー セット、パラメータ、グローバル値が表示されます。

パラメータ 詳細は、 p.120 ストリーム、セッション、およびスーパーノード **パラメータ** を参照してください。

グローバル値:詳細については、『IBM SPSS Modeler 入力、プロセス、出力ノードガイド』を参照してください。

値の表示または選択

フィールドの値は、式ビルダー、データ検査レポートを含めたシステム内の多くの場所からと、時間区分ノード内で将来の値を編集しているときに、表示できます。この機能を使用するには、データがソースまたはデータ型ノード内で完全にインスタンス化されていることが必要であり、そのようにすると、ストレージ、データ型、および値が明らかになります。詳細は、4章 [値] ダイアログ ボックスの使用 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノードを参照してください。

図 7-9

フィールド リストと選択されたフィールドに表示された値



- ▶ Clem 式ビルダーまたは時間区分ノードからフィールドの値を表示するには、表示するフィールドを選択して値ピッカー ボタンをクリックすると、選択したフィールドの値を表示するダイアログ ボックスが表示されます。次に値を選択して、その値を現在の式またはリストに貼り付けるために、[挿入] をクリックします。

図 7-10
値ピッカーボタン

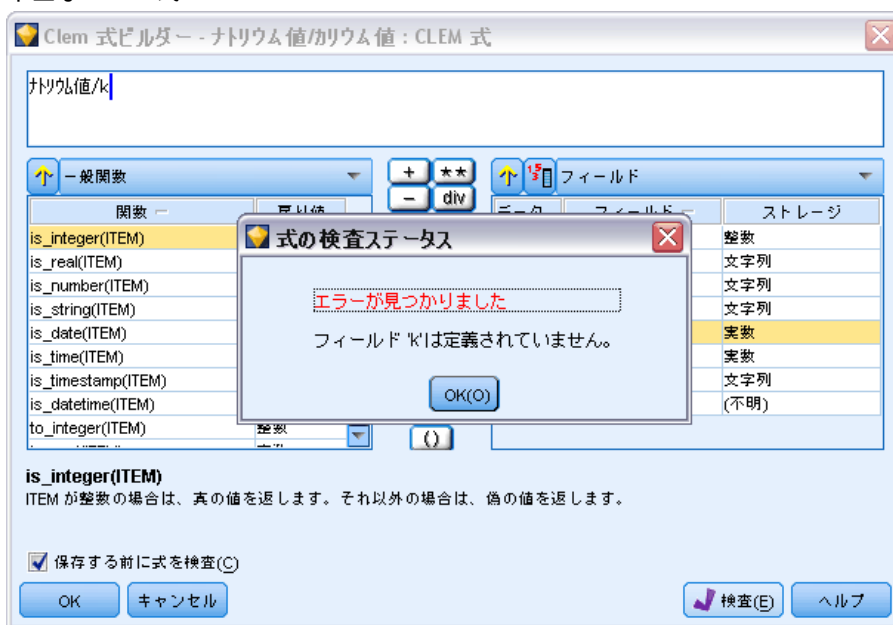


フラグ型と名義型フィールドの場合、すべての定義済みの値が表示されます。連続型（数値範囲型）フィールドの場合、最小値と最大値が表示されます。

CLEM 式の検査

Clem 式ビルダーで（右下隅にある）[検査] をクリックして、CLEM 式が正しいかどうかを検証します。検査されていない式は赤で表示されます。エラーが検出されると、原因を示すメッセージが表示されます。

図 7-11
不正な CLEM 式



検査する項目を次に示します。

- 値やフィールド名の引用符が正しいかどうか
- パラメータやグローバル変数の使用法が正しいかどうか
- 演算子が適切に使用されているかどうか

- 参照されているフィールドが存在しているかどうか
- 参照されているグローバル値が存在しているか、また定義が正しいかどうか

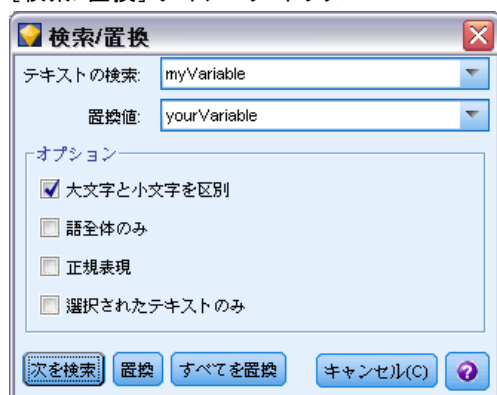
シンタックスにエラーを発見した場合は、手動で式を入力するのではなく、リストや演算子ボタンを使って式を作成してください。この方法では、フィールドや値に適切な引用符が自動的に追加されます。

検索と置換

[検索/置換] ダイアログ ボックスは、スクリプト エディタ、CLEM 式ビルダーなど、スクリプトまたは式のテキストを編集する場合、またはレポート ノードでテンプレートを定義する場合に使用できます。これらの領域のいずれかでテキストを編集する場合、**Ctrl + F** キーを押してダイアログ ボックスにアクセスし、カーソルがテキスト領域にフォーカスしていることを確認します。[フィルタ] ノードを使用している場合、たとえば、[設定] タブのテキスト領域から、または CLEM 式ビルダーのテキスト フィールドからダイアログ ボックスにアクセスできます。

図 7-12

[検索/置換] ダイアログ ボックス



- ▶ テキスト領域内にカーソルを置いて、**Ctrl + F** キーを押して [検索/置換] ダイアログ ボックスにアクセスします。
- ▶ 検索するテキストを入力するか、最近検索した項目のドロップダウン リストから選択します。
- ▶ 置換テキストがある場合は、入力します。
- ▶ [次を検索] をクリックして、検索を開始します。
- ▶ [置換] をクリックして現在の選択内容を置換するか、[すべて置換] をクリックしてすべてまたは選択したインスタンスを更新します。

- ▶ 各操作が終了すると、ダイアログ ボックスが閉じます。テキスト領域で F3 を押すと最後の検索操作が繰り返され、または Ctrl + F キーを押すとダイアログに再度アクセスします。

検索オプション

大文字と小文字を区別： 検索操作で、たとえば myvar が myVar と位置するかどうかなど、大文字と小文字を区別するかどうかを指定します。この設定に関係なく、置換テキストは常に入力したとおりに挿入されます。

語全体のみ： 検索操作が語内に埋め込まれたテキストに一致するかどうかを指定します。このオプションを選択すると、spider に関する検索は、spiderman または spider-man に一致しません。

正規表現： 正規表現のシンタックスを使用するかどうかを指定します（次項参照）。このオプションを選択すると、[語全体のみ] オプションは無効化され、その値は無視されます。

選択されたテキストのみ：[すべて置換] オプションを使用する場合、検索の範囲を制御します。

正規表現シンタックス

正規表現を使用すると、タブまたは改行文字などの特殊文字、a から d までなど文字のクラスまたは範囲、行の開始または終了などの境界について検索することができます。次の種類の表現がサポートされています。

文字の一致

文字	一致
x	文字 x
¥¥	円記号
¥0n	8 進法の値を持つ文字 0n (0 ≤ n ≤ 7)
¥0n	8 進法の値を持つ文字 0nn (0 ≤ n ≤ 7)
¥0mnn	8 進法の値を持つ文字 0mnn (0 ≤ m ≤ 3, 0 ≤ n ≤ 7)
¥xhh	16 進法の値を持つ文字 0xhh
¥uhhhh	16 進法の値を持つ文字 0xhhh
¥t	タブ文字 ('¥u0009')
¥n	改行文字 ('¥u000A')
¥r	復帰文字 ('¥u000D')
¥f	改ページ文字 ('¥u000C')
¥a	アラート (ベル) 文字 ('¥u0007')
¥e	エスケープ文字 ('¥u001B')
¥cx	xに対応する制御文字

文字クラスの一致

文字クラス	一致
[abc]	a、b、または c (単純クラス)
[^abc]	a、b、または c 以外の文字 (減法)
[a-zA-Z]	a から z または A から Z の各文字 (範囲)
[a-d[m-p]]	a から d、または m から p (統合) また、[a-dm-p] と指定することもできます。
[a-z&&[def]]	a から z、および d、e、または f (交差)
[a-z&&[^bc]]	b と c 以外の a から z (減法) また、[ad-z] と指定することもできます。
[a-z&&[^m-p]]	a から z、m から p を除く (減法) また、[a-lq-z] と指定することもできます。

事前設定された文字クラス

事前設定された文字クラス	一致
.	任意の文字 (行末に一致する場合または一致しない場合があります)
¥d	任意の数字 : [0-9]
¥D	数字以外 : [^0-9]
¥s	空白文字 : [¥t¥n¥x0B¥f¥r]
¥S	空白文字以外 : [^¥s]
¥w	語文字 : [a-zA-Z_0-9]
¥W	語文字以外 : [^¥w]

境界の一致

境界の一致	一致
^	行頭
\$	行末
¥b	語の境界
¥B	語以外の境界
¥A	入力の開始
¥Z	最後の行末以外の入力の終了
¥z	入力の終了

CLEM 言語に関するリファレンス

CLEM リファレンス概要

この項では、Control Language for Expression Manipulation (CLEM) について説明していきます。CLEM は、IBM® SPSS® Modeler ストリーム内で使われるデータの分析と操作の非常に役に立つツールです。ノード内で CLEM を使用して、条件の評価や値の新規作成からレポートへのデータ挿入まで、作業を実行できます。 [詳細は、7 章 p.113 CLEM についてを参照してください。](#)

CLEM 言語のサブセットは、また、ユーザー インターフェイスでスクリプトを使用する場合に使用することができます。これによって、同じデータ操作の大部分を自動的に行うことができます。 [詳細は、3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。](#)

CLEM 式は、値、フィールド名、演算子、および関数で構成されます。正しい構文を使って、さまざまなデータ操作を作成することができます。 [詳細は、7 章 p.116 CLEM の例 を参照してください。](#)

CLEM データ型

CLEM データ型は次のいずれかで構成できます。

- 整数
- 実数
- 文字
- 文字列
- リスト
- Fields
- 日付と時刻

引用符の使用規則

IBM® SPSS® Modeler では、CLEM 式で使われるフィールド、値、パラメータ、および文字列などを柔軟に指定することができます。次の規則に従って式を作成することをお勧めします。

- 文字列-文字列を指定する場合は、常に二重引用符を使用してください (例: "Type 2")。単一引用符を使用することもできますが、引用符で囲まれたフィールドと誤解される危険性があります。

- フィールド-スペースや他の特殊文字を入れる必要があるような場合にだけ単一引用符を使用します (例: 'Order Number')。データ セット中に単一引用符で囲まれているのに未定義のフィールドがあると、それは文字列として読み込まれてしまいます。
- パラメーター-パラメータを使用する場合は、常に単一引用符を使用します (例: '\$P-threshold')。
- 文字-常に単一の逆引用符 (') を使用します (例: stripchar('d','drugA'))。

詳細は、7 章 p.118 値とデータ型 を参照してください。これらの規則は、以降の項目で詳細に説明しています。

整数

整数は、10 進数のシーケンスとして表されます。必要に応じて、整数の前にマイナス符号 (-) を付けて負の数を表すことができます。たとえば、1234、999、-77 のように記述します。

CLEM 言語は任意の精度の整数を処理します。整数の最大サイズは使用するプラットフォームによって異なります。値が大きすぎて整数フィールドに表示できない場合は、通常、フィールドのデータ型を Real に変更すると値を正確に表示できます。

実数

実数は浮動小数点数を意味しています。実数は、1 つ以上の数値と、その後続く小数点、その後続く 1 つ以上の数値で表されます。CLEM 実数は倍精度で保持されます。

必要に応じて、実数の前にマイナス符号 (-) を付けて負の数を表すことができます。たとえば、1.234、0.999、-77.001 のように記述します。指数表記で実数を表すには、<数値> e<指数> の形式を使用します。たとえば、1234.0e5、1.7e-2 のように記述します。IBM® SPSS® Modeler アプリケーションがファイルから数値文字列を読み込んで自動的に数値に変換する場合、小数点の前に数字がない数値や小数点の後に数字がない数値も受け取ります。たとえば、999. や .11 などです。ただし、これらの形式は CLEM 式では不正です。

注：CLEM 式の実数を参照する場合、現在のストリームまたはロケールの設定に関わらず、小数点区切り文字としてピリオドを使用する必要があります。たとえば、次のとおりです。

Na > 0.6

次のようには、指定できません。

Na > 0.6

これは、[ストリームのプロパティ] ダイアログ ボックスでカンマが小数点として選択された場合でも適用されます。また、コード構文が特定のロケールまたは表記方法から独立する必要があるという一般的なガイドラインを検討します。

文字

一般的に文字（通常 **CHAR** と表記）は、CLEM 式内で文字列のテストを実行するために用いられます。たとえば、**isuppercode** 関数を使って、文字列の先頭文字が大文字かどうかを判断することができます。文字列の先頭文字に対してテストを行う必要があることを示すために、文字を使用する CLEM 式を次に示します。

```
isuppercode(subscrs(1, "MyString"))
```

CLEM 式中の特定文字のコード（場所ではなく）を表すには、単一逆引用符を `<文字>--` の形式で指定します。たとえば、``A``、``Z`` のように記述します。

注：フィールドに対する **CHAR** ストレージ タイプはありません。そのため、結果が **CHAR** となる式でフィールドが作成または置換された場合、その結果は文字列に変換されます。

文字列

基本的に、文字列は二重引用符で囲んでください。たとえば、文字列は `"c35product2"` や `"referrerID"` となります。文字列内で特別な文字を示すには、`\"$65443"` のように円記号を使用します（円記号を示すには、`\\` のように円記号を 2 つ使用してください）。文字列を単一引用符で囲むこともできますが、その場合引用符で囲まれたフィールド（`'referrerID'`）と区別できない可能性があります。詳細は、[p. 154 文字列関数](#) を参照してください。

リスト

リストは、順序付けられた要素のシーケンスであり、データ型が混在していることもあります。リストは、大カッコ（`[]`）で囲みます。たとえば、リストは `[1 2 4 16]` や `["abc" "def"]` となります。リストは IBM® SPSS® Modeler フィールドの値としては使用されません。リストは、**member** や **oneof** などの関数に引数を渡すために使用します。

Fields

CLEM 式内で、関数名以外の名前はフィールド名とみなされます。これらを **Power**、**val27**、**state_flag** のように記述できますが、または名前が数字から始まる、またはスペースなどアルファベット以外の文字（アンダースコアを

除く)を含む場合、'Power Increase'、'2nd answer'、'#101'、'\$P-NextField' のように名前を単一引用符で囲みます。

注： データ セット中に単一引用符で囲まれているのに未定義のフィールドがあると、それは文字列として読み込まれてしまいます。

日付(D)

日付の計算は、基準日に基づいて行われます。基準日は、[ストリームのプロパティ] ダイアログ ボックスで指定します。デフォルトの基準日は、1900 年 1 月 1 日です。詳細は、[5 章 p.58 ストリームのオプションの設定](#) を参照してください。

CLEM 言語は、次の日付のフォーマットをサポートします：

書式	例
DDMMYY	150163
MMDDYY	011563
YYMMDD	630115
YYYYMMDD	19630115
YYYYDDD	4 桁の西暦とそれに続く 3 桁の 1 月 1 日からの日数。-たとえば 2000032 は、2000 年の 32 番目の日にち、つまり 2000 年 2 月 1 日を表します。
DAY	現在のロケールの曜日です。-たとえば、英語の場合、Monday、Tuesday などです。
MONTH	現在のロケールの月名です。たとえば、英語の場合、January、February などです。
DD/MM/YY	15/01/63
DD/MM/YYYY	15/01/1963
MM/DD/YY	01/15/63
MM/DD/YYYY	01/15/1963
DD-MM-YY	15-01-63
DD-MM-YYYY	15-01-1963
MM-DD-YY	01-15-63
MM-DD-YYYY	01-15-1963
DD.MM.YY	15.01.63
DD.MM.YYYY	15.01.1963
MM.DD.YY	01.15.63
MM.DD.YYYY	01.15.1963
DD-MON-YY	15-JAN-63, 15-jan-63, 15-Jan-63
DD/MON/YY	15/JAN/63, 15/jan/63, 15/Jan/63

書式	例
DD.MON.YY	15.JAN.63, 15.jan.63, 15.Jan.63
DD-MON-YYYY	15-JAN-1963, 15-jan-1963, 15-Jan-1963
DD/MON/YYYY	15/JAN/1963, 15/jan/1963, 15/Jan/1963
DD.MON.YYYY	15.JAN.1963, 15.jan.1963, 15.Jan.1963
MON YYYY	Jan 2004
q Q YYYY	四半期を表す 1 桁の数字 (1-4) とそれに続く文字 Q、そして 4 桁の年です。たとえば2004年12月25日の表記は、4 Q 2004 になります。
www WK YYYY	1 年の内での週を表す 2 桁の数で、次に文字 WK と 4 桁の年が続きます。年内の週は、週の最初の日が月曜日で、また、少なくとも最初の週に 1 日以上あるという仮定の下に計算されます。

Time

CLEM 言語は、次の時間のフォーマットをサポートします：

書式	例
HHMMSS	120112, 010101, 221212
HHMM	1223, 0745, 2207
MMSS	5558, 0100
HH:MM:SS	12:01:12, 01:01:01, 22:12:12
HH:MM	12:23, 07:45, 22:07
MM:SS	55:58, 01:00
(H)H:(M)M:(S)S	12:1:12, 1:1:1, 22:12:12
(H)H:(M)M	12:23, 7:45, 22:7
(M)M:(S)S	55:58, 1:0
HH.MM.SS	12.01.12, 01.01.01, 22.12.12
HH.MM	12.23, 07.45, 22.07
MM.SS	55.58, 01.00
(H)H.(M)M.(S)S	12.1.12, 1.1.1, 22.12.12
(H)H.(M)M	12.23, 7.45, 22.7
(M)M.(S)S	55.58, 1.0

CLEM演算子

次の演算子が利用できます。

操作	コメント	優先順位 (次項参照)
or	2 つのCLEM 式間で使用されます。どちらかが真 (true) の場合、または両方が真 (true) の場合に、値を返します。	10
and	2 つのCLEM 式間で使用されます。両方が真 (true) の場合に、値を返します。	9
=	任意の比較可能な 2 つの項目間で使用されます。ITEM1 が ITEM2 と等しい場合に真が返されます。	7
==	= と同等	7
/=	任意の比較可能な 2 つの項目間で使用されます。ITEM1 \neq ITEM2 と等しくない場合に真が返されます。	7
/==	/= と同等	7
>	任意の比較可能な 2 つの項目間で使用されます。ITEM1 が厳密に ITEM2 より大きい場合に真を返します。	6
>=	任意の比較可能な 2 つの項目間で使用されます。ITEM1 が ITEM2 以上の場合に真が返されます。	6
<	任意の比較可能な 2 つの項目間で使用されます。ITEM1 が厳密に ITEM2 より小さい場合に真を返します。	6
<=	任意の比較可能な 2 つの項目間で使用されます。ITEM1 が ITEM2 以下の場合に真が返されます。	6
&&=_0	2 つの整数間に用いられます。ブール式の $INT1 \ \&\& \ INT2 = 0$ と同じになります。	6
&&/=_0	2 つの整数間に用いられます。ブール式の $INT1 \ \&\& \ INT2 = 0$ と同じになります。	6
+	2 つの数値で加算します。NUM1 + NUM2。	5
><	2 つの文字列を次のように連結します。 STRING1 >< STRING2.	5

操作	コメント	優先順位 (次項参照)
-	1つの数値をもう1つの数値から減算します。NUM1 - NUM2。1つの数値の前にも使用できます。 - NUM。	5
*	2つの数値を乗算するのに使用されます。NUM1 * NUM2。	4
&&	2つの整数間に用いられます。結果は、INT1 と INT2 のビット単位の「論理積」になります。	4
&&~	2つの整数間に用いられます。結果は、INT1 と、INT2 のビット単位の補数との、ビット単位の「論理積」になります。	4
	2つの整数間に用いられます。結果は、INT1 と INT2 のビット単位の「包括的論理和」になります。	4
~~	整数の前に用いられます。整数 INT のビット単位の補数を生成します。	4
&	2つの整数間に用いられます。結果は、INT1 と INT2 のビット単位の「排他的論理和」になります。	4
INT1 << N	2つの整数間に用いられます。N の数だけ位置を左にシフトした INT のビットパターンを生成します。	4
INT1 >> N	2つの整数間に用いられます。N の数だけ位置を右にシフトした INT のビットパターンを生成します。	4
/	1つの数値をもう1つの数値で除算するのに使用されます: NUM1 / NUM2。NUM1 / NUM2。	4
**	2つの数値間に用いられます。BASE ** POWERBASE の POWER 乗を返します。	3
rem	2つの整数間に用いられます。INT1 rem INT2 剰余 INT1 - (INT1 div INT2) * INT2 を返します。	2
div	2つの整数間に用いられます。INT1 div INT2 整数の除算を実行します。	2

演算子の優先順位

優先順位は、複数の 2 項演算子を使ったカッコで囲まれていない式などの、複雑な式の解析方法を決めるものです。例をあげると、次のようになります。

$3+4*5$

* が + の前に解析されることを示すため、 $(3+4)*5$ ではなく $3+(4*5)$ として解析します。CLEM 言語中のすべての演算子には、それに対応した優先順位があります。優先順位が低いほど、その演算子は処理リスト上で重要な意味を持ち、他の演算子よりも先に処理されます。

関数のリファレンス

IBM® SPSS® Modeler でデータを処理するために、次の CLEM 関数を利用できます。これらの関数は、フィールド作成ノードやフラグ設定ノードなど、さまざまなダイアログ ボックスにコードとして入力できます。または、Clem 式ビルダーを利用して、有効な CLEM 式を作成することができます。関数やフィールド名を覚えておく必要はありません。

関数の種類	説明
情報	フィールド値を詳しく調べる場合に用いられます。たとえば、関数 <code>is_string</code> は、データ型が文字列型のすべてのレコードに対して真を返します。
変換	新しいフィールドの作成や、ストレージ タイプの変換に用いられます。たとえば、関数 <code>to_timestamp</code> は選択されているフィールドをタイムスタンプに変換します。
比較	フィールドの値を互いに比較したり、指定した文字列と比較する場合に用いられます。たとえば、 <code><=</code> は、あるフィールドの値がもう 1 つのフィールドの値以下かどうかを比較します。
論理	<code>if</code> 、 <code>then</code> 、 <code>else</code> などの論理演算を行うために用いられます。
Numeric	フィールド値の自然ログ数の算出など、数値計算に用いられます。
三角関数	指定された角度の アークコサインの算出など、三角関数の計算に用いられます。
Probability	学生からの <code>t</code> 分布値が特定値に満たなくなる確率など、さまざまな分布を基準にして確率を返します。
ビット単位	整数をビット パターンとして操作する場合に用いられます。
Random	無作為に項目を選択したり、無作為な 数字を生成するために用いられます。

関数の種類	説明
String	指定した文字を削除する <code>stripchar</code> のように、文字列に関するさまざまな操作を行うために用いられます。
SoundEx	正しいスペルが分からない場合に、特定文字の発音方法についての音声的な仮定を基準にして、文字列を検索するために用いられます。
日付と時刻	日付、時間、タイムスタンプ フィールドに対してさまざまな操作を行うために用いられます。
シーケンス	データ セットのレコード シーケンスの詳細を調べたり、そのシーケンスに基づいた操作を行うために用いられます。
グローバル	グローバル ノードが作成したグローバル値にアクセスするために用いられます。たとえば、 <code>@MEAN</code> は、データ セット全体のフィールドのすべての値の平均を参照するために用いられます。
空白とヌル	アクセス、フラグ設定、およびユーザーが指定した空白やシステム欠損値を埋めるために用いられます。たとえば、 <code>@BLANK(FIELD)</code> は、空白があるレコードに真のフラグを設定するために用いられます。
特殊フィールド	調査対象の特定のフィールドを表すために用いられます。たとえば、 <code>@FIELD</code> は複数のフィールドを作成する場合に用いられます。

関数の表記方法について

このガイドでは、次の規約を関数中のアイテムを参照するために使用します。

表記方法	説明
BOOL	真 (true) または偽 (false) を示すブールまたはフラグ。
NUM、NUM1、 NUM2	任意の数値。
REAL、REAL1、 REAL2	1.234 または -77.01 のような任意の実数
INT、INT1、 INT2	1 または -77 のような任意の実数
CHAR	'A' のような文字コード。
STRING	"referrerID" のような文字列。
リスト	["abc" "def"] のような、アイテムのリスト。
ITEM	Customer または extract_concept のような任意の実数
DATE	start_date のような日付フィールド。ここで、値のフォーマットは DD-MON-YYYY のようになります。
TIME	power_flux のような時刻フィールド。ここで、値のフォーマットは HHMMSS のようになります。

このガイドにある関数の一覧では、関数を最初の列に、結果のタイプ（整数、文字列等）を 2 番目の列に、説明（存在する場合）を 3 番目の列に示しています。たとえば、次に **rem** 関数の説明を示します。

関数	結果	説明
INT1 rem INT2	数値	INT1 を INT2 で除算した剰余を返しますたとえば、 $INT1 - (INT1 \text{ div } INT2) * INT2$ となります。

項目をリストにする方法や、関数内で文字を指定する方法などの使用方法の詳細は、別の場所で説明されています。詳細は、[p. 136 CLEMデータ型](#) を参照してください。

情報関数

情報関数は、特定のフィールドの値に対する洞察を行うために用いられます。通常これらは、フラグ型フィールドの作成に用いられます。たとえば、**@BLANK** 関数を使って、選択したフィールドに対する値が空白のレコードを示すフラグ型フィールドを作成することができます。同様に、**is_string** などのストレージタイプ関数を使って、フィールドのストレージタイプを確認することもできます。

関数	結果	説明
@BLANK(FIELD)	ブール	上流のデータ型ノードまたは入力ノードで設定された空白処理規則（[データ型] タブ）にしたがって、値が空白のレコードに対して真を返します。この関数は、スクリプトから呼び出すことができません。詳細は、 3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
@NULL(ITEM)	ブール	値が未定義のすべてのレコードに対して真を返します。未定義の値はシステムのヌル値で、IBM® SPSS® Modeler では \$null\$ として表されます。この関数は、スクリプトから呼び出すことができません。詳細は、 3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
is_date(ITEM)	ブール	データ型が日付のすべてのレコードに対して真 (true) を返します。
is_datetime(ITEM)	ブール	データ型が日付、時間またはタイムスタンプのすべてのレコードに対して真 (true) を返します。
is_integer(ITEM)	ブール	データ型が整数のすべてのレコードに対して真 (true) を返します。
is_number(ITEM)	ブール	データ型が数値のすべてのレコードに対して真 (true) を返します。
is_real(ITEM)	ブール	データ型が実数のすべてのレコードに対して真 (true) を返します。

関数	結果	説明
is_string(ITEM)	ブール	データ型が文字列のすべてのレコードに対して真 (true) を返します。
is_time(ITEM)	ブール	データ型が時間のすべてのレコードに対して真 (true) を返します。
is_timestamp(ITEM)	ブール	データ型がタイムスタンプのすべてのレコードに対して真 (true) を返します。

変換関数

変換関数により、新規フィールドを作成し、既存のファイルのストレージタイプを変換することができます。たとえば、文字列を結合したり、切り離したりして、新しい文字列を生成できます。文字列を結合するには、演算子 \times を使用します。たとえば、フィールド **Site** の値が "BRAMLEY" である場合、" xx " \times **Site** は " xxBRAMLEY " を返します。 \times の結果は、引数が文字列でない場合でも、常に文字列となります。フィールド **V1** が **3** で **V2** が **5** である場合、 $\text{V1} \times \text{V2}$ は "**35**" を返します (数値ではなく文字列)。

変換の関数と日付や時刻の値のような、入力に特別な型が必要なその他の関数は、[ストリームのオプション] ダイアログ ボックスに指定されている現在のフォーマットに依存します。たとえば、値が Jan 2003、Feb 2003 などの文字列フィールドを日付ストレージへ変換する場合、ストリームのデフォルトの日付フォーマットとして一致する [MON YYYY] を選択します。詳細は、5 章 p. 58 [ストリームのオプションの設定](#) を参照してください。

関数	結果	説明
ITEM1 \times ITEM2	String	2 つのフィールドの値を連結し、結果の文字列を ITEM1ITEM2. の形式で返します。
to_integer(ITEM)	Integer	指定されたフィールドのストレージを整数に変換します。
to_real(ITEM)	Real	指定されたフィールドのストレージを実数に変換します。
to_number(ITEM)	数値	指定されたフィールドのストレージを数値に変換します。
to_string(ITEM)	String	指定されたフィールドのストレージを文字列に変換します。
to_time(ITEM)	Time	指定されたフィールドのストレージを時間に変換します。
to_date(ITEM)	Date	指定されたフィールドのストレージを日付に変換します。
to_timestamp(ITEM)	Timestamp	指定されたフィールドのストレージをタイムスタンプに変換します。

関数	結果	説明
to_datetime(ITEM)	日時	指定されたフィールドのストレージを日付、時間またはタイムスタンプ値に変換します。
datetime_date(ITEM)	Date	数値、文字列またはタイムスタンプの日付値を返します。数値（秒単位）を日付へ変換しなおすことができるのは、この関数だけです。ITEM が文字列の場合は、現在のデータフォーマットで文字列を解析することにより日付を作成します。この関数が正常に機能するためには、[ストリームのプロパティ] ダイアログ ボックスの [日付のフォーマット] に、正しい値が指定されていなければなりません。ITEM が数値の場合は、基準日（または紀元）からの秒数として解釈します。日付の端数は切り捨てられます。ITEM がタイムスタンプの場合は、日付をタイムスタンプの一部として返します。ITEM が日付の場合は、変更せずに返します。

比較関数

比較関数は、フィールドの値を互いに比較したり、指定した文字列と比較する場合に用いられます。たとえば、文字列が等しいかどうかは、`=` を使って確認することができます。次のように、文字列が等しいかどうかを調べます。 `Class = "class 1"`。

数値の比較を目的とする場合、`greater`（より大きい）は正の無限大に近いことを意味し、`lesser`（より小さい）は負の無限大に近いことを意味します。つまり、すべての負の数値は、すべての正の数値より小さいこととなります。

関数	結果	説明
count_equal(ITEM1, LIST)	Integer	フィールドの LIST から ITEM1 と等しい数値を返します。ITEM1 が NULL の場合は、NULL を返します。詳細は、7 章 p.124 複数フィールドの要約を参照してください。
count_greater_than(ITEM1, LIST)	Integer	フィールドの LIST から ITEM1 より大きい数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_less_than(ITEM1, LIST)	Integer	フィールドの LIST から ITEM1 より小さい数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_not_equal(ITEM1, LIST)	Integer	フィールドの LIST から ITEM1 と等しくない数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_nulls(LIST)	Integer	フィールドの LIST から NULL 値の数を返します。
count_non_nulls(LIST)	Integer	フィールドの LIST から NULL 値以外の数を返します。
date_before(DATE1, DATE2)	ブール	日付値の順序の確認に用いられます。DATE1 が DATE2 より前の場合に真 (true) を返します。

関数	結果	説明
first_index(ITEM, LIST)	Integer	フィールドの LIST から ITEMを含む最初のフィールドの索引、または値が見つからないなら 0 を返します。サポート対象は、文字列、整数、実数型のみです。詳細は、7 章 p.126 複数レスポンスデータの処理を参照してください。
first_non_null(LIST)	CLEM 式	提供されるフィールド リストの最初の非ヌル値を返します。ストレージ タイプはすべてサポート対象です。
first_non_null_index(LIST)	Integer	非ヌル値を含む特定の LIST の最初のフィールドの索引、またはすべての値がヌル 0 を返します。ストレージ タイプはすべてサポート対象です。
ITEM1 = ITEM2	ブール	ITEM1 が ITEM2 と等しい場合に真 (true) を返します。
ITEM1 /= ITEM2	ブール	2 つの文字列が異なるか、同じでも 0 の場合に真 (true) を返します。
ITEM1 < ITEM2	ブール	ITEM1 が ITEM2 より小さい場合に真 (true) を返します。
ITEM1 <= ITEM2	ブール	ITEM1 が ITEM2 以下の場合に真 (true) を返します。
ITEM1 > ITEM2	ブール	ITEM1 が ITEM2 より大きい場合に真 (true) を返します。
ITEM1 >= ITEM2	ブール	ITEM1 が ITEM2 以上の場合に真 (true) を返します。
last_index(ITEM, LIST)	Integer	フィールドの LIST から ITEMを含む最新のフィールドの索引、または値が見つからないなら 0 を返します。サポート対象は、文字列、整数、実数型のみです。詳細は、7 章 p.126 複数レスポンスデータの処理を参照してください。
last_non_null(LIST)	CLEM 式	提供されるフィールド リストの最後の非ヌル値を返します。ストレージ タイプはすべてサポート対象です。
last_non_null_index(LIST)	Integer	非ヌル値を含む特定の LIST の最後のフィールドの索引、またはすべての値がヌル 0 を返します。ストレージ タイプはすべてサポート対象です。
max(ITEM1, ITEM2)	CLEM 式	ITEM1 または ITEM2 のどちらか大きい方を返します。
max_index(LIST)	Integer	数値フィールドの LIST から最大値を含むフィールドの索引、またはすべての値がヌルなら 0 を返します。たとえば、3番目にリストされたフィールドに最大値がある場合は、インデックス値 3 を返します。複数のフィールドに最大値がある場合は、最初にリストされたもの (左端) を返します。詳細は、7 章 p.126 複数レスポンスデータの処理を参照してください。
max_n(LIST)	数値	数値フィールドの LIST から最大値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。詳細は、7 章 p.124 複数フィールドの要約を参照してください。

関数	結果	説明
member(ITEM, LIST)	ブール	ITEM が、指定された LIST のメンバーの場合に真 (true) を返します。それ以外の場合は、偽 (false) の値が返されます。また、フィールド名のリストを定義することもできます。詳細は、7 章 p.124 複数フィールドの要約を参照してください。
min(ITEM1, ITEM2)	CLEM 式	ITEM1 または ITEM2 のどちらか小さい方を返します。
min_index(LIST)	Integer	数値フィールドの LIST から最小値を含むフィールドの索引、またはすべての値がヌルなら 0 を返します。たとえば、3番目にリストされたフィールドに最小値がある場合は、インデックス値 3 を返します。複数のフィールドに最小値がある場合は、最初にリストされたもの (左端) を返します。詳細は、7 章 p.126 複数レスポンス データの処理を参照してください。
min_n(LIST)	数値	数値フィールドの LIST から最小値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。
time_before(TIME1, TIME2)	ブール	時間値の順序の確認に用いられます。TIME1 が TIME2 より前の場合に真 (true) が返されます。
value_at(INT, LIST)		オフセットが有効値の範囲外の場合 (つまり 1 以上またはリストされたフィールドの数値以下)、オフセット NT または NULL でそれぞれリストされたフィールドの値を返します。ストレージタイプはすべてサポート対象です。

論理関数

CLEM 式を使って論理演算を行うことができます。

関数	結果	説明
COND1 and COND2	ブール	この演算は論理積で、COND1 と COND2 の両方が真 (true) の場合に真 (true) の値を返します。COND1 が偽 (false) の場合、COND2 は評価されません。こうすることによって、COND2 の演算が正しいかどうかを COND1 で判断することができます。たとえば、length(Label) >=6 や Label(6) = 'x' となります。
COND1 or COND2	ブール	この演算は、(包括的) 論理和で、COND1 と COND2 のどちらかが真 (true) の場合、または両方とも真 (true) の場合に、真 (true) の値を返します。COND1 が真 (true) の場合、COND2 は評価されません。
not(COND)	ブール	この演算は論理否定で、COND が偽 (false) の場合に真 (true) の値を返します。それ以外の場合、この演算は 0 の値を返します。

関数	結果	説明
if COND then EXPR1 else EXPR2 endif	CLEM 式	この演算は条件評価です。COND が真 (true) の場合、この演算は EXPR1 の結果を返します。それ以外の場合は、EXPR2 を評価した結果を返します。
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	CLEM 式	この演算は複数の分岐を持つ条件評価です。COND1 が真 (true) の場合、この演算は EXPR1 の結果を返します。それ以外の場合で、COND2 が真ならば、この演算は EXPR2 を評価した結果を返します。それ以外の場合は、EXPR_N を評価した結果を返します。

数値関数

CLEM には、一般的に使われるさまざまな数値関数が用意されています。

関数	結果	説明
-NUM	数値	NUM を否定する場合に用いられます。対応する数値の符号を逆にした値を返します。
NUM1 + NUM2	数値	NUM1 と NUM2 を合計した値を返します。
code -NUM2	数値	NUM1 から NUM2 を減算した値を返します。
NUM1 * NUM2	数値	NUM1 を NUM2 で乗算した値を返します。
NUM1 / NUM2	数値	NUM1 を NUM2 で除算した値を返します。
INT1 div INT2	数値	整数の除算を行うために用いられます。INT1 を INT2 で除算された値を返します。
INT1 rem INT2	数値	INT1 を INT2 で除算した剰余を返しますたとえば、 $INT1 - (INT1 \text{ div } INT2) * INT2$ となります。
INT1 mod INT2	数値	この関数は廃止されました。代わりに rem 関数を使用します。
BASE ** POWER	数値	POWER 乗までべき乗した BASE を返します。BASE と POWER はどちらも任意の数値です (ただし、POWER が整数の 0 以外のいずれかのデータ型のゼロの場合、BASE はゼロ以外である必要があります)。POWER が整数の場合は、BASE のべき乗を 順次掛けていくことによって計算されます。したがって、BASE が整数の場合、結果は整数になります。POWER が整数の 0 の場合、結果は常に BASE と同じデータ型の 1 になります。POWER が整数ではない場合、結果は $\exp(\text{POWER} * \log(\text{BASE}))$ のように計算されます。
abs(NUM)	数値	NUM の絶対値を返します。この値は常に、同じデータ型の数値になります。
exp(NUM)	Real	NUM 乗までべき乗した e を返します。この e は自然対数の底です。
fracof(NUM)	Real	NUM-intof(NUM) として定義される NUM の小数部を返します。
intof(NUM)	Integer	引数を切り捨てて整数にします。NUM と同じ符号で、 abs(INT) <= abs(NUM) のような最大の絶対値を持つ整数を返します。

関数	結果	説明
log(NUM)	Real	NUM の自然対数 (底 e) を返します。NUM は、ゼロ以外でなければなりません。
log10(NUM)	Real	NUM の常用対数を返します。NUM は、ゼロ以外でなければなりません。この確率は $\log(\text{NUM})/\log(10)$ のように定義されます。
negate(NUM)	数値	NUM を否定する場合に用いられます。対応する数値の符号を逆にした値を返します。
round(NUM)	Integer	NUM 外政の数の場合は <code>intof(NUM+0.5)</code> または NUM が負の数の場合は <code>intof(NUM-0.5)</code> を指定して NUM を整数に丸めます。
sign(NUM)	数値	NUM の符号を判断するために用いられます。NUM が整数の場合、この演算は -1, 0, または 1 を返します。NUM が実数の場合、NUM が負、0、または正の値かによって、-1.0、0.0、または 1.0 を返します。
sqrt(NUM)	Real	NUM の平方根を返します。NUM は正でなければなりません。
sum_n(LIST)	数値	数値フィールドの LIST から合計値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。詳細は、7 章 p.124 複数フィールドの要約を参照してください。
mean_n(LIST)	数値	数値フィールドの LIST から平均値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。
sdev_n(LIST)	数値	数値フィールドの LIST から標準偏差を返します。フィールドのすべての値が NULL の場合は、NULL を返します。

三角関数

この項の関数は、すべて引数として角度を取るかまたは、結果として角度を返します。どちらの場合も、角度の単位 (ラジアンまたは度) は関連するストリーム オプションの設定によって制御されます。

関数	結果	説明
arccos(NUM)	Real	指定した角度のアーコサインを計算します。
arccosh(NUM)	Real	指定した角度の双曲線アーコサインを計算します。
arcsin(NUM)	Real	指定した角度のアークサインを計算します。
arcsinh(NUM)	Real	指定した角度の双曲線アークサインを計算します。
arctan(NUM)	Real	指定した角度のアーктanジェントを計算します。
arctan2(NUM_Y, NUM_X)	Real	NUM_Y / NUM_X のアーктanジェントを計算し、2つの数値の符号を使用して、象限情報を派生させます。結果は、 $-\pi < \text{ANGLE} \leq \pi$ (radians) - $180 < \text{ANGLE} \leq 180$ (degrees) の範囲内の実数になります。
arctanh(NUM)	Real	指定した角度の双曲線アーктanジェントを計算します。

関数	結果	説明
cos(NUM)	Real	指定した角度のコサインを計算します。
cosh(NUM)	Real	指定した角度の双曲線コサインを計算します。
pi	Real	この定数は、パイに最も近い値の実数です。
sin(NUM)	Real	指定した角度のサインを計算します。
sinh(NUM)	Real	指定した角度の双曲線サインを計算します。
tan(NUM)	Real	指定した角度のタンジェントを計算します。
tanh(NUM)	Real	指定した角度の双曲線タンジェントを計算します。

確率関数

確率分布で、学生からの t 分布値が特定値に満たなくなる確率など、さまざまな分布を基準にして確立が返されます。

関数	結果	説明
cdf_chisq(NUM, DF)	Real	指定した自由度のカイ 2 乗分布からの値が特定の数字より小さくなる確率を返します。
cdf_f(NUM, DF1, DF2)	Real	DF1 と DF2 の自由度の F 分布からの値が指定した数字より小さくなる確率を返します。
cdf_normal(NUM, MEAN, STDDEV)	Real	指定した平均と標準偏差の正規分布からの値が指定した数字より小さくなる確率を返します。
cdf_t(NUM, DF)	Real	指定した自由度の t 分布からの値が特定の数字より小さくなる確率を返します。

ビット単位の整数演算

これらの演算を使用すると、2 の補数値を表すビット パターンとして整数を操作できます。この場合、ビット位置 N は $2^{**}N$ の重みを持ちます。ビットは 0 から上方向に番号が付けられます。これらの演算は、整数の符号ビットが左方向に無限に拡張されているかのように処理します。つまり、最上位ビットを超えたすべての位置で、正の整数は 0 のビットを持ち、負の整数は 1 のビットを持ちます。

注：Bitwise 関数はスクリプトから呼び出せません。詳細は、3章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。

関数	結果	説明
~~ INT1	Integer	INT1 のビット単位の補数を生成します。つまり、INT1 の 0 がある各ビットが 1 になります。~~ INT = -(INT + 1) は常に true となります。この関数は、スクリプトから呼び出すことができません。詳細は、3章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
INT1 INT2	整数	この演算の結果は、INT1 と INT2 のビット単位の「包括的論理和」またはです。つまり、INT1 と INT2 のどちらかまたは両方に 1 がある各ビットが 1 になります。
INT1 & INT2	整数	この演算の結果は、INT1 と INT2 のビット単位の「排他的論理和」またはです。つまり、INT1 と INT2 のどちらかにだけ（両方ではない）1 がある各ビットが 1 になります。
INT1 && INT2	整数	INT1 と INT2 のビット単位の「論理積」を生成します。つまり、INT1 と INT2 の両方に 1 がある各ビットが 1 になります。
INT1 &&~~ INT2	整数	この演算の結果は、INT1 と、INT2 のビット単位の補数との、ビット単位の「論理積」です。つまり、INT1 のあるビット位置の値が 1 で、INT2 の同じビット位置の値が 0 の場合に、結果は 1 になります。これは INT1&&(~~INT2) と同じで、INT2 の INT1 セットのビットを消去する場合に役に立ちます。
INT << N	整数	N の数だけ位置を左にシフトした INT1 のビットパターンを生成します。N の値が負の場合は、右にシフトします。
INT >> N	整数	N の数だけ位置を右にシフトした INT1 のビットパターンを生成します。N の値が負の場合は、左にシフトします。
INT1 &&=_0 INT2	ブール	ブール式 INT1 && INT2 /== 0 と同じですが、より効率的です。
INT1 &&/_0 INT2	ブール	ブール式 INT1 && INT2 == 0 と同じですが、より効率的です。
integer_bitcount(INT)	整数	INT の 2 の補数 による結果の、1 または 0 のビットの数をカウントします。INT が負ではない場合、N は 1 のビット数になります。INT が負の場合、N は 0 のビット数になります。符号の拡張のため、負ではない整数には無限大の数の 0 のビットがあります。また、負の整数には無限大の数の 1 のビットがあります。integer_bitcount(INT) = integer_bitcount(-(INT+1)) は常に true となります。

関数	結果	説明
integer_leastbit(INT)	整数	整数 INT の最下位ビット セットのビット位置 N を返します。N は、最も大きな 2 のべき乗です。N によって INT が正確に割られます。
integer_length(INT)	整数	INT の長さのビット数を、2 の補数の整数として返します。つまり、次のように、N は $INT < (1 \ll N)$ if $INT \geq 0$ $INT \geq (-1 \ll N)$ if $INT < 0$ のような最も小さい整数です。INT が負ではない場合、符号なしの整数として INT を表すには、少なくとも N ビットのフィールドが必要です。または、INT の符号にかかわらず、INT を符号付きの整数として表すには、少なくとも N+1 ビットが必要です。
testbit(INT, N)	ブール	整数 INT の N の位置にあるビットを検定し、ビット N の状態をブール値として返します。このブール値は、1 の場合は真 (true)、0 の場合は偽 (false) になります。

乱数関数

次の関数は、無作為に項目を選択したり、乱数を生成する場合に用いられます。

関数	結果	説明
oneof(LIST)	CLEM 式	無作為 (ランダム) に選択された LIST の要素を返します。LIST 項目は [ITEM1,ITEM2,...,ITEM_N] のように入力する必要があります: フィールド名のリストも定義できることに注意してください。詳細は、7 章 p.124 複数フィールドの要約を参照してください。
random(NUM)	数値	1~NUM の範囲の、同じデータ型 (INT または REAL) の一様に分布した乱数を返します。整数を使用する場合、整数だけが返されます。実数 (10 進数) を使用する場合は、実数値が返されます (精度はストリーム オプションによって決まります)。この関数で返される可能性がある最大の乱数は、NUM になります。
random0(NUM)	数値	random(NUM) と同じ性質を持ちますが、値の範囲が 0 から始まります。関数により返される可能性がある最大の乱数値が X と等しくなることはありません。

文字列関数

CLEM では、文字列に対して次の操作を行うことができます。

- 文字列の比較
- 文字列の生成
- 文字へのアクセス

CLEM で文字列とは、"string quotes" のように、二重引用符で囲まれた文字の
ことを表しています。任意の単一の英数字が、文字 (CHAR) になります。
これらは、`z`、`A`、または `2` のように `<文字>` の形式で、単一後方引用
符を使用して CLEM 式内で宣言します。範囲外の文字、または文字列に対
する逆索引文字は、未定義の動作が生じます。

注：SQL プッシュバックを使用する文字列と使用しない文字列と比較する
と、接尾空白を含むさまざまな結果を生成する場合があります。

関数	結果	説明
allbutfirst(N, STRING)	String	STRING の先頭 N 文字、0 削除した文字列を返します。
allbutlast(N, STRING)	String	STRING の最後の文字を削除した文字列を返します。
alphabefore(STRING1, STRING2)	ブール	文字列のアルファベット順を確認するために用いられます。STRING1 が STRING2 より前にある場合に真を返します。
endstring(LENGTH, STRING)	String	指定した文字列から最後の N 文字を抽出します。文字列の長さが指定した長さよりも短い、またはそれに等しい場合は、変更されません。
hasendstring(STRING, SUBSTRING)	整数	この関数は、isendstring(SUBSTRING, STRING) と同じです。
hasmidstring(STRING, SUBSTRING)	整数	この関数は、ismidstring(SUBSTRING, STRING) (埋め込みサブ文字列) と同じです。
hasstartstring(STRING, SUBSTRING)	整数	この関数は、isstartstring(SUBSTRING, STRING) と同じです。
hassubstring(STRING, N, SUBSTRING)	整数	この関数は、issubstring(SUBSTRING, N, STRING) と同じです。N のデフォルトは 1 です。
count_substring(STRING, SUBSTRING)	整数	指定したサブ文字列が文字列内に発生する回数を返します。例をあげると、次のようになります。 count_substring("foooo.txt", "oo") は 3 を返します。
hassubstring(STRING, SUBSTRING)	整数	この関数は、issubstring(SUBSTRING, 1, STRING) と同じです。N のデフォルトは 1 です。
isalphacode(CHAR)	ブール	CHAR が、文字コードが文字である指定された文字列 (通常フィールド名) 中の文字の場合に真を返します。それ以外の場合、この関数は 0 の値を返します。たとえば、isalphacode(produce_num(1)) のようになります。

関数	結果	説明
isendstring(SUBSTRING, STRING)	整数	文字列 STRING がサブ文字列 SUBSTRING で終わる場合、この関数は、STRING 内の SUBSTRING の整数の添字を返します。それ以外の場合、この関数は 0 の値を返します。
islowercode(CHAR)	ブール	CHAR が指定された文字列（通常フィールド名）の小文字の場合に、真（true）の値を返します。それ以外の場合、この関数は 0 の値を返します。たとえば、islowercode(“)や islowercode(country_name(2)) などが有効な式になります。
ismidstring(SUBSTRING, STRING)	整数	SUBSTRING が STRING の部分文字列で、STRING の初めの文字から始まっていないか、または最後の文字で終わっていない場合、この関数は部分文字列が始まる位置の添字を返します。それ以外の場合、この関数は 0 の値を返します。
isnumbercode(CHAR)	ブール	指定された文字列（通常フィールド名）の CHAR が、文字コードが数字である文字の場合に真を返します。それ以外の場合、この関数は 0 の値を返します。たとえば、isnumbercode(product_id(2)) のようになります。
isstartstring(SUBSTRING, STRING)	整数	文字列 STRING がサブ文字列 SUBSTRING から始まる場合、この関数は添字 1 を返します。そうでない場合は、この関数は 0 の値を返します。
issubstring(SUBSTRING, N, STRING)	整数	この関数は、文字列 STRING の N 番目の文字から始めて、文字列 SUBSTRING と等しいサブ文字列を検索します。文字列が見つかった場合、一致する部分文字列が始まる位置の添字 M（整数）を返します。それ以外の場合、この関数は 0 の値を返します。N が与えられていない場合、この関数はデフォルトで 1 になります。
issubstring(SUBSTRING, STRING)	整数	この関数は、文字列 STRING の N 番目の文字から始めて、文字列 SUBSTRING と等しいサブ文字列を検索します。文字列が見つかった場合、一致する部分文字列が始まる位置の添字 M（整数）を返します。それ以外の場合、この関数は 0 の値を返します。N が与えられていない場合、この関数はデフォルトで 1 になります。
issubstring_count(SUBSTRING, N, STRING):	整数	指定した STRING 内で N 番目に発生した SUBSTRING のインデックスを返します。N 番目に発生する SUBSTRING よりも少ない場合、0 を返します。

関数	結果	説明
<code>issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)</code>	整数	この関数は <code>issubstring</code> と同じですが、添字 <code>STARTLIM</code> から、またはその前から始まり、添字 <code>ENDLIM</code> で、またはその前で終わるように、マッチングが制限されます。 <code>STARTLIM</code> 制約または <code>ENDLIM</code> 制約は、どちらかの引数に偽 (<code>false</code>) の値を指定することによって無効にできます。たとえば、 <code>issubstring_lim(SUBSTRING, N, false, false, STRING)</code> は <code>issubstring</code> と同じです。
<code>isuppercode(CHAR)</code>	ブール	この関数は <code>CHAR</code> が大文字の場合に、真 (<code>true</code>) の値を返します。それ以外の場合、この関数は 0 の値を返します。たとえば、 <code>isuppercode(")</code> や <code>isuppercode(country_name(2))</code> などが有効な式になります。
<code>last(CHAR)</code>	String	この関数は、 <code>STRING</code> の最後の文字 <code>CHAR</code> を返します (少なくとも 1 文字以上の長さがなければなりません)。
<code>length(STRING)</code>	整数	文字列 <code>STRING</code> の長さ (つまり文字列内の半角文字数) を返します。
<code>locchar(CHAR, N, STRING)</code>	整数	シンボル値フィールド中の文字の位置を識別するために用いられます。この関数は、文字列 <code>STRING</code> 中の <code>N</code> 番目の文字から、文字 <code>CHAR</code> の検索を開始します。この関数は、文字が見つかった (<code>N</code> から始まる) 位置を示す値を返します。文字が見つからない場合は 0 を返します。関数のオフセット (<code>N</code>) が無効な場合 (たとえばオフセットが文字列の長さを超えているなど)、この関数は <code>\$null\$</code> を返します。 たとえば、 <code>locchar('n', 2, web_page)</code> と指定すると、フィールド <code>web_page</code> 中の 2 番目の文字から、文字 <code>'n'</code> を検索します。 注: 指定する文字を、忘れずに、単一逆引用符で囲むようにしてください。
<code>locchar_back(CHAR, N, STRING)</code>	整数	<code>locchar</code> に似ていますが、 <code>N</code> 番目の文字から前方向に検索される点が異なります。たとえば、 <code>locchar_back('n', 9, web_page)</code> と指定すると、フィールド <code>web_page</code> の 9 番目の文字から、文字列の先頭方向に向かって検索が開始されます。関数のオフセットが無効な場合 (たとえばオフセットが文字列の長さを超えているなど)、この関数は <code>\$null\$</code> を返します。できる限り、 <code>locchar_back</code> とともに関数 <code>length(<field>)</code> を使用して、フィールドの現在の値の長さを動的に使用することをお勧めします。たとえ

関数	結果	説明
		ば、 <code>locchar_back('n', (length(web_page)), web_page)</code> となります。
<code>lowertoupper(CHAR)</code> <code>lowertoupper(String)</code>	CHAR または文字列	文字または文字列を入力にすることができ、同じデータ型の新しい項目を返すために用いられます。その際小文字はすべて同じ文字の 大文字に変換されます。たとえば、 <code>lowertoupper('a')</code> 、 <code>lowertoupper("My string")</code> 、および <code>lowertoupper(field_name(2))</code> となります。
<code>matches</code>	ブール	文字列が指定したパターンに一致する場合、真を返します。パターンは文字列リテラルにする必要があり、パターンを含むフィールド名にしてはなりません。クエスチョン マーク (?) をパターンに含めて正確に 1 つの文字に一致させることができます。アスタリスク (*) は 0 かそれ以上の文字数に一致します。リテラル クエスチョン マークまたはアスタリスクを (むしろ、ワイルドカードとして使用しないで) 一致させるために、バックslash をエスケープ文字として使用することができます。
<code>replace(SUBSTRING, NEWSUBSTRING, STRING)</code>	String	指定した String 内で、SUBSTRING のすべてのインスタンスを NEWSUBSTRING を使って置き換えます。
<code>replicate(COUNT, STRING)</code>	String	指定した回数だけコピーされた元の文字列を含む文字列を返します。
<code>stripchar(CHAR,String)</code>	String	文字列またはフィールドから、指定した文字を削除します。この関数を利用すれば、データから通貨表記などの余分な記号を削除して、単純な数字または名前を取得できます。たとえば、シンタックス <code>stripchar('\$','Cost')</code> を使用すると、すべての値からドル記号を除去した新しいフィールドが返されます。 注：指定する文字を、忘れずに、単一逆引用符で囲むようにしてください。
<code>skipchar(CHAR, N, STRING)</code>	整数	文字列 STRING の N 文字目から、CHAR 以外の文字を検索します。この関数は、見つかった文字の位置を示す整数サブ文字列を返します。N 番目以降のすべての文字が CHAR の場合は、0 を返します。関数のオフセットが無効な場合 (たとえばオフセットが文字列の長さを超えているなど)、この関数は <code>\$null\$</code> を返します。 <code>locchar</code> は、よく関数 <code>skipchar</code> と一緒に、N (文字列の検索開始点) の値を判断するために用いられます。たとえば、 <code>skipchar('s', (locchar('s', 1, "MyString")), "MyString")</code> となります。

関数	結果	説明
skipchar_back(Char, N, STRING)	整数	skipchar に似ていますが、N 番目の文字から前に戻る方向に検索される点が異なります。
startstring(LENGTH, STRING)	String	指定した文字列から最初の N 文字を抽出します。文字列の長さが指定した長さよりも短い、またはそれに等しい場合は、変更されません。
strmember(Char, STRING)	整数	locchar(Char, 1, STRING) への接続 Char が最初に発生する、または 0 の地点を示す整数の部分文字列を返します。関数に無効なオフセットがある場合（たとえば、オフセットが文字列の長さを超えている）、この関数は \$null\$ を返します。
subscrs(N, STRING)	Char	入力文字列 STRING の N 番目の文字 Char を返します。この関数は、STRING(N) という短い形式で記述することもできます。たとえば、lowertoupper("name"(1)) が有効な式となります。
substring(N, LEN, STRING)	String	文字列 SUBSTRING を返します。この文字列は、文字列 STRING の添字 N から LEN 文字分の文字列で構成されています。
substring_between(N1, N2, STRING)	String	添字 N1 から始まり、添字 N2 で終わる STRING のサブ文字列を返します。
trim(STRING)	String	指定した文字列から、文字列の前後の空白文字を削除します。
trim_start(STRING)	String	指定した文字列から、文字列の前の空白文字を削除します。
trimend(STRING)	String	指定した文字列から、文字列の後の空白文字を削除します。
unicode_char(NUM)	Char	NUM の Unicode 値を返します。
unicode_value(Char)	NUM	Char の Unicode 値を返します。
uppertolower(Char) uppertolower (STRING)	Char または文字列	文字または文字列を入力にすることができ、同じデータ型の新しい項目を返すために用いられます。その際、大文字はすべて同じ文字の小文字に変換されます。 注：文字列は二重引用符で、文字は単一の逆引用符で忘れずに指定するようにしてください。単純なフィールド名の場合は、引用符は使用しません。

SoundEx 関数

SoundEx は、サウンドは分かっているにもかかわらず正しいスペルが分からない場合に、文字列を検索するために用いられる方法です。1918年に開発されたこの方法では、特定文字の発音方法についての音声的な仮定を基準にする、似通ったサウンドの単語が検索されます。この方法は、たとえば、似通った名前のスペルや発音がさまざまに異なる場合に、データベースで名前検索を行うために使用されます。基本的な SoundEx アルゴリズムはさまざまな文献で引用されており、また、(ph や f のように、文字列の前の文字の組み合わせが、同じサウンドを持つにもかかわらず一致しないことなど) 制約があることは知られていますが、ほとんどのデータベースで何らかの形でサポートされています。

関数	結果	説明
soundex(String)	整数	指定した String の 4 文字の SoundEx コードを返します。
soundex_difference(String1, String2)	整数	2 つの文字列で SoundEx エンコードが同じ文字数を示す 0 から 4 の整数を返します。ここで、0 は類似性がないこと、また、4 は強い類似性があること、または同じ文字列であることを示します。

日付および時刻の関数

CLEM には、日付や時間を表す文字列変数の日付と時間ストレージのフィールドを操作する関数が用意されています。使用する日付と時間の書式は、ストリームごとに異なり、[ストリームのプロパティ] ダイアログ ボックスで指定します。日付と時間の関数は、現在選択されているフォーマットに従って、日付と時間の文字列を解析します。

日付に 2 桁だけを使用する (世紀を指定しない) 年を指定すると、IBM® SPSS® Modeler では、[ストリームのプロパティ] ダイアログ ボックスで指定されているデフォルトの世紀が使用されます。

注：日付と時間の関数はスクリプトから呼び出せません。詳細は、3章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。

関数	結果	説明
@TODAY	String	[ストリームのプロパティ] ダイアログ ボックスで [日/分をロールオーバー] を選択している場合、この関数は現在の日付形式を使用して、現在の日付を文字列として返します。2桁の日付形式を使用しており、[日/分をロールオーバー] を選択していない場合は、現在のサーバーの \$null\$ を返します。この関数は、スクリプトから呼び出すことができません。詳細は、3章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
to_time(ITEM)	Time	指定されたフィールドのストレージを時間に変換します。
to_date(ITEM)	Date	指定されたフィールドのストレージを日付に変換します。
to_timestamp(ITEM)	Timestamp	指定されたフィールドのストレージをタイムスタンプに変換します。
to_datetime(ITEM)	日時	指定されたフィールドのストレージを日付、時間またはタイムスタンプ値に変換します。
datetime_date(ITEM)	Date	数値、文字列またはタイムスタンプの日付値を返します数値（秒単位）の日付をへ変換しなおすことができるのは、この関数だけです。ITEM が文字列の場合は、現在のデータ フォーマットで文字列を解析することにより日付を作成します。この関数が正常に機能するためには、[ストリームのプロパティ] ダイアログ ボックスの [日付のフォーマット] に、正しい値が指定されていなければなりません。ITEM が数値の場合は、基準日（または紀元）からの秒数として解釈します。日付の端数は切り捨てられます。ITEM がタイムスタンプの場合は、日付をタイムスタンプの一部として返します。ITEM が日付の場合は、変更せずに返します。
date_before(DATE1, DATE2)	ブール	DATE1 が DATE2 より前の日付またはタイムスタンプの場合に true を返します。それ以外の場合、この関数は 0 の値を返します。
date_days_difference(DATE1, DATE2)	整数	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの日数を整数で返します。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。

関数	結果	説明
date_in_days(DATE)	整数	基準日から DATE で表される日付またはタイムスタンプまでの日数を整数で返します。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。たとえば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_in_months(DATE)	Real	基準日から DATE で表される日付またはタイムスタンプまでの月数を実数で返します。これは、1 か月を 30.4375 日と仮定した近似値になります。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。たとえば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_in_weeks(DATE)	Real	基準日から DATE で表される日付またはタイムスタンプまでの週数を実数で返します。ここでは、1 週間を 7.0 日と仮定しています。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。たとえば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_in_years(DATE)	Real	基準日から DATE で表される日付またはタイムスタンプまでの年数を実数で返します。これは、1 年 365.25 日を基準とした近似値になります。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。たとえば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_months_difference (DATE1, DATE2)	Real	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの月数を実数で返します。これは、1 か月を 30.4375 日と仮定した近似値になります。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。
datetime_date(YEAR, MONTH, DAY)	Date	YEAR、MONTH、および DAY の日付値を作成します。引数は整数でなければなりません。
datetime_day(DATE)	整数	指定されたDATE またはタイムスタンプから、日付を返します。結果は 1～31 の範囲の整数になります。
datetime_day_name(DAY)	String	指定された DAY のフルネームを返します。引数は、1 (日曜)～7 (土曜) の範囲の整数でなければなりません。
datetime_hour(TIME)	整数	TIME またはタイムスタンプから時間を返します。結果は 0～23 の範囲の整数になります。

関数	結果	説明
datetime_in_seconds(TIME)	Real	TIME に保存された秒の部分を返します。
datetime_in_seconds(DATE), datetime_in_seconds(DATETIME)	Real	現在の DATE または DATETIME と基準日（1900-01-01）からの差から集計した数値を秒に変換して返します。
datetime_minute(TIME)	整数	TIME またはタイムスタンプから分を返します。結果は 0～59 の範囲の整数になります。
datetime_month(DATE)	整数	DATE またはタイムスタンプから月を返します。結果は 1～12 の範囲の整数になります。
datetime_month_name (MONTH)	String	指定された DAY のフルネームを返します。引数は、1～12 の範囲の整数でなければなりません。
datetime_now	Timestamp	現在の時刻をタイムスタンプとして返します。
datetime_second(TIME)	整数	TIME またはタイムスタンプから秒を返します。結果は 0～59 の範囲の整数になります。
datetime_day_short_name(DAY)	String	DAY の名前を省略形で返します。引数は、1（日曜）～7（土曜）の範囲の整数でなければなりません。
datetime_month_short_name(MONTH)	String	MONTH の名前を省略形で返します。引数は、1～12 の範囲の整数でなければなりません。
datetime_time(HOUR, MINUTE, SECOND)	Time	指定された HOUR、MINUTE、および SECOND の時間値を返します。引数は整数でなければなりません。
datetime_time(ITEM)	Time	ITEM の時間値を返します。
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	Timestamp	与えられた YEAR、MONTH、DAY、HOUR、MINUTE、および SECOND のタイムスタンプ値を返します。
datetime_timestamp(DATE, TIME)	Timestamp	DATE および TIME のタイムスタンプ値を返します。
datetime_timestamp (NUMBER)	Timestamp	与えられた秒数のタイムスタンプ値を返します。
datetime_weekday(DATE)	整数	指定されたDATE またはタイムスタンプから、曜日を返します。
datetime_year(DATE)	整数	DATE またはタイムスタンプから年を返します。結果は 2002 のような整数になります。
date_weeks_difference (DATE1, DATE2)	Real	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの週数を実数で返します。ここでは、1 週間を 7.0 日と仮定しています。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。
date_years_difference (DATE1, DATE2)	Real	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの年数を実数で返します。これは、1 年 365.25 日を基準とした近似値になります。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。

関数	結果	説明
time_before(TIME1, TIME2)	ブール	TIME1 が TIME2 より前の時間またはタイムスタンプの場合に真を返します。それ以外の場合、この関数は 0 の値を返します。
time_hours_difference (TIME1, TIME2)	Real	時間またはタイムスタンプ TIME1 と TIME2 間の時間差 (時間) を実数で返します。[ストリームのプロパティ] ダイアログ ボックスで [日/分をロールオーバー] を選択している場合、TIME1 の値の方が大きいと、その値は前の日付を参照します。ロールオーバー オプションをオンにしていない場合、TIME1 の値の方が大きいと、返される値は負になります。
time_in_hours(TIME)	Real	TIME で表される時間を実数で返します。たとえば、時間のフォーマット HHMM では、式 time_in_hours('0130') は 1.5 に評価されます。TIME は時間またはタイムスタンプを示します。
time_in_mins(TIME)	Real	TIME で表される分を実数で返します。TIME は時間またはタイムスタンプを示します。
time_in_secs(TIME)	整数	TIME で表される秒を整数で返します。TIME は時間またはタイムスタンプを示します。
time_mins_difference(TIME1, TIME2)	Real	時間またはタイムスタンプ TIME1 と TIME2 間の時間差 (分) を実数で返します。[ストリームのプロパティ] ダイアログ ボックスで [日/分をロールオーバー] を選択している場合、TIME1 の値の方が大きいと、その値は前の日 (または、現在のフォーマットで分と秒だけが指定されている場合は前の時間) を参照します。ロールオーバー オプションをオンにしていない場合、TIME1 の値の方が大きいと、返される値は負になります。
time_secs_difference(TIME1, TIME2)	整数	TIME1 と TIME2 の時間またはタイムスタンプの差異を秒で返します。[ストリームのプロパティ] ダイアログ ボックスで [日/分をロールオーバー] を選択している場合、TIME1 の値の方が大きいと、その値は前の日 (または、現在のフォーマットで分と秒だけが指定されている場合は前の時間) を参照します。ロールオーバー オプションをオンにしていない場合、TIME1 の値の方が大きいと、返される値は負になります。

日付と時刻の値の変換:

変換関数および日付や時刻の値のような、入力に特別な型が必要なその他の関数は、[ストリームのオプション] ダイアログ ボックスに指定されている現在のフォーマットに依存します。たとえば、DATE という名前のフィールド名がある場合は、Jan 2003、Feb 2003 などの値で文字列として保存されており、次のように日付ストレージへ変換できます。

to_date(DATE)

この返還を行うには、ストリームのデフォルト日付フォーマットとして、一致する日付フォーマット `MON YYYY` を選択します。詳細は、5章 p.58 [ストリームのオプションの設定](#) を参照してください。

フィルター ノードを使用して文字列値を日付へ変換する例については、streams サブフォルダ内の `¥Demos` フォルダにインストールされている、ストリーム `broadband_create_models.str` を参照してください。詳細は、15章 [時系列ノードによる予測 in IBM SPSS Modeler 14.2 アプリケーション ガイド](#) を参照してください。

数値として保存される日付： 上記例の `DATE` はフィールド名、`to_date` は CLEM 関数です。数値として保存された日付がある場合は、数値が基準日（または紀元）からの秒数として解釈される `datetime_date` 関数を用いることでそれらを変換できます。

```
datetime_date(DATE)
```

日付を秒数へ（および逆）変換することで、次のように現在の日付に一定の日数をプラス、マイナスするといった計算を実行できます。

```
datetime_date((date_in_days(DATE)-7)*60*60*24)
```

シーケンス関数

一部の演算子では、イベントのシーケンス（順序）が重要になります。アプリケーションで使用できるレコード シーケンスは、以下のとおりです。

- シーケンスと時系列
- シーケンス関数
- レコード インデックスの作成
- 値の平均、合計、および比較
- 変化の把握（差分）
- @SINCE
- オフセット値
- その他のシーケンス機能

多くのアプリケーションでは、ストリームを通過している各レコードは、それぞれ個別で、他のすべてのレコードから独立したものと見なされます。通常、このような場合は、レコードの順序は重要ではありません。

ただし、問題によっては、レコード シーケンスが非常に重要になります。特に時系列の場合がそうで、レコードのシーケンスは、イベントまたは発生の順序、すなわちシーケンスを表します。各レコードは、特定の瞬間のスナップショットを示します。しかし、最も重要な情報は、瞬間的な値にあるのではなく、このような値が時間の経過に伴ってどのように変化し、動いていくのかということにあるのです。

もちろん、該当するパラメータが時間以外のものであってもかまいません。たとえば、レコードが、線からの距離について実行される分析を示している場合でも、同じ原則が適用されます。

シーケンスおよび特殊関数は、次の特徴によってすぐに判別できます。

- 関数名の最初に @ が付いている。
- 関数名が大文字である。

シーケンス関数は、ノードによって現在処理中のレコード、すでにノードを通過したレコード、あるいはまだノードに到達していないレコードを参照します。シーケンス関数は、CLEM 式の他の要素と自由に組み合わせて使用できますが、引数としての使用を制約されているものもあります。

例

ある事象が発生してから、またはある条件が真になってからの長さを知りたい場合があります。その場合は、次のように @SINCE を使用します。

@SINCE(Income > Outgoings)

この関数は、指定した条件が真 (true) であった最後のレコードのオフセットを返します。つまり、指定した条件が真 (true) であった最後のレコード以前のレコード数を返します。指定した条件が真にならなかった場合、@SINCE は @INDEX + 1 を返します。

@SINCE で使用される式の現在のレコードの値を参照したいこともあります。関数 @THIS を使用して、フィールド名が常に現在のレコードに適用されるように指定します。Concentration フィールドの値が、現在のレコードの 2 倍より大きい最後のレコードのオフセットを調べるには、次のように記述します。

@SINCE(Concentration > 2 * @THIS(Concentration))

定義により、現在のレコードに対して真 (true) である条件を @SINCE に指定する場合があります。次に例を示します。

@SINCE(ID == @THIS(ID))

この場合、@SINCE は現在のレコードに対して条件を評価しません。前のレコードと現在のレコードに対して条件を評価する場合は、同様な関数 @SINCEO を使用します。現在のレコードで条件が真 (true) の場合は、@SINCEO は 0 を返します。

注：@ 関数はスクリプトから呼び出せません。詳細は、3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。

関数	結果	説明
MEAN(FIELD)	Real	指定された FIELD または FIELDS に対して、値の平均値を返します。
@MEAN(FIELD, EXPR)	Real	現在のレコードを含めて、現在のノードが受け取った最後の EXPR レコードまでの、FIELD の値の平均値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの平均が返されます。この関数は、スクリプトから呼び出すことができません。詳細は、3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
@MEAN(FIELD, EXPR, INT)	Real	現在のレコードを含めて、現在のノードが受け取った最後の EXPR レコードまでの、FIELD の値の平均値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの平均が返されます。INT には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@DIFF1(FIELD)	Real	FIELD1 の最初の差分を返します。したがって、1 つの引数を指定する形式では、単純にフィールドの現在値と前の値の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@DIFF1(FIELD1, FIELD2)	Real	2 つの引数を指定する形式では、FIELD2 に関する FIELD1 の最初の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@DIFF2(FIELD)	Real	FIELD1 の 2 番目の差分を返します。したがって、1 つの引数を指定する形式では、単純にフィールドの現在値と前の値の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@DIFF2(FIELD1, FIELD2)	Real	2 つの引数を指定する形式では、FIELD2 に関する FIELD1 の最初の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。

関数	結果	説明
@INDEX	整数	現在のレコードのインデックスを返します。インデックスは、レコードが現在のノードに到達したときにレコードに対して割り当てられます。最初のレコードにはインデックス 1 が与えられます。インデックスは、その後の各レコードに対して 1 ずつ増やされます。
@LAST_NON_BLANK(FIELD)	CLEM 式	上流の入力ノードまたはデータ型ノードで定義されるように、空白でない FIELD の最後の値を返します。それまでに読み込んだレコードの FIELD の値がすべて空白である場合は、\$nullS を返します。ユーザー欠損値とも呼ばれる空白値は、各フィールドに個別に定義することができますことに注意してください。
@MAX(FIELD)	数値	指定された FIELD の最大値を返します。
@MAX(FIELD, EXPR)	数値	現在のレコードを含めて、現在までに受け取った過去 EXPR レコードの FIELD の最大値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。
@MAX(FIELD, EXPR, INT)	数値	現在のレコードを含めて、現在までに受け取った過去 EXPR レコードの FIELD の最大値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの最大値が返されます。INT には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@MIN(FIELD)	数値	指定された FIELD の最小値を返します。
@MIN(FIELD, EXPR)	数値	現在のレコードを含めて、現在までに受け取った過去 EXPR レコードの FIELD の最小値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。
@MIN(FIELD, EXPR, INT)	数値	現在のレコードを含めて、現在までに受け取った過去 EXPR レコードの FIELD の最小値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの最小値が返されます。INT には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。

CLEM 言語に関するリファレンス

関数	結果	説明
@OFFSET(FIELD, EXPR)	CLEM 式	<p>現在のレコードからEXPR で指定された値のオフセットにあるレコードの FIELDの値を返します。正のオフセットがすでにパスしたレコードを参照するのに対し、負のオフセットはまだ到着していないレコードに「先読み」を指定します。たとえば、@OFFSET(Status, 1) と指定すると、前のレコードの Status フィールドの値が返されます。一方、@OFFSET(Status, -4) では、値を取得するためにシーケンス内で 4 個先のレコードを「先読み (look ahead)」(つまり、このノードをまだ通過していないレコードまで) します。負の(先読み) オフセットは、定数として指定する必要があります。負のオフセットに限っては、EXPR も任意の CLEM 式であり、現在のレコードに対してオフセットを与えるために評価されます。この場合、性能を改善するために、この関数の引数が 3 個のバージョンを使用する必要があります(次の関数を参照)。この式が負ではない整数以外の値を返す場合、エラーになります。つまり、計算された後方参照を指定することは不正です。 注：自己参照の @OFFSET 関数ではリテラルの先読みを使用できません。たとえば、置換ノードでは、field1 の値を @OFFSET(field1,-2) のような式を使用して置換できません。</p>
@OFFSET(FIELD, EXPR, INT)	CLEM 式	<p>@OFFSET 関数と同じ演算を行いますが、3 番目の引数 INT に、前方参照する値の最大数を指定することができます。オフセットを式から計算することができる場合、性能を改善するために、この 3 番目の引数を使用する必要があります。 たとえば、@OFFSET(Foo, Month, 12) のような式では、システムが Foo の最後の 12 個の値だけを保持していればよいと判断できます。さもないければ、安全のため、全ての値を保存しておく必要があります。オフセットの値が定数の場合-これには負数の「先読み」用オフセットの値が定数である場合も含みます-、3 番目の引数は無意味で、この関数の引数が 2 個のバージョンを使用する必要があります。前述の 2 引数バージョンの自己参照関数に関する注意事項を参照してください。</p>
@SDEV(FIELD)	Real	指定された FIELD または FIELDS に対して、値の平均値を返します。
@SDEV(FIELD, EXPR)	Real	現在のレコードを含めて、現在のノードが受け取った最後の EXPR レコードまでの、FIELD の値の標準偏差を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの標準偏差を返します。

関数	結果	説明
@SDEV(FIELD, EXPR, INT)	Real	現在のレコードを含めて、現在のノードが受け取った最後の EXPR レコードまでの、FIELD の値の標準偏差を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの標準偏差を返します。INT には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@SINCE(EXPR)	CLEM 式	任意の CLEM 式が真 (true) の場合に、EXPR から過ぎたレコード数を返します。
@SINCE(EXPR, INT)	CLEM 式	2 番目の引数 INT には、前方参照するレコードの最大数を指定します。EXPR が真になっていない場合、INT は @INDEX+1 になります。
@SINCE0(EXPR)	CLEM 式	現在のレコードも考慮します。一方、@SINCE は現在のレコードは考慮しません。つまり @SINCE0 は、現在のレコードについて EXPR が真 (true) の場合に 0 を返します。
@SINCE0(EXPR, INT)	CLEM 式	2 番目の引数 INT には、前方参照するレコードの最大数を指定します。
@SUM(FIELD)	数値	指定された FIELD または FIELDS に対して、値の合計値を返します。
@SUM(FIELD, EXPR)	数値	現在のレコードを含めて、現在のノードが受け取った最後の EXPR レコードまでの、FIELD の値の合計値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの合計が返されます。
@SUM(FIELD, EXPR, INT)	数値	現在のレコードを含めて、現在のノードが受け取った最後の EXPR レコードまでの、FIELD の値の合計値を返します。FIELD は数値型フィールドの名前でなければなりません。EXPR は、0 より大きい整数を評価する任意の式を使用できます。EXPR を省略した場合、または EXPR の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの合計が返されます。INT には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@THIS(FIELD)	CLEM 式	現在のレコードの FIELD で指定された名前前のフィールドの値を返します。@SINCE 式でだけ使用されます。

グローバル関数

関数 @MEAN、@SUM、@MIN、@MAX、および @SDEV は、最大すべてのレコードおよび現在のレコードに機能します。しかし、現在のレコードの値とデータセット全体での値とを比較できると便利な場合もあります。グローバル ノードを使ってデータ セット全体の値を生成したら、CLEM 式でグローバル関数を使ってこれらの値にアクセスすることができます。

例をあげると、次のようになります。

@GLOBAL_MAX(Age)

は、データセット内で最も大きい Age の値を返します。一方、

(Value - @GLOBAL_MEAN(Value)) / @GLOBAL_SDEV(Value)

は、このレコードの Value とグローバル平均との差を標準偏差として示します。グローバル ノードによりグローバル値が算出されないと、グローバル値を使用することはできません。現在のすべてのグローバル値は、[ストリームのプロパティ] ダイアログ ボックスの [グローバル] タブにある [グローバル値の消去] ボタンをクリックしてキャンセルすることができます。

注：@ 関数はスクリプトから呼び出せません。詳細は、[3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド](#) を参照してください。

関数	結果	説明
@GLOBAL_MAX(FIELD)	数値	以前にグローバル ノードで生成されたように、データ セット全体の FIELD の最大値を返します。FIELD は数値型フィールドの名前でなければなりません。対応するグローバル値が設定されていない場合は、エラーが発生します。この関数は、スクリプトから呼び出すことができません。詳細は、 3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
@GLOBAL_MIN(FIELD)	数値	以前にグローバル ノードで生成されたように、データ セット全体の FIELD の最小値を返します。FIELD は数値型フィールドの名前でなければなりません。対応するグローバル値が設定されていない場合は、エラーが発生します。
@GLOBAL_SDEV(FIELD)	数値	以前にグローバル ノードで生成されたように、データ セット全体の FIELD の値の標準偏差を返します。FIELD は数値型フィールドの名前でなければなりません。対応するグローバル値が設定されていない場合は、エラーが発生します。

関数	結果	説明
@GLOBAL_MEAN(FIELD)	数値	以前にグローバル ノードで生成されたように、データ セット全体の FIELD の平均値を返します。FIELD は数値型フィールドの名前でなければなりません。対応するグローバル値が設定されていない場合は、エラーが発生します。
@GLOBAL_SUM(FIELD)	数値	以前にグローバル ノードで生成されたように、データ セット全体の FIELD の値の合計を返します。FIELD は数値型フィールドの名前でなければなりません。対応するグローバル値が設定されていない場合は、エラーが発生します。

空白値とヌル値処理関数

CLEM を使って、フィールド内の特定の値を「空白」、つまり欠損値と見なすように指定することができます。空白値を処理する関数を次に示します。

注：@ 関数はスクリプトから呼び出せません。詳細は、[3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド](#) を参照してください。

関数	結果	説明
@BLANK(FIELD)	ブール	上流のデータ型ノードまたは入力ノードで設定された空白処理規則 ([データ型] タブ) にしたがって、値が空白のレコードに対して真を返します。この関数は、スクリプトから呼び出すことができません。詳細は、 3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
@LAST_NON_BLANK(FIELD)	CLEM 式	上流の入力ノードまたはデータ型ノードで定義されるように、空白でない FIELD の最後の値を返します。それまでに読み込んだレコードの FIELD の値がすべて空白である場合は、\$null を返します。ユーザー欠損値とも呼ばれる空白値は、各フィールドに個別に定義することができることに注意してください。
@NULL(FIELD)	ブール	FIELD の値がシステム欠損値 \$null の場合 true を返し、ユーザー定義の空白地など、他のすべての値については false を返します。これらを確認する場合は、@BLANK(FIELD) および @NULL(FIELD) を使用します。
undef	CLEM 式	一般的に CLEM で \$null 値を入力するために用いられます。たとえば、置換ノードで空白値にヌルを挿入するために用いられます。

空白フィールドが、置換ノードで書き込まれる場合もあります。置換ノードおよびフィールド作成ノード（複数モードの場合）の両方で、特殊 CLEM 関数の @FIELD は、調査対象の現在のフィールドを表します。

特殊フィールド

特殊関数は、調査対象の特定のフィールドを表したり、フィールドのリストを入力として生成したりするために用いられます。たとえば、複数のフィールドを一度に作成する場合、**@FIELD** を使って「このフィールド作成操作を選択したフィールドに対して行う」ことを指示します。式 **log(@FIELD)** を使用すると、選択した各フィールドに対して、新しいログフィールドが作成されます。

注：@ 関数はスクリプトから呼び出せません。詳細は、3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。

関数	結果	説明
@FIELD	CLEM 式	式のコテキスト中に指定されているすべてのフィールドに対して処理を行います。この関数は、スクリプトから呼び出すことができません。詳細は、3 章 スクリプト内の CLEM 式 in IBM SPSS Modeler 14.2 スクリプト とオートメーション ガイド を参照してください。
@TARGET	CLEM 式	ユーザー定義の分析関数で CLEM 式を使用する場合、@TARGET は対象のフィールドを表すか、または分析される対象と予測のペアに対する「正しい値」を表します。通常この関数は、精度分析ノードで使用されます。
@PREDICTED	CLEM 式	ユーザー定義の分析関数で CLEM 式を使用する場合、@PREDICTED は、分析される対象と予測のペアに対して、予測される値を表します。通常この関数は、精度分析ノードで使用されます。
@PARTITION_FIELD	CLEM 式	現在のデータ区分フィールドの名前を置き換えます。
@TRAINING_PARTITION	CLEM 式	現在の学習用データ区分の値を返します。たとえば、データ選択ノードを使用して学習レコードを選択する場合、次の CLEM 式を使用します。 @PARTITION_FIELD = @TRAINING_PARTITION この方法は、データ内の各データ区分を表すためにどのデータが使用されているか関係なく、常に選択ノードが正しく動作することを保証します。
@TESTING_PARTITION	CLEM 式	現在のテスト用データ区分の値を返します。
@VALIDATION_PARTITION	CLEM 式	現在の検証用データ区分の値、0返します。

関数	結果	説明
@FIELDS_BETWEEN(start, end)	CLEM 式	データ中のフィールドの普通の順序（つまり、挿入）にもとづく、指定された開始フィールドと最終フィールドの間（開始、最終フィールドを含む）のフィールド名のリストを返します。詳細は、7 章 p.124 複数フィールドの要約を参照してください。
@FIELDS_MATCHING(pattern)	CLEM 式	指定したパターンに一致するフィールド名のリストを返します。クエスション マーク (?) をパターンに含めて正確に 1 つの文字に一致させることができます。アスタリスク (*) は 0 かそれ以上の文字数に一致します。リテラル クエスション マークまたはアスタリスクを（むしろ、ワイルドカードとして使用しないで）一致させるために、バックslashをエスケープ文字として使用することができます。詳細は、7 章 p.124 複数フィールドの要約を参照してください。
@MULTI_RESPONSE_SET	CLEM 式	名前の付いた複数レスポンス セットでフィールドのリストを返します。詳細は、7 章 p.126 複数レスポンス データの処理を参照してください。

リポジトリでの IBM SPSS Modeler の使用

IBM SPSS Collaboration and Deployment Services Repository について

IBM® SPSS® Modeler は IBM SPSS Collaboration and Deployment Services リポジトリと連動して使用でき、データマイニングモデルや関連する予測オブジェクトのライフサイクルの管理し、これらのオブジェクトをエンタープライズアプリケーション、ツール、ソリューションで使用することができます。このようにして共有できる SPSS Modeler オブジェクトには、ストリーム、ノード、ストリーム出力、シナリオ、プロジェクト、およびモデルがあります。オブジェクトは中央リポジトリに保管され、ここでは、アプリケーションと共有したり、拡張されたバージョン管理、メタデータ、および検索機能を使用して追跡することができます。

注： IBM® SPSS® Collaboration and Deployment Services リポジトリを利用するには、別途ライセンスが必要です。詳細は、<http://www.ibm.com/software/analytics/spss/products/deployment/cds/> を参照してください。

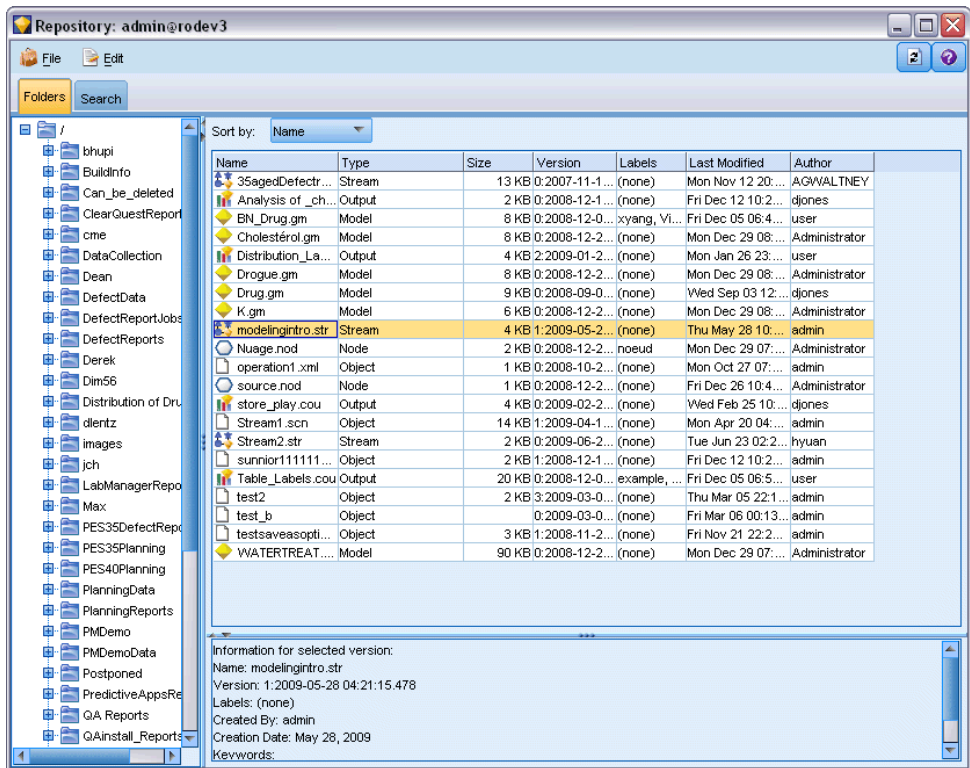
SPSS Modeler をリポジトリで使用する前に、リポジトリホストにアダプタをインストールする必要があります。このアダプタがない場合、特定の SPSS Modeler ノードまたはモデルからリポジトリのオブジェクトにアクセスしようとする、次のメッセージが表示されます。

The repository may need updating to support new node, model and output types.

アダプタインストール方法の詳細は、SPSS Modeler Deployment DVD の『SPSS Modeler Deployment のインストール』を参照してください。IBM® SPSS® Collaboration and Deployment Services Deployment Manager から SPSS Modeler リポジトリオブジェクトにアクセスする方法の詳細は、『SPSS Modeler Deployment Guide.』を参照してください。

次のセクションには、SPSS Modeler 内のリポジトリへのアクセスについての情報が記載されています。

図 9-1
IBM SPSS Collaboration and Deployment Services Repository のオブジェクト



拡張バージョン管理と検索のサポート

リポジトリには、包括的なオブジェクトのバージョン管理および検索機能があります。たとえば、ストリームを作成して、そのストリームをリポジトリに格納して、他の部門の研究者と共有する^{◆◆◆}合を考えます。後で、SPSS Modeler のストリームを更新する場合、以前のバージョンを上書きしないで、更新されたバージョンをリポジトリに追加できます。すべてのバージョンが、アクセス可能な状態で維持され、名前、ラベル、使用されているフィールドまたはその他の属性で検索できます。たとえば、純収益を入力フィールドと使用しているすべてのモデルのバージョンや、特定の作成者により作成されたすべてのモデルを検索できます(この処理を従来のファイルシステムで行うと、各バージョンごとに異なるファイル名で保存する必要がありました。またバージョン間の関連がソフトウェアから認識できない場合があります)。

シングル サインオン

シングル サインオン機能を使用すると、接続するたびにユーザー名やパスワードを入力することなく IBM® SPSS® Collaboration and Deployment Services Repository に接続することができます。ユーザーの既存のローカル ネットワーク ログインの詳細では、IBM SPSS Collaboration and Deployment Services への必要な認証が設定されています。この機能は、以下によって異なります。

- IBM SPSS Collaboration and Deployment Services を、シングル サインオン プロバイダを使用するよう設定する必要がある
- プロバイダと互換性のあるホストにログインする必要がある

詳細は、[p.179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。

IBM SPSS Collaboration and Deployment Services Repository オブジェクトの保存と展開

IBM® SPSS® Modeler で作成されるストリームは、拡張子 `.str` の付いたファイルとして、そのまま IBM® SPSS® Collaboration and Deployment Services Repository に保存できます。このように、全社規模の複数のユーザーが単一のストリームにアクセスできます。詳細は、[p.182 IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの保存](#) を参照してください。

IBM SPSS Collaboration and Deployment Services Repository でストリームを展開することもできます。展開ストリームは、拡張子 `.str` のファイルとして保存されますが、追加のメタデータを使用して、ストリームをシンクライアント アプリケーションの IBM® SPSS® Modeler Advantage で使用することができます。詳細は、[10 章 p.220 IBM SPSS Modeler Advantage でストリームを開く](#) を参照してください。

IBM® SPSS® Collaboration and Deployment Services のエンタープライズレベルの機能を利用するには、ストリームをシナリオとして展開する必要があります。シナリオとは、スコアリングのためあるいはモデル リフレッシュで使用するために IBM SPSS Collaboration and Deployment Services Repository 内で使用できる特定のノード、モデル、および追加プロパティを伴う SPSS Modeler ストリームを含んでいるファイル（拡張子 `.scn`）のことです。たとえば、新規データが利用可能になると、自己学習モデル (SLRM) を定期的なスケジュールで自動的に更新することができます。またあるいは一連のストリームを Champion-Challenger 分析を目的として展開ことができ、シナリオを比較して最も効果的な世族モデルを含むシナリオを決定します。

ストリームのシナリオとしての保存の詳細については、「[ストリーム展開のオプション](#)」（p.208）を参照してください。

シナリオとして展開されたストリームの要件

- 企業データに一貫してアクセスするには、IBM SPSS Collaboration and Deployment Services の IBM® SPSS® Collaboration and Deployment Services Enterprise View コンポーネントを介してシナリオとして展開されたストリームにアクセスする必要があります。つまり、SPSS Modeler には、ストリームの定義されたスコアリングまたはモデル作成枝内に少なくとも 1 つの Enterprise View 入力ノードが存在する必要があります。詳細は、[2 章 Enterprise View ノード in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。
- Enterprise View ノードを使用するには、まず IBM SPSS Collaboration and Deployment Services を、すでに Enterprise View ビュー、アプリケーション ビューおよびデータ プロバイダ定義 (DPD) が定義されたサイトにインストールし、設定する必要があります。詳細は、各サイトの管理者に問い合わせるか、当社 Web サイト (<http://www.ibm.com/software/analytics/spss/products/deployment/cds/>) を参照してください。
- DPD は特定の ODBC データ ソースに対して定義されます。SPSS Modeler で DPD を使用する場合、ODBC を同じ名前を持つ SPSS Modeler サーバー ホスト上で定義し、DPD で参照されるデータ ソースとして同じデータ ストアに接続する必要があります。
- また、IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver をストリームの変更または実行に使用する各コンピュータにインストールする必要があります。Windows の場合、IBM® SPSS® Modeler または IBM® SPSS® Modeler Server がインストールされているコンピュータにドライバをインストールするだけで、ドライバの設定は必要ありません。UNIX の場合、pev.sh への参照を、起動スクリプトに追加する必要があります。詳細は、[B 付録 エンタープライズ ビューノードのドライバを構成 in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド](#) を参照してください。IBM SPSS Collaboration and Deployment Services Enterprise View Driver ドライバの詳細は、各サイトの管理者に問い合わせてください。

その他の展開のオプション

IBM SPSS Collaboration and Deployment Services は、企業コンテンツ管理に次のような広範の機能を提供し、ストリームを展開またはエクスポートするための多くのメカニズムを利用することができます。

- Predictive Applications 4.x ウィザードを使用して、Predictive Applications の該当バージョンに展開するストリームをエクスポートします。詳細は、10 章 p.220 Predictive Applications 4.x ウィザード を参照してください。
- 後で IBM® SPSS® Modeler Solution Publisher ランタイムで利用するために、ストリームとモデルをエクスポートします。詳細は、2 章 IBM SPSS Modeler Solution Publisher in IBM SPSS Modeler 14.2 Solution Publisher を参照してください。
- 1 つ以上のモデルを、モデル情報をエンコードするための XML フォーマットの PMML としてエクスポートします。詳細は、10 章 p.230 PMML としてのモデルのインポートおよびエクスポート を参照してください。

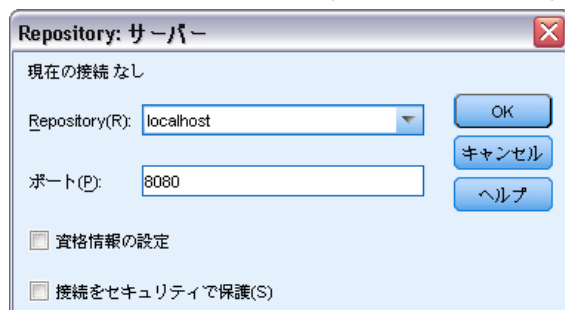
IBM SPSS Collaboration and Deployment Services Repository への接続

- ▶ IBM® SPSS® Collaboration and Deployment Services Repository に接続するには、IBM® SPSS® Modeler メニューから次の項目をクリックします。
ツール > リポジトリ > オプション...
- ▶ 必要に応じて、ログイン オプションを指定します。

設定は、各サイトまたはインストール環境によって変わってきます。特定のポート、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。

注： IBM® SPSS® Collaboration and Deployment Services リポジトリを利用するには、別途ライセンスが必要です。詳細は、<http://www.ibm.com/software/analytics/spss/products/deployment/cds/> を参照してください。

図 9-2
IBM SPSS Collaboration and Deployment Services Repository ログイン



レポジトリ：アクセスする IBM SPSS Collaboration and Deployment Services Repository のインストール。通常、これはレポジトリがインストールされているホスト サーバーの名前になります。一度に接続できるレポジトリは 1 つだけです。

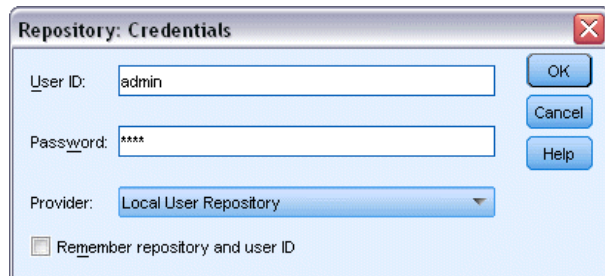
ポート：接続をホストするのに使用されるポート番号を指定します。通常は、デフォルトで 8080 です。

資格情報の設定：このボックスのチェックを解除すると、**シングル サインオン**機能を有効にし、ローカル コンピュータのユーザー名とパスワードの詳細を入力してログインします。シングル サインオンを使用できない場合、またはこのボックスをチェックしてシングル サインオンを無効にした場合（たとえば、管理者アカウントにログインした場合）、資格情報を入力するための詳細画面が表示されます。

セキュア接続の確認：SSL (Secure Sockets Layer) 接続を使用するかどうかを指定します。SSL は、ネットワークを介してセキュアなデータ送信を行うために一般的に使用されているプロトコルです。この機能を使用するには、IBM SPSS Collaboration and Deployment Services Repository をホストするサーバー側で SSL を有効にする必要があります。必要な場合、詳細を各サイトの管理者に問い合わせてください。

IBM SPSS Collaboration and Deployment Services Repository の資格情報の入力

図 9-3
IBM SPSS Collaboration and Deployment Services Repository 資格情報の入力



The screenshot shows a dialog box titled "Repository: Credentials". It has a close button (X) in the top right corner. The dialog contains three input fields: "User ID:" with the text "admin", "Password:" with masked characters "****", and "Provider:" with a dropdown menu showing "Local User Repository". To the right of the input fields are three buttons: "OK", "Cancel", and "Help". At the bottom left, there is a checkbox labeled "Remember repository and user ID" which is currently unchecked.

ユーザー ID およびパスワード：ログインに使用する有効なユーザー名とパスワードを指定します。必要があれば、詳細についてローカル システムの管理者にお問い合わせください。

プロバイダ：認証するセキュリティ プロバイダを選択します。異なるセキュリティ プロバイダを使用するよう、IBM® SPSS® Collaboration and Deployment Services Repository を設定することができます。必要な場合、詳しい情報を各サイトの管理者に問い合わせてください。

レポジトリとユーザー ID を記憶：毎回、接続時に同じ情報を再入力しなくても済むように、現在の設定をデフォルトとして保存します。

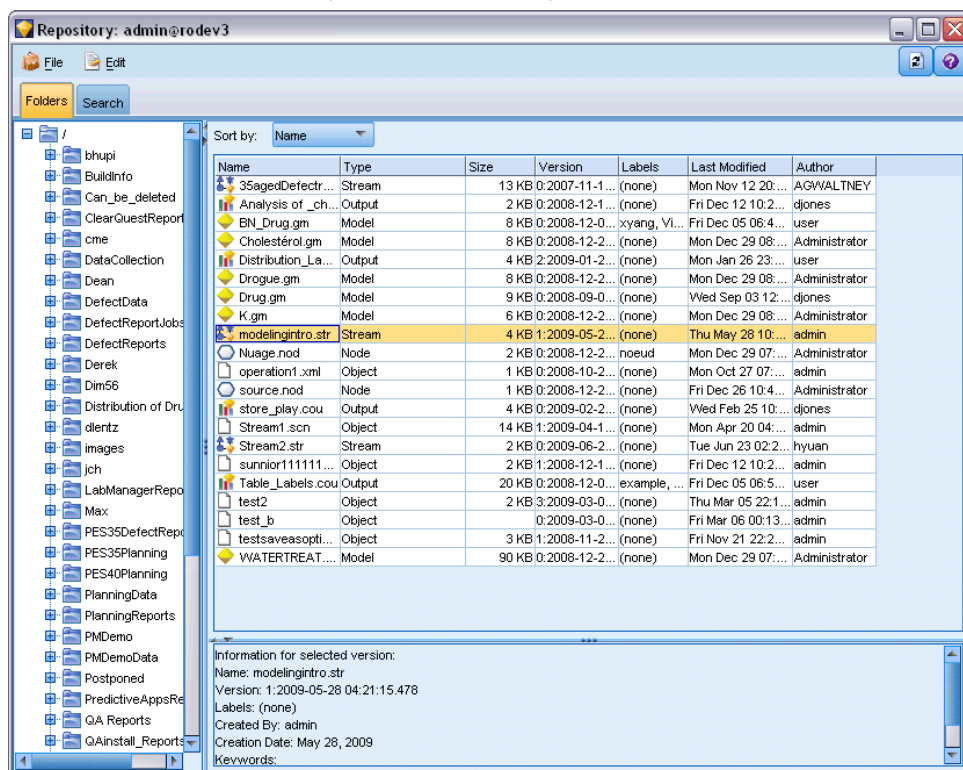
IBM SPSS Collaboration and Deployment Services Repository Repository コンテンツを参照

IBM® SPSS® Collaboration and Deployment Services Repository により、Windows Explorer のように格納した内容を参照することができます。また格納された各オブジェクトのバージョンも参照することができます。

- ▶ IBM SPSS Collaboration and Deployment Services Repository ウィンドウを開くには、IBM® SPSS® Modeler のメニューで次の各項目をクリックします。
ツール > リポジトリ > [検証(E)...]
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続を参照してください](#)。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。

図 9-4

IBM SPSS Collaboration and Deployment Services Repository コンテンツを参照

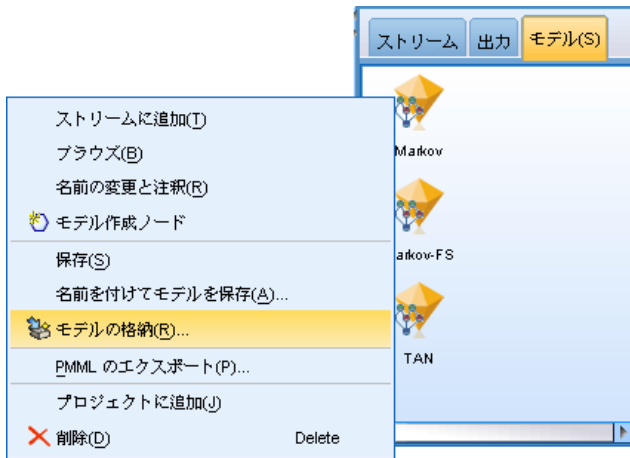


エクスプローラ ウィンドウには、最初にフォルダ階層のツリー ビューが表示されます。フォルダ名をクリックしてコンテンツを表示します。

現在の選択または検索基準に一致するオブジェクトが右側の領域にリストされ、選択されているバージョンの詳細な情報が右下の領域に表示されます。表示される属性値は、最も新しいバージョンに適用されます。

IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの保存

図 9-5
モデルの格納



リポジトリにストリーム、ノード、モデル、モデル パレット、プロジェクト、出力オブジェクトを格納できます。他のユーザーやアプリケーションはそれらにアクセスすることができます。

注： IBM® SPSS® Collaboration and Deployment Services リポジトリを利用するには、別途ライセンスが必要です。詳細は、<http://www.ibm.com/software/analytics/spss/products/deployment/cds/> を参照してください。

ストリーム出力を IBM® SPSS® Collaboration and Deployment Services Repository に、その他のユーザーが IBM® SPSS® Collaboration and Deployment Services Deployment Portal を使用してインターネットで表示できる形式で公開することもできます。詳細は、6 章 [Web に公開] in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。

オブジェクトの特性の設定

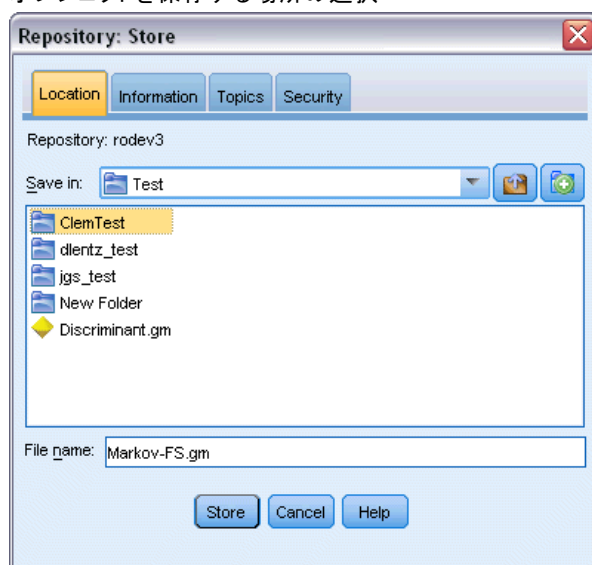
オブジェクトを保存すると、[リポジトリ：保存] ダイアログ ボックスが表示され、オブジェクトのさまざまなプロパティの値を設定できるようになります。以下を行うことができます。

- オブジェクトを保存する名前およびリポジトリ フォルダを選択する
- バージョン ラベルやその他の検索可能なプロパティ例など、オブジェクトに関する情報を追加する
- 1 つまたは複数の分類トピックをオブジェクトに割り当てる
- オブジェクトのセキュリティ オプションを設定する

次の項では、設定できるプロパティについて説明します。

オブジェクトを保存する場所の選択

図 9-6
オブジェクトを保存する場所の選択



ファイルの保存場所: オブジェクトが保存される場所、現在のフォルダが表示されます。リストでフォルダ名をダブルクリックすると、そのフォルダが現在のフォルダに設定されます。親フォルダを選択するには、[上位フォルダ] ボタンを使用します。現在のレベルにフォルダを作成するには、[新規フォルダ] ボタンを使用します。

ファイル名: オブジェクトを保存する名前。

保存: オブジェクトを現在の場所に保存します。

保存するオブジェクトに関する情報の追加

このタブのフィールドはオプションです。

図 9-7
保存するオブジェクトに関する情報の追加

The screenshot shows a dialog box titled "Repository: Store" with a close button in the top right corner. It has four tabs: "Location", "Information", "Topics", and "Security". The "Information" tab is active. Below the tabs, there are several input fields and controls:

- Author:** A text field containing "admin".
- Version Label:** A list box with three items: "LATEST" (checked), "All SLRM-TrainingData", and "Baseline". To the right of the list is an "Add..." button.
- Description:** A large empty text area.
- Keywords:** A large empty text area.
- Expiration:** Two radio buttons: "None" (selected) and "Date:". The "Date:" field contains "July 30, 2010" and has a calendar icon to its right.

At the bottom of the dialog are three buttons: "Store", "Cancel", and "Help".

作成者：リポジトリのオブジェクトを作成したユーザーのユーザー名。デフォルトでは、リポジトリ接続に使用されたユーザー名が表示されますが、ここでこの名前を変更できます。

バージョンラベル：オブジェクトのバージョンを示すラベルをリストから選択するか、[追加]をクリックして、新しいラベルを作成します。ラベルには「[」文字を使用しないでください。このオブジェクトバージョンにラベルを割り当てない場合は、チェックボックスがオンになっていないことを確認してください。詳細は、[p. 203 オブジェクトのプロパティの表示および編集](#)を参照してください。

説明：オブジェクトの説明。説明によって、オブジェクトを検索できます（注意事項を参照）。

キーワード：オブジェクトに関連し、検索に使用できる 1 つまたは複数のキーワード（注意事項を参照）。

有効期限：その日以降、オブジェクトが一般ユーザーに表示されなくなる日付。ただし、この日以降も所有者やリポジトリ管理者には表示されます。有効期限の日付を設定するには、[日付]オプションを選択して日付を入力するか、カレンダーのボタンを使用して日付を選択します。

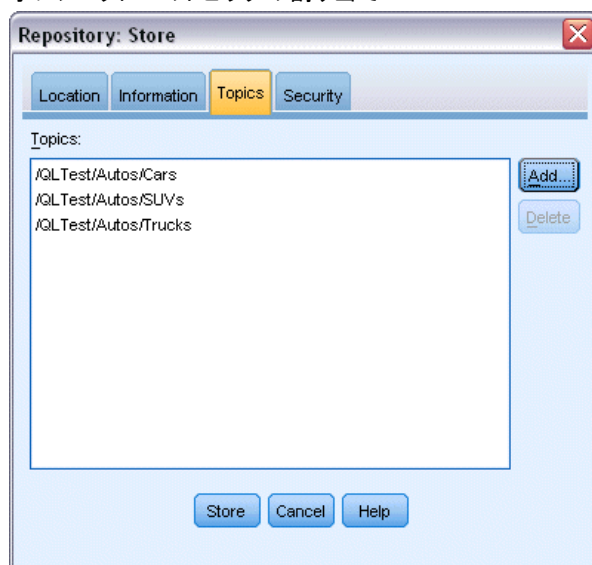
保存：オブジェクトを現在の場所に保存します。

注：[説明] フィールド [キーワード] フィールドの情報は、オブジェクトの [注釈] タブの IBM® SPSS® Modeler に入力したものとは異なるものとして扱われます。説明またはキーワードによるリポジトリ検索では、[注釈] タブからの情報は返されません。詳細は、[p.195 IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの検索](#) を参照してください。

保存したオブジェクトへのトピックの割り当て

トピックは、リポジトリに保存されているコンテンツの階層分類システムです。オブジェクトを保存するときに使用可能なトピックから選択でき、またユーザーはトピックによってオブジェクトを検索できます。使用できるトピックのリストは、適切な権限を持つリポジトリ ユーザーによって設定されます（詳細は、『Deployment Manager ユーザー ガイド』を参照してください）。

図 9-8
オブジェクトへのトピックの割り当て



トピックをオブジェクトに割り当てるには

- ▶ [追加] ボタンをクリックします。
- ▶ 使用できるトピックのリストのトピック名をクリックします。
- ▶ [OK] をクリックします。

トピックの割り当てを削除する手順は、次のとおりです。

- ▶ 割り当てたトピックのリストでトピックを選択します。

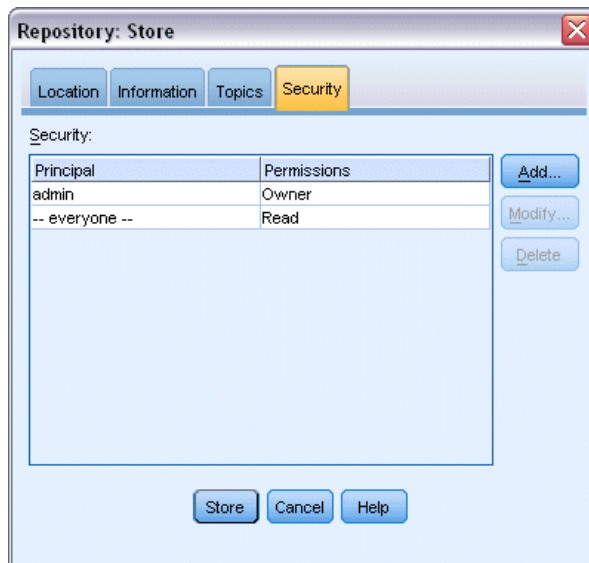
- ▶ 「削除」をクリックします。

保存したオブジェクトのセキュリティ オプションを設定

保存したオブジェクトのさまざまなセキュリティ オプションを設定または変更できます。1 つまたは複数の**主成分**（ユーザーまたはユーザーのグループ）について、次のことを実行できます。

- オブジェクトへのアクセス権限を割り当てる
- オブジェクトへのアクセス権限を変更する
- オブジェクトへのアクセス権限を削除する

図 9-9
オブジェクトのセキュリティ オプションの設定



主成分： このオブジェクトにアクセス権限があるユーザーまたはグループのリポジトリ ユーザー名。

アクセス許可： このユーザーまたはグループがオブジェクトに対して持つアクセス権限。

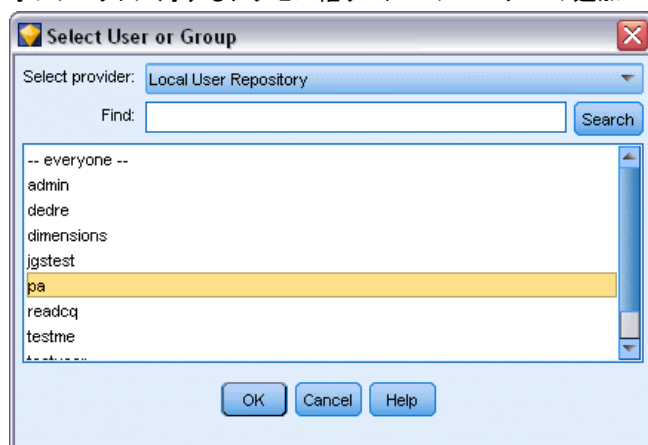
追加： 1 つまたは複数のユーザーまたはグループをこのオブジェクトのアクセス権限を持つユーザーまたはグループのリストに追加できます。 [詳細は、 p.187 アクセス権リストへのユーザーの追加 を参照してください。](#)

変更： このオブジェクトに対する選択したユーザーまたはグループのアクセス権限を変更することができます。デフォルトでは、読み取りアクセス権が付与されています。所有者、書き込み、読み取り、削除、変更という追加のアクセス権限を与えることができます。

削除： このオブジェクトに対するアクセス権リストから選択したユーザーまたはグループを削除します。

アクセス権リストへのユーザーの追加

図 9-10
オブジェクトに対するアクセス権リストへのユーザーの追加



プロバイダを選択： 認証するセキュリティ プロバイダを選択します。異なるセキュリティ プロバイダを使用するよう、IBM® SPSS® Collaboration and Deployment Services Repository を設定することができます。必要な場合、詳しい情報を各サイトの管理者に問い合わせてください。

検索： 追加するユーザーまたはグループのリポジトリ ユーザー名を入力し、[検索] を検索するとユーザー リストに名前が表示されます。一度に複数のユーザー名を追加するには、このフィールドを空白にし、[検索] をクリックすると、すべてのリポジトリ ユーザー名のリストが表示されます。

ユーザー リスト： リストから 1 つまたは複数のユーザー名を選択し、[OK] をクリックすると、そのユーザー名がアクセス権リストに追加されます。

オブジェクトに対するアクセス権限の変更

図 9-11
オブジェクトに対するアクセス権限の変更



所有者： このユーザーまたはグループに、オブジェクトに対する所有者のアクセス権限を与えます。所有者は、削除および変更のアクセス権限など、オブジェクトに対して完全なコントロール権を持ちます。

読み取り： デフォルトでは、オブジェクトの所有者ではないユーザーまたはグループには、オブジェクトに対する読み取りアクセス権限のみが与えられます。該当するチェック ボックスをオンにすると、このユーザーまたはグループの書き込み、削除および権限の変更のアクセス権限が追加されます。

ストリームの保存

リポジトリにストリームを .str ファイルとして保存できます。リポジトリから他のユーザーはそれらにアクセスすることができます。

注： リポジトリ機能をさらに利用するための、ストリーム展開の詳細については、「ストリームの展開」（ p. 206 ）を参照してください。

現在のストリームを保存するには

- ▶ メイン メニューで次の各項目をクリックします。
File > 保存 > ストリームとして保存...
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、 [p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ Repository の [保存] ダイアログ ボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。詳細は、 [p. 182 オブジェクトの特性の設定](#) を参照してください。

プロジェクトの格納

リポジトリに完全な IBM® SPSS® Modeler プロジェクトを .cpj ファイルとして保存できます。他のユーザーもプロジェクトにアクセスすることができます。

プロジェクト ファイルがその他の SPSS Modeler オブジェクトのコンテナであるため、SPSS Modeler のプロジェクトのオブジェクトを格納することを IBM® SPSS® Collaboration and Deployment Services Repository に知らせる必要があります。この操作では、[プロジェクト プロパティ] ダイアログ ボックスでの設定を使用します。詳細は、11 章 p.243 [プロジェクトのプロパティの設定](#) を参照してください。

プロジェクトをリポジトリに格納するよう設定すると、新規のオブジェクトをそのプロジェクトに追加するたびに、SPSS Modeler はそのオブジェクトを格納するよう要求するメッセージを自動的に出力します。

SPSS Modeler とのセッションを終了するときに、プロジェクト ファイルの新しいバージョンを格納して、その追加を記憶する必要があります。プロジェクト ファイルは、自動的にこのオブジェクトの最新バージョンを含めます（そして取得します）。SPSS Modeler とのセッション中にオブジェクトを追加しなかった場合は、このプロジェクト ファイルを復元してはいけません。ただし、変更したプロジェクトのオブジェクト（ストリーム、出力など）の新しいバージョンを格納する必要があります。

プロジェクトを格納する手順は、次のとおりです。

- ▶ SPSS Modeler のマネージャ パレットの [CRISP-DM] タブまたは [クラス] タブでプロジェクトを選択し、メニューから次の項目をクリックします。
File > プロジェクト > プロジェクトを格納...
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ Repository の [保存] ダイアログ ボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。詳細は、[p. 182 オブジェクトの特性の設定](#) を参照してください。

ノードの格納

リポジトリに現在のストリームの各ノード定義を .nod ファイルとして保存できます。リポジトリから他のユーザーはそのファイルにアクセスすることができます。

ノードを格納する手順は、次のとおりです。

- ▶ ストリーム キャンバスにあるノードを右クリックして、[ノードを格納] をクリックします。
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ Repository の [保存] ダイアログ ボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。詳細は、[p. 182 オブジェクトの特性の設定](#) を参照してください。

出力オブジェクトの格納

リポジトリに現在のストリームの出力オブジェクトを .nod ファイルとして保存できます。リポジトリから他のユーザーはそのファイルにアクセスすることができます。

出力オブジェクトを格納する手順は、次のとおりです。

- ▶ IBM® SPSS® Modeler のマネージャ パネルの [出力] タブでオブジェクトを選択し、メイン メニューで次の項目をクリックします。
File > 出力 > 出力を格納...
- ▶ または、[出力] タブにあるオブジェクトを右クリックして、[格納] をクリックします。
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ Repository の [保存] ダイアログ ボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。詳細は、[p. 182 オブジェクトの特性の設定](#) を参照してください。

モデルおよびモデル パレットの格納

リポジトリに各モデルを .gm ファイルとして保存できます。リポジトリから他のユーザーはそれらにアクセスすることができます。また、モデル パレットのすべての内容を .gen ファイルとしてリポジトリに保存することもできます。

モデルの格納

- ▶ IBM® SPSS® Modeler の [モデル] パレットのオブジェクトをクリックし、メインメニューで次の項目をクリックします。
File > Models > モデルの格納(R)...
- ▶ または、モデル パレットにあるオブジェクトを右クリックして、[格納] をクリックします。
- ▶ 「格納手順の完了」から先に進みます。

モデル パレットの格納

- ▶ モデル パレットの背景を右クリックします。
- ▶ ポップアップメニューで [パレットを格納] をクリックします。
- ▶ 「格納手順の完了」から先に進みます。

格納手順の完了

- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続を参照してください](#)。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ Repository の [保存] ダイアログボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。詳細は、[p. 182 オブジェクトの特性の設定を参照してください](#)。

IBM SPSS Collaboration and Deployment Services Repository からのオブジェクトの取得

IBM® SPSS® Collaboration and Deployment Services Repository に格納されたストリーム、モデル、モデル パレット、ノード、プロジェクト、出力オブジェクトを取得できます。

注：ここで説明したようにメニュー オプションを使用するほか、IBM® SPSS® Modeler ウィンドウの右上にあるマネージャ パネルの該当するタブを右クリックして、ストリーム、出力オブジェクト、モデルおよびモデルパレットを取得することができます。

- ▶ ストリームを取得するには、SPSS Modeler のメインメニューの次の項目をクリックしてください。
File > [ストリームを取得(V)...]

- ▶ モデル、モデル パレット、プロジェクト、または出力オブジェクトを取得するには、SPSS Modeler のメイン メニューで次の項目をクリックします。

File > Models > モデルの取得...

or

File > Models > モデル パレットの取得...

or

File > プロジェクト > プロジェクトの取得...

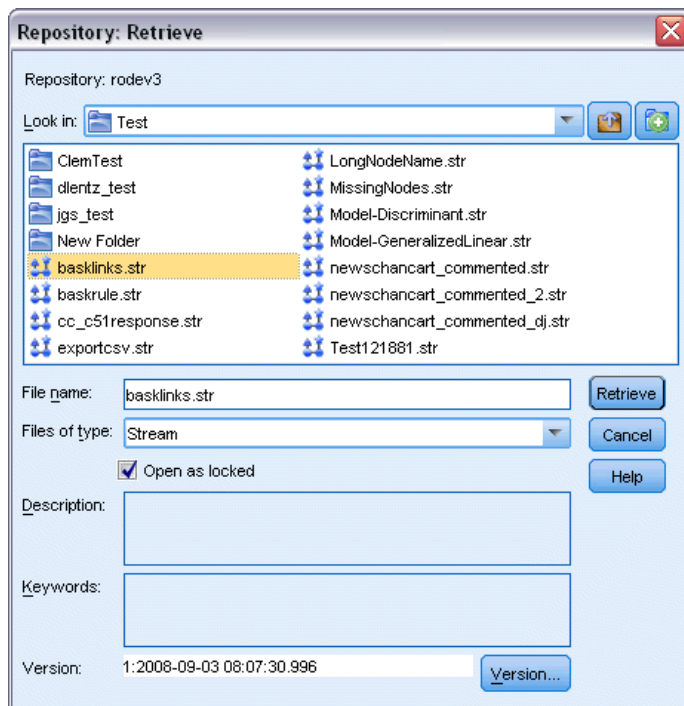
or

File > 出力 > 出力の取得...

- ▶ または、マネージャ またはプロジェクト パネル内で右クリックし、ポップアップ メニューから [取得] をクリックします。
- ▶ ノードを取得するには、SPSS Modeler のメインメニューの次の項目をクリックしてください。
[挿入(O)] > リポジトリからノード (またはスーパーノード)...
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ [リポジトリ: 取得] ダイアログで、オブジェクトを参照し、オブジェクトを選択して [取得] ボタンをクリックします。

取得するオブジェクトの選択

図 9-12
IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの取得



検索対象: 現在のフォルダのフォルダ階層が表示されます。異なるフォルダに移動するには、このリストから選択して直接移動するか、このフィールドの下にあるオブジェクト リストを使用して移動します。

[上位フォルダ] ボタン: 階層内の現在のフォルダの 1 つ上のレベルに移動します。

[新規フォルダ] ボタン: 階層内の現在のフォルダに新しいフォルダを作成します。

ファイル名: 選択したオブジェクトのリポジトリ ファイル名。そのオブジェクトを取得するには、[取得] をクリックします。

ファイルの種類: 取得するオブジェクトの種類。フォルダと共に、この種類のオブジェクトのみが、オブジェクト リストに表示されます。取得する異なる種類のオブジェクトを表示するには、リストからオブジェクトの種類を選択します。

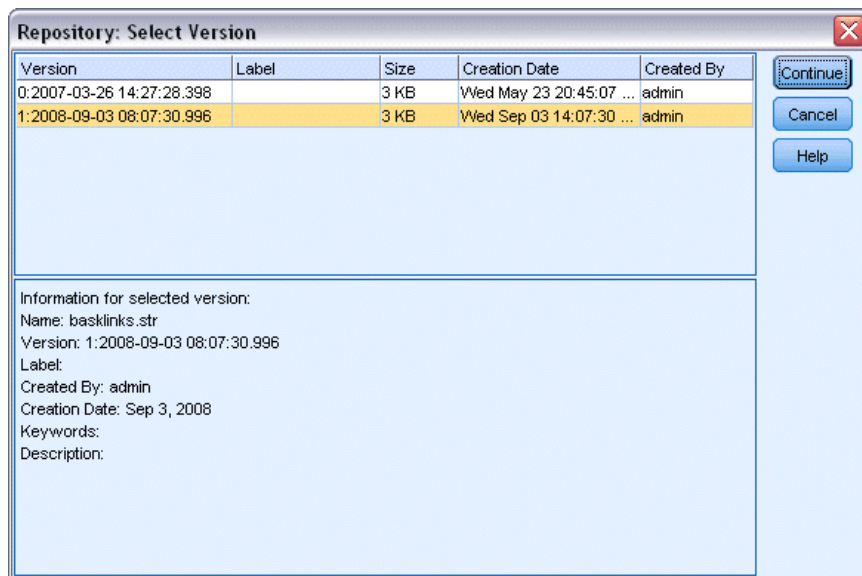
ロックして開く: デフォルトでは、オブジェクトを取得すると、リポジトリ内でロックされ、他のユーザーは更新できません。取得時にオブジェクトをロックしたくない場合、このボックスをオフにします。

説明、キーワード: オブジェクトを保存したとき、オブジェクトに関する詳細が定義されている場合、これらの詳細情報がここに表示されます。詳細は、 [p. 183 保存するオブジェクトに関する情報の追加](#) を参照してください。

バージョン: 最新版以外のオブジェクトのバージョンを取得するには、このボタンをクリックします。すべてのバージョンについての詳細な情報が表示され、必要なバージョンを選択できます。

オブジェクトのバージョンの選択

図 9-13
オブジェクトのバージョンの選択



特定のバージョンのリポジトリ オブジェクトを選択する手順は、次のとおりです。

- ▶ (オプション) 該当する列の見出しをクリックして、バージョン、ラベル、サイズ、作成日、作成ユーザーによってリストをソートします。
- ▶ 使用するオブジェクトのバージョンを選択します。
- ▶ [続行] をクリックします。

IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの検索

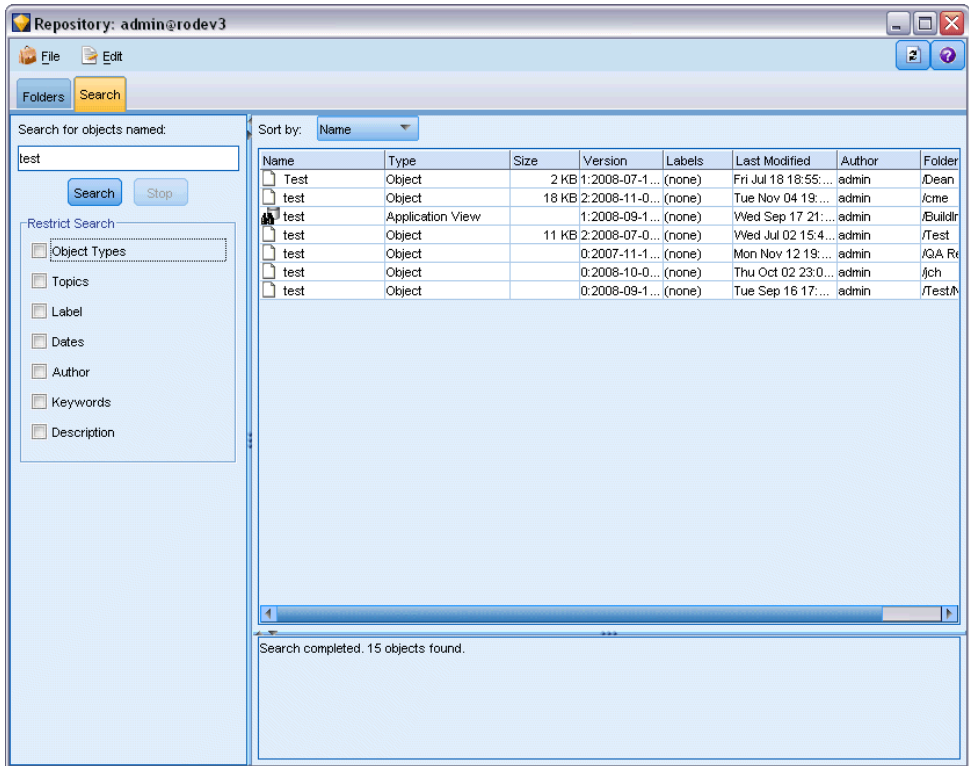
オブジェクトを、名前、フォルダ、タイプ、ラベル、日付、またはその他の基準で検索できます。

名前で検索

名前でオブジェクトを検索する手順は、次のとおりです。

- ▶ IBM® SPSS® Modeler のメインメニューで次の各項目をクリックします。
ツール > IBM® SPSS® Collaboration and Deployment Services Repository > [検証(E)...]
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ [検索] タブをクリックします。
- ▶ [検索するオブジェクトの名前] フィールドで、検索したいオブジェクトの名前を指定します。

図 9-14
名前でオブジェクトを検索



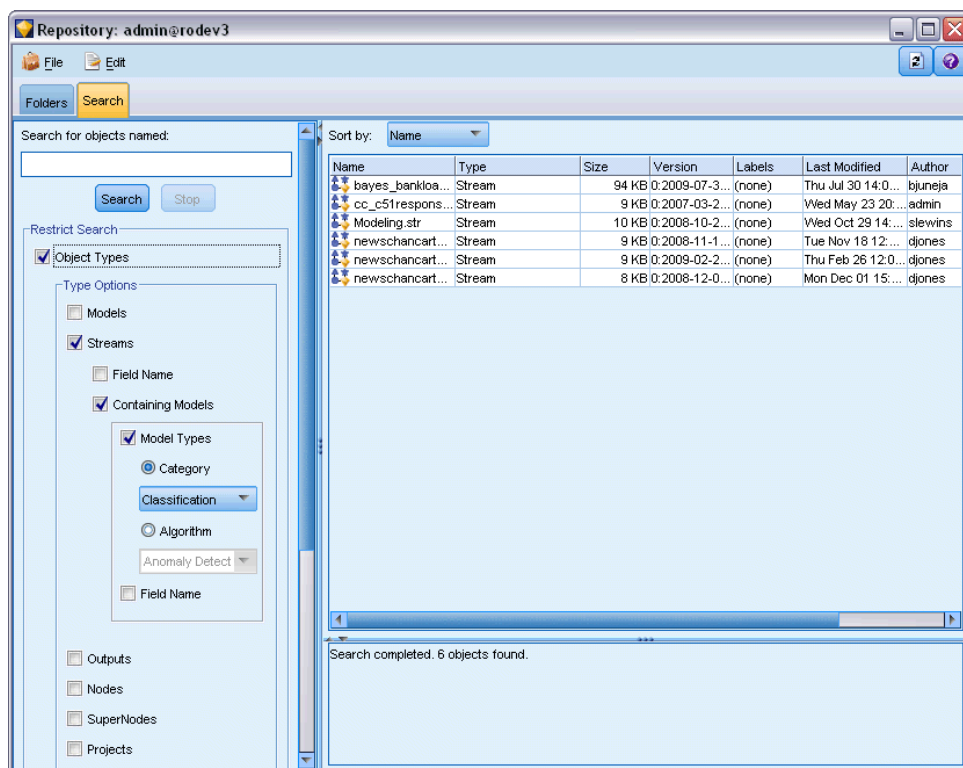
名前でオブジェクトを検索する場合、ワイルドカード文字としてアスタリスク (*) を使用すると、任意の文字列、疑問符 (?) を使用すると、1 文字に相当します。たとえば、*cluster* は、cluster という文字列を名前の一部に持っているすべてのオブジェクトに一致します。文字列 m0?_* を検索すると、M01_cluster.str と M02_cluster.str は一致しますが、M01a_cluster.str には一致しません。検索する場合は大文字と小文字の区別はされません (cluster は Cluster と CLUSTER にも一致します)。

注： オブジェクトが多数ある場合は、検索に時間がかかる場合があります。

他の基準による検索

タイトル、ラベル、日付、作成者、キーワード、インデックス付きコンテンツ、または説明に基づいて検索を実行できます。すべての指定された検索基準に一致するオブジェクトだけが検索されます。たとえば、特定のラベルが付けられ、特定の日付以降に変更された、1 つ以上のクラスタリング モデルを含むすべてのストリームを検索できます。

図 9-15
特定のタイプのモデルを含むストリームの検索



オブジェクトの種類：モデル、ストリーム、出力、ノード、スーパーノード、プロジェクト、モデルパレット、シナリオまたは他の種類のオブジェクトに対する検索を制限することができます。

- **モデル：**カテゴリ（分類、近似、クラスタリング、その他）、また **◆◆◆** は Kohonen など、特定のモデル作成のアルゴリズムによりモデルを検索できます。

使用されているフィールド名で検索することもできます。たとえば、入力フィールドまたは出力（対象）フィールドとして、収入という名前のフィールドを使用しているすべてのモデルを検索できます。

- **ストリーム：**ストリームについて、使用されるフィールド名またはストリームに含まれるモデルタイプ（カテゴリまたはアルゴリズム）で検索を制限できます。

トピック：適切な権限を持つリポジトリユーザーによって、リストセットの特定のトピックに関連するモデルを検索できます（詳細は、『Deployment Manager ユーザーガイド』を参照してください）。リストを取得するには、このボックスをオンにし、表示される [トピックを追加] ボタンを

クリックして、リストから 1 つまたは複数のトピックを選択し、[OK] をクリックします。

ラベル: 特定のオブジェクト バージョンのラベルに検索を制限します。

日付: 作成日時または修正日時を指定して、以前、以降、または指定された日付範囲の間にあるオブジェクトを検索できます。

作成者: 特定のユーザーが作成したオブジェクトに検索を制限します。

キーワード: 指定されたキーワードの検索。SPSS Modeler では、キーワードは、ストリーム、モデル、または出力オブジェクトの [注釈] タブで指定します。

説明: 説明フィールドにある特定の用語を検索します。SPSS Modeler では、説明は、ストリーム、モデル、または出力オブジェクトの [注釈] タブで指定します。複数のキーワードをセミコロンで区切って指定できます (例: `income; crop type; claim value`)。 (検索するフレーズでは空白が重要です。たとえば、スペースが 1 個の `crop type` とスペースが 2 個の `crop type` は同一ではありません。)

IBM SPSS Collaboration and Deployment Services Repository オブジェクトの変更

IBM® SPSS® Modeler から直接、IBM® SPSS® Collaboration and Deployment Services Repository の既存のオブジェクトを変更できます。以下を行うことができます。

- フォルダの作成、名前の変更または削除
- オブジェクトのロックまたはロック解除
- オブジェクトの削除

フォルダの編集、名前の変更、削除

- ▶ IBM® SPSS® Collaboration and Deployment Services Repository でフォルダの操作を実行するには、IBM® SPSS® Modeler のメニューで次の項目をクリックします。
ツール > IBM® SPSS® Collaboration and Deployment Services Repository > [検証(E)...]
- ▶ 必要に応じて、リポジトリへの接続の設定を行います。詳細は、[p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続](#) を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- ▶ [フォルダ] タブがアクティブであることを確認してください。

- ▶ 新しいフォルダを作成するには、親フォルダを右クリックして、[新規フォルダ] をクリックします。
- ▶ フォルダ名を変更するには、そのフォルダを右クリックして [フォルダ名の変更] をクリックします。
- ▶ フォルダを削除するには、そのフォルダを右クリックして [フォルダの削除] をクリックします。

IBM SPSS Collaboration and Deployment Services Repository オブジェクトのロックおよびロック解除

オブジェクトをロックして、ほかのユーザーが既存のバージョンを更新したり新しいバージョンを作成しないようにすることができます。ロックされたオブジェクトは、オブジェクト アイコン上に南京錠のシンボルで示されます。

図 9-16
ロックされたオブジェクト



オブジェクトをロックするには

- ▶ IBM® SPSS® Collaboration and Deployment Services Repository ウィンドウで目的のオブジェクトを右クリックします。
- ▶ [ロック] をクリックします。

オブジェクトをロックを解除するには

- ▶ IBM SPSS Collaboration and Deployment Services Repository ウィンドウで目的のオブジェクトを右クリックします。
- ▶ [ロック解除] をクリックします。

IBM SPSS Collaboration and Deployment Services Repository オブジェクトの削除

オブジェクトを IBM® SPSS® Collaboration and Deployment Services Repository から削除する前に、全部のバージョンを削除するか、またはあるオブジェクトの特定のバージョンを削除するのかを決める必要があります。

オブジェクトの全部のバージョンを削除するには

- ▶ IBM SPSS Collaboration and Deployment Services Repository ウィンドウで目的のオブジェクトを右クリックします。
- ▶ [オブジェクトの削除] をクリックします。

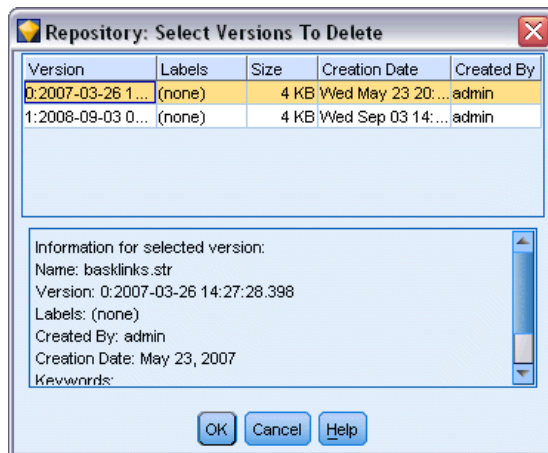
オブジェクトの最新バージョンを削除するには

- ▶ IBM SPSS Collaboration and Deployment Services Repository ウィンドウで目的のオブジェクトを右クリックします。
- ▶ [削除] をクリックします。

オブジェクトの前のバージョンを削除するには

- ▶ IBM SPSS Collaboration and Deployment Services Repository ウィンドウで目的のオブジェクトを右クリックします。
- ▶ [バージョンの削除] をクリックします。
- ▶ 削除するバージョンを選択して [OK] をクリックします。

図 9-17
削除するバージョンの選択



IBM SPSS Collaboration and Deployment Services Repository オブジェクトのプロパティの管理

IBM® SPSS® Modeler からさまざまなオブジェクトのプロパティを制御できます。以下を行うことができます。

- フォルダのプロパティを表示する

- オブジェクトのプロパティを表示および編集する
- オブジェクトのバージョン ラベルを作成、適用、および削除する

フォルダのプロパティの表示

IBM® SPSS® Collaboration and Deployment Services Repository ウィンドウにあるフォルダのプロパティを表示するには、目的のフォルダを右クリックします。[フォルダのプロパティ] をクリックします。

[全般] タブ

図 9-18
フォルダのプロパティ

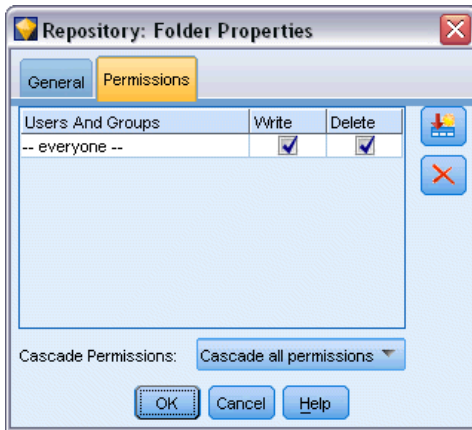


フォルダ名、作成、および修正日時を表示します。

[権限] タブ

フォルダの読み込み権限と書き込み権限を指定できます。親フォルダへのアクセスを持つすべてのユーザーとグループがリストされます。権限は階層によって異なります。たとえば、読み取り権限を持っていない場合、書き込み権限を持つことができません。書き込み権限を持っていない場合、削除権限を持つことはできません。

図 9-19
フォルダのプロパティ



ユーザーとグループ：このフォルダに対して少なくとも読み取りアクセス権を持っているリポジトリ ユーザーおよびグループが表示されます。[書き込み] チェック ボックスおよび [削除] チェック ボックスをオンにすると、このフォルダに対するこれらのアクセス権が、特定のユーザーまたはグループに割り当てられます。別のユーザーまたはグループにアクセスを割り当てるには、[権限] タブの右側にある [ユーザーとグループの追加] アイコンをクリックします。利用可能なユーザーとグループのリストは、管理者により管理されています。

権限のカスケード：子フォルダがある場合、現在のフォルダに対する変更が、子フォルダにどのように適用されるかを制御できます。

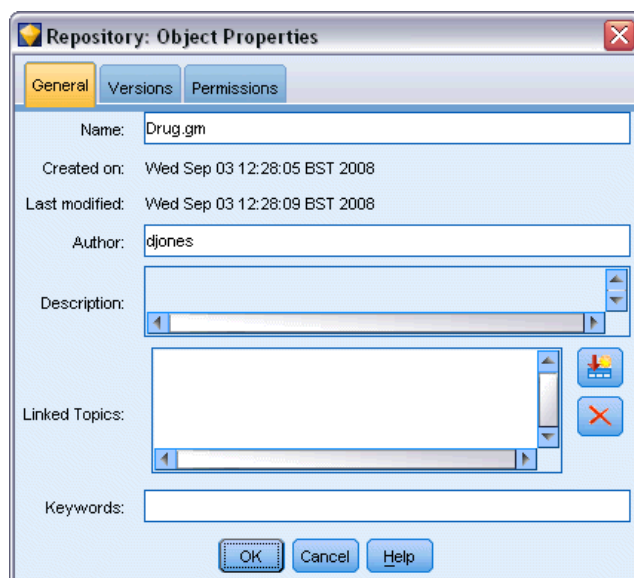
- **すべての権限をカスケードする：**現在のフォルダからすべての子、および子孫のフォルダから権限設定をカスケードします。この方法は、複数のフォルダの権限を一度に素速く設定できます。親フォルダで必要な権限を設定し、それを必要に応じてカスケードします。
- **変更のみカスケードする：**前回の変更が適用されてから行われた変更のみをカスケードします。たとえば、新しいグループが追加されたので、そのグループに「販売」ツリーにある全てのフォルダへのアクセスを与えたい場合、グループ アクセスをルートの「販売」フォルダに与えて、その変更をすべてのサブフォルダにカスケードできます。既存のサブフォルダに対するすべてのそれ以外の権限は、以前のまま変更されません。
- **カスケードしない：**変更は、現在のフォルダにのみ適用され、子フォルダにカスケードされません。

オブジェクトのプロパティの表示および編集

IBM® SPSS® Collaboration and Deployment Services Repository のオブジェクトのプロパティ ダイアログ ボックスで、プロパティの表示や編集ができます。プロパティには変更できないものもありますが、いつでも新しいバージョンを追加することで更新できます。

- ▶ IBM SPSS Collaboration and Deployment Services Repository ウィンドウで目的のオブジェクトを右クリックします。
- ▶ [オブジェクトのプロパティ] をクリックします。

図 9-20
オブジェクトのプロパティ



[全般] タブ

名前 :IBM SPSS Collaboration and Deployment Services Repository で表示したときのオブジェクトの名前です。

作成日 : オブジェクト（バージョンでない）が作成された日付です。

最終更新日 : 最新のバージョンが更新された日時です。

作成者 : ユーザーのログイン名です。

説明 : デフォルトでは、IBM® SPSS® Modeler でオブジェクトの [注釈] タブで指定された説明が含まれています。

リンクされたトピック :IBM SPSS Collaboration and Deployment Services Repository を使用すると、モデルおよび関連付けられたオブジェクトを、必要に応じてトピックごとに整理できます。使用できるトピックのリスト

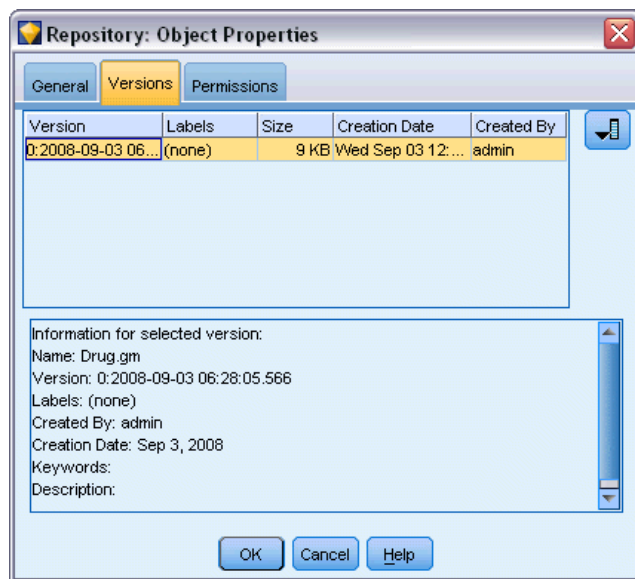
は、適切な権限を持つリポジトリ ユーザーによって設定されます（詳細は、『Deployment Manager ユーザー ガイド』を参照してください）。

キーワード： ストリーム、モデル、または出力オブジェクトの [注釈] タブでキーワードを指定します。複数のキーワードは、スペースで区切る必要があります。また、最大で 255 文字（半角）までです。（キーワードにスペースが含まれる場合は、キーワード間の区切りには疑問符を使用してください。）

[バージョン] タブ

IBM SPSS Collaboration and Deployment Services Repository に格納されたオブジェクトには、複数のバージョンがあります。[バージョン] タブは、各バージョンについての情報を表示します。

図 9-21
バージョンのプロパティ



格納されているオブジェクトの指定されたバージョンで、次のプロパティを指定または変更できます。

バージョン： バージョンの一意の識別子で、そのバージョンが格納された日時に基づいて生成されます。

ラベル： バージョンの現在のラベルです。バージョン識別子と異なり、ラベルはオブジェクトのあるバージョンから別のバージョンに移動できます。

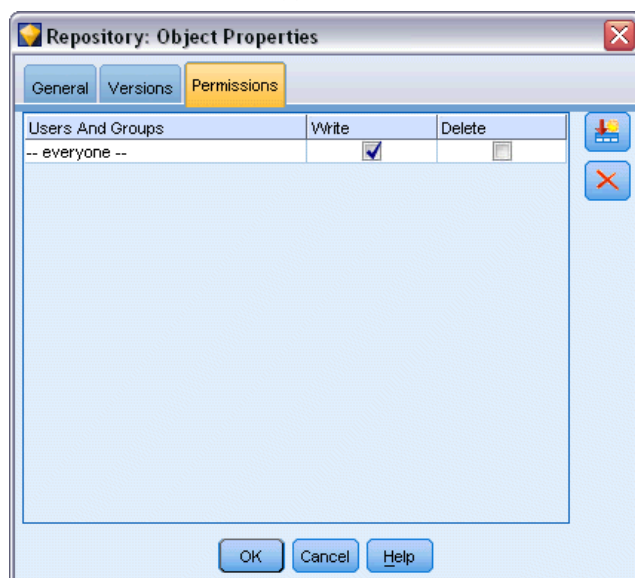
各バージョンで、ファイル サイズ、作成日時、および作成者も表示されます。

ラベルの編集： [バージョン] タブの右上にある[ラベルの編集] アイコンを使用して、格納されたオブジェクトのラベルを定義、適用または削除します。 詳細は、 [p.205 オブジェクトのバージョン ラベルの管理](#) を参照してください。

[権限] タブ

[権限] タブで、オブジェクトに対する読み取りと書き込みの権限を設定できます。現在のオブジェクトへのアクセスを持つすべてのユーザーとグループがリストされます。権限は階層によって異なります。たとえば、読み取り権限を持っていない場合、書き込み権限を持つことができません。書き込み権限を持っていない場合、削除権限を持つことはできません。

図 9-22
オブジェクトのアクセス権限



ユーザーとグループ： このオブジェクトに対して少なくとも読み取りアクセス権を持っているリポジトリ ユーザーおよびグループが表示されます。 [書き込み] チェック ボックスおよび [削除] チェック ボックスをオンにすると、このオブジェクトに対するこれらのアクセス権が、特定のユーザーまたはグループに割り当てられます。別のユーザーまたはグループにアクセスを割り当てるには、[権限] タブの右側にある [ユーザーとグループの追加] アイコンをクリックします。利用可能なユーザーとグループのリストは、管理者により管理されています。

オブジェクトのバージョン ラベルの管理

[バージョン ラベルの編集] ダイアログ ボックスで以下のことができます。

- 選択したオブジェクトへのラベルの適用
- 選択したオブジェクトからのラベルの削除
- 新しいラベルの定義およびオブジェクトへの適用

ラベルをオブジェクトに適用するには

- ▶ [使用できるラベル] リストで 1 つ以上のラベルを選択します。
- ▶ 右矢印のボタンをクリックして、選択したラベルを [適用されたラベル] リストに移動します。
- ▶ [OK] をクリックします。

オブジェクトからラベルを削除するには

- ▶ [適用されたラベル] リストで 1 つ以上のラベルを選択します。
- ▶ 左矢印のボタンをクリックして、選択したラベルを [使用できるラベル] リストに移動します。
- ▶ [OK] をクリックします。

新しいラベルを定義しオブジェクトに適用するには

- ▶ [新規ラベル] フィールドでラベル名を入力します。
- ▶ 右矢印のボタンをクリックして、新しいラベルを [適用されたラベル] リストに移動します。
- ▶ [OK] をクリックします。

ストリームの展開

ストリームをシンアプリケーションの IBM® SPSS® Modeler Advantage で使用できるようにするには、リポジトリ内でストリーム (.str ファイル) として展開する必要があります。

IBM® SPSS® Collaboration and Deployment Services のエンタープライズレベルの機能を利用するには、ストリームをリポジトリで展開する必要があります (.scn ファイル)。詳細は、[p.177 IBM SPSS Collaboration and Deployment Services Repository オブジェクトの保存と展開](#) を参照してください。

現在のストリームを展開するには ([ファイル] メニューを使用した方法)

- ▶ メイン メニューで次の各項目をクリックします。
File > 保存 > 展開

- ▶ 展開の種類を選択し、必要に応じてダイアログ ボックスの残りを入力します。
- ▶ [ストリームとして展開] をクリックして IBM SPSS Modeler Advantage で使用するためにストリームを展開するか、[シナリオとして展開] をクリックして、IBM SPSS Collaboration and Deployment Services で使用するためにストリームを展開します。
- ▶ [保存] ボタンをクリックします。詳細は、[ヘルプ] ボタンをクリックしてください。
- ▶ 「展開プロセスの完了」から先に進みます。

現在のストリームを展開するには ([ツール] メニューを使用した方法)

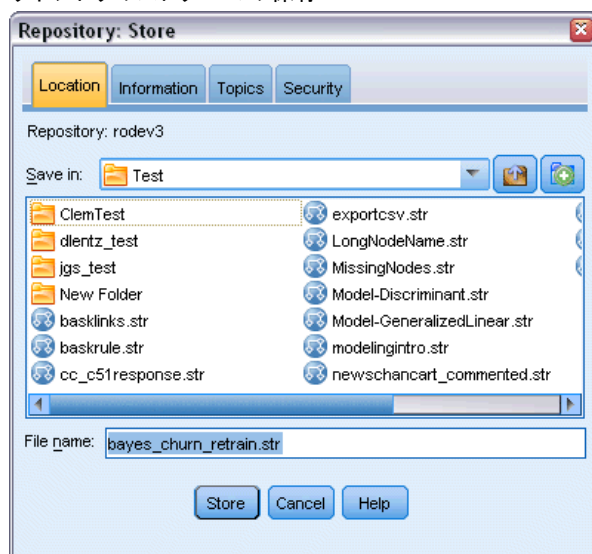
- ▶ メイン メニューで次の各項目をクリックします。
ツール > ストリームのプロパティ > Deployment
- ▶ 展開の種類を選択し、必要に応じて [展開] タブの残りを入力し、[保存] ボタンをクリックします。 [詳細は、 p. 208 ストリーム展開のオプション を参照してください。](#)

展開プロセスの完了

- ▶ 必要に応じて、リポジトリへの接続の設定を行います。 [詳細は、 p. 179 IBM SPSS Collaboration and Deployment Services Repository への接続 を参照してください。](#) 特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。

図 9-23

リポジトリのストリームの保存



- ▶ Repository の [保存] ダイアログ ボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。詳細は、[p.182 オブジェクトの特性の設定を参照してください](#)。

ストリーム展開のオプション

[ストリーム オプション] ダイアログ ボックスの [展開] タブを使用すると、ストリーム展開のオプションを指定できます。ストリームまたはシナリオとして展開できます。

ストリームとして展開すると、シンククライアント アプリケーション IBM® SPSS® Modeler Advantage でストリームを開いたり変更したりできます。ストリームは、拡張子 .str のファイルとしてリポジトリに保存されます。

シナリオとして展開すると、マルチユーザー アクセス、自動スコアリング、モデル リフレッシュ、Champion-Challenger 分析など、IBM® SPSS® Collaboration and Deployment Services で使用できる追加機能を利用できます。シナリオとして展開すると、リポジトリのストリームは拡張子 .scn のファイルとして保存されます。ストリームは Predictive Applications でも使用できるようになります。

このタブで、IBM® SPSS® Modeler がストリームに作成するストリームの説明をプレビューすることもできます。詳細は、[5 章 p.75 ストリームの説明を参照してください](#)。

注：エンタープライズ データへの一貫したアクセスを実現するために、シナリオとして展開されたストリームは IBM® SPSS® Collaboration and Deployment Services Enterprise View を使用してソース データにアクセスする必要があります。そのため、この場合は、指定された各スコアリングまたはモデル作成枝内に Enterprise View ビュー入力ノードが少なくとも 1 つ必要です。詳細は、[2 章 Enterprise View ノード in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノードを参照してください](#)。

図 9-24
ストリーム展開のオプション



展開タイプ: ストリームの展開方法を選択します。すべてのストリームは、シナリオとして展開する前に指定されたスコアリング ノードが必要です。追加の要件およびオプションは展開の種類によって異なります。

- **[<なし>]:** ストリームはリポジトリに展開されません。[ストリームの説明をプレビュー] を除くすべてのオプションが無効になります。
- **スコアリングのみ:** [保存] ボタンをクリックすると、ストリームがリポジトリに展開されます。[スコアリング ノード] フィールドで指定したノードを使用して、データをスコアリングできます。
- **モデルリフレッシュ:** [スコアリングのみ] と同じですが、[モデル作成ノード] フィールドおよび [モデルナゲット] フィールドで指定したオブジェクトを使用して、モデルをリポジトリで更新できます。

スコアリング ノード： グラフ、出力、エクスポート、またはエクスポート ノードを選択して、データのスコアリングに使用するストリーム枝を識別します。ストリームは実際いくつもの有効な枝、モデル、ターミナル ノードを含んでいますが、唯一のスコアリング枝が展開のために定義されている必要があります。ストリームを展開する最も基本的な要件です。

スコアリング パラメータ： パラメータを指定することができ、スコアリング枝が実行される際にそのパラメータを変更することができます。詳細は、[p. 211 スコアリングおよびモデル作成パラメータ](#) を参照してください。

モデル作成ノード： モデル リフレッシュについて、リポジトリのモデルを再生成または更新するために使用されるモデル作成ノードを指定します。[モデル ナゲット] に指定されたモデル ナゲットと同じ種類のモデル作成ノードである必要があります。

モデル構築のパラメータ： パラメータを指定することができ、モデル作成ノードが実行される際にそのパラメータを変更することができます。詳細は、[p. 211 スコアリングおよびモデル作成パラメータ](#) を参照してください。

モデル ナゲット： モデル リフレッシュについて、リポジトリでシナリオが更新されるたびに更新または再生成されるモデル ナゲットを指定します（通常はスケジュール化されたジョブの一環として）。モデルはスコアリング枝にある必要があります。複数のモデルがスコアリング枝に存在する場合がありますが、定義されるのは 1 つだけです。シナリオが最初に作成される場合、新規データが利用可能になると更新または再生成されるプレースホルダー モデルとなる場合があります。

ストリームとして展開： IBM SPSS Modeler Advantage でストリームを使用する場合に、このオプションをクリックします。

シナリオとして展開： IBM SPSS Collaboration and Deployment Services でストリームを使用する場合に、このオプションをクリックします（次の注意事項を参照）。

検査： 展開するために有効なストリームかどうかを確認します。シナリオとして展開する場合、ストリームには、有効な Enterprise View 入力ノードが含まれている必要があります。含まれていない場合、エラー メッセージが表示されます。

保存： ストリームが有効であれば、ストリームをシナリオとして展開します。有効でない場合、エラー メッセージが表示されます。[修正] ボタンをクリックしてエラーを修正し、もう一度実行してください。

ストリームの説明をプレビュー： SPSS Modeler がストリームに作成するストリームの説明の内容をプレビューすることができます。詳細は、[5 章 p. 75 ストリームの説明](#) を参照してください。

注：（シナリオとしての展開のみ）複数の Enterprise View ノードをモデリング枝内で使用することができます。その場合、枝内の全 エンタープライズ ノードに対する単一のデータ接続の使用が多くの場合推奨され、Champion-Challenger 分析に必要となります。

- Champion-Challenger サポートを必要としない場合、データ プロバイダ 定義 (DPD) によってのみ接続が異なる限り、様々な Enterprise View 接続を同じ枝内で使用することができます。
- これらの制限は、指定された枝内でのみ適用されます。スコアリングとモデル作成枝の間では、様々な Enterprise View 接続をそうした制限なしに使用することができます。

スコアリングおよびモデル作成パラメータ

ストリームを IBM® SPSS® Collaboration and Deployment Services に展開する場合、モデルが更新またはスコアリングされるごとに、どのパラメータを表示または編集するかを選択できます。たとえば、最大値および最小値、またはジョブが実行されるごとに変更する必要のあるその他の値を指定することができます。

図 9-25
[スコアリング パラメータ] ダイアログ ボックス



- ▶ シナリオの展開後に表示または編集できるようパラメータを表示するには、ダイアログ ボックスのリストから選択します。

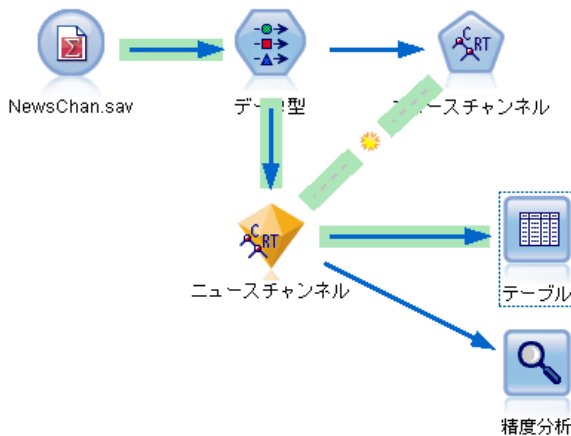
使用可能なパラメータのリストは、[ストリームのプロパティ] ダイアログボックスの [パラメータ] タブで定義されます。詳細は、5 章 p.66 [ストリームとセッション パラメータの設定](#) を参照してください。

スコアリング枝

ストリームを展開している場合、ストリームのいずれかの枝を**スコアリング枝**（スコアリング ノードを含む枝）として指定する必要があります。枝をスコアリング枝として指定すると、その枝はストリーム領域で強調表示され、スコアリング枝のナゲットへのモデル リンクとなります。この視覚的表示は、スコアリング枝がすぐに明らかになっていない、複数の枝による複雑なストリームで特に役立ちます。

注：スコアリング枝として指定できるストリーム枝は 1 つだけです。

図 9-26
スコアリング枝が強調表示されたストリーム



ストリームでスコアリング枝が既に定義されている場合、新しく枝を指定すると、それがスコアリング枝として置き換えられます。[カスタム カラー] オプションを使用して、スコアリング枝を示す色を設定できます。詳細は、12 章 p.261 表示オプションの設定を参照してください。

[ストリーム マークアップを表示/非表示] ツールバー ボタンを使用して、スコアリング枝の表示を表示または非表示にできます。

図 9-27
[ストリーム マークアップを表示/非表示] ツールバー ボタン



展開するスコアリング枝の特定

ターミナル ノードのポップアップ メニューまたは [ツール] メニューを使用して、スコアリング枝を指定できます。ポップアップ メニューを使用すると、スコアリング ノードがストリーム プロパティの [展開] タブで自動的に設定されます。

枝をスコアリング枝として指定するには (ポップアップ メニュー)

- ▶ モデル ナゲットをターミナル ノード (ナゲットから下流の処理ノードまたは出力ノード) に接続します。
- ▶ ターミナル ノードを右クリックします。
- ▶ メニューで、[スコアリング枝として使用] を選択します。

枝をスコアリング枝として指定するには ([ツール] メニュー)

- ▶ モデル ナゲットをターミナル ノード (ナゲットから下流の処理ノードまたは出力ノード) に接続します。
- ▶ メイン メニューで次の各項目をクリックします。
ツール > ストリームのプロパティ > Deployment
- ▶ [展開タイプ] リストで、必要に応じて [スコアリングのみ] または [モデルリフレッシュ] をクリックします。 [詳細は、 p.208 ストリーム展開のオプション を参照してください。](#)
- ▶ [スコアリングノード] フィールドをクリックして、リストからターミナル ノードを選択します。
- ▶ [OK] をクリックします。

モデル リフレッシュ

モデル リフレッシュは、新しいデータを使用するシナリオの既存のモデルを再構築するプロセスです。リポジトリのストリーム自体は変わりません。たとえば、アルゴリズムやストリーム固有の設定は同じままですが、モデルは新しいデータに保持され、新しいバージョンのモデルが古いモデルより優れている場合、更新されます。

リフレッシュに設定できるシナリオのモデル ナゲットは 1 つだけです。このモデルは、**リフレッシュ モデル**と呼ばれます。ストリームのプロパティの [展開] タブで [モデルリフレッシュ] オプションをクリックすると (「ストリーム展開のオプション」 (p.208) 参照)、そのとき指定したモデル ナゲットがリフレッシュ モデルとなります。モデル ナゲットのポップアップ

メニューからモデルをリフレッシュ モデルに指定することもできます。指定するには、ナゲットがあらかじめスコアリング枝にある必要があります。

ナゲットの「リフレッシュ モデル」状態をオフにすると、シナリオの展開タイプを [スコアリングのみ] に設定することとなり、ストーリーのプロパティのダイアログの [展開] タブがそれに従って更新されます。現在のスコアリング枝にあるナゲットのポップアップ メニューで [リフレッシュ モデルとして使用] オプションを使用して、この状態をオンにしたりオフにしたりすることができます。

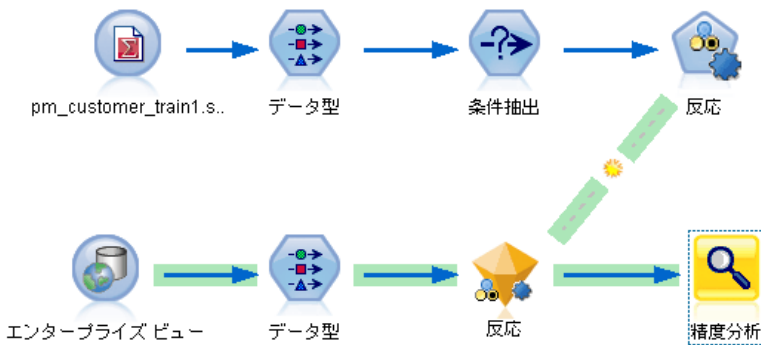
スコアリング枝のナゲットのモデル リンクを削除すると、ナゲットの「リフレッシュ モデル」状態も削除されます。[編集] メニューまたはツールバーを使用して、モデル リンクの削除を取り消すことができます。取り消した場合、ナゲットの「リフレッシュ モデル」の状態も元に戻ります。

リフレッシュ モデルの選択方法

スコアリング枝のほか、リフレッシュ モデルへのリンクもストリーム内で強調表示されます。リフレッシュ モデルに選択されるモデル ナゲット、つまり強調表示されるリンクは、ストリーム内のナゲットの数によって異なります。

ストリームで 1 つのモデル

図 9-28
ストリームに 1 つのモデルがあるスコアリング枝



リンクしたモデルがスコアリングに 1 つあり、そのように特定されている場合、そのナゲットがシナリオのリフレッシュ モデルとなります。

ストリームに複数のモデル

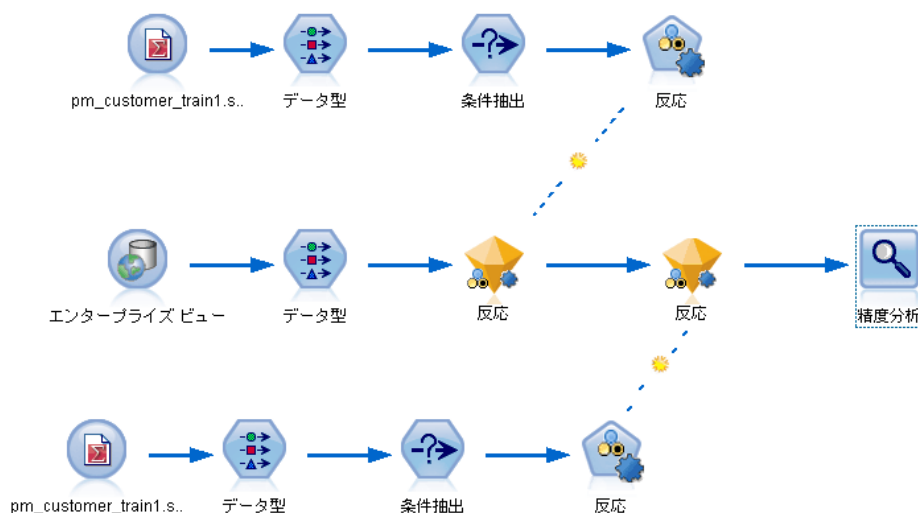
ストリームにリンクしたナゲットが複数ある場合、次のようにしてリフレッシュ モデルを選択します。

[ストリームのプロパティ] ダイアログ ボックスの [展開] タブでモデルナゲットが定義されており、それがストリーム内にある場合、そのナゲットがリフレッシュ モデルになります。

[展開] タブでナゲットが定義されていない場合、またはナゲットが定義されているがスコアリング枝にない場合、ターミナル ノードに最も近いナゲットがリフレッシュ モデルとなります。

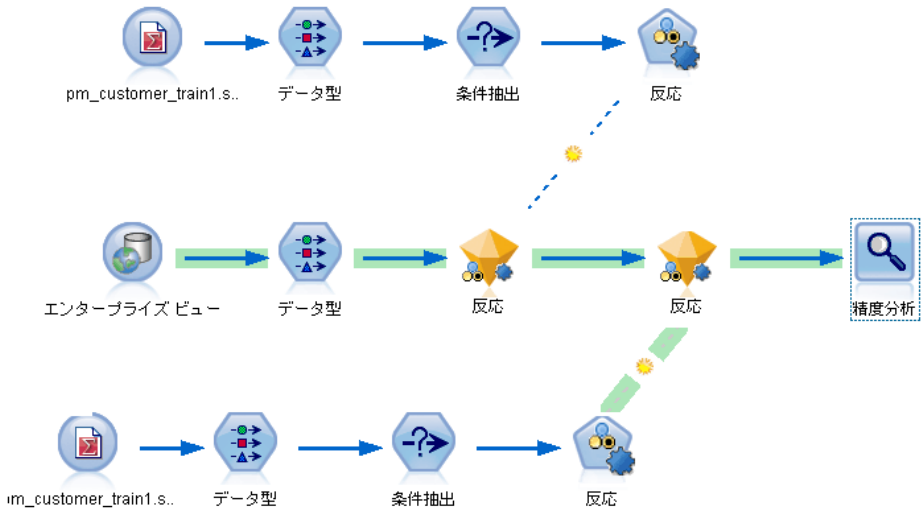
このことを説明するために、次のストリームがあると想定します。

図 9-29
ストリームに複数のモデルがあるスコアリング枝



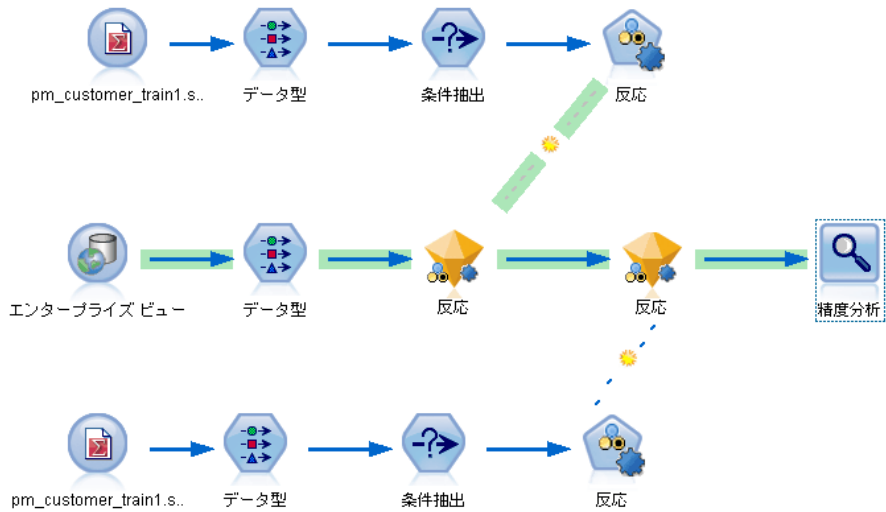
精度分析ノードを右クリックしてそのメニューを使用し、スコアリング枝を設定すると、その枝が強調表示されます。また、スコアリング枝を設定すると、精度分析ノードに最も近いモデルがリフレッシュ モデルに指定され、次のように強調表示されたリフレッシュ リンクで示されます。

図 9-30
複数のモデルおよびリフレッシュリンクを持つ強調表示されたスコアリング枝



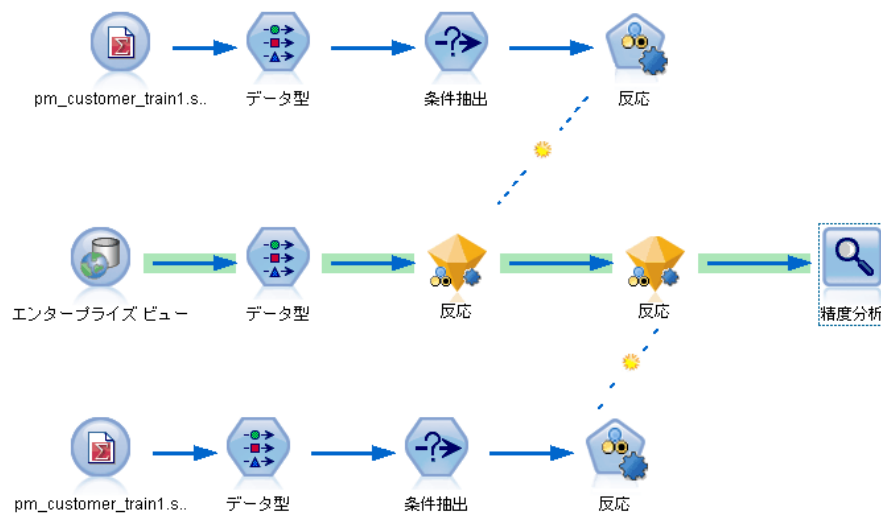
ただし、ストリーム内の他のモデル ナゲットをリフレッシュ モデルとして使用したい場合にメニューから設定すると、そのモデル リンクがリフレッシュ リンクとして使用されるよう設定されます。

図 9-31
リフレッシュリンクが最初のモデル ナゲットに切り替えられたスコアリング枝



リフレッシュ リンクである 2 つのモデル リンクを選択を解除すると、スコアリング枝のみが強調表示され、リンクは強調表示されなくなります。展開タイプは [スコアリングのみ] に設定されます。

図 9-32
複数のモデルがありリフレッシュ リンクがないスコアリング枝



注：一方のリンクを置換え状態に設定できますが、もう一方は置き換えることができません。この場合、リフレッシュ モデルに選択されているモデル ナゲットは、スコアリング枝が指定されているときにリフレッシュ リンクがあり、ターミナル ノードに最も近いモデル ナゲットです。

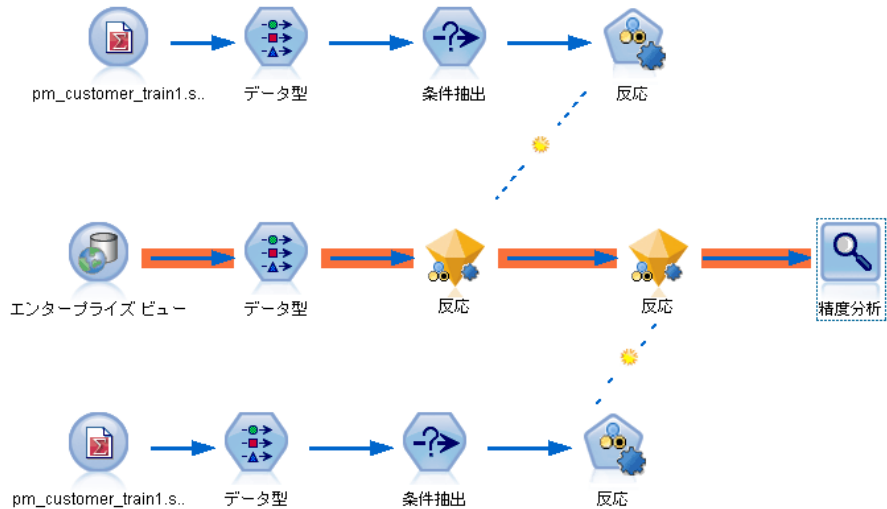
ストリームにモデルがない

ストリームにモデルがない場合、またはモデル リンクのないモデルだけがある場合、展開タイプは [スコアリングのみ] に設定されます。

エラーのあるスコアリング枝のチェック

スコアリング枝を指定すると、シナリオとして展開しているときストリーム内に Enterprise View ノードがないなど、エラーの有無がチェックされます。

図 9-33
エラーのあるスコアリング枝



エラーが見つかったら、スコアリング枝がスコアリング枝エラーの色で強調表示され、エラーメッセージが表示されます。[カスタム カラー] オプションを使用して、エラーの色を設定できます。詳細は、12 章 p.261 表示オプションの設定を参照してください。

エラーが見つかった場合、次の手順を実行してください。

- ▶ エラーメッセージの内容に従って、エラーを修正します。
- ▶ 次のいずれかを実行します。
 - ターミナル ノードを右クリックし、ポップアップメニューで [シナリオを確認] をクリックします。
 - メインメニューで次の各項目をクリックします。
ツール > ストリームのプロパティ > Deployment

そして、[確認] ボタンをクリックします。
- ▶ 必要に応じて、エラーがなくなるまでこのプロセスを繰り返します。

外部アプリケーションへのエクスポート

外部アプリケーションへのエクスポートについて

IBM® SPSS® Modeler では、データマイニング プロセス全体を、外部のアプリケーションにエクスポートするためのいくつかの機構を備えており、データの準備作業を行ってから、SPSS Modeler の外部で、ユーザー独自のアプリケーションを使用してモデルを構築できます。

前の項では、マルチユーザー アクセス、ジョブ スケジュール、その他の機能を利用するための IBM SPSS Collaboration and Deployment Services リポジトリへのストリームの展開方法について説明しました。同じように、SPSS Modeler ストリームも次のアプリケーションと組み合わせて使用できます。

- IBM® SPSS® Modeler Advantage
- Predictive Applications 5.0 アプリケーション
- Predictive Applications 4.x アプリケーション
- PMML 形式のファイルをインポートおよびエクスポートできるアプリケーション

IBM SPSS Modeler Advantage でのストリーム使用の詳細は、「IBM SPSS Modeler Advantage でストリームを開く」(p.220)のを参照してください。

Predictive Applications 5.0 で使用するストリームをエクスポートするには、指示に従ってシナリオとして展開してください。 [詳細は、9 章 p.206 ストリームの展開 を参照してください。](#)

4.x バージョンの Predictive Applications で展開するストリームのパッケージの詳細は、「Predictive Applications 4.x ウィザード」(p.220)を参照してください。

PMML 形式をサポートするアプリケーションとモデルを共有するための、PMML ファイルとしてのモデルのエクスポートおよびインポートの詳細は、「PMML としてのモデルのインポートおよびエクスポート」(p.230)を参照してください。

注：Predictive Applications 製品は、IBM® SPSS® Decision Management に置き換えられました。Predictive Applications のサポートは、SPSS Modeler の今後のリリースでは廃止されます。

IBM SPSS Modeler Advantage でストリームを開く

IBM® SPSS® Modeler ストリームは、シンクライアント アプリケーション IBM® SPSS® Modeler Advantage と組み合わせて使用できます。IBM SPSS Modeler Advantage 全体でカスタマイズされたアプリケーションを作成できますが、アプリケーション ワークフローの基本として、SPSS Modeler で作成されたストリームを使用することもできます。

IBM SPSS Modeler Advantage でストリームを開く手順は次のとおりです。

- ▶ [ストリームとして展開] オプションをクリックして、IBM® SPSS® Collaboration and Deployment Services リポジトリでストリームを展開します。 [詳細は、9 章 p.206 ストリームの展開 を参照してください。](#)

- ▶ [IBM SPSS Modeler Advantage で開く] ツールバー ボタンをクリックするか、メイン メニューで次の項目をクリックします。

File > IBM SPSS Modeler Advantage で開く

- ▶ 必要に応じて、リポジトリへの接続の設定を行います。 [詳細は、9 章 p.179 IBM SPSS Collaboration and Deployment Services Repository への接続 を参照してください。](#) 特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。

注：リポジトリ サーバーにも、IBM SPSS Modeler Advantage ソフトウェアをインストールする必要があります。

- ▶ Repository の [保存] ダイアログ ボックスで、オブジェクトを保存したいフォルダを選択し、記録したいその他の情報を指定して [保存] ボタンをクリックします。 [詳細は、9 章 p.182 オブジェクトの特性の設定 を参照してください。](#)

インストールすると、ストリームがすでに開いている IBM SPSS Modeler Advantage を起動します。SPSS Modeler のストリームは閉じます。

Predictive Applications 4.x ウィザード

注：Predictive Applications 製品は、IBM® SPSS® Decision Management に置き換えられました。Predictive Applications のサポートは、IBM® SPSS® Modeler の今後のリリースでは廃止されます。

Predictive Applications 4.x ウィザードを使用すると、PredictiveMarketing 2.0 以降を含む Predictive Applications バージョン 4.x の展開用に、SPSS Modeler からストリームをパッケージ化できます。SPSS Modeler で行ったすべてのデータ操作やモデル作成作業を、スコアリング ソリューションとしてパッケージ化して保存することができます。展開した後は、アプリケーションを使用して、SPSS Modeler モデルをキャンペーン用のソリューションに組み込むことができますようになります。

たとえば、データ アナリストが、顧客をセグメント化して silver、gold、platinum の応用価値モデルを開発し、それらを PredictiveMarketing やその他のアプリケーションで使用したいと考える場合があります。最初に SPSS Modeler 内でモデルを開発してから、PredictiveMarketing で使用するためにそのモデルをエクスポートできます。PredictiveMarketing は、SPSS Modeler モデルを内部で使用して、最高の値のオファーを提案します。

ウィザードにアクセスするには

- ▶ メイン メニューで次の各項目をクリックします。
ツール > Predictive Applications 4.x ウィザード

Predictive Applications ウィザードを使用する前に

これ以降は、IBM® SPSS® Modeler とその他の予測アプリケーションとの統合を説明する情報です。ウィザードを使ってストリームをエクスポートする前に、統合と公開の前提条件をレビューします。

統合の仕組み

通常の状態であれば、SPSS Modeler を使用して、データ マイニングを行い、予測アプリケーションが使用できるように配置機能を拡張できるようにします。

- ▶ 現在使用している予測アプリケーションで始めます。Customer View Builder のオプションを使用すると、Unified Customer View (UCV) データモデルを XML ファイルとしてエクスポートすることができます。エクスポートした XML ファイルを SPSS Modeler で活用するときのために、ファイルの場所を覚えておいてください。
- ▶ 次に、SPSS Modeler の入力ノードの設定を行い、UCV 用のフィールド（既にエクスポートした XML ファイルに表示されています）を含むすべてのデータ ソース（データベース、フラット ファイルなど）にアクセスします。UCV が参照するすべてのフィールドを含ませることも、作成中のモデルに必要な UCV の一部だけを使用することもできます。通常、モデル作成用のデータにアクセスするには、SPSS Modeler の入力ノードをいくつか使用します。

- ▶ SPSS Modeler を使用して、データ マイニング タスクのために必要なデータの結合、変換、および導出のすべてを行います。
- ▶ データ型ノードが含まれ、そのノードに UCV という名前が付いていることをストリーム内のある時点で確認してください。この UCV データ型ノードは、モデル作成の際の方向性を決めるためだけでなく、データが XML ファイルで定義されたフィールド情報と一致していることを確認するためにも使用されます。データ型ノードの設定と、既に生成した XML ファイルの属性仕様とを比較してみると良いでしょう。 [詳細は、 p.224 ステップ 3 :UCV ノードの選択 を参照してください。](#)
- ▶ 次に、SPSS Modeler を使って生成しているモデルのタイプについて検討します。ニューラル ネットワークなどの値モデルを開発している場合には、2 値の予測（たとえば解約の真/偽）を傾向としてエクスポートしたくなるかもしれません。それにより、アプリケーションによって生成されたモデルからの予測に対して比較することが可能になります。 [詳細は、 p.222 2 値の予測を傾向スコアとしてエクスポート を参照してください。](#)
- ▶ そのモデルに満足し、確信が傾向となった場合、ストリームの展開枝にターミナル ノードを追加します。テーブル ノードを使う人が多いのですが、この場合はターミナル ノードで十分です。外部のアプリケーションで表示するフィールドだけをターミナル ノードで表示するようにしてください。つまり、このターミナル ノードの前に、展開しないフィールドの除外が必要な場合があります。
- ▶ また、モデルにより生成された予測フィールドが、エクスポートされる前にインスタンス化されていることを確認してください。必要な場合、これは、生成されたモデルと、末端の「エクスポート」ノードとの間にデータ型ノードを追加することで実現できます。
- ▶ ウィザード使用前の最後のステップとして、展開のためのストリームが準備できたことを、テストの実行により確認してください。

ストリームを展開するための準備ができました。SPSS Modeler の [ツール] メニューから、Predictive Applications ウィザードにアクセスできます。リアルタイム環境での公開に必要なストリーム情報とメタデータを含む Modeler Deployment Package (.cdp) を生成するためには、本ドキュメントで説明するウィザード手順にしたがってください。

2 値の予測を傾向スコアとしてエクスポート

Predictive Applications ウィザードでエクスポートする前に、使用している IBM® SPSS® Modeler モデルのスコアリング出力が、予測アプリケーションが生成した予測と矛盾しないかどうかを検討します。多くの場合、2 値の予測を複数のモデル間で予測の強度を正確に比較することができる傾向スコアとしてエクスポートする必要があります。

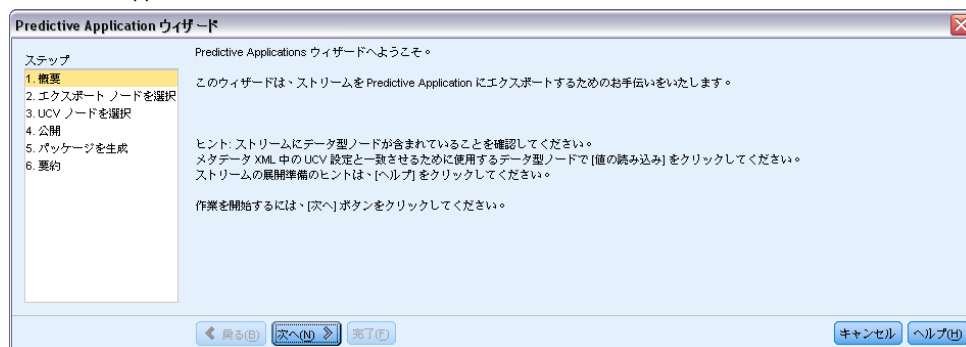
- **値モデル**は、特定の「はい」または「いいえ」の尤度をランク付けする各レコード単独の**傾向スコア**を 0.0 から 1.0 までのスケールで作成します。たとえば、解約モデルは、1.0（解約しやすい）から 0.0（解約しにくい）間のスコアを作成します。傾向スコアは確率ではないので、0.5 のスコアは必ずしも 50% 解約しやすい、または 0.25 のスコアを持つ人の 2 倍も解約しやすいわけではありません。しかし、0.4 のスコアを持つ誰かより「もっと解約しやすい」ことを意味します。傾向スコアはランク付けに使用することができます。たとえば、最も解約しやすい 10 % の顧客を発見することに使用できます。
- **オファーマodel**は予測や確信度の値を生成し、それがアプリケーションに送信されると、UCV の「仮想属性」として使用されます。その値は、数値か連続する範囲の値です。

SPSS Modeler モデルの多くは、ニューラル ネットワーク、ディジジョン ツリー、ロジスティック回帰モデルなどを含む 2 値の予測を生成することができます。値モデルを展開する場合、エクスポートの前に傾向スコアのエクスポートが必要な場合があります。傾向スコアは、モデル作成ノードの [精度分析] タブまたは生成されたモデル ナゲットの [設定] タブで有効化することができます。詳細は、[3 章 モデル作成ノードの分析オプション in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。

ステップ 1 :Predictive Applications ウィザードの概要

初めて Predictive Application ウィザードを開く場合、必要なストリーム コンポーネントを指定するために必要な指示を記載した、初期画面が表示されます。

図 10-1
Predictive Applications ウィザード



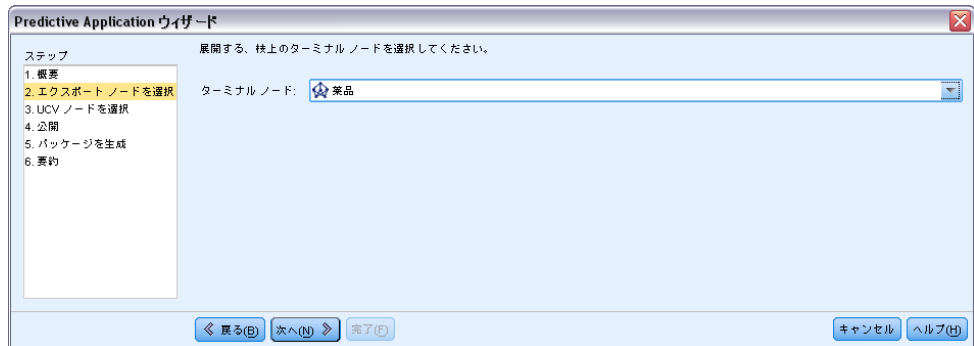
ウィザードの残りを処理すると、展開のためのパッケージ生成のプロセスが終了し、リアルタイム環境に移ります。作業を開始する前に、前述の前提条件を参照し、ストリームの展開準備ができていることを確認して

ください。詳細は、[p.221 Predictive Applications ウィザード](#)を使用する前に [を参照してください](#)。

ステップ 2 :ターミナル ノードの選択

ステップ 2 では、ストリームからターミナル ノードを指定することができ、それはスコアリング枝あるいはスコアリングが発生するストリームの一部を表しています。

図 10-2
ターミナル ノードの選択



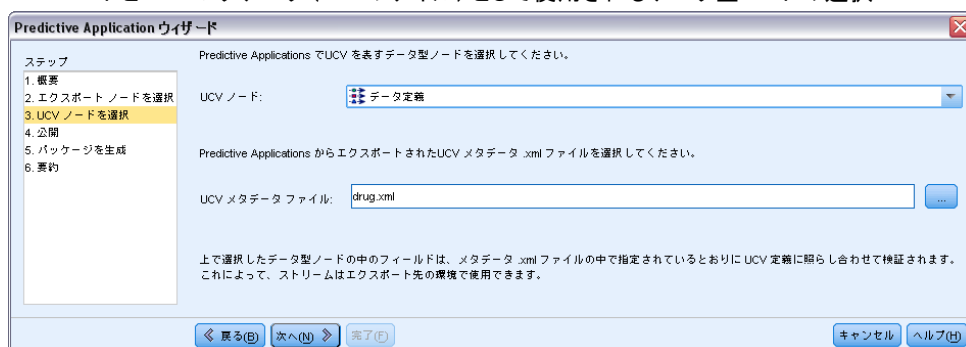
ウィザードのリストには、ストリーム中の各ターミナル ノードの名前とノードの種類しか表示されないため、同じ名前を持つターミナル ノードを区別することが重要になります。混乱を避けるために、ストリーム中のターミナル ノードが一意の名前を持つようにしてください。

また、アプリケーション環境で表示させたいフィールドだけをターミナル ノードで表示させるようにしてください。つまり、このターミナル ノードの前に、展開しないフィールドの除外が必要な場合があります。

ステップ 3 :UCV ノードの選択

ステップ 3 では、発行のための 2 つの重要情報、UCV ノードと UCF メタデータ ファイルを指定する必要があります。

図 10-3
UCV ノードと UCV メタデータ (XML ファイル) として使用されるデータ型ノードの選択



UCV ノード: UCV ノードは、ストリームからのデータ型ノードであり、すべてのデータが Unified Customer View (UCV) の定義に一致していることを確認するために使用されます。こうした仕様は、XML ファイル (Customer View Builder からエクスポート済み) に保存されます。[次へ] をクリックすると、データ型ノードにおける設定が、ここで指定した XML ファイルの内容に対して自動的にチェックされます。その際には、次の [UCV メタデータ ファイル] オプションが使われます。

UCV メタデータ ファイル: UCV メタデータ ファイルは、Customer View Builder から既已取得済みの XML ファイルです。ここで選択した XML ファイルは、アプリケーションへ展開するために必要なデータ属性を含んでいます。

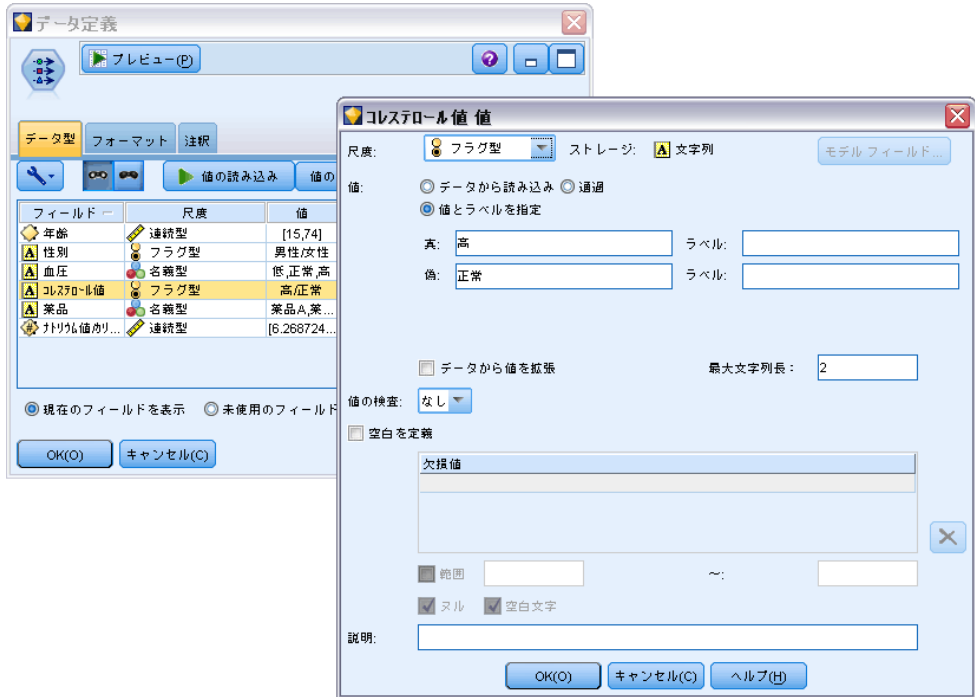
データ不一致エラー

ウィザードがデータ不一致エラーを表示した場合、IBM® SPSS® Modeler ストリームまで戻り、データ型ノードの仕様を確認します。データ型ノードのフィールド情報と Customer View Builder から生成した XML ファイルにあるフィールド情報とを比較します (XML ファイルは、Notepad などのテキスト ブラウザで開くことができます)。データ型ノードの設定は、UCV 属性と一致していますか？ たとえば、XML ファイルが、[コレステロール] という名前のフィールドが UCV に必要で、その文字列の値を次のように示しているとします。

```
<UcvAttribute Name="Cholesterol" Domain="String"/>
```

SPSS Modeler のストリームで、データ型ノードの設定をチェックして、リアルタイム環境に有益なフィールドのみが SPSS Modeler からエクスポートされていることを確認してください。

図 10-4
データ型ノードの設定と [値] サブダイアログ ボックスの記憶域の確認



記憶域をドメインへマッピング

データ型ノードの設定と UCV 仕様との間のマッピングがはっきりしない場合があります。この統合において、UCV XML ファイルで定義されたドメインは、SPSS Modeler の記憶域と等価です。これら各々についてのフィールド設定値は、エクスポート中に対応付けがなされます。

テーブル 10-1
記憶域とドメインのマッピング

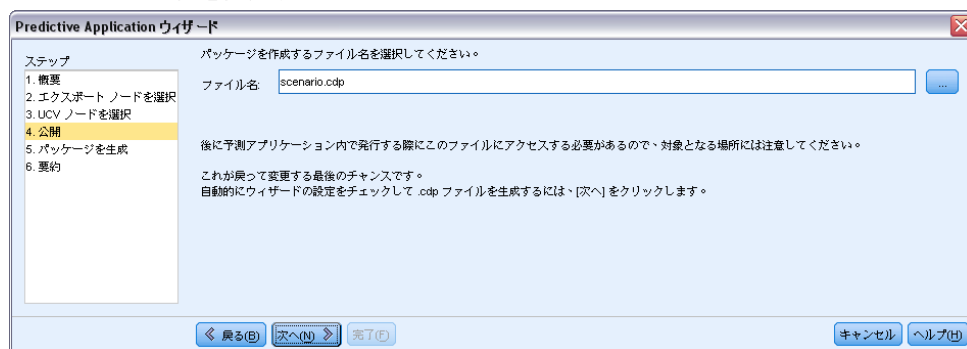
SPSS Modeler ストレージ	アプリケーション ドメイン
String	String 文字 Bit
Real	Float Double 小数
整数	Long 整数
Date	Date
Timestamp	Timestamp
Time	なし

ストレージ タイプは、フィールド作成ノードまたはフィルタ ノードで `to_integer` などの変換関数を使って変更できます。詳細は、4 章 置換ノードを使ったストレージの変換 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード を参照してください。

ステップ 4 :パッケージの指定

ステップ 4 では、ストリーム パッケージを保存する場所とその名前を決める必要があります。このステップを完了すると、多数のオブジェクトが **Modeler Deployment Package** という名前で、拡張子が `.cdp` となるバンドルとして保存されます

図 10-5
パッケージの名前を指定する

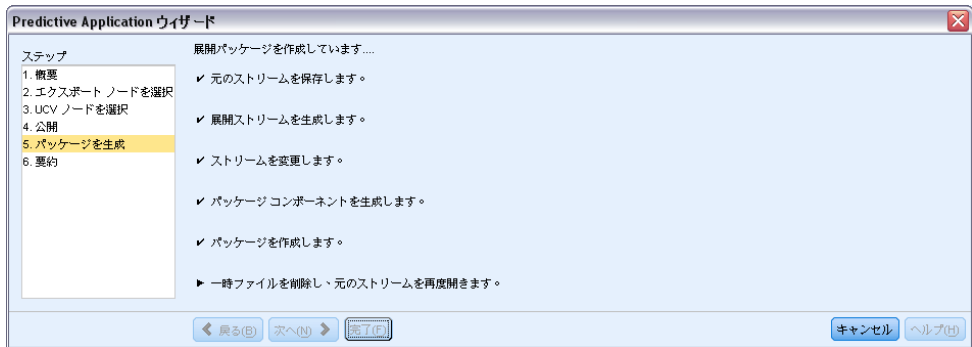


[次] をクリックして、ストリームのメタデータと仕様を自動的にチェックします。すべての仕様が正しいことがチェックされると、`.cdp` ファイルが生成されます。後でアプリケーションを発行する際に、このファイルにアクセスする必要がありますので、ターゲットの場所には注意してください。

ステップ 5 :パッケージの生成

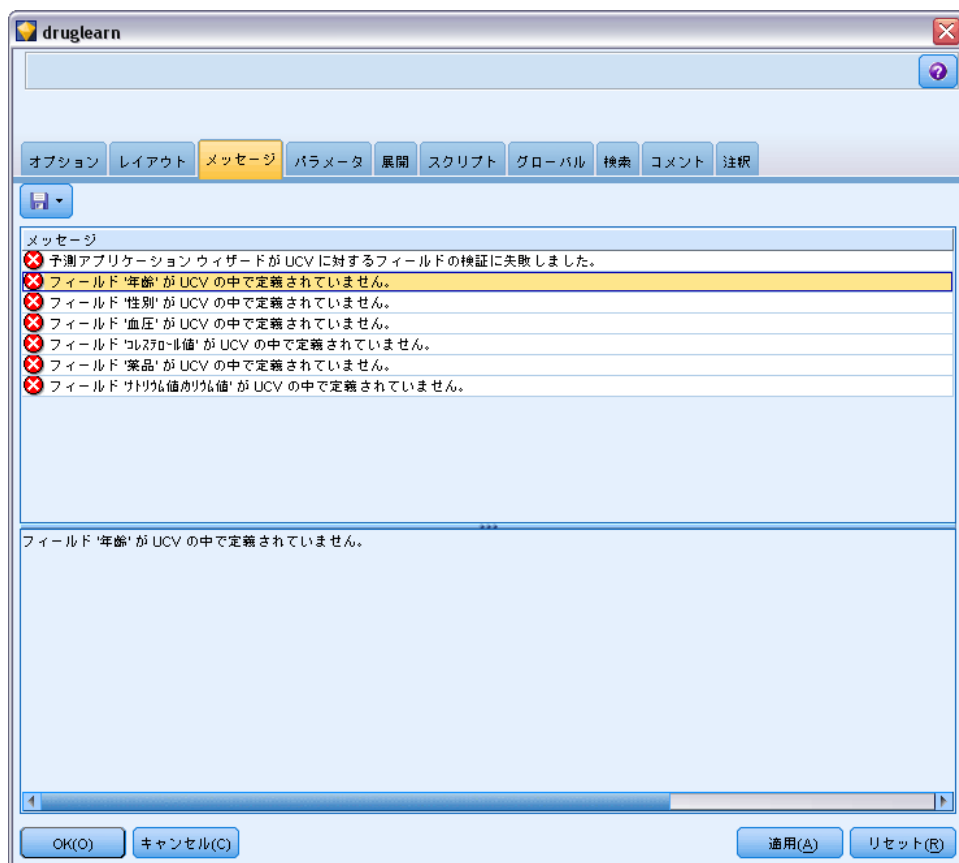
この段階で、ウィザードはストリームの情報と前の画面で作成した仕様を自動的にチェックします。

図 10-6
展開パッケージの作成



IBM® SPSS® Modeler のストリームからのフィールドの名前と型は、UCV から生成された XML ファイルを使って検証されます。フィールド情報が一致しない場合、ウィザードは自動的に終了し、対応するエラーメッセージを表示します。

図 10-7
メタデータ不一致によるエラー メッセージ



エラーが発生すると、SPSS Modeler ストリームに戻り、次のチェックを行います。

- UCV ノード（ストリームのデータ型ノード）のすべてのフィールド名が、UCV を定義する XML ファイルの中に存在することを確認します。さらに、その型（ドメイン）に矛盾がないことを確認します。
- SPSS Modeler とその他のアプリケーションはいずれも、フィールド名の大文字と小文字を区別することに注意してください。これに対して、フィールドの順番は重要ではありません。

注：生成された.cdp パッケージの大きさは、5 KB より小さい必要があります。特に、生成されたフィールドのリストが長い場合、制限に収まるようにそれらのいくつかを説明から削除する必要があります。

ステップ 6 :要約

IBM® SPSS® Modeler Deployment Package が問題なく生成されれば、SPSS Modeler を使った作業は終了です。

次のステップへ

次に、モデル (.cdp 拡張子のパッケージ ファイル) を Interaction Builder へインポートします。

PMML としてのモデルのインポートおよびエクスポート

PMML (Predictive Model Markup Language) は、モデルへの入力、データマイニングのデータの準備に使用する返還、モデル自体を定義するパラメータなど、データマイニングおよび統計モデルを説明する XML 形式です。IBM® SPSS® Modeler は PMML をインポートおよびエクスポートし、IBM® SPSS® Statistics など、この形式をサポートする他のアプリケーションとモデルを共有できるようにします。

注：PMML をエクスポートするには、IBM® SPSS® Modeler Server のライセンスが必要です。

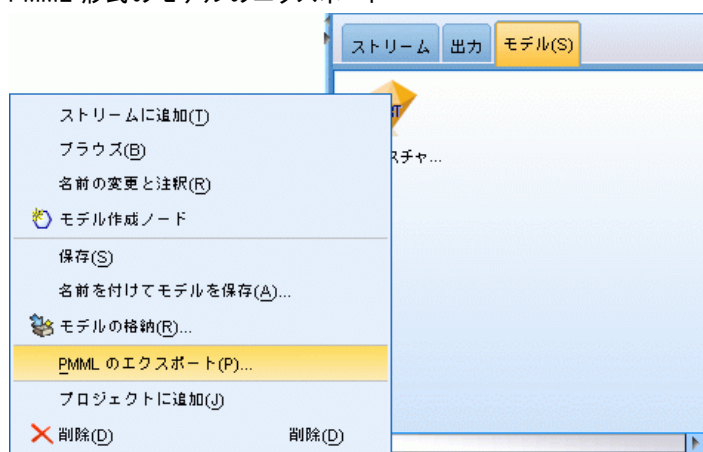
PMML の詳細は、データマイニンググループの Web サイト (<http://www.dmg.org>) を参照してください。

モデルをエクスポートするには

PMML エクスポートでは、SPSS Modeler 内で生成されたほとんどの種類のモデルがサポートされます。詳細は、[p. 233 PMML をサポートするモデルの種類](#) を参照してください。

- ▶ モデルパレットのモデルナゲットを右クリックします(または、キャンバス上のモデルナゲットをダブルクリックして、[ファイル]メニューを選択します)。
- ▶ メニューで、[PMMLのエクスポート] をクリックします。

図 10-8
PMML 形式のモデルのエクスポート



- ▶ [エクスポート] (または [保存]) ダイアログ ボックスで、対象ディレクトリとモデルの一意的な名前を指定します。

注： [ユーザー オプション] ダイアログ ボックスで、PMML エクスポートのオプションを変更できます。メイン メニューで次の各項目をクリックします。

ツール > オプション > [ユーザー オプション]

そこで [PMML] タブをクリックします。

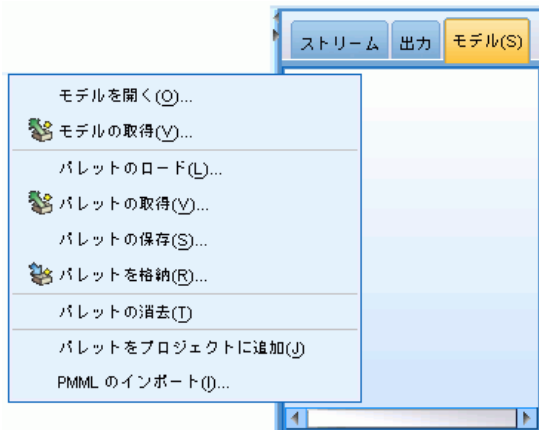
詳細は、12 章 p.266 PMML エクスポート オプション の設定 を参照してください。

PMML として保存されたモデルをインポートするには

SPSS Modeler または別のアプリケーションから PMML としてエクスポートされたモデルは、生成済みモデル パレットへインポートできます。詳細は、p.233 PMML をサポートするモデルの種類 を参照してください。

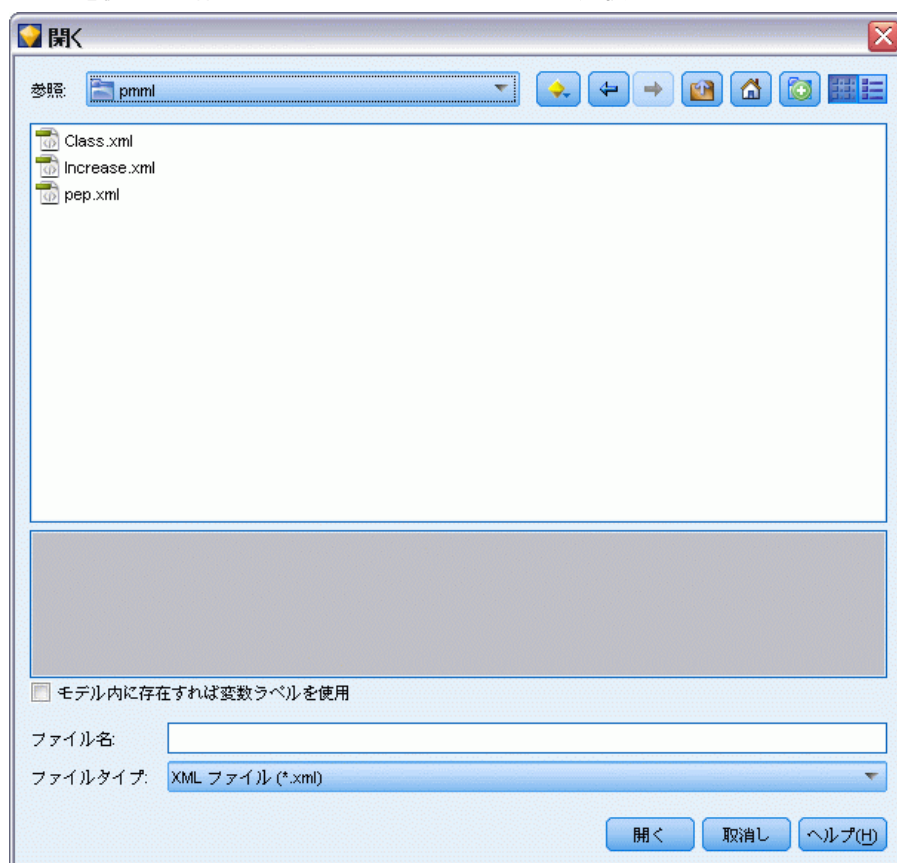
- ▶ モデル パレット内で、パレットを右クリックし、メニューから [PMML のインポート] を選択します。

図 10-9
PMML フォーマットでのモデルのインポート



- ▶ インポートするファイルを選択し、必要に応じて、変数のラベルに関するオプションを指定します。
- ▶ [開く] をクリックします。

図 10-10
PMML を使用して保存されたモデルの XML ファイルの選択



モデル内に存在すれば変数ラベルを使用：PMML が、データ ディクショナリ内の変数に対して、変数名と変数ラベル（RefID に対する Referrer ID など）の両方を指定している場合があります。元のエクスポートされた PMML に変数ラベルが存在するときに変数ラベルを使用するには、このオプションを選択します。

変数ラベル オプションを選択したにもかかわらず、PMML 内に変数ラベルがない場合、変数名は通常のように使用されます。

PMML をサポートするモデルの種類

PMML のエクスポート

SPSS Modeler モデル :IBM® SPSS® Modeler で作成された次のモデルは、PMML 4.0 としてエクスポートされます。

- C&R Tree

- QUEST
- CHAID
- 線型回帰
- ニューラル ネットワーク
- C5.0
- ロジスティック回帰
- 一般化線型
- SVM
- ベイズ ネットワーク
- Apriori
- Carma
- シーケンス
- K-Means
- Kohonen
- TwoStep
- KNN
- Statistics モデル

SPSS Modeler で作成された次のモデルは、PMML 3.2 としてエクスポートされます。

- デシジョン リスト

データベース固有モデル: データベース固有のアルゴリズムを使用して作成されたモデルの場合、PMML エクスポートは、IBM InfoSphere Warehouse モデルのみ利用可能です。Microsoft または Oracle Data Miner の Analysis Services を使用して作成されたモデルをエクスポートすることはできません。また、PMML としてエクスポートされた IBM モデルを、SPSS Modeler に再度インポートすることはできません。詳細は、[2 章 データベースモデル作成の概要 in IBM SPSS Modeler 14.2 データベース内マイニング ガイド](#) を参照してください。

PMML のインポート

SPSS Modeler では、すべての IBM® SPSS® Statistics 製品の現在のバージョンで作成された PMML モデルをインポートおよびスコアリングできます。SPSS Statistics 17.0 で生成されたモデルまたは変換 PMML と同様に、SPSS Modeler からエクスポートされたモデルもインポートおよびスコアリングできます。基本的には、次の例外を除いて、スコアリング エンジンはすべての PMML をスコアリングできます。

- Apriori、CARMA および異常値検出モデルをインポートすることはできません。
- スコアリングに使用できる場合でも、SPSS Modeler へのインポート後に PMML をブラウズすることはできません。(これには、初めに SPSS Modeler からエクスポートされたモデルは含まれません。この制限を回避するには、PMML ではなく一般化されたモデル [* .gm] としてエクスポートします。)
- PMML としてエクスポートされた IBM InfoSphere Warehouse モデルを、インポートすることはできません。
- インポート時には制限つき検証が行われますが、モデルのスコアリング試行時には完全検証が実行されます。そのため、インポートは正常に行われますが、スコアリングは失敗したり不正な結果が生成されます。

プロジェクトとレポート

プロジェクトの概要

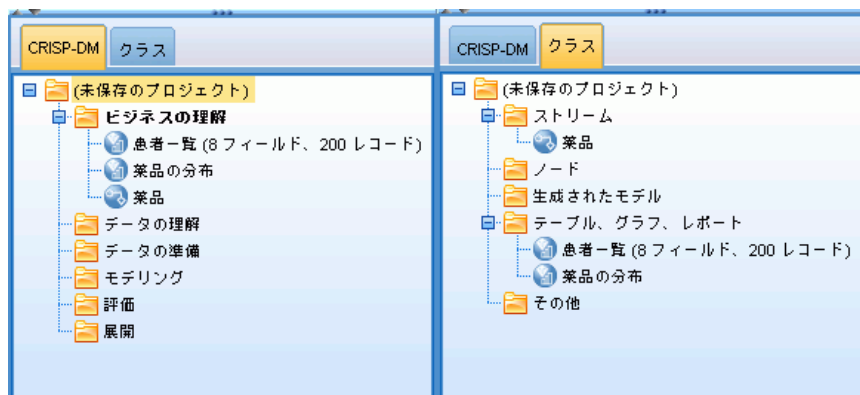
プロジェクトは、データマイニング作業に関連したファイルのグループです。プロジェクトには、データストリーム、グラフ、生成されたモデル、レポート、およびその他 IBM® SPSS® Modeler で作成したすべてが含まれます。SPSS Modeler プロジェクトの第一印象は、単に出力を編成するための手段のように思われますが、それ以外にもさまざまな用途に利用することができます。プロジェクトを使用して、次のような処理ができます。

- プロジェクトファイル内の各オブジェクトに注釈を付ける。
- CRISP-DM 手法を使用してデータマイニング作業のガイド役をする。プロジェクトには、詳細な説明と CRISP-DM によるデータマイニング作業の実例を含んだ、CRISP-DM ヘルプシステムも用意されています。
- データマイニングの目標をプレゼンテーションするための PowerPoint スライドショーや、使用するアルゴリズムに関する白書などの、SPSS Modeler 以外のオブジェクトを追加する。
- 注釈に基づいて総合レポート、および簡単な更新レポートの両方を作成する。これらのレポートは、社内イントラネットなどに公開しやすい HTML 形式で生成することができます。

注：プロジェクトパネルが SPSS Modeler ウィンドウに表示されていない場合は、[表示]メニューで [プロジェクト] をクリックします。

プロジェクトに追加したオブジェクトは、次の 2 種類の方法で表示することができます。クラスビューと CRISP-DM ビュー。プロジェクトに追加したものはすべて両方のビューに追加されます。2 つのビューを切り替えながら作業を行って、最適な編成を作成できます。

図 11-1
プロジェクトファイルの CRISP-DM ビューとクラス ビュー



CRISP-DM ビュー

CRISP-DM (Cross-Industry Standard Process for Data Mining) をサポートすることで、IBM® SPSS® Modeler プロジェクトには、業界で実証された標準のデータマイニング作業の編成手段が用意されます。CRISP-DM では、作業の開始（ビジネス要件の収集）から完了（結果の展開）までのプロセスを大きく 6 つのフェーズに分けて記述します。通常、一部のフェーズは Clementine での作業とは関係ありませんが、SPSS Modeler のプロジェクトパネルでは、プロジェクトに関連するすべての資料やリソースを集中管理、保存、および追跡できるように、6 つのフェーズがすべて用意されています。たとえば、一般的に業務の理解フェーズでは、要件の収集と各作業担当者との打ち合わせにより、データマイニングの最終目標を決定します。SPSS Modeler でデータ処理などの作業を行うわけではありません。プロジェクトパネルを利用すれば、このような打ち合わせ内容などの資料や記録を [業務の理解] フォルダに保存して、将来の作業の参考にしたり、レポートに記載したりすることができます。

図 11-2
CRISP-DM ビュー



プロジェクト パネルの CRISP-DM ビューには、データ マイニング作業のライフ サイクルを説明する独自のヘルプ システムも用意されています。SPSS Modeler から、このヘルプには [ヘルプ] メニューの [CRISP-DM ヘルプ] をクリックしてアクセスします。

注： プロジェクト パネルがウィンドウに表示されていない場合は、[表示] メニューで [プロジェクト] をクリックします。

デフォルトのプロジェクトフェーズの設定

プロジェクトに追加されたオブジェクトは、CRISP-DM のデフォルト フェーズに追加されます。このため、オブジェクトを使用するデータ マイニングフェーズに応じて、手作業でオブジェクトを編成し直す必要があります。現在作業中のフェーズを、デフォルト フォルダに設定することが賢明です。

デフォルト フォルダとして使用するフェーズを選択する手順は、次のとおりです。

- ▶ CRISP-DM ビューで、デフォルトとして設定するフェーズに対応するフォルダを右クリックします。
- ▶ メニューから、[デフォルト値に設定] をクリックします。
デフォルト フォルダは太字で表示されます。

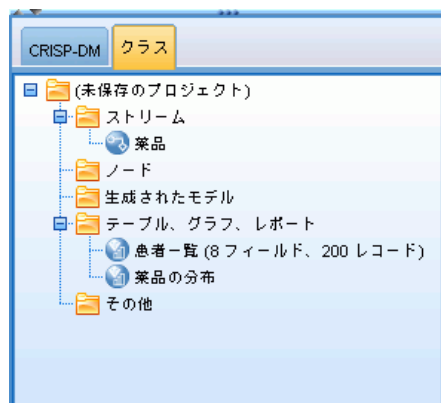
クラス ビュー

プロジェクト パネルのクラス ビューでは、IBM® SPSS® Modeler で行った作業が作成したオブジェクトの種類ごとに分類、編成されます。保存したオブジェクトは、次の任意のカテゴリに追加することができます。

- ストリーム
- ノード

- Models
- テーブル、グラフ、およびレポート
- その他（スライド ショーやデータ マイニング作業の関連資料などの非 SPSS Modeler ファイル）

図 11-3
クラス ビュー



オブジェクトをクラス ビューに追加すると、そのオブジェクトは CRISP-DM ビューのデフォルト フェーズ フォルダにも追加されます。

注： プロジェクト パネルがウィンドウに表示されていない場合は、[表示] メニューで [プロジェクト] をクリックします。

プロジェクトの作成

基本的にプロジェクトは、プロジェクトに関連付けたすべてのファイルへの参照を含んだファイルです。つまり、プロジェクト内の各項目は個別に保存されるとともに、プロジェクト ファイル (.cpj) 内に参照として保存されます。このような参照構造を採用しているため、次の点に注意してください。

- プロジェクトに追加する各項目は、プロジェクトに追加する前に、個別に保存しておく必要があります。追加する項目がまだ保存、³されていない場合は、プロジェクトに追加する前に保存するよう要求するメッセージが表示されます。
- ストリームのように個別に更新されるオブジェクトは、プロジェクト ファイル内でも更新されます。
- ファイル システムから手作業でオブジェクト（ストリーム、ノード、出力オブジェクトなど）を移動または削除すると、プロジェクト ファイル内のリンクが無効になります。

新規プロジェクトの作成

新しいプロジェクトは、IBM® SPSS® Modeler ウィンドウから簡単に作成できます。まだプロジェクトを開いていない場合は、プロジェクトの作成を開始することができます。または、すでにプロジェクトを開いている場合、それを閉じて、新しいプロジェクトを作成することもできます。

- ▶ メイン メニューで次の各項目をクリックします。
File > プロジェクト > 新規プロジェクト...

プロジェクトへの追加

プロジェクトを作成するか、または既存のプロジェクトを開いた後に、データ ストリーム、ノード、およびレポートなどのオブジェクトを追加できます。オブジェクトを追加するために、さまざまな方法を利用できます。

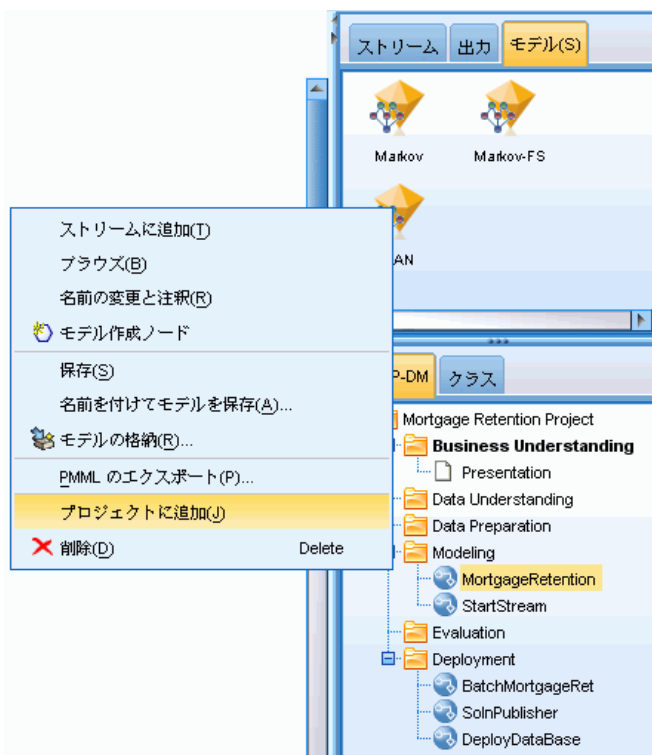
マネージャからのオブジェクトの追加

IBM® SPSS® Modeler ウィンドウの右上にあるマネージャを使って、ストリームまたは出力を追加することができます。

- ▶ 適切なマネージャ タブから、テーブルやストリームなどのオブジェクトを選択します。
- ▶ 右クリックして、[プロジェクトに追加] を選択します。
オブジェクトがすでに保存されたことがある場合は、それが適切なオブジェクト フォルダ (クラス ビュー) またはデフォルトのフェーズ フォルダ (CRISP-DM ビュー) に自動的に追加されます。
- ▶ 代わりに、オブジェクトをマネージャからプロジェクト パネルにドラッグアンド ドロップすることもできます。

注：先にオブジェクトの保存を要求するメッセージが表示されることもあります。オブジェクトを保存する際には、[保存] ダイアログ ボックスで [ファイルをプロジェクトに追加] を忘れずに選択してください。このオプションを選択すると、オブジェクトを保存した後、そのオブジェクトが自動的にプロジェクトに追加されます。

図 11-4
プロジェクトへの項目の追加



ストリーム領域からのノードの追加

[保存] ダイアログ ボックスを使用して、ストリーム領域から個別のノードを追加できます。

- ▶ ストリーム領域のノードを選択します。
- ▶ ノードを右クリックして、[ノードの保存] をクリックします。または、メインメニューで次の各項目をクリックします。
編集 > ノード > ノードの保存...
- ▶ [保存] ダイアログ ボックスで、[ファイルをプロジェクトに追加] を選択します。
- ▶ ノード名を指定して、[保存] をクリックします。

ファイルが保存され、プロジェクトにノードが追加されます。ノードは、クラス ビューの [ノード] フォルダと、CRISP-DM ビューのデフォルトのフェーズ フォルダに追加されます。

外部ファイルの追加

SPSS Modeler 以外のさまざまなオブジェクトをプロジェクトに追加することができます。これは、SPSS Modeler 内でデータマイニングプロセス全体を管理する場合に役立ちます。たとえば、データへのリンク、注釈、プレゼンテーション、およびグラフィックなどをプロジェクトに保存できます。CRISP-DM ビューでは、外部ファイルを保存するフォルダを選択できます。クラス ビューの場合、外部ファイルは [その他] フォルダにしか保存できません。

プロジェクトに外部ファイルを追加する手順は、次のとおりです。

- ▶ デスクトップからプロジェクトにファイルをドラッグ アンド ドロップします。

or

 - ▶ CRISP-DM ビューまたはクラス ビュー内の対象フォルダを右クリックします。
 - ▶ メニューで、[フォルダに追加] をクリックします。
 - ▶ ダイアログ ボックス内でファイルを選択して、[開く] をクリックします。
- SPSS Modeler プロジェクト内の選択したオブジェクトへの参照が追加されます。

プロジェクト IBM SPSS Collaboration and Deployment Services Repository への転送

すべてのコンポーネント ファイルも含めたプロジェクト全体を、IBM® SPSS® Collaboration and Deployment Services Repository へ一度に転送できます。すでに目的の場所にあるオブジェクトは、移動しません。これは逆にも機能します。プロジェクト全体を IBM SPSS Collaboration and Deployment Services Repository からローカル ファイル システムに移行できます。

注： IBM® SPSS® Collaboration and Deployment Services リポジトリを利用するには、別途ライセンスが必要です。詳細は、<http://www.ibm.com/software/analytics/spss/products/deployment/cds/> を参照してください。

プロジェクトの転送

転送するプロジェクトがプロジェクト パネル内で開かれていることを確認します。

プロジェクトを転送する手順は、次のとおりです。

- ▶ ルート プロジェクト フォルダを右クリックして、[プロジェクトの転送] をクリックします。
- ▶ メッセージが表示されたら、IBM SPSS Collaboration and Deployment Services Repository へログインします。
- ▶ プロジェクトの新しい場所を指定して、[OK] をクリックします。

プロジェクトのプロパティの設定

プロジェクトのプロパティのダイアログ ボックスを使用して、プロジェクトの内容やドキュメントをカスタマイズできます。プロジェクトのプロパティにアクセスする手順は、次のとおりです。

- ▶ プロジェクト パネル内のオブジェクトまたはフォルダを右クリックして、[プロジェクトのプロパティ] をクリックします。
- ▶ プロジェクトの基本情報を指定するには、[プロジェクト] タブをクリックします。

図 11-5
プロジェクトのプロパティの設定

作成日: 2009/09/17 14:30:03 BST

要約:

プロジェクト オブジェクト	回数
ストリーム	1
ノード	0
スーパーノード	0
生成されたモデル	0
テーブル、グラフ、およびレポート	2
その他	0
合計	3

未保存のオブジェクトを名前を付けて保存:

ローカル ファイル システム上のファイル

Repository 内のオブジェクト

プロジェクトのロード中にオブジェクト参照を更新

作成日: プロジェクトの作成日を表示します（編集不可）。

要約: プロジェクト レポートに表示される、データ マイニング プロジェクトの要約を入力できます。

内容: このプロジェクト ファイルに参照されるコンポーネントの種類と数を一覧表示します（編集不可）。

未保存のオブジェクトを名前を付けて保存: 未保存のオブジェクトをローカル ファイル システムに保存するか、またはリポジトリに格納するかを指定します。詳細は、9 章 p.175 [IBM SPSS Collaboration and Deployment Services Repository](#) についてを参照してください。

プロジェクトのロード中にオブジェクト参照を更新: プロジェクトのコンポーネントへの参照を更新する場合に、このオプションを選択します。注：プロジェクトに追加したファイルは、そのプロジェクト ファイル内に保存され

るわけではありません。プロジェクトには、各ファイルへの参照が保存されます。つまり、元のファイルを移動または削除すると、その参照もプロジェクトから削除されます。

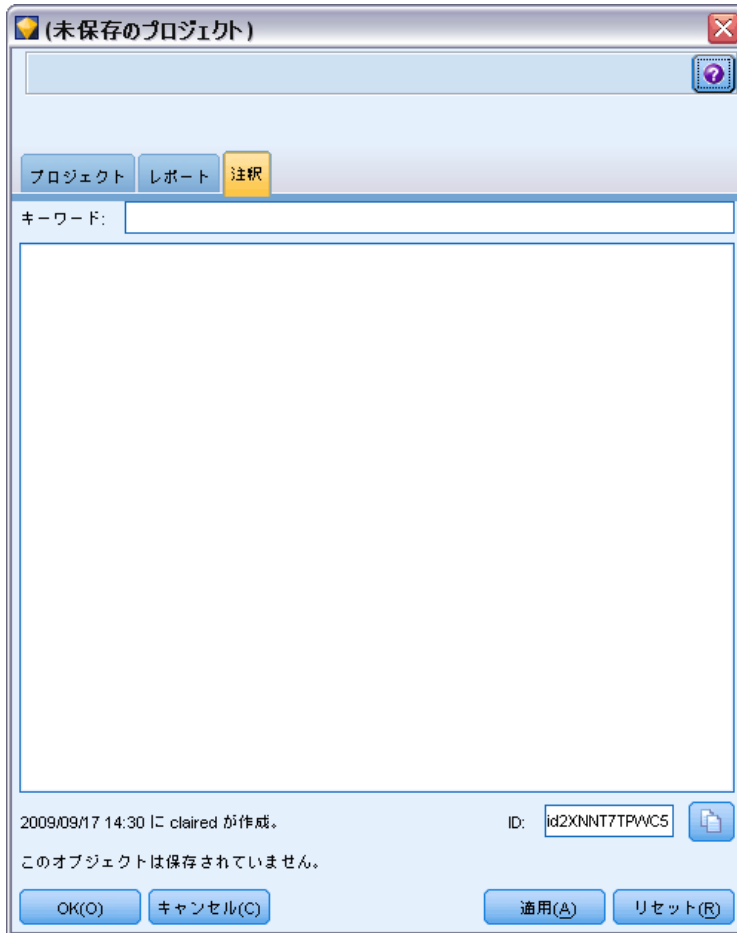
プロジェクトの注釈

プロジェクト パネルには、データ マイニング作業に関する注釈を記述するための、さまざまな方法が用意されています。プロジェクト レベルの注釈は、しばしば「大局的な」目標や決定などの追跡に使用され、フォルダやノード レベルの注釈は、詳細な説明を伝えるために使用されます。[注釈] タブには、取得できない欠損データの除外や、データ探索の過程で確立された有望な仮説などの、プロジェクト レベルの詳細を記録するために十分な領域が用意されています。

プロジェクトに注釈を付ける手順は、次のとおりです。

- ▶ CRISP-DM ビューまたはクラス ビューで、プロジェクト フォルダを選択します。
- ▶ フォルダを右クリックして、[プロジェクトのプロパティ] をクリックします。
- ▶ [注釈] タブをクリックします。

図 11-6
[プロジェクトのプロパティ] ダイアログ ボックスの [注釈] タブ



- ▶ プロジェクトを記述するキーワードとテキストを入力します。

フォルダのプロパティと注釈

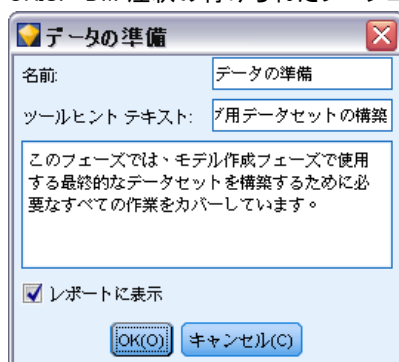
個別のプロジェクト フォルダ (CRISP-DM ビューおよびクラス ビューの両方の中) に注釈を付けることができます。CRISP-DM ビューでは、注釈機能がデータ マイニングの各フェーズに対する組織目標を記録するための、最も効果的な手段になります。たとえば、[ビジネスの理解] フォルダで注釈 ツールを使用して、「今回の調査におけるビジネス目標は、高い価値を持つ顧客の顧客離れを減らすことにある。」のような情報を記入することができます。ここに記入したテキストは、[レポートに入れる] オプションを選択すると、プロジェクト レポートに自動的に記載されます。

フォルダに注釈を付ける手順は、次のとおりです。

- ▶ プロジェクト パネル内でフォルダを選択します。
- ▶ フォルダを右クリックして、[フォルダのプロパティ] をクリックします。

CRISP-DM ビューのフォルダには、各フェーズの目的の概略や、関連するデータマイニング作業を行うためのガイドなどの注釈が付けられます。これらの注釈は、必要に応じて削除したり編集したりすることができます。

図 11-7
CRISP-DM 注釈の付けられたプロジェクトフォルダ



名前: この領域には、選択したフィールドの名前が表示されます。

ツールヒントテキスト: マウスポインタをプロジェクトフォルダ上に移動した時に表示される、カスタムのツールヒントを作成します。この機能は、CRISP-DM ビュー内で各フェーズの目標を記載したり、「作業中」や「完了」のようにフェーズのステータスを表示したりする場合に役立ちます。

[注釈] フィールド: このフィールドは、プロジェクトレポートに記載できるより長い注釈を記載するために利用されます。CRISP-DM ビューの注釈には、データマイニングの各フェーズの注釈が記載されていますが、プロジェクトの内容に応じて、自由に変更することができます。

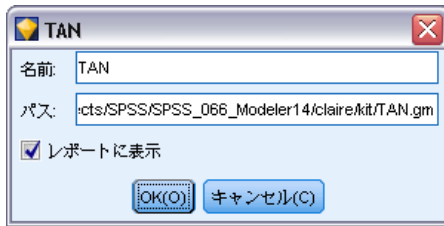
レポートに入れる: この注釈をレポートに記載するには、[レポートに入れる] を選択します。

オブジェクトのプロパティ

オブジェクトのプロパティを表示したり、プロジェクトレポートに個別のオブジェクトを記載するかどうかを選択できます。オブジェクトのプロパティにアクセスする手順は、次のとおりです。

- ▶ プロジェクト パネルで、オブジェクトを右クリックします。
- ▶ メニューで [オブジェクトのプロパティ] をクリックします。

図 11-8
[オブジェクトのプロパティ] ダイアログ ボックス



名前: この領域には、保存したオブジェクトの名前が一覧表示されます。

パス: この領域には、オブジェクトの保存場所が一覧表示されます。

レポートに入れる: 生成するレポートにオブジェクトの詳細を記載する場合に選択します。

プロジェクトを閉じる

IBM® SPSS® Modeler を終了したり、新しいプロジェクトを開いたりすると、既存のプロジェクト ファイル (.cpj) が閉じます。

製品と関連するいくつかのファイル (ストリーム、ノード、またはグラフなど) は開いたままとなる場合があります。これらのファイルを開いたままにする場合、[... これらのファイルを保存して閉じますか?] というメッセージに [いいえ] と答えてください。

プロジェクトを閉じた後に関連するファイルを変更して保存した場合、それらの変更内容は、次にプロジェクトを開いた時に反映されます。変更内容を反映しないようにするには、プロジェクトからそのファイルを削除するか、それを別のファイル名で保存してください。

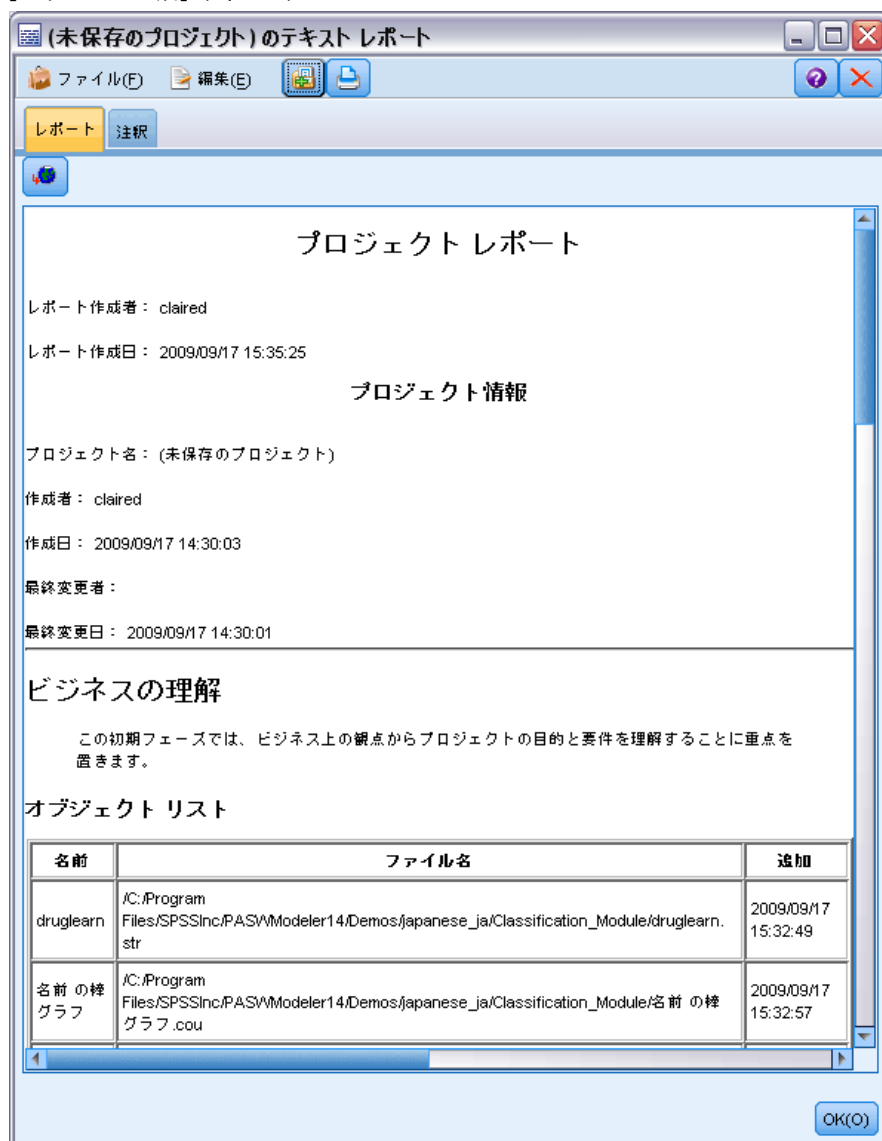
レポートの生成

プロジェクトのもっとも役立つ機能の 1 つとして、プロジェクト内の各項目や注釈に基づいてレポートを生成できる機能があります。この機能は、「CRISP-DM 手法」で説明しているように、効果的なデータ マイニングを行うために必要不可欠です。レポートは、いくつかのファイル タイプの 1 つに直接生成することも、ただちに確認できるように画面上のウィンドウに出力することもできます。この出力から、印刷や保存、または Web ブラウザ内でのレポートの表示を行うことができます。保存されたレポートは、組織内の他の部門に配布できます。

レポートは、データ マイニング プロセス中にプロジェクト ファイルから何度も生成され、プロジェクトの関係者に配布されます。レポートには、プロジェクト ファイルで参照されるオブジェクト情報や、作成した注

釈などから、必要な情報や重要な情報を記載します。レポートは、クラスビューまたは CRISP-DM ビューのどちらかに基づいて作成できます。

図 11-9
[レポートの生成] ウィンドウ



レポートを生成する手順は次のとおりです。

- ▶ CRISP-DM ビューまたはクラス ビューで、プロジェクト フォルダを選択します。
- ▶ フォルダを右クリックして、[プロジェクトのレポート] をクリックします。

- ▶ レポート オプションを指定して、[レポートの生成] ボタンをクリックします。

図 11-10
レポート オプションの選択



[レポート] ダイアログ ボックスのオプションには、目的に応じたさまざまなレポートを生成するいくつかの方法が用意されています。

出力名：レポートの出力を画面に送ることを選択した場合は、出力ウィンドウの名前を指定します。独自の名前を指定することも、IBM® SPSS® Modeler に自動的にウィンドウの命名をさせることもできます。

画面に出力：レポートを生成し、出力ウィンドウに表示する場合に選択します。出力ウィンドウからさまざまな種類のファイルへレポートをエクスポートするオプションもあります。

ファイルに出力：レポートを生成して、[ファイルの種類] リストで指定された種類のファイルとして保存する場合に選択します。

ファイル名 :生成されたレポートのファイル名を指定します。ファイルは、デフォルトで SPSS Modeler %bin ディレクトリへ保存されます。別の場所を指定するには、反復記号の [...] ボタンを使用します。

ファイルの種類 :指定できるファイルの種類を次に示します。

- **HTMLドキュメント** : レポートは、単一の HTML ファイルとして保存されます。レポートにグラフが含まれている場合は PNG ファイルとして保存され、HTML ファイルに参照されます。レポートをインターネットで公開する場合は、HTML ファイルと、そのファイルが参照する画像の両方を必ずアップロードしてください。
- **テキストドキュメント** : レポートは、単一のテキスト ファイルとして保存されます。レポートにグラフが含まれている場合は、ファイル名とパスの参照のみがレポートに含まれます。
- **Microsoft Wordドキュメント** : レポートは、単一のドキュメントとして保存され、グラフはドキュメント内に直接埋め込まれます。
- **Microsoft Excelドキュメント** : レポートは、単一のスプレッドシートとして保存され、グラフはスプレッドシート内に直接埋め込まれます。
- **Microsoft PowerPointドキュメント** : 各フェーズが新しいスライドに表示されます。どのようなグラフであっても、PowerPoint スライドに直接埋め込まれます。
- **出力オブジェクト** : SPSS Modeler 内で開くと、このファイル (.cou) は [レポート形式] の [画面に出力] と同じになります。

注 : Microsoft Office ファイルをエクスポートするには、対応するアプリケーションがインストールされている必要があります。

表題 . レポートのタイトルを指定します。

レポート構造 : [CRISP-DM] または [クラス] のどちらかを選択します。CRISP-DM ビューは、「大局的な」観点からのステータス レポートと、データ マイニングの各フェーズの詳細情報を提供します。クラス ビューはオブジェクトを基準にしたビューで、データやストリームの内部的な追跡を行う場合に適しています。

作成者 : デフォルトのユーザー名が表示されますが、変更できます。

レポートfglに表示する項目 : オブジェクトをレポートに含める方法を選択します。プロジェクト ファイルに追加されたすべての項目を記載するには、[すべてのフォルダとオブジェクト] を選択します。オブジェクトのプロパティで [レポートに入れる] が選択されているかどうかを基準にして、項目を記載することもできます。代わりに、レポートに表示しない項目を確認するために、除外することを選択した項目だけを記載することもできます ([レポートに入れる] の選択を解除する)。

選択： レポート内で [最近使用した項目] のみを選択することで、プロジェクトの更新内容を提供できるようにします。代わりに [古い項目] のパラメータを設定して、古い項目や未解決の問題を追跡することもできます。レポートで時間的なことを考慮しない場合は、[すべての項目] を選択します。

表示順： フォルダ内のオブジェクトの順序を決めるために、次のオブジェクト特性の組み合わせを選択できます。

- **タイプ：** データ型を基準にオブジェクトをグループ化します。
- **名前：** オブジェクトをアルファベット順に並べます。
- **追加日：** プロジェクトに追加された日付を使用して、オブジェクトをソートします。

生成されたレポートの保存とエクスポート

画面に生成されたレポートは、新規出力ウィンドウに表示されます。レポートに含まれるグラフは、インラインの画像として表示されます。

各ストリームのノードの合計数がレポート内に表示されます。次の見出しに数が表示され、CRISP-DM 用語集ではなく IBM® SPSS® Modeler 用語集が使用されます。

- **データリーダー：** 入力ノード。詳細は、[2 章 概要 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。
- **データライター：** エクスポート ノード。詳細は、[7 章 エクスポート ノードの概要 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。
- **モデルビルダー：** 構築ノードまたはモデル作成ノード。詳細は、[3 章 モデル作成ノードの概要 in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。
- **モデル アプライヤー：** ナゲットとも呼ばれる、生成されたモデル。詳細は、[3 章 モデル ナゲット in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。
- **出力ビルダー：** グラフ ノードまたは出力ノード。詳細は、[5 章 グラフ作成ノードの共通の機能 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。 詳細は、[6 章 出力ノードの概要 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。
- **その他：** プロジェクトに関連するその他のノード。たとえば、ノード パレットの [フィールド設定] タブまたは [レコード設定] タブで使用できるノード。

レポートを保存する手順は次のとおりです。

- ▶ [ファイル] メニューで [保存] をクリックします。

- ▶ ファイル名を指定します。
レポートは、出力オブジェクトとして保存されます。

レポートをエクスポートする手順は次のとおりです。

- ▶ [ファイル] メニューで [エクスポート] をクリックし、エクスポートするファイルの種類を選択します。
- ▶ ファイル名を指定します。
レポートは、選択した形式で保存されます。

次の種類のファイルへエクスポートできます。

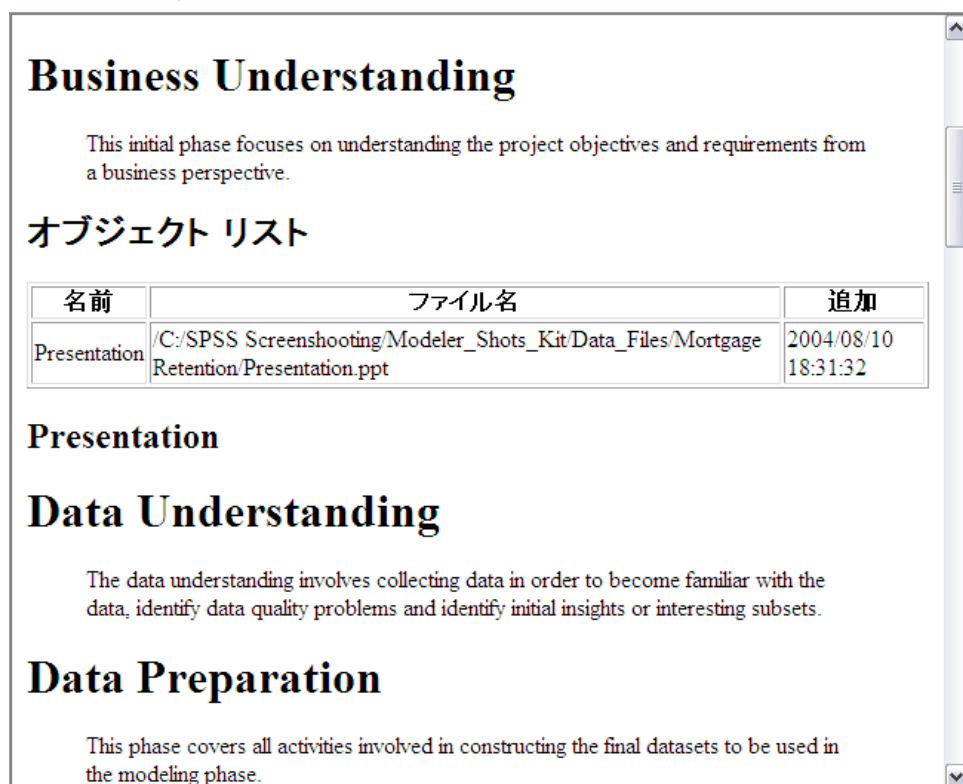
- HTML
- テキスト
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

注：Microsoft Office ファイルをエクスポートするには、対応するアプリケーションがインストールされている必要があります。

ウィンドウの上部にあるボタンを使用して、次の作業を実行します。

- レポートを印刷します。
- 外部の Web ブラウザ内で HTML としてレポートを表示します。

図 11-11
Web ブラウザに表示したレポート



Business Understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective.

オブジェクト リスト

名前	ファイル名	追加
Presentation	C:/SPSS Screenshooting/Modeler_Shots_Kit/Data_Files/Mortgage Retention/Presentation.ppt	2004/08/10 18:31:32

Presentation

Data Understanding

The data understanding involves collecting data in order to become familiar with the data, identify data quality problems and identify initial insights or interesting subsets.

Data Preparation

This phase covers all activities involved in constructing the final datasets to be used in the modeling phase.

IBM SPSS Modeler のカスタマイズ

IBM SPSS Modeler オプションのカスタマイズ

必要に応じて IBM® SPSS® Modeler をカスタマイズするためには、さまざまな操作があります。このカスタマイズは、主にメモリーの割り当て、デフォルトのディレクトリ、音声および色の使用など、特定のユーザー オプションの設定で構成されています。また、SPSS Modeler ウィンドウの下部にあるノード パレットをカスタマイズすることもできます。

IBM SPSS Modeler オプションの設定

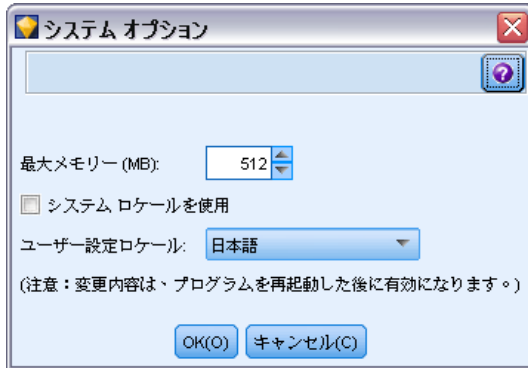
IBM® SPSS® Modeler のオプションをカスタマイズしたり、設定するには、さまざまな方法があります。

- メモリー使用量やロケールなどのシステム オプションを設定するには、[ツール>オプション] メニューの [システム オプション] をクリックします。
- フォント、最適化、および警告などのユーザー オプションを設定するには、[ツール>オプション] メニューの [ユーザー オプション] をクリックします。
- SPSS Modeler と連携するアプリケーションの場所を指定するには、[ツール>オプション] メニューの [ヘルパー アプリケーション] をクリックします。
- SPSS Modeler が使用するデフォルト ディレクトリを指定するには、[ファイル] メニューの [ディレクトリの設定] または [サーバー ディレクトリの設定] をクリックします。

システム オプション

IBM® SPSS® Modeler の使用言語またはロケールを指定するには、[ツール>オプション] メニューの [システム オプション] をクリックします。ここでは、SPSS Modeler の最大メモリー使用量を設定することもできます。このダイアログ ボックスで行った変更内容は、SPSS Modeler を再起動しないと有効にはならないことに注意してください。

図 12-1
[システム オプション] ダイアログ ボックス



最大メモリー :SPSS Modeler のメモリー使用量を MB 単位で制限する場合に選択します。一部のプラットフォームでは、リソースに制約のあるコンピュータや負荷の高いコンピュータの負担を軽減するために、SPSS Modeler はプロセスのサイズを制限しています。そのため、大量のデータを扱うと、「メモリー不足」のエラーが発生することがあります。ここに新しい値を指定して、メモリーの負荷を調整することができます。

システム ロケールを使用 : このオプションはデフォルトで選択され、[日本語] に設定されています。他の言語を指定する場合は、このオプションの選択を解除して、リストから適切な言語やロケールを選択してください。

メモリーの管理

[システム オプション] ダイアログ ボックスの [最大メモリー] の設定以外にも、メモリーの使用を最適化するためのさまざまな方法があります。

- ターミナル ノード以外のノードにキャッシュを設定して、データ ストリームの実行時に、データ ソースではなくキャッシュからデータを読み込むようにする。こうすることによって、大きなデータ セットによるメモリー負荷を減らすことができます。 [詳細は、5 章 p.54 ノードのキャッシュ オプション を参照してください。](#)
- [ストリームのプロパティ] ダイアログ ボックスの [名義型フィールドの最大メンバー数] オプションを調整する。このオプションには、名義型フィールドの最大メンバー数を指定します。これ以降のフィールドの測定レベルは、データ型不明になります。 [詳細は、5 章 p.58 ストリームのオプションの設定 を参照してください。](#)
- IBM® SPSS® Modeler のウィンドウの右下に表示されている、SPSS Modeler が使用中のメモリーおよびメモリーの総割り当て量 (xxMB/xxMB) の領域をクリックして、メモリーを解放する。この領域をクリックすると、いったん領域が暗くなって押されたような状態になります。その後領域が元の色に戻ると、SPSS Modeler で解放できるメモリーがすべて解放されます。

デフォルト ディレクトリの設定

ファイルの参照や出力に使われるデフォルト ディレクトリを指定するには、[ファイル] メニューの [ディレクトリの設定] または [サーバー ディレクトリの設定] を選択します。

- **ディレクトリの設定** :作業ディレクトリを設定します。デフォルトの作業ディレクトリは、IBM® SPSS® Modeler のインストール場所、または SPSS Modeler を起動するために用いられたコマンド ラインのパスをペースにしています。ローカル モードの場合、作業ディレクトリはクライアント側の操作や出力ファイルで使われたパスになります (相対パスで参照された場合)。
- **サーバー ディレクトリの設定** : リモート サーバーへの接続がある場合、[ファイル] メニューの [サーバー ディレクトリの設定] オプションは常に有効になっています。このオプションを使って、入力や出力に使われるすべてのサーバー ファイルやデータ ファイルのデフォルト ディレクトリを指定します。デフォルトのサーバー ディレクトリは、\$CLEO/data になります。ここで、\$CLEO は SPSS Modeler のサーバーバージョンがインストールされたディレクトリを表します。コマンド ラインを使って、このデフォルト値に優先するディレクトリを指定するには、`modelerclient` コマンド ラインの引数として `-server_directory` フラグを指定します。

[ユーザー オプション] ダイアログ ボックスの設定

IBM® SPSS® Modeler の一般オプションを設定するには、[ツール > オプション] メニューの [ユーザー オプション] を選択します。これらのオプションは、SPSS Modeler で使用するすべてのストリームに適用されます。

ここでは、次の種類のオプションを設定することができます。これらのオプションを設定するには、該当するタブをクリックしてください。

- [通知] オプション : モデルの上書きやエラー メッセージなどのオプションを設定します。
- [表示] オプション : グラフや背景の色を設定します。
- [最適化] オプション : SQL 生成やストリームのリライトなどのオプションを設定します。
- モデルを Predictive Model Markup Language (PMML) へエクスポートするときに使った PMML エクスポート オプション。
- 名前、イニシャル、メール アドレスなどのユーザまたは作成者の情報。この情報を、作成したノードまたはその他のオブジェクトの[注釈]タブに表示することができます。

数字区切り文字、時刻や日付のフォーマット、およびストリーム スクリプトなどのストリーム固有のオプションの設定は、[ストリームのプロパティ] ダイアログ ボックスから行います。このダイアログ ボックスは、[ファイル] メニューと [ツール] メニューから利用できます。

通知オプションの設定

[ユーザー オプション] ダイアログ ボックスの [通知] タブでは、IBM® SPSS® Modeler の警告および確認ウィンドウの表示や種類に関係するさまざまな設定を指定できます。また、新しい出力やモデルが生成された時の、マネージャ パネルの [出力] タブおよび [モデル] タブの動作も設定することができます。

図 12-2
[ユーザー オプション] ダイアログ ボックス、[通知] タブ



ストリーム実行フィードバック ダイアログの表示 : このオプションを選択すると、ストリームが 3 秒間実行されている場合、進捗指標のインジケータが表示されます。ダイアログ ボックスには、ストリームで作成された出力オブジェクトの詳細も表示されています。

- **完了時にダイアログを閉じる** : デフォルトでは、ストリームが実行を完了すると、このダイアログ ボックスが閉じられます。ストリームの実行が完了した後もダイアログ ボックスを表示したままにする場合は、このボックスの選択を解除 (チェックを外す) します。

ノードがファイルを上書きする時に警告 : このオプションを選択すると、ノード操作により既存のファイルが上書きされる場合に、警告のエラー メッセージが表示されます。

ノードがデータベース テーブルを上書きする時に警告 : このオプションを選択すると、ノード操作により既存のデータベース テーブルが上書きされる場合に警告のエラー メッセージが表示されます。

サウンド通知

下にあるリストを使って、イベントやエラーが発生した時に通知するサウンドを指定します。さまざまな音声やサウンドを利用できます。選択したサウンドを再生する場合は、[再生] ボタン (スピーカー) を使用します。サウンド ファイルを探して選択する場合は、[...] ボタンを使用します。

注 : SPSS Modeler のサウンドを作成するために使われる .wav ファイルは、インストール ディレクトリ中の /media/sounds ディレクトリにあります。

- **すべての音をミュート** : イベントのサウンド通知を行わない場合に選択します。

視覚通知

このグループのオプションを使って、新規項目生成時のマネージャ パネル右上の [出力] タブおよび [モデル] タブの動作を指定します。リストから [新規モデル] または [新規出力] を選択して、対応するタブの動作を設定してください。

新規モデルについては、次のモデルを使用できます。

モデルをストリームに追加 : デフォルトでは、[モデル] タブのほか、モデルが作成されるとすぐに新しいモデルがストリームに追加されます。ストリームでは、モデルが作成されたモデル作成ノードへリンクした状態で、モデルが表示されます。このボックスをオフにすると、モデルは [モデル] タブにのみ追加されます。

前のモデルを置換： デフォルトでは、[モデル] タブおよびストリーム領域のこのストリームの既存モデルを上書きします。このボックスをオフにすると、[モデル] タブおよび領域の既存モデルにモデルが追加されます。この設定は、モデルリンクのモデル置換設定によって上書きされます。詳細は、[3 章 モデルの置換 in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。

新規出力については、次のモデルを使用できます。

出力超過時に警告 [n]： [出力] タブの項目数が事前に指定された数を越えた場合に警告を表示するかどうかを選択します。デフォルトの数は 20 です。必要に応じて、この数を変更することができます。

すべてのケースで次のオプションを利用できます。

タブを選択： ストリームの実行時にオブジェクトが生成された場合に、[出力] タブまたは [モデル] タブに切り替えるかどうかを指定します。

- マネージャ パネルの該当するタブに切り替える場合は、[常時] を選択します。
- 現在領域に表示されているストリームによってオブジェクトが生成された時にだけタブに切り替える場合は、[現在のストリームが生成した場合] を選択します。
- 出力やモデルが生成されたことを、対応するタブに切り替えて知らせない場合は、[しない] を選択します。

タブを点滅： 新しい出力やモデルが生成された場合に、マネージャ パネルにある [出力] タブまたは [モデル] タブを点滅させるかどうかを選択します。

- マネージャ パネル中の対応するタブを点滅させる場合は（タブが選択されていない場合）、[選択されていない場合] を選択します。
- オブジェクトが生成されたことを、対応するタブを点滅させて知らせない場合は、[しない] を選択します。

パレットを見えるようにスクロール (新規モデルの場合)： 最新のモデルが見えるように、マネージャ パネルの [モデル] タブを自動的にスクロールするかどうかを選択します。

- スクロールする場合は、[常時] を選択します。
- 現在領域に表示されているストリームによってオブジェクトが生成された時にだけスクロールする場合は、[現在のストリームが生成した場合] を選択します。
- [モデル] タブの自動スクロールを行わない場合は、[しない] を選択します。

ウィンドウを開く (新規出力の場合)： 生成時に出力ウィンドウを自動的に開くかどうかを選択します。

- 常に新しく出力ウィンドウを開く場合は、[常時] を選択します。
- 現在領域に表示されているストリームによって出力が生成された時にだけ新しいウィンドウを開く場合は、[現在のストリームが生成した場合] を選択します。
- 生成された出力に対して自動的に新しいウィンドウを開かない場合は、[しない] を選択します。

このタブのデフォルト設定に戻す場合は、[デフォルト値] を選択します。

表示オプションの設定

[ユーザー オプション] ダイアログ ボックスの [表示] タブを使って、IBM® SPSS® Modeler のフォントや色の表示を設定することができます。

図 12-3

[ユーザー オプション] ダイアログ ボックス、[表示] タブ



ようこそダイアログが起動時に表示: ようこそダイアログ ボックスが起動時に表示されます。ようこそダイアログには、アプリケーションの例チュートリアルを起動するオプション、デモ ストリームまたは既存のストリー

ムやプロジェクトを開くオプション、または新規ストリームを作成するオプションがあります。

ストリームおよびスーパーノードのマークアップを表示： デフォルトでは、ストリームおよびスーパーノードのマークアップがあれば表示されます。マークアップはストリームのコメント、モデル リンク、およびスコアリング枝の強調表示で構成されています。

標準のフォントと色 (再起動後に有効)： このコントロール ボックス中のオプションを使って、SPSS Modeler の画面デザイン、表示色およびフォント サイズを指定します。ここで指定した内容は、いったん SPSS Modeler を終了してから再起動しないと有効にはなりません。

- **外観と使用感：** 標準のカラー スキームおよび画面デザインを使用できます。次の項目を選択できます。

SPSS 標準 (デフォルト)： IBM SPSS 製品共通のデザイン。

SPSS クラシック： 以前のバージョンの SPSS Modeler のユーザーにお馴染みのデザイン。

Windows： ストリーム領域やパレットのコントラストを強くする場合などに役立つ Windows のデザイン。

- **小さいノードのフォント サイズ：** ノード パレット内、およびストリーム領域に小さいノードを表示する際に使用するフォント サイズを指定します。
- **大きいノードのフォント サイズ：** ストリーム領域に大きいノード (標準) を表示する際に使用するフォント サイズを指定します。

注： ストリーム領域のノード サイズは、[ストリームのプロパティ] ダイアログ ボックスの [レイアウト] タブから指定することができます。

カスタム カラー： このテーブルには、さまざまな表示項目に使われる色の、現在の設定が表示されています。テーブルに表示された項目について、[色] 列の該当する行をダブルクリックし、リストから色を選択して、現在の色を変更できます。カスタム カラーを指定するには、リストの最後にスクロールして、[色...] エントリをクリックします。

グラフ カテゴリの色順序： このテーブルには、新しく作成したグラフを表示する場合に使われる色の、現在の設定が表示されています。表示されている色の順序が、グラフで使われる順番を表しています。たとえば、色のオーバーレイに使われている名義型フィールドに 4 種類の固有値がある場合、ここに記載されている先頭から 4 つの色が使われます。テーブルに表示された項目について、[色] 列の該当する行をダブルクリックし、リストから色を選択して、現在の色を変更できます。カスタム カラーを指定するには、リストの最後にスクロールして、[色...] エントリをクリックします。ここで行った変更は、それまでに作成したグラフには適用されません。

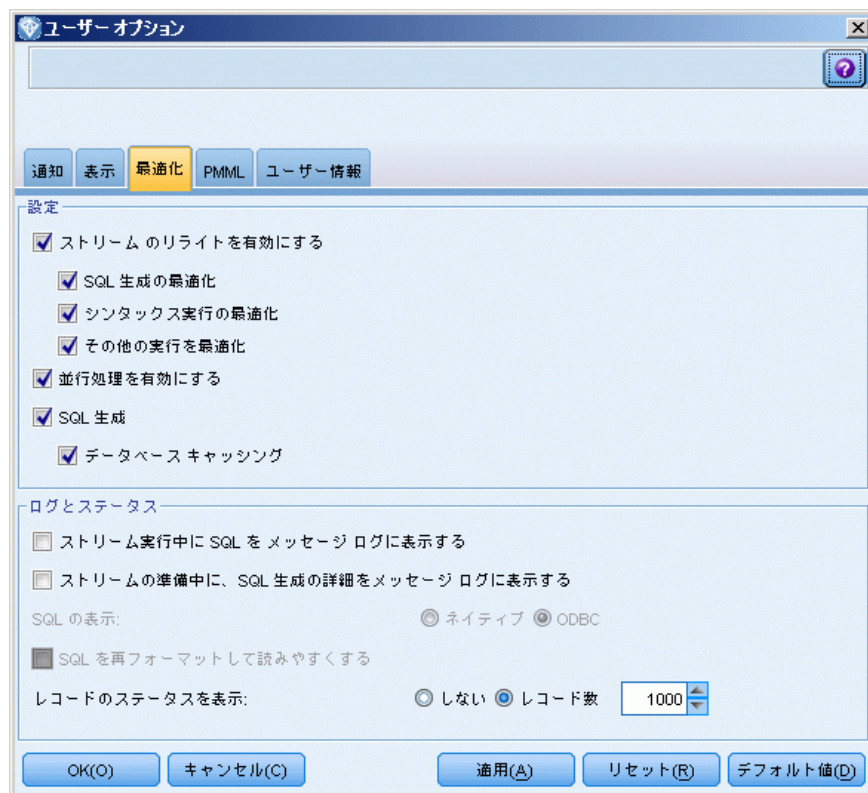
このタブのデフォルト設定に戻す場合は、[デフォルト値] を選択します。

最適化オプションの設定

[ユーザー オプション] ダイアログ ボックスの [最適化] タブを使用すると、ストリームのパフォーマンスを最適化できるようになります。IBM® SPSS® Modeler Server が使用されている場合のパフォーマンスと最適化の設定は、クライアント内で上書きされます。このタブは、SPSS Modeler Server 接続が IBM® SPSS® Modeler コンピュータで有効化されている場合にのみ使用できます。

注： データベース モデル作成および SQL 最適化では、SPSS Modeler Server 接続が SPSS Modeler コンピュータ上で可能でなければなりません。この設定を有効にすると、データベース アルゴリズムにアクセスし、SPSS Modeler から SQL を直接プッシュバック、SPSS Modeler Server にアクセスできます。現在のライセンス ステータスを検証するには、SPSS Modeler の [ヘルプ] メニューから [バージョン情報] を選択します。詳細は、3 章 p.10 IBM SPSS Modeler Server への接続を参照してください。

図 12-4
[ユーザー オプション] ダイアログ ボックス、[最適化] タブ



注：SQL プッシュバックおよび最適化がサポートされているかどうかは、使用しているデータベースの種類によって異なります。IBM® SPSS® Modeler 14.2 での使用がサポートおよびテストされているデータベースおよび ODBC ドライバの最新情報については、当社サポート サイト (<http://www.ibm.com/support>) にある製品互換性マトリクスを参照してください。

ストリームのリライトを有効にする： このオプションを選択すると、SPSS Modeler のストリームのリライトが有効になります。リライトには 2 種類あり、どちらかまたは両方を選択することができます。ストリームのリライトは、ストリーム内のノードをバックグラウンドで並べ替え、ストリームの意味論を変更することなくストリームをより効率的に操作できるようにする処理です。

- **SQL 生成の最適化：** このオプションを選択すると、より多くのストリーム操作がデータベースにプッシュバックされて SQL 生成を使用して実行されるように、ストリーム内のノードが並べ替えられます。SQL へ変換できないノードが検出されると、SQL へ変換可能な下流にあるかどうか、またそれを、ストリームの意味論に影響を与えずに安全に問題のノードの前に移動できるかどうかを調べることができます。データベースの方が SPSS Modeler より効率的に操作を実行できるだけでなく、このようなプッシュバックによって SPSS Modeler で処理するために返されるデータセットのサイズも小さくすることができます。この結果、ネットワークトラフィックを減らし、ストリーム操作を高速化することができます。SQL 最適化を有効にするには、[SQL 生成] チェックボックスをオンにする必要があります。詳細は、6 章 SQL 最適化 in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド を参照してください。
- **シンタックスの実行を最適化：** この方法のストリームのリライトでは、IBM® SPSS® Statistics シンタックスを含む複数のノードを結合する操作を効率化します。各操作それぞれを実行する代わりに、シンタックスコマンドを 1 つの操作に結合することによって最適化が行われます。
- **その他の実行を最適化：** この方法のストリームのリライトでは、データベースで処理できない操作を効率化します。最適化は、ストリームのデータ量をできるだけ早く減らすことによって実現されます。データの整合性を保ちながら、操作をデータソースの近くに置いて、下流で結合などの時間のかかる操作に使用されるデータの量を減らすように、ストリームがリライトされます。

並行処理を有効にする： マルチ プロセッサのコンピュータで実行する場合、このオプションでシステムがプロセッサ間の負荷のバランスがとれるようになり、より速いパフォーマンスとなります。複数のノードを使用、または C5.0、結合（キーごと）、ソート、ビン（ランクおよびタイル手法）、レコード集計（1 つまたは複数のキー フィールドを使用）を使用すると、平行処理のメリットが得られます。

SQL 生成：SQL 生成を有効にするには、このオプションを選択します。実行プロセスを生成する SQL コードを使用することで、ストリーム操作をデータベースにプッシュバックできるようにし、それがパフォーマンスを向上させる可能性があります。さらにパフォーマンスを向上させるには、データベースにプッシュバックする操作の数を最大限にできるように、[SQL 生成の最適化] も選択できます。ノードの操作がデータベースにプッシュバックされると、そのノードはストリームの実行時に紫色で強調表示されます。

- **データベース キャッシング：**SQL を生成するストリームをデータベース内で実行するために、ファイル システムでなくデータベース内の一時テーブルへ、データを中流でキャッシュできます。この機能を SQL 最適化と組み合わせると、パフォーマンスが著しく向上する可能性があります。たとえば、データ マイニング ビューを作成するために複数のテーブルを併合するストリームからの出力をキャッシュし、必要に応じて再使用できます。データベースのキャッシングが有効化されると、任意の非ターミナル ノードを右クリックするだけでその場所のデータがキャッシュされ、次にストリームが実行されると自動的に、データベース内にキャッシュが直接作成されます。これにより SQL が下流のノード用に生成されるようになり、パフォーマンスをさらに改善します。代わりに、たとえばポリシーまたは許可によってデータベースに書き込まれたデータが損なわれる場合、このオプションを必要に応じて無効にすることができます。データベース キャッシングまたは SQL 最適化が有効でない場合、キャッシュは、代わりにファイル システムへ書き込まれます。 [詳細は、5 章 p.54 ノードのキャッシュ オプション を参照してください。](#)

注：SQL 実装の些細な違いが原因で、データベース内で実行されるストリームは、SPSS Modeler 内で実行されるときと若干異なる結果を返すことがあります。これらの違いは、似たような理由で、データベース ベンダーによっても異なる可能性があります。

ストリーム実行中に SQL をメッセージ ログに表示する：ストリームの実行中に、生成された SQL をメッセージ ログに渡すかどうかを指定します。

ストリームの準備中に、SQL 生成の詳細をメッセージ ログに表示する：ストリームのプレビュー中に、生成される SQL のプレビューをメッセージ ログに渡すかどうかを指定します。

SQL の表示：ログ中に表示される SQL を、固有の SQL 関数、または、SPSS Modeler に生成されたとおりの {fn FUNC(...)} 形式の標準 ODBC 関数のどちらかの形式で含めるのかを指定します。前者は、実装されていない可能性がある ODBC ドライバの機能に依存しています。たとえば、このコントロールは、SQL Server に対して何の影響も及ぼしません。

SQL を再フォーマットして読みやすくする：ログに表示された SQL を読みやすくするためにフォーマットするかどうかを指定します。

レコードのステータスを表示：レコードがターミナル ノードに達したときに、そのレコードを報告する時期を指定します。N レコードごとにステータスの更新を行うことを、数字で指定します。

このタブのデフォルト設定に戻す場合は、[デフォルト値] を選択します。

PMML エクスポート オプション の設定

[PMML] タブで、IBM® SPSS® Modeler から Predictive Model Markup Language (PMML) へモデルをエクスポートする方法を制御できます。詳細は、10 章 p.230 PMML としてのモデルのインポートおよびエクスポート を参照してください。

図 12-5

[ユーザー オプション] ダイアログ ボックス、[PMML] タブ



PMML をエクスポート：PMML を対象アプリケーションと最高の状態で動作させるための各種設定をここでを行います。

- [拡張機能] を選択し、標準 PMML 相当する機能のない特殊なケースに PMML 拡張機能を使用できるようにします。ほとんどのケースで、標準 PMML と同じ結果が生成されます。
- [PMML V3.2 標準として] を選択し、PMML 標準にできる限り準拠するように PMML をエクスポートします。

標準 PMML オプション: [PMML V3.2 標準として] を選択した場合には、線型回帰モデルとロジスティック回帰モデルをエクスポートするための方法を、次の 2 つから選択できます。

- PMML <GeneralRegression> モデルとして
- PMML <Regression> モデルとして

ユーザー情報の設定

ユーザーと作成者の情報: ここで入力した情報は、作成したノードおよびその他のオブジェクトの [注釈] タブに表示されます。

ノード パレットのカスタマイズ

ストリームは、ノードを使用して構築されます。IBM® SPSS® Modeler ウィンドウの下部にあるノード パレットには、ストリームの構築に使われるすべてのノードが用意されています。 [詳細は、3 章 p.16 ノード パレット を参照してください。](#)

ノード パレットは、次の 2 つの方法で再編成することができます。

- パレット マネージャをカスタマイズする。 [詳細は、p.267 パレット マネージャのカスタマイズ を参照してください。](#)
- サブ パレットを含むパレットがノード パレットにどのように表示されるかを変更する。 [詳細は、p.272 サブパレットの作成 を参照してください。](#)

図 12-6
ノード パレットの [レコード設定] タブ



パレット マネージャのカスタマイズ

パレット マネージャを、IBM® SPSS® Modeler の使用方法に合わせてカスタマイズすることができます。たとえば、データベースの時系列データを頻繁に分析するような場合、データベース入力ノード、時間区分ノード、時系列ノード、時系列グラフ ノードがすべて一意のパレットのタブで使用

できるようにしておくくと便利です。パレット マネージャを使用すると、これらのノードをノード パレットのカスタム パレット タブを作成して容易に調整することができます。

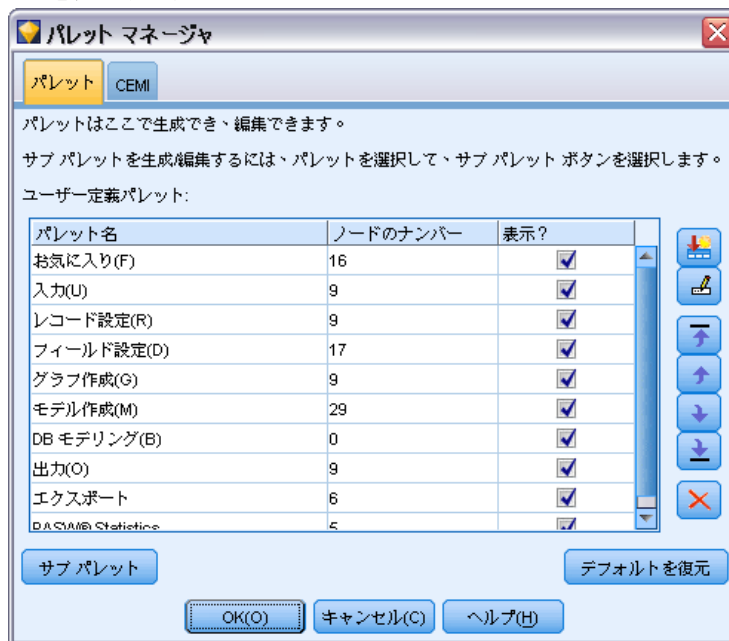
パレット マネージャを使用して、次のようなさまざまなタスクを実行することができます。

- ストリーム領域の下のノード パレットに表示されるパレット タブを制御する。
- パレット タブがノード パレットに表示される順序を変更する。
- パレット タブおよび関連するサブ パレットを作成および編集する。
- [お気に入り] タブのデフォルトのノード選択を編集する。

パレット マネージャにアクセスする手順は、次のとおりです。

- ▶ [ツール] メニューの [パレットの管理] をクリックします。

図 12-7
タブを表示する、ノード パレット上のパレット マネージャ



パレット名: ノード パレット上に表示されているかどうかに関係なく、使用できるパレット タブが表示されます。ここでは、作成したパレット タブが表示されます。 [詳細は、 p. 269 パレット タブの作成 を参照してください。](#)

ノード数: 各パレット タブに表示されるノード数。数が多いほど、タブのノードを分割するサブ パレットを作成しやすくなります。 [詳細は、p.272 サブパレットの作成 を参照してください。](#)

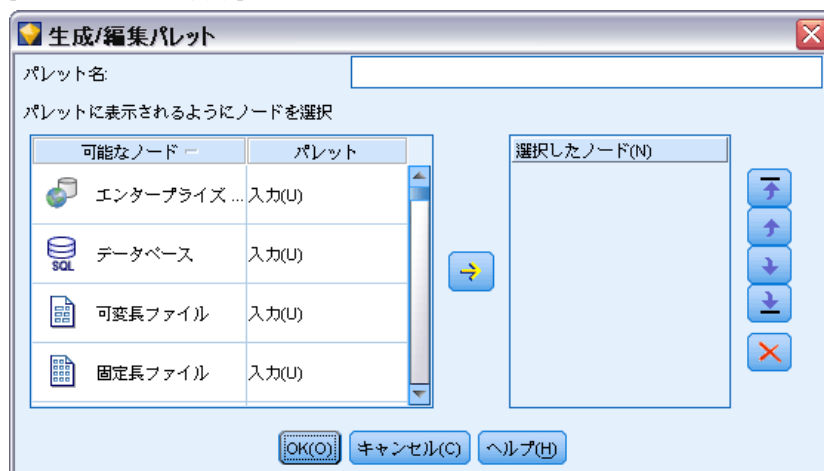
表示 ?: ノード パレット上にパレット タブを表示します。 [詳細は、p.270 ノード パレットのパレット タブの表示 を参照してください。](#)

サブ パレット: パレット タブに表示するサブ パレットを選択するには、該当する [パレット名] 強調表示し、このボタンをクリックすると [サブ パレット] ダイアログ ボックスが表示されます。 [詳細は、 p.272 サブパレットの作成 を参照してください。](#)

復元デフォルト。 このボタンをクリックすると、パレットおよびサブ パレットに行ったすべての変更および追加を完全に削除し、デフォルトのパレット設定に戻します。

パレット タブの作成

図 12-8
[パレットの作成/編集] ダイアログ ボックスでのパレット タブの作成



カスタム パレット タブを作成する手順は、次のとおりです。

- ▶ [ツール] メニューからパレット マネージャを起動します。
- ▶ [表示 ?] 列の右側にある [パレットを追加] ボタンをクリックすると、[パレットの作成/編集] ダイアログ ボックスが表示されます。
- ▶ 一意の [パレット名] に入力します。
- ▶ [使用可能なノード] 領域で、パレット タブに追加されるノードを選択します。

- ▶ [ノードを追加] の右の矢印のボタンをクリックして、強調表示されたノードを [選択されたノード] 領域に移動します。該当するすべてのノードを追加するまで、この手順を繰り返します。

該当するノードをすべて追加した後、次のようにしてパレット タブに表示される順序を変更することができます。

- ▶ 単純な矢印ボタンをクリックすると、ノードが 1 行上または下に移動します。
- ▶ 線の付いた矢印ボタンを使用すると、ノードがリストの最上位または最下位に移動します。
- ▶ ノードをパレットから削除するには、[選択されたノード] 領域の右側にある [削除] ボタンをクリックします。

ノード パレットのパレット タブの表示

IBM® SPSS® Modeler 内に使用しないオプションがある場合があります。この場合、パレット マネージャを使用して、これらのノードを含むタブを非表示にすることができます。

図 12-9
タブを表示する、ノード パレット上のパレット マネージャ



ノード パレットに表示されるタブを選択する手順は、次のとおりです。

- ▶ [ツール] メニューからパレット マネージャを起動します。

- ▶ [表示?] 列のチェック ボックスを使って、各パレット タブを表示するか非表示にするかを指定します。

ノード パレットからパレット タブを完全に削除するには、ノードを強調表示して、[表示 ?] 列の右側にある [削除] ボタンをクリックします。削除したパレット タブを復元することはできません。

注： [お気に入り] タブを除き、SPSS Modeler で提供されたデフォルトのパレット タブを削除することはできません。

ノード パレット上の表示順の変更

表示するパレット タブを選択した後、次のようにしてノード パレットに表示される順序を変更することができます。

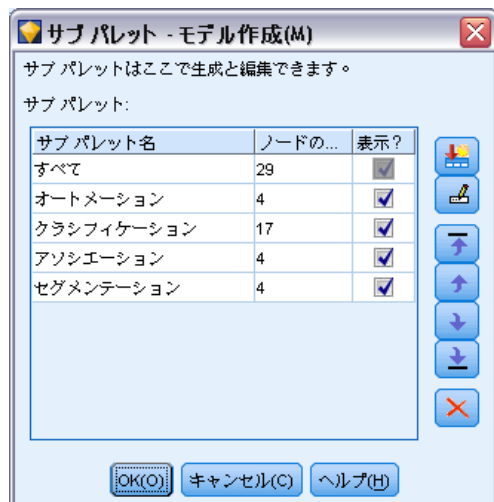
- ▶ 単純な矢印ボタンをクリックすると、パレット タブが 1 行上または下に移動します。上に移動するとノード パレットの左側に移動し、下に移動するとノード パレットの右側に移動します。
- ▶ 線の付いた矢印ボタンを使用すると、パレット タブがリストの最上位または最下位に移動します。リストの最上位にあるタブ パレットは、ノード パレットの左側に表示されます。

[パレット] タブのサブパレットの表示

ノード パレットに表示されるパレット タブを制御するのと同じように、親パレット タブで使用可能なサブ パレットを制御することができます。

図 12-10

[モデル作成パレット] タブで使用可能なサブ パレット



パレット タブで表示するサブ パレットを選択する手順は、次のとおりです。

- ▶ [ツール] メニューからパレット マネージャを起動します。
- ▶ 必要なパレットを選択します。
- ▶ [サブ パレット] ボタンをクリックすると、[サブ パレット] ダイアログ ボックスが表示されます。
- ▶ [表示?] 列のチェック ボックスを使って、各サブパレットを [パレット] タブに追加するかどうかを指定します。[すべて] のサブ パレットは常に表示され、削除することはできません。
- ▶ パレット タブからサブ パレットを完全に削除するには、サブ パレットを強調表示して、[表示 ?] 列の右側にある [削除] ボタンをクリックします。
注： [モデル作成パレット] タブで提供されているデフォルトのサブ パレットを削除することはできません。

パレット タブ上の表示順の変更

表示するサブパレットを選択した後、次のようにして親パレットタブに表示される順序を変更することができます。

- ▶ 単純な矢印ボタンをクリックすると、サブパレットが 1 行上または下に移動します。
- ▶ 線の付いた矢印ボタンを使用すると、サブパレットがリストの最上位または最下位に移動します。

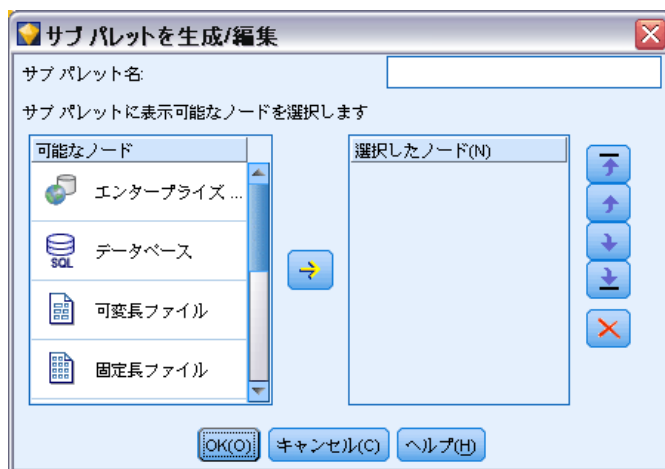
親パレット タブを選択すると、作成したサブ パレットがノード パレットに表示されます。 [詳細は、 p.273 パレット タブの表示の変更 を参照してください。](#)

サブパレットの作成

既存のノードを作成するカスタム パレット タブに追加することができるため、スクロールせずに容易に表示できないほど多くのノードを選択することができます。スクロールしなくてもすむようにするために、パレット タブに選択するノードを配置するサブパレットを作成することができます。たとえば、ストリーム作成に最も頻繁に使用するノードを含むパレット タブを作成する場合、入力ノード、フィールド設定、モデル作成および出力で選択を分割するサブパレットを作成することができます。

注： 親パレット タブに追加されたサブパレット ノードからのみ、選択することができます。

図 12-11
[サブパレットの作成/編集] ダイアログ ボックスでのサブ パレットの作成



サブパレットを作成するには

- ▶ [ツール] メニューからパレット マネージャを起動します。
- ▶ サブパレットを追加するパレットを選択します。
- ▶ [サブ パレット] ボタンをクリックすると、[サブ パレット] ダイアログ ボックスが表示されます。
- ▶ [表示 ?] 列の右側にある [サブ パレットを追加] ボタンをクリックすると、[サブ パレットの作成/編集] ダイアログ ボックスが表示されます。
- ▶ 一意の [サブパレット名] に入力します。
- ▶ [使用可能なノード] 領域で、サブパレットに追加されるノードを選択します。
- ▶ [ノードを追加] の右の矢印のボタンをクリックして、選択されたノードを [選択したノード] に移動します。
- ▶ 該当するノードを追加したら、[OK] をクリックして [サブ パレット] ダイアログ ボックスに戻ります。

親パレット タブを選択すると、作成したサブ パレットがノード パレットに表示されます。 詳細は、 [p.273 パレット タブの表示の変更](#) を参照してください。

パレット タブの表示の変更

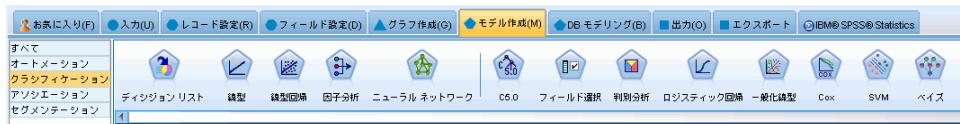
IBM® SPSS® Modeler で使用できるノード数が多すぎるため、ノード パレットの左側または右側へのスクロールが必要ない小さい画面に表示することができません。これは、[モデル作成パレット] タブで特に注意する

必要があります。スクロールを少なくするために、サブパレットに含まれるノードのみを表示することができます（該当する場合）。詳細は、[p.272 サブパレットの作成](#) を参照してください。

パレット タブに表示されるノードを変更するには、パレット タブを選択し、左側のメニューですべてのノードを表示するか、特定のサブパレットのノードのみを表示するかを選択します。

図 12-12

Classification モジュールのサブパレットを表示する [モデル作成パレット] タブ



CEMI ノードの管理

CEMI は廃止され、より柔軟で使いやすい機能セットを提供する CLEF と置き換えられました。詳細については、このリリースで提供されている『IBM® SPSS® Modeler 14.2CLEF Developer's Guide』をご参照ください。

ストリームおよびノードのパフォーマンスの考察

ストリームは、最も効果的な構成でノードを配置したり、必要に応じてノード キャッシュを有効にしたり、このセクションで説明するその他の検討事項に注意を払うことによって、パフォーマンスが最大になるように設計できます。

ここで説明される検討事項のほかに、データベースの効果的な使用、特に SQL 最適化を利用することによって、一般的に、より実質的な追加のパフォーマンス改善が実現します。詳細は、5 章 [パフォーマンスの概要 in IBM SPSS Modeler Server 14.2 管理およびパフォーマンス ガイド](#) を参照してください。

ノードの順序

SQL 最適化を使用していなくても、ストリーム内のノードの順序がパフォーマンスに影響を与えることがあります。一般的な目的は下流の処理を最小化することです、そのため、データの量を削減するノードがある場合、ストリームの初めあたりにそれらを配置します。IBM® SPSS® Modeler Server では、コンパイル時に自動的に順序変更ルールを適用し、安全であると判断した場合に特定のノードを進めることができます(この機能は、デフォルトで有効です。お使いのコンピュータでこの機能が有効であることをシステム管理者に確認してください)。

SQL 最適化を使用する場合は、可用性と効率性を最大限にするように設定します。データベースで実行できない操作がストリームに含まれているときに最適化が停止するので、ストリームの先頭に SQL 最適化の操作をグループ化するのが、もっともよい方法です。このストラテジによりデータベース内での処理が増えるので、IBM® SPSS® Modeler へ持ち込まれるデータが少なくなります。

ほとんどのデータベース内で、次の操作を実行できます。ストリームの「先頭」に次の操作をグループ化するようにしてください。

- キーによる結合
- Select
- グループ集計
- Sort
- サンプル

- レコード追加
- すべてのフィールドが選択される include (含む) モードでの、識別の操作
- 置換操作
- 標準的な算術式または文字列操作を使用した、基本的な算出操作 (どの操作がデータベースにサポートされるかに依存)
- フラグ設定

以下の操作は、ほとんどのデータベース内で実行できません。上記リストの操作の「後」に、次の操作をストリームに配置する必要があります。

- フラット ファイルのような、非データベースのデータに対する操作
 - 順序による結合
 - Balance
 - discard モードでの識別の操作、またはフィールドのサブセットのみが異種として選択された場合の識別の操作
 - 処理対象以外のレコードからのデータへのアクセスが必要な操作
 - ステート型および度数型フィールドの導出
 - 時系列ノードの操作
 - ”@” (時系列) 関数に関連する操作
 - データ型チェック モードの「警告」と「中止」
 - モデル作成、アプリケーション、および分析
- 注： ディジション ツリー、ルールセット、線型回帰、および因子生成のモデルで、SQL を生成し、データベースへプッシュバックできます。
- データを処理中の同じデータベース以外の任意の場所へのデータ出力

ノードのキャッシュ

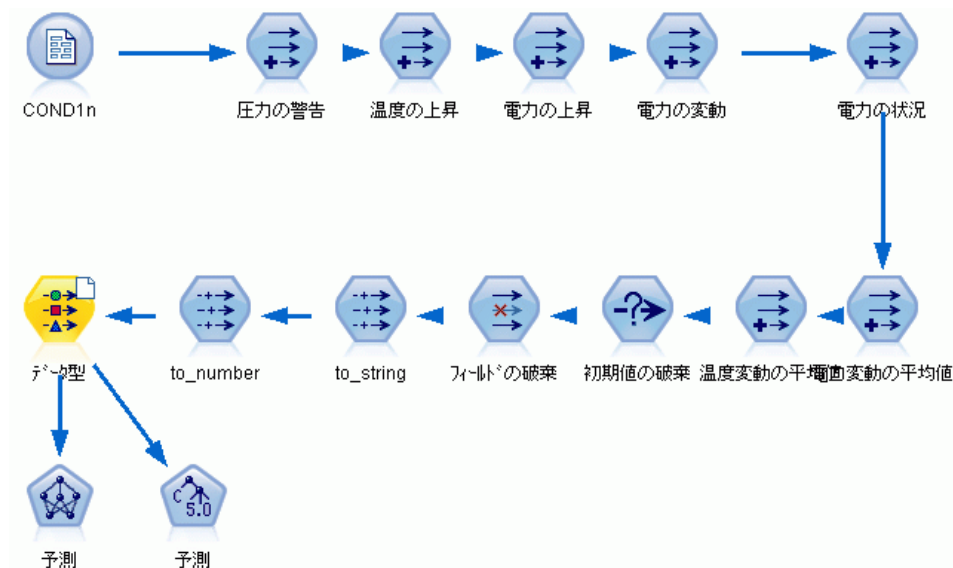
ストリームの実行を最適化するために、ターミナル ノード以外の任意のノードに**キャッシュ**を設定することができます。ノードにキャッシュを設定すると、次のデータ ストリームの実行時にそのノードを通過するデータがキャッシュされます。データがキャッシュされたら、以降のデータはデータ ソースの代わりに、キャッシュ (ディスクの一時ディレクトリに格納) から読み込まれます。

キャッシュは、ソート、結合、集計など、時間のかかる操作に最も有用です。たとえば、データベースから販売データを読み込む入力ノードと、販売地域ごとに集計するレコード集計ノードがあるとします。この場合、データ セット全体ではなく集計済みデータをキャッシュしたいので、入力ノードではなくレコード集計ノードにキャッシュを設定します。

注： 入力ノードでキャッシュする場合、IBM® SPSS® Modeler に読み込まれるように元のデータのコピーを保存するため、多くの環境でパフォーマンスは改善されません。

キャッシュを有効にしたノードの右上には、小さい文書アイコンが表示されます。データがキャッシュされると、この文書アイコンが緑に変わります。

図 13-1
新規作成フィールドを格納するデータ型ノードのキャッシュ



キャッシュを有効にするには

- ▶ ストリーム領域でノードを右クリックし、メニューの [キャッシュ] をクリックします。
- ▶ キャッシュ設定用のサブメニューから、[使用する] を選択します。
- ▶ キャッシュを解除するには、ノードを右クリックして、キャッシュ設定用のサブメニューから [使用しない] を選択します。

データベース内でのノードのキャッシュ

データベース内で実行されるストリームの場合、ファイル システムでなくデータベース内の一時テーブルへ、データを中流でキャッシュできます。この機能を SQL 最適化と組み合わせると、パフォーマンスが著しく向上する可能性があります。たとえば、データ マイニング ビューを作成するために複数のテーブルを併合するストリームからの出力をキャッシュし、必要に応じて再使用できます。すべての下流ノードで自動的に SQL を生成するようにすると、パフォーマンスがさらに向上する可能性があります。

256 文字を超える文字列でキャッシュするデータベースを使用する場合、キャッシュ ノードの上流に、フィールド値が読み込まれるデータ型ノードがあることを確認してください。それにより、一時テーブルの該当する列を適切な幅に設定し、文字列を調整します。

データベースのキャッシングを利用するには、SQL 最適化とデータベースキャッシングの両方を有効にする必要があります。Client の設定は、Server の最適化設定によって無効になるので注意してください。詳細は、[12 章 p.263 最適化オプションの設定](#)を参照してください。

データベースのキャッシングが有効にされると、任意の非ターミナルノードを右クリックするだけでその場所のデータがキャッシュされ、次にストリームが実行されるときに自動的に、データベース内にキャッシュが直接作成されます。データベースキャッシングまたは SQL 最適化が有効にされていないと、キャッシュは、代わりにファイルシステムへ書き出されます。

注：次のデータベースは、キャッシュのために一時テーブルをサポートします。DB2、Netezza、Oracle、SQL Server および Teradata。その他のデータベースでは、データベースキャッシングに通常のテーブルが使用されます。SQL コードを特定のデータベース向けにカスタマイズできます。サポートに連絡してください。

パフォーマンス : プロセス ノード

ソート:ソート ノードでは、ソートする前に入力データ セット全体を読み込む必要があります。データは、一定の限度までメモリーに格納され、超過分はディスクへ書き出されます。ソート アルゴリズムは組合せのアルゴリズムです。データは上限までメモリーに読み込まれ、高速で複合的なクイックソート アルゴリズムを使用してソートされます。すべてのデータがメモリーに収まれば、ソートはこれで完了です。そうでない場合は、マージソートのアルゴリズムが適用されます。ソート済みのデータがファイルへ書き込まれ、次のデータの塊りがメモリーに読み込まれ、ソートされ、ディスクへ書き込まれます。この過程はすべてのデータが読み込まれるまで繰り返され、その後、ソート済みのデータの塊りが併合されます。この併合も、ディスクに格納されたデータに対して繰り返して行うことが必要な場合があります。処理のピーク時には、ソート ノードが、ソート済みと未ソートの完璧な 2 つのデータ セットをディスク上に持つこととなります。

このアルゴリズムの全体実行時間は、約 $N \cdot \log(N)$ です。N は、レコード数です。メモリー内でのソートはディスクからの併合よりも高速なので、ソートにメモリーを多く割り当てるほど、実際の実行時間を減らすことができます。このアルゴリズムは、IBM® SPSS® Modeler Server の構成オプションである「メモリー使用量の乗数」で制御される、物理 RAM の一部をアルゴリズム自体に割り当てます。ソートに使用されるメモリーを増やすには、物理 RAM を増やすか、この値を増やします。使用されるメモリー部分がプロセスの作業セットより大きくなり、メモリーの一部がディ

スクヘページングされると、内部ソート アルゴリズムのメモリーアクセス パターンがランダムになり、過度のページングの原因となるので、パフォーマンスが減衰することを覚えておいてください。ソート アルゴリズムはソート ノード以外の複数のノードでも使用されますが、同じパフォーマンス ルールが適用されます。

データ分割:データ分割ノードは、ビンの境界を計算するために、レコードのビンへの割り当て前に入力データ セット全体を読み込みます。データ セットは、境界の計算中はキャッシュへ入れられ、その後、割り当て用に再スキャンされます。データ分割の手法が「固定幅」または「平均 + 標準偏差」の場合、データ セットはディスクへ直接キャッシュされます。これらの手法には直線的に増加する実行時間が必要であり、また、データ セット全体を格納するのに十分なディスク領域が必要です。データ分割手法が「ランク」または「分位」の場合、データ セットは先に説明したソート アルゴリズムを使用してソートされ、ソート済みのデータ セットがキャッシュとして使用されます。これらの手法のソートの実行時間は $M*N*\log(N)$ です。M はビンに分割されるフィールド数、N はレコード数です。また、データ セットのサイズの 2 倍のディスク領域が必要です。

生成されたビンに基づいてフィールド作成ノードを生成しておく、以後のパスのパフォーマンスが向上します。フィールド作成の速度は、データ分割に比べてかなり速くなります。詳細は、[4 章 生成されたビンのレビュー in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

キーによる結合:結合の手法が「キー」である（データベース結合と同じ）レコード結合ノードでは、入力データ セットの各レコードが、キー フィールドでソートされます。手続き中のこの部分の実行時間は $M*N*\log(N)$ です。M は入力データ セット数、N は最大の入力データ セット内のレコード数です。また、入力データ セットのすべてに加えて最大のデータ セットの 2 番目のコピーを格納するのに十分なディスク領域が必要です。結合自体の実行時間そのものは、出力データ セットのサイズに比例し、キーが一致する頻度に左右されます。最悪の場合は出力が入力のデカルト積であり、実行時間が NM に近づくことがあります。これは稀な例ですが、ほとんどの結合では一致するキーが少ない場合があります。1 つのデータ セットがほかに比べて比較的大きい場合、または、入力データがすでにキー フィールドでソートされている場合は、[最適化] タブを使用して、このノードのパフォーマンスを改善できます。詳細は、[3 章 レコード結合の最適化設定 in IBM SPSS Modeler 14.2 入力ノード、プロセス ノード、出力ノード](#) を参照してください。

レコード集計:[連続キー] オプションが設定されていないと、このノードでは、集計済みの出力を作成する前に入力データ セット全体を読み込みます（ただし、格納しません）。さらに極端な状況、つまり、集計されたデータのサイズが制限（SPSS Modeler Server 構成オプションの「メモリー使用量の乗数」で決定される）に達すると、データセットの残りの部分は、「連続

キー」のオプションが設定されたようにしてソートされ、処理されます。このオプションが設定されていると、集計済みの出力レコードが入力データが読み込まれるごとに作成されるので、データは格納されません。

重複レコード：重複レコード ノードでは、入力データセット内の一意のキーフィールドのすべてが格納されます。すべてのフィールドがキー フィールドですべてのレコードが一意であるような場合、データセット全体が格納されます。デフォルトでは、重複レコード ノードはキー フィールドのデータをソートし、各グループの最初の重複レコードを選択（または破棄）します。重複レコード キーが少ない小さなデータセットの場合、または事前にソートされているデータセットの場合、処理のスピードおよび効率を改善するオプションを選択できます。

タイプ：ある場合には、データ型ノードで値の読み取り時に入力データをキャッシュします。このキャッシュは、下流の処理で使用されます。キャッシュでは、データ セット全体を格納しても処理速度を上げるために、十分なディスク領域が必要です。

評価：評価ノードでは、分位を計算するために入力データをソートする必要があります。スコアと結果のレコード順が各ケースで異なるので、評価済みの各モデルに対してソートが繰り返されます。実行時間は、 $M*N*\log(N)$ です。M はモデル数、N はレコード数です。

パフォーマンス :モデル作成ノード

ニューラルおよび Kohonen：ニューラル ネットワークの学習アルゴリズム (Kohonen アルゴリズムも含む) により、学習データに多くの経路が作成されます。データは、一定の限度までメモリーに格納され、超過分はディスクへ書き出されます。アクセス方法がランダムで、その結果ディスク活動が過度になるため、ディスクから学習データへアクセスすることはコスト的に高くつきます。ノードのダイアログ ボックスの [モデル] タブで [速度優先で最適化] オプションを選択することですべてのデータがメモリーに格納されるようにし、これらのアルゴリズムのディスク領域の使用を無効化することができます。データを格納するのに必要なメモリー量がサーバープロセスの作業セットより大きい場合、メモリーの一部がディスクへページングされて、それによってパフォーマンスが落ちることになります。

[メモリー優先で最適化] が有効にされると、IBM® SPSS® Modeler Server 構成オプションの「モデル作成メモリー制限の割合」の値に従って、物理 RAM の指定されたパーセンテージがアルゴリズムに割り当てられます。学習ニューラル ネットワークにさらに多くのメモリーを使用するには、RAM をさらに増やすか、このオプションの値を増やすかですが、設定値が高すぎると、ページングの原因になります。

ニューラル ネットワーク アルゴリズムの実行時間は、精度のレベルに左右されます。実行時間は、ノードのダイアログ ボックスで停止条件を設定することで制御できます。

K-Means : K-Means クラスタリング アルゴリズムには、ニューラル ネットワーク アルゴリズムと同様の、メモリー使用量を制御するオプションがあります。データへのアクセスがシーケンシャルであるため、ディスクに格納されたデータへのパフォーマンスのほうが良好です。

パフォーマンス :CLEM 式

データ ストリームを調べ直す CLEM シーケンス関数（“@ 関数”）では、最長の参照を可能にする十分なデータを格納する必要があります。参照の程度に制限がない操作では、フィールドのすべての値が格納される必要があります。無制限の操作とは、たとえば `@OFFSET(Sales, Month)` のように、オフセット値がリテラルの整数ではない操作です。オフセット値がフィールド名 `Month` なので、その値は実行されるまで不明です。正確な結果を保証するために、サーバーは `Sales` フィールドのすべての値を保存しておく必要があります。上限値がわかっている場合は、たとえば `@OFFSET(Sales, Month, 12)` のように、追加の引数を指定する必要があります。この指定により、サーバーは `Sales` の 12 の最新値以降を保存しないようになります。制限付きか付きでないかにかかわらず、シーケンス関数によって、ほとんど常に SQL 生成が抑制されます。

IBM SPSS Modeler のアクセス機能

IBM SPSS Modeler のアクセス機能の概要

本リリースでは、ユーザー向けのアクセス機能が大幅に拡張されました。また、視覚障害や他の機能障害をお持ちのユーザー用のサポート機能も用意されています。このセクションでは、画面読み上げソフトウェアやキーボード ショートカットなどのアクセス機能の概要、およびこれらの機能を使った作業方法を説明していきます。

アクセス機能サポートの種類

視覚障害をお持ちの方や、キーボードを使った操作しかできないような方でも利用できるように、さまざまな代替手段が用意されています。たとえば、ストリームの構築、オプションの設定、および出力の読み込みなどの作業を、すべてマウスを使わないで行うことができます。利用できるキーボード ショートカットについては、後述します。また、IBM® SPSS® Modeler は、JAWS for Windows のような画面読み上げソフトウェアもサポートしています。さらに、コントラストを強くするために、カラー スキーマを最適化することもできます。これらのサポート内容の詳細は、以降のトピックで説明していきます。

視力の弱いユーザーのためのアクセシビリティ

ソフトウェアを使いやすくするために、IBM® SPSS® Modeler にはさまざまなプロパティが用意されています。

表示オプション

グラフの表示色を選択することができます。また、ソフトウェア自体に対して特定の Windows 設定を利用することもできます。これは、コントラストを強くするような場合に役立ちます。

- ▶ 表示オプションを設定するには、[ツール] メニューの [ユーザー オプション] をクリックします。
- ▶ [表示] タブをクリックします。このタブには、ソフトウェアの表示色、グラフの色、およびノードのフォント サイズなどを設定できます。

サウンドを通知に使用

サウンドをオンまたはオフにして、ソフトウェアの特定の操作をサウンドで知らせることができます。たとえば、ノードの作成、削除、または新規出力やモデルの生成時に、サウンドを鳴らすことができます。

- ▶ 通知オプションを設定するには、[ツール] メニューの [ユーザー オプション] をクリックします。
- ▶ [通知] タブをクリックします。

新規ウィンドウの自動起動

[ユーザー オプション] ダイアログ ボックスの [通知] タブでは、テーブルやグラフなどの新しく生成された出力を、新しいウィンドウに表示するかどうかを指定することもできます。このオプションを無効にして、必要に応じて出力ウィンドウを表示することもできます。

- ▶ これらのオプションを設定するには、[ツール] メニューの [ユーザー オプション] をクリックします。
- ▶ [通知] タブをクリックします。
- ▶ ダイアログ ボックスで、[ビジュアル通知] のリストから [新規出力] を選択します。
- ▶ [ウィンドウを開く] で、[開かない] を選択します。

ノード サイズ

ノードは、標準のサイズまたは小さいサイズで表示することができます。必要に応じて、大きさを調整してください。

- ▶ ノード サイズ オプションを設定するには、[ファイル] メニューの [ストリームのプロパティ] をクリックします。
- ▶ [レイアウト] タブをクリックします。
- ▶ [アイコン サイズ] リストで、[標準] を選択します。

視覚障害ユーザーのためのアクセシビリティ

視覚障害者向けのサポート機能は、JAWS for Windows のような画面読み上げソフトウェアを使用しているかどうか大きく依存しています。IBM® SPSS® Modeler と一緒に使用する画面読み上げソフトウェアを最大限に活用できるように、多くの設定を指定できます。

表示オプション

スクリーンリーダー（画面読み上げソフトウェア）は、画面のコントラストが強い方が、より正確に動作する傾向があります。Windows をすでに高コントラストの設定にしている場合は、それらの Windows 設定をソフトウェアで使用することもできます。

- ▶ 表示オプションを設定するには、[ツール] メニューの [ユーザー オプション] をクリックします。
- ▶ [表示] タブをクリックします。

サウンドを通知に使用

サウンドをオンまたはオフにして、ソフトウェアの特定の操作をサウンドで知らせることができます。たとえば、ノードの作成、削除、または新規出力やモデルの生成時に、サウンドを鳴らすことができます。

- ▶ 通知オプションを設定するには、[ツール] メニューの [ユーザー オプション] をクリックします。
- ▶ [通知] タブをクリックします。

新規ウィンドウの自動起動

[ユーザー オプション] ダイアログ ボックスの [通知] タブでは、新しく生成された出力を、新しいウィンドウに表示するかどうかを指定することもできます。このオプションを無効にして、必要に応じて出力ウィンドウを表示することもできます。

- ▶ これらのオプションを設定するには、[ツール] メニューの [ユーザー オプション] をクリックします。
- ▶ [通知] タブをクリックします。
- ▶ ダイアログ ボックスで、[ビジュアル通知] のリストから [新規出力] を選択します。
- ▶ [ウィンドウを開く] で、[開かない] を選択します。

キーボードのアクセス機能

製品の機能にキーボードを使用してアクセスできます。基本的には、Alt キーと他の適切なキーを同時に押してメニュー項目を選択したり（例：Alt + F キーで [ファイル] メニューを選択）、Tab キーを使ってダイアログ ボックス中のコントロール間を移動することができます。しかし、それ以外に製品のメイン ウィンドウのそれぞれに関連した固有な問題があり、ダイアログ ボックス内での操作についての便利なヒントもあります。

このセクションでは、ストリームを開く、ノード ダイアログ ボックスの使用、および出力作業などを行うために利用できる、キーボードによるアクセス機能を説明していきます。また、効率的な移動を行うために、キーボード ショートカットの一覧も記載されています。

メイン ウィンドウ内を移動するためのショートカット

大半の作業は IBM® SPSS® Modeler のメイン ウィンドウで行われます。メイン領域は**ストリーム領域**と呼ばれ、データ ストリームの作成と実行に用いられます。ウィンドウの下部にある**ノード パレット**には、利用できるノードが用意されています。パレットは、データ マイニング操作の種類に応じたタブにわかれており、それぞれに適切なノードが配置されています。たとえば、SPSS Modeler にデータを取り込むために使われるノードは、[入力] タブに配置されています。また、フィールドの作成、フィルタリング、およびデータ型の設定に使われるノードは、[フィールド設定] タブに配置されています。

ウィンドウの右側には、ストリーム、出力、およびプロジェクトを管理するためのさまざまなツールが用意されています。ウィンドウ右側の上半分には、ストリーム、出力、および生成されたモデルを管理するために用いられる**マネージャ**があります。このマネージャには、3 つのタブがあります。ストリーム、出力、生成されたモデルなどのオブジェクトにアクセスするには、適切なタブを選択して、リストからオブジェクトを選択します。メイン ウィンドウ右側の下半分には、プロジェクトの作業を行うための**プロジェクト パネル**があります。このエリアには、プロジェクトに 2 の異なるビューが存在することを反映して 2 つのタブがあります。**クラス ビュー**は、プロジェクト オブジェクトをタイプごとにソートしますが、**CRISP-DM ビュー**は、データ準備およびモデル作成などの関連するデータ マイニング フェーズごとにオブジェクトをソートします。これらの SPSS Modeler ウィンドウの各部の詳細は、ヘルプやマニュアルを参照してください。

SPSS Modeler のメイン ウィンドウで、およびストリームの構築に使用できるショートカットの一覧を次の表に示します。ダイアログ ボックスや出力で用いられるショートカットは後述します。これらのショートカットキーは、メイン ウィンドウでしか利用できないことに注意してください。

メイン ウィンドウのショートカット

ショートカット キー	Function
Ctrl+F5	フォーカスをノード パレットに移動します。
Ctrl+F6	フォーカスをストリーム領域に移動します。
Ctrl+F7	フォーカスをマネージャ パネルに移動します。
Ctrl+F8	フォーカスをプロジェクト パネルに移動します。

ノードとストリームのショートカット

ショートカット キー	Function
Ctrl+N	ストリーム領域に新しい空のストリームを作成します。
Ctrl+O	既存のストリームを選択して開くことのできる [開く] ダイアログボックスを表示します。
Ctrl+数字キー	フォーカスをウィンドウまたはペインの対応するタブに移動します。たとえば、タブのある領域やウィンドウで、Ctrl-1 を押すと左から 1 番目のタブが、Ctrl-2 を押すと 2 番目のタブが選択されます。
Ctrl+下矢印キー	ノード パレットでフォーカスをパレット タブからそのタブの最初のノードまで移動するのに使用します。
Ctrl+上矢印キー	ノード パレットでフォーカスをノードからそのパレット タブまで移動するのに使用します。
Enter	ノード パレット ([生成されたモデル] パレットの調整済みモデルも含めて) でノードが選択されている場合に、このキーを押すとノードがストリーム領域に追加されます。キャンバスでノードがすでに選択されている場合に Enter を押すと、そのノードのダイアログ ボックスが開きます。
Ctrl+Enter	パレットでノードが選択されている場合に、ノードをストリーム領域に追加します。ただし、追加したノードは選択されず、フォーカスはノード パレットのまま変わりません。
Alt+Enter	パレットでノードが選択されている場合に、ノードをストリーム領域に追加して選択します。ただしフォーカスはノード パレットのまま変わりません。
Shift+スペースキー	ノードまたはコンポーネントがパレットにフォーカスを置く場合、ノードまたはコンポーネントの選択と非選択を切り替えます。 ほかのノードまたはコンポーネントも選択されている場合、これらの選択は解除されます。
Ctrl+Shift+スペースキー	ノードまたはコメントがストリームにフォーカスを置く場合、またはノードまたはコメントがパレットにフォーカスを置く場合、ノードまたはコメントの選択と非選択を切り替えます。 そのほかの選択されたノードまたはコメントに影響はありません。
左右の矢印キー	ストリーム領域にフォーカスがある場合、ストリーム全体を画面上で平行に移動します。パレット タブにフォーカスがある場合、タブの間を順次移動します。パレット ノードにフォーカスがある場合、パレットのノード間を移動します。

ショートカット キー	Function
上下の矢印キー	ストリーム領域にフォーカスがある場合、ストリーム全体を画面上で平行に移動します。パレット ノードにフォーカスがある場合、パレットのノード間を移動します。サブパレットにフォーカスがある場合、このパレット タブに対して他のサブパレット間を移動します。
Alt+左右の矢印キー	ストリーム領域上で選択されているノードまたはコメントを、矢印キーの方向に平行に移動します。
Alt+上下の矢印キー	ストリーム領域上で選択されているノードまたはコメントを、矢印キーの方向で垂直方向に移動します。
Ctrl+A	ストリーム内のすべてのノードを選択します。
Ctrl+Q	ノードにフォーカスがある場合に、そのノードと下流にあるすべてのノードを選択し、上流にあるすべてのノードを非選択にします。
Ctrl+W	ノードにフォーカスがある場合に、そのノードと下流にあるすべての選択したノードを非選択にします。
Ctrl+Alt+D	選択したノードを複製します。
Ctrl+Alt+L	ストリームでモデル ナゲットが選択されている場合、[挿入] ダイアログを開いて .nod ファイルからストリームへ保存したモデルをロードできるようにします。
Ctrl+Alt+R	選択したノードの [注釈] タブを表示し、ノードの名前が変更できるようになります。
Ctrl+Alt+U	ユーザー入力ソース ノードを生成します。
Ctrl+Alt+C	ノードのキャッシュのオンとオフを切り替えます。
Ctrl+Alt+F	ノードのキャッシュを消去します。
Tab	ストリーム領域で、現在のストリームのすべての入力ノードおよびコメントを順次移動します。選択したサブパレットで、サブパレットのサブパレット ラベルと最初のノードの間でフォーカスを切り替えます。
Shift+Tab	Tab と同じ操作を実行しますが、順序が逆になります。
Ctrl+Tab	ストリーム領域、マネージャ パネルまたはプロジェクト パネルにフォーカスを置き、ノード パレットにフォーカスを移動します。
アルファベット キー	現在のストリームのノードにフォーカスがある状態から、名前が押したキーの文字で始まる次のノードにフォーカスを移動します。
F1	ヘルプのフォーカスに関連したトピックを開きます。
F2	ストリーム領域上で選択されているノードの接続処理を開始します。Tab キーを使って接続先のノードを選択し、Shift + スペース バーを押すと接続が行われます。

ショートカット キー	Function
F3	ストリーム領域上で選択されているノードに対するすべての接続を削除します。
F6	マネージャ パネル、プロジェクト パネルおよびノード パレット間でフォーカスを移動します。
F10	[ファイル] メニューを表示します。
Shift+F10	ノードまたはストリームのポップアップ メニューを表示
Del	キャンバスから選択したノードを削除します。
Esc	ポップアップ メニューまたはダイアログ ボックスを閉じます。
Ctrl+Alt+X	スーパーノードを展開します。
Ctrl+Alt+Z	スーパーノードをズーム インします。
Ctrl+Alt+Shift+Z	スーパーノードからズーム アウトします。
Ctrl+E	ストリーム領域上にフォーカスを置きながら、現在のストリームを実行します。

ほかにも、コピーの場合は Ctrl-C キーなど、SPSS Modeler ではさまざまな標準のショートカットキーを利用することができます。詳細は、3 章 p.25 ショートカット キーの使用 を参照してください。

ダイアログ ボックスや表のショートカット キー

ダイアログ ボックス、表、およびダイアログ ボックス中の表に関する作業を行う際に役立つさまざまなショートカットキーや画像読み上げソフトウェア用キーが用意されています。これらのキーの一覧を次に示します。

ダイアログ ボックスと Clem 式ビルダーのショートカット

ショートカット キー	Function
Alt+4	開かれているすべてのダイアログ ボックスまたは出力ウィンドウを閉じるために用いられます。出力は、マネージャ パネルの [出力] タブから取得することができます。
Ctrl+End	Clem 式ビルダー内のコントロールにフォーカスがある場合に、挿入ポイントを式の最後に移動します。
Ctrl+1	Clem 式ビルダーで、フォーカスを CLEM 式編集コントロールに移動します。
Ctrl+2	Clem 式ビルダーで、フォーカスを関数リストに移動します。
Ctrl+3	Clem 式ビルダーで、フォーカスをフィールドリストに移動します。

テーブルのショートカット

テーブルのショートカットは、出力テーブルや、データ型、フィルタ、およびレコード結合などのノードのダイアログ ボックス中にあるテーブルコントロールで用いられます。一般的には、テーブル中のセル間の移動には Tab キーを、テーブル コントロールを終了するには Ctrl-Tab キーを使用します。注： 場合によっては、画面読み上げソフトウェアがセルの内容をすぐに読み上げ始めないこともあります。矢印キーを 1 回または 2 回押すと、ソフトウェアがリセットされて、読み上げが開始されます。

ショートカット キー	Function
Ctrl+W	テーブルで選択されている行 (row) の簡単な説明を読み上げます。たとえば、「選択されている行 2 の値は、性別、フラグ型、男/女・・・」のように読み上げます。
Ctrl+Alt+W	テーブルで選択されている行 (row) の長い説明を読み上げます。たとえば、「選択されている行 2 の値は、フィールド = 性別、データ型 = フラグ型、性別 = 男/女・・・」のように読み上げます。
Ctrl+D	テーブルで選択されている領域の、簡単な説明を読み上げます。たとえば、「列 6 の 1 行目が選択されています。」のように読み上げます。
Ctrl+Alt+D	テーブルで選択されている領域の、長い説明を提供します。たとえば、「1 行 6 列が選択されています。選択されている列は、フィールド、データ型、欠損値です。選択された行は 1 です。」のようになります。
Ctrl+T	テーブルで選択されている領域の、簡単な説明を、n 供します。たとえば、「フィールド、データ型、欠損値」のようになります。
Ctrl+Alt+T	テーブルで選択されている領域の、長い説明を提供します。たとえば、「選択されている列は、フィールド、データ型、欠損値です。」のようになります。
Ctrl+R	テーブル中のレコード数を提供します。
Ctrl+Alt+R	テーブルの場合に、テーブル中のレコード数、および列名を提供します。
Ctrl+I	テーブルに対して、フォーカスのあるセルのフィールド情報または内容を読み上げます。
Ctrl+Alt+I	テーブルに対して、フォーカスがあるセルのセル情報の長い説明 (列名とセルの内容) を読み上げます。
Ctrl+G	テーブルで、簡単な一般選択情報を提供します。

ショートカット キー	Function
Ctrl+Alt+G	テーブルで、長い一般選択情報を提供します。
Ctrl+Q	テーブル中のセルを早く切り替えることができます。Ctrl-Q を押すと、テーブル内を矢印キーを使って移動すれば、その詳細な説明が「性別 = 女性」のように読み上げられます。もう一度 Ctrl-Q を選択すると、簡単な説明に切り替わります（セルの内容）。

コメントのショートカット

画面上のコメントを処理している場合、次のショートカットを使用できます。

ショートカット キー	Function
Alt+C	コメントの表示/非表示機能を切り替えます。
Alt+M	コメントが現在表示されている場合は新しいコメントを挿入し、非表示になっている場合はコメントを表示します。
Tab	ストリーム領域で、現在のストリームのすべての入力ノードおよびコメントを順次移動します。
Enter	コメントにフォーカスがある場合、編集の開始を指示します。
Alt+Enter または Ctrl+Tab	編集を終了し、編集の変更を保存します。
Esc	編集を中止します。編集集中に行われた変更は失われます。
Alt+Shift+上矢印キー	グリッドによる配置の設定がオン（またはオフ）の場合、1 グリッド セル（または 1 ピクセル）ごとにテキスト エリアの高さを縮小します。
Alt+Shift+下矢印キー	グリッドによる配置の設定がオン（またはオフ）の場合、1 グリッド セル（または 1 ピクセル）ごとにテキスト エリアの高さを拡大します。
Alt+Shift+左矢印キー	グリッドによる配置の設定がオン（またはオフ）の場合、1 グリッド セル（または 1 ピクセル）ごとにテキスト エリアの幅を縮小します。
Alt+Shift+右矢印キー	グリッドによる配置の設定がオン（またはオフ）の場合、1 グリッド セル（または 1 ピクセル）ごとにテキスト エリアの幅を拡大します。

クラスタ ビューアおよびモデル ビューアのショートカット

ショートカット キーは、クラスタ ビューアおよびモデル ビューア ウィンドウを移動するために使用できます。

全般 - クラスタ ビューアおよびモデル ビューア

ショートカット キー	Function
Tab	フォーカスを次の画面コントロールに移動します。
Shift+Tab	フォーカスを前の画面コントロールに移動します。
下矢印	ドロップダウン リストにフォーカスがある場合、リストを開くか、リストの次の項目に移動します。 メニューにフォーカスがある場合、メニューの次の項目に移動します。 サムネイル グラフにフォーカスがある場合、セット内の次のサムネイル（または最後のサムネイルにフォーカスがある場合は最初のサムネイル）に移動します。
上矢印	ドロップダウン リストにフォーカスがある場合、リストの前の項目に移動します。 メニューにフォーカスがある場合、メニューの前の項目に移動します。 サムネイル グラフにフォーカスがある場合、セット内の前のサムネイル（または最初のサムネイルにフォーカスがある場合は最後のサムネイル）に移動します。
Enter	使用中のドロップダウン リストを閉じるか、または使用中のメニューで選択を行います。
F6	ウィンドウの左側および右側のパネルでフォーカスを切り替えます。
左右の矢印キー	タブにフォーカスがある場合、メニューの前後のタブに移動します。 メニューにフォーカスがある場合、前後のメニューに移動します。
Alt+文字キー	その名前で下線の付いた文字を含むボタンまたはメニューを選択します。
Esc	使用中のメニューまたはドロップダウン リストを閉じます。

クラスタ ビューアのみ

クラスタ ビューアには、クラスタ-特徴グリッドを含むクラスタ ビューアがあります。詳細は、[11 章 クラスタ ビュー in IBM SPSS Modeler 14.2 Modeling Nodes](#) を参照してください。

モデルの要約ビューの代わりにクラスタ ビューアを選択する手順は、次のとおりです。

- ▶ **[表示]** ボタンが選択されるまで、Tab キーを繰り返し押します。

A 付録

- ▶ 下向き矢印を 2 回押して、[クラスタ] を選択します。
ここから、グリッド内の個別のセルを選択できます。
- ▶ 視覚化ツールバーの最後のアイコンが選択されるまで、Tab キーを繰り返し押しします。

図 A-1
[視覚化ツリーを表示] アイコン



- ▶ Tab キーをもう一度押し、その後スペース、そして矢印キーを押します。

次のキーボード ショートカットを使用できます。

ショートカット キー	Function
矢印キー	グリッド内の各セル間でフォーカスを移動します。右側のパネルのセルの分布表示が、フォーカスが移動するごとに変わります。
Ctrl+カンマ	セルにフォーカスがあるグリッドの列全体を選択または選択解除します。列を選択部分に追加するには、矢印キーを使用して、その列のセルに移動し、もう一度 Ctrl キーを押しながらカンマを押します。
Tab	フォーカスをグリッドから次の画面コントロールに移動します。
Shift+Tab	フォーカスをグリッドから前の画面コントロールに移動します。
F2	編集モード（ラベルおよびセル名のセルのみ）に入ります。
Enter	編集した変更を保存し、編集モード（ラベルおよびセル名のセルのみ）を終了します。
Esc	変更を保存せずに編集モード（ラベルおよびセル名のセルのみ）を終了します。

ショートカット キーの例 :ストリームの構築

キーボードや画面読み上げソフトウェアを使用した、ユーザーのストリームの構築作業を理解するために、マウスを使わないでストリームを構築する例を取り上げます。この例では、可変長ファイルノード、フィールド作成ノード、およびヒストグラム ノードがあるストリームを、次の手順にしたがって作成していきます。

- ▶ **SPSS Modeler の起動** :IBM® SPSS® Modeler を初めて起動した場合、フォーカスはノード パレットの [お気に入り] タブにあります。
- ▶ **Ctrl+下矢印キー** : フォーカスをタブ自体から、タブの内容へと移動します。

- ▶ **右矢印キー**：フォーカスを可変長ファイル ノードに移動します。
- ▶ **スペース バー**：可変長ファイル ノードを選択します。
- ▶ **Ctrl+Enter**：可変長ファイル ノードをストリーム領域に追加します。このキーの組み合わせでは、可変長ファイル ノードが選択されたままになるため、次のノードを追加すると可変長ノードに接続されます。
- ▶ **Tab**：フォーカスをノード パレットに移動します。
- ▶ **右矢印キーを 4 回**：フィールド作成ノードを移動します。
- ▶ **スペース バー**：フィールド作成ノードを選択します。
- ▶ **Alt+Enter**：フィールド作成ノードをストリーム領域に追加して、フィールド作成ノードを選択します。これで、このノードは次に追加されるノードと接続されることとなります。
- ▶ **Tab**：フォーカスをノード パレットに移動します。
- ▶ **右矢印キーを 5 回**：フォーカスがパレット中のヒストグラム ノードに移動します。
- ▶ **スペース バー**：ヒストグラム ノードを選択します。
- ▶ **Enter**：ノードをストリームに追加し、フォーカスをストリーム領域に移動します。

次の例に進むか、後で次の例を試す場合はストリームを保存します。

ショートカット キーの例 :ノードの編集

この例では、前の例で作成したストリームを使用します。このストリームは、可変長ファイル ノード、フィールド作成ノード、およびヒストグラム ノードから成り立っています。この例では、ストリーム中の 3 番目にあるヒストグラム ノードにフォーカスがある所から始まります。

- ▶ **Ctrl+右矢印キーを 2 回**：フォーカスを可変長ファイル ノードに移動します。
- ▶ **Enter**：[可変長ファイル] ダイアログ ボックスをオープンします。[ファイル] フィールドに Tab キーで移動し、テキスト ファイルのパスおよび名前を入力して、そのファイルを選択します。Ctrl+Tab キーでダイアログの下部分に移動し、Tab キーで [OK] ボタンに移動して Enter キーを押すとダイアログ ボックスが閉じます。
- ▶ **Ctrl+右矢印キー**：フォーカスを 2 番目のフィールド作成ノードに移動します。
- ▶ **Enter**：[フィールド作成ノード] ダイアログ ボックスが表示されます。Tab キーを何回か押してフィールドを選択し、フィールド作成条件を指定し

ます。Ctrl+Tab キーで [OK] ボタンまで移動し、Enter キーを押すとダイアログ ボックスが閉じます。

- ▶ **Ctrl+右矢印キー**：フォーカスを 3 番目のヒストグラム ノードに移動します。
- ▶ **Enter**：[ヒストグラム ノード] ダイアログ ボックスが表示されます。Tab キーを何回か押してフィールドを選択し、グラフ オプションを指定します。ドロップダウン リストで、下向き矢印キーを押すとリストが開いてリスト項目を強調表示し、Enter キーを押してリスト項目を選択します。-Tab キーで [OK] ボタンまで移動し、Enter キーを押すとダイアログ ボックスが閉じます。

この時点で、他のノードを追加したり、現在のストリームを実行することができます。ストリームを構築するときには、忘れずに次のヒントに従ってください。

- 手作業でノードを接続する場合、F2 キーを使って接続の始点を作成し、Tab キーを使用して終点まで移動し、Shift キーとスペースバーを使用して接続を完了します。
- ストリーム領域で選択されているノードのすべての接続を解除するには、F3 キーを使用します。
- ストリームを作成した後、Ctrl + E キーを押すと現在のストリームを実行します。

ショートカット キーの完全な一覧が用意されています。詳細は、[p. 285 メイン ウィンドウ内を移動するためのショートカット](#) を参照してください。

画面読み上げソフトウェアの使用

多数の画面読み上げソフトウェアが市場で販売されています。IBM® SPSS® Modeler は SPSS Modeler と一緒にインストールされる Java Access Bridge を使用して Windows JAWS に対応するよう設定されています。JAWS がインストール済みであれば、SPSS Modeler を起動する前に JAWS を起動するだけで利用できます。

SPSS Modeler 独自のデータマイニングプロセスのグラフィック表示のため、グラフは視覚的に使用するために最適化されています。ただし、画面読み上げソフトウェアを使用して「文字的に」表示できる出力やモデルを理解しながら、作業を行うことができます。

注：64 ビットのクライアントコンピュータでは、支援技術機能が機能しません。これは Java Access Bridge が 64 ビットの操作向けに設計されていないためです。

IBM SPSS Modeler ディクショナリ ファイルの使用

SPSS Modeler ディクショナリ ファイル (Awt. JDF) は、JAWS に含めて利用することができます。このファイルを使用するには

- ▶ SPSS Modeler のインストール ディレクトリ中の /accessibility サブディレクトリに移動して、ディクショナリ ファイル (Awt. JDF) をコピーします。
- ▶ コピーしたファイルを、JAWS スクリプトのあるディレクトリに貼り付けます。

他の JAVA アプリケーションを稼働させている場合は、Awt. JDF という名前のファイルがすでに存在していることもあります。この場合、手作業でディクショナリ ファイルを編集しないと、このファイルを使用できないこともあります。

HTML 出力での画面読み上げソフトウェアの使用

IBM® SPSS® Modeler で画面読み上げソフトウェアを使って HTML 出力を表示する場合、いくつかの問題が発生する可能性があります。次のように、多数の出力タイプが影響を受けます。

- 線形回帰ノード、ロジスティック回帰ノード、および因子分析ノードの [詳細] タブに表示する出力。
- レポート ノードの出力

これらの各ウィンドウまたはダイアログ ボックスには、出力をデフォルトのブラウザに表示するためのツールがツールバーに用意されています。このデフォルトのブラウザは、標準の画面読み上げソフトウェアをサポートしています。ブラウザに表示した出力情報を、画面読み上げソフトウェアを使って調査します。

インタラクティブ ツリー ウィンドウのアクセス

インタラクティブ ツリー ウィンドウ内でのディシジョン ツリーの標準的な表示方法は、画面読み上げソフトウェアを使用する際に問題になることがあります。アクセス可能なバージョンにアクセスするには、インタラクティブ ツリーのメニューから次のメニュー項目を選択してください。

表示 > [アクセス可能なウィンドウ(A)]

標準のツリー マップと似ていますが、JAWS が正しく読み上げることができるビューを表示します。標準的な矢印キーを使用すると、上下左右に移動できます。アクセス可能なウィンドウ内を移動すると、それに応じてインタラクティブ ツリー ウィンドウ内のフォーカスも移動します。選択を

変更するにはスペースバーを使用します。また、現在の選択を拡張するには、Ctrl+スペースを使用します。

使用上のヒント

IBM® SPSS® Modeler 環境をより使いやすくするための、さまざまなコツがあります。SPSS Modeler で作業を行う際の一般的なヒントを次に示します。

- **拡張テキストボックスの終了**：拡張テキストボックスを終了するには Ctrl+Tab を使用します。注：Ctrl-Tab は、テーブルコントロールを終了する際にも用いられます。
- **矢印キーよりも Tab キーを利用する**：ダイアログボックスのオプションを選択する場合、オプションボタン間の移動には Tab キーを使用します。ここでは、矢印キーを移動に使用することはできません。
- **ドロップダウンリスト**：ダイアログボックスのドロップダウンリストでは、Esc キーまたはスペースバーを使って項目を選択し、リストを閉じることができます。他のコントロールにタブで移動しており、閉じていないドロップダウンリストを閉じるには、Esc キーを使用することもできます。
- **実行ステータス**：大きいデータベースでストリームを実行している場合、JAWS ではストリームステータスの読み上げを遅くすることができます。ステータスレポートを更新するには、Ctrl キーを定期的に押してください。
- **ノードパレットの使用**：ノードパレットのタブを初めて参照する場合、JAWS がノード名の代わりに「グループボックス」と読み上げることがあります。この場合、Ctrl-右矢印キーまたは Ctrl-左矢印キーを使って、画面読み上げソフトウェアをリセットしてから、ノード名を読み上げてください。
- **メニューの読み上げ**：最初にメニューを開く際に、JAWS が最初のメニュー項目を読み上げないことがあります。この問題が発生している可能性がある場合は、下矢印キーを押した後に上矢印キーを押して、最初のメニュー項目を確認してください。
- **カスケードメニュー**：JAWS は、カスケードメニューのメイン（第 1 レベル）メニューは読み上げません。メニュー内を移動している際に、何も読み上げられないことがあった場合は、右矢印キーを押して子メニューの項目を読み上げてください。

また、IBM® SPSS® Text Analytics をインストールした場合、次のヒントでインタラクティブワークベンチインターフェイスがより使いやすくなります。

- **ダイアログボックスの入力**：ダイアログボックスに入力直後、Tab キーを押して最初のコントロールに焦点を当てる必要がある場合があります。

- **拡張テキスト ボックスの終了:** Ctrl+Tab キーを押して、拡張テキスト ボックスを終了し、次のコントロールに移動します。注: Ctrl-Tab は、テーブル コントロールを終了する際にも用いられます。
- **最初の文字を入力してツリー リストの要素を検索:** カテゴリ領域、抽出された結果の領域、またはライブラリ ツリーで要素を検索する場合、領域にフォーカスがある際に要素の最初の文字を入力することができます。これにより、入力した文字で始まる要素の次の出現を選択します。
- **ドロップダウン リスト:** ダイアログ ボックスのドロップダウン リストでは、スペース バーを使って項目を選択し、リストを閉じることができます。

使用上のその他のヒントは、以降のトピックで詳細に説明します。

他のソフトウェアとの相性

弊社開発チームが IBM® SPSS® Modeler で JAWS などの画面読み上げソフトウェアをテストした際に、SMS (Systems Management Server) を使用していると、JAWS による SPSS Modeler などの Java ベース アプリケーションの読み上げが正常に行われられない可能性があることがわかりました。SMS を無効化するとこの問題を解決できます。SMS の詳細は、Microsoft の Web サイトを参照してください。

JAWS および Java

JAWS のバージョンによって、Java ベースのソフトウェア アプリケーションのサポート レベルも異なります。IBM® SPSS® Modeler は最近のバージョンのすべての JAWS で利用できますが、バージョンによっては Java ベースのシステムで使用する際に小さな問題が発生する場合があります。詳細は、JAWS for Windows の Web サイト <http://www.FreedomScientific.com> を参照してください。

IBM SPSS Modeler でのグラフの使用

ヒストグラム、評価グラフ、線グラフ、および散布図のような、視覚的に表示された情報を、画面読み上げソフトウェアで確認することは困難です。ただし、Web グラフや棒グラフの場合は、出力ウィンドウで利用できるテキスト要約情報を参照することができます。

Unicode のサポート

IBM SPSS Modeler 内での Unicode のサポート

IBM® SPSS® Modeler は、IBM® SPSS® Modeler と IBM® SPSS® Modeler Server の両方で、Unicode を完全にサポートしています。これにより、多言語データベースなどの Unicode をサポートする他のアプリケーションと、ロケール固有のコード化スキーム間で変換を行ったときに生じる情報の欠落なしに、データの交換が可能になります。

- SPSS Modeler は、内部でデータを Unicode で保存しており、データベースに Unicode で保存されている多言語データを、情報の失うことなく読み取ったり、書き込んだりすることができます。
- SPSS Modeler は、UTF-8 エンコード テキスト ファイルを読み書きできます。テキスト ファイルのインポートとエクスポートのデフォルトは、ロケールのエンコードになっていますが、それ以外に UTF-8 もサポートされています。この設定は、ファイル インポートおよびエクスポート ノードで指定できます。また、[ストリームのプロパティ] ダイアログ ボックスで、デフォルトのエンコードを変更できます。[詳細は、5 章 p.58 ストリームのオプションの設定 を参照してください。](#)
- ロケールのエンコードで保存されている Statistics、SAS、およびテキスト データ ファイルは、インポート時に UTF-8 に変換され、エクスポート時に元に戻されます。任意のファイルに書き込む場合、そのロケール文字セットに対応する文字が存在しない Unicode 文字がある場合、別の文字で置換され、また、警告が表示されます。この問題は、データが、Unicode をサポートするデータ ソース（データベースまたは UTF-8 テキスト ファイル）からインポートされ、さらに、そのデータにロケールが異なる文字や、複数のロケールまたは文字セットからの文字が含まれている場合についてのみ発生します。
- IBM® SPSS® Modeler Solution Publisher のイメージは、UTF-8 でエンコードされており、プラットフォームおよびロケール間で完全に移植可能です。

Unicode について

Unicode 規格の目標は、多言語テキストの一貫性のあるエンコード法を提供することです。これにより、国、地域、およびアプリケーションを越えた共有が容易になります。Unicode 規格は現在、バージョン 4.0.1 ですが、世界中で一般的に使用されているすべての文字セットのスーパーセットの文字セットを定義しており、それぞれの文字に一意的な名前とコード ポイントを割り当てています。文字とそのコード ポイントは、ISO-10646 により

定義された Universal Character Set (UCS) のものと同一です。詳細は、[Unicode ホームページ \(http://www.unicode.org\)](http://www.unicode.org) を参照してください。

注意事項

This information was developed for products and services offered worldwide.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive,
Armonk, NY 10504-1785, U. S. A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing, Legal and Intellectual Property
Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa
242-8502 Japan.

次の文は、条項が法律と一致しないイギリスなどの国には適用されません。 本出版物は、SPSS INC., AN IBM COMPANY によって提供され、明示的および暗黙的なあらゆる保証、制限されていない場合を除く商品性や特定の目的への適合性、および無違反に関する暗黙的な保証を含む意思表示と保証を放棄します。特定の取引では明示的または暗黙的な保証の免責が許可されないため、この文が適用されない場合があります。

この情報には、技術的な誤りや誤植を含まれる場合があります。本文では変更が定期的に行われます。これらの変更は本書の次の版に組み込まれます。SPSS は、本文書に記載された製品やプログラムは予告なしに改善または変更される場合があります。

この情報内にある SPSS 以外または IBM 以外の Web サイトに対する参照は、便宜上提供されたものであり、これらの Web サイトを推奨するものではありません。これらの Web サイトの資料は、この SPSS 社製品の使用の一部ではなく、これらの Web サイトの使用は個人の責任によるものです。

IBM または SPSS に情報を送信すると、あなたに対する義務を負うことなく、適切とする方法でその情報を使用または配布する非独占的権利と IBM および SPSS 付与するものとします。

SPSS 以外の製品に関する情報は、これらの製品、公開された通知、公表されているソースの供給者から得たものです。SPSS は、それらの製品をテストしていません。また、SPSS 以外の製品に関連するパフォーマンスの正確性、互換性、またはその他の要求を確認することはできません。SPSS 以外の製品の機能に関する質問は、これらの製品の供給者にお問い合わせください。

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the

capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM' s future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

この情報には、日常の業務で使用されているデータおよびレポートの例が含まれています。それらを可能な限り詳細に説明するために、例には個人、企業、ブランド、製品の名前が含まれます。これらの名前はすべて架空のものであり、実際の名前や住所に似ているものでも、まったくの偶然によるものです。

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

商標

IBM、IBM ロゴ、ibm.com は世界各国の四方に基づく IBM 社の登録商標です。IBM の商標の現在のリストは Web サイト <http://www.ibm.com/legal/copytrade.shtml> を参照してください。

SPSS Inc., an IBM Company の SPSS の商標 は、世界各国の司法に基づく登録商標です。

Adobe、Adobe のロゴ、PostScript、および PostScript ロゴはアメリカ合衆国およびその他各国のアドビシステムズ社の登録商標または商標です。

IT Infrastructure Library は、イギリス商務局の一部である中央電子計算機局の登録商標です。

Intel、Intel のロゴ、Intel Inside、Intel Inside のロゴ、Intel Centrino、Intel Centrino のロゴ、Celeron、Intel Xeon、Intel SpeedStep、Itanium、Pentium はアメリカ合衆国およびその他各国のインテル社およびその子会社の商標または登録商標です。

Linux は、アメリカ合衆国およびその他各国の Linus Torvalds の登録商標です。

Microsoft、Windows、Windows NT、および Windows ロゴは、アメリカ合衆国およびその他合衆国のマイクロソフト社の商標です。

ITIL は、米国特許商標局の登録商標および登録共同体商標です。

UNIX は、アメリカ合衆国およびその他各国の The Open Group の登録商標です。

Cell Broadband Engine は、アメリカ合衆国およびその他各国のソニーコンピュータエンタテインメント株式会社の使用許諾に基づいて使用されています。

Java および Java ベースの商標およびロゴは、アメリカ合衆国およびその他各国のサン・マイクロシステムズ株式会社の商標です。

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

その他の製品およびサービス名は、IBM、SPSS、またはその他の企業の商標である場合があります。

索引

- 複数二分法設定
 - CLEM 式内, 126
- 科学的表記法
 - 表示書式, 60
- 非等号演算子, 147
- 小数点記号, 58
 - 数字の表示フォーマット, 60
- 文字列関数, 154
- 等号演算子, 147
- 三角関数, 151
- 並行処理
 - 有効化, 263
- 優先順位, 141
- 分布関数, 152
- 変換関数, 146
- 小数桁数
 - 表示書式, 60
- 情報関数, 145
- 指数関数, 150
- 数値関数, 150
- 日付関数, 139-140
 - date_before, 147, 160
 - date_days_difference, 160
 - date_in_days, 160
 - date_in_months, 160
 - date_in_weeks, 160
 - date_in_years, 160
 - date_months_difference, 160
 - date_weeks_difference, 160
 - date_years_difference, 160
 - @TODAY 関数, 160
- 時間関数, 139-140
 - time_before, 147, 160
 - time_hours_difference, 160
 - time_in_hours, 160
 - time_in_mins, 160
 - time_in_secs, 160
 - time_mins_difference, 160
 - time_secs_difference, 160
- 正規分布
 - 確率関数, 152
- 比較関数, 147
- 特殊関数, 173
- 確率関数, 152
- 空白文字
 - 文字列から削除する, 121, 154
- 線型回帰
 - PMML としてのエクスポート, 267
- 表示書式
 - 科学的表記, 60
 - 小数桁数, 60
 - 数値, 60
 - 通貨, 60
 - グループ化記号, 60
- 表記方法, 144
- 論理関数, 149
- 報告書
 - 生成, 248
 - 出力の保存, 92
- 差分法
 - 傾向スコアに対する確信度をカバー, 222
 - ニューラル ネットワーク モデル, 222
- 文字列, 136, 138
 - 一致, 121
 - 置換, 121
 - CLEM 式内での操作, 121
- 新機能, 6-7
- 最小化, 23
- 最適化, 263
- 欠損値, 108-109, 122
 - 処理, 106
 - 置換, 106
 - CLEM 式, 110
 - レコード, 108
- 演算子
 - 文字列の結合, 146
- 保存
 - nodes, 90
 - 出力オブジェクト, 92
 - ステート, 91
 - ストリーム, 90
 - 複数のオブジェクト, 91
- 印刷, 26
 - ストリーム, 51
- 商標, 302
- 図表
 - 出力の保存, 92
- 変数, 28
- 実数, 136-137
- 接続
 - IBM SPSS Collaboration and Deployment Services Repository へ, 179-180
 - IBM SPSS Modeler Server へ, 10, 12-13
 - サーバー クラスタ, 13
- 数値, 123, 137
- 整数, 136-137
- 文字, 136, 138
- 日付
 - 変換, 164
 - 操作, 164
- 条件, 119
- 検索
 - ストリーム内でのノード, 72

- 概要, 136
 - IBM SPSS Modeler, 8, 255
- 次数
 - 測定単位, 59
- 注釈
 - nodes, 79, 88
 - コメントへの変換, 88
 - ストリーム, 79, 88
 - フォルダ, 246
 - プロジェクト, 245
- 確率
 - 予測アプリケーション, 222
- 空白, 106, 108, 122
- 言語
 - options, 255
- 警告, 64
 - オプションの設定, 258
- 追加
 - プロジェクトに, 239
- 通知
 - オプションの設定, 258
- 関数, 139-140, 144-145, 165
 - 例, 116
 - @BLANK, 110
 - CLEM 式内, 129
 - @FIELD, 127, 173
 - @GLOBAL_MAX, 171
 - @GLOBAL_MEAN, 171
 - @GLOBAL_MIN, 171
 - @GLOBAL_SDEV, 171
 - @GLOBAL_SUM, 171
 - @PARTITION, 173
 - @PREDICTED, 127, 173
 - @TARGET, 127, 173
 - 欠損値の処理, 110
- 領域, 16
- 例
 - 概要, 4
 - アプリケーション ガイド, 3
- 値
 - CLEM 式への追加, 131
 - データ検査から表示, 131
- 式, 136
- 色
 - 設定, 261
- 日付/時刻の値, 123
- 508 準拠, 282

- abs 関数, 150
- allbutfirst 関数, 154
- allbutlast 関数, 154
- alphabefore 関数, 154
- AND 演算子, 149
- arccos 関数, 151
- arcosh 関数, 151
- arcsin 関数, 151
- arcsinh 関数, 151
- arctan 関数, 151
- arctan2 関数, 151
- arctanh 関数, 151
- attribute, 28
- @BLANK 関数, 110, 145, 172
- cdf_chisq 関数, 152
- cdf_f 関数, 152
- cdf_normal 関数, 152
- cdf_t 関数, 152
- Champion Challenger 分析, 177, 208
- CLEM, 127
 - 概要, 27, 113
 - 関数, 129
 - 例, 116
 - 式, 119, 136
 - CLEM 式の作成, 128
 - language, 136
 - データ型, 137-138
 - 式の検査, 132
- CLEM 関数
 - 三角関数, 151
 - 特殊関数, 173
 - 数値型, 150
 - 欠損値, 110
 - 無作為, 154
 - 変換, 146
 - 情報, 145
 - 比較, 147
 - 論理, 149
 - probability, 152
 - sequence, 165, 167
 - string, 154
 - グローバル, 171
 - 日付と時間, 160
 - 空白とヌル, 172
 - ビット単位, 152
 - 利用可能リスト, 143
- CLEM 式
 - parameters, 66, 120
 - performance, 281
 - テキストの検索および置換, 133
- CLEM 式内の円記号, 138
- CLEM 式の検査, 132
- Clem 式ビルダー, 288
 - 使用, 128
 - 概要, 127
 - アクセス, 128
 - テキストの検索および置換, 133
- Coordinator of Processes, 13
- COP, 13

索引

- cos 関数, 151
- cosh 関数, 151
- count_equal 関数, 124, 147
- count_greater_than 関数, 124, 147
- count_less_than 関数, 124, 147
- count_non_nulls 関数, 147
- count_not_equal 関数, 124, 147
- count_nulls 関数, 110, 124, 147
- count_substring 関数, 154
- CRISP-DM, 20, 236
 - プロジェクト ビュー, 237
- CRISP-DM プロセス モデル, 31-32
- data
 - プレビュー, 56
- date_before 関数, 147
- datetime_date 関数, 146
- DIFF 関数, 167
- @DIFF 関数, 165, 167
- div 関数, 150
- DTD, 231
- encoding, 60, 298
- endstring 関数, 154
- Enterprise View ノード, 208
- f 分布
 - 確率関数, 152
- factor, 295
- @FIELD 関数, 110, 173
- fields, 28, 136, 138
- @FIELDS_BETWEEN 関数, 110, 124, 173
- @FIELDS_MATCHING 関数, 110, 124, 173
- first_index 関数, 126, 147
- first_non_null 関数, 126, 147
- first_non_null_index 関数, 126, 147
- fracof 関数, 150
- hasendstring 関数, 154
- hasmidstring 関数, 154
- hasstartstring 関数, 154
- hassubstring 関数, 154
- HTML 出力
 - スクリーン リーダー, 295
- IBM InfoSphere Warehouse (ISW)
 - PMML のエクスポート, 233
- IBM SPSS Collaboration and Deployment Services, 177
- IBM SPSS Collaboration and Deployment Services Enterprise View, 177, 208
- IBM SPSS Collaboration and Deployment Services Repository, 175, 177
 - 接続先, 179-180
 - 参照, 181
 - 検索, 195
 - オブジェクトおよびバージョンの削除, 199
 - オブジェクトの取得, 191
 - オブジェクトの格納, 182
 - オブジェクトのプロパティ, 203
 - オブジェクトのロックおよびロック解除, 199
 - シングル サインオン, 179
 - フォルダ, 198, 201
 - プロジェクトの転送, 242
- IBM SPSS Collaboration and Deployment Services Repository オブジェクトのロック, 199
- IBM SPSS Collaboration and Deployment Services Repository
 - オブジェクトのロック解除, 199
- IBM SPSS Collaboration and Deployment Services Repository
 - からのオブジェクトの取得, 191
- IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの保存, 182
- IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの場所, 195
- IBM SPSS Modeler, 1, 15
 - 概要, 8, 255
 - options, 255
 - アクセス機能, 282
 - コマンド ラインからの実行, 9
 - ドキュメンテーション, 3
 - はじめに, 8
 - ヒントとショートカット, 101
- IBM SPSS Modeler Advantage, 177, 206
- IBM SPSS Modeler Server
 - password, 10
 - ドメイン名 (Windows) , 10
 - ホスト名, 10, 12
 - ポート番号, 10, 12
 - ユーザー ID, 10
- IBM SPSS Modeler Server 接続の追加, 12-13
- IBM SPSS Modeler Server へのログイン, 10
- IBM SPSS Statistics モデル, 41
- IBM SPSS Text Analytics, 2
- if、then、else 関数, 149
- INDEX 関数, 167
- @INDEX 関数, 165, 167
- integer_bitcount 関数, 152
- integer_leastbit 関数, 152
- integer_length 関数, 152
- intof 関数, 150
- is_date 関数, 145
- is_datetime 関数, 145
- is_integer 関数, 145
- is_number 関数, 145
- is_real 関数, 145
- is_string 関数, 145
- is_time 関数, 145

- is_timestamp 関数, 145
- isalphacode 関数, 154
- isendstring 関数, 154
- islowercode 関数, 154
- ismidstring 関数, 154
- isnumbercode 関数, 154
- isstartstring 関数, 154
- issubstring 関数, 154
- issubstring_count 関数, 154
- issubstring_lim 関数, 154
- isuppercode 関数, 154
- Java, 297
- JAWS, 282, 294, 297
- K-Means ノード
 - performance, 281
 - ラージセット, 58
- keywords
 - ノードの注釈設定, 88
- Kohonen ノード
 - performance, 280
 - ラージセット, 58
- last_index 関数, 126, 147
- LAST_NON_BLANK 関数, 167
- @LAST_NON_BLANK 関数, 165, 167, 172
- last_non_null 関数, 126, 147
- last_non_null_index 関数, 126, 147
- length 関数, 154
- locchar 関数, 154
- locchar_back 関数, 154
- log 関数, 150
- log10 関数, 150
- lowertoupper 関数, 154
- matches 関数, 154
- max 関数, 147
- MAX 関数, 167
- @MAX 関数, 165, 167
- max_index 関数, 126, 147
- max_n 関数, 124, 147
- MEAN 関数, 165, 167
- @MEAN 関数, 165, 167
- mean_n 関数, 124, 150
- member 関数, 147
- min 関数, 147
- MIN 関数, 167
- @MIN 関数, 165, 167
- min_index 関数, 126, 147
- min_n 関数, 124, 147
- mod 関数, 150
- @MULTI_RESPONSE_SET 関数, 126, 173
- negate 関数, 150
- nodes, 8
 - 有効化, 48
 - 無効化, 48, 51
 - 保存, 90
 - 削除, 45
 - 概要, 44
 - 注釈, 79, 88
 - 編集, 51
 - 追加, 45, 49
 - 重複, 51
 - 順序, 275
- IBM SPSS Collaboration and Deployment
 - Services Repository に保存, 189
- performance, 278, 280
 - オプションの設定, 51
 - カスタム サブパレットの作成, 272
 - カスタム パレットの作成, 269
 - コメントを追加, 79
 - ストリーム内の無効化, 48
 - ストリーム内の接続, 46
 - ストリーム内の迂回, 47
 - データ プレビュー, 56
 - データのプレビュー, 56
 - 接続の削除, 50
 - の検索, 72
 - パレット タブのカスタマイズ, 273
 - パレットからの削除, 270
 - パレット上の表示, 270
 - プロジェクトへの追加, 240
 - ロック, 57
 - ロード, 94
- NOT 演算子, 149
- @NULL 関数, 110, 145, 172
- OFFSET 関数, 167
- @OFFSET 関数, 165, 167
 - パフォーマンスに関する考察, 281
- oneof 関数, 154
- options, 255
 - ストリームのプロパティ, 58, 62, 64
- OR 演算子, 149
- output, 18
- parameters
 - type, 69
 - シナリオでの使用, 211
 - スコアリング, 211
 - ストリーム, 66, 68-69, 120
 - セッション, 66, 68-69, 120
 - モデル構築, 211
 - ランタイム プロパティ, 68
- @PARTITION_FIELD 関数, 173
- password
 - IBM SPSS Modeler Server, 10
- performance
 - CLEM 式, 281
 - ノード キャッシング, 54, 276
 - プロセス ノード, 278
 - モデル作成ノード, 280
- pi 関数, 151

索引

- PMML
 エキスポート オプション, 266
 モデルのインポート, 231, 233
 モデルのエキスポート, 230, 233
 PMML モデル
 線型回帰, 267
 ロジスティック回帰, 267
 power (指数) 関数, 150
 PowerPoint ファイル, 238
 @PREDICTED 関数, 173
 Predictive Applications, 208
 Predictive Applications ウィザード, 179, 220, 223
 ストリーム的前提条件, 221
 PredictiveMarketing, 220
 random 関数, 154
 random0 関数, 154
 records, 28
 regression, 295
 rem 関数, 150
 replace 関数, 154
 replicate 関数, 154
 round 関数, 150
 SAS ファイル
 encoding, 298
 SDEV 関数, 167
 @SDEV 関数, 165, 167
 sdev_n 関数, 124, 150
 set コマンド, 66, 120
 sign 関数, 150
 sin 関数, 151
 SINCE 関数, 167
 @SINCE 関数, 165, 167
 sinh 関数, 151
 skipchar 関数, 154
 skipchar_back 関数, 154
 SoftMax 法
 傾向スコアに対する確信度をカバー, 222
 ニューラル ネットワーク モデル, 222
 soundex 関数, 160
 soundex_difference 関数, 160
 SPSS Modeler Server, 1
 SQL 生成, 263
 プレビュー, 265
 ログ記録, 265
 sqrt 関数, 150
 startstring 関数, 154
 Statistics モデル, 41
 Statisticsファイル
 encoding, 298
 stripchar 関数, 154
 strmember 関数, 154
 subscrs 関数, 154
 substring 関数, 154
 substring_between 関数, 154
 SUM 関数, 167
 @SUM 関数, 165, 167
 sum_n 関数, 124, 150
 t 分布
 確率関数, 152
 tan 関数, 151
 tanh 関数, 151
 @TARGET 関数, 173
 testbit 関数, 152
 @TESTING_PARTITION 関数, 173
 THIS 関数, 167
 @THIS 関数, 165, 167
 time_before 関数, 147
 to_date 関数, 146, 160
 to_dateline 関数, 160
 to_datetime 関数, 146
 to_integer 関数, 146
 to_number 関数, 146
 to_real 関数, 146
 to_string 関数, 146
 to_time 関数, 146, 160
 to_timestamp 関数, 146, 160
 @TODAY 関数, 160
 @TRAINING_PARTITION 関数, 173
 trim 関数, 154
 trim_start 関数, 154
 trimend 関数, 154
 undef 関数, 172
 Unicode のサポート, 298
 unicode_char 関数, 154
 unicode_value 関数, 154
 uppertolower 関数, 154
 UTF-8 エンコード, 60, 298
 @VALIDATION_PARTITION 関数, 173
 value_at 関数, 126, 147
 values, 118

 アイコン
 オプションの設定, 62
 アクセシビリティ, 282, 297
 例, 292-293
 IBM SPSS Modeler のヒント, 296
 アクセス機能
 IBM SPSS Modeler の機能, 282
 アプリケーション, 29
 アプリケーションの例, 3

 [インタラクティブ ツリー] ウィンドウ
 アクセシビリティ, 295
 インポート
 PMML, 231, 233

- エクスポート
 - PMML, 230, 233
 - ストリームの説明, 77
- エラー メッセージ, 64
- オフアー モデル, 222
- オブジェクト
 - プロパティ, 247
- 出力オブジェクト
 - IBM SPSS Collaboration and Deployment Services Repository に保存, 190
- オブジェクトのプロパティ、IBM SPSS Collaboration and Deployment Services Repository, 203
- オプション
 - display, 261
 - IBM SPSS Modeler の, 255
 - PMML, 266
 - ユーザー, 257
- 展開オプション
 - シナリオ, 208
- オートメーション, 113
- カイ 2 乗分布
 - 確率関数, 152
- カスタム パレットの作成, 269
 - サブパレットの作成, 272
- 複数カテゴリ設定
 - CLEM 式内, 126
- カンマ, 58
- キャッシュ
 - 有効化, 54, 256, 276
 - 保存, 56
 - キャッシュの設定, 51
 - ノードのオプション, 54, 276
 - 取り消し, 55, 58
- キャッシュ ファイル ノード
 - ロード, 94
- キーボードのショートカット, 284-285, 288, 290
- 開く
 - nodes, 94
 - output, 94
 - ステート, 94
 - ストリーム, 94
 - プロジェクト, 239
 - モデル, 94
- クライアント
 - デフォルト ディレクトリ, 257
- クラス, 20, 236, 238
- グラフ
 - 出力の保存, 92
 - プロジェクトへの追加, 238
- グループ化記号
 - 数字の表示フォーマット, 60
- グローバル関数, 171
- グローバル値
 - CLEM 式内, 130
- ケース, 28
- コピー, 21
- コマンド ライン
 - IBM SPSS Modeler の起動, 9
- コメント
 - キーボードのショートカット, 290
 - ストリームのすべてを一覧表示, 86
 - ノードとストリームの, 79
- サイズ変更, 23
- サブパレット
 - 作成, 272
 - パレット タブからの削除, 271
 - パレット上の表示, 271
- 生成されたモデル パレット, 18
- サーバー
 - サーバーの COP の検索, 13
 - デフォルト ディレクトリ, 257
 - 接続の追加, 12
 - ログイン, 10
- システム
 - options, 255
- システム欠損値, 106
- シナリオ, 206
 - 定義, 177
 - 展開オプション, 208
- シナリオの展開, 208
- ショートカット
 - キーボード, 25, 284-285, 288, 290
 - 一般的な使用法, 101
- シングル サインオン, 11, 180
- シングル サインオン、IBM SPSS Collaboration and Deployment Services Repository, 175, 179
- シーケンス関数, 165, 167
- スクリプト, 27, 113
 - テキストの検索および置換, 133
- スクリーン リーダー, 285, 288, 294-295
 - 例, 292-293

索引

- スクロール
 - オプションの設定, 62
- 傾向スコア
 - 予測アプリケーション, 222
- スコアリング
 - 枝, 79, 212-213, 217
- ステート
 - 保存, 91
 - ロード, 94
- ストリーム, 8, 16
 - 保存, 90
 - 実行, 77
 - 構築, 43
 - 注釈, 79, 88
- IBM SPSS Collaboration and Deployment Services Repository に保存, 188
- options, 58
- 展開オプション, 208
- コメントの追加, 79
- 名前の変更, 74, 88
- ノードの無効化, 48
- ノードの接続, 46
- ノードの迂回, 47
- ノードの追加, 45, 49
- バックアップ ファイル, 90
- プロジェクトへの追加, 238, 240
- ロード, 94
- ストリーム領域
 - 設定, 62
- ストリーム名, 88
- ストリーム パラメータ, 66, 68-69, 120
- ストリーム ファイルのバックアップ
 - 復元, 90
- ストリームの実行, 77
- ストリームの説明, 75, 77
- ストリームのすべてのコメントを一覧表示, 86
- ストリームのデフォルトの文字コード, 60
- ストリームのリライト
 - 有効化, 263
- スペース
 - 文字列から削除する, 121, 154
- スーパーノード
 - parameters, 66, 120
- ズーム, 21

- セッション パラメータ, 66, 68-69, 120
- セット, 58

- ソート ノード
 - performance, 278
- 起動ダイアログ ボックス, 261

- 展開タイプ, 208
- ターミナル ノード, 44

- ツリー ベースの分析
 - 典型的なアプリケーション, 29
- ツールバー, 21
- ツールヒント
 - ノードの注釈設定, 88

- ディクショナリ ファイル, 294
- ディシジョン ツリー
 - アクセシビリティ, 295
- ディシジョン ツリー モデル
 - 傾向スコア, 222
- ディレクトリ
 - default, 257
- 一時ディレクトリ, 14
- テキスト データ ファイル
 - encoding, 298
- テキストの検索, 133
- テキストの置換, 133
- テキストの文字コード, 60
- デフォルト
 - プロジェクト フェーズ, 238
- テンプレート, 96
- テンプレート フィールド, 100
- データ ストリーム
 - 構築, 43
- データ マイニング, 28
 - 戦略, 31
 - アプリケーションの例, 42
- データ マイニングの用途, 30
- データ マッピング ツール, 95-96
- データの型, 118
 - パラメータ, 69
- データのマッピング, 99
- データ分割ノード
 - performance, 279
- データ検査ノード
 - データ マイニングでの使用, 30
 - 探索に使用, 28
- データ型ノード
 - 欠損値, 110
 - performance, 280
- テーブル, 288
 - 出力の保存, 92
 - プロジェクトへの追加, 238

- 日付と時間関数
 - datetime_date, 160
 - datetime_day, 160
 - datetime_day_name, 160
 - datetime_day_short_name, 160

- datetime_hour, 160
- datetime_in_seconds, 160
- datetime_minute, 160
- datetime_month, 160
- datetime_month_name, 160
- datetime_month_short_name, 160
- datetime_now datetime_second, 160
- datetime_time, 160
- datetime_timestamp, 160
- datetime_weekday, 160
- datetime_year, 160
- ドキュメンテーション, 3
- 時間と日付の関数, 139-140
- ドメイン名 (Windows)
 - IBM SPSS Modeler Server, 10
- 特別な文字
 - 文字列から削除する, 121
- 典型的なアプリケーション, 29
- ナゲット, 78
 - 定義, 19
- ナビゲート
 - キーボードのショートカット, 284
- 元に戻す, 21
- 法律に関する注意事項, 300
- ニューラル ネットワーク ノード
 - ラージ セット, 58
- ニューラル ネットワーク モデル
 - 傾向スコア, 222
- ニューラル ノード
 - performance, 280
- ヌル, 106, 122
- 演算子の優先順位, 141
- 文字列の連結, 146
- 名前の変更
 - nodes, 88
 - ストリーム, 74
- 日付の書式, 60, 139-140
- 知識の発見, 28
- 空白の処理
 - CLEM 関数, 172
- 接続の COP の検索, 13
- 複数の IBM SPSS Modeler セッション, 15
- ノイズを含んだデータ, 31
- 紫色のノード, 263
- 数字の表示フォーマット, 60
- 通貨の表示フォーマット, 60
- 時間のフォーマット, 60, 139-140
- ノード
 - プロジェクトへの追加, 238
- 欠損値検査ノード
 - 欠損値, 109
- ノード名, 88
- 入力ノード, 44
 - データのマッピング, 96
 - リフレッシュ, 58
- 出力ノード, 44
- 置換ノード
 - 欠損値, 110
- 評価ノード
 - performance, 280
- ノード キャッシング
 - 有効化, 54, 276
- ノード パレットの選択, 270
- ノードとストリームに名前をつける, 88
- ノードの有効化, 48
- ノードの無効化, 48, 51
- ノードの処理, 44
 - performance, 278
- ノードのロック, 57
- パラメータ
 - CLEM 式内, 130
- パレット, 16
 - カスタマイズ, 267
- パレット タブのカスタマイズ, 273
- バージョン ラベル, IBM SPSS Collaboration and Deployment Services Repository オブジェクト, 205
- ビジュアル プログラミング, 15
- ビット単位関数, 152
- ピリオド, 58
- ヒント
 - アクセス機能, 296
 - 一般的な使用法, 101
- 出力ファイル
 - 保存, 92
- フィールド
 - CLEM 式内, 130
 - 値の表示, 131
- 必須フィールド, 95, 99-100
- 時間フィールド
 - 変換, 164
- フィールドの関連付け, 95
- フィールドの関連付けの解除, 95
- フィールド選択ノード
 - 欠損値, 109

索引

- フォルダ、IBM SPSS Collaboration and Deployment Services Repository, 198, 201
- フォント, 261
- プッシュバック, 263
- プレビュー
 - ノード データ, 56
- プロジェクト, 20, 236
 - 構築, 239
 - 注釈, 245
 - CRISP-DM ビュー, 237
 - IBM SPSS Collaboration and Deployment Services Repository に保存, 189
 - IBM SPSS Collaboration and Deployment Services Repository の, 242
 - オブジェクトの追加, 240
 - オブジェクトのプロパティ, 247
 - クラス ビュー, 238
 - 閉じる, 248
 - デフォルト フォルダの設定, 238
 - 新規に作成, 240
 - フォルダのプロパティ, 246
 - プロパティの設定, 243
 - レポートの生成, 248
- プロパティ
 - データ ストリーム, 58
 - プロジェクト フォルダ, 246
 - レポート フェーズ, 248

- ホスト名
 - IBM SPSS Modeler Server, 10, 12
- ホット キー, 25
- ポート番号
 - IBM SPSS Modeler Server, 10, 12

- マウス
 - IBM SPSS Modeler で使用, 24, 46
- マウスの中央ボタン
 - シミュレート, 24, 46
- マシン学習, 28
- マネージャ, 18

- メイン ウィンドウ, 16
- メッセージ
 - 生成された SQL の表示, 265
- メモリー
 - 管理, 255-256

- モデル, 78
 - 置換, 260
 - IBM SPSS Collaboration and Deployment Services Repository に保存, 190
 - エクスポート, 266
 - プロジェクトへの追加, 238
 - リフレッシュ, 213-214
- モデル作成
 - 枝, 79
 - 値モデル, 222
 - モデル ナゲット, 78
 - [モデル] パレット, 190
 - 枝、モデル作成およびスコアリング, 79, 212-213, 217
 - モデルの置換, 260
 - モデルのリフレッシュ, 208, 213-214
 - モデル作成ノード, 34, 44
 - performance, 280
 - モデル作成パレット タブのカスタマイズ, 273

- ユーザー欠損値, 106
- ユーザー ID
 - IBM SPSS Modeler Server, 10
- ユーザー オプション, 257

- ようこそダイアログ ボックス, 261
- より大演算子, 147
- より小演算子, 147

- ラジアン
 - 測定単位, 59
- ラベル
 - 変数, 231
 - value, 231
 - 表示する, 61
- ラベル、IBM SPSS Collaboration and Deployment Services Repository
 - オブジェクト, 205
- ランタイム プロパティ, 68

- 貼り付け, 21
- リスト, 136, 138
- リフレッシュ
 - 入力ノード, 58
- 切り取り, 21

- ルール セット
 - 評価, 58
- ルール ノードの構築
 - ロード, 94
- ルールセット モデル
 - 傾向スコア, 222

- レコード
 - 欠損値, 108

- 重複レコード ノード
 - performance, 280
- レコード結合ノード
 - performance, 279
- レコード集計ノード
 - performance, 279
- 複数レスポンス セット
 - CLEM 式内, 126, 130
- レポート
 - プロジェクトへの追加, 238
 - プロパティの設定, 248

- ログ ファイル
 - 生成された SQL の表示, 265
- ロケール
 - options, 255
- ロジスティック回帰, 295
 - PMML としてのエクスポート, 267
- ロジスティック回帰モデル
 - 傾向スコア, 222
- ロード
 - nodes, 94
 - ステート, 94

- 実行を中止, 21
- 日をロールオーバー, 60