

IBM SPSS Modeler 15 入力ノード、プロセスノード、出力ノード



注：この情報とサポートされている製品をご使用になる前に、「注意事項」(p.) の一般情報をお読みください。

本版は IBM SPSS Modeler 15 , および新版で指示されるまで後続するすべてのリリースおよび変更に対して適用されます。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。

Licensed Materials - Property of IBM

© Copyright IBM Corporation 1994, 2012.

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

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 製品 | 1 |
| IBM SPSS Modeler | 1 |
| IBM SPSS Modeler Server | 2 |
| IBM SPSS Modeler Administration Console | 2 |
| IBM SPSS Modeler Batch | 3 |
| IBM SPSS Modeler Solution Publisher | 3 |
| IBM SPSS Modeler Server の IBM SPSS Collaboration and Deployment Services | 3 |
| IBM SPSS Modeler エディション | 3 |
| IBM SPSS Modeler ドキュメント | 5 |
| SPSS Modeler Professional ドキュメント | 5 |
| SPSS Modeler Premium ドキュメント | 6 |
| アプリケーションの例 | 7 |
| Demos フォルダ | 7 |

2 入力ノード 9

| | |
|--|----|
| 概要 | 9 |
| Enterprise View ノード | 10 |
| Enterprise View ノードのオプション設定 | 12 |
| Enterprise View の接続 | 13 |
| DPD の選択 | 15 |
| テーブルの選択 | 16 |
| データベース入力ノード | 16 |
| データベース ノード オプションの設定 | 19 |
| データベース接続の追加 | 20 |
| データベース接続のプリセット値の指定 | 22 |
| データベース テーブルの選択 | 25 |
| データベースの照会 | 26 |
| 可変長ファイル ノード | 28 |
| 可変長ファイル ノードのオプションの設定 | 30 |
| 固定長ファイル ノード | 32 |
| 固定長ファイル ノードのオプションの設定 | 33 |
| フィールドのストレージと形式の設定 | 35 |
| Data Collection ノード | 39 |
| Data Collection インポート ファイルのオプション | 40 |

| | |
|--|----|
| IBM SPSS Data Collection インポート メタデータのプロパティ | 44 |
| データベース接続文字列 | 45 |
| 詳細プロパティ | 46 |
| 複数プロパティ設定のインポート | 46 |
| IBM SPSS Data Collection 列インポート ノード | 48 |
| IBM Cognos BI 入力ノード | 49 |
| Cognos オブジェクトのアイコン | 49 |
| Cognos データのインポート | 50 |
| Cognos レポートのインポート | 52 |
| Cognos の接続 | 54 |
| Cognos の場所の選択 | 55 |
| データまたはレポートのパラメータの指定 | 56 |
| SAS 入力ノード | 56 |
| SAS 入力ノードのオプションの設定 | 57 |
| Excel 入力ノード | 58 |
| XML 入力ノード | 60 |
| 複数のルート要素からの選択 | 63 |
| XML ソース データの不要なスペースの削除 | 63 |
| ユーザー入力ノード | 65 |
| ユーザー入力ノードのオプションの設定 | 66 |
| 共通の入力ノード タブ | 72 |
| 入力ノードの尺度の設定 | 72 |
| 入力ノードからのフィールドのフィルタリング | 74 |

3 レコード設定ノード 76

| | |
|-----------------------|----|
| レコード設定の概要 | 76 |
| 条件抽出ノード | 77 |
| サンプル ノード | 79 |
| サンプル ノードのオプション | 80 |
| クラスタと階層の設定 | 84 |
| 階層のサンプル サイズ | 87 |
| バランス ノード | 88 |
| バランス ノードのオプション設定 | 89 |
| レコード集計ノード | 90 |
| レコード集計ノードのオプション設定 | 91 |
| RFM レコード集計ノード | 94 |
| RFM レコード集計ノードのオプション設定 | 95 |

| | |
|---------------------------|-----|
| ソートノード | 96 |
| ソートの最適化設定 | 97 |
| レコード結合ノード | 98 |
| 結合の種類 | 99 |
| 結合方法とキーの指定 | 102 |
| 部分結合のデータの選択 | 104 |
| 結合の条件の指定 | 104 |
| レコード結合ノードからのフィールドのフィルタリング | 105 |
| 入力順序とタグの設定 | 107 |
| レコード結合の最適化設定 | 109 |
| レコード追加ノード | 111 |
| 追加オプションの設定 | 112 |
| 重複レコードノード | 113 |
| 重複レコード最適化設定 | 116 |

4 フィールド設定ノード 118

| | |
|----------------------|-----|
| フィールド設定の概要 | 118 |
| 自動データ準備 | 120 |
| [フィールド] タブ | 123 |
| [設定] タブ | 124 |
| フィールド設定 | 124 |
| 日付および時刻の準備 | 126 |
| フィールドの除外 | 127 |
| 入力フィールドおよび目標フィールドの準備 | 128 |
| 構築およびフィールド選択 | 131 |
| フィールド名 | 133 |
| [分析] タブ | 134 |
| フィールド処理の要約 | 136 |
| フィールド | 138 |
| アクションの概要 | 140 |
| 予測精度 | 141 |
| [フィールド] テーブル | 142 |
| フィールド詳細 | 143 |
| アクションの詳細 | 145 |
| フィールド生成ノードの生成 | 148 |
| データ型ノード | 150 |
| 尺度 | 153 |
| 連続型データの変換 | 156 |

| | |
|------------------------------|-----|
| インスタンス化とは？ | 157 |
| データ値 | 158 |
| 欠損値の定義 | 165 |
| データ型の値の検査 | 165 |
| フィールドの役割の設定 | 167 |
| データ型属性のコピー | 169 |
| フィールド形式の [設定] タブ | 170 |
| フィールドのフィルタリングまたは名前の変更 | 173 |
| フィルタリング オプションの設定 | 174 |
| アンサンブル ノード | 180 |
| アンサンブル ノードの設定 | 181 |
| フィールド作成ノード | 185 |
| フィールド作成ノードの基本オプションの設定 | 186 |
| 複数フィールドの作成 | 188 |
| CLEM 式作成オプションの設定 | 191 |
| フィールド作成ノード (フラグ型) のオプションの設定 | 191 |
| フィールド作成ノード (セット型) のオプションの設定 | 193 |
| フィールド作成ノード (ステート型) のオプションの設定 | 195 |
| フィールド作成ノード (カウント型) のオプションの設定 | 196 |
| フィールド作成ノード (条件式型) のオプションの設定 | 197 |
| フィールド作成ノードを使用して値を再コード化する | 198 |
| 置換ノード | 199 |
| 置換ノードを使ったストレージの変換 | 200 |
| 匿名化ノード | 202 |
| 匿名化ノードのオプションの設定 | 203 |
| フィールド値の匿名化 | 205 |
| データ分類ノード | 207 |
| データ分類ノードのオプション設定 | 208 |
| 複数フィールドのデータ分類 | 210 |
| 再分類されたフィールドのストレージと尺度 | 212 |
| データ分割ノード | 212 |
| データ分割ノードのオプション設定 | 213 |
| 固定幅のデータ分割 | 215 |
| 分位 (等カウントまたは合計) | 216 |
| ケースのランク付け | 219 |
| 平均/標準偏差 | 220 |
| 最適カテゴリ化 | 221 |
| 生成されたビンのプレビュー | 223 |
| RFM 分析ノード | 225 |
| RFM 分析ノードの設定 | 226 |
| RFM 分析ノードの分割 | 228 |

| | |
|---------------------|-----|
| データ区分ノード | 229 |
| データ区分ノードのオプション | 230 |
| フラグ設定ノード | 232 |
| フラグ設定ノードのオプションの設定 | 233 |
| 再構成ノード | 233 |
| 再構成ノードのオプション設定 | 235 |
| 行列入替ノード | 236 |
| 行列入替ノードのオプション設定 | 237 |
| 時間区分ノード | 241 |
| 時間区分を指定 | 242 |
| 時間区分構築のオプション | 244 |
| 推定期間 | 247 |
| 予測(S) | 248 |
| サポートされる区分 | 250 |
| 時系列ノード | 263 |
| 時系列ノードのオプションの設定 | 264 |
| フィールド順序ノード | 265 |
| フィールド順序ノードのオプションの設定 | 265 |

5 グラフ作成ノード

269

| | |
|-----------------------------|-----|
| グラフ作成ノードの共通の機能 | 269 |
| 外観、オーバーレイ、パネル、およびアニメーション | 271 |
| [出力] タブの使用方法 | 275 |
| [注釈] タブの使用方法 | 276 |
| 3次元グラフ | 277 |
| グラフボード ノード | 278 |
| グラフボード [基本] タブ | 279 |
| グラフボード [詳細] タブ | 284 |
| 組み込まれている利用可能なグラフボード視覚化タイプ | 287 |
| Creating Map Visualizations | 294 |
| Graphboard Examples | 295 |
| グラフボードの [外観] タブ | 316 |
| テンプレート、スタイルシート、マップの位置の設定 | 318 |
| テンプレート、スタイルシート、マップ ファイルの管理 | 320 |
| マップ シェープファイルの変換と配布 | 322 |
| マップの主要コンセプト | 323 |
| マップ変換ユーティリティの使用 | 324 |
| マップ ファイルの配布 | 331 |

| | |
|---------------------|-----|
| 散布図ノード | 332 |
| 散布図ノードのタブ | 334 |
| 散布図の [オプション] タブ | 337 |
| 散布図の [外観] タブ | 339 |
| 散布図グラフの使用方法 | 340 |
| 棒グラフノード | 341 |
| 棒グラフの [プロット] タブ | 341 |
| 棒グラフの [外観] タブ | 342 |
| 棒グラフノードの使用方法 | 343 |
| ヒストグラムノード | 346 |
| ヒストグラムの [プロット] タブ | 347 |
| ヒストグラムの [オプション] タブ | 348 |
| ヒストグラムの [外観] タブ | 349 |
| ヒストグラムの使用方法 | 350 |
| 集計棒グラフノード | 351 |
| 集計棒グラフの [プロット] タブ | 351 |
| 集計棒グラフの [オプション] タブ | 352 |
| 集計棒グラフの [外観] タブ | 353 |
| 集計棒グラフの使用方法 | 355 |
| 線グラフノード | 356 |
| 線グラフの [プロット] タブ | 357 |
| 線グラフの [外観] タブ | 359 |
| 線グラフの使用方法 | 360 |
| Web グラフノード | 361 |
| Web グラフの [プロット] タブ | 363 |
| Web グラフの [オプション] タブ | 365 |
| Web グラフの [外観] タブ | 367 |
| Web グラフの使用方法 | 368 |
| 時系列ノード | 373 |
| 時系列の [プロット] タブ | 375 |
| 時系列の [外観] タブ | 377 |
| 時系列グラフの使用方法 | 378 |
| 評価ノード | 378 |
| 評価の [プロット] タブ | 383 |
| 評価の [オプション] タブ | 385 |
| 評価の [外観] タブ | 387 |
| モデル評価の結果の読み込み | 388 |
| 評価グラフの使用方法 | 390 |
| グラフの検証 | 391 |
| バンドの使用 | 392 |
| 領域の使用 | 397 |

| | |
|------------------------|-----|
| マークされた要素の使用 | 401 |
| グラフからのノードの生成 | 402 |
| 視覚化の編集 | 405 |
| 視覚化編集の一般的な規則 | 407 |
| テキストの編集と書式設定 | 408 |
| 色、パターン、破線化、および透明度の変更 | 409 |
| ポイント要素の回転と、形状および縦横比の変更 | 411 |
| グラフィック要素のサイズの変更 | 411 |
| 余白とパディングの指定 | 412 |
| 数値の書式設定 | 413 |
| 軸とスケール設定の変更 | 414 |
| カテゴリの編集 | 416 |
| 方向パネルの変更 | 418 |
| 座標システムの変換 | 419 |
| 統計量とグラフ要素の変更 | 420 |
| 凡例の位置の変更 | 424 |
| 視覚化および視覚化データのコピー | 424 |
| キーボード ショートカット | 425 |
| 表題と脚注の追加 | 425 |
| グラフのスタイルシートの使用 | 427 |
| スタイルシートの適用 | 429 |
| グラフの印刷、保存、エクスポート | 431 |

6 出力ノード

434

| | |
|-------------------|-----|
| 出力ノードの概要 | 434 |
| 出力の管理 | 435 |
| 出力を表示 | 436 |
| [Web に公開] | 437 |
| HTML ブラウザで出力結果を表示 | 440 |
| 出力のエクスポート | 441 |
| セルと列の選択 | 442 |
| テーブル ノード | 443 |
| テーブル ノードの [設定] タブ | 443 |
| テーブル ノードの [形式] タブ | 443 |
| 出力ノードの [出力] タブ | 444 |
| テーブル ブラウザ | 446 |
| クロス集計ノード | 449 |
| クロス集計ノードの [設定] タブ | 449 |

| | |
|-----------------------------------|-----|
| クロス集計ノードの [外観] タブ | 451 |
| クロス集計ノード出力ブラウザ | 452 |
| 精度分析ノード | 454 |
| 精度分析ノードの [精度分析] タブ | 454 |
| 精度分析出力ブラウザ | 457 |
| データ検査ノード | 459 |
| データ検査ノードの [設定] タブ | 461 |
| データ検査の [欠損値検査] タブ | 463 |
| データ検査出力ブラウザ | 465 |
| 変換ノード | 476 |
| 変換ノードの [オプション] タブ | 477 |
| 変換ノードの [出力] タブ | 478 |
| 変換ノードの出力ビューア | 478 |
| 記述統計ノード | 482 |
| 記述統計ノードの [設定] タブ | 483 |
| 記述統計量出力ブラウザ | 485 |
| 平均比較ノード | 487 |
| 独立したグループの平均値を比較 | 488 |
| 一対のフィールド間の平均値の比較 | 489 |
| 平均比較ノードのオプション | 489 |
| 平均値ノード出力ブラウザ | 491 |
| レポートノード | 493 |
| レポートノードの [テンプレート] タブ | 494 |
| レポートノード出力ブラウザ | 496 |
| グローバルノード | 497 |
| グローバルノードの [設定] タブ | 498 |
| IBM SPSS Statistics ヘルパー アプリケーション | 499 |

7 エクスポート ノード

501

| | |
|----------------------------|-----|
| エクスポートノードの概要 | 501 |
| データベース エクスポート ノード | 502 |
| データベース ノードの [エクスポート] タブ | 503 |
| データベース エクスポート結合オプション | 504 |
| データベース エクスポートのスキーマ オプション | 507 |
| データベース エクスポートのインデックス オプション | 511 |
| データベース エクスポートの詳細オプション | 515 |
| バルク ローダーのプログラミング | 517 |

| | |
|-------------------------------------|-----|
| ファイル エクスポート ノード | 526 |
| ファイル ノードの [エクスポート] タブ | 527 |
| IBM SPSS Data Collection エクスポート ノード | 528 |
| IBM Cognos BI エクスポート ノード | 530 |
| Cognos の接続 | 531 |
| ODBC 接続 | 532 |
| SAS エクスポート ノード | 534 |
| SAS エクスポート ノード、[エクスポート] タブ | 535 |
| Excel エクスポート ノード | 536 |
| Excel ノードの [エクスポート] タブ | 536 |
| XML エクスポート ノード | 537 |
| XML データの作成 | 539 |
| XML マッピングのレコード オプション | 539 |
| XML マッピングのフィールド オプション | 540 |
| XML マッピングのプレビュー | 541 |

8 IBM SPSS Statistics ノード 543

| | |
|--|-----|
| IBM SPSS Statistics ノードの概要 | 543 |
| Statistics ファイル ノード | 544 |
| Statistics 変換ノード | 547 |
| Statistics 変換ノードの [シンタックス] タブ | 547 |
| 利用可能なシンタックス | 550 |
| Statistics モデル ノード | 552 |
| Statistics モデル ノードの [モデル] タブ | 552 |
| Statistics モデル ノード - モデル ナゲットの要約 | 554 |
| Statistics 出力ノード | 555 |
| Statistics 出力ノードの [シンタックス] タブ | 556 |
| Statistics 出力ノードの [出力] タブ | 559 |
| Statistics エクスポート ノード | 561 |
| Statistics エクスポート ノードの [エクスポート] タブ | 562 |
| IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング | 563 |

9 スーパーノード 565

| | |
|--------------------|-----|
| スーパーノードの概要 | 565 |
| スーパーノードの種類 | 565 |
| 入カスーパーノード | 566 |
| プロセススーパーノード | 566 |
| ターミナルスーパーノード | 567 |
| スーパーノードの作成 | 568 |
| スーパーノードのネスト | 571 |
| 有効なスーパーノードの例 | 571 |
| 無効なスーパーノードの例 | 573 |
| スーパーノードのロック | 574 |
| スーパーノードのロックとロック解除 | 574 |
| ロックされたスーパーノードの編集 | 576 |
| スーパーノードの編集 | 577 |
| スーパーノードの種類の変更 | 578 |
| スーパーノードの注釈付けと名前の変更 | 578 |
| スーパーノードのパラメータ | 579 |
| スーパーノードとキャッシュ | 585 |
| スーパーノードとスクリプト | 587 |
| スーパーノードの保存とロード | 588 |

付録

A 注意事項 590

索引 593

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 と組み合わせてクライアントとし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 製品

製品と関連するソフトウェアの IBM® SPSS® Modeler ファミリの構成は次のとおりです。

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

IBM 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、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 SPSS Modeler を実行するか、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 を使用して、プログラミングの必要なく、正確な予測モデルを迅速かつ直感的に構築することができます。独自のビジュアル インターフェイスを使用すると、データ マイニング プロセスを簡単に視覚化することができます。製品に組み込まれている高度な分析の支援を受けて、データ内に隠れたパターンやトレンドを発見することができます。結果をモデル化し、ビジネスチャンスを活用してリスクを軽減できるようになり、それらに影響を与える要因を理解することができます。

SPSS Modeler は SPSS Modeler Professional および SPSS Modeler Premium の 2 つのエディションで使用できます。詳細は、[IBM SPSS Modeler エディション in IBM SPSS Modeler 15 ユーザー ガイド](#) 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。

SPSS Modeler Server は、1 つまたは複数の IBM® SPSS® Modeler のインストールと組み合わせてサーバー ホストで分散分析モードで継続的に実行する、別途ライセンスが必要な製品です。このように、メモリー集中型の操作は、クライアントコンピュータにデータをダウンロードせずにサーバー上で実行することができるため、SPSS Modeler Server は大きなデータセットに対し優れたパフォーマンスを示すことができます。IBM® SPSS® Modeler Server は、パフォーマンスと自動化のさらなる利点を提供し、SQLの最適化とデータベース内のモデリング機能をサポートしています。

IBM SPSS Modeler Administration Console

Modeler Administration Console は多くの SPSS Modeler Server 設定オプションを管理し、オプション ファイルによって設定可能なグラフィカルアプリケーションです。アプリケーションには、SPSS Modeler Server のインストールを監視、構成するコンソール ユーザー インターフェイスが用意されており、しかも、現在の SPSS Modeler Server のお客様には無料で提供されます。アプリケーションは Windows コンピュータにのみインストールできますが、サポートされる任意のプラットフォームにインストールされたサーバーを管理できます。

IBM SPSS Modeler Batch

データマイニングは、通常、対話型のプロセスですが、グラフィカル ユーザー インターフェースを必要とせず、コマンドラインから SPSS Modeler を実行することも可能です。たとえば、ユーザーの介入なしで実行する長期実行または反復的なタスクがあります。SPSS Modeler Batch は、通常のユーザーインターフェイスにアクセスせずに SPSS Modeler の完全な分析機能のサポートを提供する製品の特別バージョンです。SPSS Modeler Batch を使用するには、SPSS Modeler Server ライセンスが必要です。

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher は、外部ランタイムで実行することができ、外部アプリケーションに埋め込まれる SPSS Modeler ストリームのパッケージ版を作成することができるツールです。このように、SPSS Modeler がインストールされていない環境で使用するための完全な SPSS Modeler ストリームを公開して展開することができます。SPSS Modeler Solution Publisher は、個別のライセンスが必要とされている IBM SPSS Collaboration and Deployment Services - Scoring サービスの一部として配布されています。このライセンスを使用すると、SPSS Modeler Solution Publisher Runtime を受信し、公開されたストリームを実行することができます。

IBM SPSS Modeler Server の IBM SPSS Collaboration and Deployment Services

さまざまな IBM® SPSS® Collaboration and Deployment Services アダプタを使用すると、SPSS Modeler および SPSS Modeler Server が IBM SPSS Collaboration and Deployment Services リポジトリとインタラクティブに機能させることができます。このように、リポジトリにデプロイされた SPSS Modeler ストリームは、複数のユーザーで共有したり、またはシンククライアントアプリケーション IBM SPSS Modeler Advantage からアクセスできます。リポジトリをホストするシステム上のアダプタをインストールします。

IBM SPSS Modeler エディション

SPSS Modeler は次のエディションで使用できます。

SPSS Modeler Professional

SPSS Modeler Professional は、CRM システムで追跡する行動や対話、人口統計データ、購入行動や販売データなど、多くの構造化データを処理するために必要なすべてのツールを提供しています。

SPSS Modeler Premium

SPSS Modeler Premium は、エンティティの分析やソーシャル ネットワーキングなどの特化したデータ、又は構造化されていないテキスト データを処理するために SPSS Modeler Professional を拡張する、別途ライセンスが必要な製品です。SPSS Modeler Premium は次のコンポーネントで構成されています。

IBM® SPSS® Modeler Entity Analytics が新しい次元を IBM® SPSS® Modeler の予測分析に追加します。予測分析は過去のデータから将来の行動を予測しようとするのに対し、エンティティ分析ではレコードの中でアイデンティティの競合を解決することで現在のデータの干渉性と一貫性を改善することに焦点を当てます。アイデンティティは、個人、組織、オブジェクトまたは曖昧さの存在する他のエンティティとなります。アイデンティティの解決は、顧客関係の管理、不正行為の検出、マネーロンダリング防止、国内および国際的なセキュリティなどのさまざまなフィールドにおいて重要になります。

IBM SPSS Modeler Social Network Analysis は、関係に関する情報を、個人およびグループの社会的行動を特徴づけるフィールドに変換します。ソーシャル ネットワークの基底となる関係を説明するデータを使用して、IBM® SPSS® Modeler Social Network Analysis はネットワークの他の人の行動に影響を与えるソーシャル リーダーを識別します。また、他のネットワーク参加者に最も影響を受ける人を確認できます。これらの結果を他の指標と組み合わせることによって、予測モデルの基準となる個人の包括的なプロフィールを作成できます。この社会的情報を含むモデルは、含まないモデルに比べてパフォーマンスが高くなります。

Text Analytics for IBM® SPSS® Modeler は、高度な言語技術と Natural Language Processing (NLP) を使用して、多様な未構築のテキスト データを急速に処理し、重要なコンセプトを抽出および組織化、そしてそのコンセプトをカテゴリ別に分類します。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、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。

IBM SPSS Modeler ドキュメント

オンライン ヘルプ形式のドキュメントは、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（インストール手順を含む）は、各製品 DVD の ¥Documentation フォルダにもあります。インストール マニュアルは、Web サイト <http://www-01.ibm.com/support/docview.wss?uid=swg27023172> からダウンロードできます。

これらの形式のドキュメントは、SPSS Modeler インフォメーション センター <http://publib.boulder.ibm.com/infocenter/spssmodl/v15r0m0/> から入手できます。

SPSS Modeler Professional ドキュメント

SPSS Modeler Professional のドキュメント スイート（インストール手順を除く）は次のとおりです。

- **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 15 モデル作成ノード](#) 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。このガイドは、PDF 形式のみです。
- **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] メニューから利用できます。詳細は、[アプリケーションの例 in IBM SPSS Modeler 15 ユーザー ガイド](#) 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、サードパーティー製のアルゴリズム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 管理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 手法を使用して SPSS Modeler によるデータ マイニングを行う段階的なガイドです。
- **IBM SPSS Modeler Batch ユーザー ガイド:** バッチ モードの実行およびコマンドラインの引数の詳細を含む、IBM SPSS Modeler をバッチ モードで使用するための完全ガイド。このガイドは、PDF 形式のみです。

SPSS Modeler Premium ドキュメント

SPSS Modeler Premium のドキュメント スイート（インストール手順を除く）は次のとおりです。

- **IBM SPSS Modeler Entity Analytics ユーザー ガイド:** リポジトリのインストールと設定、エンティティ分析ノード、管理タスクについて説明した、SPSS Modeler でのエンティティ分析の使用に関する情報。
- **IBM SPSS Modeler Social Network Analysis ユーザー ガイド:** グループ分析および拡散分析を含む SPSS Modeler によるソーシャル ネットワーク分析を実行するためのガイド。

- **Text Analytics for SPSS Modeler ユーザー ガイド:** テキスト マイニング ノード、インタラクティブ ワークベンチ、テンプレート、その他のリソースについて説明した、SPSS Modeler でのテキスト分析の使用に関する情報。
- **Text Analytics for IBM SPSS Modeler 管理コンソール ユーザー ガイド:** Text Analytics for SPSS Modeler と使用するために IBM® 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。

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 作業者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、[Demos フォルダ in IBM SPSS Modeler 15 ユーザー ガイド](#) 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。

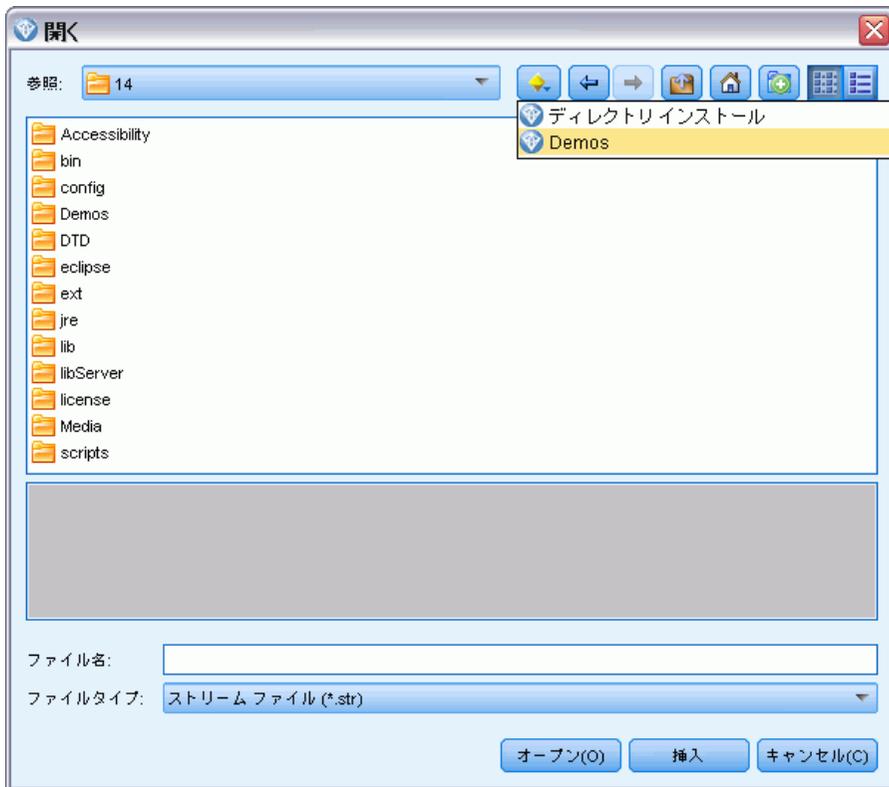
スクリプトの例: 例は、『IBM SPSS Modeler スクリプトとオートメーション ガイド』を参照してください。

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 15 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



入力ノード

概要

入力ノードを使用すると、フラット ファイル、IBM® SPSS® Statistics (.sav)、SAS、Microsoft Excel、および ODBC 準拠のリレーショナル データベースも含めたさまざまなフォーマットで格納されたデータを、インポートできるようになります。ユーザー入力ノードを使用して、合成データを生成することもできます。

[入力] パレットには、次のノードがあります。



Enterprise View ノードは、IBM® SPSS® Collaboration and Deployment Services Repository への接続を作成し、Enterprise View のデータをストリームに読み込み、他のユーザーがレポジトリからアクセスできるシナリオにモデルをパッケージ化できます。詳細は、[p.10 Enterprise View ノード](#) を参照してください。



データベース ノードは、Microsoft SQL Server、DB2、Oracle など ODBC（開放型データベース接続）を使用するさまざまなパッケージからデータをインポートするのに使用できます。詳細は、[p.16 データベース入力ノード](#) を参照してください。



可変長ノードで、可変長フィールド テキスト ファイル、つまりフィールド数は一定でも各フィールド内の文字数が異なるレコードを含むファイルから、データを読み込みます。このノードは、固定長のヘッダー テキストやある種の注釈があるファイルにも使用できます。詳細は、[p.28 可変長ファイル ノード](#) を参照してください。



固定長ノードで、固定長フィールド テキスト ファイルからデータをインポートします。ここで、ファイルのフィールドは区切られていませんが、同じ位置から始まって長さは固定されています。コンピュータ生成のデータや、旧来のシステムのデータなどは、しばしば固定長フィールド形式で保存されています。詳細は、[p.32 固定長ファイル ノード](#) を参照してください。



Statistics ファイル ノードは、同じフォーマットを使用する SPSS Statistics で使用される .sav ファイル形式のデータおよび IBM® SPSS® Modeler に保存されたキャッシュ ファイルを読み込みます。詳細は、[8 章 p.544 Statistics ファイル ノード](#) を参照してください。



IBM® SPSS® Data Collection ノードで、Data Collection Data Model に基づく市場調査ソフトウェアによって使用されるさまざまな形式の調査データをインポートします。このノードを使用するには、Data Collection Developer Library がインストールされている必要があります。詳細は、[p. 39 Data Collection ノード](#) を参照してください。



IBM Cognos BI 入力ノードは、Cognos BI データベースからデータをインポートします。詳細は、[p. 50 Cognos データのインポート](#) を参照してください。



SAS ファイル ノードで、SAS データを SPSS Modeler へインポートします。詳細は、[p. 56 SAS 入力ノード](#) を参照してください。



Excel ノードで、Microsoft Excel の各バージョンからデータをインポートします。ODBC データ ソースは不要です。詳細は、[p. 58 Excel 入力ノード](#) を参照してください。



XML 入力ノードを使用して、XML 形式のデータをストリームにインポートできます。ディレクトリの 1 つのファイルまたはすべてのファイルをインポートできます。オプションで、XML 構造を読み込むスキーマ ファイルを指定できます。詳細は、[p. 60 XML 入力ノード](#) を参照してください。



ユーザー入力ノードで、最初から、または既存のデータを変更して、合成データを作成する簡単な方法が提供されます。これは、モデル作成用の検定データセットを作成する場合などに役立ちます。詳細は、[p. 65 ユーザー入力ノード](#) を参照してください。

ストリームを開始するには、入力ノードをストリーム領域に追加します。次に、配置したノードをダブル クリックして、ダイアログ ボックスを表示します。このダイアログ ボックス内のさまざまなタブで、データの読み込み、フィールドと値の表示ができ、フィルタ、データ型、フィールドの役割、欠損値の検査などを含む多様なオプションを設定することができます。

Enterprise View ノード

Enterprise View ノードを使用すると、共有の IBM® SPSS® Collaboration and Deployment Services Repository で IBM® SPSS® Modeler セッションおよび Enterprise View 間の接続を作成および維持管理することができます。これにより、Enterprise View から SPSS Modeler ストリームヘデータを読み込むことができ、また SPSS Modeler モデルを、共有リポジトリのほかのユーザーが評価できるシナリオをにパッケージ化することができます。

シナリオとは、スコアリングのためあるいは自動モデル リフレッシュで使用するために IBM SPSS Collaboration and Deployment Services Repository に展開できる特定のノード、モデル、および追加属性を伴う SPSS Modeler ストリームを含んでいるファイルのことです。シナリオによって Enterprise View を使用すると、マルチ ユーザーの環境で、すべてのユーザーが同じデータを処理することができます。接続 は、IBM SPSS Collaboration and Deployment Services Repository での SPSS Modeler セッションから Enterprise View へのリンクです。

Enterprise Viewは、データが置かれている物理的な場所に関わらず、組織に属するデータの完全なセットです。それぞれの接続は、1 つの **アプリケーション ビュー** (Enterprise View のサブセット)、**データ プロバイダの定義** (論理アプリケーション ビュー テーブル/列を物理データ ソースにリンク)、および **環境** (どの特定の列を定義済みのビジネス セグメントに関連付けるかを判断) の特定の選択から成ります。実際のデータは 1 つまたは複数のデータベースまたはその他の外部ソースの中にありますが、エンタープライズ ビュー、アプリケーション ビューおよび DPD 定義はリポジトリに格納され、バージョンングされています。

接続が確立されると、アプリケーション ビューの **テーブル** を指定して SPSS Modeler で処理します。アプリケーション ビューの場合、テーブルは 1 つまたは複数の物理データベースにある 1 つまたは複数の物理テーブルの一部またはすべての列で構成される論理ビューです。そのため、Enterprise View ノードを使用すると、複数のデータベース テーブルのレコードを、SPSS Modeler の単一テーブルとして表示できます。

要件

- Enterprise View ノードを使用するには、まず IBM SPSS Collaboration and Deployment Services Repository を、すでに Enterprise View、アプリケーション ビューおよび 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。

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。

- 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 15 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お問い合わせください。

- 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。

Enterprise View ノードのオプション設定

[Enterprise View] ダイアログ ボックス の [データ] タブでは以下のオプションを使用できます。

- 既存のリポジトリ接続を選択する
- 既存のリポジトリ接続を編集する
- 新規リポジトリ接続を作成する
- アプリケーション ビュー テーブルを選択する

リポジトリでの作業の詳細は、『IBM® SPSS® Collaboration and Deployment Services Administrator' s Guide』を参照してください。

図 2-1

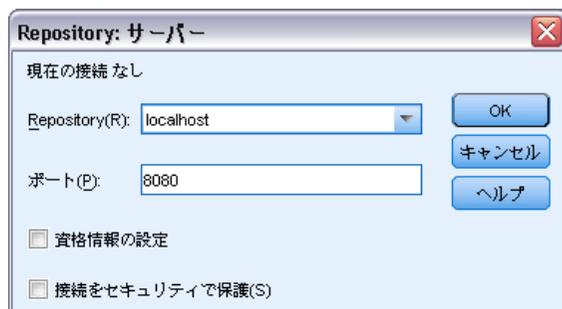
IBM SPSS Collaboration and Deployment Services Repository への接続の追加



接続: ドロップダウン リストには、既存のレポジトリ接続の選択、既存の接続の編集、あるいは接続の追加のオプションが用意されています。IBM® SPSS® Modeler を介してリポジトリにすでにログインしている場合、[接続の

[追加/編集] オプションを選択して、[Enterprise View の接続] ダイアログボックスを表示します。現在の接続の必要な詳細を定義または編集することができます。ログインしていない場合は、このオプションを選択すると、リポジトリの [ログイン] ダイアログボックスを表示します。

図 2-2
リポジトリへのログイン



リポジトリへのログインに関する詳細は、『SPSS Modeler ユーザー ガイド』を参照してください。

リポジトリへの接続が確立されると、SPSS Modeler を終了するまで接続は維持されます。同じストリーム内の他のノードと接続を共有できますが、それぞれの新規ストリームに接続を新規作成する必要があります。

ログインに成功すると、[Enterprise View の接続] ダイアログボックスが表示されます。

テーブル名: このフィールドは最初は空であり、接続を作成するまでは何も読み込めません。アクセスするアプリケーション ビュー テーブル名が既知の場合に、[テーブル名] フィールドにその名前を入力します。そうでない場合は、[選択] ボタンをクリックして、利用可能なアプリケーション ビュー テーブルを表示したダイアログボックスを開きます。

Enterprise View の接続

このダイアログボックスでは、リポジトリ接続に必要な詳細を定義または編集できます。以下の事項を指定します。

- アプリケーション ビューおよびバージョン
- 環境
- データ プロバイダの定義
- 接続の詳細

図 2-3
アプリケーション ビューの選択



接続： 既存のレポジトリ接続を一覧表示します。

- **新規接続の追加：** [オブジェクトの取得] ダイアログ ボックスを表示し、そこから検索し、レポジトリからアプリケーション ビューを選択することができます。
- **選択したテキストのコピー：** 選択した接続のコピーを作成すると、同じアプリケーション ビューを再度表示する必要はありません。
- **選択された接続の削除：** 選択した接続をリストから削除します。

接続の詳細設定： 接続ウィンドウで現在選択している接続について、アプリケーション ビュー、バージョン テーブル、環境、データ プロバイダ定義、および説明文を表示します。

- **アプリケーション ビュー：** ドロップダウンリストに、選択したアプリケーション ビューがあれば表示されます。現在のセッションで他のアプリケーション ビューに対して接続が行われた場合も、ドロップダウン リストにこれらが表示されます。隣接する [表示] ボタンをクリックして、レポジトリ内にある他のアプリケーション ビューを検索します。
- **バージョン ラベル：** 指定されたアプリケーション ビューについて定義されたすべてのバージョン ラベルがドロップダウン フィールドに一覧表示されます。バージョン ラベルは、特定のレポジトリ オブジェクト のバージョンを識別するのに役立ちます。たとえば、あるアプリケーション ビューについて 2 種類のバージョンがあるとします。ラベルを用いて、開発環境で使ったバージョンに対してラベル [テスト] を指定し、運用環境で使ったバージョンについてはラベル [運用] を指定します。適切なラベルを選択します。

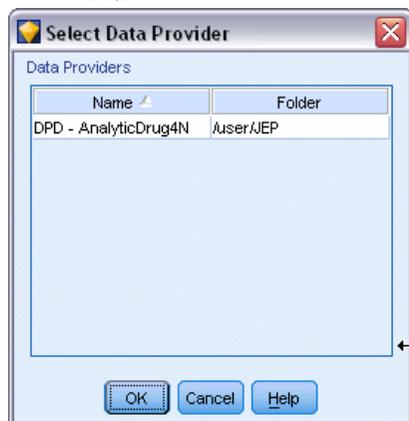
注：ラベルには “[” 文字を使用することはできません。使用した場合、テーブル名が [Enterprise View] ダイアログ ボックスの [データ] タブに表示されません。

- **環境**：有効なすべての環境がドロップダウン フィールドに一覧表示されます。環境設定によって、どの DPD が使用できるかを決定します。そのため、特定の列の指定は定義されたビジネス セグメントと関連する必要があります。たとえば、[分析] を選択すると、[分析] として定義されたアプリケーション ビューの列だけが返されます。デフォルト環境は [分析] ですが、[操作] を選択することもできます。
- **データ プロバイダ**：ドロップダウン リストには、選択したアプリケーション ビュー最大 10 個のデータ プロバイダ 定義の名前が表示されます。選択したアプリケーション ビューを参照する DPD だけが表示されます。隣接する [表示] ボタンをクリックし、現在のアプリケーション ビューに関連する DPD の名前およびパスを表示します。
- **説明**：レポジトリ接続に関する説明文。このテキストは接続名を示すために使用します。[OK] をクリックすると [接続] ドロップダウン リストおよび [Enterprise View] ダイアログ ボックスのタイトル バーにテキストが表示され、さらに Enterprise View ノードのラベルとして領域に表示されます。

DPD の選択

[データ プロバイダの選択] ダイアログ ボックスでは、現在のアプリケーション ビューを参照するすべての DPD の名前およびパスを表示します。

図 2-4
DPD の選択



アプリケーション ビューでは、複数の DPD を所有してプロジェクトのさまざまな段階をサポートすることができます。たとえば、モデルの構築に使用される履歴データがあるデータベースから取得する場合もあれば、運用データを別のデータベースから取得する場合があります。

DPD は特定の ODBC データソースに対して定義されます。IBM® 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。

- ▶ 処理する DPD を選択するには、リストに表示された DPD の名前を選択し、[OK] をクリックします。

テーブルの選択

[テーブルの選択] ダイアログ ボックスでは、現在のアプリケーション ビューを参照するすべてのテーブルを表示します。IBM SPSS Collaboration and Deployment Services Repository への接続が存在しない場合、ダイアログ ボックスは空です。

図 2-5
テーブルの選択



- ▶ 処理するテーブルを選択するには、リストに表示されたテーブルの名前を選択し、[OK] をクリックします。

データベース入力ノード

データベース入力ノードは、Microsoft SQL Server、DB2、Oracle など ODBC（開放型データベース接続）を使用するさまざまなパッケージからデータをインポートするのに使用できます。

データベースを読み書きするには、ODBC データ ソースがインストールされていて、該当するデータベースに対して必要に応じて読み取り権限や書き込み権限が設定されている必要があります。IBM® SPSS® Data Access

Pack には、この目的で使用できる ODBC ドライバが含まれています。また、これらのドライバは、IBM SPSS Data Access Pack DVD またはダウンロード サイトから入手できます。ODBC データ ソースに関する質問がある場合は、データベース管理者に問い合わせてください。

IBM® SPSS® Modeler のデータベースのサポートは 3 つの段階に分類され、それぞれデータベース ベンダーに応じて SQL プッシュバックおよび最適化のサポートについて異なるレベルを示しています。各レベルのサポートは、さまざまなシステム設定を使用して実装され、サービス契約の一環としてカスタマイズすることができます。

データベース サポートの 3 つの段階は次のようになります。

テーブル 2-1
データベース サポートの段階

| サポートの段階 | 説明 |
|---------|--|
| 段階 1 | すべての SQL プッシュバックを使用でき、データベース固有の SQL 最適化を行います。 |
| 段階 2 | 多くの SQL プッシュバックを使用でき、データベース固有でない SQL 最適化を行います。 |
| 段階 3 | SQL プッシュバックを使用できず、最適化も行いません。データベースの読み込み、および書き込みのみです。 |

サポートされている ODBC ドライバ

SPSS Modeler 15 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。

ドライバをインストールする場所

ODBC ドライバは、処理が行われる各コンピュータにインストールして設定する必要があることに注意してください。

- ローカル (スタンドアロン) モードで IBM® SPSS® Modeler を実行する場合は、ドライバをローカル コンピュータにインストールする必要があります。
- SPSS Modeler をリモートの IBM® SPSS® Modeler Server に対して分散モードで実行する場合、ODBC ドライバは SPSS Modeler Server がインストールされたコンピュータにインストールする必要があります。UNIX システムの SPSS Modeler Server を使用している場合は、本セクションの「UNIX システムの ODBC ドライバの設定」も参照してください。

- SPSS Modeler と SPSS Modeler Server の両方から同じデータ ソースにアクセスする必要がある場合、ODBC ドライバは両方のコンピュータにインストールする必要があります。
- 端末サービスを介して SPSS Modeler を実行する場合、ODBC ドライバは SPSS Modeler がインストールされた端末サービス サーバーにインストールする必要があります。
- 公開されたストリームを別のコンピュータで実行するために IBM® SPSS® Modeler Solution Publisher ランタイムを使用する場合、そのコンピュータに ODBC ドライバをインストールして設定する必要があります。

注：UNIX で SPSS Modeler Server を使用して Teradata データベースを使用する場合、Teradata ODBC ドライバとともにインストールされた ODBC ドライバマネージャを使用する必要があります。この変更を SPSS Modeler Server に行うには、コメントで示される `modelersrv.sh` スクリプトの最上位近くの `ODBC_DRIVER_MANAGER_PATH` に値を指定してください。この環境変数は、Teradata ODBC ドライバに付属する ODBC ドライバ マネージャの場所（Teradata ODBC ドライバのデフォルト インストールの `/usr/odbc/lib`）に設定する必要があります。変更を有効にするには、SPSS Modeler Server を再起動する必要があります。Teradata アクセスに対応する SPSS Modeler Server プラットフォーム、サポートされている Teradata ODBC のバージョンの詳細は、サポート サイト <http://www.ibm.com/support> を参照してください。

UNIX システムの ODBC ドライバの設定

デフォルトでは、DataDirect Driver Manager は UNIX システムの SPSS Modeler Server 向けには設定されていません。DataDirect Driver Manager ロードするよう UNIX を設定するには、次のコマンドを入力します。

```
cdmodeler_server_install_directory/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

これにより、デフォルトのリンクが削除され、DataDirect Driver Manager へのリンクを作成します。

通常、データベースからデータにアクセスするには、次の作業を行います。

- ▶ 使用するデータベースに ODBC ドライバをインストールして、データソースを構成します。
- ▶ [データベース ノード] ダイアログ ボックスで、テーブル モードまたは SQL クエリー モードを使用してデータベースに接続します。
- ▶ データベースからテーブルを選択します。

- ▶ [データベース ノード] ダイアログ ボックスのタブを使用して、データフィールドの使用タイプを変更したり、フィールドをフィルタリングすることができます。

これらの作業の詳細は、続く各トピックで説明します。

データベース ノード オプションの設定

[データベース入力ノード] ダイアログ ボックスの [データ] タブにあるオプションを使用して、データベースにアクセスし、選択したテーブルからデータを読み込むことができます。

図 2-6
テーブルを選択してデータをロード



モード: ダイアログ ボックスのコントロールを使用してテーブルに接続するには、[テーブル] を選択します。

SQL を使用して、下で選択したデータベースに問い合わせるには、[SQL クエリー] を選択します。 [詳細は、 p. 26 データベースの照会 を参照してください。](#)

データソース: テーブルおよび SQL クエリー モードのどちらでも、データソース フィールドに名前を入力するか、またはドロップダウン リストから [新規データベース接続の追加] を選択できます。

ダイアログ ボックスを使用して、データベースに接続してテーブルを選択するには、次のオプションを使用します。

テーブル名： アクセスするテーブル名が既知の場合に、[テーブル名] フィールドにその名前を入力します。そうでない場合は、[選択] ボタンをクリックして、利用可能なテーブルを表示したダイアログ ボックスを開きます。

テーブルおよび列名を引用符で囲む： クエリーをデータベースに送信するときにテーブル名と列名を引用符で囲むかどうかを指定します（たとえば、テーブル名と列名にスペースや句読点が含まれているような場合）。

- [必要に応じて] オプションを選択すると、非標準文字が含まれている場合にのみ、テーブル名とフィールド名が引用符で囲まれます。非標準文字とは、非 ASCII 文字、スペース文字、およびピリオド (.) 以外の非英数文字を指します。
- テーブル名とフィールド名を引用符で囲まない場合は、[しない] を選択します。
- すべてのテーブル名とフィールド名を引用符で囲む場合は、[常時] を選択します。

前後のスペースを削除： 文字列の前後のスペースを破棄する場合に選択します。

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。

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、IBM® 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。

データベース接続の追加

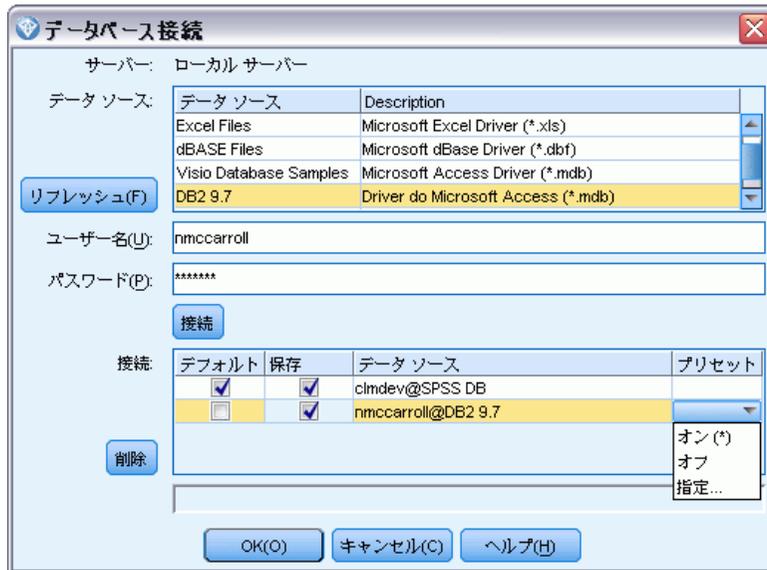
データベースを開くには、最初に、接続するデータ ソースを選択する必要があります。[データ] タブの [データ ソース] ドロップダウン リストで [新規データベース接続の追加] を選択します。

[データベース接続] ダイアログ ボックスが表示されます。

注： または、メイン メニューから次の各項目を選択して表示することもできます。

ツール > データベース(D)...

図 2-7
[データベース接続] ダイアログ ボックス



データソース: 使用できるデータ ソースの一覧が表示されています。目的のデータベースが表示されていない場合は、リストを下方へスクロールしてください。データ ソースを選択してパスワードを入力したら、[接続] をクリックします。リストを更新するには、[リフレッシュ] をクリックします。

ユーザー名: データ ソースがパスワードで保護されている場合は、ユーザー名を入力します。

パスワード: データ ソースがパスワードで保護されている場合は、パスワードを入力します。

接続: 現在接続しているデータベースが表示されます。

- **デフォルト:** オプションで、1 つの接続をデフォルトとして選択できます。これにより、データベース入力ノードまたはエクスポートノードはこの接続をデータ ソースとして事前定義します。必要に応じて編集可能です。
- **保存** オプションで、後続のセッションで再度表示する接続を選択します。
- **データソース:** 現在接続しているデータベースの接続文字列。
- **プリセット:** (* 文字を使用して) プリセット値がデータベース接続に指定されているかどうかを示します。プリセット値を指定するには、データベース接続に対応する行のこの列をクリックして、リストから [指定] を選択します。詳細は、p. 22 データベース接続のプリセット値の指定 を参照してください。

接続を除去するには、目的の接続をリストから選択し、[除去] をクリックします。

選択が完了したら、[OK] をクリックします。

データベース接続のプリセット値の指定

一部のデータベースで、データベース接続のデフォルト設定を指定できます。設定はすべてデータベース エクスポートに適用されます。

この機能をサポートするデータベースの種類は次のとおりです。

- DB2 9.1 以降で実行する IBM InfoSphere Warehouse。詳細は、[p.22 IBM DB2 InfoSphere Warehouse の設定](#) を参照してください。
- SQL Server 2008 以降 Enterprise および Developer Edition。詳細は、[p.23 SQL Server の設定](#) を参照してください。
- Oracle 10g および 11gR1 以降の Enterprise または Personal Editions。詳細は、[p.23 Oracle の設定](#) を参照してください。
- IBM Netezza、IBM DB2 for z/OS、および Teradata はすべて同様の方法でデータベースまたはスキーマに接続します。詳細は、[p.25 IBM Netezza、IBM DB2 for z/OS、および Teradata の設定](#) を参照してください。

この機能をサポートしていないデータベースまたはスキーマに接続する場合、「このデータベース接続にプリセットを設定できません」というメッセージが表示されます。

IBM DB2 InfoSphere Warehouse の設定

これらの設定は DB2 9.1 以降で実行する IBM InfoSphere Warehouse に表示されます。

テーブル スペース： エクスポートに使用するテーブルスペース。データベース管理者は、テーブルスペースを区分して作成または設定できます。データベースのエクスポートに使用する場合、(デフォルトのテーブルスペースではなく) これらのテーブルスペースのいずれかを選択することをお勧めします。

圧縮を使用： テーブルを作成し、圧縮してエクスポートします (たとえば、SQL の `CREATE TABLE MYTABLE(...) COMPRESS YES;` と同等のもの)。

更新を記録しない： テーブルの作成およびデータの挿入時にログ記録されないようにします (たとえば、SQL の `CREATE TABLE MYTABLE(...) NOT LOGGED INITIALLY;` と同等のもの)。

SQL Server の設定

これらの設定は、SQL Server 2008 以降 Enterprise および Developer Edition に表示されます。

圧縮を使用： 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル： 圧縮のレベルを選択します。

- **行：** 行レベルの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW);）。
- **ページ：** ページレベルの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE);）。

Oracle の設定

Oracle 10g の設定

これらの設定は Oracle 10g Enterprise または Personal Edition に表示されます。

圧縮を使用： 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。本バージョンのデータベースの場合、基本的な圧縮のみ可能です（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS;）。

Oracle 11gR1 の設定

これらの設定は Oracle 11g R1 Enterprise または Personal Edition に表示されます。

圧縮を使用： 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル： 圧縮のレベルを選択します。

- **デフォルト：** デフォルトの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS;）。この場合、[ダイレクトロード操作] オプションと同じ効果があります。
- **ダイレクトロード操作：** バルク（ダイレクトパス）挿入操作圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS FOR DIRECT_LOAD OPERATIONS;）。
- **すべての操作：** すべての操作の圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS FOR ALL OPERATIONS;）。

Oracle 11gR2 の設定 - 基本オプション

これらの設定は、基本オプションを使用する Oracle 11g R2 Enterprise または Personal Edition に表示されます。

圧縮を使用： 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル： 圧縮のレベルを選択します。

- **デフォルト：** デフォルトの圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS;`)。この場合、[基本] オプションと同じ効果があります。
- **基本：** 基本の圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS BASIC;`)。

Oracle 11gR2 の設定 - 高度なオプション

これらの設定は、高度なオプションを使用する Oracle 11g R2 Enterprise または Personal Edition に表示されます。

圧縮を使用： 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル： 圧縮のレベルを選択します。

- **デフォルト：** デフォルトの圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS;`)。この場合、[基本] オプションと同じ効果があります。
- **基本：** 基本の圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS BASIC;`)。
- **OLTP:** OLTP の圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS FOR OLTP;`)。
- **Query Low/High:** (Exadata サーバーのみ) QUERY の Hybrid Columnar Compression 圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY LOW;` または `CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY HIGH;`)。データ ウェアハウス環境では QUERY の圧縮が役に立ちます。HIGH は、LOW より圧縮率が高くなります。
- **Archive Low/High:** (Exadata サーバーのみ) ARCHIVE の Hybrid Columnar Compression 圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE LOW;` または `CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE HIGH;`)。長期間格納されるデータの圧縮には ARCHIVE の圧縮が役に立ちます。HIGH は、LOW より圧縮率が高くなります。

IBM Netezza、IBM DB2 for z/OS、および Teradata の設定

IBM Netezza、IBM DB2 for z/OS、または Teradata のプリセットを指定する場合、以下を選択するようメッセージが表示されます。

サーバーのスコアリング アダプタ データベース/スキーマを使用する: 選択すると、[サーバーのスコアリング アダプタ データベース/スキーマ] オプションが有効になります。

サーバーのスコアリング アダプタ データベース/スキーマ: ドロップダウン リストから、該当する接続を選択します。

データベース テーブルの選択

データ ソースに接続したら、特定のテーブルまたはビューからフィールドをインポートできます。[データベース] ダイアログ ボックスの [データ] タブで、[テーブル名] フィールドにテーブル名を入力するか、または [選択] をクリックして利用できるテーブルとビューの一覧を表示し、そこから選択します。

図 2-8
現在接続しているデータベースからテーブルを選択



テーブル所有者の表示: 指定したデータ ソースがテーブルにアクセスするためにテーブルの所有者を指定する必要がある場合に、選択します。指定する必要がないデータ ソースの場合は、このオプションの選択を解除してください。

注：通常、SAS データベースと Oracle データベースでは、テーブル所有者を表示する必要があります。

テーブル/ビュー：インポートするテーブルまたはビューを選択します。

表示：現在接続しているデータ ソースの列の一覧を表示します。次のいずれかのオプションを選択して、利用できるテーブルのビューをカスタマイズできます。

- データベース ユーザーが作成した通常のデータベース テーブルを表示する場合は、[ユーザー テーブル] をクリックします。
- システムが所有するテーブル（インデックス詳細などのデータベースに関する情報を提供するテーブル）を表示するには、[システム テーブル] をクリックします。このオプションは、Excel データベースで使用されるタブを表示する場合に必要になります（独立した Excel 入力ノードも利用可能です）。[詳細は、 p. 58 Excel 入力ノード を参照してください。](#)
- 1 つまたは複数の通常のテーブルを含むクエリーに基づいた仮想テーブルを表示するには、[表示] をクリックします。
- 既存のテーブルに対してデータベースで作成されたシノニムを表示するには、[シノニム] をクリックします。

名前/所有者フィルタ：これらのフィールドを使用すると、名前または所有者で表示されたテーブルのリストにフィルタを適用できます。たとえば、SYS と入力すると、この所有者のテーブルのみがリストに表示されます。ワイルドカード検索では、アンダースコア () を任意の 1 文字に、パーセント記号 (%) を 0 個以上の文字の並びに表現するために使用できます。

デフォルト値に設定：現在の設定を現在のユーザーのデフォルトとして保存します。この設定は、将来、ユーザーが同じデータ ソース名とユーザー ログインに対して新しいテーブル選択ダイアログ ボックスを開いた場合にだけ、復元されます。

データベースの照会

データ ソースに接続したら、SQL クエリーを使用してフィールドをインポートできます。メイン ダイアログ ボックスから、接続モードとして [SQL クエリー] を選択します。ダイアログ ボックスにクエリー エディタ ウィンドウが追加されます。クエリー エディタを使用して、結果セットがデータ ストリームに読み込まれる SQL クエリーの作成やロードなどの作業を行うことができます。

複数の SQL クエリーを指定する場合、セミコロン (;) で区切って、複数の SELECT 文がないようにします。

クエリー エディタ ウィンドウをキャンセルして閉じるには、接続モードとして [テーブル] を選択します。

図 2-9
SQL クエリを使用したデータのロード



SQL クエリに SPSS Modeler ストリーム パラメータ（ユーザ定義変数の一種）を含めることができます。詳細は、[p.27 SQL クエリのストリームパラメータの使用](#) を参照してください。

クエリのロード:以前保存したクエリをロードするファイル ブラウザを表示する場合にクリックします。

クエリの保存:現在のクエリを保存できる [クエリの保存] ダイアログ ボックスを表示する場合にクリックします。

デフォルト値のインポートダイアログ ボックスで選択されているテーブルと列を使って自動的に構築された、例の SQL SELECT 文をインポートする場合にクリックします。

消去: 作業領域の内容を消去します。作業を最初からやり直す場合に使用します。

SQL クエリのストリーム パラメータの使用

フィールドをインポートする SQL クエリを作成する場合、以前定義された SPSS Modeler ストリーム パラメータを含めることができます。詳細は、[5 章 ストリームとセッション パラメータの設定 in IBM SPSS](#)

[Modeler 15 ユーザー ガイド](#) を参照してください。すべての種類のパラメータをサポートできます。

次の表では、SQL クエリーのストリーム パラメータのいくつかの例をどのように解釈するかについて示しています。

テーブル 2-2
ストリーム パラメータの例

| ストリーム パラメータ名 (例) | ストレージ | ストリーム パラメータ値 | 解釈 |
|------------------|-----------|------------------------|--------------------------------|
| PString | String | ss | ' ss' |
| PInt | Integer | 5 | 5 |
| PReal | Real | 5.5 | 5.5 |
| PTime | Time | 23:05:01 | t{ '23:05:01' } |
| PDate | Date | 2011-03-02 | d{ '2011-03-02' } |
| PTimeStamp | TimeStamp | 2011-03-02 23:05:01 | ts{ '2011-03-02 23:05:01' } |
| PColumn | Unknown | IntValue | IntValue |

SQL クエリーでは、CLEM 式と同じ方法でストリーム パラメータを指定します。主に '\$P-<parameter_names>' を使用しますが、<parameter_name> はストリーム パラメータに定義されている名前を示します。

フィールドを参照する場合、ストレージ タイプは Unknown に指定、パラメータ値は必要に応じて引用符で囲む必要があります。SQL クエリーを入力した場合の例は以下のとおりです。

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

上記は以下のように評価されます。

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

PColumn パラメータを使用して IntValue フィールドを参照する場合、以下のようにクエリーを指定して同じ結果を取得する必要があります。

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

可変長ファイル ノード

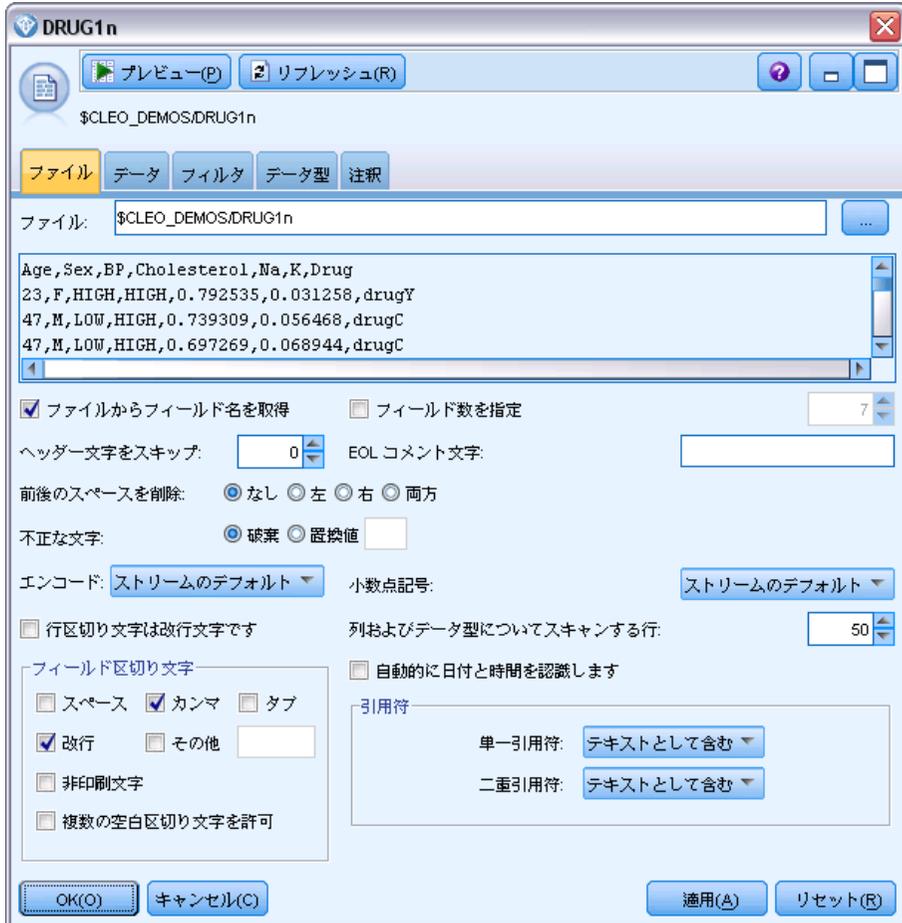
可変長ファイル ノードを使用すると、区切りテキスト ファイルとも呼ばれる可変長フィールド テキスト ファイル（フィールド数は一定だが各フィールド中の文字数が異なるレコードを含むファイル）からデータをインポートできます。このノードは、固定長のヘッダー テキストや特定の種類の注釈があるファイルにも使用できます。レコードは 1 回に 1 つずつ読み込まれ、ファイル全体が読み込まれるまでストリームを通過します。

区切りテキスト データの読み込み時の注意

- レコードは、各行の終わりで改行文字で区切る必要があります。改行文字は、(フィールド名または値内など) 他の目的で使用することはできません。理想的には前後のスペースを削除してスペースを少なくする必要がありますが、それほど重要ではありません。オプションで、ノードによって前後のスペースが削除される場合があります。
- フィールドは、カンマまたは区切り文字としてのみ使用されるそのほかの文字によって区切られる必要があります。区切り文字としてのみ使用される文字はフィールド名または値には使用されません。フィールドを区切ることができない場合、フィールド名またはテキスト値に二重引用符が含まれていなければ、すべてのテキスト フィールドを二重引用符で囲むことができます。フィールド名または値に二重引用符が使用されている場合、値内の別の場所に単一引用符が使用されていなければ、代わりに単一引用符ですべてのテキスト フィールドで囲むことができます。単一引用符および二重引用符のいずれも使用することができない場合、テキスト値を修正して区切り文字、単一引用符または二重引用符のいずれかを削除または置換する必要があります。
- ヘッダー行を含む各行には、同じ数のフィールドが含まれています。
- 最初の行にはフィールド名が含まれています。最初の行にフィールド名が含まれない場合、[ファイルからフィールド名を取得] のオプションを解除すると、データ セット中の各フィールドに フィールド 1、フィールド 2 のような一般名が付けられます。
- 2 行目にはデータの最初のレコードが含まれている必要があります。空白の行またはコメントはありません。
- 数値には、千単位の区切り文字またはグループ化記号を使用することはできません。たとえば、3,000.00 にはカンマを使用しません。小数点は、必要な場合にのみ使用します。
- 日付や時間の値には、DD/MM/YYYY または HH:MM:SS など [ストリーム オプション] ダイアログ ボックスで認識される形式のいずれかの形式を使用する必要があります。ファイル内のすべての日付/時間フィールドは同じ形式を使用するのが理想的です。また、日付を含むフィールドでは該当するフィールド内のすべての値に同じ形式を使用する必要があります。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

可変長ファイル ノードのオプションの設定

図 2-10
[可変長ファイル ノード] ダイアログ ボックス



ファイル。 ファイルの名前を指定します。ファイル名を入力するか、[...] ボタンをクリックしてファイルを選択します。ファイルを選択すると、下部にファイルパスが表示されます。また、ファイルの内容が、下のパネルに区切り文字付きで表示されます。

データ ソースから表示されたサンプル テキストは、次のコントロールを使ってコピー、貼り付けすることができます。EOL コメント文字およびユーザー指定の区切り文字。コピーと貼り付けには、それぞれ Ctrl-C および Ctrl-V を使用します。

ファイルからフィールド名を取得： デフォルトで選択されているこのオプションは、データ ファイル中の最初の行を列のラベルとして取り扱います。最初の行が見出しではない場合は、このオプションを解除すると、データ

セット中の各フィールドにフィールド 1、フィールド 2 のような数字の付けられた一般名が与えられます。

フィールド数を指定：各レコードのフィールドの数を指定します。レコードが改行文字で終了していれば、フィールド数が自動的に検出されます。フィールド数を自分で設定することもできます。

ヘッダー文字をスキップ：最初のレコードの先頭で無視する文字数を指定します。

EOL コメント文字：データ内で注釈を示す文字（# や ! など）を指定します。データ ファイル内でこれらの文字がある場所から次の改行文字のある場所までは、すべて注釈になります。ただし、その改行文字は注釈に含まれずに無視されます。

前後のスペースを削除：インポート時に文字列の前後のスペースを破棄する場合に選択します。

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。

不正な文字：データ入力から不正な文字を除去する場合に、[破棄] を選択します。不正な文字を指定した記号（1 文字だけ）で置換する場合は、[置換値] を選択します。ヌル（0）文字または指定されたエンコード方法に存在しない任意の文字が不正な文字になります。

文字コード：使用するテキストのエンコード方法を指定します。サーバー デフォルト、システム デフォルト、UTF-8 から選択できます。

- 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 IBM SPSS Modeler 内での Unicode のサポート in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- デフォルトは、[ストリーム プロパティ] ダイアログ ボックスで指定されます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

小数点記号。データ ソース内で使用されている小数点記号を選択します。[ストリームのデフォルト] は、[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] タブで選択された文字です。これを使用しない場合は、[ピリオド(.)] または [カンマ(,)] を選択すると、その文字を小数点区切り文字として、このダイアログ ボックス中のすべてのデータを読み込みます。

行の区切り文字は改行文字である：フィールド区切りの代わりに、改行文字を行の区切り文字として使用するには、このオプションを選択します。たとえば、行が折り返して表示される行に奇数の区切り文字がある場合役立つ場合があります。このオプションを選択した場合、[区切り文字] リストの [改行] は選択できません。

区切り文字：このコントロール用に表示されたチェック ボックスを使用して、カンマ (,) などの、ファイル内のフィールドの境界を定義する文字を指定できます。複数の区切り文字を使用するレコードの場合、「, |」のように複数の区切り文字を指定することもできます。デフォルトの区切り文字はカンマです。

注：カンマが、桁区切り文字としても定義されている場合、ここでのデフォルト設定は使用されません。この場合、カンマはフィールドの区切り文字と桁区切り記号の両方であるため、区切り文字リストから【その他】を選択します。次に、手動で入力フィールドにカンマを指定します。

隣接する複数の空白文字を単一の区切り文字として認識する場合は、【複数の空白区切り文字を許可】を選択します。たとえば、あるデータ値の後に 4 つのスペースが続き、その後に別のデータ値が続いている場合は、5 つのフィールドではなく、2 つのフィールドとして扱われます。

列およびタイプのスキャン行：指定したデータ型をスキャンする行および列数を指定します。

自動的に日付と時刻を認識する：IBM® SPSS® Modeler がデータ エントリを日付または時刻として自動的に認識できるようにするには、このチェック ボックスを選択します。たとえば、07-11-1965 などのエントリを日付として識別し、02:35:58 を時刻として認識します。ただし、07111965 や 023558 のようなあいまいなエントリは、数値の間に区切り文字がないため、整数として表示されます。

注：以前のバージョンの SPSS Modeler のデータ ファイルを使用する場合に考えられるデータ上の問題を回避するために、13 より前のバージョンで保存された情報についてはデフォルトでこのボックスがオフになります。

引用符：ドロップダウン リストを使用して、インポート時に単一引用符および二重引用符をどのように取り扱うかを指定できます。すべての引用符を【破棄】、フィールド値として引用符を【テキストとして含める】、または【ペアで破棄】を選択して、引用符のペアを組み合わせることで破棄することができます。対応する引用符がない場合は、エラー メッセージが表示されます。【破棄】と【ペアで破棄】では、フィールド値を文字列として（引用符なしで）保存します。

このダイアログ ボックスで作業中は、任意の時点で【リフレッシュ】をクリックすると、フィールドがデータ ソースから再ロードされます。これは、入力ノードへのデータ接続を変更したり、ダイアログ ボックス内のタブ間を行き来して作業を行うような場合に役立ちます。

固定長ファイル ノード

固定長ファイル ノードを使用して、固定長フィールド テキスト ファイル（各フィールドは区切られていないが、同じ位置から始まる固定長であるファイル）からデータをインポートすることができます。コンピュータ生成のデータや、旧来のシステムのデータなどは、しばしば固定長フィールド

形式で保存されています。固定長ファイル ノードの [ファイル] タブを使用すると、データ中の列の位置や長さを簡単に指定することができます。

固定長ファイル ノードのオプションの設定

固定長ファイル ノードの [ファイル] タブを使用して、列の位置やレコードの長さを指定し、データを IBM® SPSS® Modeler に取り込むことができます。ダイアログ ボックスの中央にあるプレビュー領域では、クリックしてフィールド間の区切りを指定する矢印を追加できます。

図 2-11
固定長フィールド データの列の指定



ファイル。 ファイルの名前を指定します。ファイル名を入力するか、[...] ボタンをクリックしてファイルを選択します。ファイルを選択すると、ファイルパスが表示されます。また、ファイルの内容が、下のパネルに区切り文字付きで表示されます。

データ プレビュー領域を使用して、列の位置と長さを指定することができます。プレビュー ウィンドウの上部にあるルーラーは、変数の長さを測定し、フィールドの区切りを指定するために役立ちます。区切り点を指定するには、フィールドの上部にあるルーラーの部分をクリックします。区切り点を移動するには、それをドラッグします。区切り点を削除するには、データ プレビュー領域外にそれをドラッグ アンド ドロップします。

- 各区切り点により、下のテーブルに自動的に新規フィールドが追加されます。
- 矢印で示されている開始点は、下のテーブルの [開始位置] 列に自動的に追加されます。

改行文字のスキップ：各レコードの末尾の改行文字をスキップする場合に選択します。

ヘッダー行をスキップ：最初のレコードの先頭で無視する行数を指定します。列見出しを無視する場合などに役立ちます。

レコード長：各レコードの文字数を指定します。

フィールド：このデータ ファイルに対して定義したすべてのフィールドが表示されます。フィールドを定義するには、次の 2 つの方法があります。

- 上のデータ プレビュー領域を使って、対話的にフィールドを指定する。
- 下のテーブルに空のフィールド行を追加して、フィールドを手作業で指定する。フィールドの右側にあるボタンをクリックして、新規フィールドを追加します。次に、空のフィールドに、フィールド名、開始位置、および長さを入力します。これらのオプションを指定すると、データ プレビュー領域に矢印が自動的に追加され、それを使用して簡単に調整することができます。

以前に定義したフィールドを除去するには、削除するフィールドを一覧から選択し、赤い削除ボタンをクリックします。

開始位置：フィールドの最初の文字の位置を指定します。たとえば、レコードの 2 番目のフィールドが 16 文字目から始まる場合は、16 を開始位置として入力します。

長さ：各フィールドの最も長い値の文字数を指定します。これによって、次のフィールドとの分割点が決まります。

前後のスペースを削除：インポート時に文字列の前後のスペースを破棄する場合に選択します。

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。

不正な文字：データ入力から不正な文字を除去する場合に、[破棄] を選択します。不正な文字を指定した記号 (1 文字だけ) で置換する場合は、[置換値] を選択します。ヌル (0) 文字または現在のエンコード中に存在しない任意の文字が不正な文字になります。

文字コード: 使用するテキストのエンコード方法を指定します。サーバー デフォルト、システム デフォルト、UTF-8 から選択できます。

- 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 IBM SPSS Modeler 内での Unicode のサポート in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- デフォルトは、[ストリーム プロパティ] ダイアログ ボックスで指定されます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

小数点記号。 データ ソース内で使用されている小数点記号を選択します。[ストリームのデフォルト] は、[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] タブで選択された文字です。これを使用しない場合は、[ピリオド(.)] または [カンマ(,)] を選択すると、その文字を小数点区切り文字として、このダイアログ ボックス中のすべてのデータを読み込みます。

自動的に日付と時刻を認識する: SPSS Modeler がデータ エントリを日付または時刻として自動的に認識できるようにするには、このチェック ボックスを選択します。たとえば、07-11-1965 などのエントリを日付として識別し、02:35:58 を時刻として認識します。ただし、07111965 や 023558 のようなあいまいなエントリは、数値の間に区切り文字がないため、整数として表示されます。

注：以前のバージョンの SPSS Modeler のデータ ファイルを使用する場合に考えられるデータ上の問題を回避するために、13 より前のバージョンで保存された情報についてはデフォルトでこのボックスがオフになります。

スキャン行: 指定したデータ型をスキャンする行数を指定します。

このダイアログ ボックスで作業中は、任意の時点で [リフレッシュ] をクリックすると、フィールドがデータ ソースから再ロードされます。これは、入力ノードへのデータ接続を変更したり、ダイアログ ボックス中のタブ間を行き来して作業を行うような場合に役立ちます。

フィールドのストレージと形式の設定

固定長、可変長、XML 入力およびユーザー入力の各入力ノードで [データ] タブにあるオプションを使用すると、データが IBM® SPSS® Modeler 内に読み込まれたり作成されたりする場合のフィールドのストレージ タイプを指定できます。また、固定長、可変長およびユーザー入力ノードの場合、フィールド フォーマットおよびその他のメタデータも指定できます。

ほかのソースから読み込まれたデータの場合、ストレージは自動的に決定されますが、置換ノードまたはフィールド作成ノード内で `to_integer` のような変換関数を使用することで変更できます。

図 2-12
インポート時のストレージのデータ型とフィールド形式の上書き



フィールド: 現在のデータ セットのフィールドを表示および選択するには、[フィールド] 列を使用します。

上書き: [上書き] 列のチェック ボックスを選択して、[ストレージ] および [入力フォーマット] 列のオプションを有効にします。

データ ストレージ

ストレージは、フィールド中へのデータの格納方法を表しています。たとえば、1 と 0 の値をとるフィールドは整数データを格納します。これは、データの使用法を記述してストレージに影響を与えない、測定レベルとは異なります。たとえば、値 1 と 0 をとる整数フィールドの測定レベルをフラグ型に設定することができます。これは、普通 1=True および 0=False の値を取ります。ストレージはソースで確定する必要があります

が、測定レベルはストリームのどこでもデータ型ノードを使って変更できます。詳細は、[4 章 p.153 尺度](#) を参照してください。

指定できるストレージ タイプを次に示します。

- **文字列**：非数値データ（別名、英数字データ）を含むフィールドに使用されます。文字列には、fred、Class 2、または 1234 など、いかなる形の文字も含むことができます。注意を要するのは、文字列内の数字は計算には使えないことです。
- **整数**：値が整数で示されるフィールドです。
- **実数**：値は数字で示され、小数点を含むことがあります（整数に限定されません）。[ストリーム プロパティ] ダイアログ ボックスで表示フォーマットを指定し、データ型ノード([フォーマット] タブ) の各フィールドに上書きすることができます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- **日付**：年、月、日など、標準フォーマットで指定された日付の値です（たとえば、「2007-09-26」の表示）。[ストリーム プロパティ] ダイアログ ボックスで特定のフォーマットを指定します。
- **時間**：期間として測定される時間です。たとえば、[ストリーム プロパティ] ダイアログ ボックスで指定した現在の時間フォーマットに応じて、1 時間 26 分 38 秒続くサービスコールを「01:26:38」と表示することができます。
- **タイムスタンプ**：たとえば、2007-09-26 09:04:00 のように、[ストリーム プロパティ] ダイアログ ボックスの現在の日付および時間のフォーマットに応じた、日付と時間の両方のコンポーネントを含む値です。タイムスタンプ値は二重引用符で囲み、日付と時間の別々の値としてでなく、単一の値として解釈されるようにする必要があります(たとえば、ユーザー入力ノードで値を入力する場合に適用されます)。

ストレージの変換：また、置換ノードで `to_string` や `to_integer` などのさまざまな変換関数を使ってフィールドのストレージを変換することもできます。詳細は、[4 章 p.200 置換ノードを使ったストレージの変換](#) を参照してください。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、[ストリームのプロパティ] 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 [MON YYYY] を選択します。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。変換関数は、フィールド作成ノードのフィールド作成計算時の一時変換でも、利用できます。また、フィールド作成ノードを使って、カテゴリ値を含む文字列フィールドの読み取りなど、他の操作も実行できます。詳細は、[4 章 p.198 フィールド作成ノードを使用して値を再コード化する](#) を参照してください。

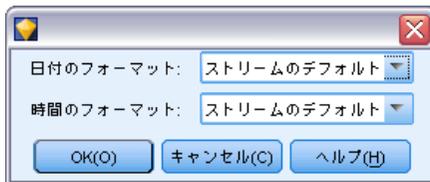
混在データの読み込み：注意を要するのは、数値ストレージ（整数、実数、時間、タイムスタンプ、または日付のいずれか）を含むフィールドで読み込む場合、数値以外の値は、ヌル値またはシステム欠損値に設定されることです。これは、一部のアプリケーションと異なり、SPSS Modeler では、1 つのフィールド内でストレージ タイプが混在することは許されないためです。これを回避するには、入力ノードでストレージ タイプを変更するか、必要な場合は外部アプリケーションで、データが混ざり合ったフィールドを文字列として読み込む必要があります。

フィールド入力形式（固定長ファイル、可変長ファイル、ユーザー入力ノードのみ）

文字列と整数以外のすべてのストレージのデータ型の場合、ドロップダウン リストを使用して選択したフィールドに対して、フォーマットのオプションを指定できます。たとえば、さまざまなロケールからのデータを結合する場合、1 つのフィールドの小数点区切り文字としてピリオドの指定が必要な場合がありますが、カンマが必要な場合もあります。

入力ノードで指定した入力オプションにより [ストリームのプロパティ] ダイアログ ボックスで指定したフォーマット オプションは無効になりますが、その後のストリームには反映されません。これらは、データの内容にしたがって入力を正確に解析するように作成されています。指定されたフォーマットは、データが SPSS Modeler に読み込まれるときにそのデータを解析するのに使用され、SPSS Modeler に読み込まれたあとでそのデータをフォーマットする方法を指定するものではありません。ストリームの他の場所でフィールドごとにフォーマット オプションを指定するには、データ型ノードの [フォーマット] タブを使用します。詳細は、4 章 p. 170 フィールド形式の [設定] タブ を参照してください。

図 2-13
タイムスタンプ フィールドの時間および日付のフォーマットを指定する



オプションはストレージ タイプによって異なります。たとえば、実数の場合、小数点区切り文字として [ピリオド(.)] または [カンマ(,)] を選択できます。タイムスタンプ フィールドでは、ドロップダウン リストで [指定] を選択すると、個別のダイアログ ボックスが表示されます。詳細は、4 章 p. 171 フィールド形式のオプションの設定 を参照してください。

すべてのストレージ タイプに対して、[ストリームのデフォルト] を指定し、インポートにストリームのデフォルト設定を使用できます。ストリームの設定は、[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。詳細は、5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。

付加オプション

他にも [データ型] タブを使用して指定できるさまざまなオプションがあります。

- まだ現在のノードを通じて接続されていないデータ（学習データなど）のストレージ設定を表示するには、[未使用のフィールド設定を表示] を選択します。古いフィールドを消去するには、[消去] をクリックします。
- このダイアログ ボックスで作業中は、任意の時点で [リフレッシュ] をクリックすると、フィールドがデータ ソースから再ロードされます。これは、入力ノードへのデータ接続を変更したり、ダイアログ ボックス中のタブ間を行き来して作業を行うような場合に役立ちます。

Data Collection ノード

Data Collection 入力ノードは、IBM Corp. の市場調査ソフトウェアが使用する IBM® SPSS® Data Collection Survey Reporter Developer Kit に基づいて、調査データをインポートします。この形式は、調査時に集められた質問に対する実際の回答である **ケース データ** と、ケース データがどのように収集され、構成されたかを説明する **メタデータ** とを区別します。メタデータは、質問テキスト、変数名とその説明、複数のレスポンスの変数定義、テキスト文字列の翻訳、ケース データの構造の定義などの情報から構成されます。

注： このノードには Data Collection Survey Reporter Developer Kit が 必要です。IBM Corp. の Data Collection ソフトウェア製品とともに配布されます。詳細は、Data Collection の Web ページ <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-repo> を参照してください。Developer Kit のインストールとは別に、追加の設定を行う必要はありません。

コメント

- 調査データは、フラットな表形式の VDATA フォーマットから、またはメタデータ ソースが含まれている場合、階層上の HDATA フォーマットのソースから読み込まれます（Data Collection 4.5 以上が必要）。
- データ型は、メタデータの情報を使用して、自動的にインスタンス化されます。
- 調査データが IBM® SPSS® Modeler へインポートされると、各回答者を 1 レコードにして、質問がフィールドとして提供されます。

Data Collection インポート ファイルのオプション

Data Collection ノードの [ファイル] タブを使用して、インポートするメタデータとケース データのオプションを指定できます。

図 2-14

Data Collection 入力ノードのファイル オプション



メタデータの設定

注： 使用できるプロバイダのファイルの種類についての完全なリストを表示するには、IBM® SPSS® Data Collection ソフトウェアに付属するData Collection Survey Reporter Developer Kit をインストールする必要があります。詳細は、Data Collection Web ページ

<http://www.ibm.com/software/analytics/spss/products/data-collection/survey-repo>を参照してください。

メタデータ プロバイダ :調査データは、Data Collection Survey Reporter Developer Kit ソフトウェアにサポートされる多くのフォーマットからインポートできます。利用できるプロバイダの種類には、次のものがあります。

- **DataCollectionMDD** : 質問定義ファイル (.mdd) からメタデータを読み込みます。これは、標準の Data Collection Data Model フォーマットです。
- **ADO データベース** : ADO ファイルからケース データとメタデータを読み取ります。メタデータが含まれた .adoinfo の名前と場所を指定します。この DSC の内部名は mrADODsc です。
- **In2data データベース** : In2data のケース データとメタデータを読み取ります。この DSC の内部名は mrI2dDsc です。
- **Data Collection ログ ファイル** :標準の Data Collection ログ ファイルからメタデータを読み込みます。通常、ログ ファイルにはファイル名拡張子の .tmp が付いています。ただし、別のファイル名拡張子が付いているログ ファイルがある場合もあります。必要な場合にはこのファイルの名前を変更して、ファイル名拡張子の .tmp を付けることもできます。この DSC の内部名は mrLogDsc です。
- **Quancept 定義ファイル** :メタデータを Quancept スクリプトに変換します。Quancept .qdi ファイルの名前を指定します。この DSC の内部名は mrQdiDrsDsc です。
- **Quanvert Database** : Quanvert のケース データとメタデータを読み込みます。.qvinf または .pkd ファイルの名前と場所を指定します。この DSC の内部名は mrQvDsc です。
- **Data Collection Participation データベース** : プロジェクトの Sample テーブルと History Table テーブルを読み込み、これらのテーブル内の列に対応する派生カテゴリ変数を作成します。この DSC の内部名は mrSampleReportingMDSC です。
- **Statistics ファイル** :IBM® SPSS® Statistics.sav ファイルからケース データとメタデータを読み込みます。SPSS Statistics 内での分析用に、SPSS Statistics.sav ファイルへケース データを書き出します。ta from an SPSS Statistics.sav ファイルからのメタデータを .mdd ファイルへ書き出します。この DSC の内部名は mrSavDsc です。
- **Surveycraft ファイル** : SurveyCraft のケース データとメタデータを読み込みます。SurveyCraft .vq ファイルの名前を指定します。この DSC の内部名は mrSCDsc です。

- **Data Collection スクリプト ファイル**: mrScriptMetadata ファイル内のメタデータを読み取ります。通常、これらのファイルには、ファイル名拡張子の .mdd または .dms が付いています。この DSC の内部名は mrScriptMDSC です。
- **Triple-S XML ファイル**: XML フォーマットの Triple-S ファイルのメタデータを読み込みます。この DSC の内部名は mrTripleSDsc です。

メタデータのプロパティ: オプションで、[プロパティ] を選択して、インポートする調査のバージョンや、使用する言語、コンテキスト、ラベルの種類を指定します。詳細は、[p. 44 IBM SPSS Data Collection インポート メタデータのプロパティ](#) を参照してください。

ケース データの設定

注: 使用できるプロバイダのファイルの種類についての完全なリストを表示するには、Data Collection ソフトウェアに付属する Data Collection Survey Reporter Developer Kit をインストールする必要があります。詳細は、Data Collection Web ページ <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-repo> を参照してください。

ケース データ設定の取得: .mddファイルだけからメタデータを読む場合は、[ケース データ設定の取得] をクリックして、どのケース データ源が選択したメタデータに関連付けられているのかを、指定されたソースにアクセスするのに必要な特定の設定とともに判断します。このオプションは、.mdd ファイルに対してだけ利用できます。

ケース データ プロバイダ: 次のプロバイダの種類がサポートされます。

- **ADO データベース**: Microsoft ADO インターフェイスを使用して、ケースデータを読み取ります。ケース データには [OLE-DB UDL] を選択して、[ケース データの URL] フィールドに接続文字列を指定します。詳細は、[p. 45 データベース接続文字列](#) を参照してください。このコンポーネントの内部名は mrADODsc です。
- **区切りテキスト ファイル (Excel)**: Excel で出力できるような、カンマ区切り形式 (.CSV) ファイルからケース データを読み込みます。内部名は mrCsvDsc です。
- **Data Collection データ ファイル**: ネイティブの Data Collection データ フォーマット ファイルからケース データを読み込みます (Data Collection 4.5 以前)。内部名は mrDataFileDsc です。
- **In2data データベース**: In2data データベース ファイル (.i2d) からケース データとメタデータを読み込みます。内部名は mrI2dDsc です。
- **Data Collection ログ ファイル**: 標準の Data Collection ログ ファイルからケース データを読み込みます。通常、ログ ファイルにはファイル名拡張子の .tmp が付いています。ただし、別のファイル名拡張子が付いて

いるログ ファイルがある場合もあります。必要な場合にはこのファイルの名前を変更して、ファイル名拡張子の .tmp を付けることもできます。内部名は mrLogDsc です。

- **Quantum データファイル:** Quantum フォーマットの ASCII ファイル (.dat) からケース データを読み込みます。内部名は mrPunchDsc です。
- **Quancept データファイル:** Quancept .drs、.drz、または .dru ファイルからケース データを読み込みます。内部名は mrQdiDrsDsc です。
- **Quanvert データベース:** Quanvert の qvinfo または .pkd ファイルからケース データを読み込みます。内部名は mrQvDsc です。
- **Data Collection データベース (MS SQL Server):** ケース データをリレーショナルな Microsoft SQL Server データベースへ読み込みます。[詳細は、 p. 45 データベース接続文字列 を参照してください。](#) 内部名は mrRdbDsc2 です。
- **Statistics ファイル:** SPSS Statistics.sav ファイルからケース データを読み込みます。内部名は mrSavDsc です。
- **Surveycraft ファイル:** SurveyCraft の .qdt ファイルからケース データを読み込みます。.vq ファイルと .qdt ファイルの両方とも、両ファイルへの読み取りと書き込みアクセス権限付きで、同じディレクトリ内に存在している必要があります。これは、SurveyCraft の使用時にデフォルトで 2 つのファイルが作成される方法ではありません。したがって、ファイルの 1 つは SurveyCraft データをインポートするために移動する必要があります。内部名は mrScDsc です。
- **Triple-S Data ファイル:** 固定長またはカンマ区切り形式の Triple-S データファイルから、ケース データを読み込みます。内部名は mr TripleDsc です。
- **Data Collection XML :** Data Collection XML データ ファイルからケース データを読み込みます。通常、このフォーマットはある場所から別の場所へケース データを転送するのに使用できます。内部名は mrXmlDsc です。

ケース データ タイプ: ケース データがファイル、フォルダ、OLE-DB UDL、または ODBC DSN から読み込まれたかどうかを指定し、それに従ってダイアログ ボックスのオプションが更新されます。有効なオプションは、プロバイダの種類によって異なります。データベース プロバイダの場合は、OLE-DB または ODBC 接続にオプションを指定できます。[詳細は、 p. 45 データベース接続文字列 を参照してください。](#)

ケース データ プロジェクト: Data Collection データベースからケース データを読み込むときに、プロジェクトの名前を入力できます。その他のケース データのデータ型については、この設定を空白のままにしておく必要があります。

変数インポート

システム変数インポート: 面談の状態（進行中、終了、終了日など）を示す変数など、システム変数をインポートするかどうかを指定します。[なし]、[すべて]、または[共通]を選択することができます。

「コード」変数インポート: カテゴリ変数の、自由回答形式の「その他」の回答に使用されるコードを示す変数のインポートを制御します。

「SourceFile」変数インポート: スキャンされたレスポンスの画像のファイル名を含む変数のインポートを制御します。

複数レスポンス変数をインポート: 複数レスポンス変数を複数のフラグ型フィールド（複合二分セット）としてインポートできます。これは新しいストリームのデフォルトの方法です。IBM® SPSS® Modeler 12.0 以前のリリースで作成されたストリームでは、値をカンマで区切った複数の回答を単一のフィールドにインポートしていました。古い方法は現在もサポートされており、以前と同じように既存のストリームを実行できますが、古いストリームを更新して新しい方法を使用することをお勧めします。 [詳細は、p. 46 複数プロパティ設定のインポート](#) を参照してください。

IBM SPSS Data Collection インポート メタデータのプロパティ

IBM® SPSS® Data Collection 調査データのインポート時に、インポートする調査のバージョンや、使用する言語、コンテキスト、ラベルの種類を指定します。一度に 1 つの言語、コンテキスト、およびラベルの種類のみをインポートできます。

図 2-15

IBM SPSS Data Collection インポート メタデータのプロパティ



バージョン: 調査の各バージョンは、ケース データの特定のセットを収集するために使用されるメタデータのスナップショットと見なすことができます。質問には変更が加えられるので、複数のバージョンが作成されることがあります。最新バージョン、すべてのバージョン、または特定のバージョンをインポートできます。

- **すべてのバージョン**：利用可能なすべてのバージョンの組み合わせ（スーパーセット）を使用する場合に、このオプションを選択します。（これは、スーパーバージョンと呼ばれることもあります。）バージョン間に矛盾がある場合は、最新バージョンが優先されます。たとえば、カテゴリのラベルがバージョン間で異なる場合、最新バージョン内のテキストが使用されます。
- **最新のバージョン**：最新のバージョンを使用する場合に、このオプションを使用します。
- **バージョンの指定**：特定の調査バージョンを使用する場合に、このオプションを使用します。

すべてのバージョンを選択することが、役立つ場合があります。たとえば、複数のバージョンのケース データをエクスポートする予定で、あるバージョンで収集されたケース データが別のバージョンで有効でないことになる、変数とカテゴリの定義に変更があるときなどです。ケース データをエクスポートするバージョンすべてを選択することは、バージョンの違いが原因の有効性のエラーが発生することなく、異なるバージョンで収集されたケース データを同時にエクスポートできる、ということです。ただし、バージョンの変更に応じて、何らかの有効性のエラーが引き続き発生する可能性があります。

言語：質問と関連テキストは、メタデータ内に複数の言語で格納できます。調査にデフォルトの言語を使用することも、特定の言語を指定することもできます。ある項目が指定された言語で利用できない場合、デフォルトが使用されます。

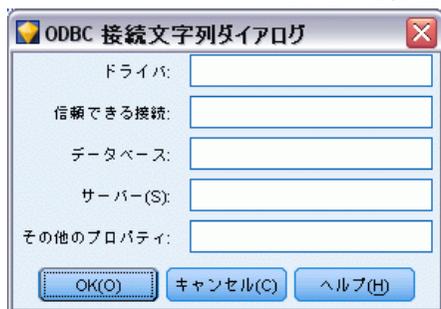
コンテキスト：使用するユーザー コンテキストを選択します。ユーザー コンテキストで、表示されるテキストが制御されます。たとえば、質問のテキストを表示するには [質問] を選択し、データを分析するときの表示に適した短いテキストを表示するには [分析] を選択します。

ラベル型：定義されたラベルの種類を一覧表示します。デフォルトは、[質問] ユーザー コンテキストで質問のテキストに使用され、[分析] ユーザー コンテキストで変数の説明に使用される [ラベル] です。その他のラベルの種類は、指示、説明などのために定義されます。

データベース接続文字列

OLE-DB または ODBC 経由でデータベースからケース データをインポートするために IBM® SPSS® Data Collection ノードを使用する場合は、[ファイル] タブから [編集] を選択して [接続文字列] ダイアログ ボックスを利用します。このダイアログ ボックスで、接続を微調整するために、プロバイダに渡す接続文字列をカスタマイズできます。

図 2-16
IBM SPSS Data Collection インポート接続文字列



詳細プロパティ

明示してログインすることが必要なデータベースからケース データをインポートするために IBM® SPSS® Data Collection ノードを使用する場合は、[高度な設定] を選択して、データ ソースにアクセスするためのユーザー ID とパスワードを提示します。

図 2-17
IBM SPSS Data Collection インポート詳細プロパティ



複数プロパティ設定のインポート

複数レスポンス変数は、変数の各値に個別のフラグ型フィールドを持つ複合二分セットとして、IBM® SPSS® Data Collection からインポートできます。たとえば、回答者が訪れたことのある博物館をリストから選択するよう質問された場合、セットには表示されたそれぞれの博物館に個別のフラグ型フィールドが含まれます。

図 2-18
複数回答の質問

質問 14 どの博物館または美術館に行ったことがありますか？ または行きたいですか？
該当するものをすべて選んでください。

| | |
|----------------------|--------------------------|
| 国立科学博物館..... | <input type="checkbox"/> |
| デザイン博物館..... | <input type="checkbox"/> |
| テキスタイル/ファッション協会..... | <input type="checkbox"/> |
| 考古学博物館..... | <input type="checkbox"/> |
| 国立美術館..... | <input type="checkbox"/> |
| 北部美術館..... | <input type="checkbox"/> |
| その他 (記入してください)..... | <input type="checkbox"/> |
| なし..... | <input type="checkbox"/> |

データをインポートした後、[フィルタ] タブを含むノードの複数回答のセットを追加または編集することができます。詳細は、4 章 p.177 [複数レスポンス セット編集](#) を参照してください。

図 2-19
[複数レスポンス セット] ダイアログ ボックス

複数レスポンス セット

セット型:

| 含めま... | グループ | ステー... |
|-------------------------------------|---------------------------|--------|
| <input type="checkbox"/> | \$Respondent_Origin | (編集済み) |
| <input type="checkbox"/> | \$DataCollection_Status | (編集済み) |
| <input checked="" type="checkbox"/> | \$museums | (編集済み) |
| <input type="checkbox"/> | \$order##First##.Column | (編集済み) |
| <input type="checkbox"/> | \$order##Second##.Colu... | (編集済み) |
| <input type="checkbox"/> | \$order##Third##.Column | (編集済み) |
| <input type="checkbox"/> | \$order##Fourth##.Column | (編集済み) |
| <input type="checkbox"/> | \$order##Fifth##.Column | (編集済み) |

新規...
編集(E)...
削除(D)

情報設定

名前: \$museums

ラベル: Museums and galleries visited or plans to visit

データ型: 複数二分法設定

カウントした値: 1

設定のフィールド:

- museums.National_Museum_of_Science
- museums.Museum_of_Design
- museums.Institute_of_Textiles_and_Fashion
- museums.Archeological_Museum
- museums.National_Art_Gallery
- museums.Northern_Gallery
- museums.Other

OK(O) キャンセル(C)

単一フィールドへの複数回答のインポート（以前のリリースで作成されたストリームの場合）

IBM® SPSS® Modeler の以前のリリースでは、前述のように複数回答をインポートするのではなく、カンマで値を区切って単一フィールドにインポートしていました。この方法は現在もサポートされており、既存のストリームがサポートされていますが、既存のストリームを更新して新しい方法を使用することをお勧めします。

IBM SPSS Data Collection 列インポート ノード

IBM® SPSS® Data Collection データの列が、次の表に要約したように、IBM® SPSS® Modeler へ読み込まれます。

| Data Collection 列のタイプ | SPSS Modeler ストレージ | 測定レベル |
|--|-------------------------------------|-----------------|
| ブール型フラグ (yes/no) | String | フラグ型 (0 と 1 の値) |
| カテゴリ | String | 名義 |
| 日付またはタイム スタンプ | Timestamp | 連続型 |
| 倍精度 (指定された範囲内の浮動小数点数値) | Real | 連続型 |
| 長形 (指定された範囲内の整数値) | Integer | 連続型 |
| テキスト (自由形式のテキスト記述) | String | Typeless |
| レベル (質問内のグリッドまたはループを示す) | VDATA 内にはなく、SPSS Modeler へインポートされない | |
| オブジェクト (走り書きのテキストを示すファクシミリや音声の録音などのバイナリ データ) | SPSS Modeler へインポートされません | |
| なし (不明なデータ型) | SPSS Modeler へインポートされません | |
| Respondent.Serial 列 (一意の ID を各回答者に関連付ける) | Integer | Typeless |

メタデータから読み込んだ値のラベルと実際の値の間に潜在する矛盾を避けるために、すべてのメタデータ値が小文字に変換されます。たとえば、値のラベル E1720_years は e1720_years へ変換されます。

IBM Cognos BI 入力ノード

IBM Cognos BI 入力ノードを使用すると、Cognos BI データまたは 1 つの表のレポートをデータ マイニング セッションに取り込むことができます。このように、Cognos のビジネス インテリジェンス機能を IBM® SPSS® Modeler の予測分析機能とを統合できます。関連する ディメンションを使用してモデル化されたリレーショナル (DMR) データおよび OLAP データをインポートできます。

Cognos サーバー接続から、まずデータまたはレポートをインポートする場所を選択します。場所には Cognos モデルと、モデルに関連するすべてのフォルダ、クエリー、レポート、ビュー、ショートカット、URL、ジョブ定義が含まれます。Cognos モデルはビジネス ルール、データの説明、データの関係、ビジネスの次元のおよび階層、その他の管理タスクを定義します。

データをインポートしている場合、選択したパッケージからインポートするオブジェクトを選択します。インポートできるオブジェクトには、クエリーの件名 (データベース テーブルを示す) または各クエリーの項目 (テーブルの列を示す) があります。詳細は、[p. 49 Cognos オブジェクトのアイコン](#) を参照してください。

パッケージのフィルタが定義されている場合、フィルタをインポートできます。インポートするフィルタがインポートされるデータに関連している場合、データがインポートされる前にフィルタが適用されます。注：インポートするデータは UTF-8 形式でなければなりません。

レポートをインポートする場合、1 つまたは複数のレポートを含む、パッケージ、またはパッケージ内のフォルダを選択します。インポートするこのレポートを選択します。注：1 つのリストのレポートだけをインポートできます。複数のリストはサポートされていません。

パラメータが定義されている場合、データ オブジェクトまたはレポートに対しては、オブジェクトまたはレポートをインポートする前にこれらのパラメータの値を指定することができます。

Cognos オブジェクトのアイコン

次の表に示すように Cognos BI データベース殻インポートできるさまざまなオブジェクト タイプが、異なるアイコンで表示されます。

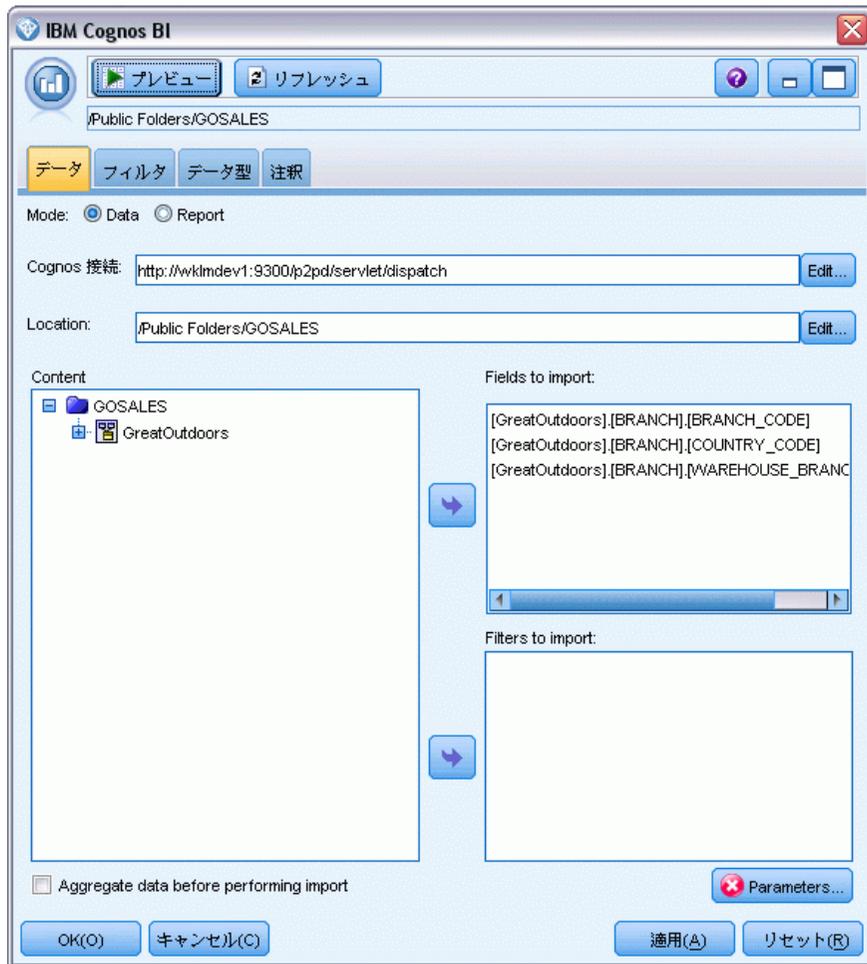
テーブル 2-3
Cognos オブジェクトのアイコン

| アイコン | オブジェクト |
|---|------------|
|  | パッケージ |
|  | 名前空間 |
|  | クエリーの件名 |
|  | クエリーの項目 |
|  | 測定の次元 |
|  | 指標 |
|  | 次元 |
|  | レベルの階層 |
|  | レベル |
|  | フィルタ |
|  | レポート |
|  | スタンドアロンの計算 |

Cognos データのインポート

IBM Cognos BI のデータベースからデータをインポートするには、[モード] が [データ] に設定され、ダイアログ ボックスを以下の様に入力します。

図 2-20
Cognos データのインポート



接続 :[編集] ボタンをクリックするとダイアログ ボックスが表示され、データまたはレポートをインポートする Cognos 接続の詳細を定義できます。IBM® SPSS® Modeler 経由ですでに Cognos サーバーにログインしている場合、現在の接続の詳細を編集することもできます。詳細は、[p. 54 Cognos の接続](#) を参照してください。

位置 : Cognos サーバー接続を確立したら、フィールドの隣の [編集] ボタンをクリックし、コンテンツをインポートするパッケージのリストを表示します。詳細は、[p. 55 Cognos の場所の選択](#) を参照してください。

内容 : 選択したパッケージの名前が、パッケージに関連する名前空間とともに表示されます。名前空間をダブルクリックすると、インポートできるオブジェクトが表示されます。さまざまなオブジェクト タイプが、

異なるアイコンで示されます。詳細は、[p. 49 Cognos オブジェクトのアイコン](#) を参照してください。

インポートするオブジェクトを選択するには、オブジェクトを選択して、2 つの右側の矢印の上部をクリックし、[インポートするフィールド] ペインにオブジェクトを移動します。クエリーの件名を選択すると、そのクエリー項目がすべてインポートされます。クエリーの件名をダブルクリックすると展開され、各クエリー項目を 1 つまたは複数選択できます。Ctrl キー（各項目を選択）、Shift キー（項目のブロックを選択）および Ctrl + A キー（すべての項目を選択）で複数の項目を選択できます。

（パッケージにフィルタが定義されている場合）適用するフィルタを選択するには、[内容] ペインのフィルタに移動し、右側の 2 つの矢印のうち下の矢印をクリックして [適用するフィルタ] ペインにフィルタを移動します。Ctrl キー（各フィルタを選択）、Shift キー（フィルタをまとめて選択）で複数の項目を選択できます。

インポートするフィールド: SPSS Modeler にインポートして処理するデータベース オブジェクトが表示されます。特定のオブジェクトを必要としなくなった場合、そのフィールドをクリックして左方向矢印をクリックすると、[内容] ペインに移動します。[内容] と同じ方法で複数の項目を選択できます。

適用するフィルタ: インポートされる前にデータに適用することを選択したフィルタが一覧表示されます。特定のフィルタを必要としなくなった場合、そのフィールドをクリックして左方向矢印をクリックすると、[内容] ペインに移動します。[内容] と同じ方法で複数の項目を選択できます。

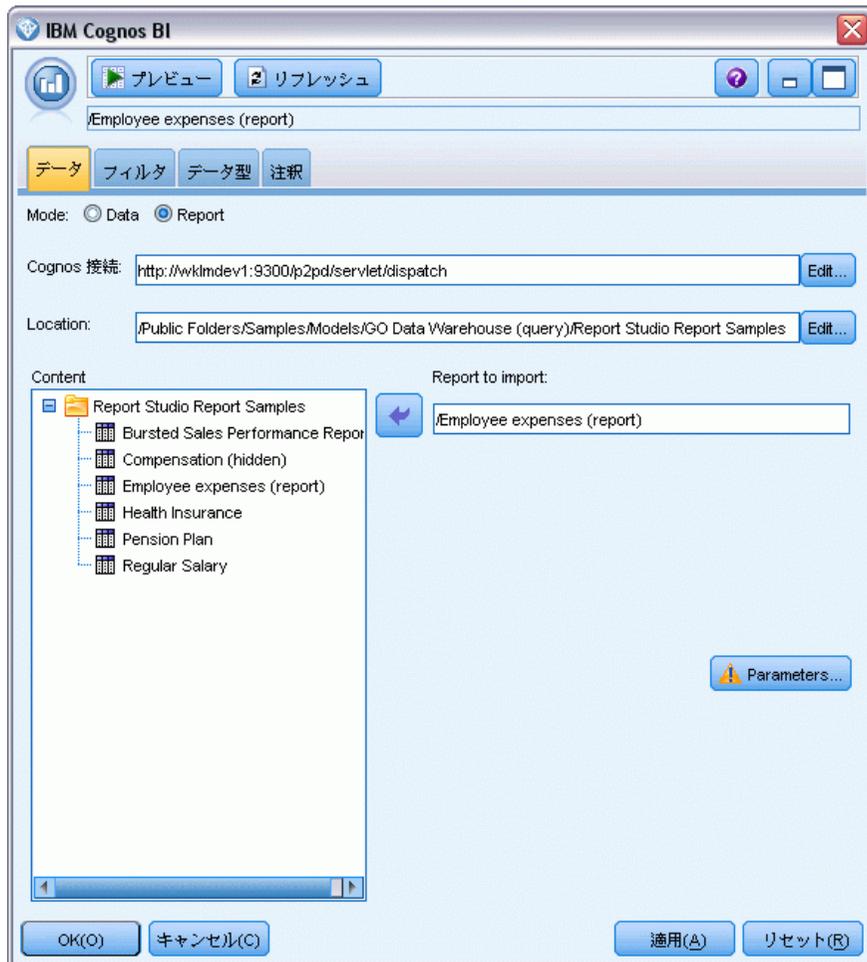
パラメータ: このボタンが有効な場合、選択したオブジェクトにパラメータが定義されます。パラメータを使用して、データをインポートする前に調整を行うことができます（たとえば、パラメータ化された計算を実行できます）。パラメータが定義されていてもデフォルトが指定されていない場合、ボタンは警告の三角形を表示します。ボタンをクリックして、パラメータを表示し、必要に応じて編集します。ボタンが無効になっている場合、レポートにはパラメータが定義されません。

インポートする前にデータを集計する: 未処理のデータではなく集計済みのデータをインポートする場合、このボックスをオンにします。

Cognos レポートのインポート

IBM Cognos BI のデータベースから定義済みデータをインポートするには、[モード] が [レポート] に設定され、ダイアログ ボックスを以下の様に入力します。注：1 つのリストのレポートだけをインポートできません。複数のリストはサポートされていません。

図 2-21
Cognos レポートのインポート



接続 [編集] ボタンをクリックするとダイアログ ボックスが表示され、データまたはレポートをインポートする Cognos 接続の詳細を定義できます。IBM® SPSS® Modeler 経由ですでに Cognos サーバーにログインしている場合、現在の接続の詳細を編集することもできます。詳細は、[p. 54 Cognos の接続](#) を参照してください。

位置 : Cognos サーバー接続を確立したら、フィールドの隣の [編集] ボタンをクリックし、コンテンツをインポートするパッケージのリストを表示します。詳細は、[p. 55 Cognos の場所の選択](#) を参照してください。

内容 : レポートを含む、選択したパッケージやフォルダの名前を表示します。特定のレポートに移動して選択し、右側の矢印をクリックして [インポートするレポート] フィールドにレポートを移動します。

インポートするレポート:SPSS Modeler へのインポートに選択されたレポートを示します。レポートを必要としなくなった場合、そのフィールドをクリックして左方向矢印をクリックすると、[内容] ペインに移動し、または異なるレポートをこのフィールドに移動します。

パラメータ: このボタンが有効な場合、選択したレポートにパラメータが定義されます。パラメータを使用して、レポートをインポートする前に調整を行うことができます（たとえば、レポート データの開始日および終了日を指定するなど）。パラメータが定義されていてもデフォルトが指定されていない場合、ボタンは警告の三角形を表示します。ボタンをクリックして、パラメータを表示し、必要に応じて編集します。ボタンが無効になっている場合、レポートにはパラメータが定義されません。

Cognos の接続

[Cognos の接続] ダイアログ ボックスを使用すると、データ オブジェクトをインポートまたはエクスポートする Cognos BI サーバーを選択できます。

図 2-22
Cognos サーバーの選択

Cognos 接続

Cognos サーバー URL:

モード: 資格情報を設定 匿名接続を使用

名前空間:

ユーザー名:

パスワード:

デフォルトとして保存

OK キャンセル ヘルプ

Cognos サーバーの URL:インポートまたはエクスポートする Cognos BI サーバーの URL を入力します。Cognos BI サーバーの IBM Cognos Configuration の「外部ディスクパッチャー URI」環境プロパティの値です。使用する URL が分からない場合、Cognos システム管理者にお問い合わせください。

モード: Cognos 名前空間、ユーザー名、パスワードを使用して（管理者として）ログインする場合、[資格情報の設定] を選択します。ユーザー資格情報を使用せずにログインする場合は [匿名接続を使用] を選択します。この場合、他のフィールドには入力されません。

名前空間: サーバーへのログインに使用する Cognos セキュリティ認証プロバイダを指定します。認証プロバイダを使用して、ユーザー、グループ、役割を定義および維持し、認証プロセスを制御します。

ユーザー名: サーバーにログインする Cognos ユーザー名を入力します。

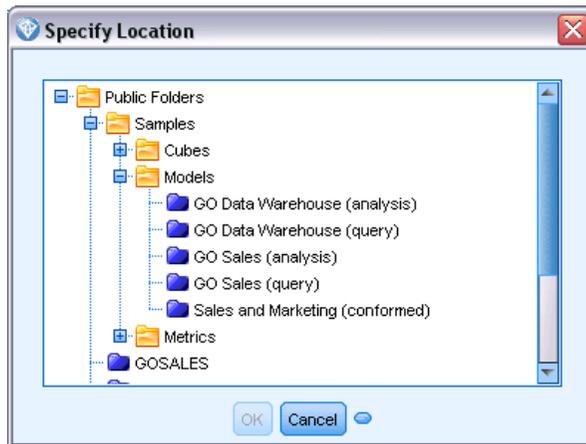
パスワード: 指定したユーザー名に対応するパスワードを入力します。

デフォルトとして保存: これらの設定をデフォルトとして保存し、ノードを開くたびに再度入力する必要がないようにします。

Cognos の場所の選択

[場所の指定] ダイアログ ボックスを使用すると、データをインポートする Cognos パッケージ、またはレポートをインポートするパッケージを選択できます。

図 2-23
Cognos の場所の選択



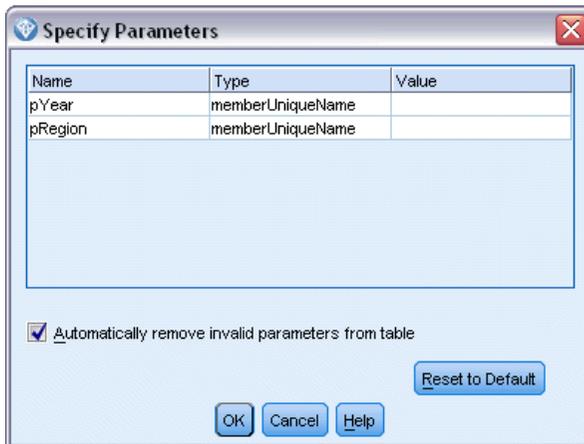
公開フォルダ: データをインポートする場合、選択したサーバから使用可能なパッケージとフォルダが一覧表示されます。使用するパッケージを選択し、[OK] をクリックします。Cognos BI 入力ノードごとに選択できるパッケージは 1 つだけです。

レポートをインポートする場合、これは、選択したサーバーから使用可能なレポートを含むフォルダやパッケージが一覧表示されます。パッケージまたはレポート フォルダを選択し、[OK] をクリックします。レポート フォルダには各レポートのほか他のレポート フォルダも含まれますが、Cognos BI 入力ノードごとにパッケージまたはレポート フォルダを 1 つだけ選択できます。

データまたはレポートのパラメータの指定

Cognos BI にパラメータが定義されている場合、データ オブジェクトまたはレポートに対しては、オブジェクトまたはレポートをインポートする前にこれらのパラメータの値を指定することができます。レポートのパラメータの例は、レポートの内容の開始日および終了日です。

図 2-24
Cognos パラメータ



名前： Cognos BI データベースで指定されるパラメータ名。

タイプ： パラメータの説明。

値： パラメータに割り当てる値。値を入力または編集するには、テーブルのセルをダブルクリックします。値はここでは検証されていないため、無効な値が実行時に検出されます。

無効なパラメータをテーブルから自動的に削除する： このオプションはデフォルトで選択され、データ オブジェクトまたはレポート内にある無効なパラメータを削除します。

SAS 入力ノード

注： この機能は、SPSS Modeler Professional および SPSS Modeler Premium で使用できます。

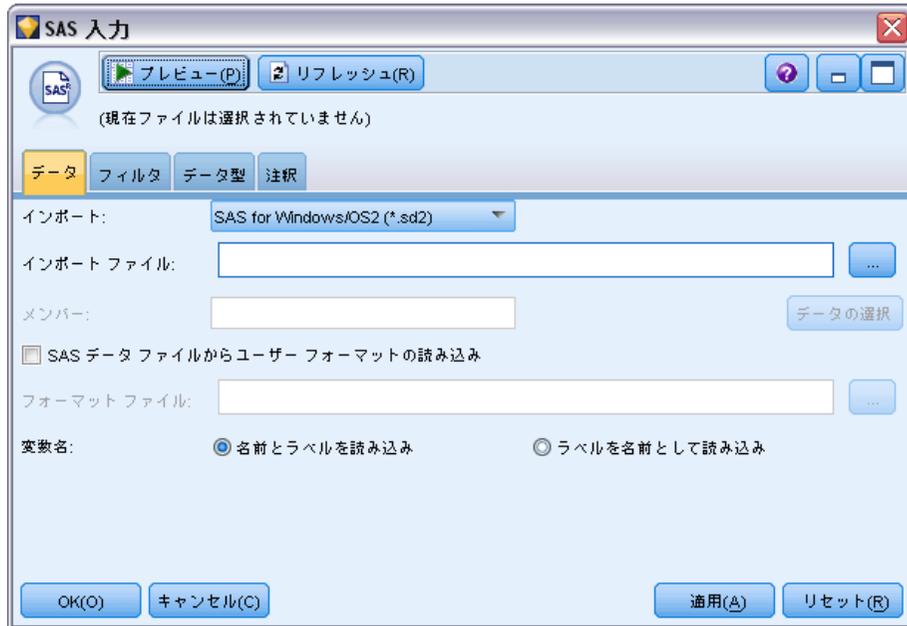
SAS 入力ノードを使用すると、SAS データをデータ マイニング セッションに取り込むことができます。次の 4 種類のファイルをインポートできます。

- SAS for Windows/OS2 (.sd2)
- SAS for UNIX (.ssd)

- SAS 移送ファイル (.tpt)
- SAS バージョン 7/8/9 (.sas7bdat)

データのインポート時は、変数はすべて保持され、変数の型は変更されません。また、すべてのケースが選択されます。

図 2-25
SAS ファイルのインポート



SAS 入カノードのオプションの設定

インポート: 移送する SAS ファイルの種類を選択します。[SAS for Windows/OS2 (.sd2)]、[SAS for UNIX (.SSD)]、[SAS トランスポートファイル (.tpt)]、または [SAS バージョン 7/8/9 (.sas7bdat)] を選択することができます。

インポート ファイル: ファイルの名前を指定します。ファイル名を入力するか、または [...] ボタンをクリックしてファイルの場所を指定します。

メンバー: 上で選択した SAS 転送ファイルからインポートするメンバーを選択します。メンバー名を入力するか、または [選択] をクリックしてファイル内のメンバーを指定します。

SAS データ ファイルからユーザー フォーマットを読み込む: ユーザー フォーマットを読み込む場合に選択します。SAS ファイルでは、データとデータフォーマット (変数ラベルなど) を別々のファイルに保存します。通常は、フォーマットもインポートします。データ セットが大きい場合は、このオプションの選択を解除すると、メモリを節約することができます。

フォーマット ファイル：フォーマット ファイルが必要な場合は、このテキスト ボックスが使用可能になります。ファイル名を入力するか、または[...] ボタンをクリックしてファイルの場所を指定します。

変数名：SAS ファイルからのインポート時に変数名とラベルを処理する方法を選択します。ここで含めることを選択したメタデータは、IBM® SPSS® Modeler 内での作業を通じて維持され、SAS 内での使用のために再びエクスポートされる可能性があります。

- **名前とラベルを読み込む：**変数名とラベルの両方を SPSS Modeler に読み込むために選択します。デフォルトでは、このオプションが選択され、変数名がデータ型ノードに表示されます。ラベルは、[ストリームのプロパティ] ダイアログ ボックスで指定したオプションに応じて、式ビルダー、グラフ、モデル ブラウザ、その他の種類の出力中に表示できます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- **ラベルを名前として読み込む：**短いフィールド名ではなく、SAS ファイルの詳細な変数ラベルを使用する場合に選択します。このラベルは SPSS Modeler の変数名として使用します。

Excel 入力ノード

Excel 入力ノードでは、Microsoft Excel のどのバージョンからでもデータをインポートできます。

図 2-26
Excel 入力ノード



ファイルの種類： インポートする Excel ファイルを選択します。

インポート ファイル： インポートするスプレッドシートの名前と場所を指定します。

名前付き範囲を使用： Excel ワークシート内で定義された名前付きのセルの範囲を指定できるようになります。省略記号ボタン ([...]) をクリックして、利用できる範囲のリストから選択します。名前付きの範囲が使用されると、その他のワークシートとデータ範囲の設定が以後適用不能になり、結果として無効になります。

ワークシートを選択： インデックスまたは名前のどちらかで、インポートするワークシートを指定します。

- **インデックスによるグループ：** インポートするワークシートのインデックス値を、最初のワークシートの 0 から始まり、2 番目のワークシートは 1 というように指定します。
- **名前順：** インポートするワークシートの名前を指定します。省略記号ボタン ([...]) をクリックして、利用できるワークシートのリストから選択します。

ワークシートの範囲： 最初の空白でない行から、または明示したセルの範囲のデータをインポートできます。

- **範囲の始点は最初に値を含む行**：最初の空白でないセルに位置決めして、これをデータ範囲の上左隅として使用します。
- **セルの明示的な範囲**：行と列で明示した範囲を指定できるようになります。たとえば、Excel の範囲 A1:D5 を指定するには、最初のフィールドで A1 と、2 番目のフィールドで D5（または、R1C1 および R5C4）と入力します。指定した範囲内のすべての行が、空白行も含めて、返されます。

空白行：複数の空白行があると、そこで [読み込みを停止] することを選択するか、その空白行も含めてワークシートの最後まですべてのデータを読み込むために [空白行を返す] を選択することができます。

最初の行に列名を指定：指定された範囲の最初の行がフィールド（列）名として使用されることを示します。選択されないと、フィールド名が自動的に生成されます。

フィールドのストレージと尺度

Excel から値を読み込むときに、数値のストレージ付きのフィールドの尺度は連続型として読み取られ、文字列フィールドは名義型として読み取られます。[データ型] タブで尺度（連続型と名義型）を変更できますが、ストレージは自動的に決定されます。ただし、必要に応じて、置換ノードまたはフィールド作成ノードで `to_integer` などの変換関数を使用して、ストレージを変更することもできます。詳細は、[p. 35 フィールドのストレージと形式の設定](#) を参照してください。

デフォルトでは、数値と文字列値の混在は数字として読み込まれます。つまり、すべての文字列値は、IBM® SPSS® Modeler 内でヌル（システム欠損値）に設定されます。このことは、SPSS Modeler は Excel と異なり、フィールド内でのストレージ型の混在を許可していないために発生します。これを避けるには、Excel スプレッドシート内でセルのフォーマットを手動で [テキスト] に設定します。そのようにすると、すべての値（数字も含めて）が文字列として読み込まれます。

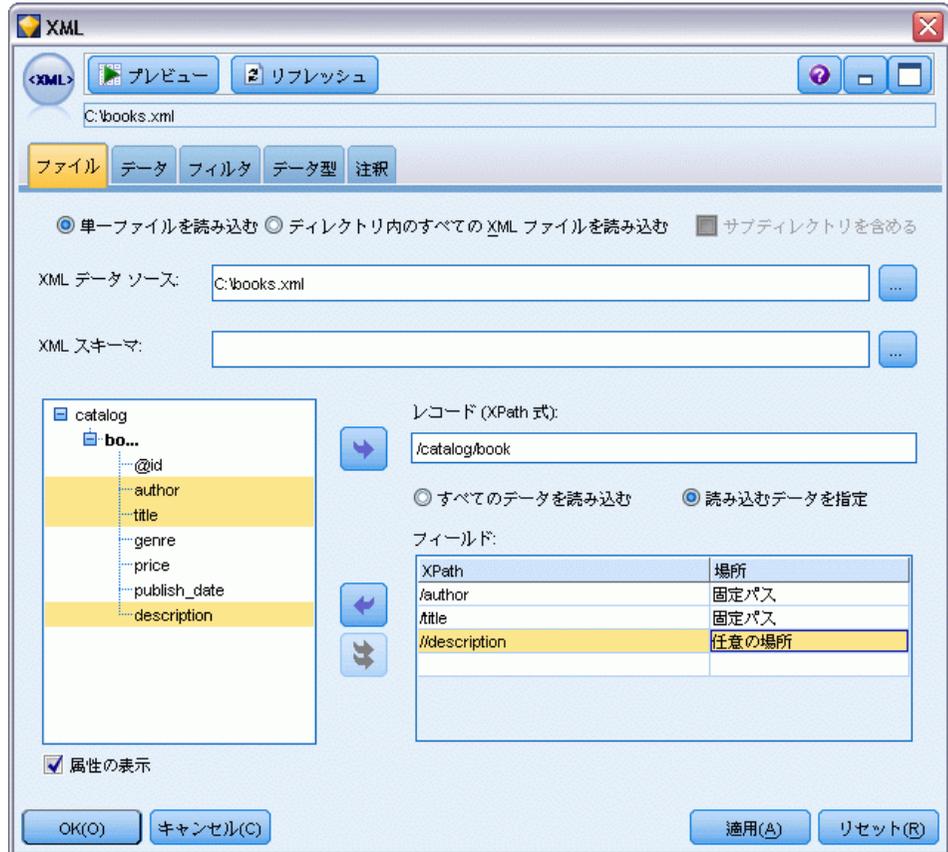
XML 入力ノード

注：この機能は、SPSS Modeler Professional および SPSS Modeler Premium で使用できます。

XML 入力ノードを使用して、ファイルから IBM® SPSS® Modeler ストリームにデータを XML 形式でインポートできます。XML はデータ交換に使用する標準言語で、多くの組織がこうした目的で選択する形式です。たとえば、国税庁は、オンラインで送信され、XML 形式のデータがある確定申告書のデータの分析が必要な場合があります。

XML データを SPSS Modeler ストリームにインポートすると、ソースに幅広い予測分析機能を実行できます。XML データがテーブル形式で解析されます。それぞれの列は、XML の要素および属性の入れ子のレベルに対応します。XML 項目は XPath 形式で表示されます (<http://www.w3.org/TR/xpath20/> を参照)。

図 2-27
XML データのインポート



単独のファイルを読み込む: デフォルトでは、SPSS Modeler は単独のファイルを読み込みます。ファイルは [XML データソース] フィールドで指定します。

ディレクトリ内のすべての XML ファイルを読み込む: 特定のディレクトリのすべての XML ファイルを読み込む場合、このオプションを選択します。表示される [ディレクトリ] フィールドで場所を指定します。[サブディレクトリを含める] チェックボックスをオンにし、指定したディレクトリのすべてのサブディレクトリから XML ファイルを読み込みます。

XML データソース: インポートする XML ソース ファイルの完全パスとファイル名を入力するか、[ブラウズ] ボタンを使用してファイルを検索します。

XML スキーマ: (オプション) XML 構造を読み込む XSD または DTD の完全パスまたはファイル名を指定するか、[ブラウズ] ボタンを使用してこのファイルを検索します。このフィールドを空白にすると、構造が XML 入力ファイルから読み込まれます。XSD ファイルまたは DTD ファイルには複数のルート要素があります。この場合、フォーカスを異なるフィールドに変更すると、使用するルート要素を選択するダイアログが表示されます。 [詳細は、 p.63 複数のルート要素からの選択 を参照してください。](#)

XML 構造: XML ソース ファイル ([XML スキーマ] フィールドで指定している場合はスキーマ) の構造を示す階層ツリー。レコードの境界を定義する場合、要素を選択して右矢印をクリックし、項目を [レコード] フィールドにコピーします。

属性の表示: [XML 構造] フィールドの XML 要素の属性を、表示または非表示にします。

レコード (XPath 式): [XML 構造] フィールドからコピーした要素の XPath シンタックスを示します。この要素は XML 構造内で強調表示され、レコードの境界を定義します。入力ファイルにこの要素が出現するごとに、新しいレコードが作成されます。このフィールドが空白である場合、ルート下の最初の子要素がレコードの境界として使用されます。

すべてのデータを読み込む: デフォルトでは、ソース ファイルのすべてのデータがストリームに読み込まれます。

読み込むデータを指定: 各要素、属性、または両方をインポートする場合、このオプションを選択します。このオプションを選択すると、[フィールド] テーブルで、インポートするデータを指定できます。

フィールド: [読み込むデータを指定] オプションを選択している場合、このテーブルには、インポートするよう選択された要素と属性が表示されます。要素または属性の XPath シンタックスを XPath 列に直接入力することも、XML 構造の要素または属性を選択して、右方向矢印ボタンをクリックして、項目をテーブルにコピーすることもできます。要素のすべての子要素および属性をコピーするには、XML 構造の要素と以下を選択し、二重矢印のボタンをクリックします。

- **XPath:** インポートする項目の XPath シンタックス。
- **位置:** インポートする項目の XML 構造の位置。[固定パス] には、XML 構造で強調表示された要素に関連する項目のパス (または強調表示される要素がない場合、ルートの下での最初の子要素) が表示されます。[任意の位置] は、XML 構造の任意の場所の指定された名前前の項目を示します。[ユーザー指定] は、XPath 列に位置を直接入力する場合に表示されます。

複数のルート要素からの選択

正しい XML ファイルが設定できるルート要素は 1 つだけですが、XSD または DTD ファイルには複数のルートを含むことができます。いずれかのルートが XML ソース ファイルと一致する場合、そのルートが使用されます。一致しない場合は、使用するルートを選択する必要があります。

図 2-28
複数のルート要素からの選択



表示するルートを選択: 使用するルート要素を選択します。デフォルトは、XSD 構造または DTD 構造の最初のルート要素です。

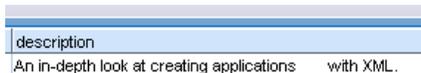
XML ソース データの不要なスペースの削除

XML ソース データの改行は、[CR][LF] 文字の組み合わせで実行できます。これらの改行は、次のように文字列の途中で発生する場合があります。

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

Web ブラウザなど、いくつかのアプリケーションでファイルが開いている場合、これらの改行が表示できない場合があります。ただし、データが XML 入力ノードを介してストリームに読み込まれると、改行は次のようにスペース文字に変換されます。

図 2-29
改行がスペースとして表示されている XML レコード



置換ノードを使用して不要なスペースを削除することにより、これを修正することができます。

図 2-30
スペースを削除する設定の置換ノード



スペースを削除する例を次に示します。

- ▶ 置換ノードを XML ノードに接続します。
- ▶ 置換ノードを開いてフィールド ピッカーを使用し、不要なスペースのあるフィールドを選択します。
- ▶ [置換] を [条件を指定] に設定、[条件] を [true] に設定します。
- ▶ [置換文字列] フィールドに、`replace(" ", "", @FIELD)` を入力して [OK] をクリックします。
- ▶ テーブル ノードを置換ノードに接続してストリームを実行します。

テーブル ノード出力に、テキストが次のように表示されます。

図 2-31
不要なスペースが削除された XML レコード

```
description  
An in-depth look at creating applications with XML.
```

ユーザー入力ノード

ユーザー入力ノードを利用すれば、最初から、または既存のデータを変更して、合成データを簡単に作成できます。これは、モデル作成用の検定データセットを作成する場合などに役立ちます。

最初からデータを作成

ユーザー入力ノードは [入力] パレットにあり、直接ストリーム領域に追加することができます。

- ▶ ノード パレットの [入力] タブをクリックします。
- ▶ ユーザー入力ノードをストリーム領域上にドラッグ アンド ドロップするか、またはダブル クリックします。
- ▶ 追加したユーザー入力ノードをダブル クリックしてダイアログ ボックスを表示し、フィールドと値を指定します。

注： [入力] パレットから選択されたユーザー入力ノードには、フィールドやデータ情報などは何も指定されていません。そのため、合成データをすべて最初から作成することができます。

既存のデータソースからのデータの生成

図 2-32
ストリーム ノードから生成されたユーザー入力ノード



ストリーム中の任意の非ターミナル ノードから、ユーザー入力ノードを生成することもできます。

- ▶ ストリーム内の、ノードを置換する位置を決めます。
- ▶ データをユーザー入力ノードに取り込むノードを右クリックして、メニューから「ユーザー入力ノードの生成」を選択します。
- ▶ ユーザー入力ノードがそのノードに関連付けられているすべての下流プロセスとともに表示され、そこに存在していた既存のノードと置換されます。生成されたユーザー入力ノードは、メタデータからデータ構造およびフィールドのデータ型情報（利用できる場合）をすべて継承します。

注： まだデータがストリーム中のすべてのノードを通過していない場合、ノードは完全にインスタンス化されていないため、ユーザー入力ノードで置換する際にストレージおよびデータ値を利用できない可能性があります。

ユーザー入力ノードのオプションの設定

ユーザー入力ノードのダイアログ ボックスには、合成データの値を入力したり、データ構造を定義するために使用できるさまざまなツールが用意されています。生成されたノードの「データ」タブのテーブルには、元のデータ ソースのフィールド名が表示されます。「入力」パレットから追加したノードの場合、テーブルには何も表示されません。このテーブルのオプションを使用して、次のような作業を実行できます。

- テーブルの右にある、「新規フィールドの追加」ボタンを使用したフィールドの追加。
- 既存のフィールド名の変更。
- 各フィールドのデータ ストレージの指定。
- 値の指定。
- 表示されているフィールドの順序を変更。

データの入力

各フィールドに対して値を指定したり、テーブルの右側にある値ピッカー ボタンを使って、オリジナルのデータ セットから値を挿入することができます。値の指定方法の詳細は、次の規則を参照してください。フィールドを空白のままにしておくこともできます。このようなフィールドには、システムの（\$null\$）ヌル値が入れられます。

図 2-33
生成されたユーザー入力ノードでの、フィールドのストレージタイプの指定



文字列値を指定するには、次のようにスペースで区切って値の列に入力します。

Fred Ethel Martin

スペースを含む文字列は、次のように二重引用符で囲まれます。

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

数値型フィールドの場合は、次のように複数の値を同じ方法（間にスペースを入れて一覧表示する方法）で入力できます。

10 12 14 16 18 20

または、同じ一連の数値を、範囲（10, 20）とその間のステップ（2）を使って指定することもできます。この方法を使用する場合は、次のように入力します。

10,20,2

上記の 2 通りの方法は、次のように一方を他方に埋め込んで組み合わせることができます。

15 7 10,20,2 21 23

この入力では、次の値が生成されます。

1 5 7 10 12 14 16 18 20 21 23

日付と時間の値は、[ストリーム プロパティ] ダイアログ ボックスで選択した現在のデフォルト形式を使用して、次のように入力することができます。

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

2007-03-14 2007-03-15 2007-03-16

日付と時間の両方のコンポーネントを含むタイムスタンプ値の場合は、次のように二重引用符を使用する必要があります。

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

詳細は、後述のデータ ストレージに関するコメントを参照してください。

データの生成： ストリームを実行するとレコードがどのようにして生成されるかを指定できます。

- **すべての組み合わせ：** フィールド値の考えられるあらゆる組み合わせを含む記録を生成するため、各フィールド値は複数のレコードに表示されます。これにより、予想以上に多くのデータが生成されることがあるので、このノードにサンプル ノードを付け加えるとよいでしょう。
- **順番に：** データ フィールド値が指定されている順序でレコードを作成します。各フィールド値のみが 1 件のレコードに表示されます。生成されるレコード数は、1 つのフィールドの最大の値に等しくなります。フィールドの値がその最大数未満の場合、未定義 (\$null\$) の値が挿入されます。

たとえば、次のように入力すると、下の表に示すようなレコードが生成されます。

- **年齢：** 30,60,10
- **BP：** LOW (低)
- **コレステロール：** NORMAL HIGH (正常 高)
- **薬品：** (空白のまま)

[データの生成] を [すべての組み合わせ] に設定：

| Age (年齢) | BP | Cholesterol (コレステロール) | Drug (薬品) |
|----------|---------|-----------------------|-----------|
| 30 | LOW (低) | NORMAL | \$null\$ |
| 30 | LOW (低) | HIGH (高) | \$null\$ |
| 40 | LOW (低) | NORMAL | \$null\$ |
| 40 | LOW (低) | HIGH (高) | \$null\$ |
| 50 | LOW (低) | NORMAL | \$null\$ |

| Age (年齢) | BP | Cholesterol (コレステロール) | Drug (薬品) |
|----------|---------|-----------------------|-----------|
| 50 | LOW (低) | HIGH (高) | \$null\$ |
| 60 | LOW (低) | NORMAL | \$null\$ |
| 60 | LOW (低) | HIGH (高) | \$null\$ |

[データの生成] を [順番に] に設定 :

| Age (年齢) | BP | Cholesterol (コレステロール) | Drug (薬品) |
|----------|----------|-----------------------|-----------|
| 30 | LOW (低) | NORMAL | \$null\$ |
| 40 | \$null\$ | HIGH (高) | \$null\$ |
| 50 | \$null\$ | \$null\$ | \$null\$ |
| 60 | \$null\$ | \$null\$ | \$null\$ |

データストレージ

ストレージは、フィールド中へのデータの格納方法を表しています。たとえば、1 と 0 の値をとるフィールドは整数データを格納します。これは、データの使用法を記述してストレージに影響を与えない、測定レベルとは異なります。たとえば、値 1 と 0 をとる整数フィールドの測定レベルをフラグ型に設定することができます。これは、普通 1=True および 0=False の値を取ります。ストレージはソースで確定する必要がありますが、測定レベルはストリームのどこでもデータ型ノードを使って変更できます。詳細は、[4 章 p.153 尺度](#) を参照してください。

指定できるストレージ タイプを次に示します。

- **文字列**：非数値データ（別名、英数字データ）を含むフィールドに使用されます。文字列には、fred、Class 2、または 1234 など、いかなる形の文字も含むことができます。注意を要するのは、文字列内の数字は計算には使えないことです。
- **整数**：値が整数で示されるフィールドです。
- **実数**：値は数字で示され、小数点を含むことがあります（整数に限定されません）。[ストリーム プロパティ] ダイアログ ボックスで表示フォーマットを指定し、データ型ノード([フォーマット] タブ) の各フィールドに上書きすることができます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- **日付**：年、月、日など、標準フォーマットで指定された日付の値です（たとえば、「2007-09-26」の表示）。[ストリーム プロパティ] ダイアログ ボックスで特定のフォーマットを指定します。

- **時間**：期間として測定される時間です。たとえば、[ストリーム プロパティ] ダイアログ ボックスで指定した現在の時間フォーマットに応じて、1 時間 26 分 38 秒続くサービスコールを「01:26:38」と表示することができます。
- **タイムスタンプ**：たとえば、2007-09-26 09:04:00 のように、[ストリーム プロパティ] ダイアログ ボックスの現在の日付および時間のフォーマットに応じた、日付と時間の両方のコンポーネントを含む値です。タイムスタンプ値は二重引用符で囲み、日付と時間の別々の値としてでなく、単一の値として解釈されるようにする必要があります(たとえば、ユーザー入力ノードで値を入力する場合に適用されます)。

ストレージの変換：また、置換ノードで `to_string` や `to_integer` などのさまざまな変換関数を使ってフィールドのストレージを変換することもできます。詳細は、4 章 p.200 [置換ノードを使ったストレージの変換](#) を参照してください。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、[ストリームのプロパティ] 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 [MON YYYY] を選択します。詳細は、5 章 [ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。変換関数は、フィールド作成ノードのフィールド作成計算時の一時変換でも、利用できます。また、フィールド作成ノードを使って、カテゴリ値を含む文字列フィールドの読み取りなど、他の操作も実行できます。詳細は、4 章 p.198 [フィールド作成ノードを使用して値を再コード化する](#) を参照してください。

混在データの読み込み：注意を要するのは、数値ストレージ（整数、実数、時間、タイムスタンプ、または日付のいずれか）を含むフィールドで読み込む場合、数値以外の値は、ヌル値またはシステム欠損値に設定されることです。これは、一部のアプリケーションと異なり、IBM® SPSS® Modeler では、1 つのフィールド内でストレージ タイプが混在することは許されないためです。これを回避するには、入力ノードでストレージ タイプを変更するか、必要な場合は外部アプリケーションで、データが混ざり合ったフィールドを文字列として読み込む必要があります。

注： インスタンス化されている場合、生成されたユーザー入力ノードには、すでにストレージ情報が入力ノードから収集されて存在していることもあります。インスタンス化されていないノードには、ストレージタイプまたは使用タイプの情報はありません。

図 2-34
生成されたユーザー入力ノードでの、フィールドのストレージタイプの指定



値の指定規則

シンボル値フィールドの場合は、次のように複数の値の間にスペースを入れる必要があります。

HIGH MEDIUM LOW

数値型フィールドの場合は、次のように複数の値を同じ方法（間にスペースを入れて一覧表示する方法）で入力できます。

10 12 14 16 18 20

または、同じ一連の数値を、範囲（10, 20）とその間のステップ（2）を使って指定することもできます。この方法を使用する場合は、次のように入力します。

10,20,2

上記の 2 通りの方法は、次のように一方を他方に埋め込んで組み合わせることができます。

1 5 7 10,20,2 21 23

この入力では、次の値が生成されます。

1 5 7 10 12 14 16 18 20 21 23

共通の入力ノード タブ

次のオプションは、すべての入力ノードで、それぞれ適切なタブをクリックして指定することができます。

- **[データ] タブ:** デフォルトのストレージ タイプを変更するために使用されます。
- **[フィルタ] タブ:** データ フィールドの除外や名前の変更などに使用されます。このタブは、フィルタ ノードと同じ機能を提供します。詳細は、[4 章 p.174 フィルタリング オプションの設定](#) を参照してください。
- **[データ型] タブ:** 尺度の設定に使用します。このタブは、データ型ノードと同じ機能を提供します。
- **[注釈] タブ:** すべてのノードで使用されるこのタブには、ノード名の変更、カスタム ツールヒントの提供、および長い注釈の保存などのオプションが用意されています。詳細は、[5 章 \[注釈\] in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

入力ノードの尺度の設定

フィールドのプロパティは、入力ノードまたは、個別のデータ型ノードで指定できます。どちらのノードでも機能は同じです。次のプロパティが使用可能です。

- **フィールド:** IBM® SPSS® Modeler 内のデータの値やフィールド ラベルを指定するには、フィールド名をダブルクリックします。たとえば、IBM® SPSS® Statistics からインポートされるフィールド メタデータを表示したり、データ型ノードで変更したりできます。同様に、フィールドの新しいラベルとそれらの値を作成できます。データ型ノードで指定したラベルは、[ストリームのプロパティ] ダイアログ ボックスの選択内容に応じて、SPSS Modeler 全体にわたって表示されます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- **尺度:** 測定レベルで、特定フィールドのデータの特性を記述するために使用します。フィールドの詳細がすべてわかっている場合には、**完全にインスタンス化済み**とされます。詳細は、[4 章 p.153 尺度](#) を参照してください。

注：フィールドの測定レベルは、ストレージ タイプとは異なります。フィールドのストレージは、データが文字列、整数、実数、数値、日付、時間、またはタイムスタンプで保管されることを示しています。

- **値:** この列を使用して、データ セットからデータ値を読み込むオプションを指定したり、[指定] オプションを使用して、測定レベルおよび値を設定する他のダイアログ ボックスを表示することができます。値を読み込まないでフィールドを渡すこともできます。詳細は、[4 章 p.158 データ値](#) を参照してください。

- **欠損値**：フィールドの欠損値の処理方法を指定します。詳細は、4章 p.165 [欠損値の定義](#) を参照してください。
- **検査**：列には、フィールドの値が指定された値または範囲内に収まっているかどうかを検査するオプションを設定できます。詳細は、4章 p.165 [データ型の値の検査](#) を参照してください。
- **役割**。フィールドがマシン学習プロセスの [入力] (予測フィールド) または [対象] (予測されるフィールド) のどちらになるかをモデル作成ノードに指示するために、使用されます。[両方] および [なし] も役割として利用できます。さらに、レコードを学習用、検定用、および検証用の独立したサンプルに分割するために使用されるフィールドを示す [データ区分] も利用できます。値 [分割] は各モデルがフィールドの可能な値それぞれに作成されるように設定します。詳細は、4章 p.167 [フィールドの役割の設定](#) を参照してください。

詳細は、4章 p.150 [データ型ノード](#) を参照してください。

図 2-35
[データ型] タブのオプション



入力ノードでインスタンス化する時期

フィールドのデータ ストレージと値を学習するには、2 種類の方法があります。この**インスタンス化**は、最初に IBM® SPSS® Modeler にデータを取り込んだ時に入力ノードで、またはデータ ストリームにデータ型ノードを挿入した時に行われます。

入力ノードでのインスタンス化は、次のような場合に役立ちます。

- データ セットが小さい場合。
- 式ビルダーを使用して新規フィールドの作成を計画している場合（インスタンス化により、式ビルダーでフィールドの値を利用できるようになります）。

一般的に、データ セットがさほど大きくなく、後でストリームにフィールドを追加する予定がない場合は、入力ノードでインスタンス化するのが便利です。

入力ノードからのフィールドのフィルタリング

[入力ノード] ダイアログ ボックスの [フィルタ] タブで、初期のデータ調査結果に基づいて下流の操作から一部のフィールドを除外することができます。このことは、たとえば、データ中に重複するフィールドがある場合や、すでにデータをよく理解しており、不要なフィールドを除外したい場合などに役立ちます。または、後の時点で別個のフィルタ ノードをストリームに追加することもできます。どちらの場合でも、機能は同じです。 [詳細は、4 章 p.174 フィルタリング オプションの設定 を参照してください。](#)

図 2-36
入力ノードからのフィールドのフィルタリング



レコード設定ノード

レコード設定の概要

レコード設定ノードは、レコード レベルでデータを変更するために使用されます。これらの操作は、特定のビジネス ニーズに合わせてデータを調整できるので、データ マイニングのデータの理解およびデータの準備フェーズ中に重要です。

たとえば、データ検査ノード（出力パレット）を使用したデータ検査の結果に基づいて、過去 3 か月の顧客購入レコードを結合するように決定できます。レコード結合ノードを使用して、Customer ID（顧客 ID）などのキー フィールドの値を基準にしてレコードを結合できます。または、Web サイトのアクセス件数に関する情報を含むデータベースに 100 万件以上のデータが蓄積されていて管理不能になっていることもあります。その場合、サンプリング ノードを使用して、モデリング用のデータのサブセットを選択することもできます。

[レコード設定] パレットには、次のノードがあります。



条件抽出ノードで、特定の条件に基づいて、データ ストリームからレコードのサブセットを選択したり破棄したりできます。たとえば、特定の営業地域に関連するレコードを選択できます。詳細は、[p. 77 条件抽出ノード](#) を参照してください。



サンプル ノードでは、レコードのサブセットを選択します。層化サンプル、クラスタ サンプル、非無作為（構造化）サンプルなど、さまざまなサンプルの種類がサポートされています。サンプリングは、パフォーマンスの向上、および分析のための関連するレコードまたはトランザクションのグループの選択に役に立ちます。詳細は、[p. 79 サンプル ノード](#) を参照してください。



バランス ノードで、データ セットが指定した条件に合うように、データ セットの不均衡を修正します。バランス式で、指定した比率によって条件が真 (true) の場合に、レコードの比率を調整します。詳細は、[p. 88 バランス ノード](#) を参照してください。



レコード集計ノードで、一連の入力レコードを要約集計された出力レコードに置き換えます。詳細は、[p. 90 レコード集計ノード](#) を参照してください。



リーセンシ、フリクエンシ、マネタリー (RFM) のレコード集計ノードを使用すると、顧客の過去のトランザクション データを取得、未使用のデータを削除、残りのトランザクション データをすべて単一行に結合することができます。これにより、最後のトランザクションの時期、トランザクション数、これらのトランザクションの合計金額が一覧表示されます。詳細は、[p. 94 RFM レコード集計ノード](#) を参照してください。



ソート ノードで、1 つまたは複数のフィールド値に基づいて、レコードを昇順または降順にソートします。詳細は、[p. 96 ソート ノード](#) を参照してください。



レコード結合ノードは、複数の入力レコードを取得し、入力フィールドの全部または一部を含む 1 つの出力レコードを作成します。この機能は、内部顧客データと購入人口データのような、異なるソースからのデータを結合する場合に役立ちます。詳細は、[p. 98 レコード結合ノード](#) を参照してください。



レコード追加ノードで、レコードのセットを連結します。レコード追加ノードは、構造が似ていながらデータが異なるデータセットを組み合せる場合に役立ちます。詳細は、[p. 111 レコード追加ノード](#) を参照してください。



重複レコード ノードで、重複レコードを削除します。その場合、最初の重複するレコードをデータ ストリームに渡すか、または、最初のレコードを破棄して、その後の重複レコードをデータ ストリームに渡します。詳細は、[p. 113 重複レコード ノード](#) を参照してください。

[レコード設定] パレットの多くのノードでは、CLEM 式を使用する必要があります。CLEM に精通している場合は、フィールドに式を入力できます。ただし、すべての式フィールドには CLEM 式ビルダーを開くボタンが用意されていて、これを使用すると、自動的に式が作成されるようになります。詳細は、[7 章 Clem 式ビルダー in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

図 3-1
[Clem 式ビルダー] ボタン



条件抽出ノード

条件抽出ノードを使用すると、BP (血圧) = "HIGH" (高) などの特定の条件に基づいて、データ ストリームからレコードのサブセットを選択したり破棄したりすることができます。

図 3-2
条件抽出ノードのダイアログ ボックス



モード: 条件を満たすレコードを、データ ストリームに入れるか、データ ストリームから除外するかを指定します。

- **含める:** 選択条件を満たすレコードを入れる場合に選択します。
- **破棄:** 選択条件を満たすレコードを除外する場合に選択します。

条件: 各レコードのテストに使用される選択条件が表示されます。CLEM 式を使用して指定します。ウィンドウに式を入力するか、またはウィンドウの右側にある計算機 (Clem 式ビルダー) ボタンをクリックして表示される Clem 式ビルダーを使用します。

次のような条件に基づいてレコードを破棄する場合、

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

デフォルトでは、条件抽出ノードがすべての選択フィールドに Null 値を持つレコードも破棄します。こうしたレコードが破棄されないよう、次の条件を元の条件に追加します。

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

条件抽出ノードは、レコードの一部を選択するためにも使用されます。通常、この操作にはサンプリング ノードなど、別のノードを使用します。ただし、用意されているパラメータよりも複雑な条件を指定する場合は、条件抽出ノードを使って独自の条件を作成します。たとえば、次のような条件を作成できます。

```
BP = "高" および 無作為 (10) <= 4
```

この条件では、高血圧を示すレコードの約 40% が選択され、詳細な分析のために下流に渡されます。

サンプル ノード

サンプル ノードを使用して、分析のためにレコードのサブセットを選択、または破棄するレコードの割合を指定することができます。層化サンプル、クラスタ サンプル、非無作為（構造化）サンプルなど、さまざまなサンプルの種類がサポートされています。サンプリングを使用する理由は、次のとおりです。

- データのサブセットのモデルを推定してパフォーマンスを向上する。サンプルから推定されたモデルは、完全なデータセットから取得したモデルと同じくらい正確で、向上したパフォーマンスによってこれまで試すことがなかったさまざまな方法を試すことができる場合、より正確になります。
- オンライン ショッピングのカートのすべてのアイテムを選択または特定の隣接地域のすべての資産を選択するなど、分析のために換算するレコードまたはトランザクションのグループを選択する。
- 品質評価、不正防止、またはセキュリティの対象となる無作為検査の単位またはケースを識別する。

注： 検証の目的でデータを学習サンプルおよび検定サンプルに分割する場合、データ区分ノードを代わりに使用することができます。 [詳細は、4 章 p. 229 データ区分ノード を参照してください。](#)

サンプルの種類

クラスタ化サンプル 個々の単位ではない、サンプル グループまたはクラスタ。たとえば、生徒ごとに 1 つのレコードを持つデータ ファイルがあるとし、学校ごとにクラスタ化し、サンプル サイズが 50% の場合、学校の 50% が選択され、選択されたそれぞれの学校からすべての生徒が取得されます。選択されない学校の生徒は却下されます。平均的には、およそ 50% の生徒が抽出されることが期待されますが、学校の規模が異なるため、割合は正確でない場合があります。同様に、トランザクション ID によってショッピング カートのアイテムをクラスタ化し、選択されたトランザクションのすべてのアイテムが含まれていることを確認します。町ごとの資産をクラスタ化する例については、`complexsample_property_str` のサンプル ストリームを参照してください。

層化サンプル: 母集団の重複しないサブグループまたは階層内のサンプルを独立して選択します。たとえば、男性および女性を等しい割合でサンプリングされ、または都市部の人口の中ですべての地域または社会経済的グループが表示されるようにすることができます。また、各階層の異なるサンプル サイズを指定することもできます（たとえば、元のデータの

1 つのグループが実際より低く評価された場合)。町ごとの資産を層化する例については、`complexsample_property.str` のサンプル ストリームを参照してください。

体系的または n 件ごとのサンプリング 無作為の選択を取得するのが難しい場合、単位を体系的（固定された間隔で）または連続してサンプリングすることができます。

抽出重み付け。 重みのサンプリングは、複雑なサンプルを引き出す際に自動的に計算され、サンプルされた各単位が元のデータに表示される「度数」にほとんど対応します。そのため、サンプルの重みの合計で、元のデータのサイズを推定する必要があります。

サンプリング フレーム

サンプリング フレームによって、サンプルまたは調査に含まれるケースの可能性のあるソースを定義します。たとえば生産ラインから外れる項目のサンプリングを行う場合、母集団の各単一メンバーを識別し、サンプルにメンバーの 1 つを含めることができます。可能性のあるすべてのケースにアクセスできない場合がよくあります。たとえば、選挙が実行された後まで、選挙で誰が投票するのかを確認できません。この場合、一部の人々が投票せず、名簿を確認した時点で登録されていない人々が投票する場合がありますが、サンプリング フレームとして選挙人名簿を使用します。サンプリング フレームに含まれない人は、サンプリングされる可能性はありません。サンプリング フレームが評価しようとしている母集団に本質的に十分近いかどうかは、それぞれの実際のケースで処理する必要のある問題です。

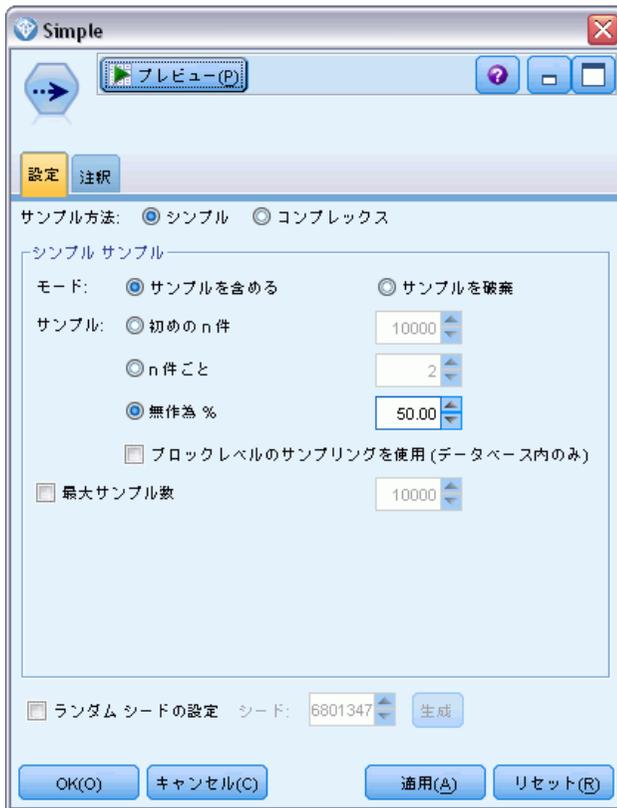
サンプル ノードのオプション

要件に応じてシンプルまたは複雑な方法を選択できます。

シンプルなサンプリングのオプション

シンプルな方法を使用すると、レコードの無作為な割合を選択、連続するレコードを選択または n 件ごとのレコードを選択することができます。

図 3-3
シンプルなサンプリングのオプション



モード: 次のモードに対して、レコードを渡す（含める）か、または破棄（除外）するかを選択します。

- **サンプルを含める:** データ ストリームの選択されたレコードを含め、他のレコードをすべて破棄します。たとえば、モードを [サンプルを含める] に設定し、[n 件ごと] に 5 を指定した場合、最大サンプル サイズになるまで 5 件ごとに 1 つのレコードがデータ ストリームに追加され、データセットが元のサイズの 5 分の 1 のサイズとなります。このモードはデータのサンプリングする際のデフォルト モードで、複雑な方法を使用する場合に唯一使用できるモードです。
- **サンプルを破棄:** 選択されたレコードを破棄し、他のすべてのレコードを含めます。たとえば、[サンプルを破棄] モードで [n 件ごと] を 5 に設定すると、5 件ごとに 1 つのレコードが破棄（除外）されます。このモードはシンプルな方法でのみ使用できます。

サンプル: 次のいずれかのサンプリング手法を選択します。

- **初めの n 件**：連続したデータ サンプリングを使用する場合に選択します。たとえば、サンプルの最大サイズが 10000 に設定されている場合、最初の 10,000 件のレコードが選択されます。
- **n 件ごと**：n 件ごとにデータを通過させるか破棄することにより、サンプル データを選択します。たとえば、n が 5 に設定されている場合、5 件ごとのレコードが選択されます。
- **無作為 %**：データを任意のパーセンテージでサンプリングする場合に選択します。たとえば、20% に設定すると、選択したモードに従って、データの 20% がデータ ストリームに渡されるか、または破棄されます。このフィールドに、サンプリングのパーセンテージを指定します。[ランダム シードの設定] から、シードの値を指定することもできます。

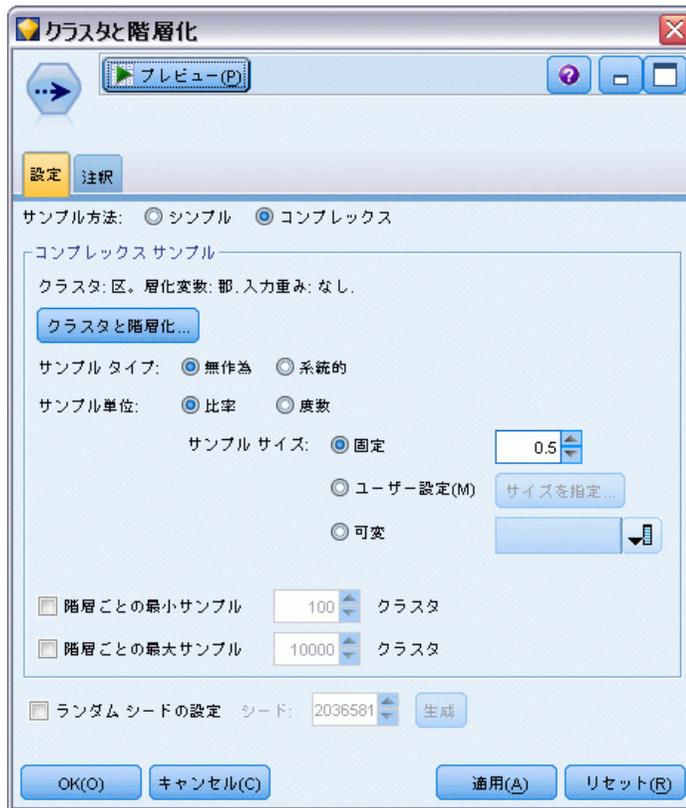
ブロックレベルのサンプリングを使用 (データベース内のみ)：このオプションは、Oracle データベースまたは IBM DB2 データベースでデータベース内マイニングを実行するときは無作為パーセント抽出を選択した場合にのみ有効です。こうした環境では、ブロックレベルのサンプリングがより効果的です。

最大サンプル数：サンプルに含めるレコードの最大数を指定します。[サンプルを含める] および [初めの n 件] が選択されている場合、このオプションは無効になります。また、[無作為 %] オプションが使用されている場合、この設定によって特定のレコードが選択されません。たとえば、データ セットに 1 千万件のレコードがあり、最大サンプル数 300 万件のレコードという設定でレコードの 50% を選択すると、最初の 600 万件のレコードだけに 50% の選択の可能性があるが、残りの 400 万件のレコードからは選択されないということになります。この制限を回避するためには、**複雑なサンプリング方法**を選択し、クラスタまたは階層変数を指定せずに 300 万件のレコードから無作為のサンプルを要求します。

複雑なサンプリングのオプション

複雑なサンプリングのオプションを使用すると、クラスタ化サンプル、層化サンプル、重み付けされたサンプルを他のオプションとともに、サンプルをより詳細に制御することができます。

図 3-4
複雑なサンプリングのオプション



クラスと階層 必要に応じて、クラス フィールド、層化フィールド、および入力重みフィールドを指定することができます。詳細は、[p. 84 クラスと階層の設定](#) を参照してください。

サンプルタイプ:

- **無作為:** 各階層内で無作為にクラスまたはレコードを選択します。
- **体系的:** 固定された間隔でレコードを選択します。このオプションは、ランダム シードに応じて最初のレコードの位置が変更することを除き、n 件ごとの方法と同じように動作します。n の値は、サンプル サイズまたは割合に基づいて自動的に決定されます。

サンプル単位: 基本的なサンプル単位として割合または度数を選択することができます。

サンプルサイズ: 次のいくつかの方法でサンプル サイズを指定することができます。

- **固定:** 全体のサンプル サイズを度数または割合として指定することができます。

- **ユーザー指定：**各サブグループまたは階層のサンプル サイズを指定することができます。このオプションは、層化フィールドが [クラスタ] および [層化] サブ ダイアログ ボックスで指定されている場合にのみ使用できます。
- **変数。**ユーザーは、各サブグループまたは階層のサンプル サイズを定義するフィールドを指定することができます。このフィールドには、特定の階層内の各レコードの同じ値が含まれています。たとえば、サンプルが地域ごとに層化されている場合、地域 = Surrey のすべてのレコードは同じ値を持つ必要があります。フィールドは数値型で、その値は選択されたサンプル単位に一致する必要があります。単位が割合の場合、値は 0 より大きく 1 より小さくなります。単位が度数の場合、最小値は 1 です。

階層ごとの最小サンプル：最小レコード数を指定します (クラスタ フィールドが指定されている場合は、最小クラスタ数が指定されます)。

階層ごとの最大サンプル：レコードまたはクラスタの最大数を指定します。クラスタまたは層化フィールドを指定せずにこのオプションを選択した場合、指定されたサイズの無作為または体系的サンプルが選択されます。

ランダム シードの設定：無作為なパーセンテージに基づいてレコードをサンプリングまたはデータ区分している場合、このオプションで、別のセッションに同じ結果を複製できるようになります。乱数ジェネレータに使用される開始値を指定することで、ノードが実行されるごとに毎回同じレコードが割り当てられることが保証されます。自動的に無作為な値を生成するには、希望のシード値を入力するか、[生成] ボタンを入力します。このオプションが選択されないと、ノードが実行されるごとに異なるサンプルが生成されます。

注： データベースから読まれるレコードに [ランダム シードの設定] オプションを使用する場合は、ノードが実行されるごとに同じ結果を保証するために、サンプリングに先行して、ソート ノードが必要になる可能性があります。この理由は、ランダム シードがレコードの順序に依存しているためです。各レコードがリレーショナル データベース内で同じ位置に留まる保証はありません。 [詳細は、 p.96 ソート ノード を参照してください。](#)

クラスタと階層の設定

複雑なサンプルを引き出す場合、[クラスタ] および [階層] ダイアログボックスを使用して、クラスタ、階層および重みフィールドを選択することができます。

図 3-5
クラスタおよび層化フィールドの設定



クラスタ: クラスタ レコードに使用するカテゴリ フィールドを指定します。レコードは、含まれるクラスタと含まれないクラスタがある所属クラスタに基づいてサンプリングされます。ただし、指定のクラスタのレコードが含まれる場合、すべてのクラスタが含まれます。たとえば、ショッピング カート内の商品の関連を分析する場合、トランザクション ID によってアイテムをクラスタ化し、選択されたトランザクションのすべてのアイテムが含まれていることを確認します。同時に販売されるアイテムに関する情報を破棄するレコードのサンプリングの代わりに、トランザクションをサンプリングして、選択されたトランザクションのすべてのレコードが保存されます。

階層化する: レコードを層化するために使用するカテゴリ フィールドを指定し、サンプルが母集団の重複していないサブグループまたは階層内で独立して選択されるようにします。たとえば、性別によって層化された 50% のサンプルを選択する場合、男女それぞれに 2 つの 50% サンプルが選択されます。たとえば、階層は社会経済的なグループ、職業のカテゴリ、年齢グループ、民族グループで、対象となるサブグループの適切なサンプルサイズを確認することができます。元のデータセットで、男性の 3 倍の女性がいる場合、この比率は各グループとは別にサンプリングすることによって保存されます。複数の層化フィールドを指定することもできます (たとえば、地域内の製品ラインのサンプリング、またはその逆)。

注： 欠損値 (ヌルまたはシステム欠損値、空白文字列、空白文字、空欄またはユーザー定義の欠損値) のあるフィールドによって層化する場合、階層のカスタム サンプル サイズを指定することはできません。欠損値または空白値を含むフィールドによって層化する際にユーザー定義のサンプル サイズを使用する場合、上流でそれらの値を入力する必要があります。詳細は、[6 章 欠損値を含むフィールドの処理 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

入力重みのみ使用: サンプリングの前にレコードを重み付けするために使用するフィールドを指定します。たとえば、重みフィールドに 1 ~ 5 の値が含まれる場合、5 の重み付けをされたレコードは 5 倍選択される可能性があります。このフィールドの値は、ノードに生成された最後の出力の重みに上書きされます（次の段落を参照）。

新規出力重み: 入力重みフィールドが指定されていない場合、最後の重みが書き込まれるフィールド名を指定します。（入力重みが指定されている場合、その値は前述のとおり最後の重みに置き換えられますが、作成される個別の出力重みフィールドはありません。）出力重みの値は、元のデータのサンプリングされた各レコードによって示されるレコード数を示します。重みの値の合計によって、サンプル サイズの推定が得られます。たとえば、無作為の 10% のサンプルが選択された場合、出力重みはすべてのレコードに対して 10 となり、サンプリングされた各レコードが元のデータの 10 件のレコードを表すことを示します。層化サンプルまたは重み付けされたサンプルでは、出力重みの値は各階層のサンプルの割合によって異なります。

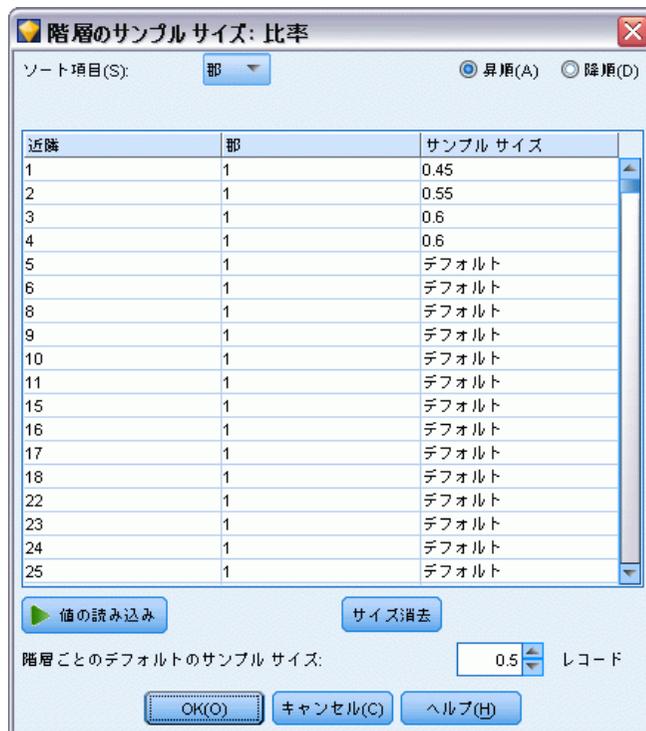
コメント

- クラスタ化されたサンプルは、サンプリングする割合の詳細なリストを取得できないが、特定のグループまたはクラスタの詳細なリストを取得できる場合に役に立ちます。無作為サンプルによって連絡が取れない被験者のリストを作成する場合にも使用されます。たとえば、国内のすべての地域に散在する農家を選択するよりも、1 つの地域のすべての農家を訪問することが簡単です。
- クラスタ フィールドおよび層化フィールドの両方を選択して、各階層内のクラスタを独立してサンプリングすることができます。たとえば、地域によって層化し、地域内の町ごとにクラスタ化した資産価値をサンプリングすることができます。これによって、町の独立したサンプルを各地域内から引き出すことができます。サンプルに含まれる町と含まれない町がある場合、含まれる町については、町内のすべての資産が含まれます。
- 各クラスタ内から単位の無作為サンプルを選択するために、2 つのサンプル ノードを結びつけることができます。たとえば、前述のように地域によって層化された町を最初にサンプリングすることができます。その後 2 番目のサンプル ノードを適用し、層化フィールドとして町を選択すると、各町ごとのレコードの割合をサンプリングすることができます。
- フィールドを組み合わせることでクラスタを一意に識別する必要がある場合、フィールド作成ノードを使用して新規フィールドを生成することができます。たとえば、複数の店舗でトランザクションの同じサンプリングシステムを使用している場合、店舗とトランザクション ID を連結する新規フィールドを取得できます。

階層のサンプル サイズ

層化サンプルを引き出す場合、デフォルトのオプションは、各階層のレコードまたはクラスタの同じ割合をサンプリングすることです。あるグループが 3 の要素で別のグループを上回る場合、通常サンプルの同じ比率をサンプルに保持します。ただしそうでない場合は、各階層ごとにサンプル サイズを個別に指定することができます。

図 3-6
階層のサンプル サイズの指定



[階層のサンプル サイズ] ダイアログ ボックスには層化フィールドの各値が一覧表示され、その階層のデフォルトを上書きすることができます。複数の層化フィールドを選択する場合、値の考えられる組み合わせがすべて表示され、たとえば各市内の各民族グループまたは各地域内の町ごとのサイズを指定することができます。サンプル ノードの現在の設定で決められたように、サイズは割合または度数として指定されています。

階層のサンプル サイズを指定するには

- ▶ サンプル ノードで [複雑] を選択し、1 つまたは複数の層化フィールドを選択します。詳細は、[p. 84 クラスタと階層の設定](#) を参照してください。
- ▶ [ユーザー設定] を選択し、[サイズを指定] を選択します。

- ▶ [階層のサンプル サイズ] ダイアログ ボックスで、左下の[値の読み込み] ボタンをクリックして表示を指定します。上流のソースまたはデータ型 ノードの値をインスタンス化する必要がある場合があります。 [詳細は、4 章 p.157 インスタンス化とは？ を参照してください。](#)
- ▶ 行をクリックして、該当する階層のデフォルト サイズを上書きします。

サンプル サイズの注意

異なる階層に異なる分散が含まれている場合、たとえばサンプル サイズを標準偏差に比例させる場合に、ユーザー設定のサンプル サイズが役に立ちます。(階層内のケースがより異なる場合、より多くのケースをサンプリングして標本サンプルを取得する必要があります。)または階層が小さい場合、より高いサンプルの割合を使用して、最小観察数が含まれていることの確認が必要な場合があります。

注： 欠損値（ヌルまたはシステム欠損値、空白文字列、空白文字、空欄またはユーザー定義の欠損値）のあるフィールドによって層化する場合、階層のカスタム サンプル サイズを指定することはできません。欠損値または空白値を含むフィールドによって層化する際にユーザー定義のサンプル サイズを使用する場合、上流でそれらの値を入力する必要があります。 [詳細は、6 章 欠損値を含むフィールドの処理 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。](#)

バランス ノード

バランス ノードを使って、データ セットの不均衡を修正し、指定したテスト基準を満たすことができます。たとえば、データ セットに 低 と 高 という 2 つの値しかなく、ケースの 90% が 低 で、残りの 10% が 高 である場合を考えてみましょう。このようにデータが偏っている場合、低の結果だけが学習され、より少ない高の結果は無視される傾向があるため、多くのモデリング手法で問題となります。低と高がほぼ同数で、データのバランスがとれていれば、モデルで 2 つのグループを区別するパターンを発見できる可能性が高くなります。このような場合にバランス ノードを使用して、低の結果を含むケースを減らすバランス式を作成します。

バランスの調整で実際に行われるのは、指定の条件に従ったレコードの複製と破棄です。適用する条件がないレコードは、すべて通過します。この処理はレコードの複製と破棄から成り立っているため、下流の操作では元のデータ シーケンスが失われます。データ ストリームにバランス ノードを追加する前に、シーケンスに関連する値を必ず作成しておきます。

注： バランス ノードは、分布図やヒストグラムから自動的に作成することができます。たとえば、データのバランスを調整して、棒グラフで表示されているように、カテゴリ型フィールドのすべてのカテゴリ全体の等しい割合を表示することができます。

例: RFM ストリームを構築して、以前のマーケティング キャンペーンに肯定的に反応して最近の顧客を識別する場合、販売会社のマーケティング部門ではバランス ノードを使用して、データ内の真 (true) と偽 (false) の回答間の差異のバランスを調整します。

バランス ノードのオプション設定

図 3-7
バランス ノードの設定



レコード バランス式: 現在のバランス式を一覧表示します。各式には、ソフトウェアに「条件が真 (true) の場合に、指定した比率だけレコードを増やす」ことを指示する比率と条件が含まれています。1.0 より低い比率を指定している場合は、指定されたレコードの比率が減少することを表しています。たとえば、処方薬が薬品 Y のレコード数を減らしたい場合、比率 0.7 で条件が Drug = "薬品Y" のバランス式を作成することができます。この式では、薬品 Y が処方薬であるレコードの数が、下流のすべての操作で 70% に減らされます。

注： 減少用のバランス比率は、小数点以下第4位まで指定できます。比率を 0.0001 未満に設定すると、結果が正しく算出されずエラーが発生します。

- テキスト フィールドの右にあるボタンをクリックすると、**条件が作成** されます。このボタンにより、新しい条件を入力するための空の行が挿入されます。条件の CLEM 式を作成するには、Clem 式ビルダー ボタンをクリックします。
- **式を削除する**には、赤い削除ボタンをクリックします。
- **式をソートする**には、赤い上矢印および下矢印ボタンを使用します。

学習データのみをバランス: ストリーム内にデータ区分フィールドがある場合、このオプションによって学習用データ区分のデータのみをバランスを調整します。特に、不均衡な検定用区分または検証用区分を要求する、調整された傾向スコアを生成する場合に役立ちます。詳細は、[3 章 傾向スコア in IBM SPSS Modeler 15 モデル作成ノード](#) を参照してください。ストリーム内にデータ区分フィールドがない場合（または複数のデータ区分フィールドが指定されている場合）、このオプションは無視され、すべてのデータのバランスが調整されます。

レコード集計ノード

レコード集計は、データ セットのサイズを減らすために頻繁に用いられるデータ準備作業です。レコード集計を行う前に、データのクリーニングを、特に欠損値に注目して行う必要があります。レコード集計を実行すると、欠損値に関する潜在的な有益情報が失われてしまう可能性があります。詳細は、[6 章 欠損値の概要 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

レコード集計ノードを使用すると、一連の入力レコードを要約（集計された出力レコード）に置き換えることができます。たとえば、次のような入力レコードのセットがあるとします。

| Age（年齢） | Sex（性別） | Region（地域） | Branch | Sales（販売額） |
|---------|---------|------------|--------|------------|
| 23 | M | S | 8 | 4 |
| 45 | M | S | 16 | 4 |
| 37 | M | S | 8 | 5 |
| 30 | M | S | 5 | 7 |
| 44 | M | N | 4 | 9 |
| 25 | M | N | 2 | 11 |
| 29 | F | S | 16 | 6 |
| 41 | F | N | 4 | 8 |
| 23 | F | N | 6 | 2 |
| 45 | F | N | 4 | 5 |
| 33 | F | N | 6 | 10 |

上記のレコードを、Sex と Region をキー フィールドにして集計することができます。次に、[平均値] モードで Age フィールドを集計し、[合計] モードで Sales フィールドを集計します。[レコード集計ノード] ダイ

アログ ボックスで [フィールドにレコード度数を含める] を選択します。集計の出力結果は次のようになります。

| Age (平均年齢) | Sex (性別) | Region (地域) | Sales (販売額) | レコード件数 |
|------------|----------|-------------|-------------|--------|
| 35.5 | F | N | 25 | 4 |
| 29 | F | S | 6 | 1 |
| 34.5 | M | N | 20 | 2 |
| 33.75 | M | S | 20 | 4 |

たとえば、このことから、北部地域の 4 名の女性販売スタッフの平均年齢が 35.5 歳で、合計販売金額が 25 単位です。

注： 集計モードを指定しないと、Branch などのフィールドは自動的に無視されます。

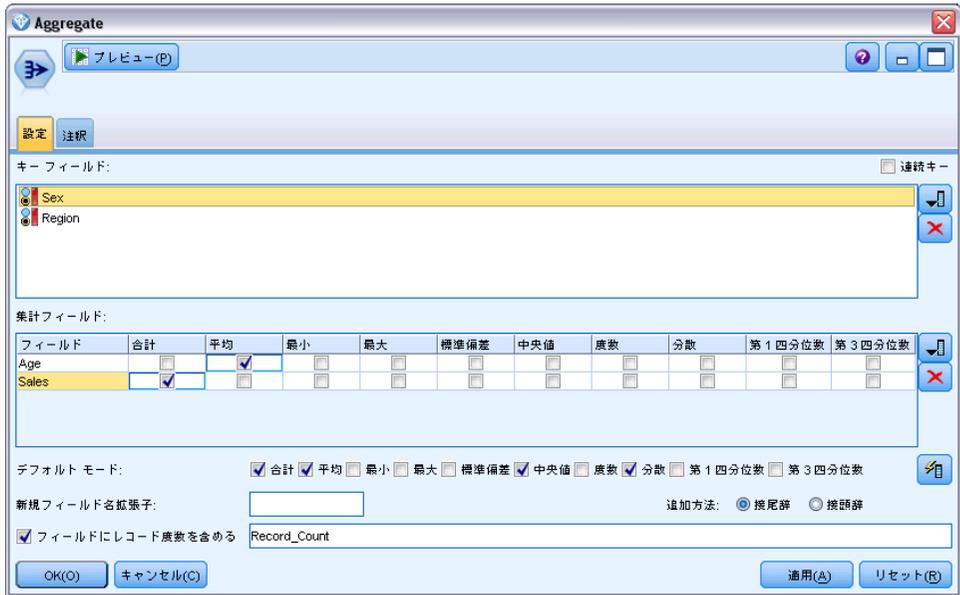
レコード集計ノードのオプション設定

[レコード集計] ノードで以下を指定します。

- レコード集計のカテゴリとして使用する 1 つまたは複数のキー フィールド。
- レコード集計値を計算する 1 つまたは複数の集計フィールド。
- 各レコード集計フィールドに出力する 1 つまたは複数のレコード集計モード (集計の種類)。

新しく追加されたフィールドに使用するデフォルトの集計モードを指定することもできます。

図 3-8
レコード集計ノードのダイアログ ボックス



キーフィールド：集計にカテゴリとして使用できるフィールドが一覧表示されます。連続型（数値型）フィールドとカテゴリフィールドの両方がキーとして使用できます。複数のフィールドを選択した場合は、値が組み合わされて、レコードを集計するためのキー値が生成されます。集計レコードは、それぞれ一意のキーフィールドに対して1つずつ生成されます。たとえば、キーフィールドがSexとRegionの場合、一意なMとFの、および地域NとSのそれぞれの組み合わせに対して集計レコードが作成されます（4つの一意な組み合わせ）。キーフィールドを追加するには、ウィンドウの右側にあるフィールドピッカーボタンをクリックします。

連続キー：同じキー値を持つすべてのレコードが入力にグループ化されている場合（たとえば、入力にキーフィールドにソートされる場合）、このオプションを選択します。このオプションを選択すると、パフォーマンスが向上します。

集計フィールド：選択されたレコード集計のモードのほか、集計される値のフィールドを表示します。リストにフィールドを追加するには、右側にあるフィールドピッカーボタンを使用します。利用できる集計関数を次に示します。

注：数値型以外のフィールドの適用できないフィールドがあります（たとえば、日付/時刻フィールドの[合計]）。選択した集計フィールドで使用できないモードが無効になります。

- **合計**：キー フィールドの各組み合わせの合計値を返す場合に選択します。合計は、欠損値のないすべてのケースに対する変数の値の合計です。
- **平均値**：キー フィールドの各組み合わせの平均値を返す場合に選択します。平均値は、中心傾向の尺度であり、算術平均です（ケース数で割った合計）。
- **最小**：キー フィールドの各組み合わせの最小値を返す場合に選択します。
- **最大**：キー フィールドの各組み合わせの最大値を返す場合に選択します。
- **標準偏差**：キー フィールドの各組み合わせの標準偏差を返す場合に選択します。標準偏差平均の周りの散らばり度です。変動測定の平方根に等しくなります。
- **中央値**：キー フィールドの各組み合わせの中央値を返す場合に選択します。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。50 番目のパーセンタイルまたは 2 番目の四分位でもあります。
- **度数**：キー フィールドの各組み合わせの null 値以外の度数を返す場合に選択します。
- **分散**：キー フィールドの各組み合わせの分散値を返す場合に選択します。分散は、平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。
- **第一四分位**：キー フィールドの各組み合わせの第一四分位（25 番目のパーセンタイル）を返す場合に選択します。
- **第三四分位**：キー フィールドの各組み合わせの第三四分位（75 番目のパーセンタイル）を返す場合に選択します。

デフォルト モード：新しく追加したフィールドに対して、デフォルトで使用する集計モードを指定します。同じ集計モードを頻繁に使用しているような場合は、ここでそれらのモードを選択し、右側にある [すべてに適用] ボタンをクリックすると、選択したモードが上記のリストに表示されているすべてのフィールドに適用されます。

新規フィールド名拡張子：重複する集計フィールドに対して、「1」や「new」などの接頭辞や接尾辞を追加する場合に選択します。たとえば、接尾辞オプションを選択して「1」を指定すると、Age フィールドに対する最小値の集計の結果、フィールド名 Age_Min_1 が追加されます。注：_Min や Max_ などの集計拡張子は実行された集計の種類を表し、自動的に新規フィールドに追加されます。付ける拡張子の種類に応じて、[接尾辞] または [接頭辞] を選択してください。

フィールドにレコード度数を含める：デフォルトで、各出力レコードに Record_Count という追加フィールドを含める場合に選択します。このフィールドは、各集計レコードを作成するために集計された入力レコード

数を示します。このフィールド名を自分で指定するには、その名前を編集フィールドに入力してください。

注： 集計の実行時にシステムのヌル値は除外されますが、レコードの度数には含まれます。一方、空白値は、集計レコードとレコード度数の両方に含まれます。空白値を除外するには、置換ノードを使用して空白値をヌル値に置き換えます。また、条件抽出ノードを使用して、空白値を削除することもできます。

パフォーマンス

集計操作は、並行処理を有効にすると、有利になる可能性があります。 [詳細は、5 章 ストリームの最適化オプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

RFM レコード集計ノード

リーセンシ、フリクエンシ、マネタリー (RFM) のレコード集計ノードを使用すると、一意の顧客 ID をキーとして使用し、顧客の過去のトランザクション データを取得、未使用のデータを削除、残りのトランザクション データをすべて単一行に結合することができます。これにより、最後のトランザクションの時期 (リーセンシ)、トランザクション数 (フリクエンシ)、これらの取引の合計金額 (マネタリー) が一覧表示されます。

レコード集計を行う前には、データのクリーニングを、特に欠損値に注目して行う必要があります。 [詳細は、6 章 欠損値の概要 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

RFM レコード集計ノードを使用してデータを識別および変換すると、RFM 分析ノードを使用してより詳細な分析を実行することができます。 [詳細は、4 章 p.225 RFM 分析ノード](#) を参照してください。

データ ファイルが RFM レコード集計ノードによって実行されると、ファイルには対象値が含まれません。そのため、データを C5.0 または CHAID などのモデル作成ノードによる詳細な分析の入力として使用する前に、(たとえば顧客 ID と一致させることによって) 他の顧客のデータと結合する必要があります。 [詳細は、p.98 レコード結合ノード](#) を参照してください。

IBM® SPSS® Modeler の RFM レコード集計ノードおよび RFM 分析ノードを設定して独立した分割を使用します。最新性、頻度、金額値の各尺度のデータを、これらの値および尺度に関係なくランク付けし、分割します。

RFM レコード集計ノードのオプション設定

図 3-9
RFM レコード集計設定

最新の相対値を計算: 最新の購買日が計算される日付を指定します。入力する 固定日付またはシステムで設定された 今日の日付のいずれかです。今日の日付はデフォルトで入力され、ノードが実行されると自動的に更新されます。

連続する ID: データ ストリーム中で同じ ID を持つすべてのレコードが一緒に表示されるようにデータをソートしている場合、このオプションを選択すると処理を高速化することができます。データがあらかじめソートされていない場合（またはわからない場合）は、このオプションは選択しないでください。この場合、ノードが自動的にデータをソートします。

ID: 顧客およびトランザクションを識別するために使用するフィールドを指定します。選択できるフィールドを表示するには、右側のフィールドピッカー ボタンを使用します。

日付: リーセンシを計算するために使用される日付フィールドを選択します。選択できるフィールドを表示するには、右側のフィールド ピッカー ボタンを使用します。

入力として使用するには、適切なフォーマットの日付、時間、またはタイムスタンプのストレージのフィールドが必要です。たとえば、Jan 2007、Feb 2007 などのような値を持つ文字列フィールドがある場合、置換ノードおよび `to_date()` 関数を使用してこれらのフィールドを日付フィールドに変換することができます。詳細は、4 章 p.200 [置換ノードを使ったストレージの変換](#) を参照してください。

値: 顧客のトランザクションの全体の金額を計算するために使用するフィールドを選択します。選択できるフィールドを表示するには、右側のフィールド ピッカー ボタンを使用します。注：これは、数値である必要があります。

新規フィールド名拡張子: 「12_month」などの接尾辞または接頭辞を、新しく生成されたリーセンシ、フリクエンシ、マネタリー フィールドに適用します。付ける拡張子の種類に応じて、[\[接尾辞\]](#) または [\[接頭辞\]](#) を選択してください。たとえば、複数の期間を調べる場合に役立ちます。

以下の値を持つレコードを破棄: 必要な場合、RFM の合計を計算する場合に使用されないトランザクションの詳細の最小値を指定することができます。値の単位は、選択された [\[値\]](#) フィールドに関連します。

最近のトランザクションのみを含める: 大規模なデータベースを分析する場合、最新のレコードのみが使用されるよう指定することができます。次のように特定の日付または最新の期間内のいずれかに記録された日付を選択できます。

- **トランザクション データ:** データが分析に含まれた後のトランザクションの日付を指定します。
- **最後のトランザクション:** レコードが分析に含まれる後の [\[最新の相対値を計算\]](#) の日付からさかのぼった期間の数および種類（日、週、月または年数）を指定します。

2 番目に最近のトランザクションの日付を保存: 各顧客の 2 番目に最近のトランザクションの日付を知りたい場合は、このオプションを選択します。また、[\[3 番目に最近のトランザクションの日付を保存\]](#) チェック ボックスも選択できます。たとえば、かなり前に多くの取引を行っているが最近では 1 回の取引しかない顧客を識別するために役に立ちます。

ソート ノード

ソート ノードを使用すれば、レコードを 1 つまたは複数のフィールド値に基づいて昇順または降順にソートすることができます。ソート ノードは、たとえば最も一般的なデータ値のレコードを表示および選択するために頻繁に使用されます。一般的に、まずレコード集計ノードを使用して

データを集計し、次にソート ノードを使用して集計データをレコード度数の降順にソートします。この結果をテーブルに表示してデータを探索し、上位 10 位までの顧客のレコードを選択するような決定を下します。

図 3-10
ソートノードのダイアログ ボックス



ソート項目： ソート キーとして使用するすべてのフィールドがテーブル内に表示されます。数値のキー フィールドが、ソートに最も適しています。

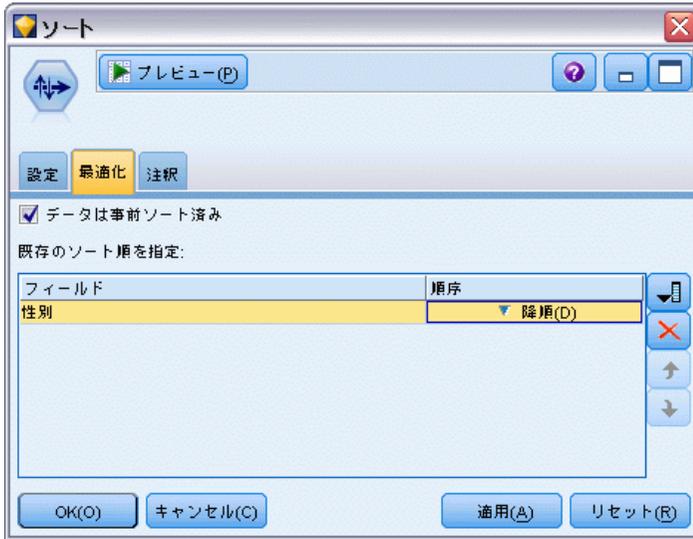
- このリストに**フィールドを追加する**には、右側にあるフィールド ピッカー ボタンを使用します。
- **並び順を選択する**には、テーブルの [順序] 列にある [昇順] または [降順] の矢印をクリックします。
- **フィールドを削除する**には、赤い削除ボタンをクリックします。
- **式をソートする**には、赤い上矢印および下矢印ボタンを使用します。

デフォルトのソート順： 新規フィールドが追加された場合にデフォルトで使用するソート順序を、[昇順] または [降順] で選択します。

ソートの最適化設定

すでにいくつかのキー フィールドでソートされているデータを処理している場合は、残りのデータをシステムが効率よくソートできるように、すでにソートされたフィールドを指定できます。たとえば、Age (降順) と Drug (昇順) でソートするときに、データがすでに Age (降順) でソート済みであることがわかっているとします。

図 3-11
最適化の設定



データは事前ソート済み： データがすでに 1 つ以上のフィールドでソート済みかどうかを指定します。

既存のソート順を指定： すでにソートされているフィールドを指定します。[フィールドの選択] ダイアログ ボックスを使用して、フィールドをリストに追加します。[順序] 列で、各フィールドが昇順でソートされているか、または降順でソートされているかを指定します。複数のフィールドを選択する場合は、正しいソート順で一覧になっていることを確認してください。リストの右にある矢印を使用して、フィールドを正しい順序で配置します。正しい既存のソート順の指定に誤りがあると、ストリームを実行するときに、ソートが指定どおりではないレコード番号を示した、エラーが表示されます。

注： ソート速度は、並行処理を有効にすると、有利になる可能性があります。詳細は、5 章 [ストリームの最適化オプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

レコード結合ノード

レコード結合ノードでは、複数の入力レコードから、入力フィールドの全部または一部を含む 1 つの出力レコードが作成されます。この機能は、内部顧客データと購入人口データのような、異なるソースからのデータを結合する場合に役立ちます。以下の方法で、データを結合できます。

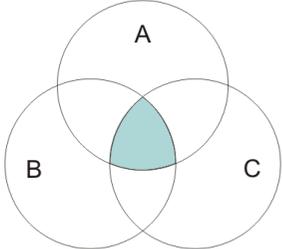
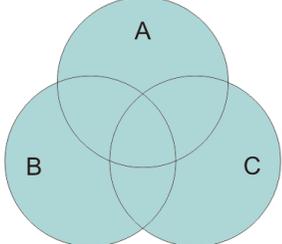
- **順序による結合**では、最も小さいデータ ソース中のデータがなくなるまで、入力順序にしたがってすべてのソースから対応するレコードを結合します。ソート ノードを使ってデータをソートしている場合に、このオプションが重要になります。
- **キー フィールドを使った結合**では、Customer ID (顧客 ID) などのキー フィールドを使用して、あるデータ ソース中のレコードと他のデータ ソース中のレコードとの結合方法を指定します。内部結合、完全外部結合、部分外部結合、および逆結合など、さまざまな種類の結合を利用できます。 [詳細は、 p.99 結合の種類 を参照してください。](#)
- **条件による結合**では、結合を行うために満たす必要のある条件を指定できます。ノードで直接条件を指定するか、Clem 式ビルダーを使用して条件を作成できます。

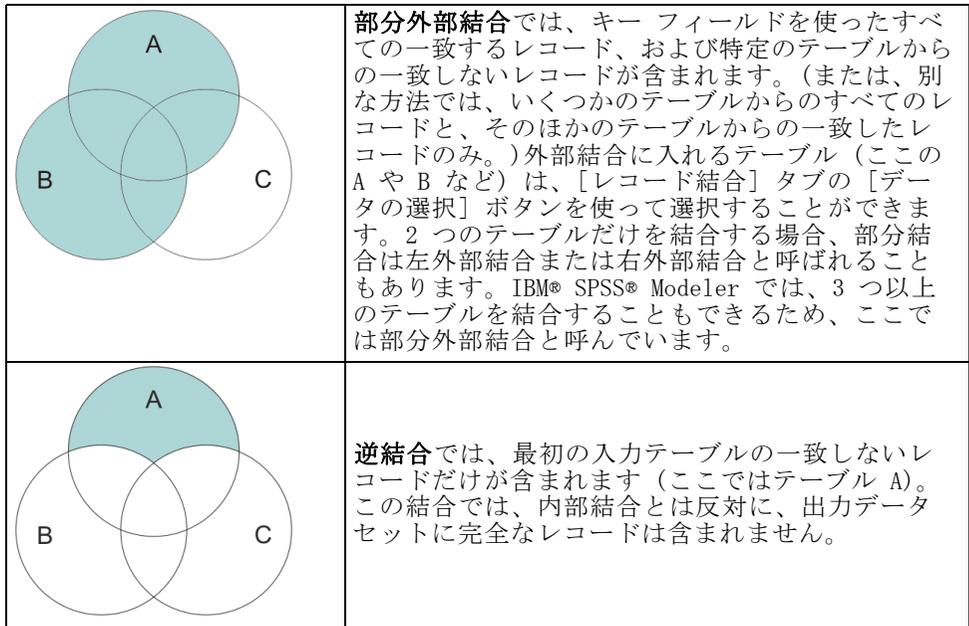
結合の種類

データ結合にキー フィールドを使用する場合、まずどのレコードを除外して、どのレコードを対象にするかを検討することをお勧めします。後述するように、さまざまな種類の結合手段があります。

基本的な結合の種類としては、内部結合と外部結合があります。これらの方法は、Customer ID などのキー フィールドの共通する値に基づいて、関連データ セットからテーブルを併合するために頻繁に使用されます。内部結合によって、制限のない併合が行われ、完全なレコードのみが含まれるデータ セットが出力されます。外部結合の場合も結合データからの完全なレコードが含まれますが、それ以外に 1 つまたは複数の入力テーブルから固有のデータを入れることもできます。

利用できる各種の結合手段の詳細は、後述します。

| | |
|---|---|
|  | <p>内部結合では、キー フィールドの値がすべての入力テーブルで共通のレコードだけが含まれます。つまり、出力データ セットには、一致しないレコードは含まれません。</p> |
|  | <p>完全外部結合では、入力テーブルからの一致するレコードと一致しないレコードの両方のレコード (すべてのレコード) が含まれます。左外部結合および右外部結合は、部分外部結合と呼ばれています。</p> |



たとえば、あるデータ セット内の農場についての情報があり、農業関連の保険金請求が別のデータ セットにある場合に、[レコード結合] オプションを使用して、最初のソースのレコードを 2 番目のソースに照合できます。

農場サンプル内の顧客が保険金請求をファイリングしているかどうかを判断するには、内部結合オプションを使用して、すべての ID が 2 つのサンプルで一致する箇所を示すリストを返します。

図 3-12
内部結合のサンプル出力

| | id | name | region | farmsize | rainfall | landquality | farmincome | maincrop | claimtype | claimvalue |
|---|-------|---------|-----------|----------|----------|-------------|------------|----------|-----------|-------------|
| 1 | id604 | name604 | southwest | 1860.000 | 103.0... | 3.000 | 625251.000 | potatoes | decomm... | 281082.0... |
| 2 | id605 | name605 | north | 1700.000 | 46.000 | 8.000 | 621148.000 | wheat | decomm... | 122006.0... |
| 3 | id620 | name620 | north | 880.000 | 74.000 | 6.000 | 426988.000 | rapeseed | arable_de | 118885.0... |

完全外部結合オプションを使用すると、入力テーブルから一致するレコードと一致しないレコードの両方が返されます。システム欠損値 (\$null\$) が、不完全な値に対して使用されます。

図 3-13
完全外部結合のサンプル出力

| | id | name | region | farmsize | rainfall | landquality | farmincome | maincrop | claimtype | claimvalue |
|---|-------|----------|-----------|----------|----------|-------------|------------|----------|-----------|------------|
| 1 | id601 | \$null\$ | \$null\$ | \$null\$ | \$null\$ | \$null\$ | \$null\$ | \$null\$ | decomm... | 74703.1C |
| 2 | id602 | name602 | north | 1780.000 | 42.000 | 9.000 | 734118.000 | maize | \$null\$ | \$nul |
| 3 | id604 | name604 | southwest | 1860.000 | 103.0... | 3.000 | 625251.000 | potatoes | decomm... | 281082.0 |
| 4 | id605 | name605 | north | 1700.000 | 46.000 | 8.000 | 621148.000 | wheat | decomm... | 122006.0 |
| 5 | id606 | \$null\$ | \$null\$ | \$null\$ | \$null\$ | \$null\$ | \$null\$ | \$null\$ | arable_de | 122135.0 |

部分外部結合では、指定されたテーブルから一致しないレコードと同様に、キー フィールドを使用して一致したすべてのレコードが含まれます。テーブルには、最初のデータ セットから一致したレコードと同様に、ID フィールドから一致したすべてのレコードが表示されます。

図 3-14
部分外部結合のサンプル出力

| | id | claimtype | claimvalue | name | region | farmsize | rainfall | landquality | farmincome | maincrop |
|---|-------|-----------|-------------|---------|-----------|----------|----------|-------------|--------------|----------|
| 1 | id602 | \$null\$ | \$null\$ | name602 | north | 1780.000 | 42.000 | 9.000 | 734118.000 | maize |
| 2 | id604 | decomm... | 281082.0... | name604 | southwest | 1860.000 | 103.0... | 3.000 | 625251.000 | potatoes |
| 3 | id605 | decomm... | 122006.0... | name605 | north | 1700.000 | 46.000 | 8.000 | 621148.000 | wheat |
| 4 | id607 | \$null\$ | \$null\$ | name607 | southeast | 1820.000 | 29.000 | 6.000 | 211605.000 | maize |
| 5 | id608 | \$null\$ | \$null\$ | name608 | southeast | 1640.000 | 108.0... | 7.000 | 1167040.0... | maize |
| 6 | id609 | \$null\$ | \$null\$ | name609 | southwest | 1600.000 | 101.0... | 5.000 | 756755.000 | wheat |
| 7 | id615 | \$null\$ | \$null\$ | name615 | midlands | 920.000 | 86.000 | 6.000 | 442554.000 | potatoes |
| 8 | id618 | \$null\$ | \$null\$ | name618 | southeast | 1180.000 | 98.000 | 3.000 | 368646.000 | maize |

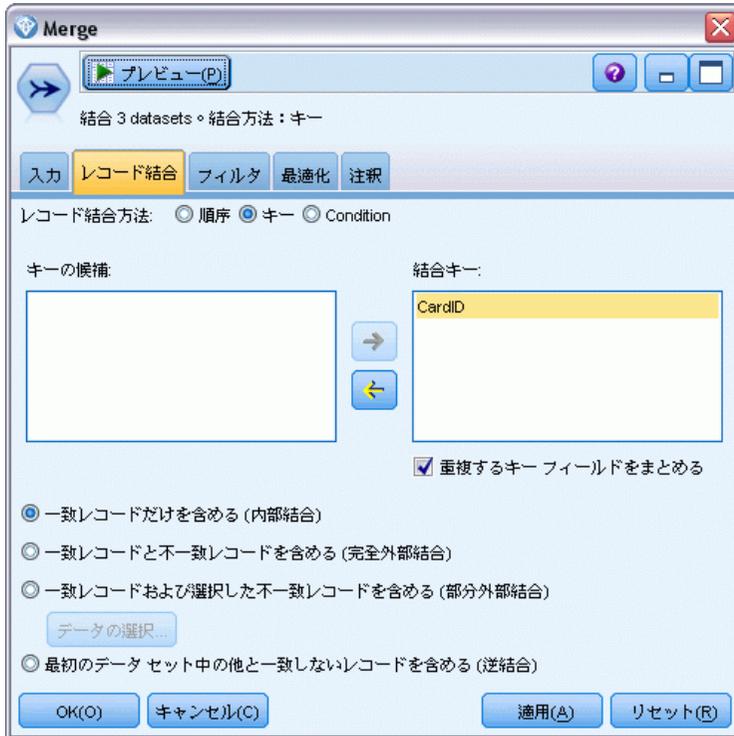
逆結合オプションを使用する場合は、最初の入力テーブルで一致しないレコードのみが返されます。

図 3-15
逆結合のサンプル出力

| | id | name | region | farmsize | rainfall | landquality | farmincome | maincrop |
|---|-------|---------|-----------|----------|----------|-------------|--------------|----------|
| 1 | id602 | name602 | north | 1780.000 | 42.000 | 9.000 | 734118.000 | maize |
| 2 | id607 | name607 | southeast | 1820.000 | 29.000 | 6.000 | 211605.000 | maize |
| 3 | id608 | name608 | southeast | 1640.000 | 108.0... | 7.000 | 1167040.0... | maize |
| 4 | id609 | name609 | southwest | 1600.000 | 101.0... | 5.000 | 756755.000 | wheat |
| 5 | id615 | name615 | midlands | 920.000 | 86.000 | 6.000 | 442554.000 | potatoes |
| 6 | id618 | name618 | southeast | 1180.000 | 98.000 | 3.000 | 368646.000 | maize |
| 7 | id619 | name619 | north | 840.000 | 64.000 | 8.000 | 457552.000 | potatoes |

結合方法とキーの指定

図 3-16
[レコード結合] タブを使った結合方法の設定



レコード結合方法 : [順序] または [キー] を選択して、レコードの結合方法を指定します。[キー] を選択すると、ダイアログ ボックスの下半分が有効になります。

- **順序** : 各入力からの n 番目のレコードを結合して、n 番目の出力レコードを生成するような方法で、レコードを結合します。一致する入力レコードがなくなると、それ以上の出力レコードは生成されません。つまり、作成されるレコード数は一番小さいデータ セットのレコード数と等しくなります。
- **キー** : Transaction ID などのキー フィールドに同じ値を持つレコードが結合されます。これは、データベースの「等結合」と同じ処理です。キーの値が複数あるような場合は、可能なすべての組み合わせが返されます。たとえば、同じキーフィールドの値 A を持つ複数のレコードが、それぞれ別のフィールドでは値 B、C、および D を持つ場合、結合後のフィールドでは A と B、A と C、および A と D のように、個別のレコードを組み合わせた結果が生成されます。

注：キーによるレコード結合では、ヌル値は同一とみなされず、結合されません。

- **条件**：結合の条件を指定します。詳細は、[p. 104 結合の条件の指定](#)を参照してください。

キーの候補：すべての入力データソースのフィールド名に完全に一致するフィールドだけが表示されます。このリストからフィールドを選択して矢印ボタンをクリックすると、レコードを結合するためのキーフィールドにそのフィールドが追加されます。複数のキーフィールドを使用することもできます。フィルタノードまたは入力ノードの [フィルタ] タブを使用して、一致しない入力フィールドの名前を変更できます。

結合キー：すべての入力データソースから、キーフィールドの値に基づいたレコードの結合に使用するすべてのフィールドが表示されます。リストからキーを削除するには、該当するキーを選択して矢印ボタンをクリックし、そのキーを [キーの候補] リストに戻します。複数のキーフィールドが選択されている場合は、次のオプションが有効になります。

重複するキーフィールドをまとめる：上で複数のキーフィールドが選択されている場合に、その名前を持つ出力フィールドを 1 つだけにします。以前のバージョンの IBM® SPSS® Modeler からストリームをインポートした場合を除いて、このオプションはデフォルトで有効になります。このオプションの選択を解除した場合、[レコード結合ノード] ダイアログボックスの [フィルタ] タブを使って、重複するキーフィールドの名前を変更するか、または除外する必要があります。

一致レコードだけを含める (内部結合)：完全なレコードだけを結合する場合に選択します。

一致レコードと不一致レコードを含める (完全外部結合)：「完全外部結合」を実行する場合に選択します。キーフィールドの値がすべての入力テーブルに存在しない場合、不完全なレコードはそのまま保持されます。未定義値 (\$null\$) は、キーフィールドに追加され、出力レコードに含まれます。

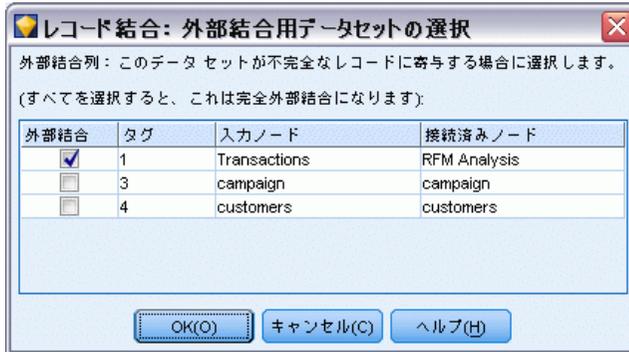
一致レコードおよび選択した不一致レコードを含める (部分外部結合)：サブダイアログボックスで選択したテーブルに対して、「部分外部結合」を実施する場合に選択します。不完全なレコードを保持するテーブルを指定するには、[データの選択] をクリックします。

最初のデータセット中の他と一致しないレコードを含める (逆結合)：最初のデータセットの不一致レコードだけを下流に渡す、逆結合を実施する場合に選択します。[入力] タブの矢印を使って、入力データセットの順序を指定することができます。この結合では、出力データセットに完全なレコードは含まれません。詳細は、[p. 99 結合の種類](#)を参照してください。

部分結合のデータの選択

部分外部結合の場合、不完全なレコードを保持するテーブルを選択する必要があります。たとえば、顧客テーブルからのすべてのレコードを保持しながら、住宅ローン テーブル中の一致するレコードだけを保持することができます。

図 3-17
部分結合または外部結合用のデータの選択

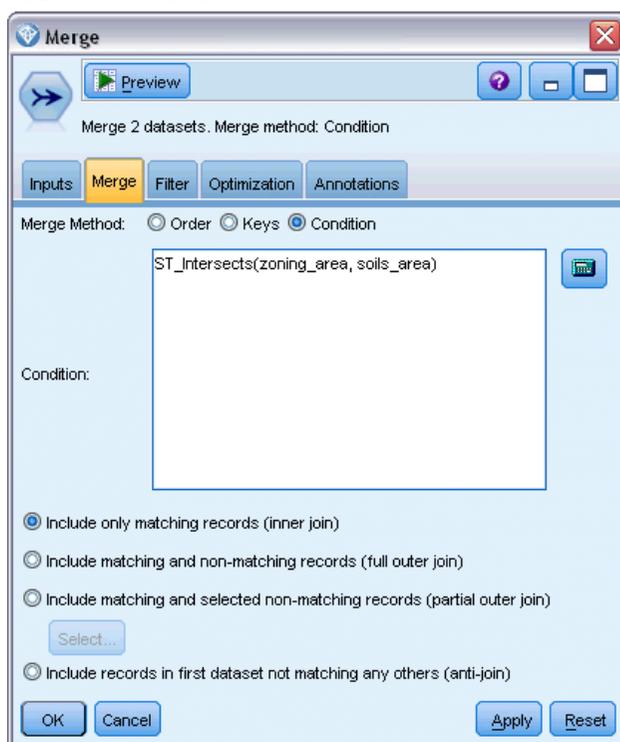


[外部結合] 列:[外部結合] 列では、そのまま全体を含めるデータ セットを選択します。部分結合の場合、ここで選択したデータ セットに対して、重複レコードや不完全レコードが保持されます。 [詳細は、 p. 99 結合の種類 を参照してください。](#)

結合の条件の指定

結合の方法を **[条件]** に設定して、結合を行うために満たす必要のある条件を指定できます。

図 3-18
結合の条件の設定



[条件] フィールドに条件を直接入力、またはフィールドの右側の計算機の記号をクリックして C1em 式ビルダーを使用し、条件を作成することができます。詳細は、7 章 [C1em 式ビルダー in IBM SPSS Modeler 15 ユーザーガイド](#) を参照してください。

レコード結合ノードからのフィールドのフィルタリング

レコード結合ノードでは、複数のデータソースを結合した結果として発生する重複フィールドをフィルタリングしたり、名前を変更することができます。フィルタリングオプションを選択するには、ダイアログボックスの [フィルタ] タブをクリックします。

図 3-19
レコード結合ノードでのフィルタリング



このオプションは、フィルタ ノードのオプションとほとんど変わりありません。ただし、ここに記載されていなくても、[フィルタ]メニューで利用できるオプションもあります。詳細は、4 章 p. 173 フィールドのフィルタリングまたは名前の変更を参照してください。

フィールド: 現在接続しているデータ ソースからの入力フィールドを表示します。

タグ: データ ソース リンクに関連付けられたタグ名（または数字）が表示されます。このレコード結合ノードへのアクティブなリンクを変更するには、[入力] タブをクリックします。

入力ノード: データを結合する入力ノードを表示します。

接続済みノード: レコード結合ノードに接続しているノードのノード名を表示します。複雑なデータ マイニング作業では、しばしば複数の結合または追加操作が行われ、同じ入力ノードが含まれることがあります。接続されたノード名を表示することによって、この問題を解消することができます。

フィルタ: 入力フィールドと出力フィールド間の現在の接続が表示されます。アクティブな接続は、正常な矢印で表示されます。赤い X が付けられている接続は、そのフィールドがフィルタリングされていることを表します。

フィールド :結合または追加の後に、出力フィールドを表示します。重複するフィールド名は赤で表示されます。重複するフィールドを無効にするには、上の [フィルタ] フィールドをクリックします。

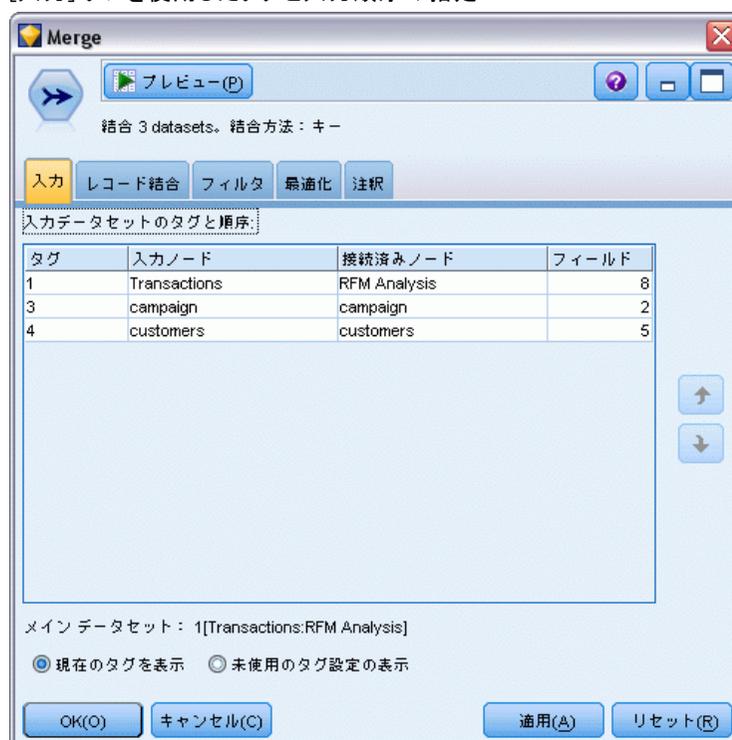
現在のフィールドを表示 : キー フィールドとして使用されているフィールドの情報を表示します。

未使用のフィールド設定を表示 :現在使用されていないフィールドの情報を表示します。

入力順序とタグの設定

[レコード結合ノード] ダイアログ ボックスまたは [レコード追加ノード] ダイアログ ボックスの [入力] タブを使用して、データ ソースの入力順序を指定したり、各ソースのタグ名を変更することができます。

図 3-20
[入力] タブを使用したタグと入力順序の指定



入力データセットのタグと順序： 完全なレコードだけを結合または追加する場合に選択します。

- **タグ：**各入力データ ソースの現在のタグ名を表示します。タグ名または**タグ**は、結合または追加操作のデータ リンクを一意に識別するための手段を提供します。たとえば、複数の水道管を流れている水が、ある場所で 1 本の水道管に合流するような状況を想像してください。IBM® SPSS® Modeler 内のデータも同じように流れていきます。また、合流点では、さまざまなデータ ソース間でしばしば複雑なやり取りが行われます。タグは、レコード結合ノードやレコード追加ノードで入力（先ほどの例の「水道管」にあたる）を管理し、ノードが保存されたり切断された場合でも、リンクをそのまま保持して簡単に識別できるようにするための手段を提供します。

レコード結合ノードまたはレコード追加ノードにデータ ソースを追加した場合、ノードの接続順序を示す数字を使用して、デフォルトのタグが自動的に作成されます。この順序は、入力または出力データ セット中のフィールドの順序とは関係ありません。デフォルトのタグを変更するには、[タグ] 列に新しい名前を入力します。

- **入力ノード：**データを結合する入力ノードを表示します。
- **接続済みノード：**レコード結合ノードまたはレコード追加ノードに接続しているノードのノード名を表示します。複雑なデータ マイニング作業では、しばしば複数の結合操作が行われ、同じ入力ノードが含まれることがあります。接続されたノード名を表示することによって、この問題を解消することができます。
- **フィールド：**各データ ソース中のフィールド数を表示します。

現在のタグを表示： レコード結合ノードまたはレコード追加ノードで現在使用されているタグを表示する場合に選択します。現在のタグは、データが流れているノードへのリンクを識別します。水道管にたとえると、現在のタグは現在水が流れ込んでいる水道管を示します。

未使用のタグ設定の表示：以前レコード結合ノードまたはレコード追加ノードへの接続に使用されたことがあり、現在データ ソースに接続されていないタグ、またはリンクを表示する場合に選択します。先ほどの例にたとえると、これは給水システム内にまだ現存しているけれども、水が流れていない水道管を表しています。これらの「水道管」を新しいソースに接続することも、削除することもできます。ノードから未使用のタグを削除するには、[消去] をクリックします。この操作を行うと、未使用のタグがすべて削除されます。

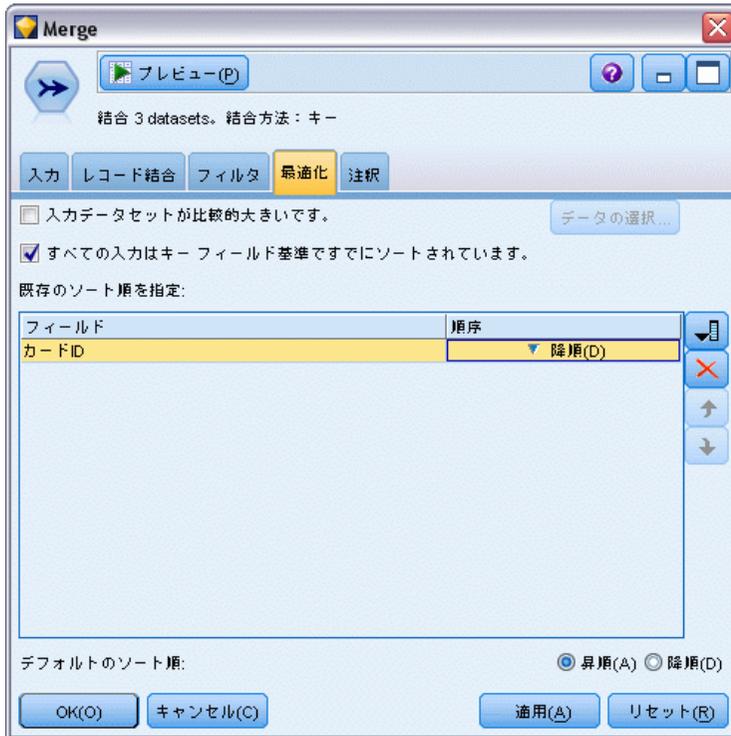
図 3-21
レコード結合ノードからの未使用のタグの削除



レコード結合の最適化設定

一定の状況でデータをより効率的にレコード結合できるように、システムには 2 つのオプションが用意されています。これらのオプションによって、1 つの入力データセットがほかのデータセットより著しく大きい場合や、データが、レコード結合に使用するキーフィールドのすべて、または一部によってすでにソート済みの場合に、レコード結合を最適化できるようになります。

図 3-22
最適化の設定



入力データセットが比較的大きい：入力データ セットの 1 つがほかのデータ セットよりもかなり大きいことを示すために選択します。システムにより小さいほうのデータ セットがメモリー内にキャッシュされ、その後、大きいデータ セットをキャッシングもソートもしないで処理することで、レコード結合が実行されます。通常、このような結合は、星状のスキーマや類似した設計を使用して設計されたデータとともに使用します。星状のスキーマには、たとえばトランザクション形式のデータ内のような、共用データの大きな中央テーブルがあります。このオプションを選択する場合は、[選択] をクリックして大きなデータ セットを指定します。ただ 1 つの大きなデータ セットのみを選択できるように注意してください。次の表に、この方法を使用して最適化できる結合を要約します。

| 結合の種類 | 大規模入力データ セット用に最適化できるか？ |
|-------|---------------------------------|
| Inner | はい |
| 部分 | 大規模データ セット内に不完全レコードがない場合に、「はい」。 |
| Full | いいえ |
| 逆結合 | 大規模データ セットが最初の入力の場合に、「はい」。 |

すべての入力はキー フィールド基準ですでにソートされている： 入力データが、レコード結合に使用するキー フィールドのうちの 1 つ以上ですでにソートされていることを示すために選択します。入力データ セットがすべてソート済みであることを確認します。

既存のソート順を指定： すでにソートされているフィールドを指定します。[フィールドの選択] ダイアログ ボックスを使用して、フィールドをリストに追加します。レコード結合に使用されている ([レコード結合] タブで指定された) キー フィールドの中からのみ、選択できます。[順序] 列で、各フィールドが昇順でソートされているか、または降順でソートされているかを指定します。複数のフィールドを選択する場合は、正しいソート順で一覧になっていることを確認してください。リストの右にある矢印を使用して、フィールドを正しい順序で配置します。正しい既存のソート順の指定に誤りがあると、ストリームを実行するときに、ソートが指定どおりではないレコード番号を示した、エラーが表示されます。

データベースが使用する照合方式の大文字と小文字の区別によっては、入力がデータベースによってソートされている場合最適化が正しく機能しないことがあります。たとえば、一方の入力が大文字と小文字を識別し、もう一方が大文字と小文字を区別しない 2 つの入力がある場合、ソートの結果が異なる可能性があります。最適化を結合した場合、レコードがソートされた順で処理される場合があります。その結果、異なる照会方法で入力がソートされた場合、結合ノードはエラーを報告し、ソートが一貫しないレコード番号を表示します。すべての入力が 1 つのソースである場合、または相互に包括的な照合によってソートされている場合、レコードは正常に結合できます。

注：レコード結合の速度は、並行処理を有効にすると、有利になる可能性があります。詳細は、[5 章 ストリームの最適化オプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

レコード追加ノード

レコード追加ノードを使用すると、一連のレコードを連結することができます。異なるソースからのレコードを結合するレコード結合ノードとは異なり、レコード追加ノードでは、1 つのソースからレコードがなくなるまですべてのレコードが読み込まれて下流に渡されます。次に、最初の入力または主入力と同じデータ構造 (レコード数やフィールド数など) を使用して、次のソースからレコードが読み込まれます。主ソースのフィールド数が他の入力ソースのフィールド数より多い場合は、不完全な値に対してはシステム ヌル文字列 (\$null\$) が使用されます。

レコード追加ノードは、構造が似ていてもデータが異なるデータ セットを組み合わせる場合に役立ちます。たとえば、3 月の売り上げデータ ファイルと 4 月の売り上げデータ ファイルのように、期間の異なるファイルにトランザクション データが保存されている場合を想定します。これ

らのファイルの構造は同じである（同じフィールドが同じ順序で並んでいる）と仮定すれば、レコード追加ノードを使用して両方のファイルを1つの大きなファイルに結合することができます。このファイルを、後ほど分析に利用します。

注： ファイルを追加するには、フィールドの尺度が同じでなければなりません。たとえば、名義型フィールドを尺度が連続型のフィールドの追加することができません。

図 3-23
フィールドの一致を名前で表示した [レコード追加ノード] ダイアログ ボックス



追加オプションの設定

フィールド一致基準： 追加するフィールドを一致させるために使用する
方法を選択します。

- **位置：** メイン データ ソース中のフィールドの位置を基準にしてデータ セットを追加する場合に選択します。この方法を使用する場合、正しく追加を行うためには、あらかじめデータをソートしておく必要があります。
- **名前：** 入力データ セット中のフィールド名を基準にしてデータ セットを追加する場合に選択します。また、フィールド名の一致に際して大文字と小文字を区別する場合は、[大文字と小文字を区別] を選択します。

出力フィールド：レコード追加ノードに接続されている入力ノードが一覧表示されます。リストの最初のノードが主入力ソースになります。表示されているフィールドをソートするには、列見出しをクリックしてください。ソートを行っても、データ セット中のフィールドの並びは変わりません。

フィールド入力元：メイン データ セット中のフィールドを基準にして出力フィールドを生成する場合は、[メイン データ セットのみ] を選択します。メイン データ セットは、[入力] タブで指定されている主入力のことです。すべての入力データ セットに渡って一致フィールドがあるかどうかに関係なく、すべてのデータ セット中のすべてのフィールドに対して、出力フィールドを選択する場合は、[すべてのデータ セット] を選択します。

フィールドに入力データセットを入れてレコードにタグ付け：各レコードのソースデータ セットを示す値を持つフィールドを、出力ファイルに追加する場合に選択します。テキスト フィールドに名前を指定します。デフォルトのフィールド名は、入力 です。

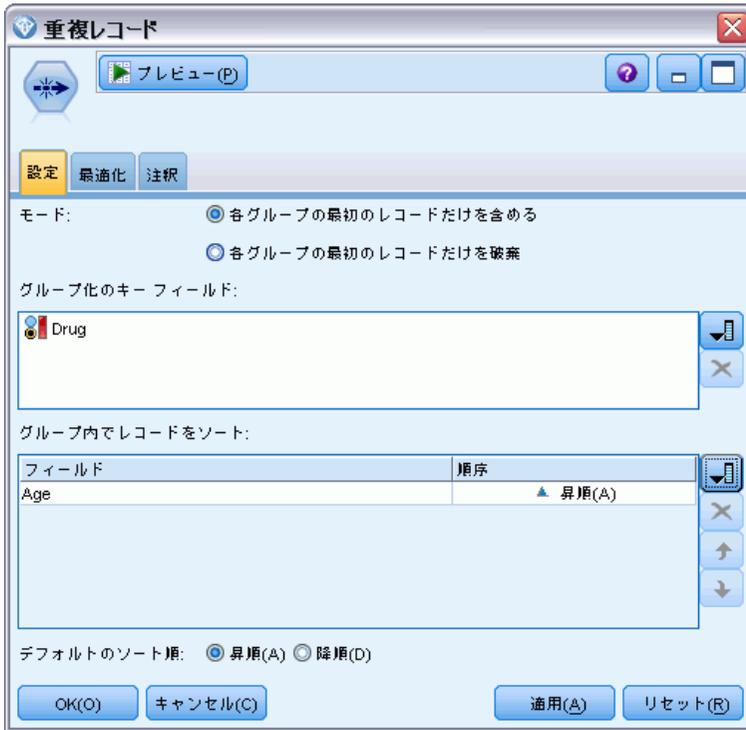
重複レコード ノード

データセット内の重複レコードは、データ マイニングを開始する前に削除する必要があります。たとえば、マーケティング データベースで、異なる住所または会社情報を持つ個人が複数回出現する場合があります。重複レコード ノードを使用して、データ セットの重複レコードを検索または削除できます。

重複レコード ノードを使用して、最初のレコードをデータ ストリームに渡して重複レコードを削除したり、最初のレコードを破棄して重複レコードをデータ ストリームに渡すことによって重複レコードを検索することもできます。

また、返された結果のキー値のソート順を定義できます。特定の行を各キーに返したい場合、上流でソート ノードを使用するのではなく、重複レコードノードでレコードを並べ替える必要があります。

図 3-24
重複レコード ノードのダイアログ ボックス



モード: 最初のレコードを含めるか除外（破棄）するかを指定します。

- **各グループの最初のレコードだけを含める:** データ ストリームの最初のレコードを含め、重複レコードを削除します。
- **各グループの最初のレコードだけを破棄:** 見つかった最初の重複レコードを破棄して、その後の同一レコードをすべてデータ ストリームに渡します。このオプションは、データ内の重複を検出しておき、後からストリームを調べる場合に役立ちます。

グループ化のキー フィールド: レコードが同一であるかどうかを判断するために使われるフィールドを表示します。以下を行うことができます。

- このリストにフィールドを追加するには、右側にあるフィールド ピッカー ボタンを使用します。
- フィールドを削除するには、赤い X（削除） ボタンをクリックします。

グループ内でレコードをソート: 各キー値でレコードをどのように並べ替えるか、昇順で並べ替えるか、降順で並べ替えるかを決定するために使用するフィールドを表示します。以下を行うことができます。

- このリストにフィールドを追加するには、右側にあるフィールド ピッカー ボタンを使用します。

- フィールドを削除するには、赤い X (削除) ボタンをクリックします。
- フィールドを移動するには、上下 2 つのボタンがあり、複数のフィールドごとに並べ替える場合に使用します。

デフォルトのソート順： デフォルトで、昇順または降順のどちらかで並べ替えるかを指定します。

重複レコード ノード内のレコードを並べ替え

重複レコード ノード内で [グループ内でレコードをソート] オプションを使用すると、各キーに特定の行を返すことができます。先行したソート ノードを使用する必要はありません。たとえば、処方薬使用者の年齢について、次のデータがあるとします。

| Age (年齢) | Drug (薬品) |
|----------|-----------|
| 50 | Drug A |
| 71 | Drug B |
| 44 | Drug A |
| 65 | Drug X |
| 39 | Drug A |
| 75 | Drug C |
| 72 | Drug Y |
| 57 | Drug X |
| 79 | Drug Y |
| 69 | Drug C |
| 74 | Drug B |
| 85 | Drug Y |
| 69 | Drug X |

各グループの最初のレコードだけを含めるモードを設定し、薬品の最も古いユーザーを検出するには、キーフィールドに [薬品] を使用し、ソート フィールドに [年齢] を使用して、降順に設定します。ソートの選択内容は、指定された [薬品] の多くの行のうち、どれが返されるかと指定するため、入力順は結果に影響を与えません。最終的なデータ出力は次のようになります。

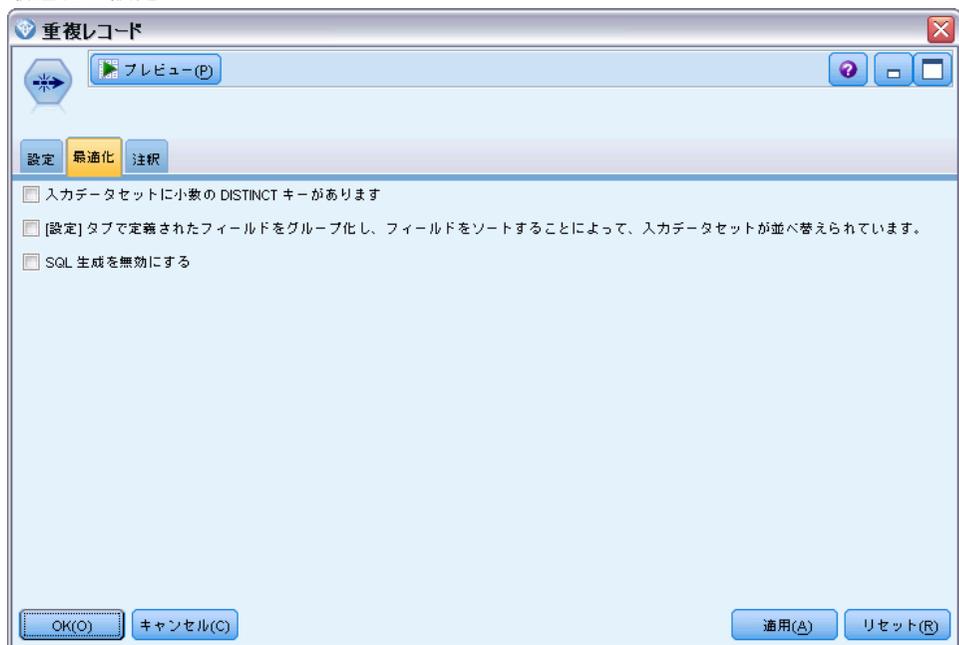
| Age (年齢) | Drug (薬品) |
|----------|-----------|
| 50 | Drug A |
| 74 | Drug B |
| 75 | Drug C |
| 69 | Drug X |
| 85 | Drug Y |

重複レコード最適化設定

処理しているデータのレコード数が少ない場合、またはすでにソートされている場合、IBM® SPSS® Modeler がデータをより効率的に処理できるようにする方法を最適化できます。

注：[入力データの重複レコード キーが少ない] を選択するか、ノードに SQL 生成を使用する場合、異なるキー値内の行を返すことができます。異なるキー内で返される行を制御するには、[設定] タブで [グループ内でレコードをソート] を使用して、ソート順を指定する必要があります。最適化オプションは、[設定] タブでソート順を指定している限り、重複レコード ノードによって出力された結果には影響を与えません。

図 3-25
最適化の設定



入力データの重複レコード キーが少ない： キー フィールドのレコードが少ないまたは一意の値が少ない場合に指定します。このオプションを選択すると、パフォーマンスが向上します。

入力データセットが [設定] タブでフィールドをグループ化およびソートして並べ替えられる： [設定] タブの [グループ内でレコードをソート] に表示されたすべてのフィールドでデータがすでにソートされている場合、データの降順または昇順が同じ場合にのみこのオプションを選択します。このオプションを選択すると、パフォーマンスが向上します。

SQL 生成を無効にする： ノードの SQL 生成を無効にする場合に選択します。

フィールド設定ノード

フィールド設定の概要

初期データを探索した後に、解析の準備として、データの選択、クリーニング、または構成が必要になるでしょう。[フィールド設定] パレットには、このデータ変換と準備作業に役立つさまざまなノードが用意されています。

たとえば、フィールド作成ノードを使用して、現在データには含まれていない属性を新しく作成できます。または、たとえばデータ分割ノードを使用して、対象の分析用に自動的にフィールド値を再コード化することができます。データ型ノードは頻繁に使用され、測定レベル、値、モデリングの役割をデータ セット内の各フィールドに割り当てることができます。この操作は、欠損値の処理と下流のモデル作成において役に立ちます。

[フィールド設定] パレットには次のようなノードがあります。



自動データ準備 (ADP) ノードでは、データ分析、固定値の識別、問題のあるまたは役に立たない可能性のあるフィールドのスクリーニング、必要に応じた新しい属性の取得、詳細なスクリーニングおよびサンプリング手法を使用したパフォーマンスの向上などを行うことができます。完全に自動化された方法でノードを使用し、ノードで固定値を選択および適用できます。または必要に応じて変更の作成および承認、拒否または修正の前に変更をプレビューできます。 [詳細は、 p. 120 自動データ準備 を参照してください。](#)



データ型ノードで、フィールドのメタデータとプロパティを指定します。たとえば、各フィールドに、測定レベル（連続型、名義型、順序型、またはフラグ）を指定し、欠損値とシステム ノルの処理のためのオプションを設定し、モデル作成の目的に対するフィールドの役割を設定し、フィールドと値のラベルを指定し、フィールドの値を指定します。 [詳細は、 p. 150 データ型ノード を参照してください。](#)



フィルタ ノードで、1 つの入力ノードから他の 1 つの入力ノードへ、フィールドをフィルタリング（破棄）し、フィールド名を変更し、また、フィールドを関連付けます。 [詳細は、 p. 173 フィールドのフィルタリングまたは名前の変更 を参照してください。](#)



フィールド作成ノードで、1 つまたは複数の既存フィールドから、データ値を変更するか、新しいフィールドを作成します。これで、タイプ式、フラグ、名義、ステート、カウント、および条件式の各フィールドが作成されます。 [詳細は、 p. 185 フィールド作成ノード を参照してください。](#)



アンサンブル ノードでは、2 つまたはそれ以上のモデル ナゲットを組み合わせて 1 つのモデルよりもより正確な予測を取得します。詳細は、[p. 180 アンサンブル ノード](#) を参照してください。



置換ノードで、フィールド値の置換やストレージの変更を行います。`@BLANK(@FIELD)` のような、CLEM 条件に基づいて値を置換することができます。また、すべての空白値やヌル値を特定の値に置換することもできます。置換ノードは、データ型ノードと一緒に使用される場合が多く、欠損値の置き換えが行われます。詳細は、[p. 199 置換ノード](#) を参照してください。



匿名化ノードは、フィールド名や値の下流の表示方法を変換し、元のデータを隠します。これは、他のユーザーが顧客名やその他の詳細情報をなどの重要情報を使用してモデルを構築できるようにする場合に有用です。詳細は、[p. 202 匿名化ノード](#) を参照してください。



データ分類ノードにより、あるカテゴリ値のセットが別のセットに変換されます。データ分類ノードは、カテゴリを再編成したり、分析用のデータをグループ化しなおす場合に役立ちます。詳細は、[p. 207 データ分類ノード](#) を参照してください。



データ分割ノードで、既存の 1 つまたは複数の連続型（数値範囲）フィールドの値に基づいて、自動的に新しい名義型（セット型）フィールドを作成します。たとえば、連続型収入フィールドを、平均からの偏差による収入グループを含む、新しいカテゴリ フィールドに変換することができます。新規フィールドのビンを作成すると、分割点に基づいてフィールド作成ノードを生成することができます。詳細は、[p. 212 データ分割ノード](#) を参照してください。



リーセンシ、フリクエンシ、マネタリー（RFM）の分析ノードを使用すると、最後に購入したのがどのくらい最近か（リーセンシ）、どのくらい頻繁に購入するか（フリクエンシ）、トランザクション全体でいくら消費したか（マネタリー）を検証することによって、最も良い顧客となると考えられるのはどの顧客かを量的に決定することができます。詳細は、[p. 225 RFM 分析ノード](#) を参照してください。



データ区分ノードで、モデル構築の学習、テスト、および検証の各ステージ用に、データを独立したサブセットに分割するデータ区分フィールドが生成されます。詳細は、[p. 229 データ区分ノード](#) を参照してください。



フラグ設定ノードで、1 つ以上の名義型フィールドに定義されたカテゴリ値に基づいた、複数のフラグ型フィールドが派生します。詳細は、[p. 232 フラグ設定ノード](#) を参照してください。



再構成ノードで、名義型またはグラフ型フィールドを、これから別のフィールドの値で埋めることができるフィールドのグループへ変換します。たとえば、credit、cash、および debit の値の payment type という名前のフィールドがある場合、3 つの新しいフィールド (credit、cash、debit) が作成されます。その各々には、実際の支払の値を含めることができます。 [詳細は、p.233 再構成ノード](#) を参照してください。



行列入替ノードで、レコードがフィールドになり、フィールドがレコードになるように、行内と列内のデータを交換します。 [詳細は、p.236 行列入替ノード](#) を参照してください。



時間区分ノードで、時系列データのモデル作成用に区分を指定し、必要に応じてラベルを作成します。値の間隔が均等に空けられていない場合は、レコード間に一律の間隔をとる必要に応じて、値を充填したり集計したりできます。 [詳細は、p.241 時間区分ノード](#) を参照してください。



時系列ノードにより、以前レコードのフィールドのデータを含む、新規フィールドが作成されます。時系列ノードは、多くの場合、時系列データなどの継続的なデータに使用されます。時系列ノードを使用する前に、ソート ノードを使用して、データをソートしておくこともできます。 [詳細は、p.263 時系列ノード](#) を参照してください。



フィールド順序ノードで、下流のフィールド表示に使用する順序を定義します。この順序は、テーブル、リスト、およびフィールドピッカーなど、さまざまな場所のフィールドの表示に適用されます。この操作は、さまざまなデータセットにおいて、特定のフィールドをより参照しやすくする場合に役立ちます。 [詳細は、p.265 フィールド順序ノード](#) を参照してください。

これらのノードの大半は、データ検査ノードが作成する検査レポートから直接生成することができます。 [詳細は、6 章 p.475 データ準備用のその他のノードの生成](#) を参照してください。

自動データ準備

分析に向けてデータを準備することは、プロジェクトにおいて最も重要な手順の 1 つですが、従来は最も時間を消費する手順の 1 つでもありました。自動データ準備 (ADP) は、データ分析および修正の特定、問題となる、または有用でないと考えられるフィールドの除外、必要に応じた新しい属性の取得、高度なスクリーニング手法を用いたパフォーマンスの改善を行い、タスクを処理します。完全に**自動化**した方法でアルゴリズムを使用して、修正を選択または適用したり、**インタラクティブ**な方法を使用して、必要に応じて変更を実行、承認または拒否する前に変更をプレビューすることができます。

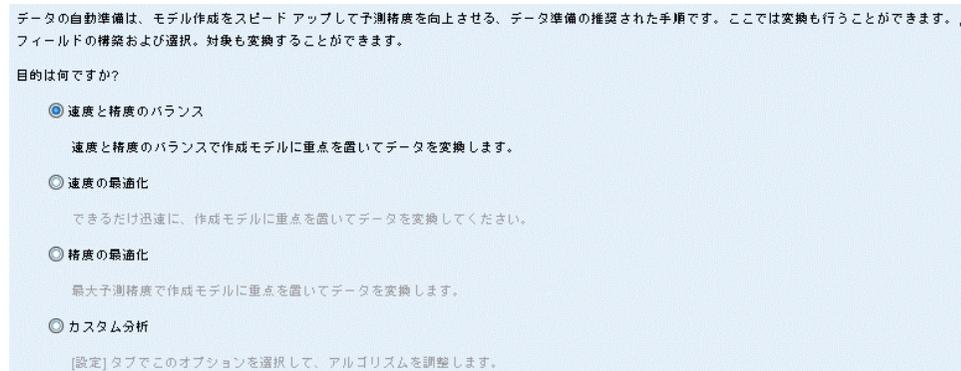
ADP を使用すると、実行する統計の概念の事前情報を必要とせず、モデルを迅速かつ用意に作成できるよう、データを準備することができます。モデルはより迅速に構築およびスコアリングするようになります。また、ADP を使用すると、モデル更新やチャンピオン/チャレンジャーなど、自動モデル作成プロセスの強固さをより向上させます。

注 :ADP で分析用のフィールドを準備する場合、古いフィールドの既存の値およびプロパティを置き換えるのではなく、調整または変換を含む新しいフィールドを作成します。古いフィールドは高度な分析には使用されません。役割は [なし] に設定されます。

例: 世帯主の保険請求を調査するためのリソースが制限されている保険会社が、不正請求の恐れのある疑いを区別するためのモデルを作成したいと考えています。モデルを作成する前に、自動データ準備を使用して、モデル作成のためのデータを準備します。変換が適用される前に提案される変換を確認できる必要があるため、自動データ準備をインタラクティブ モードで使用します。

自動車産業グループは、さまざまな個人用自動車の売り上げを記録します。採算ベースを上回るモデルおよび下回るモデルを特定できるように、自動車の売り上げと自動車の特性との関係を確立したいと考えます。自動データ準備を使用して分析用のデータを準備し、準備「前」および準備「後」のデータを使用してモデルを作成し、結果がどのように異なるかを確認します。

図 4-1 自動データ準備の [目的] タブ



目的は？ 自動データ準備では、ほかのアルゴリズムがモデルを構築し、それらのモデルの予測精度を改善できる速度に影響を与えるような、データ準備の手順を推奨します。このような手順には、フィールドの変換、構築および選択が含まれます。目標も変換することができます。データ準備プロセスで重点を置く必要があるモデル作成の優先度を指定できます。

- **速度および精度のバランス:** このオプションでは、モデル作成アルゴリズムによってデータが処理される速度と、予測の精度の両方に同等の優先度を指定するよう、データを準備します。
- **速度の最適化:** このオプションでは、モデル作成アルゴリズムによってデータが処理される速度に優先度を与えるよう、データを準備します。大きいデータセットを処理する場合、または迅速な回答を求めている場合は、このオプションを選択します。
- **精度の最適化:** このオプションでは、モデル作成アルゴリズムによる予測生成の精度に優先度を与えるよう、データを準備します。
- **カスタム分析。** [設定] タブでアルゴリズムを手動で修正する場合、このオプションを選択します。継続して [設定] タブのオプションに変更を行うも、その他の目的と互換性がない場合、この設定が自動的に選択されます。

ノードの学習

ADP ノードはプロセス ノードとして実装され、データ型ノードと同じように機能します。ADP ノードの**学習**は、データ型ノードのインスタンス化に対応しています。分析が実行されると、上流データが変更されない限り、高度な分析を行わずに指定された変換がデータに適用されます。データ型ノードやフィルタ ノード同様、ADP ノードの接続が解除されても、データモデルや変換は記憶され、再接続された場合に再度学習する必要がなくなります。必要に応じて、一般データのサブセットのデータ モデルについて学習し、実データに使用するためにコピーまたは展開することができます。

ツールバーの使用

ツールバーを使用すると、データ分析の表示を実行および更新し、元のデータと組み合わせて使用できるノードを生成できます。

図 4-2
自動データ準備 - ツールバー



- **ノードの生成** フィルタ ノードまたはフィールド生成ノードを生成できます。このメニューは、分析が [分析] タブに表示されている場合にのみ使用できます。

フィルタ ノードは、変換された入力 フィールドを削除します。ADP ノードがデータセットに元の入力フィールドを残すよう設定する場合、元のセットの入力フィールドを復元し、フォール度ごとにスコアフィールドを解釈できます。たとえば、これはさまざまな入力に対してスコア フィールドのグラフを生成したい場合に役立ちます。

フィールド生成ノードは、元のデータセットと目標の単位を復元します。ADP ノードに範囲型目標を再調整する分析が含まれている場合（[入力および目標の順位] パネルで Box-Cox 再調整を選択）のみ、フィールド生成ノードを生成できます。目標が範囲型でない場合、または Box-Cox 再調整が選択されていない場合、フィールド生成ノードは生成できません。詳細は、[p. 148 フィールド生成ノードの生成](#) を参照してください。

- **表示** [分析] タブに表示される項目を制御するオプションが用意されています。ここでは、グラフ編集コントロールや、メイン パネルおよびリンク ビュー双方の表示選択が用意されています。
- **プレビュー** 入力データに適用される変換のサンプルが表示されます。詳細は、[5 章 ノードのデータをプレビューする in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- **データの分析** 現在の設定を使用して分析を開始し、[分析] タブに結果を表示します。
- **分析のクリア** 既存の分析を削除します（現在の分析が存在する場合にのみ使用できます）。

ノードの状態

矢印、または分析が行われたかどうかを示すアイコンをクリックすることによって、IBM® SPSS® Modeler 領域に ADP ノードの状態が示されます。

[フィールド] タブ

図 4-3



モデルを作成する前に、対象フィールドや入力フィールドを指定する必要があります。いくつかの例外を除いて、すべてのモデル ノードは、上流のデータ型ノードからのフィールド情報を使用します。データ型ノードを使用して入力フィールドおよび対象フィールドを選択する場合、このタブで何も変更する必要はありません

データ型ノードの設定を使用： このオプションを選択すると、上流のデータ型ノードからのフィールド情報がこのノードで使用されます。これはデフォルトです。

ユーザー設定を使用： このオプションを選択すると、上流のデータ型ノードからのフィールド情報ではなく、ここで指定したフィールド情報がこのノードで使用されます。このオプションを選択した後に、必要に応じて以下のフィールドを指定します。

対象： 1 つまたは複数の対象フィールドが必要なモデルの場合に、対象フィールドを選択します。これは、データ型ノードのフィールドの役割を [対象] に設定するのと似ています。

入力： 入力フィールドを選択してください。これは、データ型ノードのフィールドの役割を [入力] に設定するのと似ています。

[設定] タブ

[設定] タブは、アルゴリズムがデータをどのように処理するかを調整するために変更できる、複数グループの設定で構成されています。その他の目的と互換性のないデフォルト設定に変更を行うと、[目的] タブが自動的に更新され、[分析のカスタマイズ] オプションを選択します。

フィールド設定

図 4-4
自動データ準備 - フィールド設定

目的を変更してもフィールド設定に影響はありません。

度数フィールドを使用

重みフィールドを使用

モデル作成から除外されたフィールドの処理方法:

未使用フィールドを除外

未使用フィールドの方向を "なし" に設定

入力フィールドが既存の分析に一致しない場合:

実行を停止して既存の分析を保存

既存の分析をクリアして、新しいデータを分析

度数フィールドを使用：フィールドを度数の重みとして選択できます。集計データを使用している場合など、学習データのレコードがそれぞれ複数の単位を示す場合に使用します。フィールド値は、レコードごとに示した単位数です。詳細は、[3 章 度数フィールドと重みフィールドの使用 in IBM SPSS Modeler 15 モデル作成ノード](#) を参照してください。

重みフィールドを使用：フィールドをケースの重みとして選択できます。ケースの重みを使用して、出力フィールドのレベル間の分散における相違を処理します。詳細は、[3 章 度数フィールドと重みフィールドの使用 in IBM SPSS Modeler 15 モデル作成ノード](#) を参照してください。

モデル作成から除外されたフィールドの処理方法：除外されたフィールドの処理方法を指定します。データを除外するか、[役割] を [なし] に設定するかを選択できます。

受信フィールドが既存の分析と一致しない場合：学習済み ADP ノードを実行するとき 1 つまたは複数の必須入力フィールドが受信データセットにならない場合の処理方法を指定します。

- **実行を停止して既存の分析を保存する：**実行プロセスを停止し、現在の分析情報を保存して、エラーを表示します。
- **既存の分析をクリアして新しいデータを分析する：**既存の分析をクリアし、受信データを分析して、推奨されている変換をそのデータに適用します。

日付および時刻の準備

図 4-5
自動データ準備の日付および時刻の準備設定

モデル作成の日付と時刻を準備

期間を計算

基準日までの経過時間を計算

基準日

本日の日付
 固定日付

日付: 2009-04-21

期間（日数）の単位

自動
 固定単位

単位: 年

基準時刻までの経過時間を計算

基準時刻

現在の時刻
 固定時刻

時間: 10:36:38

期間（時間数）の単位

自動
 固定単位

単位: 時間

周期的時間要素の取得

日付から取得:

年 月 日

時刻から取得:

時間 分 秒

多くのモデル作成アルゴリズムは、日付や時刻の詳細を直接処理することはできません。これらの設定を使用して、既存データの日付および時刻から、モデル入力として使用できる新しい期間データを取得できます。日付および時刻を含むフィールドは、日付または時間のストレージタイプで事前定義する必要があります。元の日付および時間フィールドは、自動データ準備に従うモデル入力としては推奨されません。

モデル作成の日付と時刻を準備: このオプションを選択解除すると、他のすべての [日付および時刻の準備] コントロールが無効になりますが選択は維持されます。

基準日までの経過時間を計算: 日付を含む各変数の基準日以降の年/月/日の数を生成します。

- **基準日:** 入力データの日付情報に関して、期間を計算する日付を指定します。[今日の日付] を選択すると、ADP が実行されている場合、現在のシステムの日付が常に使用されます。特定の日付を使用するには、[固定日付] を選択して、該当する日付を入力します。ノードが初

めて作成された場合、[固定日付] フィールドには、現在の日付が自動的に入力されます。

- **期間(日数)の単位:** ADP が自動的に期間 (日数) の単位を決定するかどうかを指定するか、年、月、または日付の [固定単位] を選択します。

基準時刻までの経過時間を計算: 時刻を含む各変数の基準日以降の時/分/秒の数を生成します。

- **基準時刻:** 入力データの時間情報に関して、期間を計算する時刻を指定します。[現在の時刻] を選択すると、ADP が実行されている場合、現在のシステムの時刻が常に使用されます。特定の時刻を使用するには、[固定時刻] を選択して、該当する時刻を入力します。ノードが初めて作成された場合、[固定時刻] フィールドには、現在の時刻が自動的に入力されます。
- **期間(時間数)の単位:** ADP が自動的に期間 (時間) の単位を決定するかどうかを指定するか、時間、分、または秒の [固定単位] を選択します。

周期的時間要素の取得: これらの設定を使用して、1 つの日付または時刻フィールドを 1 つまたは複数のフィールドに分割します。たとえば、3 つの日付のチェックボックスをすべてオンにすると、入力日付フィールド「1954-05-23」は 1954、5、および 23 の 3 つのフィールドに分割されます。それぞれ [フィールド名] パネルで定義された接尾辞を使用すると、元の日付フィールドは無視されます。

- **日付から取得:** 日付フィールドについて、年、月、日付またはそれらの組み合わせを取得するかどうかを指定します。
- **時刻から取得:** 時刻フィールドについて、時間、分、秒またはそれらの組み合わせを取得するかどうかを指定します。

フィールドの除外

図 4-6
自動データ準備のフィールドの除外設定

連続型フィールドは常に除外されます。

低品質の入力フィールドを除外する

入力フィールドを除外する

欠損値が多すぎるフィールドを除外する

欠損値の最大パーセンテージ: %

一意のカテゴリが多すぎる名義フィールドを除外する

カテゴリの最大数:

単一カテゴリに値が多すぎるカテゴリフィールドを除外する

単一カテゴリの最大パーセンテージ: %

品質の悪いデータは、予測の精度に影響を与える場合があります。そのため、入力フィールドに適切な品質レベルを指定することができます。定数または 100% 欠損値であるすべてのフィールドは、自動的に除外されます。

品質の悪い入力フィールドを除外: このオプションを選択解除すると、その他すべての [フィールドを除外] コントロールが無効になりますが選択は維持されます。

欠損値の多いフィールドの除外: 欠損値が指定された割合を超えて含まれるフィールドは、高度な分析から除外されます。0 以上 100 以下の値を指定しますが (0 はオプションの選択解除を示す)、すべての欠損値を含むフィールドは自動的に除外されます。デフォルトは 50 です。

一意のカテゴリの名義フィールドの除外: カテゴリ数が指定された数を超えて含まれるフィールドは、高度な分析から除外されます。正の整数を指定します。デフォルトは 100 です。ID、住所、名前などのモデル作成からレコード特有の情報を含むフィールドを自動的に削除する場合に役立ちます。

単一カテゴリの値が多いカテゴリフィールドの除外: 指定された割合を超えるレコードが含まれるカテゴリを持つ順序型フィールドおよび名義型フィールドは、高度な分析から除外されます。0 以上 100 以下の値を指定しますが (0 はオプションの選択解除を示す)、定数フィールドは自動的に除外されます。デフォルトは 95 です。

入力フィールドおよび目標フィールドの準備

処理しているデータで完全な状態のものがないため、分析を実行する前にいくつかの設定を調整する必要があります。たとえば、外れ値の削除、欠損値の処理方法の指定、タイプの調整などです。

注： このパネルで値を変更した場合、[目的] タブが自動的に更新され、[カスタム分析] オプションが選択されます。

図 4-7
自動データ準備 - 入力と目標の設定

モデル作成用の入力フィールドおよび対象フィールドを準備

データ型を調整してデータ品質を改善

| 入力 | 対象 |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> 数値型フィールド (順序型および連続型) の種類を変更 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 最小のカテゴリが最初に、最大カテゴリが最後になるよう名義フィールドの順序を変更 |
| <input type="checkbox"/> | <input type="checkbox"/> 連続型フィールドの外れ値を置換 (共通スケールに投入された場合の入力フィールドに推奨) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 連続型フィールド: 欠損値を平均値と置換 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 名義フィールド: 欠損値を最頻値と置換 |
| <input type="checkbox"/> | <input type="checkbox"/> 順序型フィールド: 欠損値を中央値に置換 |

順序型フィールドの値の最大数:

連続型フィールドの値の最大数:

外れ値の分割値: (標準偏差)

外れ値の置換方法: 分割値に置換 値の削除

連続型フィールドを変更

共通スケールにすべての連続型入力フィールドを投入する (フィールド構築を実行する場合に高く推奨)

最終平均値: 最終標準偏差:

最終平均値: 最終標準偏差:

モデル作成の入力フィールドの目標フィールドを準備する: パネルのすべてのフィールドをオンまたはオフに切り替えます。

タイプを調整してデータ品質を改善する: 入力フィールドおよび目標フィールドについて、それぞれいくつかのデータ変換を指定できます。目標フィールドの値を変更する必要がないためです。たとえば、収入の予測 (単位: ドル) は、ログで測定される予測よりもより意味があります。また、目標フィールドに欠損値がある場合、欠損値を入力する予測ゲインがないため、入力フィールドに欠損値を入力すると、一部のアルゴリズムは、欠損する情報を処理することができます。

外れ値の分割値など、これらの変換の追加設定は、目標フィールドと入力フィールドに共通しています。

入力フィールドと目標フィールドのいずれか、または両方に次の設定を選択できます。

- **数値型フィールドのタイプを調整する:** 測定レベルが順序型の数値型フィールドを連続型に変換できるかどうか、またはその逆を指定します。最小しきい値および最大しきい値を指定して、変換を制御できます。
- **名義型フィールドの並べ替え:** 数値型 (セット型) フィールドを最小カテゴリから最大カテゴリの順に並べ替えます。
- **連続型フィールドの外れ値を置換:** 外れ値を置き換えるかどうかを指定します。次の [外れ値の置換方法] オプションと組み合わせて使用します。

- **連続型フィールド:欠損値を平均値に置換**連続型（範囲型）フィールドの欠損値を置き換えます。
- **名義フィールド:欠損値を最頻値に置換**名義型（セット型）フィールドの欠損値を置き換えます。
- **順序型フィールド:欠損値を中央値に置換**順序型（順序セット型）フィールドの欠損値を置き換えます。

順序フィールドの値の最大数: 順序型（順序セット型）フィールドを連続型（範囲型）を再定義するしきい値を指定します。デフォルトは 10 です。そのため、順序型フィールドに 10 を超えるカテゴリがある場合、連続型（範囲型）に再定義されます。

連続型フィールドの値の最小数: 連続型（範囲型）フィールドを順序型（順序セット型）を再定義するしきい値を指定します。デフォルトは 5 です。そのため、連続型フィールドに 5 を超えるカテゴリがある場合、順序型（順序セット型）に再定義されます。

外れ値の分割値: 標準偏差で測定される、外れ値の分割基準を指定します。デフォルトは 3 です。

外れ値の置換方法: 外れ値を、分割値でトリム化 (coerce) して置き換えるか、外れ値を削除して欠損値として設定するかを選択します。欠損値に設定した外れ値は、上記で選択された欠損値処理の設定にしたがって処理されます。

すべての連続型入力フィールドを共通尺度に設定: 連続型入力フィールドを正規化するには、このチェック ボックスをオンにして、正規化方法を選択します。デフォルトは [zスコア変換] で、デフォルトが 0 の [最終平均値]、デフォルトが 1 の [最終標準偏差] を指定できます。また、[最小/最大変換] を使用して最小値と最大値を指定できます（デフォルトはそれぞれ 0 および 100 です）。

このフィールドは、[選択および構築] 設定で [フィールド構築の実行] を選択する場合に特に役立ちます。

Box-Cox 変換で連続型目標を再調整する: 連続型（スケールまたは範囲型）目標フィールドを正規化するには、このチェック ボックスをオンにします。Box-Cox 変換は、[最終平均値] のデフォルト値が 0、[最終標準偏差] のデフォルト値が 1 です。

注： 目標を正規化する場合、目標の次元が変換されます。この場合、フィールド生成ノードを生成して、逆変換を適用し、変換された単位を、高度な処理で認識可能な形式に戻す必要があります。 [詳細は、p. 148 フィールド生成ノードの生成](#) を参照してください。

構築およびフィールド選択

データの予測精度を向上させるために、入力フィールドを変換したり、既存フィールドに基づいて新しいフィールドを構築できます。

注：このパネルで値を変更した場合、[目的] タブが自動的に更新され、[カスタム分析] オプションが選択されます。

図 4-8
自動データ準備 - 変換、構築、選択の設定

入力フィールドを変換、構築および選択して予測精度を向上させる

カテゴリ入力フィールド

まばらなカテゴリを結合して対象との関連を最大化 p 値:

監視された結合が除外された後 1 つのカテゴリのみを持つフィールドを入力します。

対象がない場合、カウントに基づいてまばらなカテゴリを結合

順序型フィールド 名義型フィールド カテゴリ内のケースの最小割合 (%):

連続型入力フィールド

予測精度を保持する場合は連続型フィールドを分割 (カテゴリ対象にのみ使用可)

p 値:

分割が除外された後 1 つのカテゴリのみを持つ入力フィールド

フィールド選択と構築

フィールド選択を実行 p 値:

対象が連続型の場合、フィールド選択が連続型入力フィールドに、そしてカテゴリ入力フィールドに適用されます。

フィールド構築を実行

対象が連続型であるまたは対象がない場合にフィールド構築が連続型入力フィールドに適用されます。

入力フィールドを変換、構築および選択して予測精度を改善する: パネルのすべてのフィールドをオンまたはオフに切り替えます。

まばらなカテゴリを結合して目標との関連性を最大化: 目標と関連して処理する変数の数を減らして、より節約的なモデルを作成します。必要に応じて、デフォルト値の 0.05 から確率値を変更します。

すべてのカテゴリが 1 つのカテゴリに結合される場合、予測値としての値がないため、元のバージョンのフィールドおよび派生した化されたフィールドは除外されます。

目標がない場合、度数に基づいてまばらなカテゴリを結合する: 目標フィールドがないデータを処理する場合、順序型 (順序セット型) フィールドおよび名義型 (セット型) フィールドのいずれか、または両方のまばらなカテゴリを結合できます。結合するカテゴリを特定するデータのケースまたはレコードの最小パーセントを指定します。デフォルトは 10 です。

カテゴリは、次の規則に従って結合されます。

- 2 値フィールドの結合は実行しない。

- 結合時にカテゴリが 2 つしかない場合、結合は停止する。
- 元のカテゴリがない場合、または結合時にカテゴリが作成されない場合、指定したケースの最小パーセントより少ない場合、結合は停止する。

予測精度を保持しながら連続型フィールドを分割: データにカテゴリ型目標が含まれている場合、強い関連を持つ連続型入力フィールドを分割して、処理のパフォーマンスを向上させることができます。必要に応じて、等質サブグループの確率値をデフォルト値の 0.05 から変更します。

カテゴリ化操作によって特定フィールドに単一ビンが生成される場合、予測値としての値がないため、元のバージョンのフィールドおよびカテゴリ化されたフィールドは除外されます。

注：IBM® SPSS® Modeler のその他の部分で使用されるADP のカテゴリ化は最適カテゴリ化とは異なります。最適カテゴリ化では、エントロピー情報を使用して、連続型変数をカテゴリ変数に変換します。最適カテゴリ化では、データを並べ替え、メモリ内にすべて保存する必要があります。ADP では、等質サブグループを使用して、連続型変数を分割します。ADP カテゴリ化では、データを並べ替え、メモリ内にすべて保存する必要はありません。等質サブグループの方法を使用して連続型変数をカテゴリ化すると、カテゴリ化したあとのカテゴリ数は、常に目標内のカテゴリ数と等しいか少なくなります。

フィールド選択を実行: 相関係数が低いフィールドを削除します。必要に応じて、デフォルト値の 0.05 から確率値を変更します。

このオプションは、目標が連続型連続型入力フィールドに、カテゴリ型入力フィールドに適用されます。

フィールド構築の実行: 複数の既存フィールドの組み合わせから新しいフィールドを取得します（既存フィールドはモデル作成から削除されます）。

このオプションは、目標が連続型の場合または目標がない場合にのみ、連続型入力フィールドに適用されます。

フィールド名

図 4-9
自動データ準備のフィールドの名前付け設定

変換および構築されたフィールド

変換された対象フィールドの名前の拡張子:

変換された入力フィールドの名前の拡張子:

構築された機能のルート名:

日付と時間から計算した期間

日付から計算した期間の名前の拡張子

年: 月: 日:

時間から計算した期間の名前の拡張子

時間: 分: 秒:

日付と時間から抽出したサイクル要素

日付から抽出したサイクル要素の名前の拡張子

年: 月: 日:

時間から抽出したサイクル要素の名前の拡張子

時間: 分: 秒:

新しいフィールドや変換されたフィールドを用意に特定できるようにするために、ADP は新しい基本名、接頭辞または接尾辞を作成し、適用します。それらの名前を修正して、ニーズおよびデータにより関連付けることができます。別のラベルを指定する場合、下流のデータ型ノードで指定する必要があります。

変換され構築されたフィールド。 変換された目標フィールドおよび入力フィールドの適用する名前の拡張子を指定します。

ADP ノードでは、何も入力されない文字列フィールドを設定すると、未使用フィールドの処理方法によってはエラーが発生する場合があります。[設定] タブの [フィールド設定] パネルで、[モデル作成から除外されたフィールドの処理方法] が [未使用フィールドを除外] に設定された場合、入力フィールドおよび目標フィールドの名前の拡張子を設定しないこともできます。元のフィールドが除外され、変換されたフィールドが上書き保存されます。この場合、新しく変換されたフィールドは元のフィールドと同じ名前がつけられます。

ただし、[未使用フィールドの方向を 'なし' に設定] を選択し、その後空欄または Null を選択すると、重複したフィールド名を作成するため、目標フィールドおよび入力フィールドの名前の拡張子によりエラーが発生する場合があります。

さらに、[選択および構築]設定を使用して、構築されるフィールドに適用する接頭辞名を指定します。数値の接尾辞をこの接頭辞のルート名に追加して、新しい名前を作成します。番号の形式は、次のように、取得された新しいフィールドの数によって異なります。

- 1 から 9 の構築済みフィールドの名前は、feature1 から feature9 です。
- 10 から 99 の構築済みフィールドの名前は、feature01 から feature99 です。
- 100 から 999 の構築済みフィールドの名前は、feature001 から feature999 です。

これにより、構築されたフィールドは、フィールド数に関係なく、合理的な順序で並べ替えられます。

日付と時刻から計算した期間：日付および時刻から算出した期間に適用する名前の拡張子を指定します。

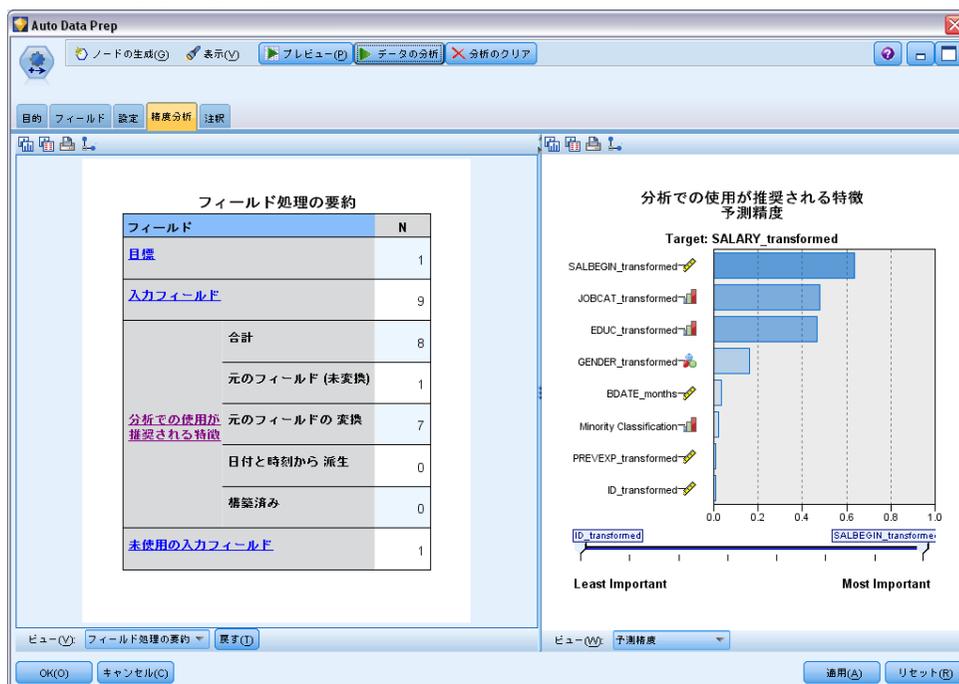
日付と時刻から抽出したサイクル要素：日付および時刻から算出した周期的要素に適用する名前の拡張子を指定します。

[分析] タブ

- ▶ [目的] タブ、[フィールド] タブ、[設定] タブで行った変更など、ADP 設定に問題がない場合、[データを分析] をクリックしてください。アルゴリズムにより設定がデータ入力に適用され、[分析] タブに結果が表示されます。

[分析] タブには、データの処理の概要を示すテーブル形式の出力およびグラフィック出力が含まれ、スコアリング用のデータをどのように修正または改善するかについての推奨事項が表示されます。これらの推奨事項を確認し、承認したり拒否したりすることができます。

図 4-10
自動データ準備の [分析] タブ



[分析] タブは 2 つのパネルで構成されています。左側はメイン ビュー、右側はリンク ビューまたは補助ビューです。メイン ビューには、次の 3 種類があります。

- フィールド処理の要約 (デフォルト)。詳細は、p. 136 フィールド処理の要約 を参照してください。
- フィールド： 詳細は、p. 138 フィールド を参照してください。
- アクションの概要。 詳細は、p. 140 アクションの概要 を参照してください。

リンク/補助ビューには、次の 4 種類あります。

- 予測の精度 (デフォルト)。 詳細は、p. 141 予測精度 を参照してください。
- フィールド テーブル。 詳細は、p. 142 [フィールド] テーブル を参照してください。
- フィールド詳細。 詳細は、p. 143 フィールド詳細 を参照してください。
- アクションの詳細。 詳細は、p. 145 アクションの詳細 を参照してください。

ビュー間のリンク

メイン ビューで、表内の下線付きテキストは、リンク ビューの表示を制御します。テキストをクリックすると、特定のフィールド、一連のフィールドまたは処理中のステップに関する詳細を取得できます。最後に選択したリンクは濃い色で表示されます。これにより、2 つのビュー パネルのコンテンツ間の接続を特定できます。

ビューのリセット

元の分析に関する推奨事項を再度表示し、[分析] ビューに行った変更を取り消す場合、メイン ビュー パネルの一番下にある [リセット] をクリックしてください。

フィールド処理の要約

図 4-11
フィールド処理の要約

| フィールド | N |
|------------------------------|---|
| 目標 | 1 |
| 入力フィールド | 9 |
| 合計 | 8 |
| 元のフィールド (未変換) | 1 |
| 分析での使用が推奨される特徴 元のフィールドの変換 | 7 |
| 日付と時刻から派生 | 0 |
| 構築済み | 0 |
| 未使用の入力フィールド | 1 |

[フィールド処理の要約] 表には、フィールドの状態や構築フィールド数への変更など、処理に対する全体の影響の射影したスナップショットが表示されます。

モデルは実際に構築されていないため、データ準備の前後に予測精度船体の変更に対する測定またはグラフはありません。その代わりに、推奨された各予測の予測精度についてのグラフを表示できます。

表には、次の情報が表示されます。

- 目標フィールド数。
- 元（入力）予測フィールド数。
- 分析およびモデル作成に使用が推奨される予測値。ここには、推奨フィールド数の合計、元の（変換されていない）推奨フィールド数、変換された推奨フィールド数（中間バージョンのフィールド、日付/時刻フィールドから算出したフィールド、構築されたフィールドは除外）、日付/時刻フィールドから算出した推奨フィールド数、構築された推奨予測フィールド数が表示されます。
- 元の形式でも、派生フィールドとしても、あるいは構築された予測値に対する入力としても、いかなる形式でも使用が推奨されない入力予測値の数。

[フィールド] 情報に下線がある場合、クリックするとリンク ビューに詳細が表示されます。[目標]、[入力フィールド]、および[未使用の入力フィールド]の詳細は、[フィールド テーブル] リンク ビューに表示されます。詳細は、[p. 142 \[フィールド\] テーブル を参照してください](#)。[分析の使用が推奨されるフィールド] は、[予測精度] リンク ビューに表示されます。詳細は、[p. 141 予測精度 を参照してください](#)。

フィールド

図 4-12
Fields

| フィールド | | | |
|---|---|---|-------|
| 目標 | | | |
| 名前 | 種類 | | |
| SALARY |  | | |
| 機能 <input type="checkbox"/> テーブルに非推奨フィールドを追加する(D) | | | |
| 使用するバージョン | 名前 | 種類 | 予測べき乗 |
| 変換 | SALBEGIN |  | 0.64 |
| 変換 | JOB CAT |  | 0.48 |
| 変換 | EDUC |  | 0.47 |
| 変換 | GENDER |  | 0.16 |
| 変換 | BDATE_Duration Months |  | 0.03 |
| 元のフィールド | MINORITY |  | 0.02 |
| 変換 | PREVEXP |  | 0.01 |

[フィールド] メイン ビューには、処理済みフィールドと、ADP が下流モデルにそれらのフィールドの使用を推奨するかどうかを表示します。任意のフィールドについての推奨事項を上書きできます。たとえば、構築済みフィールドを除外する、または ADP が除外を推奨するフィールドを追加するなどです。フィールドが変換された場合、推奨された変換を受け入れるか、元のバージョンを使用するかを決定できます。

[フィールド] ビューは、2 つのテーブルで構成されています。1 つは目標フィールドについてのテーブル、もう 1 つは処理されたまたは作成された予測フィールドについてのテーブルです。

[目標] テーブル

[目標] テーブルには、目標がデータに定義されているかどうかだけが表示されます。

テーブルには、次の 2 つの列があります。

- **名前**：目標フィールドの名前またはラベルです。フィールドが変換された場合でも、元の名前が常に使用されます。
- **測定レベル**。測定レベルを示すアイコンが表示されます。マウス ポインタをアイコンの上に停止させると、データについて説明するラベル（連続型、順序型、名義型など）が表示されます。
目標が変換されると、[尺度] 列には、最終的な変換バージョンが反映されます。注：目標の変換をオフにすることはできません。

予測値テーブル

[予測値] テーブルが常に表示されます。テーブルの各行は、フィールドを示します。デフォルトでは、行は予測精度の高い順に並んでいます。

通常のフィールドの場合、元の名前は常に行の名前として使用されます。元のバージョンおよび派生バージョンの日付/時刻フィールドがテーブルの各行に表示されます。また、テーブルには構築済み予測フィールドも表示されます。

テーブルに表示される変換されたバージョンのフィールドは、常に最終バージョンを示します。

デフォルトでは、推奨されたフィールドのみが、[予測値] テーブルに表示されます。残りのフィールドを表示するには、テーブルの上にある [テーブルに非推奨フィールドを追加する] ボックスを選択します。これらのフィールドは、テーブルの一番下に表示されます。

テーブルには、次の列が表示されます。

- **使用バージョン**：フィールドを下流で使用するかどうか、推奨された変換を使用するかどうかを制御するドロップダウン リストが表示されます。デフォルトでは、ドロップダウン リストには推奨事項が反映されます。
変換された通常の予測値の場合、ドロップダウン リストには [変換済み]、[変換前]、[使用しない] の 3 つの選択肢があります。
変換されていない通常の予測値の場合、[変換前] および [使用しない] の選択肢があります。
派生した日付/時刻フィールドおよび構築済み予測フィールドの場合、[変換済み] および [使用しない] の選択肢があります。
元の日付フィールドの場合、ドロップダウン リストは無効となり、[使用しない] に設定されます。
注：変換前バージョンと変換済みバージョンの両方の予測フィールドの場合、[変換前] と [変換済み] でバージョンを変更すると、自動的にそれらのフィールドの [尺度] および [予測精度] の設定が更新されます。

- **名前**：各フィールドの名前はリンクになっています。名前をクリックすると、フィールドに関する詳細情報がリンク ビューに表示されます。詳細は、[p.143 フィールド詳細](#) を参照してください。
- **測定レベル**。データ型を示すアイコンが表示されます。マウス ポインタをアイコンの上に停止させると、データについて説明するラベル（連続型、順序型、名義型など）が表示されます。
- **予測精度**：ADP が推奨するフィールドについての予測精度のみが表示されます。この列は、目標が定義されている場合にのみ表示されます。予測精度は 0 ～ 1 で、値が大きいほど、予測精度が「良い」ことを示します。一般的に、予測精度は ADP 分析の予測を比較するのに役立ちますが、予測精度の値を分析間で比較することはできません。

アクションの概要

図 4-13
アクションの概要

アクションの要約

| アクション |
|-------------------------------|
| テキスト フィールド |
| 日付および時刻のフィールド |
| 特徴のスクリーニング |
| チェック タイプ |
| 外れ値 |
| 欠損値 |
| 目標 |
| カテゴリ型フィールド |
| 連続型フィールド |

自動データ準備で実行された各アクションについて、入力予測フィールドは変換および/または除外されます。ステップを通過したフィールドは、次のステップで使用されます。最後のステップまで通過したフィールドがモデル作成に推奨されます。変換された入力予測フィールドおよび構築されたフィールドは除外されます。

各予測フィールドの予測精度は、目標が連続型かカテゴリかに応じて、線型回帰、または naïve Bayes から算出されます。

[フィールド] テーブル

図 4-15
フィールド テーブル

入力フィールド

| 名前 | 種類 |
|----------|---------|
| ID | 続行 |
| GENDER | 設定 |
| BDATE | 続行 |
| EDUC | 順序付けセット |
| JOB CAT | 順序付けセット |
| SALBEGIN | 続行 |
| JOB TIME | 続行 |
| PREVEXP | 続行 |
| MINORITY | 順序付けセット |

[フィールド処理の要約] メイン ビューで [目標]、[予測フィールド]、[未使用の予測フィールド] をクリックすると表示され、[フィールド テーブル] ビューには関連するフィールドを示す単純なテーブルが表示されます。

テーブルには、次の 2 つの列があります。

- **名前**：予測フィールド名です。

目標フィールドの場合、目標が変換されている場合でも、フィールドの元の名前またはラベルが使用されます。

変換されたバージョンの通常のパredictフィールドの場合、フィールド名は [設定] タブの [フィールド名] パネルでの `_transformed` など接尾辞の選択内容を反映します。

日付と時刻から算出したフィールドの場合、最終的に変換されたバージョンの名前が使用されます。たとえば、`bdate_years` です。

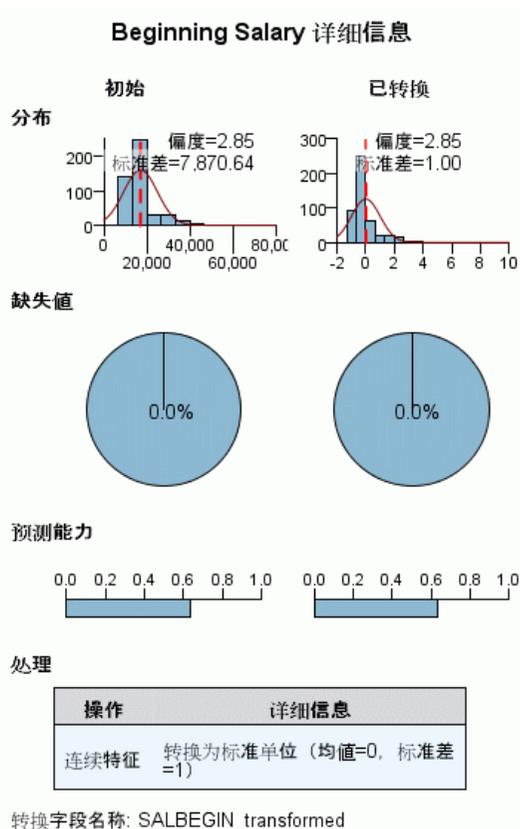
構築済み予測フィールドの場合、Predictor1 など、構築済み予測フィールドの名前が使用されます。

- **尺度:** データ型を示すアイコンが表示されます。

目標フィールドの場合、[尺度] は常に変換されたバージョンが反映されます（目標フィールドが変換されている場合）。たとえば、順序型（順序セット型）から連続型（範囲型、スケール）への変更、またはその逆も同様です。

フィールド詳細

図 4-16
フィールド詳細



[フィールド] メイン ビューで [名前] をクリックすると表示され、[フィールド詳細] ビューには選択したフィールドの分布、欠損値、予測精度グラフ（該当する場合）が表示されます。また、必要に応じて、フィールドの処理履歴や変換フィールドの名前も表示されます。

各図表セットについて、2 つのバージョンが並んで表示され、変換が適用されたフィールドと適用されていないフィールドを比較します。変換されたバージョンのフィールドがない場合、元のバージョンの図表のみが表示されます。派生した日付/時刻フィールドおよび構築済み予測フィールドの場合、新しい予測フィールドの図表のみ表示されます。

注： カテゴリ数が多すぎるためにフィールドが除外された場合、処理の履歴のみが表示されます。

分布図

連続型フィールドの分布は、正規曲線が重なり、平均値を表す垂直参照線を使用したヒストグラムで表示されます。カテゴリ フィールドは棒グラフで表示されます。

ヒストグラムには、標準偏差や歪度を示すラベルがつけられています。ただし、値の数が 2 以下の場合、または元のフィールドの分散が 10 ～ 20 より小さい場合、歪度は表示されません。

図表の上にマウスポインタを停止させると、ヒストグラムの平均値、またはカテゴリのレコード数合計の度数またはパーセンテージを棒グラフで表示します。

欠損値のグラフ

円グラフは、変換が適用された場合、変換が適用されていない場合の欠損値の割合を比較します。グラフのラベルはパーセンテージを示します。

ADP が欠損値の処理を実行した場合、変換後の円グラフには置換値、つまり欠損値の変わりに使用される値もラベルで表示します。

グラフにマウスポインタを停止させると、全体のレコード数の欠損値数と全体の割合が表示されます。

予測精度グラフ

推奨フィールドについて、棒グラフに変換前後の予測精度が表示されます。目標フィールドが変換されると、予測精度は変換後の目標フィールドについて計算されます。

注： 目標が定義されていない場合、またはメイン ビュー パネルで目標をクリックした場合、予測精度のグラフは表示されません。

グラフの上のマウスポインタを停止させると、予測精度の値が表示されます。

処理履歴表

表には、変換されたバージョンのフィールドがどのように取得されたかを示されます。ADP によって行われた処理が、実行順に表示されます。ただし、特定のステップにおいては、特定のフィールドに対して複数の処理が実行されている場合があります。

注：この表は、変換されていないフィールドには表示されません。

表内の情報は、2 つまたは 3 列に分けて表示されます。

- **アクション:** アクションの名前たとえば、連続型予測フィールドです。詳細は、[p. 145 アクションの詳細](#) を参照してください。
- **詳細:** 実行された処理のリスト(例: 標準単位への変換)。
- **関数:** 構築された予測フィールドにのみ表示され、「 $.06*age + 1.21*height$ 」など、入力フィールドの線型結合が表示されます。

アクションの詳細

図 4-17
ADP 分析 - アクションの詳細

ステップ 9: 連続型フィールド

| 変換 | 特徴の数 | 基準 | |
|-----------|------|-----|------|
| | | 平均値 | 標準偏差 |
| 標準の単位への変換 | 5 | 0 | 1 |

| 特徴空間の構築 | N |
|----------------------|---|
| 特徴の構築 | 0 |
| 目標との関連性の低さにより除外される特徴 | 1 |
| 分割後定数項となったため除外される特徴 | 0 |

[アクションの概要] メイン ビューで下線の付いた [アクション] を選択した場合に表示されます。[アクションの詳細] リンク ビューには、実行された各アクションのアクション固有の情報およびおよび共通情報が表示されます。アクション固有の詳細情報が最初に表示されます。

各アクションについて、説明が、リンク ビューの一番上にタイトルとして表示されます。アクション固有の詳細がタイトルの下に表示され、派生予測フィールド数、フィールドの再計算、目標の変換、結合または並べ替えられたカテゴリ、構築または除外された予測フィールドの詳細が含まれる場合があります。

各アクションが処理されるごとに、予測フィールドが除外されたり結合されたりするなどの処理中に使用される予測フィールド数が変わる場合があります。

注：アクションが無効になった場合、または指定された目標がなかった場合、[アクションの概要] メイン ビューでアクションがクリックされた場合、アクションの詳細の代わりにエラー メッセージが表示されます。

アクション数は 9 つですが、すべての分析で、すべての処理が行われるわけではありません。

テキストフィールド テーブル

テーブルには、次の数が表示されます。

- 削除される空白値
- 分析から除外された予測フィールド

日付および時刻の予測フィールド テーブル

テーブルには、次の数が表示されます。

- 日付および時刻予測フィールドから算出した期間
- 日付および時刻の要素
- 派生した日付および時刻の予測フィールドの合計

期間（日数）が計算された場合、基準日または基準時刻が脚注として表示されます。

予測フィールドスクリーニング テーブル

テーブルには、処理から除外された次の予測フィールドの数が表示されます。

- 定数
- 欠損値の多い予測フィールド
- 単一カテゴリのケース数が多い予測フィールド
- カテゴリ数の多い名義型フィールド（セット）
- 除外された予測フィールドの合計

尺度テーブルのチェック

テーブルには再計算されたフィールド数を、次の項目に分けて表示します。

- 連続型として計算された順序型フィールド（順序セット型）
- 順序型フィールドとして計算された連続型フィールド
- 再計算の合計

連続型または順序型である入力フィールド（対象または予測フィールド）がない場合、脚注として表示されます。

外れ値テーブル

テーブルには、外れ値の処理方法の数が表示されます。

- [設定] タブの [入力と目標の準備] パネルの設定に応じて、外れ値が検出されトリム化された連続型フィールドの数、または外れ値が検出され欠損値に設定された外れ値の連続型フィールドの数。
- 外れ値を処理した後定数項となったために除外される連続型フィールドの数。

1 つの脚注には外れ値の分割値、連続型である入力フィールド（目標または予測フィールド）がない場合、別の脚注が表示されます。

欠損値テーブル

テーブルには欠損値を置換したフィールド数を、次の項目に分けて表示します。

- 対象：目標が指定されていない場合、この行は表示されません。
- 予測値：名義型（セット型）、順序型（順序セット型）、連続型に分割して表示されます。
- 置換された欠損値の合計数。

目標テーブル

テーブルには、目標が変換されたかどうかについて、次のように表示されます。

- 正規性への Box-Cox 変換。指定の基準（平均および標準偏差）およびラムダを示す列に分割されます。
- 安定性を向上させるために並べ替えられた目標カテゴリ。

カテゴリ型予測フィールド テーブル

テーブルには、次のようなカテゴリ型予測フィールドの数が表示されます。

- カテゴリが安定性が低いものから高いものの順に並べ替えられている。
- 目標との関連性を最大化するためにカテゴリが結合されている。
- まばらなカテゴリを処理するためにカテゴリが結合されている。
- 目標との関連性の低さにより除外されている。
- 結合後定数項となったため除外されている。

カテゴリ予測フィールドがない場合、脚注が表示されます。

連続型予測フィールド テーブル

連続型フィールド テーブルには、2 つのテーブルがあります。一方には変換に関する次の数値のいずれかが表示されます。

- 標準の単位に変換された予測フィールド値。また、変換された予測フィールドの数、指定された平均値、標準偏差が表示されます。
- 共通範囲にマッピングされた予測フィールド値。また、指定された最小値や最大値のほか、min-max 変換を使用して変換された予測フィールド数も表示されます。
- 分割された予測フィールド値と分割された予測フィールド数。

もう一方のテーブルには、予測領域構築の詳細が、次のような予測フィールド数で表示されます。

- 構築済み。
- 目標との関連性の低さにより除外されている。
- 分割後定数項となったため除外されている。
- 構築後定数項となったため除外されている。

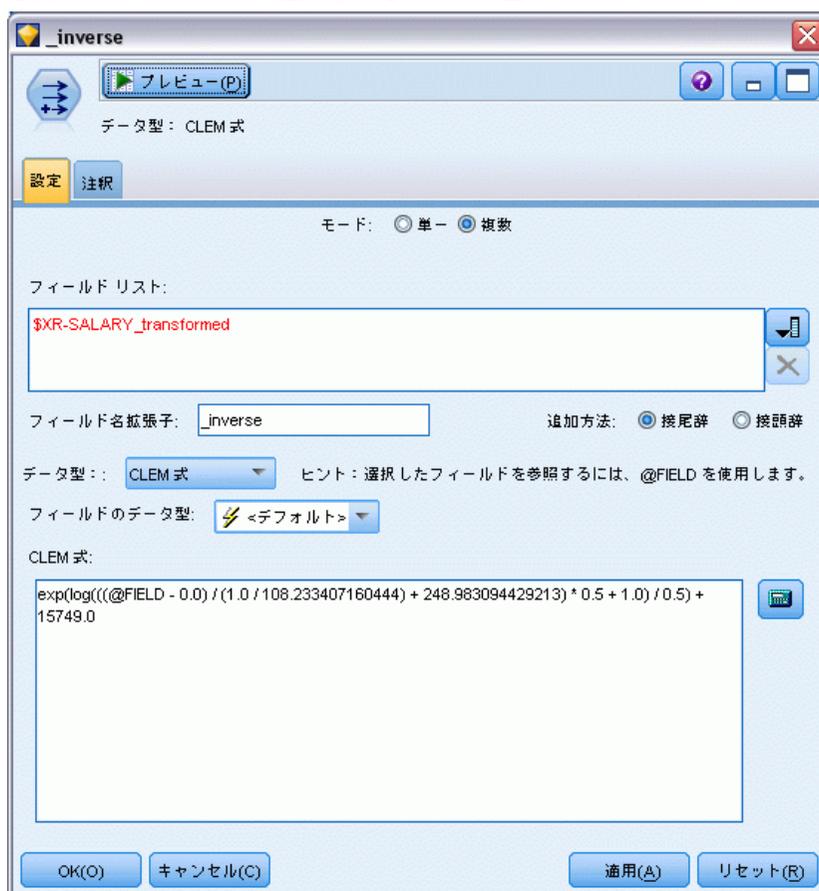
入力となっている連続型予測フィールドがない場合、脚注が表示されます。

フィールド生成ノードの生成

フィールド生成ノードを生成すると、目標の逆変換がスコア フィールドに適用されます。デフォルトで、ノードは自動モデラー（自動分類または自動数値）またはアンサンブル ノードで作成されたスコア フィールドの名前を投入します。尺度型（範囲型）目標が変換されると、スコア フィールドは変換された単位で表示されます（たとえば、\$ ではなく $\log(\$)$ ）。結果を解釈して使用するために、予測値を元の尺度に変換する必要があります。

注： ADP ノードに範囲型目標を再調整する分析が含まれている場合（[入力および目標の順位] パネルで Box-Cox 再調整を選択）のみ、フィールド生成ノードを生成できます。目標が範囲型でない場合、または Box-Cox 再調整が選択されていない場合、フィールド生成ノードは生成できません。

図 4-18
自動データ準備ノードから生成されたフィールド生成ノード



フィールド生成ノードは複数モードで作成され、式に @FIELD を使用して、必要に応じて変換された目標を追加できます。たとえば、次の詳細情報を使用します。

- 対象フィールド名 : 応答
- 変換された対象フィールド名 : response_transformed
- スコア フィールド名 : \$XR-response_transformed

新しいフィールド \$XR-response_transformed_inverse を作成する、フィールド作成ノードを生成します。

注 : 自動モデラーまたはアンサンブル ノードを使用していない場合、フィールド生成ノードを編集して、モデルの適切なスコア フィールドを変換する必要があります。

連続型目標の正規化

デフォルトでは、[入力と目標の準備] パネルで [Box-Cox 変換で連続型目標を再調整する] チェック ボックスをオンにすると、目標を変換し、モデル作成の目標となる新しいフィールドが作成されます。たとえば、元の目標が [response] の場合、新しい目標は [response_transformed] となります。ADP ノードの下流モデルは、この目標を自動的に選択します。

元の目標によっては、この操作で問題が発生する場合があります。たとえば、目標が [年齢] だった場合、新しい目標の値は [年] ではなく、変換されたバージョンの [年] となります。認識可能な単位ではないため、スコアを確認できず解釈もできないということになります。こうした場合、変換された単位を以前の単位に戻す、逆変換を適用できます。このためには、次の手順を実行します。

- ▶ [データを分析] をクリックして、ADP 分析を実行した後、[生成] メニューから [フィールド生成ノード] を選択します。
- ▶ モデル領域のナゲットの後にフィールド生成ノードを投入します。

フィールド生成ノードがスコア フィールドを元の次元に復元し、予測値が元の [年] の値となるようにします。

デフォルトでは、フィールド生成ノードは、アンサンブル化されたモデルの自動モデラーで生成されたスコア フィールドを変換します。個別モデルを作成している場合、フィールド生成ノードを編集して、実際のスコア フィールドから取得する必要があります。モデルを評価する場合、変換済み目標をフィールド生成ノードの [派生元] フィールドに追加する必要があります。これにより、同じ逆変換が目標に適用され、下流の評価ノードまたは分析ノードが、これらのノードをメタデータではなくフィールド名を使用する限り、変換済みデータを適切に使用します。

変換前の名前も復元する場合、存在する場合はフィルタ ノードを使用して、元の目標フィールドを削除し、目標フィールドおよびスコア フィールドの名前を変更できます。

データ型ノード

フィールドのプロパティは、入力ノードまたは、個別のデータ型ノードで指定できます。どちらのノードでも機能は同じです。次のプロパティが使用可能です。

- **フィールド**: IBM® SPSS® Modeler 内のデータの値やフィールド ラベルを指定するには、フィールド名をダブルクリックします。たとえば、IBM® SPSS® Statistics からインポートされるフィールド メタデータを表示したり、データ型ノードで変更したりできます。同様に、フィールドの

新しいラベルとそれらの値を作成できます。データ型ノードで指定したラベルは、[ストリームのプロパティ] ダイアログ ボックスの選択内容に応じて、SPSS Modeler 全体にわたって表示されます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

- **尺度**：測定レベルで、特定フィールドのデータの特性を記述するために使用します。フィールドの詳細がすべてわかっている場合には、完全にインスタンス化済みとされます。詳細は、[p. 153 尺度](#) を参照してください。

注：フィールドの測定レベルは、ストレージ タイプとは異なります。フィールドのストレージは、データが文字列、整数、実数、数値、日付、時間、またはタイムスタンプで保管されることを示しています。

- **値**：この列を使用して、データ セットからデータ値を読み込むオプションを指定したり、[指定] オプションを使用して、測定レベルおよび値を設定する他のダイアログ ボックスを表示することができます。値を読み込まないでフィールドを渡すこともできます。詳細は、[p. 158 データ値](#) を参照してください。
- **欠損値**：フィールドの欠損値の処理方法を指定します。詳細は、[p. 165 欠損値の定義](#) を参照してください。
- **検査**：列には、フィールドの値が指定された値または範囲内に収まっているかどうかを検査するオプションを設定できます。詳細は、[p. 165 データ型の値の検査](#) を参照してください。
- **役割**。フィールドがマシン学習プロセスの [入力] (予測フィールド) または [対象] (予測されるフィールド) のどちらになるかをモデル作成ノードに指示するために、使用されます。[両方] および [なし] も役割として利用できます。さらに、レコードを学習用、検定用、および検証用の独立したサンプルに分割するために使用されるフィールドを示す [データ区分] も利用できます。値 [分割] は各モデルがフィールドの可能な値それぞれに作成されるように設定します。詳細は、[p. 167 フィールドの役割の設定](#) を参照してください。

図 4-19
データ型ノードのオプション



他にもデータ型ノード ウィンドウを使用して指定できるさまざまなオプションがあります。

- データ型ノードがインスタンス化されたら、ツール メニュー ボタンを使って、[単一フィールドを無視] を選択することができます。[単一フィールドを無視] を選択すると、1 つの値だけを持つフィールドが自動的に無視されます。
- データ型ノードがインスタンス化されたら、ツール メニュー ボタンを使って、[ラージセットを無視] を選択することができます。[ラージセットを無視] を選択すると、メンバー数の多いセットが自動的に無視されます。
- データ型ノードがインスタンス化されたら、ツール メニュー ボタンを使って、[連続型整数を順序型に変換] を選択することができます。 [詳細は、 p. 156 連続型データの変換 を参照してください。](#)
- ツール メニュー ボタンを使って、フィルタ ノードを生成して選択したフィールドを除外することができます。
- サングラスの形をしたトグル ボタンを使って、すべてのフィールドのデフォルトを [読み込み順] または [通過] に設定することができます。入力ノードの [データ型] タブのデフォルトでは、フィールドを通過しますが、データ型ノード自体はデフォルトで値を読み込みます。

- **[値の消去]** ボタンを使って、このノード中でフィールド値に対して行った変更内容を消去し（継承されない値）、上流の操作から値を読み込み直すことができます。このオプションは、上流からの特定のフィールドに対して行った変更をリセットする場合に役立ちます。
- **[すべての値の消去]** ボタンを使用して、ノードに読み込まれたすべてのフィールドの値をリセットすることができます。このオプションにより、すべてのフィールドに対して [値] 列に **[読み込み]** が設定されます。このオプションは、すべてのフィールドに対して値をリセットし、上流の操作から値とデータ型を読み込み直す場合に役立ちます。
- **コンテキスト メニュー** を使って、1 つのフィールドから別のフィールドへの属性の **[コピー]** を選択できます。 [詳細は、 p. 169 データ型属性のコピー を参照してください。](#)
- **[未使用のフィールド設定を表示]** オプションを使って、データ中にすでに存在していない、または過去にこのデータ型ノードに接続されていたフィールドのデータ型の設定を表示することができます。これは、変更されたデータ セットのデータ型ノードを再利用する場合に役立ちます。

尺度

尺度（以前の「データ型」または「使用タイプ」）、IBM® SPSS® Modeler におけるデータ フィールドの使用法を記述します。入力ノードまたはデータ型の [データ型] タブで尺度を設定することができます。たとえば、値 1 と 0 をとる整数フィールドの測定レベルをフラグ型に設定することができます。これは、普通 1=True および 0=False の値を取ります。

ストレージと尺度の比較： フィールドの尺度は、ストレージ タイプとは異なります。これは、データが文字列、整数、実数、日付、時間、またはタイムスタンプのどれで保存されるかを示すことに注意してください。データ型ノードを使用してストリームの任意のポイントでデータ型を変更できますが、一方、ストレージは、SPSS Modeler へのデータ読み込時において入力で決定する必要があります。 [詳細は、 2 章 p. 35 フィールドのストレージと形式の設定 を参照してください。](#)

いくつかのモデリング ノードは、それらの [フィールド] タブ上のアイコンによって入力フィールドおよび対象フィールドに対して許可される測定レベルの種類を示します。

測定レベルのアイコン

| アイコン | 尺度 |
|---|---------|
|  | Default |
|  | 連続型 |

| アイコン | 尺度 |
|---|----------|
|  | カテゴリ |
|  | Flag |
|  | 名義 |
|  | 順序 |
|  | Typeless |

使用できる測定レベルは、次の通りです。

- **デフォルト**: ストレージ タイプおよび値が不明なデータ（たとえば、まだ読み込まれていないため）は、[<デフォルト>] と表示されます。
- **連続**: 0-100 や 0.75-1.25 のように、数値の範囲を記述するために用いられます。連続値は、整数、実数、または日付/時間になります。
- **カテゴリ**: 実際の DISTINCT 値の数字が不明な場合に、文字列値に用いられます。これは**インスタンス化されていないデータ型**で、すべてのストレージに関する情報やデータの使用方法に関する情報がわかっていないことを示します。データが読み込まれると、[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た名義型フィールドの最大メンバー数に応じて、データ型は **フラグ型**、**名義型**、または**データ型不明**になります。
- **フラグ型**: true と false、Yes と No または 0 と 1 など、特性の有無を示す 2 つの値を持つデータに使用されます。使用される値は異なる場合がありますが、一方の値は常に「true」値として、もう一方の値は「false」値として割り当てられる必要があります。データは、テキスト、整数、実数、日付/時間、またはタイムスタンプを表します。
- **名義**: 複数の DISTINCT 値を持つデータを記述するために用いられ、それぞれが「small/medium/large」のようなセットのメンバーとして扱われます。本バージョンの - では、名義型には任意のストレージ（数値、文字列、または日付/時間）を利用できます。測定レベルを名義型にしても、自動的に値が文字列に変わることはないことに注意してください。
- **順序**: 固有の順位のある複数の DISTINCT 値を持つデータを記述するために用いられます。たとえば、給与区分や満足度ランキングは、順序型データとして類別できます。順序は、データ要素の普通のソート順により定義されます。たとえば、1,3,5 は一連の整数のデフォルトのソート順ですが、HIGH, LOW, NORMAL（アルファベットの昇順）は、文字列の順序です。可視化、モデル設定および IBM® SPSS® Statistics など順位データを DISTINCT 型として認識する他のアプリケーションへの出力するため、順序型の測定レベルによりカテゴリ データ セットを順位データとして定義できます。順序型フィールドは、名義型フィールドを使用できる場合はいつでも使用できます。また、任意の

ストレージ型（実数、整数、文字列、日付、時間など）のフィールドは順序型として定義できます。

- **データ型不明**：上記のいずれのデータ型にもあてはまらない、値が1つのフィールド、または定義した最大値より大量のメンバーがある名義型のデータに用いられます。この測定レベルは、不明にしないと、データ型が多くメンバー（アカウント番号など）を使って設定されてしまうような場合に効果的です。フィールドに**データ型不明**を選択すると、方向が自動的に**なし**に設定され、**レコードID**が唯一の代替となります。デフォルトのセットの最大サイズは、250の一意な値です。この数値は、**ツール**メニューの**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**タブで調整または無効化することができます。

手作業で尺度を指定するか、またはソフトウェアにデータを読み込ませ、その値に基づいて尺度を判断させることができます。

また、カテゴリ データとして処理する必要のあるいくつかの連続型データフィールドがある場合、それらを変換するオプションを選択することができます。 [詳細は、 p.156 連続型データの変換 を参照してください。](#)

自動入力を使用するには

- ▶ データ型ノードまたは入力ノードの**データ型**タブのいずれかで、すべてのフィールドに対して**値**列を**読み込み**に設定します。これで、下流にあるすべてのノードでメタデータが利用できるようになります。ダイアログボックスにある**サングラス**ボタンを使えば、すべてのフィールドを簡単に**読み込み**または**通過**に設定することができます。
- ▶ **値の読み込み** ボタンをクリックして、データ ソースから直接値を読み込むこともできます。

フィールドの尺度を手作業で設定するには

- ▶ テーブル中のフィールドを選択します。
- ▶ **尺度**列のドロップダウン リストで、フィールドの尺度を選択します。
- ▶ Ctrl-A または Ctrl キーを押しながらクリックして、複数のフィールドを選択してから、ドロップダウン リストで尺度を選択することもできます。

図 4-20
測定レベルの手動設定



連続型データの変換

カテゴリ データを連続型として処理すると、データの品質に重大な影響があります。たとえば、それが対象フィールドである場合、2 値モデルでなく回帰モデルを作成します。この影響を回避するために、整数の範囲を、順序型またはフラグ型などのカテゴリ型に変換できます。

- ▶ 道具の記号の付いた [操作および生成] メニュー ボタンから、[連続型を順序型に変換] を選択します。[変換値] ダイアログが表示されます。

図 4-21
[変換値] ダイアログ



- ▶ 自動的に変換される範囲のサイズを指定します。これは、入力したサイズまでの範囲に適用されます。
- ▶ [OK] をクリックします。影響を受ける範囲型は、フラグ型または順序型に変換され、データ型ノードの [データ型] タブに表示されます。

変換の結果

- 整数のストレージを持つ連続型フィールドを順序型に変更すると、下限値および上限値を拡張して、下限値から上限値までの整数値のすべてを含みます。たとえば、範囲が 1 ~ 5 の場合、値のセットは 1、2、3、4、5 です。
- 連続型フィールドをフラグ型フィールドに変更すると、下限値および上限値はフラグ型フィールドの偽の値および真の値となります。

インスタンス化とは？

インスタンス化は、データ フィールドのストレージ タイプや値などの情報を読み込む、または指定するプロセスです。システム リソースを最適化するために、インスタンス化を行う作業はユーザーが指示する必要があります。入力ノードの [データ型] タブでオプションを指定するか、またはデータ型ノードにデータを流すことによって、ソフトウェアに値の読み込みを指示します。

- 不明なデータ型のデータは、**インスタンス化されていないデータ**とも呼ばれます。ストレージ タイプと値が不明なデータ型は、[データ型] タブの [尺度] 列に [＜デフォルト＞] と表示されます。
- 文字列や数値などフィールドのストレージに関する一部の情報がわかっている場合、そのデータは「**部分的にインスタンス化されている**」と言います。**カテゴリ**や**連続型**は、部分的にインスタンス化された尺度です。たとえば、**カテゴリ**はフィールドがシンボル型であること（ただしそれが名義型、順序型またはフラグ型のどちらであるかわからないこと）を示します。
- データ型の詳細が、値も含めてすべて明らかな場合、この列には**完全にインスタンス化された尺度**（名義型、順序型、フラグ型、または連続型）が表示されます。注：連続型は、部分的にインスタンス化されたデータ フィールドと、完全にインスタンス化されたフィールドの両方で使用されます。連続型データは、整数または実数になります。

データ型ノードでデータ ストリームを実行中、インスタンス化されていないデータ型は、初期のデータ値を基にして、部分的にインスタンス化されます。すべてのデータがノードを通過したら、それらのデータは値が [＜通過＞] に設定されている場合を除いて完全にインスタンス化されます。実行が中断された場合は、データは部分的にインスタンス化されたままになります。[データ型] タブがインスタンス化されたら、ストリーム中のその時点でフィールドの値は固定化されます。つまり、ストリームを再実行した場合も、上流の変更は特定のフィールドの値に影響しないということです。新しいデータや追加の操作に基づいて、値を変更または更新するには、[データ型] タブで変更するか、またはフィールドの値を [＜読み込み＞] または [＜読み込み +>] に設定する必要があります。

インスタンス化する場合

一般的に、データ セットがさほど大きくなく、後でストリームにフィールドを追加する予定がない場合は、入力ノードでインスタンス化するのが便利です。ただし、次の場合には、別のデータ型ノードでインスタンス化するほうが便利です。

- データ セットが巨大で、ストリームがデータ型ノードの前でサブセットをフィルタリングしている場合。
- ストリーム中でデータをフィルタリングしている場合。
- ストリーム中でデータが結合または追加されている場合。
- 処理の過程で新しいデータ フィールドが作成される場合。

データ値

[データ型] タブの [値] 列を使って、データから値を自動的に読み込んだり、個別のダイアログ ボックスで尺度と値を指定することができます。

図 4-22
データ値の読み込み、通過、または指定方法の選択



このドロップダウン リストで利用できるオプションを利用して、自動入力に関する次の指示を行うことができます。

| オプション | 関数 |
|-----------|--|
| <Read> | ノードの実行時にデータが読み込まれます。 |
| <読み込み+> | データが読み込まれ、現在のデータ（存在している場合）に追加されます。 |
| <Pass> | データは読み込まれません。 |
| <Current> | 現在のデータ値を保持します。 |
| 指定... | 値や尺度オプションを指定するための個別のダイアログ ボックスが表示されます。 |

データ型ノードを実行するか、[値の読み込み] をクリックすると、選択内容に応じてデータ ソースから値が自動的に読み込まれ、データ型が判断されます。これらの値は、[指定] オプションを使うか、または [フィールド] 列のセルをダブル クリックして、手作業で入力することもできます。

データ型ノードでフィールドを変更した場合、ダイアログ ボックスのツールバーにある次のボタンを使って、値情報をリセットすることができます。

- [値の消去] ボタンを使って、このノード中でフィールド値に対して行った変更内容を消去し（継承されない値）、上流の操作から値を読み込み直すことができます。このオプションは、上流からの特定のフィールドに対して行った変更をリセットする場合に役立ちます。
- [すべての値の消去] ボタンを使用して、ノードに読み込まれたすべてのフィールドの値をリセットすることができます。このオプションにより、すべてのフィールドに対して [値] 列に [読み込み] が設定されます。このオプションは、すべてのフィールドに対して値をリセットし、上流の操作から値と尺度を読み込み直す場合に役立ちます。

[値] ダイアログ ボックスの使用

[データ型] タブの [値] または [欠損値] 列をクリックすると、事前定義された値のドロップダウン リストが表示されます。このリストの [指定] オプションを選択すると個別のダイアログ ボックスが表示され、選択したフィールドの値の読み込み、指定、ラベル付け、処理のオプションを設定できます。

図 4-23
データ値のオプションの設定

薬品値

尺度: ストレージ: モデルフィールド...

値: データから読み込み 通過
 値とラベルを指定

| 値 | ラベル |
|-----|-----|
| 薬品A | |
| 薬品B | |
| 薬品C | |
| 薬品X | |
| 空白 | |

データから値を拡張 最大文字列長:

値の検査:

空白を定義

欠損値

範囲 ~:

ヌル 空白文字

説明:

OK(O) キャンセル(C) ヘルプ(H)

コントロールの大半は、すべての種類のデータに共通しています。ここでは、これらの共通のコントロールを説明していきます。

尺度: 現在選択されている測定レベルを表示します。データ利用目的に応じて設定を変更することができます。たとえば、`day_of_week` というフィールドに個別の曜日を表す数字が格納されている場合に、各カテゴリを個別に調査する棒グラフ ノードを作成するために、名義型データに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、選択した尺度の影響を受けません。ストレージ タイプを変更するには、可変長ノードまたは固定長ノードの [データ] タブを使用するか、または置換ノードの変換関数を使用します。

モデル フィールド: モデル ナゲットのスコアリングの結果として生成されたフィールドの場合、モデル フィールドの詳細も表示することができます。詳細には、モデル作成時のフィールドの役割（予測値、確率、傾向など）や対象フィールド名も含まれています。

値： 選択したフィールドの値を決める方法を選択します。ここで選択した内容は、[データ型ノード] ダイアログ ボックスの [値] 列で行った選択内容に優先します。値の読み込みに関する選択項目には、次のようなものがあります。

- **データから読み込み：** ノードの実行時に値を読み込む場合に選択します。このオプションは、[読み込み] と同じです。
- **通過：** 現在のフィールドのデータを読み込まない場合に選択します。このオプションは、[通過] と同じです。
- **値とラベルの指定：** このオプションは選択したフィールドの値とラベルを指定するために使用します。現在のフィールドに対する知識に基づいて値を指定します。このオプションは、値の検査とともに用いられます。このオプションを選択すると、フィールドの種類に応じた独自のコントロールが有効になります。値とラベルのオプションは、以降の各項目で個別に説明しています。注： 尺度がデータ型不明または<デフォルト>のフィールドに対して値またはラベルを指定することはできません。
- **データから値を拡張：** 現在のデータに、ここで入力した値を追加する場合に選択します。たとえば、field_1 の範囲が (0, 10) の場合に、値の範囲として (8, 16) を入力した場合、元の最小値を除去せずに 16 を追加して範囲が拡張されます。新しい範囲は (0, 16) になります。このオプションを選択すると、自動入力オプションが自動的に [読み込み+] に設定されます。

値の検査： 値が指定した連続型、フラグ型、または名義型の規則にしたがっているかどうかの検査方法を選択します。このオプションは、[データ型ノード] ダイアログ ボックスの [検査] 列に対応しており、ここでの設定内容がダイアログ ボックスでの設定に優先されます。[値を指定] オプションとともに用いられるこのオプションを利用すれば、データ中の値が予期した値かどうかを検査することができます。たとえば、値に 1, 0 と指定した後、検査オプションを使用して、1 と 0 以外のすべての値を持つレコードを破棄することができます。

空白を定義： データの欠損値や空白値を定義する場合に選択します。

- **欠損値テーブル：** 特定の値 (99 または 0 など) を空白として定義できます。この値は、フィールドのストレージ タイプに適切なものでなければなりません。
- **範囲：** 欠損値の範囲を指定する目的で使用されます。たとえば、1 - 17 歳、および 65 歳を上回る年齢です。境界値が空欄のままの場合、その範囲は無制限になります。たとえば、下限に 100 が指定されていて上限がない場合、100 以上のすべての値は欠損値になります。各境界値は範囲にふくまれます。たとえば、下限が 5 で上限が 10 の範囲は、範囲定義に 5 と 10 も含まれます。欠損値範囲は、任意のストレージ タイプに対して定義できます。これには、日付/時刻および

文字列もふくまれます（値が範囲内かどうかを決定するために、アルファベット順のソート順序が使用されます）。

- **ヌル/空白文字**：システムのヌル値（データ中で \$null\$ と表示）および空白文字（表示されない文字を持つ値）を空白値として定義できます。内部で別に格納され、特定のケースで別に処理されるにもかかわらず、データ型ノードも空白文字列のような空の文字列を分析のために扱います。詳細は、6 章 欠損値の概要 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。

注：未定義または \$null\$ として空白をコード化するには、置換ノードを使用する必要があります。

説明：このテキスト ボックスでフィールド ラベルを指定します。ラベルは、[ストリームのプロパティ] ダイアログ ボックスの選択内容に応じて、グラフ、テーブル、出力およびモデル ブラウザなどの様々な場所で表示されます。

連続型データの値およびラベルの指定

連続型尺度は数値型フィールドに使用されます。連続型ノードには、3 種類のストレージ タイプがあります。

- Real
- Integer
- 日付と時刻

これらの 3 種類の連続型フィールドを編集するには、同じダイアログ ボックスが使用されます。ただし、ストレージ タイプは参照用のみ表示されます。

図 4-24
連続型の値とラベルを指定するオプション

尺度: ストレージ:

値: データから読み込み 通過
 値とラベルを指定

下限:

上限:

値の指定

次のコントロールは連続型フィールドに固有のもので、値の範囲を指定するのに使用します：

下限：値の範囲の下限を指定します。

上限 :値の範囲の上限を指定します。

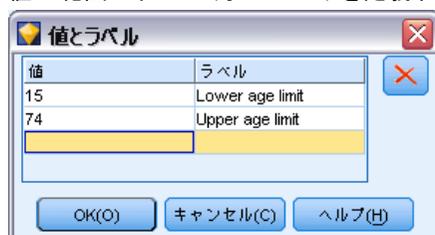
ラベルの指定

集計範囲フィールドの任意の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。

値とラベル サブダイアログ ボックス

集計範囲フィールドの [値を指定] ダイアログ ボックスの [ラベル] をクリックして、範囲中の任意の値のラベルを指定できる新しいダイアログ ボックスを開きます。

図 4-25
値の範囲のラベル (オプション) を定義する



このテーブルの [値] と [ラベル] 列で値とラベルのペアを定義できます。現在定義されているペアがここに表示されます。空のセルをクリックして値と対応するラベルを入力すると、新しいラベルのペアを追加できます。注：このテーブルに値/値とラベルのペアを追加しても、フィールドに新しい値は追加されません。その代わりに、フィールド値のメタデータが作成されるだけです。

データ型ノードで指定したラベルは、[ストリームのプロパティ] ダイアログ ボックスの選択内容に応じて、ツールヒントや出力ラベルなどとしてさまざまな場所に表示されます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

数値型データおよび順序型の名前とラベルの指定

名義型 (セットが他) および順序型 (順序セット型) の尺度は、データの値がセットのメンバーとして個別に使われることを表しています。セット型では、文字列、整数、実数、または日付/時刻のストレージタイプを利用することができます。

図 4-26
名義型の値とラベルを指定するオプション

尺度: ストレージ:

値: データから読み込み 通過
 値とラベルを指定

| 値 | ラベル |
|-----|---------------------|
| 薬品A | Lisinopril |
| 薬品B | Metoprolol |
| 薬品C | Hydrochlorothiazide |
| 薬品X | Amlodipine |
| 薬品Y | |

次のコントロールは名義型フィールドおよび順序型フィールドに固有のもので、値とラベルを指定するのに使用します。

値: テーブルの [値] 列により、現在のフィールドに対する知識に基づいて値を指定できます。このテーブルを使ってフィールドに期待値を入力し、[値の検査] ドロップダウン リストを使ってこれらの値に対するデータセットの整合性を検査することができます。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。

ラベル: [ラベル] 列によりセット内の各値にラベルを指定できます。ラベルは、[ストリームのプロパティ] ダイアログ ボックスの選択内容に応じて、グラフ、テーブル、出力およびモデル ブラウザなどの様々な場所で表示されます。詳細は、5 章 [ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

フラグ型の値の指定

フラグ型フィールドは、2 つの DISTINCT 値を持つデータを表示するために用いられます。フラグ型では、文字列、整数、実数、または日付/時刻のストレージ タイプを利用することができます。

図 4-27
フラグ型フィールドの値の指定用オプション

尺度: ストレージ:

値: データから読み込み 通過
 値とラベルを指定

真: ラベル:

偽: ラベル:

真 (true): 条件を満たす場合のフィールドのフラグ値を指定します。

偽 (false): 条件を満たさない場合のフィールドのフラグ値を指定します。

ラベル: フラグ型フィールドの各値のラベルを指定します。ラベルは、[ストリームのプロパティ] ダイアログ ボックスの選択内容に応じて、グラフ、テーブル、出力およびモデル ブラウザなどの様々な場所で表示されます。詳細は、5 章 [ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

欠損値の定義

[データ型] タブの [欠損値] 列は、欠損値の処理がフィールドに定義されているかどうかを示します。次のように設定できます。

オン(*): 欠損値の処理がこのフィールドに定義されていることを示します。下流の置換ノードを使用して、または [指定] オプションを使用した明示的な指定によって (下記参照) 行うことができます。

オフ: フィールドに欠損値の処理が定義されていません。

指定: このフィールドで欠損値として見なされる明示的な値を宣言できるダイアログを表示します。

詳細は、6 章 [欠損値の概要 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

データ型の値の検査

各フィールドの [検査] オプションを有効にすると、そのフィールド中のすべての値を調べて、値が現在のデータ型または [値を指定] ダイアログ ボックスで指定した値に適合しているかどうかを検査されます。これは、1 回の操作で、データ セットをクリーン アップして、データ セットのサイズを減らすために役立ちます。

図 4-28
選択したフィールドの検査オプションの選択



[データ型ノード] ダイアログ ボックス内の [検査] 列の設定によって、データ型、I制限を超えた値が検出されたときの処理方法が決まります。フィールドの検査の設定を変更するには、そのフィールドの [検査] 列にあるドロップダウン リストを使用します。すべてのフィールドに対して検査を行うには、[フィールド] 列をクリックして Ctrl-A を押します。次に [検査] 列中の任意のフィールドのドロップダウン リストを使用してください。

[検査] では、次の設定を行うことができます。

なし：値は検査されずに通過します。これは、デフォルトの設定です。

無効：制限外の値をシステムのヌル値 (\$null\$) に変更します。

強制：完全にインスタンス化された尺度のフィールドに対して、その値が指定した範囲を超えていないかが検査されます。指定範囲外の値は、次のルールに基づいて各尺度における有効な値に変換されます。

- フラグ型の場合、真 (true) または偽 (false) 以外の値は、偽 (false) の値に変換されます。
- セット型 (名義型または順序型) の場合は、すべての未知の値は、セットの値の最初のメンバーに変換されます。
- 範囲の上限より大きい数値は、上限値に置き換えられます。

- 範囲の下限より小さい数値は、下限値に置き換えられます。
- 範囲が設定されているにもかかわらず、ヌル値が検出された場合は、その範囲の中間の値が指定されます。

破棄： 不正な値が検出された場合は、レコード全体が破棄されます。

警告： すべてのデータを読み込む際に、不正な項目数がカウントされ、件数が [ストリームのプロパティ] ダイアログ ボックスにレポートされます。

中止： 最初に不正な値が検出された時点で、ストリームの実行が終了します。エラーは、[ストリームのプロパティ] ダイアログ ボックスに報告されます。

フィールドの役割の設定

フィールドの役割は、フィールドがモデル構築でどのように使用されるかを指定します。たとえば、フィールドが入力か対象（予測対象）か、などです。

注：データ区分、度数、レコード ID の役割は、それぞれ 1 つのフィールドだけに適用できます。

図 4-29
データ型ノードのフィールドの役割オプションの設定



次の役割があります。

入力： このフィールドは、マシン学習に対する入力（予測変数フィールド）として使用されます。

対象: このフィールドは、マシン学習の出力または対象（モデルが予測しようとするフィールドの 1 つ）として使用されます。

両方: このフィールドは、Apriori ノードで入力と出力の両方として使用されます。他のすべてのモデル作成ノードでは、このフィールドは無視されます。

なし: このフィールドはマシン学習では無視されます。尺度が [データ型不明] に設定されているフィールドは、自動的に [役割] 列が [なし] に設定されます。

データ区分: データを、学習、テスト、および（オプションで）検証の各目的用の異なるサンプルに分割するためのフィールドを示します。このフィールドは、([フィールド値] ダイアログ ボックスで定義されているように) 2 個 または 3 個の値を取るセット型としてインスタンス化されなければなりません。最初の値は、学習用サンプル、2 番目の値はテスト用サンプル、3 番目に値がある場合は、検証用サンプルを表しています。その他の値は無視され、フラグ フィールドを使用できません。分析でデータ区分を使用するには、適切なモデル構築または分析ノードの [モデルのオプション] タブでデータ分割を有効にする必要があることに注意してください。データ分割が有効になっている場合、データ区分フィールドではヌル値をもつレコードは無視されます。複数のデータ区分フィールドがストリーム内で定義されていた場合、該当する各モデリング ノードの [フィールド] タブで単一のデータ区分フィールドを指定する必要があります。使用するデータ中に適切なフィールドがまだ存在していない場合、データ区分ノードまたはフィールド作成ノードを使用すると新規に作成できます。詳細は、[p. 229 データ区分ノード](#) を参照してください。

分割: (名義型、順序型、フラグ型フィールドのみ) フィールドの可能な値それぞれにモデルが作成されるよう指定します。詳細は、[3 章 分割モデルの作成 in IBM SPSS Modeler 15 モデル作成ノード](#) を参照してください。

度数: (数値型フィールドのみ) この役割を設定すると、フィールド値をレコードの度数の重みの因子として使用できます。この機能は C&R ツリー、CHAID、QUEST および線型モデルにのみサポートされます。他のすべてのノードはこの役割を無視します。度数の重みは、この機能をサポートするモデル作成ノードの [フィールド] タブで [度数の重みを使用] オプションを設定すると使用できます。詳細は、[3 章 度数フィールドと重みフィールドの使用 in IBM SPSS Modeler 15 モデル作成ノード](#) を参照してください。

レコード ID: 一意のレコード ID として使用されます。この機能は多くのノードによって無視されますが、線型モデルによってサポートされており、IBM Netezza データベース内マイニング ノードに必要です。

データ型属性のコピー

値、検査オプション、および欠損値などのデータ型の属性を、あるフィールドから別のフィールドに簡単にコピーすることができます。

- ▶ 属性をコピーするフィールドを右クリックします。
- ▶ コンテキスト メニューから、[コピー] を選択します。
- ▶ 属性を変更するフィールドを右クリックします。
- ▶ コンテキスト メニューから、[形式を選択して貼り付け] を選択します。注: Ctrl キーを押しながらクリックするか、またはコンテキスト メニューの [フィールドの選択] オプションを使用することにより、複数のフィールドを選択することもできます。

新しいダイアログ ボックスが表示されます。ここから、貼り付ける特定の属性を選択することができます。複数のフィールドに貼り付ける場合は、ここで選択したオプションがすべての対象フィールドに適用されます。

次の属性を貼り付け: あるフィールドから別のフィールドに貼り付ける属性を、下のリストから選択します。

- **データ型:** 尺度貼り付ける場合に選択します。
- **値:** フィールドの値を貼り付ける場合に選択します。
- **欠損値:** 欠損値の設定を貼り付ける場合に選択します。
- **検査:** 値の検査オプションを貼り付ける場合に選択します。
- **役割:** フィールドの役割を貼り付ける場合に選択します。

フィールド形式の [設定] タブ

図 4-30
データ型ノードの [フォーマット] タブ



[テーブル] 中の [フォーマット] タブとデータ型ノードは、現在のまたは未使用フィールドのリストおよび各フィールドの書式設定オプションを示します。フィールド フォーマット テーブルの各列の説明を次に示します。

フィールド： 選択したフィールド名が表示されます。

書式： この列のセルをダブルクリックすると、ダイアログ ボックスが表示され、個別にフィールドの書式設定を指定できます。詳細は、[p. 171 フィールド形式のオプションの設定](#) を参照してください。ここで指定した書式設定は、ストリームのプロパティ全体で指定した書式設定に上書きされます。

注： Statistics エクスポートと Statistics 出力ノードは、メタデータにフィールドごとの書式設定を含む .sav ファイルを出力します。IBM® SPSS® Statistics.sav ファイル形式でサポートされない設定がフィールドごとの書式に指定された場合、ノードは SPSS Statistics のデフォルト形式を使用します。

表示位置： この列を使用して、テーブルの列内で値をどのように表示するかを指定します。デフォルトの設定は [自動] で、シンボル値を左寄せ、数値を右寄せで表示します。デフォルト値を変更するには、[左]、[右]、または [中央] を選択します。

列幅： デフォルトでは、フィールドの値に基づいて列幅が自動的に計算されます。自分で値を指定するには、テーブルセルをクリックしてドロップダウンリストから新規の幅を選択します。ここに表示されない任意の幅を入力するには、[フィールド] または [フォーマット] 列のテーブルセルをダブルクリックして、フィールド設定のサブダイアログボックスを開きます。代わりに、セルを右クリック、μで [フォーマットの設定] を選択することもできます。

現在のフィールドを表示： デフォルトでは、このダイアログボックスには現在アクティブなフィールドが表示されます。使われていないフィールドを表示する場合は、代わりに [未使用のフィールド設定を表示] を選択します。

コンテキストメニュー： このタブで利用できるコンテキストメニューには、さまざまな選択項目やオプションが用意されています。

- **すべて選択：** すべてのフィールドを選択します。
- **選択解除：** 選択を解除します。
- **フィールドの選択：** データ型またはストレージの種類を基準にしてフィールドを選択します。選択できるオプションには、[カテゴリ型を選択]、[連続型を選択] (数値)、[データ型不明の選択]、[文字列の選択]、[数字の選択]、または [日付/時間の選択] があります。 [詳細は、 p.153 尺度 を参照してください。](#)
- **フォーマットの設定：** フィールドごとに日付、時刻や小数のオプションを指定するには、サブダイアログボックスを表示します。
- **表示位置の設定：** 選択したフィールドの表示位置を設定します。[自動]、[中央]、[左]、または [右] を選択することができます。
- **列幅の設定：** 選択したフィールドのフィールド幅を設定します。データから幅を読み込むには [自動判別] を指定します。あるいはフィールド幅を 5、10、20、30、50、100 または 200 に設定することもできます。

フィールド形式のオプションの設定

フィールド形式は、データ型またはテーブルノード [フォーマット] タブからサブダイアログボックスで指定します。このダイアログボックスの表示前に複数のフィールドを選択した場合、選択した最初のフィールドの設定がすべてに使用されます。ここで指定した後 [OK] をクリックして、これらの設定を [フォーマット] タブで選択したすべてのフィールドに適用します。

図 4-31
1 つ以上のフィールドにフォーマット オプションを設定する



次の各オプションがフィールドごとに利用可能です。これらの設定の多くは、[ストリームのプロパティ] ダイアログ ボックスでも設定できます。詳細は、5 章 [ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。フィールド レベルの設定は、ストリームのデフォルト設定をオーバーライドします。

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。

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）。

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、またはエクスポートするときに使用する、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。このエクスポート フォーマットは実数型のストレージを持つフィールドにのみ適用されます。

表示位置：列内で値をどのように表示するかを指定します。デフォルトの設定は[自動]で、シンボル値を左寄せ、数値を右寄せで表示します。デフォルト値を変更するには、[左]、[右]、または[中央]を選択します。

列幅：デフォルトでは、フィールドの値に基づいて列幅が自動的に計算されます。リスト ボックスの右にある矢印を使用して、5 つの間隔における任意の幅を指定できます。

フィールドのフィルタリングまたは名前の変更

ストリーム内の任意の場所でフィールドの名前変更またはフィールドを除外することができます。たとえば、患者（レコード レベル データ）の カリウム値（フィールド レベル データ）を重要視していない場合、[K]（カリウム値）フィールドを除外できます。個々のフィルタ ノード、または入力ノードか出力ノードの [フィルタ] タブを使用して実行することができます。どのノードからアクセスしているかに関係なく、機能は同じです。

- 可変長、固定長、および Statistics および XML ファイルなどの入力ノードから、IBM® SPSS® Modeler にデータを読み込むフィールドの名前を変更したり、フィルタリングを行うことができます。
- フィルタ ノードを使って、ストリーム中の任意の場所でフィールドの名前を変更したり、フィルタリングすることができます。
- Statistics エクスポート ノード、Statistics 変換ノード、Statistics モデル ノードおよび Statistics 出力ノードから、IBM® SPSS® Statistics の命名規則に従ったフィールドをフィルタリングしたり、名前を変更することができます。 [詳細は、8 章 p.563 IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング を参照してください。](#)
- 上記のノードいずれかの [フィルタ] タブを使用して、複数のレスポンス セットを定義または編集することができます。 [詳細は、p.177 複数レスポンス セット編集 を参照してください。](#)
- 最後にフィルタ ノードを使用して、ある入力ノードから別のノードへフィールドをマップすることができます。 [詳細は、5 章 データ ストリームのマッピング in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。](#)

図 4-32
フィルタ ノードのオプションの設定



フィルタリング オプションの設定

[フィルタ] タブ中のテーブルには、入力時の各フィールド名、および出力時の各フィールド名が表示されます。このテーブル中のオプションを使って、重複しているフィールドや、下流の操作で不要なフィールドの、フィールド名を変更したり、フィールドをフィルタリングすることができます。

- **フィールド**: 現在接続しているデータ ソースからの入力フィールドを表示します。
- **フィルタ**: すべての入力フィールドのフィルタ ステータスが表示されます。フィルタリングされているフィールドはこの列に、フィールドが下流に渡されないことを示す赤い X 印が表示されます。選択したフィールドの [フィルタ] 列をクリックすると、フィルタリングの設定と解除を行えます。また、Shift キーを押しながら選択するところにより、複数のフィールドのオプションを同時に設定することができます。
- **フィールド**: フィルタ ノードを離れたフィールドを表示します。重複した名前は赤で表示されます。フィールド名を編集するには、この列をクリックして新しい名前を入力します。または、[フィルタ] 列をクリックして重複するフィールドを無効にすることにより、フィールドを除去することもできます。

テーブル中のすべての列は、列見出しをクリックしてソートすることができます。

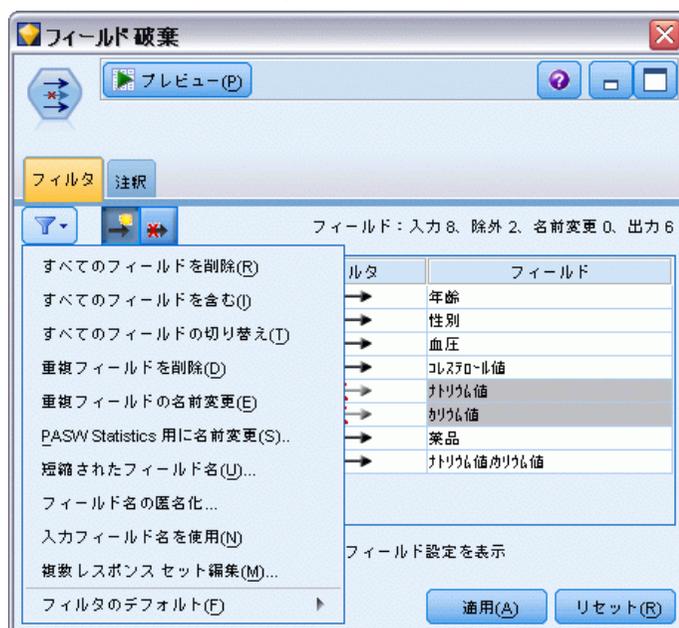
現在のフィールドを表示 : フィルタ ノードに接続しているデータ セットのフィールドを表示する場合に選択します。このオプションはデフォルトで選択されており、フィルタ ノードを使用するもっとも一般的な方法となっています。

未使用のフィールド設定を表示 : フィルタ ノードに接続したことがある（今は接続していない）データ セットのフィールドを表示する場合に選択します。このオプションは、フィルタ ノードをあるストリームから別のストリームにコピーしたり、フィルタ ノードを保存して再ロードするような場合に適しています。

[フィルタ] ボタンのメニュー

ダイアログ ボックスの左上にある [フィルタ] ボタンをクリックして、多くのショートカットやその他のオプションを提供するメニューにアクセスします。

図 4-33
フィルタ ノードのメニュー オプション



次の処理を選択できます。

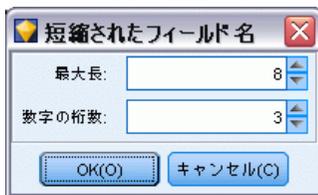
- すべてのフィールドを除去
- すべてのフィールドを含める
- すべてのフィールドの切り替え
- 重複を除去注：このオプションを選択すると、最初のフィールドも含めてすべての重複名が除去されます。

- その他のアプリケーションに準拠させるため、フィールドおよび複数のレスポンスセットの名前を変更 [詳細は、8 章 p.563 IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング を参照してください。](#)
- フィールド名の短縮
- フィールドおよび複数レスポンス セット名の匿名化
- 入力フィールド名を使用
- 複数レスポンス セット編集 [詳細は、p.177 複数レスポンス セット編集 を参照してください。](#)
- デフォルト フィルタ ステートの設定

ダイアログ ボックスの上部にある矢印のトグル ボタンを使用して、デフォルトでフィールドを含める、またはフィールドを破棄するかを指定することもできます。これは、大きいデータ セットで、ほんのわずかなフィールドだけが下流で必要な場合に役立ちます。たとえば保持しておきたいフィールドのみを選択して、その他のすべてのフィールドを（個別にではなく、破棄するフィールドをすべて選択して）破棄するよう指定できます。

フィールド名の短縮

図 4-34
フィールド名の短縮ダイアログ ボックス



[フィルタ] ボタンのメニュー ([フィルタ] タブの左上部) から、フィールド名の短縮を選択できます。

最大長： フィールド名の長さを \$ 限するための文字数を指定します。

数字の桁数： フィールド名を短縮するとフィールド名が重複する場合は、名前をさらに短縮して数字を追加することにより区別します。使用する数字の桁数を指定できます。矢印ボタンを使って、数を調節してください。

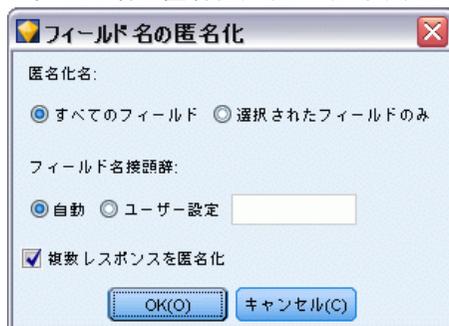
たとえば、次の表は、デフォルトの設定（最大文字数 = 8 および数字の桁数 = 2）を使用すると医療データ セット内のフィールド名がどのように短縮されるのかを示しています。

| フィールド名 | 短縮されたフィールド名 |
|-----------------|-------------|
| Patient Input 1 | Patien01 |
| Patient Input 2 | Patien02 |

| フィールド名 | 短縮されたフィールド名 |
|------------|-------------|
| Heart Rate | HeartRat |
| BP | BP |

フィールド名の匿名化

図 4-35
フィールド名の匿名化ダイアログ ボックス



[フィルタ] タブの左上部にある [フィルタ] ボタン メニューをクリックして [フィールド名の匿名化] を選択し、[フィルタ] タブを含むノードのフィールド名を匿名化します。匿名化したフィールド名は、一意の数字の基づいた値の接頭辞を含む文字列から成り立っています。

匿名化名: [選択したフィールドのみ] を選択し、[フィルタ] タブですでに選択したフィールド名のみを匿名化します。デフォルトは [すべてのフィールド] で、全フィールド名を匿名化します。

フィールド名接頭辞: 匿名化したフィールド名に対するデフォルトの接頭辞は anon_ です。異なる接頭辞を希望する場合、[ユーザー設定] を選択し、独自の接頭辞を入力します。

複数レスポンス セットを匿名化: フィールドと同じ方法で、複数レスポンスセットの名前を匿名化します。詳細は、[p. 177 複数レスポンス セット編集](#) を参照してください。

元のフィールド名で保存するには、[フィルタ] ボタン メニューから [入力したフィールド名を使用] を選択します。

複数レスポンス セット編集

[フィルタ] タブの左上部にある [フィルタ] ボタン メニューをクリックして [フィールド名の匿名化] を選択し、[フィルタ] タブを含むノードのフィールド名を匿名化します。

図 4-36
[複数レスポンス セット] ダイアログ ボックス



たとえば、どの博物館に行くのか、またはどの雑誌を読むのかに関する調査の回答を求める場合など、複数レスポンス セットを使用して、各ケースに複数の値を持つデータを記録します。複数レスポンス セットは、Data Collection 入力ノードまたは Statistics ファイル 入力ノードを使用して IBM® SPSS® Modeler にインポートし、フィルタ ノードを使用して SPSS Modeler に定義することができます。

- ▶ [新規作成] をクリックして新しい複数レスポンス セットを作成するか、または [編集] をクリックして既存の設定を変更します。

図 4-37
複数レスポンス セットを編集

複数レスポンス セット 編集 Smurveys

情報設定

名前: \$museums

ラベル: Museums and galleries visited or plans to visit

データ型: 複数二分法設定
 複数カテゴリ設定

カウントした値: 1

設定のフィールド:

- museums.National_Museum_of_Science
- museums.Museum_of_Design
- museums.Institute_of_Textiles_and_Fashion
- museums.Archeological_Museum

オリジナル値を使用

OK(O) キャンセル(C)

名前とラベル: セットの名前と説明を指定します。

タイプ: 複数回答の質問を、次の 2 つのいずれかの方法で処理することができます。

- **複数二分法設定:** 個々のフラグ型フィールドがそれぞれの回答に対して作成されます。10 冊の雑誌がある場合、10 のフラグ型フィールドがあり、それぞれに真 (true) または偽 (false) を表す 0 または 1 などの値が含まれます。カウントされた値を使用して、真 (true) としてカウントされる値を指定することができます。この方法は、回答者が適用されるすべてのオプションを選択できるようにしたい場合に役に立ちます。
- **複数カテゴリ設定:** 名義型フィールドが、各回答に対して回答者からの最大回答数まで作成されます。各名義型フィールドには、「Time」の 1、「Newsweek」の 2、および「PC Week」の 3 などの考えられる回答を表す値が含まれます。この方法は、たとえば回答者にもっともよく読む 3 冊の雑誌を選択するよう質問する場合など、回答数を制限する場合に最も役に立ちます。

設定のフィールド: 右側のアイコンを使用して、フィールドを追加または削除します。

図 4-38
複数回答の質問

質問 14 どの博物館または美術館に行ったことがありますか？ または行きたいですか？
該当するものをすべて選んでください。

| | |
|----------------------|--------------------------|
| 国立科学博物館..... | <input type="checkbox"/> |
| デザイン博物館..... | <input type="checkbox"/> |
| テキスタイル/ファッション協会..... | <input type="checkbox"/> |
| 考古学博物館..... | <input type="checkbox"/> |
| 国立美術館..... | <input type="checkbox"/> |
| 北部美術館..... | <input type="checkbox"/> |
| その他 (記入してください)..... | <input type="checkbox"/> |
| なし..... | <input type="checkbox"/> |

コメント

- すべての複数レスポンス セットのすべてのフィールドは同じ保存タイプを持たなければいけません。
- セットは含まれるフィールドとは異なります。たとえば、セットを削除しても含まれるフィールドは削除されず、これらのフィールド間のリンクのみが削除されます。セットは削除のポイントから上流では表示されますが、下流では表示されません。
- フィールドの名前がフィルタ ノードを使用して (タブで直接または [フィルタ] メニューの [IBM® SPSS® Statistics 用に名前変更]、[短縮]、または [匿名化] を選択) 変更された場合、複数レスポンス セットで使用されるフィールドへの参照も更新されます。ただし、フィルタ ノードで削除される複数レスポンス セットのフィールドは、複数レスポンス セットからは削除されません。このようなフィールドは、ストリーム内では表示されませんが、複数レスポンス セットによって参照されます。このことは、たとえばエクスポート時に検討されます。

アンサンブル ノード

アンサンブル ノードでは、2 つまたはそれ以上のモデル ナゲットを組み合わせて個々のモデルのいずれかから取得するよりも、より正確な予測を取得します。複数モデルの予測を組み合わせることにより、個々のモデルの制限を回避でき、全体の精度がより高くなります。こうして組み合わせられたモデルは通常、少なくとも最良のモデルと同じくらい、あるいはしばしばそれ以上のパフォーマンスを実現します。

ノードのこの組み合わせは、自動分類、自動数値および自動クラスターの自動モデル作成ノードで自動的に作成されます。 [詳細は、5 章 自動化モデル作成ノード in IBM SPSS Modeler 15 モデル作成ノード を参照してください。](#)

アンサンブル ノードを使用した後、分析ノードまたは評価ノードを使用して、各入力モデルと組み合わせた結果の制度を比較することができます。これを実行するには、[アンサンブル モデルにより生成されたフィールドを除外] オプションがアンサンブル ノードの [設定] タブで選択されていないことを確認します。

出力フィールド

各アンサンブル ノードは、結合したスコアを含むフィールドを生成します。名前は、特定の対象フィールドに基づき、フィールドの尺度（フラグ型、名義（セット）型、または連続型（範囲）型）によって \$XF_、\$XS_、または \$XR_ の接頭辞が付きます。たとえば、対象フィールドがフラグ型で response という名前の場合、出力フィールド名は \$XF_response となります。

確信度または傾向フィールド フラグ型フィールドおよび名義型フィールドの場合、次の表で説明されているとおり、追加の確信度または傾向フィールドがアンサンブル方法に基づいて作成されます。

| アンサンブル法 | フィールド名 |
|--|----------------|
| Voting 確信度-重み付き票決 行傾向-重み付き票決 調整済み傾向-重み付き票決 最高確信度勝ち取り | \$XFC_<field> |
| 平均行傾向 | \$XFRP_<field> |
| 平均調整済み行傾向 | \$XFAP_<field> |

アンサンブル ノードの設定

集合体の対象フィールド: 2 つまたはそれ以上の上流モデルで対象フィールドとして使用される単一フィールドを選択します。上流モデルはフラグ型、名義型または連続型対象を使用することができますが、少なくとも 2 つのモデルが同じ対象を共有してスコアを結合する必要があります。

アンサンブル モデルにより生成されたフィールドを除外: 出力から、アンサンブル ノードに使用する個々のモデルで生成されたすべての追加フィールドを削除します。すべての入力モデルの結合スコアにのみ関心がある場合、このチェック ボックスを選択します。たとえば分析ノードまたは評価ノードを使用して結合スコアのと各入力モデルの制度を比較する場合、このオプションが選択解除されていることを確認します。

図 4-39
対象として選択された連続型フィールドを持つアンサンブル ノード



使用可能な設定は、対象として選択されたフィールドの尺度によって異なります。

連続型対象

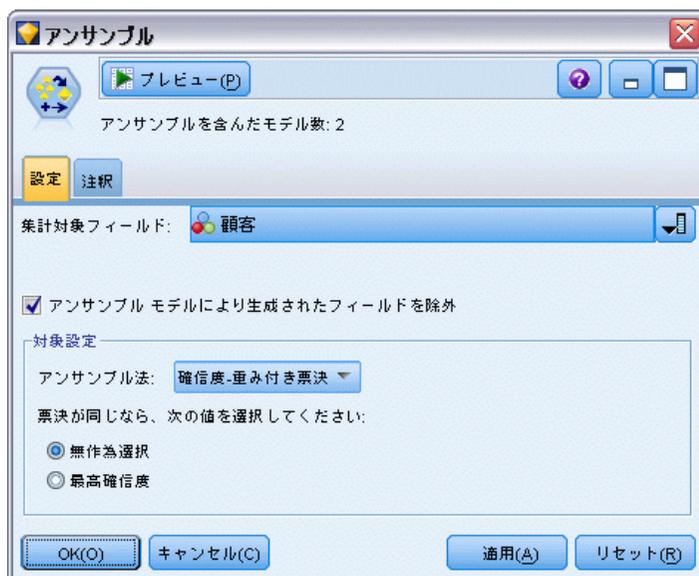
連続型対象の場合、スコアは平均が算出されます。これはスコアの結合にのみ使用できる方法です。

スコアまたは推定を平均化する場合、アンサンブル ノードでは標準誤差の計算を使用して、測定されたまたは推定された値と真の値との間の差異を算出し、これらの推定がどれくらい近いかを示します。デフォルトでは、新しいモデルに標準誤差の計算が生成されます。ただし、再生成する場合など、既存のモデルのチェック ボックスの選択を解除することができます。

カテゴリ対象

カテゴリ対象の場合、それぞれの予測値が選択される回数を集計し、最も高い合計数を持つ値を選択することによって動作する**票決**など、多くの方法がサポートされています。たとえば、5 つのモデルのうち 3 つがはいと予測し、残り 2 つがいいえを予測する場合、はいが 3 対 2 の票決で勝ちます。また、各予測の確信度または傾向値に基づいて、票決に**重み付け**することができます。重みは集計され、最も大きな合計の値が再度選択されます。最後の予測の確信度は、勝った値の重みの合計をアンサンブルに含まれるモデルの数で割った値です。

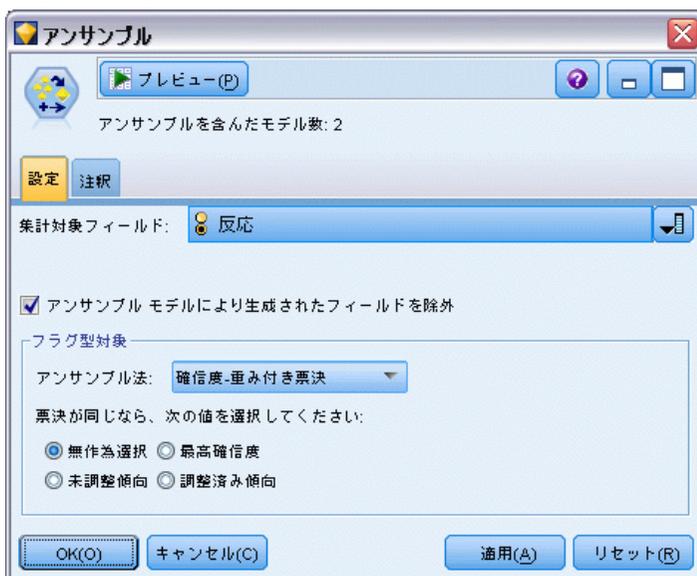
図 4-40
対象として選択された名義型フィールドを持つアンサンブル ノード



すべてのカテゴリ フィールド: フラグ型フィールドおよび名義型フィールドの場合、次の方法がサポートされます。

- Voting
- 確信度-重み付き票決
- 最高確信度勝ち取り

図 4-41
対象として選択されたフラグ型フィールドを持つアンサンブル ノード



フラグ型フィールドのみ: フラグ型フィールド場合のみ、傾向に基づいた次の方法も使用できます。

- 行傾向-重み付き票決
- 調整済み傾向-重み付き票決
- 平均行傾向
- 平均調整済み傾向

可否同数: 票決方法の場合、可否同数の解決方法を指定することができます。

- **無作為選択:** 可否同数の値の 1 つが無作為に選択されます。
- **最高確信度:** 最高確信度で予測された可否同数の値が勝ちます。これは、予測されたすべての値の最高確信度と必ずしも同じとは限りません。
- **行または調整済み傾向 (フラグ型フィールドのみ):** 絶対傾向が次のように計算されている場合の、最大絶対傾向によって予測された可否同数の値。

$$\text{abs}(0.5 - \text{傾向}) * 2$$

または、調整済み傾向の場合は次のようになります。

$$\text{abs}(0.5 - \text{調整済み傾向}) * 2$$

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 15 [モデル作成ノード](#) を参照してください。

フィールド作成ノード

IBM® SPSS® Modeler の強力な機能の 1 つとして、データ値を変更して、既存のデータから新しいフィールドを作成できることがあげられます。長期間にわたるデータマイニングプロジェクトでは、Web ログデータから顧客 ID を抽出したり、トランザクションデータや人口統計データから顧客の生涯価値を作成するような、新しいデータの作成が一般的に行われます。これらの変換はすべて、さまざまなフィールド設定ノードを使って行うことができます。

さまざまなノードで、新規フィールドを作成することができます。



フィールド作成ノードで、1 つまたは複数の既存フィールドから、データ値を変更するか、新しいフィールドを作成します。これで、タイプ式、フラグ、名義、ステート、カウント、および条件式の各フィールドが作成されます。詳細は、[p.185 フィールド作成ノード](#) を参照してください。



データ分類ノードにより、あるカテゴリ値のセットが別のセットに変換されます。データ分類ノードは、カテゴリを再編成したり、分析用のデータをグループ化しなおす場合に役立ちます。詳細は、[p.207 データ分類ノード](#) を参照してください。



データ分割ノードで、既存の 1 つまたは複数の連続型（数値範囲）フィールドの値に基づいて、自動的に新しい名義型（セット型）フィールドを作成します。たとえば、連続型収入フィールドを、平均からの偏差による収入グループを含む、新しいカテゴリフィールドに変換することができます。新規フィールドのビンを作成すると、分割点に基づいてフィールド作成ノードを生成することができます。詳細は、[p.212 データ分割ノード](#) を参照してください。



フラグ設定ノードで、1 つ以上の名義型フィールドに定義されたカテゴリ値に基づいた、複数のフラグ型フィールドが派生します。詳細は、[p.232 フラグ設定ノード](#) を参照してください。



再構成ノードで、名義型またはグラフ型フィールドを、これから別のフィールドの値で埋めることができるフィールドのグループへ変換します。たとえば、credit、cash、および debit の値の payment type という名前のフィールドがある場合、3 つの新しいフィールド (credit、cash、debit) が作成されます。その各々には、実際の支払の値を含めることができます。詳細は、[p.233 再構成ノード](#) を参照してください。



時系列ノードにより、以前レコードのフィールドのデータを含む、新規フィールドが作成されます。時系列ノードは、多くの場合、時系列データなどの継続的なデータに使用されます。時系列ノードを使用する前に、ソート ノードを使用して、データをソートしておくこともできます。詳細は、[p.263 時系列ノード](#) を参照してください。

フィールド作成ノードの使用

フィールド作成ノードを使って、1 つまたは複数の既存フィールドから、6 種類の新規フィールドを作成することができます。

- **CLEM 式**：新規フィールドは任意の CLEM 式の結果です。
- **フラグ型**：新規フィールドは、指定した条件を示すフラグ型です。
- **名義**：新規フィールドは名義型であり、そのメンバーは指定した値のグループです。
- **ステート**：新規フィールドは、2 つの状態（ステート）のどちらかです。これら 2 つの状態は指定した条件によって切り替えられます。
- **度数**：新規フィールドは、条件を満たした回数を基準に決まります。
- **条件式**：新規フィールドは、条件の値に応じて、2 つの式のどちらかの値になります。

これらのノードには、それぞれの [フィールド作成ノード] ダイアログボックスに、特別な一連のオプションが用意されています。これらのオプションは、以降の各項目で説明しています。

フィールド作成ノードの基本オプションの設定

フィールド作成ノードのダイアログ ボックスの上部には、必要なフィールド作成ノードのデータ型を選択するさまざまなオプションが用意されています。

図 4-42
フィールド作成ノード ダイアログ ボックス



モード: 複数のフィールドを作成するかどうかによって、[単一] または [複数] を選択します。[複数] を選択した場合、ダイアログ ボックスには複数の派生フィールド用のオプションが表示されます。

派生フィールド: 単純なフィールド作成ノードに対して、作成して各レコードに追加するフィールド名を指定します。デフォルトのフィールド名は DeriveN です。ここで、N は現在のセッション中でここまでに作成されたフィールド作成ノード数を表します。

データ型: ドロップダウン リストから、CLEM 式や名義型などのフィールド作成ノードのデータ型を選択します。それぞれのデータ型に対応したダイアログ ボックスで指定した条件に基づいて、新規フィールドが作成されます。

ドロップダウン リストでオプションを選択すると、各データ型のプロパティに応じて、フィールド作成ノードのダイアログ ボックスに異なるオプションが表示されます。

フィールドのデータ型: 連続型、カテゴリ型、またはフラグ型など、新しい派生フィールドの尺度を選択します。このオプションは、すべてのフィールド作成ノードで共通です。

注：新規フィールドの作成には、しばしば特殊関数や数式が必要なことがあります。これらの式の作成を支援するために、すべての種類のフィールド作成ノードのダイアログ ボックスから利用することができ、式の検査機能やすべての CLEM 式の一覧を提供する CLEM 式ビルダーが用意されています。詳細は、7 章 CLEM について in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。

複数フィールドの作成

フィールド作成ノード内でモードを【複数】に設定すると、同じノード内で同じ条件に基づいて複数のフィールドを作成できます。この機能を利用すれば、データ セット中の複数のフィールドで同じ変換処理を行う場合に、時間を節約することができます。たとえば、初任給と以前の経歴に基づいて現在の給与を予測する回帰モデルを作成する場合、3 つの非対称変数すべてに対数変換を適用するために役立ちます。各変換に新しいフィールド作成ノードを追加する代わりに、すべてのフィールドに一度に同じ関数を適用できます。新しいフィールドの作成元となるすべてのフィールドを選択し、フィールドのカッコ内で @FIELD 関数を使用してフィールド作成式を入力します。

注：@FIELD 関数は、複数のフィールドを同時に作成するために重要なツールです。このツールを使用すれば、正確なフィールド名を指定することなく現在のフィールドの内容を参照することができます。たとえば、複数のフィールドに対して対数変換を適用するために用いられる CLEM 式は、 $\log(@FIELD)$ になります。

図 4-43
複数フィールドの作成



[複数] モードを選択すると、ダイアログ ボックスに次のオプションが表示されます。

フィールド リスト: フィールド ピッカーを使って、新規フィールドを作成するフィールドを選択します。選択した各フィールドに対して、1 つの出力フィールドが生成されます。注： 選択するフィールドがそれぞれ同じストレージ タイプである必要はありません。ただし、すべてのフィールドに対して条件が有効でなければ、フィールド作成操作は失敗してしまいます。

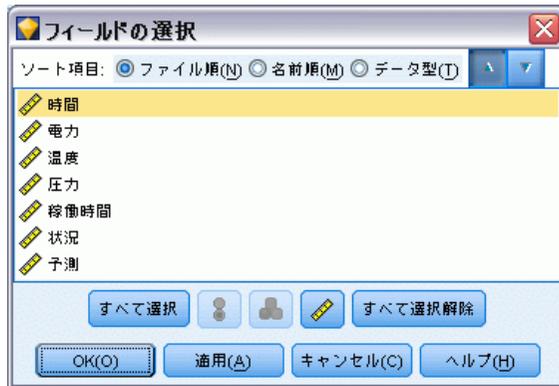
フィールド名拡張子: 新しいフィールド名、E追加する拡張子を入力します。たとえば、Current Salary の対数を含む新しいフィールドの場合、フィールド名に拡張子 log_ を追加すると、log_Current Salary が生成されます。ラジオ ボタンを使用して、拡張子をフィールドの接頭辞（先頭）として追加するか、接尾辞（最後）として追加するかを指定します。デフォルトのフィールド名は DeriveN です。ここで、N は現在のセッション中でここまでに作成されたフィールド作成ノード数を表します。

単一モードのフィールド作成ノードと同様に、新しいフィールドの作成に使用する CLEM 式を作成する必要があります。選択したフィールド作成操作の種類に応じて、条件を作成するためのさまざまなオプションがあります。これらのオプションは、以降の各項目で説明しています。CLEM 式を作成するには、CLEM 式フィールドに直接入力するか、または計算機ボタンをクリックして Clem 式ビルダーを使用してください。複数のフィールドの操作を参照する場合は、@FIELD 関数を使用してください。

複数フィールドの選択

フィールド作成ノード（複数モード）、レコード集計ノード、ソート ノード、線グラフ ノード、時系列ノードなどの、複数の入力フィールドに対して操作を実行するすべてのノードで、次に示す [フィールドの選択] ダイアログ ボックスを使って複数のフィールドを簡単に選択することができます。

図 4-44
複数フィールドの選択



ソート項目： 次のオプションを使って、表示する利用可能フィールドをソートすることができます。

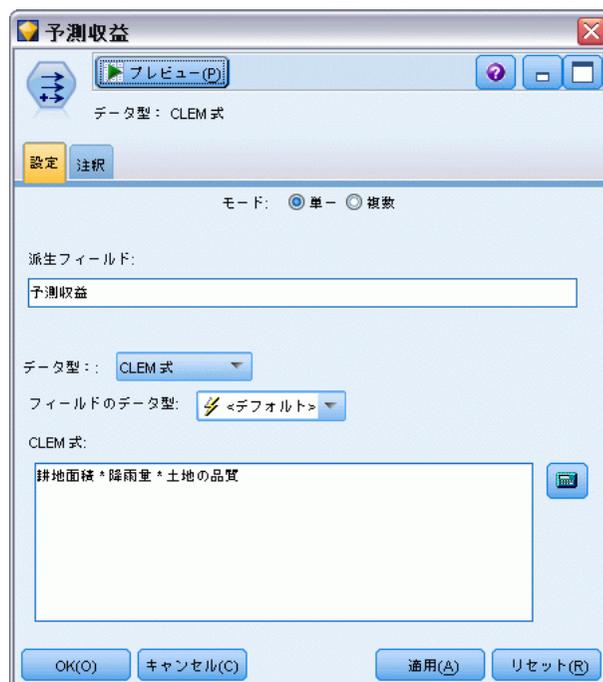
- **ファイル順：** データ ストリームから現在のノードにデータが渡された順序にフィールドをソートします。
- **名前：** アルファベット順でフィールドをソートします。
- **データ型：** 尺度でソートしたフィールドを表示します。特定の尺度のフィールドを選択する場合に役立ちます。

リストからフィールドを 1 回に 1 つずつ選択するか、または Shift キーまたは Ctrl キーを押しながら複数のフィールドを選択します。また、リストの下ボタンを使って、尺度に基づいて複数のフィールドを選択したり、テーブル中のすべてのフィールドを選択または選択解除することができます。

CLEM 式作成オプションの設定

フィールド作成ノード（CLEM 式型）は、CLEM 式の結果を基にして、データセット内の各レコード用に新規フィールドを作成します。この式は条件式ではありません。条件式に基づいて値を作成するには、フラグ型または条件式型のフィールド作成ノードを使用します。

図 4-45
フィールド作成ノード（CLEM 式型）のオプションの設定



CLEM 式 :CLEM 言語を使用して新規フィールドの値を作成するための式を指定します。詳細は、7 章 CLEM について in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。

フィールド作成ノード（フラグ型）のオプションの設定

フィールド作成ノード（フラグ型）は、高血圧や使用されていない顧客アカウントなどの、特定の条件を示すために使用されます。フラグ フィールドは各レコードに対して作成され、真（true）の条件を満たす場合に、フィールドに真（true）のフラグ値が追加されます。

図 4-46
フラグ型フィールドの作成



真 (true) の値： 下のフィールドで指定した条件と一致するレコードの、フラグ型フィールドに入れる値を指定します。デフォルトは T です。

偽 (false) の値： 下のフィールドで指定した条件と一致しないレコードの、フラグ型フィールドに入れる値を指定します。デフォルトは F です。

真 (true) となる条件： 各レコードの特定の値を評価して、真 (True) の値または 偽 (False) の値 (上記) を与えるために使用する CLEM 条件を指定します。偽 (false) ではない数値に対しては、真 (true) の値がレコードに与えられることに注意してください。

注：空文字列を返すには、"" のように中に何も入れずに引用符を 2 つ指定します。空文字列は、たとえばテーブル中で真 (true) の値をより際立たせるために、偽 (true) の値として使用されます。同様に、数値として扱われる値を文字列値として使用する場合にも、引用符で囲む必要があります。

例

IBM® SPSS® Modeler 12.0 以前のリリースでは、値をカンマで区切った複数の回答を単一のフィールドにインポートしていました。

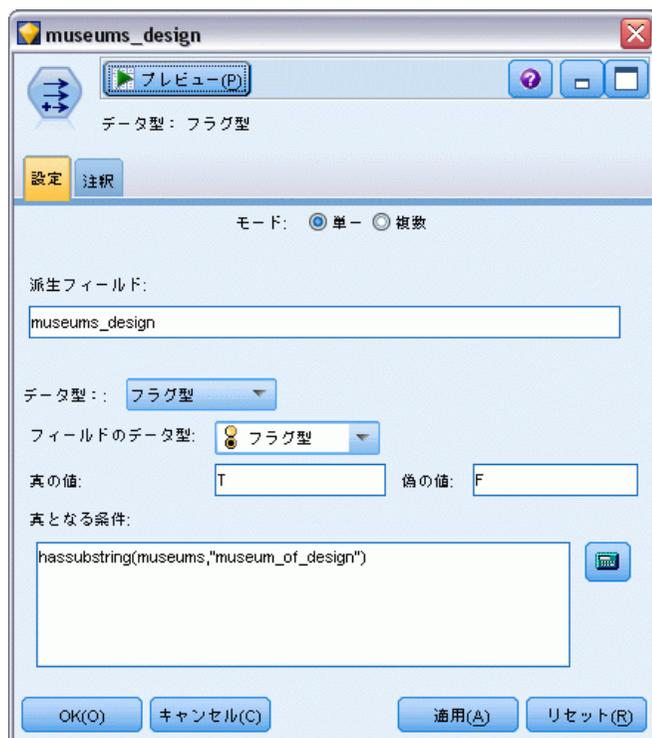
| |
|--|
| museums |
| museum_of_design,institute_of_textiles_and_fashion |
| museum_of_design |

| |
|---|
| museums |
| archeological_museum |
| \$null\$ |
| national_art_gallery, national_museum_of_science, other |

分析の目的でこのデータを準備するには、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、**hassubstring** 関数を使用できます。

```
hassubstring(museums,"museum_of_design")
```

図 4-47
hassubstring 関数を使用したフラグ型フィールドの作成



フィールド作成ノード (セット型) のオプションの設定

フィールド作成ノード (セット型) は、複数の CLEM 条件のセットを実行して、各レコードがどの条件を満たすかを判断するために使用します。各レコードが条件を満たすと、値 (どの条件のセットを満たしたかを示す) が新しく作成されたフィールドに追加されます。

図 4-48
セット型フィールド作成ノードの使用



デフォルト値：一致する条件がない場合に、新規フィールドで使用する値を指定します。

フィールド設定値：特定の条件を満たした場合に、新規フィールドに入れる値を指定します。リスト中の各値には、対応する条件があり、この条件は隣接する列で指定します。

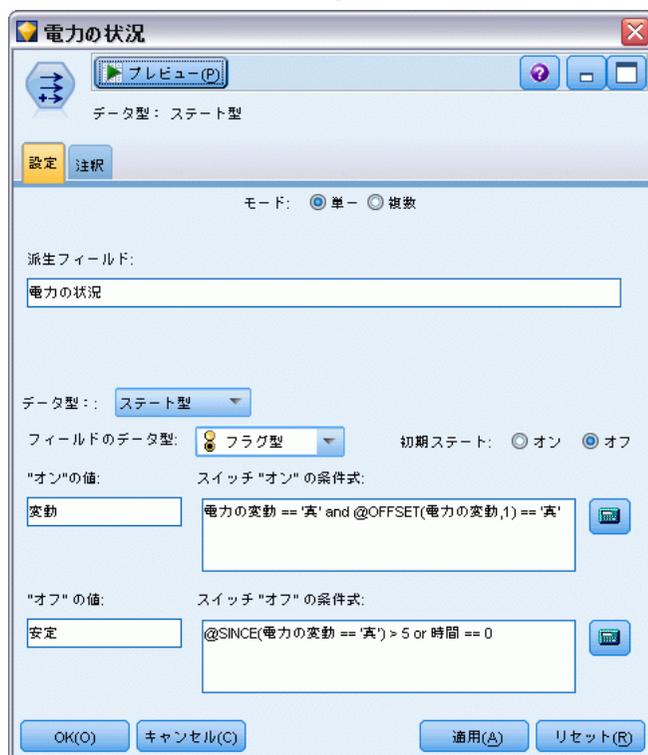
値の設定条件 (真の場合に値を設定)：セット型フィールド中の各メンバーに対する条件を指定します。Clem 式ビルダーを使って、利用可能な関数とフィールドを選択します。矢印および削除ボタンを使って、条件を並べ替えたり、削除することができます。

データ セット内の特定のフィールドの値が、条件に該当するかどうかを調べます。各条件が調べられると、どの条件を満たしたかを示すために（満たした場合）、上のフィールドで指定した値が新規フィールドに割り当てられます。どの条件も満たさなかった場合は、デフォルト値が使用されます。

フィールド作成ノード（ステート型）のオプションの設定

フィールド作成ノード（ステート型）は、フィールド作成ノード（フラグ型）に似ています。フィールド作成ノード（フラグ型）は、現在のレコードが 1 つの条件を満たすかどうかを示す値を表示しますが、フィールド作成ノード（ステート型）では、2 つの独立した条件をどのように満たすかに応じてフィールドの値が変化します。つまり、各条件を満たすと値が変わります（オンまたはオフになる）。

図 4-49
セット型フィールド作成ノードの使用



初期ステート：新規フィールドの各レコードに初期ステートとしてオンまたはオフのどちらの値を与えるかを選択します。この値は、各条件を満たす場合に変わることに注意してください。

「オン」の値：オンの条件を満たす場合の新規フィールドの値を指定します。

スイッチ「オン」の条件式：条件が真（true）の場合にステートをオンに変更する CLEM 条件を指定します。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

「オフ」の値：オフの条件を満たす場合の新規フィールドの値を指定します。

スイッチ「オフ」の条件式：条件が偽（false）の場合にステートをオフに変更する CLEM 条件を指定します。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

注：空文字列を指定するには、"" のように中に何も入れずに引用符を 2 つ指定します。同様に、数値として扱われる値を文字列値として使用する場合にも、引用符で囲む必要があります。

フィールド作成ノード（カウント型）のオプションの設定

フィールド作成ノード（カウント型）は、データセット内の数値型フィールドの値に一連の条件を適用するために使用します。各条件を満たすたびに、作成されたカウント型フィールドの値が設定された増分の値ずつ増やされます。このデータ型のフィールド作成ノードは時系列データの場合に役に立ちます。

図 4-50
フィールド作成ノード ダイアログ ボックスのカウント オプション



初期値：実行時に新規フィールド用に使用する値を設定します。初期値は数値の定数にする必要があります。値を増やしたり減らすには、矢印ボタンを使用します。

増分条件 :CLEM 条件を指定します。この条件を満たす場合に、[増分] で指定した数ずつ、作成された値が増加します。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

増分 : カウントを増やすための値を設定します。数値の定数または CLEM 式の結果のどちらかを使用できます。

リセット条件 : この条件を満たす場合に、作成値を初期値にリセットする条件を指定します。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

フィールド作成ノード (条件式型) のオプションの設定

フィールド作成ノード (条件式 (If-Then) 型) は一連の If-Then-Else 文を使用して、新規フィールドの値を作成します。

図 4-51
フィールド作成ノード (条件式型) の使用



If : 実行時に各レコードを評価する CLEM 条件を指定します。条件が真 (数値のときは偽以外) の場合は、Then の式によって下のフィールドに指定された値が新規フィールドに与えられます。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

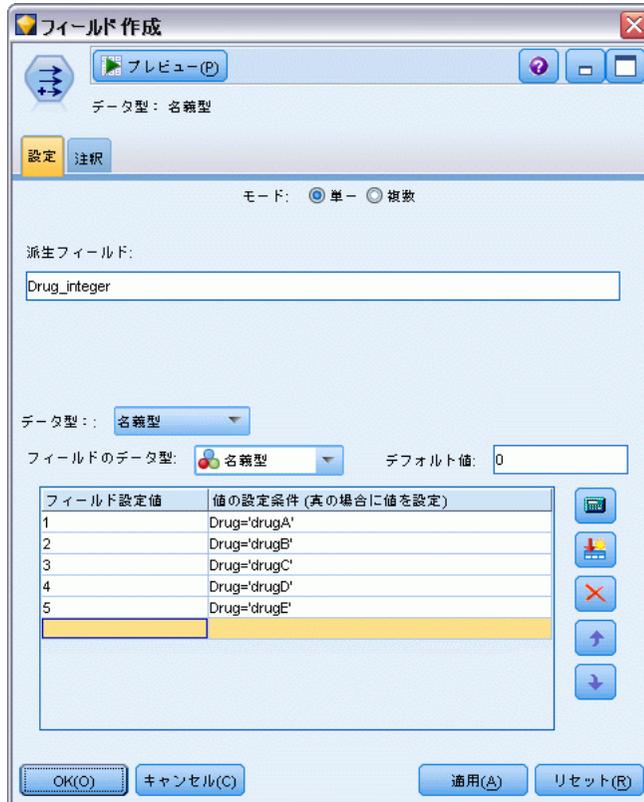
Then : 上の If 文が真（または偽以外）の場合に新規フィールドに入れる値または CLEM 式を指定します。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

Else : 上の If 文が 偽 (false) の場合に新規フィールドに入れる値または CLEM 式を指定します。Clem 式ビルダーを表示するには、計算機ボタンをクリックします。

フィールド作成ノードを使用して値を再コード化する

フィールド作成ノードは、たとえばカテゴリ型の値を持つ文字列フィールドを数値名義（セット）型フィールドに変換して値を再コード化するために使用できます。

図 4-52
文字列の値を再コード化する



- ▶ [データ型] については、該当するフィールドのタイプ（名義型、フラグ型など）を選択します。
- ▶ 値を再コード化するための条件を指定します。たとえば、Drug='drugA' と Drug='drugB' の値を、それぞれ 1 と 2 に設定することができます。

置換ノード

置換ノードは、フィールド値の置換やストレージの変更に用いられます。**@BLANK(FIELD)**のような、CLEM 条件に基づいて値を置換することができます。また、すべての空白値やヌル値を特定の値に置換することもできます。置換ノードは、データ型ノードと組み合わせて、欠損値を置き換えるためによく用いられます。たとえば、**@GLOBAL_MEAN** のような式を指定することによって、空白をフィールドの平均値で置き換えることができます。この式は、すべての空白値をグローバル ノードで算出した平均値に置き換えます。

図 4-53
置換ノードのダイアログ ボックス



対象フィールド: フィールド ピッカー (テキスト フィールドの左側にあるボタン) を使って、値を調査、置換するデータ セットのフィールドを選択します。デフォルトでは、下の [条件] および [置換値] の指定、および関連する式に基づいて、値が置換されます。また、下の [置換] オプションを使って、置換の代替手法を選択することもできます。

注： ユーザーが定義した値で置換するフィールドを複数選択する場合は、フィールドのデータ型が同じでなければなりません (すべて数値またはすべてシンボル値)。

置換：次のいずれかの方法を使って、選択したフィールドの値を置換する場合に選択します。

- **条件を指定**：このオプションを選択すると、[条件] フィールドと CLEM 式ビルダーがアクティブになります。これを使って、指定した値と置換するための条件として使用する式を作成します。
- **常時**：選択したフィールドのすべての値を置換します。たとえば、このオプションを使って所得 (income) のストレージを文字列に変換するには、次の CLEM 式を使用します (`to_string(income)`)。)
- **空白値**：選択したフィールドの、ユーザーが指定したすべての空白値を置換します。空白値を選択するには、標準の条件 `@BLANK(@FIELD)` が使用されます。注：入力ノードまたはデータ型ノードを使って、空白値を定義することができます。
- **ヌル値**：選択したフィールドの、すべてのヌル値を置換します。Null 値を選択するには、標準の条件 `@NULL(@FIELD)` が使用されます。
- **空白値とヌル値**：選択したフィールドの空白値とシステムヌル値の両方を置換します。このオプションは、ヌルが欠損値として定義されているかどうか不明な場合などに役立ちます。

条件 [条件を指定] オプションを選択した場合に、このオプションを指定することができます。このテキスト ボックスには、選択したフィールドを評価するための CLEM 式を指定します。CLEM 式ビルダーを表示するには、計算機ボタンをクリックします。

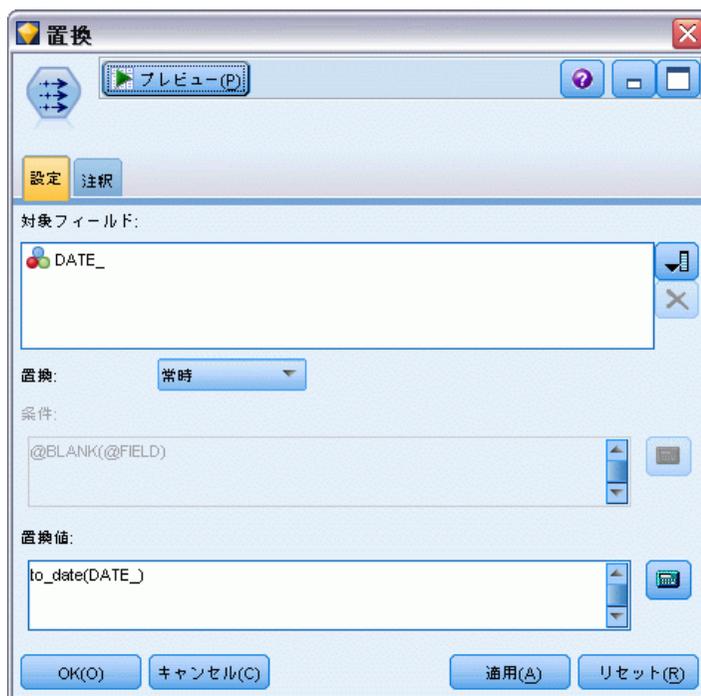
置換値：選択したフィールドに新しい値を設定する CLEM 式を指定します。テキスト ボックスに `undef` と入力することにより、値をヌル値で置き換えることもできます。CLEM 式ビルダーを表示するには、計算機ボタンをクリックします。

注：選択したフィールドが文字型の場合は、文字型の値で置換する必要があります。文字型フィールドに対して、置換値にデフォルトの 0 または他の数値を使用すると、エラーが発生してしまいます。

置換ノードを使ったストレージの変換

置換ノードの [置換] 条件を使えば、単一または複数のフィールドのストレージ タイプを簡単に変換することができます。たとえば、関数 `to_integer` を使用すれば、収入 (income) フィールドのストレージを、文字列から整数に変換することができます。次に、この CLEM 式の例を示します。 `to_integer(income)`。

図 4-54
置換ノードを使ったフィールドのストレージの変換



Clem 式ビルダーを使えば、利用できる変換関数を表示したり、CLEM 式を自動的に作成することができます。[関数] ドロップダウン リストで、[変換] を選択すると、ストレージ変換関数が一覧表示されます。利用できる変換関数を次に示します。

- to_integer(ITEM)
- to_real(ITEM)
- to_number(ITEM)
- to_string(ITEM)
- to_time(ITEM)
- to_timestamp(ITEM)
- to_date(ITEM)
- to_datetime(ITEM)

日付と時刻の値の変換： 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 [MON YYYY]

を選択します。詳細は、5 章 [ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

変換関数は、フィールド作成ノードのフィールド作成計算時の一時変換でも、利用できます。また、フィールド作成ノードを使って、カテゴリ値を含む文字列フィールドの読み取りなど、他の操作も実行できます。詳細は、[p.198 フィールド作成ノードを使用して値を再コード化する](#) を参照してください。

匿名化ノード

ノードのモデル下流に含まれるデータと連携している場合、匿名化ノードで、フィールド名、フィールド値のどちらかまたは両方を隠すことができます。これにより、権限を持たないユーザーが従業員の記録や患者の治療記録など機密データを閲覧する危険なく、生成されたモデルを自由に(たとえばテクニカル サポートへ)分散させることができます。

ストリームの匿名化ノードの場所によっては、他のノードに変更する必要があります。たとえば選択ノードの上流に匿名化ノードを挿入する時、選択ノードの選択基準を匿名化された値に実行する場合に変更する必要があります。

匿名化で使用される方法は、様々な要素によって決まります。フィールド名、および連続型尺度以外のすべてのデータ値について、データは次のフォーマットの文字列に置換されます。

`prefix_Sn`

`prefix_` はユーザー指定の文字列またはデフォルトの文字列 `anon_`、および `n` は 0 で始まり各一意値まで増加する整数の値です (たとえば `anon_S0`、`anon_S1` など)。

数値の範囲は文字列より整数または実数値に対応しているため、範囲型のフィールド値を変換する必要があります。フィールド値はその範囲を異なる範囲に変換することによってのみ匿名化することができ、元のデータを隠します。範囲内にある値 x の変換は、次のように行われます。

$A*(x + B)$

ここでの意味は次の通りです。

A は、0 より大きい換算係数です。

B は値に追加する翻訳オフセットです。

例

換算係数 A が 7 に、翻訳オフセット B が 3 に設定されているフィールド AGE の場合、AGE の値は次のように変換されます。

7*(AGE + 3)

匿名化ノードのオプションの設定

ここでは、どのフィールドで値をより下流に隠すかを選択することができます。

データ フィールドを匿名化ノードから上流にインスタンス化した後、匿名化処理を実行することができます。データ型ノードまたは入力ノードの [データ型] タブで [値の読み込み] をクリックすると、データをインスタンス化することができます。

図 4-55
匿名化オプションの設定



フィールド: 現在のデータ セットのフィールドの一覧を表示します。フィールド名がすでに匿名化されている場合、匿名化された名前がここで表示されます。

尺度: フィールドの尺度。

匿名化値: 1 つ以上のフィールドを選択、この列をクリックして[はい]を選択し、デフォルトの接頭辞 `anon_` を使用してフィールド値を匿名化します。[指定] を選択してダイアログ ボックスを表示して独自の接頭辞を入力、あるいは連続型のフィールド値の場合は、フィールド値の変換が乱数またはユーザー定義の値を使用するのか指定します。連続型フィールド タイプまたは連続型でないフィールド タイプは、同じ操作では指定できません。各フィールドタイプで個別に指定する必要があります。

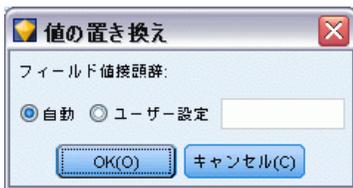
現在のフィールドを表示： 匿名化 ノードに接続しているデータ セットのフィールドを表示する場合に選択します。このオプションは、デフォルトで選択されています。

未使用のフィールド設定を表示： ノードに接続したことがある（今は接続していない）データ セットのフィールドを表示する場合に選択します。このオプションは、ノードをあるストリームから別のストリームにコピーしたり、ノードを保存して再ロードするような場合に適しています。

フィールド値の匿名化方法の指定

[値を置換] ダイアログ ボックスで、フィールド値の匿名化にデフォルトの接頭辞を使用するか、またはユーザー指定の接頭辞を使用するか選択することができます。ダイアログ ボックスの [OK] をクリックし、選択したフィールドに対し、[設定] タブの [はい] を選択して値の匿名化の設定を変更します。

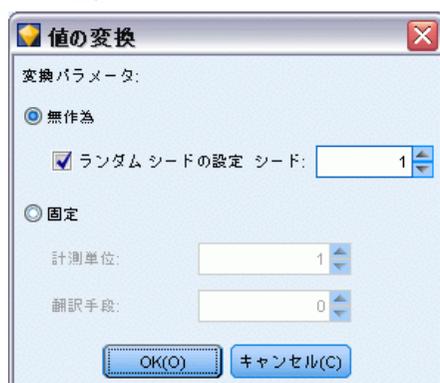
図 4-56
[値を置換] ダイアログ ボックス



フィールド値の接頭辞： 匿名化したフィールド値に対するデフォルトの接頭辞は `anon_` です。異なる接頭辞を希望する場合、[ユーザー設定] を選択し、独自の接頭辞を入力します。

[値の変換] ダイアログ ボックスは、連続型フィールドに対してのみ表示され、フィールド値の変換では乱数またはユーザー定義の値使用するのが指定することができます。

図 4-57
[値の変換] ダイアログ ボックス



無作為: [無作為] オプションを選択し、変換に乱数を使用します。デフォルトでは、[ランダム シードの設定] が選択されます。[シード] フィールドで値を指定するか、デフォルト値を使用します。

固定: [固定] オプションを選択し、変換に乱数を使用します。

- **スケール:** フィールドが変換中に複製される数です。最小値は 1 で、最大値は通常 10 ですが、あふれを防止するために低くなる場合があります。
- **翻訳:** 変換中、フィールド値に追加される数です。最小値は 0 で、最大値は通常 1000 ですが、あふれを防止するために低くなる場合があります。

フィールド値の匿名化

[設定] タブで匿名化のために選択されたフィールドには、匿名化された値が含まれます。

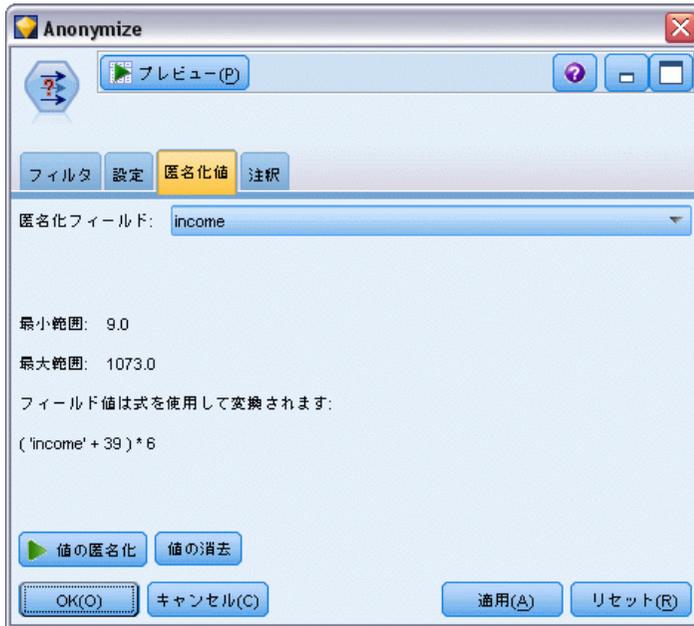
- 匿名化ノードを含むストリームを実行する場合
- 値をプレビューする場合

値をプレビューするには、[値の匿名化] タブの [値の匿名化] ボタンをクリックします。ドロップダウン リストから色を選択します。

尺度が連続型の場合、次の項目が表示されます。

- 元の範囲の最小値および最大値
- 値の変換に使用された方程式

図 4-58
フィールド値の匿名化



尺度が連続型以外のものである場合、画面にはそのフィールドの元の値および匿名化された値が表示されます。

図 4-59
フィールド値の匿名化



黄色の背景色で表示された場合、前回値が匿名化されたため選択されたフィールドの設定が変更されたか、匿名化された値が正常でないなど、匿名化ノードのデータ上流に変更が行われたことを示します。値の現在の設定が表示された場合、再度 **[値の匿名化]** ボタンをクリックし、現在の設定にしたがって値の新しい設定を生成します。

匿名化値：選択したフィールドに匿名化された値を作成し、テーブルに表示します。連続型フィールドにランダム シードを使用している場合、繰り返しこのボタンをクリックすると、クリックごとに異なる値のセットが作成されます。

値の消去：テーブルから元の値および匿名化された値を消去します。

データ分類ノード

データ分類ノードにより、あるカテゴリ値のセットを別のセットに変換することができます。データ分類ノードは、カテゴリを再編成したり、分析用のデータをグループ化しなおす場合に役立ちます。たとえば、**[商品]** の値を、キッチン用品、バス用品、および電化製品の 3 種類のグループに分類しなおすことができます。しばしばこの操作は、値をグループ化してデータ分類ノードを生成することにより、棒グラフ ノードから直接行われます。[詳細は、5 章 p. 343 棒グラフ ノードの使用方法を参照してください。](#)

データ分類は、1 つまたは複数のシンボル値フィールドに対して実行することができます。また、既存のフィールドを新しい値で置き換えたり、新規フィールドを生成することもできます。

データ分類ノードを使用する前に、その作業により適している他のフィールド設定ノードがないかどうかを検討してください。

- 数値範囲型（ランクやパーセントイルなど）をセット型に自動的に変換するには、データ分割ノードを使用する必要があります。[詳細は、p. 212 データ分割ノードを参照してください。](#)
- 数値範囲型をセット型に手作業で分類するには、フィールド作成ノードを使用する必要があります。たとえば、給与の値を特定の給与範囲カテゴリに分類する場合は、フィールド作成ノードを使って各カテゴリを手作業で定義する必要があります。
- Mortgage_type のようなカテゴリ フィールドの値に基づいて 1 つまたは複数のフィールドを作成するには、フラグ設定ノードを使用する必要があります。
- カテゴリ型フィールドを数値型ストレージに変換するには、フィールド作成ノードを使用できます。たとえば、No と Yes 値を、それぞれ 0 と 1 に変換することができます。[詳細は、p. 198 フィールド作成ノードを使用して値を再コード化するを参照してください。](#)

データ分類ノードのオプション設定

データ分類ノードの使用には、次の 3 つのステップがあります。

- ▶ まず、複数のフィールドを再分類するのか、または 1 つのフィールドを再分類するのかを選択します。
- ▶ 次に、分類したものを既存のフィールドに再コード化するのか、または新しいフィールドを作成するのかを選択します。
- ▶ 最後に、データ分類ノードのダイアログ ボックスのオプションを使って、セットを適切にマップします。

図 4-60
データ分類ノードのダイアログ ボックス



モード: 1 つのフィールドのカテゴリを再分類する場合、[単一] を選択します。複数のフィールドを同時に変換する場合は、[複数] を選択します。

データ分類先: 元の名義型フィールドをそのまま保持して、再分類した値を含む追加フィールドを作成する場合は、**[新規フィールド]** を選択します。元のフィールドの値に新しい分類値を上書きする場合は、**[既存のフィールド]** を選択します。これは本質的には、「置換」操作になります。

モードおよび置換オプションを指定したら、ダイアログ ボックスの下半分に表示されるオプションを使って変換フィールドを選択し、新しい分類値を指定する必要があります。これらのオプションは、選択したモードによって異なります。

データ分類フィールド: 右にあるフィールド選択ボタンを使用して、1 つ（単一モード）または複数（複数モード）のカテゴリ型フィールドを選択します。

新規フィールド名: 記録値を入れる新しい名義型フィールド名を指定します。このオプションは、単一モードで **[新規フィールド]** が選択されている場合にだけ利用できます。**[既存のフィールド]** が選択されている場合は、元のフィールド名がそのまま保持されます。複数モードで作業を行う場合は、このオプションに代わって、各新規フィールドに追加する拡張子を指定するオプションが表示されます。 [詳細は、 p. 210 複数フィールドのデータ分類](#) を参照してください。

データ分類値: このテーブルにより、古いセット値からここに指定した値へのマッピングを明確に行うことができます。

- **元の値:** 選択したフィールドの既存の値が表示されます。
 - **新しい値:** 新しいカテゴリ値を入力するか、またはドロップダウン リストから選択します。分布図からの値を使用して、データ分類ノードを自動的に生成する場合、これらの値はドロップダウン リストに含まれます。これにより、既存の値を既知の値のセット素早くに関連付けることができます。たとえば、ネットワークやロケールに基づいて、医療機関は異なる方法で診断をグループ化することがあります。合併または買収の後、すべての機関は新規のみならず既存のデータさえも一貫した方法で分類することが要求されます。長いリストから各対象値を手作業で入力しなくても、値の基本リストから IBM® SPSS® Modeler に読み込み、**[診断]** フィールドの分布図を実行し、分布図から直接このフィールドのデータ分類ノード（値）を生成できます。このプロセスにより、**[新しい値]** ドロップダウン リストのすべての対象 **[診断]** 値が利用可能になります。
- ▶ 上で選択した、1 つまたは複数のフィールドの元の値を読み込むには、**[取得]** をクリックします。
- ▶ 元の値を、まだマップされていないフィールドの **[新しい値]** 列に貼り付けるには、**[コピー]** をクリックします。マップされていない元の値が、ドロップダウン リストに追加されます。

- ▶ [新しい値] 列のすべての指定内容を消去するには、[新規消去] をクリックします。注：このオプションをクリックしても、ドロップダウン リストの値は消去されません。
- ▶ 元のそれぞれの値に対して、連続する整数を自動的に生成するには、[自動] をクリックします。この場合、整数値（1.5、2.5 などの実数値ではない）しか生成することはできません。

図 4-61
[自動分類] ダイアログ ボックス



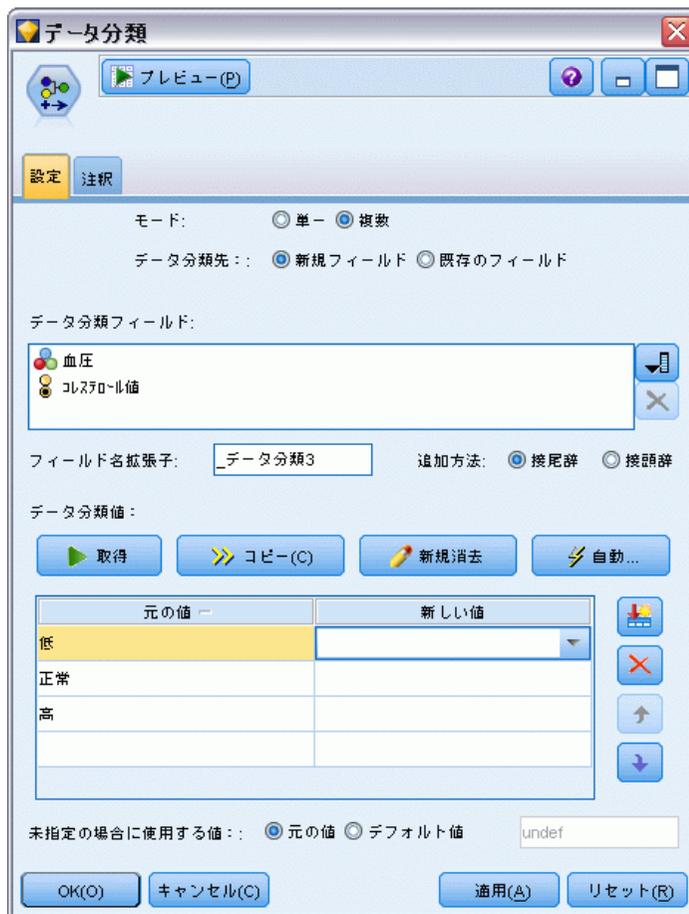
たとえば、商品名に対する連続した商品 ID 番号や、大学の講義番号などを自動的に生成することができます。この機能は、IBM® SPSS® Statistics の自動再コード変換に対応しています。

未指定の場合に使用する値：このオプションは、新しいフィールドで未指定の値を置換するために用いられます。[元の値] を選択して元の値をそのまま保持することも、デフォルト値を指定することもできます。

複数フィールドのデータ分類

複数フィールドのカテゴリ値を同時にマップするには、[複数] モードを選択します。複数モードにすると、[データ分類] ダイアログ ボックスには、次の設定項目が表示されます。

図 4-62
複数フィールドのデータ分類を行う場合のダイアログ ボックス オプション



データ分類フィールド： 右側にあるフィールド選択ボタンを使用して、変換するフィールドを選択します。すべてのフィールドを一度に選択することも、名義型やフラグ型のように、同じ種類のフィールドだけを選択することもできます。

フィールド名拡張子： 複数のフィールドを同時に再コード化する場合は、個別のフィールド名を指定するよりも、新しいフィールドすべてに共通の拡張子を付ける方が効率的です。「_recode」のような拡張子を指定して、拡張子を元のフィールド名の前に付けるか、または後に付けるかを選択してください。

再分類されたフィールドのストレージと尺度

データ分類ノードでは、常に再コード化操作により名義型フィールドが作成されます。そのため、状況によっては [既存のフィールド] の再分類を行う際に、フィールドの尺度が変更されてしまうこともあります。

新しいフィールドのストレージ（データがどのように使用されるかで、1 なく、データがどのように格納されるか）は、次の [設定] タブの設定に基づいて算出されます。

- 未指定の値に対してデフォルト値を使用するように設定されている場合、新しい値とデフォルト値の両方が調べられ、適切なストレージタイプが決められます。たとえば、すべての値が整数と判断された場合は、フィールドのストレージタイプは整数になります。
- 未指定の値に対して元の値を使用するように設定されている場合、ストレージタイプは元のフィールドのストレージに基づいて決められます。すべての値が、元のフィールドのストレージとして解析された場合は、そのストレージがそのまま保持されます。それ以外の場合は、古い値と新しい値の両方を考慮して、もっとも適切なストレージタイプが決められます。たとえば、整数セット { 1, 2, 3, 4, 5 } を「4 => 0, 5 => 0」で再分類すると、新しい整数セット { 1, 2, 3, 0 } が生成されますが、「4 => “3 以上、5 => “3 以上」で再分類すると、文字列セット { “1”, “2”, “3”, “3 以上” } が生成されます。

注：元のデータ型がインスタンス化されていない場合は、新しいデータ型もインスタンス化されません。

データ分割ノード

データ分割ノードにより、既存の 1 つまたは複数の連続型（数値範囲）フィールドの値に基づいて、自動的に新しい名義型フィールドを作成することができます。たとえば、連続型収入フィールドを、平均からの同じ偏差による収入グループを含む、新しいカテゴリフィールドに変換することができます。または、2 つのフィールド間の当初のアソシエーションの強度を保存するために、カテゴリの「スーパーバイザ」フィールドを選択できます。

データ分割は、次を含む多くの理由で、有用です。

- **アルゴリズムの要件**：Naive Bayes やロジスティック回帰などの一定のアルゴリズムには、カテゴリ入力が必要です。
- **パフォーマンス**：多項ロジスティックなどのアルゴリズムは、入力フィールドの異なる値の数が減らされると、より適正に実行されます。たとえば、各ビンの当初の値ではなく、中央値または平均値を使用します。
- **データのプライバシー**：給与などの慎重な扱いが必要な個人情報、プライバシーを保護するために、実際の数字でなく、一定の範囲内の数字として報告できます。

多くのビン メソッドが利用できるため、新規フィールドのビンを作成したら、分割点に基づいてフィールド作成ノードを生成することができます。

データ分割ノードを使用する前に、その作業により適している他の技法がないかどうかを検討してください。

- あらかじめ定義された給与範囲など、カテゴリの分割点を手作業で指定するには、フィールド作成ノードを使用します。詳細は、[p. 185 フィールド作成ノード](#) を参照してください。
- 既存のセットの新しいカテゴリを作成するには、データ分類ノードを使用します。詳細は、[p. 207 データ分類ノード](#) を参照してください。

欠損値の処理

データ分割ノードは欠損値を次のように処理します。

- **ユーザー定義の空白**: 変換時に、空白として指定された欠損値が含まれます。たとえば、データ型ノードを使って空白値を示すために -99 を指定した場合、この値がデータ分割処理に含まれます。データ分割処理中に空白値を無視するには、置換ノードを使って空白値をシステムのヌル値に置き換える必要があります。
- **システム欠損値 (\$null\$)**: データ分割処理時にヌル値は無視され、変換後もヌル値のまま保持されます。

[設定] タブには、利用できる技術に関するオプションが用意されています。[表示] タブには、以前にこのノードに流されたデータに対して確立された分割点が表示されます。

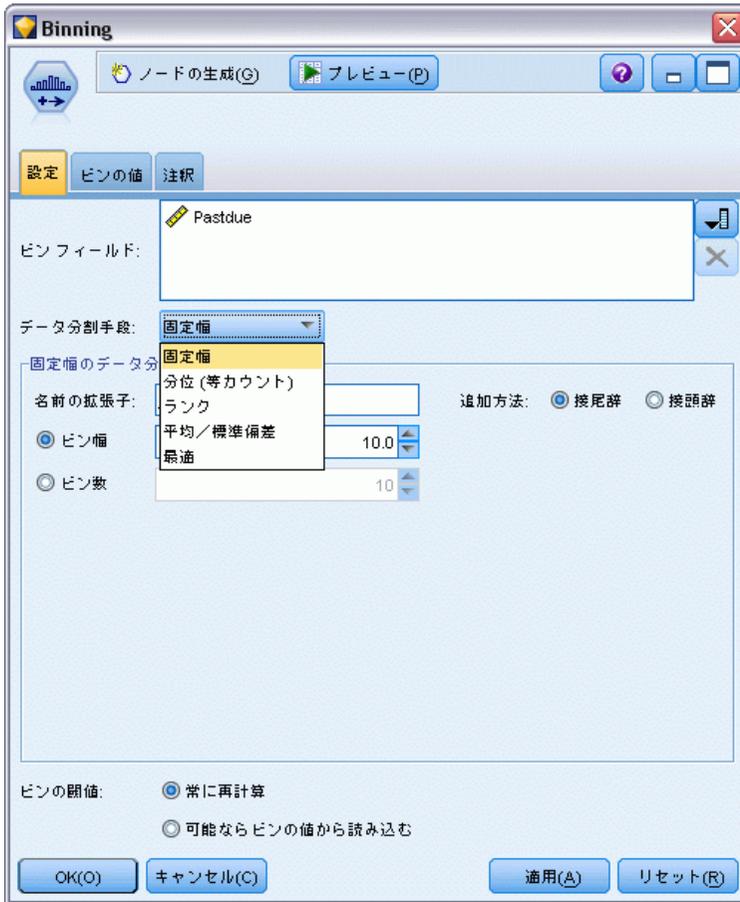
データ分割ノードのオプション設定

データ分割ノードで次の手法を利用して、自動的にビン（カテゴリ）を生成することができます。

- 固定幅のデータ分割
- 分位（等カウントまたは合計）
- 平均と標準偏差
- ランク
- カテゴリの「スーパーバイザ」フィールドに関連する最適化

このダイアログ ボックスの下半分に表示されるオプションは、上で選択したデータ分割方法によって異なります。

図 4-63
データ分割ノードのダイアログボックスの [設定] タブ



ビン フィールド: 変換保留中の連続型（数値範囲）フィールドがここに表示されます。データ分割ノードにより、複数のフィールドを同時にデータ分割することができます。右側のボタンを使って、フィールドを追加または除去してください。

データ分割手段。 新規フィールドのビン（カテゴリ）の分割点を判断する方法を選択します。後続のトピックで、各ケースで使用できるオプションについて説明します。

ビンの閾値: ビンの閾値をどのように計算するかを指定します。

- **常に再計算:** 分割点およびビンの割り当てを、ノード実行時に常に再計算します。
- **可能な場合は [ビン値] タブから読み込む:** 分割点およびビンの割り当てを必要に応じて計算します（新しいデータを追加する場合など）。

次のトピックでは、利用できるデータ分割方法とオプションについて説明していきます。

固定幅のデータ分割

データ分割手段として [固定幅] を選択した場合、ダイアログ ボックスには新しい種類のオプション群が表示されます。

図 4-64
データ分割ノードのダイアログ ボックス：[設定] タブの固定幅データ分割オプション

名前拡張子： フィールドの生成に使用する拡張子を指定します。デフォルトの拡張子は「_BIN」になります。また、拡張子をフィールド名の先頭に追加するか（**接頭辞**）、または最後に追加するか（**接尾辞**）を指定することもできます。たとえば、「income_BIN」という名前の新規フィールドを生成することができます。

ビン幅： ビンの「幅」を算出するために使用する値（整数または実数）を指定します。たとえば、デフォルト値の 10 を使って、フィールド [年齢] を分割することができます。[年齢] フィールドの範囲は 18-65 のため、生成されるビンは次のようになります。

テーブル 4-1
18-65 の範囲の年齢のビン

| ビン 1 | ビン 2 | ビン 3 | ビン 4 | ビン 5 | ビン 6 |
|-------------|-------------|-------------|-------------|-------------|-------------|
| >=13 to <23 | >=23 to <33 | >=33 to <43 | >=43 to <53 | >=53 to <63 | >=63 to <73 |

ビンの開始点は、検出されたもっとも低い値から、ビン幅（指定された）の半分を減算したものになります。たとえば、上記の例のビンでは、次の計算によりビンの開始点が算出されます。18 [もっとも低いデータ値] - 5 [0.5 × (ビン幅 10)] = 13.

ビン数： このオプションを使って、新規フィールドの固定幅ビン（カテゴリ）数を判断するために使用する整数を指定します。

ストリーム中でデータ分割ノードを実行した後に、データ分割ノードの [プレビュー] タブをクリックして、生成されたビンの閾値を表示できます。詳細は、[p. 223 生成されたビンのプレビュー](#) を参照してください。

分位（等カウントまたは合計）

データ分割手段は、検出されたレコードを 100 分位（または 4 分位、10 分位他）グループへの分割に使用できる名義型フィールドを作成します。そのため、各グループは同じ番号のレコードを含むか、または各グループの合計が等しくなります。レコードは、指定されたビン フィールドの値に基づいて昇順でランク付けされます。したがって、選択されたビンの変数の一番低い値はランク 1 に割り当てられ、次のレコード セットはランク 2 というように割り当てられます。各ビンの閾値は、データと使用されている分位方法に基づいて自動的に生成されます。

図 4-65
データ分割ノードのダイアログ ボックス：[設定] タブの等カウント データ分割オプション

分位名の拡張子：標準のパーセンタイルを使って生成されるフィールドに対して使用する拡張子を指定します。デフォルトの拡張子は、_TILE に N を付けたものになります。N は分位数です。また、拡張子をフィールド名の先頭に追加するか（接頭辞）、または最後に追加するか（接尾辞）を指定することもできます。たとえば、「income_TILE4」という名前の新規フィールドを生成することができます。

ユーザー設定の分位の拡張子：カスタム分位範囲に使用する拡張子を指定します。デフォルトは [_TILEN] です。この場合、N が顧客番号で置換されることがないことに注意してください。

利用できる分位を次に示します。

- **4 分位：**それぞれが 25% のケースを含む、4 のビンを生成します。
- **5 分位：**それぞれが 20% のケースを含む、5 のビンを生成します。
- **10 分位：**それぞれが 10% のケースを含む、10 のビンを生成します。
- **20 分位：**それぞれが 5% のケースを含む、20 のビンを生成します。

- **パーセンタイル**：それぞれが 1% のケースを含む、100 のビンを生成します。
- **カスタム N**：ビンの数を指定するために選択します。たとえば、3 を指定すると、それぞれが 33.3 % のケースを含む 3 つのカテゴリ (2 つの分割点) が生成されます。

データ内の離散型の値が指定された分位数より少ない場合は、すべての分位が使用されません。このような場合、新しい分布は元のデータ分布を反映する可能性があります。

分位方法：レコードをビンに割り当てるときに使用する方法を指定します。

- **レコード件数**：各便に等しい数のレコードが割り当てられるよう調べます。
- **合計値**：各ビンの中の値の合計が等しくなるようなビンにレコードを割り当てよう調べます。たとえば営業成績を対象とする場合、この方法を使用して見通しをレコードごとの値に基づき、最上位のビンで最も高い見通しの値により 10 分位グループに割り当てることができます。たとえば、製薬会社は、医師を書いた処方箋の数に基づいて 10 分位にランク付けするかもしれません。各 10 分位にはほぼ同数の処方箋が含まれますが、これらの処方箋に貢献している医師の数は、10 分位に集中したほとんどの処方箋を書く医師により、同じではありません。

タイ：分割点の両側の値が同じ場合、結果は「タイ」状態になります。たとえば、10 分位を割り当てており、10 % 以上のレコードがビンフィールドに対して同じ値を持っている場合、閾値を一方または他方に合わせない限り、その値はすべて同じビンに適合しません。タイを次のビンに持ち上げるか現在のビンにとどめておくことができますが、一部のビンが想定値以上の値を持つことになった場合でも、同じ値を持つすべてのレコードが同じビンに分類されるようにします。次のビンの閾値も、タイを解決する方法に従って同じ数値のセットに対して別のやり方で値を割り当てられるように、結果に合わせて調整することもできます。

- **隣へ追加**：タイ値を次のビンに移動するように選択します。
- **現在のまま保持**：タイ値を現在の (低い) ビンに保持します。この方法は生成されるビンの数を結果として少なくします。
- **無作為割当**：タイ値をビンに無作為に割り当てるときに選択します。各ビンのレコード数が等しい数になるようにします。

例:レコード カウントによる分位

以下に示す表は、単純化されたフィールド値が、どの程度レコード カウントによって分位された場合に 4 分位にランク付けされるかを示しています。この結果は選択したタイオプションによって異なります。

| 値 | 隣へ追加 | 現在のまま保持 |
|----|------|---------|
| 10 | 1 | 1 |
| 13 | 2 | 1 |
| 15 | 3 | 2 |
| 15 | 3 | 2 |
| 20 | 4 | 3 |

ビンあたりの項目数は、次のように算出されます。

値の総数 / 分位数

この例では、ビンあたりの望ましい項目数は 1.25 になります (5 つの値 / 4 分位)。値 13 (値番号 2) は望ましいカウント閾値の 1.25 にまたがっているため、選択されているタイ オプションに応じて扱いが異なります。[隣へ追加] の場合、この値はビン 2 に割り当てられます。[現在のまま保持] の場合、この値はビン 1 に割り当てられ、ビン 4 に割り当てられる値の範囲は、既存のデータ値の範囲外の値になってしまいます。その結果、3 つのビンのみが生成され、各ビンの閾値はその結果に従って調整されます。

図 4-66
生成されたビンの閾値



注: 分位によるデータ分割スピードは、並行処理を可能にします。詳細は、5章 [ストリームの最適化オプションの設定 in IBM SPSS Modeler 15 ユーザーガイド](#) を参照してください。

ケースのランク付け

データ分割手段として [ランク] を選択した場合、ダイアログボックスには新しい種類のオプション群が表示されます。

図 4-67
データ分割ノードのダイアログボックス: [設定] タブのランク付けオプション



ランク付けでは、指定されたオプションに応じて、ランク、ランクの比率、および数値フィールドのパーセンタイル値を含む新規フィールドが作成されます。

ランク順序:[昇順] (もっとも小さい値が 1 となる) または [降順] (もっとも大きい値が 1 となる) を選択します。

ランク: 上で指定したように、ケースを昇順または降順でランク付けする場合に選択します。新規フィールドの値の範囲は 1-N となります。ここで、N は元のフィールド中の不連続値の数を表します。同順位の値には、そのランクの平均が与えられます。

ランクの比率: 新規フィールドの値が、ランクを非欠損ケースの重みの合計で除算した値になるように、ケースをランク付けする場合に選択します。ランクの比率は 0 - 1 の範囲の値になります。

ランクのパーセンテージ: 各ランクが、有効な値を持つレコード数で除算された後、100 倍されます。ランクのパーセンテージは、1-100 の範囲の値になります。

拡張: すべてのランクオプションで、カスタム拡張子を作成し、拡張子をフィールド名の先頭に追加するか (接頭辞)、または最後に追加するか (接尾辞) を指定することができます。たとえば、「income_P_RANK」という名前の新規フィールドを生成することができます。

平均/標準偏差

データ分割手段として [平均/標準偏差] を選択した場合、ダイアログ ボックスには新しい種類のオプション群が表示されます。

図 4-68
データ分割ノードのダイアログ ボックス: [設定] タブの平均/標準偏差オプション

この方法では、指定したフィールドの分布の平均および標準偏差の値に基づいて、バンド カテゴリを持つ 1 つまたは複数の新規フィールドが生成されます。以下から、使用する偏差数を選択してください。

名前の拡張子: フィールドの生成に使用する拡張子を指定します。デフォルトの拡張子は「_SDBIN」になります。また、拡張子をフィールド名の先頭に追加するか (接頭辞)、または最後に追加するか (接尾辞) を指定することもできます。たとえば、「income_SDBIN」という名前の新規フィールドを生成することができます。

- **+/- 1 標準偏差:**3 つのビンを作成する場合に選択します。
- **+/- 2 標準偏差:**5 つのビンを作成する場合に選択します。
- **+/- 3 標準偏差:**7 つのビンを作成する場合に選択します。

たとえば、[+/-1 標準偏差] を選択すると、以下のように計算されて 3 つのビンが生成されます。

| ビン 1 | ビン 2 | ビン 3 |
|---------------------------------------|---|---------------------------------------|
| $x < (\text{Mean} - \text{Std. Dev})$ | $(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$ | $x > (\text{Mean} + \text{Std. Dev})$ |

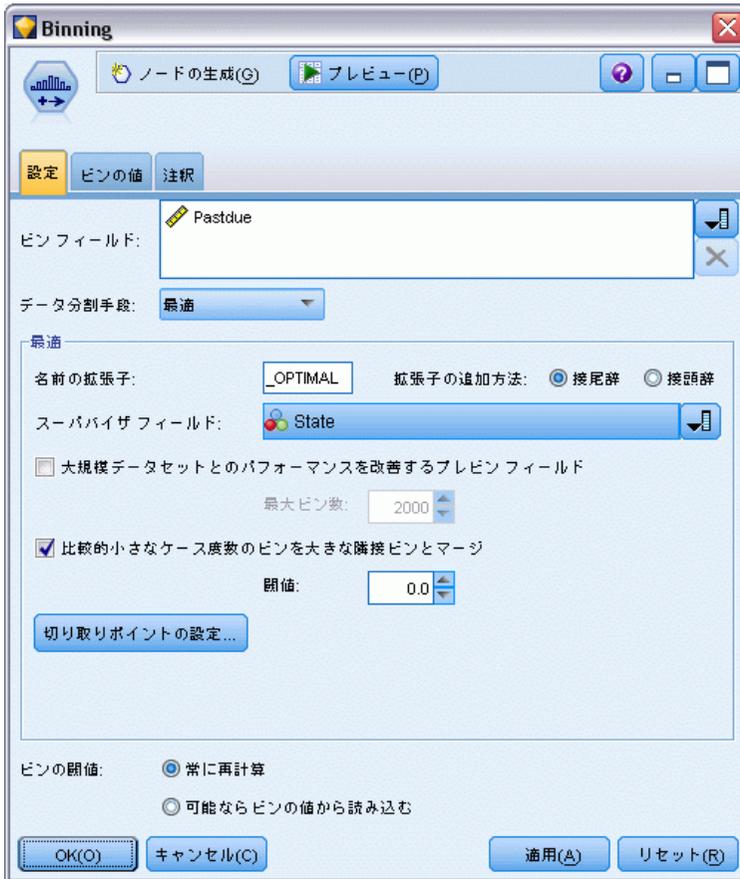
正規分布では、68% のケースが平均の 1 標準偏差に、95% が 2 標準偏差に、そして 99% が 3 標準偏差にあてはまります。ただし、標準偏差に基づいてバンド カテゴリを作成すると、一部のビンが実際のデータ範囲外に定義されたり、取り得るデータ値の範囲外（たとえば、負の給与範囲など）になる可能性があることに注意してください。

最適カテゴリ化

データ分割対象のフィールドが別のカテゴリ フィールドと強力に関連付けられている場合は、2 つのフィールド間の当初のアソシエーションの強度を保つ方法でビンを作成するために、カテゴリ フィールドを「スーパバイザ」フィールドとして選択することができます。

たとえば、クラスタ分析を使用して、住宅ローンの不履行率に基づき、最初のクラスタの最も高い割合で状態を分類するとします。この場合、「スーパバイザ」フィールドとしてのモデルによって生成されたビン フィールドや所属クラスタ フィールドとして「期日経過率」および「請戻権喪失率」を選択します。

図 4-69
最適なまたは監視されるデータ分割のオプション



名前の拡張子：生成されるフィールドに使用する拡張子と、それをフィールド名の先頭（接頭辞）または最後（接尾辞）に付け加えるかどうかを指定します。たとえば、「pastdue_OPTIMAL」および「infoclosure_OPTIMAL」という名前の新規フィールドを生成することができます。

スーパーバイザフィールド：ビンの作成に使用されるカテゴリ フィールド。

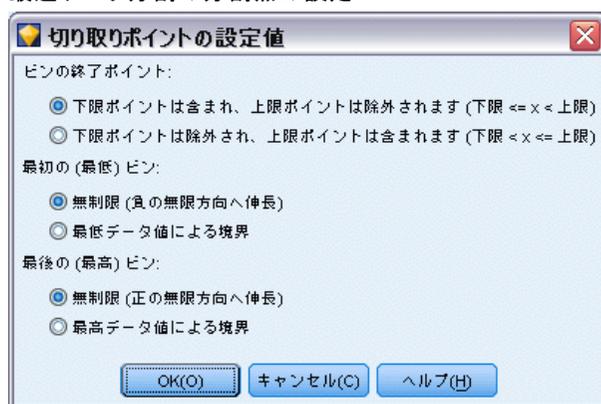
大規模データセットのパフォーマンスを向上させるビン前のフィールド：最適なデータ分割を簡略化するのに前処理を使用すべきかどうかを示します。単純な、監視されないデータ分割方法を使用して大量のビンへ値が振り分けられ、各ビン内で値はその平均により表現されて、監視されるデータ分割に進む前に、重みに応じてケースが調整されます。実際的な問題として、この方法は、正確さを犠牲にしても速度が速いほうを採用し、大規模データセット向けに推奨されます。このオプションが使用される場合に、事前処理後変数が終了するビンの最大数を指定することもできます。

比較的小さなケース度数のビンを大きな隣接ビンとマージ： このオプションが有効な場合、隣接のビンに対するサイズの比率（ケース数）が指定された閾値よりも小さい場合ビンが結合されることを示します。閾値が大きいほど結合が多くなります。

分割点の設定

[分割点の設定] ダイアログ ボックスで、最適データ分割アルゴリズムの詳細設定を指定することができます。これらの設定は、アルゴリズムに対象フィールドを使用してビンを計算する方法を指示します。

図 4-70
最適データ分割の分割点の設定



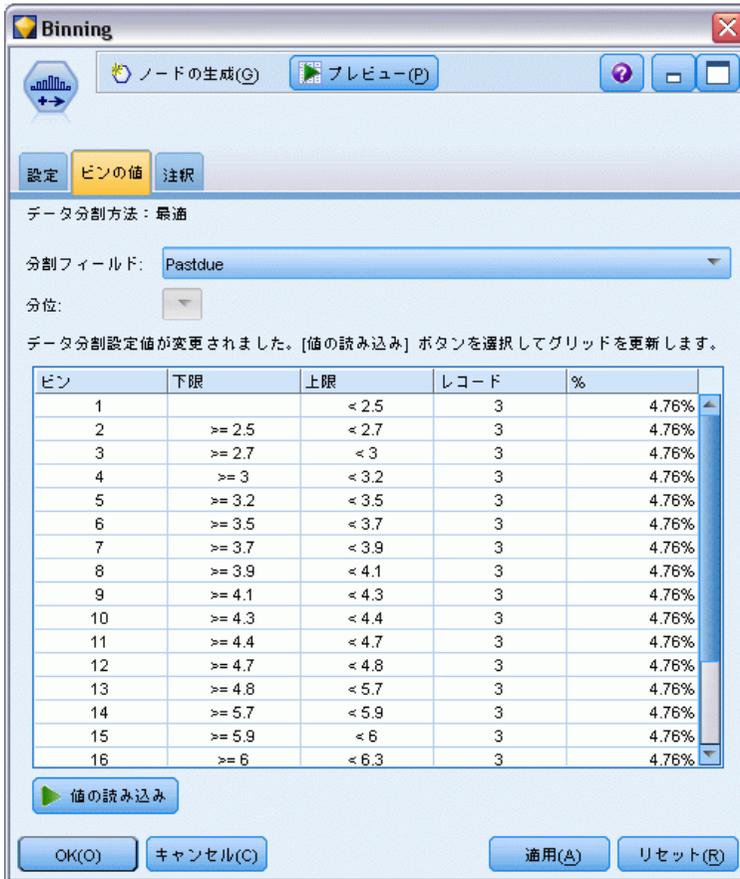
ビンの終点： 下位または上位の終点が包括的（下位 $\leq x$ ）か、排他的（下位 $< x$ ）かを指定できます。

最初のビンと最後のビン： 最初と最後のビン両方に対し、ビンに境界がないか（正または負の方向に無限に拡張）、最低または最高のデータポイントによる境界があるかを指定できます。

生成されたビンのプレビュー

データ分割ノードの [ビンの値] タブで、生成されたビンの閾値を表示できます。[生成] メニューを使用して、あるデータセットから別のデータセットへこの閾値を適用するのに使用できる、フィールド作成ノードも生成できます。

図 4-71
データ分割ノードのダイアログ ボックス: [ビンの値] タブ



分割フィールド: ドロップダウン リストから、表示するフィールドを選択します。表示されるフィールド名は、元のフィールド名を使用しています。

分位: ドロップダウン リストを使って、10 や 100 などの、表示する分位数を選択します。このオプションは、分位方法（等カウントまたは等合計）でビンが生成された場合にだけ利用できます。

ビンの閾値: 生成されたビンごとに、ここに閾値が表示されます。そのビンに分類されたレコード数も表示されます。最適化されたデータ分割方法の場合のみ、各ビンのレコード数が、全体の割合として表示されます。閾値は、ランク付けデータ分割方法を使用中の場合に適用できません。

値の読み込み: データ セットからビンに分けられた値を読み込みます。ストリームに新しいデータが流されると、閾値は上書きされます。

フィールド生成ノードの生成

[生成] メニューを使用して、現在の閾値に基づいたフィールド作成ノードを作成できます。確立したビンの閾値を、あるデータ セットから別のデータ セットに適用する場合に、このオプションが役立ちます。また、大きいデータ セットに対して作業を行う場合、いったんこれらの分割点 が明らかになれば、データ分割操作よりもフィールド作成操作の方がより効率的に、速く実行できます。

RFM 分析ノード

リーセンシ、フリクエンシ、マネタリー (RFM) の分析ノードを使用すると、最後に購入したのがどのくらい最近か (リーセンシ)、どのくらい頻繁に購入するか (フリクエンシ)、トランザクション全体でいくら消費したか (マネタリー) を検証することによって、最も良い顧客となると考えられるのはどの顧客かを量的に決定することができます。

RFM 分析の推論は、製品またはサービスを購入する顧客がサイド購入する可能性が高いということです。カテゴリ化された顧客データは、多くのビンに分割され、分割基準は必要に応じて調整されます。それぞれのビンで、顧客はスコアに割り当てられます。これらのスコアは結合され、全体の RFM スコアを提供します。このスコアは、それぞれの RFM パラメータに作成されたビンの顧客の所属を表します。この分割されたデータは、たとえば最も頻繁に取引し、支出の高い顧客を識別することによってニーズを満たすことができます。また、詳細なモデル作成および分析のためにストリーム内に渡される場合があります。

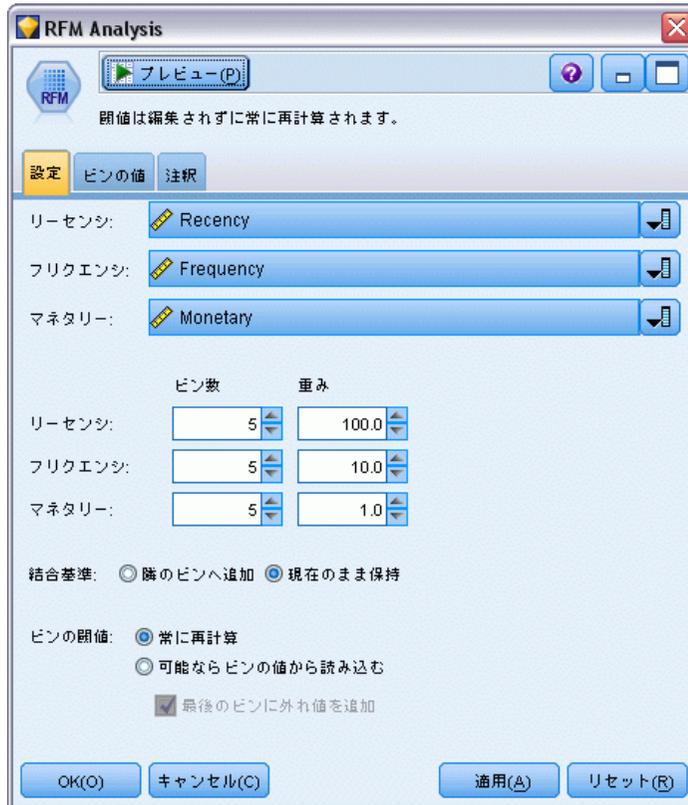
ただし、RFM スコアを分析しタンク付けする機能は役に立つツールですが、使用する場合は特定の要素に注意する必要があります。高くランク付けされた対象の顧客を勧誘する場合がありますが、これらの顧客の過剰な勧誘は不快感を呼び、取引の繰り返しが実際は失敗してしまう恐れがありますので注意してください。また、低いスコアの顧客は無視することはせず、より良い顧客を開拓することができることを記憶しておく価値があります。それに対し、市場によっては高いスコアだけが必ずしも良好な販売の可能性を反映するわけではありません。たとえば、リーセンシを表すビン 5 の最近購入した顧客は、車やテレビなど効果で長持ちする商品を販売する者にとっては、対象となる顧客ではありません。

注：データの保存方法によっては、RFM 分析ノードを RFM レコード集計ノードに先行してデータを使用可能な形式に変換する必要があります。たとえば、入力データは顧客ごとに 1 行の顧客のフォーマットである必要がありますが、顧客のデータがトランザクション フォームである場合、上流で RFM レコード集計ノードを使用してリーセンシ、フリクエンシ、マネタリーのフィールドを作成する必要があります。詳細は、[3 章 p.94 RFM レコード集計ノード](#) を参照してください。

IBM® SPSS® Modeler の RFM レコード集計ノードおよび RFM 分析ノードを設定して独立した分割を使用します。最新性、頻度、金額値の各尺度のデータを、これらの値および尺度に関係なくランク付けし、分割します。

RFM 分析ノードの設定

図 4-72
RFM 分析オプションの設定



リーセンシ： フィールド ピッカー（テキスト ボックスの右側にあるボタン）を使って、リーセンシのフィールドを選択します。このフィールドは日付、タイムスタンプまたは単純な数値です。日付またはタイムスタンプが最も新しいトランザクションの日付を示す場合、最も高い値が最新のものと見なします。数値が指定されている場合、数値は最新のトランザクションから経過した時間を表し、最も低い値が最新であると見なします。

注： RFM レコード集計ノードが RFM 分析ノードに先行する場合、RFM レコード集計ノードに生成されたリーセンシ、フリクエンシ、マネタリーのフィールドが RFM 分析ノードの入力として選択されます。

度数: フィールド ピッカーを使用して、使用するフリクエンシのフィールドを選択します。

マネタリー: フィールド ピッカーを使用して、使用するマネタリーのフィールドを選択します。

ビン数。 それぞれの 3 つの出力タイプに対し、作成するビン数を指定します。デフォルトは 5 です。

注： ビン数の最小値は 2 で、最大値は 9 です。

重み: デフォルトでは、スコア計算時最も高い重要度がリーセンシのデータに与えられ、次にフリクエンシ、マネタリーの順に与えられます。必要に応じて、これらのフィールドに影響する重みを修正して、高い重要度を与えるフィールドを変更します。

RFM スコアは次のように計算されます。(リーセンシ スコア x リーセンシの重み) + (フリクエンシ スコア x フリクエンシの重み) + (マネタリー スコア x マネタリーの重み)

タイ: 同じ (タイ) のスコアがどのように分割されるかを指定します。オプションは次のとおりです。

- **隣へ追加:** タイ値を次のビンに移動するように選択します。
- **現在のまま保持:** タイ値を現在の (低い) ビンに保持します。この方法は生成されるビンの数を結果として少なくします。(デフォルトの設定です。)

ビンの閾値: ノードが実行された場合に RFM スコアおよびビンの割り当てが常に再計算されるかどうか、必要な場合にのみ計算されるか (たとえば、新しいデータが追加されて場合) を指定します。[可能ならビンの値から読み込む] を選択すると、[ビンの値] タブでさまざまなビンの上限および下限の分割点を編集することができます。

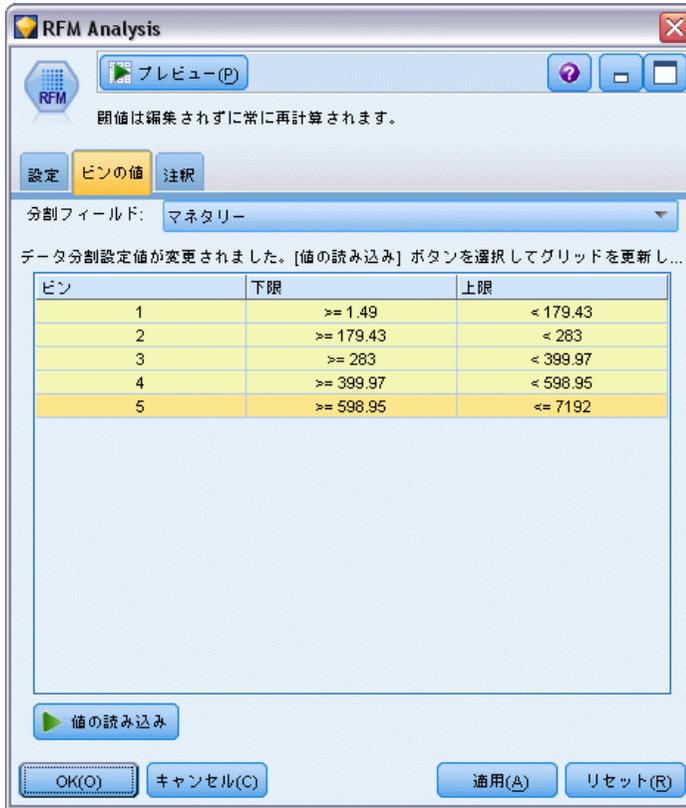
実行時、RFM 分析ノードは処理されていないリーセンシ、フリクエンシ、マネタリーのフィールドを分割し、次の新しいフィールドをデータセットに追加します。

- リーセンシ スコア。リーセンシのランク (ビン値)
- フリクエンシ スコア。フリクエンシのランク (ビン値)
- マネタリー スコア。マネタリーのランク (ビン値)
- RFM スコア。リーセンシ、フリクエンシ、マネタリー スコアの重みの合計

終わりのビンに外れ値を追加: このチェック ボックスを選択した場合、最も低いビンの下にあるレコードが最も低いビンに追加され、最も高いビンの上にあるレコードは最も高いビンに追加されます。選択していない場合は、ヌル値が与えられます。このチェック ボックスは、[可能ならビンの値から読み込む] が選択されている場合にのみ使用できます。

RFM 分析ノードの分割

図 4-73
RFM 分析のビンの値の設定



[ビンの値] タブを使用すると、生成されたビンの閾値を表示でき、またある場合は修正こともあります。

注： このタブでは、[設定] タブの [可能ならビンの値から読み込む] が選択されている場合にのみ、値を修正できます。

分割フィールド： ドロップダウン リストから、ビンに分割するフィールドを選択します。[設定] タブで選択された値を使用できます。

ビンの値のテーブル： 生成された各ビンの閾値がここに表示されます。[設定] タブで [可能ならビンの値から読み込む] を選択すると、関連するセルをダブルクリックして、各ビンの上限および下限の分割点を修正することができます。

値の読み込み： データセットから分割された値を読み込み、ビンの値のテーブルを作成します。[設定] タブで [常に再計算] を選択した場合は、新しいデータがストリームで実行されると、ビンの閾値が上書きされます。

データ区分ノード

データ区分ノードは、モデル構築の学習、テスト、および検証の各ステージ用に、データを独立したサブセット（サンプル）に分割するフィールドを生成するために使用されます。1 組のサンプルをモデルの生成に使用し、別の組のサンプルで生成したモデルをテストすることにより、そのモデルが、このデータに似た性質を持つより大きなデータセットにどの程度適用できるかについての良い目安を得ることができます。

データ区分ノードは、役割が **[データ区分]** に設定された名義型フィールドを生成します。適当なフィールドが既に存在している場合、データ型ノードを使用すると、そのフィールドをデータ区分として指定できます。この場合、新しいデータ区分ノードは必要ありません。2 つまたは 3 つの値を持つインスタンス化された設定フィールドをデータ区分として使用できますが、名義型フィールドは使用できません。 [詳細は、p. 167 フィールドの役割の設定](#) を参照してください。

単一のストリーム内で複数のデータ区分フィールドを定義できますが、その場合、データ分割を使用する各モデリング ノードごとに **[フィールド]** タブでデータ区分フィールドを 1 つだけ選択しなければなりません。(1 つのデータ区分だけが存在している場合、データ分割を有効にすると、そのデータ区分が必ず自動的に使用されます)。

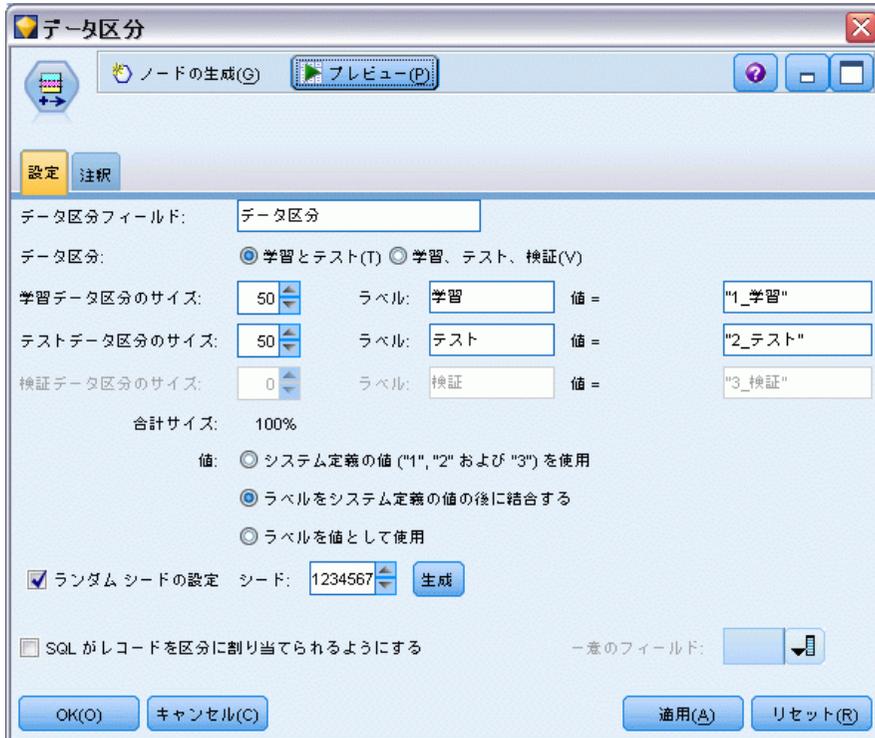
データ分割を有効にする: 分析でデータ区分を使用するには、適当なモデル構築または分析ノードの **[モデルのオプション]** タブでデータ分割を有効にする必要があります。このオプションの選択を解除すると、フィールドを削除しないでデータ分割を無効にできます。

データ範囲や場所などの他のいくつかの基準に基づいてデータ区分フィールドを生成する場合、フィールド作成ノードを使用できます。 [詳細は、p. 185 フィールド作成ノード](#) を参照してください。

例: RFM ストリームを構築して、以前のマーケティング キャンペーンに肯定的に応答した最近の顧客を識別する場合、販売会社のマーケティング部門ではデータ区分ノードを使用して、データを学習用データ区分および検定用データ区分に分割します。

データ区分ノードのオプション

図 4-74
データ区分ノードのダイアログ ボックス: [設定] タブ



データ区分フィールド: そのノードにより生成されるフィールドの名前を指定します。

データ区分: データを、2 組（学習とテスト）または 3 組（学習、テスト、および検証）のサンプルに分割できます。

- **学習とテスト:** データを 2 つのサンプルに区分し、一方のサンプルを使用して学習し、もう一方のサンプルを使用してテストできるようになります。
- **学習、テスト、および検証:** データを 3 つのサンプルに区分し、1 つのサンプルを使用してモデルを学習し、2 つ目のサンプルを使用してモデルのテストと調整を行い、3 つめのサンプルを使用して結果を検証できるようになります。3 組にすると、結果として各データ区分のサイズが小さくなりますが、作業するデータ セットが非常に大きければ、ほとんどの場合に最適です。

データ区分のサイズ: 各データ区分の相対的なサイズを指定します。各データ区分のサイズの合計が 100% より小さい場合、データ区分に含まれないレコードは、破棄されます。たとえば、ユーザーが 1000 万個のレコードを持っており、学習データ区分のサイズを 5%、テストを 10% に指定した

場合、そのノードの実行後、およそ 500,000 個が学習レコードに、100 万個がテストレコードに割り当てられ、残りは破棄されます。

値： データ中の各データ区分サンプルを表すために使用される値を指定します。

- **システム定義の値（「1」、「2」および「3」）を使用：** 各データ区分を表すのに整数値を使用します。たとえば、全てのレコードが、学習データ区分に含まれる全てのサンプルは、データ区分フィールドの値が 1 になります。これにより、ロケール間でのデータのポータビリティが保証され、データ区分フィールドが別の場所にインストールされた場合にも、ソートの順序が維持されます（1 は学習区分を表します）。ただし、値の解釈にやや手間がかかります。
- **ラベルをシステム定義の値の後に結合する：** ラベルを整数値に結合します。たとえば、学習データ区分レコードの値は 1_Training になります。こうすることにより、人間がデータを見たとき、それぞれの値が何を表しているかが解りやすくなります。しかも、ソートの順序も維持されたままです。ただし、値は、特定のロケールに固有になります。
- **ラベルを値として使用：** 整数値を持たないラベルを使用します。たとえば、学習 です。この場合、ラベルを編集して値を指定できるようになります。ただし、データはロケール固有になり、データ区分列の再インストールすると、値はインストール先の言語のソート順序で処理されるため、本来の「意味上の」順序とは一致しなくなる場合があります。

ランダムシードの設定： 無作為なパーセンテージに基づいてレコードをサンプリングまたはデータ区分している場合、このオプションで、別のセッションに同じ結果を複製できるようになります。乱数ジェネレータに使用される開始値を指定することで、ノードが実行されるごとに毎回同じレコードが割り当てられることが保証されます。自動的に無作為な値を生成するには、希望のシード値を入力するか、[生成] ボタンを入力します。このオプションが選択されないと、ノードが実行されるごとに異なるサンプルが生成されます。

注： データベースから読まれるレコードに [ランダムシードの設定] オプションを使用する場合は、ノードが実行されるごとに同じ結果を保証するために、サンプリングに先行して、ソートノードが必要になる可能性があります。この理由は、ランダムシードがレコードの順序に依存しているためです。各レコードがリレーショナルデータベース内で同じ位置に留まる保証はありません。詳細は、3 章 p.96 ソートノードを参照してください。

SQL を有効化してレコードをデータ区分に割り当てる：（層 1 のデータベースのみ）SQL プッシュバックを使用して、レコードをデータ区分に割り当てます。[一意のフィールド] ドロップダウンから、一意の値を持つフィールド（ID フィールドなど）を選択肢、レコードが無作為にかつ繰り返し割り当てられるようにします。

データベースの階層については、データベース入力ノードの説明に記載されています。詳細は、2 章 p.16 データベース入力ノード を参照してください。

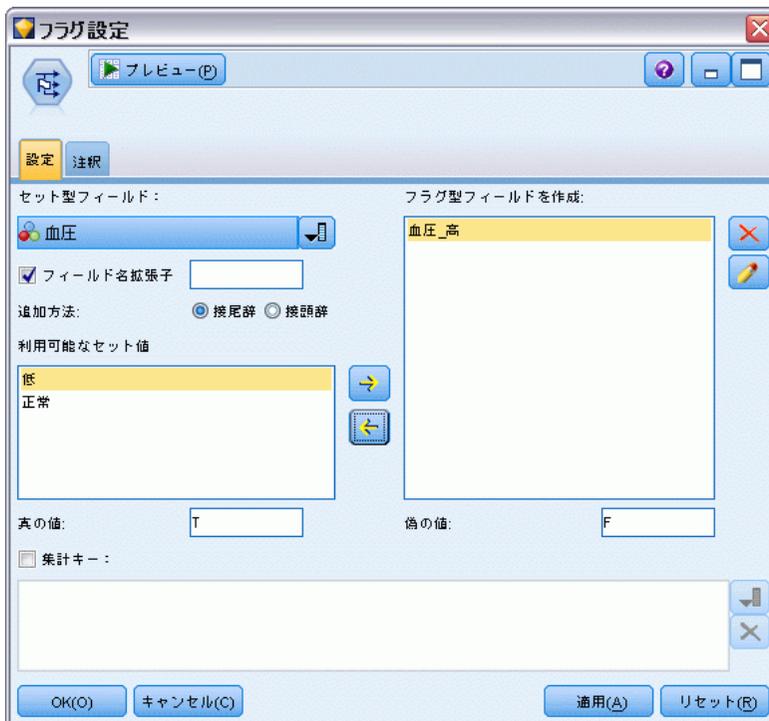
条件抽出ノードの生成

[データ区分] ノードの [生成] メニューを使用すると、各データ区分ごとに [条件抽出ノード] を自動的に生成できます。たとえば、すべてのレコードを学習データ区分に選択して、このデータ区分のみを使用して、さらに評価または分析を続けることができます。

フラグ設定ノード

フラグ設定ノードは、1 つ以上の名義型フィールド用に定義されたカテゴリ値を基にして、フラグ型フィールドを派生させるために使用します。たとえば、データセットには、高、正常、低 という値を持つ名義型フィールド BP（血圧）を含まれる場合があります。データの操作を簡単にするために、高血圧用のフラグ型フィールドを作成し、そこで患者が高血圧であるかどうかを示すことができます。

図 4-75
高血圧のフラグ型フィールドの作成



フラグ設定ノードのオプションの設定

セット型フィールド：「名義型」（セット型）の尺度を持つすべてのデータフィールドが表示されます。リストから 1 つのフィールドを選択して、セット内の値を表示します。これらの値の中から選択して、フラグフィールドを作成することができます。利用可能な名義型フィールドとその値を表示するには、上流のソースまたはデータ型ノードを使用して、データを完全にインスタンス化する必要があります。詳細は、[p.150 データ型ノード](#) を参照してください。

フィールド名拡張子：新しく作成するフラグフィールドの接頭辞または接尾辞として追加する拡張子を指定する場合に選択します。デフォルトでは、フィールド名_フィールド値のように、元のフィールド名にフィールドの値を組み合わせた新規フィールド名が自動的に作成されます。

利用できるセット値：上のフィールドで選択したセット内の値が表示されます。フラグを生成する対象になる 1 つ以上の値を選択します。たとえば、フィールド blood_pressure の値が高、中、および低の場合、高を選択して右のリストに追加することができます。この場合、高血圧を示す値があるレコードに対して、フラグフィールドが作成されます。

フラグ型フィールドを作成：新しく作成されたフラグ型フィールドのリストが表示されます。フィールド名拡張子コントロールを使って、新しいフィールドの命名に関するオプションを指定することができます。

真 (true) の値：フラグを設定するときにノードが使用する真 (true) の値を指定します。デフォルトの値は、T です。

偽 (false) の値：フラグを設定するときにノードが使用する偽 (false) の値を指定します。デフォルトの値は、F です。

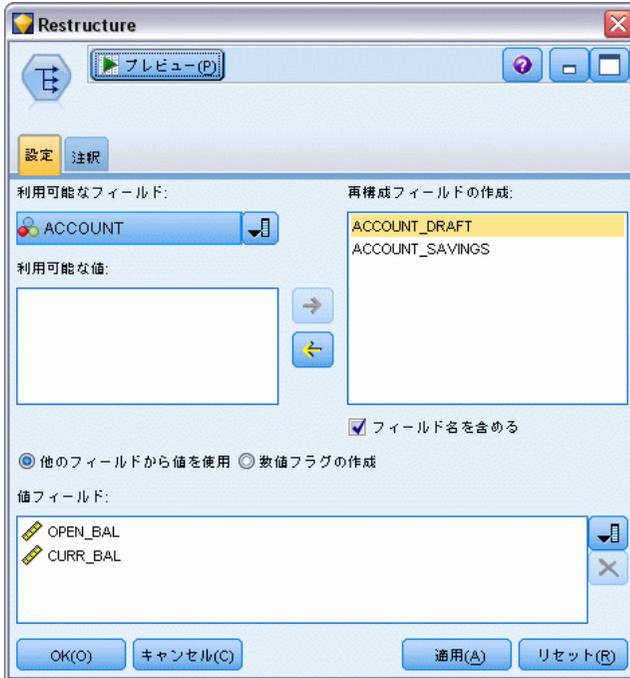
集計キー：下のフィールドで指定するキーフィールドを基にしてレコードをグループ化する場合に選択します。[集計キー] を選択している場合には、真に設定されたレコードが 1 つでもあると、グループ内のすべてのフラグ型フィールドがオンになります。フィールドピッカーを使って、レコードを集計するために使用するキーフィールドを指定してください。

再構成ノード

再構成ノードを使用し、名義型フィールドまたはフラグ型フィールドの値に基づいて複数のフィールドを生成することができます。新たに生成されたフィールドは、他のフィールドからの値、または数値フラグ (0 または 1) を含むことができます。このノードの機能は、フラグ設定ノードの機能と似ています。ただし、より柔軟なノードです。このノードを使用すると、他のフィールドからの値を使用して (数値フラグを含む) どのようなタイプのフィールドでも生成できるようになります。したがって、レコード集計や下流の他のノードを使用する操作が可能になります。(フラグ設定

ノードにより、ワン ステップでフィールドのレコード集計が可能となり、これはフラグ型フィールドを生成する場合に便利です。)

図 4-76
再構成されたアカウントのフィールドを生成



たとえば、次のようなデータ セットは、預金と手形という値を持つ名義型フィールド アカウトを含んでいます。開設時の残高と現在の残高は各アカウントに記録され、顧客は各タイプの複数のアカウントを持っています。特定のアカウント タイプを持つ顧客がいるか、いた場合、各アカウント タイプの残高はいくらか、ということを知りたくなります。再構成ノードを使用して、アカウントの各値に対してフィールドを生成し、値として Current_Balance（現在の残高） を選択します。新しい各フィールドには、あるレコードの現在の残高が書き込まれます。

テーブル 4-2
再構成前のサンプル データ

| CustID | アカウント | Open_Bal | Current_Bal |
|--------|---------|----------|-------------|
| 12701 | ドラフト(D) | 1000 | 1005.32 |
| 12702 | 貯金 | 100 | 144.51 |
| 12703 | 貯金 | 300 | 321.20 |
| 12703 | 貯金 | 150 | 204.51 |
| 12703 | ドラフト(D) | 1200 | 586.32 |

テーブル 4-3
再構成後のサンプル データ

| CustID | アカウント | Open_Bal | Current_Bal | Account_Draft_Current_Bal | Account_Savings_Current_Bal |
|--------|----------|----------|-------------|---------------------------|-----------------------------|
| 12701 | ドラフト (D) | 1000 | 1005.32 | 1005.32 | \$null\$ |
| 12702 | 貯金 | 100 | 144.51 | \$null\$ | 144.51 |
| 12703 | 貯金 | 300 | 321.20 | \$null\$ | 321.20 |
| 12703 | 貯金 | 150 | 204.51 | \$null\$ | 204.51 |
| 12703 | ドラフト (D) | 1200 | 586.32 | 586.32 | \$null\$ |

レコード集計ノードと共に再構成ノードを使用

多くのケースで、再構成ノードとレコード集計ノードを一对として使用したい場合があります。先の例では、ある顧客 (ID 12703) が 3 つのアカウントを持っていました。レコード集計ノードを使用して各アカウントタイプの全体残高を計算します。キーとなるフィールドは CustID であり、レコード集計フィールドは新たに再構成されたフィールド、Account_Draft_Current_Bal と Account_Savings_Current_Bal です。結果を次の表に示します。

テーブル 4-4
再構築とレコード集計後のサンプルデータ

| CustID | Record_Count | Account_Draft_Current_Bal_Sum | Account_Savings_Current_Bal_Sum |
|--------|--------------|-------------------------------|---------------------------------|
| 12701 | 1 | 1005.32 | \$null\$ |
| 12702 | 1 | \$null\$ | 144.51 |
| 12703 | 3 | 586.32 | 525.71 |

再構成ノードのオプション設定

利用可能フィールド: 「名義型」 (セット型) のまた「フラグ型」は尺度を持つすべてのデータ フィールドが表示されます。リストから 1 つのフィールドを選択してセット型 (またはフラグ型) に値を表示します。次に、この値の中から再構成フィールドを生成するための値を選択します。利用可能なフィールドとその値を表示するには、上流のソースまたはデータ型ノードを使用して、データを完全にインスタンス化する必要があります。 [詳細は、 p. 150 データ型ノード を参照してください。](#)

利用可能な値: 上のフィールドで選択したセット内の値が表示されます。再構成フィールドを生成する対象になる 1 つ以上の値を選択します。たとえば、フィールド 血圧 の値が高、中、および低の場合、高を選択して右のリストに追加することができます。これで、高 の値を持つレコードに対して指定した値 (以下参照) を持つフィールドが生成されます。

再構成フィールドの作成：新しく作成された再構成フィールドのリストが表示されます。デフォルトでは、フィールド名_フィールド値のように、元のフィールド名にフィールドの値を組み合わせた新規フィールド名が自動的に作成されます。

フィールド名を含める：新しいフィールド名から接頭辞としての元のフィールド名を削除するために選択解除します。

他のフィールドから値を使用：再構築されたフィールドに書き込まれる値を持つ 1 つ以上のフィールドを指定します。フィールド ピッカー ボタンを使用して 1 つ以上のフィールドを選択します。選択された各フィールドに対して、新しいフィールドが 1 つ生成されます。フィールド名の値が、再構成されたフィールド名にレコード追加されます。たとえば BP_High_Age または BP_Low_Age となります。各新規のフィールドは、元の値フィールドのデータ型を引き継いでいます。

値フラグの作成：選択すると、他のフィールドからの値を使用せずに、新規のフィールドに数値フラグ (0=偽、1=真) を書き込みます。

行列入替ノード

デフォルトでは、列とフィールドおよび行は、レコードと観測値が書き込まれます。必要に応じて、行列入替ノードを使用して、行にあるデータと列にあるデータを入れ替えて、フィールドをレコードに、レコードをフィールドにすることができます。たとえば、各時系列データが列ではなく行に書き込まれている時系列データを持っている場合、分析前にそのデータを入れ替えることができます。

図 4-77
行列入替ノードの [設定] タブ



行列入替ノードのオプション設定

新規フィールド名

新しいフィールド名は、指定された接頭辞 (Prefix) に基づいて自動的に作成できます。または、既存のデータ内のフィールドからフィールド名を読み込むことができます (Read)。

接頭辞を使用： このオプションは、新しいフィールド名を指定した接頭辞 (Field1、Field2、など) に基づいて自動的に生成します。必要に応じて接頭辞をカスタマイズできます。このオプションを使用する場合は、元のデータの行数に無関係に、生成するフィールド数を指定する必要があります。たとえば、[フィールド数] が 100 に設定されると、最初の 100 行を超えるすべてのデータは破棄されます。元のデータが 100 行に満たない場合は、一部のフィールドはヌルとなります。(必要に応じてフィールド数を増やすことができますが、この設定の目的は、100 万のレコードを

100 万のフィールドに行列入替することを防ぐことにあり、もしそのような入れ替えを行うと管理不能となります。)

たとえば、行内に系列データがあり、各月の独立したフィールド（列）があると想定します。各系列が個別のフィールドに、各月が行になるように、行列の入れ替えを実行できます。

図 4-78
行内の連続する元のデータ

| | Jan | Feb | Mar | Apr |
|---|-----|-----|-----|-----|
| 1 | 1 | 3 | 5 | 7 |
| 2 | 2 | 4 | 6 | 8 |

図 4-79
列内の連続する入れ替えられたデータ

| | Month | Field1 | Field2 |
|---|-------|--------|--------|
| 1 | Jan | 1 | 2 |
| 2 | Feb | 3 | 4 |
| 3 | Mar | 5 | 6 |
| 4 | Apr | 7 | 8 |

注：この結果を作成するために、[新規フィールド数] オプションが 100 から 2 へ変更され、[行 ID 名] が [ID] から [月] に変更されました（下記参照）。

フィールドから読み込み：既存のフィールドからフィールド名を読み込みます。このオプションを使用すると、新規フィールド数が指定された最大数を上限として、データによって決定します。選択されたフィールドの各値

は、出力データの新しいフィールドとなります。選択されたフィールドには（整数、文字列、日付など）任意のストレージタイプがありますが、フィールド名の重複を避けるために、選択されたフィールドの各値は一意である必要があります（つまり、値の数は行数と一致する必要があります）。フィールド名が重複している場合、警告が表示されます。

図 4-80
既存のフィールドからフィールド名を読み込む



- **値の読み込み:** 選択されたフィールドがインスタンス化されていない場合、このオプションを選択して新しいフィールド名のリストを編成します。フィールドがすでにインスタンス化されている場合、この手順は必要ではありません。
- **読み込む値の最大数:** データからフィールド名を読み込む場合、上限値を指定してあまりに多いフィールドが作成されることを回避します。（前述のとおり、100 万のレコードを 100 万のフィールドに入れ替えると、管理できない結果が生じます。）

たとえば、データの最初の列が各シリーズの名前を指定している場合、これらの値を入れ替えられたデータのフィールド名として使用することができます。

図 4-81
単一行内の連続する元のデータ

| | date | 1949-01-01 | 1949-02-01 | 1949-04-01 | 1949-05-01 | 1949-06-01 | 1949-07-01 | 1949-08-01 |
|---|------------|------------|------------|------------|------------|------------|------------|------------|
| 1 | passengers | 112.000 | 118.000 | 129.000 | 121.000 | 135.000 | 148.000 | 148.000 |

図 4-82
列内の連続する入れ替えられたデータ

| | Date | passengers |
|----|------------|------------|
| 1 | 1949-01-01 | 112.000 |
| 2 | 1949-02-01 | 118.000 |
| 3 | 1949-04-01 | 129.000 |
| 4 | 1949-05-01 | 121.000 |
| 5 | 1949-06-01 | 135.000 |
| 6 | 1949-07-01 | 148.000 |
| 7 | 1949-08-01 | 148.000 |
| 8 | 1949-09-01 | 136.000 |
| 9 | 1949-10-01 | 119.000 |
| 10 | 1949-11-01 | 104.000 |
| 11 | 1949-12-01 | 118.000 |
| 12 | 1950-01-01 | 115.000 |
| 13 | 1950-02-01 | 126.000 |
| 14 | 1950-03-01 | 141.000 |
| 15 | 1950-04-01 | 135.000 |
| 16 | 1950-05-01 | 125.000 |
| 17 | 1950-06-01 | 149.000 |
| 18 | 1950-07-01 | 170.000 |
| 19 | 1950-08-01 | 170.000 |
| 20 | 1950-09-01 | 158.000 |

入れ替え :デフォルトでは、連続型（数値範囲）フィールドのみが入れ替わります（整数または実数）。オプションで、数値型フィールドのサブセットを選択するか、代わりに文字列フィールドを入れ替えることができます。ただし、すべての入れ替えられたフィールドは同じストレージタイプ（@数型または文字列のいずれか）である必要があります。それは入力

フィールドを混在させると、各出力列内で混在した値が生成され、フィールドのすべての値が同じストレージである必要があるという規則に反するためです。その他のストレージタイプ（日付、時間、タイムスタンプ）を入れ替えることはできません。

- **すべての数値型：**すべての数値型フィールド（整数または実数ストレージ）を入れ替えます。出力の行数は、下のデータの数値型フィールド数に一致します。
- **すべての文字列：**すべての文字列フィールドを入れ替えます。
- **ユーザー指定：**数値型フィールドのサブセットを選択することができます。出力の行数は、選択したフィールド数に一致します。注：このオプションは、数値型フィールドにのみ使用できます。

行 ID 名：ノードで作成された行 ID フィールドの名前を指定します。このフィールドの値は、下のデータのフィールド名によって決まります。

ヒント：行から列へ時系列データを入れ替える際、下のデータに各測定値の期間にラベルを付ける日付、月、または年などの行が含まれる場合、データの最初の行のラベルを含めるのではなく、これらのラベルを IBM® SPSS® Modeler にフィールド名として読み込みます（前述の例で説明したとおり、下のデータの月または日付をフィールド名として表示します）。これにより、各列のラベルと値の混在を回避します（ストレージタイプが列内で混在しないため、数値を文字列として読み込むことを強制します）。

時間区分ノード

時間区分ノードを使用すると、区分を指定し、推測や予測のための時系列モデルの作成または時系列ノードで使用される時系列データのラベルを生成できるようになります。全時間区分がサポートされ、範囲は秒から数年の間を指定することができます。たとえば、2005 年 1 月 3 日から始まる毎日の測定値の時系列データを持っているとします。その日に始まるレコードにラベルを付け、2 番目の行に 1 月 4 日のラベルを付け、以降このようにラベルを付けることができます。周期性も指定できます。たとえば、1 週間 5 日、または 1 日 8 時間。

さらに、推測に使用するレコードの範囲を指定することができます。系列の最初の値を除外するか、ホールドアウトを指定するかを選択することができます。そうすると、わかっている値と該当する期間の推測値を比較するために、時系列データの最新レコードをホールドアウトしてモデルをテストできるようになります。

また、今後どのくらいの期間の予測を行うのかを指定することができ、下流の時系列モデル作成ノードによる予測で使用する将来値を指定することができます。

この時間区分ノードは、各レコードに一意の整数を割り当てる TimeIndex フィールドと共に、指定された区分と周期に適切な形式で、TimeLabel フィールドを生成します。追加フィールド数も、選択した周期や間隔（測定値が範囲内に収まる分や秒など）に従って生成されます。

測定値が均等に配置されるように、必要な値を埋め込んだり、または集計したりできます。時系列データをモデル作成する方法では、欠損値を空の行で示し、各測定間を均一な区分とすることが求められます。データがこの要件を満たしていない場合、このノードがデータをそのように変更することができます。

コメント

- 周期は実際の時間と一致しない可能性があります。たとえば、標準的な 1 週間 5 日を基にした時系列は、金曜日から月曜日までの間のギャップを 1 日として扱います。
- 時間区分ノードは、各時系列がフィールドまたは列に置かれ、測定値が各行に置かれることを前提としています。必要に応じて、この要件に一致するようにデータの行列入替をすることができます。 [詳細は、p.236 行列入替ノード を参照してください。](#)
- 均等な間隔でない時系列に対しては、各測定値に対応する日付または時間を示すフィールドを指定することができます。入力として使用するには、適切なフォーマットの日付、時間、またはタイムスタンプのフィールドが必要です。必要な場合には、文字列ラベル フィールドなどの既存フィールドを、置換ノードを使用してこのフォーマットへ変換できます。 [詳細は、p.200 置換ノードを使ったストレージの変換 を参照してください。](#)
- 生成されたラベルとインデックス フィールドを表示する場合、値ラベルの表示をオンにしておくに役に立ちます。たとえば、月別データ用に生成された値を持つテーブルを表示させる場合、ツールバーにある値ラベル アイコンをクリックし、1、2、3、... などではなく、1 月、2 月、3 月... と表示させることができます。

図 4-83
値ラベル アイコン



時間区分を指定

[区分] タブを使用すると、時系列を構築、またはラベリングのための区間と周期性を指定できるようになります。具体的な設定は選択した区分によって異なります。たとえば、[1 日あたりの時間数] を選択すると、週あたりの日数、週の始まりの日、各日の時間数、各日の始まりの時間を指定できます。 [詳細は、p.250 サポートされる区分 を参照してください。](#)

図 4-84
時の時系列の時間区分設定

時間区分

周期性: 24

プレビュー(P)

区間 ビルド 推定 予測 注釈

時間区分: 1日あたりの時間数

週あたりの日数: 7 週の開始: 月曜日

1日の時間数: 24 1日の開始: 00:00

最初のレコードからラベル付けを開始 データからビルド

年: 2000 月: 1月 日: 1

時間: 00:00

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

日付のフォーマット: YYYY-MM-DD 時間のフォーマット: HH:MM:SS

OK(O) キャンセル(C) 適用(A) リセット(R)

時系列のラベリングまたは構築

レコードに連続してラベルを付ける (Label) か、または指定された日付、タイムスタンプ、または時間フィールドに基づいて系列を構築するか (Create) を指定することができます。

- **最初のレコードからラベル付けを開始**：開始データおよび時間またはそのいずれかを指定し、レコードに連続してラベルを付けます。たとえば、日あたりの時間にラベルを付けると、時系列が始まる日と時 (hour) を、それ以降各時間に 1 つのレコードを配置するように指定できます。ラベル付けを別にして、この方法は元のデータを変更しません。その代わりに、レコードが既に均等に配列されており、各測定値間の区分が均一であることを前提としています。欠損している測定値はデータ内に空の行で示されていなければなりません。
- **データからビルド**：均等な間隔でない時系列に対しては、各測定値に対応する日または時間を示すフィールドを指定することができます。入力として使用するには、適切なフォーマットの日付、時間、またはタイムスタンプのフィールドが必要です。たとえば、Jan 2000、Feb 2000 などの値の文字列フィールドがある場合、これを置換ノードを使用して、日

付フィールドへ変換できます。詳細は、[p. 200 置換ノードを使ったストレージの変換](#) を参照してください。[データからビルド] オプションも、必要に応じて埋め込みをしたりレコードを集計したりして、日付を、指定した区分に一致させます。たとえば、週を「ロール アップ」して月としたり、欠損レコードを空白値あるいは挿入値で置き換えたりします。[ビルド] タブ上のレコードを埋め込みまたは集計するための関数を指定することができます。詳細は、[p. 244 時間区分構築のオプション](#) を参照してください。

新規フィールド名拡張子：このノードによって生成されたすべてのフィールドに適用される接頭辞または接尾辞を指定できるようになります。たとえば、デフォルトの接頭辞 \$TI_ を使用すると、このノードによって生成されたフィールドの名前は、\$TI_TimeIndex、\$TI_TimeLabel のようになります。

日付のフォーマット：このノードによって生成された [TimeLabel] フィールドのフォーマットを、現在の区分に指定します。このオプションは現在の選択によっては利用できない場合があります。

時間のフォーマット：このノードによって生成された [TimeLabel] フィールドのフォーマットを、現在の区分に指定します。このオプションは現在の選択によっては利用できない場合があります。

時間区分構築のオプション

時間区分ノードの [ビルド] タブを使用して、指定した区分に合わせるためにフィールドの集計や埋め込みのオプションを指定できます。これらの設定は [データから構築] オプションが [区間] タブで選択されている場合にのみ、適用されます。たとえば、週と月が混在するデータがあるとし、月区分を均一にするために週の値を「ロール アップ」して集計することができます。または、区間を週単位に設定し、欠損のある週に対して空白値を挿入、または指定した埋め込み関数を使用して欠損値を挿入して時系列を埋め込むことができます。

データを埋め込みまたは挿入する場合、既存のデータまたはタイムスタンプ フィールドは、生成された TimeLabel および TimeIndex フィールドによって効果的に置換され、出力から除外されます。不明なフィールドも除外されます。時間を持続時間として測定するフィールド（サービスコールの開始時間ではなくサービス コールを測定するフィールドなど）は保持され、タイムスタンプ フィールドではなく、時間フィールドとしてできるだけ長期間内部に保存されます。詳細は、[2 章 p. 35 フィールドのストレージと形式の設定](#) を参照してください。その他のフィールドは [ビルド] タブで指定したオプションに基づいて集計されます。

図 4-85
時間区分ノード：[ビルド] タブ



- **デフォルトのフィールドと関数の使用** すべてのフィールドを、前述のように日付、タイムスタンプおよび不明のフィールドの例外を使用し、必要に応じて集計または埋め込むように指定します。デフォルトの関数は、フィールドの尺度を基に適用されます。たとえば、連続型フィールドは、平均値を使用して集計され、一方名義型フィールドはモードを使用します。ダイアログ ボックスの下部で、1 つ以上の尺度のデフォルトを変更することができます。
- **フィールドと関数を指定**：フィールドを埋め込むまたは集計するように指定し、それぞれに使用する関数を指定できるようになります。選択されていないフィールドは出力から除外されます。右側にあるアイコンを使用してフィールドを追加、またはテーブルから削除するか、あるいは該当する列のセルをクリックして、そのフィールドに使用する集計または埋め込み関数を変更し、デフォルトをオーバーライドします。不明のフィールドはこのリストから除外され、テーブルに加えることはできません。

デフォルト：さまざまなタイプのフィールドにデフォルトで使われる集計および埋め込み関数を指定します。これらのデフォルトは、[デフォルトを使用] が選択されている場合に適用され、また初期デフォルトとしてすべてのテーブルに追加される新規フィールドに適用されます。（デフォルトを変

更しても、テーブル内の既存の設定は変更されませんが、それ以降追加されるフィールドには変更が適用されます)。

集計関数：利用できる集計関数を次に示します。

- **連続：**連続型フィールドに利用できる関数には、平均、合計、最頻値、最小値、および **最大値** があります。
- **名義：**オプションには、[最頻値]、[初めのn件]、および [最終] があります。「初めのn件」とは、集計グループでの初めのヌルでない値（日付別に格納される）意味し、「最終」はこのグループの最後のヌルでない値を意味します。
- **フラグ型：**オプションには、[いずれかが真 (true) の場合は真 (true)]、[最頻値]、[初めのn件]、および [最終] があります。

パディング関数：利用できる埋め込み関数を次に示します。

- **連続：**オプションには、[空白] と、生成される時間直前 の3 つの最も最近のヌルでない値の平均値を意味する [最も最近使用したポイントの平均] があります。値が 3 つない場合は、新規の値は空白となります。最近の値には実際の値のみ含まれ、先に生成された値はヌルでない値の検索時には含まれません。
- **名義：**[空白] および [最も最近の値]。「最も最近」とは、生成される時間前の最も最近のヌルでない値を言います。また、実際の値のみが最近の値の検索時に考慮されます。
- **フラグ型：**オプションには、[空白]、[真]、および [偽] があります。

結果のデータセット内の最大レコード数：生成されるレコード数に対して上限を指定します。特に時間区分が「秒」に設定される場合、この制限をしないとレコード数がきわめて大きくなります（意図するしないに関わらず）。たとえば、わずか 2 つの値の時系列は（2000 年 1 月 1 日から 2001 年 1 月 1 日までの場合）、秒で埋め込んだ場合は 31,536,000 レコードとなります（60 秒 x 60 分 x 24 時間 x 365 日）。指定した最大値を超えた場合、システムは処理を中止し、警告メッセージを表示します。

カウント型フィールド

値を集計または埋め込む場合、新規の [度数] フィールドが生成され、新規レコードを決定する時点で含まれるレコード数を示します。たとえば、4 週間の値が 1 か月に集計された場合、カウントは 4 となります。埋め込まれたレコードに対しては、カウントは 0 です。フィールド名は、[度数] であり、他に [区分] タブで接頭辞と接尾辞が指定されます。

推定期間

時間区分ノードの [推定] タブを使用して、ホールドアウトと同様に、モデル推定に使用されるレコードの範囲を指定できます。これらの設定値は下流のモデル作成ノードで必要に応じて上書きできますが、ここで指定するほうが、各ノードに個別に指定するよりも便利です。

図 4-86
時間区分ノードの [推定] タブ



推定の開始: データの先頭で推定期間を開始することも、予測で制限付きに使用できる古い値を排除することもできます。データによっては、予測の正確さを著しく失うこともなく、推定期間を短縮してパフォーマンスを上げることも、データの準備に費やす時間を減らすこともできることが判明する可能性もあります。

推定の終了: データの終わりまですべてのレコードを使用してモデルを推定することも、モデルを評価するために最新のレコードを「ホールドアウト」することもできます。後者の場合、モデルの効果を計測するための観察値と予測値が比較できるようになり、効率的にすでに既知の値を「予測」することになります。

予測(S)

時間区分ノードの [予測] タブを使用して、予測するレコード数と、下流の時系列モデル作成ノードによる予測で使用する、将来値を指定できます。これらの設定値は下流のモデル作成ノードで必要に応じて上書きできますが、ここで指定するほうが、各ノードに個別に指定するよりも便利です。

図 4-87
時間区分ノードの [予測] タブ



レコードを将来へ拡張： 推定期間を超えて予測するレコード数を指定します。[推定] タブで指定されるホールドアウト数により、これらのレコードは、「予測フィールド」である場合も、ない場合もあります。

未来インディケータフィールド： レコードに予測データが含まれるかを表示する、生成されたフィールドのラベルラベルのデフォルト値は、\$TI_Futureです。

予測で使用する将来の値 予測する各レコードに対し（ホールドアウトは除く）、予測フィールド（役割を入力に設定）を使用する場合、各予測の予測期間に対し、推定値を指定する必要があります。手動で値を指定することも、リストから選択することもできます。

- フィールド：** フィールド選択ボタンをクリックし、予測として使用するフィールドを選択します。ここで選択したフィールドは、モデル作成で使用されることも、使用されないこともあります。フィールドを予測フィールドとして実際に使用するには、下流のモデル作成ノードで選択する必要があります。このダイアログボックスは将来の値を指定する便利な場所であり、下流のモデル作成ノードによって共有できるので、各ノードで将来の値を個別に指定しなくても済みます。また、利用できるフィールドの一覧が [ビルド] タブでの選択に制約される可能性があります。たとえば、[フィールドと関数を指定する] が [ビルド] タブで選択されている場合、集計されない、または値が埋め込まれないフィールドはストリームから削除され、モデル作成に使用できません。

注： 将来の値がストリーム内で以後使用されないフィールドに指定された場合（削除されたか、[ビルド] タブで選択が更新されたことが原因）、そのフィールドは [予測] タブに赤で表示されます。

- 値：** 各フィールドに対し、関数のリストから選択するか、または [指定] をクリックして手動で入力または事前に定義された値から選択することができます。予測フィールドが、管理するまたは事前に検知できる項目と関連する場合、値を手動で入力する必要があります。たとえば、部屋の予約数に基づいてホテルの翌月の収益を予測する場合、当月実際に取得した予約数を指定することができます。それに対し、予測フィールドが株価など管理外のものに関連する場合、最も最近使用した値や最近使用したポイントの平均などの関数を使用することができます。

利用できる関数は、フィールドの尺度によって異なります。

| 尺度 | 関数 |
|---------------------|--|
| 連続型フィールドまたは名義型フィールド | Blank 最近使用したポイントの平均 最も最近使用した値 Specify |
| フラグ型フィールド | Blank 最も最近使用した値 True False Specify |

最近使用したポイントの平均 - 最近使用した 3 つのデータポイントの平均から将来の値を計算します。

最も最近使用した値 - 最も最近使用したデータポイントの値に将来の値を設定します。

真/偽 – 指定された真または偽にフラグ型フィールドの将来の値を設定します。

指定 – 手動で将来の値を指定するため、または事前に定義されたリストから値を選択するためダイアログ ボックスを開きます。

図 4-88
予測フィールド用の将来の値の指定



将来の値

ここでは、下流の時系列モデル作成ノードによる予測に使用する将来の値を指定することができます。これらの設定値は下流のモデル作成ノードで必要に応じて上書きできますが、ここで指定するほうが、各ノードに個別に指定するよりも便利です。

手動で値を入力できます。または、ダイアログ ボックスの右側にある選択ボタンをクリックして、現在のフィールドに定義された値のリストから値を選択できます。詳細は、[7 章 値の表示または選択 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

指定できる将来の値の数は、将来に時系列を拡張するレコードの数に対応しています。

サポートされる区分

時間区分ノードでは、周期的(季節性など)または非周期的な期間と同様に、秒から年までのすべての期間区分がサポートされます。[区分] タブの時間区分フィールドで、区分を指定します。

Periods

[期間] を選択し、その他の指定されている区分に一致しない既存の循環しない時系列にラベルを付けます。この時系列は既に正しい順番になっており、各測定値間の区分は均一でなければなりません。[データからビルド] オプションは、この区分が選択された場合に利用できます。

図 4-89
循環しない期間の時間区分設定

時間区分: 期間

最初のレコードからラベル付けを開始 データからビルド

期間: 1

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

サンプルの出力

レコードは指定された開始値からインクリメンタルにラベルが付けられます (Period 1、Period 2...)。新規フィールドは次のように作成されます。

| \$TI_TimeIndex (整数) | \$TI_TimeLabel (文字列) | \$TI_Period (整数) |
|---------------------|----------------------|------------------|
| 1 | Period 1 | 1 |
| 2 | Period 2 | 2 |
| 3 | Period 3 | 3 |
| 4 | Period 4 | 4 |
| 5 | Period 5 | 5 |

循環する期間

[循環する期間] を選択し、循環を繰り返し、標準の区分に適合しない既存の時系列にラベルを付けます。たとえば、会計年度内で 10 か月しか期間がない場合はこのオプションを使用できます。この時系列は既に正しい順番になっており、各測定値間の区分は均一でなければなりません。([データからビルド] オプションは、この区分が選択された場合に利用できます)。

図 4-90
循環する期間の時間区分設定

時間区分: 循環する期間

サイクルあたりの期間数: 12

最初のレコードからラベル付けを開始 データからビルド

サイクル: 1

期間: 1

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

サンプルの出力

レコードは、指定した循環の開始時点および期間 (Cycle 1、Period 1、Cycle 1、Period 2...) に基づいてインクリメンタルにラベルが付けられます。たとえば、循環ごとの期間数を 3 に設定すると、新規フィールドは次のように生成されます。

| \$TI_TimeIndex (整数) | \$TI_TimeLabel (文字列) | \$TI_Cycle (整数) | \$TI_Period (整数) |
|---------------------|----------------------|-----------------|------------------|
| 1 | Cycle 1, Period 1 | 1 | 1 |
| 2 | Cycle 1, Period 2 | 1 | 2 |
| 3 | Cycle 1, Period 3 | 1 | 3 |
| 4 | Cycle 2, Period 1 | 2 | 1 |
| 5 | Cycle 2, Period 2 | 2 | 2 |

Years

数年間にわたり、開始年を指定して、連続するレコードをラベルするか、[データからビルド] を選択してタイムスタンプまたは各レコードの年次を識別する日付フィールドを指定します。

図 4-91
連続年の時系列の時間区分設定

時間区分: 年数

最初のレコードからラベル付けを開始 データからビルド

年: 2000

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (文字列) | \$TI-Year (整数) |
|---------------------|----------------------|----------------|
| 1 | 2000 | 2000 |
| 2 | 2001 | 2001 |
| 3 | 2002 | 2002 |
| 4 | 2003 | 2003 |
| 5 | 2004 | 2004 |

Quarters

四半期の時系列に対して、会計年度が開始する月を指定できます。開始年を指定して、連続するレコードをラベルするか、[データからビルド] を選択して、各レコードの四半期および年次を識別するタイムスタンプまたは日付フィールドを選択します。

図 4-92
四半期の時系列の時間区分設定

時間区分: 四半期数

会計年度の開始: 1月

最初のレコードからラベル付けを開始 データからビルド

年: 2000 四半期: 1

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

サンプルの出力

会計年度が 1 月に開始する場合、新規フィールドは次のように生成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (文字列) | \$TI-Year (整数) | \$TI-Quarter (ラベル付き整数) |
|---------------------|----------------------|----------------|------------------------|
| 1 | Q1 2000 | 2000 | 1 (Q1) |
| 2 | Q2 2000 | 2000 | 2 (Q2) |
| 3 | Q3 2000 | 2000 | 3 (Q3) |
| 4 | Q4 2000 | 2000 | 4 (Q4) |
| 5 | Q1 2001 | 2001 | 1 (Q1) |

年度が 1 月以外から開始する場合、新規フィールドは次のように生成されます (7 月から会計年度が始まるとする)。各四半期の月を識別するラベルを表示するには、ツールバーのアイコンをクリックして値の表示をオンにします。

図 4-93
値ラベル アイコン



| \$TI-TimeIndex (整数) | \$TI-TimeLabel (文字列) | \$TI-Year (整数) | \$TI-Quarter (ラベル付き整数) |
|---------------------|----------------------|----------------|------------------------|
| 1 | Q1 2000/2001 | 1 | 1 (Q1 Jul-Sep) |
| 2 | Q2 2000/2001 | 1 | 2 (Q2 Oct-Dec) |
| 3 | Q3 2000/2001 | 1 | 3 (Q3 Jan-Mar) |
| 4 | Q4 2000/2001 | 1 | 4 (Q4 Apr-Jun) |
| 5 | Q1 2001/2002 | 2 | 1 (Q1 Jul-Sep) |

Months

開始年度と月を選択して連続レコードのラベルを付けるか、または [データからビルド] を選択して、各レコードの月を示すタイムスタンプまたは日付フィールドを選択します。

図 4-94
連続月の時系列の時間区分設定

時間区分: 月数

最初のレコードからラベル付けを開始 データからビルド

年: 2000 月: 1月

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (日付) | \$TI-Year (整数) | \$TI-Months (ラベル付き整数) |
|---------------------|---------------------|----------------|-----------------------|
| 1 | Jan 2000 | 2000 | 1 (January) |
| 2 | Feb 2000 | 2000 | 2 (February) |
| 3 | Mar 2000 | 2000 | 3 (March) |
| 4 | Apr 2000 | 2000 | 4 (April) |
| 5 | May 2000 | 2000 | 5 (May) |

週 (非周期的)

週の時系列の場合、サイクルが開始する曜日を選択することができます。

月数、四半期数、年数が週の数と同じである必要がないため、週の数是非周期的のみになりますので注意してください。ただし、タイム スタンプの押されたデータは、容易に非周期的モデルの週レベルで集計または埋め込むことができます。

図 4-95
連続週時系列の時間区分設定

時間区分: 週数 (非定期的)

週の開始: 月曜日

最初のレコードからラベル付けを開始 データからビルド

年: 2000 月: 1月 日: 1

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

日付のフォーマット: YYYY-MM-DD

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (日付) | \$TI-Week (整数) |
|------------------------|------------------------|----------------|
| 1 | 1999-12-27 | 1 |
| 2 | 2000-01-03 | 2 |
| 3 | 2000-01-10 | 3 |
| 4 | 2000-01-17 | 4 |
| 5 | 2000-01-24 | 5 |

1 週間の \$TI-TimeLabel フィールドは週の最初の日を表示します。前の表では、ラベルは 2000 年 1 月 1 日から開始します。ただし、週は月曜日から始まり、2000 年 1 月 1 日は土曜日となります。そのため、1 月 1 日を含む週は 1999 年 12 月 27 日から開始、これが最初のポイントのラベルとなります。

日付のフォーマットによって、\$TI-TimeLabel フィールドに作成される文字列を決定します。

週あたりの日数

週で循環する毎日の測定に対して、1 週間あたりの日数と各週の開始日を指定することができます。開始日を指定して連続レコードのラベルを付けるか、または [データからビルド] を選択して、各レコードの日付を示すタイムスタンプまたは日付フィールドを選択します。

図 4-96
連続日の時系列の時間区分設定

時間区分: 曜日数

週あたりの日数: 7 週の開始: 月曜日

最初のレコードからラベル付けを開始 データからビルド

年: 2000 月: 1月 日: 1

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

日付のフォーマット: YYYY-MM-DD

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (日付) | \$TI-Week (整数) | \$TI-Day (ラベル付き整数) |
|---------------------|---------------------|----------------|--------------------|
| 1 | Jan 5 2005 | 1 | 3 (Wednesday) |
| 2 | Jan 5 2006 | 1 | 4 (Thursday) |
| 3 | Jan 5 2007 | 1 | 5 (Friday) |
| 4 | Jan 5 2010 | 2 | 1 (Monday) |
| 5 | Jan 5 2011 | 2 | 2 (Tuesday) |

注：週は常に最初の期間に対して 1 で開始し、カレンダーに基づいて循環させてはいけません。したがって、週 52 はその後に、週 53、54... と続きます。この週は、連続して週の数が増加するだけで、年次の週に影響を及ぼしません。

日数 (非周期的)

通常の週サイクルに適合しない毎日の測定がある場合は、非周期的日数を選択します。開始日を指定して連続レコードのラベルを付けるか、または [データからビルド] を選択して、各レコードの日付を示すタイムスタンプまたは日付フィールドを選択します。

図 4-97
連続日の時系列の時間区分設定 (非周期的)

時間区分: **日数 (非周期的)**

最初のレコードからラベル付けを開始 データからビルド

年: 月: 日:

新規フィールド名拡張子: 追加方法: 接頭辞 接尾辞

日付のフォーマット:

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (日付) |
|---------------------|---------------------|
| 1 | Jan 5 2005 |
| 2 | Jan 5 2006 |
| 3 | Jan 5 2007 |

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (日付) |
|---------------------|---------------------|
| 4 | Jan 5 2008 |
| 5 | Jan 5 2009 |

1 日あたりの時間数

日サイクルで行われる毎時測定に対しては、週あたりの日数、1日の時間数（8時間の勤務時間など）、週の開始日、毎日の開始時間を指定できます。時間は、24時間システムの分単位までで指定します（例：14:05 = 2:05 P.M.）。

図 4-98
時の時系列の時間区分設定

時間区分: 1日あたりの時間数

週あたりの日数: 7 週の開始: 月曜日

1日の時間数: 24 1日の開始: 00:00

最初のレコードからラベル付けを開始 データからビルド

年: 2000 月: 1月 日: 1

時間: 00:00

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

日付のフォーマット: YYYY-MM-DD 時間のフォーマット: HH:MM:SS

開始日と時間を指定して連続レコードのラベルを付けるか、または [データからビルド] を選択して、各レコードの日付と時間を識別するタイムスタンプフィールドを選択します。

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (タイムスタンプ) | \$TI-Day (ラベル付き整数) | \$TI-Hour (ラベル付き整数) |
|---------------------|--------------------------|--------------------|---------------------|
| 1 | Jan 5 2005 08:00:00 | 3 (Wednesday) | 8 (8:00) |
| 2 | Jan 5 2005 09:00:00 | 3 (Wednesday) | 9 (9:00) |
| 3 | Jan 5 2005 10:00:00 | 3 (Wednesday) | 10 (10:00) |

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (タイムスタンプ) | \$TI-Day (ラベル付き整数) | \$TI-Hour (ラベル付き整数) |
|---------------------|--------------------------|--------------------|---------------------|
| 4 | Jan 5 2005 11:00:00 | 3 (Wednesday) | 11 (11:00) |
| 5 | Jan 5 2005 12:00 | 3 (Wednesday) | 12 (12:00) |

時間数 (非周期的)

通常の日サイクルに適合しない毎時の測定がある場合は、このオプションを選択します。開始時間を指定して連続レコードのラベルを付けるか、または [データからビルド] を選択して、各レコードの時間を示すタイムスタンプまたは時間フィールドを選択します。

図 4-99
年次の時系列の時間区分設定

時間区分: 時間数 (非周期的)

最初のレコードからラベル付けを開始 データからビルド

時間: 00:00

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

時間は24時間時計とし (13:00 = 1:00 P.M.)、ラップ (24時の後が25時) させません。

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (文字列) | \$TI-Hour (ラベル付き整数) |
|---------------------|----------------------|---------------------|
| 1 | 8:00 | 8 (8:00) |
| 2 | 9:00 | 9 (9:00) |
| 3 | 10:00 | 10 (10:00) |
| 4 | 11:00 | 11 (11:00) |
| 5 | 12:00 | 12 (12:00) |

1日あたりの分数

日サイクルの分ごとの測定については、週あたりの日数、週の開始日、日あたりの時間数、1日の開始時間を指定できます。時間は24時間時計で指定し、コロンを使用して分、秒まで指定されます (例: 2:05:17 P.M. =

14:05:17)。分の増分（1分ごと、2分ごと...、増分は60の公約数でなければなりません）も指定できます。

図 4-100
日あたりの分の時間区分設定

開始日と時間を指定して連続レコードのラベルを付けるか、または[データからビルド]を選択して、各レコードの日付と時間を識別するタイムスタンプフィールドを選択します。

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (タイムスタンプ) | \$TI-Minute |
|---------------------|--------------------------|-------------|
| 1 | 2005-01-05 08:00:00 | 0 |
| 2 | 2005-01-05 08:01:00 | 1 |
| 3 | 2005-01-05 08:02:00 | 2 |
| 4 | 2005-01-05 08:03:00 | 3 |
| 5 | 2005-01-05 08:04:00 | 4 |

分数 (非周期的)

通常の日サイクルに適合しない毎分の測定がある場合は、このオプションを選択します。分の増分（1分ごと、2分ごと...、増分は60の公約数でなければなりません）も指定できます。

図 4-101
分の時間区分設定 (非周期的)

開始時間を指定して連続レコードのラベルを付けるか、または [データからビルド] を選択して、各レコードの時間を識別するタイムスタンプまたは時間フィールドを指定します。

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (文字列) | \$TI-Minute |
|---------------------|----------------------|-------------|
| 1 | 8:00 | 0 |
| 2 | 8:01 | 1 |
| 3 | 8:02 | 2 |
| 4 | 8:03 | 3 |
| 5 | 8:04 | 4 |

- TimeLabel の文字列は、時と分の間にコロンを使用して生成されます。時間はラップしません。24 時の後に 25 時とはしません。
- ダイアログ ボックスで指定された値により、分は増加します。たとえば、増分が 2 であれば、TimeLabel は 8:00、8:02、... と増加し、分は、0、2、... と増加します。

1日あたりの秒数

日サイクルの秒区分の測定については、週あたりの日数、週の開始日、日あたりの時間数、1 日の開始時間を指定できます。時間は 24 時間時計で指定し、コロンを使用して分、秒まで指定されます (例: 2:05:17 P.M. = 14:05:17)。秒の増分 (1 秒ごと、2 秒ごと...、増分は 60 の公約数でなければなりません) も指定できます。

図 4-102
日あたりの秒の時間区分設定

時間区分: 1日あたりの秒数 ▼ 増分: 1 ▼

週あたりの日数: 7 ▼ 週の開始: 月曜日 ▼

1日の時間数: 24 ▼ 1日の開始: 00:00

最初のレコードからラベル付けを開始 データからビルド

年: 2000 ▼ 月: 1月 ▼ 日: 1 ▼

時間: 00:00:00

新規フィールド名拡張子: \$TI_ 追加方法: 接頭辞 接尾辞

日付のフォーマット: YYYY-MM-DD ▼ 時間のフォーマット: HH.MM.SS ▼

日と時間を指定して連続レコードのラベル付けを開始するか、または [データからビルド] を選択して、各レコードの日付と時間を指定するタイムスタンプ フィールドを選択します。

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (タイムスタンプ) | \$TI-Minute | \$TI-Second |
|---------------------|--------------------------|-------------|-------------|
| 1 | 2005-01-05 08:00:00 | 0 | 0 |
| 2 | 2005-01-05 08:00:01 | 0 | 1 |
| 3 | 2005-01-05 08:00:02 | 0 | 2 |
| 4 | 2005-01-05 08:00:03 | 0 | 3 |
| 5 | 2005-01-05 08:00:04 | 0 | 4 |

秒数 (非周期的)

通常の日サイクルに適合しない毎秒の測定がある場合は、このオプションを選択します。秒の増分 (1 秒ごと、2 秒ごと...、増分は 60 の公約数でなければなりません) を指定できます。

図 4-103
秒の時間区分設定 (非周期的)

時間区分: 増分:

最初のレコードからラベル付けを開始 データからビルド

時間:

新規フィールド名拡張子: 追加方法: 接頭辞 接尾辞

時間を指定して連続レコードのラベル付けを開始するか、または [データからビルド] を選択して、各レコードの時間を識別するタイムスタンプまたは時間フィールドを指定します。

サンプルの出力

新規フィールドは次のように作成されます。

| \$TI-TimeIndex (整数) | \$TI-TimeLabel (文字列) | \$TI-Minute | \$TI-Second |
|---------------------|----------------------|-------------|-------------|
| 1 | 8:00:00 | 0 | 0 |
| 2 | 8:00:01 | 0 | 1 |
| 3 | 8:00:02 | 0 | 2 |
| 4 | 8:00:03 | 0 | 3 |
| 5 | 8:00:04 | 0 | 4 |

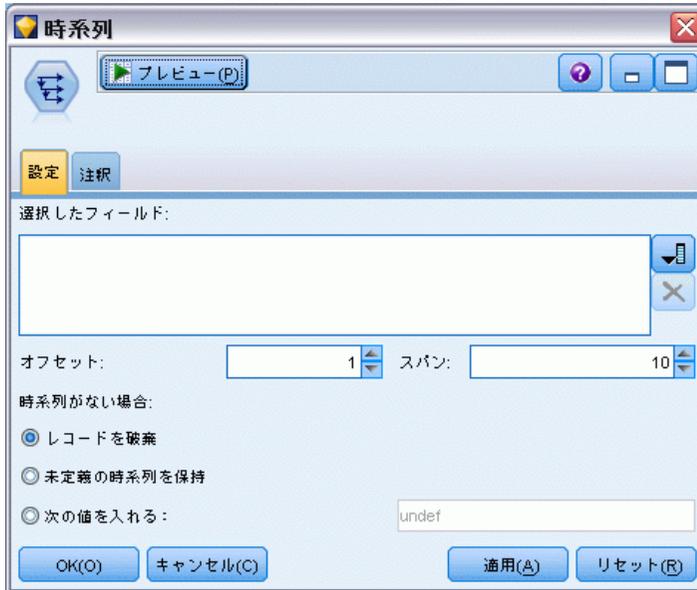
- TimeLabel の文字列は、時と分の間、分と秒の間にコロンを使用して生成されます。時間はラップしません。24 時の後に 25 時とはしません。
- 数値による秒の増分はすべて増分として指定されます。増分が 2 であれば、TimeLabel は 08:00:00、08:00:02、... と増加し、秒は、0、2、... と増加します。

時系列ノード

時系列ノードは、多くの場合、時系列データなどの継続的なデータに使用されます。時系列ノードを使用して、前のレコードのフィールドのデータを含む新規フィールドを作成します。時系列ノードを使用する際には、あらかじめ特定のフィールドでソートされたデータがあれば便利です。このためには、ソート ノードを使用します。

時系列ノードのオプションの設定

図 4-104
時系列ノードのダイアログ ボックス



選択したフィールド： フィールド ピッカー（テキスト ボックスの右側にあるボタン）を使って、時系列データを取得するフィールドを選択します。選択した各フィールドを使って、データ セット中のすべてのレコードに対する新しいフィールドが作成されます。

オフセット。 時系列フィールド値を抽出する最新レコードが、現在のレコードのいくつ前にあるかを指定します。たとえば、[オフセット] を 3 に設定すると、各レコードがこのノードを通過するときに、3 つ前のレコードのフィールド値が現在のレコードに追加されます。[スパン] の設定を使用して、いくつ前のレコードまでの値を抽出するか指定します。オフセット値を調整するには、矢印を使用します。

スパン： 値を抽出する元になる前のレコードの数を指定します。たとえば、[オフセット] を 3 に設定し、[スパン] を 5 に設定した場合は、ノードを通過する各レコードに対し、[選択したフィールド] リストで指定した各フィールドごとに 5 つのフィールドが追加されます。つまり、ノードがレコード 10 を処理するときには、レコード 7 からレコード 3 までのフィールドが追加されます。スパン値を調整するには、矢印を使用します。

時系列がない場合： 時系列値がないレコードの処理方法を、次の 3 つのオプションから選択します。このようなレコードは、時系列として使用する前のレコードがない、データ セットの先頭数レコードなどがあてはまります。

- **レコードを破棄：** 選択したフィールドで時系列値を利用できない場合は、そのレコードを破棄します。

- **未定義の時系列を保持**：時系列値がない場合もレコードを保持します。この場合、時系列値には未定義の値が入れられ、`$null$` として表示されます。
- **次の値を入れる**：時系列値がないレコードに対して使用する値または文字列を指定します。デフォルトは、システムヌル値の `undef` です。ヌル値は文字列 `$null$` で表されます。

置換値を選択する場合は、適切な処理を行うために、次の規則にしたがってください。

- 選択するフィールドは、それぞれ同じストレージタイプでなければなりません。
- 選択したすべてのフィールドのストレージタイプが数値の場合、置換値は整数でなければなりません。
- 選択したすべてのフィールドのストレージタイプが実数の場合、置換値は実数でなければなりません。
- 選択したすべてのフィールドのストレージタイプがシンボル値の場合、置換値は文字列でなければなりません。
- 選択したすべてのフィールドのストレージタイプが日付/時間の場合、置換値は日付/時間フィールドでなければなりません。

上記の条件を満たさない場合は、時系列ノードの実行時にエラーが発生してしまいます。

フィールド順序ノード

フィールド順序ノードにより、下流のフィールドを表示するために使用する順序を定義することができます。この順序は、テーブル、リスト、およびフィールドピッカーなど、さまざまな場所のフィールドの表示に適用されます。この操作は、さまざまなデータセットにおいて、特定のフィールドをより参照しやすくする場合などに役立ちます。

フィールド順序ノードのオプションの設定

ファイルを並べ替えるには、次の 2 つの方法があります。ユーザー指定の順序と自動ソートです。

ユーザー指定の順序

[ユーザー指定の順序] を選択すると、フィールド名とデータ型のテーブルが有効になります。このテーブルから、すべてのフィールドを参照したり、矢印ボタンを使って独自の並び順を作成することができます。

図 4-105
目的のフィールドが最初に表示されるように並び替え



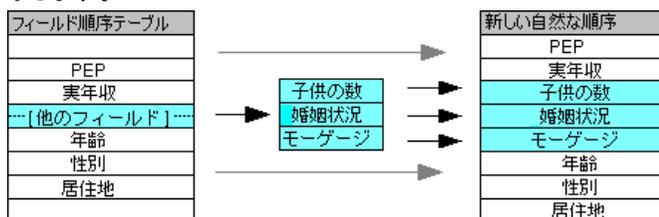
フィールドを並び替えるには

- ▶ テーブル中のフィールドを選択します。複数のフィールドを選択するには、Ctrl キーを押しながらフィールドを選択します。
- ▶ 単純な矢印ボタンをクリックすると、フィールドが 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、フィールドがリストの最上位または最下位に移動します。
- ▶ ここに含まれていないフィールドの順序を指定するには、[他のフィールド] で示されている区切り行を上または下に移動します。

他のフィールド： 区切り行 [他のフィールド] は、テーブルを 2 つに分割するために用いられています。

- この区切り行の上に表示されているフィールドは、このノードの下流におけるフィールド表示の並び順の一番上に表示されます（テーブルに表示されているように）。
- この区切り行の下に表示されているフィールドは、このノードの下流におけるフィールド表示の並び順の一番下に表示されます（テーブルに表示されているように）。

図 4-106
フィールドの並び順において「他のフィールド」がどのように組み込まれるかを、以下の図に示します。



- ここに表示されていない他のフィールドは、区切り行の位置で示される、「最上位のフィールド」と「最下位のフィールド」の間に配置されます。

他のカスタム オプションには、次のようなものがあります。

- 各列見出しの矢印をクリックすると、その列が昇順または降順にソートされます（[データ型]、[名前]、または[ストレージ]）。列を基準にソートする場合、ここに指定されていないフィールド（[他のフィールド] 行）は、最後に普通の順序でソートされます。
- フィールド順序ノードから未使用のフィールドをすべて削除するには、[未使用を消去] をクリックします。未使用のフィールドは、テーブル中に赤で表示されます。これは、上流の操作でそのフィールドが削除されたことを表しています。
- 任意の新規フィールド（稲妻のアイコンは、新規または未指定のフィールドを示します）の順序を指定します。[OK] または [適用] をクリックすると、アイコンが消えます。

注：ユーザー指定の順序を適用した後に、上流でフィールドが追加された場合、その新規フィールドはカスタム リストの最後に追加されます。

自動ソート

ソート用パラメータを指定する場合は、[自動ソート] を選択します。ダイアログ ボックスに、自動ソート用のオプションが表示されます。

図 4-107
自動ソート オプションを使ったすべてのフィールドの並び替え



ソート項目：フィールド順序ノードに読み込まれるフィールドのソート方法を選択します。矢印ボタンは、ソートが昇順に行われるか、または降順に行われるかを示しています。目的の並び順を選択してください。

- 名前
- データ型
- ストレージ

自動ソートが行われた後に、フィールド順序ノードの上流にフィールドが追加された場合、そのフィールドも設定内容に応じて自動的に適切な位置に配置されます。

グラフ作成ノード

グラフ作成ノードの共通の機能

IBM® SPSS® Modeler に取り入れたデータを調べるために、データ マイニングのさまざまなフェーズでグラフやチャートが使用されます。たとえば、散布図ノードや棒グラフ ノードをデータ ソースに接続して、データの型や分布を知ることができます。その後、レコードやフィールドを操作して、モデル作成操作のデータを準備できます。また、新しく作成されたフィールド間の分布や相関関係を確認する場合にも、グラフがよく用いられます。

[グラフ] パレットには次のノードがあります。



グラフボード ノードでは、単一のノードにさまざまな種類のグラフを提供しています。このノードを使用して、検証するデータ フィールドを選択肢、選択したデータに使用できるグラフを選択できます。選択したフィールドに適していないグラフの種類は、自動的に除外されます。詳細は、[p. 278 グラフボード ノード](#) を参照してください。



散布図ノードで、数値フィールド間の関係が示されます。プロットは、点（散布図）または折れ線を使用して作成できます。詳細は、[p. 332 散布図ノード](#) を参照してください。



棒グラフ ノードで、ローンの種類や性別など、シンボル値（カテゴリ）の出現頻度を表示します。通常、棒グラフ ノードを使用してデータの不均衡を表示しますが、そのデータはモデルの作成前にバランス ノードを使って修正できます。詳細は、[p. 341 棒グラフ ノード](#) を参照してください。



ヒストグラム ノードでは、数値フィールドの値の出現頻度が示されます。多くの場合、ヒストグラム ノードは、操作やモデルの構築前にデータを調べるために使用されます。棒グラフ ノードと同様、ヒストグラム ノードにより、データ内の不均衡がしばしば明らかになります。詳細は、[p. 347 ヒストグラムの \[プロット\] タブ](#) を参照してください。



集計棒グラフ ノードで、他の数値フィールドの値に相対的な数値フィールドの値の棒グラフを表示します（集計棒グラフ ノードでは、ヒストグラムに似たグラフが作成されます）。集計棒グラフは、値が時間の経過とともに変化する変数やフィールドを表示する場合に役立ちます。3次元グラフを使って、分布をカテゴリ別に表示するシンボル値軸を追加することもできます。詳細は、[p. 351 集計棒グラフの \[プロット\] タブ](#) を参照してください。



線グラフ ノードでは、1 つの X フィールドに対して複数の Y フィールドを表示するプロットが作成されます。Y フィールドは色付きの線でプロットされ、それぞれ [スタイル] フィールドを [ライン] に、[X モード] フィールドを [ソート] に設定した散布図 ノードに相当します。線グラフは、複数の変数の変動を長期にわたって調査するときに役立ちます。詳細は、[p. 356 線グラフ ノード](#) を参照してください。



Web グラフ ノードで、複数のシンボル値 (カテゴリ) フィールドの値の関係の強さが示されます。このグラフでは、接続の強さを示すためにさまざまな幅の線が使用されます。Web グラフ ノードを使用して、たとえば、E コマース サイトで購入されたさまざまな商品の関係を調査できます。詳細は、[p. 361 Web グラフ ノード](#) を参照してください。



時系列ノードで、時系列データの 1 つ以上のセットを表示します。通常、最初に時間区分ノードを使用して TimeLabel フィールドを作成します。このフィールドは、x 軸にラベルを付けるために使用されます。詳細は、[p. 373 時系列ノード](#) を参照してください。



評価ノードは、予測モデルの評価と比較に用いられます。評価グラフで、モデルが特定の結果をどの程度予測するかを表示します。それによって、予測値と予測の信頼度に基づいたレコードがソートされます。そして、レコードが等サイズ (分位) のグループに分割され、各分位のビジネスに関する基準の値が、高い方から降順でプロットされます。プロットには、複数のモデルが異なる線で示されます。詳細は、[p. 378 評価ノード](#) を参照してください。

グラフ作成ノードをストリームに追加すると、ノードをダブルクリックしてオプションを指定するダイアログ ボックスを開くことができます。大部分のグラフでは、1 つまたは複数のタブにさまざまな固有のオプションが用意されています。また、すべてのグラフに共通なタブ オプションも数多く用意されています。これらの共通オプションの詳細は、次の各の内容を参照してください。

グラフ作成ノードにオプションを環境設定すると、ダイアログ ボックス内またはストリームの一部としてグラフを実行できます。生成されたグラフ ウィンドウ内で、データを選択するか領域を指定し、効果的にデータの「サブセットを作成」することで、フィールド作成 (設定とフラグ) ノードと条件抽出ノードを生成することができます。たとえば、強力なこの機能を使用し、外れ値を識別してそれを除外することができます。

外観、オーバーレイ、パネル、およびアニメーション

オーバーレイと外観

外観（および重ね書き）は、視覚化に次元数を追加します。外観の効果（グループ化、クラスタ化、積み上げ）は、視覚化タイプ、フィールドまたは変数の種類、およびグラフ要素の種類と統計に依存します。たとえば、色についてカテゴリ フィールドを使用して、散布図のポイントをグループ化し、または積み上げ棒グラフ内の積み上げを作成します。また、色について連続した数値の範囲を使用して、散布図の各ポイントの範囲の値を示します。

外観およびオーバーレイを検証して、ニーズを満たすものを探する必要があります。次の説明で、適切な外観およびオーバーレイを選択できます。

注：すべての外観またはオーバーレイが、すべての視覚化の種類に使用できるわけではありません。

- **色：** 色がカテゴリ フィールドによって定義されている場合、個々のカテゴリに基づいて、各カテゴリに 1 色ずつ、視覚化を分割します。色が連続する数値の範囲を表す場合、範囲型フィールドの値に基づいて色が異なります。グラフィック要素（バーまたはボックスなど）が複数のレコード/ケースを表し範囲型フィールドが色に対して使用されている場合、色は範囲型フィールドの平均値によって異なります。
- **形状：** 形状は、視覚化をさまざまな形状の要素に分割するカテゴリ変数によって定義され、1 つのカテゴリに 1 つの形状となります。
- **透明：** 透過性がカテゴリ フィールドによって定義されている場合、個々のカテゴリに基づいて、各カテゴリに透過性レベルを 1 つずつ、視覚化を分割します。透過性が連続する数値の範囲を表す場合、範囲型フィールドの値に基づいて透過性が異なります。グラフィック要素（バーまたはボックスなど）が複数のレコード/ケースを表し範囲型フィールドが透過性に対して使用されている場合、色は範囲型フィールドの平均値によって異なります。最大値では、グラフィック要素は完全に透明です。最小値では、グラフィック要素は完全に不透明です。
- **データラベル：** データ ラベルはフィールドの種類によって定義され、その値はグラフィックに適用されるラベルを作成するために使用されます。
- **サイズ：** サイズがカテゴリ フィールドによって定義されている場合、個々のカテゴリに基づいて、各カテゴリに 1 つのサイズずつ、視覚化を分割します。サイズが連続する数値の範囲を表す場合、範囲型フィールドの値に基づいてサイズが異なります。グラフィック要素（バーまたはボックスなど）が複数のレコード/ケースを表し範囲型フィールドがサイズに対して使用されている場合、サイズは範囲型フィールドの平均値によって異なります。

図 5-1
色のオーバーレイの外観を使用したグラフ

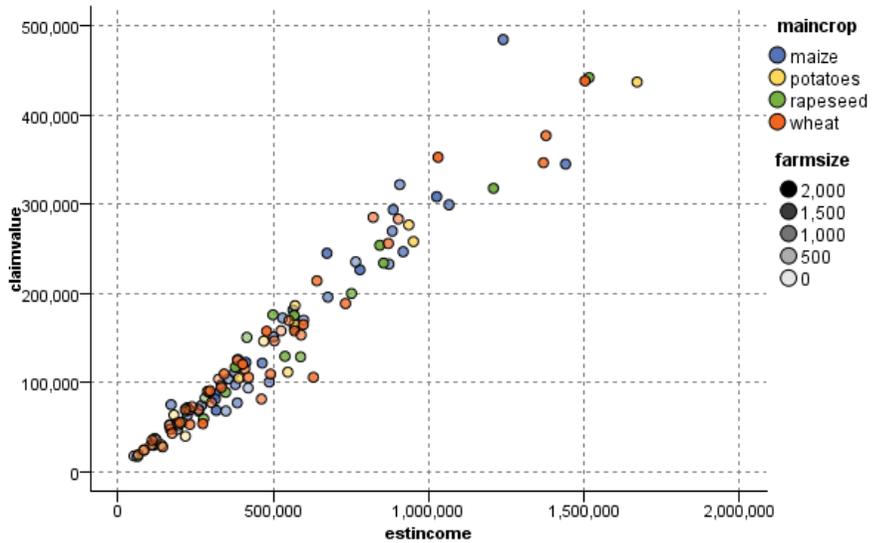
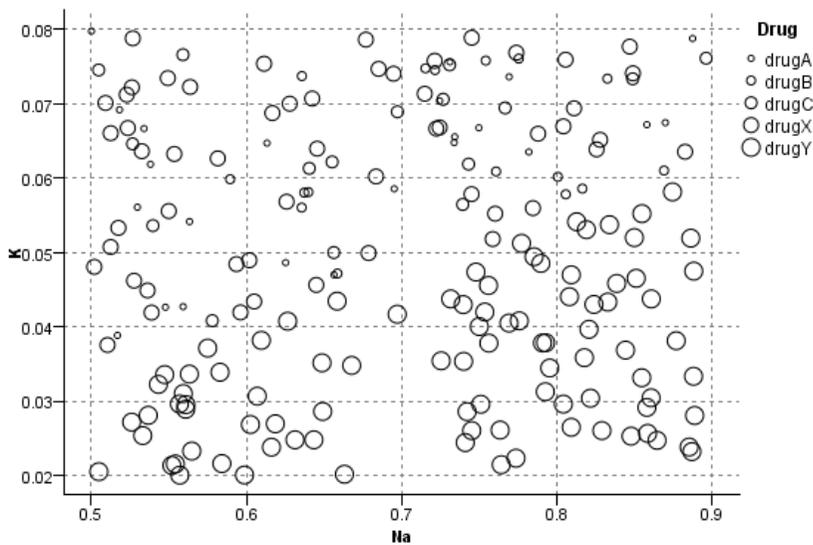


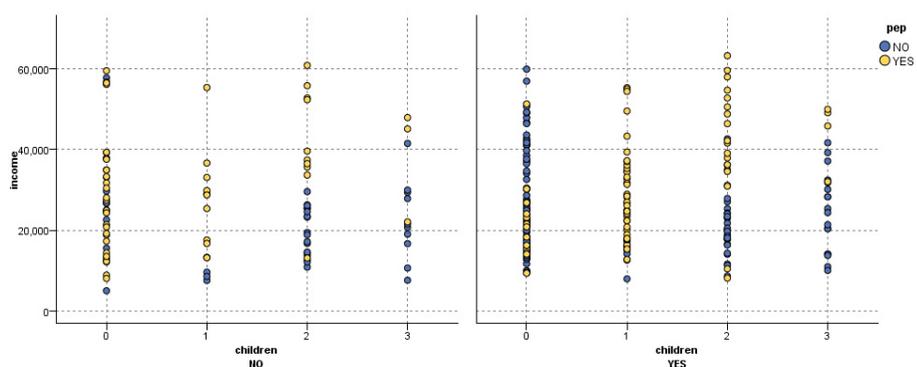
図 5-2
サイズのオーバーレイの外観を使用したグラフ



パネルとアニメーション

パネル: ファセットとも呼ばれるパネリングによって、グラフのテーブルを作成します。1つのグラフがパネリングフィールドの各カテゴリに生成されますが、すべてのパネルは同時に表示されます。パネリングは、視覚化がパネルフィールドの条件に従っているかどうかを確認する場合に役立ちます。たとえば、ヒストグラムを性別によってパネリングし、度数分布が男性と女性で等しいかどうかを確認することができます。つまり、給与が性差の影響を受けているかどうかを確認することができます。パネリング向けに単一のカテゴリフィールドを選択します。

図 5-3
結婚(既婚/未婚)によるパネルを使用したグラフ



アニメーション: アニメーションは、複数のグラフがアニメーションフィールドの値から作成されるという点類似していますが、これらのグラフは同時に表示されません。検証モードのコントロールを使用して、一連のグラフの出力およびフリップをアニメーション化します。さらに、パネリングとは異なり、アニメーションはカテゴリフィールドを必要としません。値が自動的に範囲に分割される連続フィールドを指定することができます。検証モードのアニメーションコントロールでさまざまなサイズの範囲を設定することができます。すべての視覚化でアニメーションを使用できるわけではありません。

図 5-4
3つのカテゴリがある変数の散布図のアニメーション - 低血圧のスライド

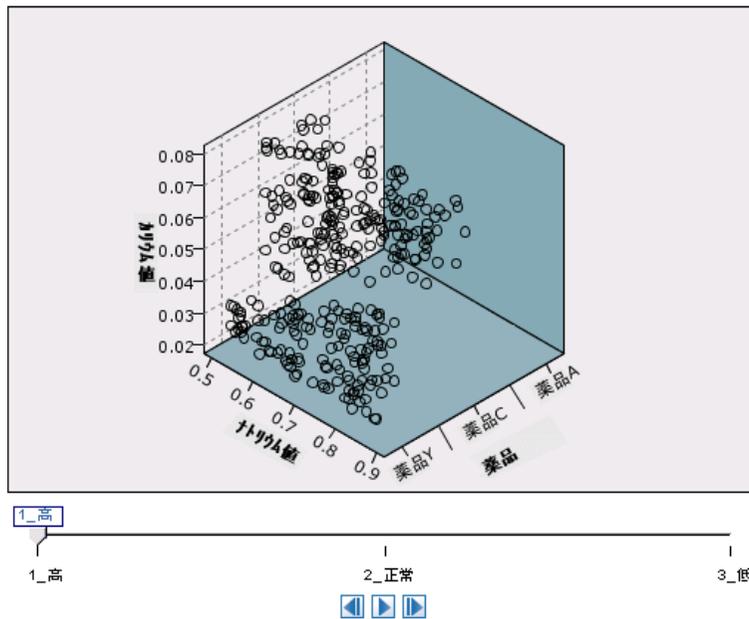


図 5-5
3つのカテゴリがある変数の散布図のアニメーション - 正常な血圧のスライド

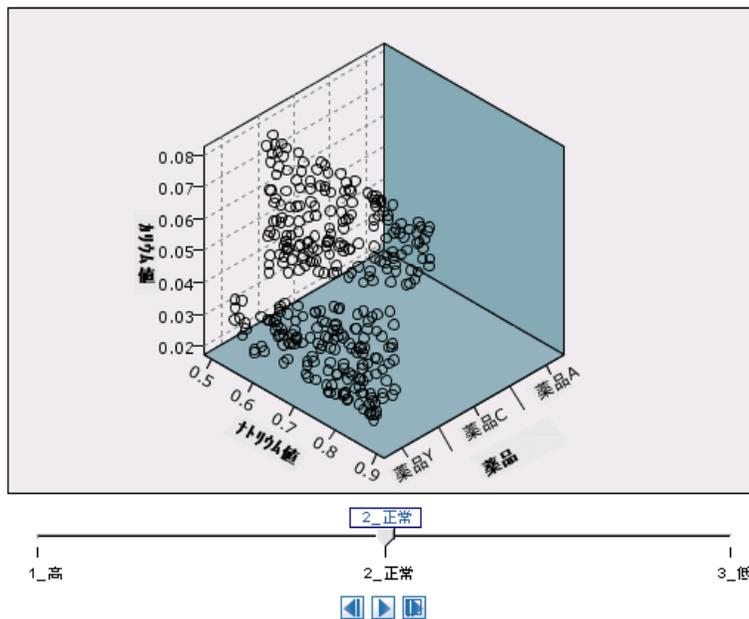
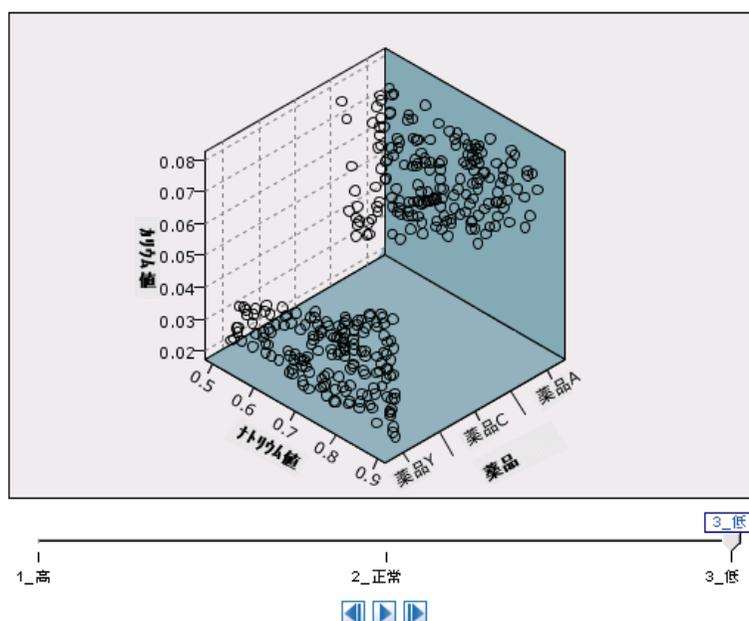


図 5-6
3つのカテゴリがある変数の散布図のアニメーション - 高血圧のスライド



[出力] タブの使用方法

すべての種類のグラフに対して、生成されたグラフのファイル名や表示に関する次のオプションを指定することができます。

注：棒グラフ ノードのグラフには、追加の設定があります。

出力名：ノードの実行時に生成されるグラフの名前を指定します。[自動]は、出力を生成するノードの名前に基づいて名前を選択します。[ユーザー設定]で別の名前を指定することもできます。

画面に出力：グラフを生成し、新しいウィンドウに表示する場合に選択します。

ファイルに出力：出力をファイルとして保存する場合に選択します。

- **出力グラフ**：グラフ形式の出力を生成する場合に選択します。棒グラフ ノードでのみ使用できます。
- **出力テーブル**：テーブル形式の出力を生成する場合に選択します。棒グラフ ノードでのみ使用できます。

- **ファイル名** :生成されたグラフまたはテーブルのファイル名を指定します。[...] ボタンを使って、ディレクトリを参照しながら特定のファイルを指定することもできます。
- **ファイルの種類** :ドロップダウン リストでファイルの種類を指定します。[出力テーブル] オプションが指定されている棒グラフ ノードを除く、すべてのグラフ作成ノードで、使用可能なグラフ ファイルの種類は次のとおりです。
 - ビットマップ (.bmp)
 - PNG (.png)
 - 出力 オブジェクト (.cou)
 - JPEG (.jpg)
 - HTML (.html)
 - その他の IBM® SPSS® Statistics アプリケーションで使用する ViZml ドキュメント (.xml)棒グラフ ノードの [出力テーブル] オプションでは、使用できるファイルの種類は次のとおりです。
 - データ (タブ区切り) (*.tab)
 - データ (カンマ区切り) (*.csv)
 - HTML (.html)
 - 出力 オブジェクト (.cou)

出力のページ分割 :出力を HTML で保存する場合、各 HTML ページのサイズを制御するためにこのオプションが有効になります(棒グラフ ノードのみに該当します)。

1 ページの行数 :[出力のページ分割] を選択する場合、各 HTML ページの長さを決めるためにこのオプションが有効になります。デフォルトは 400 行です(棒グラフ ノードのみに該当します)。

[注釈] タブの使用方法

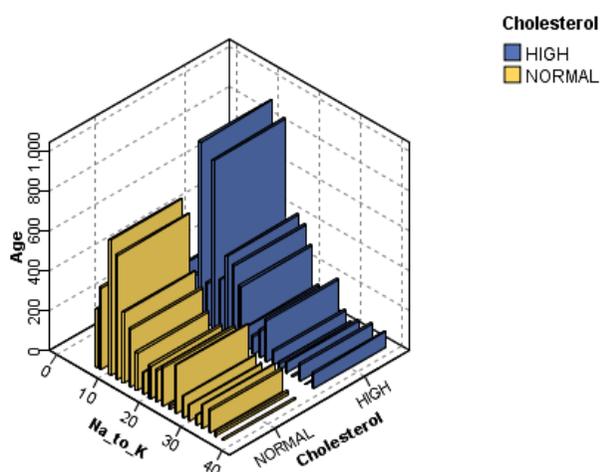
すべてのノードで使用されるこのタブには、ノード名の変更、カスタム ツールヒントの提供、および長い注釈の保存などのオプションが用意されています。詳細は、5 章 [注釈] in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。

3 次元グラフ

IBM® SPSS® Modeler の散布図および集計棒グラフには、情報を 3 次元に表示する機能があります。この機能は、モデル作成のためにサブセットを選択したり、新規フィールドを作成するために、データを視覚化する場合に役立ちます。

3 次元グラフを作成すると、グラフをクリックしてマウスでドラッグし、回転させてあらゆる角度で表示することができます。

図 5-7
x、y、z 軸を持つ集計棒グラフ



SPSS Modeler で 3 次元グラフを作成するには、次の 2 種類の方法があります。情報を 3 番目の軸にプロットする方法（真の 3 次元グラフ）と、グラフを 3 次元効果で表示する方法です。どちらの方法も、散布図や集計棒グラフで利用することができます。

情報を 3 番目の軸にプロットするには

- ▶ [グラフ作成ノード] ダイアログ ボックスで、[散布図] タブをクリックします。
- ▶ z 軸のオプションを有効にするには、[3 次元] ボタンをクリックします。
- ▶ フィールド ピッカー ボタンを使って、z 軸のフィールドを選択します。場合によっては、シンボル値フィールドしか選択できないこともあります。フィールド ピッカーが、適切なフィールドを表示します。

グラフに 3 次元効果を追加するには

- ▶ グラフを作成したら、出力ウィンドウの [グラフ] タブをクリックします。
- ▶ [3 次元] ボタンをクリックして、3 次元グラフ ビューに切り替えます。

グラフボード ノード

グラフボード ノードでは、1 つのノードで多数の異なるグラフ出力（棒グラフ、円グラフ、ヒストグラム、散布図、heatmap など）から選択することができます。まず、最初のタブで、必要なデータ フィールドを選択すると、データに使用できるグラフの種類を選択肢が表示されます。選択したフィールドに適していないグラフの種類は、自動的に除外されます。[詳細] タブでは、詳細な、またはより高度なグラフ オプションを定義できます。

注： グラフボード ノードを編集したりグラフの種類を選択したりするには、データを含むストリームにグラフボード ノードを接続する必要があります。

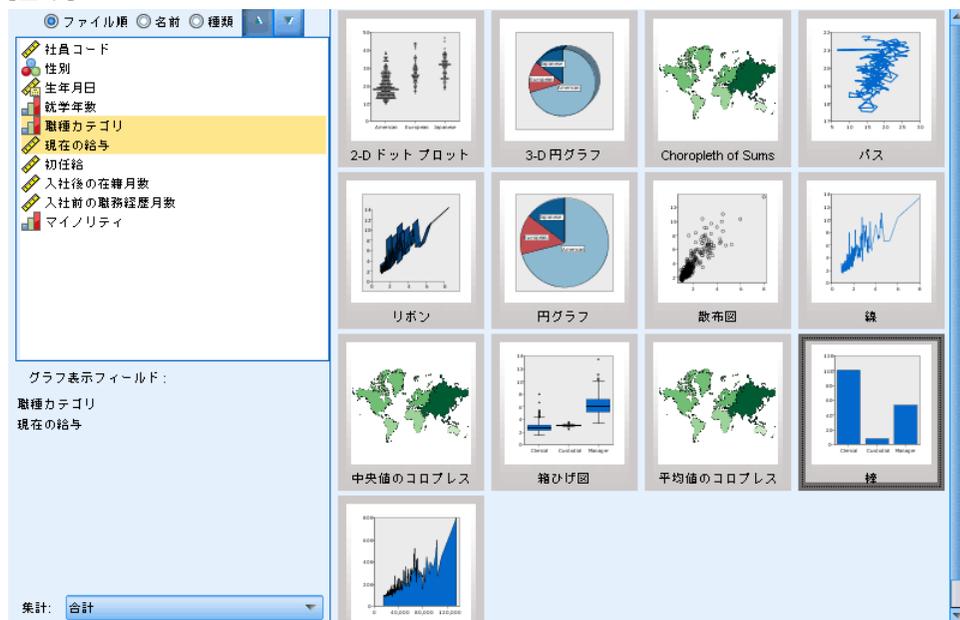
使用できる視覚化テンプレート（およびスタイルシート、およびマップ）を制御できる 2 つのボタンがあります。

管理: 視覚化テンプレート、スタイルシート、およびマップをコンピュータで管理します。視覚化テンプレート、スタイルシート、およびマップをローカル マシンでインポート、エクスポート、名前変更、および削除できます。詳細は、[p. 320 テンプレート、スタイルシート、マップ ファイルの管理](#) を参照してください。

位置： 視覚化テンプレート、スタイルシート、およびマップを保存する場所を変更します。現在の場所は、ボタンの右側に表示されます。詳細は、[p. 318 テンプレート、スタイルシート、マップの位置の設定](#) を参照してください。

グラフボード [基本] タブ

図 5-8
[基本] タブ



データを良好に表示できる視覚化の種類がわからない場合、[基本] タブを使用します。データを選択する場合、データに適切な視覚化のサブセットが提供されます。詳細は、「[Graphboard Examples](#)」(p.295) を参照してください。

- ▶ リストから 1 つまたは複数のフィールド (変数) を選択します。複数のフィールドを選択するには、Ctrl キーを押したままクリックします。

フィールドの測定レベルによって、使用できる視覚化の種類が決まります。リストでフィールドを右クリックし、オプションを選択して測定レベルを変更できます。使用できる測定レベルの種類に関する詳細は、「[フィールド \(変数\) タイプ](#)」(p.281) を参照してください。

- ▶ 視覚化の種類を選択します。使用できる種類の詳細は、[組み込まれている利用可能なグラフボード視覚化タイプ](#) p.287 . を参照してください。
- ▶ 特定の視覚化について、要約統計を選択することができます。統計が度数の統計量か連続フィールドから計算されているかによって、異なる統計のサブセットを使用できます。使用できる統計も、テンプレートによって異なります。使用できる統計の完全リストは次のステップの後にあります。

- ▶ オプションの外観およびパネル フィールドなど、さらにオプションを定義する場合、[詳細] をクリックします。詳細は、[p. 284 グラフボード \[詳細\] タブ](#) を参照してください。

連続型フィールドから計算された要約統計

- **Mean.** 中心傾向の測定値。観測値の合計をケース数で割った算術平均。
- **Median.** ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。
- **モード.** 最も多く出現する値。複数の値が最高の頻度で出現する場合は、それぞれが最頻値となります。
- **最小.** 数値型変数をとる最も小さい値。
- **最大.** 数値型変数の最大値。
- **範囲:** 最小値と最大値の差異。
- **中間域:** 範囲の中間、つまり最小値との差が最大値との差と等しい値です。
- **Sum.** 欠損値のないすべてのケースに対する変数の値の合計または全体。
- **累積合計:** 値の累積合計。各グラフィック要素は、サブグループの合計と、以前のグループすべての総合計を示します。
- **パーセント合計:** 合計したフィールドに基づいたサブグループ内の、全グループの合計に対するパーセンテージ。
- **累積パーセント合計:** 合計したフィールドに基づいたサブグループ内の、全グループの合計に対する累積パーセンテージ。各グラフィック要素は、サブグループのパーセンテージと、以前のグループすべての総パーセンテージを示します。
- **Variance.** 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **標準偏差.** 平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **標準誤差.** サンプル間で検定統計量の値がどの程度ばらついているかを測ったもの。ある統計量のサンプル分布の標準偏差です。たとえば、平均値の標準誤差はサンプル平均の標準偏差です。
- **尖度.** 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、分布の極値まで両裾が薄く

なることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。

- **歪度**：分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

次の領域の統計で、サブグループごとに複数のグラフィック要素が作成される場合があります。区間、領域、または辺のグラフィック要素を使用する場合、領域統計では、範囲を示すグラフィック要素が作成されます。他のすべてのグラフィック要素は、2 つの要素を生成しします。一方は範囲の始点を示し、もう一方は範囲の終点を示します。

- **領域:範囲**：最小値と最大値の間の値の範囲。
- **領域:平均値の 95 % の信頼区間**：母集団の平均値を含む 95 % の確率をもつ値の範囲。
- **領域:個別の 95 % の信頼区間**：このケースが指定された予測値を含む 95 % の確率をもつ値の範囲。
- **領域:平均値の上下 1 標準偏差**：平均値の上下 1 標準偏差間の値の範囲。
- **領域:平均値の上下 1 標準誤差**：平均値の上下 1 標準誤差間の値の範囲。

度数ベースの要約統計

- **度数**：行/ケースの数。
- **累積度数**：行/ケースの累積数。各グラフィック要素は、サブグループの度数と、以前のグループすべての度数合計を示します。
- **度数のパーセント**：行/ケースの総合系に対する、各サブグループ内の行/ケースのパーセンテージ。
- **度数の累積パーセント**：行/ケースの総合系に対する、各サブグループ内の行/ケースの累積パーセンテージ。各グラフィック要素は、サブグループのパーセンテージと、以前のグループすべての総パーセンテージを示します。

フィールド (変数) タイプ

フィールドのタイプおよびデータ型を示すアイコンが、変数リストの変数の隣に表示されます。また、アイコンは複数レスポンス セットも示します。

| 測定レベル | データ型 | | | |
|-------|------|-----|----|----|
| | 数値 | 文字列 | 日付 | 時間 |

| | | | | |
|---------|---|---|---|---|
| 連続型 |  | 利用不可 |  |  |
| 順序付けセット |  |  |  |  |
| Set |  |  |  |  |

| | |
|--------------------|---|
| 複数レスポンス セット、複数カテゴリ |  |
| 複数レスポンス セット、複合二分 |  |

尺度

フィールドの尺度は、視覚化作成時に重要です。以下は尺度の説明です。リストでフィールドを右クリックし、オプションを選択して尺度を一時的に変更できます。多くの場合、フィールド、カテゴリ、連続の 2 つの幅広い分類のみを考慮する必要があります。

カテゴリ。 値やカテゴリの数が限られているデータ（性別や宗教など）。カテゴリ フィールドは、文字列（英数字）変数でも、数値コードを使用してカテゴリを表す数値フィールドでもかまいません（たとえば、0 = 男性、1 = 女性など）。質的データともいいます。セット型、順序セット型、フラグはすべてカテゴリ型フィールドです。

-
-
-

連続: 間隔または比率尺度について測定したデータです。ただし、データ値は、値の順序と値の間の距離を示します。たとえば、\$72,195 の給料は、\$52,398 の給料より高く、2 つの値の距離は \$19,797 です。また、量的データ、スケールデータ、数値範囲データとも呼ばれます。

カテゴリ フィールドは、視覚化のカテゴリを定義し、通常、それぞれのグラフィック要素を描画、またはグラフィック要素をグループ化します。連続フィールドは、カテゴリ フィールドのカテゴリ内で要約されることがよくあります。たとえば、性別分類内にある収入のデフォルトの視覚化は、男性の平均収入と女性の平均収入を表示します。連続型フィールドの行値は、散布図としてプロットすることもできます。たとえば、散布図各ケースの現在の給与および初期の給与を表示できます。カテゴリ フィールドを使用して、ケースを性別ごとに分類できます。

データの型

尺度は、データ型を決定するフィールドのプロパティだけではありません。フィールドは、特定のデータ型でも保存されます。データ型は、文字列（文字など、非数値型データ）、数値（実数）、日付です。尺度と異なり、変数のデータ型を一時的に変更することはできません。データを元のデータ セットに保存する方法を変更する必要があります。

多重回答グループ(U)

一部のデータ ファイル、**マルチアンサー セット** と呼ばれる特殊な「フィールド」をサポートします。マルチアンサー セットは、通常の意味での「フィールド」ではありません。マルチアンサー セットは、複数のフィールドを使用して質問に対するレスポンスを記録し、回答者は複数の回答をすることができます。マルチアンサー セットはカテゴリ フィールドのように扱われ、カテゴリ フィールドで可能なことのほとんどを複数レスポンス セットでも行うことができます。

マルチアンサー セットは、多重二分グループの場合と、多重カテゴリ セットの場合があります。

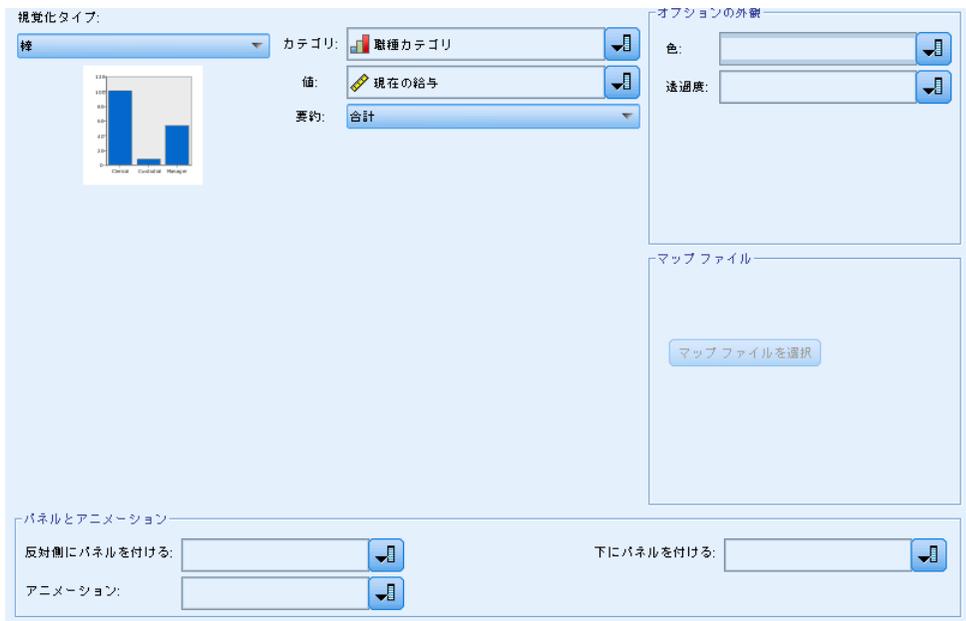
多重二分グループ 多重二分グループは、通常、値がはい/いいえ、あり/なし、チェック/チェックなしのフィールドで構成されています。フィールドは厳密に二分というわけではありませんが、セット内のすべてのフィールドは同じようにコード化されます。

たとえば、調査で「次のどのニュース ソースを信頼しますか?」という質問に 5 つの回答が設定されているとします。回答者は、各選択肢の隣のボックスをチェックして、複数の選択肢を選択することができます。5 つの回答は、データ ファイルで 5 つのフィールドとなります。No（チェックなし）は 0、Yes（チェック済み）は 1 にコード化されます。

多重カテゴリ グループ。多重カテゴリ グループは、複数のフィールドで構成され、すべて同じ方法でコード化されます。多くのレスポンス カテゴリがある場合があります。たとえば、「あなたの民族の財産を最もよく表す国民性を最大 3 つ挙げてください」という調査項目があるとします。非常に多くの回答が考えられますが、コード化のため、最も一般的な国民性 40 件に制限され、他の回答は「その他」のカテゴリに分類されます。データ ファイルでは、3 つの選択すると 3 つのフィールドに該当します。それぞれに 41 件のカテゴリがあります（40 は国民性をコード化、1 つは「その他」カテゴリ）。

グラフボード [詳細] タブ

図 5-9
[詳細] タブ



作成する視覚化の種類がわかっている場合、オプションの外観、パネルを追加する場合または視覚化をアニメーション化する場合、[詳細] タブを使用します。詳細は、「[Graphboard Examples](#)」(p. 295) を参照してください。

- ▶ [基本] タブで視覚化の種類を選択する場合、その種類が表示されます。[基本] タブで選択しない場合は、ドロップダウン リストから選択します。視覚化の種類の詳細は、[組み込まれている利用可能なグラフボード視覚化タイプ](#) p. 287 . を参照してください。
- ▶ 視覚化の種類に必要なフィールド (変数) を指定するコントロールが、視覚化のサムネイル イメージのすぐ右側にあります。これらのフィールドのすべてを指定する必要があります。
- ▶ 特定の視覚化について、要約統計を選択することができます。棒グラフなど、透過度の表示にこれらの要約オプションのいずれかを使用することができます。要約統計の詳細は、[グラフボード \[基本\] タブ](#) p. 279 . を参照してください。
- ▶ 1 つまたは複数のオプションの表示を選択することができます。視覚化に他のフィールドを追加することができる次元性を追加することができます。たとえば、フィールドを使用すると、散布図のポイントのサイズを変更することができます。オプションの外観に関する詳細は、「[外観、オー](#)

「[オーバーレイ、パネル、およびアニメーション](#)」（p. 271）を参照してください。透過度の表示は、スクリプトにはサポートされていません。

- ▶ マップの視覚化を作成している場合、[マップファイル] グループが使用されるマップ ファイルを表示します。デフォルトのマップ ファイルがある場合、このファイルが表示されます。マップ ファイルを変更するには、[マップファイルを選択] をクリックして [マップの選択] ダイアログ ボックスを表示します。このダイアログ ボックスでデフォルトのマップ ファイルを指定することもできます。詳細は、[p. 285 マップ視覚化のためのマップ ファイルの選択](#) を参照してください。
- ▶ 1 つまたは複数のパネリングまたはアニメーションを選択することができます。パネリングおよびアニメーションのオプションに関する詳細は、「[外観、オーバーレイ、パネル、およびアニメーション](#)」（p. 271）を参照してください。

マップ視覚化のためのマップ ファイルの選択

マップの視覚化テンプレートを選択する場合、マップを描画するための地理的情報を定義するマップ ファイルが必要です。デフォルトのマップ ファイルがある場合、マップの視覚化に使用されます。異なるマップ ファイルを選択するには、[詳細] タブの [マップファイルを選択] をクリックして [マップの選択] ダイアログ ボックスを表示します。

[マップの選択] ダイアログ ボックスを使用して、プライマリ マップ ファイルと参照マップ ファイルを選択します。マップ ファイルは、マップを描画するための地理的情報を定義します。アプリケーションは標準 マップ ファイルとともにインストールされます。他に使用する ERI シェープファイルがある場合、まずシェープファイルを SMZ ファイルに変換する必要があります。詳細は、[p. 322 マップ シェープファイルの変換と配布](#) を参照してください。マップを変換した後、[テンプレート ピッカー] ダイアログ ボックスの [管理...] をクリックし、[マップを選択] ダイアログ ボックスで使用できるよう、管理システムにマップをインポートします。

マップ ファイルを指定する際に考慮するポイントは以下のとおりです。

- すべてのマップ テンプレートには少なくとも 1 つのマップ ファイルが必要です。
- 通常、マップ ファイルはマップのキー属性をデータ キーにリンクさせます。
- テンプレートにデータ キーにリンクするマップ キーを必要としない場合、参照マップの要素を描画するための座標（緯度や経度など）を指定する参照マップ ファイルおよびフィールドが必要です。
- オーバーレイ マップ テンプレートには、プライマリ マップ ファイルと参照マップ ファイルが必要です。参照マップが最初に描画されます。

属性や特徴など、マップの用語に関する詳細は、[マップの主要コンセプト](#)（ p. 323 ）を参照してください。

マップ ファイル: 管理システムのマップ ファイルを選択できます。事前にインストールされたマップ ファイルやインポートしたマップ ファイルがあります。マップ ファイル管理に関する詳細は、[「テンプレート、スタイルシート、マップ ファイルの管理」](#)（ p. 320 ）を参照してください。

マップ キー: マップ ファイルからデータ キーにリンクさせるキーとして使用する属性を指定します。

このマップ ファイルと設定をデフォルトとして保存: 選択したマップ ファイルをデフォルトとして使用します。デフォルトのマップ ファイルを指定した場合、マップの視覚化を作成するごとにマップ ファイルを指定する必要はありません。

データ キー: [テンプレート ピッカーの詳細] タブに表示されるものと同じ値が表示されます。選択した特定のマップ ファイルによりキーを変更する必要がある場合に便利です。

視覚化のすべてのマップのフィーチャを表示: このオプションを選択すると、マッチするデータ キーの値がない場合でもマップ内のすべてのフィーチャが表示されます。データのあるフィーチャのみを表示する場合は、このオプションをオフにします。[マッチしないマップ キー] リストに表示されたマップ キーで指定されたフィーチャは、視覚化には表示されません。

マップ値とデータ値を比較: マップ キーとデータ キーはお互いにリンクし、マップの視覚化を作成します。これらの 2 つのキーの値は一致しなければなりません。一致しない場合はマップの視覚化を作成できません。[比較] をクリックして、データ キーとマップ キーの値が一致しているかどうかをテストします。表示されたアイコンにより、比較の状態が分かります。これらのアイコンについては、下記のとおりです。比較が実行され、マップ キーの値が一致しないデータ キーの値がある場合、データ キーの値が[一致しないデータキー] リストに表示されます。[一致しないデータキー] リストでは、マッチするデータキー値がないマップ キーを確認することもできます。[視覚化ですべてのマップフィーチャを表示] がオフの場合、これらのマップ キー値で指定されたフィーチャは表示されません。

テーブル 5-1
比較アイコン

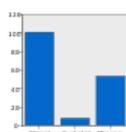
| アイコン | 説明 |
|---|---|
|  | 比較が実行されていません。[比較] をクリックする前のデフォルトの状態です。データ キーとマップ キーの値が一致しているか分からないため、注意して続行する必要があります。 |
|  | 比較が実行され、データ キーとマップ キーの値が一致しています。データ キーのすべての値で、マップ キーに識別された一致フィーチャがあります。 |

| アイコン | 説明 |
|---|--|
|  | 比較が実行され、データ キーとマップ キーの値が一致していないものがあります。一部のデータ キーについて、マップ キーに識別された一致フィーチャがありません。注意して続行する必要があります。続行すると、マップの視覚化に含まれないデータ値があります。 |
|  | 比較が実行され、一致しているデータ キーとマップ キーの値がありません。続行してもマップが表示されないため、異なるデータ キーまたはマップ キーを選択する必要があります。 |

組み込まれている利用可能なグラフボード視覚化タイプ

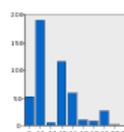
様々な視覚化タイプを作成することができます。組み込まれている次のタイプは、[基本] タブおよび [詳細] タブのどちらでも使用できます。テンプレートの一部の説明（特にマップ テンプレート）は、[特殊テキスト] を使用して [詳細] タブで指定されたフィールド（変数）を識別します。

テーブル 5-2
利用可能なグラフ タイプ



棒
連続する数値範囲の要約統計量を計算し、カテゴリ フィールドについて、カテゴリごとの結果を棒グラフで表示します。

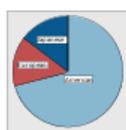
必須: カテゴリ フィールドおよび連続型フィールド。



度数の棒グラフ

カテゴリ フィールドのカテゴリごとに、行またはケースの比率を棒グラフで表示します。分布グラフのノードを使用して、このグラフを作成することもできます。このノードには追加のオプションもあります。詳細は、[p. 341 棒グラフ ノードを参照してください](#)。

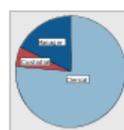
必須: 1 つのカテゴリ フィールド。



円グラフ

連続する数値フィールドの合計値を計算し、その合計値のカテゴリ フィールドのカテゴリごとの比率を、円を分割して表示します。

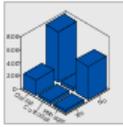
必須: カテゴリ フィールドおよび連続型フィールド。



度数の円グラフ

カテゴリ フィールドのカテゴリごとに、行またはケースの比率を円を分割して表示します。

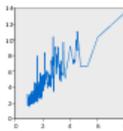
必須: 1 つのカテゴリ フィールド。



3-D 棒

連続する数値フィールドの要約統計量を計算し、2つのカテゴリフィールドまたはカテゴリ変数の間のカテゴリの交差結果を表示します。

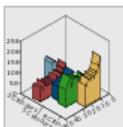
必須: カテゴリ フィールドおよび連続型フィールドのペア。



線(L)

別のフィールドの値についてフィールドの要約統計量を計算し、値を接続する線を引きます。プロットグラフのノードを使用して、折れ線グラフを作成することもできます。このノードには追加のオプションもあります。[詳細は、p. 332 散布図ノードを参照してください。](#)

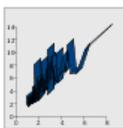
必須: 任意の種類のフィールドのペア。



3-D 面

あるフィールドの値とプロットして、カテゴリフィールドごとに分割して表示します。カテゴリごとに面要素が表示されます。

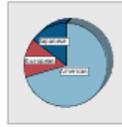
必須: カテゴリ フィールドと任意の種類のフィールドのペア。



リボン

別のフィールドの値についてフィールドの要約統計量を計算し、値を接続するリボンを引きます。帯は、基本的には 3-D 効果付きの線です。本格的な 3-D グラフではありません。

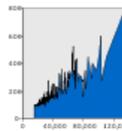
必須: 任意の種類のフィールドのペア。



3-D 円

3-D 効果が追加されている以外は、円グラフと同じです。

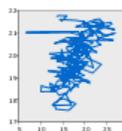
必須: カテゴリ フィールドおよび連続型フィールド。



面

別のフィールドの値についてフィールドの要約統計量を計算し、値を接続する面を描きます。エリアが、下部の色づけされたスペースの線と似ているという点で、線とエリアの誤差は最小となります。ただし、色外観を使用すると、線が分割されたり、エリアが積み上げられた状態になります。

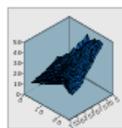
必須: 任意の種類のフィールドのペア。



パス

フィールドの値を他の値とプロットして、元のデータセットの出現順に値を線をつないで表示します。順番は、パスとライン間の大きな違いです。

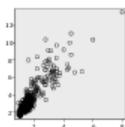
必須: 任意の種類のフィールドのペア。



表面

3つのフィールドの値をそれぞれの値とプロットして、値同士を面でつないで表示します。

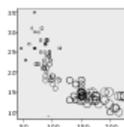
必須: 任意の種類の 3つのフィールド。



散布図

フィールドの値を、他の値とプロットして表示します。フィールド間の関係を（存在する場合に）際立たせることができます。プロットグラフのノードを使用して、散布図を作成することもできます。このノードには追加のオプションもあります。詳細は、[p. 332 散布図ノード](#)を参照してください。

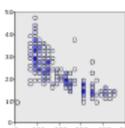
必須: 任意の種類のフィールドのペア。



バブルプロット

基本的な散布図と同様に、フィールドの値を、他の値とプロットして表示します。散布図との違いは、3つ目のフィールドの値が、各ポイントのサイズを変更するのに使われる点です。

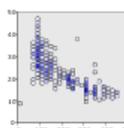
必須: 任意の種類の 3 つのフィールド。



ビン分割散布図

基本的な散布図と同様に、フィールドの値を、他の値とプロットして表示します。散布図との違いは、類似値がグループのビンに分割され、色またはサイズの外観を使用してそれぞれのビンに含まれるケース数が示される点です。

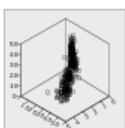
必須: 連続型フィールドのペア。



六角ビン分割散布図

ビン分割散布図の説明を参照してください。違いは、基礎となるビンの形で、丸ではなく六角形をしています。結果として作成される六角ビン分割散布図は、分割散布図と似たものになります。ただし、基礎となるビンの形状のために、ビンごとの値の数はグラフごとに異なります。

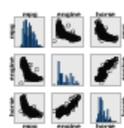
必須: 連続型フィールドのペア。



3-D 散布図

3つのフィールドを、相互にプロットして表示します。フィールド間の関係を（存在する場合に）際立たせることができます。プロットグラフのノードを使用して、3-D 散布図を作成することもできます。このノードには追加のオプションもあります。詳細は、[p. 332 散布図ノード](#)を参照してください。

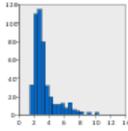
必須: 任意の種類の 3 つのフィールド。



散布図行列 (SPLOM)

あるフィールドの値を、各フィールドの他の値とプロットして表示します。SPLOM は、散布図のテーブルに似ています。SPLOM には、フィールドごとのヒストグラムも含まれます。

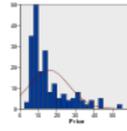
必須: 複数の連続型フィールド。



ヒストグラム(S)

フィールドの度数分布表を表示します。ヒストグラムは、分布の種類を決定したり、分布が歪んでいるかを確認するのに役立つ場合があります。ヒストグラム グラフのノードを使用して、このグラフを作成することもできます。このノードには追加のオプションもあります。詳細は、[p. 347 ヒストグラムの \[プロット\] タブ](#) を参照してください。

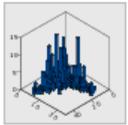
必須: 任意の種類の 1 つのフィールド。



正規分布を示すヒストグラム

正規分布の曲線を重ね合わせて、連続型フィールドの度数分布表を表示します。

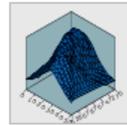
必須: 1 つの連続型フィールド。



3-D ヒストグラム

1 組の連続型フィールドの度数分布表を表示します。

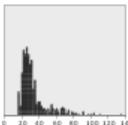
必須: 連続型フィールドのペア。



3 次元密度

1 組の連続型フィールドの度数分布表を表示します。棒ではなく面を使用して分布を表示する以外は、3-D ヒストグラムに似ています。

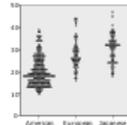
必須: 連続型フィールドのペア。



ドットプロット

ケースまたは行を表示し、x 軸上の異なるデータ点に積み上げます。このグラフは、データの分布を表している点で、ヒストグラムに似ていますが、特定のビン (値の範囲) の集計数ではなく、それぞれのケースまたは行を表示しています。

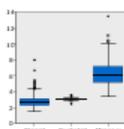
必須: 任意の種類の 1 つのフィールド。



2-D ドットプロット

各ケースまたは各行を表示し、カテゴリフィールドのカテゴリごとに y 軸上の異なるデータ点に積み上げます。

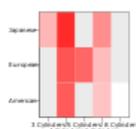
必須: カテゴリ フィールドおよび連続型フィールド。



箱ひげ図

カテゴリ フィールドのカテゴリごとに連続型フィールドの 5 種類の統計量 (最小値、最初の 4 分位、中央値、3 番目の 4 分位、最大値) を計算します。結果は箱ひげ図またはスキーマの要素として表示されます。箱ひげ図は、連続データの分布がカテゴリ間でどのように変化するかを確認するのに役立つ場合があります。

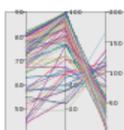
必須: カテゴリ フィールドおよび連続型フィールド。



ヒート マップ

2 つのカテゴリ フィールドのカテゴリの交差結果に関して、連続型フィールドの平均値を計算します。

必須: カテゴリ フィールドおよび連続型フィールドのペア。



平行モデル

フィールドごとに平行軸を作成し、データ内にあるすべての行またはケースのフィールドの値に線を引きます。

必須: 複数の連続型フィールド。



度数のコロプレス

カテゴリ型フィールドの各カテゴリの度数 ([データキー]) を計算し、カテゴリに対応するマップ フィーチャの度数を示すために彩度を使用するマップを描画します。

必須: 1 つのカテゴリ フィールド。キーが [データキー] カテゴリであるマップ ファイル。



平均値/中央値/合計のコロプレス

カテゴリ型フィールド ([データキー]) の各カテゴリに対する連続型フィールド ([色]) の平均値、中央値、連続フィールドの平均値、中央値を計算し、カテゴリに対応するマップ フィーチャで計算された統計を示すために彩度を使用するマップを描画します。

必須: カテゴリ フィールドおよび連続型フィールド。キーが [データキー] カテゴリであるマップ ファイル。



値のコロプレス

別のカテゴリ型フィールド ([データキー]) で分割した値に対応するマップ フィーチャのカテゴリ型フィールド ([色]) の値を示すために色を使用するマップを描画します。各フィーチャの [色] フィールドに複数のカテゴリ値がある場合、中央値が使用されます。

必須: カテゴリ型フィールドのペア。キーが [データキー] カテゴリであるマップ ファイル。



度数のコロプレスの座標

コロプレス マップの点を描画するための座標を指定する 2 つの連続型フィールド（経度と緯度）があるという点を除き、度数のコロプレスと類似しています。

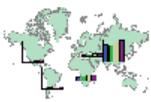
必須: カテゴリ フィールド、および連続型フィールドのペア。キーが [データキー] カテゴリであるマップファイル。



値のコロプレスの座標

コロプレス マップの点を描画するための座標を指定する 2 つの連続型フィールド（経度と緯度）があるという点を除き、値のコロプレスと類似しています。

必須: カテゴリ フィールド、および連続型フィールドのペア。キーが [データキー] カテゴリであるマップファイル。



マップの棒グラフ

連続型フィールド（値）の要約統計量を計算し、各マップ フィーチャ（データキー）のカテゴリ フィールド（カテゴリ）の各カテゴリに対する結果を各マップ フィーチャの中央に棒グラフとして表示します。

必須: カテゴリ フィールドおよび連続型フィールドのペア。キーが [データキー] カテゴリであるマップファイル。



平均値/中央値/合計のコロプレスの座標

コロプレス マップの点を描画するために座標を指定する 2 つの連続型フィールド（経度と緯度）があるという点を除き、平均値/中央値/合計のコロプレスと類似しています。

必須: カテゴリ フィールドおよび 3 つの連続型フィールド。キーが [データキー] カテゴリであるマップ ファイル。



マップの度数の棒グラフ

各マップ フィーチャ（データキー）のカテゴリ フィールド（カテゴリ）の各カテゴリにおける行/ケースの割合を計算し、マップと、各マップ フィーチャの中央に棒グラフを描画します。

必須: カテゴリ型フィールドのペア。キーが [データキー] カテゴリであるマップ ファイル。



マップの度数の円グラフ

各マップ フィーチャ（データキー）のカテゴリ フィールド（カテゴリ）の各カテゴリにおける行/ケースの割合を計算し、マップと、各マップ フィーチャの中央に円グラフのスライスとして割合を描画します。

必須: カテゴリ型フィールドのペア。キーが [データキー] カテゴリであるマップ ファイル。



マップの円グラフ

各マップ フィーチャ (データキー) のカテゴリ フィールド (カテゴリ) の各カテゴリにおける連続型フィールドの合計 (値) を計算し、マップと、各マップ フィーチャの中央に円グラフのスライスとして合計を描画します。

必須: カテゴリ フィールドおよび連続型フィールドのペア。キーが [データキー] カテゴリであるマップ ファイル。



マップの折れ線グラフ

各マップ フィーチャ (データキー) の別のフィールド (X) に対する連続型フィールドの要約統計量 (Y) を計算し、マップと、各マップ フィーチャの中央に値を繋ぐ折れ線グラフを描画します。

必須: カテゴリ フィールドと任意の種類のフィールドのペア。キーが [データキー] カテゴリであるマップ ファイル。



参照マップの座標

点の座標を示す連続型フィールド (経度および緯度) を使用してマップおよび点を描画します。

必須: 範囲フィールドのペア。マップ ファイル。



参照マップの矢印

マップと、各矢印の始点 (始点の経度と始点の緯度) および終点 (終点の経度と終点の緯度) が描画されます。データのレコード/ケースはマップ内の矢印に対応します。

必須: 4 つの連続型フィールド。マップ ファイル。



ポイント オーバーレイ マップ

参照マップを描画し、点フィーチャを連続型フィールド (色) で色づけして別のポイント マップに重ねます。

必須: カテゴリ型フィールドのペア。キーが [データキー] カテゴリであるポイント マップ ファイル。参照マップ ファイル。



ポリゴン オーバーレイ マップ

参照マップを描画し、ポリゴン フィーチャを連続型フィールド (色) で色づけして別のポリゴン マップに重ねます。

必須: カテゴリ型フィールドのペア。キーが [データキー] カテゴリであるポリゴン マップ ファイル。参照マップ ファイル。



ライン オーバーレイ マップ

参照マップを描画し、ライン フィーチャを連続型フィールド (色) で色づけして別のライン マップに重ねます。

必須: カテゴリ型フィールドのペア。キーが [データキー] カテゴリであるライン マップ ファイル。参照マップ ファイル。

Creating Map Visualizations

For many visualizations, you have to make only two choices: fields (variables) of interest and a template to visualize those fields. No additional choices or actions are required. Map visualizations require at least one additional step: select a map file that defines the geographic information for the map visualization.

The basic steps for creating a simple map are the following:

- ▶ Select the fields of interest on the Basic tab. For information about the type and number of fields required for different map visualizations, see [組み込まれている利用可能なグラフボード視覚化タイプ](#) p.287 .
- ▶ Select a map template.
- ▶ Click the Detailed tab.
- ▶ Check that the **Data Key** and other required dropdown lists are set to the correct fields.
- ▶ In the Map Files group, click **Select a Map File**.
- ▶ Use the Select Maps dialog box to choose the map file and map key. The map key's values must match the values for of the field specified by **Data Key**. You can use the **Compare** button to compare these values. If you select an overlay map template, you will also need to choose a reference map. The reference map is not keyed to the data. It is used as the background for the main map. For more information about the Select Maps dialog box, see [マップ視覚化のためのマップ ファイルの選択](#) p.285 .
- ▶ Click **OK** to close the Select Maps dialog box.

- ▶ In the Graphboard Template Chooser, click **Run** to create the map visualization.

Graphboard Examples

This section includes several different examples to demonstrate the available options. The examples also provide information for interpreting the resulting visualizations.

These examples use the stream named `graphboard.str`, which references the data files named `employee_data.sav`, `customer_subset.sav`, and `worldsales.sav`. These files are available from the Demos folder of any IBM® SPSS® Modeler Client installation. This can be accessed from the SPSS Modeler program group on the Windows Start menu. The `graphboard.str` file is in the streams folder.

It is recommended that you read the examples in the order presented. Subsequent examples build on previous ones.

Example: Bar Chart with a Summary Statistic

We will create a bar chart that summarizes a continuous numeric field/variable for each category of a set/categorical variable. Specifically, we will create a bar chart that shows the mean salary for men and women.

This and several of the following examples use Employee data, which is a hypothetical dataset containing information about a company's employees.

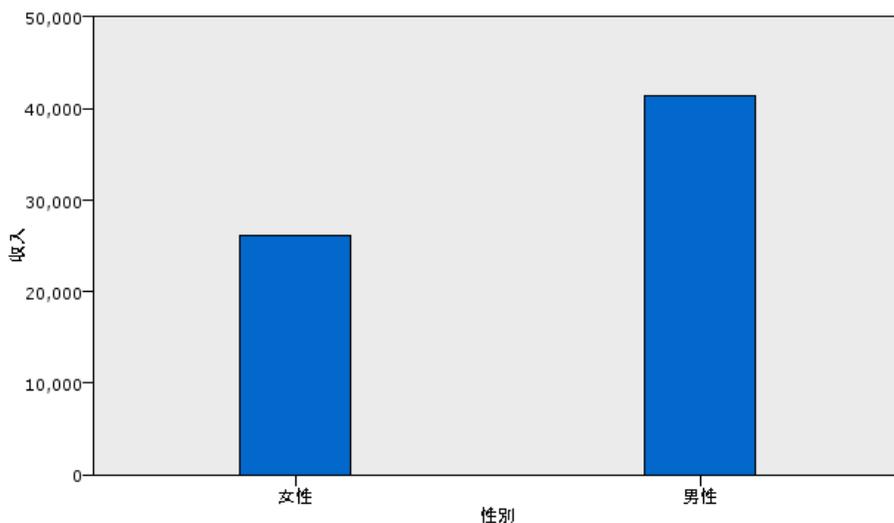
- ▶ Add a Statistics File source node that points to `employee_data.sav`.
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Gender and Current Salary. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select **Bar**.
- ▶ From the Summary drop-down list, select **Mean**.

図 5-11
Basic tab selections, bar chart with a summary statistic



- ▶ Click Run.
- ▶ On the resulting display, click the “Display field and value labels” toolbar button (the second of the group of two in the center of the toolbar).

図 5-12
Bar chart with a summary statistic



We can observe the following:

- Based on the height of the bars, it is clear that the mean salary for males is greater than the mean salary for females.

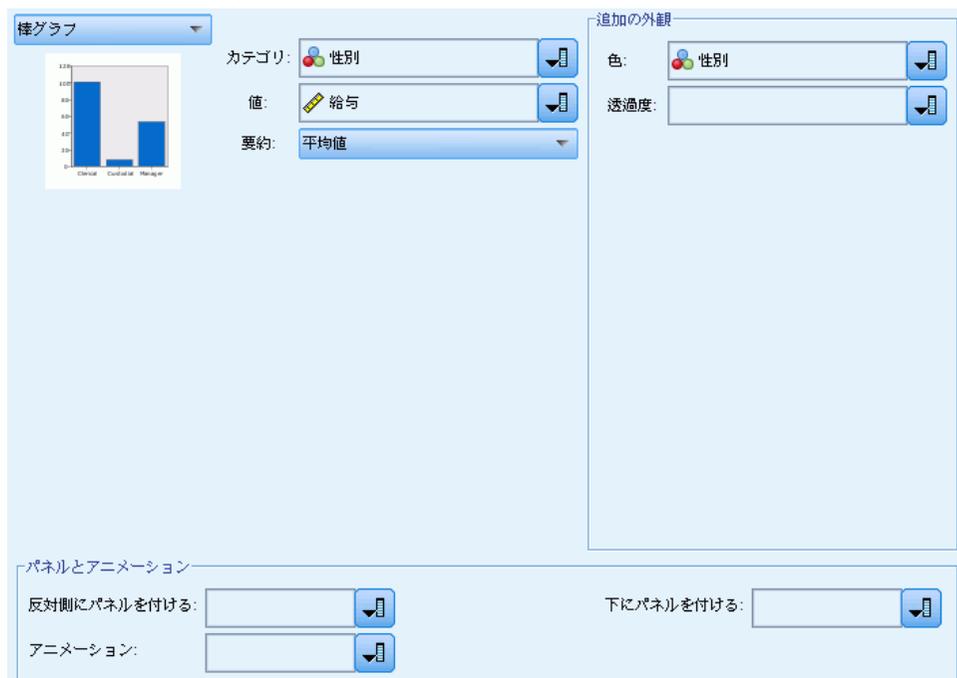
Example: Clustered Bar Chart with a Summary Statistic

We will now create a clustered bar chart to see whether the difference in mean salary between males and females is subject to job type. Perhaps females, on average, make more than males for certain job types.

Note: This example uses Employee data.

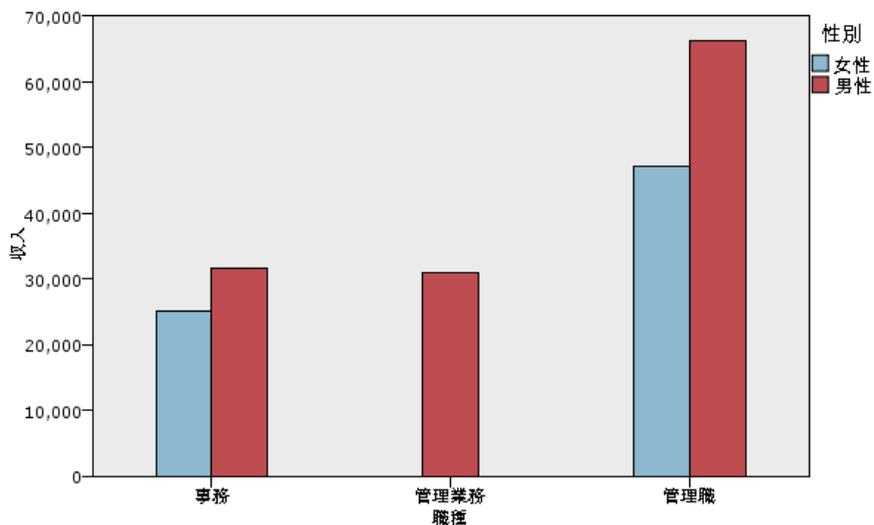
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Employment Category and Current Salary. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select **Bar**.
- ▶ From the Summary list, select **Mean**.
- ▶ Click the Detailed tab. Note that your selections from the previous tab are reflected here.
- ▶ In the Optional Aesthetics group, choose gender from the Color drop-down list.

図 5-13
Detailed tab selections, clustered bar chart



▶ Click Run.

図 5-14
Clustered bar chart



We can observe the following:

- The difference in mean salaries for each job type does not appear to be as great as it was in the bar chart that compared the mean salaries for all males and females. Perhaps there are varying number of males and females in each group. You could check this by creating a bar chart of counts.
- Regardless of job type, the mean salary for males is always greater than the mean salary for females.

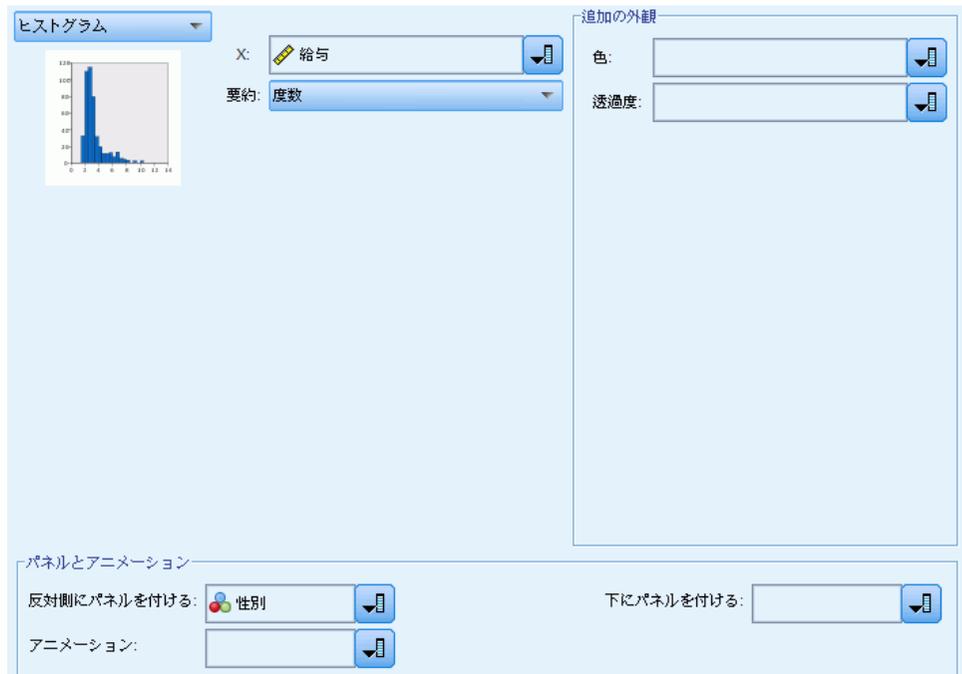
Example: Paneled Histogram

We will create a histogram that is paneled by gender so we can compare the frequency distributions of salary for men and women. The frequency distribution shows how many cases/rows lie within specific salary ranges. The paneled histogram can help us further analyze the difference in salaries between genders.

Note: This example uses Employee data.

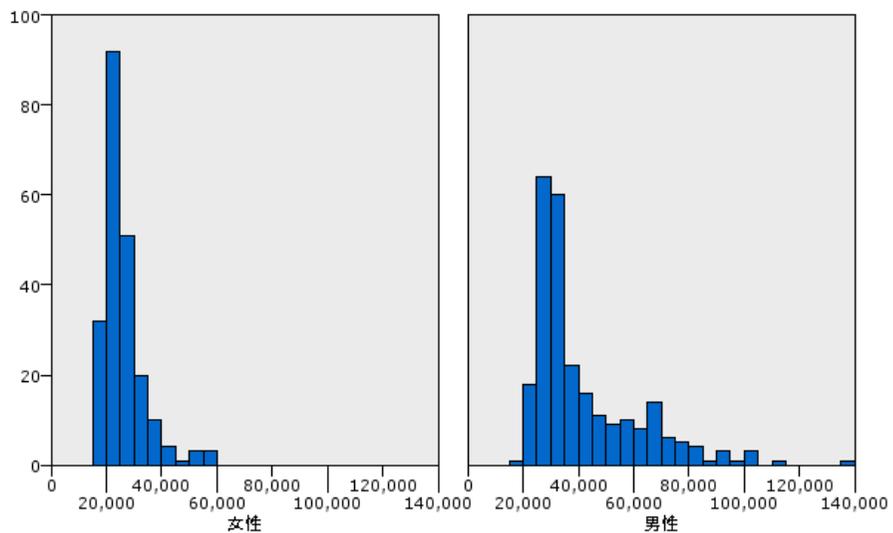
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Current Salary.
- ▶ Select **Histogram**.
- ▶ Click the Detailed tab.
- ▶ In the Panels and Animation group, choose gender from the Panel Across drop-down list.

図 5-15
Detailed tab selections, paneled histogram



▶ Click Run.

図 5-16
Paneled histogram



We can observe the following:

- Neither frequency distribution is a normal distribution. That is, the histograms do not resemble bell curves, as they would if the data were normally distributed.
- The taller bars are on the left side of each graph. Therefore, for both men and women, more make lower rather than higher salaries.
- The frequency distributions of salary among men and women are not equal. Notice the shape of the histograms. There are more men who make higher salaries than woman who make higher salaries.

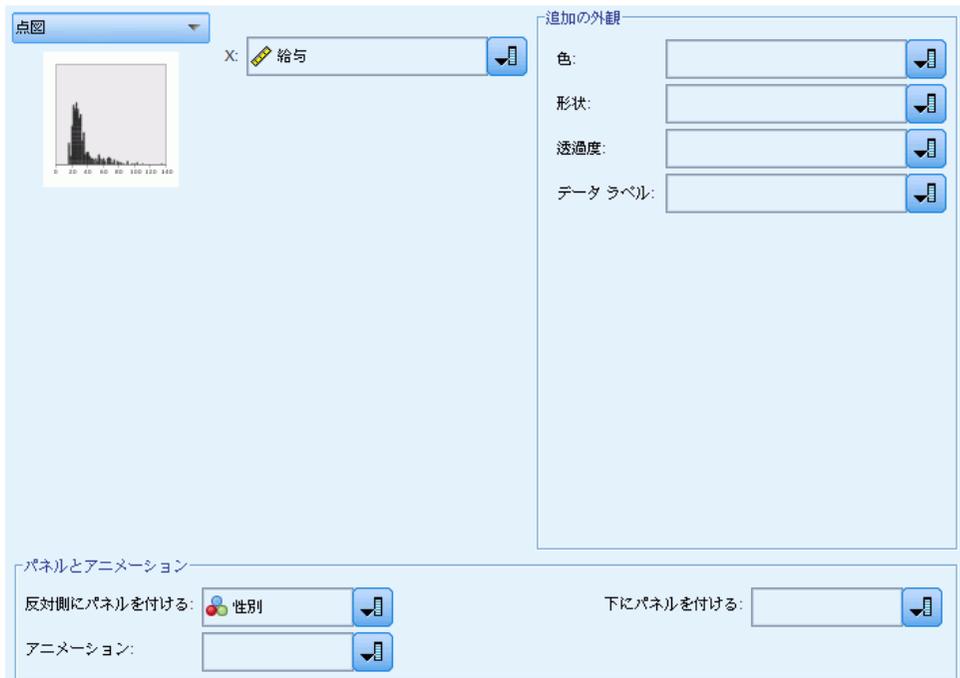
Example: Paneled Dot Plot

Like a histogram, a dot plot shows the distribution of a continuous numeric range. Unlike a histogram, which shows counts for binned ranges of data, a dot plot shows every row/case in the data. Therefore, a dot plot provides more granularity compared to the histogram. In fact, using a dot plot may be the preferred starting point when analyzing frequency distributions.

Note: This example uses Employee data.

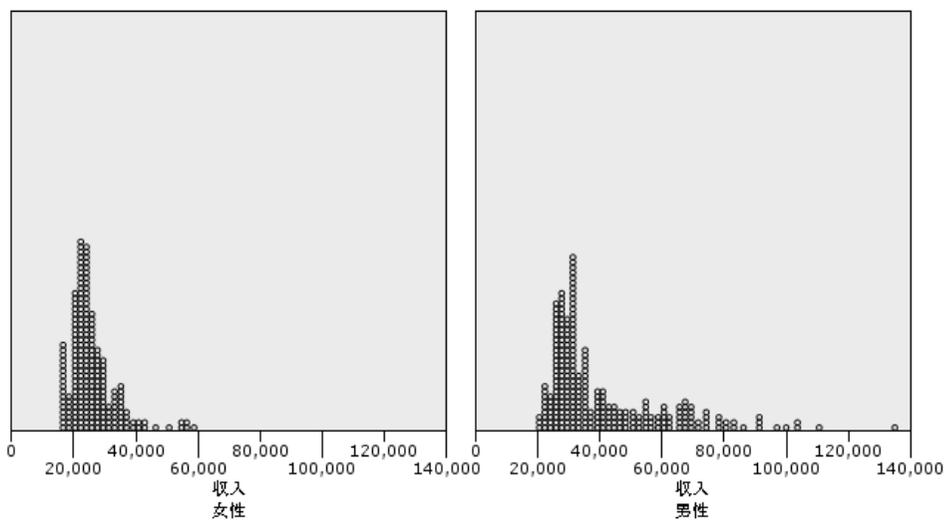
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Current Salary.
- ▶ Select **Dot Plot**.
- ▶ Click the Detailed tab.
- ▶ In the Panels and Animation group, choose gender from the Panel Across drop-down list.

図 5-17
Detailed tab selections, paneled dot plot



- ▶ Click Run.
- ▶ Maximize the resulting output window to view the plot more clearly.

図 5-18
Paneled dot plot



Compared to the histogram (see [Example: Paneled Histogram](#) p.299), we can observe the following:

- The peak at 20,000 that appeared in the histogram for females is less dramatic in the dot plot. There are many cases/rows concentrated around that value, but most of those values are closer to 25,000. This level of granularity is not apparent in the histogram.
- Although the histogram for males suggests that the mean salary for males gradually declines after 40,000, the dot plot shows that the distribution is fairly uniform after that value, to 80,000. At any one salary value in that range, there are three or more males who earn that particular salary.

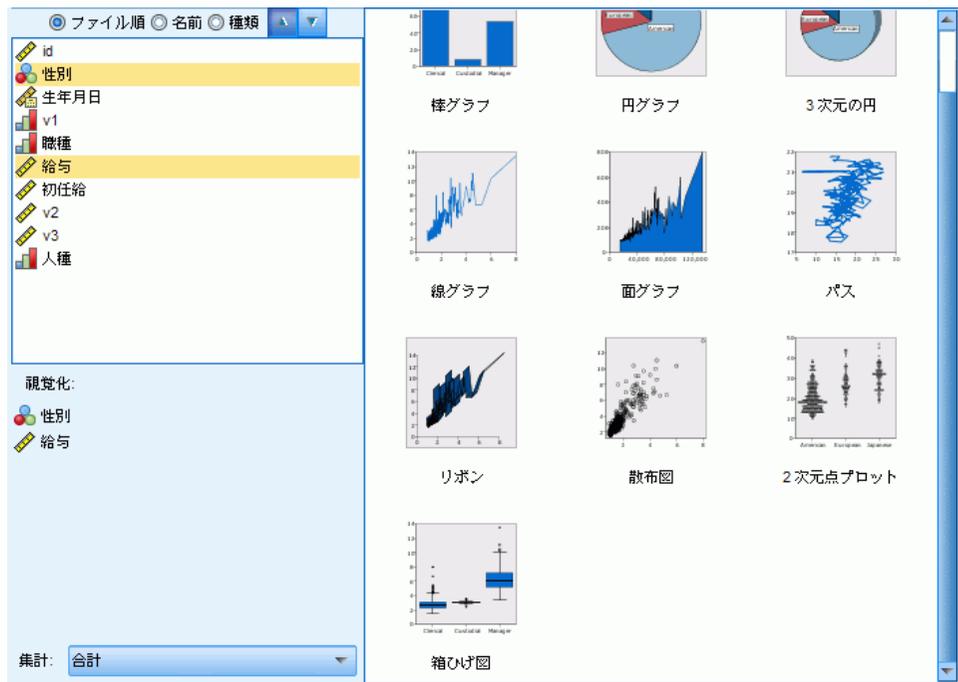
Example: Boxplot

A boxplot is another useful visualization for viewing how the data are distributed. A boxplot contains several statistical measures that we will explore after creating the visualization.

Note: This example uses Employee data.

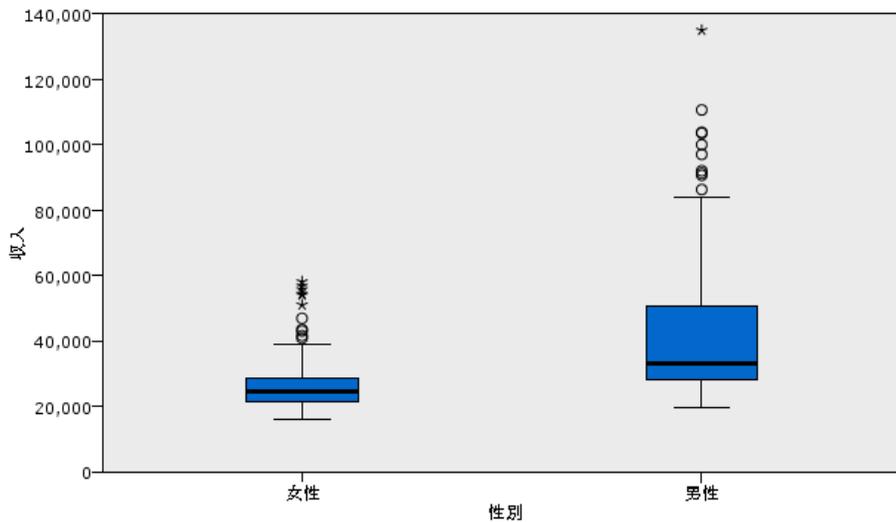
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Gender and Current Salary. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select **Boxplot**.

図 5-19
Basic tab selections, boxplot



▶ Click Run.

図 5-20
Boxplot



Let's explore the different parts of the boxplot:

- The dark line in the middle of the boxes is the median of salary. Half of the cases/rows have a value greater than the median, and half have a value lower. Like the mean, the median is a measure of central tendency. Unlike the mean, it is less influenced by cases/rows with extreme values. In this example, the median is lower than the mean (compare to [Example: Bar Chart with a Summary Statistic](#) p.295). The difference between the mean and median indicates that there are a few cases/rows with extreme values that are elevating the mean. That is, there are a few employees who earn large salaries.
- The bottom of the box indicates the 25th percentile. Twenty-five percent of cases/rows have values below the 25th percentile. The top of the box represents the 75th percentile. Twenty-five percent of cases/rows have values above the 75th percentile. This means that 50% of the case/rows lie within the box. The box is much shorter for females than for males. This is one clue that salary varies less for females than for males. The top and bottom of the box are often called **hinges**.
- The T-bars that extend from the boxes are called **inner fences** or **whiskers**. These extend to 1.5 times the height of the box or, if no case/row has a value in that range, to the minimum or maximum values. If the data are distributed normally, approximately 95% or the data are expected to lie between the inner fences. In this example, the inner fences extend less for females compared to males, another indication that salary varies less for females than for males.
- The points are **outliers**. These are defined as values that do not fall in the inner fences. Outliers are extreme values. The asterisks or stars are **extreme outliers**. These represent cases/rows that have values more than three times the height of the boxes. There are several outliers for both females and males. Remember that the mean is greater than the median. The greater mean is caused by these outliers.

Example: Pie Chart

We will now use a different dataset to explore some other visualization types. The dataset is `customer_subset`, which is a hypothetical data file that contains information about customers.

We will first create a pie chart to check the proportion of customers in different geographic regions.

- ▶ Add a Statistics File source node that points to `customer_subset.sav`.

- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Geographic indicator.
- ▶ Select Pie of Counts.

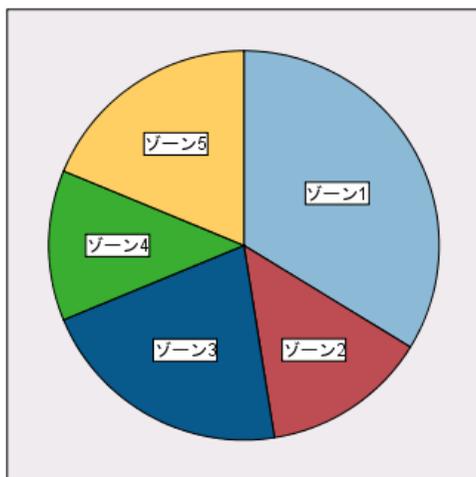
図 5-21

Basic tab selections, pie chart

The screenshot displays a software interface for data visualization. On the left, a list of fields is shown, with 'アメリカの地域' (Region of America) selected. Below this list, the 'グラフ表示フィールド:' (Graph display field) is set to 'アメリカの地域', and the '集計:' (Aggregation) is set to '度数' (Count). On the right, four chart types are presented in a grid: 'ビン度数の棒' (Bar chart), 'ビン度数の円' (Pie chart), 'ヒストグラム' (Histogram), and 'ドットプロット' (Dot plot). The 'ビン度数の円' chart is highlighted with a dashed border, indicating it is the selected option.

- ▶ Click Run.

図 5-22
Pie chart



We can observe the following:

- Zone 1 has more customers than in each of the other zones.
- Customers are equally distributed among the other zones.

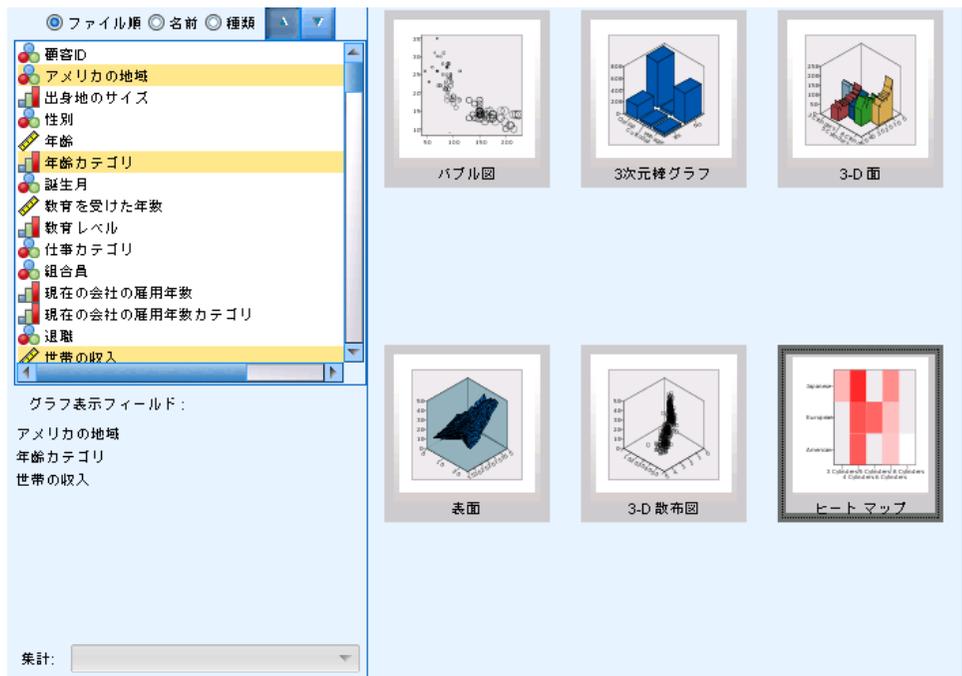
Example: Heat Map

We will now create a categorical heat map to check the mean income for customers in different geographic regions and age groups.

Note: This example uses `customer_subset`.

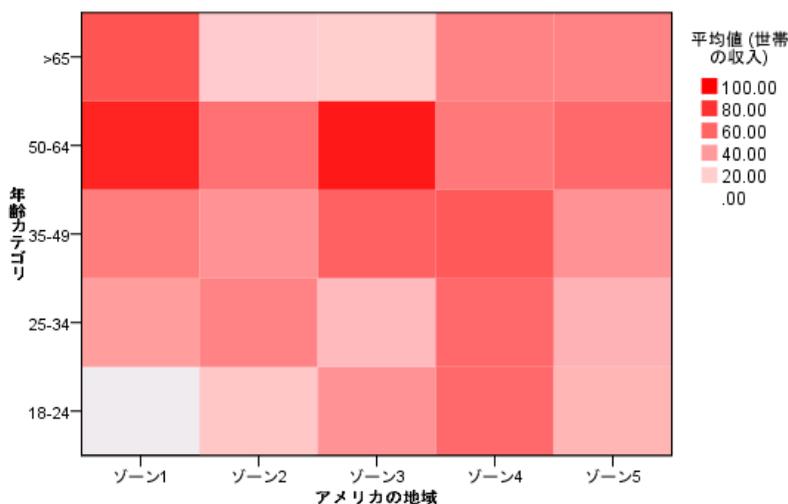
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Geographic indicator, Age category, and Household income in thousands, in that order. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select **Heat Map**.

図 5-23
Basic tab selections, heat map



- ▶ Click Run.
- ▶ On the resulting output window, click the “Display field and value labels” toolbar button (the right-hand one of the two in the center of the toolbar).

図 5-24
Categorical heat map



We can observe the following:

- A heat map is like a table that uses colors instead of numbers to represent the values for the cells. Bright, deep red indicates the highest value, while gray indicates a low value. The value of each cell is the mean of the continuous field/variable for each pair of categories.
- Except in Zone 2 and Zone 5, the group of customers whose age is between 50 and 64 have a greater mean household income than those in other groups.
- There are no customers between the ages of 25 and 34 in Zone 4.

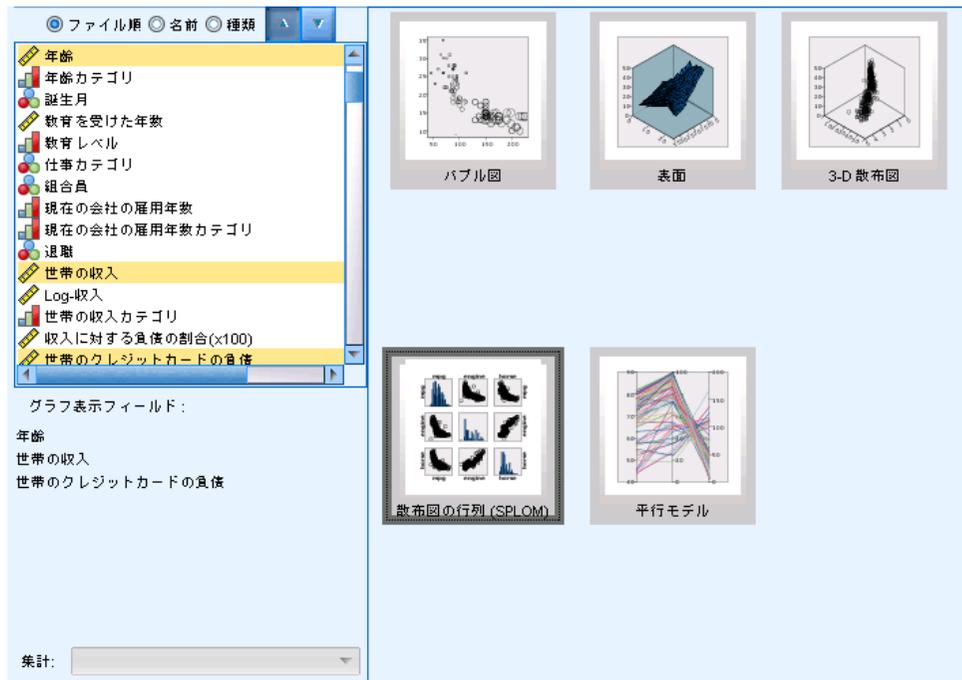
Example: Scatterplot Matrix (SPLOM)

We will create a scatterplot matrix of several different variables so that we can determine whether there are any relationships among the variables in the dataset.

Note: This example uses `customer_subset`.

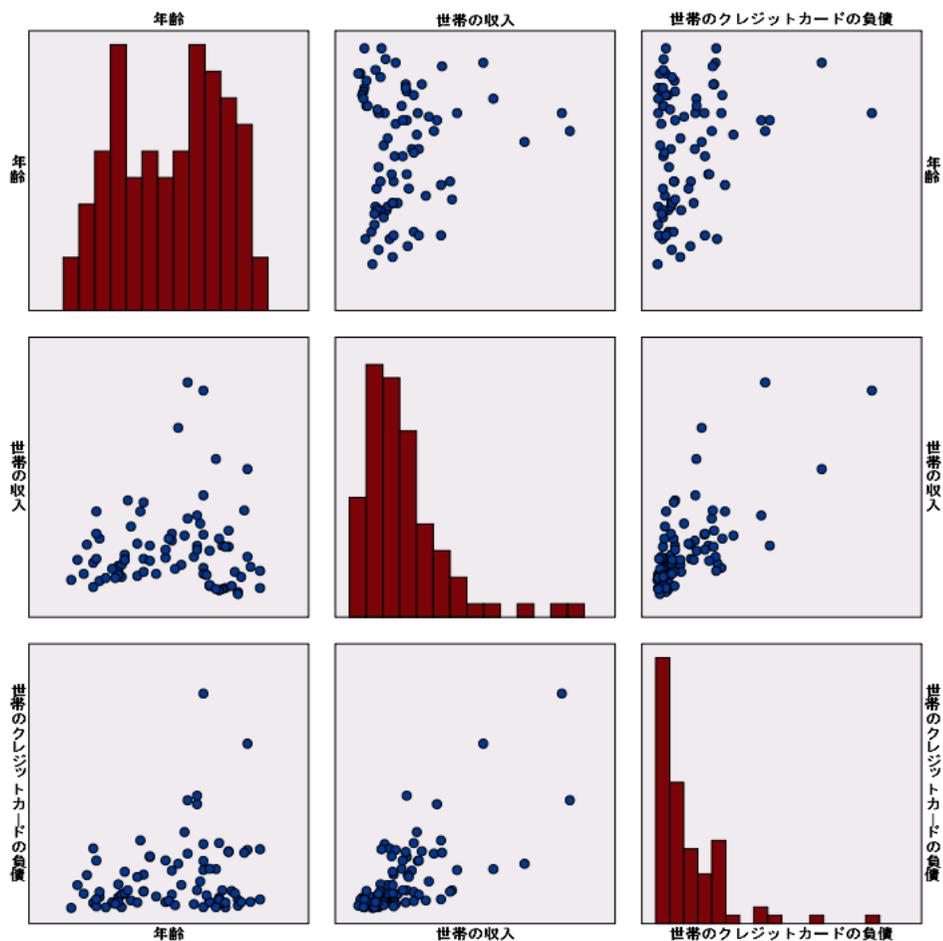
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Age in years, Household income in thousands, and Credit card debt in thousands. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select SPLOM.

図 5-25
Basic tab selections, SPLOM



- ▶ Click Run.
- ▶ Maximize the output window to view the matrix more clearly.

図 5-26
Scatterplot matrix (SPLOM)



We can observe the following:

- The histograms displayed on the diagonal show the distribution of each variable in the SPLOM. The histogram for age appears in the upper left cell, that for income in the center cell, and that for creddebt in the lower right cell. None of the variables appears normally distributed. That is, none of the histograms resembles a bell curve. Also, note that the histograms for income and creddebt are positively skewed.

- There does not seem to be any relationship between age and the other variables.
- There is a linear relationship between income and creddebt. That is, creddebt increases as income increases. You may want to create individual scatterplots of these variables and the other related variables to explore the relationships further.

Example: Choropleth (Color Map) of Sums

We will now create a map visualization. Then, in the subsequent example, we will create a variation of this visualization. The dataset is `worldsales`, which is a hypothetical data file that contains sales revenue by continent and product.

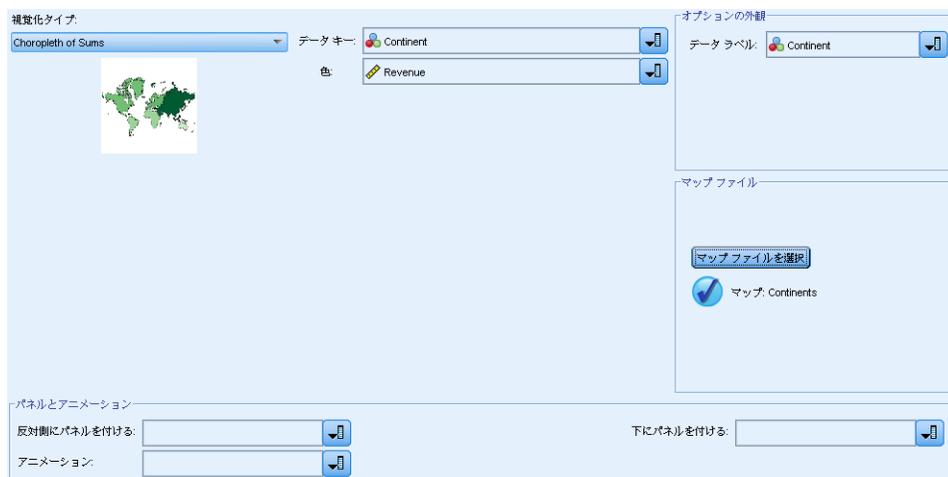
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Continent and Revenue. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select **Choropleth of Sums**.
- ▶ Click the Detailed tab.
- ▶ In the Optional Aesthetics group, choose Continent from the Data Label drop-down list.
- ▶ In the Map Files group, click **Select a Map File**.
- ▶ In the Select Maps dialog box, check that **Map** is set to Continents and **Map Key** is set to CONTINENT.
- ▶ In the Compare Map and Data Values groups, click **Compare** to ensure the map keys match the data keys. In this example, all of data key values have matching map keys and features. We can also see that there is no data for Oceania.

図 5-27
Select Maps dialog box



- ▶ In the Select Maps dialog box, click OK.

図 5-28
Basic tab selections, Choropleth of Sums



- ▶ Click Run.

図 5-29
Choropleth of Sums



From that map visualization, we can easily see that revenue is highest in North America and lowest in South America and Africa. Each continent is labeled because we used Continent for the data label aesthetic.

Example: Bar Charts on a Map

This example shows how revenue breaks down by product in each continent.

Note: This example uses `worldsales`.

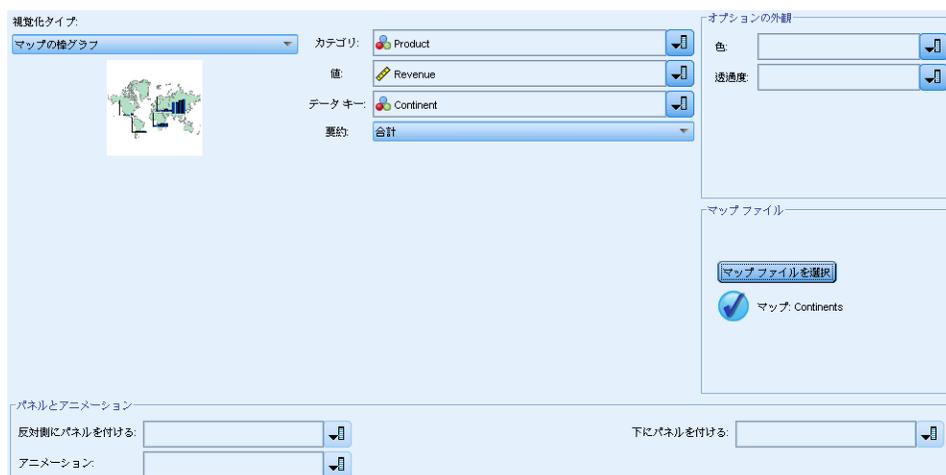
- ▶ Add a Graphboard node and open it for editing.
- ▶ On the Basic tab, select Continent, Product, and Revenue. (Use Ctrl+Click to select multiple fields/variables.)
- ▶ Select **Bars on a Map**.
- ▶ Click the Detailed tab.

When using more than one field of a specific type, it is important to check that each field is assigned to the correct slot.

- ▶ From the Categories drop-down list, choose Product.
- ▶ From the Values drop-down list, choose Revenue.
- ▶ From the Data Key drop-down list, choose Continent.
- ▶ From the Summary drop-down list, choose Sum.
- ▶ In the Map Files group, click **Select a Map File**.
- ▶ In the Select Maps dialog box, check that **Map** is set to Continents and **Map Key** is set to CONTINENT.
- ▶ In the Compare Map and Data Values groups, click **Compare** to ensure the map keys match the data keys. In this example, all of data key values have matching map keys and features. We can also see that there is no data for Oceania.
- ▶ On the Select Maps dialog box, click OK.

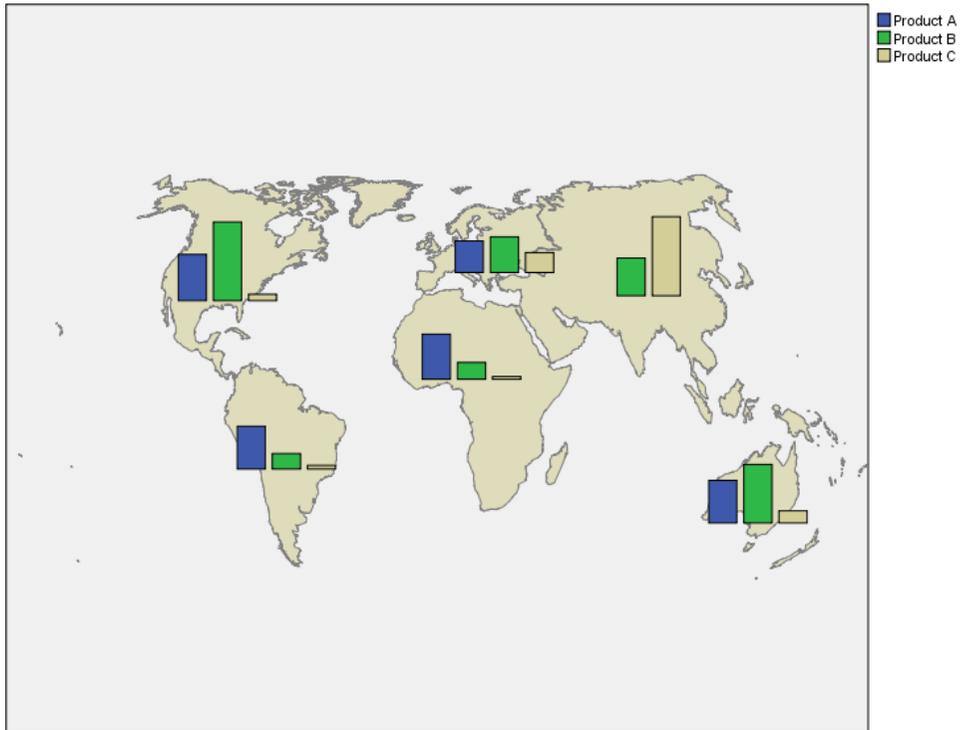
図 5-30

Basic tab selections, Bars on a Map



- ▶ Click **Run**.
- ▶ Maximize the resulting output window to see the display more clearly.

図 5-31
Bar Charts on a Map



We can observe the following:

- The distribution of total revenue across products is very similar in South America and Africa.
- Product C generates the least revenue everywhere except in Asia.
- There is no or minimal revenue from Product A in Asia.

グラフボードの [外観] タブ

外観のオプションはグラフ作成前に指定できます。

図 5-32
グラフボード ノードの [外観] タブの設定



一般的な外観オプション

タイトル: グラフのタイトルに使用するテキストを入力します。

サブタイトル: グラフのサブタイトルに使用するテキストを入力します。

解説: グラフの解説に使用するテキストを入力します。

サンプリング: 大きいデータ セットのための手法を指定します。最大データ セット サイズを使用するか、またはデフォルトのレコード数を使用することができます。[サンプル] を選択すると、大きいデータ セットに対するパフォーマンスが向上します。代わりに、[すべてのデータを使用] を選択して、すべてのデータ ポイントをプロットすることもできます。ただし、この場合ソフトウェアのパフォーマンスが大幅に低下する可能性があります。

スタイルシートの外観オプション

使用できる視覚化テンプレート（およびスタイルシート、およびマップ）を制御できる 2 つのボタンがあります。

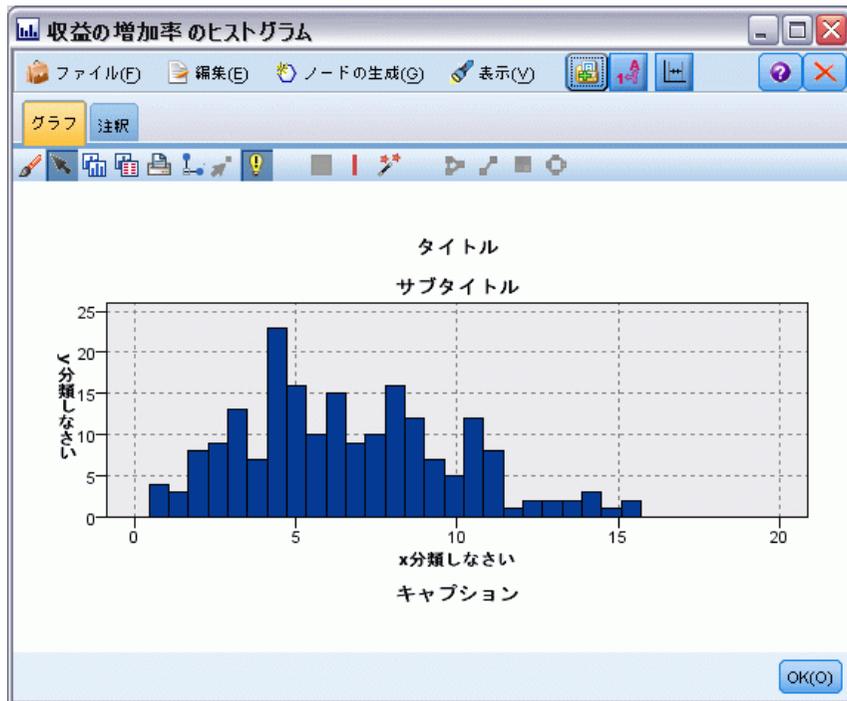
管理: 視覚化テンプレート、スタイルシート、およびマップをコンピュータで管理します。視覚化テンプレート、スタイルシート、およびマップをローカル マシンでインポート、エクスポート、名前変更、および削除でき

ます。詳細は、[p. 320 テンプレート、スタイルシート、マップ ファイルの管理](#) を参照してください。

位置：視覚化テンプレート、スタイルシート、およびマップを保存する場所を変更します。現在の場所は、ボタンの右側に表示されます。詳細は、[p. 318 テンプレート、スタイルシート、マップの位置の設定](#) を参照してください。

次の例では、表示オプションがグラフ内のどこに表示されるかを示します(注：すべてのグラフですべてのオプションが使用されるわけではありません)。

図 5-33
さまざまなグラフ表示オプションの場所



テンプレート、スタイルシート、マップの位置の設定

視覚化テンプレート、視覚化スタイルシート、マップ ファイルは、特定のローカル フォルダまたは IBM® SPSS® Collaboration and Deployment Services Repository に保存されています。テンプレート、スタイルシート、マップを選択する場合、この場所に組み込まれたものだけが表示されます。すべてのテンプレート、スタイルシート、マップ ファイルを 1 つの場所に保存することによって、IBM SPSS アプリケーションはそれらに容易にアクセスできるようになります。この場所に追加のテンプレート、スタイル

シート、マップ ファイルを追加する方法については、「[テンプレート、スタイルシート、マップ ファイルの管理](#)」（ p. 320 ）を参照してください。

テンプレート、スタイルシート、マップ ファイルの位置の設定するには

- ▶ テンプレート、スタイルシート、マップ ファイルのダイアログボックスで、[場所...] をクリックして [テンプレートとスタイルシート] ダイアログ ボックスが表示されます。
- ▶ テンプレート、スタイルシート、マップ ファイルのデフォルトの場所について、オプションを選択します。

ローカル コンピュータ :テンプレート、スタイルシート、マップ ファイルは、ローカル コンピュータの特定のローカル フォルダにあります。Windows XP の場合、このフォルダは C:\Documents and Settings\<user>\Application Data\SPSSInc\Graphboard となります。フォルダは変更できません。

IBM® SPSS® Collaboration and Deployment Services Repository: テンプレート、スタイルシート、マップ ファイルは、IBM SPSS Collaboration and Deployment Services Repository のユーザー指定のフォルダにあります。特定のフォルダを指定するには、[フォルダ] を選択します。詳細は、「[テンプレート、スタイルシート、マップ ファイルの場所として IBM SPSS Collaboration and Deployment Services Repository を使用](#)」（ p. 319 ）を参照してください。

- ▶ [OK] をクリックします。

テンプレート、スタイルシート、マップ ファイルの場所として IBM SPSS Collaboration and Deployment Services Repository を使用

視覚化テンプレートおよびスタイルシートを IBM® SPSS® Collaboration and Deployment Services Repository に保存できます。この場所は、IBM SPSS Collaboration and Deployment Services Repository の固有のフォルダです。これがデフォルトの場所として設定されている場合、この場所のテンプレート、スタイルシート、マップ ファイルを選択できます。

IBM SPSS Collaboration and Deployment Services Repository のフォルダをテンプレート、スタイルシート、マップ ファイルとして設定するには

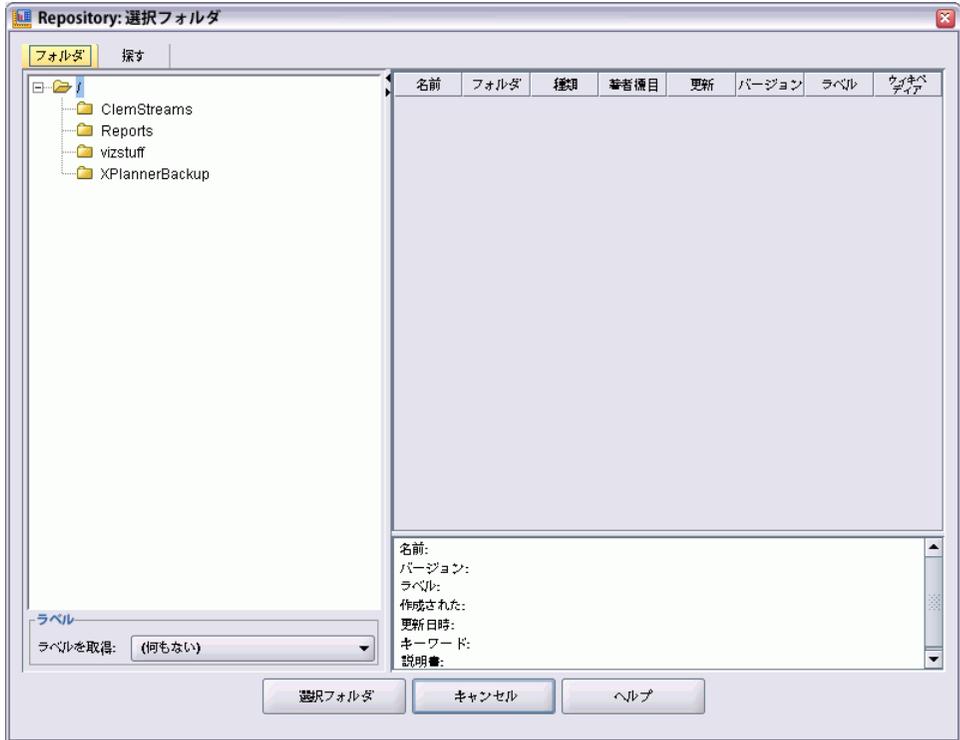
- ▶ [場所] ボタンのあるダイアログ ボックスで、[場所...] をクリックします。
- ▶ [IBM® SPSS® Collaboration and Deployment Services Repository] を選択します。
- ▶ [フォルダ] をクリックします。

注： IBM SPSS Collaboration and Deployment Services Repository に接続していない場合、接続情報を求めるメッセージが表示されます。

- ▶ [フォルダの選択] ダイアログ ボックスで、テンプレート、スタイルシート、マップ ファイルが保存されるフォルダを選択します。

図 5-34

[フォルダの選択] ダイアログ ボックス



- ▶ 必要に応じて、[ラベルの取得] からラベルを選択します。ラベルの付いたテンプレート、スタイルシート、マップ ファイルのみが表示されます。
- ▶ 特定のテンプレート、スタイルシート、マップ ファイルを含むフォルダを検索する場合、[検索] タブでテンプレート、スタイルシート、マップ ファイルを検索できます。[フォルダの選択] ダイアログ ボックスでは、検索されたテンプレート、スタイルシート、マップ ファイルが保存されているフォルダを自動的に選択します。
- ▶ [フォルダの選択] をクリックします。

テンプレート、スタイルシート、マップ ファイルの管理

[テンプレート、スタイルシート、マップ ファイルの管理] ダイアログ ボックスを使用して、コンピュータ上のローカルの場合でテンプレート、スタイルシート、マップ ファイルを管理できます。このダイアログ ボックスを使用して、コンピュータのローカルの場合で、視覚化テンプレ

ト、スタイルシート、マップ ファイルのインポート、エクスポート、名前変更、削除が可能です。

- ▶ テンプレート、スタイルシート、マップ ファイルを選択したいいずれかのダイアログ ボックスで、[管理...] をクリックします。

[テンプレート、スタイルシート、マップ ファイルの管理] ダイアログ ボックス

[テンプレート] タブには、ローカル テンプレートがすべて表示されています。[スタイルシート] タブには、サンプル データによる視覚化の例のほか、ローカル スタイルシートがすべて表示されています。スタイルシートのいずれかを選択して、視覚化の例にスタイルを適用することができます。詳細は、[p. 429 スタイルシートの適用](#) を参照してください。[マップ] タブには、ローカル マップ ファイルがすべて表示されています。このタブには、サンプル値、マップ作成時に指定した場合は子マント、マップのプレビューも表示されます。

次のボタンは、現在有効なタブで作動します。

インポート。 視覚化テンプレート、スタイルシート、マップ ファイルをファイル システムからインポートします。テンプレート、スタイルシート、マップ ファイルをインポートすると、IBM SPSS アプリケーションでそれらを使用できます。別のユーザーがテンプレート、スタイルシート、マップ ファイルを送付した場合、テンプレート、スタイルシート、マップ ファイルをインポートしてから、アプリケーションで使用します。

エクスポート。 視覚化テンプレート、スタイルシート、マップ ファイルをファイル システムからエクスポートします。別のユーザーに送信する場合、テンプレート、スタイルシート、マップ ファイルをエクスポートします。

名前の変更 : 選択した視覚化テンプレート、スタイルシート、マップ ファイルの名前を変更します。すでに使用されている名前に変更することはできません。

マップ キーのエクスポート: マップ キーをコンマ区切り (CSV) ファイルとしてエクスポートします。このボタンは [マップ] タブでのみ有効です。

削除。 選択した視覚化テンプレート、スタイルシート、マップ ファイルを削除します。複数のテンプレート、スタイルシート、マップ ファイルを選択するには、Ctrl + クリック で選択できます。削除するとやり直しができないため、慎重に操作してください。

マップ シェープファイルの変換と配布

Graphboard Template Chooser を使用して、視覚化テンプレートと SMZ ファイルの組み合わせからマップを作成できます。SMZ ファイルは、マップを描画するための地理的情報（国境など）を含む点で ESRI シェープファイル（SHP ファイル形式）と類似していますが、SMZ ファイルはマップの視覚化向けに最適化されています。Graphboard Template Chooser は、SMZ ファイルとともにあらかじめインストールされています。マップの視覚化に使用する既存の ERI シェープファイルがある場合、まずマップ変換ユーティリティを使用して、シェープファイルを SMZ ファイルに変換する必要があります。マップ変換ユーティリティは、1 つの層を含む点、ポリライン、またはポリゴン（シェープ タイプ 1、3 および 5）の ESRI シェープファイルをサポートしています。

ESRI シェープファイルを変換するほか、マップ変換ユーティリティを使用して、マップの詳細レベルを変更、フィーチャ ラベルを変更、フィーチャを結合、フィーチャを移動することができます。マップ変換ユーティリティを使用して、既存の SMZ ファイルを変更することもできます。

プリインストールされた SMZ ファイルの編集

- ▶ SMZ ファイルを管理システムからエクスポートします。詳細は、[p. 320 テンプレート、スタイルシート、マップ ファイルの管理](#) を参照してください。
- ▶ マップ変換ユーティリティを使用して、エクスポートされた SMZ ファイルを開き、編集します。ファイルを別名で保存することをお勧めします。詳細は、[p. 324 マップ変換ユーティリティの使用](#) を参照してください。
- ▶ SMZ ファイルを管理システムにインポートします。詳細は、[p. 320 テンプレート、スタイルシート、マップ ファイルの管理](#) を参照してください。

マップ ファイルのその他のリソース

マッピングのニーズをサポートする SHP ファイル形式の地理空間データは、多くの非公開ソースおよび公開ソースから利用できます。無料のデータを探す場合は、各国政府の Web サイトを確認してください。この製品のテンプレートの多くは、GeoCommons (<http://www.geocommons.com>) および U.S. Census Bureau (<http://www.census.gov>) から取得された公開して使用できるデータに基づいています。アメリカ合衆国、州、ローカルの地理空間データの細野化のソースは、米地質調査所 (<http://www.geodata.gov>) です。

重要:IBM 以外の製品に関する情報は、それらの製品の供給業者、公開済みの発表、または公開で使用できるソースから取得しています。IBM は、それらの製品のテストは行っておらず、IBM 以外の製品に関連する性能、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給業者に通知する必要があります。IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、決してそれらの Web サイトを推奨するものではありません。この IBM プログラムに付属する通知ファイルに記載のない限り、それらの Wrb サイトの資料はこの IBM プログラムの資料の一部ではなく、資料のサイトの使用は自身の責任によるものとします。

マップの主要コンセプト

シェープファイルに関連する主要なコンセプトについて理解すると、マップ変換ユーティリティを効率的に使用できるようになります。

シェープファイルは、マップを描画するための地理的情報を提供します。マップ変換ユーティリティがサポートしているシェープファイルは、次の 3 種類です。

- **ポイント:** 都市など、点の場所を指定します。
- **ポリライン:** 川など、パスや場所を指定します。
- **ポリゴン:** 国など、境界のある地域およびその場所を指定します。

一般的には、ポリゴン シェープファイルを使用します。コロプレス マップはポリゴン シェープファイルから作成します。コロプレス マップは、各ポリゴン（領域）内の値を示すために色を使用します。ポイント シェープファイルおよびポリライン シェープファイルは通常、ポリゴン シェープファイルに重ねられます。例で示すのは、米国の都市のポリゴン シェープファイルに重なられた米国の都市のポイント シェープファイルです。

シェープファイルは、**フィーチャ**で構成されています。フィーチャは、各地理的エンティティです。たとえば、フィーチャは、国、州、都市などとなります。シェープファイルには、フィーチャに関するデータも含まれます。これらのデータは、**属性**に含まれます。属性は、データ ファイルのフィールドまたは変数に類似しています。フィーチャの**マップ キー**である属性が少なくとも 1 つ必要です。マップ キーは、国名や州名などのラベルとなります。マップ キーは、マップの視覚化を作成するデータファイルの変数/フィールドにリンクします。

SMZ ファイルのキー属性のみを保持することができます。マップ変換ユーティリティは、追加属性の保存には対応していません。様々なレベルで集約する場合、複数の SMZ ファイルを作成する必要があります。たとえば、

米国の州と地域を集約する場合、別個の SMZ ファイルが必要です。州を指定するキーを持つファイルと、地域を指定するキーを持つファイルです。

マップ変換ユーティリティの使用

マップ変換ユーティリティの開始方法。

- ▶ メニューから次の項目を選択します。
ツール > マップ変換ユーティリティ

マップ変換ユーティリティには、4 つの主要な画面（ステップ）があります。1 つのステップには、マップ ファイルの編集に関する詳細なコントロールを示すサブステップが含まれています。

ステップ 1 - 保存先とソース ファイルの選択

まず、ソース マップ ファイルと、変換されるマップ ファイルの保存先を選択します。シェープファイルの .shp および .dbf ファイルが必要です。

変換する .shp (ESRI) ファイルまたは .smz ファイルを選択する。 コンピュータの既存のマップ ファイルを参照します。SMZ ファイルに変換し、SMZ ファイルとして保存するファイルです。シェープファイルの .dbf ファイルは、.shp ファイルに一致するベース ファイル名と同じ場所に保存する必要があります。.shp ファイルの属性情報が含まれるため、.dbf ファイルが必要です。

変換されるマップ ファイルの保存先およびファイル名を設定する。 元のマップ ソースから作成される SMZ ファイルのパスとファイル名を入力します。

ステップ 2 - マップ キーの選択

SMZ ファイルに含むマップ キーを選択します。マップの表示に影響を与えるいくつかのオプションを変更することができます。マップ変換ユーティリティのその後の手順には、マップのプレビューが含まれます。選択する表示オプションを使用して、マップ プレビューを生成します。

プライマリ マップ キーを選択する。 マップ内のフィーチャを指定し、ラベルを付けるプライマリ キーとなる属性を選択します。たとえば、世界地図のプライマリ キーは国名を指定する属性となります。プライマリ キーはデータをマップのフィーチャにリンクさせるため、選択する属性の値（ラベル）はデータの値と一致する必要があります。属性を選択すると、ラベルが表示されます。これらのラベルを変更する必要がある場合、後の手順で変更することができます。

追加のキーを選択する。 プライマリ マップ キーのほか、生成した SMZ ファイルに含める他のキー属性を確認します。たとえば、属性には翻訳されたラベルが含まれる場合があります。データを他の言語でコード化する

場合、これらの属性を保持する必要があります。プライマリ キーと同じフィーチャを示す追加キーのみを選択できます。たとえば、プライマリ キーが米国州のフルネームを示す場合、州名の略称など、米国の州を示す属性キーのみを選択できます。

自動的にマップを平滑化する。 ポリゴンを含むシェープファイルには通常、多くのデータ ポイントと多くの詳細データが含まれているため、統計マップの視覚化ができません。詳細情報が多すぎると、混乱し、パフォーマンスに悪影響がある場合があります。平滑化によって、詳細のレベルを軽減し、マップを一般化することができます。マップがきれいに表示され、表示速度が速くなります。マップを自動的に平滑化すると、最大角度は 15 度、維持するパーセンテージは 99% となります。これらの設定に関する詳細については、[\[マップの平滑化\]](#) (p.326) を参照してください。後で追加の平滑化を適用することができます。

同じフィーチャの接触しているポリゴン間の境界を削除する。 フィーチャには、重要なメイン フィーチャの内部に境界を持つサブフィーチャがある場合があります。たとえば、世界地図には、各大陸内に国の境界線があります。このオプションを選択すると、内部の境界線はマップ上に表示されません。世界地図の例では、このオプションを選択すると、国の境界線は削除されますが、大陸の境界はそのまま表示されます。

ステップ 3 - マップの編集

マップの基本オプションを指定したら、より詳細なマップを編集できるようになります。これらの変更はオプションです。マップ変換ユーティリティのこの手順では、関連するタスクについて説明し、変更を検証できるようなマップのプレビューを表示します。タスクは、シェープファイルの種類（ポイント、ポリライン、ポリゴン）または座標によって使用できない場合があります。

すべてのタスクには、マップ変換ユーティリティの左側に次のような共通コントロールがあります。

マップにラベルを表示する。 デフォルトでは、フィーチャ ラベルはプレビューに表示されません。ラベルを表示するよう選択することができます。ラベルはフィーチャを特定できますが、プレビュー マップの直接選択したものと対応する場合があります。たとえば、フィーチャ ラベルを編集している場合など、必要に応じてこのオプションをオンにします。

マップ プレビューの色を指定する。 デフォルトでは、プレビュー マップはエリアを一色で表示します。すべてのフィーチャが同じ色です。各マップ フィーチャにさまざまな色を指定することができます。このオプションによって、マップ内のフィーチャを区別することができます。フィーチャを結合し、プレビューで新しいフィーチャがどのように表示されるかを確認する場合に役立ちます。

すべてのタスクには、マップ変換ユーティリティの右側に次のような共通コントロールがあります。

元に戻す。 不要な変更を行った場合、[元に戻す] をクリックして前の状態に戻すことができます。最高 100 回変更を取り消すことができます。

マップの平滑化

ポリゴンを含むシェープファイルには通常、多くのデータポイントと多くの詳細データが含まれているため、統計マップの視覚化できません。詳細情報が多すぎると、混乱し、パフォーマンスに悪影響がある場合があります。平滑化によって、詳細のレベルを軽減し、マップを一般化することができます。マップがきれいに表示され、表示速度が速くなります。このオプションは、ポイント マップおよびポリライン マップに対しては使用できません。

最大角度: 1 ~ 20 の最大角度は、ほとんど線型であるポイントのセットを平滑化する許容度を指定します。値が大きいと線型平滑化の許容度が高くなり、点が削除され、より一般化されたマップが作成されます。線型平滑化を適用する場合、マップ変換ユーティリティはマップ内の 3 点によって形成された内角をチェックします。180 - 角度が指定された値を下回る場合、マップ変換ユーティリティは中間点を削除します。つまり、マップ変換ユーティリティは、3 点によって形成される線がほぼ直線かどうかを確認します。ほぼ直線である場合、マップ変換ユーティリティは直線として扱い、中間点を削除します。

維持パーセント: 90 ~ 100 度の維持パーセントは、マップが平滑化された場合に位置する土地エリアの面積を指定します。このオプションは、フィーチャに島が含まれる場合など、複数のポリゴンを含むフィーチャのみに影響を与えます。フィーチャの合計面積からポリゴンを差し引いた値が元の面積の指定された割合より低い場合、マップからポリゴンを削除します。フィーチャのすべてのポリゴンを削除することはありません。つまり、適用される平滑化の面積に関係なく、フィーチャには常に少なくとも 1 つのポリゴンがあります。

最大角度と維持パーセントを選択した後、[適用] をクリックします。プレビューは、平滑化の変更によって更新します。再度マップを平滑化する場合、生活化が必要なレベルに達するまで繰り返します。平滑化には制限があります。繰り返し平滑化を行うと、追加の平滑化がマップに適用できなくなります。

フィーチャ ラベルの編集

必要に応じてフィーチャ ラベルを編集し（予期されたデータに一致させる）、マップ内のラベルを再配置することもできます。ラベルを変更する必要がないと思う場合でも、マップから視覚化を作成する前にレビューする必要があります。デフォルトではプレビューにラベルが表示されないため、**[マップにラベルを表示]** を選択して表示する必要があります。

キー。 レビューまたは編集するフィーチャ ラベルを含むキーを選択します。

フィーチャ。 このリストには、選択したキーに含まれるフィーチャ ラベルが表示されます。ラベルを編集するには、リスト内のラベルをダブルクリックします。ラベルがマップ上に表示されている場合、マップ プレビューで直接フィーチャ ラベルをダブルクリックします。ラベルを実際のデータ ファイルと比較する場合、**[比較]** をクリックします。

X/Y。 このテキスト ボックスには、マップの選択されたフィーチャのラベルの現在の中心点が表示されます。マップの座標にユニットが表示されます。ローカルの、カルテシアン座標（国家平面座標系など）または地理的座標（ x が経度で y が緯度）の場合があります。ラベルの新しい位置を示す座標を入力します。ラベルが表示されている場合、マップのラベルをクリック アンド ドラッグして移動します。テキスト ボックスが新しい位置に更新されます。

比較。 データ ファイルに特定のキーのフィーチャ ラベルと一致させるデータ値を含む場合、**[比較]** をクリックして、**[外部データ ソースと比較]** ダイアログ ボックスを表示します。このダイアログ ボックスで、データ ファイルを開き、値をマップ キーのフィーチャ ラベルの値と比較することができます。

[外部データ ソースとの比較] ダイアログ ボックス

[外部データ ソースとの比較] ダイアログ ボックスを使用して、タブ区切り値ファイル（拡張子 .txt）またはカンマ区切り値ファイル（拡張子 .csv）を開くことができます。ファイルを開くと、データ ファイル内のフィールドを選択して、特定のマップ キーのフィーチャ ラベルと比較することができます。その後マップ ファイルの相違点を修正することができます。

データ ファイルのフィールド。 フィーチャ ラベルと比較する値のフィールドを選択します。.txt ファイルまたは .csv ファイルの最初の行にフィールドの説明的ラベルが含まれる場合、**[最初の行を列ラベルとして使用]** をオンにします。説明的ラベルが含まれない場合、各フィールドはデータ ファイルの場所によって識別されます（「列 1」「列 2」など）。

比較するキー。 データ ファイル フィールド値と比較するフィーチャ ラベルのマップ キーを選択します。

比較。 値を比較する場合にクリックします。

比較の結果。 デフォルトでは、[比較の結果] テーブルに、データ ファイルの一致しないフィールド値のみ表示されます。アプリケーションは、挿入されたスペースまたは欠落しているスペースをチェックして、関連するフィーチャ ラベルを見つけようとします。[マップ ラベル] 列のドロップダウン リストをクリックして、マップ ファイルのフィーチャ ラベルを表示されるフィールド値と一致させます。マップ ファイルに該当するフィーチャ ラベルがない場合、[不一致のままにする] を選択します。すでにフィーチャ ラベルに一致する値でもすべてのフィールド値を表示する場合、[不一致のケースのみを表示] をオフにします。1 つまたは複数の一致を上書きする場合にこの操作を行います。

フィールド値にフィーチャを一致させるには、各フィーチャを 1 回だけ使用できます。複数のフィーチャを 1 つのフィールド値に一致させる場合、フィーチャを結合して新しく結合されたフィーチャをフィールド値に一致させます。フィーチャ結合の詳細は、「[フィーチャの結合](#)」(p. 328) を参照してください。

フィーチャの結合

フィーチャの決堂は、マップ内に広い領域を作成する場合に役立ちます。たとえば、州の地図を変換する場合、州（この例ではフィーチャ）をより大きい北部、南部、東部、西部の地域に結合することができます。

キー。 結合するフィーチャを識別できるフィーチャ ラベルを含むマップ キーを選択します。

フィーチャ。 結合する最初のフィーチャをクリックします。結合する他のフィーチャを Ctrl を押しながらクリックします。フィーチャはマップ プレビューでも選択されます。リストからフィーチャを選択するほか、マップ プレビュー内で直接フィーチャをクリックおよび Ctrl キーを押しながらクリックすることができます。

結合するフィーチャを選択した後、[結合] をクリックして [名前と結合したフィーチャ] ダイアログ ボックスを表示し、新しいフィーチャにラベルを適用することができます。フィーチャを結合して結果が期待したものであることを確認した後、[マップ プレビューの色を指定する] をオンにする場合があります。

フィーチャを結合した後、新しいフィーチャのラベルを移動することもできます。[フィーチャ ラベルの編集] タスクで移動できます。 [詳細は、p. 327 フィーチャ ラベルの編集](#) を参照してください。

[名前と結合したフィーチャ] ダイアログ ボックス

[名前と結合したフィーチャ] ダイアログ ボックスを使用して、ラベルを新しく結合したフィーチャに割り当てることができます。

[ラベル] テーブルには、マップ ファイルの各キーの情報が表示され、各キーのラベルを割り当てることができます。

新規ラベル。 特定のマップ キーに割り当てられる結合されたフィーチャの新しいラベルを入力します。

キー。 新しいラベルを割り当てられるマップ キー。

古いラベル。 新しいフィーチャに結合されるフィーチャのラベル。

接触するポリゴン間の境界を削除。 結合したフィーチャから境界を削除します。たとえば、州を地域に結合した場合、このオプションを指定すると各州の周りの境界線を削除します。

フィーチャの移動

マップ内のフィーチャを移動できます。本土や島しょ部など、フィーチャを一緒に配置する場合に役立ちます。

キー。 移動するフィーチャを識別できるフィーチャ ラベルを含むマップ キーを選択します。

フィーチャ。 結合するフィーチャをクリックします。フィーチャはマップ プレビューで選択されます。マップ プレビューで直接フィーチャをクリックできます。

X/Y。 このテキスト ボックスには、マップの選択されたフィーチャの中心点が表示されます。マップの座標にユニットが表示されます。ローカルの、カルテシアン座標（国家平面座標系など）または地理的座標（ x が経度で y が緯度）の場合があります。フィーチャの新しい位置を示す座標を入力します。マップ上のフィーチャをクリック アンド ドラッグして移動することができます。テキスト ボックスが新しい位置に更新されます。

フィーチャの削除

マップ内の不要なフィーチャを削除できます。マップの視覚化に不要なフィーチャを削除したい場合に役立ちます。

キー。 削除するフィーチャを識別できるフィーチャ ラベルを含むマップ キーを選択します。

フィーチャ。 削除するフィーチャをクリックします。複数のフィーチャを同時に削除する場合、Ctrl を押しながらフィーチャをクリックします。フィーチャはマップ プレビューでも選択されます。リストからフィーチャ

を選択するほか、マップ プレビュー内で直接フィーチャをクリックおよび Ctrl キーを押しながらクリックすることができます。

各要素を削除

フィーチャ全体を削除するほか、湖や小さい島など、フィーチャを構成する各要素を削除することができます。このオプションは、ポイントマップに対しては使用できません。

要素。 削除する要素をクリックします。複数の要素を同時に削除する場合、Ctrl を押しながら要素をクリックします。要素はマップ プレビューでも選択されます。リストからフィーチャを選択するほか、マップ プレビュー内で直接要素をクリックおよび Ctrl キーを押しながらクリックすることができます。要素名のリストが説明的でない（各要素にフィーチャ内の番号が割り当てられている）ため、マップ プレビューの選択をチェックして、該当する要素が選択されていることを確認します。

投影法の設定

マップの投影法は、三次元の地球を二次元で表示する方法を指定します。すべての投影法で歪みが発生します。ただし、世界地図を表示するか、ローカルな地図を表示するかによって適切な投影法があります。元のフィーチャの形を維持する投影法もあります。形を維持する投影法は正角投影法です。このオプションは、地理的座標（経度と緯度）を示すマップにのみ使用できます。

マップ変換ユーティリティのオプションと異なり、投影法はマップ視覚化が作成された後に変更できます。

投影法。 マップ投影法を選択します。グローバルまたは半球のマップを作成する場合、ローカル、メルカトル、ヴィンケル図法を使用できます。小さいエリアの場合、ローカル、ランベルト正角円錐図法、または横メルカトル図法を使用します。すべての投影法は、測地系の WGS83 楕円体を使用しています。

- **ローカル**投影法は、マップが国家平面座標系などのローカル座標系で作成された場合に使用されます。これらの座標系は、地理的座標（経度と緯度）ではなくカルテシアン座標によって定義されます。ローカル投影法では、水平線および垂直線はカルテシアン座標系で等間隔になっています。ローカル投影法は正角ではありません。
- **メルカトル図法**は世界地図向けの正角投影法です。水平線および垂直線が直線で、お互いに直角に交差しています。メルカトル図法は、北極および南極に近づくほど拡大されるため、北極または南極を含む地図には使用できません。北極および南極に近付くと、歪みが最大になります。

- **ヴィンケル図法**は世界地図向けの非正角投影法です。正角ではありませんが、形とサイズのバランスが良い図法です。赤道とグリニッジ子午線を除いて、すべての線がカーブしています。世界地図に北極または南極が含まれている場合、この図法が適しています。
- 名前が示す通り、**ランベルト正角円錐図法**は正角図法で、南北に対して東西が長い大陸または小さい土地の集合の地図に使用します。
- **横メルカトル図法**は、大陸または小さい取りの集合体の地図に向けた正角図法です。この投影法は、東西に比べて南北が長い土地に使用します。

ステップ 4 - 終了

この時点で、マップ ファイルを説明するコメントを追加し、マップ キーからサンプル データ ファイルを作成することもできます。

マップ キー。 マップ ファイルに複数のキーがある場合、プレビューに表示するフィーチャ ラベルのマップ キーを選択します。マップからデータ ファイルを作成する場合、これらのラベルはデータ値に使用されます。

コメント。 マップを説明し、元のシェープファイルのソースなど、ユーザーに関連する追加情報を提供するコメントを入力します。コメントは Graphboard Template Chooser の管理システムに表示されます。

フィーチャ ラベルからデータ セットを作成。 表示されたフィーチャ ラベルからテキスト データ ファイルを作成します。[参照...] をクリックすると、場所をファイル名を指定することができます。.txt 拡張子を追加すると、ファイルはタブ区切りファイルとして保存されます。.csv 拡張子を追加すると、ファイルはコンマ区切りファイルとして保存されます。拡張子がない場合は、CSV がデフォルトとなります。

マップ ファイルの配布

マップ変換ユーティリティの最初のステップで、変換した SMZ ファイルを保存する場所を選択しています。Graphboard Template Chooser の管理システムにマップを追加する場合があります。管理システムに保存すると、同じコンピュータで実行する IBM SPSS 製品でマップを使用することができますようになります。

マップを他のユーザーに配布するには、SMZ を送信する必要があります。管理システムを使用してマップをインポートすることができます。ステップ 1 で指定した場所のファイルを送信することができます。管理システムにあるファイルを送信する場合、次の手順でまずエクスポートする必要があります。

- ▶ テンプレート ピッカーで [管理...] をクリックします。

- ▶ [マップ] タブをクリックします。
- ▶ 配布するマップを選択します。
- ▶ [エクスポート...] をクリックして、ファイルを保存する場所を選択します。

マップ ファイルを他のユーザーに送信することができるようになりました。ユーザーはこのプロセスを遡って、管理システムにマップをインポートする必要があります。

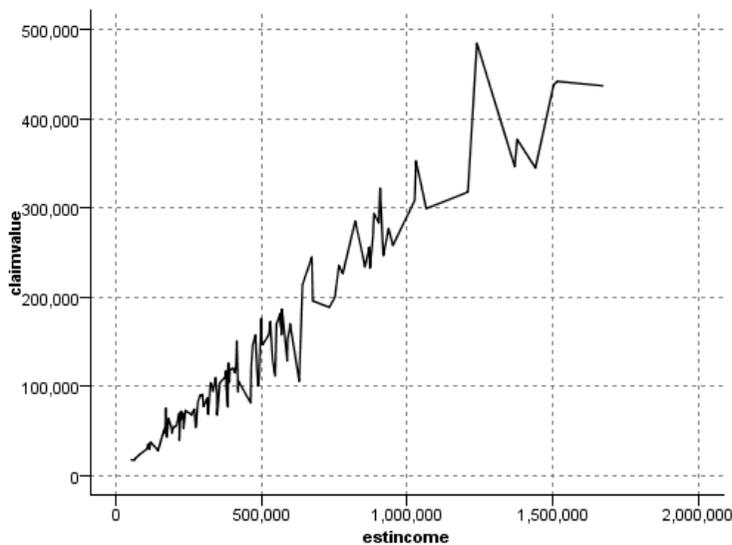
散布図ノード

散布図ノードでは、数値フィールド間の相関が示されます。散布図と呼ばれる点を使用したプロットを作成したり、折れ線を使用することができます。ダイアログ ボックスで [X モード] を指定すると、3 種類の折れ線を作成することができます。

X モード = ソート

[X モード] を [ソート] に設定すると、x 軸にプロットするフィールドの値に基づいてデータがソートされます。この場合、グラフ上で左から右へ進む 1 つの線が生成されます。名義型をオーバーレイとして使用すると、グラフの左から右へと進む異なる色の複数の線が生成されます。

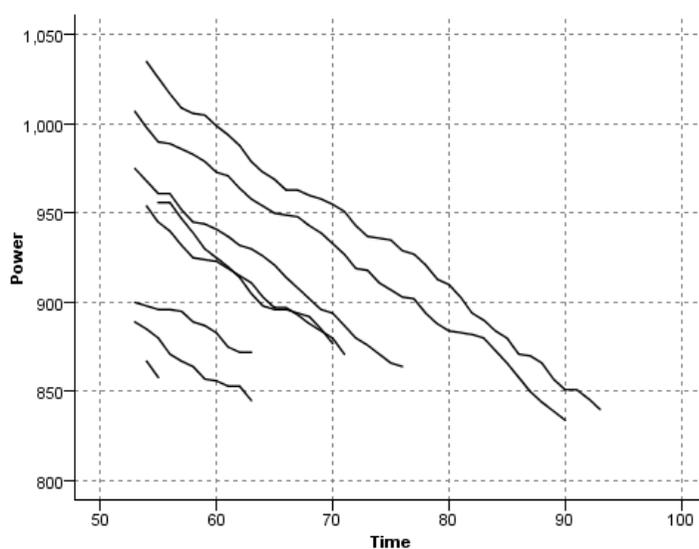
図 5-35
[X モード] を [ソート] に設定した折れ線



X モード = オーバーレイ

[X モード] を [オーバーレイ] に設定すると、同じグラフ上で複数の折れ線が作成されます。オーバーレイ プロットでデータをソートすることはできません。x 軸の値が増え続ける限り、データは 1 つの線にプロットされますが、値が減少すると新しい線が作成されます。たとえば、x 値が 0 から 100 に変化する場合、y 値は 1 つの線にプロットされます。しかし、x 値が 100 を下回ると、最初の線のほかに新しい線がプロットされます。このため、最終的にはグラフに多数のプロットが描画されることもあります。これは、連続する複数の y 値を比較する場合に便利です。このタイプのプロットは、連続する 24 時間単位の電力需要など、定期的な時間コンポーネントを持つデータに適しています。

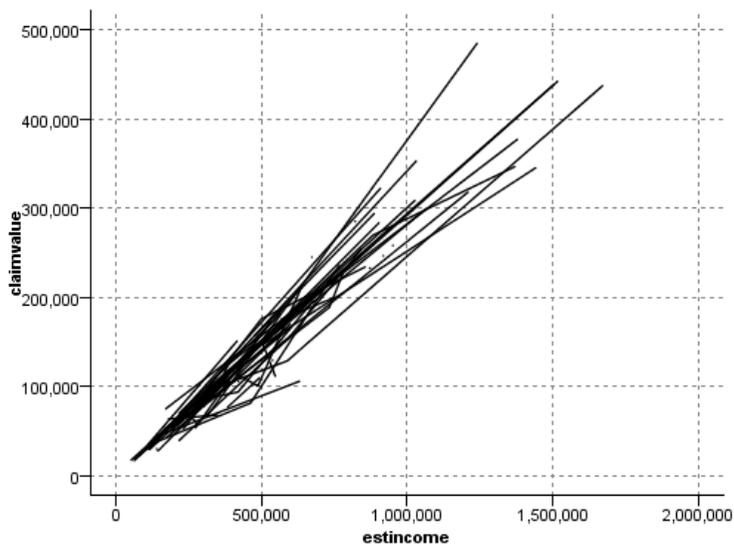
図 5-36
[X モード] を [オーバーレイ] に設定した折れ線



X モード = 読み込み順

[X モード] を [読み込み順] に設定すると、x 値と y 値はデータ ソースから読み込まれた順にプロットされます。このオプションは、傾向、つまりデータの順序に依存するパターンを調べるときに使用する、時系列コンポーネントを持つデータに適しています。このタイプの散布図を作成する前に、データをソートする必要がある場合があります。また、[X モード] を [ソート] と [読み込み順] に設定した場合の 2 つの類似プロットを比較して、パターンがどの程度ソートに依存しているかを調べるときにも役立ちます。

図 5-37
最初 [X モード] を [ソート] で、次に [読み込み順] で再実行した折れ線



グラフボード ノードを使用して散布図や折れ線を生成することもできます。ただし、このノードでは、選択肢のオプション数が多くなります。詳細は、[p. 287](#) 組み込まれている利用可能なグラフボード視覚化タイプを参照してください。

散布図ノードのタブ

散布図は、X フィールドの値に対する Y フィールドの値を表しています。多くの場合、これらのフィールドはそれぞれ従属変数と独立変数に対応しています。

図 5-38
散布図ノードの [プロット] タブの設定



Xフィールド: リストから、横の x 軸に表示するフィールドを選択します。

Yフィールド: リストから、縦の y 軸に表示するフィールドを選択します。

Zフィールド: [3次元グラフ] ボタンをクリックすると、z 軸に表示するフィールドをリストから選択できます。

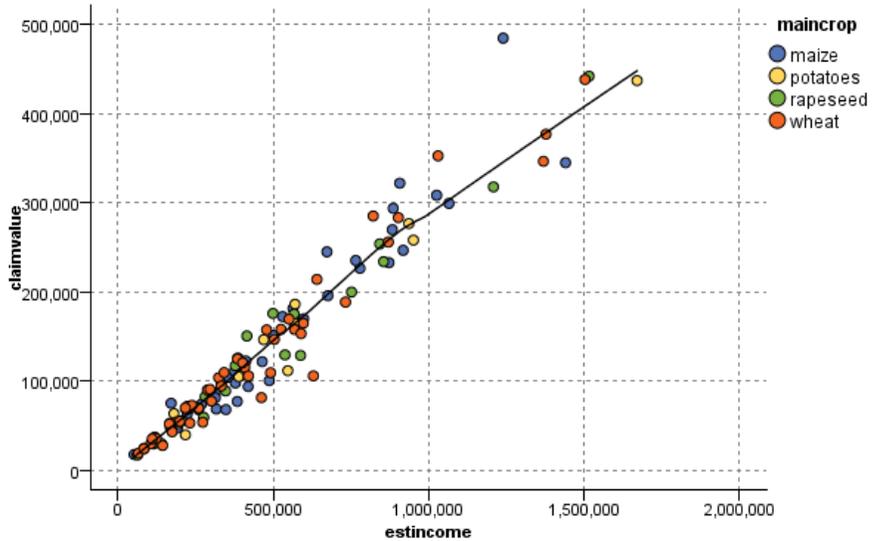
オーバーレイ: データ値のカテゴリを描くにはさまざまな方法があります。たとえば、申請者による主作物の伸びによる「推定所得」と「申請値」の値を示すために、主作物を色のオーバーレイとして使用することができます。詳細は、p. 271 外観、オーバーレイ、パネル、およびアニメーションを参照してください。

オーバーレイタイプ: オーバーレイ関数が表示されるか、平滑化が表示されるかを指定します。平滑化とオーバーレイ関数は、常に y の関数として計算されます。

- **なし**: オーバーレイを表示しません。
- **平滑化**: 局所的に重みを付けたインタラクティブな強力な最小 2 乗法 (LOESS) を使用して計算された平滑化フィットラインを表示します。この方法は散布図内の狭い領域に焦点をあてて、一連の回帰を効果的に

計算します。これで、滑らかな曲線を作成するために後に結合される、一連の「局所的な」回帰線が作成されます。

図 5-39
LOESS 平準化オーバーレイによるプロット



- **関数**: 実際の値と比較する既知の関数を指定する場合に選択します。たとえば、実際の値と予測値を比較する場合は、 $y = x$ という関数をオーバーレイとしてプロットできます。[y] テキスト ボックスで関数を指定してください。デフォルトの関数は $y = x$ ですが、 x の代わりに、2 次関数や任意の式などのあらゆる種類の関数を指定することもできます。

注： パネルまたはアニメーション グラフでオーバーレイ関数を使用することはできません。

プロットのオプションを設定したら、[実行] をクリックして、ダイアログボックスから直接プロットを実行できます。[オプション] タブを使って、区分け、X モード、およびスタイルなどを指定することもできます。

散布図の [オプション] タブ

図 5-40
散布図ノードの [オプション] タブの設定



スタイル: プロットのスタイルとして [ポイント] または [線] のどちらかを選択します。[線] を選択すると、[X モード] コントロールが有効になります。[ポイント] を選択すると、プラス記号 (+) をデフォルトのポイント形状として使用します。いったんグラフを作成したら、ポイントの形状およびサイズを変更することができます。

X モード: 折れ線グラフの場合は、[X モード] フィールドを選択して、折れ線のスタイルを定義する必要があります。[ソート]、[オーバーレイ]、または [読み込み順] を選択します。[オーバーレイ] または [読み込み順] を選択した場合、最初の n レコードのサンプリングに使用する最大データセット サイズを指定する必要があります。それ以外の場合は、デフォルトの 2,000 レコードが用いられます。

自動 X 範囲: この軸に沿ったデータ中の値の範囲全体を使用します。指定した [最小] および [最大] に基づいて値の一部を明示的に使用する場合は、選択を解除してください。この範囲は、値を入力するか矢印を使用して指定します。デフォルトでは、グラフの構築を高速化するために、自動範囲のオプションが選択されています。

自動 Y 範囲: この軸に沿ったデータ中の値の範囲全体を使用します。指定した [最小] および [最大] に基づいて値の一部を明示的に使用する場合は、選択を解除してください。この範囲は、値を入力するか矢印を使用して指定します。デフォルトでは、グラフの構築を高速化するために、自動範囲のオプションが選択されています。

自動 Z 範囲: [プロット] タブで 3 次元グラフが指定されている場合のみ。この軸に沿ったデータ中の値の範囲全体を使用します。指定した [最小] および [最大] に基づいて値の一部を明示的に使用する場合は、選択を解除してください。この範囲は、値を入力するか矢印を使用して指定します。デフォルトでは、グラフの構築を高速化するために、自動範囲のオプションが選択されています。

ジッタ: 拡散としても知られているジッタリングは、多くの値が繰り返されるデータ セットのポイント プロットに役立ちます。値の分布を明確にするため、ジッタを使用して実際の値の周囲に無作為 (ランダム) にポイントを分散できます。

前のバージョンの IBM® SPSS® Modeler ユーザーに対する注意 : 散布図に使用するジッタ値は、本リリースの SPSS Modeler では異なるメトリックを使用しています。前のバージョンでは、実際の数字が値になりましたが、今回はフレーム サイズの比率に変更されています。つまり、古いストリームで使われている拡散値は大きすぎる可能性があります。このリリースでは、ゼロ以外の拡散値は 0.2 に変換されます。

プロットするレコードの最大数: 大きいデータ セットのプロット手法を指定します。最大データ セット サイズを使用するか、またはデフォルト値 (2,000 レコード) を使用することができます。[ピン] または [サンプル] を選択すると、大きいデータ セットに対するパフォーマンスが向上します。代わりに、[すべてのデータを使用] を選択して、すべてのデータ ポイントをプロットすることもできます。ただし、この場合ソフトウェアのパフォーマンスが大幅に低下する可能性があります。

注 : [X モード] を [オーバーレイ] または [読み込み順] に設定した場合、これらのオプションは無効になり、最初の n レコードだけが使用されます。

■ **ピン:** データ セットに格納されているレコード数が、指定した数より大きい場合に、分割を有効にします。分割を行うと、グラフが細かいグリッドに分割されてから、プロットや各グリッド セルに現れるポイント数のカウントが実際に行われます。最終的なグラフでは、ビン重心 (ビン中のすべてのポイントの位置の平均) でセルごとに 1 つのポイントがプロットされます。プロットされたシンボルの大きさは、その領域中にあるポイント数を示しています (サイズをオーバーレイとして使用しない場合)。重心とサイズでポイント数を表すことにより、密集領域への過度のプロット (画一的な色の集合) やシンボルの羅列 (人工的な重心パターン) を避けることができます。そのため、分割されたプロットは大きいデータ セットを表すための最適な方法となっています。このようなシンボルの羅列は、特定のシンボル (特にプラ

ス記号 [+]) が、生データ中に存在しない密集領域を生成するような競合がある場合に発生します。

- **サンプル** :ここに入力した数のレコードまで、無作為にデータのサンプリングを行います。デフォルトは 2,000 です。

散布図の [外観] タブ

外観のオプションはグラフ作成前に指定できます。

図 5-41
散布図ノードの [外観] タブの設定

The image shows a software dialog box titled "散布図" (Scatter Plot). It has a tabbed interface with five tabs: "プロット" (Plot), "オプション" (Options), "外観" (Appearance), "出力" (Output), and "注釈" (Annotations). The "外観" tab is currently selected and highlighted in yellow. The dialog contains several input fields and radio buttons for configuring the appearance of the scatter plot. At the bottom, there are five buttons: "OK(O)", "実行" (Execute), "キャンセル(C)", "適用(A)" (Apply), and "リセット(R)" (Reset).

散布図

プロット オプション **外観** 出力 注釈

タイトル:

サブタイトル:

解説:

Xラベル: 自動 ユーザー設定

Yラベル: 自動 ユーザー設定

Zラベル: 自動 ユーザー設定

グリッドの表示

OK(O) 実行 キャンセル(C) 適用(A) リセット(R)

タイトル : グラフのタイトルに使用するテキストを入力します。

サブタイトル : グラフのサブタイトルに使用するテキストを入力します。

解説 : グラフの解説に使用するテキストを入力します。

Xラベル : 自動的に生成された x 座標 (水平) ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

Yラベル : 自動的に生成された y 座標 (垂直) ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

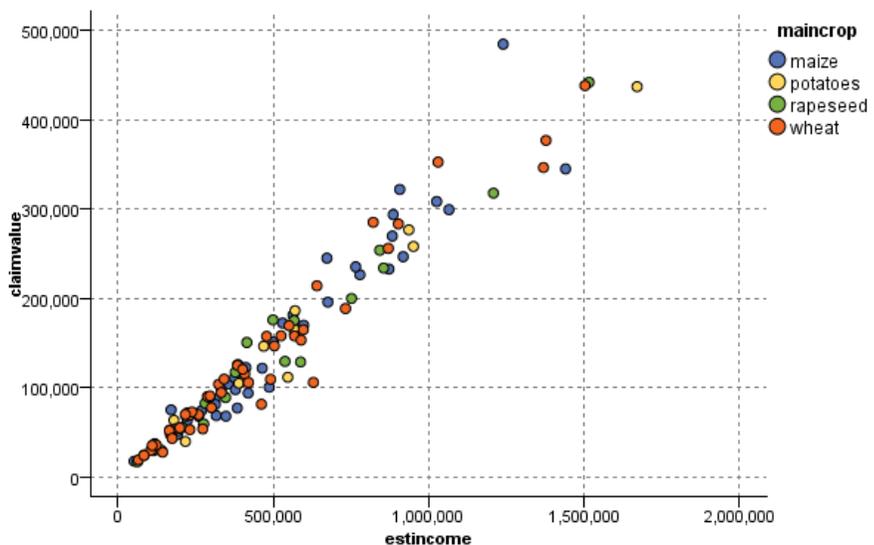
Z ラベル: 3 次元のグラフのみで利用可能で、自動的に生成された z 座標ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

グリッドの表示: デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

散布図グラフの使用方法

散布図および線グラフは、基本的に Y に対する X のプロットです。たとえば、農業助成金申請における不法行為を調べる場合 (IBM® SPSS® Modeler インストール ディレクトリの Demos フォルダにある fraud.str を参照)、申請者の申告している所得とニューラル ネットワークによる推定所得をプロットすることができます。主要作物の種類などをオーバーレイすることにより、申請内容 (値または数字) と作物の種類の間にある関係があるかどうかを描き出すことができます。

図 5-42
推定所得と申請値間の関係を、主要作物の種類をオーバーレイとして描画



散布図グラフ、線グラフ、および評価グラフは、X に対する Y の関係を 2 次元で表すため、領域を定義、要素をマーク、またはバンドを描画することによってグラフを簡単に操作することができます。これらの領域、バン

ド、または要素で表示されるデータのノードを生成することもできます。
詳細は、 [p. 391 グラフの検証](#) を参照してください。

棒グラフ ノード

棒グラフまたはテーブルは、ローンの種類や性別など、データ セット内のシンボル値（非数値）の出現頻度を示します。一般に棒グラフ ノードは、モデルの作成前にバランス ノードを使って修正できる、データの不均衡を表す場合に使用されます。バランス ノードは、棒グラフ ウィンドウの [ノードの生成] メニューを使って自動的に生成することができます。

グラフボード ノードを使用して棒グラフを作成することもできます。ただし、このノードでは、選択肢のオプション数が多くなります。 [詳細は、 p. 287 組み込まれている利用可能なグラフボード視覚化タイプ](#) を参照してください。

注： 数値の出現頻度を表すには、ヒストグラム ノードを使います。

棒グラフの [プロット] タブ

図 5-43
棒グラフ ノードの [プロット] タブの設定



作図 :棒グラフの種類を選択します。選択したフィールドの棒グラフを表示する場合は、[**選択したフィールド**]を選択します。データセットのフラグ型フィールドの真 (true) の値の棒グラフを表示するには、[**すべてのフラグ (真の値)**]を選択します。

フィールド :値の棒グラフを表示する名義型またはフラグ型フィールドを選択します。数値として明示的に設定されていないフィールドだけがリストに表示されます。

オーバーレイ :色のオーバーレイとして使用する名義型またはフラグ型フィールドを選択し、指定したフィールドの各値内の値の棒グラフを表示します。たとえば、マーケティング キャンペーンの回答者 (pep) を子供の数 (children) のオーバーレイとして使用し、家族サイズの回答を描画することができます。詳細は、[p. 271 外観、オーバーレイ、パネル、およびアニメーション](#)を参照してください。

色で正規化 :グラフの幅全体を占めるようにすべてのバーを表示する場合に選択します。オーバーレイした値は各バーの比率と等しくなるため、カテゴリを簡単に比較できます。

ソート :棒グラフに値を表示する方法を選択します。アルファベット順に並べる場合は [**アルファベット順**]を選択します。また、出現頻度の降順に並べる場合は [**出現頻度順**]を選択します。

プロポーションナル スケール :最大カウントを持つ値がプロットの幅全体を占めるように値の棒グラフを表示する場合に選択します。他のすべての棒はこの値を基準に表示されます。このオプションを選択しないと、バーは各値の合計カウントに従って表示されます。

棒グラフの [外観] タブ

外観のオプションはグラフ作成前に指定できます。

図 5-44
[外観] タブの設定



タイトル： グラフのタイトルに使用するテキストを入力します。

サブタイトル： グラフのサブタイトルに使用するテキストを入力します。

解説： グラフの解説に使用するテキストを入力します。

Xラベル： 自動的に生成された x 座標（水平）ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

Yラベル： 自動的に生成された y 座標（垂直）ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

グリッドの表示： デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

棒グラフ ノードの使用方法

棒グラフ ノードは、データ セット中のシンボル値の分布を表すために用いられます。棒グラフは、操作ノードの前段階で、データの調査と不均衡の修正を行うためにしばしば利用されます。たとえば、子供がいない回答

者のインスタンスが、他の種類の回答者よりも頻繁に発生しているような場合、後のデータマイニング操作でより有益なルールを作成するために、これらのインスタンスを減らすことができます。このような不均衡を調査、修正するために、棒グラフノードが役立ちます。

棒グラフノードは、データを分析するグラフやテーブルを両方作成するという点で通常と異なります。

図 5-45
マーケティングキャンペーンの回答者で子供がいる人、いない人の数を表す棒グラフ

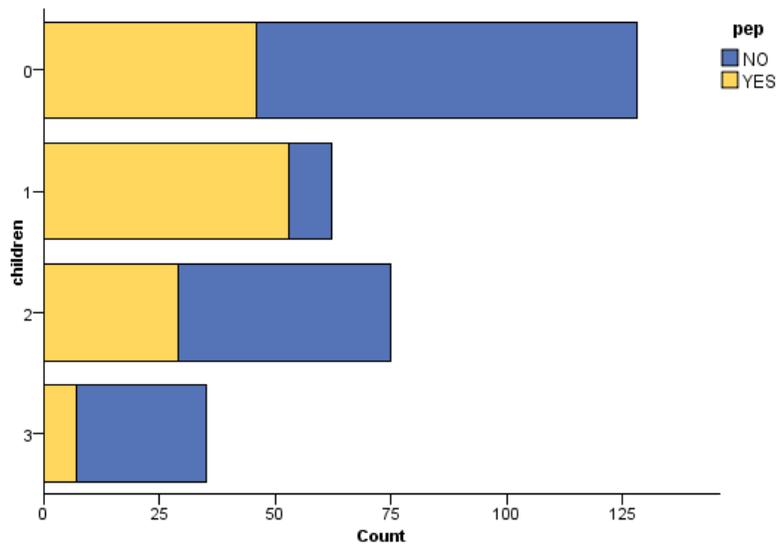


図 5-46
マーケティング キャンペーン の回答者 で子供が いる人、いない 人の割合 を表す 棒グラフ
テーブル



棒グラフ テーブルや棒グラフを作成して結果を調査したら、メニュー オプションを使って値のグループ化、値のコピー、およびノードの生成などを行って、データを準備することができます。さらに、グラフやテーブルの情報を、MS Word や MS PowerPoint など他のアプリケーションで使用するためにコピーまたはエクスポートすることができます。詳細は、p. 431 [グラフの印刷、保存、エクスポート](#) を参照してください。

棒グラフ テーブルから値を選択してコピーするには

- ▶ 値のセットを選択するには、行の上でマウス ボタンを押したままドラッグします。[編集] メニューから [すべて選択] を選択して、すべての値を選択することもできます。
- ▶ [編集] メニューから [テーブルをコピー] または [テーブルをコピー (フィールド名を含む)] を選択します。
- ▶ クリップボードまたは目的のアプリケーションに貼り付けます。

注： バーが直接コピーされることはありません。代わりにテーブルの値がコピーされます。つまり、オーバーレイされた値はコピー先のテーブルに表示されません。

棒グラフ テーブルから値をグループ化するには

- ▶ Ctrl キーを押しながら、グループ化する値を選択します。
- ▶ [編集] メニューから、[グループ化] を選択します。

注： 値をグループ化またはグループ化を解除する場合、[グラフ] タブのグラフは、自動的に再描画されて変更を表示します。

次の作業を行うこともできます。

- 棒グラフリストのグループ名を選択し、[編集] メニューから [グループ解除] を選択して値のグループ化を解除する。
- 棒グラフリストのグループ名を選択し、[編集] メニューから [グループの編集] を選択してグループを編集する。このオプションを選択すると、値をグループに追加、削除できるダイアログ ボックスが表示されます。

図 5-47

[グループの編集] ダイアログ ボックス



[ノードの生成] メニューのオプション

[ノードの生成] メニューのオプションを使って、データのサブセットの選択、フラグ型フィールドの作成、値の再グループ化、値の再分類、グラフまたはテーブルのデータの平均化などの作業を行うことができます。これらの操作により、データの準備ノードが生成され、それがストリーム領域に配置されます。生成されたノードを使用するには、これを既存のストリームと接続します。詳細は、[p. 402 グラフからのノードの生成](#) を参照してください。

ヒストグラム ノード

ヒストグラム ノードでは、数値フィールドの値の出現頻度が示されます。多くの場合、ヒストグラム ノードは、操作やモデルの構築前にデータを調べる場合に使用されます。棒グラフ ノードと同様、ヒストグラム ノードはデータの不均衡を調べる場合にもよく使用されます。グラフボード ノードを使用してヒストグラムを生成することもできますが、このノードで

は、選択肢のオプションが多くなります。詳細は、[p. 287 組み込まれている利用可能なグラフボード視覚化タイプ](#)を参照してください。

注：シンボル値フィールドの値の出現頻度を表すには、棒グラフ ノードを使用, μ てください。

ヒストグラムの [プロット] タブ

図 5-48
ヒストグラム ノードの [プロット] タブの設定

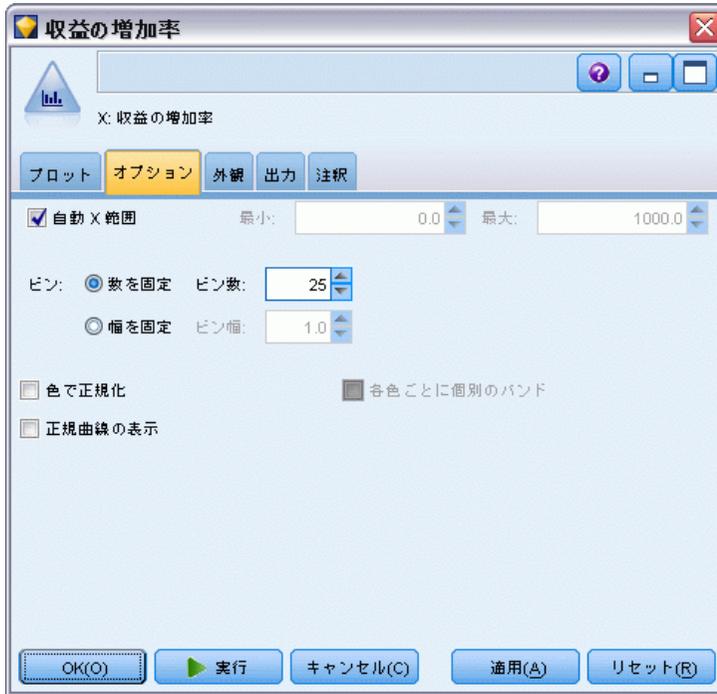


フィールド: 値の棒グラフを表示させる数値フィールドを選択します。シンボル値（カテゴリ）として明示的に定義されていないフィールドだけがリストに表示されます。

オーバーレイ: 指定したフィールドの値のカテゴリを示すシンボル値フィールドを選択します。[オーバーレイ フィールド] でフィールドを選択すると、ヒストグラムは選択したフィールドの各カテゴリを色で表す積み重ねグラフに変換されます。ヒストグラム ノードを使用すると、色、パネル、アニメーションの 3 つのオーバーレイがあります。詳細は、[p. 271 外観、オーバーレイ、パネル、およびアニメーション](#)を参照してください。

ヒストグラムの [オプション] タブ

図 5-49
ヒストグラム ノードの [オプション] タブの設定



自動 X 範囲：この軸に沿ったデータ中の値の範囲全体を使用します。指定した [最小] および [最大] に基づいて値の一部を明示的に使用する場合は、選択を解除してください。この範囲は、値を入力するか矢印を使用して指定します。デフォルトでは、グラフの構築を高速化するために、自動範囲のオプションが選択されています。

ビン：[数を固定] または [幅を固定] のいずれかを選択します。

- 表示するには、[数を固定] を選択します。バーの幅は、指定する範囲とビン数によって決まります。[ビン数] オプションで、グラフで使用するビン数を指示します。矢印を使用して、数を調節してください。
- また、固定幅のバーを持つグラフを作成するには、[幅を固定] を選択します。ビンの数は、指定した幅と値の範囲によって決まります。[ビンの幅] オプションでバーの幅を指示します。

色で正規化：すべてのバーを同じ高さに揃え、オーバーレイした値を各バーの全ケースに対する割合（パーセント）として表示する場合に選択します。

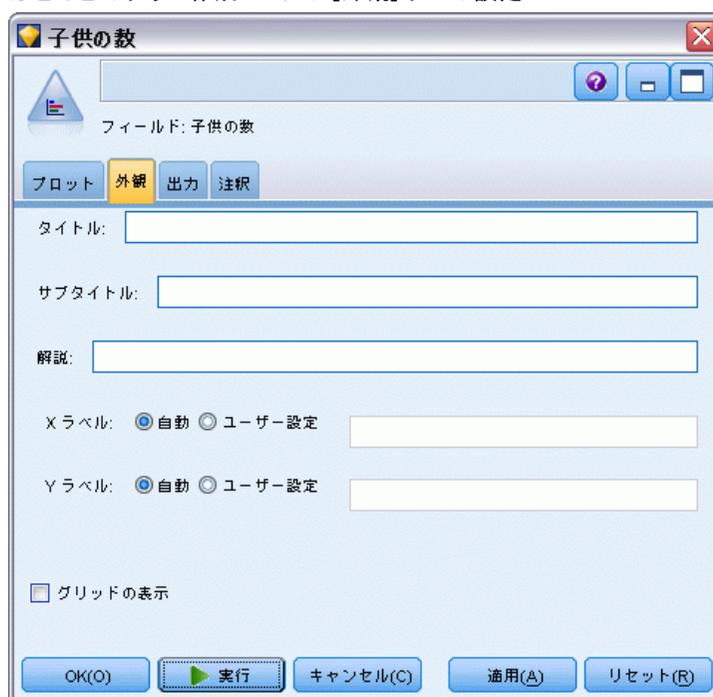
正規曲線の表示：データの平均や変数を表示するグラフに、正規曲線を追加します。

各色ごとに個別のバンド:オーバーレイした各値をグラフ上に個別のバンドとして表示する場合に選択します。

ヒストグラムの [外観] タブ

外観のオプションはグラフ作成前に指定できます。

図 5-50
ほとんどのグラフ作成ノードの [外観] タブの設定



タイトル: グラフのタイトルに使用するテキストを入力します。

サブタイトル: グラフのサブタイトルに使用するテキストを入力します。

解説: グラフの解説に使用するテキストを入力します。

Xラベル: 自動的に生成された x 座標 (水平) ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

Yラベル: 自動的に生成された y 座標 (垂直) ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

グリッドの表示: デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

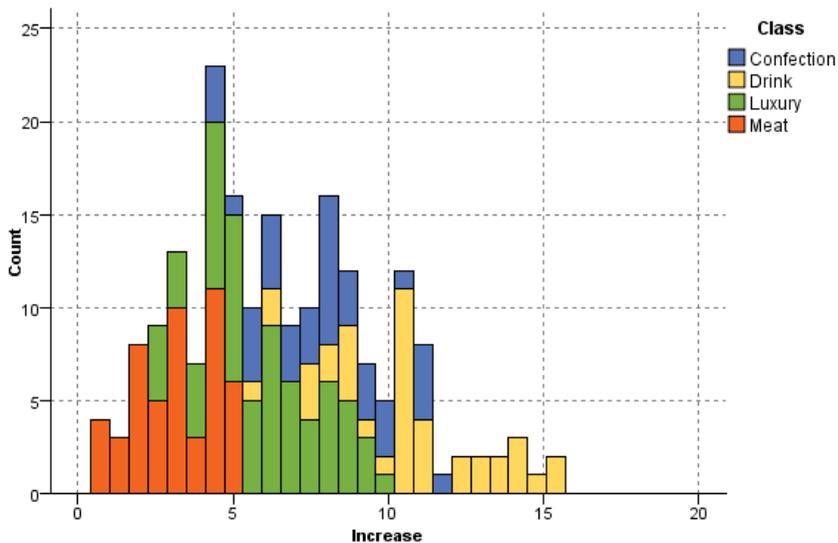
ヒストグラムの使用方法

ヒストグラムでは、x 軸の範囲である数値フィールドの値の棒グラフが示されます。ヒストグラムは、集計棒グラフと同様に動作します。集計棒グラフは、単一のフィールドの値の頻度ではなく、ある数値フィールドの値の「他のフィールドの値に対する」分布を表しています。

グラフを作成したら、結果を調べてバンドを定義し、x 軸に沿って値を分割したり、領域を定義したりできます。また、グラフ内で要素をマークすることもできます。詳細は、[p. 391 グラフの検証](#) を参照してください。

[ノードの生成] メニューのオプションを使用すると、グラフ内、具体的にはバンド、領域、マークされた要素内のデータを使用して、バランスノード、条件抽出ノード、フィールド作成ノードを作成することができます。この種のグラフは、操作ノードの前段階において、ストリームで使用するグラフからバランスノードを生成して、データの調査と不均衡の修正を行うために頻繁に使用されます。また、フィールド作成ノード（フラグ型）を生成して各レコードがどのバンドに該当するかを表すフィールドを追加したり、条件抽出ノードを生成して特定のセットまたは値の範囲内のすべてのレコードを選択することもできます。特定のデータのサブセットをさらに詳細に調査するような場合に、このような操作が役立ちます。詳細は、[p. 402 グラフからのノードの生成](#) を参照してください。

図 5-51
プロモーションによる購入の増加の分布をカテゴリ別に表したヒストグラム



集計棒グラフ ノード

集計棒グラフはヒストグラムと似ていますが、1つのフィールドの値の出現頻度ではなく、別のフィールドの値と連関がある1つの数値フィールドの値の棒グラフが示される点が異なります。集計棒グラフは、値が時間の経過とともに変化する変数やフィールドを表示する場合に役立ちます。3次元グラフを使って、分布をカテゴリ別に表示するシンボル値軸を追加することもできます。2次元の集計グラフが積み上げ棒グラフで、使用されている場合はオーバーレイで表示されます。詳細は、[p. 271 外観、オーバーレイ、パネル、およびアニメーション](#)を参照してください。

集計棒グラフの [プロット] タブ

図 5-52
集計棒グラフ ノードの [プロット] タブの設定



集計棒グラフ :表示するフィールドを選択します。このフィールドの値は、[対象フィールド] で指定するフィールドの値の範囲にわたって収集されます。シンボル値として定義されていないフィールドだけがリストに表示されます。

対象フィールド [初期データの収集] で指定するフィールドの表示に使用する値のフィールドを選択します。

フィールド :3次元グラフを作成する際に有効になります。カテゴリ別の集計フィールドを表示するために用いられる名義型またはフラグ型フィールドを選択します。

演算 :集計棒グラフの各バーによって表される対象を選択します。オプションには、[合計]、[平均]、[最大]、[最小]、および[標準偏差]が含まれています。

オーバーレイ :選択したフィールドの値のカテゴリを示すシンボル値フィールドを選択します。オーバーレイフィールドを選択すると、集計棒グラフが変換され、複数のバーがカテゴリ別に色分けされて表示されます。このノードには、色、パネル、アニメーションの3つのオーバーレイがあります。詳細は、[p.271 外観、オーバーレイ、パネル、およびアニメーション](#)を参照してください。

集計棒グラフの [オプション] タブ

図 5-53
集計棒グラフ ノードの [オプション] タブの設定



自動 X 範囲 : この軸に沿ったデータ中の値の範囲全体を使用します。指定した [最小] および [最大] に基づいて値の一部を明示的に使用する場合は、選択を解除してください。この範囲は、値を入力するか矢印を使用して指

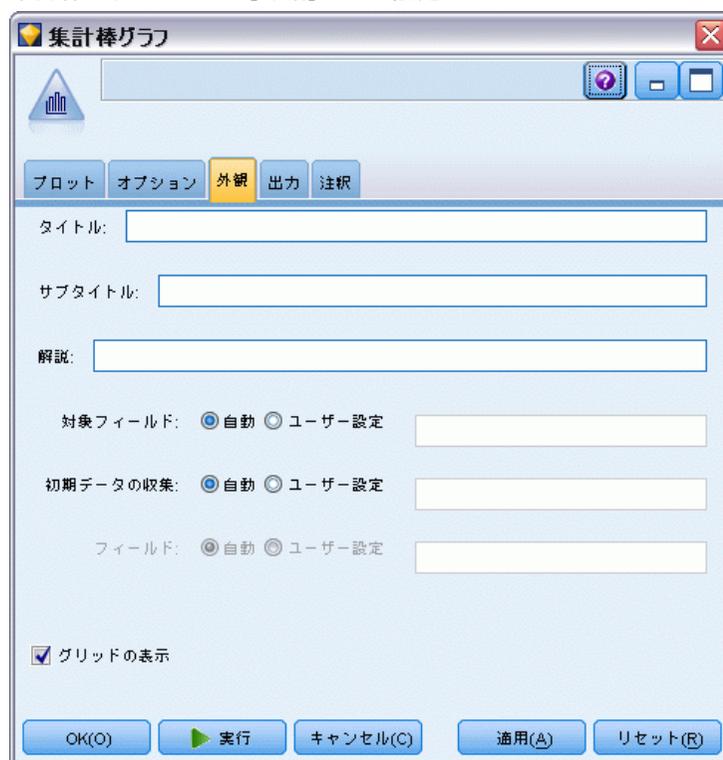
定します。デフォルトでは、グラフの構築を高速化するために、自動範囲のオプションが選択されています。

ビン : [数を固定] または [幅を固定] のいずれかを選択します。

- 表示するには、[数を固定] を選択します。バーの幅は、指定する範囲とビン数によって決まります。[ビン数] オプションで、グラフで使用するビン数を指示します。矢印を使用して、数を調節してください。
- また、固定幅のバーを持つグラフを作成するには、[幅を固定] を選択します。ビンの数は、指定した幅と値の範囲によって決まります。[ビンの幅] オプションでバーの幅を指示します。

集計棒グラフの [外観] タブ

図 5-54
集計棒グラフ ノードの [外観] タブの設定



外観のオプションはグラフ作成前に指定できます。

タイトル： グラフのタイトルに使用するテキストを入力します。

サブタイトル： グラフのサブタイトルに使用するテキストを入力します。

解説： グラフの解説に使用するテキストを入力します。

対象フィールドラベル :自動生成されたラベルを承認するか、[ユーザー設定]を選択してラベルを指定します。

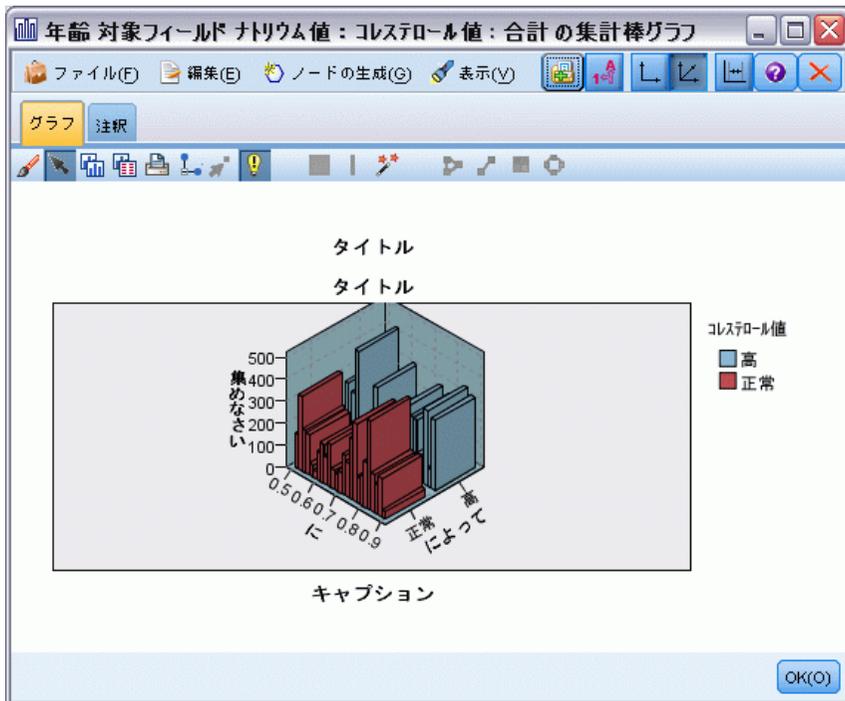
初期データの収集ラベル :自動生成されたラベルを承認するか、[ユーザー設定]を選択してラベルを指定します。

フィールドラベル :自動生成されたラベルを承認するか、[ユーザー設定]を選択してラベルを指定します。

グリッドの表示 : デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

次の例では、表示オプションが 3-D バージョンのグラフ内のどこに表示されるかを示します

図 5-55
3-D 集計棒グラフのグラフ外観オプションの位置



集計棒グラフの使用法

集計棒グラフは、単一のフィールドの値の頻度ではなく、ある数値フィールドの値の「他のフィールドの値に対する」分布を表しています。ヒストグラムは、集計棒グラフと同様に動作します。ヒストグラムでは、x 軸の範囲である数値フィールドの値の棒グラフが示されます。

グラフを作成したら、結果を調べてバンドを定義し、x 軸に沿って値を分割したり、領域を定義したりできます。また、グラフ内で要素をマークすることもできます。詳細は、[p. 391 グラフの検証](#)を参照してください。

[ノードの生成]メニューのオプションを使用すると、グラフ内、具体的にはバンド、領域、マークされた要素内のデータを使用して、バランスノード、条件抽出ノード、フィールド作成ノードを作成することができます。この種のグラフは、操作ノードの前段階において、ストリームで使用するグラフからバランスノードを生成して、データの調査と不均衡の修正を行うために頻繁に使用されます。また、フィールド作成ノード（フラグ型）を生成して各レコードがどのバンドに該当するかを表すフィールドを追加したり、条件抽出ノードを生成して特定のセットまたは値の範囲内のすべてのレコードを選択することもできます。特定のデータのサブセットをさらに詳細に調査するような場合に、このような操作が役立ちます。詳細は、[p. 402 グラフからのノードの生成](#)を参照してください。

図 5-56

ナトリウム値/カリウム値の合計と年齢、およびコレステロール値の上限と正常値を表す 3 次元集計棒グラフ

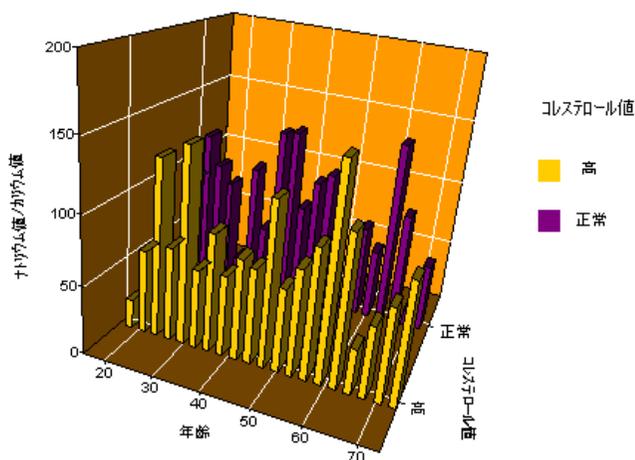
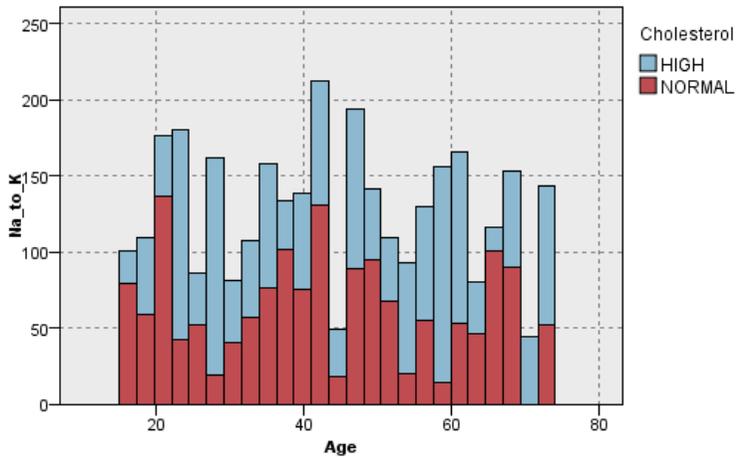


図 5-57
z 軸は表示せずにコレステロールを色のオーバーレイとして使用した集計棒グラフ



線グラフ ノード

線グラフは、1 つの X フィールドに対して複数の Y フィールドを表示する特殊な種類のプロットです。Y フィールドは色線でプロットされ、それぞれ [スタイル] フィールドを [ライン] に、[X モード] フィールドを [ソート] に設定した散布図ノードと等しくなります。時系列データがあり、時間の経過に伴うさまざまな変数の変動を調査するような場合に、線グラフが役立ちます。

線グラフの [プロット] タブ

図 5-58
線グラフ ノードの [プロット] タブの設定



X フィールド : リストから、横の x 軸に表示するフィールドを選択します。

Y フィールド : X フィールドの値の範囲にわたって表示させる、1 つ以上のフィールドをリストから選択します。複数のフィールドを選択するには、フィールド ピッカー ボタンを使用してください。リストからフィールドを除去する場合は、削除ボタンをクリックします。

オーバーレイ : データ値のカテゴリを描くにはさまざまな方法があります。たとえば、アニメーション オーバーレイを使って、データ中の各値を示す複数の散布図を表示することができます。これは、カテゴリが 10 個以上ある場合に役立ちます。カテゴリ数が 15 を超えると、パフォーマンスが低下する可能性があります。詳細は、[p.271 外観、オーバーレイ、パネル、およびアニメーション](#) を参照してください。

正規化 : すべての Y 値をグラフの 0~1 の範囲で増減させる場合に選択します。正規化を使用することで、各系列の値の範囲内での差異が原因で不明確になりかねない各線の間関係が明らかになります。また、同じグラフ上に複数の線をプロットする場合や、隣り合ったパネル内でプロットを比較する場合に、正規化をお勧めします。(正規化は、すべてのデータ値が似たような範囲内に収まる場合は不要です。)

図 5-59

時間の経過による発電装置の変動を示す標準の線グラフ（正規化を行わないと、圧力のプロットは参照できないことに注意してください）

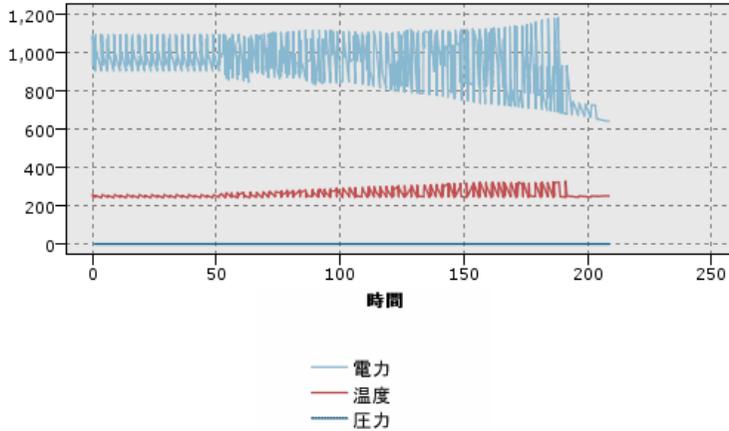
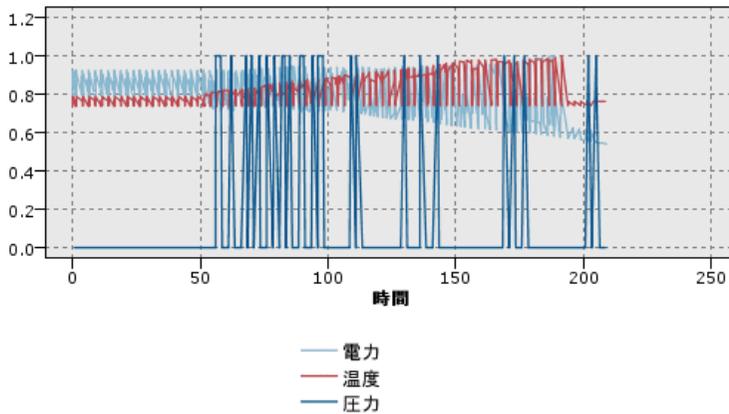


図 5-60

圧力のプロットを表示する正規化された線グラフ



オーバーレイ機能 : 実際の値と比較する既知の関数を指定する場合に選択します。たとえば、実際の値と予測値を比較する場合は、 $y = x$ という関数をオーバーレイとしてプロットできます。[y] テキスト ボックスで関数を指定してください。デフォルトの関数は $y = x$ ですが、 x の代わりに、2 次関数や任意の式などのあらゆる種類の関数を指定することもできます。

注： パネルまたはアニメーション グラフでオーバーレイ関数を使用することはできません。

レコード数が次の値より大きい場合 :大きいデータ セットのプロット手法を指定します。最大データ セット サイズを使用するか、またはデフォルト値 (2,000 ポイント) を使用することができます。[ビン] または [サンプル] を選択すると、大きいデータ セットに対するパフォーマンスが向上します。代わりに、[すべてのデータを使用] を選択して、すべてのデータ ポイントをプロットすることもできます。ただし、この場合ソフトウェアのパフォーマンスが大幅に低下する可能性があります。

注 : [X モード] を [オーバーレイ] または [読み込み順] に設定した場合、これらのオプションは無効になり、最初の n レコードだけが使用されます。

- **ビン** :データ セットに格納されているレコード数が、指定した数より大きい場合に、分割を有効にします。分割を行うと、グラフが細かいグリッドに分割されてから、プロットや各グリッド セルに現れる接続数のカウントが実際に行われます。最終的なグラフでは、ビン重心 (ビン中のすべての接続の位置の平均) でセルごとに 1 つの接続がプロットされます。
- **サンプル** :ここに指定した数のレコードまで、無作為にデータのサンプリングを行います。

線グラフの [外観] タブ

外観のオプションはグラフ作成前に指定できます。

図 5-61
ほとんどのグラフ作成ノードの [外観] タブの設定



タイトル： グラフのタイトルに使用するテキストを入力します。

サブタイトル： グラフのサブタイトルに使用するテキストを入力します。

解説： グラフの解説に使用するテキストを入力します。

Xラベル： 自動的に生成された x 座標（水平）ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

Yラベル： 自動的に生成された y 座標（垂直）ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

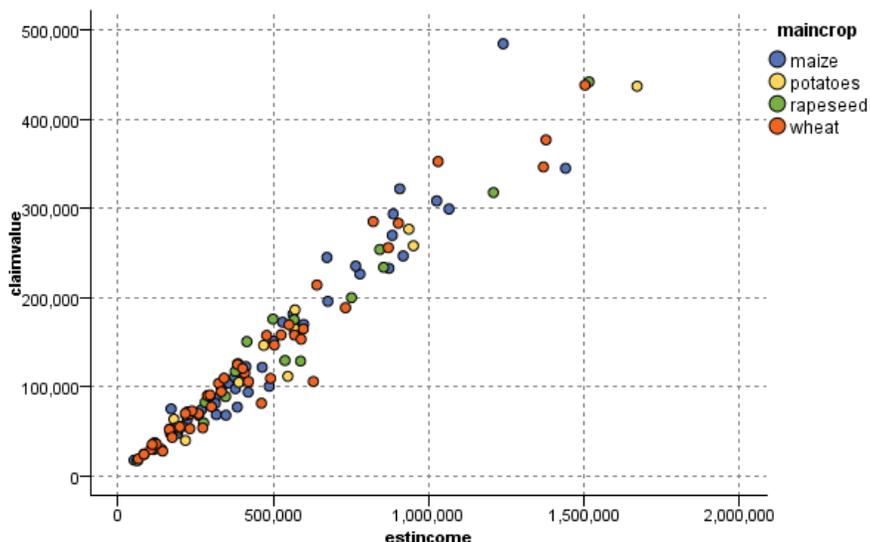
グリッドの表示： デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

線グラフの使用方法

散布図および線グラフは、基本的に Y に対する X のプロットです。たとえば、農業助成金申請における不法行為を調べる場合（IBM® SPSS® Modeler インストール ディレクトリの Demos フォルダにある fraud.str を参照）、

申請者の申告している所得とニューラル ネットワークによる推定所得をプロットすることができます。主要作物の種類などをオーバーレイすることにより、申請内容（値または数字）と作物の種類の間があるかどうかを描き出すことができます。

図 5-62
推定所得と申請値間の関係を、主要作物の種類をオーバーレイとして描画

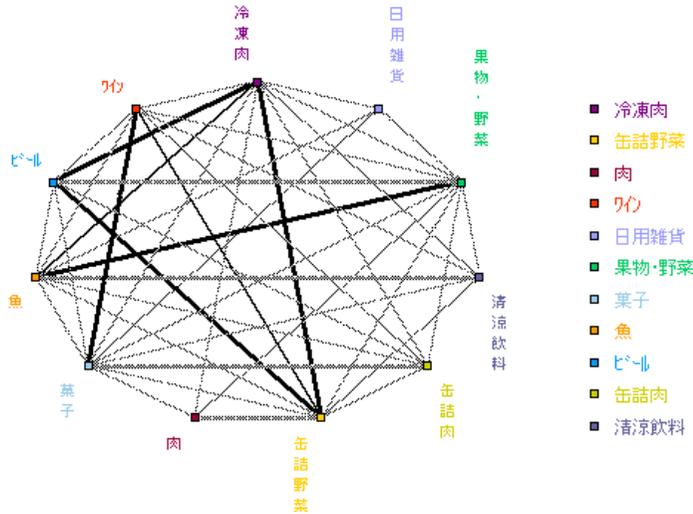


散布図グラフ、線グラフ、および評価グラフは、X に対する Y の関係を 2 次元で表すため、領域を定義、要素をマーク、またはバンドを描画することによってグラフを簡単に操作することができます。これらの領域、バンド、または要素で表示されるデータのノードを生成することもできます。詳細は、[p. 391 グラフの検証](#) を参照してください。

Web グラフ ノード

Web グラフ ノードでは、2 つ以上のシンボル値フィールドの値の相関の強さが示されます。このグラフでは相関の強さがさまざまな線の種類で示されます。Web グラフ ノードを使って、E コマース サイトや従来の小売店で購入されたさまざまな商品の関係を調査することができます。

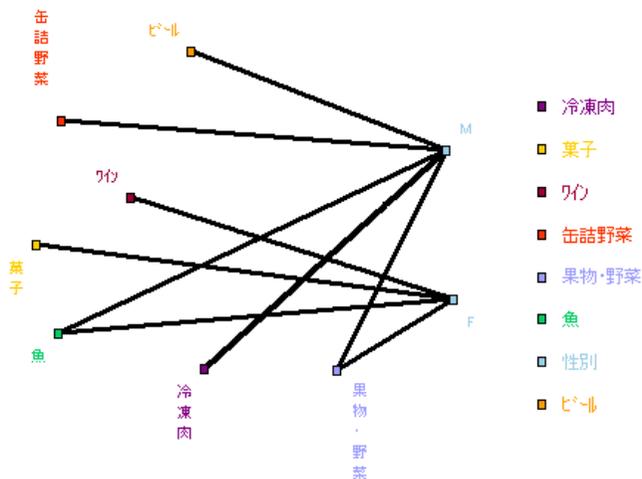
図 5-63
食料雑貨の購入の関係を表した Web グラフ



Web グラフ

Web グラフ ノードは、シンボル値フィールド間の相関の強さが示される点で MultiWeb グラフ ノードと似ています。ただし、Web グラフ ノードには、1 つの終点フィールドに対する 1 つ以上の始点フィールドからの相関だけが表示されます。これらの相関は一方方向です。

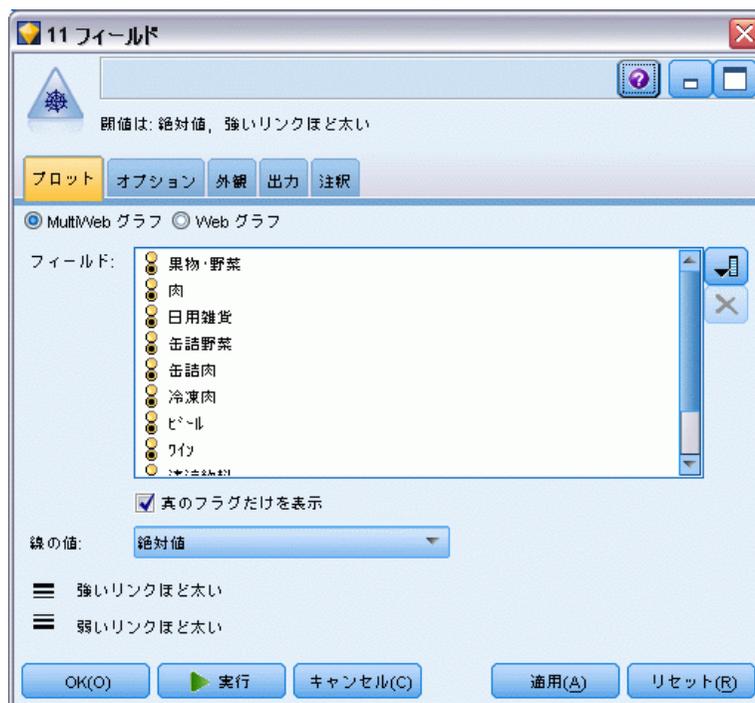
図 5-64
購入された食料雑貨品と性別の関係を表示した Web グラフ



MultiWeb グラフ ノードと同様、グラフではさまざまな線の種類を使用して相関の強さが示されます。Web グラフ ノードを使って、性別と特定商品の購入傾向との関連などを調査することができます。

Web グラフの [プロット] タブ

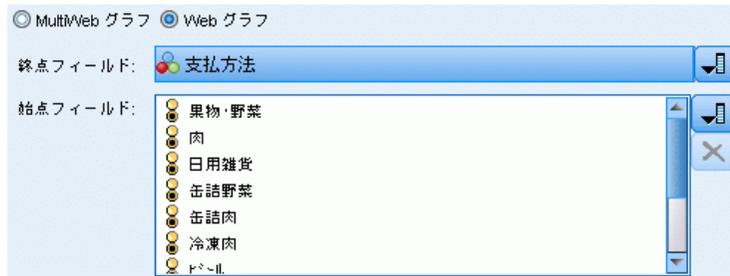
図 5-65
Web グラフ ノードの [プロット] タブの設定



MultiWeb グラフ :指定したすべてのフィールド間の相関の強さを表す Web グラフを作成する場合に選択します。

Web グラフ :複数のフィールドとある 1 つのフィールドの値（例：性別や宗教など）の相関の強さを表す Web グラフを作成する場合に選択します。このオプションを選択すると、[終点フィールド] が有効になります。また、下にある [フィールド] コントロール名が、[始点フィールド] に変わります。

図 5-66
Web グラフのオプション



終点フィールド (Web グラフの場合) : Web グラフで使用するフラグ型または名義型フィールドを選択します。数値として明示的に設定されていないフィールドだけがリストに表示されます。

フィールド/始点フィールド : Web グラフを作成するためのフィールドを選択します。数値として明示的に設定されていないフィールドだけがリストに表示されます。フィールド ピッカー ボタンを使って複数のフィールドを選択するか、またはフィールドの種類を選択します。

注： Web グラフの場合、このコントロールは始点フィールドを選択するために使用します。

真 (true) のフラグだけを表示 : フラグ型フィールドが真 (true) のフラグだけを表示する場合に選択します。このオプションを選択すると、Web グラフの表示が単純化されます。またこのオプションは、正の値の出現が特に重要となるデータによく使用されます。

線の値 : ドロップダウン リストから閾値のタイプを選択します。

- [絶対値] を選択すると、閾値はそれぞれ一对の値を持つレコードの数に基づいて設定されます。
- [パーセント (全体)] には、Web グラフで示された値の各ペアの頻度の割合として、リンクで表されているケースの絶対数が表示されます。
- また、[パーセンテージ (小さいフィールド/値)] と [パーセンテージ (大きいフィールド/値)] には、割合の評価に使うフィールドと値が示されます。たとえば、[薬品] フィールドに薬品Y の値を持つレコードが 100 個あり、[血圧] フィールドに低い値を持つレコードが 10 個しかない場合を考えてみましょう。このとき、薬品Y と低い両方の値を持つレコードが 7 個あると、このレコードの割合は小さいフィールド ([血圧] フィールド) または大きいフィールド ([薬品] フィールド) のどちらを参照するかに応じて、それぞれ 70% または 7% となります。

注： Web グラフの場合、上記の 3 番目のオプションは利用できません。代わりに、[パーセンテージ (“終点”フィールド/値)] と [パーセンテージ (“始点”フィールド/値)] を選択することができます。

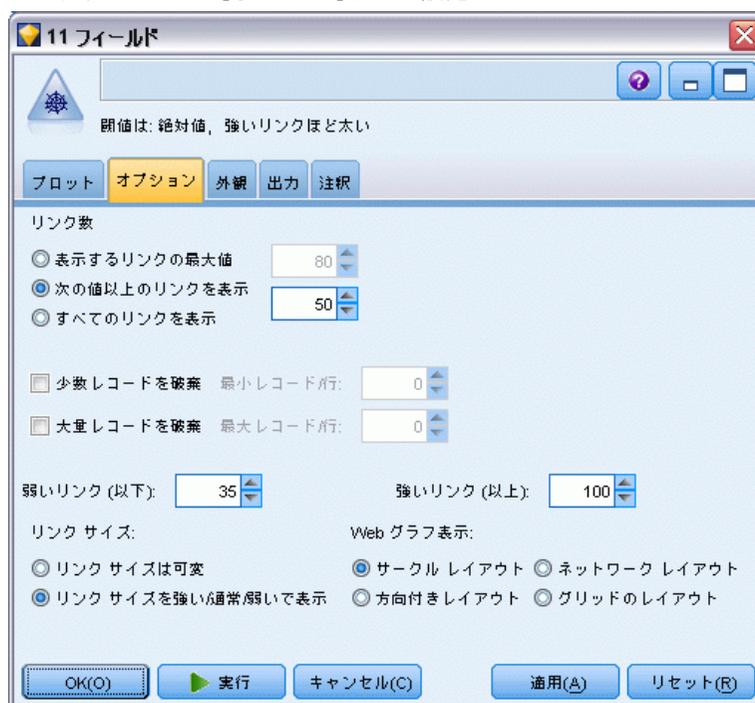
強い相関ほど太い :デフォルトで選択されています。フィールド間のリンクを表示する標準的な方法です。

弱い相関ほど太い :太線で表示されているリンクの意味を逆にする場合に選択します。このオプションは、不正行為の検出や外れ値の調査などで頻繁に使用されます。

Web グラフの [オプション] タブ

Web グラフ ノードの [オプション] タブには、出力グラフをカスタマイズするためのさまざまなオプションが用意されています。

図 5-67
Web グラフ ノードの [オプション] タブの設定



リンク数 :出力グラフに表示されるリンク数を制御するために使われるオプションを次に示します。[弱い関係ほど太い] や [強い関係ほど太い] など、一部のオプションは、出力グラフ ウィンドウでも使用することができます。最終的なグラフでスライダーを使って、表示するリンク数を調整することもできます。

- **表示するリンクの最大値** :出力グラフに表示する最大リンク数を指定します。矢印を使用して、値を調節してください。

- **次の値以上のリンクを表示** :Web グラフにリンクを表示する場合の最小値を指定します。矢印を使用して、値を調節してください。
- **すべてのリンクを表示** :最小値や最大値の設定に関係なく、すべてのリンクを表示します。フィールド数が多い場合にこのオプションを選択すると、処理時間が増加する可能性があります。

少数レコードを破棄 :リンクをサポートするレコード数が少ない場合、リンクを無視するときに選択します。[最小レコード/行] フィールドに数値を入力して、このオプションの閾値を設定します。

大量レコードを破棄 :強くサポートされているリンクを無視するときに選択します。[最大レコード/行] フィールドに数値を入力します。

弱いリンク(以下) :弱いリンク(点線)と通常のリンク(実線)の閾値を示す数値を指定します。これより小さな値のリンクはすべて弱いリンクと見なされます。

強いリンク(以上) :強いリンク(太線)と通常のリンク(実線)の閾値を示す数値を指定します。これより大きな値のリンクはすべて強いリンクと見なされます。

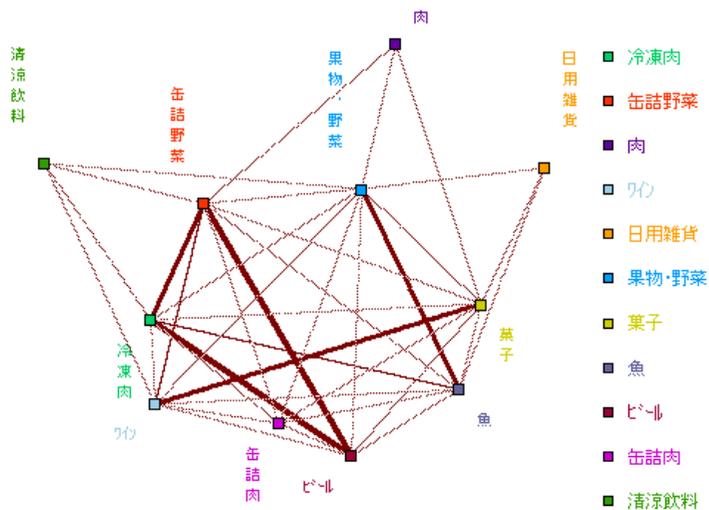
リンク サイズ :リンクのサイズを制御するオプションを指定します。

- **リンク サイズは可変** :このオプションを選択すると、実際のデータ値に基づいて接続強度の変動を反映する、リンク サイズの範囲が表示されます。
- **リンク サイズは固定** :このオプションを選択すると、3 種類の接続強度(強い、中間、弱い)が表示されます。これらのカテゴリの閾値は、上記のほかにも最終的なグラフで指定することができます。

Web グラフ表示 :Web グラフの表示の種類を選択します。

- **サークル レイアウト** :標準の Web グラフ表示を選択します。
- **ネットワーク レイアウト** :もっとも強いリンクをグループ化するアルゴリズムの使用を選択します。このオプションは、空間的な差異や重み付けられた線を使って、強いリンクを強調することを目的にしています。
- **方向付きレイアウト** :[分布図] タブから [終点フィールド] 選択をその方向に対するフォーカスとして使用する、Web グラフ表示を作成します。
- **グリッドのレイアウト** :規則的なスペースを含むグリッド パターンにレイアウトされた Web グラフ表示を作成します。

図 5-68
冷凍食品と缶詰野菜から他の食料雑貨への強い接続を表した Web グラフ



Web グラフの [外観] タブ

図 5-69
Web グラフ ノードの [外観] タブの設定



外観のオプションはグラフ作成前に指定できます。

タイトル： グラフのタイトルに使用するテキストを入力します。

サブタイトル： グラフのサブタイトルに使用するテキストを入力します。

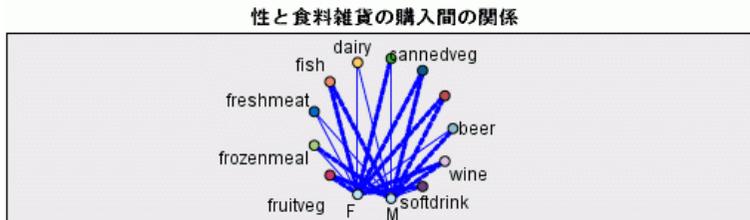
解説： グラフの解説に使用するテキストを入力します。

凡例を表示： 凡例の表示を指定することができます。多数のフィールドのあるプロットの場合、凡例を非表示にするとプロットの外観を改善できます。

ラベルをノードとして使用： ラベルを隣接表示するのではなく、ラベル テキストを各ノード内に表示します。プロットするフィールド数が少ない場合、グラフが読みやすくなります。

図 5-70

ラベルをノードとして表示する Web グラフ



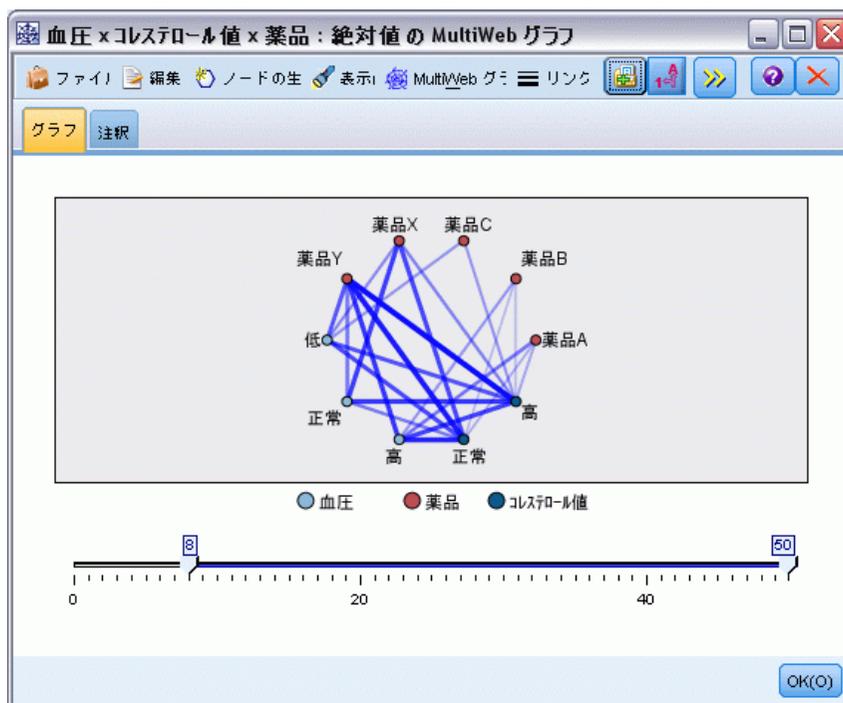
Web グラフの使用方法

Web グラフ ノードは、2 つ以上のシンボル値フィールドの値の相関の強さを表示するために用いられます。接続は、接続の強度を示すさまざまな種類の線としてグラフに表示されます。たとえば、Web グラフ ノードを使用して、コレステロール レベル、血圧、および患者の治療に効果的な薬品の相関を調べることができます。

- 強い接続は太い線で表されます。これは、2 つの値が強く関係しており、より詳細に調査する必要があることを示しています。
- 中程度の接続は、中間の線で表示されます。
- 弱い接続は点線で表されます。
- 2 つの値間に線がない場合は、これら 2 つの値が同じレコード内には完全に存在していないか、またはこれらの組み合わせの発生回数が、[Web グラフ ノード] ダイアログ ボックスで指定された閾値を下回っていることを意味しています。

Web グラフ ノードを生成すると、さまざまなオプションを使用してグラフの表示を調整したり、詳細な分析のためのノードを生成したりできるようになります。

図 5-71
 正常な血圧と薬品 X、高コレステロールと薬品 Y などの強い相関を表す Web グラフ



MultiWeb グラフ ノードと Web グラフ ノードでは、次の操作ができます。

- Web グラフの表示レイアウトを変更する。
- ポイントを隠して、表示を単純化する。
- 線の種類を制御する閾値を変更する。
- 「選択された」相関を示すために、値間の線を強調表示する。
- 1 つ以上の選択されたレコードに条件抽出ノードを生成するか、または Web グラフ内の 1 つ以上の相関に関連付けられたフラグ型フィールド作成ノードを生成する。

ポイントを調整するには

- **移動** : ポイントをマウスでクリックし、新しい場所にドラッグして移動します。これによって、新しい場所を反映して Web グラフが再描画されます。
- **非表示** : Web グラフのポイント上を右クリックし、コンテキストメニューから [非表示] または [非表示にして再計算] を選択して、ポイントを隠します。[非表示] を選択すると、選択したポイントとこのポイントに関連付けられた線が隠されます。[非表示にして再計算] を選択する

と、変更を反映して Web グラフが再描画されます。このとき、手動による移動は取り消されます。

- **表示**： グラフ ウィンドウの [Web グラフ] メニューから、[すべて表示] または [すべて表示して再計算] を選択して、隠したすべてのポイントを表示します。[すべて表示して再計算] を選択すると、前に隠したすべてのポイントとその相関を含めて Web グラフが再描画されます。

線を選択または「強調表示」するには

選択した行は赤で強調表示されます。

- ▶ 1 行を選択して、その行を左クリックします。
- ▶ 複数の行を選択するには、次のいずれかを実行します。
 - カーソルを使用して、選択したい行のポイントの周りで円を描きます。
 - Ctrl キーを押しながら、選択する各行を左クリックします。

グラフの背景をクリックするか、グラフ ウィンドウの Web メニューから [選択を解除] を選択して、選択するすべての行の選択を解除します。

異なるレイアウトを使って Web グラフを表示するには

- ▶ [Web グラフ] メニューから、[サークルレイアウト]、[ネットワークレイアウト]、[方向付きレイアウト]、または[グリッドのレイアウト] を選択し、グラフのレイアウトを変更します。

リンク スライダーをオンまたはオフにするには

- ▶ [表示] メニューから、[リンクスライダー]、 δ 'I ' δ , μ , \bar{U} , $\cdot B$

単一の相関に対してレコードを選択したり、フラグを設定するには

- ▶ 対象となる相関を表す線を右クリックします。
- ▶ コンテキスト メニューから [リンクの条件抽出ノード生成] または [リンクのフィールド作成ノード生成] を選択します。

適切なオプションと指定した条件を備えた条件抽出ノードまたはフィールド作成ノードが、自動的にストリーム領域に追加されます。

- 条件抽出ノードでは、指定された相関に該当するレコードがすべて選択されます。
- フィールド作成ノードでは、データ セット全体のレコードにおいて、選択された相関が真 (true) であるかどうかを示すフラグが生成されます。フラグ フィールドには、低_薬品C や 薬品C_低 など、相関関係にある 2 つの値を下線で連結した名前が付けられます。

関連グループのレコードを選択したり、フラグを設定するには

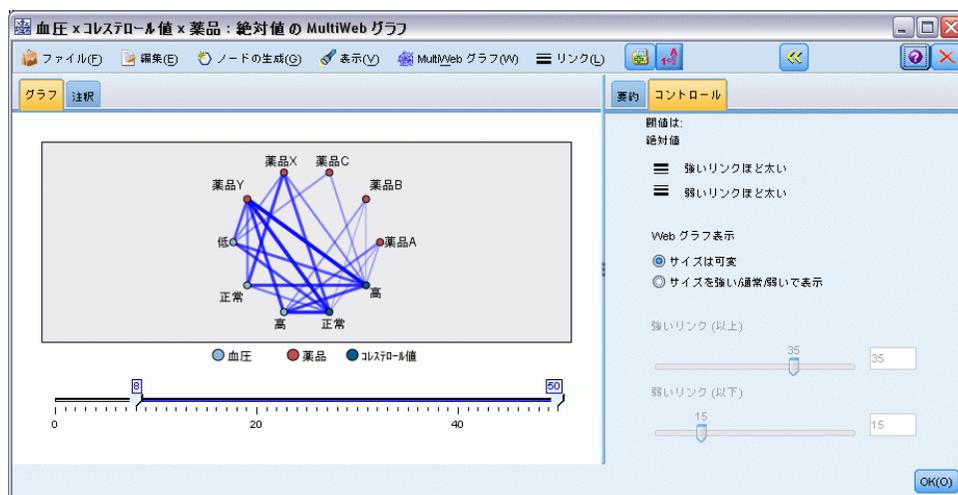
- ▶ Web グラフで、対象となる相関を表す線を選択します。
- ▶ グラフ ウィンドウの [ノードの生成] メニューで、[条件抽出ノード (AND)]、[条件抽出ノード (OR)]、[フィールド作成ノード (AND)]、または [フィールド作成ノード (OR)] を選択します。
 - “OR” ノードでは条件が分岐されます。つまり、このノードは選択したいずれかの相関に該当するレコードに適用されます。
 - “AND” ノードでは条件が結合されます。つまり、このノードは選択したすべての相関に該当するレコードだけに適用されます。選択した相関が相互に排他的な場合は、エラーが発生します。

選択が完了すると、適切なオプションと指定した条件を備えた条件抽出ノードまたはフィールド作成ノードが、自動的にストリーム領域に追加されます。

Web グラフの閾値の調整

Web グラフを作成したら、最低限表示する線を変更するために、ツールバーやスライダを使用して、線のスタイルを制御する閾値を調整することができます。また、ツールバーの黄色い二重矢印ボタンをクリックして [Web グラフ] ウィンドウを展開し、閾値に関する付加オプションを表示することもできます。次に、[コントロール] タブをクリックして、付加オプションを表示します。

図 5-72
ウィンドウを展開して表示および閾値オプションを表示



閾値は:[Web グラフ ノード] ダイアログ ボックス での作成中に選択した閾値のタイプが表示されます。

強い相関ほど太い :デフォルトで選択されています。フィールド間のリンクを表示する標準的な方法です。

弱い相関ほど太い :太線で表示されているリンクの意味を逆にする場合に選択します。このオプションは、不正行為の検出や外れ値の調査などで頻繁に使用されます。

Web グラフ表示 :出力グラフ中のリンクのサイズを制御するオプションを指定します。

- **サイズは可変** :このオプションを選択すると、実際のデータ値に基づいて接続強度の変動を反映する、リンク サイズの範囲が表示されます。
- **サイズを強い/通常/弱いで表示** :このオプションを選択すると、3 種類の接続強度（強い、中間、弱い）が表示されます。これらのカテゴリの閾値は、上記のほかにも最終的なグラフで指定することができます。

強いリンク (以上) :強いリンク（太線）と通常のリンク（実線）の閾値を示す数値を指定します。これより大きな値のリンクはすべて強いリンクと見なされます。スライダを使って値を調整するか、またはフィールドに値を入力します。

弱いリンク (以下) :弱いリンク（点線）と通常のリンク（実線）の閾値を示す数値を指定します。これより小さな値のリンクはすべて弱いリンクと見なされます。スライダを使って値を調整するか、またはフィールドに値を入力します。

Web グラフの閾値を調整したら、Web グラフ ツールバーにある Web グラフメニューを使用して、新しい閾値で Web グラフを再計算または再描画することができます。もっともよく意味のあるパターンが現れる設定を見つけたら、グラフ ウィンドウの [Web グラフ] メニューから [親ノードの更新] を選択して、Web グラフ ノード（親ノードとも呼ばれます）の元の設定を更新することができます。

Web グラフの概要の作成

強い、中間、および弱いリンクを記載した Web グラフの概要ドキュメントを作成するには、ツールバーの黄色い二重矢印ボタンをクリックして [Web グラフ] ウィンドウを展開します。次に、[要約] タブをクリックして、各種類のリンクのテーブルを表示します。テーブルは、トグル ボタンを使って展開したり、閉じることができます。

図 5-73
血圧、コレステロール、および薬品の種類の相関を一覧表示した Web グラフの概要

| 要約 | | |
|--------|-----------------|-----------------|
| コントロール | | |
| 強いリンク | | |
| リンク | フィールド 1 | フィールド 2 |
| 47 | コレステロール値 = "高" | 薬品 = "薬品 Y" |
| 44 | コレステロール値 = "正常" | 薬品 = "薬品 Y" |
| 42 | 血圧 = "高" | コレステロール値 = "正常" |
| 38 | 血圧 = "高" | 薬品 = "薬品 Y" |
| 37 | 血圧 = "正常" | コレステロール値 = "高" |
| 36 | 血圧 = "正常" | 薬品 = "薬品 X" |
| 中間のリンク | | |
| リンク | フィールド 1 | フィールド 2 |
| 35 | 血圧 = "高" | コレステロール値 = "高" |
| 34 | コレステロール値 = "正常" | 薬品 = "薬品 X" |
| 33 | 血圧 = "低" | コレステロール値 = "正常" |
| 31 | 血圧 = "低" | コレステロール値 = "高" |
| 30 | 血圧 = "低" | 薬品 = "薬品 Y" |
| 23 | 血圧 = "正常" | 薬品 = "薬品 Y" |
| 23 | 血圧 = "高" | 薬品 = "薬品 A" |
| 22 | 血圧 = "正常" | コレステロール値 = "正常" |
| 20 | コレステロール値 = "高" | 薬品 = "薬品 X" |
| 18 | 血圧 = "低" | 薬品 = "薬品 X" |
| 16 | 血圧 = "低" | 薬品 = "薬品 C" |
| 16 | コレステロール値 = "高" | 薬品 = "薬品 C" |
| 16 | 血圧 = "高" | 薬品 = "薬品 B" |
| 弱いリンク | | |
| リンク | フィールド 1 | フィールド 2 |
| 12 | コレステロール値 = "高" | 薬品 = "薬品 A" |
| 11 | コレステロール値 = "正常" | 薬品 = "薬品 A" |
| 8 | コレステロール値 = "高" | 薬品 = "薬品 B" |
| 8 | コレステロール値 = "正常" | 薬品 = "薬品 B" |

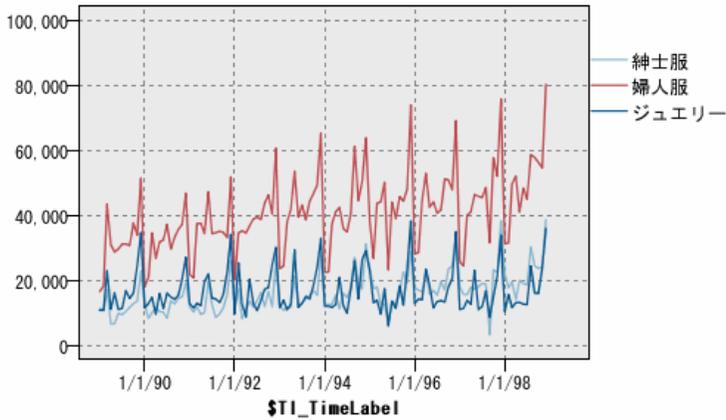
要約を印刷するには、Web グラフ ウィンドウのメニューから次を選択します。

ファイル > 要約の印刷

時系列ノード

時系列ノードを使用すると、長時間にわたり連続して繰り返しプロットすることができます。一連のプロットには数値が含まれていなければならない、周期が一定の時間の領域にプロットすることを前提とします。通常、時系列ノードの前に時間区分ノードを使用して TimeLabel フィールドを生成します。このノードはデフォルトでグラフに x 軸のラベルを指定するのに使用されます。詳細は、4 章 p.241 時間区分ノード を参照してください。

図 5-74
男女の衣類や宝石の売り上げを長期間プロット

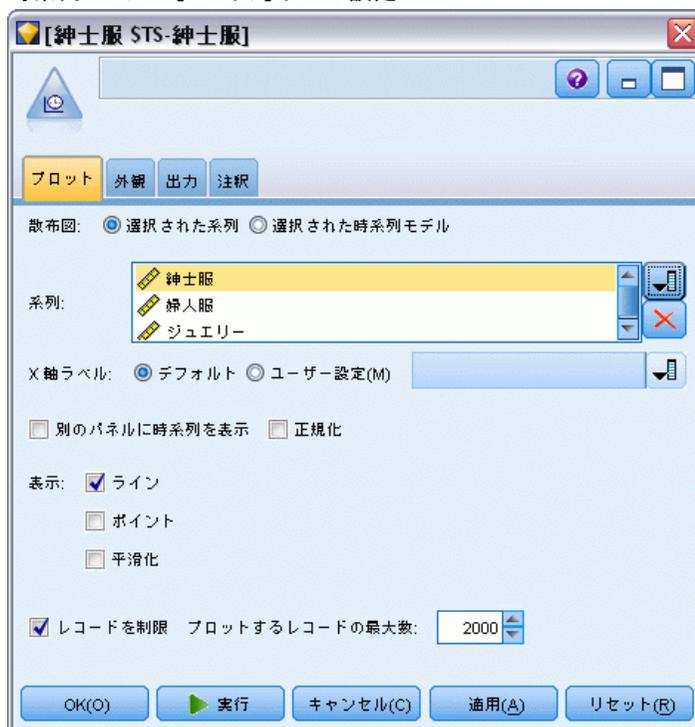


干渉およびイベントの作成

コンテキストメニューからフィールド作成（フラグまたは名義型）ノードを生成し、時系列からイベントおよび干渉フィールドを作成することができます。たとえば、鉄道会社のストライキが発生した場合イベントフィールドを作成することができ、そこでイベントが発生した場合はドライブステートを真（true）、発生しない場合は偽（false）となります。干渉フィールドの場合、たとえば価格の値上げに対して、フィールド作成（カウント型）を使用し、古い価格に対しては 0、新しい価格に対しては 1 を指定して、値上げの日付を識別することができます。詳細は、4 章 p.185 フィールド作成ノード を参照してください。

時系列の [プロット] タブ

図 5-75
時系列ノードの [プロット] タブの設定



プロット : 時系列データのプロット方法の選択を提供します。

- **選択された系列** : 選択された時系列に対するプロット値。信頼区間のプロット時このオプションを選択する場合、**[正規化]** チェック ボックスを解除します。
- **選択された時系列モデル** : 時系列モデルとともに使用し、このオプションはすべての関連フィールド（信頼区間同様、実際のおよび予測された値）を 1 つ以上の選択された時系列に対してプロットします。このオプションは、ダイアログ ボックスのその他のオプションを無効化します。信頼区間をプロットする場合、このオプションが優先されます。

系列 : プロット対象の時系列データを含む 1 つ以上のフィールドを選択します。このデータは数値でなければなりません。

X 軸ラベル : デフォルトのラベルを選択するか、単一のフィールドを選択して、プロットで x 軸のラベルとして使用します。**[デフォルト]** を選択すると、上流の時間区分ノード、または時間区分ノードが存在しない場合は連続する整数から生成された TimeLabel フィールドを使用します。 [詳細は、4 章 p.241 時間区分ノード を参照してください。](#)

別のパネルに時系列を表示 :別のパネルにそれぞれの時系列を表示するかどうかを指定します。または、パネルを選択しない場合は、すべての時系列が同じグラフ上にプロットされ、平準化は利用できません。すべての時系列を同一グラフにプロットする場合、各時系列は別々の色で表されます。

正規化 :すべての Y 値をグラフの 0~1 の範囲で増減させる場合に選択します。正規化を使用することで、各系列の値の範囲内での差異が原因で不明確になりかねない各線の間関係が明らかになります。また、同じグラフ上に複数の線をプロットする場合や、隣り合ったパネル内でプロットを比較する場合に、正規化をお勧めします。(正規化は、すべてのデータ値が似たような範囲内に収まる場合は不要です。)

表示 :グラフに表示する要素を 1 つ以上選択します。[ライン]、[ポイント]、[(LOESS) 平準化] から選択できます。平準化は、時系列を別のパネルに表示している場合に使用できます。デフォルトでは、ライン要素が選択されています。グラフ ノードを実行する前に、少なくとも 1 つのプロット要素を必ず選択してください。何も選択しないと、プロットするものを選択してない旨を告げるエラーが返されます。

レコードを制限 :プロットされるレコードを制限したい場合は、このオプションを選択します。データ ファイルの先頭部分から読み取った、プロットされるレコードの数を、[プロットするレコードの最大数] オプションで指定します。デフォルトでは、2,000 に設定されています。データ ファイルの最後の n 個のレコードをプロットしたい場合は、このノードを実行する前にソートノードを使用してレコードを時間の降順に並べ替えることができます。

時系列の [外観] タブ

図 5-76
時系列ノードの [外観] タブの設定



外観のオプションはグラフ作成前に指定できます。

タイトル: グラフのタイトルに使用するテキストを入力します。

サブタイトル: グラフのサブタイトルに使用するテキストを入力します。

解説: グラフの解説に使用するテキストを入力します。

Xラベル: 自動的に生成された x 座標（水平）ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

Yラベル: 自動的に生成された y 座標（垂直）ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

グリッドの表示: デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

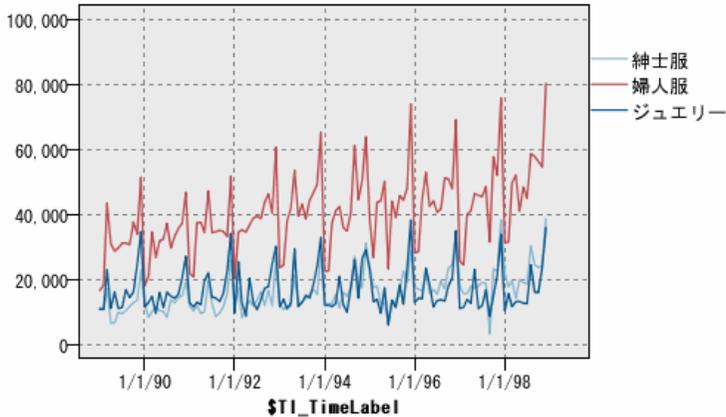
レイアウト: 時系列についてのみ、時間値を横軸に表示するか縦軸に表示するかを指定することができます。

時系列グラフの使用方法

時系列グラフを生成すると、さまざまなオプションを使用してグラフの表示を調整したり、詳細な分析のためのノードを生成したりできるようになります。 [詳細は、 p. 391 グラフの検証 を参照してください。](#)

図 5-77

男女の衣類や宝石の売り上げを長期間プロット



時系列を作成し、バンドを定義して結果を調べると、[ノードの生成] メニューとコンテキスト メニューのオプションを使用して、バランス ノード、条件抽出ノード、またはフィールド作成ノードを生成できるようになります。 [詳細は、 p. 402 グラフからのノードの生成 を参照してください。](#)

評価ノード

評価ノードを使用すると、簡単に予測モデルを評価および比較して、アプリケーションに最適なモデルを選択できます。評価グラフでは、各モデルの特定の結果の予測方法が示されます。また評価グラフでは、レコードは予測フィールドと予測の確信度に基づいてソートされて等しいサイズのグループ (分位) に分割され、分位ごとにビジネスに関する基準の値が高い方から順番にプロットされます。プロットには、複数のモデルが異なる線で示されます。

結果は、特定の値または値の範囲を**ヒット**として定義することで処理されます。通常、ヒットはある種の成功 (顧客への販売など) や対象となるイベント (特定の医療診断など) を示します。ダイアログ ボックスの [オプション] タブでヒット基準を定義したり、次のようなデフォルトのヒット基準を使用することができます。

- **フラグ型出力フィールドの場合**、ヒットはそのまま真の値に対応しています。

- 一方**名義型**出力フィールドの場合は、セットの最初の値がヒットを表します
- **連続型**出力フィールドの場合、フィールドの範囲の中間より大きい値がヒットになります。

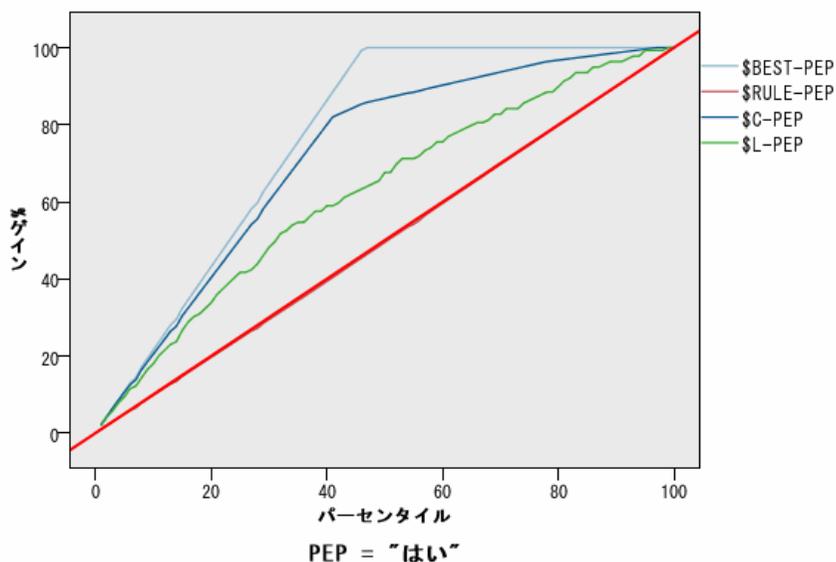
評価グラフには 5 種類あり、それぞれ強調される評価基準は異なります。

ゲイン グラフ

増加率は各分位で発生する総ヒット数の割合として定義され、(分位内のヒット数 / 総ヒット数) × 100% によって計算されます。

図 5-78

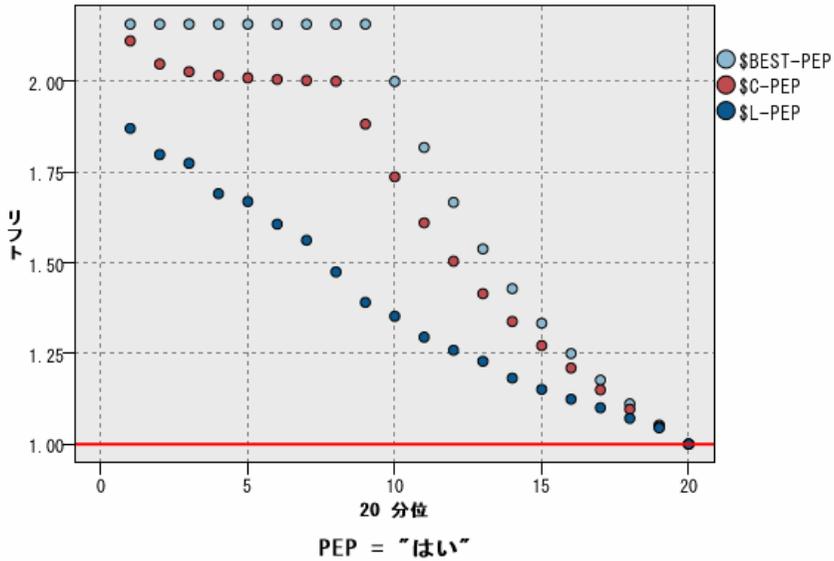
基準、ベストライン、およびビジネス ルールが表示されたゲイン グラフ(累積)



リフト グラフ

リフトでは、各分位でヒットしたレコードの割合（パーセント）が、トレーニング データ内の全ヒットの割合と比較されます。これは、(分位内のヒット数 / 分位内のレコード数) / (総ヒット数 / 総レコード数) で算出されます。

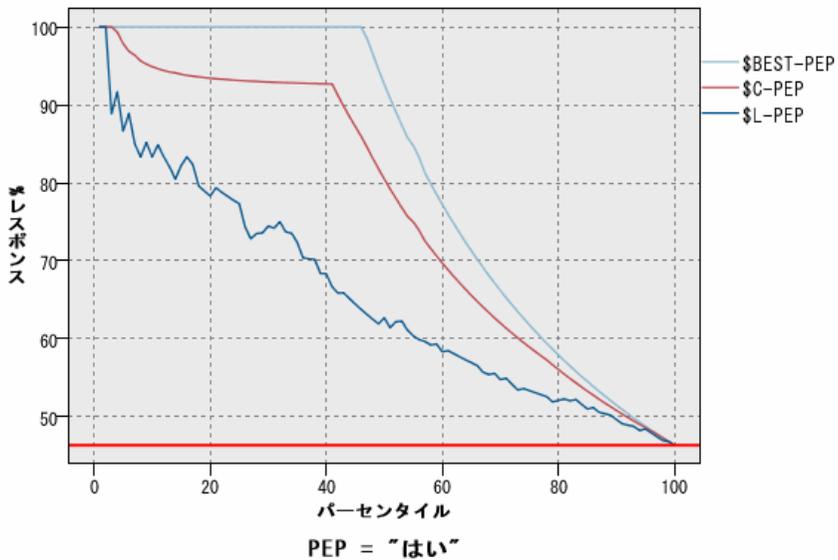
図 5-79
ポイントおよびベストラインを使用するリフトグラフ(累積)



レスポンス グラフ

レスポンスは分位内のヒットしたレコードの割合であり、(分位内のヒット数 / 分位内のレコード数) × 100% によって計算されます。

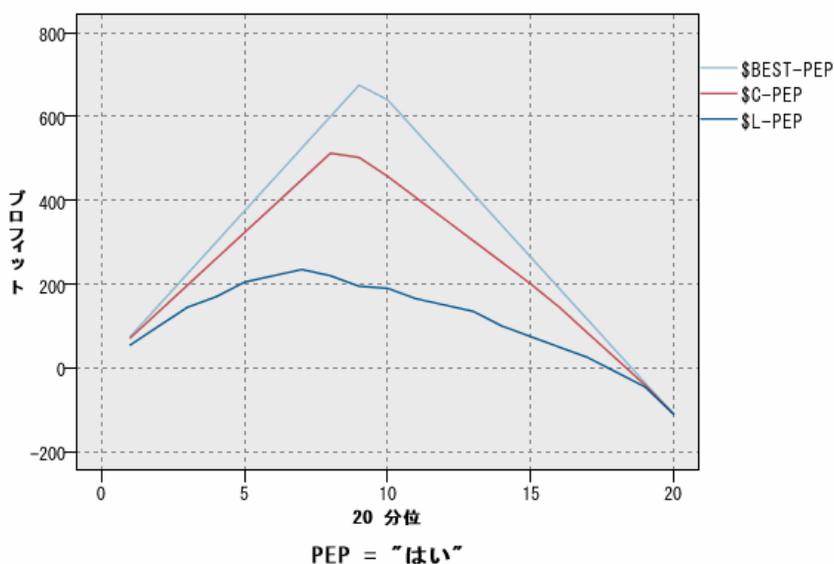
図 5-80
ベストラインが表示されたレスポンスグラフ(累積)



プロフィット グラフ

プロフィットは、各レコードの**収益**から、そのレコードの**コスト**を引いた値と等しくなります。分位のプロフィットは、その分位の全レコードのプロフィットを合計したものです。収益はヒットだけに適用されることを前提としますが、コストはすべてのレコードに適用されます。また、プロフィットとコストは固定にすることも、データのフィールドで定義することもできます。プロフィットは、(分位内のレコードの合計収益 - 分位内のレコードの合計コスト) によって計算されます。

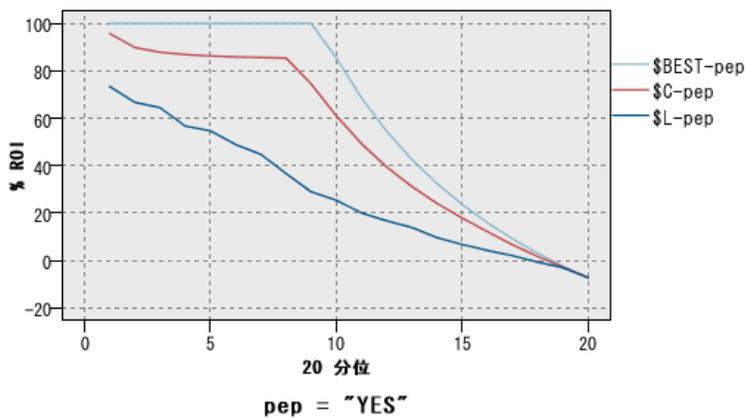
図 5-81
ベストラインが表示されたプロフィット グラフ (累積)



ROI グラフ

ROI (投資収益) は、収益とコストを定義するという点でプロフィット グラフと似ていますが、ROI では分位のプロフィットとコストが比較されます。(分位のプロフィット / 分位のコスト) × 100% によって計算されます。

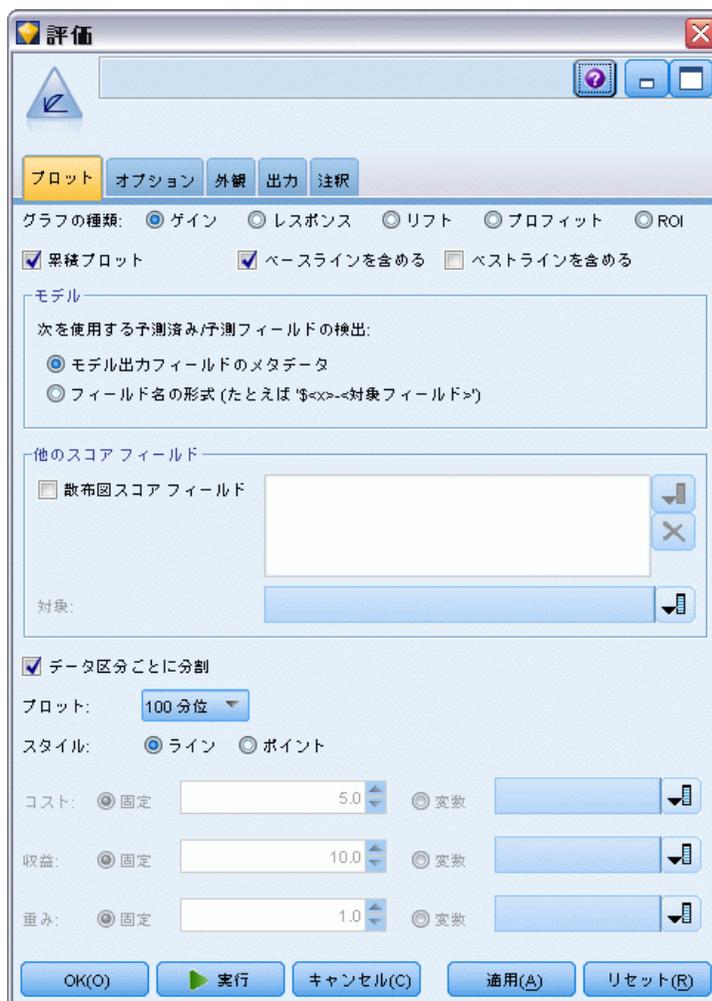
図 5-82
ベストラインが表示されたROI グラフ (累積)



評価グラフは、各ポイントが対応する分位とそれより上位にあるすべての分位の値と等しくなるように、累積で表すこともできます。累積グラフは、通常モデルの全体的な性能を表す場合場合に役に立ちます。一方、非累積グラフは、そのモデルにおける特定の問題領域を表す場合場合に役に立ちます。

評価の [プロット] タブ

図 5-83
評価ノードの [プロット] タブの設定



グラフの種類 :次に示すタイプの 1 つを選択します。[ゲイン]、[レスポンス]、[リフト]、[利益]、または [ROI]（投資収益）。

累積プロット :累積グラフを作成する場合に選択します。累積グラフの値は、各分位とそれより上位にあるすべての分位に対してプロットされます。

ベースラインを含める :完全にランダムなヒットの分布で、確信度が無関係になることを示すため、プロットにベースラインを含める場合に選択します（[ベースラインを含む] は、利益グラフと ROI グラフでは使用できません）。

ベストラインを含める :プロットにベストラインを含めて、完全な確信度（ヒットがケースの 100%）を示す場合に選択します。

次を使用する予測済み/予測フィールドの検出 :[モデル出力フィールドのメタデータ] を選択してそれらのメタデータを使用するグラフ内で予測済みフィールドを検索するが、[フィールド名の形式] を選択して名前フィールドを検索します。

散布図スコア フィールド :このチェック ボックスを選択すると、スコア フィールド ピッカーが有効になります。次に、1 つ以上の範囲、または連続したスコア フィールド（厳密には予測モデルではないが、ヒットになる傾向の点でレコードをランク付けするのに役立つ可能性のあるフィールド）を選択します。評価ノードでは、1 つ以上のスコア フィールドの任意の組み合わせを 1 つ以上の予測モデルと比較できます。典型的な例は、複数の RFM フィールドを最適な予測モデルと比較することです。

対象 :フィールド ピッカーを使用して対象フィールドを選択します。2 つ以上の値を持つインスタンス化されたフラグまたは名義型フィールドを選択してください。

注： この対象フィールドは、スコア フィールド（予測モデルが独自の対象を定義）のみに該当するため、[オプション] タブでヒット条件がユーザー設定されている場合は無視されます。

データ区分データによる分割 :レコードを、学習、テスト、および検定用の各サンプルに分割するためにデータ区分フィールドが使用される場合、このオプションを選択すると、各データ区分ごとに、別々の評価グラフが表示されます。 [詳細は、4 章 p.229 データ区分ノード を参照してください。](#)

注： データ区分を分割する場合、データ区分フィールドにあるヌル値を持つレコードは、評価から除外されます。データ区分ノードは、ヌル値を生成しないため、データ区分ノードを使用している場合は、問題になりません。

作図 :ドロップダウン リストから、グラフにプロットする分位のサイズを選択します。[4 分位]、[5 分位]、[10 分位]、[20 分位]、[100 分位]、または [1000 分位] を選択します。

スタイル :[ライン] または [ポイント] を選択します。

プロフィット グラフと ROI グラフ :プロフィット グラフと ROI グラフの場合は、コスト、収益、および重みをコントロールできます。

- **コスト** : 各レコードに関連付けるコストを指定します。[固定] または [変数] を選択することができます。固定コストの場合はコストの値を指定してください。可変コストの場合は、フィールド ピッカー ボタンをクリックして、コスト フィールドとして使用するフィールドを選択してください。
- **収益** : ヒットを表し各レコードに関連付ける収益を指定します。[固定] または [変数] を選択することができます。固定収益の場合は収益値を指定してください。可変収益の場合は、フィールド ピッカー ボ

タンをクリックして、収益フィールドとして使用するフィールドを選択してください。

- **重み**：データのレコードが複数のユニットからなる場合は、出現頻度の高い重みを使用して結果を調整できます。[固定] または [変数] を選択して、各レコードに関連付ける重みの種類を指定します。重みを固定する場合は、重みの値（レコードごとのユニット数）を指定してください。重みを変数にする場合は、フィールド ピッカー ボタンをクリックして、重みフィールドとして使用するフィールドを選択してください。

評価の [オプション] タブ

評価グラフノードの [オプション] タブでは、グラフに表示するヒット、スコアリング基準、およびビジネス ルールなどを定義することができます。また、モデルの評価結果をエクスポートするためのオプションも設定することができます。

図 5-84
評価ノードの [オプション] タブの設定



ユーザー定義のヒット: ヒットを示す条件を自分で指定する場合に選択します。このオプションは、対象フィールドの種類および値の並びから推論する代わりに対象の結果を定義する場合に役立ちます。

- **条件**: [ユーザー定義のヒット] を選択した場合に、ヒット条件の CLEM 式を指定する必要があります。たとえば、@TARGET = "YES" は「Yes」の値を持つ対象フィールドをヒットとして評価する場合の有効な式になります。指定された条件は、すべての対象フィールドに対して使用されます。条件を作成するには、フィールドに式を直接入力するか、または Clem 式ビルダーを使って条件式を生成します。データがインスタンス化されている場合、Clem 式ビルダーから直接値を挿入することができます。

ユーザー定義のスコア: 分位に割り当てる前に、ケースのスコアリングに使用する条件を指定する場合に選択します。デフォルトのスコアは、予測値と確信度から算出されます。[式] フィールドを使って、スコアリング式を作成します。

- **式**: スコアリングに使用する CLEM 式を指定します。たとえば、範囲 0 ~ 1 の数値出力が、低い値の方が高い値よりも良好であると順序付ける場合、ヒットを @TARGET < 0.5、関連するスコアを 1 - @PREDICTED と定義することができます。スコア式の結果は数値でなければなりません。条件を作成するには、フィールドに式を直接入力するか、または Clem 式ビルダーを使って条件式を生成します。

ビジネス ルールを含める: 注目する基準を表すルール条件を指定します。たとえば、mortgage = "Y" and income >= 33000 であるすべてのケースのルールを表示する場合があります。ビジネス ルールは、グラフに描画され、キーにルールとしてラベル付けされます。

- **条件**: 出力グラフのビジネス ルールとして定義するために使用する CLEM 式を指定します。フィールドに式を直接入力するか、または Clem 式ビルダーを使って条件式を生成します。データがインスタンス化されている場合、Clem 式ビルダーから直接値を挿入することができます。

結果をファイルへエクスポート: 区切り文字で区切られたモデルの評価結果をテキスト ファイルへエクスポートする場合に選択します。このファイルを読み込んで、計算した値の特殊な分析を行うことができます。次のエクスポート オプションを設定します。

- **ファイル名**: 出力ファイルのファイル名を入力します。[...] ボタンを使って、目的のフォルダを指定することもできます。
- **区切り文字**: カンマやスペースなど、フィールドの区切り文字として使用する文字を入力します。

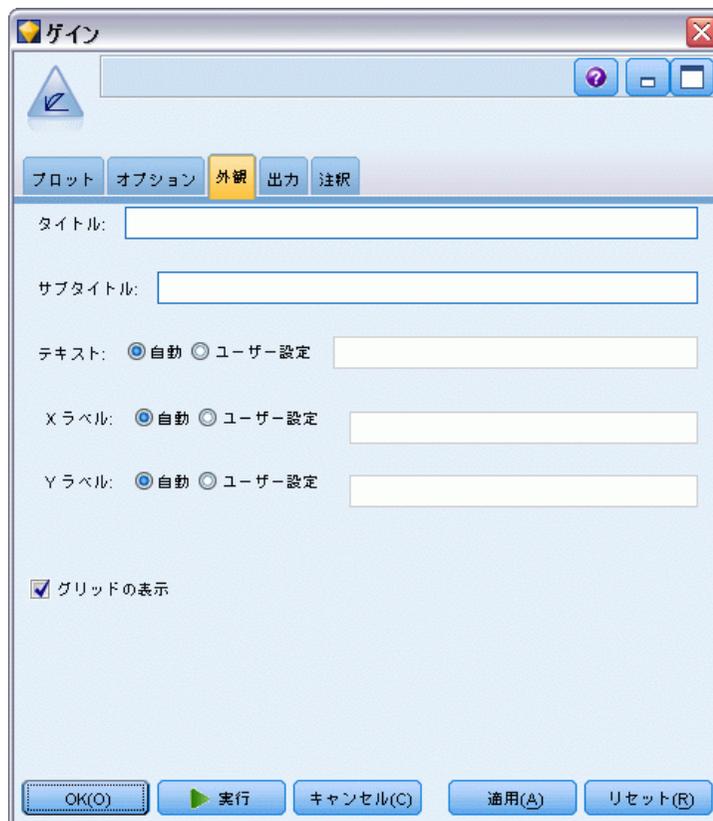
フィールド名を含める: 出力ファイルの最初の行にフィールド名を表示する場合に選択します。

各レコードの後に改行を入れる: 各レコードを新しい行で表示する場合に選択します。

評価の [外観] タブ

外観のオプションはグラフ作成前に指定できます。

図 5-85
評価ノードの [外観] タブの設定



タイトル: グラフのタイトルに使用するテキストを入力します。

サブタイトル: グラフのサブタイトルに使用するテキストを入力します。

テキスト: 自動生成されたテキスト ラベルを承認するか、[ユーザー設定] を選択してラベルを指定します。

Xラベル: 自動的に生成された x 座標 (水平) ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

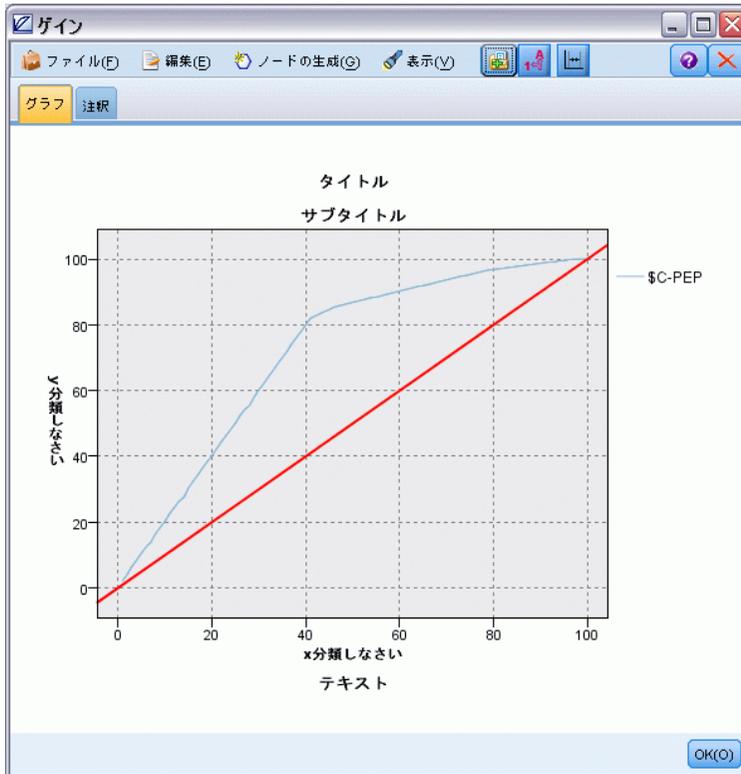
Yラベル: 自動的に生成された y 座標 (垂直) ラベルを承認するか、[ユーザー設定] を選択してユーザー設定ラベルを指定します。

グリッドの表示: デフォルトで選択されているこのオプションは、散布図またはグラフの背景にグリッドを表示する場合に使用します。グリッドを使用すると、領域やバンドの分割点を簡単に決めることができます。グリッ

ド線は、グラフの背景が白でない限り、常に白で表示されます。背景が白の場合は、灰色で表示されます。

次の例では、表示オプションがグラフ内のどこに表示されるかを示します

図 5-86
評価グラフのグラフ外観オプションの位置



モデル評価の結果の読み込み

評価グラフの解釈は、ある程度グラフの種類によって異なりますが、すべての評価グラフに共通する特性もあります。累積グラフの場合、特にグラフの左側において、上部の線ほどすぐれたモデルであることを意味しています。また、複数モデルを比較するとき線がクロスすることがよくあります。つまり、グラフの一部であるモデルの線が上になっており、グラフの別の部分で異なるモデルの線が上になっているような場合です。このような場合に使用するモデルを選択するときは、サンプルのどの部分が必要かを考慮（x 軸上のポイントを定義）しなければなりません。

ほとんどの非累積グラフは非常に似ています。すぐれたモデルの場合、非累積グラフはグラフの左側が高く、右側は低くなります(非累積グラフがこのぎり形になっている場合は、プロットする分位数を減らしてグラフを

再実行すると、滑らかにすることができます)。グラフの左側が落ち込んでいる場合や右側が鋭くとがっている場合は、モデルのその領域の予測精度が低いことを意味しています。また、グラフ全体を通して線が平坦な場合は、モデルからはほとんど情報が得られないことを意味しています。

ゲイン グラフ:累積ゲイン グラフは、常に左から右に向かって進み、0% で始まり 100% で終わります。適切なモデルの場合、ゲイン グラフは 100% に向けて急勾配で上昇し、その後、水平状態になります。左下から右上に 45 度の線を描くモデルからは情報が得られません (このモデルは、[ベースラインを含める] を選択した場合にグラフに表示されます)。

リフト グラフ:累積リフト グラフは、左から右に向かって進み、1.0 より上から始まって 1.0 に近づくにつれ徐々に下降します。グラフの右端がデータセット全体を表します。つまり、データ全体のヒット数に対する累積分位内のヒット数の比率は 1.0 です。すぐれたモデルの場合、リフト グラフは左端で 1.0 のかなり上から始まったまま、右に移動しても高い状態を維持し、グラフの右側で 1.0 に向かって急激に下降します。グラフ全体にわたって線が 1.0 付近に留まっているモデルからは、情報が得られません ([ベースラインを含める] を選択すると、グラフの 1.0 のレベルに水平線が参照線として表示されます)。

レスポンス グラフ:累積レスポンス グラフはリフト グラフと非常に似ていますが、尺度が異なります。レスポンス グラフは、通常 100% 付近から始まり、グラフの右端で全体的な応答率 (総ヒット数 / 総レコード数) に達するまで徐々に下降します。すぐれたモデルの場合、線は左側の 100% 付近から始まったまま、右に移動しても高い状態を維持し、グラフの右側で全体的な応答率に向かって急激に下降します。グラフ全体にわたって線が全体的な応答率のレベルに留まっているグラフからは、情報が得られません ([ベースラインを含める] を選択すると、グラフの全体的な応答率のレベルに水平線が参照線として表示されます)。

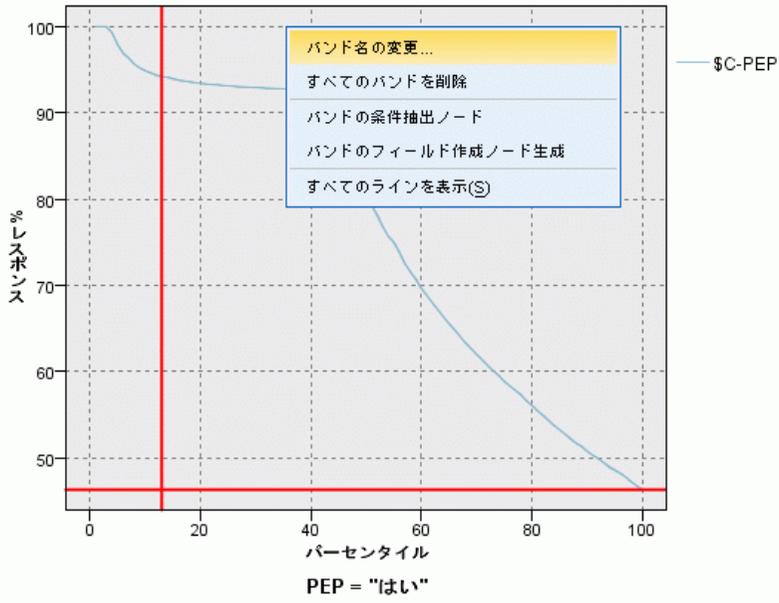
プロフィット グラフ:累積プロフィット グラフでは、左から右に向かって、選択したサンプルのサイズの増加に伴う利益の合計が示されます。通常、プロフィット グラフは 0 付近から始まり、中央の山形または台形の部分に向かって徐々に増加し、グラフの右端に向かって下降します。すぐれたモデルの場合、グラフの中央付近にくっきりとした山形が見られます。線が比較的真っ直ぐで、適用するコストや収益の構造によって上昇または下降したり、平坦になったりするモデルからは、情報が得られません。

ROI グラフ:累積 ROI (投資収益) グラフは、応答グラフおよびリフト グラフと似ていますが、尺度が異なります。ROI グラフは通常 0% 付近から始まり、データ セット全体の ROI に達するまで徐々に下降します (負の値になることもあります)。すぐれたモデルの場合、線は 0% 付近から始まったまま、右に移動しても高い状態を保ち、グラフの右側で全体的な ROI 値に向かって急激に下降します。線が全体的な ROI 値の付近に留まっているモデルからは情報が得られません。

評価グラフの使用方法

マウスを使った評価グラフの探索は、ヒストグラムや集計棒グラフの場合と同じように行うことができます。x 軸は、20 分位や 10 分位など、指定した分位によるモデル スコアを表しています。

図 5-87
評価グラフの作業

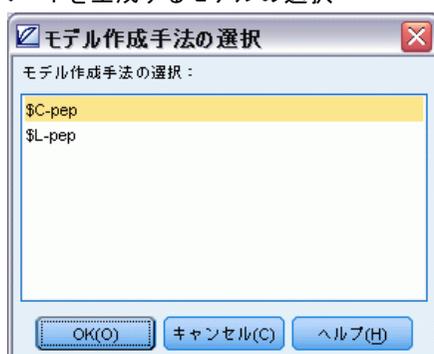


分割アイコンを使って軸を等しいバンドに自動分割するオプションを設定して、ヒストグラムのように x 軸をバンドに分割することができます。詳細は、[p. 391 グラフの検証](#) を参照してください。バンドの境界を手作業で編集するには、[編集] メニューの [グラフバンド] を選択します。

評価グラフを作成し、バンドを定義して結果を調べたら、[ノードの生成] メニューとコンテキストメニューのオプションを使用して、グラフ中の選択項目に基づいて自動的にノードを生成できるようになります。詳細は、[p. 402 グラフからのノードの生成](#) を参照してください。

評価グラフからノードを生成する場合、グラフで利用できるすべてのモデルから 1 つのモデルを選択するように指示するメッセージが表示されます。

図 5-88
ノードを生成するモデルの選択



モデルを選択して、[OK] をクリックすると、ストリーム領域に新しいノードが生成されます。

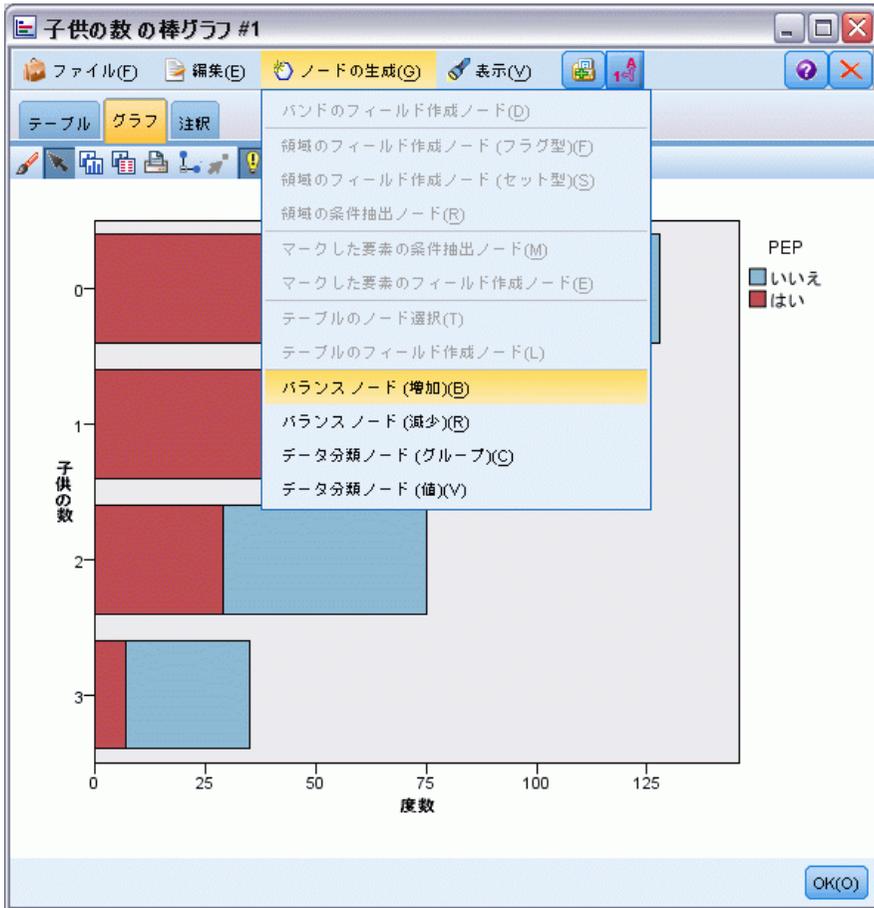
グラフの検証

編集モードを使用するとグラフのレイアウトおよび外観を編集できますが、検証モードを使用するとグラフに表示されたデータおよび値を分析的に検証できます。検証の主な目的は、データを分析し、バンド、領域およびマークを使用して値を識別し、条件抽出ノード、フィールド作成ノード、バランスノードを生成します。検証モードを選択するには、メニューから [表示] → [検証] を選択します (またはツールバー アイコンをクリックします)。

一部のグラフはすべての検証ツールを使用できますが、1 つのツールのみ受け入れるグラフもあります。検証モードには次の内容が含まれます。

- x 軸の尺度に沿って値を分割するために使用されるバンドの定義および編集。 [詳細は、 p. 392 バンドの使用 を参照してください。](#)
- 四角形の領域内にある値のグループを識別するために使用する領域の定義および編集。 [詳細は、 p. 397 領域の使用 を参照してください。](#)
- 条件抽出ノードまたはフィールド作成ノードを生成するために使用できた値を手作業で選択する要素のマークおよびマーク解除。 [詳細は、 p. 401 マークされた要素の使用 を参照してください。](#)
- バンド、領域、マークされた要素、ストリームで使用する Web リンクによって識別される値を使用したノードの生成。 [詳細は、 p. 402 グラフからのノードの生成 を参照してください。](#)

図 5-89
生成メニューを表示したグラフ



バンドの使用

x 軸上に尺度フィールドを持つグラフでは、垂直なバンドのラインを描いて x 軸上の値の範囲を分割できます。グラフに複数のパネルがある場合、1 つのパネルに書かれたバンドのラインは他のパネルにも同様に表示されます。

すべてのグラフでバンドを使用できるわけではありません。バンドを使用できるグラフには、ヒストグラム、棒グラフ、分布図、プロット（線、散布図、時間など）、コレクション、評価グラフがあります。パネルを含むグラフでは、バンドがすべてのパネルに表示されます。また SPLOM では、フィールド/変数のバンドが描かれた軸が表示されるため、水平なバンドのラインが表示される場合があります。

図 5-90
3つのバンドを使用したグラフ

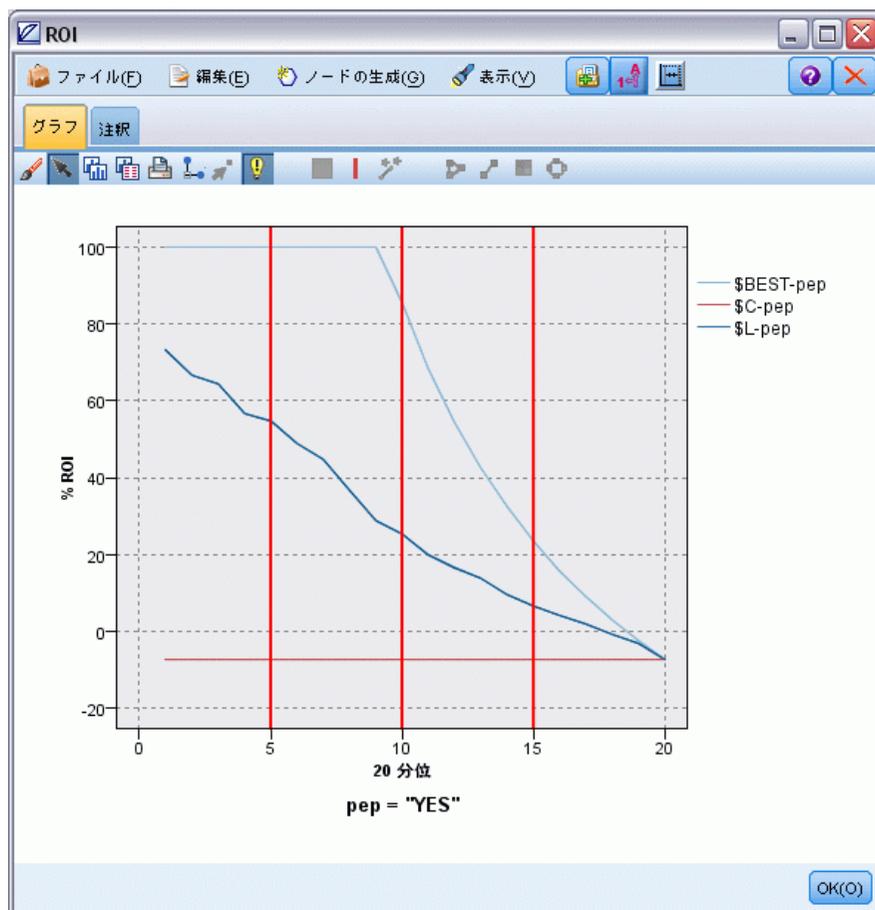
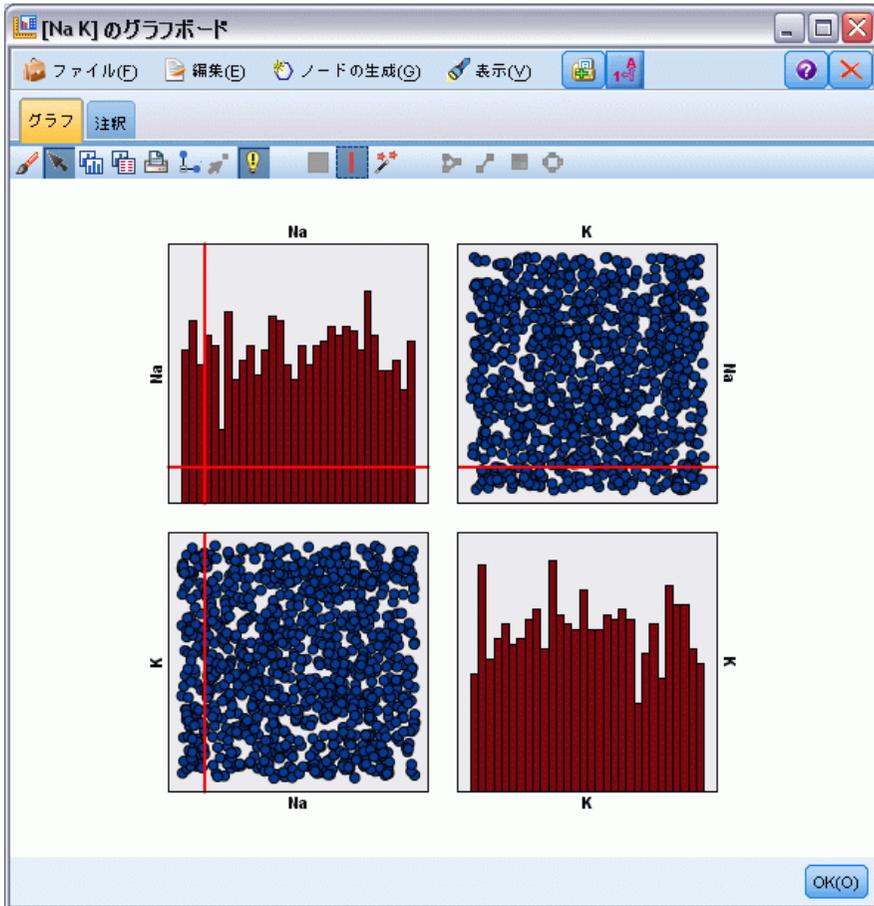


図 5-91
バンドを使用した SPLOM



バンドの定義

バンドのないグラフで、バンドのラインを追加するとグラフを 2 つのバンドに分割します。バンドのラインの値は、下限値としても参照される、グラフを左から右へ読んだ場合の 2 番目のバンドの開始点を表します。同様に 2 つのバンドを使用したグラフでは、バンドのラインを追加するとこれらのバンドの 1 つを 2 つに分割し、3 つのバンドを作成します。デフォルトでは、バンドにはバンドN という名前が付けられます。N には、x 軸上で左から右へ順番にバンド数が割り当てられます。

バンドを定義すると、バンドをドラッグ アンド ドロップして x 軸に再配置できます。バンドの内部を右クリックして、特定のバンドのノードの名前の変更、削除、生成などのタスクのショートカットをさらに表示できます。

バンドを定義する手順は次のとおりです。

- ▶ 検証モードであることを確認します。メニューから [表示] → [検証モード] を選択します。
- ▶ 検証モードのツールバーで、[バンドを描画] ボタンをクリックします。

図 5-92

[バンドを描画] ツールバー ボタン



- ▶ バンドを使用するグラフで、バンドのラインを定義する x 軸の値のポイントをクリックします。

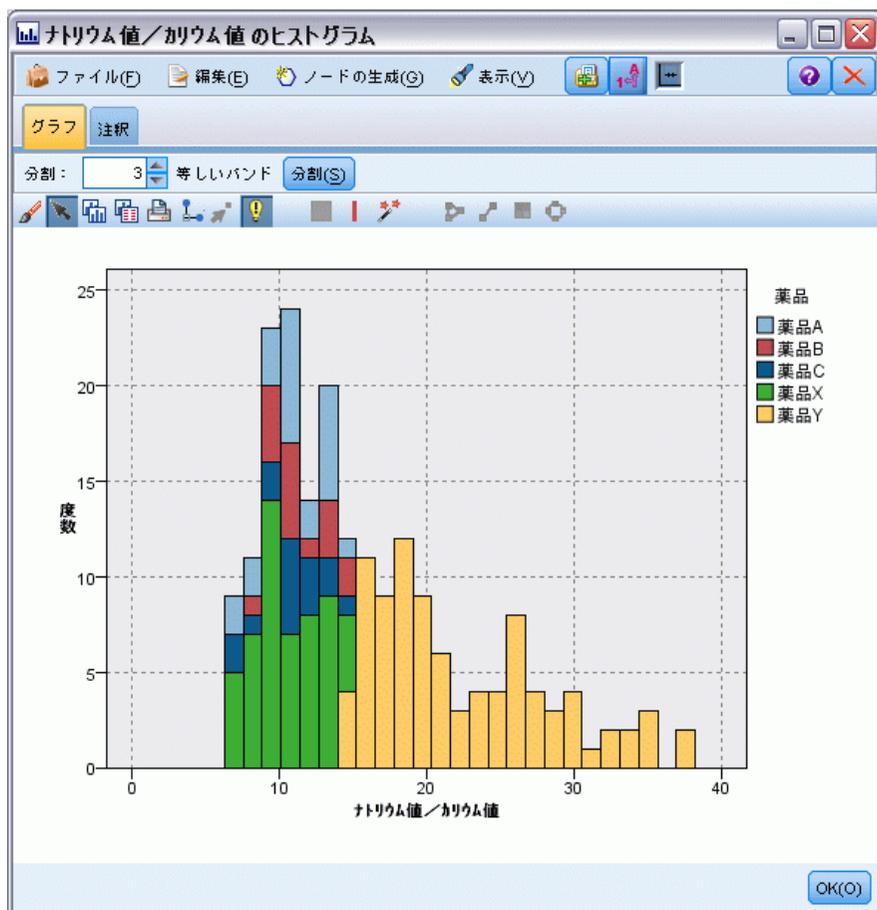
注：代わりに、[グラフをバンドに分割] ツールバー アイコンをクリックして、必要な等しいバンドの数を入力し、[分割] をクリックします。

図 5-93

バンドに分割するオプションを表示する分割アイコン



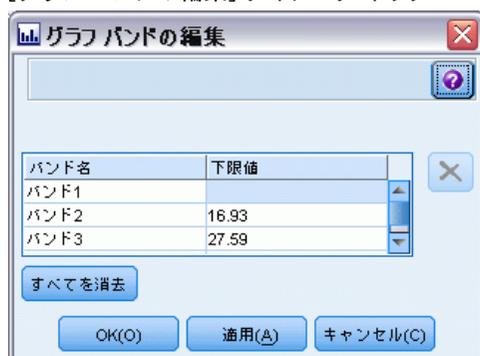
図 5-94
バンドを有効化した、等しいバンドを作成するツールバー



バンドの編集、名前の変更、削除

[グラフバンドの編集] ダイアログ ボックスまたはグラフのコンテキストメニューで既存のバンドのプロパティを編集できます。

図 5-95
[グラフバンドの編集] ダイアログ ボックス



バンドを編集する手順は次のとおりです。

- ▶ 検証モードであることを確認します。メニューから [表示] → [検証モード] を選択します。
- ▶ 検証モードのツールバーで、[バンドを描画] ボタンをクリックします。
- ▶ メニューから [編集] → [グラフバンド] を選択します。[グラフバンドの編集] ダイアログボックスが表示されます。
- ▶ グラフ (SPLOM グラフなど) に複数のフィールドがある場合、必要なフィールドをドロップダウンリストで選択できます。
- ▶ 名前および下限を入力して新しいバンドを追加します。Enter キーを押して、新しい行を開始します。
- ▶ [下限] 値を調整してバンドの境界を編集します。
- ▶ 新しいバンド名を入力して、バンドの名前を変更します。
- ▶ テーブルから削除するラインを選択、[削除] ボタンをクリックして、バンドを削除します。
- ▶ [OK] をクリックして、変更を適用し、ダイアログボックスを閉じます。

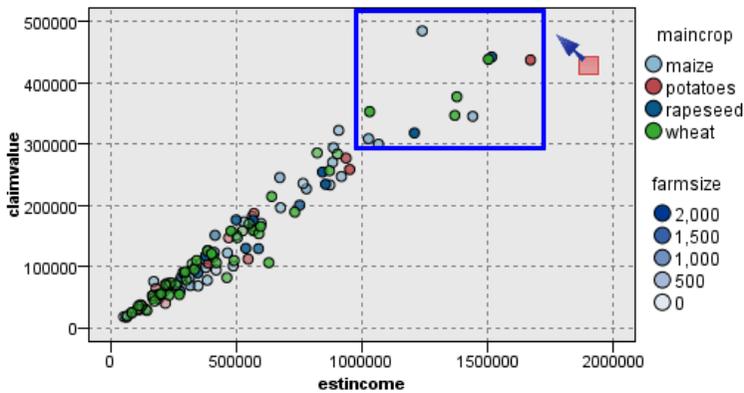
注：代わりに、バンドのラインを右クリックしてコンテキストメニューからオプションを選択し、グラフのバンドを直接削除したり名前を変更することもできます。

領域の使用

2 つの尺度 (または領域) 軸を持つグラフでは、領域を描画して、描画した四角形の領域内に値をグループ化します。**領域**とは、X と Y の最大値と最小値で示されるグラフの領域のことです。グラフに複数のパネルがある場合、1 つのパネルに書かれた領域は他のパネルにも同様に表示されます。

すべてのグラフで領域を使用できるわけではありません。領域を使用できるグラフには、プロット（線、散布図、バブル、時間など）、SPLOM、コレクションがあります。これらの領域は X 領域および Y 領域に描画されます。そのため、1 次元、3 次元またはアニメーション プロットでは定義できません。パネルを含むグラフでは、領域がすべてのパネルに表示されます。散布図行列 (SPLOM) の場合、1 つの尺度フィールドのみ表示されるため、対応する領域は対角プロットでなく対応する上位プロットに表示されます。

図 5-96
申請値の高い領域の定義



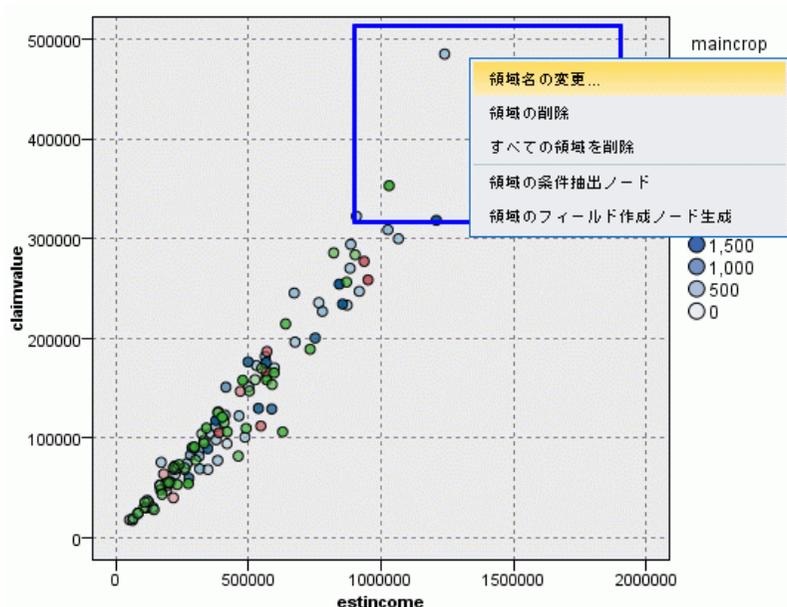
領域の定義

領域を定義する場合、値のグループを作成します。デフォルトでは、新しい領域は領域 <N> と命名されます。N には、すでに作成された領域数に対応した数字が入ります。

領域を定義すると、領域のラインを右クリックして基本のショートカットを取得できます。ただし、ライン上ではなく領域の内部を右クリックして、特定の領域の条件抽出ノードおよびフィールド選択ノードの名前の変更、削除、生成などのタスクのショートカットをさらに表示できます。

特定の領域または複数の領域の 1 つに含まれているかどうかを基準にして、レコードのサブグループを選択できるようになります。また、フィールド作成ノードを作成し、領域に含まれているかどうかを基準にレコードにフラグを設定して、レコードの領域情報を組み込むこともできます。詳細は、[p. 402 グラフからのノードの生成](#) を参照してください。

図 5-97
高い申請値を持つ領域の調査



領域を定義する手順は、次のとおりです。

- ▶ 検証モードであることを確認します。メニューから [表示] → [検証モード] を選択します。
- ▶ 検証モードのツールバーで、[領域を描画] ボタンをクリックします。

図 5-98
[領域を描画] ツールバー ボタン



- ▶ 領域を使用するグラフで、マウスをクリックおよびドラッグして、四角形の領域を描画します。

領域の編集、名前の変更、削除

[グラフ領域の編集] ダイアログ ボックスまたはグラフのコンテキスト メニューで既存のバンドのプロパティを編集できます。

図 5-99
定義した領域のプロパティの指定



領域を編集する手順は、次のとおりです。

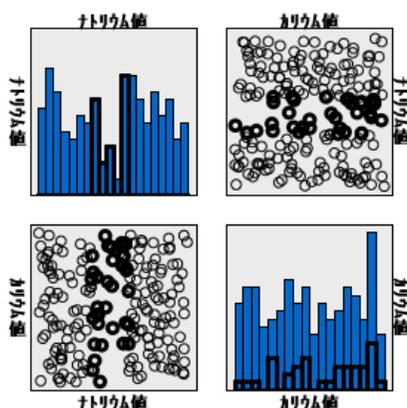
- ▶ 検証モードであることを確認します。メニューから [表示] → [検証モード] を選択します。
- ▶ 検証モードのツールバーで、[領域を描画] ボタンをクリックします。
- ▶ メニューから [編集] → [グラフ領域] を選択します。[グラフ領域の編集] ダイアログ ボックスが表示されます。
- ▶ グラフ (SPLOM など) に複数のフィールドがある場合、[フィールド A] 列および [フィールド B] 列に領域のフィールドを定義する必要があります。
- ▶ 名前を入力、フィールド名を選択 (該当する場合) および各フィールドの上限および下限を定義して、新しいラインに新しい領域を追加します。Enter キーを押して、新しい行を開始します。
- ▶ A および B の [最小] 値および [最大] 値を調整して既存の領域の境界を編集します。
- ▶ テーブル内の領域の名前を変更して、領域名の変更を行います。
- ▶ テーブルから削除するラインを選択、[削除] ボタンをクリックして、領域を削除します。
- ▶ [OK] をクリックして、変更を適用し、ダイアログ ボックスを閉じます。

注：代わりに、領域のラインを右クリックしてコンテキストメニューからオプションを選択して、グラフの領域を直接削除したり名前を変更することもできます。

マークされた要素の使用

グラフ内のバー、スライス、ポイントなどの要素をマークできます。時系列グラフ、線グラフ、評価グラフのラインはフィールドを参照するため、それ以外のグラフのライン、領域、面をマークできません。要素をマークすると、要素で表示されたすべてのデータは基本的に強調表示されます。同じケースが複数の場所に表示されるグラフ（SPLOM など）では、マークはブラシと同義です。バンド内および領域内であっても、グラフの要素をマークできます。要素をマーク後に編集モードに戻っても、マークは表示されたままです。

図 5-100
SPLOM の要素のマーク



グラフ内の要素をクリックして要素をマークおよびマーク解除できます。最初に要素をクリックしてマークすると、要素は境界線の色で濃く表示され、マークされていることが示されます。要素をサイドクリックすると、境界の色が消え、要素のマークが解除されます。複数の要素をマークするには、Ctrl キーを押したまま要素をクリックするか、「マジックワンド」を使用してマークするそれぞれの要素の上でマウスをドラッグします。Ctrl キーを押さずに別の領域または要素をクリックすると、以前マークされた要素はすべてクリアされます。

グラフ内のマークされた要素から条件抽出ノードおよびフィールド作成ノードを生成できます。詳細は、[p. 402 グラフからのノードの生成](#)を参照してください。

要素をマークする手順は、次のとおりです。

- ▶ 検証モードであることを確認します。メニューから [表示] → [検証モード] を選択します。
- ▶ 検証モードのツールバーで、[要素をマーク] ボタンをクリックします。

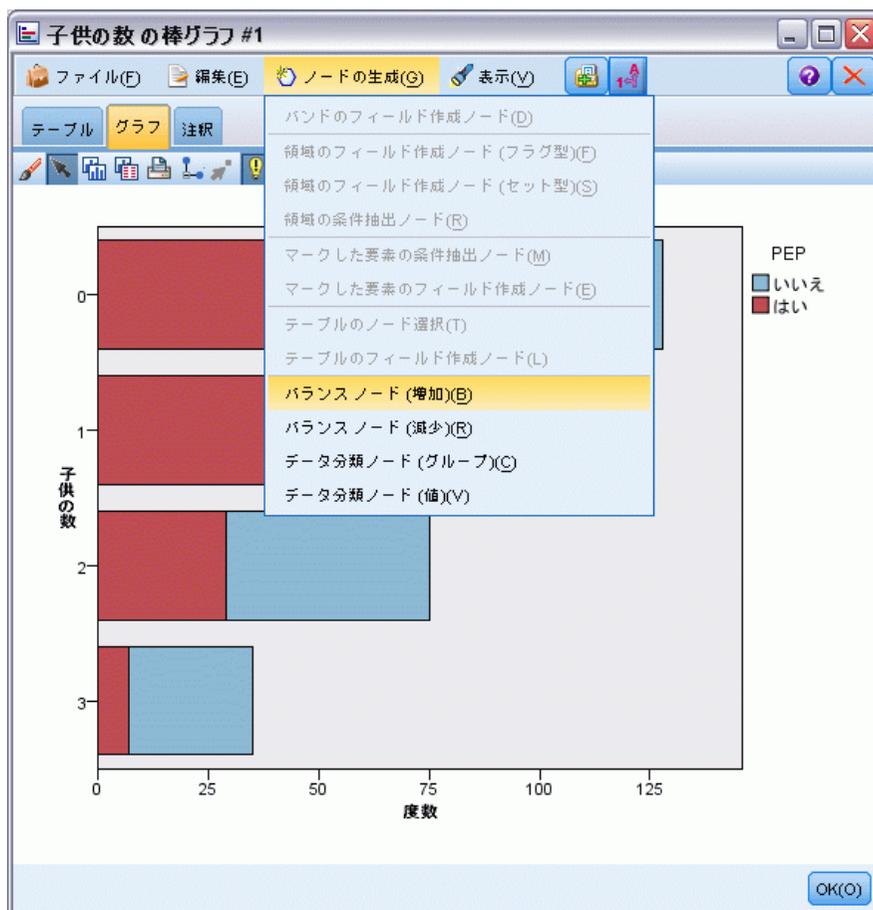
- ▶ 必要な要素をクリックするか、マウスでクリック アンド ドラッグして複数の要素を含む領域周辺に線を描画します。

グラフからのノードの生成

IBM® SPSS® Modeler のグラフで提供される最も強力な機能の 1 つは、グラフまたはグラフ内で選択した要素からノードを生成する機能です。たとえば、時系列グラフから、データを選択するか領域を指定し、効果的にデータの「サブセットを作成」することで、フィールド作成（設定とフラグ）ノードと条件抽出ノードを生成することができます。たとえば、強力なこの機能を使用し、外れ値を識別してそれを除外することができます。

バンドを描画すると、フィールド作成ノードも生成できます。2 つの尺度軸を持つグラフでは、グラフ内に描画された領域からフィールド作成ノードまたは条件抽出ノードを生成できます。マークされた要素を持つグラフでは、これらの要素からフィールド作成ノード、条件抽出ノードを作成でき、またフィルタ ノードを作成できる場合もあります。バランスノード生成は、度数の分布を示すグラフで可能です。

図 5-101
生成メニューを表示したグラフ



ノードを生成すると、ストリーム領域に直接配置され、ノードを既存のストリームに接続できます。条件抽出ノード、フィールド作成ノード、バランスノード、フィルタノード、データ分類ノードはグラフから生成できます。

条件抽出ノード

条件抽出ノードを生成して、領域内のレコードの追加および領域外または下流の処理の逆となる全レコードの除外を検定できます。

- **バンドの場合** : 該当するバンド内のレコードを追加または除外する条件抽出ノードを生成できます。条件抽出ノードで使用するバンドを選択する必要があるため、[バンドのみの条件抽出ノード] はコンテキストメニューでのみ使用できます。

- **領域の場合** :該当する領域内のレコードを追加または除外する条件抽出ノードを生成できます。
- **マークされた要素の場合** :マークされた要素または Web グラフのリンクに対応するレコードを取得する条件抽出ノードを生成できます。

フィールド作成ノード

フィールド作成ノードは、領域、バンド、およびマークされた要素から生成できます。すべてのグラフで、フィールド作成ノードを作成できます。評価グラフの場合、モデル選択のダイアログ ボックスが表示されます。Web グラフの場合、[フィールド作成ノード(「および」)] および [フィールド作成ノード(「または」)] を生成できます。

- **バンドの場合** :[バンドの編集] ダイアログ ボックスにカテゴリ名として表示されたバンド名を使用して、軸上にマークされた各区分のカテゴリを作成するフィールド作成ノードを生成できます。
- **領域の場合** :領域内のレコードには T に設定、全領域外のレコードには F に設定されるフラグを持つ in_region というフラグ型フィールドを作成するフィールド作成ノード ([フィールド作成ノード(フラグ型)]) を生成できます。また、領域と呼ばれる各レコードの新しいフィールドを含む、各領域の値を持つセットを作成するフィールド作成ノード ([フィールド作成ノード(セット型)]) も生成できます。そのフィールド作成ノードではレコードが分類される領域の名前を値として取得します。どの領域にも属さないレコードの場合は、デフォルトの領域名が表示されます。値の名前は、[領域の編集] ダイアログ ボックスに表示された領域名となります。
- **マークされた要素の場合** :マークされたすべての要素は真、それ以外のすべてのレコードは偽であるフラグを集計するフィールド作成ノードを生成できます。

バランス ノード

バランス ノードを生成して、共通する値の頻度を減らす ([バランス ノード(減少)] メニュー オプションの使用)、またはあまり出現しない値の発生を増加させる ([バランス ノード(増加)] メニュー オプションの使用) など、データの不均衡を修正できます。バランス ノードの生成は、ヒストグラム、点グラフ、集計グラフ、度数の棒グラフ、度数の円グラフ、および線グラフなど、度数の分布を示すグラフで有効です。

フィルタ ノード

フィルタ ノードを生成して、グラフ内でマークされたラインまたはノードに基づいてフィールドの名前を変更またはフィルタリングできます。評価グラフの場合、最良の適合線はフィルタ ノードを生成しません。

データ分類ノード

データ分類ノードを生成して、値を再分類できます。このオプションは、分布図に使用されます。グループ中の選択内容に応じて (Ctrl + [テーブル] タブでグループを選択)、表示されたフィールドの特定の値を再分類する**グループ**のデータ分類ノードを生成できます。また、**値**のデータ分類ノードを生成して、各種会社の金融データを分析のために結合するためにデータを標準的な値のセットに再分類するなど、多数の値からなる既存のセットにデータを再分類できます。

注：値があらかじめ定義されている場合、SPSS Modeler にフラット ファイルとして読み込み、棒グラフを使ってすべての値を表示できます。次に、このフィールドのデータ分類 (値) ノードをチャートから直接生成します。これにより、すべての対象値を分類ノードの [新しい値] 列 (ドロップダウン リスト) に分類します。

グラフからのノードの生成

グラフ出力ウィンドウの [ノードの生成] メニューを使用してノードを生成できます。生成されたノードはストリーム領域に配置されます。このノードを使用するには、これを既存のストリームと接続します。

グラフからノードを生成する手順は、次のとおりです。

- ▶ 検証モードであることを確認します。メニューから [表示] → [検証モード] を選択します。
- ▶ 検証モードのツールバーで、[領域] ボタンをクリックします。
- ▶ ノードの定義に必要なバンド、領域、およびマークされた要素を定義します。
- ▶ [ノード生成] メニューから、作成するノードの種類を選択します。作成できるノードのみが有効です。

注：代わりに、右クリックした後、コンテキスト メニューから該当する生成オプションを選択して、グラフから直接ノードを生成することもできます。

視覚化の編集

探索的分析モードでは、視覚化によって表現されたデータや値を分析的に検討することができます。一方、編集モードでは、視覚化のレイアウトや外観を変更することができます。たとえば、フォントおよび色を変更して、会社のスタイル ガイドに準拠することができます。編集モードを選択するには、メニューから [表示] → [編集モード] を選択します (またはツールバー アイコンをクリックします)。

編集モードには、視覚化のレイアウトのさまざまな要素に影響を与えるいくつかのツールバーがあります。使用しないツールバーがある場合は、ツールバーを隠し、グラフが表示されるダイアログ ボックス内のスペースを大きくすることができます。ツールバーを選択または選択を解除するには、[表示] メニューの対応するツールバー名をクリックします。

注：視覚化に更なる詳細を追加するには、タイトル、脚注、軸ラベルを適用することができます。詳細は、[p. 425 表題と脚注の追加](#) を参照してください。

編集モードには、視覚化を編集するためのオプションがいくつか用意されています。以下を行うことができます。

- テキストの編集および書式の設定。
- 塗りつぶしの色、透明度、および枠やグラフ要素のパターンの変更。
- 境界線と線の色とダッシュの形の変更。
- ポイント要素の回転、形状および縦横比の変更。
- グラフィック要素（棒やポイントなど）のサイズの変更。
- 境界やパディングを使用した項目の周囲のスペースの調整。
- 番号の書式の指定。
- 軸とスケールの設定の変更。
- カテゴリ軸のカテゴリのソート、除外、終了。
- パネルの方向の設定。
- 座標システムの変換。
- 統計量、グラフ要素の種類、および衝突変更子の変更。
- 凡例の位置の変更。
- 視覚化スタイルシートの適用。

次のトピックでは、これらのさまざまなタスクを実行する方法を説明します。また、グラフの編集に関する一般的なルールもお読みください。

編集モードに切り替える方法

- ▶ メニューから次の項目を選択します。
表示 > 編集モード

視覚化編集の一般的な規則

編集モード

すべての編集は編集モードで行われます。編集モードを有効にするには、メニューから次の項目を選択します。

表示 > 編集モード

選択

編集で使用可能なオプションは、選択によって異なります。何が選択されたかに応じて、異なるツールバーおよびプロパティ パレットのオプションが有効になります。有効にされた項目のみが現在の選択に適用されます。たとえば、軸を選択した場合は、プロパティ パレットで [スケール]、[大分割の目盛り]、および [小分割の目盛り] タブが有効になります。

視覚化で項目を選択する際のヒントをいくつか示します。

- 項目をクリックして選択します。
- 1 回クリックして、(散布図内のポイントまたは棒グラフ内の棒など) グラフィック要素を選択します。最初の選択の後で、もう一度クリックしてグラフ要素のグループまたは単一のグラフ要素に選択を絞り込みます。
- Esc キーを押すと、すべてが選択解除されます。

Palettes

視覚化で項目が選択されている場合、さまざまなパレットが選択状態を反映するために更新されます。パレットには、選択に対する編集を行うコントロールが含まれます。パレットはツールバーの場合と制御用の UI とタブを持ったパネルの場合があります。パレットは非表示にすることができますので、編集時に必要なパレットが表示されているか確認してください。[表示] メニューで、現在表示されているパレットを確認してください。

ツールバー パレットまたは他のパレットの左側で空いている場所をクリックしてからドラッグして、パレットを再配置することができます。視覚的な反応により、パレットを移動することができる場所が示されます。ツールバー以外のパレットの場合は、閉じるボタンをクリックしてパレットを非表示にしたり、取り外しボタンをクリックしてパレットを別のウィンドウに表示することもできます。ヘルプボタンをクリックすると、そのパレットのヘルプが表示されます。

自動設定

一部の設定には、[-auto-] オプションがあります。これは、自動的な値が適用されることを示しています。どの自動設定が使用されるかは、それぞれの視覚化およびデータ値により異なります。値を入力して、自動設定を上書きすることもできます。自動設定値を復元する場合は、現在の値を削除し、Enter キーを押します。設定には [-auto-] が再び表示されます。

項目の削除/非表示

視覚化では、さまざまな項目の除外/非表示を行うことができます。たとえば、凡例または軸ラベルを隠すことができます。項目を削除するには、選択して Del キーを押します。削除が許可されない項目の場合は、何も起きません。誤って項目を削除した場合は、Ctrl + Z キーで、削除を取り消すことができます。

State

いくつかのツールバーでは、現在の選択の状態が反映されますが、反映されないものもあります。プロパティ パレットは常に状態を反映します。ツールバーが状態を反映しない場合は、ツールバーを説明するトピックにこの点が説明されています。

テキストの編集と書式設定

テキストはその場で編集でき、テキスト ブロック全体のフォーマットを変更できます。ただし、データ値に直接リンクされているテキストは編集できません。たとえば、ラベルの内容は元となるデータから引き出されるので、目盛ラベルの文字列を編集することはできません。しかし、視覚化ではすべてのテキストの書式設定が可能です。

その場でテキストを編集する方法

- ▶ テキスト ブロックをダブルクリックします。これで、すべてのテキストが選択されます。テキストを編集している間は視覚化のほかの部分を変更することはできないので、すべてのツールバーが無効になります。
- ▶ 入力して既存のテキストを置き換えます。テキストをもう一度クリックすると、カーソルを表示できます。カーソルを適切な位置に合わせて、追加のテキストを入力します。

テキストのフォーマットの方法

- ▶ テキストを含む枠を選択します。テキストをダブルクリックしないでください。

- ▶ フォント ツールバーを使用してテキストをフォーマットします。ツールバーが有効になっていない場合は、テキストを含む「枠」だけを確実に選択します。テキスト自体を選択した場合は、ツールバーが無効になります。

図 5-102
フォント ツールバー



次のフォントを変更できます。

- 色
- ファミリー (Arial や Verdana など)
- サイズ (pc などの異なる単位を示さないかぎり、単位は pt)
- 太さ (W)
- テキスト枠に対する位置合わせ

書式は枠の中のすべてのテキストに適用されます。テキストの特定ブロック内の個々の文字または単語の書式は変更できません。

色、パターン、破線化、および透明度の変更

視覚化では、さまざまな項目が、塗りつぶしと枠線を持っています。最もわかりやすい例は、棒グラフの棒です。棒の色は塗りつぶし色です。棒には、棒を取り囲む黒い境界線があります。

視覚化では、そのほかにも塗りつぶし色を持った項目があります。塗りつぶし色が透明であると、そこに塗りつぶし色があることがわかりません。たとえば、軸ラベル内のテキストを考えてみてください。このテキストは「浮動」テキストであるかのように表示されますが、実際には透明の塗りつぶし色の枠内に表示されるのです。軸ラベルを選択すると、その枠を表示できます。

視覚化全体を囲んでいる枠を含めて、視覚化におけるすべての枠は、塗りつぶし色と枠線のスタイルを持つことができます。また、すべての塗りつぶしには、関連する調節可能な不透明/透明レベルがあります。

色、パターン、破線化、および透明度の変更する方法

- ▶ フォーマットする項目を選択します。たとえば、棒グラフ内の棒またはテキストを含む枠を選択します。視覚化がカテゴリ変数やフィールドで分割されている場合、個別のカテゴリに対応するグループを選択することもできます。これで、グループに割り当てられたデフォルトの外観を変更できるようになります。たとえば、積み上げ棒グラフ内の積み上げグループの 1 つの色を変更できます。

- ▶ 塗りつぶし色、境界線色、または塗りつぶしパターンを変更するには、カラー ツールバーを使用します。

図 5-103
色ツールバー

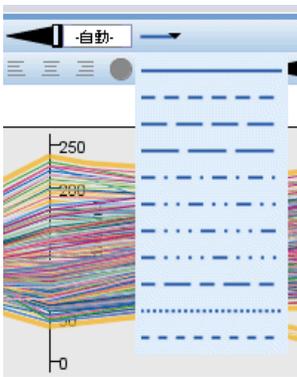


注： このツールバーは、現在の選択のステート（状態）を反映しません。

色や塗りつぶしを変更するために、ボタンをクリックして表示されているオプションを選択したり、ドロップダウンの矢印をクリックして別のオプションを選択したりすることができます。色については、白地に赤い対角線が通っているように見える色が 1 つあります。これは透明色です。たとえばこれを使用して、ヒストグラム内の棒の境界線を隠すことができます。

- 最初のボタンは、塗りつぶし色をコントロールします。
 - 2 番目のボタンは、枠線の色をコントロールします。
 - 3 番目のボタンは、塗りつぶしパターンをコントロールします。塗りつぶしパターンには境界線の色を使用します。したがって、塗りつぶしパターンは境界線の色が目に見える色のときのみ、表示されます。
 - 4 つ目のコントロールは、塗りつぶし色とパターンの非透明度を制御するスライダーとテキストボックスです。このパーセントが低いと非透明度が低く、透明度が高いことになります。100% の場合は完全に非透明です。
- ▶ 枠線または線のダッシュを変更するには、ライン ツールバーを使用します。

図 5-104
ライン ツールバー



注： このツールバーは、現在の選択のステート（状態）を反映しません。

ほかのツールバーと同様に、ボタンをクリックして表示されたオプションを選択するか、ドロップダウン矢印をクリックして別のオプションを選択します。

ポイント要素の回転と、形状および縦横比の変更

ポイント要素を回転したり、さまざまな定義済みの形状を割り当てたり、縦横比（高さに対する幅の割合）を変更したりすることができます。

ポイント要素の変更方法

- ▶ ポイント要素を選択します。個々のポイント要素の形状と縦横比は、回転または変更できません。
- ▶ マーカー ツールバーを使用して、ポイントを修正します。

図 5-105
マーカー ツールバー



- 最初のボタンで、ポイントの形状を変更できます。ドロップダウン矢印をクリックして、定義済みの形を選択します。
- 2番目のボタンを使用すると、ポイントを特定のコンパス位置に回転できます。ドロップダウンの矢印をクリックし、希望の位置へコンパスの針をドラッグします。
- 3番目のボタンを使用すると、縦横比を変更できます。ドロップダウンの矢印をクリックし、表示された四辺形をクリックアンドドラッグします。四辺形の形状で縦横比が表現されます。

グラフィック要素のサイズの変更

視覚化では、グラフ要素のサイズを変更することができます。変更できるものは、棒、線、およびポイントです。グラフィック要素のサイズが変数またはフィールドで定められている場合、指定されているサイズは「最小」の値です。

グラフィック要素のサイズの変更方法

- ▶ サイズ変更するグラフィック要素を選択します。
- ▶ スライダを使用するか、マーカー ツールバーにあるオプションに特定のサイズを入力します。異なる単位が示されていないかぎり、単位はピクセルです（単位の省略形については下記を参照してください）。また、パーセンテージ（30% など）も指定できます。これは、グラフィック要素が使用可能なスペースの中で指定されたパーセンテージを使用することを意味します。利用可能な領域は、グラフ要素の種類や個々の視覚化によって異なります。

テーブル 5-3
有効な単位の省略形

| 略語 | 単位 |
|----|---------|
| cm | センチメートル |
| 入力 | インチ |
| mm | ミリメートル |
| pc | パイカ |
| pt | ポイント |
| px | ピクセル |

図 5-106
シンボル ツールバーでのサイズの制御



余白とパディングの指定

視覚化で枠の周りや内部に空白がありすぎる場合や、逆になさ過ぎる場合、余白とパディングの設定を変更することができます。余白は、枠とそれを取り囲む他の項目との間の空間量です。パディングは、枠線と枠の「内容」との間の空間量です。

余白およびパディングの指定方法

- ▶ 余白およびパディングを指定する枠を選択します。これは、テキスト枠、凡例を囲む枠、または棒やポイントなどのグラフィック要素を表示するデータ枠のいずれであってもかまいません。
- ▶ プロパティ パレット上の [余白] タブを使用して、設定を指定します。cm または in などの別の単位を指定しないかぎり、すべてのサイズはピクセル数です。

図 5-107
[余白] タブ



数値の書式設定

連続軸の目盛りラベルや、数値を表示するデータ値ラベルの数値の書式を指定できます。たとえば、目盛りラベルに表示される数値が 1000 単位になるように設定することができます。

数値書式を指定する方法

- ▶ 数値を含む連続軸の目盛りラベル、またはデータ値ラベルを選択します。
- ▶ プロパティ パレットで [書式] タブをクリックします。

図 5-108
[書式] タブ

| 尺度 | 余白 | 大分割の目盛り | 小分割の目盛り | 形式 |
|------|----|---------------|---------------|--|
| 接頭辞: | | 整数の最小桁数: -自動- | 少数の最小桁数: -自動- | 科学的表記: -自動- <input type="checkbox"/> -ve の括弧 |
| 接尾辞: | | 整数の最大桁数: -自動- | 少数の最大桁数: -自動- | 尺度変更: -自動- <input checked="" type="checkbox"/> グループ化 |

- ▶ 目的の数値書式設定オプションを選択します。

接頭辞 : 数値の先頭に表示する文字。たとえば、数値が U.S. ドルを通貨とする給与を表している場合は、ドル記号 (\$) を入力します。

接尾辞 : 数値の最後に表示する文字。たとえば、数値がパーセントを表している場合は、パーセント記号 (%) を入力します。

整数の最小桁数 : 小数を含む数字の表記の整数部分として表示される最小桁数。実際の値に最小桁数分の桁がない場合、値の整数部分には 0 が追加されます。

整数の最大桁数 : 小数を含む数字の表記の整数部分として表示される最大桁数。実際の値が最大桁数より大きい桁の場合、値の整数部分はアスタリスク (*) により置き換えられます。

小数の最小桁数 : 小数を含む数字の表記または科学的表記の小数部分として表示される最小桁数。実際の値に最小桁数分の桁がない場合、値の小数部分には 0 が追加されます。

小数の最大桁数 : 小数を含む数字の表記または科学的表記の小数部分として表示される最大桁数。実際の値が最大桁数より大きい桁の場合、値の小数は適切な桁数に丸められます。

科学的 : 数値を科学的表記法で表示するかどうか。科学的表記は、非常に大きなまたは小さな数値に便利です。-auto- を使用すると、アプリケーションが科学的表記が適切かどうか判定します。

尺度変更 : 元の値を分割する数値である、目盛りの単位です。数値が大きく桁数が多いため、目盛りラベルが長くなるのを避けたいときに、目盛りの単位を変更できます。目盛りラベルの数値の書式を変更する場合、軸の表題を編集して、数値をどのように解釈すべきかを説明してください

い。たとえば、スケール軸に給与を表示するケースで、ラベルが 30,000、50,000、および 70,000 にあるとします。目盛りの単位として「1000」を入力すると、30、50、70 と表示されます。その際、スケール軸の表題を編集し、「1000 ドル単位」というテキストを含めるようにします。

-ve の括弧 : マイナスの値を表示するときに括弧で囲むかどうか。

グループ化 : 数値の桁区切り文字を表示するかどうか。ユーザーのコンピュータの現在のロケールにより、数値のグループ化に使用される文字が決定されます。

軸とスケール設定の変更

軸とスケールの修正には複数のオプションがあります。

軸およびスケール設定の変更方法

- ▶ 軸の任意の部分を選択します（軸ラベルや軸ラベル文字列など）。
- ▶ プロパティ パレット上の [スケール]、[大分割の目盛り]、および [小分割の目盛り] タブを使用して、軸およびスケール設定を変更します。

図 5-109
プロパティ パレット



[スケール] タブ

注 : データが事前に集計されていない場合、グラフの [スケール] タブは表示されません（ヒストグラムなど）。

タイプ : スケールが直線状か、または変換されているかを指定します。スケールを変換すると、データの理解、または統計的な推測に必要な想定役に立ちます。散布図では、独立および従属の変数またはフィールド間の関係が線状でない場合に、変換したスケールが使用される可能性があります。スケールの変換は、歪んだヒストグラムをより対称的にして正規分布に似るようにするために使用できます。ただし、データが表示されているスケールのみを変換できるのであって、実際のデータを変換するものではありません。

- **線型**：線型の変換されていないスケールを指定します。
- **対数**：常用対数に変換したスケールを指定します。0 および負の値を扱うために、この変換では対数関数の修正版を使用します。” safe log” 関数は $\text{sign}(x) * \log(1 + \text{abs}(x))$ と定義されます。そのため、`safeLog(-99)` は次の式と等しくなります。

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$

- **べき乗**：0.5 のべき乗を使用して、べき乗変換スケールを使用します。負の値を扱うために、この変換ではべき乗関数の修正版を使用します。” safe log” 関数は $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$ と定義されます。そのため、`safePower(-100)` は次の式と等しくなります。

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

最小/最大/適切な下限/適切な上限：スケールの範囲を指定します。[下限バンドの最適化] および [上限バンドの最適化] を選択すると、データに基づいて適切なスケールが選択されます。通常、最大および最小の実際のデータ値よりも大きいか小さい包括的な値であるため、最小値および最大値は、「適切な」値です。たとえば、データ範囲が 4 から 92 までの場合、スケールの下限および上限バンドの適切化の値は実際データの最小値と最大値ではなく、0 と 100 にすることができます。小さすぎて重要な項目を隠すような範囲を設定しないように注意してください。また、[ゼロを含める] オプションが選択された場合は、明示して最大値と最小値を設定できないことにも注意してください。

下の境界/上の境界：軸の下端/上端に境界を作成します。境界は、選択された軸に直角の位置で表示されます。cm または in などの別の単位を指定しないかぎり、単位はピクセルです。たとえば、垂直軸の [上の境界] を 5 に設定すると、5 ピクセルの水平の境界線がデータ枠の一番上に沿って表示されます。

逆：スケールが逆に表示されるかどうかを指定します。

ゼロを含める：尺度に 0 が含まれる必要があることを示します。このオプションは一般的に、一番短い棒の最上部付近から値が始まるのではなく、0 から棒が始まるように棒グラフで使用されます。このオプションを選択した場合は、スケール範囲にユーザー指定の最小値および最大値は設定できなくなるために、[最小] および [最大] は無効になります。

[大分割の目盛り] タブ/[小分割の目盛り] タブ

目盛りまたは目盛りマークは軸に表示される線です。これらは、特定の間隔またはカテゴリで値を示します。**大分割の目盛り**は、ラベル付きの目盛りです。これらはほかの目盛りマークよりも長くなります。**小分割の目盛り**は、大分割の目盛り間に表示される目盛りです。一部のオプションは目盛りの種類に特有なものですが、ほとんどのオプションは大分割の目盛りにも小分割の目盛りにも利用できます。

目盛りの表示：大分割または小分割の目盛りをグラフに表示するかどうか指定します。

グリッドの表示：大分割または小分割の目盛りにグリッド線が表示されるかどうかを指定します。**グリッド線**は、軸から軸までグラフ全体に引かれる線です。

位置：軸に相対的な目盛りマークの位置を指定します。

長さ：目盛りマークの長さを指定します。cm または in などの別の単位を指定しないかぎり、単位はピクセルです。

基点：大分割の目盛りだけに適用されます。最初の大分割目盛りに表示される値を指定します。

デルタ：大分割の目盛りだけに適用されます。大分割目盛り間の差を指定します。つまり n をデルタ値とすると、大分割の目盛りは n 番目の値ごとに表示されます。

分割：小分割の目盛りだけに適用されます。大分割目盛り間の小分割の数を指定します。小分割の目盛り数は分割数の数より 1 つ少なくなります。たとえば、大分割の目盛りが 0 と 100 に存在しているとします。「2」を小分割の目盛りによる区分の数として指定すると、実際には小分割の目盛りが 50 に 1 つ追加され、0 ~ 100 の範囲を 2 つの区分に分割します。

カテゴリの編集

カテゴリ軸のカテゴリを、次のような方法で編集することができます。

- カテゴリを表示するソート順を変更する。
- 特定のカテゴリを除外する。
- データセットに出現しないカテゴリを追加する。
- 小さなカテゴリを 1 つのカテゴリに集約または結合する。

カテゴリのソート順の変更方法

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットに軸上のカテゴリを表示します。

注：パレットが表示されない場合、有効化されていることを確認してください。IBM® SPSS® Modeler の [表示] メニューから、[カテゴリ] を選択します。

- ▶ [カテゴリ] パレットで、ドロップダウン リストからソートのオプションを選択します。

ユーザー指定：パレットに表示された順序に基づいてカテゴリをソートします。矢印ボタンを使用して、カテゴリをリストの最上部、上下、最下部に移動します。

データ：データセットの発生する順序に基づいてカテゴリをソートします。

名前：パレットに表示された名前を使用して、アルファベット順にカテゴリをソートします。ツールバーが値やラベルのどちらを表示するかを選択によって、値またはラベルのいずれかとなります。

値：パレットにある括弧内に表示された値を使用して、基礎となるデータの値順にカテゴリを並び替えます。メタデータのあるデータ ソース (IBM® SPSS® Statistics データ ファイルなど) のみがこのオプションをサポートします。

統計値：各カテゴリの計算された統計値に基づいてカテゴリをソートします。統計値には、度数、パーセンテージ、平均値が含まれます。このオプションは、統計値がグラフに使用されている場合にのみ使用できます。

カテゴリを追加する方法

デフォルトでは、データセットに出現するカテゴリのみが使用できます。必要に応じて、視覚化にカテゴリを追加することができます。

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットに軸上のカテゴリを表示します。

注：パレットが表示されない場合、有効化されていることを確認してください。SPSS Modeler の [表示] メニューから、[カテゴリ] を選択します。

- ▶ [カテゴリ] パレットで、[カテゴリの追加] ボタンをクリックします。

図 5-110
[カテゴリの追加] ボタン



- ▶ [新しいカテゴリの追加] ダイアログ ボックスで、カテゴリ名を入力します。
- ▶ [OK] をクリックします。

特定のカテゴリの除外方法

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットに軸上のカテゴリを表示します。

注：パレットが表示されない場合、有効化されていることを確認してください。SPSS Modeler の [表示] メニューから、[カテゴリ] を選択します。

- ▶ [カテゴリ] パレットで、[含める] リストのカテゴリ名を選択し、[X] ボタンをクリックします。カテゴリに戻すには、[除外] リストで名前を選択し、リストの右側にある矢印をクリックします。

小さなカテゴリを集約または結合する方法

個別に表示する必要のない小さいカテゴリを結合することができます。たとえば、円グラフにカテゴリがたくさんあるのであれば、10 パーセント未満のカテゴリを集約することを検討してみてください。集約できるのは、相加的な統計のみです。例えば、平均値は相加的でないため、平均値同士を合計することはできません。したがって、平均値を使用してカテゴリを結合または集約することはできません。

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットに軸上のカテゴリを表示します。

注：パレットが表示されない場合、有効化されていることを確認してください。SPSS Modeler の [表示] メニューから、[カテゴリ] を選択します。

- ▶ [カテゴリ] パレットで、[閉じる] を選択し、割合を指定します。合計の割合が指定された数値を下回るカテゴリは、1 つのカテゴリに結合されます。割合は、グラフに表示された統計値に基づきます。度数および合計の統計量でのみ、カテゴリを閉じることができます。

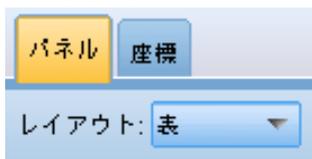
方向パネルの変更

視覚化でパネルを使用している場合、パネルの方向を変更することができます。

パネルの方向の変更方法

- ▶ 視覚化の任意の一部を選択します。
- ▶ プロパティ パレットで [パネル] タブをクリックします。

図 5-111
[パネル] タブ



- ▶ [レイアウト] から次のオプションを選択します。

テーブル： パネルをテーブルのように並べます。そこには、個々の値に割り当てられた行または列があります。

行列入替：パネルをテーブルのように並べますが、元の行と列を入れ替えます。このオプションは、グラフ自体の行列入れ替えと同じではありません。このオプションを選択したときは、x 軸と y 軸は変更されないことに注意してください。

リスト：パネルをリストのように並べ、各セルが値の組み合わせを表します。列と行は、もう個々の値に割り当てられません。このオプションを使用すると、必要に応じてパネルが折り返されます。

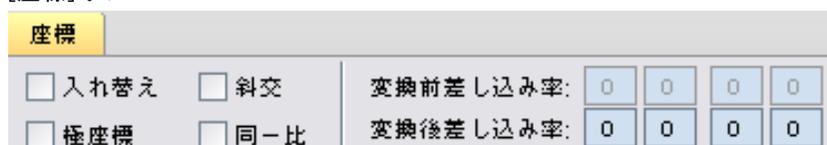
座標システムの変換

多くの視覚化は平面、長方形、または座標システムで表示されます。座標システムは、必要に応じて変換することができます。たとえば、極座標変換を座標システムに適用して、斜めの影付き効果を追加して、軸を置き換えることができます。これらの変換が現在の視覚化に既に適用されている場合は、変換を取り消すこともできます。たとえば、円グラフが極座標システムに描かれているとします。必要に応じて、極変換を取り消して、円グラフを単一の積み上げ棒グラフとして長方形の座標システムに表示することができます。

座標システムを変換する方法

- ▶ 変換する座標システムを選択します。個別のグラフの周りの枠を選択して、座標システムを選択します。
- ▶ プロパティ パレットで [座標] タブをクリックします。

図 5-112
[座標] タブ



- ▶ 座標システムに適用したい変換を選択します。変換を取り消すために、変換の選択を解除することもできます。

置き換え：軸の方向の変更は、**行と列の入れ換え**と呼ばれます。これは 2-D の視覚化で縦と横の軸を入れ替えるのと似ています。

極座標：極座標変換は特定の角度およびグラフの中心部からの特定の距離で、グラフ要素を描きます。円グラフは、個々の棒を特定の角度で描いた極変換による 1-D の視覚化です。レーダー図は、特定の角度およびグラフの中心部からの特定の距離でグラフ要素を描いた、極変換による 2-D の視覚化です。3-D の視覚化には、奥行き表す次元が追加されます。

斜交：斜交変換は 3-D 効果をグラフ要素に追加します。この変換によりグラフ要素に奥行きが追加されますが、この奥行きは純粹に裝飾的なものです。この奥行きが特定のデータ値の影響を受けることはありません。

同一比：同じ比率を適用すると、それぞれの尺度上での同じ距離がデータ値における同じ差異を指定するようになります。たとえば、両尺度における 2 cm はデータ値での 1000 の差をあらわします。

変換前差し込み率：変換後に軸がクリップしている場合は、変換前に差し込みをグラフに追加することができます。座標システムに変換が適用される前に、差し込みによって各次元が一定のパーセント縮みます。最小 x、最大 x、最小 y、最大 y の順番で各次元を制御することができます。

変換後差し込み率：グラフの縦横比を変更する場合は、変換の適用後にグラフを差し込みを追加することができます。座標システムに変換が適用された後に、差し込みによって各次元が一定のパーセント縮みます。グラフが変換されていない場合でも、これらの差し込みを適用することができます。最小 x、最大 x、最小 y、最大 y の順番で各次元を制御することができます。

統計量とグラフ要素の変更

を別の種類に変換したり、グラフ要素の描画に使用される統計量を変更したり、グラフ要素が重なった際に何が起こるのかを決定する衝突変更子を指定したりすることができます。

グラフ要素を変換する方法

- ▶ 変換したいグラフ要素を選択します。
- ▶ プロパティ パレットで [要素] タブをクリックします。

図 5-113
[要素] タブ



- ▶ [種類] リストから新しいグラフ要素の種類を選択します。

| グラフ要素の種類 | 説明 |
|----------|--|
| ポイント | 特定のデータ ポイントを示すマーカー。ポイント要素は、散布図やその他関連する視覚化で使用されます。 |
| 区間 | 特定のデータ値で描画され、始点と別のデータ値との間の領域を埋める四角形の領域。区間要素は、棒グラフおよびヒストグラムで使用されます。 |

| グラフ要素の種類 | 説明 |
|----------|--|
| 線 | データ値を接続する線。 |
| パス | データセットに表示された順にデータ値を接続する線。 |
| 領域 | 線と始点の間の領域を満たし、データ要素を接続する線。 |
| 多角形 | データ領域を囲む多角形。多角形要素は、分割された散布図またはマップで使用できます。 |
| スキーマ | 外れ値を示すひげとマーカーを持つボックスで構成された要素。スキーマ要素はボックスプロットに使用されます。 |

統計量を変更する方法

- ▶ 統計量を変更したいグラフ要素を選択します。
- ▶ プロパティ パレットで [要素] タブをクリックします。
- ▶ [要約] ドロップダウン リストから新しい統計量を選択します。統計量を選択するとデータが集計されることに注意してください。集計されていないデータを視覚化で表示するには、[要約] リストから[(統計なし)]を選択します。

連続型フィールドから計算された要約統計

- **Mean.** 中心傾向の測定値。観測値の合計をケース数で割った算術平均。
- **Median.** ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。
- **モード.** 最も多く出現する値。複数の値が最高の頻度で出現する場合は、それぞれが最頻値となります。
- **最小.** 数値型変数にとる最も小さい値。
- **最大.** 数値型変数の最大値。
- **範囲:** 最小値と最大値の差異。
- **中間域:** 範囲の中間、つまり最小値との差が最大値との差と等しい値です。
- **Sum.** 欠損値のないすべてのケースに対する変数の値の合計または全体。
- **累積合計:** 値の累積合計。各グラフィック要素は、サブグループの合計と、以前のグループすべての総合計を示します。
- **パーセント合計:** 合計したフィールドに基づいたサブグループ内の、全グループの合計に対するパーセンテージ。

- **累積パーセント合計** : 合計したフィールドに基づいたサブグループ内の、全グループの合計に対する累積パーセンテージ。各グラフィック要素は、サブグループのパーセンテージと、以前のグループすべての総パーセンテージを示します。
- **Variance.** 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **標準偏差.** 平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **標準誤差.** サンプル間で検定統計量の値がどの程度ばらついているかを測ったもの。ある統計量のサンプル分布の標準偏差です。たとえば、平均値の標準誤差はサンプル平均の標準偏差です。
- **尖度.** 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。
- **歪度.** 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

次の領域の統計で、サブグループごとに複数のグラフィック要素が作成される場合があります。区間、領域、または辺のグラフィック要素を使用する場合、領域統計では、範囲を示すグラフィック要素が作成されます。他のすべてのグラフィック要素は、2 つの要素を生成します。一方は範囲の始点を示し、もう一方は範囲の終点を示します。

- **領域 : 範囲** : 最小値と最大値の間の値の範囲。
- **領域 : 平均値の 95 % の信頼区間** : 母集団の平均値を含む 95 % の確率をもつ値の範囲。
- **領域 : 個別の 95 % の信頼区間** : このケースが指定された予測値を含む 95 % の確率をもつ値の範囲。
- **領域 : 平均値の上下 1 標準偏差** : 平均値の上下 1 標準偏差間の値の範囲。
- **領域 : 平均値の上下 1 標準誤差** : 平均値の上下 1 標準誤差間の値の範囲。

度数ベースの要約統計

- **度数** :行/ケースの数。
- **累積度数** :行/ケースの累積数。各グラフィック要素は、サブグループの度数と、以前のグループすべての度数合計を示します。
- **度数のパーセント** :行/ケースの総合系に対する、各サブグループ内の行/ケースのパーセンテージ。
- **度数の累積パーセント** :行/ケースの総合系に対する、各サブグループ内の行/ケースの累積パーセンテージ。各グラフィック要素は、サブグループのパーセンテージと、以前のグループすべての総パーセンテージを示します。

衝突変更子を指定する方法

衝突変更子により、グラフ要素同士が重なった際に何が起こるのかが決定されます。

- ▶ 衝突変更子を指定したいグラフ要素を選択します。
- ▶ プロパティ パレットで [要素] タブをクリックします。
- ▶ [変更子] ドロップダウン リストから、衝突変更子を選択します。-auto-を使用すると、アプリケーションがグラフ要素の種類と統計量に対して適切な衝突変更子を判定します。

重ね描き :複数のグラフ要素が同じ値の場合に、それぞれの上部に描画します。

積み上げ :同じデータ値を持った通常は重ね描きされる複数のグラフ要素を積み上げます。

避ける :重ね描きする代わりに、グラフ要素を同じ値に表示されている他のグラフ要素の隣に移動します。グラフ要素は左右対称に並べられます。つまり、グラフ要素はそれぞれ中心点の両側に移動されます。避ける機能は、クラスタ化とよく似ています。

積み重ね :重ね描きする代わりに、グラフ要素を同じ値に表示されている他のグラフ要素の隣に移動します。グラフ要素は左右非対称に並べられます。つまり、一番下のグラフ要素が尺度上の特定の値に位置した状態で、グラフ要素はそれぞれの上に積み上げられます。

ジッター (正規) :同じ値のグラフ要素を、正規分布を使用してランダムに再配置します。

ジッター (一様) :同じ値のグラフ要素を、一様分布を使用してランダムに再配置します。

凡例の位置の変更

グラフに凡例が含まれる場合、凡例は通常グラフの右側に表示されます。必要に応じてこの位置を変更できます。

凡例の位置の変更方法

- ▶ 凡例を選択します。
- ▶ プロパティ パレットで [凡例] タブをクリックします。

図 5-114
[凡例] タブ



- ▶ 位置を選択します。

視覚化および視覚化データのコピー

[全般] パレットには、視覚化とそのデータをコピーするためのボタンが含まれます。

図 5-116
[視覚化のコピー] ボタン



視覚化のコピー : この操作により、視覚化がクリップボードにイメージとしてコピーされます。複数のイメージ フォーマットが利用可能です。イメージを他のアプリケーションに貼り付ける際は、[形式を選択して貼り付け] オプションを選択して、貼り付けで利用可能なイメージ フォーマットから任意のものを選択することができます。

図 5-117
[視覚化データのコピー] ボタン



視覚化データのコピー : この操作により、視覚化を描画するとき使用される基のデータがコピーされます。データはクリップボードにテキスト形式または HTML テキスト形式でコピーされます。データを他のアプリケー

ションに貼り付ける際は、[形式を選択して貼り付け] オプションを選択して、これらのフォーマットのうちのどちらかを選択することができます。

キーボード ショートカット

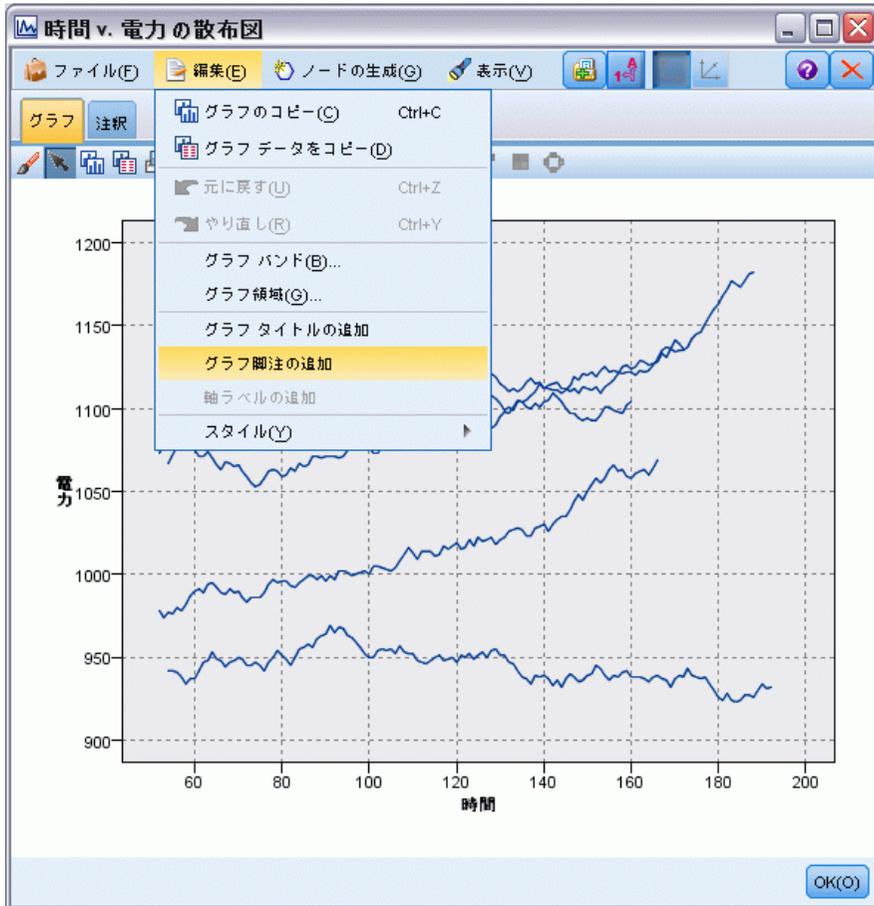
テーブル 5-4
キーボード ショートカット

| ショートカット キー | 関数 |
|----------------|---------------------|
| Ctrl + スペース キー | 探索的分析モードと編集モードの切り替え |
| Del | 視覚化の項目の削除 |
| Ctrl + Z | 元に戻す |
| Ctrl + Y | やり直し (R) |
| F2 | グラフの項目を選択するための外枠の表示 |

表題と脚注の追加

すべてのグラフのタイプで、独自のタイトル、脚注、または軸ラベルを追加して、グラフに表示されているものを識別することができます。

図 5-118
グラフの脚注の追加



グラフへのタイトルの追加

- ▶ メニューから [編集] → [グラフタイトルの追加] を選択します。〈TITLE〉 と入力されているテキスト ボックスが、グラフの上に表示されます。
- ▶ 編集モードであることを確認します。メニューから [表示] → [編集モード] を選択します。
- ▶ 〈TITLE〉 テキストをダブルクリックします。
- ▶ 希望のタイトルを入力し、Return キーを押します。

グラフへの脚注の追加

- ▶ メニューから [編集] → [グラフ脚注の追加] を選択します。〈FOOTNOTE〉 と入力されているテキスト ボックスが、グラフの下に表示されます。

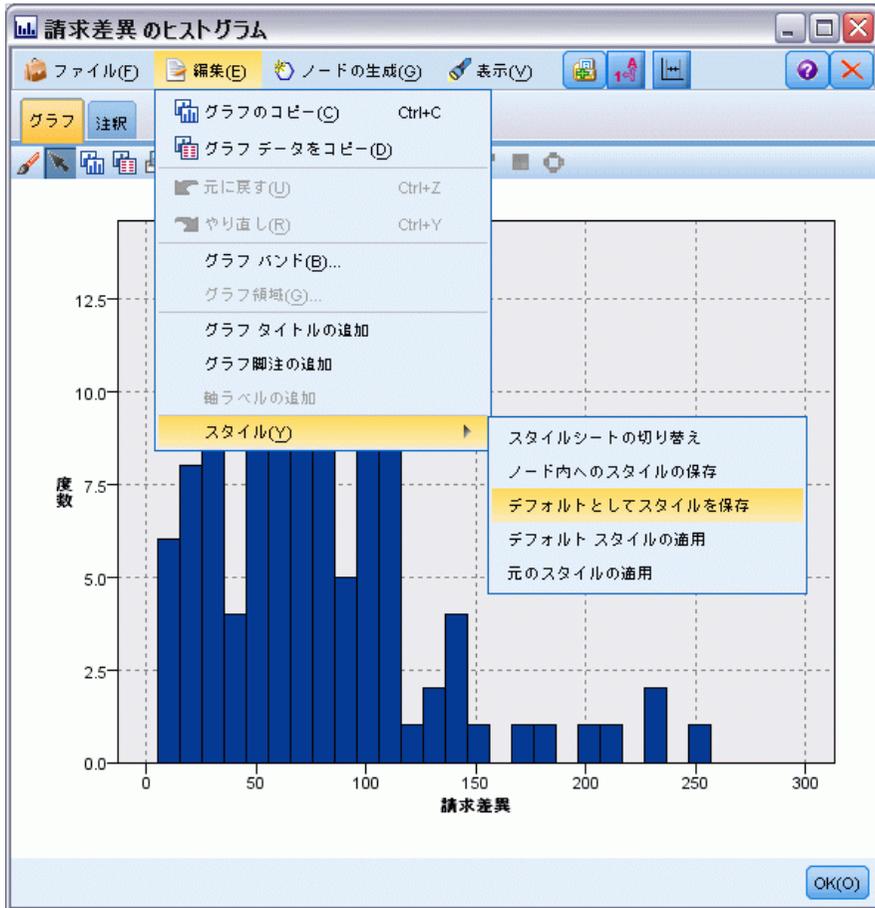
- ▶ 編集モードであることを確認します。メニューから [表示] → [編集モード] を選択します。
- ▶ <FOOTNOTE> テキストをダブルクリックします。
- ▶ 希望のタイトルを入力し、Return キーを押します。

グラフのスタイルシートの使用

色、フォント、記号、線の太さなど、グラフの表示に関する基本的な情報はスタイルシートで管理されます。IBM® SPSS® Modeler が提供するデフォルトのスタイルシートがありますが、必要に応じて変更することができます。たとえば、プレゼンテーション用のグラフで使用する企業の色を設定を行うことができます。詳細は、[p. 405 視覚化の編集](#) を参照してください。

グラフ作成ノードでは、編集モードを使用してグラフの外観のスタイルを変更できます。[編集] → [スタイル] メニューを使用すると、その後現在のグラフノードから生成するすべてのグラフに適用するスタイルシートとして変更を保存できるほか、SPSS Modeler を使用して生成するすべてのグラフの新しいデフォルトのスタイルシートとして変更を保存することもできます。

図 5-119
グラフスタイルの選択



[編集] メニューの [スタイル] オプションで利用可能なスタイルシート オプションが 5 つあります。

- **スタイルシートを切り替え** グラフの外観を変更できる、さまざまな種類の保存されたスタイルシートのリストが表示されます。 [詳細は、p. 429 スタイルシートの適用](#) を参照してください。
- **ノード内へのスタイルの保存** :このオプションでは、選択したグラフのスタイルに対する変更を保存し、それは現在のストリームの同じグラフノードから作成したすべての将来のグラフに適用されます。
- **デフォルトとしてスタイルを保存** :このオプションでは、選択したグラフのスタイルに対する変更を保存し、それはあらゆるストリームのあらゆるグラフノードから作成したすべての将来のグラフに適用されます。このオプションを選択すると、[デフォルトスタイルの適用] を選択して同じスタイルを使用するいかなる既存のグラフも変更することができます。

- **デフォルトスタイルの適用**:このオプションを使用すると、選択したグラフのスタイルをデフォルトとして保存したスタイルに変更します。
- **元のスタイルの適用**:このオプションを使用すると、元のデフォルトとして提供されているスタイルに再度変更します。

スタイルシートの適用

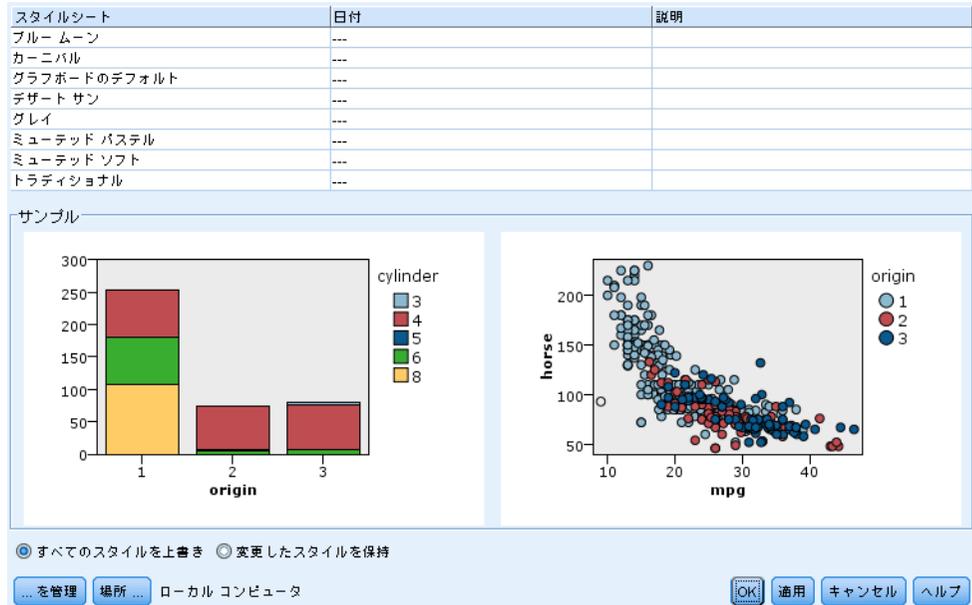
視覚化のスタイル上のプロパティを指定する視覚化スタイルシートを適用することができます。たとえば、スタイルシートはその他のオプションから、フォント、ダッシュ、色を定義できます。ある程度まで、スタイルシートでは、手動で実行する必要がある、編集用のショートカットが用意されています。ただし、スタイルシートは「スタイル」の変更に限られています。凡例の場所または尺度領域などのその他の変更は、スタイルシートに保存されません。

スタイルシートの適用方法

- ▶ メニューから次の項目を選択します。
編集(E) > スタイル > スタイルシートを切り替え
- ▶ [スタイルシートの切り替え] ダイアログ ボックスを使用して、スタイルシートを選択します。
- ▶ ダイアログを閉じずに[適用] をクリックして、スタイルシートを視覚化に適用します。[OK] をクリックして、スタイルシートを適用し、ダイアログ ボックスを閉じます。

[スタイルシートの切り替え/選択] ダイアログ ボックス

図 5-120
[スタイルシートの切り替え] ダイアログ ボックス



ダイアログ ボックスの上部のテーブルには、現在使用できる視覚化スタイルシートがすべて表示されています。事前にインストールされているスタイルシートがありますが、IBM® SPSS® Visualization Designer (別製品) で作成されているスタイルシートもあります。

ダイアログ ボックスの下部には、サンプル データを含む視覚化の例が表示されています。スタイルシートのいずれかを選択して、視覚化の例にスタイルを適用します。これらの例を使用して、実際の視覚化にスタイルシートがどのように影響を与えるかを確認できます。

ダイアログ ボックスには、次のオプションもあります。

既存スタイル : デフォルトでは、スタイルシートが視覚化のすべてのスタイルを上書きできます。この動作は変更できます。

- **すべてのスタイルを上書き** : スタイルシート適用時、現在の編集セッションの間に視覚化で変更されたスタイルなど、視覚化のすべてのスタイルが上書きされます。
- **変更したスタイルを保持** : スタイルシート適用時、現在の編集セッションの間に視覚化で変更されなかったスタイルのみが上書きされます。現在の編集セッションで変更されたスタイルは保持されます。

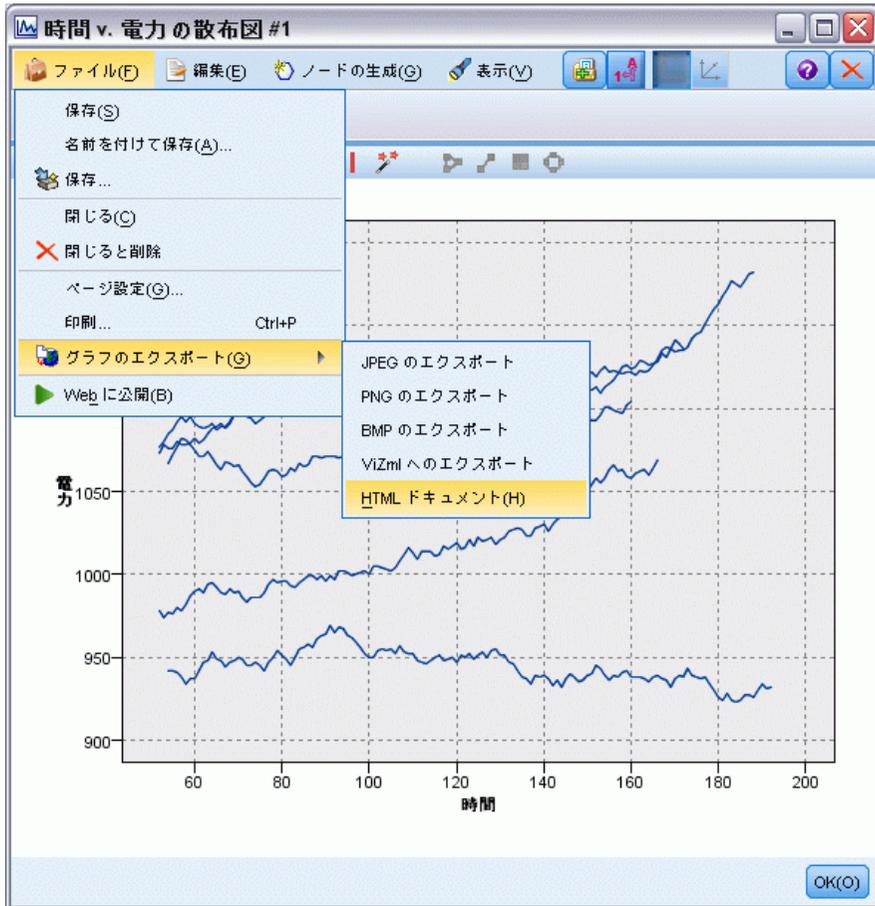
管理: 視覚化テンプレート、スタイルシート、およびマップをコンピュータで管理します。視覚化テンプレート、スタイルシート、およびマップをローカル マシンでインポート、エクスポート、名前変更、および削除できます。詳細は、[p. 320 テンプレート、スタイルシート、マップ ファイルの管理](#) を参照してください。

位置 : 視覚化テンプレート、スタイルシート、およびマップを保存する場所を変更します。現在の場所は、ボタンの右側に表示されます。詳細は、[p. 318 テンプレート、スタイルシート、マップの位置の設定](#) を参照してください。

グラフの印刷、保存、エクスポート

各グラフには、グラフの保存、印刷、またはグラフを他のフォーマットでエクスポートできるようにする多くのオプションが付属しています。これらのオプションのほとんどは [ファイル] メニューから利用できます。さらに、[編集] メニューから選択して、他のアプリケーションで使用するグラフまたはグラフ内のデータをコピーすることができます。

図 5-121
グラフ ウィンドウの [ファイル] メニューとツールバー



印刷

- ▶ グラフを印刷するには、[印刷] メニューを選択するかボタンを使用します。印刷する前に、[ページ設定(U)] と [印刷プレビュー(W)] を使用して、印刷オプションを設定したり、出力をプレビューすることができます。

グラフの保存

- ▶ グラフを IBM® SPSS® Modeler の出力ファイル (*.cou) に保存するには、[ファイル] → [保存(S)] または [ファイル] → [名前を付けて保存] を選択します。

or

グラフをリポジトリに保存するには、[ファイル] → [出力を格納] を選択します。

グラフのコピー

- ▶ MS Word や MS PowerPoint など他のアプリケーションで使用するグラフをコピーするには、[編集] → [グラフのコピー] を選択します。

データのコピー

- ▶ MS Excel や MS Word など他のアプリケーションで使用するデータをコピーするには、[編集] → [データのコピー] を選択します。デフォルトでは、データの形式は HTML になります。貼り付け時に他の形式オプションを表示するには、貼り付け先のアプリケーションの [形式を選択して貼り付け (P)] を使用します。

グラフのエクスポート

[グラフのエクスポート] オプションを使用すると、ビットマップ (bmp)、JPEG (jpg)、PNG (png)、HTML (html)、ViZml ドキュメント (xml) のいずれかの形式でグラフをエクスポートして他の IBM® SPSS® Statistics アプリケーションで使用することができます。

- ▶ グラフをエクスポートするには、[ファイル] → [グラフのエクスポート] を選択して、形式を選択します。

テーブルのエクスポート

[テーブルのエクスポート] オプションを使用すると、タブ区切り (tab)、カンマ区切り (csv)、HTML (html) のいずれかの形式でテーブルをエクスポートできます。

- ▶ テーブルをエクスポートするには、[ファイル] → [テーブルのエクスポート] を選択して、形式を選択します。

出力ノード

出力ノードの概要

出力ノードを利用すれば、データやモデルに関する情報を取得できます。出力ノードは、他のソフトウェア ツールのインターフェイスに対応した、さまざまなフォーマットでデータをエクスポートできるメカニズムも備えています。

次の出力ノードを利用できます。



テーブル ノードで、データがテーブル形式で表示されます。このデータは、ファイルにも書き込めます。この機能は、データの値を調査したり、データを読みやすいフォーマットでエクスポートする必要がある場合に役立ちます。 [詳細は、 p. 443 テーブルノード を参照してください。](#)



クロス集計ノードで、フィールド間の関係を示すテーブルを作成します。一般的にこのノードは、2 つのシンボル値フィールドの関係を示す場合によく使用されますが、フラグ型フィールド間または数値型フィールド間の関係を示すこともできます。 [詳細は、 p. 449 クロス集計ノード を参照してください。](#)



精度分析ノードで、予測モデルの能力を評価して正確な予測を生成します。分析ノードでは、1 つ以上のモデル ナゲットについて、予測値と実際値をさまざまな方法で比較します。また、分析ノードでは予測モデル同士を比較できます。 [詳細は、 p. 454 精度分析ノード を参照してください。](#)



データ検査ノードでは、欠損値、外れ値、および極値に関する情報の他、各フィールドの要約統計量、ヒストグラムや棒グラフを含む、データを広範に検査するための手段を提供しています。結果は把握しやすい行列形式で表示され、ソートしたり、フルサイズのグラフやデータ準備ノードを生成することができます。 [詳細は、 p. 459 データ検査ノード を参照してください。](#)



変換ノードによって、選択フィールドに適用する前に変換の結果を選択し、視覚的に確認することができます。 [詳細は、 p. 476 変換ノード を参照してください。](#)



記述統計ノードでは、数値型フィールドに関する基本的な集計情報が提供されます。このノードで、個々のフィールドの要約統計量とフィールド間の相関が計算されます。 [詳細は、 p. 482 記述統計ノード を参照してください。](#)



平均比較ノードでは、独立したグループ間で、または関連するフィールドのペア間で著しい違いがあるかどうかを調べるために、平均を比較します。たとえば、販売促進活動の前後で平均収益を比較したり、販売促進活動を受けなかった顧客と受けた顧客からの収益を比較することができます。詳細は、[p. 487 平均比較ノード](#) を参照してください。



レポート ノードで、固定テキスト、およびデータやデータから導かれた他の式を含む、フォーマット済みレポートを作成します。レポートの書式は、固定テキストとデータの出力構成を定義するテキスト テンプレートを使用して指定します。テンプレート内の HTML タグを使用し、また [出力] タブでオプションを設定することで、カスタムのテキスト書式設定を提供できます。テンプレート内の CLEM 式を使用して、データ値やその他の条件出力を含めることができます。詳細は、[p. 493 レポートノード](#) を参照してください。



グローバル ノードで、データを走査し、CLEM 式で使用できる要約値を算出します。たとえば、グローバル ノードを使用して、[年齢] という名前のフィールドの統計量を算出し、次に CLEM 式に `@GLOBAL_MEAN(年齢)` 関数を挿入して年齢の全体的な平均を算出することができます。詳細は、[p. 497 グローバルノード](#) を参照してください。

出力の管理

出力マネージャは、IBM® SPSS® Modeler のセッション中に生成されるチャート、グラフ、テーブルを表示します。マネージャ内でダブルクリックすると、いつでも再表示することができます。該当するストリームやノードに戻る必要はありません。

出力マネージャを表示するには

- ▶ [表示(V)] メニューを開き、[マネージャ(M)] を選択します。[出力] タブをクリックします。

図 6-1
出力マネージャ



出力マネージャから、次のことができます。

- ヒストグラム、評価グラフ、およびテーブルなどの既存の出力オブジェクトを表示する。
- 出力オブジェクトの名前を変更する。
- 出力オブジェクトをディスク、または IBM® SPSS® Collaboration and Deployment Services Repository（可能ならば）に保存する。
- 現在のプロジェクトに出力ファイルを追加する。
- 現在のセッションから未保存の出力オブジェクトを削除する。
- 出力オブジェクトを IBM SPSS Collaboration and Deployment Services Repository（可能ならば）に保存、またはそこから出力オブジェクトを取得する。

オプションにアクセスするには、[出力] タブの任意の場所で右クリックします。

出力を表示

オンスクリーン出力は、出力ブラウザ ウィンドウに表示されます。出力ブラウザ ウィンドウは、それ自体に出力を印刷または保存、または別のフォーマットで出力をエクスポートできるメニュー セットを備えています。具体的なオプションは、出力のタイプによって大きく異なることに注意してください。

データの印刷、保存、エクスポート： 以下に詳しく説明します。

- 出力を印刷するには、[印刷(P)] メニュー オプションを選択するかボタンを使用します。印刷する前に、[ページ設定(U)] と [印刷プレビュー(W)] を使用して、印刷オプションを設定したり、出力をプレビューすることができます。
- 出力を IBM® SPSS® Modeler 出力ファイル (.cou) に保存する場合、[ファイル] メニューから [保存] または [名前を付けて保存] を選択します。
- テキストや HTML などの他の形式で出力を保存するには、[ファイル] メニューから [エクスポート] を選択します。 [詳細は、 p. 441 出力のエクスポート を参照してください。](#)
- ほかのユーザーが IBM® SPSS® Collaboration and Deployment Services Deployment Portal を使用して出力を表示できるよう共有リポジトリに出力を保存するには、[ファイル] メニューから [Web に公開] を選択します。このオプションには、別途 IBM® SPSS® Collaboration and Deployment Services のライセンスが必要なことに注意してください。

セルと列の選択： [編集(E)] メニューには、現在の出力形式に合わせて、選択、選択解除、セルや列のコピーに関するさまざまなオプションが含まれています。 [詳細は、 p. 442 セルと列の選択 を参照してください。](#)

ノードの生成： [ノードの生成] メニューでは、出力ブラウザの内容に応じて新しいノードを生成することができます。オプションは、出力形式や、現在選択されている出力内の項目によって大きく異なります。特殊な出力形式のノード生成オプションの詳細については、該当する出力の説明書を参照してください。

[Web に公開]

[Web に公開] を選択して、特定の種類のストリームの出力を IBM® SPSS® Collaboration and Deployment Services の基礎を形成する中央共有 IBM® SPSS® Collaboration and Deployment Services Repository に公開できます。このオプションを使用すると、出力を表示する必要があるそのほかのユーザーはインターネット アクセスおよび IBM SPSS Collaboration and Deployment Services アカウントを使用して出力を表示できます。IBM® SPSS® Modeler をインストールする必要はありません。

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。

次のテーブルに、[Web に公開] 機能をサポートする SPSS Modeler ノードを示します。これらのノードからの出力は、IBM SPSS Collaboration and Deployment Services Repository に出力オブジェクト (.cou) 形式で保存され、IBM® SPSS® Collaboration and Deployment Services Deployment Portal で直接表示できます。

その他の種類の出力は、関連するアプリケーション（ストリーム オブジェクトの場合は SPSS Modeler）が、ユーザーのコンピュータにインストールされている場合にのみ表示できます。

テーブル 6-1
Web に公開をサポートしているノード

| ノードの種類 | ノード |
|-----------------------|---------------|
| グラフ作成 | all |
| 出力 | テーブル |
| | クロス集計 |
| | データ検査 |
| | 変換 |
| | 平均 |
| | 分析 |
| | 統計値 |
| | レポート (HTML) |
| IBM® SPSS® Statistics | Statistics 出力 |

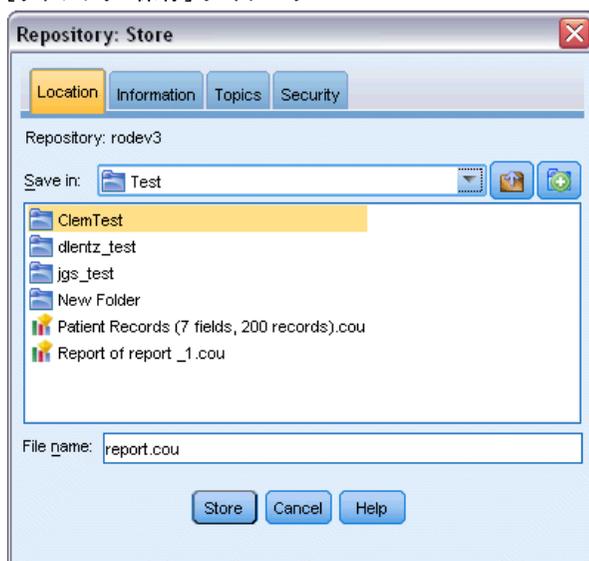
出力を Web に公開する

出力を Web に公開する手順は次のとおりです。

- ▶ IBM® SPSS® Modeler では、テーブルに表示されたノードの 1 つを実行します。ノードの 1 つを実行すると、新しいウィンドウで出力オブジェクトを作成します（テーブル、マトリックス、レポート オブジェクトなど）。
- ▶ [出力オブジェクト] ウィンドウから次の項目を選択します。
File > [Web に公開]
注：標準の Web ブラウザで使用するために単純な HTML ファイルをエクスポートする場合、[ファイル] メニューから [エクスポート] を選択した後、[HTML] を選択します。
- ▶ IBM® SPSS® Collaboration and Deployment Services Repository への接続 詳細は、[9 章 リポジトリへの接続 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

正常に接続されると、さまざまなストレージ オプションを提供する [リポジトリ: 保存] ダイアログが表示されます。詳細は、[9 章 リポジトリのオブジェクトの保存 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

図 6-2
[リポジトリ: 保存] ダイアログ



- ▶ ストレージ オプションを選択したら、[保存] をクリックします。

公開出力の Web 表示

この機能を使用するには、IBM SPSS Collaboration and Deployment Services アカウントのセット アップが必要です。表示するオブジェクトタイプの関連するアプリケーション（IBM® SPSS® Modeler または IBM® SPSS® Statistics）がインストールされている場合、出力はブラウザではなくアプリケーションで表示されます。

注 :IBM® SPSS® Collaboration and Deployment Services 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。

公開出力を Web で表示する手順は次のとおりです。

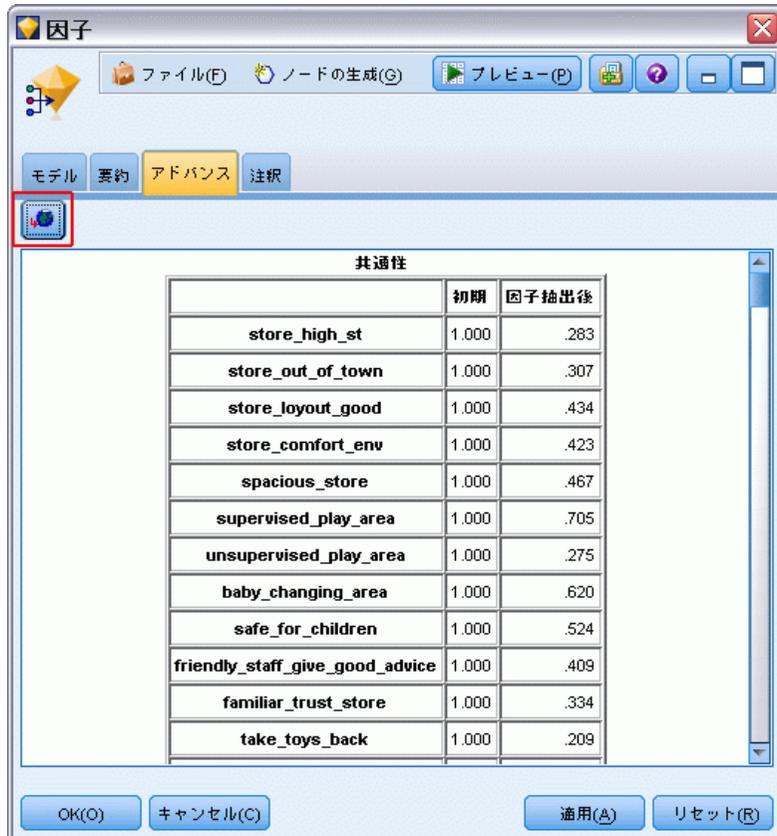
- ▶ ブラウザに `http://<repos_host>:<repos_port>/peb` を指定します。
repos_host および repos_port は IBM SPSS Collaboration and Deployment Services ホストのホスト名およびポート番号です。
- ▶ IBM SPSS Collaboration and Deployment Services アカウントのログインの詳細を入力してください。
- ▶ [コンテンツ リポジトリ] をクリックします。

- ▶ 表示するオブジェクトに移動または検索します。
- ▶ オブジェクト名をクリックします。グラフなど一部のオブジェクト タイプについて、オブジェクトをブラウザで表示するときに遅延が生じる場合があります。

HTML ブラウザで出力結果を表示

線形、ロジスティックおよび因子分析モデル ナゲットの [詳細] タブで、Internet Explorer などの各ブラウザに情報を表示することができます。情報は HTML として出力され、保存して、企業のイントラネットやインターネット サイトなどあらゆる場所で再利用することができます。

図 6-3
モデル ナゲットの [詳細] タブの [起動] ボタン



ブラウザで情報を表示するには、モデル ナゲットの [詳細] タブ ダイアログ ボックスの左上にあるモデル アイコンの下 [起動] ボタンをクリックします。

出力のエクスポート

出力ブラウザ ウィンドウで、テキストや HTML などの他のフォーマットに出力をエクスポートするよう選択できます。エクスポート フォーマットは、出力形式によって大きく異なりますが、一般にその出力を生成するために使用したノードで [ファイルに保存] を選択すると、利用できるファイル形式オプションは似たものとなります。

出力をエクスポートするには

- ▶ 出力ブラウザで、[ファイル(F)] メニューを開き、[エクスポート] を選択します。次に生成するファイル形式を選択します。
 - **タブ区切り (*.tab)**: データ値を含む、フォーマット済みのテキスト ファイルを生成します。このスタイルは、他のアプリケーションにインポートできる、プレーン テキストで表した情報を生成する場合に役立ちます。このオプションは、テーブル ノード、クロス集計、および平均値ノードで使用できます。
 - **カンマ区切り (*.dat)**: データ値を含む、カンマで区切られたテキスト ファイルを生成します。このスタイルでは、表計算アプリケーションやデータ分析アプリケーションにインポートできるフォーマットのデータ ファイルをすばやく生成できます。このオプションは、テーブル ノード、クロス集計、および平均値ノードで使用できます。
 - **移行タブ区切り (*.tab)**: このオプションは、[タブ区切り] オプションとまったく同じですが、行がフィールドを表し、列がレコードを表すように、データの行列入替が行われます。
 - **移行カンマ区切り (*.dat)**: このオプションは、[カンマ区切り] オプションとまったく同じですが、行がフィールドを表し、列がレコードを表すように、データの行列入替が行われます。
 - **HTML (*.html)**: ファイルに HTML フォーマットの出力データを書き込みます。

セルと列の選択

図 6-4
テーブル ブラウザ ウィンドウ

| | ID | 名前 | 居住地 | 耕地面積 | 降雨量 | 土地の品質 | 農場の収... | 主作物 | 申請内容 |
|----|-------|---------|-----|------|-----|-------|-------------|--------|-----------|
| 1 | id602 | name602 | 北部 | 1780 | 42 | 9 | 734118.0... | トウモロコシ | arable_de |
| 2 | id606 | name606 | 南東部 | 1580 | 42 | 7 | 445785.0... | トウモロコシ | arable_de |
| 3 | id607 | name607 | 南東部 | 1820 | 29 | 6 | 211605.0... | トウモロコシ | arable_de |
| 4 | id608 | name608 | 南東部 | 1640 | 108 | 7 | 1167040.... | トウモロコシ | arable_de |
| 5 | id610 | name610 | 南東部 | 600 | 80 | 6 | 267928.0... | 小麦 | arable_de |
| 6 | id611 | name611 | 南東部 | 980 | 38 | 6 | 222703.0... | トウモロコシ | arable_de |
| 7 | id613 | name613 | 南東部 | 440 | 86 | 3 | 115544.0... | ジャガイモ | arable_de |
| 8 | id614 | name614 | 南東部 | 1260 | 90 | 8 | 900243.0... | トウモロコシ | arable_de |
| 9 | id616 | name616 | 中部 | 1660 | 36 | 9 | 490617.0... | 大豆 | arable_de |
| 10 | id620 | name620 | 北部 | 880 | 74 | 6 | 426988.0... | 大豆 | arable_de |
| 11 | id621 | name621 | 南西部 | 1160 | 105 | 4 | 299274.0... | トウモロコシ | arable_de |
| 12 | id622 | name622 | 南東部 | 1500 | 61 | 7 | 687736.0... | 小麦 | arable_de |
| 13 | id623 | name623 | 南東部 | 1260 | 17 | 8 | 170279.0... | トウモロコシ | arable_de |
| 14 | id626 | name626 | 中部 | 1580 | 109 | 8 | 1286430.... | 小麦 | arable_de |
| 15 | id627 | name627 | 南東部 | 500 | 93 | 3 | 102720.0... | 大豆 | arable_de |
| 16 | id628 | name628 | 南東部 | 880 | 15 | 5 | 70439.800 | 小麦 | arable_de |
| 17 | id630 | name630 | 中部 | 680 | 81 | 4 | 221391.0... | ジャガイモ | arable_de |
| 18 | id636 | name636 | 南東部 | 1160 | 21 | 8 | 185939.0... | ジャガイモ | arable_de |
| 19 | id637 | name637 | 中部 | 940 | 106 | 6 | 622450.0... | トウモロコシ | arable_de |
| 20 | id638 | name638 | 中部 | 1480 | 64 | 6 | 586185.0... | 小麦 | arable_de |

テーブル ノード、クロス集計ノード、平均値ノードを含む多くのノードがテーブル形式の出力を生成します。セルの選択、テーブル全体または一部のクリップボードへのコピー、現在の選択を基に新規ノードの生成、およびテーブルの保存と印刷を含む同じような方法でこれらの出力テーブルを表示し、操作することができます。

セルの選択: セルを選択するには、そのセルをクリックします。複数のセルを範囲として選択するには、目的の範囲の一方の角をクリックした後、その範囲の対角までマウスをドラッグしてマウス ボタンを放します。列全体を選択するには、列見出しをクリックします。複数の列を選択するには、Shift キーまたは Ctrl キーを押しながら列見出しをクリックします。

新しく選択すると、その前の選択は取り消されます。Ctrl キーを押しながら新しく選択すると、その前の選択は取り消されず、既存の選択項目に新しい選択項目が追加されます。この方法を使用して、テーブル内の連続していない複数の領域を選択できます。[編集] メニューにも [すべて選択(A)] と [選択解除] があります。

列の並び替え： テーブル ノードと平均値ノードの出力ブラウザで、列の見出しをクリックして目的の位置にドラッグすると、テーブルの列を移動することができます。列は 1 回に 1 つだけ移動することができます。

テーブル ノード

テーブル ノードでは、データ内の値を一覧表示するテーブルを作成します。ストリーム内のすべてのフィールドおよびすべての値が含まれ、データ値を容易に調査したり読みやすい形式でデータ値をエクスポートすることができます。オプションで、特定の条件を満たすレコードを強調表示することができます。

テーブル ノードの [設定] タブ

図 6-5
テーブル ノード :[設定] タブ



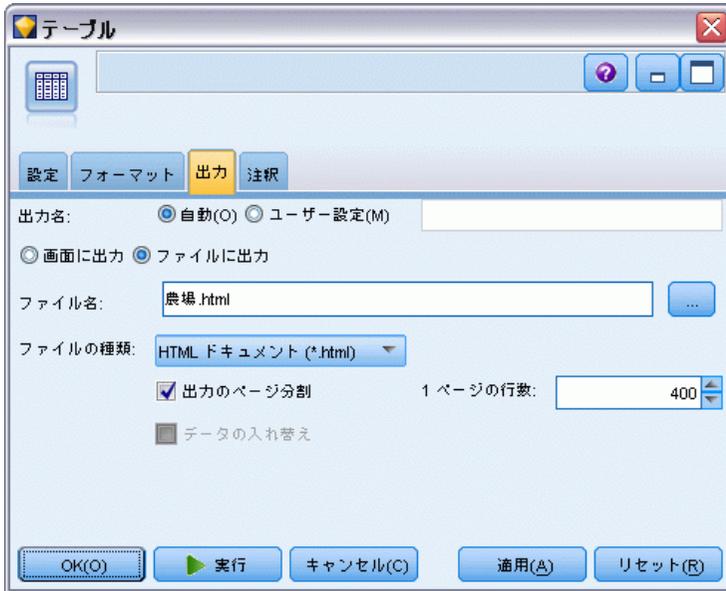
レコードの強調表示： テーブル内のレコードを強調表示するために、対象のレコードを真にする CLEM 式を入力します。このオプションは、[画面に出力] が選択されている場合に有効になります。

テーブル ノードの [形式] タブ

[形式] タブには、フォーマットをフィールド単位で指定するための設定が用意されています。このタブは、データ型ノードと共用です。 [詳細は、p. 170 フィールド形式の \[設定\] タブ を参照してください。](#)

出力ノードの [出力] タブ

図 6-6
出力ノードの [出力] タブ



テーブル形式の出力を生成するノードについては、[出力] タブで形式と出力先を指定することができます。

出力名： ノードの実行時に生成される出力の名前を指定します。[自動] は、出力を生成するノードの名前に基づいて名前を選択します。[ユーザー設定] で別の名前を指定することもできます。

画面に出力 (デフォルト) : オンラインで表示するための出力を作成します。出力ノードを実行すると、出力オブジェクトはマネージャ ウィンドウの [出力] タブに表示されます。

ファイルに出力： ノードの実行時に、出力をファイルに保存します。このオプションを選択した場合は、ファイル名を入力して (または、ファイル選択 ボタンを使ってディレクトリを参照し、ファイル名を指定して)、ファイルフォーマットを選択してください。ファイル形式には、特定の出力形式に使用できないものがありますので注意してください。

データはシステム デフォルトの文字コード形式で出力され、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 IBM SPSS Modeler 内の Unicode のサポート in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

- **データ(タブ区切り) (*.tab)**: データ値を含む、フォーマット済みのテキスト ファイルを生成します。このスタイルは、他のアプリケーションにインポートできる、プレーン テキストで表した情報を生成する場合に役立ちます。このオプションは、テーブル ノード、クロス集計、および平均値ノードで使用できます。
- **データ(カンマ区切り) (*.dat)**: データ値を含む、カンマで区切られたテキスト ファイルを生成します。このスタイルでは、表計算アプリケーションやデータ分析アプリケーションにインポートできるフォーマットのデータ ファイルをすばやく生成できます。このオプションは、テーブル ノード、クロス集計、および平均値ノードで使用できます。
- **HTML (*.html)**: ファイルに HTML フォーマットの出力データを書き込みます。テーブル式出力(テーブル ノード、クロス集計、または平均値ノードから)の場合は、一連の HTML ファイルに フィールド名を含む内容パネルとHTML テーブルのデータが格納されます。テーブルの行数が [1 ページの行数] の設定値を超えた場合、1 つのテーブルが複数の HTML ファイルに分割されることがあります。このような場合は、すべてのテーブル ページに対するリンクが内容パネルに記載され、これを使って各テーブルに移動できます。テーブル形式以外の出力の場合は、ノードの結果を含む HTML ファイルが 1 つ作成されます。

注: HTML 出力に最初のページの書式だけが含まれている場合は、[出力のページ分割] を選択し、1 ページにすべての出力が収まるように [1 ページの行数] の設定を調整してください。または、レポート ノードのようなノードの出力テンプレートにカスタムの HTML タグがある場合は、フォーマットの種類として [カスタム] を選択してください。
- **テキストファイル (*.txt)**: 出力データを含むテキスト ファイルを生成します。このスタイルは、ワープロやプレゼンテーション ソフトウェアなど、他のアプリケーションにインポートできる出力データを生成する場合に便利です。このオプションは、一部のノードでは利用できません。
- **出力オブジェクト (*.cou)**: このフォーマットで保存された出力オブジェクトは、IBM® SPSS® Modeler で開いて表示したり、プロジェクトへの追加、IBM® SPSS® Collaboration and Deployment Services Repository を使用して公開し、追跡することができます。

出力ビュー: 平均値ノードに対してデフォルトで、シンプルまたは詳細のどちらのモードで出力を表示するかを指定できます。このビューは、生成された出力をブラウズするときに切り替えることができます。 [詳細は、p. 491 平均値ノード出力ブラウザ](#) を参照してください。

書式: レポート ノードの場合に、出力を自動的にフォーマットするか、またはテンプレート中の HTML を使ってフォーマットするかを選択することができます。テンプレート中の HTML を使ってフォーマットする場合は、[カスタム] を選択します。

表題. レポート ノードに対して、レポート出力の上部に表示するタイトルテキストを指定することができます。

挿入文字列を強調表示： レポート ノードの場合に、このオプションを選択すると、レポート テンプレート中の CLEM 式が生成した文字列が強調表示されます。詳細は、[p. 494 レポート ノードの \[テンプレート\] タブ](#) を参照してください。フォーマットに [カスタム] を指定した場合、このオプションの使用はお勧めできません。

1 ページの行数： レポート ノードの場合に、出力レポートの自動フォーマット時に、各ページに入れる行数を指定します。

データの入れ替え： このオプションは、エクスポート前にデータの行列を入れ替えて、行がフィールドを、列がレコードを表すようにします。

注： 大きなテーブルでは、特にリモート サーバーを使用しているような場合に、上記のオプションを利用するとパフォーマンスが低下する場合があります。そのような場合は、ファイル出力ノードを使用すれば、パフォーマンスを向上することができます。詳細は、[7 章 p. 526 ファイル エクスポート ノード](#) を参照してください。

テーブル ブラウザ

図 6-7
テーブル ブラウザ ウィンドウ

| | ID | 名前 | 居住地 | 耕地面積 | 降雨量 | 土地の品質 | 農場の収... | 主作物 | 申請内容 |
|----|-------|---------|-----|------|-----|-------|-------------|--------|-----------|
| 1 | id602 | name602 | 北部 | 1780 | 42 | 9 | 734118.0... | トウモロコシ | arable_de |
| 2 | id606 | name606 | 南東部 | 1580 | 42 | 7 | 445785.0... | トウモロコシ | arable_de |
| 3 | id607 | name607 | 南東部 | 1820 | 29 | 6 | 211605.0... | トウモロコシ | arable_de |
| 4 | id608 | name608 | 南東部 | 1640 | 108 | 7 | 1167040... | トウモロコシ | arable_de |
| 5 | id610 | name610 | 南東部 | 600 | 80 | 6 | 267928.0... | 小麦 | arable_de |
| 6 | id611 | name611 | 南東部 | 980 | 38 | 6 | 222703.0... | トウモロコシ | arable_de |
| 7 | id613 | name613 | 南東部 | 440 | 86 | 3 | 115544.0... | ジャガイモ | arable_de |
| 8 | id614 | name614 | 南東部 | 1260 | 90 | 8 | 900243.0... | トウモロコシ | arable_de |
| 9 | id616 | name616 | 中部 | 1660 | 36 | 9 | 490617.0... | 大豆 | arable_de |
| 10 | id620 | name620 | 北部 | 880 | 74 | 6 | 426988.0... | 大豆 | arable_de |
| 11 | id621 | name621 | 南西部 | 1160 | 105 | 4 | 299274.0... | トウモロコシ | arable_de |
| 12 | id622 | name622 | 南東部 | 1500 | 61 | 7 | 687736.0... | 小麦 | arable_de |
| 13 | id623 | name623 | 南東部 | 1260 | 17 | 8 | 170279.0... | トウモロコシ | arable_de |
| 14 | id626 | name626 | 中部 | 1580 | 109 | 8 | 1286430... | 小麦 | arable_de |
| 15 | id627 | name627 | 南東部 | 500 | 93 | 3 | 102720.0... | 大豆 | arable_de |
| 16 | id628 | name628 | 南東部 | 880 | 15 | 5 | 70439.800 | 小麦 | arable_de |
| 17 | id630 | name630 | 中部 | 680 | 81 | 4 | 221391.0... | ジャガイモ | arable_de |
| 18 | id636 | name636 | 南東部 | 1160 | 21 | 8 | 185939.0... | ジャガイモ | arable_de |
| 19 | id637 | name637 | 中部 | 940 | 106 | 6 | 622450.0... | トウモロコシ | arable_de |
| 20 | id638 | name638 | 中部 | 1480 | 64 | 6 | 586185.0... | 小麦 | arable_de |

テーブル ブラウザは、テーブル形式のデータを表示し、そこで、セルの選択やコピー、列の並べ替え、テーブルの印刷などの標準操作ができるようになります。詳細は、[p. 442 セルと列の選択](#) を参照してください。これらは、ノードでデータをプレビューする際に実行できる同じ操作です。詳細は、[5 章 ノードのデータをプレビューする in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

テーブル データのエクスポート: 次の項目を選択して、テーブル ブラウザからデータをエクスポートできます。

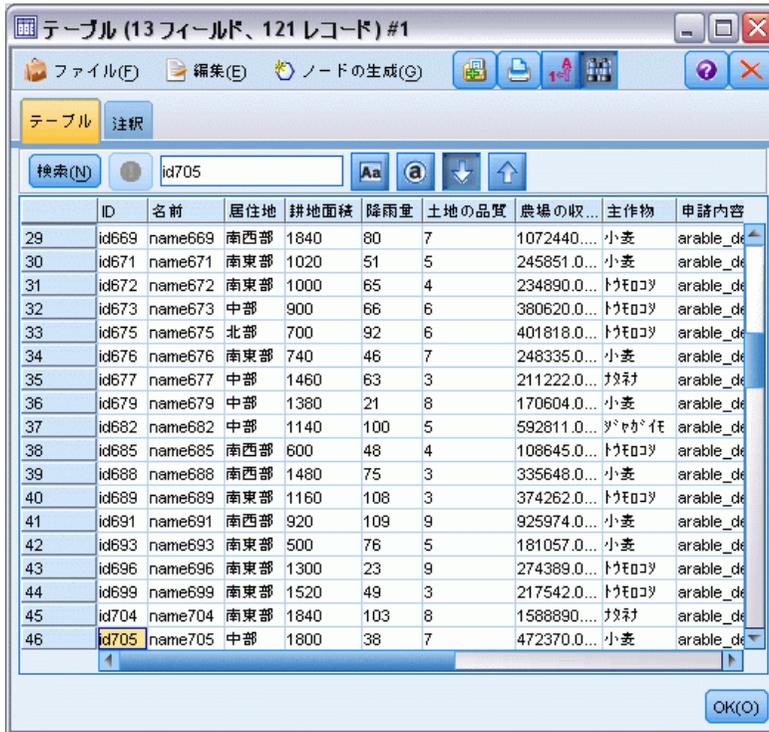
File > エクスポート

詳細は、[p. 441 出力のエクスポート](#) を参照してください。

データはシステム デフォルトの文字コード形式で出力され、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 IBM SPSS Modeler 内の Unicode のサポート in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

テーブルの検索: メイン ツールバーの検索ボタン (双眼鏡のアイコン) をクリックすると、検索用ツールバーが表示されます。ここから、テーブル中の特定の値を検索することができます。テーブル中を前方または後方に検索することができます。また、大文字と小文字を区別するかどうかを指定することもできます ([Aa] ボタン)。検索処理を中断するには、検索中断ボタンをクリックします。

図 6-8
検索コントロールが有効になっているテーブル



ノードの生成：[ノード生成]メニューには、ノード生成操作に関するメニュー項目があります。

- **条件抽出ノード(レコード)**：テーブル内で選択したセルに対するレコードを選択する条件抽出ノードを生成します。
- **条件抽出ノード(AND)**：テーブル内で選択したすべての値を含むレコードを選択する条件抽出ノードを生成します。
- **条件抽出ノード(OR)**：テーブル内で選択したいずれかの値を含むレコードを選択する条件抽出ノードを生成します。
- **フィールド作成ノード(レコード)**：新しいフラグ型フィールドを作成する、フィールド作成ノードを生成します。テーブル内で選択したセルのレコードには T、その他のレコードには F がフラグ型フィールドに含まれます。

- **フィールド作成ノード (AND)**: 新しいフラグ型フィールドを作成する、フィールド作成ノードを生成します。テーブル内で選択した値をすべて含むレコードには T、その他のレコードには F がフラグ型フィールドに含まれます。
- **フィールド作成ノード (OR)**: 新しいフラグ型フィールドを作成する、フィールド作成ノードを生成します。テーブル内で選択した値をすべて含むレコードには T、その他のレコードには F がフラグ型フィールドに含まれます。

クロス集計ノード

クロス集計ノードでは、フィールド間の関係を示すテーブルを作成できます。一般的にこのノードは、2つのカテゴリ フィールド（フラグ型、名義型、順序型）の関係を示す場合によく使われますが、連続型（数値範囲）型フィールド間の関係を示す場合にも使用できます。

クロス集計ノードの [設定] タブ

[設定] タブでは、クロス集計の構造に関する設定を指定することができます。

図 6-9
クロス集計ノード:[設定] タブ



フィールド: 次のオプションで、フィールド選択タイプを選択します。

- **選択**：行列の [行] と [列] に 1 つずつカテゴリ フィールドを選択できます。行列の行と列は、選択したカテゴリ フィールドの値のリストで定義されます。行列のセルには、下で選択する項目の要約統計量が入ります。
- **すべてのフラグ (真の値)**：データ中の各フラグ型フィールドの 1 行 1 列から成る行列を要求します。行列のセルには、フラグの組み合わせごとに、両方に対して真 (true) となるレコードの合計数が入ります。たとえば、[パン購入に対応する行とチーズ購入に対応する列がある場合、その行と列が交差するセルには、パン購入とチーズ購入の両方に対して真 (true) となるレコード数が含まれます。
- **すべての数値**：各数値フィールドの 1 行 1 列で構成される行列を要求します。行列のセルは、対応する一対のフィールドのクロスする値の積の合計を表します。つまり、行列の各セルでは、レコードごとに行フィールドと列フィールドの値が乗算され、レコード全体の合計が計算されます。

欠損値を含める：ユーザーによる欠損値 (空白) とシステムによる欠損値 (ヌル) が、行と列の出力に含まれます。たとえば、値 N/A が選択されたフィールドのユーザー欠損として定義されていた場合、N/A とラベル付けされた別の列は、その他のカテゴリのようにそのテーブルに含まれます。このオプションが選択解除されると、その頻度に関わらず N/A 列は除外されます。 [詳細は、6 章 欠損値の概要 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。](#)

注：欠損値を含めるオプションは、選択したフィールドがクロス集計されている場合にのみ適用されます。空白値はヌルにマップされ、モードが [選択]、コンテンツが [関数] に設定されている場合は、関数フィールドの集計から除外され、モードが [すべての数値] に設定されている場合は、すべての数値の集計から除外されます。

セルの内容：上で [選択] フィールドを選択した場合、行列のセルで使う統計量を指定できます。度数の統計量を選択するか、またはオーバーレイ フィールド を選択して、行フィールドと列フィールドの値に基づいた数値フィールドの値を集計します。

- **クロス表**：セル値は、対応する値の組み合わせが含まれているレコードの度数やパーセンテージです。[外観] タブの設定を使って、どのクロス表要約を使用するかを指定することができます。グローバル カイ 2 乗値も有意値と共に表示されます。 [詳細は、p. 452 クロス集計 ノード出力ブラウザ を参照してください。](#)
- **関数**：要約関数を選択すると、適切な行値と列値を含むケースに対して、セル値が、選択したオーバーレイ フィールド値の関数になります。たとえば、行フィールドが地域で列フィールドが製品、そしてオーバーレイ フィールドが収益の場合、北東地区行と部品列中のセルには、北東地域で販売された部品の収益の合計 (または平均、最小、また

は最大) が格納されます。デフォルトの要約関数は [平均] になります。関数フィールドを要約する他の関数を選択できます。オプションには、[平均]、[合計]、[標準偏差]、[最大]、および [最小] が含まれています。

クロス集計ノードの [外観] タブ

[外観] タブでは、行列のソートや強調表示に関する設定や、クロス表行列に表示する統計量に関する設定を指定することができます。

図 6-10
クロス集計ノード :[外観] タブ



行と列： 行列中の行および列見出しを制御します。デフォルトは [未ソート] です。行や列見出しをソートするには、[昇順] または [降順] を選択してください。

オーバーレイ： 行列中の極値を強調表示します。値は、セルの度数（クロス表行列）または計算された値（関数行列）に基づいて強調表示されます。

- **上位を強調表示：** 行列の上位 N 個の値を強調表示（赤）できます。強調表示する値の数を指定します。
- **下位を強調表示：** 行列の下位 N 個の値を強調表示（緑）できます。強調表示する値の数を指定します。

注：これらの強調表示オプションでは、同じ値が複数ある場合、実際に強調表示したい数より多くの値が強調表示されることがあります。たとえば、ゼロ値を持つセルが 6 つある行列の場合、[最下位を強調表示] に 5 を指定すると、6 つのゼロ値すべてが強調表示されます。

クロス表セルの内容：クロス表に対して、クロス表のクロス集計の行列に含める要約統計量を指定できます。これらのオプションは、[すべての数値] または [関数] オプションのどちらかが [設定] タブで選択されている場合には利用できません。

- **度数：**セルには、対応する行の値を持つ列値があるレコード数が含まれます。これは、デフォルトのセルの内容です。
- **期待度数：**行と列の間には何も関係がないと仮定して、セルにはレコード数に対する期待値が含まれます。期待値は次の式を基準に算出されます。

$$p(\text{行の値}) * p(\text{列の値}) * \text{総レコード数}$$

- **残差：**確認された値と期待値との差。
- **行のパーセンテージ：**対応する列の値を持つ行値があるすべてのレコードのパーセンテージが含まれます。行内のパーセントの合計は 100 になります。
- **列のパーセンテージ：**対応する行の値を持つ列値があるすべてのレコードのパーセンテージが含まれます。列内のパーセントの合計は 100 になります。
- **合計パーセンテージ：**列値と行値の組み合わせを持つすべてのレコードのパーセンテージが含まれます。行列全体のパーセントの合計は 100 になります。
- **行および列合計を含める：**列および行の合計に対して、行列に行と列を加えます。
- **設定値の適用：**(出力ブラウザのみ) 出力ブラウザを閉じたり再度開いたりすることなく、クロス集計ノード出力の外観を変更することができます。出力ブラウザの [設定値の適用] を使用して変更し、このボタンをクリックして [クロス集計] タブを選択し、変更の影響を確認します。

クロス集計ノード出力ブラウザ

行列 ブラウザにはクロス表形式のデータが表示され、セルの選択、行列全体または一部のクリップボードへのコピー、選択された行列に基づく新しいノードの生成、および行列の保存と印刷など、行列に関するさまざまな操作を実行できます。クロス集計ブラウザは、Oracle の Naive Bayes モデルなどのモデルの出力を表示するためにも使用されます。

図 6-11
クロス集計ブラウザ

response_01

| ownpc | | 0 | 1 | 合計 |
|-------|----|----------|---------|------|
| 0 | 度数 | 1611 | 225 | 1836 |
| | 予測 | 1682.510 | 153.490 | 1836 |
| | 残差 | -71.510 | 71.510 | 0 |
| 1 | 度数 | 2971 | 193 | 3164 |
| | 予測 | 2899.490 | 264.510 | 3164 |
| | 残差 | 71.510 | -71.510 | 0 |
| 合計 | 度数 | 4582 | 418 | 5000 |
| | 予測 | 4582 | 418 | 5000 |
| | 残差 | 0 | 0 | 0 |

セルの内容: フィールドのクロス集計 (欠損値を含む)
カイ 2 乗 = 57.452, df = 1, 確率 = 0

OK(O)

[ファイル] と [編集] メニューは、印刷、保存、出力のエクスポート、データの選択とコピーの通常のオプションを提供します。 [詳細は、p. 436 出力を表示 を参照してください。](#)

カイ 2 乗: 2 つのカテゴリ フィールドのクロス集計表の場合、そのテーブルの下に大域的な Pearson カイ 2 乗も表示されます。この検定は、2 つのフィールドが関係がないことの確率を、確認されたカウントと関係がない場合の期待値との差に基づいて示します。たとえば、顧客満足度と店舗の場所の間に関係がない場合、同様の満足度を全店舗に対して期待します。しかし、特定の店舗における顧客が、一貫して他の店舗よりも高い満足度を表明している場合は、一致性に疑念を抱くかもしれません。その差が大きくなればなるほど、ただのサンプリングのミスであったという確率は小さくなります。

- カイ 2 乗テストは、2 つのフィールドが関係がなく、その中で、確認された頻度と期待頻度との間の差は、ただの見込みであることの確率を示します。この確率が非常に小さな場合は「一般的に 5 % 未満」で 2 つのフィールドの関係は有意であると言われます。
- 1 列と 1 行しかない場合（一元カイ 2 乗テスト）、自由度はセル数マイナス 1 です。二元カイ 2 乗テストの場合、自由度は行数マイナス列数の 1 倍マイナス 1 です。

- カイ 2 乗の統計量を解釈する場合、すべての期待されるセルの自由度が 5 未満であるかどうか注意します。
- カイ 2 乗テストは、2 つのフィールドがクロス表の場合にのみ利用できます。([すべてのフラグ] または [すべての数値] が [設定] タブで選択されている場合、このテストは表示されません)。

[ノードの生成] メニュー： [ノード生成] メニューには、ノード生成操作に関するメニュー項目があります。これらの操作は、クロス表行列でしか利用できません。また、行列中のセルを最低 1 つ以上選択している必要があります。

- **条件抽出ノード：** 行列で選択した任意のセルと一致するレコードを選択する、条件抽出ノードを生成します。
- **フィールド作成ノード (フラグ型)：** 新しいフラグ型フィールドを作成する、フィールド作成ノードを生成します。行列内で選択された任意のセルに一致するレコードには T、その他のレコードには F がフラグ型フィールドに含まれます。
- **フィールド作成ノード (セット型)：** 新しい名義型フィールドを作成する、フィールド作成ノードを生成します。名義型フィールドには、行列中の選択されたセルのそれぞれの連続セットに対して、1 つのカテゴリが含まれます。

精度分析ノード

精度分析ノードでは、正確な予測を生成するモデルの能力を評価することができます。分析ノードでは、1 つ以上のモデル ナゲットについて、予測値と実際値 (対象フィールド) をさまざまな方法で比較します。精度分析ノードを使用して、予測モデル同士を比較することもできます。

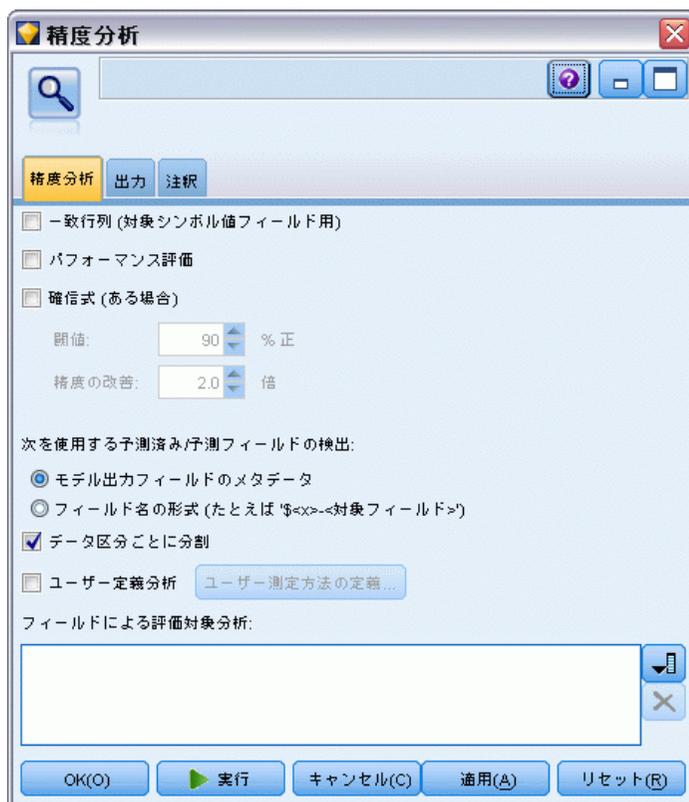
精度分析ノードを実行すると、実行ストリーム中のそれぞれのモデル ナゲットに対して、[要約] タブの [精度分析] に分析結果の要約が自動的に追加されます。精度分析の詳細な結果は、マネージャ ウィンドウの [出力] タブに表示されます。また、直接ファイルに書き込むこともできます。

注：精度分析ノードが予測値と実際値を比較するため、この 2 つの値は監視モデル (対象フィールドが必要なモデル) でのみ有用です。クラスタリング アルゴリズムなどの非監視モデルについては、比較のベースとして利用できる実際の結果はありません。

精度分析ノードの [精度分析] タブ

[精度分析] タブでは、精度分析の詳細を指定することができます。

図 6-12
精度分析ノード:[精度分析] タブ



一致行列 (シンボル対象またはカテゴリ対象): カテゴリ対象 (フラグ型、名義型、または順序型) の各生成 (予測) フィールドとその対象フィールド間の一致パターンを示します。実際値で構成される行と予測値で構成される列から成るテーブルが作成されます。各セルには、そのパターンを含むレコード数が表示されます。これは、予測時の系統誤差を判別する場合に役に立ちます。異なるモデルによって生成された複数の生成フィールドが、同じ出力フィールドに関連している場合は、これらのフィールドが一致する場合と一致しない場合がカウントされ、合計に表示されます。一致する場合は、別の正/誤統計が表示されます。

パフォーマンス評価: カテゴリ出力を行うモデルのパフォーマンス評価統計量を表示します。この統計量は、出力フィールドの各カテゴリに対して報告され、そのカテゴリに属するレコードを予測するためにモデルの平均情報量 (ビット数) を測定します。分類の難しさを考慮して、まれなカテゴリについて正確な予測を行うために、一般的なカテゴリの予測時よりも高いパフォーマンス評価インデックスが与えられます。あるカテゴリに関するモデルの予測が推量にすぎない場合は、そのカテゴリのパフォーマンス評価インデックスは 0 になります。

確信式 (ある場合): 確信度フィールドを生成するモデルの場合に、確信度の値およびその値と予測値の関係に関する統計量が報告されます。この項目には 2 つの設定があります。

- **閾値:** 精度が指定されたパーセントに達する確信度レベルを報告します。
- **精度の改善:** 指定した因子によって精度が改善される確信度レベルを報告します。たとえば、全体的な精度が 90% で、このオプションを 2.0 に設定した場合、報告される値は、95% の精度を達成するのに必要な確信度です。

次を使用する予測済み/予測フィールドの検出: 予測フィールドが元の対象フィールドにどのように一致するかを指定します。

- **モデル出力フィールドのメタデータ:** モデル フィールド情報に基づいて、予測フィールドを対象に一致させます。予測フィールドの名前が変更されている場合でも一致は可能です。予測フィールドのモデル フィールド情報は、データ型ノードを使用して、[値] ダイアログ ボックスからアクセスすることができます。 [詳細は、4 章 p.159 \[値\] ダイアログ ボックスの使用 を参照してください。](#)
- **フィールド名フォーマット:** 名前の表記方法に基づいて、フィールドに一致します。たとえば、レスポンスという名の対象に C5.0 モデル ナゲットが生成した予測値が、\$C-response というフィールド内にある必要があります。

データ区分によって分割: レコードを、学習、テスト、および検定用の各サンプルに分割するために データ区分フィールドが使用される場合、このオプションを選択すると、各データ区分ごとに、別々の結果が表示されます。 [詳細は、4 章 p.229 データ区分ノード を参照してください。](#)

注： データ区分を分割する場合、データ区分フィールドにあるヌル値を持つレコードは、分析から除外されます。データ区分ノードは、ヌル値を生成しないため、データ区分ノードを使用している場合は、問題になりません。

ユーザー定義分析: 各自のモデル評価に使用する独自の分析計算式を指定できます。各レコードに対して何を計算するのかを CLEM 式を使って指定し、さらにレコード ベルのスコアを全体的なスコアに組み込む方法を指定します。また、@TARGET 関数と @PREDICTED 関数を使用して、対象 (実際の出力) 値と予測値をそれぞれ参照します。

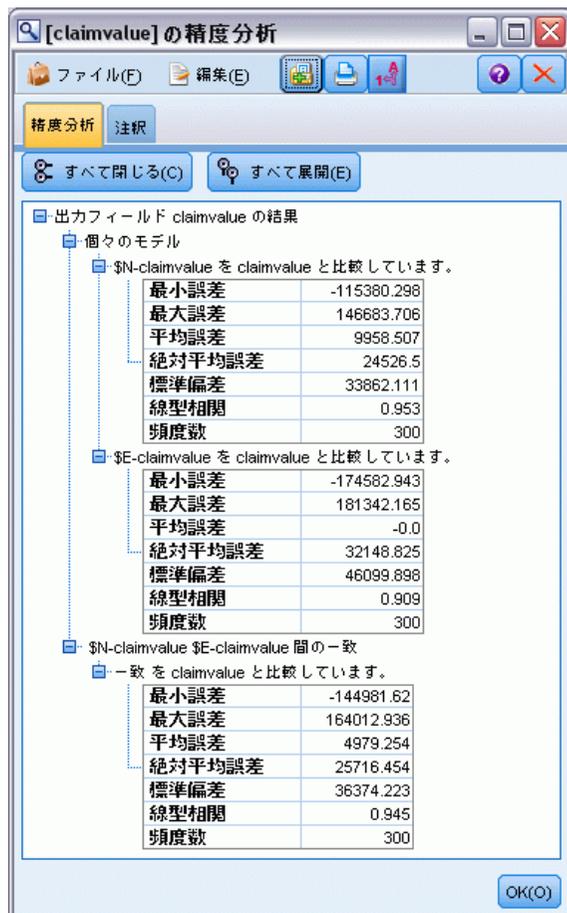
- **If:** 条件に応じて異なる計算を使う必要がある場合は、条件式を指定します。
- **Then:** IF 条件が真 (true) の場合に実行する計算式を指定します。
- **Else:** IF 条件が偽 (false) の場合に実行する計算式を指定します。
- **Use:** 個々のスコアから全体的なスコアを算出するための統計量を選択します。

フィールドによる評価対象分析： 精度分析の対象として使用できるカテゴリフィールドを表示します。全体的な分析に加えて、各対象フィールドのカテゴリごとに個別に分析を行うこともできます。

精度分析出力ブラウザ

精度分析出力ブラウザでは、精度分析ノードの実行結果を参照することができます。通常の保存、エクスポート、および印刷操作は、[ファイル]メニューから行うことができます。 [詳細は、 p. 436 出力を表示を参照してください。](#)

図 6-13
精度分析出力ブラウザ



精度分析の出力を初めて参照する場合、結果は展開されて表示されています。参照し終わった後に結果を隠すには、項目の左側にある拡張をコントロールを使って目的の結果を閉じるか、または [すべて閉じる] ボタンをク

リックしてすべての結果を非表示にします。閉じた結果をもう一度表示するには、項目の左側にある拡張コントロールを使用するか、または [すべて展開] ボタンをクリックしてすべての結果を表示してください。

出力フィールドの結果： 精度分析の出力には、各出力フィールドに対するセクションが含まれています。ここには、生成されたモデルが作成した、対応する予測フィールドが格納されています。

比較： 出力フィールド セクション内には、出力フィールドと関連付けられた各予測フィールドに対するサブセクションがあります。カテゴリ出力フィールドの場合、このセクションの最上位レベルにあるテーブルには、正/誤の予測数とパーセント、およびストリーム中の総レコード数が表示されます。数値出力フィールドの場合、このセクションには次の情報が表示されます。

- **最小誤差：** 最小誤差を表示します（観測値と予測値の差異）。
- **最大誤差：** 最大誤差を表示します。
- **平均誤差：** すべてのレコードの平均誤差を表示します。これは、モデル中に系統**バイアス**（過小な推定よりも過大な推定を行う傾向がある、またはその逆）があるかどうかを示します。
- **絶対平均誤差：** すべてのレコードの誤差の絶対値の平均を表示します。この値は、方向に関係ない絶対値の平均誤差を示しています。
- **標準偏差：** 誤差の標準偏差を表示します。
- **線型相関：** 予測値と実際の値の間の線型相関を表示します。統計量は $-1.0 \sim 1.0$ の範囲の値になります。 $+1.0$ に近い値は強い正の相関を表し、そのため、高い予測値が高い実際値と関連付けられ、低い予測値が低い実際値と関連付けられます。 -1.0 に近い値は強い負の相関を表し、高い予測値が低い実際値と、低い実際値が高い予測値と関連付けられます。 0.0 に近い値は弱い相関を表し、予測値と実際の値は多かれ少なかれ独立しています。注：空白のエントリは、実際の値または予測値が定数であるため、この場合は線型相関を計算できないことを示します。
- **頻度数：** 精度分析に使用したレコード数を表示します。

一致行列： カテゴリ出力フィールドに対して、精度分析オプションで一致行列を要求している場合、ここに行列が表示されます。行は実際の観測値を、列は予測値を表します。テーブル中のセルは、それぞれの予測値と実際の値の組み合わせのレコード数を示しています。

パフォーマンス評価： カテゴリ出力フィールドに対して、精度分析オプションでパフォーマンス評価統計量を要求している場合、ここにパフォーマンス評価の結果が表示されます。各出力カテゴリが、パフォーマンス評価統計量とともに表示されます。

確信度値レポート： シンボル値の出力フィールドに対して、精度分析オプションで確信度値を要求している場合、その値がここに表示されます。モデルの確信度値に対して、次の統計量が報告されます。

- **範囲**：ストリーム データ中のレコードに対する確信度値の集計範囲（最小値と最大値）を表示します。
- **正解の平均値**：正しく分類されたレコードの確信度の平均値を表示します。
- **誤りの平均値**：誤って分類されたレコードの確信度の平均値を表示します。
- **常に正解**：予測が常に正解で、ケースのパーセンテージがこの基準を満たす確信度の閾値を表示します。
- **常に不正解**：予測が常に不正解で、ケースのパーセンテージがこの基準を満たす確信度の閾値を表示します。
- **精度 X% 以上**：精度が X% 以上の確信度レベルを表示します。X はほぼ、[精度分析] オプションの [閾値] で指定された値になります。一部のモデルやデータセットでは、オプションに指定された正しい閾値を与える確信度値を選択できません（閾値近くで同じ確信度値を持つ類似ケースのクラスタのためなど）。報告される閾値は、単一の確信度値閾値から取得できる、指定された精度基準に最も近い値になります。
- **分割 X 以上**：データセットの総体的な精度よりも X 倍良好な精度の確信度値を表示します。X は、[精度分析] オプションの [精度の改善] で指定された値になります。

間的一致：ストリーム中に、同じ出力フィールドを予測する 2 つ以上の生成されたモデルがある場合、モデルが生成した予測間の一一致に関する統計量も表示されます。これには、予測が一致したレコード数とパーセンテージ（カテゴリ型出力フィールドの場合）、または誤差要約統計量（連続型出力フィールドの場合）が含まれます。カテゴリ フィールドの場合、モデルが同意した（同じ予測値を生成）レコードのサブセットの実際の値と比較した予測の精度分析が含まれます。

データ検査ノード

データ検査ノードは、IBM® SPSS® Modeler に取り込むデータを広範に検査するための手段を提供し、把握しやすい行列形式で表示され、簡単にソートしたり、グラフやデータ準備ノードを生成することができます。

図 6-14
データ検査ブラウザ

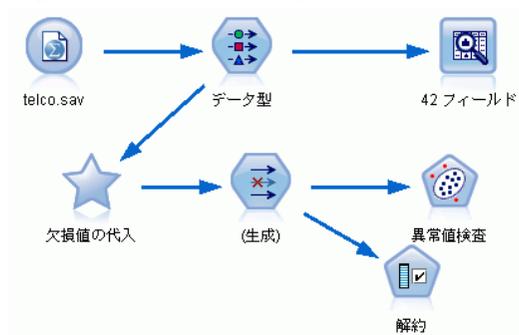


- [検査] タブでは要約統計量、ヒストグラム、棒グラフを含むレポートを表示し、データにの予備調査を高めるのに有効です。レポートでは、フィールド名の前にストレージアイコンも表示します。
- 検査レポートの [欠損値検査] タブは、外れ値、極値、および欠損値に関する情報を示し、これらの値を処理するためのツールを提供します。

データ検査ノードの使用

データ検査ノードは、入力ノードやインスタンス化されたデータ型ノードの下流に直接接続することができます。また、結果に基づき、多くのデータ準備ノードを生成することもできます。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、7 章 [分析用のデータの準備 \(データ検査\) in IBM SPSS Modeler 15 アプリケーションガイド](#) を参照してください。

図 6-15
欠損値スーパーノードによるストリーム



データのスクリーニングまたはサンプリング 「大容量データ」を処理する際、最初の検査が特に効果的であるため、サンプル ノードを使用し、レコードのサブ セットのみを選択することによって最初の検査にかかる処理時間を短縮することができます。また、データ検査ノードは、分析の調査段階でフィールド選択や異常値検出ノードなどと連携して使用することができます。

データ検査ノードの [設定] タブ

[設定] タブで、検査のための基本パラメータを指定することができます。

図 6-16
データ検査ノード:[設定]タブ



デフォルト: ストリームにノードを単純に接続し、[実行] をクリックして、デフォルト設定に基づいて全フィールドに対する検査レポートを生成することができます。たとえば次の通りです。

- データ型ノードの設定がない場合は、レポートにはすべてのフィールドが含まれます。
- インスタンス化されているかどうかにかかわらず、データ型の設定がある場合は、すべての [入力]、[対象]、および [両方] フィールドが含まれます。[対象] フィールドが 1 つある場合は、それがオーバーレイフィールドとして使用されます。複数の [対象] フィールドが指定されている場合は、デフォルトのオーバーレイは指定されません。

ユーザー設定フィールドを使用: このオプションを使用すると、フィールドを手動で選択できます。右側にあるフィールド選択ボタンを使って、フィールドを個別に、またはデータ型により選択します。

オーバーレイフィールド。 オーバーレイ フィールドは、検査レポートに表示されるサムネイル グラフの描画に使用されます。連続型 (数値範囲型) フィールドの場合、二変数の統計値 (共分散および相関) も計算されます。データ型ノード設定に基づいて対象フィールドがひとつ表示された場合、上記に表示されている通り、デフォルトのオーバーレイ フィールドと

して使用されます。代わりに、[ユーザー設定フィールドを使用] を選択し、オーバーレイを指定することもできます。

表示： グラフが出力可能か指定したり、デフォルトで表示される統計値を選択したりすることができます。

- **グラフ：** 選択された各フィールドに、データの必要に応じて棒グラフ、ヒストグラム、散布図のいずれかのグラフを表示します。グラフは、最初のレポートではサムネイルで表示されますが、フルサイズのグラフおよびグラフ ノードも生成されます。 [詳細は、 p. 465 データ検査出力ブラウザ を参照してください。](#)
- **基本/詳細統計量：** デフォルトで出力表示される統計量のレベルを指定します。この設定で最初の表示を定義しますが、すべての統計がこの設定と関係なく出力できます。 [詳細は、 p. 467 統計の表示 を参照してください。](#)

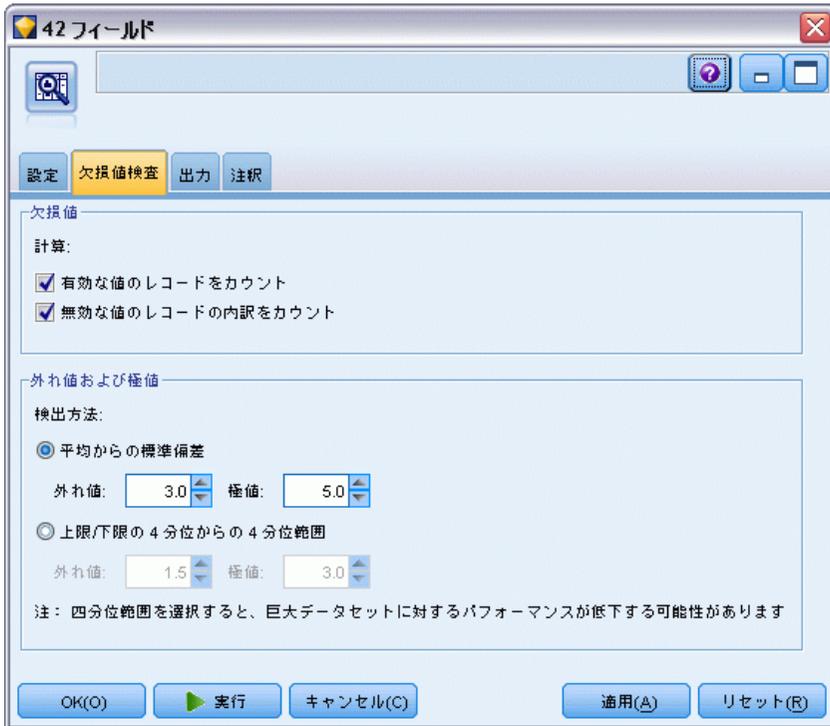
中央値と最頻値： レポートの全フィールドに対し、中央値と最頻値を計算します。大容量のデータセットでは、他の統計より計算に時間がかかるため、これらの統計により多くの処理時間がかかることがあります。中央値のみの場合、報告された値は、いくつかのケースのうち（データセット全体ではなく）2000 レコードのサンプルに基づいています。このサンプリングは、メモリが拡張されない場合、フィールドあたりの基準で行われます。サンプリングが実行されている場合、結果は出力（中央値のみではなくサンプル中央値）などでラベル付けされます。中央値以外の統計値はすべて、常に完全なデータセットを使用して計算されます。

空またはデータ型不明のフィールド： インスタンス化されたデータとともに使用した場合、データ型不明のフィールドは検査レポートには含まれていません。データ型不明のフィールド（空のフィールド含む）を含むには、データ型ノードのいずれかの上流で [すべての値の消去] を選択します。これにより、データはインスタンス化されず、すべてのフィールドがレポートに含まれます。たとえば、すべてのフィールドの完全なリストを取得したり、空のフィールドを除外するフィルタ ノードを生成する場合、このオプションが役に立ちます。 [詳細は、 p. 473 欠損データを含むフィールドのフィルタリング を参照してください。](#)

データ検査の [欠損値検査] タブ

データ検査ノードの [欠損値検査] タブは、欠損値や外れ値、極値を処理するためのオプションを提供します。

図 6-17
データ検査ノードの [欠損値検査] タブ



欠損値

- **有効な値のレコードをカウント**：各評価フィールドに対して、有効な値を持つレコード数を表示する場合に選択します。ヌル（未定義の）値、空白値、ホワイトスペースや空の文字列は、常に無効な値として処理されます。
- **無効な値のレコードの内訳のカウント**：各フィールドに対して、各種の不正な値を持つレコード数を表示する場合に選択します。

外れ値および極値

外れ値および極値の検出方法次の 2 つの方法がサポートされています。

平均からの標準偏差：平均からの標準偏差に基づいて、外れ値や極値を検出します。たとえば、フィールドの平均 100 で標準偏差が 10 である場合、3.0 を指定して、70 未満および 130 を超える値が外れ値として扱われているかを表示することができます。

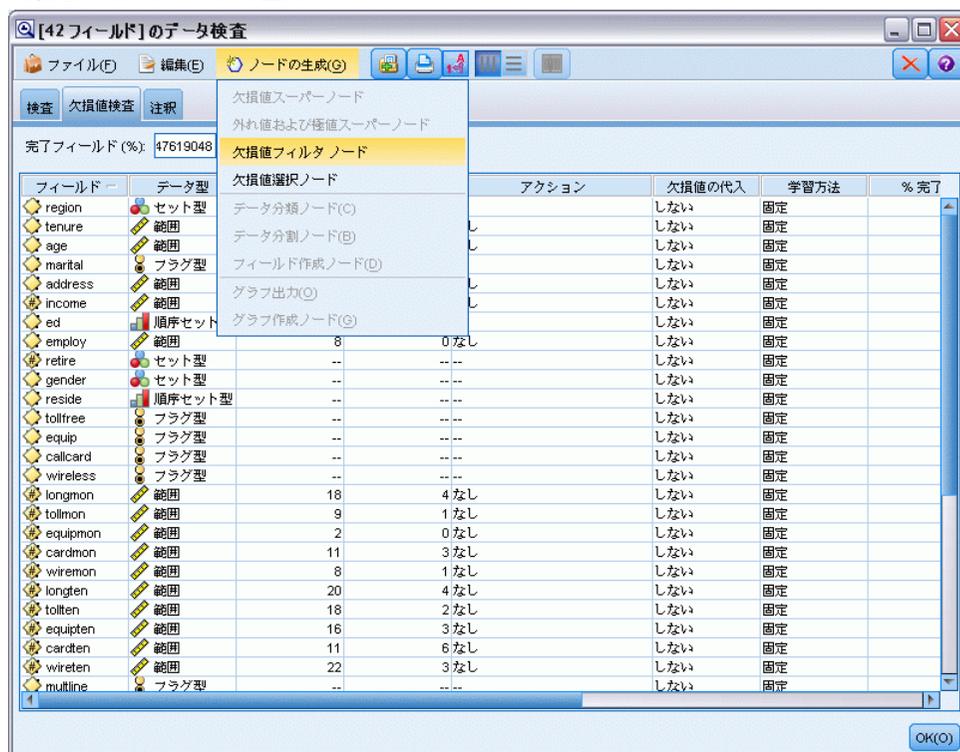
4 分位範囲：4 分位範囲に基づいて外れ値や極値を検出、その範囲は 2 つの中央分位が存在する範囲です（25 ～ 75 分位の間）。たとえば、デフォルト設定の 1.5 に基づき、外れ値の下限閾値は $Q1 - 1.5 * IQR$ となり、

上限閾値は $Q3 + 1.5 * IQR$ となります。このオプションを使用すると、大容量のデータセットに対するパフォーマンスが遅くなります。

データ検査出力ブラウザ

データ検査ブラウザは、データの概要を把握するための強力なツールです。[欠損値検査] タブでは外れ値、極値、欠損値についての情報を表示しますが、[検査] タブでは、全フィールドのサムネイル グラフやストレージアイコン、統計を表示します。初期グラフおよび要約統計量に基づいて、数値フィールドの記録、新規フィールドの作成、または名義型フィールドの値の再分類などの作業を行うことができます。また、さまざまな機能を使ってデータを視覚化し、より詳細にデータを探索することもできます。[ノードの生成] メニューを使用して、検査レポート ブラウザから直接多くのノードを作成し、データの変換や視覚化に使用することができます。

図 6-18
欠損値フィルタ ノードの生成



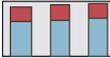
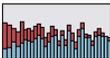
- 列のヘッダをクリックして列をソートしたり、ドラッグ アンド ドロップを使って列の並べ替えることができます多くの標準出力処理も、サポートされています。詳細は、[p. 436 出力を表示](#)を参照してください。

- [尺度] 列または [一意] 列のフィールドをダブル クリックして、フィールドの値と範囲を参照する。
- ツールバーまたは [編集] メニューをクリックし、値のレベルを表示または非表示にしたり、表示する統計値を選択することができます。 [詳細は、 p. 467 統計の表示 を参照してください。](#)
- フィールド名の左側にあるストレージ アイコンを確認します。ストレージは、フィールド中へのデータの格納方法を表しています。たとえば、1 と 0 の値をとるフィールドは整数データを格納します。これは、データの使用法を記述してストレージに影響を与えない、測定レベルとは異なります。 [詳細は、 2 章 p. 35 フィールドのストレージと形式の設定 を参照してください。](#)

グラフの表示および生成

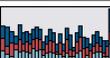
オーバーレイ フィールドが選択されていない場合は、[検査] タブで棒グラフ（名義型またはフラグ型）またはヒストグラム（連続型）を表示します。

図 6-19
オーバーレイ フィールドなしの検査結果の例

| フィールド | グラフ | データ型 | 最小値 | 最大値 | 平均値 | 標準偏差 | 歪度 | カテゴリ数 | 有効 |
|-------|--|------|-----|-----|--------|--------|-------|-------|------|
| 地域 |  | セット型 | 1 | 3 | -- | -- | -- | 3 | 1000 |
| 期間 |  | 範囲 | 1 | 72 | 35.526 | 21.360 | 0.112 | -- | 1000 |

名義型またはフラグ型フィールドのオーバーレイの場合、グラフはオーバーレイの値で色分けされます。

図 6-20
名義型フィールドのオーバーレイの検査結果例

| フィールド | グラフ | データ型 | 最小値 | 最大値 | 平均値 | 標準偏差 | 歪度 | カテゴリ数 | 有効 |
|-------|---|------|-----|-----|--------|--------|-------|-------|------|
| 地域 |  | セット型 | 1 | 3 | -- | -- | -- | 3 | 1000 |
| 期間 |  | 範囲 | 1 | 72 | 35.526 | 21.360 | 0.112 | -- | 1000 |

連続型フィールドのオーバーレイの場合、1 次元の棒グラフやヒストグラムの代わりに、2 次元の散布図が生成されます。この場合、x 軸がオーバーレイ フィールドにマップされ、すべての x 軸を同じ尺度で参照することができます。

図 6-21
連続型フィールドのオーバーレイの検査結果例

| フィールド | グラフ | データ型 | 最小値 | 最大値 | 平均値 | 相関 | 相関 T | 相関 T 自由度 |
|-------|---|------|-----|-----|--------|-------|--------|----------|
| 地域 |  | セット型 | 1 | 3 | -- | -- | -- | -- |
| 期間 |  | 範囲 | 1 | 72 | 35.526 | 0.490 | 17.768 | 998.000 |

- フラグ型または名義型フィールドの場合、カーソルをバーの上当てると、ツール ヒントの潜在的な値またはラベルを表示します。
- フラグ型または名義型フィールドの場合、ツールバーを使用して、サムネイル グラフの位置を水平方向から垂直方向に切り替えることができます。
- フルサイズのグラフをサムネイルから生成するには、サムネイルをダブル クリックするか、サムネイルを選択して [ノードの生成] メニューから [グラフの出力] を選択します。注： サンプルされたデータに基づくサムネイル グラフの場合、元のデータ ストリームが開かれていれば、生成されるグラフにはすべてのケースが含まれます。

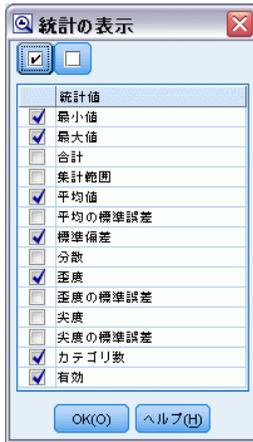
出力を作成したデータ検査ノードはストリームに接続している場合にのみグラフを生成できます。

- 一致するグラフ ノードを生成するには、[検査] タブで 1 つ以上のフィールドを選択し、[ノードの生成] メニューから [グラフノード] を選択します。結果ノードがストリーム領域に追加され、そのノードを使用して、ストリームが実行されるごとにグラフを再作成することができます。
- オーバーレイ セットに 100 個を超える値がある場合、警告メッセージが表示され、オーバーレイは入れられません。

統計の表示

[統計の表示] ダイアログ ボックスで、[検査] タブに表示する統計を選択することができます。初期設定は、データ検査ノードで指定されています。詳細は、[p. 461 データ検査ノードの \[設定\] タブ](#) を参照してください。

図 6-22
統計の表示



最小値. 数値型変数がとる最も小さい値。

最大. 数値型変数の最大値。

合計. 欠損値のないすべてのケースに対する変数の値の合計または全体。

範囲. 数値型変数の最大値と最小値の差。

平均. 中心傾向の測定値。観測値の合計をケース数で割った算術平均。

平均値の標準誤差. 同一の分布から取り出したサンプル間で平均値がどの程度ばらついているかを測ったもの。観測された平均と仮説された値を比較するために使うことができます (すなわち、差と標準誤差の比率が -2 より小さいか $+2$ より大きい場合に、2 つの値は異なっていると結論付けることができます)。

標準偏差. 平均の周りの散らばり度です。分散の平方根に等しくなります。標準偏差は元の変数と同じ単位で表されます。

分散 (信頼性分析). 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

歪度. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

歪度の標準誤差. 標準誤差に対する歪度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか $+2$ より大きい場合は、正規性を棄却することができます)。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

尖度. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。

尖度の標準誤差. 標準誤差に対する尖度の比率は、正規性の検定として使うことができます（すなわち、比率が -2 より小さいか $+2$ より大きい場合は、正規性を棄却することができます）。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します（箱型の一様分布の裾のようになります）。

一意的. あらゆるタイプの他のすべての効果を調整することによって、すべての効果を同時に評価します。

有効. システム欠損値またはユーザー欠損として定義された値をもたない有効なケース。

中央値. ケースの中央付近にある値です。50 パーセントایلです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最頻値. 最も多く出現する値。複数の値が最高の頻度で出現する場合は、それぞれが最頻値となります。

パフォーマンス改善のために中央値と最頻値はデフォルトで抑制されていますが、データ検査ノードの [設定] タブで選択することができます。 [詳細は、 p. 461 データ検査ノードの \[設定\] タブ を参照してください。](#)

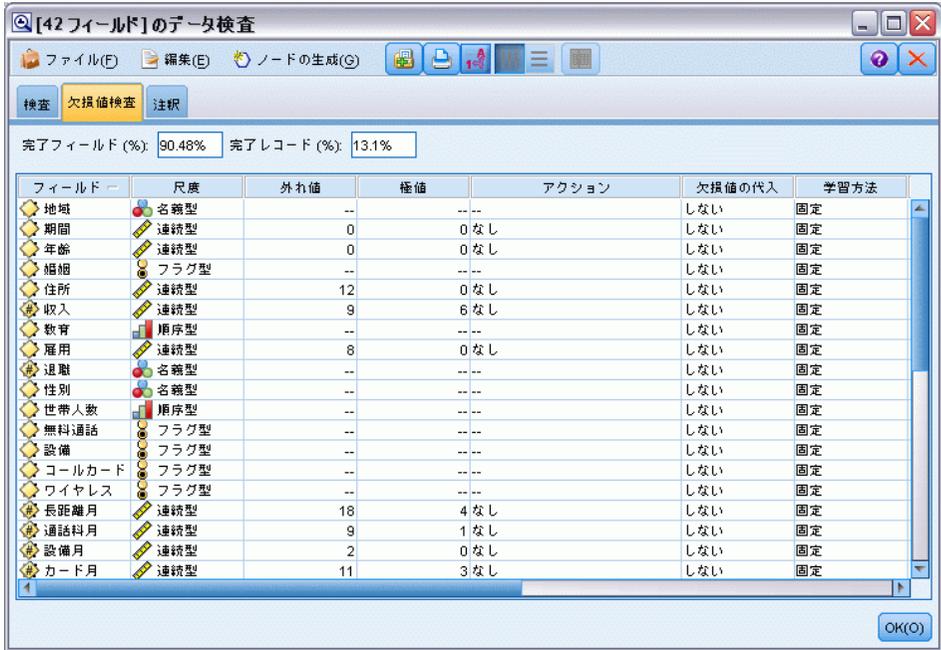
オーバーレイの統計

連続型（数値範囲型）のオーバーレイ フィールドが使用されている場合、次の統計も利用可能です。

共分散. 2 変数間の標準化されていない関連度。偏差の積和を $N-1$ で割ったものです。

データ検査ブラウザの [欠損値検査] タブ

図 6-23
データ検査ブラウザの品質レポート



データ検査ブラウザの [欠損値検査] タブは、データ品質分析の結果を表示し、外れ値、極値および欠損値の処理を指定します。

値の代入

検査レポートは各フィールドの完全なレコードの割合を有効値、ヌル値、空白値の数とともに一覧表示します。必要時にこのオプションを選択して、特定のフィールドに欠損値を代入し、スーパー ノードを生成してこれらの変換に適用することができます。

- ▶ [欠損値の代入] 列では、欠損値が存在した場合に代入する値の種類を指定します。このオプションを選択して、空白値、ヌル値または両方を代入し、代入する値を選択するユーザー設定条件や式を指定します。

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 p. 159 [値] ダイアログ ボックスの使用 を参照してください。

- ▶ [方法] の列では、使用する方法を指定します。

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。

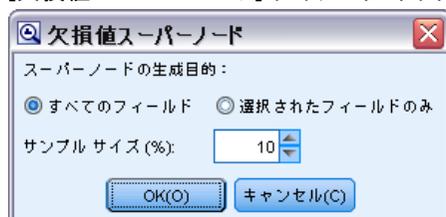
式：ユーザー設定の式を指定することができます。たとえば、値をグローバル値設定ノードで作成されたグローバル変数と置き換えることができます。

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。

- ▶ 欠損値スーパー ノードを生成するには、メニューから次の通り選択します。

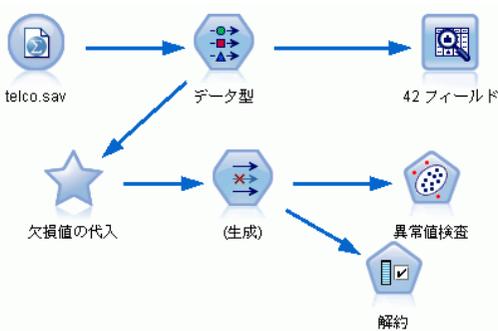
ノードの生成 > 欠損値スーパーノード

図 6-24
[欠損値スーパー ノード] ダイアログ ボックス



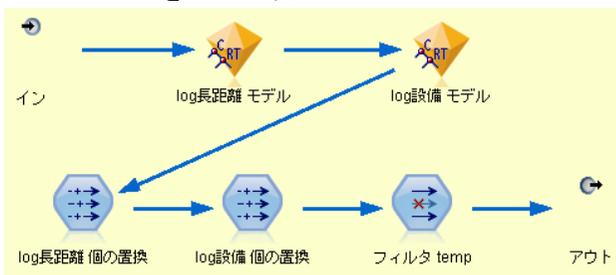
- ▶ [すべてのフィールド] または [選択されたフィールドのみ] を選択し、必要があればサンプル サイズを指定します。(指定のサンプルは割合で、デフォルトで全レコードの 10% がサンプルとなります。)
- ▶ [OK] をクリックして、生成されたスーパーノードをストリーム領域に追加します。
- ▶ スーパーノードをストリームに接続させ、変換を適用させます。

図 6-25
スーパーノードのストリームへの追加



スーパーノード内では、モデル ナゲット、置換、および置換ノードの組み合わせが必要に応じて使用されます。スーパーノードを編集して [ズームイン] をクリックして、どのように動作するか確認することができます。また、スーパーノード内の特定ノードを追加、編集、削除して、動作を調整することができます。

図 6-26
スーパーノードをズームイン



外れ値および極値の処理

検査レポートは多くの外れ値および極値を一覧表示し、データ検査ノードで指定された [検出] オプションに基づいて、各フィールドに一覧表示されます。詳細は、[p. 463 データ検査の \[欠損値検査\] タブ](#) を参照してください。必要時にこのオプションを選択して、特定のフィールドにこ

これらの値を強制、削除、無効化し、スーパーノードを生成してこれらの変換に適用することができます。

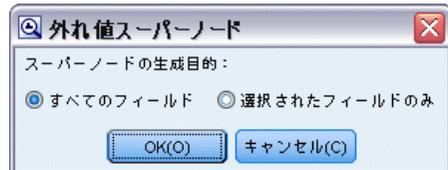
- ▶ [アクション] 列では、必要に応じて特定のフィールドに対する外れ値および極値の処理を指定します。

外れ値および極値の処理には、次のアクションが有効です。

- **強制**：外れ値および極値を、極端とはみなされない直近の値と置換します。たとえば、外れ値が標準偏差 3 を上回るまたは下回ると定義されている場合、外れ値はこの領域内の最大値または最低値と置き換えられます。
- **破棄**：指定されたフィールドに対し、範囲外の値または極値を含むレコードを破棄します。
- **無効**：外れ値および極値を、ヌル値またはシステム欠損値と置き換えます。
- **外れ値の強制/極値の廃棄**：極値のみを破棄します。
- **外れ値の強制/極値の無効化**：極値のみを無効にします。

- ▶ スーパーノードを生成するには、メニューから次の通り選択します。
ノードの生成 > 外れ値および極値スーパーノード

図 6-27
[外れ値スーパーノード] ダイアログ ボックス



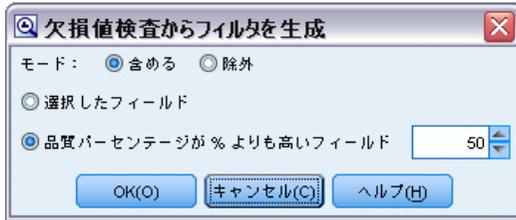
- ▶ [すべてのフィールド] または [選択したフィールドのみ] を選択し、[OK] をクリックして、生成されたスーパーノードをストリーム領域に追加します。
- ▶ スーパーノードをストリームに接続させ、変換を適用させます。

必要に応じて、スーパーノードを編集し、ズームインして表示または変更することができます。スーパーノードでは、必要に応じて、一連の条件抽出ノードおよび置換ノードを使用して、値を破棄、強制または無効化します。

欠損データを含むフィールドのフィルタリング

データ検査ブラウザから、欠損値検査の結果に基づいて新しいフィルタノードを生成することができます。

図 6-28
[欠損値検査からフィルタを生成] ダイアログ ボックス



モード: 指定したフィールドに対する操作として、[含める] または [除外] を選択します。

- **選択したフィールド:** フィルタ ノードは、欠損値検査テーブルで選択されたフィールドを含めるか、または除外します。たとえば、[% 完了の列] でテーブルをソートし、Shift キーを押したままクリックして、最小の完了フィールドを選択し、これらのフィールドを除外するフィルタ ノードを生成することができます。
- **品質パーセンテージが % よりも高いフィールド:** フィルタ ノードは、完全なレコードの割合が指定した閾値よりも大きい場合に、フィールドを含めるか、または除外します。デフォルトの閾値は 50 % です。

空のまたはデータ型不明フィールドのフィルタリング

データ値がインスタンス化された後、データ型不明または空のフィールドは検査結果および IBM® SPSS® Modeler の多くのその他の出力結果から除外されます。これらのフィールドは、モデル作成のためには無視されますが、データを拡大または拡散させる場合があります。その場合、データ検査ブラウザを使用して、フィルタ ノードを生成し、ストリームからこれらのノードを削除することができます。

- ▶ 空のまたはデータ型不明のフィールドを含むすべてのフィールドが、検査に含まれていることを確かめるには、上流のノースまたはデータ型ノードで [すべての値の消去] を選択するか、値をすべてのフィールドの <パス> に設定します。
- ▶ データ検査ブラウザでは、[% 完了] 列でソートし、ゼロの有効な値（またはその他の閾値）を含むフィールドを選択し、[ノード生成] メニューでストリームに追加することのできるフィルタ ノードを生成します。

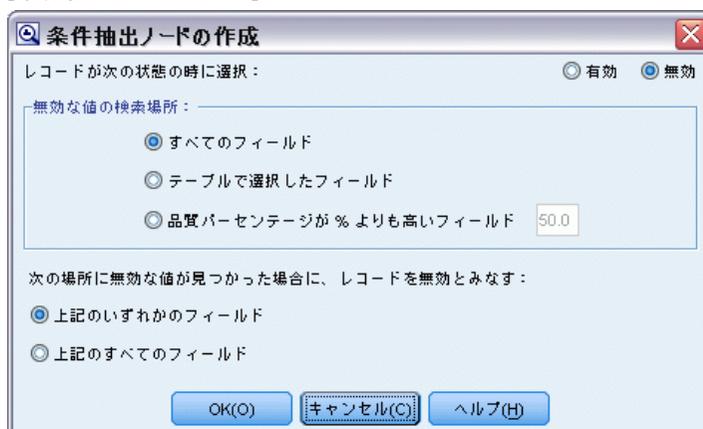
欠損データを含むレコードの選択

データ検査ブラウザから、欠損値検査の結果に基づいて新しい条件抽出ノードを生成することができます。

- ▶ データ検査ブラウザで、[品質] タブを選択します。

- ▶ メニューから次の項目を選択します。
ノードの生成 > 欠損値選択ノード

図 6-29
[条件抽出ノードの生成] ダイアログ ボックス



レコードが次の状態の時に選択： レコードが [有効] または [無効] の場合に保持することを指定します。

無効な値の検索場所： 無効な値の検索場所を指定します。

- **すべてのフィールド：** 条件抽出ノードは、すべてのフィールドに対して無効な値があるかどうかを検査します。
- **テーブルで選択したフィールド：** 条件抽出ノードは、現在欠損値検査出力テーブルで選択されているフィールドだけを検査します。
- **品質パーセンテージが % よりも高いフィールド：** 完全なレコードの割合が、指定した閾値よりも大きい場合に、条件抽出ノードはフィールドを検査します。デフォルトの閾値は 50 % です。

次の場所に無効な値が見つかった場合に、レコードを無効と見なす： レコードを無効と見なす条件を指定します。

- **上記のいずれかのフィールド：** 上で指定されたいずれかのフィールドに、そのレコードの不正な値があった場合に、レコードが不正と見なされます。
- **上記のすべてのフィールド：** 上で指定されたすべてのフィールドに、そのレコードの不正な値があった場合にだけ、レコードが不正と見なされます。

データ準備用のその他のノードの生成

データ分類ノード、データ分割ノード、フィールド作成ノードなど、データの準備に使用するさまざまなノードを、データ検査ブラウザから直接生成することができます。次に例を示します。

- `claimvalue` と `farmincome` の値に基づいて新しいフィールドを作成するには、検査レポートからこれらのフィールドを選択し、[生成] メニューの [フィールド作成] を選択します。ストリーム領域に新しいノードが追加されます。
- 同様に、検査結果に基づいてより詳細な分析を行うために、パーセントイルに基づいたビンに `farmincome` を記録することもできます。データ分割ノードを生成するには、表示されているフィールド行を選択した後、[生成] メニューの [データ分割] を選択します。

ノードが生成されて、ストリーム領域に追加されたら、そのノードをストリームに接続した後、選択したフィールドに関するオプションを指定する必要があります。

変換ノード

入力フィールドの正規化は、回帰、ロジスティック回帰、判別分析など、既存のスコアリング時技術を使用する前の重要なステップです。これらの技術は、多くの生データ ファイルにとっては真 (`true`) ではないデータの正常な分散に関する仮説を実行します。実際のデータの処理に対する 1 つのアプローチは、より正常な分散へ生データの要素を移動させる変換を適用することです。また、正規化フィールドは、用意にお互いを比較することができます。たとえば、収入および年齢は生データ ファイルではまったく異なるスケールですが、正規化されるとそれぞれの関連する影響を、容易に解釈することができます。

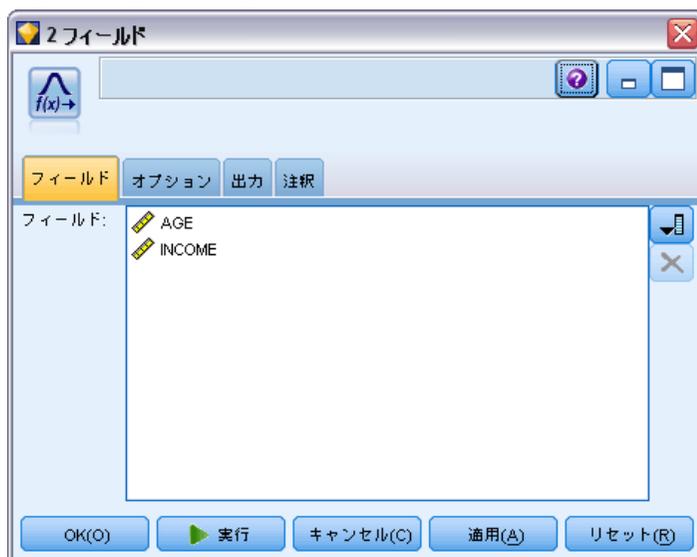
変換ノードを使用すると、出力ビューアで、最高の変換に対する評価を高速かつ視覚的に実行することができます。変数が正常に分散しているか確認したり、必要に応じて適用する変換を選択することができます。複数のフィールドを選択し、フィールドごとに 1 つの変換を実行することができます。

フィールドに対し優先的な変換を選択した後、フィールド作成ノードまたは置換ノードを生成して、変換を実行し、これらのノードをストリームを接続することができます。置換ノードは既存のフィールドを変換しますが、フィールド作成ノードを使用すると新規フィールドを作成できます。詳細は、[p. 481 グラフの生成](#) を参照してください。

変換ノードの [フィールド] タブ

[フィールド] タブでは、可能な変換を表示し適用させるために使用するデータのフィールドを指定します。数値的フィールドのみが変換できます。フィールド選択ボタンをクリックし、表示されたリストから 1 つ以上の数値フィールドを選択します。

図 6-30
変換ノード:[フィールド] タブ



変換ノードの [オプション] タブ

[オプション] タブでは、実行したい変換の種類を指定することができます。このタブを選択すると、利用可能な変換をすべて含むことができ、または個別に変換を選択することができます。

個別に選択する場合は、逆変換またはログ変換のためにデータをオフセットする数を入力することもできます。この処理は、ゼロを多く含むデータの大部分が、平均値や標準偏差にバイアスがある状況において役に立ちます。

たとえば、ゼロの値をいくつか含む [BALANCE] という名前のフィールドがあり、逆変換を実行したいと仮定します。不要なバイアスを避けるには、**逆数** ($1/x$) を選択し、[データオフセットの使用] フィールドに 1 を入力します。(このオフセットは、IBM® SPSS® Modeler の @OFFSET シーケンス機能が実行するものとは関連しません。)

図 6-31
変換ノード:[オプション] タブ



すべての公式：利用可能な変換が計算され出力結果に表示されることを指示します。

公式の選択：計算され出力結果に表示される異なる変換を選択することができます。

- **逆数 (1/x)**：逆変換が出力結果に表示されることを指示します。
- **対数 (log n)**：対数_n変換が出力結果に表示されることを指示します。
- **対数 (log 10)**：対数₁₀変換が出力結果に表示されることを指示します。
- **指数**：指数変換 (e^x) が出力結果に表示されることを指示します。
- **平方根**：平方根変換が出力結果に表示されることを指示します。

変換ノードの [出力] タブ

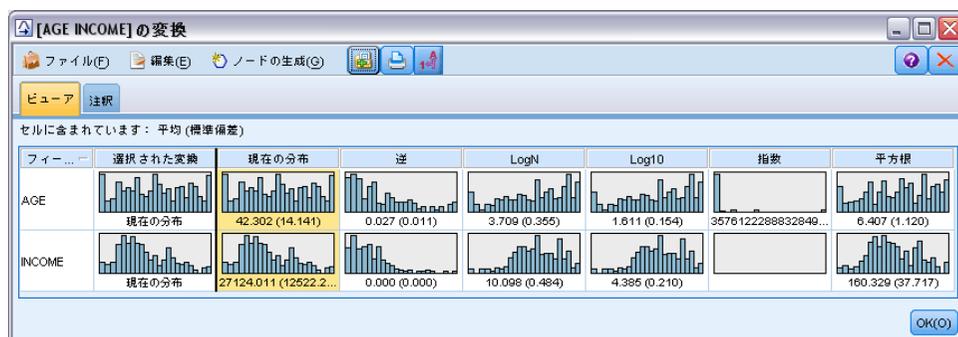
[出力] タブを使用すると、出力フォーマットや位置を指定することができます。結果をスクリーンに表示、または結果を標準ファイル形式の 1 つに送ることができます。詳細は、[p.444 出力ノードの \[出力\] タブ](#) を参照してください。

変換ノードの出力ビューア

出力ビューアを使用すると、変換ノード実行の結果を表示することができます。ビューアは、フィールドあたり複数の変換をサムネイルで表示し、フィールドをすばやく比較することができます。[ファイル] メニューのこ

のオプションを使用して、出力結果を保存、エクスポート、印刷することができます。詳細は、[p. 436 出力を表示](#) を参照してください。

図 6-32
フィールドごとに使用可能な変換の表示



(選択した変換以外の) 各変換に対し、判例が形式の下に表示されます。

Mean (Standard deviation)

変換ノードの作成

出力ビューアから、データの準備を開始することができます。たとえば、[年齢]のフィールドを正規化して、正常な分布を仮定するロジスティック回帰または判別分析などのスコアリング技術を使用したい場合があります。初期のグラフおよび要約統計に基づき、特定の分布に従って (.log など) [年齢] フィールドを変換することを決定します。優先的な分布を選択した後、スコアリングに使用する標準化された変換によって、フィールド作成ノードを生成します。

出力ビューアから、次のフィールド操作ノードを生成することができます。

- フィールド作成
- 置換

置換ノードは既存のフィールドを変換しますが、フィールド作成ノードを使用すると、希望の変換によって新しいフィールドを作成することができます。ノードはスーパーノードの形式で領域に設置されます。

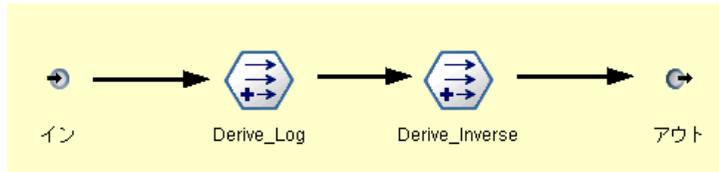
異なるフィールドで同じ変換を選択した場合、フィールド作成ノードまたは置換ノードには、変換が適用されるすべてのフィールドの変換タイプに関する式が含まれます。たとえば、次のようなフィールドと変換を選択して、フィールド作成ノードを生成したとします。

| フィールド | 変換 |
|-------|-------|
| AGE | 現在の分布 |

| | |
|----------|-------|
| INCOME | ログ |
| OPEN_BAL | 逆数(N) |
| バランス | 逆数(N) |

次のノードは、スーパーノードに含まれています。

図 6-33
領域内のスーパーノード



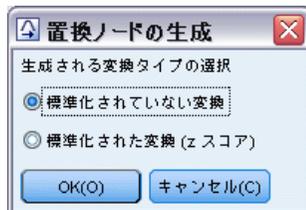
この例では、Derive_Log ノードには [INCOME] フィールドの対数式、Derive_Inverse ノードには OPEN_BAL および [BALANCE] フィールドに逆数式が含まれています。

ノードを生成するには

- ▶ 出力ビューアの各フィールドに対し、必要な変換を選択します。
- ▶ [ノード生成] メニューから、必要に応じてフィールド作成ノードまたは置換ノードを選択します。

選択すると、[フィールド作成ノードの生成] または [置換ノードの生成] ダイアログ ボックスが必要に応じて表示されます。

図 6-34
標準化された変換または標準化されていない変換の選択



[標準化されていない変換] または [標準化された変換 (z スコア)] を必要に応じて選択します。[標準化された変換] オプションは、変換に z スコアを適用します。z スコアは、標準偏差の変数の平均からの距離を現る機能として、値を表示します。たとえば、対数変換を [年齢] フィールドに適用し、標準化された変換を選択した場合、生成されたノードの最終式は次のようになります。

$(\log(\text{AGE}) - \text{Mean}) / \text{SD}$

ノードが生成され、ストリーム領域に表示されると、

- ▶ ストリームに接続されます。
- ▶ スーパーノードの場合、必要に応じてノードをダブル クリックし、コンテンツを表示させます。
- ▶ 必要に応じてフィールド作成ノードまたは置換ノードをダブル クリックし、選択したフィールドのオプションを変更します。

グラフの生成

出力ビューアで、サムネイル ヒストグラムから、フルサイズのヒストグラム出力を生成することができます。

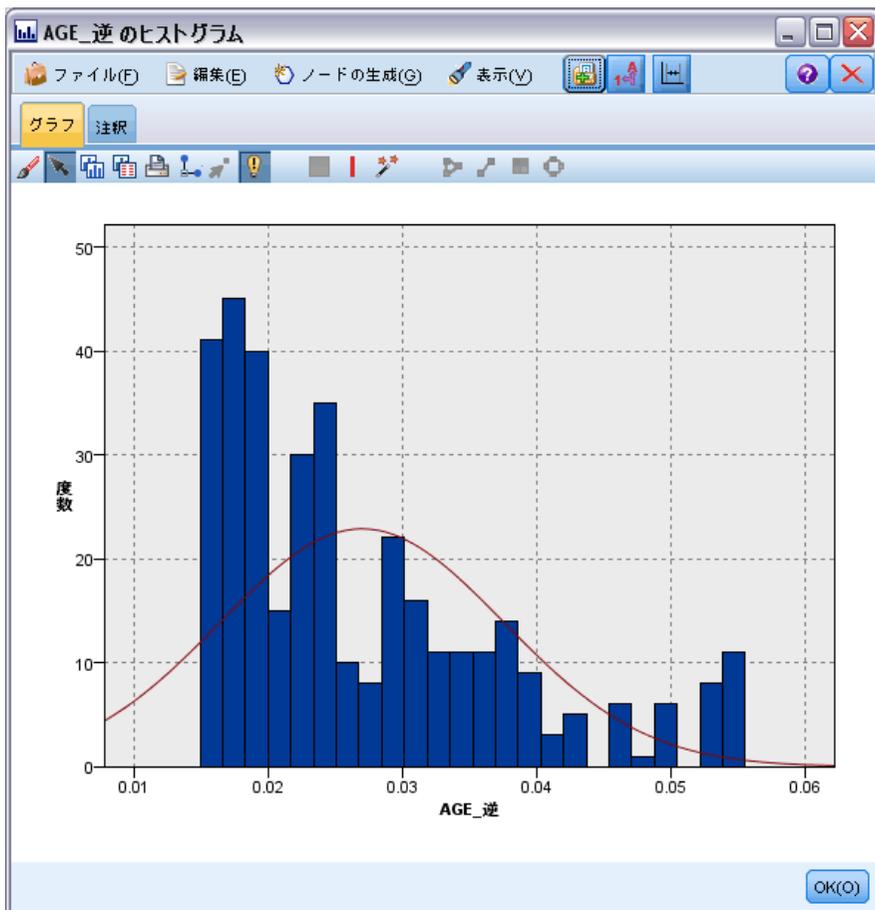
グラフを生成するには

- ▶ 出力ビューア中のサムネイル グラフをダブル クリックします。
- or
- ▶ 出力ビューア中のサムネイル グラフを選択します。
 - ▶ [生成] メニューの [グラフ出力] を選択します。

正規分布曲線を含むヒストグラムが表示されます。ヒストグラムの表示により、利用可能な各変換がどれほど近く、正規分布に一致するか比較することができます。

注： 出力を作成した変換ノードはストリームに接続している場合にのみグラフを生成できます。

図 6-35
正規分布曲線を含む変換ヒストグラム



その他の操作

出力ビューアから、次のことができます。

- フィールド列で出力グリッドのソート
- 出力結果を HTML ファイルにエクスポート [詳細は、 p. 441 出力のエクスポート を参照してください。](#)

記述統計ノード

記述統計ノードでは、数値型フィールドに関する基本的な集計情報を得ることができます。このノードから、個々のフィールドの要約統計量とフィールド間の相関についての情報を取得できます。

記述統計ノードの [設定] タブ

図 6-36
記述統計ノード:[設定] タブ



検証： 要約統計量を個別に算出するフィールドを選択します。複数のフィールドを選択できます。

統計量： 報告する統計量を選択します。利用できるオプションには、[カウント]、[平均]、[合計]、[最小]、[最大]、[集計範囲]、[分散]、[標準偏差]、[平均の標準誤差]、[中央値]、または[最頻値]があります。

相関関係： 相関させるフィールドを選択します。複数のフィールドを選択できます。相関フィールドを選択すると、出力に各 [検証] フィールドと相関フィールド間の相関が表示されます。

相関の設定： 出力における相関強度を表示するオプションを指定できます。

相関の設定

IBM® SPSS® Modeler では、重要な関係を強調するために、相関を詳細レベルで特徴づけることができます。**相関関係**は、2 つの連続型（数値範囲）フィールド間の相関の強さを測定します。これは、-1.0~1.0 の範囲の値を取ります。+1.0 に近い値は強い正の相関を表し、あるフィールドの大きい値が別のフィールドの大きい値と関連付けられ、小さい値が別の小さい値と関連付けられます。-1.0 に近い値は強い負の相関を示し、あるフィールドの大きい値が別のフィールドの小さい値と、そして小さい値が別の

大きい値と関連付けられます。0.0 に近い値は弱い相関を示し、2 つのフィールドの値は、独立しているといえます。

相関ラベルの表示、カテゴリを規定する閾値の変更、各範囲に使われるラベルの変更をコントロールできます。相関値を特徴付ける方法は、問題のドメインに大きく依存しているため、状況に応じて範囲とラベルをカスタマイズすることができます。

図 6-37
[相関設定] ダイアログ ボックス



出力に相関の強さラベルを表示： このオプションは、デフォルトで選択されています。出力に詳細ラベルを表示しない場合は、このオプションの選択を解除してください。

相関強度： 相関強度の定義とラベル付けには次の 2 つのオプションがあります。

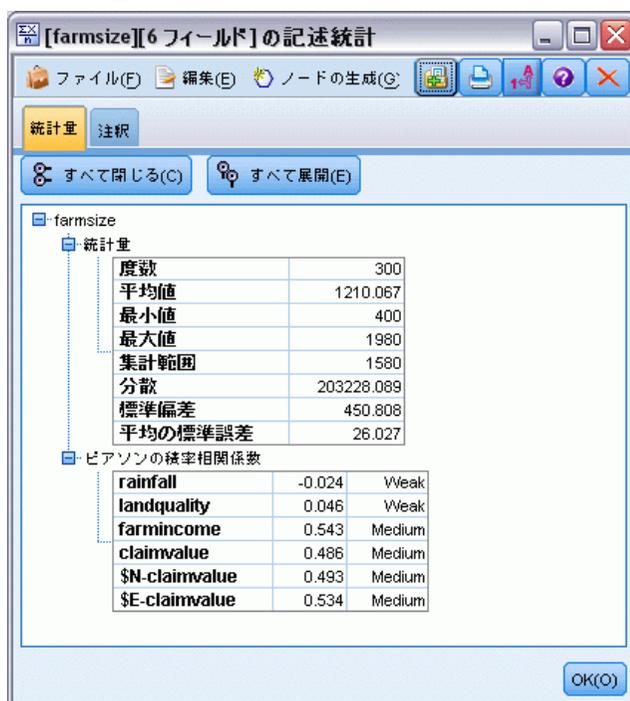
- **重要度 (1-p) による相関強度の定義：** 重要度を基に相関にラベル付けをします。1 マイナス有意、または平均値の差が確率のみで説明できる 1 マイナス確率として定義されます。この値が 1 に近いほど、2 つのフィールドが「独立していない」、つまり、これらの中に何らかの関係が存在することを示します。重要度を基に相関にラベル付けは、データにおけるバラつきを考慮に入れて通常絶対値をお勧めしています。たとえば、定数 0.6 は、1 つのデータ セットでは高い有意性を示しますが、その他のすべてにおいては有意性はありません。デフォルトでは、0.0~0.9 (絶対値) の重要度の値は [低い]、0.9~0.95 は [中間]、0.95~1.0 は [高い] のラベルが付けられます。
- **絶対値による相関強度の定義：** Person の相関係数の絶対値を基に相関にラベルを付けます。前述したように、この範囲は-1 から 1 までです。この測定値の絶対値が 1 に近ければ近いほど相関が強くなります。デフォルトでは、0.0~0.3333 (絶対値) の相関は [弱い]、0.3333~0.6666 の相関は [中間]、0.6666~1.0 の相関は [強い] のラベルが付けられます。ただし、指定された値の有意性をデータセットからその他に一般化するのは困難です。そのためほとんどのケースにおいて、絶対値ではなく確率に基づいて相関を定義することをお勧めしています。

記述統計量出力ブラウザ

記述統計ノード出力ブラウザには、統計分析の結果が表示されます。ここで、フィールドの選択、選択に基づく新しいノードの生成、結果の保存と印刷などの操作を実行できます。通常の保存、エクスポート、および印刷関連オプションは [ファイル] メニューから、通常の編集オプションは [編集] メニューから利用できます。詳細は、[p. 436 出力を表示](#) を参照してください。

記述統計量の出力を初めて参照する場合、結果は展開されて表示されています。参照し終わった後に結果を隠すには、項目の左側にある拡張をコントロールを使って目的の結果を閉じるか、または [すべて閉じる] ボタンをクリックしてすべての結果を非表示にします。閉じた結果をもう一度表示するには、項目の左側にある拡張コントロールを使用するか、または [すべて展開] ボタンをクリックしてすべての結果を表示してください。

図 6-38
記述統計量出力ブラウザ



出力には、各 [検証] フィールド用のセクションが含まれています。このセクションには、要求された統計量がテーブルで表示されています。

- **度数**：フィールドの有効な値を持つレコード数。
- **平均値**：すべてのレコードに渡るフィールドの平均値。
- **合計**：すべてのレコードに渡るフィールドの合計値。

- **最小**：フィールドの最小値。
- **最大**：フィールドの最大値。
- **範囲**：最小値と最大値の差異。
- **分散**：フィールドの値における変動の測定値。この値は、各値と総合平均間の差異を二乗して、すべての値を合計したら、それをレコード数で除算して算出します。
- **標準偏差**：フィールドの値中の変動の測定値で、分散の平方根として算出されます。
- **平均の標準誤差**：平均が新しいデータに適用されることを前提にしている場合の、フィールドの平均の推定中の不確実性の測定値です。
- **中央値**：フィールドの中央値、つまりデータの上半分と下半分を分割する場所にある値です（フィールドの値を基準にして）。
- **モード**：データ中に最も多く出現する単一値です。

相関係数：相関関係フィールドを指定した場合、出力にも [検証] フィールドと各相関関係フィールド間の Pearson の積率相関係数、および相関関係値の詳細ラベル（オプション）を記載するセクションが含まれます。詳細は、[p. 483 相関の設定](#) を参照してください。

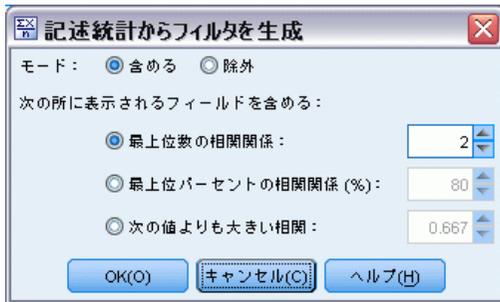
[ノードの生成] メニュー：[ノード生成] メニューには、ノード生成操作に関するメニュー項目があります。

- **フィルタ**：他のフィールドと相関していないフィールドや相関の弱いフィールドを取り除くフィルタ ノードを生成します。

統計量からのフィルタ ノードの生成

図 6-39

[統計量からフィルタを生成] ダイアログ ボックス



記述統計量出力ブラウザから生成されたフィルタ ノードでは、他のフィールドとの相関関係に基づいてフィールドがフィルタリングされます。絶対値の順序で相関関係をソートして、最大の相関関係（ダイアログ ボックスで設定する基準に基づく）を取り出し、これらの最大相関関係に現れるすべてのフィールドを通過させるフィルタを作成します。

モード: 相関関係の選択方法を決定します。[含める]を選択すると、指定した相関関係に現れるフィールドが保持されます。[除外]を選択すると、フィールドがフィルタリングされます。

次の所に表示されるフィールドを含める/除外する: 相関関係を選択する基準を定義します。

- **最上位数の相関関係:** 指定数の相関関係を選択し、それらの相関関係のいずれかに現れるフィールドを含めるか、または除外します。
- **最上位パーセントの相関関係 (%):** 指定したパーセンテージ (n%) の相関関係を選択し、それらの相関関係のいずれかに現れるフィールドを含めるか、または除外します。
- **次の値よりも大きい相関:** 指定された閾値よりも絶対値が大きい相関関係を選択します。

平均比較ノード

平均比較ノードでは、独立したグループ間で、または関連するフィールドのペア間で著しい違いがあるかどうかを調べるために、平均を比較します。たとえば、販売促進活動の実施前後に平均収入を比較、または販売促進活動を受け容れない顧客からの売上げの比較ができます。

2 つの異なる方法で、以下のデータを基に平均値を比較できます。

- **フィールド内のグループ間で比較:** 独立グループを比較するには、テストフィールドとグループ化フィールドを選択します。たとえば、販売促進活動の対象から「ホールドアウト」顧客のサンプルを除外でき、ホールドアウトグループとその他のすべてと平均収入を比較できます。この場合、各顧客の収入を示す単一のテストフィールドを、フラグまたはかれらがオファーを受けるかどうかを示す名義型フィールドで指定します。サンプルは各レコードが 1 つのグループ他に割り当てられること、また 1 つのグループの特定のメンバーをその他のグループの特定のメンバーに結びつける方法がない、という意味で独立しています。また、2 つ以上の値で名義型フィールドも指定し、複数のグループの平均値を比較できます。実行時には、ノードは選択したフィールドで一元 ANOVA テストを計算します。2 つのフィールドグループしかない場合、一元 ANOVA の結果は本質的に独立したサンプル t テストと同じです。詳細は、[p. 488 独立したグループの平均値を比較](#) を参照してください。
- **フィールドのペア間で比較:** 2 つの関連フィールドの平均値を比較する場合、有意の結果を得るために、そのグループ群は何らかの方法でペアを組みます。たとえば、販売促進活動の実施前後の同じグループの顧客からの平均収入を比較するか、または夫婦ペア間サービスの利用率を比較して差が見られるかどうか確認します。各レコードには、独立しているが関連する有意性を比較できる 2 つの方法を含みます。実行時に、ノードは一組のサンプル t テストを指定された各フィー

ルドペアで計算します。詳細は、p. 489 一对のフィールド間の平均値の比較を参照してください。

独立したグループの平均値を比較

平均比較ノードで [フィールド内のグループ間] を選択し、2 つ以上の独立グループを比較します。

図 6-40
1 フィールド内のグループ間の平均を比較



グループ化フィールド: 比較したいグループにレコードを区切った、たとえばオファーを受け取る/受け取らない人などのように区切った、2 つ以上の重複レコード値を持つ数値フラグまたは名義型フィールドを選択します。試験フィールド数に無関係に、ただひとつグループ化するフィールドを選択できます。

テストフィールド: 試験の対象となる測定値を含む 1 以上の数値型フィールドを選択します。独立した試験を選択した各フィールドに実施します。たとえば、やり方、資金、運動の盛り上げに関する特定の販売促進活動の影響をテストすることができます。

一対のフィールド間の平均値の比較

平均値ノードで [一対のフィールド間] を選択し、独立したフィールド間の平均値を比較します。何らかの方法で有意となるようにフィールド同士を関連付ける必要があります（たとえば、販売促進活動の前後の収入など）。複数のフィールド ペアを選択することもできます。

図 6-41
一対のフィールド間の平均値の比較



フィールド 1: 比較したい最初の測定値を含む数値フィールドを選択します。前後の検討において、これは「前」のフィールドにあたります。

フィールド 2: 比較したい 2 番目のフィールドを選択します。

追加: 選択したペアをテスト型フィールド ペア リストに追加します。

フィールドの選択を繰り返し、必要な複数のペアをリストに追加します。

関連の設定: ラベル付けのオプションを設定して関連強度を指定することができます。詳細は、[p.483 関連の設定](#) を参照してください。

平均比較ノードのオプション

[オプション] タブを選択すると結果に対して重要度 高、境界、または重要度 低とラベル付けするのに使用する閾値 P を設定できます。また、各ランク付けラベルの編集もできます。重要度 高は、パーセンテージで測定さ

れ、大雑把に 1 マイナス結果を取得する確率（2 つのフィールドの平均値の違いなど）と定義され、観察された確率のみによる結果とほぼ同程度、またはもっと極端と定義します。たとえば、0.95 を超える p 値は、確率のみで結果を説明できるとすると、5 % 未満の確率であることを示します。

図 6-42
重要度設定



重要度のラベル：出力時に各フィールドペアまたはグループにラベル付けするラベルの編集ができます。デフォルトのラベルは、[重要度 高]、[境界]、[重要度 低] です。

分割値：各ランクの閾値を指定します。一般的に、0.95 を超える p 値は、重要度 高とランク付けられ、0.9 未満は、重要度 低とランク付けされます。しかしこれらの閾値を必要に応じて調整することができます。

注：重要度の測定はさまざまなノードで利用できます。具体的な演算、使用する対象フィールドおよび入力フィールドのデータ型はノードによって異なりますが、すべてがパーセンテージによる測定であるため、値は比較することができます。

平均値ノード出力ブラウザ

平均値出力ブラウザは、クロス集計データを表示し、同時にテーブルの1行の選択やコピー、任意の列によるソート、テーブルの保存タ印刷を含む標準の演算を実行できます。 [詳細は、 p. 436 出力を表示 を参照してください。](#)

表の具体的な情報は、比較のタイプ（フィールド内のグループまたは独立したフィールド）によって異なります。

ソート項目： 出力を特定の列でソートできるようになります。上下の矢印をクリックしてソートの方向を変更します。または、任意の列見出しをクリックして、列ごとのソートが可能です。（列内のソートの方向を変更するには、もう一度クリックします。）

表示： [シンプル] または [詳細] を選択して表示の詳細度をコントロールできます。詳細表示には、シンプル表示からのすべての情報が含まれますが、その他に詳細な内容が提供されます。

フィールド内でグループを比較する平均値出力

フィールド内のグループを比較する場合、グループ化するフィールド名が出力テーブルの上の方に表示され、平均値や関連する統計値が各グループ別に別々に表示されます。テーブルには、各テストフィールド用に独立した行が含まれています。

図 6-43
フィールド内のグループを比較

| フィールド | Standard* | New Promoti... | 重要度 |
|------------------------------------|-----------|----------------|-------|
| \$ spent during promotional period | 1566.389 | 1637.500 | 0.976 |

また、次の列も表示されます。

- **フィールド：** 選択したテスト フィールドの名前が表示されます。

- **グループ別の平均値**：グループ化するフィールドの各カテゴリの平均値が表示されます。たとえば、特別なオファーを受けた人（新規販売促進活動対象）と受けなかった人（標準）を比較できます。詳細表示で、標準偏差、標準誤差、度数も表示できます。
- **重要度**：重要度の値とラベルを表示します。詳細は、[p. 489 平均比較ノードのオプション](#) を参照してください。

詳細出力

詳細表示では、次の列が追加表示されます。

- **F-検定**：このテストは、グループ間の分散と各グループ内の分散の比率を基にします。平均値がすべてのグループで同じとすると、両方とも同じ母集団分散の予測であるため、F 率がほぼ 1 になると期待されます。この比率が大きくなると、グループ間の分散が大きくなり、有意差が存在する可能性がより大きくなります。
- **自由度**：自由度を表示します。

フィールドのペアを比較する平均値出力

独立したフィールドを比較する場合、出力テーブルには選択したフィールドペアの行が含まれます。

図 6-44
フィールド ペアの比較

| フィールド 1 | フィールド... | 平均 1* | 平均 2* | 相関 | 平均差* | 重要度 |
|--------------|-------------------|---------|---------|-----------------|--------|----------------------|
| Triglyceride | Final triglyce... | 138.438 | 124.375 | -0.286 Weak | 14.062 | 0.751 Unimportant |
| Weight | Final weight | 198.375 | 190.312 | 0.996 Strong | 8.062 | 1.000 Important |

- **フィールド 1/2**：各ペアの 1 番目と 2 番目のフィールド名を表示します。詳細表示で、標準偏差、標準誤差、度数も表示できます。
- **平均値 1/2**：各フィールドの平均値を個別に表示します。

- **相関**：相関強度は、2 つの連続型（数値範囲）フィールド間の関係の強さを測定します。+1.0に値が近いほど正の関連が強いことを示し、-1.0に近いほど負の関連が強いことを示します。 [詳細は、 p. 483 相関の設定 を参照してください。](#)
- **平均差**：2 つのフィールドの平均値の差を表示します。
- **重要度**：重要度の値とラベルを表示します。 [詳細は、 p. 489 平均比較ノードのオプション を参照してください。](#)

詳細出力

詳細出力には次の列が追加されます。

95 % の信頼区間：真（true）の平均値が、この母集団のこのサイズでのすべての可能なサンプルの 95 % が含まれる範囲の上下の境界。

T-検定 :t 統計値は、平均値の差をその標準誤差で除算すると得られます。この統計値の絶対値が大きくなると、平均値が異なる可能性が大きくなります。

自由度：統計値の自由度を表示します。

レポート ノード

レポート ノードを使って、固定テキスト、およびデータやデータから導かれた他の式を含む、フォーマット済みレポートを作成できます。レポートの書式は、固定テキストとデータの出力構成を定義するテキスト テンプレートを使用して指定します。テンプレート中の HTML タグを使用し、[出力] タブで設定を指定することにより、カスタムの書式設定を利用することができます。テンプレート中の CLEM 式を使用して、データ値および他の条件出力がレポートに入れられます。

レポート ノードの代替

レポート ノードは通常、一定の条件を満たすすべてのレコードなど、ストリームからのレコードやケース出力を表示するために使用されます。この点で、レポート ノードは構造化の度合いが少ない、テーブル ノードの代替物だと考えられます。

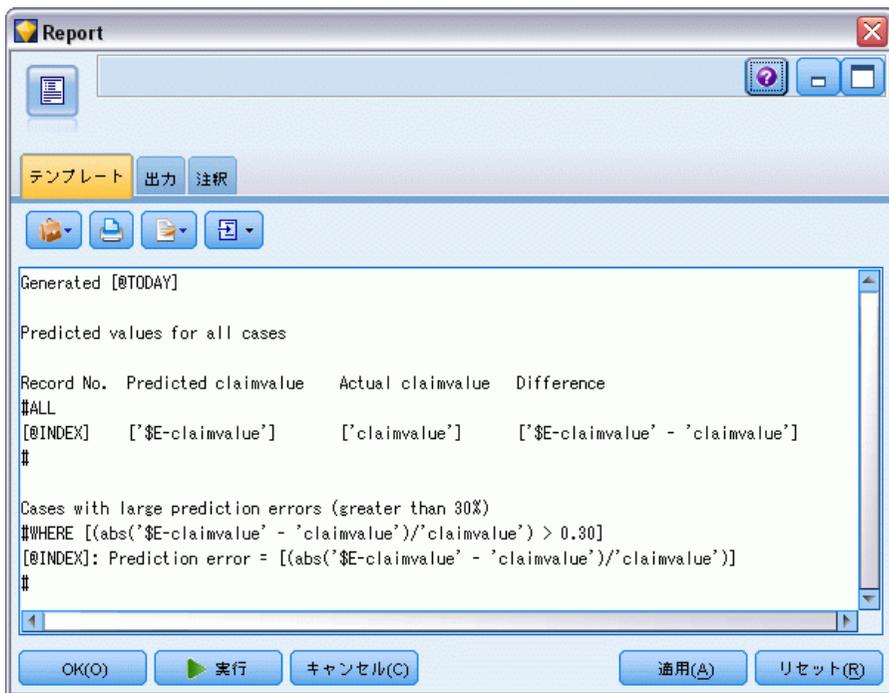
- データ自体でなくストリーム内で定義されたフィールド情報やその他の情報（データ型ノード内で指定されたフィールド定義など）をレポートに表示しようとする場合は、代わりにスクリプトを使用できます。 [詳細は、 6 章 データ型ノード レポート in IBM SPSS Modeler 15 スクリプトとオートメーション ガイド を参照してください。](#)
- ストリームに生成されたモデル、テーブル、およびグラフの集合として複数の出力オブジェクトを含むレポートを生成するには、そしてそのレポートがテキスト、HTML、および Microsoft Word/Office を含む複数

の形式で出力できる場合は、IBM® SPSS® Modeler プロジェクトが使用できます。詳細は、11 章 プロジェクトの概要 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。

- スクリプトを使用しないでフィールド名のリストを作成するには、すべてのレコードを廃棄するサンプル ノードに先導された、テーブル ノードが使用できます。これにより、行のないテーブルが作成され、エクスポート時に行列が入れ替えられて、1 列のフィールド名のリストが作成されます。(このためには、テーブル ノードの [出力] タブで、[データの入れ替え] を選択します。)

レポート ノードの [テンプレート] タブ

図 6-45
レポート ノード :[テンプレート] タブ



テンプレートの作成： レポートの内容を定義するには、レポート ノードの [テンプレート] タブでテンプレートを作成します。テンプレートはテキスト行から構成されており、各行にレポートの内容を指定します。内容行の範囲を示す特殊なタグも含まれています。各内容行では、その行をレポートに送る前に、大カッコ ([]) で囲まれている CLEM 式が評価されます。テンプレートでは、1 行に付き次の 3 種類の範囲を指定できます。

固定： マークが付けられていない行は固定と見なされます。固定行は、行に含まれる他の式を評価した後に、レポートに 1 行だけ書き込まれます。たとえば、次のような行があるとします。

```
This is my report, printed on [@TODAY]
```

この場合、上記の文字列と現在の日付を記載した 1 行がレポートに書き込まれます。

グローバル (ALL を反復)： 入力データの各レコードに 1 回ずつ、特殊なタグ #ALL と # で囲まれている行がレポートに書き込まれます。CLEM 式（かっこで囲まれている）が、各出力行の現在のレコードに基づいて評価されます。たとえば、次のような行があるとします。

```
#ALL
For record [@INDEX], the value of AGE is [AGE]
#
```

この場合、各レコードに対して、レコード番号と年齢を示す 1 行が記載されます。

すべてのレコードのリストを生成するには

```
#ALL
[Age] [Sex] [Cholesterol] [BP]
#
```

条件式 (WHERE を反復)： 指定した条件が真 (true) であるレコードごとに 1 回ずつ、特殊なタグ #WHERE <状況> と # で囲まれている行がレポートに書き込まれます。条件は CLEM 式で指定します。(WHERE 条件内では、大かっこは省略できます。) たとえば、次のような行があるとします。

```
#WHERE [SEX = 'M']
Male at record no. [@INDEX] has age [AGE].
#
```

この場合、性別が M の値を持つ各レコードに対して 1 行がファイルに書き込まれます。完全なレポートには、入力データにテンプレートを適用して定義された固定行、グローバル行、および条件行が含まれます。

[出力] タブから、結果の表示や保存に関する設定を指定することができます。この設定は、さまざまな種類の出力ノードに共通しています。 [詳細は、p. 444 出力ノードの \[出力\] タブ を参照してください。](#)

HTML または XML 形式データを出力する

両方の形式でレポートを書くために、HTML または XML タグを直接テンプレートに含めることができます。たとえば、次のテンプレートは HTML タブを生成します。

This report is written in HTML.
Only records where Age is above 60 are included.

```
<HTML>
<TABLE border="2">
  <TR>
    <TD>Age</TD>
    <TD>BP</TD>
    <TD>Cholesterol</TD>
    <TD>Drug</TD>
  </TR>

  #WHERE Age > 60
  <TR>
    <TD>[Age]</TD>
    <TD>[BP]</TD>
    <TD>[Cholesterol]</TD>
    <TD>[Drug]</TD>
  </TR>

#
</TABLE>
</HTML>
```

レポート ノード出力ブラウザ

レポート ブラウザには、生成されたレポートの内容が表示されます。通常の保存、エクスポート、および印刷関連オプションは [ファイル] メニューから、通常の編集オプションは [編集] メニューから利用できます。詳細は、[p. 436 出力を表示](#) を参照してください。

図 6-46
レポート ブラウザ

Report のレポート

ファイル(F) 編集(E)

レポート 注釈

Prediction Report

Generated 2009-11-11

Predicted values for all cases

| Record No. | Predicted claimvalue | Actual claimvalue | Difference |
|------------|----------------------|-------------------|-------------|
| 1 | 100031.774 | 74703.100 | 25328.674 |
| 2 | 219399.708 | 245354.000 | -25954.292 |
| 3 | 74674.031 | 84213.000 | -9538.969 |
| 4 | 192439.185 | 281082.000 | -88642.815 |
| 5 | 186574.464 | 122006.000 | 64568.464 |
| 6 | 134760.974 | 122135.000 | 12625.974 |
| 7 | 64797.597 | 68969.200 | -4171.603 |
| 8 | 352105.619 | 485011.000 | -132905.381 |
| 9 | 231023.034 | 160904.000 | 70119.034 |
| 10 | 86104.837 | 90350.600 | -4245.763 |
| 11 | 69092.825 | 63494.000 | 5598.825 |
| 12 | 317729.972 | 389421.000 | -71691.028 |
| 13 | 41828.464 | 30377.000 | 11451.464 |
| 14 | 272413.717 | 322365.000 | -49951.283 |

OK(O)

グローバル ノード

グローバル ノードで、データを走査し、CLEM 式で使用できる要約値を算出します。たとえば、グローバル ノードを使用して、[年齢] という名前のフィールドの統計量を算出し、次に CLEM 式に `@GLOBAL_MEAN(age)` 関数を挿入して年齢の全体的な平均を算出することができます。詳細は、[8 章 CLEM リファレンス概要 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

グローバルノードの [設定] タブ

図 6-47
グローバルノード : [設定] タブ



作成するグローバル値： グローバル値を使用可能にするフィールドを選択します。複数のフィールドを選択できます。各フィールドに対して、目的の統計量がフィールド名の隣の列で選択されていることを確認し、計算する統計量を指定します。

- **平均：** すべてのレコードに渡るフィールドの平均値。
- **合計：** すべてのレコードに渡るフィールドの合計値。
- **最小：** フィールドの最小値。
- **最大：** フィールドの最大値。
- **標準偏差：** 標準偏差フィールドの値中の変動の測定値で、分散の平方根として算出されます。

デフォルトの処理： ここで選択したオプションが、上のグローバル値リストに新規フィールドを追加する際に使用されます。デフォルトの統計量セットを変更するには、必要に応じて統計量を選択または解除してください。[適用] ボタンを使用して、リスト中のすべてのフィールドにデフォルトの処理を適用することもできます。

実行前にすべてのグローバル値を消去： 新しい値を計算する前にすべてのグローバル値を取り除く場合に、このオプションを選択します。このオプションを選択しない場合、古い値が新しく計算された値に置き換えられますが、再計算されなかったグローバル値はそのまま使用できます。

実行後に作成されたグローバル値のプレビューを表示：このオプションを選択した場合、実行後に [ストリームのプロパティ] ダイアログ ボックスの [グローバル値] タブが現れ、計算されたグローバル値が表示されます。詳細は、[5 章 ストリームのグローバル値の表示 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

IBM SPSS Statistics ヘルパー アプリケーション

IBM® SPSS® Statistics の互換性バージョンがコンピュータにインストールされ、ライセンス供与されると、IBM® SPSS® Modeler を設定し、Statistics 変換ノード、Statistics モデル ノード Statistics 出力ノード、または Statistics エクスポート ノードを使用して、SPSS Statistics の機能によりデータを処理することができます。

- ▶ SPSS Modeler で SPSS Statistics や他のアプリケーションを利用できるようにするには、次のメニューを選択します。

ツール > オプション > ヘルパー アプリケーション

IBM SPSS Statistics Interactive: Statistics エクスポート ノードで作成されたデータ ファイルで直接 SPSS Statistics を起動する場合に使用するコマンドのフルパスと名前を入力します (例: C:\Program Files\IBM\SPSS\Statistics\stats.exe)。詳細は、[8 章 p.561 Statistics エクスポート ノード](#) を参照してください。

接続 :SPSS Statistics サーバーが IBM® SPSS® Modeler Server と同じサーバーに存在している場合、2 つのアプリケーション間の接続を有効にすることができます。有効にすると、分析中にサーバー上にデータを配置することにより、効率を向上することができます。[サーバー] を選択すると、下の [ポート] オプションが有効になります。デフォルトの設定は [ローカル] です。

ポート :SPSS Statistics サーバーのポートを指定します。

IBM SPSS Statistics ライセンス ロケーション ユーティリティ:SPSS Modeler が Statistics 変換ノード、Statistics モデル ノードおよび Statistics 出力ノードを使用できるようにするには、ストリームが実行されているコンピュータに SPSS Statistics をインストールしライセンスを与えられている必要があります。さらに、リモート SPSS Modeler Server に対して分散型モードで実行している場合、SPSS Statistics クライアントのコピーを SPSS Modeler クライアント コンピュータにインストールしライセンス認証する必要があります。

- ローカル (スタンドアロン) モードで SPSS Modeler を実行中の場合、SPSS Statistics のライセンスが付与されたドライバをローカル コンピュータに搭載されている必要があります。このボタンをクリック

して、ライセンスに使用するローカル SPSS Statistics インストールの場所を指定します。

- また、リモート SPSS Modeler Server に対して分散モードで実行する場合、サーバー コンピュータに SPSS Statistics のライセンス バージョンを搭載し、サーバーにライセンス設定を行う必要があります。ライセンス設定を行うには、Windows の場合、SPSS Modeler Serverbin ディレクトリに移動して、コマンド プロンプトで次の文を実行します。

```
statisticsutility -location=<path to IBM SPSS Statistics Server license file>/bin
```

また、UNIX の場合は、次を実行します。

```
./statisticsutility -location=<path to IBM SPSS Statistics Server license file>/bin
```

この場合、<path to SPSS Statistics Server license file> は、ライセンスが与えられた SPSS Statistics サーバーのインストール ディレクトリを示します。

ローカル コンピュータに SPSS Statistics のライセンス認証されたコピーがない場合でも、ライセンス認証された SPSS Statistics サーバーに対して Statistics ファイル ノードを実行できますが、他の SPSS Statistics ノードを実行するとエラー メッセージが表示されます。

コメント

SPSS Statistics プロシージャ ノードの実行に何か問題がある場合は、次の事柄を確認してください。

- SPSS Modeler で使用されているフィールド名が 8 文字 (SPSS Statistics 12.0 より前のバージョン) より長い場合、64 文字 (SPSS Statistics 12.0 以降のバージョン) より長い場合、または不正な文字が含まれている場合は、SPSS Statistics に読み込む前にフィールド名を変更するか、または名前を切り詰める必要があります。詳細は、[8 章 p. 563 IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング](#) を参照してください。
- SPSS Modeler の後に SPSS Statistics をインストールした場合、上記で説明したとおり、SPSS Statistics のライセンスの場所を指定する必要があります。

エクスポート ノード

エクスポート ノードの概要

エクスポート ノードは、他のソフトウェア ツールのインターフェイスに対応した、さまざまなフォーマットでデータをエクスポートできるメカニズムも備えています。

次のエクスポート ノードを利用できます。



データベース エクスポート ノードで、データを ODBC 対応のリレーショナル データ ソースに書き込みます。ODBC データ ソースに書き込むには、データ ソースが存在し、そのデータ ソースに対する書き込み権限を取得している必要があります。詳細は、[p. 502 データベース エクスポート ノード](#) を参照してください。



ファイル ノードでは、データが区切り文字で区切られたテキスト ファイルへ出力されます。このことは、他の分析ソフトウェアや表計算ソフトウェアに読み込めるフォーマットでデータをエクスポートする場合に、役立ちます。詳細は、[p. 526 ファイル エクスポート ノード](#) を参照してください。



Statistics エクスポート ノードでは、IBM® SPSS® Statistics .sav フォーマットでデータを出力します。.sav ファイルは、SPSS Statistics Base およびその他の製品で読み込むことができます。このフォーマットは、IBM® SPSS® Modeler のキャッシュ ファイルでも使用されます。詳細は、[8 章 p. 561 Statistics エクスポート ノード](#) を参照してください。



IBM® SPSS® Data Collection エクスポート ノードは、Data Collection の市場調査ソフトウェアで使用する形式でデータを出力します。このノードを使用するには、Data Collection Data Library がインストールされている必要があります。詳細は、[p. 528 IBM SPSS Data Collection エクスポート ノード](#) を参照してください。



SAS エクスポート ノードで、SAS または SAS 互換ソフトウェア パッケージで読み込むデータを、SAS フォーマットで出力できます。3 つの SAS ファイル形式が利用可能です。SAS for Windows/OS2、SAS for UNIX、または SAS バージョン 7/8 詳細は、[p. 534 SAS エクスポート ノード](#) を参照してください。



Excel エクスポート ノードは、Microsoft Excel 形式 (.xls) でデータを出力します。オプションで、ノードが実行されるときに自動的に Excel が起動し、エクスポートするファイルを開けるように選択できます。詳細は、[p. 536 Excel エクスポート ノード](#) を参照してください。



XML エクスポート ノードでは、XML 形式のファイルにデータを出力します。オプションで、エクスポートしたデータをストリームに読み込む XML 入力ノードを作成できます。詳細は、[p. 537 XML エクスポート ノード](#) を参照してください。

データベース エクスポート ノード

データベース ノードを使用して、データを ODBC 準拠の関連データ ソースに書き込むことができます。詳細は、データベース入力ノードに記載されています。詳細は、[2 章 p. 16 データベース入力ノード](#) を参照してください。

データベースにデータを書き込むには、次の作業を行います。

- ▶ 使用するデータベースに ODBC ドライバをインストールして、データソースを構成します。
- ▶ データベース ノードの [エクスポート] タブで、書き込み先のデータソースとテーブルを指定します。新規テーブルを生成するか、既存のテーブルにデータを挿入します。
- ▶ 必要に応じてオプションを追加します。

これらの作業の詳細は、続く各トピックで説明します。

データベース ノードの [エクスポート] タブ

図 7-1
データベース エクスポート ノードの [エクスポート] タブ



データソース： 選択したデータ ソースが表示されます。名前を入力するか、ドロップダウン リストから選択します。リスト中に目的のデータベースが見つからない場合は、[\[新規データベース接続の追加\]](#) を選択して、[\[データベース接続\]](#) ダイアログ ボックスから目的のデータベースを検索してください。[詳細は、2 章 p. 20 データベース接続の追加 を参照してください。](#)

テーブル名： データの送信先のテーブル名を入力します。[\[テーブルへ挿入\]](#) オプションを選択した場合、[\[選択\]](#) ボタンをクリックして、データベース中の既存のテーブルを選択することができます。

テーブルの作成： 新規データベース テーブルを作成する場合、または既存のデータベース テーブルを上書きする場合に選択します。

テーブルへ挿入： 既存のデータベース テーブルにデータを新しい行として挿入する場合に選択します。

テーブルを結合： (可能な場合) 選択したデータベースの列を、該当する入力データフィールドの値で更新します。このオプションを選択すると [\[結合\]](#) ボタンが有効となり、入力データフィールドをデータベース列に関連付けできるダイアログが表示されます。

既存のテーブルを削除： 新しいテーブルの作成時に、同じ名前を持つ既存のテーブルを削除する場合に選択します。

既存の行を削除： テーブルへの挿入時に、エクスポート前に既存の行をテーブルから削除する場合に選択します。

注： 上記の 2 種類のオプションのどちらも選択されていない場合、ノードの実行時に「上書き警告」というメッセージが表示されます。この警告メッセージを表示しない場合は、[ユーザー オプション] ダイアログ ボックスの [通知] タブにある、[ノードがデータベース テーブルを上書きする時に警告] を選択します。詳細は、12 章 [通知オプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

デフォルトの文字列サイズ： 上流にあるデータ型ノードでデータ型不明とされたフィールドは、データベースに文字列フィールドとして書き込まれます。ここには、データ型不明フィールドに使う文字列のサイズを指定します。

[スキーマ] をクリックしてダイアログ ボックスを開きます。ここでさまざまなエクスポート オプション（この機能をサポートするデータベースを設定、SQL データ型をフィールドに設定し、データベース インデックス用の第 1 入力フィールド キーを指定します。詳細は、[p. 507 データベースエクスポートのスキーマ オプション](#) を参照してください。

[インデックス] をクリックして、エクスポートされたテーブルのインデックスを生成するオプションを指定してデータベースのパフォーマンスを高めます。詳細は、[p. 511 データベース エクスポートのインデックス オプション](#) を参照してください。

バルク ロードおよびデータベースのコミットに関するオプションを指定するには、[詳細] をクリックします。詳細は、[p. 515 データベース エクスポートの詳細オプション](#) を参照してください。

テーブルおよび列名を引用符で囲む： データベースに CREATE TABLE 文を送信する場合に使用するオプションを選択します。スペースまたは非標準文字を含むテーブルや列名は引用符で囲む必要があります。

- **必要に応じて：** 引用符が必要かどうかを IBM® SPSS® Modeler が個別に判断して、自動的に引用符で囲む場合に選択します。
- **常時：** テーブル名と列名を常に引用符で囲む場合に選択します。
- **しない：** 引用符を使用しない場合に選択します。

データのインポートノードを生成： 指定したデータ ソースとテーブルにエクスポートしたように、データベース入力ノードを生成する場合に選択します。実行時に、ストリーム領域にこのノードが追加されます。

データベース エクスポート結合オプション

入力データのフィールドを、対象データベース テーブルの列にマッピングできます。入力データ フィールドがデータベース列に関連付けされると、ストリーム実行時に列の値が入力データ値に置き換えられます。データベース内の関連付けられていない入力フィールドは、変わらないままです。

図 7-2
入力データ フィールドのデータベース列への関連付け



フィールドのマップ：入力データ フィールドとデータベース列とのマッピングを指定します。データベースの列と名前が同じ入力データ フィールドは自動的にマッピングされます。

- **マップ**： ボタンの左側にあるフィールド リストで選択した入力データ フィールドを右側のリストで選択したデータベース列にマッピングします。一度に複数のフィールドをマッピングできますが、両方のリストで選択するエントリ数は同じでなければなりません。
- **マップ解除**： 選択した 1 つまたは複数のデータベース列のマッピングを解除します。ダイアログ右側のテーブルでフィールド列またはデータベース列を選択すると、このボタンが有効になります。
- **追加**： ボタンの左側にあるフィールド リストで選択した 1 つまたは複数の入力データ フィールドを右側のリストに追加すると、マッピングできます。左側のリストでフィールドを選択し、右側のリストにその名前のフィールドがない場合、このボタンが有効になります。このボタンをクリックすると、選択したフィールドを、同じ名前の新しいデータベース列にマッピングします。単語 <NEW> がデータベース列の名前の後に表示され、新しいフィールドであることを示します。

行を結合：トランザクション ID などのキー フィールドに同じ値を持つレコードが結合されます。これは、データベースの「等結合」と同じ処理です。キー値はこれらのプライマリ キーでなければなりません。つまり一意でなければならず、Null 値を含むことはできません。

- **キーの候補** :すべての入力データ ソースから共通するすべてのフィールドが表示されます。このリストから 1 つまたは複数のフィールドを選択して矢印ボタンをクリックすると、レコードを結合するためのキー フィールドとしてそのフィールドが追加されます。対応するマッピングされたデータベース列を持つマップ フィールドは、キーとしてのごきますが、新しいデータベース列（名前の後に <NEW> と表示）は使用できません。
- **結合キー** : すべての入力データ ソースから、キー フィールドの値に基づいたレコードの結合に使用するすべてのフィールドが表示されます。リストからキーを削除するには、該当するキーを選択して矢印ボタンをクリックし、そのキーを [キーの候補] リストに戻します。複数のキー フィールドが選択されている場合は、次のオプションが有効になります。
- **データベース内にあるレコードのみを含める** : 部分結合を実行します。レコードがデータベースとストリーム内にある場合、マップされたフィールドが更新されます。
- **レコードをデータベースに追加** : 外部結合を実行します。データベース内に同じレコードがある場合は、ストリーム内のすべてのレコードが結合され、レコードがデータベース内にない場合は追加されます。

入力データ フィールドを新しいデータベース列にマッピングするには

- ▶ [マップフィールド] 下の、左側のリストの入力フィールド名をクリックします。
- ▶ [追加] ボタンをクリックして、マッピングを完了します。

入力データ フィールドを既存のデータベース列にマッピングするには

- ▶ [マップフィールド] 下の、左側のリストの入力フィールド名をクリックします。
- ▶ 右側の [データベース列] の列名をクリックします。
- ▶ [マップ] ボタンをクリックして、マッピングを完了します。

マッピングを解除するには

- ▶ [フィールド] 下の右側のリストで、マッピングを削除するフィールドの名前をクリックします。
- ▶ [マップ解除] ボタンをクリックします。

リストのフィールドの選択を解除するには

- ▶ Ctrl キーを押したまま、フィールド名をクリックします。

データベース エクスポートのスキーマ オプション

データベース エクスポートの [スキーマ] ダイアログ ボックスでは、データベース（これらのオプションをサポートするデータベース）へのエクスポートのオプションを設定し、フィールドの SQL データ型を設定し、第 1 入力フィールドを指定し、エクスポート時に作成した CREATE TABLE 文をカスタマイズすることができます。

図 7-3
データベース エクスポートの [スキーマ] ダイアログ ボックスの例



ダイアログ ボックスは次の部分に分けられます。

- 上部のセクション（表示されている場合）には、これらのオプションをサポートするデータベースへのエクスポートのオプションが表示されません。InfoSphere Warehouse データベースに接続していない場合、このセクションは表示されません。
- 中央部のテキスト フィールドは、CREATE TABLE コマンドの生成に使用するテンプレートを表示し、デフォルトでは次の形式となっています。

CREATE TABLE <table-name> <(table columns)>

- 下部にあるテーブルを使用して、各フィールドの SQL データ型を指定し、後述するように第 1 入力フィールド キーを指定できます。ダイアログ ボックスが、テーブルの仕様に基づいて自動的に <table-name> と <(table columns)> パラメータを生成します。

データベース エクスポート オプションの設定

このセクションが表示された場合、データベースへのエクスポートのさまざまな設定を指定できます。この機能をサポートするデータベースの種類は次のとおりです。

- DB2 9.1 以降で実行する IBM InfoSphere Warehouse。詳細は、[p. 509 IBM DB2 InfoSphere Warehouse のオプション](#) を参照してください。
- SQL Server 2008 以降 Enterprise および Developer Edition。詳細は、[p. 509 SQL Server のオプション](#) を参照してください。
- Oracle 10g および 11gR1 以降の Enterprise または Personal Editions。詳細は、[p. 510 Oracle のオプション](#) を参照してください。

CREATE TABLE 文のカスタマイズ

このダイアログ ボックスのテキスト フィールド部分から、CREATE TABLE 文にデータベース専用オプションを追加することができます。

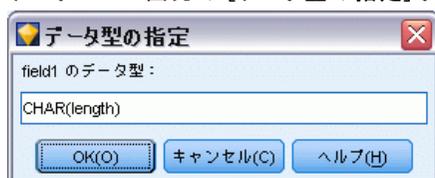
- ▶ [CREATE TABLE コマンドのカスタマイズ] のチェック ボックスを選択し、テキスト ウィンドウを有効にします。
- ▶ データベース専用オプションを文に追加します。<table-name> と (<table-columns>) パラメータのテキストを保持するようにしてください。なぜなら、それらは IBM® SPSS® Modeler によって実テーブル名と列の定義に置き換えられるからです。

SQL データ型の設定

デフォルトでは、SPSS Modeler によってデータベース サーバーが自動的に SQL データ型を割り当てることができます。データ型の自動割り当てを上書きするには、フィールドに対応する行を探して、スキーマ テーブルの [データ型] 列にあるドロップダウン リストから、目的のデータ型を選択します。Shift キーを押しながらクリックすると、複数の行を選択できます。

長さ、精度、または尺度引数を受け取るデータ型の場合 (BINARY、VARBINARY、CHAR、VARCHAR、NUMERIC、および NUMBER)、データベース サーバーに自動的に長さを割り当てさせるのではなく、自分で長さを指定する必要があります。たとえば、長さには VARCHAR(25) のような意味のある値を指定することにより、SPSS Modeler 中のストレージ タイプを確実に上書きすることができます。自動割り当てに優先させるには、[データ型] ドロップダウン リストで [指定] を選択し、データ型定義を適切な SQL タイプ定義文と置換します。

図 7-4
データベース出力の [データ型の指定] ダイアログ ボックス



もっとも簡単な方法は、まず目的の定義に似ているデータ型を選択し、次に [指定] を選択してその定義を編集することです。たとえば、SQL データ型に `VARCHAR(25)` を設定するには、まず [データ型] ドロップダウン リストでデータ型を `VARCHAR(length)` に設定してから、次に [指定] を選択してテキスト長に 25 を設定します。

第 1 入力キー

エクスポートされたテーブルに 1 つ以上の列は、一意の値またはすべての行の値の組合せでなければならず、適用する各フィールドの [第 1 入力フィールド キー] チェック ボックスをオンにして指示します。ほとんどのデータベースは、無効な第 1 入力フィールド キー を制限し、自動的に第 1 入力フィールド キーにインデックスを生成する方法で、テーブルの修正を禁止しています。(オプションで、インデックス ダイアログ ボックスでその他のフィールドのインデックスを生成できます)。詳細は、[p. 511 データベース エクスポートのインデックス オプション](#) を参照してください。

IBM DB2 InfoSphere Warehouse のオプション

テーブル スペース: エクスポートに使用するテーブルスペース。データベース管理者は、テーブルスペースを区分して作成または設定できます。データベースのエクスポートに使用する場合、(デフォルトのテーブルスペースではなく) これらのテーブルスペースのいずれかを選択することをお勧めします。

フィールドごとにデータを区分する: データ区分に使用する入力フィールドを指定します。

圧縮を使用: テーブルを作成し、圧縮してエクスポートします (たとえば、SQL の `CREATE TABLE MYTABLE(...) COMPRESS YES;` と同等のもの)。

SQL Server のオプション

圧縮を使用: 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル：圧縮のレベルを選択します。

- **行**：行レベルの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW);）。
- **ページ**：ページレベルの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE);）。

Oracle のオプション

Oracle 10g の設定

圧縮を使用：選択した場合は、圧縮によるエクスポート用のテーブルを作成します。本バージョンのデータベースの場合、基本的な圧縮のみ可能です（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS;）。

Oracle 11gR1 の設定

圧縮を使用：選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル：圧縮のレベルを選択します。

- **デフォルト**：デフォルトの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS;）。この場合、[ダイレクトロード操作] オプションと同じ効果があります。
- **ダイレクトロード操作**：バルク（ダイレクトパス）挿入操作圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS FOR DIRECT_LOAD OPERATIONS;）。
- **すべての操作**：すべての操作の圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS FOR ALL OPERATIONS;）。

Oracle 11gR2 の設定 - 基本オプション

圧縮を使用：選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル：圧縮のレベルを選択します。

- **デフォルト**：デフォルトの圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS;）。この場合、[基本] オプションと同じ効果があります。
- **基本**：基本の圧縮を有効にします（例：SQL の CREATE TABLE MYTABLE(...) COMPRESS BASIC;）。

Oracle 11gR2 の設定 - 高度なオプション

圧縮を使用： 選択した場合は、圧縮によるエクスポート用のテーブルを作成します。

圧縮レベル： 圧縮のレベルを選択します。

- **デフォルト：** デフォルトの圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS;`)。この場合、[基本] オプションと同じ効果があります。
- **基本：** 基本の圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS BASIC;`)。
- **OLTP：** OLTP の圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS FOR OLTP;`)。
- **Query Low/High:** (Exadata サーバーのみ) QUERY の Hybrid Columnar Compression 圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY LOW;` または `CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY HIGH;`)。データ ウェアハウス環境では QUERY の圧縮が役に立ちます。HIGH は、LOW より圧縮率が高くなります。
- **Archive Low/High:** (Exadata サーバーのみ) ARCHIVE の Hybrid Columnar Compression 圧縮を有効にします (例: SQL の `CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE LOW;` または `CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE HIGH;`)。長期間格納されるデータの圧縮には ARCHIVE の圧縮が役に立ちます。HIGH は、LOW より圧縮率が高くなります。

データベース エクスポートのインデックス オプション

インデックス ダイアログ ボックスで、IBM® SPSS® Modeler からエクスポートされたデータベース テーブルにインデックスを作成できるようになります。フィールド セットに含ませるフィールドを指定し、必要に応じて `CREATE INDEX` コマンドをカスタマイズします。

図 7-5
データベース出力の [インデックス] ダイアログ ボックス



ダイアログ ボックスには次の 2 つの部分があります。

- 最上部のテキスト フィールドは、**CREATE INDEX** コマンドの生成に使用するテンプレートを表示し、デフォルトは次のフォーマットとなっています。

CREATE INDEX <index-name> ON <table-name>

- ダイアログ ボックス下部にあるテーブルで、生成する各インデックスの仕様を追加することができます。各インデックスに対しては、インデックス名とフィールドまたは含める列を指定します。ダイアログ ボックスが、自動的に <index-name> と <table-name> パラメータを生成します。たとえば、フィールド empid と deptid に関する単一のインデックスに対して生成された SQL はこのようになります。

CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID)

複数のインデックスを生成するために複数の行を追加することができます。独立した **CREATE INDEX** コマンドが各行に生成されます。

CREATE INDEX コマンドのカスタマイズ

オプションで、**CREATE INDEX** コマンドをすべてのインデックスに対して、または特定のインデックスのみに対してカスタマイズできます。これによって、必要に応じて自由に、固有のデータベースの要件またはオプションを統合し、すべてのインデックスまたは特定のインデックスのみに対してカスタマイズの結果を適用することができるようになります。

- ダイアログ ボックスの最上部にある [CREATE INDEX コマンドをカスタマイズ] を選択し、これ以降追加されるすべてのインデックスに使用するテンプレートを変更します。変更結果は、既にテーブルに追加されているインデックスに対して自動的に適用されます。
- テーブルで 1 つ以上の行を選択し、ダイアログ ボックスの最上部にある [選択したインデックスを更新] をクリックして現在のカスタマイズした結果を選択したすべての行に適用します。
- 各行の [カスタマイズ] チェック ボックスをオンにして、そのインデックのみに対してコマンド テンプレートを修正します。

<index-name> と <table-name> パラメータの値が、テーブルの仕様に基いてダイアログ ボックスによって自動的に生成され、直接編集できないことに注意してください。

BITMAP KEYWORD : Oracle のデータベースを使用中に、テンプレートをカスタマイズして、標準のインデックスではなく、次のようにビットマップ インデックスを生成することができます。

CREATE BITMAP INDEX <index-name> ON <table-name>

ビットマップ インデックスは、少数の重複レコード値を持つ列のインデックス生成に役立ちます。生成された SQL はこのようになります。

CREATE BITMAP INDEX MYTABLE_IDX1 ON MYTABLE(COLOR)

UNIQUE キーワード : 多くのデータベースが、CREATE INDEX コマンドの UNIQUE キーワードをサポートしています。これは、基本的なテーブルにおける第 1 入力フィールド キーの制約に類似した、一意性の制約を強制します。

CREATE UNIQUE INDEX <index-name> ON <table-name>

第 1 入力フィールド キーとして実際に指定されたフィールドに対して、この指定は必要でないことに注意してください。ほとんどのデータベースは、第 1 入力フィールド キーとして指定されたフィールドのインデックスを、CREATE TABLE コマンド内に自動的に生成します。したがって、これらのフィールドで明示的に生成するインデックスは必要ありません。詳細は、[p. 507 データベース エクスポートのスキーマ オプション](#) を参照してください。

FILLFACTOR キーワード : 一部のインデックス用の物理パラメータを精密に調整することができます。たとえば、SQL Server では、ユーザーがインデックス サイズを、(初期生成後に) 将来テーブルに変更が加えられたときのメンテナンス コストに対してトレード オフすることができます。

CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID) WITH FILLFACTOR=20

その他のコマンド

- 指定された名前を持つ既存のインデックスがある場合、インデックスの生成は失敗します。失敗は警告として処理され、それ以降のインデックスの生成を許可し、それから、すべてのインデックスを試みた後にメッセージ ログ内にエラーとして再報告されます。 [詳細は、5 章 ストリーム操作メッセージの表示 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- 最高のパフォーマンスを得るためには、データがテーブルにロードされた後にインデックスを生成すべきです。インデックスには最低 1 行は含まれていなければなりません。
- ノードを実行する前に、メッセージ ログ内の生成された SQL をプレビューすることができます。 [詳細は、6 章 生成された SQL のプレビュー in IBM SPSS Modeler Server 15 管理およびパフォーマンス ガイド](#) を参照してください。
- データベースに書き込まれる一時テーブル（つまり、ノード キャッシュが有効なとき）に対しては、第 1 入力フィールド キーを指定するオプションは利用できません。ただし、データが下流ノードで使用される頻度に合わせて、システムが一時テーブルにインデックスを生成します。たとえば、キャッシュされたデータがその後 DEPT 列ごとに結合された場合、キャッシュされているこの列のテーブルをインデックスする意味があります。 [詳細は、5 章 ノードのキャッシュ オプション in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

インデックスおよびクエリーの最適化

一部のデータベースの管理システムでは、データベースのテーブルの生成、ロード、インデックスの作成が行われた後に、オブティマイザーがインデックスを利用して新規のテーブル上でのクエリーの高速化ができるようになるまで、あるステップが必要になります。たとえば Oracle では、コストベースのクエリー オプティマイザーが、テーブルのインデックスをクエリーの最適化に使用できるようにする前に、テーブルの解析を要求します。Oracle の ODBC プロパティ ファイル（ユーザーには見えません）には、これを実行する次のようなオプションが含まれます。

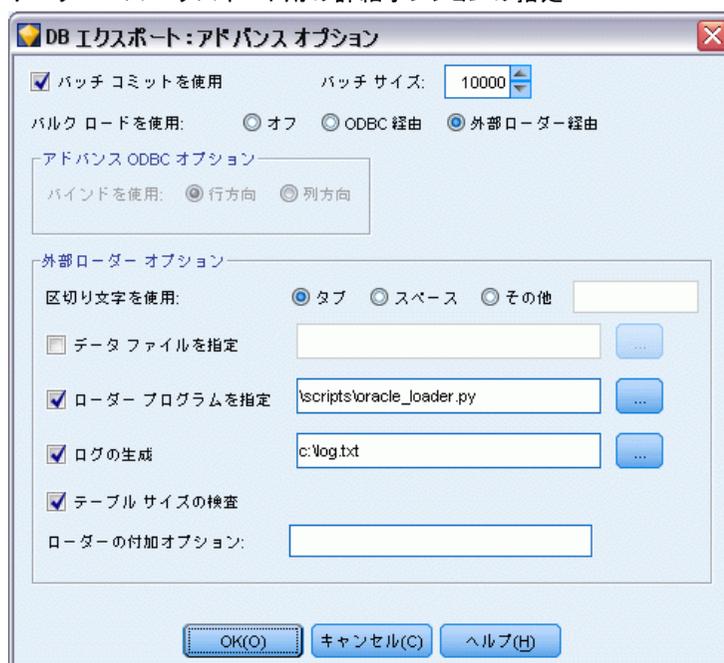
```
# Defines SQL to be executed after a table and any associated indexes  
# have been created and populated table_analysis_sql, 'ANALYZE TABLE  
table-name  
COMPUTE STATISTICS
```

このステップは、Oracle 内にテーブルが生成されるときに必ず実行されます（第 1 入力フィールド キーまたはインデックスが定義済みか否かにかかわらず）。必要に応じて、その他のデータベース用の ODBC プロパティ ファイルを同様にカスタマイズすることができます。サポートにお問い合わせください。

データベース エクスポートの詳細オプション

データベース エクスポート ノードのダイアログ ボックスで [アドバンス] ボタンをクリックすると、結果をデータベースにエクスポートするための技術的な詳細を設定できるダイアログ ボックスが表示されます。

図 7-6
データベース エクスポート用の詳細オプションの指定



バッチ コミットを使用： データベースへの行ごとのコミットを無効にする場合に選択します。

バッチ サイズ： メモリーにコミットする前に、データベースに送信するレコード数を指定します。この値を減らすと、データの整合性は向上しますが、転送速度が遅くなります。この値を調整して、データベースのパフォーマンスを最適な状態にすることができます。

InfoSphere Warehouse オプション： InfoSphere Warehouse データベース (IBM DB2 9.7 以降) に接続している場合にのみ表示されます。[更新を記録しない] を選択すると、テーブルを作成したりデータを挿入したりした場合にイベントがログ記録されないようにします。

バルク ロードを使用： IBM® SPSS® Modeler からデータベースに直接データをバルク ロードするための方法を選択します。どのバルク ロードが特定のシナリオに適切かを選択するには、実験が必要です。

- **ODBC 経由**：通常のデータベースへのエクスポートと比べて大幅に効率的な、ODBC API を使った複数行挿入を実施する場合に選択します。下にあるオプションから、行方向または列方向のバインドを選択してください。
- **外部ローダー経由**：データベース固有のカスタム バルク ロード プログラムを使用する場合に選択します。このオプションを選択すると、下にある [外部ローダー オプション] が有効になります。

詳細 ODBC オプション：これらのオプションは、[ODBC 経由] を選択した場合にだけ利用できます。この機能はすべての ODBC ドライバに対応しているわけではないことに注意してください。

- **行方向**：データベースへのデータのロードに、`SQLBulkOperations` コールを使用する場合に選択します。一般的に行方向のバインドは、データをレコード単位に挿入するパラメータ化された `INSERT` 文を使用する場合に比べて、速度を向上することができます。
- **列方向**：データベースへのデータのロードに、列方向のバインドを使用する場合に選択します。列方向のバインドは、各データベースの列（パラメータ化された `INSERT` 文中の）を `N` 値の配列にバインドすることにより、パフォーマンスを向上します。`INSERT` を 1 回実行すると、データベースに `N` 行が挿入されます。この方法により、大幅にパフォーマンスを向上することができます。

外部ローダー オプション [外部ローダー経由] を選択すると、データ セットをファイルにエクスポートしたり、そのファイルからデータをデータベースにロードするための外部ローダー プログラムの指定と実行を行うためのオプションを設定できるようになります。SPSS Modeler は多くの一般的なデータベース システムの外部ローダーと連携できます。このソフトウェアには、テクニカル ドキュメントと一緒にスクリプトがいくつか付属しており、それらは `scripts` サブディレクトリの下にあります。この機能を使用するには、Python 2.7 を SPSS Modeler または IBM® SPSS® Modeler Server と同じマシンにインストールし、`python_exe_path` パラメータを `options.cfg` ファイルに設定する必要があります。詳細は、[p. 517 バルクローダーのプログラミング](#) を参照してください。

- **区切り文字を使用**：エクスポートするファイルで使用する区切り文字を指定します。タブ文字で区切る場合は [タブ] を、スペースで区切る場合は [スペース] を選択します。カンマ (,) などの他の文字を指定する場合は、[その他] を選択します。
- **データファイルの指定**：バルク ロード時に書き込まれる、データ ファイルのパスを入力する場合に選択します。デフォルトでは、サーバーの `temp` ディレクトリに一時ファイルが作成されます。
- **ローダー プログラムの指定**：バルク ロード プログラムを指定する場合に選択します。デフォルトでは、SPSS Modeler インストール ディレクトリ中の `/scripts` サブディレクトリから、データベースに対して実行される Python スクリプトが検索されます。このソフトウェアには、テ

クニカル ドキュメントと一緒にスクリプトがいくつか付属しており、それらは `scripts` サブディレクトリの下にあります。

- **ログの生成**：指定したディレクトリにログを生成する場合に選択します。ログ ファイルには、バルク ロード操作が失敗した場合などに役立つエラー情報が含まれています。
- **テーブル サイズの検査**：SPSS Modeler からエクスポートされる行数に対応してテーブル サイズを増やすために、テーブル検査を実施する場合に選択します。
- **ローダーの付加オプション**：ローダー プログラムに対する引数を指定します。スペースを含む引数には二重引用符を使用してください。

円記号を前に付けることにより、オプションの引数中に二重引用符を指定することができます。たとえば、`-comment "This is a \"comment\""` として指定されたオプションには、`-comment` フラグおよび `This is a "comment"` のように表示されるコメントが含まれます。

円記号を 1 つ指定するには、もう 1 つの円記号を付けて指定します。たとえば、`-specialdir "C:\\Test Scripts\\"` として指定されたオプションには、フラグ `-specialdir` および `C:\\Test Scripts\\` のように表示されるディレクトリが含まれます。

バルク ローダーのプログラミング

データ エクスポート ノードは、[詳細オプション] ダイアログ ボックスにバルク ロードのオプションがあります。バルク ローダー プログラムを使用して、データをテキストからデータベースにロードすることができます。

オプション [Use bulk loading - via external loader] は IBM® SPSS® Modeler を次の 3 つに構成します。

- 要求されたすべてのデータベース テーブルを生成する。
- データベースをテキスト ファイルにエクスポートする。
- バルク ローダー プログラムを起動し、このファイルからデータをデータベースにロードする。

一般に、バルク ローダー プログラムはデータベース ロード ユーティリティそのもの（たとえば、Oracle の `sqlldr` ユーティリティ）ではありませんが、正しい引数を形成し、データベース固有の補助ファイル（コントロール ファイルなど）を生成し、次にデータベース ロード ユーティリティを起動する小型のスクリプトまたはプログラムです。以下のセクションの情報は、既存のバルク ローダーを編集できます。

またはバルク ロードの独自のプログラムを作成することもできます。詳細は、[p. 523 バルク ローダー プログラムの開発](#) を参照してください。

バルク ロードのスク립ト

SPSS Modeler には、Python スクリプトを使用して実装されるさまざまなデータベースのさまざまなバルク ロダーが用意されています。**[外部ローダー経由]** オプションを選択してデータベース エクスポート ノードを含むストリームを実行すると、SPSS Modeler は ODBC 経由でデータベース テーブルを（必要に応じて）作成し、IBM® SPSS® Modeler Server を実行するホストの一時ファイルにデータをエクスポートし、バルク ロード スクリプトを起動します。このスクリプトは DBMS ベンダーが提供するユーティリティを実行し、一時ファイルのデータをデータベースにアップロードします。

注：SPSS Modeler のインストールには Python ランタイム インタープリターが含まれていないため、Python を別途インストールする必要があります。詳細は、[p. 515 データベース エクスポートの詳細オプション](#) を参照してください。

以下のデータベースのスクリプトが提供されます（SPSS Modeler インストール ディレクトリの %scripts フォルダ）。

テーブル 7-1
提供されるバルク ロード スクリプト

| データベース | スクリプト名 | |
|-------------|--------------------|---|
| IBM DB2 | db2_loader.py | 詳細は、 p. 518 データの IBM DB2 データベースへのバルク ロード を参照してください。 |
| IBM Netezza | netezza_loader.py | 詳細は、 p. 519 データの IBM Netezza データベースへのバルク ロード を参照してください。 |
| Oracle | oracle_loader.py | 詳細は、 p. 520 データの Oracle データベースへのバルク ロード を参照してください。 |
| SQL Server | mssql_loader.py | 詳細は、 p. 522 データの SQL Server データベースへのバルク ロード を参照してください。 |
| Teradata | teradata_loader.py | 詳細は、 p. 522 データの Teradata データベースへのバルク ロード を参照してください。 |

データの IBM DB2 データベースへのバルク ロード

以下のポイントによって、IBM® SPSS® Modeler からのバルク ロードをデータベース エクスポートの [詳細オプション] ダイアログ ボックスの [外部ローダー] オプションを使用する IBM DB2 データベースに設定することができます。

DB2 コマンドライン プロセッサ (CLP) ユーティリティをインストールする

スクリプト db2_loader.py が DB2 LOAD コマンドを起動します。コマンドライン プロセッサ (UNIX の場合は db2、Windows の場合は db2cmd) を db2_loader.py が実行されるサーバー (通常、IBM® SPSS® Modeler Server を実行するホスト) にインストールします。

ローカル データベースのエイリアス名が実際のデータベース名と同じかどうかを確認する

DB2 ローカル データベースのエイリアスは、ローカルまたはリモート DB2 インスタンスのデータベースを参照する DB2 クライアント ソフトウェアによって使用される名前です。ローカル データベースのエイリアスがリモート データベースの名前と異なる場合、追加のローダー オプションを指定します。

```
-alias <local_database_alias>
```

たとえば、ホスト GALAXY のリモート データベースの名前は STARS ですが、SPSS Modeler Server を実行するホストの DB2 ローカル データベースのエイリアスは STARS_GALAXY です。追加のローダー オプションを使用

```
-alias STARS_GALAXY
```

ASCII 文字以外のデータ エンコード

ASCII 形式でないデータをバルク ロードしている場合、db2_loader.py の設定セクションのコードページ変数が正しく設定されている必要があります。

空白の文字列

空白の文字列はデータベースに NULL 値としてエクスポートされます。

データの IBM Netezza データベースへのバルク ロード

以下のポイントによって、IBM® SPSS® Modeler からのバルク ロードをデータベース エクスポートの [詳細オプション] ダイアログ ボックスの [外部ローダー] オプションを使用する IBM Netezza データベースに設定することができます。

Netezza nzload ユーティリティをインストールする

スクリプト netezza_loader.py が Netezza ユーティリティ nzload を起動します。nzload を、netezza_loader.py を実行するサーバーにインストールして正しく設定します。

非 ASCII データのエクスポート

エクスポートに ASCII 形式でないデータが含まれている場合、`-encoding UTF8` を [DB エクスポートの詳細オプション] ダイアログ ボックスの [追加のローダー オプション] フィールドに追加する必要があります。ASCII 以外のデータは、正しくアップロードする必要があります。

日付、時刻、タイムスタンプ形式のデータ

ストリームのプロパティで、日付の形式を [DD-MM-YYYY] に、時刻の形式を [HH:MM:SS] に設定します。詳細は、[5 章 ストリームのオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

空白の文字列

空白の文字列はデータベースに NULL 値としてエクスポートされます。

データを既存のテーブルに挿入する場合のストリームおよび対象テーブルの列の異なる順序

ストリームの列の順序が対象テーブルの列の順序と異なる場合、日付の値が誤った列に挿入されます。フィールドの順序変更ノードを使用して、ストリームの列の順序が対象テーブルの順序と一致するようにします。詳細は、[4 章 p.265 フィールド順序ノード](#) を参照してください。

nzload の進捗状況の追跡

SPSS Modeler をローカル モードで実行している場合、`-sts` を [DB エクスポートの詳細オプション] ダイアログ ボックスの [追加のローダー オプション] フィールドに追加して、nzload ユーティリティで開いたコマンド ウィンドウに 1000 行ごとのステータス メッセージが表示されるようにします。

データの Oracle データベースへのバルク ロード

以下のポイントによって、IBM® SPSS® Modeler からのバルク ロードをデータベース エクスポートの [詳細オプション] ダイアログ ボックスの [外部ローダー] オプションを使用する Oracle データベースに設定することができます。

Oracle sqlldr ユーティリティをインストールする

スクリプト `oracle_loader.py` が Oracle ユーティリティ `sqlldr` を起動します。`sqlldr` は Oracle Client に自動的に含まれません。`sqlldr` は、`oracle_loader.py` が実行されるサーバーにインストールされます。

データベース SID またはサービス名を指定します

データをローカル以外の Oracle サーバーにエクスポートしている場合、またはローカルの Oracle サーバーに複数のデータベースがある場合、[DB エクスポートの詳細オプション] ダイアログ ボックスの [追加のローダー オプション] フィールドで以下を指定し、SID またはサービス名で渡す必要があります。

```
-database <SID>
```

oracle_loader.py の設定セクションの編集

UNIX（およびオプションで、Windows）システムで、oracle_loader.py スクリプトの始めの設定セクションを編集します。ここで、ORACLE_SID、NLS_LANG、TNS_ADMIN および ORACLE_HOME 環境変数の値を、sqlldr ユーティリティのフル パスとともに必要に応じて指定できます。

日付、時刻、タイムスタンプ形式のデータ

ストリームのプロパティで、通常は日付の形式を [YYYY-MM-DD] に、時刻の形式を [HH:MM:SS] に設定する必要があります。詳細は、[5 章 ストリームのオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

上記と異なる日付および時刻の形式を使用する必要がある場合、Oracle のマニュアルを参照し、oracle_loader.py スクリプト ファイルを編集してください。

ASCII 文字以外のデータ エンコード

ASCII 形式でないデータをバルク ロードしている場合、環境変数 NLS_LANG が正しく設定されている必要があります。Oracle ローダー ユーティリティ sqlldr によって読み込まれます。たとえば、Windows の Shift-JIS の NLS_LANG に対する正しい値は Japanese_Japan.JA16SJIS です。NLS_LANG の詳細については、Oracle のマニュアルを参照してください。

空白の文字列

空白の文字列はデータベースに NULL 値としてエクスポートされます。

データの SQL Server データベースへのバルクロード

以下のポイントによって、IBM® SPSS® Modeler からのバルクロードをデータベースエクスポートの [詳細オプション] ダイアログボックスの [外部ローダー] オプションを使用する SQL Server データベースに設定することができます。

SQL Server bcp.exe ユーティリティをインストールする

スクリプト `mssql_loader.py` が SQL Server ユーティリティ `bcp.exe` を起動します。`bcp.exe` は、`mssql_loader.py` が実行されるサーバーにインストールされます。

区切り文字としてのスペースの使用は無効

[DB エクスポートの詳細オプション] ダイアログボックスでは、スペースを区切り文字として使用しないでください。

テーブルサイズを確認オプションの推奨

[DB エクスポートの詳細オプション] ダイアログボックスで [テーブルサイズを確認] オプションを有効にすることをお勧めします。バルクロードプロセスの障害は常に検知されるわけではないため、このオプションを有効にして追加のチェックを行い、適切な数の行が読み込まれていることを確認します。

空白の文字列

空白の文字列はデータベースに NULL 値としてエクスポートされます。

データの Teradata データベースへのバルクロード

以下のポイントによって、IBM® SPSS® Modeler からのバルクロードをデータベースエクスポートの [詳細オプション] ダイアログボックスの [外部ローダー] オプションを使用する Teradata データベースに設定することができます。

Teradata fastload ユーティリティをインストールする

スクリプト `teradata_loader.py` が Teradata ユーティリティ `fastload` を起動します。`fastload` を、`teradata_loader.py` を実行するサーバーにインストールして正しく設定します。

データのバルク ロードは空のテーブルにのみ可能

バルク ロードのターゲットとしては空のテーブルだけが使用できます。バルク ロード前の対象テーブルにデータが含まれている場合、操作は失敗します。

日付、時刻、タイムスタンプ形式のデータ

ストリームのプロパティで、日付の形式を [YYYY-MM-DD] に、時刻の形式を [HH:MM:SS] に設定します。詳細は、[5 章 ストリームのオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

空白の文字列

空白の文字列はデータベースに NULL 値としてエクスポートされます。

Teradata プロセス ID (tdpid)

デフォルトで、fastload は `tdpid=dbc` によってデータを Teradata システムにエクスポートします。通常、`dbccop1` が Teradata サーバーの IP アドレスと関連する HOSTS ファイルにエントリがあります。異なるサーバーを使用する場合、[DB エクスポートの詳細オプション] ダイアログ ボックスの [追加のローダー オプション] フィールドで以下を指定し、このサーバーの `tdpid` を渡します。

```
-tdpid <id>
```

テーブル名および列名のスペース

テーブル名または列名にスペースが含まれる場合、バルク ロード操作は失敗します。可能な場合は、テーブル名または列名の名前を変更してスペースを削除してください。

バルク ローダー プログラムの開発

このトピックでは、IBM® SPSS® Modeler から実行して、テキスト ファイルのデータをデータベースに読み込むバルク ローダー プログラムの開発方法について説明します。

Python を使用したバルク ローダー プログラムの作成

デフォルトでは、SPSS Modeler は、データベースのタイプに基づいてデフォルトのバルク ローダー プログラムを検索します。[テーブル 7-1 p. 518](#) を参照してください。

スクリプト test_loader.py を使用して、バッチ ロード プログラムの開発を支援します。詳細は、p. 526 バルク ロード プログラムのテスト を参照してください。

バルク ロード プログラムに渡されるオブジェクト

SPSS Modeler は、バルク ロード プログラムに渡される 2 つのファイルを作成します。

- **データ ファイル:** 読み込まれるデータ (テキスト形式) が含まれます。
- **スキーマ ファイル:** 列の名前と種類を説明する XML ファイルで、データ ファイルをどのように書式設定するかに関する情報が提供されます (フィールド間の区切り文字として使用される文字など)。

また、SPSS Modeler は、バルク ロード プログラム起動時に、テーブル名、ユーザー名、パスワード名など他の情報を引数として渡します。

注 : SPSS Modeler に正常に完了したことを通知するために、バルク ロード プログラムはスキーマ ファイルを削除する必要があります。

バルク ロード プログラムに渡される引数

プログラムに渡される引数は次のとおりです。

テーブル 7-2
バルク ロード プログラムに渡される引数

| 引数 | 説明 |
|--------------|------------------------------------|
| schemafile | スキーマ ファイルのパス。 |
| data file | データ ファイルのパス。 |
| servername | DBMS サーバーの名前。空白の場合あり。 |
| databasename | DBMS サーバー内のデータベースの名前。空白の場合あり。 |
| username | データベースにログインするユーザー名。 |
| password | データベースにログインするパスワード。 |
| tablename | 読み込むテーブルの名前。 |
| ownername | テーブル所有者の名前 (スキーマ名)。 |
| logfile | ログファイルの名前 (空白の場合、ログ ファイルは作成されません)。 |
| rowcount | データセット内の行数。 |

[DB エクスポートの詳細オプション] ダイアログボックスの [追加のロード オプション] フィールドで指定されたオプションは、これらの標準引数の後でバルク ロード プログラムに渡されます。

データ ファイルの形式

データはテキスト形式でデータ ファイルに書き込まれ、各フィールドは [DB エクスポートの詳細オプション] ダイアログボックスで指定された区切り文字で区切られます。以下に、タブ区切りのデータ ファイルがどのように表示されるかの例を示します。

```
48 F HIGH NORMAL 0.692623 0.055369 drugA
15 M NORMAL HIGH 0.678247 0.040851 drugY
37 M HIGH NORMAL 0.538192 0.069780 drugA
35 F HIGH HIGH 0.635680 0.068481 drugA
```

ファイルは、IBM® SPSS® Modeler Server で使用されるローカル エンコードで作成されます (SPSS Modeler Server に接続していない場合は SPSS Modeler)。いくつかの形式は、SPSS Modeler ストリーム設定で制御されます。

スキーマ ファイルの形式

スキーマ ファイルは、データ ファイルを記述する XML ファイルです。以下に、先行するデータ ファイルに伴う例を示します。

```
<?xml version="1.0" encoding="UTF-8"?>
<DBSCHEMA version="1.0">
  <table delimiter="\t" commit_every="10000" date_format="YYYY-MM-DD" time_format="HH:MM:SS"
  append_existing="false" delete_datafile="false">
    <column name="Age" encoded_name="416765" type="integer"/>
    <column name="Sex" encoded_name="536578" type="char" size="1"/>
    <column name="BP" encoded_name="4250" type="char" size="6"/>
    <column name="Cholesterol" encoded_name="43686F6C65737465726F6C" type="char" size="6"/>
    <column name="Na" encoded_name="4E61" type="real"/>
    <column name="K" encoded_name="4B" type="real"/>
    <column name="Drug" encoded_name="44727567" type="char" size="5"/>
  </table>
</DBSCHEMA>
```

次の表は、スキーマ ファイルの <table> および <column> 要素の属性を示します。

テーブル 7-3

<table> 要素の属性

| 属性 | 説明 |
|--------------|---|
| delimiter | フィールド区切り文字 (TAB は \t として示されます)。 |
| commit_every | バッチ サイズの間隔 ([DB エクスポートの詳細オプション] ダイアログボックスで指定されているとおり) |
| date_format | 日付の表示に使用する形式。 |
| time_format | 時刻の表示に使用する形式。 |

| 属性 | 説明 |
|-----------------|--|
| append_existing | 読み込まれるテーブルにデータが含まれている場合は true。そうでない場合は false。 |
| delete_datafile | バルク ロード プログラムがロードの完了時にデータ ファイルを削除するかどうかは true。 |

テーブル 7-4
 <column> 要素の属性

| 属性 | 説明 |
|--------------|--|
| name | 列名。 |
| encoded_name | データ ファイルと出力と同じエンコードに 2 桁の 16 進数で変換される列の名前。 |
| type | 列のデータ型。integer、real、char、time、date、および datetime のいずれか。 |
| size | char データ型の場合、列の最大幅の文字数。 |

バルク ローダー プログラムのテスト

IBM® SPSS® Modeler インストール ディレクトリの %scripts フォルダに含まれるテスト スクリプト test_loader.py i を使用して、バルク ロードをテストできます。このテストは、SPSS Modeler で使用するバルク ロード プログラムまたはスクリプトを開発、デバッグまたはトラブルシューティングする場合に役立ちます。

テスト スクリプトを使用するには、以下の手順に従います。

- ▶ test_loader.py スクリプトを実行して、スキーマ ファイルとデータ ファイルをファイル schema.xml および data.txt にコピーし、Windows バッチファイル (test.bat) を作成します。
- ▶ test.bat ファイルを編集して、テストするバルク ローダー プログラムまたはスクリプトを選択します。
- ▶ コマンド シェルから test.bat ファイルを実行して、バルク ロード プログラムまたはスクリプトをテストします。

注： test.bat を実行してもデータは実際にデータベースには読み込まれません。

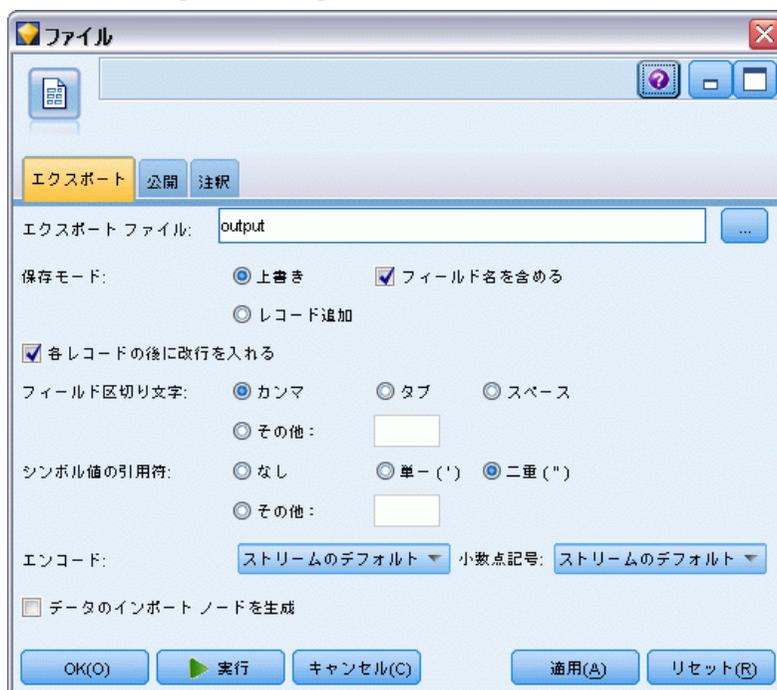
ファイル エクスポート ノード

ファイル エクスポート ノードを使用すると、データを区切り文字で区切った形式でテキスト ファイルに書き込むことができます。他の分析ソフトや表計算ソフトに読み込める形式でデータをエクスポートする場合に役に立ちます。

注：IBM® SPSS® Modeler がキャッシュ ファイルに古いキャッシュ形式を使用しないため、古いキャッシュ形式のファイルを書き込むことはできません。SPSS Modeler のキャッシュ ファイルは IBM® SPSS® Statistics.sav 形式で保存され、Statistics エクスポート ノードを使用して書き込みできます。詳細は、8 章 p.561 Statistics エクスポート ノードを参照してください。

ファイル ノードの [エクスポート] タブ

図 7-7
ファイル ノード、[エクスポート] タブ



エクスポート ファイル： ファイルの名前を指定します。ファイル名を入力するか、またはファイル選択ボタン をクリックしてファイルの場所を指定します。

保存モード：[上書き] を選択すると、指定したファイル中の既存のデータが上書きされます。[レコード追加] を選択すると、出力データが既存ファイルの末尾に追加され、既存データはそのまま保存されます。

- **フィールド名を含める**：このオプションを選択すると、出力ファイルの 1 行目にフィールド名が書き込まれます。このオプションは、保存モードで [上書き] を選択した場合にだけ利用できます。

各レコードの後に改行を入れる：このオプションを選択すると、各レコードが出力ファイル中の新しい行に書き込まれます。

フィールド区切り文字：生成するテキスト ファイルで、フィールド値の間に挿入する文字列を選択します。[カンマ]、[タブ]、[スペース]、または [その他] を選択することができます。[その他] を選択した場合は、テキスト ボックスに適切な区切り文字を入力してください。

シンボル値の引用符：シンボル値フィールドの値に対して使用する引用符の種類を選択します。[なし] (値に引用符を付けない)、[単一(')]、[二重(")]、または [その他] を選択できます。[その他] を選択した場合は、テキスト ボックスに適切な引用文字を入力してください。

文字コード：使用するテキストのエンコード方法を指定します。サーバー デフォルト、システム デフォルト、UTF-8 から選択できます。

- 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 IBM SPSS Modeler 内での Unicode のサポート in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。
- デフォルトは、[ストリーム プロパティ] ダイアログ ボックスで指定されます。詳細は、[5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

小数点記号：データ中で小数点記号をどのように表すかを指定します。

- **ストリームのデフォルト：**現在のストリームのデフォルト設定で定義された小数点区切り文字が使用されます。これは、通常、コンピュータのロケールの設定で定義された小数点区切り文字になります。
- **ピリオド(.)：**小数点区切り文字として、ピリオドを使用します。
- **カンマ(,)：**小数点区切り文字として、カンマを使用します。

データのインポートノードを生成：エクスポートされたデータ ファイルを読み込む、可変長入力ノードを自動的に生成する場合に選択します。詳細は、[2 章 p. 28 可変長ファイル ノード](#) を参照してください。

IBM SPSS Data Collection エクスポート ノード

IBM® SPSS® Data Collection エクスポート ノードは、Data Collection Data Model に基づき、Data Collection の市場調査ソフトウェアで使用する形式でデータを保存します。このフォーマットは、調査中に収集された質問に対する実際の回答であるケース データを、ケース データが収集され整理されたメタデータと区別します。メタデータは、質問テキスト、変数名とその説明、複数のレスポンス セット、種々のテキストの翻訳、ケース データの構造の定義などの情報から構成されます。詳細は、[2 章 p. 39 Data Collection ノード](#) を参照してください。

図 7-8
IBM SPSS Data Collection エクスポート ノードの [エクスポート] タブ



注：このノードには、Data Collection ソフトウェアと共に配布される Data Collection Data Model バージョン 4.0 以降が必要です。詳細は、Data Collection Web ページ <http://www.ibm.com/software/analytics/spss/products/data-collection/> を参照してください。Data Model のインストールとは別に、追加の設定を行う必要はありません。

メタデータ ファイル：メタデータが保存される質問定義ファイルの名前 (.mdd) を指定します。デフォルトの質問は、フィールドのデータ型情報に基づいて作成されます。たとえば、名義型 (セット型) フィールドは、定義された各値の質問テキストおよび個別のチェック ボックスとして使用するフィールド説明を含む単一の質問として表されます。

メタデータ結合：メタデータが既存のバージョンを上書きするか、既存のメタデータと結合するかを指定します。結合オプションを選択した場合、ストリームを実行するごとに新しいバージョンが作成されます。これにより、変更が行われるごとに質問のバージョンを記録することができます。各バージョンは、ケース データの特定のセットを収集するために使用されるメタデータのスナップショットと見なすことができます。

システム変数を使用：システム変数がエクスポートされた .mdd ファイルに含まれるかどうかを指定します。Respondent.Serial、Respondent.Origin、DataCollection.StartTime などの変数が含まれます。

ケース データの設定： ケース データがエクスポートされる IBM® SPSS® Statistics データ (.sav) ファイルを指定します。変数名および値名のすべての制限がここで適用されるため、たとえば [フィルタ] タブに切り替えて、[フィルタ] オプション メニューで「SPSS Statistics 用に名前を変更する」オプションを使用し、フィールド名の無効な文字を修正する必要があります。

データのインポート ノードを生成： エクスポートされたデータ ファイルを読み込む Data Collection 入力ノードを、自動的に生成する場合に選択します。

複数レスポンス設定： ストリームに定義された複数のレスポンス セットは、ファイルがエクスポートされると自動的に保存されます。[フィルタ] タブで、ノードの複数のレスポンス セットを表示および編集することができます。詳細は、4 章 p.177 [複数レスポンス セット編集](#) を参照してください。

IBM Cognos BI エクスポート ノード

IBM Cognos BI エクスポート ノードを使用して、IBM® SPSS® Modeler ストリームから Cognos BI にデータを UTF-8 形式でエクスポートできます。こうすることによって、Cognos BI は SPSS Modeler からの変換データまたはスコアリング データを使用できます。たとえば、Cognos BI Report Studio を使用して、予測値や確信度値など、エクスポートされたデータに基づいてレポートを作成できます。レポートは Cognos BI サーバーに保存し、Cognos BI ユーザーに配布できます。

注： 関連データだけをエクスポートし、OLAP データはエクスポートしません。

データを Cognos BI にエクスポートするには、次を指定する必要があります。

- Cognos 接続 - Cognos BI サーバーへの接続
- ODBC 接続 - Cognos BI サーバーが使用する Cognos データ サーバーへの接続

Cognos の接続内では、使用する Cognos データソースを指定します。このデータソースは、ODBC データソースと同じログインを使用する必要があります。

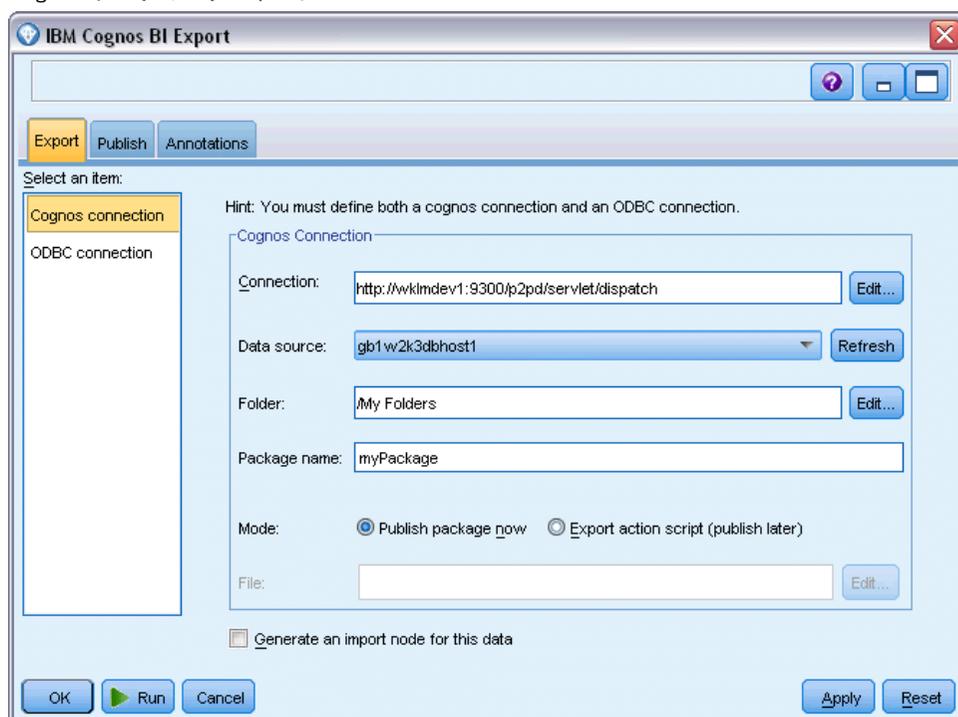
実際のストリーム データをデータ サーバーにエクスポートし、パッケージ メタデータを Cognos BI サーバーにエクスポートします。

その他のエクスポート ノードと同様に、ノード ダイアログ ボックスの [公開] タブを使用し、IBM® SPSS® Modeler Solution Publisher で展開するストリームを公開できます。

Cognos の接続

エクスポートに使用する Cognos BI サーバーへの接続を指定します。この手順では、メタデータを Cognos BI サーバーの新しいパッケージにメタデータをエクスポートし、ストリーム データは Cognos データ サーバーにエクスポートされます。

図 7-9
Cognos データのエクスポート



接続 :[編集] ボタンをクリックするとダイアログ ボックスが表示され、データをエクスポートする Cognos BI サーバーの URL とその他の詳細を定義できます。IBM® SPSS® Modeler 経由ですでに Cognos BI サーバーにログインしている場合、現在の接続の詳細を編集することもできます。詳細は、[2 章 p. 54 Cognos の接続](#) を参照してください。

データソース : データをエクスポートしている Cognos データ ソース (通常はデータベース) の名前。ドロップダウン リストには、現在の接続でアクセスできる Cognos データ ソースがすべて表示されます。リストを更新するには、[リフレッシュ] ボタンをクリックします。

フォルダ : エクスポート パッケージを作成する Cognos BI サーバーのフォルダのパスと名前。

パッケージ名: エクスポートされたメタデータを含む指定フォルダ内のパッケージの名前。単一のクエリ サブジェクトの新しいパッケージでなければなりません。既存のパッケージへはエクスポートできません。

モード: エクスポートの実行方法を指定します。

- **パッケージを今すぐ公開:** (デフォルト) **[実行]** をクリックするとすぐにエクスポート操作が実行されます。
- **アクション スクリプトをエクスポート:** エクスポートを後で実行する XML スクリプトを作成します (Framework Manager を使用するなど)。 **[ファイル]** フィールドにスクリプトのパスおよびファイル名を入力するか、 **[編集]** ボタンを使用して、スクリプト ファイルの名前および場所を指定します。

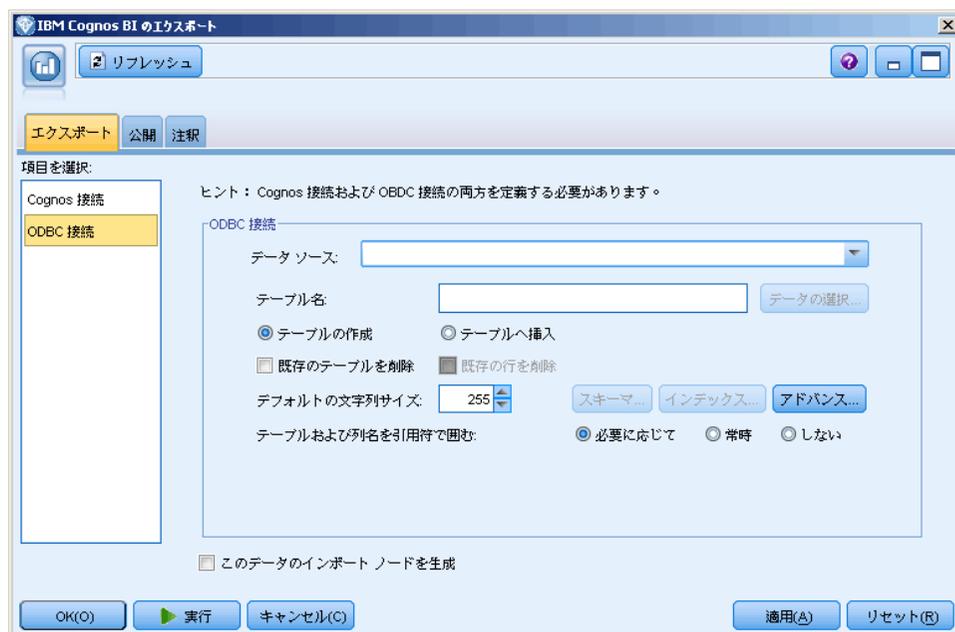
データのインポート ノードを生成: 指定したデータ ソースとテーブルにエクスポートしたように、入力ノードを生成する場合に選択します。 **[実行]** をクリックすると、ストリーム領域にこのノードが追加されます。

ODBC 接続

ここで、ストリーム データをエクスポートする Cognos データ サーバー (データベース) への接続を指定します。

注: ここで指定するデータ ソースが、 **[Cognos 接続]** パネルで指定したものと同一データソースを示さなければなりません。また、Cognos 接続データソースが ODBC データソースと同じログインを使用していることを確認する必要があります。

図 7-10
Cognos データのエクスポート



データ ソース： 選択したデータ ソースが表示されます。名前を入力するか、ドロップダウン リストから選択します。リスト中に目的のデータベースが見つからない場合は、[新規データベース接続の追加] を選択して、[データベース接続] ダイアログ ボックスから目的のデータベースを検索してください。詳細は、2 章 p. 20 データベース接続の追加 を参照してください。

テーブル名： データの送信先のテーブル名を入力します。[テーブルへ挿入] オプションを選択した場合、[選択] ボタンをクリックして、データベース中の既存のテーブルを選択することができます。

テーブルの作成： 新規データベース テーブルを作成する場合、または既存のデータベース テーブルを上書きする場合に選択します。

テーブルへ挿入： 既存のデータベース テーブルにデータを新しい行として挿入する場合に選択します。

テーブルを結合： (可能な場合) 選択したデータベースの列を、該当する入力データフィールドの値で更新します。このオプションを選択すると [結合] ボタンが有効となり、入力データフィールドをデータベース列に関連付けできるダイアログが表示されます。

既存のテーブルを削除： 新しいテーブルの作成時に、同じ名前を持つ既存のテーブルを削除する場合に選択します。

既存の行を削除： テーブルへの挿入時に、エクスポート前に既存の行をテーブルから削除する場合に選択します。

注： 上記の 2 種類のオプションのどちらも選択されていない場合、ノードの実行時に「上書き警告」というメッセージが表示されます。この警告メッセージを表示しない場合は、[ユーザー オプション] ダイアログ ボックスの [通知] タブにある、[ノードがデータベーステーブルを上書きする時に警告] を選択します。詳細は、[12 章 通知オプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド](#) を参照してください。

デフォルトの文字列サイズ： 上流にあるデータ型ノードでデータ型不明とされたフィールドは、データベースに文字列フィールドとして書き込まれます。ここには、データ型不明フィールドに使う文字列のサイズを指定します。

[スキーマ] をクリックしてダイアログ ボックスを開きます。ここでさまざまなエクスポート オプション（この機能をサポートするデータベース）を設定、SQL データ型をフィールドに設定し、データベース インデックス用の第 1 入力フィールド キーを指定します。詳細は、[p. 507 データベースエクスポートのスキーマ オプション](#) を参照してください。

[インデックス] をクリックして、エクスポートされたテーブルのインデックスを生成するオプションを指定してデータベースのパフォーマンスを高めます。詳細は、[p. 511 データベース エクスポートのインデックス オプション](#) を参照してください。

バルク ロードおよびデータベースのコミットに関するオプションを指定するには、[詳細] をクリックします。詳細は、[p. 515 データベース エクスポートの詳細オプション](#) を参照してください。

テーブルおよび列名を引用符で囲む： データベースに CREATE TABLE 文を送信する場合に使用するオプションを選択します。スペースまたは非標準文字を含むテーブルや列名は引用符で囲む必要があります。

- **必要に応じて：** 引用符が必要かどうかを IBM® SPSS® Modeler が個別に判断して、自動的に引用符で囲む場合に選択します。
- **常時：** テーブル名と列名を常に引用符で囲む場合に選択します。
- **しない：** 引用符を使用しない場合に選択します。

データのインポートノードを生成： 指定したデータ ソースとテーブルにエクスポートしたように、入力ノードを生成する場合に選択します。[実行] をクリックすると、ストリーム領域にこのノードが追加されます。

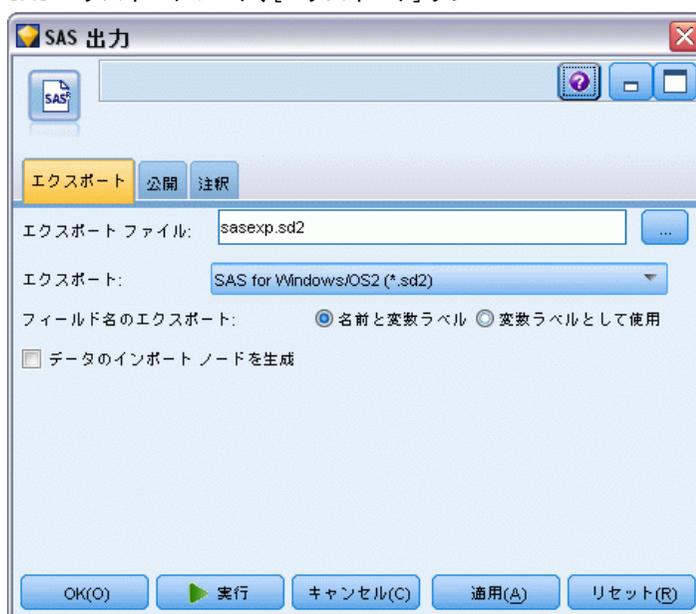
SAS エクスポート ノード

注： この機能は、SPSS Modeler Professional および SPSS Modeler Premium で使用できます。

SAS エクスポート ノードを使用すれば、データを SAS 形式で書き込んで、SAS または SAS 互換ソフトウェア パッケージで読み込むことができます。次の 3 種類の SAS ファイル形式でエクスポートできます。SAS for Windows/OS2、SAS for UNIX、または SAS バージョン 7/8。

SAS エクスポート ノード、[エクスポート] タブ

図 7-11
SAS エクスポート ノード、[エクスポート] タブ



エクスポート ファイル： ファイルの名前を指定します。ファイル名を入力するか、またはファイル選択ボタン をクリックしてファイルの場所を指定します。

エクスポート： エクスポートするファイル フォーマットを指定します。[SAS for Windows/OS2]、[SAS for UNIX] または、[SAS バージョン 7/8] を指定することができます。

フィールド名のエクスポート： SAS で使用するために、IBM® SPSS® Modeler からフィールド名とラベルをエクスポートするオプションを選択します。

- **名前と変数ラベル** :SPSS Modeler のフィールド名とフィールドラベルをエクスポートしてください。名前は SAS の変数名としてエクスポートされ、ラベルは SAS の変数ラベルとしてエクスポートされます。
- **変数ラベルとして使用** :SPSS Modeler フィールド名を SAS の変数ラベルとして使用します。SPSS Modeler は SAS 変数名では無効であるフィールド名の文字を使用できます。SAS で無効な名前が作成されることを防止するには、代わりに [名前と変数ラベル] を選択します。

データのインポートノードを生成： エクスポートされたデータ ファイルを読み込む SAS 入力ノードを、自動的に生成する場合に選択します。 [詳細は、2 章 p. 56 SAS 入力ノード を参照してください。](#)

Excel エクスポート ノード

Excel エクスポート ノードは、Microsoft Excel 形式 (.xls) でデータを出力します。オプションで、ノードが実行されるときに自動的に Excel が起動し、エクスポートするファイルを開けるように選択できます。

Excel ノードの [エクスポート] タブ

図 7-12
Excel エクスポート ノード、[エクスポート] タブ



ファイル名： ファイル名を入力するか、またはファイル選択ボタン をクリックしてファイルの場所を指定します。デフォルトのファイル名は excelexp.xls です。

ファイルの種類： エクスポートする Excel ファイルを選択します。

ファイルの新規作成： 新しい Excel ファイルを作成します。

既存ファイルに挿入 [セルで開始] フィールドで指定されたセル以降の内容が置き換えられます。スプレッドシートの他のセルは、元の内容が残されます。

フィールド名を含める：フィールド名がワークシートの最初の行に表示されるかどうかを指定します。

セルの開始点：最初のエクスポート レコードに使用されるセルの場所（または、[フィールド名を含める] がオンの場合、最初のフィールド名）。データは最初のセルの右側から下に向かって入力されます。

ワークシートを選択：データをエクスポートするワークシートを指定します。インデックスまたは名前のどちらかで、ワークシートを指定します。

■ **インデックスによるグループ**：ファイルを新規作成する場合、エクスポートするワークシートを示す 0 ～ 9 の値を指定します。最初のワークシートは 0、2 番目のワークシートは 1 というように指定します。ワークシートがすでにこの位置にある場合にのみ、10 以上の値を使用できます。

■ **名前順**：新しいファイルを作成している場合、ワークシートに使用される名前を指定します。既存のファイルに挿入している場合、ワークシートがあればデータはそのワークシートに挿入され、ない場合はこの名前を持つ新しいワークシートが作成されます。

Excel を起動する：ノードが実行されるときに、Excel が自動的に起動され、エクスポート ファイルが開くようにするかどうかを指定します。IBM® SPSS® Modeler Server に対して分散モードで実行中の場合は、この出力はサーバーのファイル システムに保存され、クライアント上で Excel はエクスポートされたファイルのコピーで起動されることに注意してください。

データのインポート ノードを生成：エクスポートされたデータ ファイルを読み込む Excel 入力 ノードを、自動的に生成する場合に選択します。 [詳細は、2 章 p.58 Excel 入力ノード を参照してください。](#)

XML エクスポート ノード

XML エクスポート ノードを使用して、UTF-8 エンコードを使用し、データを XML 形式で出力できます。オプションで、エクスポートしたデータをストリームに読み込む XML 入力ノードを作成できます。

図 7-13
XML データのエクスポート



XML エクスポート ファイル： データをエクスポートする XML ファイルの完全パスとファイル名です。

XML スキーマを使用： スキーマまたは DTD を使用して、エクスポートするデータの構造を制御します。制御することによって [マップ] ボタンが有効になります。

スキーマまたは DTD を使用しない場合、次のデフォルト構造がエクスポート データに使用されます。

```
<records>
  <record>
    <fieldname1>value</fieldname1>
    <fieldname2>value</fieldname2>
    :
    <fieldnameN>value</fieldnameN>
  </record>
  <record>
    :
  </record>
  :
</records>
```

フィールド名のスペースは、アンダースコアに置き換えられます。たとえば、「My Field」は「<My_Field>」となります。

マップ：XML スキーマの使用を選択した場合、このボタンによって XML 構造のどの部分を使用して新しいレコードを開始するかを指定できるダイアログが表示されます。詳細は、[p. 539 XML マッピングのレコード オプション](#) を参照してください。

マップしたフィールド：マップされたフィールド数を示します。

データのインポートノードを生成：エクスポートされたデータ ファイルをストリームに読み込む XML 入力ノードを、自動的に生成する場合に選択します。詳細は、[2 章 p. 60 XML 入力ノード](#) を参照してください。

XML データの作成

XML 要素が指定されると、フィールド値が要素タグ内に入力されます。

```
<element>value</element>
```

属性がマップされると、フィールド値が属性の値として指定されます。

```
<element attribute="value">
```

フィールドが `<records>` 要素の上の要素にマップされる場合、フィールドが一度だけ記述され、すべてのレコードの定数となります。この要素の値は最初のレコードに由来します。

Null 値が書き込む場合、空白の内容を指定します。要素の場合は、次のようになります。

```
<element></element>
```

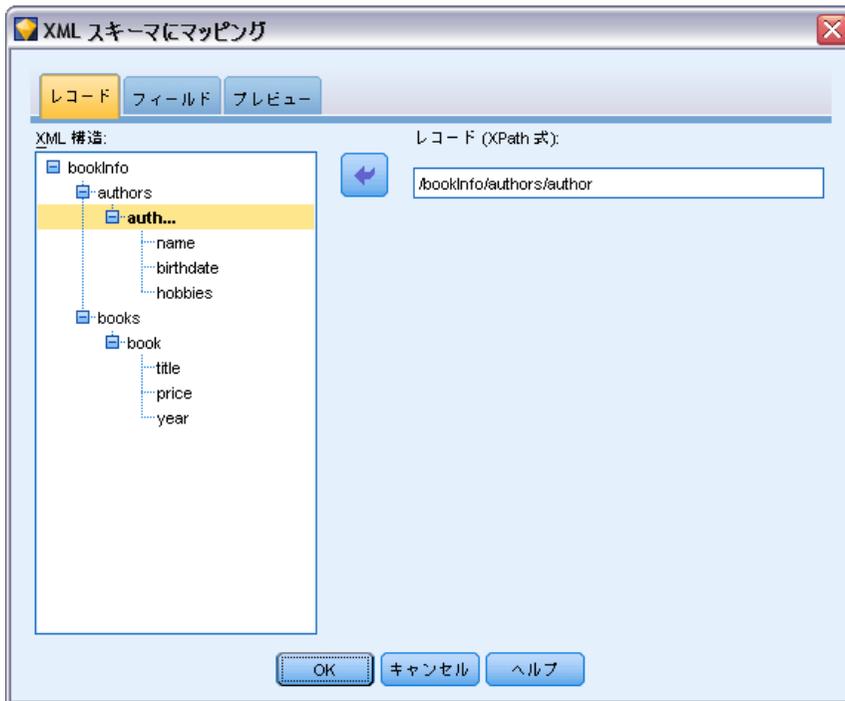
属性の場合は、次のようになります。

```
<element attribute="">
```

XML マッピングのレコード オプション

[レコード] タブを使用して、各新規レコードの開始に使用する XML 構造の部分を指定できます。スキーマに正しくマッピングするために、レコード区切り文字を指定する必要があります。

図 7-14
XML マッピング レコード



XML 構造: 前の画面で指定された XML スキーマの構造を示す階層ツリー。

レコード (XPath 式): レコード区切り文字を設定するには、XML 構造の要素を選択し、右方向矢印ボタンをクリックします。入力データにこの要素が出現するごとに、新しいレコードが出力ファイルに作成されます。

注：XML 構造のルート要素を選択すると、レコードを 1 つだけ書き込むことができ、他のすべてのレコードがスキップされます。

XML マッピングのフィールド オプション

スキーマ ファイルを使用する場合、[フィールド] タブを使用して、データセットのフィールドを XML 構造の要素または属性にマッピングします。

要素名または属性名と一致するフィールド名は、要素名または属性名が一意である限り、自動的にマッピングされます。そのため、要素および属性の名前がいずれも **field1** である場合、自動マッピングは行われません。**field1** という名前の項目が構造に 1 つだけある場合、ストリーム内に同じ名前を持つフィールドが自動的にマッピングされます。

図 7-15
XML マッピング フィールド



フィールド: モデル内にあるフィールドのリスト。1 つまたは複数のフィールドをマッピングのソース部分として選択します。リストの一番下のボタンを使用してすべてのフィールドを選択したり、特定の測定レベルであるすべてのフィールド選択できます。

XML 構造: XML 構造の要素をマップ対象として選択します。マッピングを作成するには、[マップ] をクリックします。その後、マッピングが表示されます。この方法でマッピングされたフィールド数は、このリストの下に表示されます。

マッピングを削除するには、XML 構造リストの項目を選択し、[マップ解除] をクリックします。

属性の表示: XML 構造の XML 要素の属性があれば、表示または非表示にします。

XML マッピングのプレビュー

[プレビュー] タブで、[更新] をクリックすると、書き込まれる XML のプレビューが表示されます。

マッピングが不適切である場合、[レコード] タブまたは [フィールド] タブに戻ってエラーを修正し、[更新] をもう一度クリックして結果を表示します。

IBM SPSS Statistics ノード

IBM SPSS Statistics ノードの概要

IBM® SPSS® Modeler およびそのデータ マイニング機能を補完するために、IBM® SPSS® Statistics では、詳細な統計分析とデータ管理を実行する機能を提供します。

互換性があり、ライセンスされた SPSS Statistics がインストールされている場合、SPSS Modeler から接続し、SPSS Modeler でサポートされていない複雑で、た段階に及ぶデータ操作および分析を実行できます。高度なユーザーの場合、コマンド シンタックスを使用して分析を詳細に変更できるオプションもあります。バージョンの互換性に関する詳細については、リリース ノートを参照してください。

可能な場合、ノード パレットの指定された部分に SPSS Statistics ノードが表示されます。

注：SPSS Statisticsの変換ノード、モデル ノードまたは出力ノードを使用する前に、データ型ノードでデータをインスタンス化することをお勧めします。また、これは AUTORECODE シンタックス コマンドを使用する場合に必要です。

SPSS Statistics パレットには次のノードがあります。



Statistics ファイル ノードは、同じフォーマットを使用する SPSS Statistics で使用される .sav ファイル形式のデータおよび SPSS Modeler に保存されたキャッシュ ファイルを読み込みます。詳細は、[p.544 Statistics ファイル ノード](#) を参照してください。



Statistics 変換ノードは、SPSS Modeler のデータ ソースに対する SPSS Statistics シンタックス コマンドの選択を行います。このノードは、ライセンスが与えられた SPSS Statistics のコピーが必要です。詳細は、[p.547 Statistics 変換ノード](#) を参照してください。



Statistics モデル ノードを使用すると、PMML を作成する SPSS Statistics プロシージャを実行してデータを分析および使用することができます。このノードは、ライセンスが与えられた SPSS Statistics のコピーが必要です。詳細は、[p.552 Statistics モデル ノード](#) を参照してください。



Statistics 出力ノードを使用すると、SPSS Statistics プロシージャを呼び出し、SPSS Modeler データを分析することができます。さまざまな SPSS Statistics 分析プロシージャにアクセスできます。このノードは、ライセンスが与えられた SPSS Statistics のコピーが必要です。詳細は、[p.555 Statistics 出力ノード](#) を参照してください。



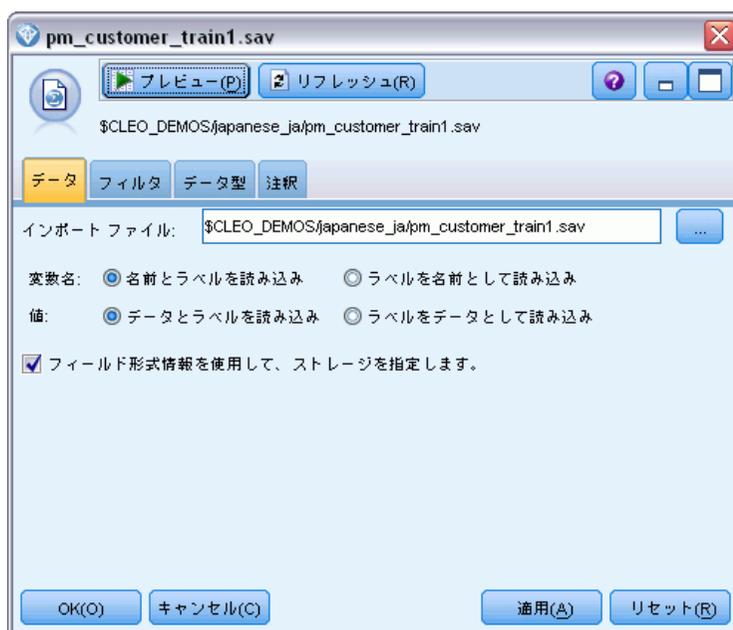
Statistics エクスポート ノードでは、SPSS Statistics.sav フォーマットでデータを出力します。.sav ファイルは、SPSS Statistics Base およびその他の製品で読み込むことができます。このフォーマットは、SPSS Modeler のキャッシュ ファイルでも使用されます。詳細は、[p.561 Statistics エクスポート ノード](#) を参照してください。

注：SPSS Statistics のライセンスがシングル ユーザーのみで、複数の枝でストリームを実行し、それぞれの枝に SPSS Statistics ノードがある場合、ライセンス エラーが発生する場合があります。このエラーは、一方の枝の SPSS Statistics セッションがもう一方のセッションを開始しようとする前に終了する場合に発生します。可能な場合、SPSS Statistics ノードを持つ複数の枝が並行して一考されないようストリームを再度設計する必要があります。

Statistics ファイル ノード

Statistics ファイル ノードを使用すると、保存された IBM® SPSS® Statistics ファイル (.sav) からデータを直接読み込むことができます。このフォーマットは、旧バージョンの IBM® SPSS® Modeler からキャッシュ ファイルを置換するために使用されます。保存されているキャッシュ ファイルをインポートする場合は、SPSS Statistics ファイル ノードを使用します。

図 8-1
.sav ファイルのインポート



インポート ファイル： ファイルの名前を指定します。ファイル名を入力するか、[...] ボタンをクリックしてファイルを選択します。ファイルを選択すると、ファイル パスが表示されます。

変数名 :SPSS Statistics.sav ファイルからインポートするときの、変数名とラベルの処理方法を選択します。ここで含めることを選択したメタデータは、SPSS Modeler 内での作業を通じて維持され、SPSS Statistics 内での使用のために再びエクスポートされる可能性があります。

- **名前とラベルを読み込む：** 変数名とラベルの両方を SPSS Modeler に読み込むために選択します。デフォルトでは、このオプションが選択され、変数名がデータ型ノードに表示されます。ラベルは、[ストリームのプロパティ] ダイアログ ボックスで指定したオプションに応じて、グラフやモデル ブラウザ、その他のタイプの出力中に表示できます。デフォルトでは、出力中のラベル表示は無効になっています。 [詳細は、5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。](#)
- **ラベルを名前として読み込む：** 短いフィールド名ではなく、SPSS Statistics.sav ファイルの詳細な変数ラベルを使用する場合に選択します。このラベルは SPSS Modeler の変数名として使用します。

値 :SPSS Statistics.sav ファイルからインポートするときの、変数とラベルの処理方法を選択します。ここで含めることを選択したメタデータは、SPSS Modeler 内での作業を通じて維持され、SPSS Statistics 内での使用のために再びエクスポートされる可能性があります。

- **データとラベルを読み込む**： 実際の値と値ラベルの両方を SPSS Modeler に読み込むために選択します。デフォルトでは、このオプションが選択され、値自体がデータ型ノードに表示されます。値ラベルは、[ストリームのプロパティ] ダイアログ ボックスで指定したオプションに応じて、式ビルダー、グラフ、モデル ブラウザ、その他の種類の出力中に表示できます。 [詳細は、5 章 ストリームの一般的なオプションの設定 in IBM SPSS Modeler 15 ユーザー ガイド を参照してください。](#)
- **ラベルをデータとして読み込む**： 値を表すために使用される数値コードまたはシンボル値コードの代わりに、.sav ファイルの値ラベルを使用する場合に選択します。たとえば、男性および女性をそれぞれ 1 または 2 の値で表す性別フィールドを持つデータに対してこのオプションを選択すると、フィールドが文字列に変換され、男性と女性が実際の値でインポートされます。

このオプションを選択する前に、SPSS Statistics データ中の欠損値を検討しておくことが大切です。たとえば、数値フィールドで欠損値 (0=回答なし、-99=不明) にだけラベルを使用している場合、このオプションを選択すると値ラベル「回答なし」および「不明」だけがインポートされ、フィールドが文字列に変換されます。このような場合は、値自体をインポートして、データ型ノードに欠損値を設定する必要があります。

フィールド形式情報を使用して、ストレージを指定します： このボックスをオンにすると、.sav ファイル形式で整数のフィールド値 (SPSS Statistics の変数ビューで Fn.0 と指定されているフィールド) は、整数ストレージを使用してインポートされます。文字列を除くすべてのフィールド値は、実数としてインポートされます。

このボックスがオフ (デフォルト) の場合、.sav ファイル形式で整数かどうかに関係なく、文字列を除くすべてのフィールド値は実数としてインポートされます。

複数レスポンス設定： ストリームに定義された複数のレスポンス セットは、SPSS Statistics ファイルがエクスポートされると自動的に保存されます。[フィルタ] タブで、ノードの複数のレスポンス セットを表示および編集することができます。 [詳細は、4 章 p.177 複数レスポンス セット編集 を参照してください。](#)

Statistics 変換ノード

Statistics 変換ノードを使用すると、IBM® SPSS® Statistics コマンド シンタックスを使用してデータ変換を完了することができます。これにより、IBM® SPSS® Modeler がサポートしていない多くの変換を完了することができます。さらに単一ノードからの数多くのフィールド作成など、複雑で他段階におよぶ変換を自動化することができます。それは、さらなる分析のためにデータが SPSS Modeler に返されることを除いて Statistics 出力ノードに似ている一方、出力ノードではデータはグラフやテーブルなど要求された出力オブジェクトとして返されます。

このノードを使用するには、互換性のあるバージョンの SPSS Statistics をインストールし、ライセンス認証する必要があります。[詳細は、6 章 p.499 IBM SPSS Statistics ヘルパー アプリケーション を参照してください。](#) 互換性に関する詳細については、リリース ノートを参照してください。

必要な場合は、[フィルタ] タブを使用してフィールドをフィルタリングまたは名前を変更して、SPSS Statistics の命名規則に従うことができます。[詳細は、p.563 IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング を参照してください。](#)

シンタックスの参照： 特定の SPSS Statistics 手順の詳細は、SPSS Statistics ソフトウェアに付属する『SPSS Statistics Command Syntax Reference』を参照してください。[シンタックス] タブでガイドを表示するには、[シンタックス エディタ] オプションを選択し、[SPSS Statistics シンタックス ヘルプの起動] ボタンをクリックします。

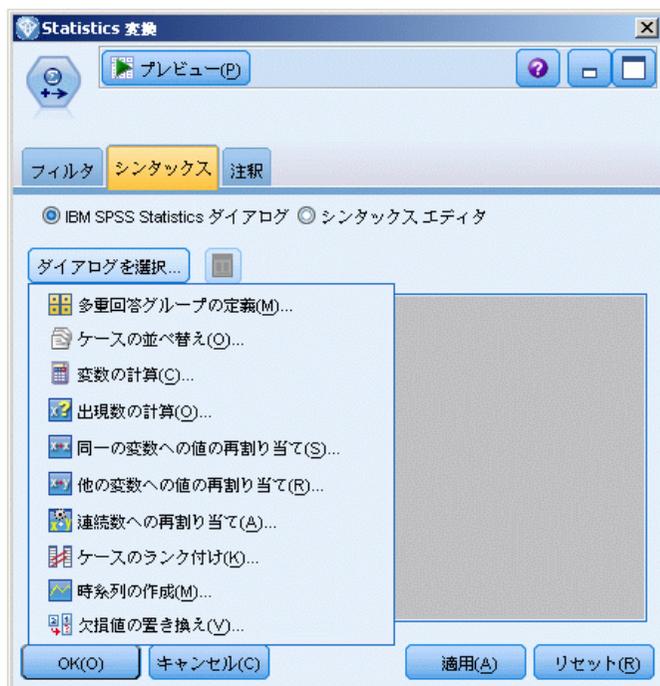
注： このノードではすべての SPSS Statistics シンタックスをサポートしているわけではありません。[詳細は、p.550 利用可能なシンタックス を参照してください。](#)

Statistics 変換ノードの [シンタックス] タブ

IBM SPSS Statistics ダイアログ オプション

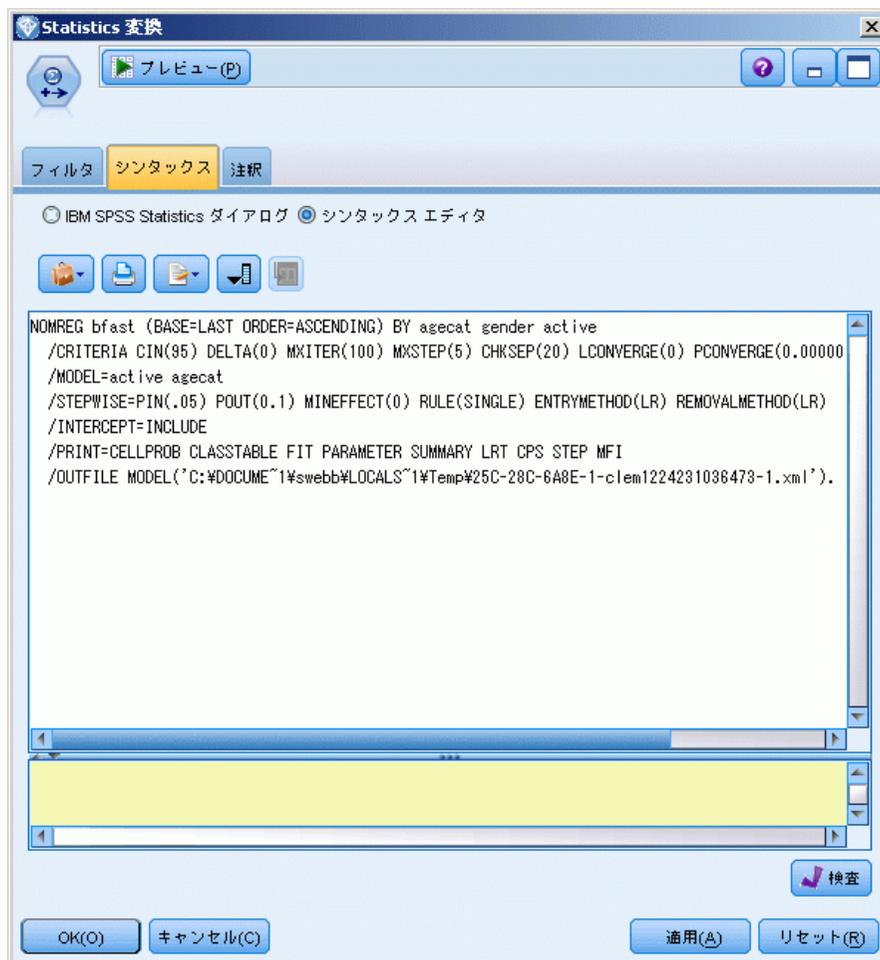
IBM® SPSS® Statistics シンタックスについてよく知らない場合、IBM® SPSS® Modeler でシンタックスを作成する最も簡単な方法は、[IBM SPSS Statistics ダイアログ] オプションを選択し、手順のダイアログ ボックスを選択、ダイアログボックスを入力して [OK] をクリックします。その後、SPSS Modeler で使用している SPSS Statistics ノードの [シンタックス] タブにシンタックスが配置されます。その後、ストリームを実行して手順より出力を取得することができます。

図 8-2
Statistics 変換 ノード、ダイアログの選択



IBM SPSS Statistics シンタックス エディタのオプション

図 8-3
Statistics 変換 ノード、シンタックス エディタ



検査：ダイアログ ボックスの上部に シンタックス コマンドを入力し、このボタンを使用して入力内容を有効にします。いかなる無効なシンタックスも、ダイアログ ボックスの下部で識別されます。

チェックのプロセスに時間がかからないようにするために、シンタックスを有効にする時、データ セット全体ではなくデータの代表的なサンプルをチェックして入力があることを確認します。

利用可能なシンタックス

IBM® SPSS® Statistics の古いシンタックスが数多くある場合や、SPSS Statistics のデータ準備機能を理解している場合、Statistics 変換ノードを利用して、多くの既存の変換を実行することができます。原則的に、ノードを使用すると、予測可能な方法でデータを変換することができます。たとえば、ループ コマンドの実行や、データの追加、ソート、フィルタリング、選択によって変換することができます。

実行可能なコマンドの例

■ 二系分布に従った乱数の計算

```
COMPUTE newvar = RV.BINOM(10000,0.1)
```

■ ある変数を新しい変数に再コード

```
RECODE Age (Lowest thru 30=1) (30 thru 50=2) (50 thru Highest=3) INTO AgeRecoded
```

■ 欠損値を置換

```
RMV Age_1=SMEAN(Age)
```

SPSS Statistics 変換ノードがサポートする Statistics シンタックスは、次のテーブルに一覧表示されています。

コマンド名

```
ADD VALUE LABELS  
APPLY DICTIONARY  
AUTORECODE  
BREAK  
CD  
CLEAR MODEL PROGRAMS  
CLEAR TIME PROGRAM  
CLEAR TRANSFORMATIONS  
COMPUTE  
COUNT  
CREATE  
DATE  
DEFINE-!ENDDEFINE  
DELETE VARIABLES  
DO IF  
DO REPEAT  
ELSE  
ELSE IF  
END CASE  
END FILE  
END IF  
END INPUT PROGRAM  
END LOOP
```

コマンド名

END REPEAT
EXECUTE
FILE HANDLE
FILE LABEL
FILE TYPE-END FILE TYPE
FILTER
FORMATS
IF
INCLUDE
INPUT PROGRAM-END INPUT PROGRAM
INSERT
LEAVE
LOOP-END LOOP
MATRIX-END MATRIX
MISSING VALUES
N OF CASES
NUMERIC
PERMISSIONS
PRESERVE
RANK
RECODE
RENAME VARIABLES
RESTORE
RMV
SAMPLE
SELECT IF
SET
SORT CASES
SORT CASES
STRING
SUBTITLE
TEMPORARY
TITLE
UPDATE
V2C
VALIDATEDATA
VALUE LABELS
VARIABLE ATTRIBUTE
VARSTOCASES
VECTOR

Statistics モデル ノード

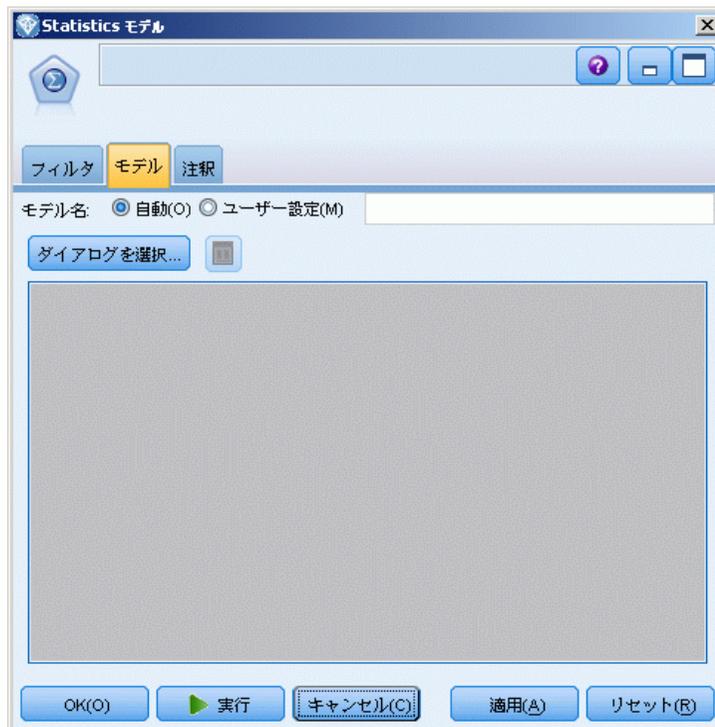
Statistics モデル ノードを使用すると、PMML を作成する IBM® SPSS® Statistics プロシージャを実行してデータを分析および使用することができます。作成したモデル ナゲットを IBM® SPSS® Modeler ストリーム内で通常の方法で使用し、スコアリングを行うことができます。

このノードを使用するには、互換性のあるバージョンの SPSS Statistics をインストールし、ライセンス認証する必要があります。詳細は、6 章 p.499 IBM SPSS Statistics ヘルパー アプリケーション を参照してください。互換性に関する詳細については、リリース ノートを参照してください。

利用可能な SPSS Statistics 分析プロシージャは、ライセンスの種類によって異なります。

Statistics モデル ノードの [モデル] タブ

図 8-4
Statistics モデル ノードの [モデル] タブ



モデル名: ターゲットまたは ID フィールド（その指定がない場合はモデル タイプ）に基づいてモデル名を生成、またはカスタム名を指定することができます。

ダイアログの選択： 選択し、実行できる IBM® SPSS® Statistics プロシージャのリストを表示します。PMML を生成し、ライセンス付与されるこれらのプロシージャのみが表示され、ユーザー指定のプロシージャは含まれません。

- ▶ 該当するプロシージャをクリックすると、関連する SPSS Statistics ダイアログが表示されます。
- ▶ [SPSS Statistics] ダイアログで、プロシージャの詳細を入力します。
- ▶ [OK] をクリックして Statistics モデル ノードに戻ります。SPSS Statistics シンタックスが [モデル] タブに表示されます。

図 8-5
[モデル] タブに表示されたシンタックス



- ▶ クエリーを変更する場合など、[SPSS Statistics] ダイアログにいつでも戻るには、プロシージャ選択ボタンの右側の [SPSS Statistics ダイアログ表示] ボタンをクリックします。

Statistics モデル ノード – モデル ナゲットの要約

Statistics モデル ノードを実行する場合、関連する IBM® SPSS® Statistics プロシージャを実行し、スコアリングするために IBM® SPSS® Modeler ストリームで使用できるモデル ナゲットを生成できます。

図 8-6
Statistics モデル ナゲット、[要約] タブ



モデル ナゲットの [要約] タブには、フィールド、構築の設定、およびモデル推定プロセスについての情報が表示されます。結果は、特定の項目をクリックすると開いたり閉じたりできるツリーで表示されます。詳細は、[3 章 モデル ナゲットの要約/情報 in IBM SPSS Modeler 15 モデル作成ノード](#) を参照してください。

[モデルの表示] をクリックすると、SPSS Statistics 出力ビューアの修正された形式で結果が表示されます。このビューアの詳細は、[の SPSS Statistics のマニュアル](#)を参照してください。

通常のエクスポート、および印刷オプションは、[ファイル] メニューから行うことができます。詳細は、6 章 p.436 出力を表示 を参照してください。

図 8-7
Statistics モデル ナゲット、[アドバンス] タブ

The screenshot shows the SPSS Statistics Output Viewer window. The left pane contains a tree view with the following items: 出力, Log, 名義回帰分析, 表題, 警告, 処理したケースの, **モデル適合情報**, 疑似 R2 乗, 尤度比検定, and パラメータ推定値. The main area displays the following tables:

モデル適合情報

| モデル | モデル当てはめ基準 -2 対数尤度 | 尤度比検定 | | |
|------|----------------------|---------|-----|------|
| | | カイ 2 乗 | 自由度 | 有意確率 |
| 切片のみ | 1.152E3 | | | |
| 最終 | 817.208 | 334.770 | 133 | .000 |

疑似 R2 乗

| | |
|-------------|------|
| Cox と Snell | .284 |
| Nagelkerke | .412 |
| McFadden | .285 |

尤度比検定

| 効果 | モデル当てはめ基準 縮小モデルの -2 対数尤度 | 尤度比検定 | | |
|----|-----------------------------|---------|-----|------|
| | | カイ 2 乗 | 自由度 | 有意確率 |
| 切片 | 817.208 | .000 | 0 | . |
| 地域 | 817.709 | .501 | 2 | .778 |
| 期間 | 1.024E3 | 206.452 | 71 | .000 |
| 年齢 | 905.353 | 88.145 | 59 | .008 |
| 婚姻 | 817.685 | .478 | 1 | .490 |

カイ 2 乗統計量は最終モデルと縮小モデルとの間の -2 対数尤度の差です。縮小モデルを作成するには、最終尤度の数値より縮小モデル推定値の尤度基準値を参照してください。

ピボットテーブルが表示されています H: 149, W: 341 pt.

Statistics 出力ノード

Statistics 出力ノードを使用すると、IBM® SPSS® Statistics プロシージャを呼び出し、IBM® SPSS® Modeler データを分析することができます。結果はブラウザ ウィンドウに表示したり、SPSS Statistics 出力ファイルフォーマットで保存することができます。SPSS Modeler からさまざまな SPSS Statistics 分析プロシージャにアクセスできます。

このノードを使用するには、互換性のあるバージョンの SPSS Statistics をインストールし、ライセンス認証する必要があります。詳細は、6 章 p.499 IBM SPSS Statistics ヘルパー アプリケーション を参照してください。互換性に関する詳細については、リリース ノート を参照してください。

必要な場合は、[フィルタ] タブを使用してフィールドをフィルタリングまたは名前を変更して、SPSS Statistics の命名規則に従うことができます。詳細は、[p. 563 IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング](#) を参照してください。

シンタックスの参照： 特定の SPSS Statistics 手順の詳細は、SPSS Statistics ソフトウェアに付属する『SPSS Statistics Command Syntax Reference』を参照してください。[シンタックス] タブでガイドを表示するには、[シンタックス エディタ] オプションを選択し、[SPSS Statistics シンタックス ヘルプの起動] ボタンをクリックします。

Statistics 出力ノードの [シンタックス] タブ

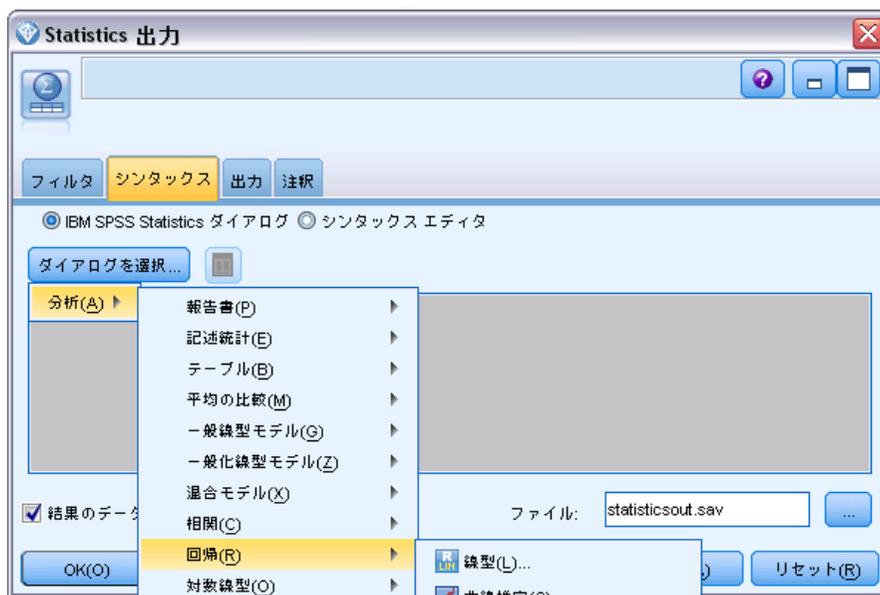
このタブを使用して、データの分析に使用する IBM® SPSS® Statistics プロシージャのシンタックスを作成します。シンタックスは、次の 2 つの部分で構成されています。文とそれに関連する**オプション**です。文には、実行する分析または操作と使用するフィールドを指定します。オプションには、表示する統計量、保存する生成フィールドなど、その他のすべてを指定します。

IBM SPSS Statistics ダイアログ オプション

SPSS Statistics シンタックスについてよく知らない場合、IBM® SPSS® Modeler でシンタックスを作成する最も簡単な方法は、[IBM SPSS Statistics ダイアログ] オプションを選択し、手順のダイアログ ボックスを選択、ダイアログボックスを入力して [OK] をクリックします。その後、SPSS Modeler で使用している SPSS Statistics ノードの [シンタックス] タブにシンタックスが配置されます。その後、ストリームを実行して手順より出力を取得することができます。

オプションで、Statistics ファイル入力ノードを生成し、データをインポートすることができます。これはたとえば、出力を表示するほか、スコアなどのフィールドをアクティブ データセットに書き込む場合に役立ちます。

図 8-8
Statistics 出力ノード、ダイアログの選択



シンタックスを作成する手順は、次のとおりです。

- ▶ [ダイアログを選択] ボタンをクリックします。
- ▶ 次のいずれかのオプションを選択してください。
 - **分析**: SPSS Statistics 分析メニューの内容が表示されます。使用する手順を選択してください。
 - **その他**: SPSS Statistics のカスタム ダイアログ ビルダーで作成されたダイアログ、[分析] メニューに表示されず、ライセンスが付与されているその他の SPSS Statistics ダイアログが表示されます。該当するダイアログがない場合、このオプションは表示されません。

注： [自動データ準備] ダイアログは表示されません。

新しいフィールドを作成する SPSS Statistics カスタム ダイアログがある場合、Statistics 出力ノードはターミナル ノードであるため、これらのフィールドを SPSS Modeler で使用することはできません。

- ▶ 任意で [生成データのインポートノードを生成] ボックスをオンにして、生成データを別のストリームにインポートできる Statistics ファイル入力ノードを作成します。このノードは画面の領域状に表示され、データは [ファイル] フィールドで指定された .sav ファイルに含まれます (デフォルトの場所は SPSS Modeler インストール ディレクトリです)。

シンタックス エディタのオプション

度数を使用するプロシージャに作成したシンタックスを保存する手順は、次のとおりです。

- ▶ ツールバーの最初のボタン [ファイル オプション] ボタンをクリックします。
- ▶ メニューから [保存] または [名前を付けて保存] を選択します。
- ▶ ファイルを .sps ファイルの形式で保存します。

以前作成したシンタックス ファイルを使用するには、シンタックス エディタの現在の内容を置き換えます。

- ▶ ツールバーの最初のボタン [ファイル オプション] ボタンをクリックします。
- ▶ メニューから [開く] を選択します。
- ▶ .sps ファイルを選択すると、その内容が出力ノードの [シンタックス] タブに貼り付けられます。

現在の内容を置き換えずに以前保存したシンタックスを挿入する手順は、次のとおりです。

- ▶ ツールバーの最初のボタン [ファイル オプション] ボタンをクリックします。
- ▶ メニューから [挿入] を選択します。
- ▶ .sps ファイルを選択すると、その内容が出力ノードのカーソルで指定されたポイントに貼り付けられます。

図 8-9
Statistics 出力ノード、シンタックス エディタ



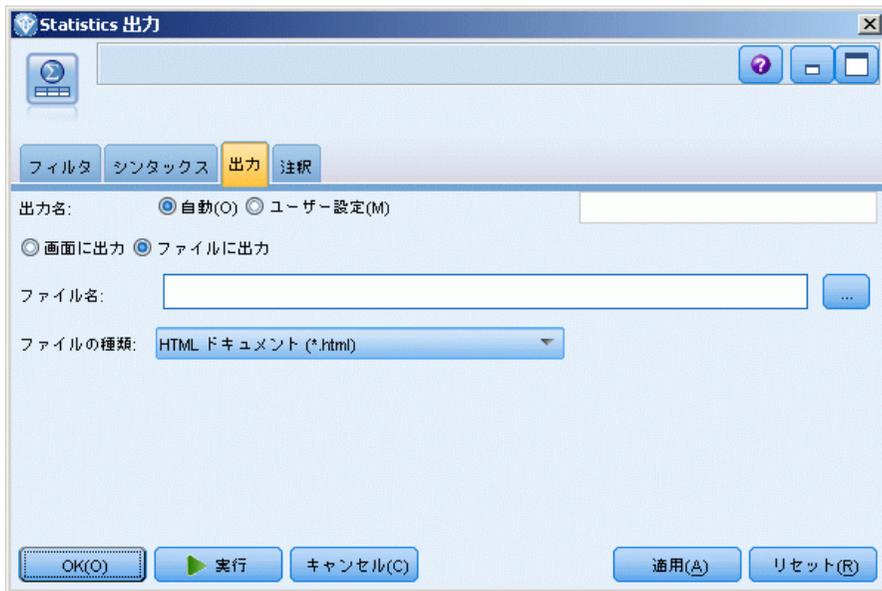
- ▶ 任意で [生成データのインポート ノードを生成] ボックスをオンにして、生成データを別のストリームにインポートできる Statistics ファイル入力ノードを作成します。このノードは画面の領域状に表示され、データは [ファイル] フィールドで指定された .sav ファイルに含まれます（デフォルトの場所は SPSS Modeler インストール ディレクトリです）。

[実行] をクリックすると、結果が SPSS Statistics 出力ビューアに表示されます。このビューアの詳細は、の SPSS Statistics のマニュアルを参照してください。

Statistics 出力ノードの [出力] タブ

[出力] タブを使用すると、出力フォーマットや位置を指定することができます。結果をスクリーンに表示、または結果を使用可能なファイル形式の 1 つに送ることができます。

図 8-10
Statistics 出力ノードの [出力] タブ



出力名： ノードの実行時に生成される出力の名前を指定します。[自動] は、出力を生成するノードの名前に基づいて名前を選択します。[ユーザー設定] で別の名前を指定することもできます。

画面に出力 (デフォルト) :オンラインで表示するための出力を作成します。出力ノードを実行すると、出力オブジェクトはマネージャ ウィンドウの [出力] タブに表示されます。

ファイルに出力： ノードの実行時に、出力をファイルに保存します。このオプションを選択した場合は、[ファイル名] フィールドにファイル名を入力して (または、ファイル選択 ボタンを使ってディレクトリを参照し、ファイル名を指定して)、ファイルフォーマットを選択してください。

ファイルの種類： 出力を送信するファイルの種類を選択します。

- **HTML ドキュメント (*.html)：** HTML 形式で出力を書き込みます。
- **SPSS Statistics ビューア ファイル (*.spv)：** 出力を IBM® SPSS® Statistics 出力ビューアで読み取れる形式で書き込みます。
- **SPSS Statistics Web レポート ファイル (*.spw)：** 出力を SPSS Statistics Web レポート形式で書き込みます。この形式は IBM SPSS Collaboration and Deployment Services リポジトリに公開し、Web ブラウザで表示できます。詳細は、6 章 p.437 [Web に公開] を参照してください。

Statistics エクスポート ノード

Statistics エクスポート ノードを使用すると、IBM® SPSS® Statistics.sav 形式でデータをエクスポートできます。SPSS Statistics.sav ファイルは、SPSS Statistics Base およびその他のモジュールで読み込むことができます。このフォーマットは、IBM® SPSS® Modeler キャッシュ ファイルでも使用されます。

SPSS Statistics の変数名は 最大 8 文字（半角）で、スペース、\$、- など、一部の文字を使用できないという制約があるため、SPSS Modeler フィールド名を SPSS Statistics の変数名にマップするとエラーが発生することがあります。この問題に対処するには、次の 2 通りの方法があります。

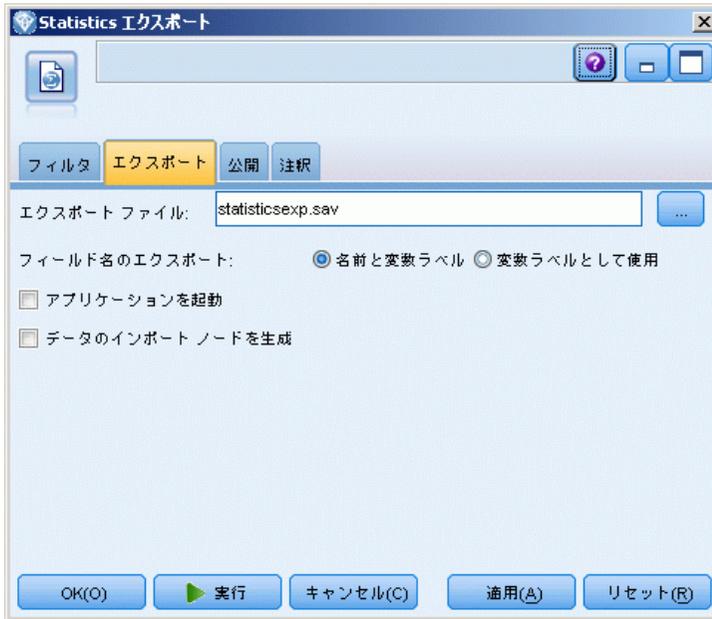
- [フィルタ] タブをクリックして、SPSS Statistics 変数名の要件に準拠したフィールド名に変更することができます。詳細は、[p.563 IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング](#) を参照してください。
- SPSS Modeler でフィールド名とラベルをエクスポートします。

注：SPSS Modeler は .sav ファイルを Unicode UTF-8 形式で書き込みます。リリース 16.0 以降の SPSS Statistics でサポートしているのは Unicode UTF-8 形式だけです。データの破損の可能性を回避するために、Unicode エンコードで保存した .sav ファイルはリリース 16.0 以前の SPSS Statistics で使用することはできません。詳細は、SPSS Statistics のヘルプを参照してください。

複数レスポンス設定： ストリームに定義された複数のレスポンス セットは、ファイルがエクスポートされると自動的に保存されます。[フィルタ] タブで、ノードの複数のレスポンス セットを表示および編集することができます。詳細は、[4 章 p.177 複数レスポンス セット編集](#) を参照してください。

Statistics エクスポート ノードの [エクスポート] タブ

図 8-11
Statistics エクスポート ノード、[エクスポート] タブ



エクスポート ファイル： ファイルの名前を指定します。ファイル名を入力するか、またはファイル選択ボタン をクリックしてファイルの場所を指定します。

フィールド名のエクスポート： IBM® SPSS® Modeler から IBM® SPSS® Statistics.sav ファイルへエクスポートする際の変数名とラベルの処理方法を選択します。

- **名前と変数ラベル** :SPSS Modeler のフィールド名とフィールドラベルをエクスポートしてください。名前は SPSS Statistics の変数名としてエクスポートされ、ラベルは SPSS Statistics の変数ラベルとしてエクスポートされます。
- **変数ラベルとして使用** :SPSS Modeler フィールド名を SPSS Statistics の変数ラベルとして使用します。SPSS Modeler は SPSS Statistics 変数名では無効であるフィールド名の文字を使用できます。無効な SPSS Statistics 名を作成しないように、[名前を変数ラベルとして利用する] を選択するか、フィールド名を調整するための [フィルタ] タブを利用します。

アプリケーションを起動 :SPSS Statistics がインストールされている場合、このオプションを選択することにより、保存したデータ ファイルに対してこれらのプログラムを直接起動できます。アプリケーションを起動するためのオプションは、[ヘルパー アプリケーション] ダイアログ ボックスで指

定する必要があります。詳細は、[6 章 p.499 IBM SPSS Statistics ヘルパー アプリケーション](#) を参照してください。外部プログラムで起動しないで、単に SPSS Statistics.sav ファイルを作成する場合は、このオプションの選択を解除してください。

データのインポート ノードを生成： エクスポートされたデータ ファイルを読み込む、Statistics ファイル ノードを自動的に生成する場合に選択します。詳細は、[p.544 Statistics ファイル ノード](#) を参照してください。

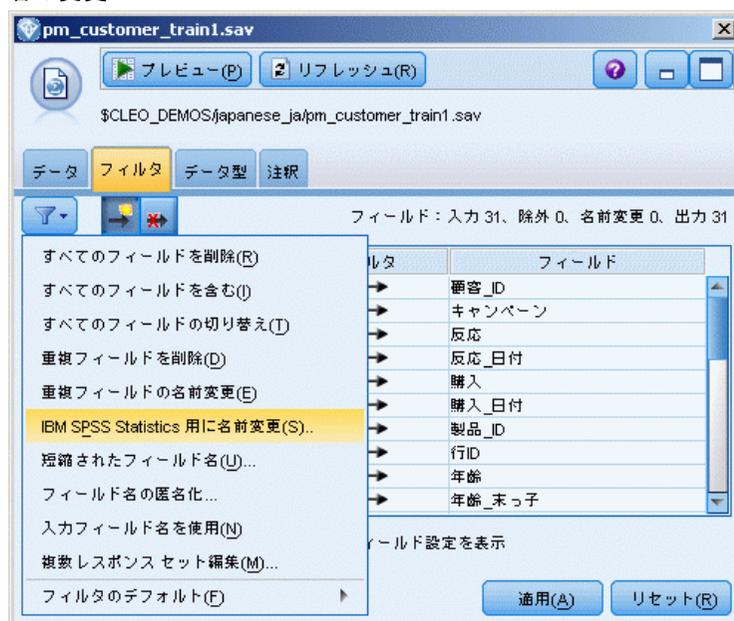
IBM SPSS Statistics 用のフィールドの名前変更またはフィルタリング

IBM® SPSS® Modeler から IBM® SPSS® Statistics などの外部アプリケーションにデータをエクスポートまたは展開する前に、フィールド名を変更したり、調節しなければならないこともあります。Statistics 変換、Statistics 出力、および Statistics エクスポート ダイアログ ボックスには、[フィルタ] タブが用意されており、ここからこの処理を簡単に行うことができます。

[フィルタ] タブの機能に関する基本的な説明は、他の場所で説明されています。詳細は、[4 章 p.174 フィルタリング オプションの設定](#) を参照してください。ここでは、SPSS Statistics にデータを読み込む際のヒントを説明していきます。

図 8-12

Statistics ファイルノードの [フィルタ] タブを使った、IBM SPSS Statistics 用のフィールド名の変更

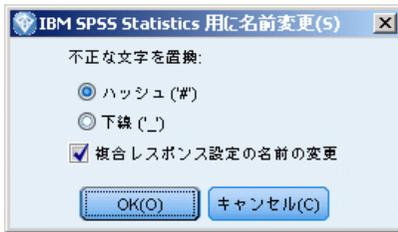


SPSS Statistics の命名規則に準ずるよう、ファイル名を調整する手順は次のとおりです。

- ▶ [フィルタ] タブで、[フィルタ オプション メニュー] ツールバー ボタン (ツールバーの最初のボタン) をクリックします。
- ▶ [SPSS Statistics 用に名前変更] を選択します。

図 8-13

フィールド名の変更



- ▶ [SPSS Statistics 用に名前変更] ダイアログで、ファイル名の不正な文字を、[ハッシュ (#)] 文字または [アンダースコア (_)] に置き換えるよう選択できます。

複数レスポンス セットの名前を変更 : Statistics ファイル入力ノードを使用して、SPSS Modeler にインポートできる複数レスポンス セットの名前を変更する場合、このオプションを選択します。調査の回答など、ケースごとに複数の値があるデータを記録するのに使用されます。

スーパーノード

スーパーノードの概要

IBM® SPSS® Modeler の視覚的なプログラミング インターフェイスは非常に簡単に習得できますが、その理由の 1 つとして、各ノードの機能が明確に定義されている点が挙げられます。しかし、複雑な処理を行うには、長い一連のノードが必要となることがあります。この結果、ストリーム領域が複雑になってストリーム ダイアグラムの追跡が困難になることがあります。ストリームが長く複雑になるのを防ぐには、次の 2 つの方法があります。

- 処理シーケンスを複数のストリームに分割して、1 つのストリームが別のストリームに送られるようにします。たとえば、1 番目のストリームがデータ ファイルを作成し、それを 2 番目のストリームで入力ファイルとして使用します。次に 2 番目のストリームがデータ ファイルを作成し、3 番目のストリームがそれを入力ファイルとして使用します。これらの複数のスクリプトをプロジェクトに保存して管理することができます。プロジェクトは、複数のストリームや出力の編成手段を提供しています。しかし、プロジェクト ファイルには、それに含まれるオブジェクトの参照だけが格納されているため、依然として複数のストリーム ファイルを管理する必要があります。
- 複雑なストリームで作業を行う際のより効率的な方法として、スーパーノードを作成することができます。

スーパーノードが、データ ストリームのセクションをカプセル化し、複数のノードを 1 つのノードにグループ化します。この機能を使用することで、データ マイニングには多くの利点が生じます。

- ストリームが簡潔で管理しやすくなります。
- ノードを組み合わせて、そのビジネスに固有のスーパーノードを作成できます。
- 複数のデータ マイニング プロジェクトで再利用するため、スーパーノードをライブラリにエクスポートすることができます。

スーパーノードの種類

スーパーノードは、データ ストリームでは星形のアイコンで表されます。アイコンの影によって、スーパーノードの種類とそのストリームの入出力方向が示されます。

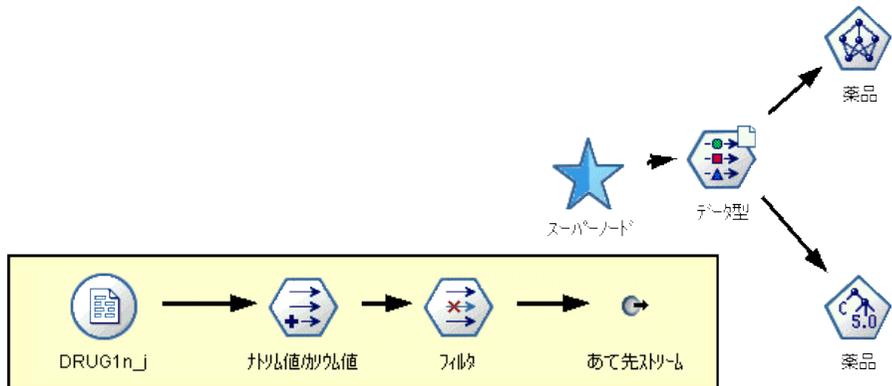
スーパーノードには、次の 3 種類があります。

- 入力スーパーノード
- プロセス スーパーノード
- ターミナル スーパーノード

入力スーパーノード

入力スーパーノードには、通常の入力ノードと同じようなデータ ソースが含まれており、通常の入力ノードを使用できる任意の場所で使用できます。入力スーパーノードのアイコンには左側に影が付けられています。これは、左側が「閉じられている」ことを表しており、データはスーパーノードから常に下流に流れることを示しています。

図 9-1
ストリームに課されたズームイン バージョンを持つ入力スーパーノード

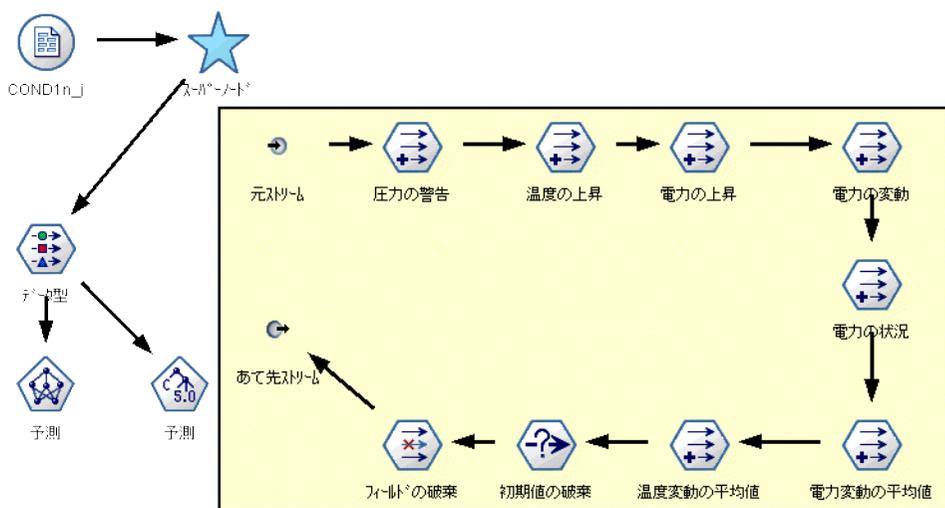


入力スーパーノードの右側には 1 つだけ接続があり、スーパーノードからデータが出力されストリームに流れていくことを表しています。

プロセス スーパーノード

プロセス スーパーノードにはプロセス ノードしか入れることはできません。また、アイコンに影は付けられていません。これは、データがこの種類のスーパーノードに入力されたり、操作スーパーノードから出力されたりすることを表しています。

図 9-2
ストリームに課されたズームイン バージョンを持つプロセス スーパーノード



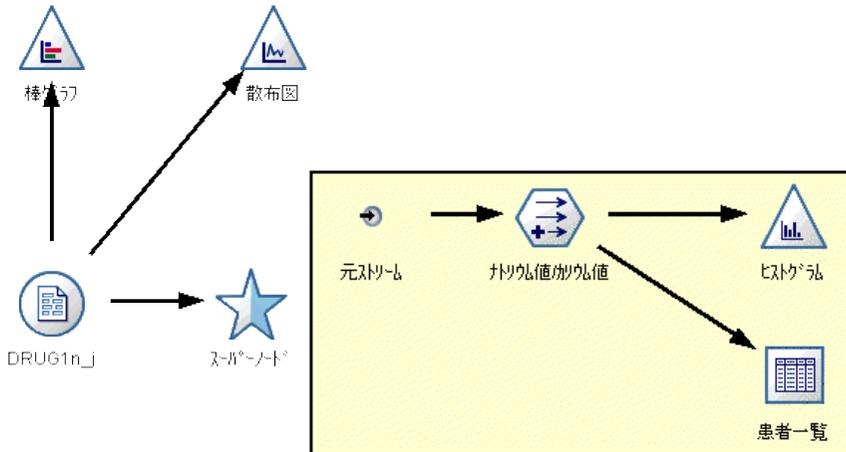
プロセス スーパーノードには、左側と右側の両方に接続点があり、データはスーパーノードに入力された後、出力されて下流のストリームに流れることを表しています。スーパーノードにはストリーム フラグメントを追加したり、余分なストリームを入れることもできますが、両方の接続点はインとアウトを接続する単一のパス上になければなりません。

注： プロセス スーパーノードは、「操作スーパーノード」と呼ばれることもあります。

ターミナル スーパーノード

ターミナル スーパーノードには、1 つ以上のターミナル ノード（散布図、テーブルなど）を入れることができ、ターミナル ノードと同じように使用することができます。ターミナル スーパーノードのアイコンには右側に影が付けられており、これは右側が「閉じられている」ことを示しています。つまり、データは常にターミナル スーパーノードに入力されることを示しています。

図 9-3
ストリームに課せられたズームイン バージョンを持つターミナル スーパーノード



ターミナル スーパーノードの左側には 1 つだけ接続点があり、データはストリームからスーパーノードに入力され、スーパーノード内で処理が終了することを表しています。

ターミナル スーパーノードには、スーパーノード内のターミナル ノードの実行順序を指定するためのスクリプトを入れることもできます。 [詳細は、p.587 スーパーノードとスクリプト](#) を参照してください。

スーパーノードの作成

スーパーノードを作成すると、単一のノード中に複数のノードがカプセル化されるため、データ ストリームが「縮小」されます。ストリーム領域上にストリームを作成またはロードしたら、さまざまな方法でスーパーノードを作成することができます。

複数選択

スーパーノードを作成するもっとも簡単な方法は、カプセル化するすべてのノードを選択することです。

- ▶ マウスを使ってストリーム領域上の複数のノードを選択します。Shift キーを押しながらクリックして、ストリームまたはストリームの一部を選択することもできます。注： 選択する複数のノードは連続したストリームまたは分岐したストリームでなければなりません。隣接しないまたは接続されていない複数のノードを選択することはできません。
- ▶ 次のいずれかの方法で、選択したノードをカプセル化します。
 - ツールバーのスーパーノード アイコン（星形）をクリックします。

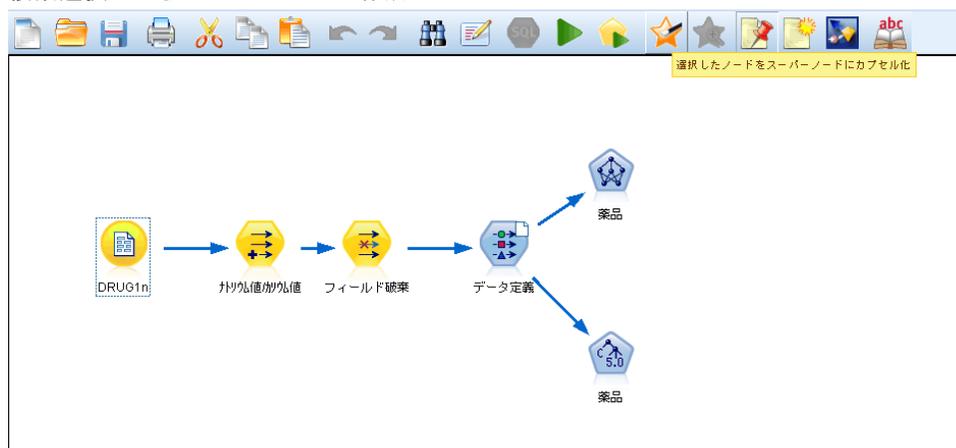
- スーパーノードを右クリックして表示されるコンテキストメニューから、次の各項目を選択します。

[スーパーノードの作成(U)] > 始点の選択

- [スーパーノード] メニューから次の各項目を選択します。

[スーパーノードの作成(U)] > 始点の選択

図 9-4
複数選択によるスーパーノードの作成



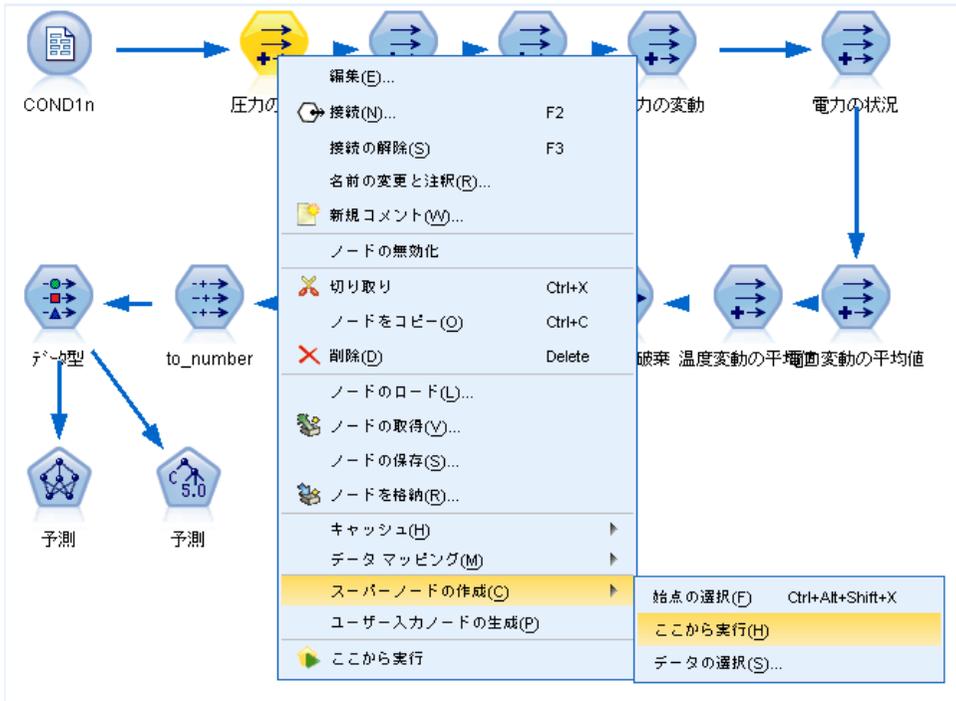
これらの 3 種類のオプションはすべて、ノードをスーパーノードにカプセル化して、その内容に応じた種類（入力、プロセス、またはターミナル）を表す影を付けます。

単一選択

単一のノードを選択して、メニュー オプションからスーパーノードの始点と終点を指定するか、または選択したノードの下流にあるすべてのノードをカプセル化して、スーパーノードを作成することもできます。

- ▶ カプセル化を開始する始点となるノードをクリックします。
- ▶ [スーパーノード] メニューから次の各項目を選択します。
[スーパーノードの作成(U)] > [ここから実行(H)]

図 9-5
単一選択でコンテキストメニューを使ったスーパーノードの作成



カプセル化するストリーム中のノードの始点と終点を選択して、より対話的にスーパーノードを作成することもできます。

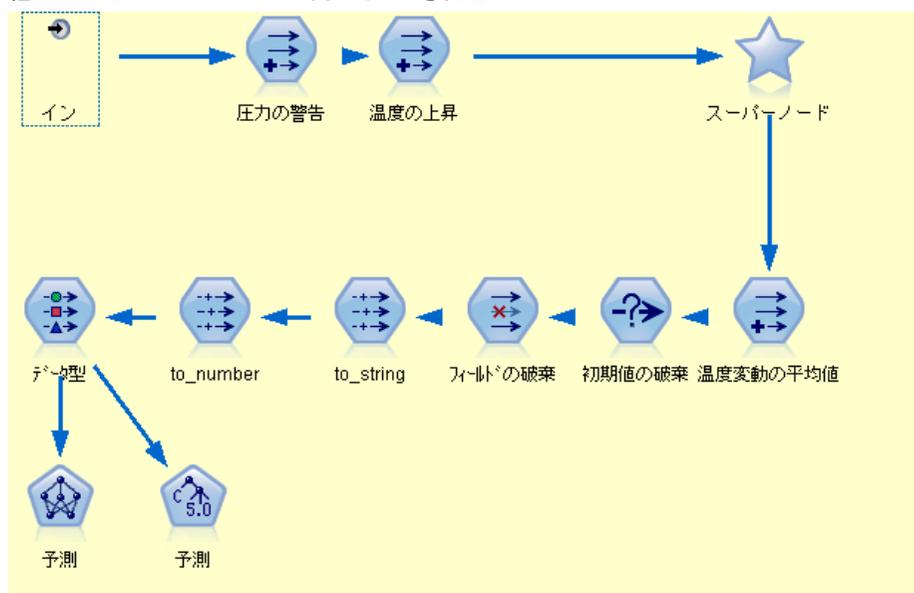
- ▶ スーパーノードに入れる最初または最後のノードをクリックします。
- ▶ [スーパーノード] メニューから次の各項目を選択します。
[スーパーノードの作成(U)] > 選択(S)...
- ▶ 代わりに、目的のノードを右クリックして表示されるコンテキストメニューを使用することもできます。
- ▶ カーソルがスーパーノード アイコンの形に変わります。これは、ストリーム中のもう一方のノードを選択する必要があることを示します。スーパーノード フラグメントの「もう一方の端」（上流または下流）に移動して、適切なノードをクリックします。選択した 2 つのノード間にある、すべてのノードがスーパーノードの星形アイコンに置き換わります。

注： 選択する複数のノードは連続したストリームまたは分岐したストリームでなければなりません。隣接しないまたは接続されていない複数のノードを選択することはできません。

スーパーノードのネスト

スーパーノードを他のスーパーノードに入れる（ネストする）ことができます。ネストされたスーパーノードには、それぞれの種類のスーパーノード（入力、プロセス、およびターミナル）と同じ規則が適用されます。たとえば、ネ스팅しているプロセス スーパーノード中のすべてのスーパーノードを連続してデータが流れていないと、そのスーパーノードはプロセス スーパーノードを維持できません。ネストされているスーパーノードのいずれかがターミナル スーパーノードだった場合、データがその階層中を連続して流れることができません。

図 9-6
他のプロセス スーパーノード内にネストされたプロセス スーパーノード



ターミナル スーパーノードと入力スーパーノードは、他の種類のスーパーノードをネストすることができます。ただし、同じようにスーパーノードを作成する際の基本的な規則が適用されます。

有効なスーパーノードの例

IBM® SPSS® Modeler で作成したものの大部分は、スーパーノード中にカプセル化することができます。有効なスーパーノードの例を次に示します。

図 9-7
有効なストリーム フローに 2 つの接続がある有効なプロセス スーパーノード

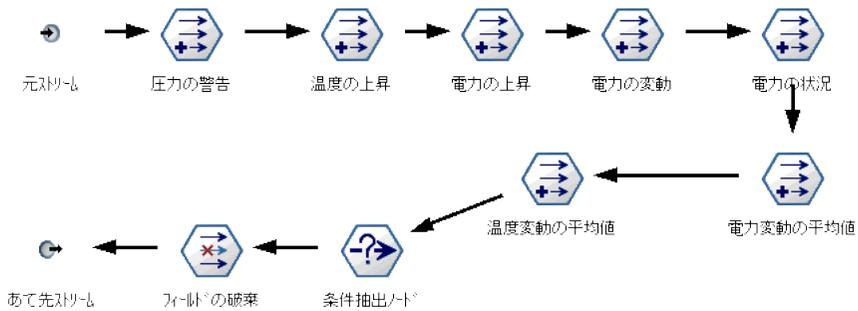


図 9-8
生成されたモデルをテストするために用いる個別のストリームを含む、有効なターミナル スーパーノード

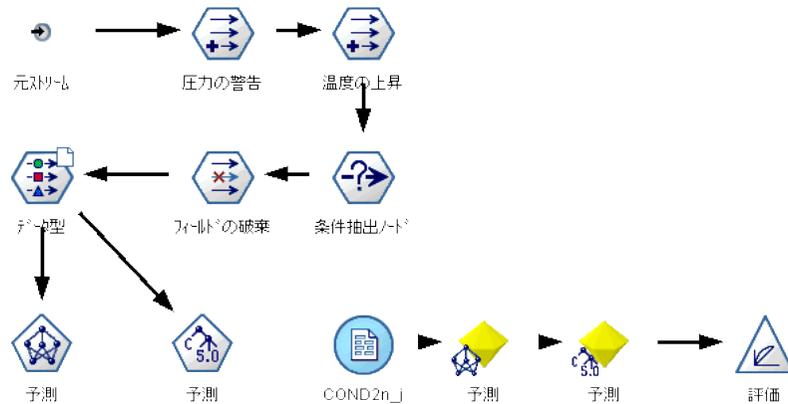
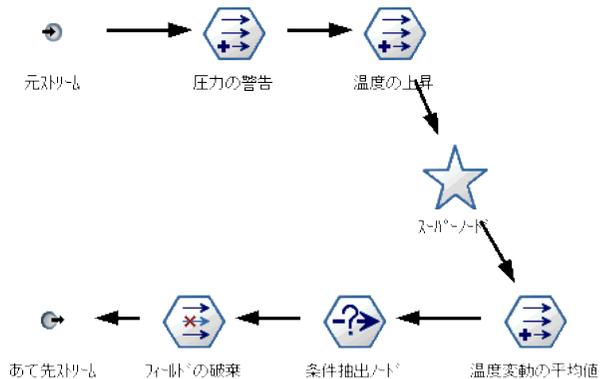


図 9-9
ネストされたスーパーノードを含む有効なプロセス スーパーノード



無効なスーパーノードの例

有効なスーパーノードを作成するためにもっとも重要なことは、スーパーノード接続をデータが線状に流れていくようにすることです。2つの接続がある場合（プロセススーパーノード）、データは入力接続から始まり、出力接続まで流れていかなければなりません。同じように、入力スーパーノードでは、入力ノードから外側のストリームにデータを戻す、データの出力点まで流れていかなければなりません。

図 9-10

無効な入力スーパーノード：入力ノードがデータフローパスに接続されていない

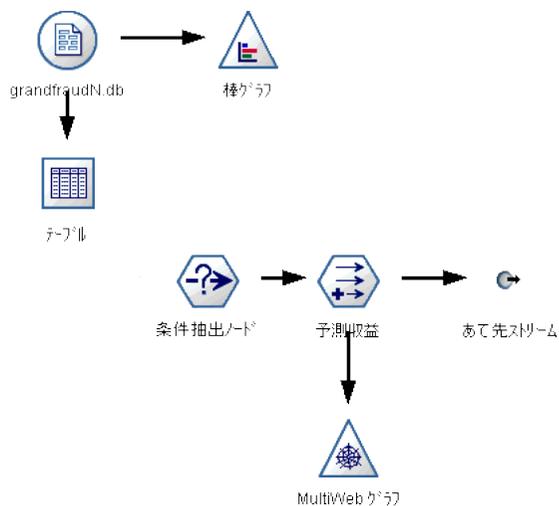
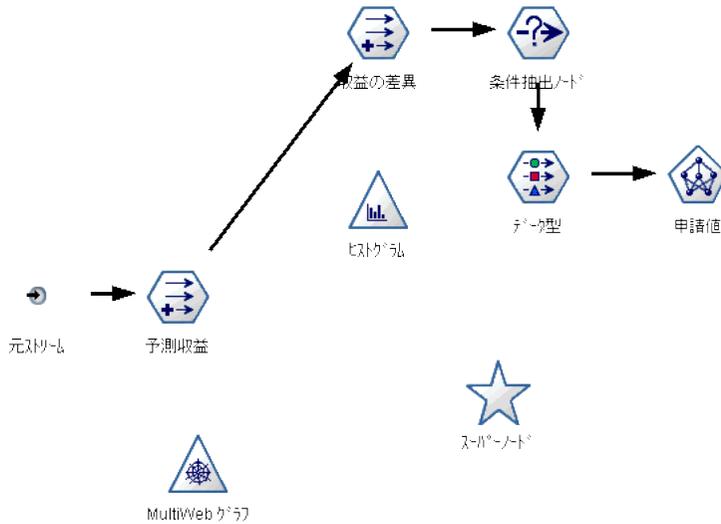


図 9-11
無効なターミナル スーパーノード : ネストされたスーパーノードがデータ フロー パスに接続されていない



スーパーノードのロック

スーパーノードを作成すると、パスワードを使用してスーパーノードをロックし、修正されるのを防ぎます。たとえば、IBM® SPSS® Modeler の質問を設定する経験があまりない組織内のユーザーが使用できるよう、固定値のテンプレートとしてストリームまたはストリームの一部を作成する場合に、スーパーノードを作成します。

スーパーノードがロックされている場合も、定義されたパラメータの [パラメータ] タブに値を入力したり、パスワードを入力せずにロックされたスーパーノードを実行することができます。

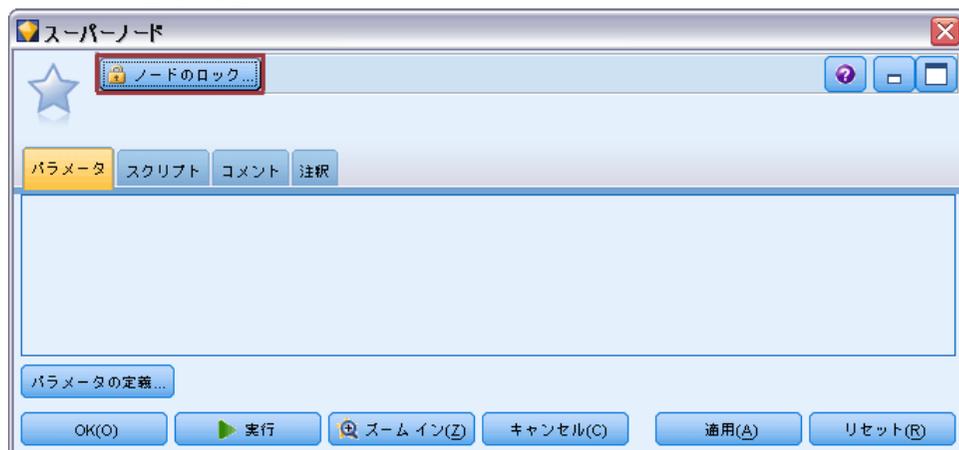
注： ロックおよびロックの解除は、スクリプトを使用して実行することはできません。

スーパーノードのロックとロック解除

警告 : 消失したパスワードを復元することはできません。

3 つのタブのうちどのタブからもスーパーノードをロックまたはロック解除できます。

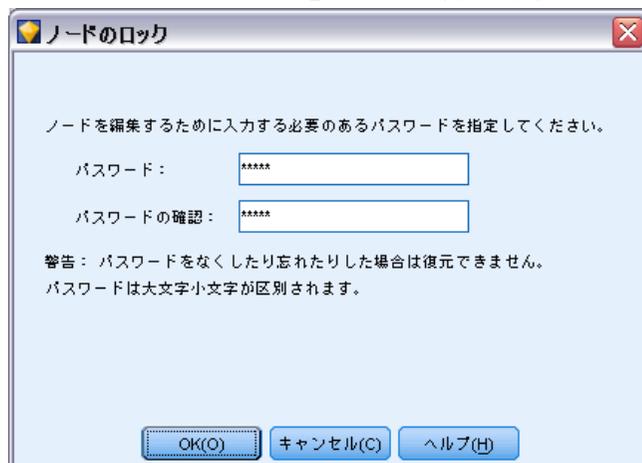
図 9-12
スーパーノードのロック



[ノードのロック] をクリックします。

パスワードを入力して確定します。

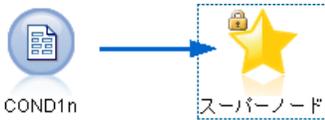
図 9-13
スーパーノードのパスワードを入力して確定します。



- ▶ [OK] をクリックします。

パスワード保護されたスーパーノードは、ストリーム領域のスーパーノードのアイコンの左上に小さい南京錠のシンボルで示されます。

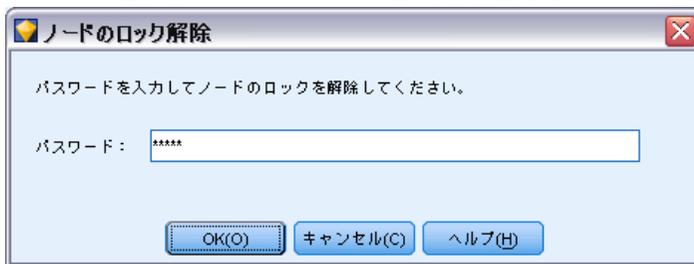
図 9-14
ストリームの一部としてのロックされた入カスーパーノード



スーパーノードのロック解除

- ▶ パスワード保護を永続的に解除するには、[ノードのロック解除] をクリックします。パスワードを要求するプロンプトが表示されます。

図 9-15
パスワードを入力してスーパーノードをロック解除する

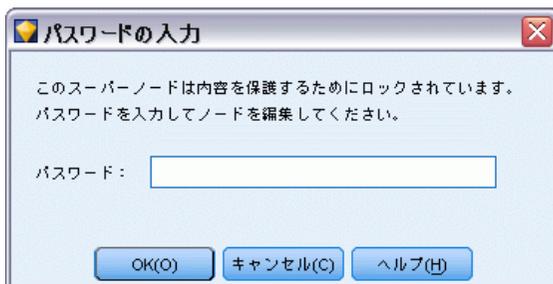


- ▶ パスワードを入力して [OK] をクリックします。スーパーノードはパスワード保護されず、ストリーム内のアイコンの隣に南京錠のシンボルは表示されなくなります。

ロックされたスーパーノードの編集

パラメータを定義するか、ズーム インしてロックされたスーパーノードを表示する場合、パスワード入力を要求するプロンプトが表示されます。

図 9-16
パスワードを入力してスーパーノードをズーム インまたは編集する



- ▶ パスワードを入力して [OK] をクリックします。

スーパーノードのあるストリームが終了するまで必要に応じて、パラメータ定義を編集したりズーム インやズーム アウトしたりできます。

これでパスワード保護を解除したわけではなく、スーパーノードを使用できるようにアクセスできるにすぎません。詳細は、[p. 574 スーパーノードのロックとロック解除](#) を参照してください。

スーパーノードの編集

スーパーノードを作成すると、ズーム インすることによってより近くでスーパーノードを検証することができます。スーパーノードがロックされている場合は、パスワードを要求するプロンプトが表示されます。詳細は、[p. 576 ロックされたスーパーノードの編集](#) を参照してください。

スーパーノードの内容を表示するには、IBM® SPSS® Modeler ツールバーのズーム イン アイコンを使用するか、または次の方法を利用します。

- ▶ スーパーノードを右クリックします。
- ▶ コンテキスト メニューから、[ズーム イン] を選択します。

選択したスーパーノードの内容が、ストリームまたはストリーム フラグメントを流れるデータを表すコネクタとともに、少し異なる SPSS Modeler 環境に表示されます。ここでは、さまざまな作業を行うことができます。

- スーパーノードの種類（入力、プロセス、またはターミナル）の変更。
- パラメータの作成またはパラメータの値の編集。パラメータはスクリプトや CLEM 式中使用されます。
- スーパーノードおよびそのサブ ノードのキャッシュ オプションの指定。
- スーパーノード スクリプトの作成と変更（ターミナル スーパーノードの場合）。

スーパーノードの種類の変更

状況によっては、スーパーノードの種類を変更する方が良いこともあります。このオプションは、スーパーノードにズーム インしている場合にだけ利用でき、このレベルのスーパーノードにだけ適用されます。スーパーノードには、次の 3 種類があります。

| | |
|---------------|----------------------------------|
| 入力スーパーノード | 1 つの出力接続があります。 |
| プロセス スーパーノード | 2 個の接続 : 入力接続と出力接続の 2 つの接続があります。 |
| ターミナル スーパーノード | 1 つの入力接続があります。 |

スーパーノードの種類を変更するには

- ▶ スーパーノードにズーム インしていることを確認してください。
- ▶ [スーパーノード] メニューの [スーパーノードの種類] を選択し、次に適切な種類を選択します。

スーパーノードの注釈付けと名前の変更

ストリームに表示されるスーパーノードの名前を変更したり、プロジェクトやレポートで使われる注釈を付けることができます。これらの作業を行うには、次の手順に従ってください。

- ▶ スーパーノードを右クリックして (ズーム アウトされる)、[名前の変更と注釈] を選択します。
- ▶ 代わりに、[スーパーノード] メニューの [名前の変更と注釈] を選択することもできます。このオプションは、ズーム インおよびズーム アウトの両方のモードで利用できます。

どちらの場合でも、[注釈] タブが選択された状態でダイアログ ボックスが表示されます。このタブのオプションを使って、ストリーム領域上に表示される名前を変更したり、スーパーノードの操作に関する説明や情報などを入力します。

図 9-17
スーパーノードの注釈の追加



スーパーノードによるコメントの使用

コメントされたノードまたはナゲットからスーパーノードを作成する場合、スーパーノードにコメントを表示したい場合にスーパーノードを作成する選択の中にコメントを含める必要があります。選択からコメントを削除する場合、スーパーノードが作成されるとコメントはストリーム上で接続が解除されたままになります。

コメントを含むスーパーノードを拡張する場合、スーパーノードが作成される前にコメントを元の場所に戻します。

コメント付きオブジェクトを含むスーパーノードを拡張したにもかかわらず、スーパーノードにコメントが含まれない場合、オブジェクトは元の場所に戻されますが、コメントは再接続されません。

スーパーノードのパラメータ

IBM® SPSS® Modeler では、Minvalue のようなユーザー独自の変数を設定することができます。変数の値は、スクリプトまたは CLEM 式で指定できます。これらの変数は、**パラメータ**と呼ばれます。パラメータは、ストリーム、セッション、およびスーパーノードに対して設定することができます。スーパーノードに対して設定されたパラメータは、スーパーノードまたはスーパーノードにネストされているノードで CLEM 式を作成する際に

利用できます。ネストされたスーパーノードに対して設定されたパラメータを、その親スーパーノードで利用することはできません。

スーパーノードのパラメータを作成して設定するには、次の 2 つのステップがあります。

- スーパーノードのパラメータを定義します。
- 次に、スーパーノードの各パラメータの値を設定します。

これらのパラメータは、カプセル化された任意のノードの CLEM 式中で利用できます。

スーパーノードのパラメータの定義

スーパーノードのパラメータは、ズーム イン モードとズーム アウトモードのどちらでも定義することができます。定義されたパラメータは、カプセル化されたすべてのノードに適用されます。スーパーノードのパラメータを定義するには、まず [スーパーノード] ダイアログ ボックスの [パラメータ] タブに移動する必要があります。次のいずれかの方法で、ダイアログ ボックスを表示してください。

- ストリーム中のスーパーノードをダブル クリックする。
- [スーパーノード] メニューから [パラメータ設定] を選択する。
- スーパーノードにズーム インしている場合に、コンテキスト メニューから [パラメータ設定] を選択する。

ダイアログ ボックスを開くと、今までに定義されているパラメータが [パラメータ] タブに表示されます。

新しいパラメータを定義するには

- ▶ [パラメータの定義] ボタンをクリックして、ダイアログ ボックスを開きます。

図 9-18
スーパーノードのパラメータの定義



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。

スーパーノード パラメータの値の設定

スーパーノードのパラメータを定義したら、CLEM 式やスクリプトを使ってパラメータに値を指定することができます。

スーパーノードのパラメータを指定するには

- ▶ スーパーノードのアイコンをダブルクリックして、[スーパーノード] ダイアログ ボックスを開きます。
- ▶ [スーパーノード] メニューから [パラメータ設定] を選択することもできます。
- ▶ [パラメータ] タブをクリックします。注： このダイアログ ボックス中のフィールドは、このタブの [パラメータの定義] ボタンをクリックして定義されたフィールドです。
- ▶ 作成した各パラメータの値を、テキスト ボックスに入力します。たとえば、minvalue に特定の閾値を設定することができます。値を設定したパラメータは、たとえばこの閾値以上または以下のレコードを選択するなど、さまざまな操作に利用することができます。

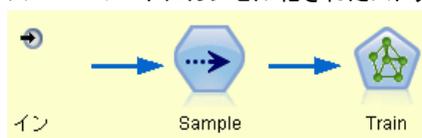
図 9-19
スーパーノードのパラメータの指定



スーパーノード パラメータを使ったノードのプロパティへのアクセス

スーパーノード パラメータは、カプセル化されているノードのプロパティ（スロット パラメータとも呼ばれます）を定義するために使用することもできます。たとえば、無作為のデータ サンプルを使用して、カプセル化されたニューラル ノードを一定時間に渡って学習するようにスーパーノードを設定する場合を考えてみましょう。パラメータを使って、学習時間の長さやサンプルの割合を指定することができます。

図 9-20
スーパーノードにカプセル化されたストリーム フラグメント



この例では、スーパーノードにサンプル ノードSampleおよびニューラル ノードTrainがあると仮定します。ノードのダイアログ ボックスを使って、サンプル ノードの [サンプル] を [無作為%] に、ニューラル ノードの [停止条件] を [時間] に設定します。これらのオプションを指定したら、パラメータを使ってノードのプロパティにアクセスし、スーパーノードに関する特定の値を指定することができます。[スーパーノード] ダイアログ ボックスで、[パラメータの定義] をクリックして、次のパラメータを作成します。

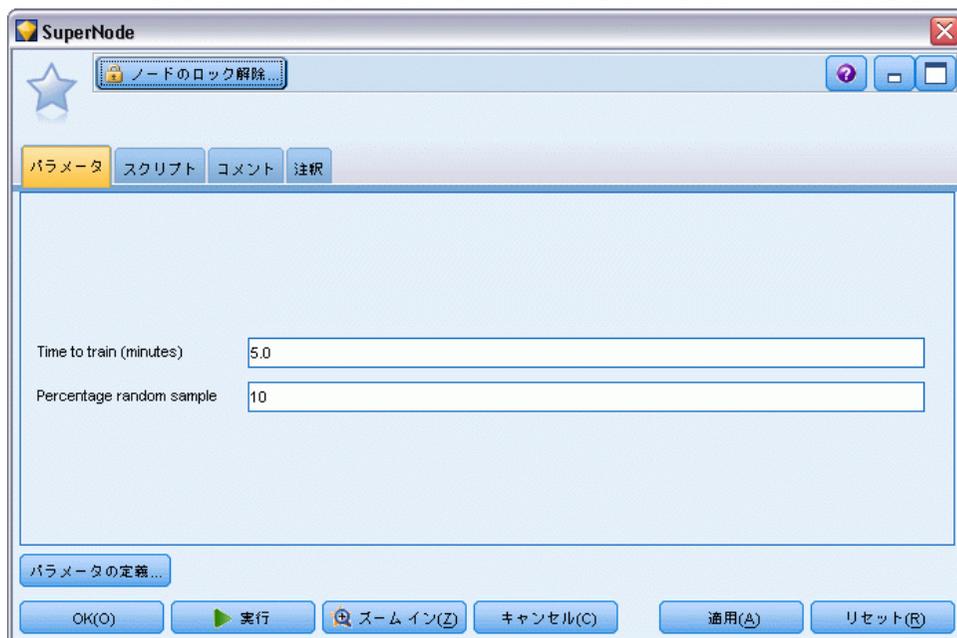
図 9-21
ノードのプロパティにアクセスするためのパラメータの定義



注：Sample.random のようなパラメータ名は、ノードのプロパティを参照するために正しいシンタックスを使用しています。ここで、Sample はノード名を、random はノードのプロパティを表しています。詳細は、9 章 [プロパティ参照の概要 in IBM SPSS Modeler 15 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。

これらのパラメータを定義したら、各ノードのダイアログ ボックスを開くことなく、サンプル ノードとニューラル ノードのプロパティの値を簡単に変更することができます。代わりに、単に [スーパーノード] メニューの [パラメータ設定] を選択して、[スーパーノード] ダイアログ ボックスの [パラメータ] タブを表示し、[無作為 %] と [時間] に新しい値を指定することもできます。この方法は、モデル構築を何回も繰り返してデータを探索しているような場合に役立ちます。

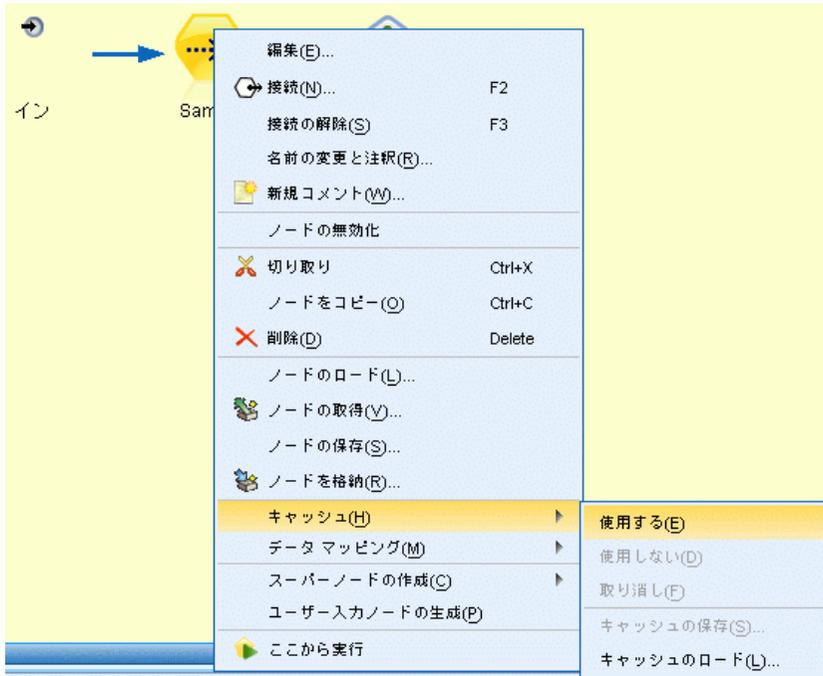
図 9-22
[スーパーノード] ダイアログ ボックスの [パラメータ] タブでのノードのプロパティ値の指定



スーパーノードとキャッシュ

スーパーノードないでは、ターミナル ノードを除くすべてのノードでキャッシュを設定することができます。キャッシュを制御するには、ノードを右クリックして [キャッシュ] コンテキスト メニューから、適切なオプションを選択します。このメニュー オプションは、スーパーノード外から、およびスーパーノード内のカプセル化されたノードで利用することができます。

図 9-23
スーパーノードの [キャッシュ] オプションの選択



次に、スーパーノードのキャッシュに関するガイドラインをいくつか示します。

- スーパーノードでカプセル化されたノードのキャッシュが有効になっている場合は、そのスーパーノードもキャッシュが有効になります。
- スーパーノードのキャッシュを無効にすると、そのスーパーノードでカプセル化されている「すべての」ノードのキャッシュが無効になります。
- スーパーノードのキャッシュを有効にすると、キャッシュ可能な最後のサブノードを実際にキャッシュできるようになります。つまり、最後のサブノードが条件抽出ノードならば、その条件抽出ノードのキャッシュが有効になります。また、最後のサブノードが（キャッシュできない）ターミナルノードの場合は、キャッシュをサポートする次の上流ノードのキャッシュが有効になります。
- スーパーノードのサブノードにキャッシュを設定すると、キャッシュされるノードからの上流操作（ノードの追加や編集など）により、キャッシュが取り消されます。

スーパーノードとスクリプト

IBM® SPSS® Modeler のスクリプト言語を使用すると、ターミナル スーパーノードの内容を操作、実行する簡単なプログラムを作成できます。たとえば、複雑なストリームの実行順序を指定することができます。散布図ノードの前に実行する必要があるグローバル ノードがスーパーノード中にある場合は、グローバル ノードを先に実行するスクリプトを作成することができます。このノードが計算する平均や標準偏差などの値は、散布図ノードを実行するときに使用します。

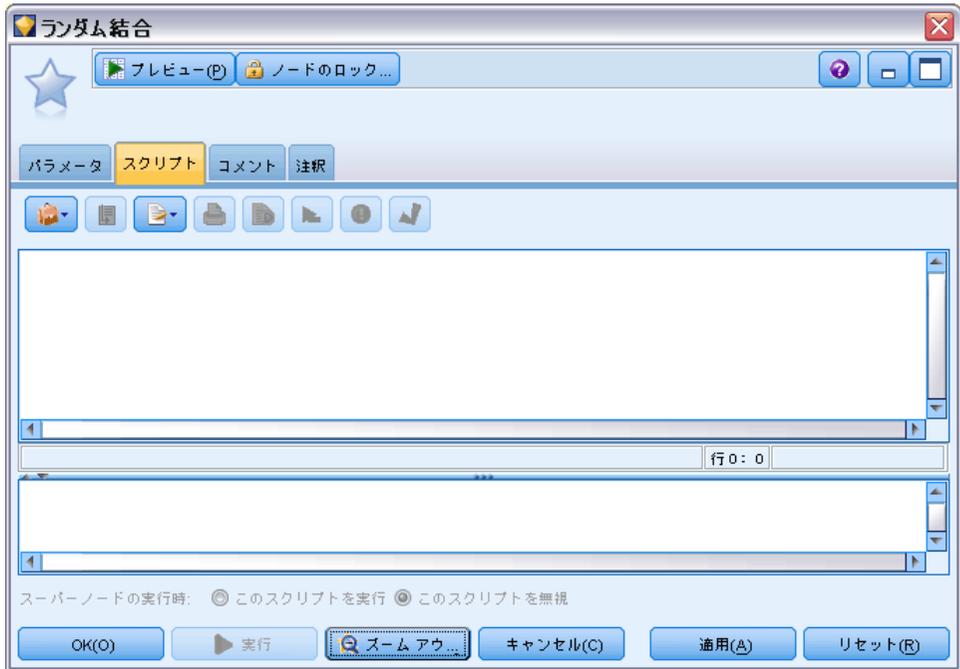
[スーパーノード] ダイアログ ボックスの [スクリプト] タブは、ターミナル スーパーノードの場合にだけ利用することができます。

ターミナル スーパーノードのスクリプト ダイアログ ボックスを表示するには

- ▶ スーパーノード領域を右クリックして、[スーパーノード スクリプト] を選択します。
- ▶ 代わりに、ズーム イン モードおよびズーム アウト モードの両方で、[スーパーノード] メニューの [スーパーノード スクリプト] を選択することもできます。

注： スーパーノード スクリプトは、ダイアログ ボックスで [このスクリプトを実行] を選択した場合にだけ、ストリームまたはスーパーノードとともに実行されます。

図 9-24
スーパーノード用のスクリプトの作成



スクリプティングに固有のオプションと SPSS Modeler 内での使用方法は、SPSS ModelerDVD にある『Scripting, Automation, and CEMI Reference』で解説されています。詳細は、[2 章 スクリプトの概要 in IBM SPSS Modeler 15 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。

スーパーノードの保存とロード

スーパーノードの利点の 1 つとして、保存して他のストリームで再使用できることがあげられます。スーパーノードの保存やロードには、拡張子 `.slb` を使用することに注意してください。

スーパーノードを保存するには

- ▶ スーパーノードをズーム インします。
- ▶ [スーパーノード] メニューから [スーパーノードの保存] を選択します。
- ▶ ダイアログ ボックスで、フィールド名とディレクトリを指定します。
- ▶ 保存したスーパーノードを、現在のプロジェクトに追加するかどうかを選択します。
- ▶ [保存] をクリックします。

スーパーノードをロードするには

- ▶ IBM® SPSS® Modeler ウィンドウの [挿入] メニューから、[スーパーノード] を選択します。
- ▶ 現在のディレクトリからスーパーノード ファイル (.slb) を選択するか、別のディレクトリのファイルを指定します。
- ▶ [ロード] をクリックします。

注： インポートしたスーパーノードのパラメータには、すべてデフォルト値があります。パラメータを変更する場合は、ストリーム領域でスーパーノードをダブルクリックしてください。

注意事項

この情報は、世界各国で提供される製品およびサービス向けに作成されています。

IBMはこのドキュメントで説明する製品、サービス、機能は他の国では提供していない場合があります。現在お住まいの地域で利用可能な製品、サービス、および、情報については、お近くの IBM の担当者にお問い合わせください。IBM 製品、プログラム、またはサービスに対する参照は、IBM 製品、プログラム、またはサービスのみが使用することができることを説明したり意味するものではありません。IBM の知的所有権を侵害しない機能的に同等の製品、プログラム、またはサービスを代わりに使用することができます。ただし、IBM 以外の製品、プログラム、またはサービスの動作を評価および確認するのはユーザーの責任によるものです。

IBMは、本ドキュメントに記載されている内容に関し、特許または特許出願中の可能性があります。本ドキュメントの提供によって、これらの特許に関するいかなる権利も使用者に付与するものではありません。ライセンスのお問い合わせは、書面にて、下記住所に送ることができます。

IBM Director of Licensing, IBM Corporation, North Castle Drive,
Armonk, NY 10504-1785, U. S. A.

2 バイト文字セット (DBCS) 情報についてのライセンスに関するお問い合わせは、お住まいの国の IBM Intellectual Property Department に連絡するか、書面にて下記宛先にお送りください。

神奈川県大和市下鶴間1623番14号 日本アイ・ビー・エム株式会社 法務・知的財産 知的財産権ライセンス渉外

以下の条項は、イギリスまたはこのような条項が法律に反する他の国では適用されません。 International Business Machines は、明示的または黙示的に関わらず、第三者の権利の侵害しない、商品性または特定の目的に対する適合性の暗黙の保証を含むがこれに限定されない、いかなる保証なく、本出版物を「そのまま」提供します一部の州では、特定の取引の明示的または暗示的な保証の免責を許可していないため、この文が適用されない場合があります。

この情報には、技術的に不適切な記述や誤植を含む場合があります。情報については変更が定期的に行われます。これらの変更は本書の新版に追加されます。IBM は、本書に記載されている製品およびプログラムについて、事前の告知なくいつでも改善および変更を行う場合があります。

IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、決してそれらの 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 IBM 製品の資料に含まれるものではなく、これらの Web サイトの使用はお客様の責任によるものとします。

IBM はお客様に対する一切の義務を負うことなく、自ら適切と考える方法で、情報を使用または配布することができるものとします。

本プログラムのライセンス取得者が (i) 別途作成されたプログラムと他のプログラム (本プログラムを含む) との間の情報交換および (ii) 交換された情報の相互利用を目的とした本プログラムに関する情報の所有を希望する場合、下記住所にお問い合わせください。

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

上記のような情報は、該当する条項および条件に従い、有料で利用できるものとします。

本ドキュメントに記載されている許可されたプログラムおよびそのプログラムに使用できるすべてのライセンス認証された資料は、IBM Customer Agreement、IBM International Program License Agreement、および当社とかわした同等の契約の条件に基づき、IBM によって提供されます。

ここに記載されているパフォーマンスデータは、すべて管理環境下で確認されたものです。そのため、他の操作環境で得られた結果は大きく異なる可能性があります。開発レベルのシステムで測定が行われている場合があり、これらの測定値は一般に利用可能なシステムと同じであることを保証するものではありません。また、測定値が推定値である可能性があり、実際の結果は異なる場合があります。本ドキュメントのユーザーは、特定の環境に適したデータを検証する必要があります。

IBM 以外の製品に関する情報は、それらの製品の供給業者、公開済みの発表、または公開で使用できるソースから取得しています。IBM は、それらの製品のテストは行っておらず、IBM 以外の製品に関連する性能、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給業者に通知する必要があります。

IBM の将来の方向性または意向に関する記述については、予告なく変更または取り消すことがあり、目的や目標のみを示すものです。

この情報には、日常の業務処理で用いられるデータや報告書の例が含まれています。できる限り詳細に説明するため、例には、個人、企業、ブランド、製品などの名前が使用されています。これらの名称はすべて架空のものであり、実際の企業で使用される名称および住所とは一切関係ありません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーのイラストが表示されない場合があります。

商標

IBM、IBM ロゴ、および ibm.com、SPSS は、世界の多くの国で登録された IBM Corporation の商標です。IBM の商標の現在のリストは、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

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、および Pentium は、米国およびその他の国の Intel Corporation または関連会社の商標または登録商標です。

Linux は、米国およびその他の国における Linus Torvalds の登録商標です。

Microsoft、Windows、Windows NT、および Windows のロゴは、米国およびその他の国における Microsoft 社の商標です。

UNIX は、米国およびその他の国における The Open Group の登録商標です。

Java およびすべての Java ベースの商標およびロゴは、米国およびその他の国の Sun Microsystems, Inc. の商標です。

その他の製品名およびサービス名等は、IBM または他の会社の商標です。



索引

- 周期的時間要素
 - 自動データ準備, 126
- 複数二分法設定, 177
- 自動日付認識, 32, 35
- 小数点記号, 30, 32, 170
 - 数字の表示フォーマット, 173
 - ファイル エクスポート ノード, 527
- 最大値関数
 - 時系列の集計, 246
- 最小値関数
 - 時系列の集計, 246
- 最頻値関数
 - 時系列の集計, 246
- 衝突変更子, 420
- 要約統計量
 - データ検査ノード, 459
- 一致行列
 - 精度分析ノード, 454
- 並行処理
 - 結合, 109
 - 並べ替え, 98
 - レコード集計ノード, 94
- 会計年度
 - 時間区分ノード, 253
- 信頼区間
 - 平均比較ノード, 492-493
- 内部結合, 99
- 合計関数
 - 時系列の集計, 246
- 外部結合, 99
- 実行順序
 - 設定, 587
- 小数桁数
 - 表示書式, 172
- 平均関数
 - 時系列の集計, 246
- 推定期間, 247
- 換算係数, 89
- 日付認識, 32, 35
- 最終関数
 - 時系列の集計, 246
- 未定義値, 102
- 標準偏差
 - 記述統計出力, 485
 - グローバル ノード, 498
 - データ分割ノード, 220
- 毎時測定
 - 時間区分ノード, 258-259
- 相関係数, 483
 - 記述統計出力, 485
 - 絶対値, 483
 - 有意, 483
- probability, 483
 - 平均比較ノード, 493
 - 詳細ラベル, 483
- 空白関数
 - 時系列の埋め込み, 246
- 自動入力, 155, 158
- 表示書式
 - 科学的表記, 172
 - 小数桁数, 172
 - 数値, 172
 - 通貨, 172
 - グループ化記号, 172
- 部分結合, 99, 104
- 中央値
 - 記述統計出力, 485
- 分割点
 - データ分割ノード, 212
- 合計値, 91
- 周期性
 - 時系列データ, 241
- 報告書
 - 出力の保存, 444
- 変数名
 - データのエクスポート, 502, 527, 535, 562
- 引用符
 - テキスト ファイルのインポート, 32
 - データベースのエクスポート, 503
- 散布図, 289, 332, 356
 - 3-D, 289
 - 六角ビン分割, 289
 - ビン分割, 289
- 時系列, 263
- 最小値
 - 記述統計出力, 485
 - グローバル ノード, 498
- 期待値
 - クロス集計ノード, 451
- 極座標, 419
- 欠損値, 118, 159, 165
 - 処理, 470
 - 置換, 470
 - クロス集計テーブルで, 450
 - レコード集計ノード内, 90
- 空白値
 - クロス集計テーブルで, 450
- 空白行
 - Excel ファイル, 58
- 自由度
 - クロス集計ノード, 452
 - 平均比較ノード, 492-493
- 視覚化
 - 数値書式, 413

索引

- 透過度, 409
- 余白, 412
- 編集, 405
- 軸, 414
- text, 408
- 色およびパターン, 409
- カテゴリ, 416
- グラフとチャート, 269
- コピー, 424
- 座標システムの変換, 419
- スケール, 414
- ダッシュ, 409
- 行と列の入れ換え, 416, 418-419
- 凡例の位置, 424
- バディング, 412
- パネル, 416, 418
- ポイントの縦横比, 411
- ポイントの回転, 411
- ポイントの形, 411
- 編集モード, 405
- 逆結合, 99
- 透過度
 - 視覚化, 271
- 重要度
 - 平均値の比較, 489
 - 平均比較ノード, 492-493
- 作成
 - 新規フィールド, 185-186
- 保存
 - output, 436
 - 出力オブジェクト, 435, 444
- 凡例
 - position, 424
- 分位
 - データ分割ノード, 216
- 分散
 - 記述統計出力, 485
- 列幅
 - フィールド用, 170
- 削除
 - 出力オブジェクト, 435
 - 視覚化スタイルシート, 320
 - 視覚化テンプレート, 320
 - フィールドのマップ, 320
- 区分
 - 時系列データ, 241
- 収益
 - 評価グラフ, 384
- 合計
 - 記述統計出力, 485
 - グローバル ノード, 498
- 商標, 592
- 図表
 - 出力の保存, 444
- 変換
 - reclassify, 207, 212
 - 再コード化, 207, 212
- 外観
 - 視覚化, 271
- 実数, 162
- 実行
 - 順序の指定, 587
- 密度
 - 3-D, 290
- 属性
 - マップ内, 323
- 干渉
 - 作成, 374
- 平均
 - 記述統計出力, 485
 - グローバル ノード, 498
 - データ分割ノード, 220
- 度数
 - 記述統計出力, 485
 - データ分割ノード, 216
- 役割
 - フィールドの指定, 72, 150, 167
- 接続
 - IBM SPSS Collaboration and Deployment Services Repository へ, 10
- 整数, 162
- 日付
 - フォーマットの設定, 170, 172
- 昇順, 96
- 時間
 - フォーマットの設定, 170
- 書式
 - data, 35, 170
- 最大
 - 記述統計出力, 485
 - グローバル ノード, 498
- 有意
 - 相関強度, 483
- 期間
 - 時間区分ノード, 250
- 条件
 - 一連の条件の指定, 196
 - 結合の条件の指定, 104
- 検査
 - 初期データ検査, 459
 - データ検査ノード, 459
- 検索
 - テーブル ブラウザ, 446
- 残差
 - クロス集計ノード, 451
- 破棄
 - fields, 173

- 空白, 470
 - クロス集計テーブルで, 450
- 範囲, 153
 - 欠損値, 159
- 結合, 98-99, 102
 - 部分外部, 104
- 表示
 - ブラウザの HTML 出力結果, 440
- 追加
 - records, 90
- 配置
 - フィールド用, 170
- 重複
 - fields, 98, 174
 - records, 113
- 閾値
 - ビン閾値の表示, 223
- 降順, 96
- 例
 - 概要, 7
 - アプリケーション ガイド, 5
- 値
 - フィールドおよび値ラベル, 72, 150
- 形
 - 視覚化, 271
- 色
 - 視覚化, 271
- 偽 (false) の値, 164
- 散布図行列 (SPLOM), 289
- 真 (true) の値, 164
- 行 (ケース) の選択, 77
- 第 1 入力フィールド
 - データベース エクスポート ノード, 509
- 一元 ANOVA
 - 平均比較ノード, 488
- 平均/標準偏差
 - フィールドの分割に使用, 220
- 日付/時刻, 153
- 10 分位ビン, 216
- 100 分位ビン, 216
- 2-D ドット プロット, 290
- 20 分位ビン, 216
- 3 次元密度, 290
- 3 次元グラフ, 277
- 3 次元グラフの回転, 277
- 3-D 散布図, 289
- 3-D 円グラフ, 288
- 3-D 棒グラフ, 288
- 3-D 面グラフ
 - 説明, 288
- 3-D ヒストグラム, 290
- 4 分位ビン, 216
- 5 分位ビン, 216
- ADO データベース
 - インポート, 41
- ANOVA
 - 平均比較ノード, 488
- auto- 設定, 408
- bar chart
 - example, 295, 297
- BITMAP インデックス
 - データベース テーブル, 513
- boxplot
 - example, 303
- choropleth
 - example, 312
- CLEM 式, 76
- Clem 式ビルダー, 76
- clustered bar chart
 - example, 297
- Cognos、IBM Cognos BI を参照, 54
- color map
 - example, 312
- costs
 - 評価グラフ, 384
- CREATE INDEX コマンド, 511
- CRISP-DM
 - データの理解, 9
- CRISP-DM プロセス モデル
 - データの準備, 118
- CSV データ
 - インポート, 41
- DAT ファイル
 - 保存, 444
 - エクスポート, 441, 536
- data
 - 匿名化, 202
 - 検査, 459
 - 検証, 459
 - 準備, 76
 - 理解, 76
 - storage, 199-200
 - ストレージ タイプ, 159
- Data Collection 調査データ
 - インポート, 39-40
- Data Collection 入力ノード, 39-40
 - メタデータ ファイル, 41
 - ログ ファイル, 41
- distribution, 346
- dot plot
 - example, 301
- DPD, 10
- employee_data.sav データ ファイル, 547
- Enterprise View ノード, 10
- EOL 文字, 30
- ESRI ファイル, 322

索引

- events
 - 作成, 374
- Excel
 - IBM SPSS Modeler からの起動, 536
- Excel インポート ノード
 - 出力から生成, 536
- Excel エクスポート ノード, 536
- Excel 入力ノード, 58
- Excel ファイル
 - エクスポート, 536
- extension
 - 派生フィールド, 188
- F 統計量
 - 平均比較ノード, 492
- fields
 - データの匿名化, 202
 - 行と列の入れ換え, 236-237
 - 複数の選択, 190
 - 並び替え, 265
 - フィールドおよび値ラベル, 162
 - 複数フィールドの作成, 188
- FILLFACTOR キーワード
 - データベース テーブルのインデックス作成, 513
- frequencies
 - データ分割ノード, 216
- hassubstring 関数, 192
- HDATA フォーマット
 - Data Collection 入力ノード, 39
- heat map
 - example, 307
- histogram
 - example, 299
- HTML
 - 出力の保存, 445
- HTML 出力
 - ブラウザ表示, 440
 - レポート ノード, 495
- IBM Cognos BI エクスポート ノード, 54, 530-532
- IBM Cognos BI 入力ノード, 49, 54-56
 - アイコン, 49
 - データのインポート, 50
 - レポートのインポート, 52
- IBM SPSS Collaboration and Deployment Services Repository
 - 接続先, 10
 - テンプレート、スタイルシート、マップ ファイルの場所として使用する, 319
- IBM SPSS Data Collection エクスポート ノード, 528
- IBM SPSS Data Collection 入力ノード, 48
 - language, 44
 - 多重回答グループ, 46
 - データベース接続の設定, 45-46
 - ラベル データ型, 44
- IBM SPSS Modeler, 1
 - ドキュメンテーション, 5
- IBM SPSS Statistics
 - IBM SPSS Modeler からの起動, 499, 555, 562
 - 有効なフィールド名, 563
 - ライセンスの位置, 499
- IBM SPSS Statistics データ ファイル
 - 調査データのインポート, 41
- IBM SPSS Statistics ノード, 543
- IBM SPSS Statistics 出力ノード
 - [出力] タブ, 559
- IBM SPSS Statistics モデル, 552
 - 高度なナゲットの詳細, 554
 - について, 552
 - モデル オプション, 552
 - モデル ナゲット, 554
- if-then-else 文, 197
- In2data データベース
 - インポート, 41
- language
 - IBM SPSS Data Collection 入力ノード, 44
- LOESS 平準化
 - 散布図ノード, 335
- lowess 平準化「LOESS 平準化」を参照
 - 散布図ノード, 335
- map visualization
 - example, 312
- map visualizations
 - creating, 294
- MDD ドキュメント
 - インポート, 41
- means
 - 比較, 487-489, 491
- Microsoft Excel 入力ノード, 58
- mode
 - 記述統計出力, 485
- n 件ごとのサンプリング, 80
- ODBC
 - IBM Cognos BI エクスポート ノードの接続, 532
 - データベース入力ノード, 16
 - バルク ロードによる, 515, 517
- ODBC エクスポート ノード : データベース エクスポート ノードを参照, 502
- Oracle, 16
- output
 - 保存, 436
 - 印刷, 436
 - HTML, 440
 - エクスポート, 441
 - から新規ノードを生成, 436

- p 値
 - 重要度, 489
- parameters
 - IBM Cognos BI, 56
 - スーパーノード, 580, 582
 - スーパーノードの設定, 579
 - ノードのプロパティ, 583
- Pearson の積率相関係数
 - 記述統計出力, 485
 - 平均比較ノード, 493
- Pearson のカイ 2 乗
 - クロス集計ノード, 452
- performance
 - 結合, 109
 - サンプリング データ, 79
 - データ分割ノード, 223
 - フィールド作成ノード, 223
 - 並べ替え, 98
 - レコード集計ノード, 94
- pie chart
 - example, 305
- Python
 - バルク ロード スクリプト, 515, 517
- Quancept データ
 - インポート, 41
- Quantum データ
 - インポート, 41
- Quanvert データベース
 - インポート, 41
- range
 - 記述統計出力, 485
- recency
 - 相対日付の設定, 95
- record
 - 度数, 91
 - length, 33
 - ラベル, 167
- records
 - 行と列の入れ換え, 236-237
- RFM 分析ノード
 - 概要, 225
 - 設定, 226
 - 独立した分割, 94, 226
 - ネストされた分割, 94, 226
 - 値の分割, 228
- RFM レコード集計ノード
 - 概要, 94
 - オプションの設定, 95
 - 独立した分割, 94, 226
 - ネストされた分割, 94, 226
- ROI
 - 図表, 378, 388
- SAS
 - インポート オプションの設定, 57
 - SAS エクスポート ノード, 534-535
 - SAS 入力ノード
 - .sd2 (SAS) ファイル, 56
 - .ssd (SAS) ファイル, 56
 - .tpt (SAS) ファイル, 56
 - トランスポート ファイル, 56
 - .sav ファイル, 544
 - scatterplot matrix
 - example, 309, 314
 - .sd2 (SAS) ファイル, 56
 - .slb ファイル, 588
 - smoother
 - 散布図ノード, 335
 - SMZ ファイル
 - 作成, 322
 - 削除, 320
 - 概要, 322
 - インポート, 320
 - エクスポート, 320
 - 名前の変更, 320
 - プリインストール, 322
 - プリインストールされた SMZ ファイルの編集, 322
 - SourceFile 変数
 - IBM SPSS Data Collection 入力ノード, 44
 - SPLOM, 289
 - example, 309, 314
 - SPSS Modeler Server, 2
 - SQL クエリー
 - データベース入力ノード, 16, 19, 26-27
 - .ssd (SAS) ファイル, 56
 - statistics
 - 説明, 280, 421
 - クロス集計ノード, 449
 - データ検査ノード, 459
 - 視覚化における編集, 421
 - Statistics エクスポート ノード, 561
 - [エクスポート] タブ, 562
 - Statistics 出力ノード, 555
 - [シンタックス] タブ, 556
 - Statistics 変換ノード, 547
 - オプションの設定, 547
 - [シンタックス] タブ, 547
 - 利用可能なシンタックス, 550
 - Statistics ファイル ノード, 544
 - storage, 159
 - 変換, 198-200
 - Surveycraft データ
 - インポート, 41
- t 検定
 - 独立サンプル, 488
 - 一対のサンプル, 489
 - 平均比較ノード, 488-489, 493

索引

text

data, 28, 32
区切り, 28

TimeIndex フィールド
時間区分ノード, 244

TimeLabel フィールド
時間区分ノード, 244

timestamp, 153

.tpt (SAS) ファイル, 56

Triple-S データ
インポート, 41

type, 35

UNIQUE キーワード
データベース テーブルのインデックス作成,
513

UTF-8 文字コード, 31, 35, 528

values

設定, 159
フィールドおよび値ラベル, 159
読み取り, 158

VDATA フォーマット

Data Collection 入力ノード, 39

Web グラフ ノード, 361

[外観] タブ, 367

Web グラフの概要, 372

[オプション] タブ, 365

グラフの使用, 368

スライド, 370

閾値の調整, 371

[プロット] タブ, 363

ポイントの調整, 369

リンク スライド, 370

リンクの定義, 365

レイアウトの変更, 370

Web グラフの方向付きレイアウト, 366

Web グラフのネットワーク レイアウト, 366

Web に公開, 437

weights

評価グラフ, 384

XLS ファイル

エクスポート, 536

XML 出力

レポート ノード, 495

XML エクスポート ノード, 537

XML 入力ノード, 60

XPath のシンタックス, 60

アイコン、IBM Cognos BI, 49

アニメーション

視覚化, 273

アプリケーションの例, 5

アンサンプル ノード

スコアの結合, 180

出力フィールド, 180

いずれかが真 (true) の場合は真関数

時系列の集計, 246

インスタンス化, 72, 150, 153, 157-158

入力ノード, 74

インポート

視覚化スタイルシート, 320

スーパーノード, 588

視覚化テンプレート, 320

データ、IBM Cognos BI, 50

フィールドのマップ, 320

レポート、IBM Cognos BI, 52

エクスポート

output, 441

視覚化スタイルシート, 320

スーパーノード, 588

視覚化テンプレート, 320

フィールドのマップ, 320

エクスポート ノード, 501

出力オブジェクトの名前の変更, 435

オプション

IBM SPSS Statistics, 499

結合オプション、データベース エクスポート,
504

オーバーラップ マップ, 293-294

カイ 2 乗

クロス集計ノード, 452

等カウント

データ分割ノード, 216

カウント型フィールド

時系列の埋め込みまたは集計, 246

時間区分ノード, 246

監視カテゴリ化, 221

多重カテゴリ グループ, 177

カテゴリ データ, 154, 156

カラー マップ, 291-292

カンマ, 32, 170

カンマ区切りファイル

保存, 444

エクスポート, 441, 536

キャッシュ

スーパーノード, 585

キャッシュ ファイル ノード, 544

連続キー, 91

キー フィールド, 91, 233

- キーによる方法, 98
- 開く
 - 出力オブジェクト, 435
- クエリー
 - データベース入力ノード, 16, 19
- クエリー エディタ
 - データベース入力ノード, 26-27
- クラスタ, 423
- クラスタ化サンプル, 79-80, 84
- グラフ
 - 時系列, 373
 - 作図, 332
 - 保存, 431
 - 印刷, 431
 - 検証, 391
 - 脚注, 425
 - 領域, 397
 - 棒, 341
 - [出力] タブ, 275
 - [注釈] タブ, 276
 - 3次元イメージの回転, 277
 - 3-D, 277
 - title, 425
 - Web, 361
 - エクスポート, 431
 - 集計棒グラフ, 351
 - 評価グラフ, 378
 - 線グラフ, 356
 - グラフィック要素のサイズ, 411
 - グラフボード, 278
 - コピー, 431
 - 編集されたレイアウトの保存, 427
 - スタイルシート, 427
 - デフォルトの色設定, 427
 - データ検査から生成, 475
 - 出力の保存, 444
 - 領域の削除, 399
 - ノードの生成, 402
 - バンド, 392
 - ヒストグラム, 346
 - 軸ラベル, 425
 - レイアウト変更の保存, 427
- グラフ要素
 - 衝突変更子, 423
 - 変換, 420
 - 変更, 420
 - タイプ, 420
- 平行座標グラフ, 291
- 表面グラフ, 288
- 円グラフ, 287
 - 度数, 287, 292
 - 3-D, 288
 - マップ上, 292-293
- 帯グラフ, 288
- 棒グラフ, 287
 - 度数, 287, 292
 - 3-D, 288
 - マップ上, 292
- 面グラフ, 288
 - 3-D, 288
- 形状グラフ オーバーレイ, 271
- 色グラフ オーバーレイ, 271
- グラフ タイプ
 - グラフボード, 287
- 集計棒グラフ ノード, 351
 - [外観] タブ, 353
 - [オプション] タブ, 351-352
 - グラフの使用, 355
- 棒グラフ ノード, 341
 - [外観] タブ, 342
 - グラフの使用, 343
 - テーブルの使用, 343
 - [プロット] タブ, 341
- 線グラフ ノード, 356
 - [外観] タブ, 359
 - グラフの使用, 360
 - [プロット] タブ, 357
- グラフからのノードの生成, 402
 - データ分類ノード, 405
 - 条件抽出ノード, 403
 - バランス ノード, 404
 - フィルタ ノード, 404
 - フィールド作成ノード, 404
- グラフ内の透過度, 271
- グラフの検証, 391
 - 領域, 397
 - グラフ バンド, 392
 - 要素のマーク, 401
 - マジック ワンド, 401
- グラフの編集
 - グラフィック要素のサイズ, 411
- グラフの領域, 397
- グラフ内のアニメーション, 271, 273
- グラフのオーバーレイ, 271
- グラフ内のバンド, 392
- グラフ内のマジック ワンド, 401
- グラフ作成ノード, 269
 - 分布, 341
 - 評価, 378
 - Histogram, 346
 - Web, 361
 - アニメーション, 271, 273
 - オーバーレイ, 271
 - 時系列グラフ, 373
 - 集計棒グラフ, 351
 - 線グラフ, 356

索引

- グラフボード, 278
- パネル, 271, 273
- プロット, 332
- グラフボード
 - グラフ タイプ, 287
- グラフボード ノード, 278
 - [外観] タブ, 316
- グループ化記号
 - 数字の表示フォーマット, 172
- 多重回答グループ
 - 複数二分法設定, 177
 - 視覚化, 281
 - 削除, 177
 - 定義, 177
 - Data Collection 入力ノード, 39-40
 - IBM SPSS Data Collection 入力ノード, 46, 48
 - IBM SPSS Statistics 入力ノード, 546
 - 多重カテゴリ グループ, 177
- クロス集計出力
 - テキストとして保存, 445
- クロス表
 - クロス集計ノード, 449, 451
- クロス集計ノード, 449
 - 強調表示, 451
 - [出力] タブ, 444
 - [外観] タブ, 451
 - [設定] タブ, 449
 - クロス表, 451
 - 行と列のソート, 451
 - 列パーセント, 451
 - 行パーセント, 451
 - 出力ブラウザ, 452
- クロス集計ブラウザ
 - [ノードの生成] メニュー, 452
- グローバル値, 497
- グローバル ノード, 497
 - [設定] タブ, 498

- ゲイン グラフ, 378, 388
- 避ける, 423
- ケース データ
 - Data Collection 入力ノード, 39-40
- ケースのランク, 219

- コミット サイズ, 515
- コメント
 - スーパーノードによる使用, 579
- コメント文字
 - 可変長ファイル内, 30
- コプレス マップ, 291-292
- コード変数
 - IBM SPSS Data Collection 入力ノード, 44
 - 自動再コード化, 207-208
 - 再コード化, 207-208, 212
 - 文字コード, 31, 35, 528

- サイズ
 - 視覚化, 271
- サイズ グラフ オーバーレイ, 271
- 計算された期間
 - 自動データ準備, 126
- 調整された傾向スコア
 - データのバランス, 89
- サンプリング データ, 87
- サンプリング フレーム, 79
- 非無作為サンプル, 79-80
- 体系的サンプル, 79-80
- 学習サンプル
 - データ区分, 229-230
 - バランス, 89
- 層化サンプル, 79-80, 84, 87
- 検定サンプル
 - データ区分, 229-230
- サンプル ノード
 - クラスタ化サンプル, 79-80, 84
 - サンプリング フレーム, 79
 - 非無作為サンプル, 79-80
 - 体系的サンプル, 79-80
 - 無作為サンプル, 79-80
 - 層化サンプル, 79-80, 84, 87
 - 階層のサンプル サイズ, 87
 - 重み付けされたサンプル, 84

- 同じ値
 - データ分割ノード, 216
- シェープファイル, 322
- システム欠損値, 470
 - クロス集計テーブルで, 450
- システム変数
 - IBM SPSS Data Collection 入力ノード, 44
- 座標システム
 - 変換, 419
- ジッタリング, 337
- ジッター, 423
- シナリオ, 10
- [シンタックス] タブ
 - Statistics 出力ノード, 556
- シード値
 - サンプリングとレコード, 84, 231

- スキーマ
 - データベース エクスポート ノード, 507
- スクリプト
 - スーパーノード, 587

- 傾向スコア
 - データのバランス, 89
- スコアリング
 - 評価グラフ ノードのオプション, 385
- スタイルシート
 - 削除, 320
 - インポート, 320
 - エクスポート, 320
 - 名前の変更, 320
- 視覚化スタイルシート
 - 削除, 320
 - 場所, 318
 - 適用, 429
 - インポート, 320
 - エクスポート, 320
 - 名前の変更, 320
- ストリーム パラメータ, 26-27
- ストレージのフォーマット, 35
- 循環する期間
 - 時間区分ノード, 251
- スーパーノード, 565
 - 作成, 568
 - 保存, 588
 - 編集, 577
 - キャッシュの作成, 585
 - スクリプト, 587
 - 入力スーパーノード, 566
 - ズーム イン, 577
 - タイプ, 565
 - ターミナル スーパーノード, 567
 - によるコメントの使用, 579
 - パスワード保護, 574, 576
 - パラメータの設定, 579
 - プロセス スーパーノード, 566
 - 入れ子, 571
 - ロック, 574
 - ロック解除, 574
 - ロード, 588
- スーパーノードのパラメータ, 580, 582-583
- スーパーノードのロック, 574
- スーパーノードのロック解除, 574
- ズーム, 577

- セット
 - 変換, 208, 210
 - フラグ型への変換, 232-233
- セット型, 153
- セット型からフラグ型への変換, 232-233
- セルの範囲
 - Excel ファイル, 58

- ソート
 - あらかじめソートされたフィールド, 116
 - 重複レコード ノード, 115
 - ソート ノード
 - 最適化設定, 97
 - 概要, 96

- 使用タイプ, 35, 153
- タイムスタンプ ストレージ フォーマット, 36, 69
- タグ, 98, 107
- ダミー コーディング, 232

- 小数点つき順位, 219
- 偏ったデータ, 88

- テキスト
 - 文字コード, 31, 35, 528
- テキスト ファイル, 28
 - エクスポート, 536
- テスト サンプル
 - データ区分, 229-230
- テンプレート
 - 削除, 320
 - インポート, 320
 - エクスポート, 320
 - 名前の変更, 320
 - レポート ノード, 494
- 視覚化テンプレート
 - 削除, 320
 - 場所, 318
 - インポート, 320
 - エクスポート, 320
 - 名前の変更, 320
- データ
 - 集計, 90
 - ストレージ, 36, 69
- データ型属性, 169
- データ区分, 229-230
 - 評価グラフ, 385
 - 精度分析ノード, 454
- データ品質
 - データ検査ブラウザ, 470
- 市場調査データ
 - Data Collection 入力ノード, 39
 - IBM SPSS Data Collection 入力ノード, 46
 - インポート, 40, 48
- 自動データ準備
 - 予測精度, 141
 - 特徴選択, 131
 - 構築, 131
 - 目的, 120
 - fields, 123
 - アクションの概要, 140

索引

- アクションの詳細, 145
- 日付と時刻の準備, 126
- 連続型目標の正規化, 130, 148
- 入力の準備, 128
- 目標の準備, 128
- ビューのリセット, 136
- ビュー間のリンク, 136
- フィールド分析, 138
- フィールド設定, 124
- [フィールド] テーブル, 142
- フィールド処理の要約, 136
- フィールドの詳細, 143
- フィールドの選択, 131
- フィールドの除外, 127
- 未使用フィールドの除外, 124
- フィールドの名前付け, 133
- フィールド生成ノードの生成, 148
- モデル ビュー, 134
- 四半期データ
 - 時間区分ノード, 253
- 時系列データ
 - 推定期間, 247
 - 区分, 242
 - 定義, 241-242, 244, 247
 - 集計, 241, 244
 - データから構築, 244
 - パディング, 241, 244
 - ホールドアウト, 247
 - ラベリング, 241-242, 244, 247
- 合成データ
 - ユーザー入力ノード, 65
- 名義データ, 154, 163
- 年次データ
 - 時間区分ノード, 252
- 月次データ
 - 時間区分ノード, 254
- 要約データ, 90
- 調査データ
 - Data Collection 入力ノード, 39
 - インポート, 40, 46, 48
- 連続データ, 154, 156, 162
- 遅延データ, 263
- 順序データ, 154, 163
- 週データ
 - 時間区分ノード, 255
- データ セットの結合, 111
- データ ソース
 - データベース接続, 20
- データ プロバイダの定義, 10
- データ ラベル
 - 視覚化, 271
- データの再構成, 233
- データ値の変更, 185
- データの削減, 77, 79
- データの検証
 - データ検査ノード, 459
- データの型, 33, 72, 118, 150, 153
 - インスタンス化, 157
- 時系列データの集計, 244
- 入力データの順序, 107
- データのエクスポート
 - DAT ファイル, 536
 - Excel, 536
 - IBM Cognos BI エクスポート ノード, 54, 530-532
 - IBM SPSS Statistics へ, 561
 - SAS 形式, 534
 - text, 536
 - XML 形式, 537
 - データベースへ, 502
 - フラット ファイル形式, 526
- データの順序付け, 96, 265
- データ型属性のコピー, 169
- 連続データのサンプリング, 80
- データ型のチェック, 165
- データ型不明のデータ, 155
- データの組み合わせ, 111
 - 複数ファイルからの, 98
- 時系列データの埋め込み, 244
- データ型の割り当て, 72, 118, 150
- データの入れ替え, 236-237
- 時系列データのロール アップ, 244
- データ分割ノード
 - 等合計, 216
 - 概要, 212
 - 平均/標準偏差のビン, 220
 - オプションの設定, 213
 - 等カウント, 216
 - 最適な, 221
 - 固定幅のデータ分割, 215
 - ビンのプレビュー, 223
 - ランク, 219
- データ分類ノード, 208, 210
 - 概要, 207, 212
 - 棒グラフから生成, 343
- データ区分ノード, 229-230
- データ検査ノード, 459
 - [出力] タブ, 444
 - [設定] タブ, 461
- データ型ノード
 - 列調整, 170
 - 列幅, 170
 - 概要, 150
 - オプションの設定, 153, 156
 - 名義データ, 163
 - 連続データ, 162
 - 順序データ, 163
 - データ型のコピー, 169

- 空白の処理, 159
- 値の消去, 72
- [フォーマット] タブ, 170
- フラグ型フィールド データ型, 164
- モデル作成役割の設定, 167
- 自動データ準備ノード, 120
- データ区分フィールド, 72, 150, 167, 229-230
- データ検査ブラウザ
 - グラフの生成, 475
 - ノードの生成, 475
 - [ファイル] メニュー, 465
 - メニューの編集..., 465
- データベース
 - サポートの段階, 17
 - バルク ロード, 515, 517
- データベース接続
 - 定義, 20
 - プリセット値, 22
- データベース エクスポート ノード, 502
 - [エクスポート] タブ, 503
 - 結合オプション, 504
 - スキーマ, 507
 - データ ソース, 503
 - 入力データ
 - フィールドのデータベース列へのマッピング, 504
 - テーブル名, 503
 - テーブルのインデックスを作成, 511
- データベース テーブルのインデックス作成, 511
- データベース テーブルの上書き, 503
- 段階、データベースのサポート, 17
- データベース入力ノード, 16
 - SQL クエリー, 19
 - クエリー エディタ, 26-27
 - テーブルおよびビューの選択, 25
- テーブル
 - 結合, 99
 - テキストとして保存, 445
 - 出力の保存, 444
- テーブル ノード, 443
 - 列調整, 170
 - 列幅, 170
 - [出力] タブ, 444
 - [設定] タブ, 443
 - 出力の設定, 443
 - [フォーマット] タブ, 170
- テーブル ブラウザ
 - 検索, 446
 - セルの選択, 442, 446
 - 列の並び替え, 442, 446
 - [ノードの生成] メニュー, 446
- テーブル形式の出力
 - セルの選択, 442
 - 列の並び替え, 442
- ドキュメンテーション, 5
- ドット プロット, 290
 - 2-D, 290
- トランスポート ファイル
 - SAS 入力ノード, 56
- 最適なデータ分割, 221
- 不均衡なデータ, 88
- 大規模なデータベース, 76
 - データ検査の実行, 459
- 不完全なレコード, 102
- 法律に関する注意事項, 590
- 局所的に重みを付けた 最小 2 乗回帰法
 - 散布図ノード, 335
- 順序による結合, 98
- ヌル, 159, 470
 - クロス集計テーブルで, 450
- ヌル値
 - クロス集計テーブルで, 450
 - 混在データ, 37, 70
- 連続型目標の正規化, 130, 148
- 集計用の四分位値, 91
- 集計用の標準偏差, 91
- 平均の標準誤差
 - 記述統計出力, 485
- 集計用の中央値, 91
- 集計用の分散値, 91
- 集計用の平均値, 91
- 集計用の度数値, 91
- 集計用の最大値, 91
- 集計用の最小値, 91
- 欠損値の処理, 118
- 視覚化の編集, 405
 - 数値書式, 413
 - 自動設定, 407
 - 透過度, 409
 - 余白, 412
 - 選択, 407
 - 軸, 414
 - 3-D 効果の追加, 419
 - text, 408
 - 色およびパターン, 409

索引

- カテゴリ, 416
- カテゴリの結合, 416
- カテゴリの除外, 416
- カテゴリのソート, 416
- カテゴリを閉じる, 416
- 座標システムの変換, 419
- スケール, 414
- ダッシュ, 409
- 行と列の入れ換え, 418-419
- 凡例の位置, 424
- パディング, 412
- パネル, 418
- ポイントの縦横比, 411
- ポイントの回転, 411
- ポイントの形, 411
- ルール, 407
- 値の正規化
 - グラフ作成ノード, 357, 376
- 出力の印刷, 436
- 名前の変更
 - エクスポート用のフィールド, 563
 - 視覚化スタイルシート, 320
 - 視覚化テンプレート, 320
 - フィールドのマップ, 320
- 普通の順序
 - 変更, 265
- 期間の計算
 - 自動データ準備, 126
- 毎日の測定
 - 時間区分ノード, 256-257
- 空白の処理, 72, 150, 159
 - データ分割ノード, 213
 - 値の置換, 199
- 複数の入力, 98
- 値の検査, 165
- 値の消去, 72
- 値の選択, 392, 397, 401
- 分の増加
 - 時間区分ノード, 259-260
- 列の順序
 - テーブル ブラウザ, 442, 446
- 秒の増分
 - 時間区分ノード, 261-262
- 小数点以下のエクスポート, 172
- 集計用のキー値, 91
- 値のグループ化, 343
- 視覚化のコピー, 424
- 文字列のストレージ フォーマット, 36, 69
- 実数のストレージ フォーマット, 36, 69
- 整数のストレージ フォーマット, 36, 69
- 日付のストレージ フォーマット, 36, 69
- 時間のストレージ フォーマット, 36, 69
- 変数のタイプ
 - 視覚化, 281
- 列方向のバインド, 515
- 行方向のバインド, 515
- 数字の表示フォーマット, 172
- 通貨の表示フォーマット, 172
- 時間のフォーマット, 172
- 要素のマーク, 397, 401
- 一意のレコード, 113
- 平均比較ノード, 487
 - 重要度, 489
 - [出力] タブ, 444
 - 独立したグループ, 488
 - 一対のフィールド, 489
 - 出力ブラウザ, 491
- 時間区分ノード, 242, 244, 247
 - 概要, 241
- 条件抽出ノード
 - 概要, 77
 - Web グラフ リンクからの生成, 370
 - グラフからの生成, 402
- 精度分析ノード, 454
 - [出力] タブ, 444
 - [分析] タブ, 454
- 行列入替ノード, 236
 - フィールド名, 237
 - 数値型フィールド, 237
 - 文字列フィールド, 237
- 記述統計ノード, 482
 - 相関係数, 483
 - [出力] タブ, 444
 - [設定] タブ, 483
 - statistics, 483
 - 相関ラベル, 483
- 再構成ノード, 233, 235
 - レコード集計ノードと共に, 235
- 匿名化ノード
 - 概要, 202
 - オプションの設定, 203
 - 匿名化された値の作成, 205
- 固定長ノード
 - 自動日付認識, 35
 - 概要, 32
 - オプションの設定, 33
- 散布図ノード, 332
 - [外観] タブ, 339
 - [オプション] タブ, 337
 - グラフの使用, 340
 - [プロット] タブ, 334
- 時系列ノード, 264, 373
 - 概要, 263
 - [外観] タブ, 377
 - グラフの使用, 378
 - [プロット] タブ, 375
- 入力ノード
 - 概要, 9

- Enterprise View ノード, 10
- Excel 入力ノード, 58
- IBM Cognos BI 入力ノード, 49, 54-56
- SAS 入力ノード, 56
- Statistics ファイル ノード, 544
- XML 入力ノード, 60
- データ型のインスタンス化, 74
- データベース入力ノード, 16
- 固定長ノード, 32
- 可変長ファイル ノード, 28
- ユーザー入力ノード, 65-66
- 出力ノード, 434, 443, 449, 454, 459, 482, 493, 497, 555
 - [出力] タブ, 444
 - Web に公開, 437
- 変換ノード, 476
- 置換ノード
 - 概要, 199
- 評価ノード, 378
 - [外観] タブ, 387
 - [オプション] タブ, 385
 - グラフの使用, 390
 - スコア式, 385
 - 結果の読み込み, 388
 - ビジネス ルール, 385
 - ヒット条件, 385
 - [プロット] タブ, 383
- ノードのカプセル化, 568
- ノードのプロパティ, 583

- パス グラフ, 288
- パネル, 271, 273
 - 視覚化, 273
- パネル グラフ オーバーレイ, 271, 273
- パフォーマンス評価統計量, 454
- バブル プロット, 289
- バランス係数, 89
- バランス ノード
 - 概要, 88
 - オプションの設定, 89
 - グラフからの生成, 402
- バルク ロード, 515, 517
- パレット
 - 非表示, 407
 - 移動, 407
 - 表示する, 407

- 箱ひげ図, 291
- ビジネス ルール
 - 評価グラフ ノードのオプション, 385
- ヒストグラム, 290
 - 3-D, 290
- ヒストグラム ノード, 346
 - [外観] タブ, 349
 - グラフの使用, 350
 - [プロット] タブ, 347-348
- ヒット
 - 評価グラフ ノードのオプション, 385
- ピリオド, 170
- 六角ビン分割散布図, 289
- ビン分割散布図, 289
 - 六角ビン, 289
- ヒート マップ, 291

- 出力ファイル
 - 保存, 444
- ファイル エクスポート ノード, 526
 - [エクスポート] タブ, 527
- 可変長ファイル ノード, 28
 - 自動日付認識, 32
 - オプションの設定, 30
- ファセット化
 - 視覚化, 273
- フィルタ ノード
 - 概要, 173
 - オプションの設定, 174
 - 多重回答グループ, 177
- フィーチャ
 - マップ内, 323
- フィールド
 - フィールドおよび値ラベル, 72, 150
 - 区切り文字, 32
 - フィールド属性, 169
 - フィールド名, 176
 - 匿名化, 177
 - データのエクスポート, 502, 527, 535, 562
- 複数フィールド
 - 選択, 190
- フィールド作成 CLEM 式, 191
- フィールド タイプ, 72, 150
 - 視覚化, 281
- 可変長フィールド テキスト データ, 28
- 固定長フィールド テキスト データ, 32
- フィールド名の匿名化, 177
- フィールドの記憶域
 - 変換, 198
- フィールド値の置換, 199
- フィールド名の短縮, 174, 176
- フィールドの方向, 72, 150, 167
- 未使用フィールドの除外
 - 自動データ準備, 124
- 複数フィールドの作成, 188
- フィールドの関連付け, 504
- フィールドのフィルタリング, 105, 173
- IBM SPSS Statistics の, 563

索引

- フィールドのマッピング
 - 削除, 320
 - 場所, 318
 - Graphboard Template Chooser での選択, 285
 - インポート, 320
 - エクスポート, 320
 - 名前の変更, 320
- フィールド生成ノード
 - 度数, 196
 - 概要, 185
 - 条件式 (If-Then), 197
 - formula, 191
 - set, 193
 - Web グラフ リンクからの生成, 370
 - オプションの設定, 186
 - グラフからの生成, 402
 - ステート, 195
 - 自動データ準備からの生成, 148
 - データ分割ノードからの生成, 223
 - 値の再割り当て, 198
 - ビンから生成, 212
 - フィールド ストレージを変換する, 198
 - 複数フィールドの作成, 188
 - フラグ型, 191
- フィールド設定ノード, 118
 - データ検査から生成, 475
- フィールド順序ノード, 265
 - オプションの設定, 265
 - 自動ソート, 267
 - ユーザー指定の順序, 265
- 科学的表示フォーマット, 172
- 出力フォーマット, 444
- フォーマット ファイル, 57
- 欠損値検査ブラウザ
 - 条件抽出ノードの生成, 474
 - フィルタ ノードの生成, 473
- 精度分析ブラウザ
 - 解釈, 457
- 記述統計ブラウザ
 - 解釈, 485
 - [ノードの生成] メニュー, 485
 - フィルタ ノードの生成, 486
- フラグ型, 153, 164
- フラグ型データ, 154
- フラグの生成, 233, 235
- フラグ設定ノード, 232-233
- フラット ファイル, 28
- プリセット値、データベース接続, 22
- 連関プロット, 361
- プロパティ
 - node, 583
 - フィールド用, 170
- プロフィット グラフ, 378, 388
- フロー マッピング, 293
- 並べ替え
 - fields, 265
 - records, 96
 - あらかじめソートされたフィールド, 97
- ベスト ライン
 - 評価グラフ ノードのオプション, 383
- ヘルパー アプリケーション, 499
- ベースライン
 - 評価グラフ ノードのオプション, 383
- ポイント プロット, 332, 356
- ホールドアウト
 - 時系列モデル作成, 247
- マッピング
 - 平滑化, 324, 326
 - 投影法, 330
 - 配布, 331
 - ESRI シェープファイルの変換, 322
 - 矢印あり, 293
 - 点あり, 292-293
 - カラー, 291-292
 - 円グラフ, 292-293
 - 棒グラフ, 292
 - シニング, 324, 326
 - 重ね書き, 293-294
 - フィーチャ ラベル, 327
 - フィーチャの削除, 329
 - フィーチャの異動, 329
 - フィーチャの結合, 328
 - 折れ線グラフ, 293
 - 各要素を削除, 330
 - 座標マッピング, 292-293
- マッピング シェープファイル
 - Graphboard Template Chooser での仕様, 322
 - コンセプト, 323
 - タイプ, 323
 - プリインストールされた SMZ マッピングの編集, 322
- マッピング変換ユーティリティ, 322, 324
- マネージャ
 - [出力] タブ, 435
 - 出力マネージャ, 435
- 重み付けされたサンプル, 84
- 積み重ね, 423

- メイン データ セット, 112
- メタデータ, 72, 150, 159
 - Data Collection 入力ノード, 39-40
- 初めの n 件関数
 - 時系列の集計, 246
- メンバー (SAS インポート)
 - 設定, 57

- 最も最近関数
 - 時系列の埋め込み, 246
- モデル
 - データの匿名化, 202
- モデル オプション
 - Statistics モデル ノード, 552
- モデル ビュー
 - 自動データ準備, 134
- モデル作成の役割
 - フィールドの指定, 72, 150, 167
- モデルの評価, 378, 454
- モデル内の使用のためのデータの隠匿, 202
- 最も最近の平均関数
 - 時系列の埋め込み, 246

- ユーザー欠損値, 470
 - クロス集計テーブルで, 450
- ユーザー入力ノード
 - 概要, 65
 - オプションの設定, 66

- ラベル, 164
 - 視覚化, 271
 - 設定, 72, 150, 159, 162-163, 165
 - インポート, 58, 545
 - エクスポート, 535, 562
- 変数ラベル
 - Statistics エクスポート ノード, 561
 - Statistics ファイル ノード, 544
- 値ラベル
 - Statistics ファイル ノード, 544
- ラベル データ型
 - IBM SPSS Data Collection 入力ノード, 44
- ラベル フィールド
 - 出力のラベリング レコード, 167
- ランダム シード値
 - レコードのサンプリング, 84, 231
- ランダム シードの設定
 - レコードのサンプリング, 84, 231

- 区切り文字, 30, 32, 515
- 区切りテキスト データ, 28
- 偏りのないデータ, 88
- リフト グラフ, 378, 388

- リンク
 - Web グラフ ノード, 365

- 折れ線グラフ, 288, 332, 356
 - マップ上, 293
- レコード
 - 結合, 98
- 重複レコード ノード
 - 最適化設定, 116
 - 概要, 113
 - レコードの並べ替え, 115
- レコードの平均値, 90
- レコードの連結, 111
- レコードの集計, 233
- レコード操作ノード
 - 時間区分ノード, 241
- レコード結合ノード, 99
 - 最適化設定, 109
 - 概要, 98
 - オプションの設定, 102, 104
 - フィールドのタグ付け, 107
 - フィールドのフィルタリング, 105
- レコード設定ノード, 76
- レコード追加ノード
 - 概要, 111
 - オプションの設定, 112
 - フィールドの一致, 112
 - フィールドのタグ付け, 107
- レコード集計ノード
 - 並行処理, 94
 - 概要, 90
 - performance, 94
 - オプションの設定, 91
- レスポンス グラフ, 378, 388
- 測定レベル, 72, 150
 - 視覚化, 281
 - 定義, 153
 - 視覚化の変更, 279
- 測定レベルの変換, 156
- 品質レポート
 - データ検査ブラウザ, 470
- レポート ノード, 493
 - [出力] タブ, 444
 - [テンプレート] タブ, 494
- レポート ブラウザ, 496

- ワークシート
 - Excel からのインポート, 58

- 連関をプロットする, 361