

**IBM SPSS Modeler 16**  
**ユーザー・ガイド**

**IBM**

**お願い**

本書および本書で紹介する製品をご使用になる前に、183 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM(r) SPSS(r) Modeler バージョン 16 リリース 0 モディフィケーション 0、および新しい版で明記されていない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

**原典：** IBM SPSS Modeler 16 User's Guide

**発行：** 日本アイ・ビー・エム株式会社

**担当：** トランスレーション・サービス・センター

# 目次

## 第 1 章 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 . . . . .	2
IBM SPSS Modeler Solution Publisher . . . . .	2
IBM SPSS Modeler Server の IBM SPSS Collaboration and Deployment Services . . . . .	2
IBM SPSS Modeler エディション . . . . .	3
IBM SPSS Modeler のドキュメント . . . . .	3
SPSS Modeler Professional のドキュメント . . . . .	3
SPSS Modeler Premium のドキュメント . . . . .	4
アプリケーションの例 . . . . .	5
Demos フォルダー . . . . .	5

## 第 2 章 IBM SPSS Modeler 16 の新機能 7

## 第 3 章 IBM SPSS Modeler の概要 . . . 11

はじめに . . . . .	11
IBM SPSS Modeler の開始 . . . . .	11
コマンド・ラインからの起動 . . . . .	12
IBM SPSS Modeler Server に接続中 . . . . .	12
一時ディレクトリの変更 . . . . .	14
複数の IBM SPSS Modeler セッションの起動 . . . . .	15
IBM SPSS Modeler インターフェースについて . . . . .	15
IBM SPSS Modeler ストリーム領域 . . . . .	16
ノード・パレット(N) . . . . .	17
IBM SPSS Modeler マネージャ . . . . .	18
IBM SPSS Modeler プロジェクト . . . . .	19
IBM SPSS Modeler ツールバー . . . . .	20
ツールバーのカスタマイズ . . . . .	21
IBM SPSS Modeler ウィンドウのカスタマイズ ストリームのアイコン・サイズの変更 . . . . .	22
IBM SPSS Modeler でのマウスの使用方法 . . . . .	23
ショートカット・キーの使用 . . . . .	23
印刷中 . . . . .	24
IBM SPSS Modeler の自動化 . . . . .	25

## 第 4 章 データ・マイニングについて . . . 27

データ・マイニングの概要 . . . . .	27
データへの評価 . . . . .	28
データ・マイニングの戦略的方法 . . . . .	29
CRISP-DM プロセス・モデル . . . . .	30
モデルの種類 . . . . .	31
データ・マイニングの例 . . . . .	36

## 第 5 章 ストリームの構築 . . . . . 37

ストリーム構築の概要 . . . . .	37
データ・ストリームの作成 . . . . .	37
ノードの作業 . . . . .	38

ストリームの作業 . . . . .	43
ストリームの説明 . . . . .	55
ストリームの実行 . . . . .	57
モデルの処理 . . . . .	57
ノードおよびストリームへのコメントおよび注釈 の追加 . . . . .	58
データ・ストリームの保存 . . . . .	64
ファイルのロード . . . . .	65
データ・ストリームのマッピング . . . . .	66
ヒントとショートカット . . . . .	68

## 第 6 章 欠損値の処理 . . . . . 71

欠損値の概要 . . . . .	71
欠損値の処理 . . . . .	72
欠損値を含むレコードの処理 . . . . .	72
欠損値を含むフィールドの処理 . . . . .	72
欠損値の代入または置換 . . . . .	73
欠損値用 CLEM 関数 . . . . .	73

## 第 7 章 CLEM 式の作成 . . . . . 75

CLEM について . . . . .	75
CLEM の例 . . . . .	75
値とデータ型 . . . . .	77
式と条件 . . . . .	78
ストリーム、セッション、およびスーパーノード・ パラメーター . . . . .	78
文字列の処理 . . . . .	79
空白および欠損値の処理 . . . . .	80
数値の処理 . . . . .	80
時間と日付の処理 . . . . .	80
複数フィールドの要約 . . . . .	81
複数回答データの処理 . . . . .	82
Clem 式ビルダー . . . . .	83
Clem 式ビルダーへのアクセス . . . . .	83
CLEM 式の作成 . . . . .	83
関数の選択 . . . . .	83
フィールド、パラメーター、およびグローバル変 数の選択 . . . . .	84
値の表示または選択 . . . . .	85
CLEM 式の検査 . . . . .	85
検索と置換 . . . . .	85

## 第 8 章 CLEM 言語リファレンス . . . . . 89

CLEM リファレンス概要 . . . . .	89
CLEM データ型 . . . . .	89
整数値 . . . . .	89
実数 . . . . .	90
文字 . . . . .	90
文字列 . . . . .	90
リスト . . . . .	90
フィールド . . . . .	91

日付	91
時間	92
CLEM 演算子	92
関数のリファレンス	94
関数の表記方法について	95
情報関数	96
変換関数	97
比較関数	97
論理関数	99
数値関数	100
三角関数	101
確率関数	101
ビット単位の整数演算	102
乱数関数	103
文字列関数	103
SoundEx 関数	108
日付および時刻の関数	108
シーケンス関数	112
グローバル関数	117
空白値とヌル値処理関数	118
特殊フィールド	119

## 第 9 章 リポジトリでの IBM SPSS Modeler の使用 . . . . . 121

IBM SPSS Collaboration and Deployment Services Repository について	121
リポジトリ・オブジェクトの保存と展開	122
リポジトリへの接続	123
リポジトリへの資格情報の入力	123
リポジトリ・コンテンツを参照	123
リポジトリのオブジェクトの保存	124
オブジェクトの特性の設定	124
ストリームの保存	126
プロジェクトの格納	127
ノードの格納	127
出力オブジェクトの格納	128
モデルおよびモデル・パレットの格納	128
リポジトリからのオブジェクトの取得	129
取得するオブジェクトの選択	129
オブジェクトのバージョンの選択	130
リポジトリのオブジェクトの検索	130
リポジトリのオブジェクトを変更	131
フォルダーの編集、名前の変更、削除	131
リポジトリ・オブジェクトのロックおよびロック解除	132
リポジトリのオブジェクトを削除	132
リポジトリ・オブジェクトのプロパティの管理	133
フォルダーのプロパティの表示	133
オブジェクトのプロパティの表示および編集	133
オブジェクトのバージョン・ラベルの管理	135
ストリームの展開	135
ストリーム展開のオプション	136
スコアリングブランチ	138

## 第 10 章 外部アプリケーションへのエクスポート . . . . . 141

外部アプリケーションへのエクスポートについて	141
IBM SPSS Modeler Advantage でストリームを開く	141
PMML としてのモデルのインポートおよびエクスポート	142
PMML をサポートするモデルの種類	143

## 第 11 章 プロジェクトとレポート . . . . . 145

プロジェクトの概要	145
CRISP-DM ビュー	145
クラス・ビュー	146
プロジェクトの作成	146
新規プロジェクトの作成	146
プロジェクトへの追加	147
プロジェクト IBM SPSS Collaboration and Deployment Services Repository への転送	148
プロジェクトのプロパティの設定	148
プロジェクトの注釈	149
オブジェクトのプロパティ	149
プロジェクトを閉じる	150
レポートの生成	150
生成されたレポートの保存とエクスポート	152

## 第 12 章 IBM SPSS Modeler のカスタマイズ . . . . . 153

IBM SPSS Modeler オプションのカスタマイズ	153
IBM SPSS Modeler オプションの設定	153
システム・オプション	153
デフォルト・ディレクトリの設定	154
「ユーザー・オプション」ダイアログ・ボックスの設定	154
ノード・パレットのカスタマイズ	158
パレット・マネージャのカスタマイズ	158
パレット・タブの表示の変更	161
CEMI ノードの管理	161

## 第 13 章 ストリームおよびノードのパフォーマンスの考察 . . . . . 163

ノードの順序	163
ノードのキャッシュ	164
パフォーマンス : プロセス・ノード	165
パフォーマンス : モデル作成ノード	166
パフォーマンス : CLEM 式	167

## 第 14 章 IBM SPSS Modeler のアクセシビリティ . . . . . 169

IBM SPSS Modeler のアクセシビリティの概要	169
アクセシビリティサポートの種類	169
視力の弱いユーザーのためのアクセシビリティ	169
視覚障害ユーザーのためのアクセシビリティ	170
キーボード・アクセシビリティ	171
画面読み上げソフトウェアの使用	178
使用上のヒント	179
他のソフトウェアとの相性	180

JAWS および Java. . . . .	180	特記事項. . . . .	183
IBM SPSS Modeler でのグラフの使用 . . . . .	180	商標 . . . . .	184
<b>第 15 章 Unicode のサポート . . . . .</b>	<b>181</b>	索引 . . . . .	<b>185</b>
IBM SPSS Modeler 内での Unicode のサポート	181		



---

## 第 1 章 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。IBM 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。

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。「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。詳しくは、「<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 ファミリー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、コンピューターにインストールし、そのコンピューターで実行します。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・インターフェースを使用すると、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。詳しくは、トピック 3 ページの『IBM SPSS Modeler エディション』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・アーキテクチャーを使用し、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ー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・インターフェー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 コンピューター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に、コマンドラインから SPSS Modeler を実行することも可能です。例えば、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・インターフェー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し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 アダプターを使用すると、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、リポジトリに展開された 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ー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** により、IBM SPSS Modeler の予測分析に新たな次元が追加されます。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。

**IBM SPSS Modeler Text Analytics** 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ー別に分類します。抽出されたコンセプトとカテゴリーを、人口統計のような既存の構造化データと組み合わせ、IBM SPSS Modeler の豊富なデータ・マイニング・ツールを適用する方法で、焦点を絞ったより良い決定を下すことができます。

---

## IBM SPSS Modeler のドキュメント

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* フォルダーにもあります。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=swg27038316> 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/v16r0m0/> から入手できます。

## SPSS Modeler Professional のドキュメント

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 ユーザーズ・ガイド:** SPSS Modeler の使用方法への全体的な入門です。データ・ストリームの作成方法、欠損値の処理方法、CLEM 式の作成方法、プロジェクトとレポートでの作業方法、および IBM SPSS Collaboration and Deployment Services、Predictive Applications、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**」。各種形式のデータの読み取り、処理、および出力に使用するすべてのノードの説明です。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についての説明です。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。
- **IBM SPSS Modeler アルゴリズム・ガイド:** IBM 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**」。このガイドの例では、特定のモデル作成手法および技法について、簡単に対象を絞って紹介します。本ガイドのオンライン バージョンは、「ヘルプ」メニューからも利用できます。詳しくは、トピック 5 ページの『アプリケーションの例』を参照してください。
- 「**IBM SPSS Modeler スクリプトとオートメーション**」。スクリプトによるシステムの自動化に関する情報です。ノードおよびストリームの操作に使用できるプロパティを含めて説明します。
- **IBM SPSS Modeler 展開ガイド:** IBM 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、IBM 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・パーティー・プログラムを統合できます。
- 「**IBM SPSS Modeler データベース内 マイニング・ガイド**」。サード・パーティー製アルゴリズムを使用してご使用のデータベースの能力を利用してパフォーマンスを向上させ、分析機能の範囲を拡張する方法に関する情報を示します。
- **IBM SPSS Modeler Server 管理およびパフォーマンス・ガイド:** IBM SPSS Modeler Server の設定および管理の方法についての情報があります。
- **IBM SPSS Modeler Administration Console ユーザー・ガイド:** IBM SPSS Modeler Server をモニターおよび設定するためのコンソール・ユーザー・インターフェースのインストールおよび使用についての情報があります。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 CRISP-DM ガイド**」。SPSS Modeler でのデータ・マイニングに対する CRISP-DM 方法の使用に関するステップバイステップのガイドです。
- 「**IBM SPSS Modeler Batch ユーザーズ・ガイド**」。IBM SPSS Modeler をバッチ・モードで使用するための完全ガイドで、バッチ・モードでの実行およびコマンド・ライン引数の詳細について説明します。このガイドは、PDF 形式のみです。

## SPSS Modeler Premium のドキュメント

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**」。SPSS Modeler でエンティティ分析を使用する場合の情報。リポジトリのインストールと構成、エンティティ分析ノード、および管理用タスクについて説明します。
- 「**IBM SPSS Modeler Social Network Analysis ユーザー・ガイド**」。SPSS Modeler でソーシャル・ネットワーク分析を行うためのガイド。グループ分析、拡散分析などについて説明します。

- 「**SPSS Modeler Text Analytics ユーザーズ・ガイド**」。SPSS Modeler でテキスト分析を使用する場合の情報。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、テンプレートなどについて説明します。
- **IBM SPSS Modeler Text Analytics Administration Console ユーザー・ガイド**：SPSS Modeler Text Analytics とともに使用される IBM SPSS Modeler Server をモニターおよび設定するためのコンソール・ユーザー・インターフェー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・ディレクトリーの *Demos* フォルダーにインストールされています。詳しくは、トピック『*Demos* フォルダー』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 フォルダー

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・ディレクトリーの *Demos* フォルダーにインストールされています。このフォルダーには、Windows 「スタート」メニューの IBM SPSS Modeler 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ーの一覧から「*Demos*」を選択してアクセスすることもできます。



---

## 第 2 章 IBM SPSS Modeler 16 の新機能

IBM SPSS Modeler は、本リリースで以下の機能を追加しています。

**モンテカルロ・シミュレーション。** 新規のシミュレーション入力ノードであり、幅広い統計分布を使用して最初から合成データを作成する簡単な方法を提供します。また、この新規の適合ノードは、履歴変数の分布や履歴変数の間の関係を反映して事前構成された入力ノードを自動的に構築できます。シミュレーション評価ノードは、シミュレーションされた分析ストリームから得られるフィールドを評価するように設計されたターミナル・ノードであり、有用な分布図や相関図を提供します。

**Python スクリプト。** ユーザー・インターフェースのプロセスの自動化に使用される IBM SPSS Modeler のスクリプトであり、古いスクリプト・モードのサポート継続に加えて Python 言語の使用が可能になりました。Python は、豊富で簡潔なシンタックス、エラー処理、そして強力なアドオン・モジュールを含む多数の機能を備えた、有名で人気のある言語です。注: 「ツール」 > 「ストリームのプロパティ」の「スクリプト」タブは、現在「実行」という名前になっています。

**ループおよび条件式実行。** これらの新規オプションを使用すると、スクリプトをコーディングしなくとも、ストリームの簡単なループおよび条件式実行が可能になります。これらの新規オプションは、「ツール」 > 「ストリームのプロパティ」 > 「実行」をクリックするか、またはストリーム内のノードを右クリックして「ループ/条件付き実行」オプションを選択することで見つかります。

**スペース-時間-ボックス (STB) ノード。** より高度な分析をサポートするために、場所とタイムスタンプのデータのピンを作成します。ハンガアウト・モードの場合、このノードでは、エンティティーが存在する場所と時間も識別されます。追加の式ビルダーの関数では、ジオハッシュおよび STB の重心の抽出がサポートされます。

**Entity Analytics の機能強化。** 個別のエンティティーの解決に加えて、エンティティー間の  $n$  次数の関係を識別できるようになりました。ストリーミング・ノードによるリアルタイムの更新、解決済みエンティティーのフラット化 (重複レコード・ノード)、およびエンティティー・リポジトリに送られるデータの匿名化に対する追加のサポートが提供されています。注: 以前のローカルの SolidDB データベースは DB2 に置き換えられています。

**新規の受信者操作特性 (ROC) 評価ノードのグラフの種類、および分析ノードの曲線下の領域 (AUC) および Gini メトリック。** 2 値の対象をサポートします。

**複合レコードを作成するための新規の重複レコード・ノード・オプション。** これにより、重複レコードを破棄するのではなく、グループ化される各フィールドに対する集計の方法を指定できます (最初の値、最後の値、値の連結、など)。レコード集計ノードは、通常、データを上位レベルに要約するために使用しますが、この新規オプションは、重複 (エンティティーの解決によって識別されたものなど) のフラット化のために使用します。

**TM1 入力ノードおよびエクスポート・ノード。** TM1 入力ノードによって TM1 Cube Views にアクセスすることや、TM1 エクスポート・ノードを使用してデータをスコアリングして既存の TM1 キューブに戻すことができます。

**集計式およびウィンドウ集計関数。** 組み込み集計関数 (平均、合計、など) やデータベース集計ユーザー定義関数を組み入れた、レコード集計ノードのカスタム集計式を作成できます。派生式では、ウィンドウ集計関数 (移動平均など) を必要とするフィールドを派生できます。組み込みおよびデータベース提供のウィンドウ集計関数が使用可能です。

**IBM Netezza Analytics のデータベース内マイニングの機能強化。** 新規の Netezza Two Step アルゴリズム、Netezza Analytic モデルを管理 (削除、名前変更、など) するための追加の「ヘルパー・アプリケーション」オプション、回帰ツリー、デシジョン・ツリー、Kmeans、および TwoStep 用のモデル・ビューアーに対するサポート。注: 新機能には INZA 3.0 が必要です。

**新規の R ノードおよび R 用のカスタム・ダイアログ・ビルダー。** SPSS Modeler 15 フィックスパック 2 で導入された R モデル構築ノードおよびモデル・ナゲットに加え、本リリースでは、R プロセスおよび R 出力という 2 つの新規ノードが追加されました。R プロセス・ノードによって、SPSS Modeler ストリームからデータを取得し、R スクリプトを使用してそのデータに変換を適用することができます。R 出力ノードによって、独自のカスタム R スクリプトを使用してデータ分析を実行したり、モデル・スコアリングの結果を要約したりすることができます。

分析のテキスト出力やグラフィック出力を生成できます。この出力は、ファイルに送信したり、R 出力ノードの出力ブラウザで表示したりできます。カスタム・ダイアログ・ビルダーを使用すると、R プログラムの抽象化やパラメーター化を可能にするためのカスタム・モデル構築、プロセス、出力ノード・タイプおよびモデル・ナゲット (フィールド・ピッカー、テキスト、番号、ラジオ・ボタン、サブダイアログ・コントロールを含む) を作成できます。ノードをインストールする前や、他の SPSS Modeler ユーザーとノードの使用を共有する前に、ノード・タイプ、宛先パレット、およびノード・アイコンを選択できます。カスタム・ダイアログ・ビルダーは「ツール」メニューから起動します。注: この機能を使用するには、SPSS Modeler - Essentials for R がインストールされている必要があります。

**データベース内の R。** R ノードに対して SQL プッシュバックがサポートされています。Netezza、SAP HANA、および Oracle データベースの場合、これらのデータベースの R サポートを活用することにより提供されます。注: データベースには、ベンダー提供の適切な R 拡張子がインストールされている必要があります。

**ストリーミング時系列プロセス・ノード。** IBM InfoSphere Streams、IBM SPSS Collaboration and Deployment Services スコアリング・サービス、または IBM SPSS Modeler Solution Publisher によるリアルタイムの展開を提供するための時系列モデルを単一ステップで作成およびスコアリングします。

「**プレビュー**」ボタン: 一部のモデル・ナゲットには「プレビュー」ボタンがあります。データベース・ソースと組み合わせて使用する場合、プレビューで SQL プッシュバックを視覚化できます。Analytic Server データ・ソースと組み合わせて使用する場合、「プレビュー」ボタンを使用して潜在的な大規模データの移動に関する情報を受け取ります。

**自動分類ノード、自動数値ノード、および自動クラスター・ノードの新規 Analytic Server オプション。** IBM SPSS Analytic Server に対してストリームを実行する場合、分割を有効 (分割モデル機能を使用している場合はこのオプションを使用します) にして実行するか、「特に大きなデータセット」オプション (分割は無視され、モデリングの目的は「ビッグ・データ」に設定されます) を使用して実行するかを選択できます。SPSS Modeler Server に組み込み可能な自動分類モデル、自動数値モデル、および自動クラスター・モデルに対するスコアリングがサポートされるようになりました。

**スコアリング・アダプター・サポートの強化。** データベース・スコアリング・アダプターによるテキスト・マイニング・モデルのスコアリングに対するサポート。加えて、DB2 LUW および Oracle 用の新規データベース・スコアリング・アダプター。

「データ・ビュー」入力ノード。「データ・ビュー」ノードにより、「分析データ・ビュー」からのデータを SPSS Modeler ストリームに読み込むことができます。「分析データ・ビュー」は、IBM SPSS Collaboration and Deployment Services 6 で統合データ・ビューを作成するための新しい方法です。

#### その他の機能強化。

- データベース接続プリセットの Teradata クエリー・バンド化。
- 更新された GLMM ノード・オプション。
- 暗号化および圧縮された .sav ファイルに対するサポート。
- IBM SPSS Collaboration and Deployment Services を必要としない SPSS Modeler Server のシングル・サインオン (SSO) に対するサポート。
- シングル・サインオンのデータベース層のサポート。
- zDB2 のサンプル・ノード用の SQL プッシュバック。
- R ノードは、SPSS Modeler の基本インストールの一環としてインストールされるようになりました。
- Modeler Adapters for IBM SPSS Collaboration and Deployment Services は IBM Installation Manager によってインストールされます。
- 分割モデルの更新。
- ロシア語のローカリゼーションに対するサポート。



## 第 3 章 IBM SPSS Modeler の概要

### はじめに

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

### IBM SPSS Modeler の開始

アプリケーションを起動するには、以下のメニューをクリックします。

「スタート」 > 「すべてのプログラム」 > 「IBM SPSS Modeler 16」 > 「IBM SPSS Modeler 16」

数秒後にメイン・ウィンドウが表示されます。

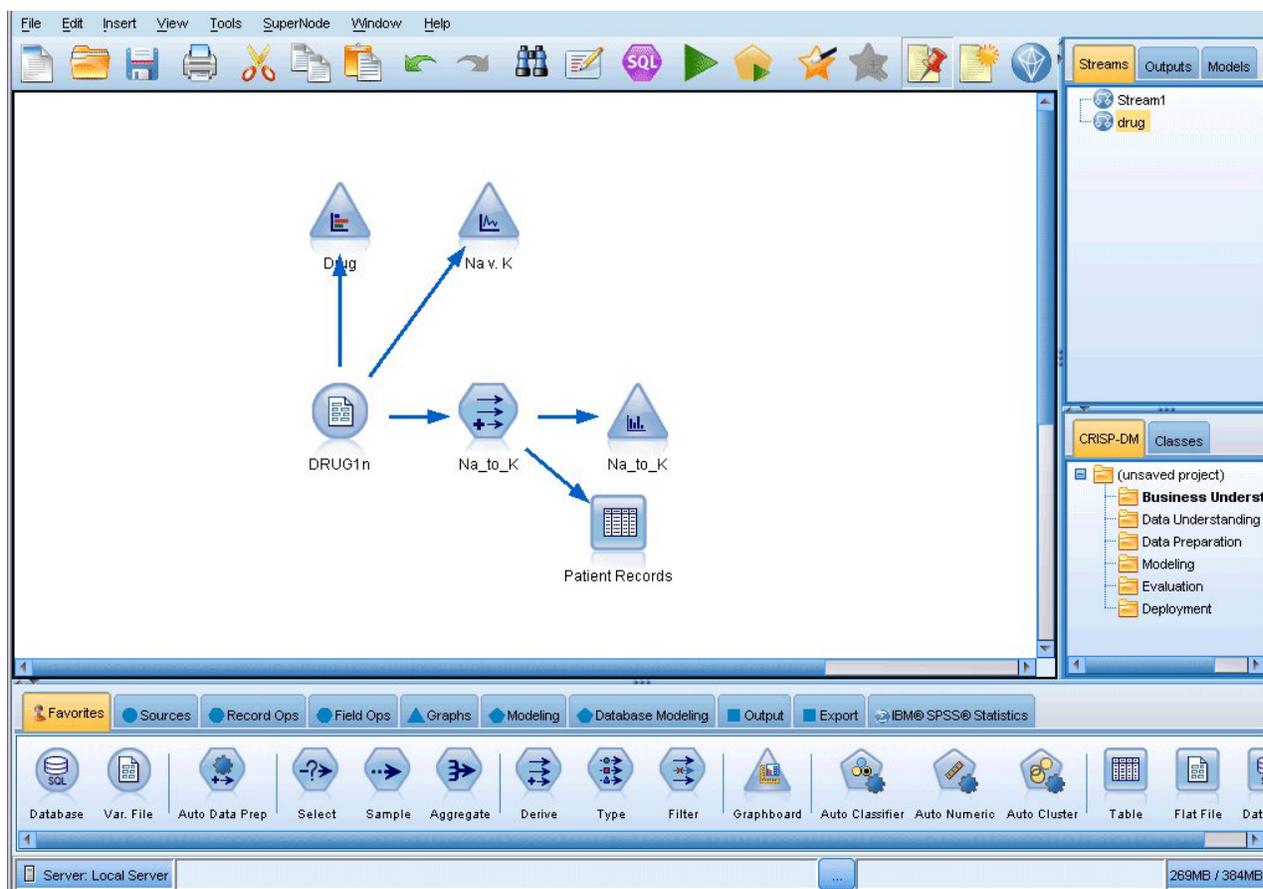


図 1. IBM SPSS Modeler のメイン・アプリケーション・ウィンドウ

## コマンド・ラインからの起動

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

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

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

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

## IBM SPSS Modeler Server に接続中

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

サーバーに接続する場合は、接続するサーバー名を手動で入力するか、以前定義した名前を選択できます。ただし、IBM SPSS Collaboration and Deployment Services を使用する場合、「サーバーへのログイン」ダイアログ・ボックスからサーバーまたはサーバー・クラスターのリストを使用して検索することができます。ネットワーク上で実行する Statistics サービスを介して参照する機能は、Coordinator of Processes で使用できます。

サーバーに接続するには

1. 「ツール」メニューの「**サーバーへのログイン**」をクリックします。「サーバーへのログイン」ダイアログ・ボックスが開きます。または、IBM SPSS Modeler ウィンドウの接続ステータス領域をダブルクリックします。
2. ダイアログ・ボックスで、ローカル・サーバーのコンピューターに接続するオプションを指定するか、テーブルから接続を選択します。
  - 「**追加**」または「**編集**」をクリックして、接続を追加または編集します。詳しくは、トピック 13 ページの『IBM SPSS Modeler Server 接続の追加および編集』を参照してください。
  - 「**検索**」をクリックして、SPSS COP のサーバーまたはサーバー・クラスターにアクセスします。詳しくは、トピック 14 ページの『IBM SPSS Collaboration and Deployment Services のサーバーの検索』を参照してください。

「**サーバー・テーブル**」。このテーブルには、一連の定義済みのサーバー接続が含まれています。このテーブルには、デフォルト接続、サーバー名、説明、ポート番号が表示されています。既存の接続を選択または検索、あるいは新しい接続を手動で追加することができます。特定のサーバーをデフォルト接続として設定するには、接続のテーブルの「デフォルト」列のチェック・ボックスを選択します。

「**デフォルト・データ・パス**」。サーバー・コンピューター上のデータに使用されるパスを指定します。「...」ボタンをクリックして、目的の場所を指定することもできます。

「**資格情報の設定**」。このボックスのチェック・マークを外した状態にして、**シングル・サインオン**機能を有効にします。これにより、ローカル・コンピューターのユーザー名およびパスワードの詳細を使用してサーバーへのログインを試みます。この でシングル・サインオンを使用できない場合、またはこ

のボックスをチェックしてシングル・サインオンを無効にした場合 (例えば、管理者アカウントにログインした場合)、資格情報を入力するための次のフィールドが表示されます。

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

「パスワード」。指定したユーザー名に関連付けられたパスワードを入力します。

「ドメイン」。サーバーにログオンするために使用するドメインを指定します。サーバー・コンピューターが クライアント・コンピューターとは異なる Windows ドメインにある場合にのみ、ドメイン名が必要です。

3. 「OK」をクリックして、接続を完了します。

サーバーとの接続を切断するには

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

## IBM SPSS Modeler Server 接続の追加および編集

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

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

サーバー接続を追加するには

1. 「ツール」メニューの「サーバーへのログイン」をクリックします。「サーバーへのログイン」ダイアログ・ボックスが開きます。
  2. ダイアログ・ボックスで、「追加」をクリックします。「サーバーへのログイン: サーバーの追加/編集」ダイアログ・ボックスが表示されます。
  3. サーバー接続の詳細を入力して「OK」をクリックします。接続が保存され、「サーバーへのログイン」ダイアログ・ボックスに戻ります。
- **サーバー。** 利用できるサーバーを指定するか、またはリストから選択します。サーバー・コンピューターは、英数字の名前 (例えば、*myserver*)、または、サーバー・コンピューターに割り当てられた IP アドレス (例えば、202.123.456.78) で識別できます。
  - **ポート。** サーバーが待機しているポート番号を入力します。デフォルトのポート番号がうまく動作しない場合は、システム管理者に問い合わせ正しいポート番号を取得してください。
  - **説明。** サーバー接続の説明をオプションで入力します。
  - **「セキュア接続を確保 (SSL の使用)」。** SSL (Secure Sockets Layer) 接続を使用するかどうかを指定します。SSL は、ネットワークを介してセキュアなデータ送信を行うために一般的に使用されているプロトコルです。この機能を使用するには、IBM SPSS Modeler Server をホストするサーバー側で SSL を有効にする必要があります。必要な場合、詳細を各サイトの管理者に問い合わせてください。

サーバー接続を編集するには

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

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

サーバー接続を手動で入力する代わりに、IBM SPSS Collaboration and Deployment Services で使用できる Coordinator of Processes (COP) を介してネットワークで使用可能なサーバーまたはサーバー・クラスターを選択できます。サーバー・クラスターは、Coordinator of Processes が処理要求に応答するのに最適なサーバーを決定するサーバーのグループです。

「サーバーへのログイン」ダイアログ・ボックスで手動でサーバーを追加することもできますが、使用可能なサーバーの検索では、正しいサーバー名およびポート番号を把握していなくてもサーバーに接続できます。この情報は自動的に提供されます。この情報は、自動的に提供されます。ただし、ユーザー名、ドメインおよびパスワードなどの、正しいログオン情報が必要です。

注：Coordinator of Processes 機能へアクセスしていない場合、接続するサーバー名を手動で入力したり、以前定義した名前を選択することができます。詳しくは、トピック 13 ページの『IBM SPSS Modeler Server 接続の追加および編集』を参照してください。

サーバーおよびクラスターを検索するには

1. 「ツール」メニューの「**サーバーへのログイン**」をクリックします。「サーバーへのログイン」ダイアログ・ボックスが開きます。
2. ダイアログ・ボックスで「**検索**」をクリックすると、「サーバーの検索」ダイアログ・ボックスが表示されます。Coordinator of Processes をブラウズしようとする際に IBM SPSS Collaboration and Deployment Services にログオンしていない場合、ログオンを指示するメッセージが表示されます。
3. リストからサーバーまたはサーバー・クラスターを選択します。
4. 「**OK**」をクリックしてダイアログ・ボックスが閉じられ、選択した接続が「サーバーへのログイン」ダイアログ・ボックスのテーブルに追加されます。

## 一時ディレクトリーの変更

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

1. 新規ディレクトリー *spss* およびそのサブディレクトリー *servertemp* を作成します。
2. IBM SPSS Modeler のインストール・ディレクトリー中の */config* にある、*options.cfg* を編集します。次のファイルの *temp\_directory* パラメーターを編集して読み込みます。*temp\_directory*,  
"C:/spss/servertemp"
3. この作業を行った後は、IBM SPSS Modeler Server サーバー・サービスを再起動する必要があります。再起動するには、Windows コントロール・パネルの「管理ツール」にある「**サービス**」タブをクリックしてください。サービスを停止した後、再び開始すると変更内容が有効になります。また、マシンを再起動しても、サービスが再開されます。

これで、新しいディレクトリーに一時ファイルが作成されるようになります。

注: この作業を行う際によくあるエラーは、間違った種類のスラッシュを使用してしまうことです。普通のスラッシュを使用してください。

## 複数の IBM SPSS Modeler セッションの起動

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

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

1. 以下のメニューをクリックします。

「スタート」 > 「すべてのプログラム」 > 「IBM SPSS Modeler 16」

2. IBM SPSS Modeler 16 のショートカット (アイコンで表示) で、右クリックして「プロパティ」を選択します。

3. 「対象」テキスト・ボックスで、文字列の終わりに `-noshare` を追加します。

4. Windows の Explorer で、次の項目を選択します。

「ツール」 > 「フォルダー・オプション...」

5. 「ファイル タイプ」タブで、「IBM SPSS Modeler ストリーム」オプションを選択し、「詳細」をクリックします。

6. 「ファイル タイプの編集」ダイアログ・ボックスで、「IBM SPSS Modeler で開く」を選択し、「編集」をクリックします。

7. 「アクションの実行に使用するアプリケーション」テキスト・ボックスで、`-stream` 引数の前に `-noshare` を追加します。

---

## IBM SPSS Modeler インターフェースについて

IBM SPSS Modeler の使いやすいインターフェースでは、データ・マイニング処理の各ポイントで、特定ビジネスの専門知識が必要となります。予測、分類、セグメント化、関連性検出などのモデリング・アルゴリズムによって、強力かつ正確なモデルが作成されます。モデルの結果は簡単に展開でき、データベース、IBM SPSS Statistics、およびさまざまなアプリケーションで使用することができます。

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

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

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

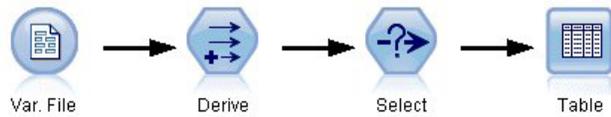


図2. 単純なストリーム

## IBM SPSS Modeler ストリーム領域

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

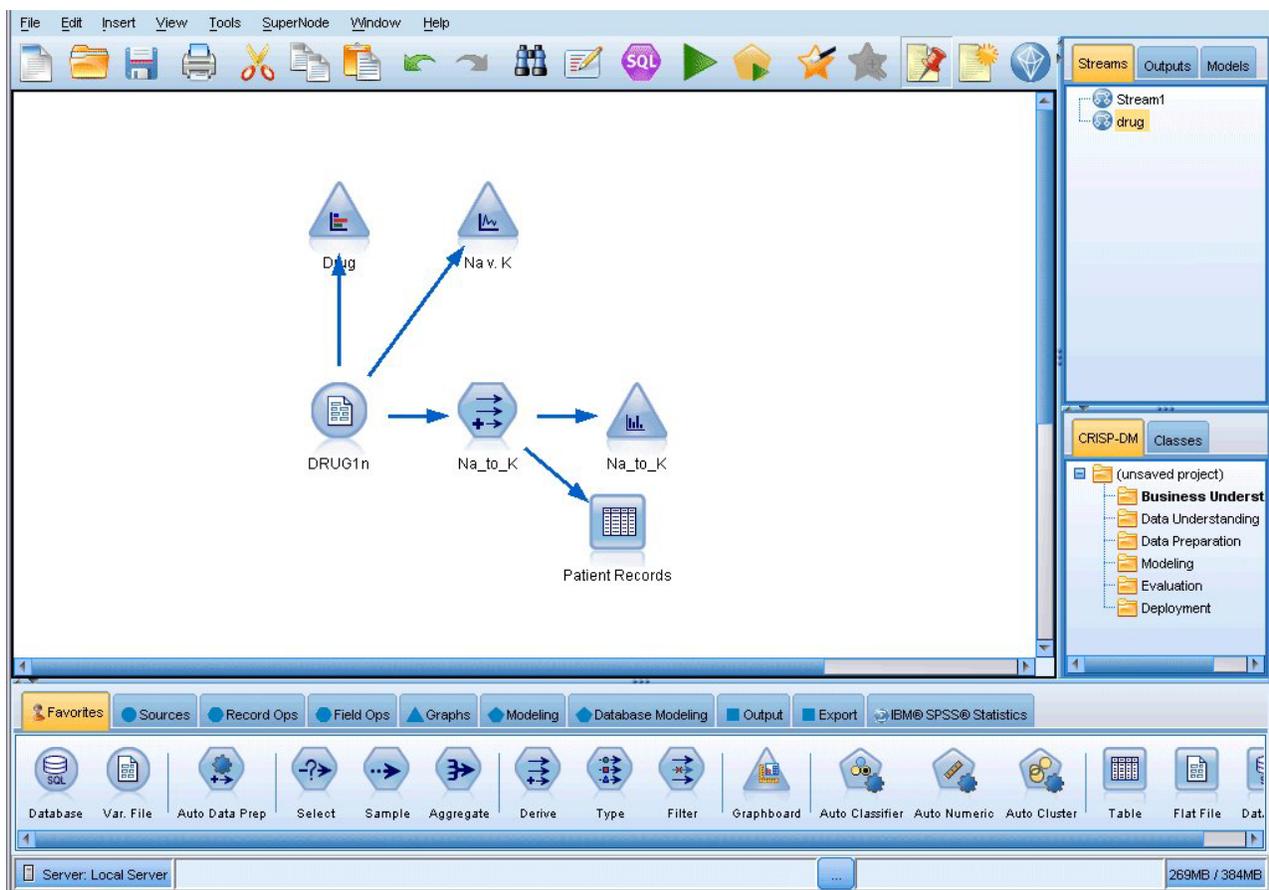


図3. IBM SPSS Modeler 作業領域 (デフォルトのビュー)

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

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

## ノード・パレット(N)

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

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

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

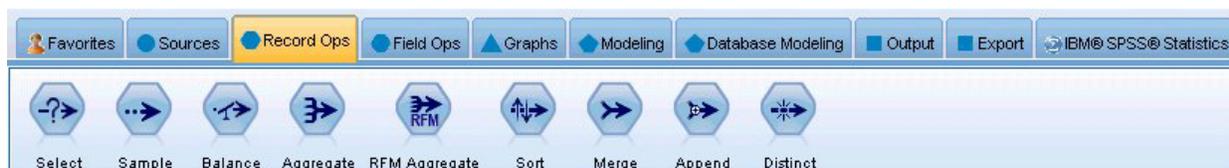


図4. ノード・パレットの「レコード設定」タブ

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

- 「ソース」。IBM SPSS Modeler にデータを入力するノード。
- 「レコード操作」。選択、結合、追加など、データレコードの操作を実行するノード。
- 「フィールド操作」。フィルタリング、新規フィールドの派生、特定フィールドの測定の尺度の特定など、データフィールドの操作を実行するノード。
- 「グラフ」。モデルの作成前後にデータをグラフィカル表示するノード。グラフには、散布図、ヒストグラム、Web グラフ・ノード、および評価グラフなどがあります。
- **モデリング**。ニューラル・ネットワーク、ディジション・ツリー、クラスタリング・アルゴリズム、およびデータのシーケンス化などの、IBM SPSS Modeler で利用できる優れたモデル作成アルゴリズムを表すノードです。
- **データベース・モデル作成**：Microsoft SQL Server、IBM DB2、および Oracle データベースと Netezza データベースで利用できるモデル作成アルゴリズムを表すノードです。
- **出力**：IBM SPSS Modeler で表示できるさまざまなデータ、グラフ、モデルの結果の出力を作成します。
- **エクスポート**：IBM SPSS Data Collection または Excel などの外部アプリケーションで表示できるさまざまな出力を作成します。
- **IBM SPSS Statistics**。IBM SPSS Statistics 手続きを実行するほか、IBM SPSS Statistics との間でデータのインポートまたはエクスポートを行います。

IBM SPSS Modeler に慣れてきたら、パレットの内容を使いやすくカスタマイズすることができます。

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

## IBM SPSS Modeler マネージャ

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

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



図5. 「ストリーム」タブ



図6. 「出力」タブ

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

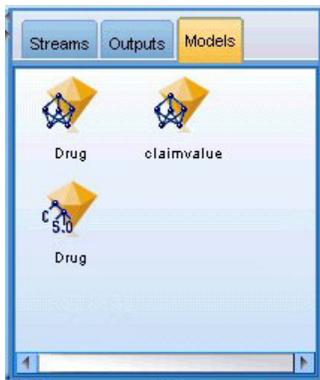


図7. モデル・ナゲットを含む「モデル」タブ

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

## IBM SPSS Modeler プロジェクト

ウィンドウの右下には、データ・マイニング・プロジェクト (データ・マイニング・タスクに関連するファイルのグループ) の作成と管理に使用するプロジェクト パネルがあります。IBM SPSS Modeler で作成したプロジェクトを表示するには、クラス・ビューと CRISP-DM ビューの 2 つの方法が使用できます。

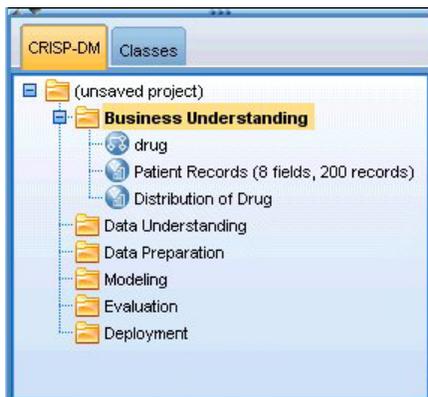


図8. CRISP-DM ビュー

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

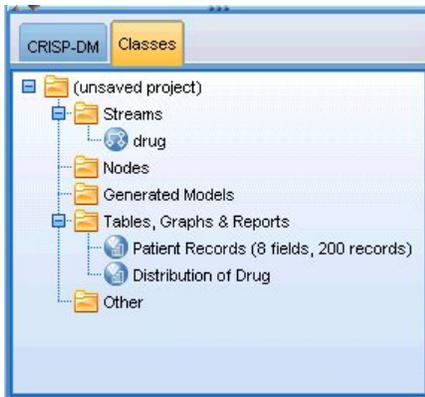


図9. クラス・ビュー

「クラス」タブでは、IBM SPSS Modeler で行った作業内容を、作成したオブジェクトの種類に応じて編成することができます。このビューは、データ、ストリーム、およびモデルの現状を把握するときに便利です。

## IBM SPSS Modeler ツールバー

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

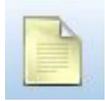
	新規ストリームの作成		ストリームを開く
	ストリームの保存		現在のストリームの印刷
	切り取ってクリップボードに移動		クリップボードにコピー
	貼り付けの選択		最後の操作を元に戻す
	やり直し		ノードの検索
	ストリームのプロパティを編集		SQL 生成をプレビュー
	ストリームを実行		選択したストリームを実行
	ストリームの中止 (ストリームの実行中にだけ利用可能)		スーパーノードの追加



ズーム・イン (スーパーノード専用)



ズーム・アウト (スーパーノード専用)



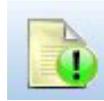
ストリームにマークアップがありません



コメントの挿入



ストリーム・マークアップの非表示 (ある場合)



非表示のストリーム・マークアップを表示



IBM SPSS Modeler Advantage でストリームを開く

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

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

## ツールバーのカスタマイズ

ツールバーは、次のようにさまざまな観点から変更できます。

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

ツールバー表示をオンまたはオフにするには

1. メイン・メニューで次の各項目をクリックします。

「表示」 > 「ツールバー」 > 「表示」

ツールヒントまたはアイコンのサイズ設定を変更するには

1. メイン・メニューで次の各項目をクリックします。

「表示」 > 「ツールバー」 > 「カスタマイズ」

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

## IBM SPSS Modeler ウィンドウのカスタマイズ

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

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

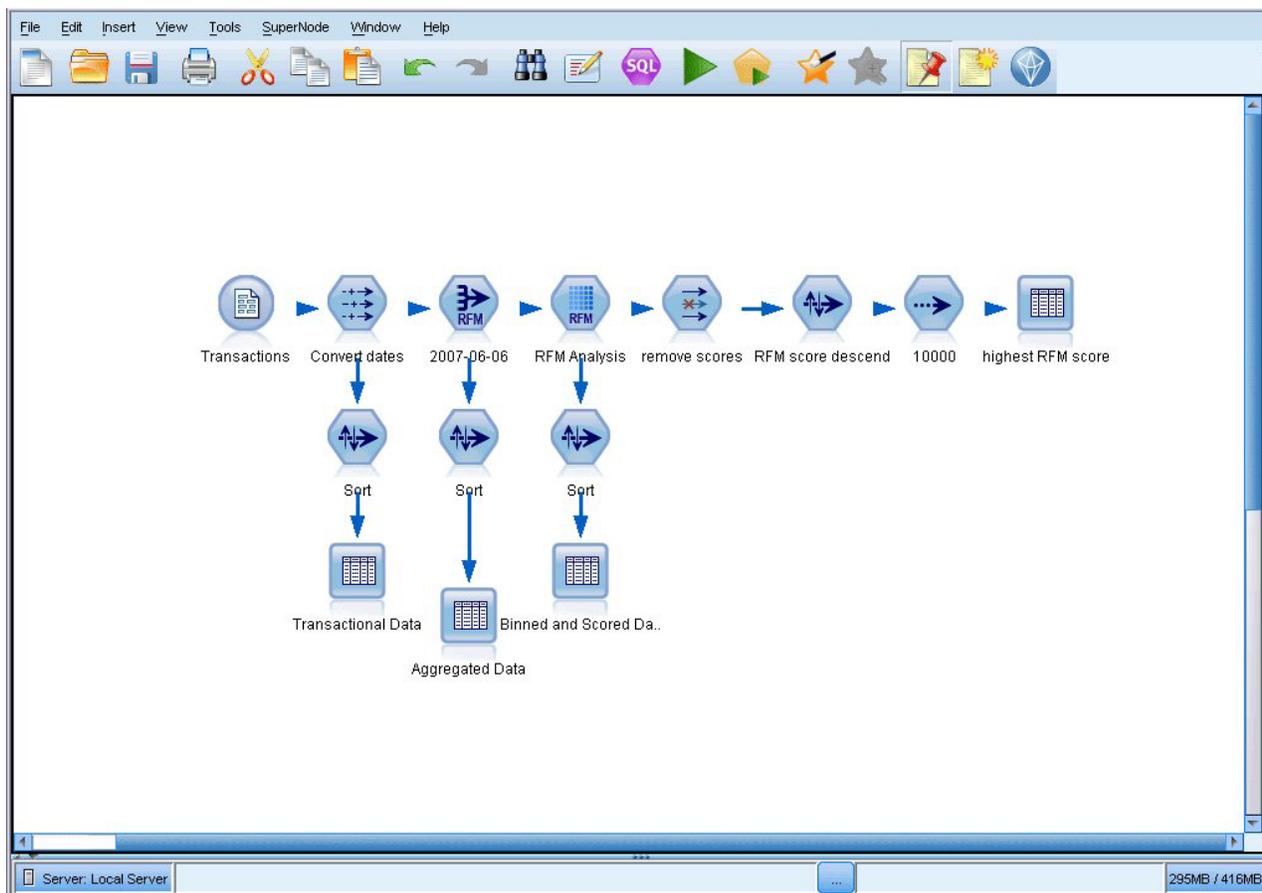


図 10. 最大化されたストリーム キャンバス

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

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

「表示」 > 「ストリーム・マークアップ」

## ストリームのアイコン・サイズの変更

以下の方法で、ストリームのアイコンのサイズを変更できます。

- ストリーム・プロパティの設定から
- ストリームのポップアップ・メニューから
- キーボードの使用

ストリーム ビュー全体を標準アイコンサイズの 8% ~ 200% のいずれかのサイズに設定します。

ストリーム全体のサイズ変更 (ストリーム・プロパティから)

1. メイン・メニューから次の各項目を選択します。

「ツール」 > 「ストリームのプロパティ」 > 「オプション」 > 「レイアウト」

2. 「アイコンのサイズ」メニューからサイズを選択します。
3. 結果を表示するには「適用」をクリックします。
4. 「OK」をクリックして、変更を保存します。

ストリーム全体のサイズ変更 (メニューから)

1. 領域のストリームの背景を右クリックします。
2. 「アイコン・サイズ」を選択し、該当するサイズを選択します。

ストリーム全体のサイズ変更 (キーボードを使用)

1. Ctrl + [-] キーを押すと、ズームアウトします。
2. Ctrl + [+] キーを押すと、ズーム・インします。

この機能は、複雑なストリームの全体のビューを表示するのに適しています。ストリームの印刷に必要なページ数を最小限にすることもできます。

## IBM SPSS Modeler でのマウスの使用方法

IBM SPSS Modeler でよく使われるマウス操作を次に示します。

- **シングルクリック**。マウスの右または左ボタンを使用して、メニューからオプションを選択したり、ポップアップ・メニューを開いたり、さまざまな標準のコントロールやオプションにアクセスすることができます。ノードをクリックしたままマウスを動かしてドラッグすれば、ノードを移動できます。
- **ダブルクリック**。マウスの左ボタンを使用してダブルクリックすると、ノードをストリーム領域に置いて、既存のノードを編集できるようになります。
- **中央ボタンのクリック**。マウスの中央ボタンをクリックして、カーソルをドラッグすることにより、ストリーム領域のノードを接続します。マウスの中央ボタンをダブルクリックすると、ノードの接続が解除されます。中央ボタンがないマウスを使用している場合は、代わりに Alt キーを押しながらクリックしたり、ドラッグしてください。

## ショートカット・キーの使用

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

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

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

表1. サポートしているショートカット・キー

ショートカット・キー	関数
Ctrl+A	すべてを選択する
Ctrl+X	切り取り
Ctrl+N	新規ストリーム
Ctrl+O	ストリームを開く
Ctrl+P	印刷
Ctrl+C	コピー
Ctrl+V	貼り付け
Ctrl+Z	元に戻る
Ctrl+Q	選択したノードの下流にあるすべてのノードを選択
Ctrl+W	下流のすべてのノードの選択を解除 (Ctrl+Q と切り替わる)
Ctrl+E	選択したノードから実行
Ctrl+S	現在のストリームを保存
Alt+矢印キー	選択したストリーム領域上のノードを矢印の方向に移動
Shift+F10	選択したノードのポップアップ・メニューを表示

表2. 古いホットキーに対応するショートカット・キー

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

## 印刷中

IBM SPSS Modeler では、次のオブジェクトを印刷できます。

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

- 注釈 (出力の「注釈」タブを使用して)

オブジェクトを印刷するには、次のようにします。

- プレビューを行わないでオブジェクトを印刷するには、ツールバーの「印刷」ボタンをクリックします。
- 印刷前に印刷の設定を行うには、「ファイル」メニューの「ページ設定」を選択します。
- 印刷前にプレビューを行う場合は、「ファイル」メニューの「印刷プレビュー」を選択します。
- 標準の印刷ダイアログ・ボックスに、選択されているプリンタのオプションを表示して、さまざまなオプションを設定するには、「ファイル」メニューの「印刷」を選択します。

---

## IBM SPSS Modeler の自動化

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

- **Control Language for Expression Manipulation (CLEM)** は、IBM SPSS Modeler ストリーム中を流れるデータの分析と操作を行うための強力な言語です。CLEM を使用すれば、経費と収入データから利益を算出するような簡単な操作から、Web ログ・データを有益な情報を含む一連のフィールドやレコードに変換するような複雑な操作まで、さまざまなストリーム操作を行うことができます。
- **スクリプト**は、ユーザー・インターフェースのプロセスを自動化する強力なツールです。スクリプトは、マウスやキーボードを使用して実行するのと同じような操作を実行できます。また、出力を指定して生成されたモデルを操作することができます。



---

## 第 4 章 データ・マイニングについて

---

### データ・マイニングの概要

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

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

#### モデル作成手法

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

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

#### データの操作と発見

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

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

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

#### 典型的なアプリケーション

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

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

**信用情報のスコアリング:** 個人の信用履歴を、ローン審査に使用する。

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

**医学研究:** 医学的な証拠に基づいて適切な処置を提案するディシジョン・ルールを作成する。

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

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

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

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

用語

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

---

## データへの評価

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

データが使用できることを確認してください。

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

データが関連する属性をカバーしているかどうかの確認

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

このような状況を評価するために、データの総合的な検査を行うことができます。まず始めに、データ・ソースにデータ検査ノードを接続して実行し、レポートを生成、調査することも検討してください。

ノイズを含んだデータの注意

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

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

データが十分にあることを確認してください。

データ・マイニングでは、データ・セットのサイズは必ずしも重要ではありません。データ・セットの標本性、可能性のある結果がデータ・セットの標本範囲内に含まれるかどうか、および変数の組み合わせの方がはるかに重要です。

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

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

#### データのエキスパートの探索

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

---

## データ・マイニングの戦略的方法

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

- 解決すべき重要な問題は何か?
- どのようなデータ・ソースを使用できるか、データのどの部分が現在の問題に関連しているか?
- データ・マイニングを開始する前に、どのような事前処理とデータの事前処理が必要か?
- どのようなデータ・マイニング手法を使用できるか?
- データ・マイニング分析の結果をどのように評価するか?
- データ・マイニングから最大限の情報を得るにはどうするか?

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

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

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

---

## CRISP-DM プロセス・モデル

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

次に 6 つのフェーズを説明します。

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

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

2 番目の主要な点は、データ・マイニングの反復特性です。ユーザーの仕事が、単にデータ・マイニング・プロジェクトを計画し実行して、自分のデータを片付けてから家に帰るだけ、ということとはほとんどありません。データ・マイニングを使用して顧客の要望を解決するというのが当面の目標です。データ・マイニングの 1 サイクルから知識を得ると、多くの場合は顧客の要望を判別してそれを満たすための新しい問い、新しい課題、新しい機会が生じます。これらの新しい問い、課題、機会は、データを再度マイニングすることで解決できます。このマイニングと新しい機会の識別プロセスは、ビジネスについての考え方の一部になり、ビジネス戦略全体の基礎になります。

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

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

---

## モデルの種類

IBM SPSS Modeler 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。「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。

IBM SPSS Modeler アプリケーション ガイド では、これらの手法の多くの例が、モデル作成プロセスの概要とともに提供されています。このガイドは、オンライン・チュートリアル、および PDF 形式で使用できます。詳しくは、トピック 5 ページの『アプリケーションの例』を参照してください。

モデル作成方法は、次の 3 つのカテゴリに分けられます。

- 分類
- アソシエーション
- セグメンテーション

### 分類モデル

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

組織は分類モデルを活用して、既知の結果に基づく予測に役立てることができます。例えば、顧客が購入するか立ち去るか、または特定の取引が既知の詐欺パターンに当てはまるかどうか、です。モデル作成手法には、マシン学習、ルール算出、サブグループ識別、統計的手法、および多重モデル生成が含まれます。

### 分類ノード



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



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



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



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



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



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



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



線型回帰モデルは、対象と 1 つまたは複数の予測値との線型の関係に基づいて連続型対象を予測します。



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



フィールド選択ノードで、（欠損値の割合などの）諸基準に基づいて入力フィールドをスクリーニングして削除にかけ、指定した目標に相対的な残りの入力フィールドの重要度をランク付けします。例えば、数百の潜在的入力フィールドを含むデータセットがあるとして、患者予後のモデリングにはどれが役に立つのでしょうか？



判別分析によって、ロジスティック回帰より厳密な仮説を立てることができますが、これらの仮説が一致した場合、ロジスティック回帰分析に対する様々な代替あるいは補足になります。



ロジスティック回帰は、入力フィールドの値に基づいてレコードを分類する統計手法です。線型回帰と似ていますが、数値範囲ではなくカテゴリー対象フィールドを使用します。



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



一般化線型混合モデル (GLMM) は線型モデルを拡張したため、対象が非正規分布となる場合があります。指定されたリンク関数を介して因子および共変量に線形に関連し、観測が相関できるようになりました。対象を非正規分布とすることができ、観測を相関させることができます。一般化線型混合モデルは、単純な線型回帰から非正規の時系列データの複雑なマルチレベルに至るまで、さまざまなモデルをカバーします。



Cox 回帰ノードを使用すると、打ち切りレコードの存在下でイベントまでの時間のデータの生存モデルを構築します。モデルは、対象のイベントが入力変数の指定の値で指定の時間 ( $t$ ) に発生する確率を予測する生存関数を作成します。



サポート・ベクター・マシン (SVM) ノードを使用すると、オーバーフィットすることなく、データを 2 つのグループのいずれかに分類することができます。SVM は、非常に多数の入力フィールドを含むデータセットなど、広範なデータセットを処理することができます。



ベイズ・ネットワーク・ノードを使用すると、観測された情報および記録された情報を実際の知識を組み合わせることによって確率モデルを作成し、発生 of 尤度を確立できます。ノードは主に分類に使用される Tree Augmented Naïve Bayes (TAN) および Markov Blanket ネットワークに焦点を当てています。



SLRM (自己学習応答モデル) ノードを使用するとモデルを構築でき、単一または少数の新しいケースを使用して全データを使用するモデルの保持をすることなく、モデルの再見積もりを行うことができます。



時系列ノードは、時系列から指数平滑法、1 変量の自己回帰型統合移動平均法 (ARIMA)、および多変量 ARIMA (または転送関数) モデルを推測し、将来のパフォーマンスの予測を作成します。時系列ノードは、時間区分ノードによって常に先行される必要があります。



$k$  が整数である場合、 $k$  最近隣 (KNN) ノードは、新しいケースを、予測領域の新しいケースに最も近い  $k$  個のオブジェクトのカテゴリまたは値と関連付けます。類似したケースはお互いに近く、類似していないケースはお互いに離れています。

## アソシエーション・モデル

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

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

## アソシエーション・ノード



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



CARMA モデルは、入力または対象フィールドを指定しなくても、データからルール・セットを抽出します。Apriori とは対照的に、CARMA ノードでは、前提条件サポートだけでなく、ルール・サポート (前提条件と結果の両方のサポート) を対象とした構築の設定が可能です。これは、生成されたルールをさまざまな用途で活用できることを意味します。例えば、この休暇シーズンに販売促進する項目を結果とする、商品またはサービス (前提条件) のリストを見つけることができます。



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

## セグメント化モデル

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

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

## セグメント化ノード



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



K-Means ノードで、データ・セットが異なるグループ（つまりクラスター）へ、クラスタリングされます。この方法で、固定数のクラスターを定義し、クラスターにレコードを繰り返し割り当てて、これ以上調整してもモデルが改善されなくなるまで、クラスターの中心を調整します。K-means では、結果を予測するのではなく、入力フィールドのセット内のパターンを明らかにするために、「非監視学習」として知られるプロセスが使用されます。



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



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



異常値検出ノードは、「正常な」データのパターンに合致しない異常ケース (外れ値) を識別します。このノードで、外れ値が既知のパターンに当てはまらなかったり、何を探しているのかははっきりしなかったりする場合でも、外れ値を識別できます。

## データベース内マイニング・モデル

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

## IBM SPSS Statistics モデル

コンピューターに IBM SPSS Statistics をインストールしライセンスが付与されている場合、IBM SPSS Modeler 内から特定の IBM SPSS Statistics ルーチンにアクセスおよび実行して、モデルを作成およびスコアリングできます。

## 詳細情報

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

---

## データ・マイニングの例

データ・マイニングについて学習する現実的で最良の方法は、最初に例を使うことです。「*IBM SPSS Modeler アプリケーション・ガイド*」には、多数のアプリケーション例が記載されています。これらの例により、具体的なモデルの作成方法と技術について、目的別に簡潔に説明しています。詳しくは、トピック 5 ページの『*アプリケーションの例*』を参照してください。

---

## 第 5 章 ストリームの構築

---

### ストリーム構築の概要

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

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



可変長ノード : データをデータ・ソースから読み込むために使用します。



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



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



テーブル・ノード : 操作結果を画面に表示するために使用します。

---

### データ・ストリームの作成

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

- ストリーム領域にノードを追加する。
- ノードを接続してストリームを形成する。
- 任意のノードまたはストリーム・オプションを指定する。
- ストリームを実行する。

このセクションでは、ノードを使ったより複雑なデータ・ストリームの作成方法を詳細に説明していきます。また、ノードやストリームのオプションや設定方法も取り上げていきます。IBM SPSS Modeler に付属のデモ・データ (プログラム インストール・ディレクトリーの *Demos* フォルダ) を使ったストリーム構築の例は、5 ページの『アプリケーションの例』を参照してください。

## ノードの作業

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

スーパーノードには、いくつかの種類があります。入力ノードはデータをストリームに投入し、ノード・パレットの「入力」タブにあります。プロセス・ノードは、各データ・レコードおよびフィールドで操作を行い、パレットの「レコード設定」タブおよび「フィールド設定」タブに表示されます。出力ノードは、さまざまなデータの出力、グラフ、モデル結果を生成し、ノード・パレットの「グラフ」タブ、「出力」タブ、「エクスポート」タブに表示されます。モデル作成ノードは統計アルゴリズムを使用してモデル・ナゲットを作成し、ノード・パレットの「モデル作成」タブ、(有効化されている場合は)「DB モデリング」タブに表示されます。詳しくは、トピック 17 ページの『ノード・パレット(N)』を参照してください。

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

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

注：ノード・パレットはユーザー設定できます。詳しくは、トピック 158 ページの『ノード・パレットのカスタマイズ』を参照してください。

## ストリームへのノードの追加

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

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

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

### ノードの削除

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

## ストリーム内のノードの接続

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

ダブルクリックでノードを追加、接続するには

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

マウスの中央ボタンを使用してノードを接続するには

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

ノードを手作業で接続するには

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

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

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

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

## ストリーム内でのノードの迂回

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

例えば、新規フィールドを作成し、フィールドのフィルタリングを行った後、結果をヒストグラムとテーブルに表示するストリームを考えてみましょう。フィールドをフィルタリングする前にもデータのグラフとテーブルを参照したい場合、ストリームに新規にヒストグラム・ノードとテーブル・ノードを追加するか、またはフィルター・ノードを迂回することができます。フィルター・ノードを迂回すると、グラフとテーブルはフィールド作成ノードから直接接続されます。フィルター・ノードの接続は解除されます。

ノードを迂回するには

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

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

## ストリーム内でのノードの無効化

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

例えば、複数のフィールドをフィルタリングするストリームがあり、減らされたデータ・セットによりモデルを作成するとします。それとは別に、フィールドをフィルタリングせずに同じモデルを作成して、モデルの結果が改善されるかどうかを確認する場合は、フィルター・ノードを無効化できます。フィルター・ノードを無効化すると、モデル作成ノードへの接続は、フィールド作成ノードからデータ型ノードに直接通過します。

ノードを無効化するには

1. ストリーム領域で、無効にするノードを右クリックします。
2. ポップアップ・メニューの「ノードの無効化」をクリックします。

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

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

## 既存の接続へのノードの追加

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

1. マウスの中央ボタンを使用して、接続の矢印をクリックし、ノードの挿入先までドラッグします。中央ボタンがない場合は、Alt キーを押しながらクリック・アンド・ドラッグすることもできます。
2. ボタンを押したまま挿入するノードまで接続をドラッグし、マウス・ボタンを離します。

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

## ノード間の接続の削除

2 つのノード間の接続を削除する手順は、次のとおりです。

1. 接続の矢印を右クリックします。
2. メニューで「接続の削除」をクリックします。

ノードとのすべての接続を削除するには、次のいずれかを実行します。

- ノードを選択して、F3 を押します。
- ノードを選択して、メイン・メニューで次の各項目をクリックします。

「編集」 > 「ノード」 > 「切断」

## ノードのオプションの設定

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

- 選択したノードのダイアログ・ボックスを開くには、「編集」をクリックします。
- あるノードから別のノードに手作業で接続するには、「接続」をクリックします。

- 選択したノードに対するすべてのリンクを削除するには、「**接続の解除**」をクリックします。
- 「編集」ダイアログ・ボックスの「注釈」タブを表示するには、「**名前の変更と注釈**」をクリックします。
- 「**新規コメント**」をクリックして、ノードに関連するコメントを追加します。詳しくは、トピック 58 ページの『ノードおよびストリームへのコメントおよび注釈の追加』を参照してください。
- 「**ノードの無効化**」をクリックして、処理時にノードを「非表示」にします。処理する場合にノードを再度表示するには、「**ノードの有効化**」をクリックします。詳しくは、トピック 40 ページの『ストリーム内でのノードの無効化』を参照してください。
- ストリーム領域からノードを削除するには、「**切り取り**」または「**削除**」をクリックします。注：「**切り取り**」をクリックした場合、そのノードを貼り付けることができます。ただし、「**削除**」を選択した場合は、貼り付けることはできません。
- 接続を持たないノードのコピーを作成するには、「**コピー**」をクリックします。コピーしたノードは、既存のストリームに追加することも、新しいストリームに追加することもできます。
- 以前に保存したノードを開いて、そのオプション設定を現在選択しているノードに取り込むには、「**ノードのロード**」をクリックします。注：この操作を行うには、ノードの種類が同じでなければなりません。
- 「**ノードの取得**」をクリックして、接続した IBM SPSS Collaboration and Deployment Services Repository からノードを取得します。
- ノードの詳細をファイルに保存するには、「**ノードの保存**」をクリックします。保存したノード定義は、同じ種類のノードにしかロードできません。
- 「**ノードを格納**」をクリックして、接続した IBM SPSS Collaboration and Deployment Services Repository の選択したノードを格納します。
- 選択したノードのキャッシュに関するオプションを設定するには、「**キャッシュ**」をクリックしてメニューを表示します。
- データの新規ソースへのマッピングや必須フィールドの指定などのオプションを設定するには、「**データ・マッピング**」をクリックしてメニューを表示します。
- 現在のストリームにスーパーノードを作成するには、「**スーパーノードの作成**」をクリックしてメニューを表示します。
- 選択したノードを置き換えるには、「**ユーザー入力ノードの生成**」をクリックします。このノードは、元のノードと同じフィールドを持ちます。
- 選択したノードの下流にあるすべてのターミナル・ノードを実行するには、「**ここから実行**」をクリックします。

## ノードのキャッシュ・オプション

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

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

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

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

キャッシュを有効にするには

1. ストリーム領域でノードを右クリックし、メニューの「**キャッシュ**」をクリックします。
2. キャッシュ設定用のサブメニューから、「**使用する**」を選択します。
3. キャッシュを解除するには、ノードを右クリックして、キャッシュ設定用のサブメニューから「**使用しない**」を選択します。

データベース内でのノードのキャッシュ

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

55文字より長い文字列でデータベース・キャッシュを使用している場合、どちらかの上流のキャッシュノードから、フィールドの値が読み込まれているそのタイプのノードがあるようにするか、または *options.cfg* ファイルの *default\_sql\_string\_length* パラメータを用いて文字列の長さを設定します。それにより、一時テーブルの該当する列を適切な幅に設定し、文字列を調整します。

データベースのキャッシングを利用するには、SQL 最適化とデータベース・キャッシングの両方を有効にする必要があります。Client の設定は、Server の最適化設定によって無効になるので注意してください。詳しくは、トピック 47 ページの『ストリームの最適化オプションの設定』を参照してください。

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

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

キャッシュを取り消すには

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

1. ストリーム領域でノードを右クリックし、メニューの「**キャッシュ**」をクリックします。
2. キャッシュ設定用のサブメニューから、「**取り消し**」を選択します。

キャッシュを保存するには

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

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

キャッシュをロードするには

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

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

## ノードのデータをプレビューする

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

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

表示される行のデフォルト数は 10 行です。ただし、ストリームのプロパティで変更できます。詳しくは、トピック 44 ページの『ストリームの一般的なオプションの設定』を参照してください。

「ノードの生成」メニューから、複数の種類のノードを作成できます。

## ノードをロックする

その他のユーザーがストリームの 1 つまたは複数のノード設定を修正できないようにするために、スーパーノードという特別な種類のノードにノードをカプセル化して、パスワード保護を適用してノードをロックすることができます。

## ストリームの作業

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

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

このタブから、次のような作業を行うことができます。

- ストリームへのアクセス。
- ストリームの保存。
- 現在のプロジェクトへのストリームの保存。
- ストリームを閉じる。
- 新規ストリームを開く。
- IBM SPSS Collaboration and Deployment Services リポジトリでストリームの保存と検索を行う (ユーザーのサイトで利用可能な場合)。詳しくは、トピック 121 ページの『IBM SPSS Collaboration and Deployment Services Repository について』を参照してください。

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

## ストリームのオプションの設定

現在のストリームに適用するさまざまなオプションを指定することができます。また、これらのオプションをすべてのストリームにデフォルトとして適用するよう保存することができます。オプションは次の通りです。

- **全般:** ストリームで使用する記号やテキストのエンコードなどのその他のオプション。詳しくは、トピック『ストリームの一般的なオプションの設定』を参照してください。
- **日時:** 日付と時刻の表現形式に関連するオプション。詳しくは、トピック 46 ページの『ストリームの日時オプションの設定』を参照してください。
- **数値書式:** 数値表現の形式を制御するオプション。詳しくは、トピック 46 ページの『ストリームの数値形式オプションの設定』を参照してください。
- **最適化:** ストリームのパフォーマンスを最適化するためのオプション。詳しくは、トピック 47 ページの『ストリームの最適化オプションの設定』を参照してください。
- **ロギングとステータス:** SQLのロギングおよび記録の状態を制御するオプション。詳しくは、トピック 48 ページの『ストリームの SQL ロギングおよびレコード ステータスオプションの設定』を参照してください。
- **レイアウト:** 領域上のストリームのレイアウトに関連するオプション。詳しくは、トピック 49 ページの『ストリームのレイアウト・オプションの設定』を参照してください。

ストリーム・オプションを設定するには

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

また、「ツール」メニューから次の各項目を選択することもできます。

「ストリームのプロパティ」 > 「オプション」

**ストリームの一般的なオプションの設定:** 一般的なオプションは、現在のストリームのさまざまな面に適用されるその他のオプションのセットです。

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

**グループ化記号:** 数字の表示形式で、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 のコンマ)。オプションには、なし、ピリオド、コンマ、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)。

**エンコード:** テキスト・エンコードについて、ストリームのデフォルトの方法を指定します。(注：可変長ファイル・ソース・ノードとフラット・ファイル・エクスポート・ノードだけに適用されます。多くのデータ・ファイルには埋め込み文字コード情報があります)。システム・デフォルトまたは UTF-8 のいずれかを選択できます。システム・デフォルトは、Windows のコントロール・パネル、または分散モードで動作している場合はサーバー・コンピュータから指定できます。詳しくは、トピック 181 ページの『IBM SPSS Modeler 内での Unicode のサポート』を参照してください。

**ルールセットの評価：** ルールセット・モデルの評価方法を決定します。デフォルトでは、ルールセットは「**評決**」を使用して個別のルールからの予測を組み合わせ、最終的な予測を決定します。ルールセットが該当する最初のルールを使用するように設定するには、「**最初のヒット**」を選択します。このオプションは、ディシジョン・リスト・モデルには適用されません。このモデルでは、アルゴリズムに定義されたように、最初のヒットが常に使用されます。

**データ・プレビューで表示する行の最大数：** データのプレビューがノードに必要な場合に表示する行の最大数。詳しくは、トピック 43 ページの『ノードのデータをプレビューする』を参照してください。

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

**Kohonen および K-Means モデリングのセット・サイズを制限:** Kohonen ネットワークおよび K-Means モデリングに使用される名義型フィールドの最大メンバー数を指定する場合に選択します。デフォルトは 20 です。以降のフィールドは無視され、フィールドに問題があることを示す警告が発せられます。

互換性を維持するために、このオプションは IBM SPSS Modeler のバージョン 14 で置き換えられた旧式のニューラル・ネットワーク・ノードにも適用されること、古いストリームには現在でもこのノードが含まれている場合があることに注意してください。

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

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

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



図 11. フィールドラベルと値ラベルの切り替えに使われるツールバー・アイコン

**実行時間を表示：** ストリームを実行した後、「実行時間」タブにストリーム・ノードの各実行時間を表示します。詳しくは、トピック 50 ページの『ノード実行時間の表示』を参照してください。

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

**ストリームの日時オプションの設定：** これらのオプションは、現在のストリーム内のさまざまな日付と時刻の表現に使用する書式を指定します。

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

**日付の形式：** 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形式を選択します。

**時間の形式：** 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形式を選択します。

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

**基準日付 (1 月 1 日)：** CLEM 日付関数で単一の日付を扱う場合に使用する基準の年 (月日は常に 1 月 1 日) を選択します。

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

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

**ストリームの数値形式オプションの設定：** これらのオプションは、現在のストリーム内のさまざまな数値の表現に使用する書式を指定します。

**数字の表示形式：** 標準 (#####.###)、科学的 (#.###E+###)、または通貨表記形式 (\$###.##) から選択できます。

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

**計算単位：** CLEM 式で使用する三角関数の角度単位として、「ラジアン」または「度数」を選択します。詳しくは、トピック 101 ページの『三角関数』を参照してください。

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

**ストリームの最適化オプションの設定：** 最適化設定を使用して、ストリームのパフォーマンスを最適化することができます。IBM SPSS Modeler Server が使用されている場合のパフォーマンスと最適化の設定は、クライアントの設定より優先されます。

注：データベース・モデル作成および SQL 最適化では、IBM SPSS Modeler Server 接続が IBM SPSS Modeler コンピューター上で可能でなければなりません。この設定を有効にすると、データベース・アルゴリズムにアクセスし、IBM SPSS Modeler から SQL を直接プッシュバック、IBM SPSS Modeler Server にアクセスできます。現在のライセンスステータスを確認するには、IBM SPSS Modeler メニューから次を選択します。

「ヘルプ」 > 「バージョン情報」 > 「その他の詳細」

接続が有効な場合、「ライセンス ステータス」タブにオプション「サーバーの有効化」が表示されます。

詳しくは、トピック 12 ページの『IBM SPSS Modeler Server に接続中』を参照してください。

注：SQL プッシュバックおよび最適化がサポートされているかどうかは、使用しているデータベースの種類によって異なります。IBM SPSS Modeler 16 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 ドライバーの最新情報については、当社サポート・サイト (<http://www.ibm.com/support>) を参照してください。

**ストリームのリライトを有効にする：** このオプションを選択すると、IBM SPSS Modeler のストリームのリライトが有効になります。リライトには 4 種類あり、そのうちの 1 つまたはすべてあるいは任意の組み合わせを選択することができます。ストリームのリライトは、ストリーム内のノードをバックグラウンドで並べ替え、ストリームの意味論を変更することなくストリームをより効率的に操作できるようにする処理です。

- **SQL 生成の最適化：** このオプションを選択すると、より多くのストリーム操作がデータベースにプッシュバックされて SQL 生成を使用して実行されるように、ストリーム内のノードが並べ替えられます。SQL へ変換できないノードが検出されると、SQL へ変換可能な下流にあるかどうか、またそれを、ストリームの意味論に影響を与えずに安全に問題のノードの前に移動できるかどうかを調べることができます。データベースの方が IBM SPSS Modeler より効率的に操作を実行できるだけでなく、このようなプッシュバックによって IBM SPSS Modeler で処理するために返されるデータ・セットのサイズも小さくすることができます。この結果、ネットワーク トラフィックを減らし、ストリーム操作を高速化することができます。SQL 最適化を有効にするには、「SQL 生成」チェックボックスをオンにする必要があります。
- **CLEM 式の最適化：** このオプションを選択すると、最適化プログラムは、ストリームの実行前に事前処理できる CLEM 式を検索して処理速度を高めることができます。単純な例として、 $\log(\text{salary})$  という式が指定されると、最適化プログラムは、実際の salary 値を計算してそれを処理のために引き渡します。このようにすると、SQL プッシュバックと IBM SPSS Modeler Server の両方のパフォーマンスを向上させることができます。
- **シンタックスの実行を最適化：** この方法のストリームのリライトでは、IBM SPSS Statistics シンタックスを含む複数のノードを結合する操作を効率化します。各操作それぞれを実行する代わりに、シンタックス・コマンドを 1 つの操作に結合することによって最適化が行われます。
- **その他の実行を最適化：** この方法のストリームのリライトでは、データベースで処理できない操作を効率化します。最適化は、ストリームのデータ量をできるだけ早く減らすことによって実現されます。データの整合性を保ちながら、操作をデータ・ソースの近くに置いて、下流で結合などの時間のかかる操作に使用されるデータの量を減らすように、ストリームがリライトされます。

**並行処理を有効にする**：マルチ プロセッサのコンピューターで実行する場合、このオプションでシステムがプロセッサ間の負荷のバランスがとれるようになり、より速いパフォーマンスとなります。C5.0、レコード結合 (キーによる)、ソート、分割 (順位付け法分位法)、およびレコード集計 (1 つ以上のキー・フィールドを使用) の複数または個別のノードを使用すると、並行処理の恩恵を受けることができます。

**SQL 生成**：SQL 生成を有効にするには、このオプションを選択します。実行プロセスを生成する SQL コードを使用することで、ストリーム操作をデータベースにプッシュバックできるようにし、それがパフォーマンスを向上させる可能性があります。さらにパフォーマンスを向上させるには、データベースにプッシュバックする操作の数を最大限にできるように、「**SQL 生成の最適化**」も選択できます。ノードの操作がデータベースにプッシュバックされると、そのノードはストリームの実行時に紫色で強調表示されます。

- **データベース・キャッシング**：SQL を生成するストリームをデータベース内で実行するために、ファイル・システムでなくデータベース内の一時テーブルへ、データを中流でキャッシュできます。この機能を SQL 最適化と組み合わせると、パフォーマンスが著しく向上する可能性があります。例えば、データ・マイニング・ビューを作成するために複数のテーブルを併合するストリームからの出力をキャッシュし、必要に応じて再使用できます。データベースのキャッシングが有効化されると、任意の非ターミナル・ノードを右クリックするだけでその場所のデータがキャッシュされ、次にストリームが実行されると自動的に、データベース内にキャッシュが直接作成されます。これにより SQL が下流のノード用に生成されるようになり、パフォーマンスをさらに改善します。代わりに、例えばポリシーまたは許可によってデータベースに書き込まれたデータが損なわれる場合、このオプションを必要に応じて無効にすることができます。データベース・キャッシングまたは SQL 最適化が有効にされていないと、キャッシュは、代わりにファイル・システムへ書き出されます。詳しくは、トピック 41 ページの『ノードのキャッシュ・オプション』を参照してください。
- **緩和された変換を使用**：単一の形式で保存されている場合、文字列から数値、または数値から文字列へのデータの変換を可能にします。例えば、データが文字列としてデータベースに保存されているが、実際に意味のある数値が含まれている場合、プッシュバックが発生したときに使用するようデータを変換することができます。

注：SQL 実装の些細な違いが原因で、データベース内で実行されるストリームは、IBM SPSS Modeler 内で実行されるときと若干異なる結果を返すことがあります。これらの違いは、似たような理由で、データベース・ベンダーによっても異なる可能性があります。

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

**ストリームの SQL ロギングおよびレコード ステータスオプションの設定**：これらの設定には、ストリームによって生成される SQL 文の表示、およびストリームで処理されたレコード数の表示を制御するさまざまなオプションが含まれています。

**ストリーム実行中に SQL を メッセージ・ログに表示する**：ストリームの実行中に、生成された SQL をメッセージ・ログに渡すかどうかを指定します。

**ストリームの準備中に、SQL 生成の詳細をメッセージ・ログに表示する**：ストリームのプレビュー中に、生成される SQL のプレビューをメッセージログに渡すかどうかを指定します。

**SQL の表示**：ログに出力する SQL に、ネイティブの SQL 関数を含めるか、IBM SPSS Modeler によって生成される {fn FUNC(...)} 形式の標準 ODBC 関数を含めるかを指定します。前者は、実装されていない可能性がある ODBC ドライバーの機能に依存しています。例えば、このコントロールは、SQL Server に対して何の影響も及ぼしません。

**SQL を再フォーマットして読みやすくする** : ログに表示された SQL を読みやすくするためにフォーマットするかどうかを指定します。

**レコードのステータスを表示** : レコードがターミナル・ノードに達したときに、そのレコードを報告する時期を指定します。N レコードごとにステータスの更新を行うことを、数字で指定します。

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

**ストリームのレイアウト・オプションの設定**: これらの設定は、ストリーム領域の表示と使用方法に関連する多くのオプションを提供します。

**ストリーム領域の最小幅** : ストリーム領域の最小幅をピクセル単位で指定します。

**ストリーム領域の最小の高さ** : ストリーム領域の最小の高さをピクセル単位で指定します。

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

**アイコン名の文字数** : ストリーム領域上のノード名の文字数制限を指定します。

**アイコン・サイズ** : ストリーム ビュー全体を標準アイコンサイズの 8% ~ 200% のいずれかのサイズに設定します。

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

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

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

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

**Analytic Server のストリームのプロパティ**: これらの設定は、Analytic Server を操作するための多くのオプションを提供します。

#### **Analytic Server の外部で処理するレコードの最大数**

Analytic Server のデータ・ソースから SPSS Modeler Server にインポートするレコードの最大数を指定します。

#### **Analytic Server でノードを処理できない場合の通知**

この設定により、Analytic Server に送信されるストリームに Analytic Server で処理できないノードが含まれている場合の動作が決まります。警告を出してストリームの処理を続行するか、またはエラーをスローして処理を停止するかを指定します。

#### **分割モデル・ストレージ設定**

### モデルが次のサイズ (MB) を超過した場合に参照により Analytic Server 内に分割モデルを保管 (Store split models by reference on Analytic Server when model size (MB) exceeds)

通常、モデル・ナゲットはストリームの一部として保管されます。多くの分割を含む分割モデルでは大きなナゲットが生成されることがあり、そのナゲットをストリームと Analytic Server の間で往復させるとパフォーマンスに影響を及ぼす可能性があります。解決策として、分割モデルが指定のサイズを超過した場合はそのモデルを Analytic Server に保管し、SPSS Modeler 内のナゲットにそのモデルへの参照を含めます。

### 実行の完了後、参照によりモデルを保管する Analytic Server 内のデフォルトのフォルダー

分割モデルを保管する Analytic Server 内の場所のデフォルト・パスを指定します。このパスは、有効な Analytic Server のプロジェクト名で開始する必要があります。

### プロモートされたモデルを保管するフォルダー(P)

「プロモートされた」モデルを保管する場所のデフォルト・パスを指定します。プロモートされたモデルは、SPSS Modeler セッションの終了時にクリーンアップされません。

## ストリーム操作メッセージの表示

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

ストリーム メッセージを表示するには

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

また、「ツール」メニューから次の各項目を選択することもできます。

「ストリームのプロパティ」 > 「メッセージ」

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

「ユーザー・オプション」ダイアログ・ボックスで、SQL 最適化とログ記録オプションが有効な場合、生成された SQL についての情報も表示されます。詳しくは、トピック 47 ページの『ストリームの最適化オプションの設定』を参照してください。

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

## ノード実行時間の表示

「メッセージ」タブで、実行時間を表示するように選択することができます。その場合、ストリーム内の、IBM SPSS Modeler Server で実行されたすべてのノードの個別の実行時間を表示することができます。他の領域 (R または Analytic Server) で実行されたストリームについては、実行時間が正確でない場合があります。

注：この機能を使用するには、「オプション」タブの「全般」設定の「実行時間を表示」チェック・ボックスをオンにします。

ノード実行時間の表では、列は次のようになります。列見出しをクリックし、エントリーを昇順または降順にソートします (例えば、最も実行時間の長いノードを表示)。

**ターミナル・ノード:** ノードが属するブランチの識別子。識別子は、ブランチの末端のターミナル・ノードの名前です。

**ノード・ラベル:** 実行時間が参照するノードの名前。

**ノード ID:** 実行時間が参照するノードの一意の識別子。この識別子は、ノードが作成されるシステムによって生成されます。

**実行時間:** このノードの実行にかかる時間 (秒)。

## ストリームとセッション・パラメーターの設定

パラメーターは、CLEM 式とスクリプトで使用するために定義できます。実際のところ、パラメーターはユーザー定義の変数であり、保存されて、現在のストリーム、セッション、またはスーパーノードで継続します。さらに、スクリプトを使用する場合と同様に、ユーザー・インターフェースからもアクセスできます。例えば、ストリームを保存すると、そのストリームに設定されているパラメーターも保存されます (これは、ローカル・スクリプト変数と異なる点です。ローカル・スクリプト変数は、宣言されたスクリプト内でのみ使用できます)。多くの場合、パラメーターはスクリプトで使用され、スクリプト内でハードコーディングする必要のないフィールドや値の情報を提供することにより、スクリプトの動作を制御します。

パラメーターの有効範囲は、それがどこで設定されたかによって異なります。

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

パラメーターは、スーパーノード用にも設定できます。この場合、スーパーノード内にカプセル化されたノードでだけ表示できます。

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

1. ストリーム・パラメーターを設定するには、次のメニュー項目を選択してください。

「ツール」 > 「ストリームのプロパティ」 > 「パラメーター」

2. セッション・パラメーターを設定するには、「ツール」メニューの「セッション・パラメーターの設定」を選択します。

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

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

**ロング ネーム :** 作成したパラメーター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ー内のように格納されるかを示します。例えば、「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 が除去されます。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。日付のパラメーターには、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。** 各パラメーターの現在の値が表示されます。必要に応じてパラメーターを調整してください。日付のパラメーターには、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。他の形式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 of 尺度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、パラメーター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・インターフェースを通じてだけ、パラメーターに設定できます。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ーを使用可能なパラメーター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ーを削除するには、削除ボタン (X マーク) を使用します。

## パラメーター値のライタイム・プロンプトの指定

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

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

**これらのプロンプトの無効化** ストリーム実行時にこれらのプロンプトを表示しない場合は、このボックスをオンにします。パラメーターが定義されているストリーム・プロパティまたはセッション・プロパティのダイアログ・ボックスの「**プロンプト?**」チェック・ボックスをオンにすると、プロンプトを再度表示することができます。詳しくは、トピック 51 ページの『ストリームとセッション・パラメーターの設定』を参照してください。

## パラメーター・タイプの値の制約の指定

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

**データ型：** 現在選択されている測定 of 尺度を表示します。IBM SPSS Modeler でのパラメーターの利用目的に応じて、値を変更することができます。

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

このダイアログ・ボックスの下半分は、「**データ型**」で選択した測定 of 尺度によって異なります。

## 連続型測定 of 尺度

**下限** :パラメーター値の下限を指定します。

**上限** :パラメーター値の上限を指定します。

**ラベル**: 集計範囲フィールドの任意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。

## 名義型測定 of 尺度

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

## フラグ型測定 of 尺度

**真 (true)** :条件を満たす場合のパラメーターのフラグ値を指定します。

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

**ラベル**: フラグ型フィールドの値のラベルを指定できます。

## ストリーム展開のオプション

「ストリームのプロパティ」ダイアログ・ボックスの「展開」タブで、モデル・リフレッシュ、自動ジョブ・スケジュール、または IBM Analytical Decision Management のさらなる活用を目的に、IBM SPSS Collaboration and Deployment Services 内のシナリオとして、ストリームを展開するためのオプションを指定できます。いずれのストリームにも、その展開の前に、指定されたスコアリング・ブランチが必要です。追加の要件およびオプションは、展開の種類によって異なります。詳しくは、トピック 122 ページの『リポジトリ・オブジェクトの保存と展開』を参照してください。

## ストリームのループ実行

「ストリームのプロパティ」ダイアログ・ボックスの「実行」タブを使用すると、ループ条件を設定して、現在のストリームの繰り返しタスクを自動化することができます。

これらの条件を設定すると、それをスクリプトへの導入として使用できます。これは、より良いスクリプトを作成するためのベースとして使用する目的などで変更可能なストリームの基本スクリプトがスクリプト・ウィンドウに取り込まれるためです。詳しくは、トピック 117 ページの『グローバル関数』を参照してください。

### ストリームのループの設定

1. 「ファイル」メニューから、「ストリームのプロパティ」を選択します (またはマネージャー・パネルの「ストリーム」タブからストリームを選択し、ポップアップ・メニューを右クリックして「ストリームのプロパティ」をクリックします)。
2. 「実行」タブをクリックします。
3. 「ループ / 条件式実行」実行モードを選択します。
4. 「ループ」タブをクリックします。

また、「ツール」メニューから次の各項目を選択することもできます。

「ストリームのプロパティ」 > 「実行」

さらに別の方法としては、ノードを右クリックして、コンテキスト・メニューから以下をクリックします。

「ループ / 条件式実行」 > 「ループ設定の編集」

**反復。** この行番号の値は編集できませんが、テーブルの右側にあるボタンを使用することで反復の追加や削除を行ったり、反復を上下に移動させたりすることができます。

**テーブル・ヘッダー。** テーブル・ヘッダーには、反復キーと、ループの設定時に作成したすべての反復変数が反映されます。

## ストリームのグローバル値の表示

「ストリームのプロパティ」ダイアログ・ボックスの「グローバル」タブでは、現在のストリームに設定されているグローバル値を表示することができます。グローバル値は、グローバル・ノードを使用して作成され、選択したフィールドの平均、合計、または標準偏差などの統計量を判断するために用いられます。

グローバル・ノードが実行されると、ストリーム操作内でこれらの値をさまざまに利用することができます。詳しくは、トピック 117 ページの『グローバル関数』を参照してください。

ストリームのグローバル値を表示するには

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

また、「ツール」メニューから次の各項目を選択することもできます。

「ストリームのプロパティ」 > 「グローバル」

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

## ストリーム内のノード検索

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

ストリーム内のノードを検索するには

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

また、「ツール」メニューから次の各項目を選択することもできます。

「ストリームのプロパティ」 > 「検索」

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

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

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

**次のキーワードを含む：**このボックスをチェックして、1 つまたは複数のキーワードを入力して、ノード・ダイアログ・ボックスの「注釈」タブの「キーワード」フィールド入力されたテキストを含むノードを検索します。入力したキーワードは完全に一致する必要があります。あるいは、複数のキーワードで検索する場合、キーワードをセミコロンで区切ります (例えば、proton;neutron と入力すると、これらのキーワードのいずれかを含むすべてのノードが検出されます)。詳しくは、トピック 63 ページの『注釈』を参照してください。

**次の注釈を含む：**このボックスをチェックして、1 つまたは複数のキーワードを入力して、ノード・ダイアログ・ボックスの「注釈」タブのメイン・テキスト領域にこのテキストを含むノードを検索します。検索では大文字と小文字の区別を行わず、複数の単語をテキストの単一部分として扱います。詳しくは、トピック 63 ページの『注釈』を参照してください。

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

**次の ID と一致：**このボックスをチェックしてノード ID を入力し、この ID を持つ特定のノードを検索します (このオプションを選択すると、前述のオプションはすべて無効になります。)。ノード ID はノード作成時にシステムによって割り当てられ、ID を使用してスクリプトまたは自動化の目的でノードを参照できます。ノード ID は 1 つだけ入力してください。完全に一致する必要があります。詳しくは、トピック 63 ページの『注釈』を参照してください。

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

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

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

## ストリーム名の変更

「ストリームのプロパティ」ダイアログ・ボックスの「注釈」タブでは、ストリームを説明する注釈を追加したり、ストリームの名前を作成することができます。これらのオプションは特に、プロジェクト・ペインに追加されたストリームのレポートを生成する際に役立ちます。詳しくは、トピック 63 ページの『注釈』を参照してください。

## ストリームの説明

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

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

### ストリームの一般情報

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

### 説明とコメント

このセクションには次が表示されます。

- ストリームの注釈 (63 ページの『注釈』を参照)
- 特定のノードに接続していないコメント
- ストリームのモデル作成ブランチおよびスコアリングブランチのノードに接続するコメント

### スコアリング情報

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

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

### モデル作成情報

ストリームのモデル作成ブランチに関連する情報が表示されます。

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

## ストリームの説明のプレビュー

「ストリームのプロパティ」ダイアログ・ボックスでオプションをクリックして、Web ブラウザーでストリームの説明の内容を表示することができます。説明の内容は、ダイアログ・ボックスの「展開」タブで指定するオプションによって異なります。詳しくは、トピック 136 ページの『ストリーム展開のオプション』を参照してください。

ストリームの説明を表示するには

1. IBM SPSS Modeler のメイン・メニューで次の各項目をクリックします。

「ツール」 > 「ストリームのプロパティ」 > 「展開」

2. 展開タイプ、指定のスコアリング・ノードおよびスコアリング・パラメーターを設定します。
3. 展開タイプが「モデルのリフレッシュ」の場合、オプションで次を選択できます。
  - モデル作成ノードとモデル構築パラメーター
  - ストリームのスコアリングブランチのモデル・ナゲット
4. 「ストリームの説明をプレビュー」 ボタンをクリックします。

## ストリームの説明のエクスポート

ストリームの説明の内容を HTML ファイルにエクスポートできます。

ストリームの説明をエクスポートするには

1. メイン・メニューで次の各項目をクリックします。

「ファイル」 > 「ストリームの説明をエクスポート」

2. HTML ファイルの名前を入力して、「保存」をクリックします。

## ストリームの実行

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

- 「ツール」メニューから「実行」を選択します。
- ツールバーの「実行...」ボタンをクリックします。これらのボタンを利用すれば、ストリーム全体を実行することも、単に選択したターミナル・ノードだけを実行することもできます。詳しくは、トピック 20 ページの『IBM SPSS Modeler ツールバー』を参照してください。
- ターミナル・ノードを右クリックし、ポップアップ・メニューから「実行」をクリックして、1 つのデータ・ストリームを実行します。
- 非ターミナル・ノードを右クリックし、ポップアップ・メニューから「ここから実行」をクリックして、1 つのデータ・ストリームの一部を実行します。これにより、選択したノードの後のこれらの操作のみを実行できます。

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

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

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

## モデルの処理

ストリームにモデル作成ノード（ノード・パレットの「モデル作成」タブまたは「DB モデリング」タブのモデル作成ノード）が含まれている場合、ストリーム実行時に**モデル・ナゲット**が作成されます。モデル・

ナゲットは、**モデル**のコンテナです。つまり、入力データに対して予測を生成できる一連のルール、式または方程式で、予測分析の中心となるものです。



図 12. モデル・ナゲット

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

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

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

## ノードおよびストリームへのコメントおよび注釈の追加

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

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

ストリームまたはスーパーノードのすべてのコメントを一覧表示し、リスト内のコメントの順序を変更、コメント テキストを編集、コメントの前景または背景の色を変更できます。詳しくは、トピック 62 ページの『ストリームのコメントの一覧表示』を参照してください。

「ストリームのプロパティ」ダイアログ・ボックス、ノードのダイアログ・ボックス、またはモデル・ナゲットのウィンドウの「注釈」タブを使用して、説明をテキスト注釈の形式でストリーム、ノード、ナゲットに追加することもできます。これらの説明は、「注釈」タブが開いているときのみ表示できますが、ストリームの注釈を画面上のコメントとして表示することもできます。詳しくは、トピック 63 ページの『注釈』を参照してください。

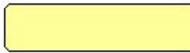
## コメント

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

注：ストリームの注釈を画面上のコメントとして表示することもできますが、これらをノードまたはナゲットに接続することはできません。詳しくは、トピック 62 ページの『注釈のコメントへの変換』を参照してください。

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

表3. コメントおよび注釈のテキスト・ボックスのモード

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

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

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

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

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

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

スーパーノードにコメントを追加することもできます。

**コメントに関する操作:** コメントに多くの操作を行うことができます。次の操作が可能です。

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

独立したコメントを追加するには

1. ストリームに何も選択されていないことを確認します。
2. 次のいずれかを実行します。
  - メイン・メニューで次の各項目をクリックします。

#### 「挿入」 > 「新規コメント」

- ストリームの背景を右クリックし、ポップアップ・メニューで「新規コメント」をクリックします。
  - ツールバーにある「新規コメント」ボタンをクリックします。
3. コメント テキストを入力 (またはクリップ ボードからテキストを貼り付け) ます。
  4. ストリームのノードをクリックして、コメントを保存します。

コメントをノードまたはナゲットに接続するには

1. ストリーム領域で 1 つまたは複数のノードまたはナゲットを選択します。
2. 次のいずれかを実行します。
  - メイン・メニューで次の各項目をクリックします。

#### 「挿入」 > 「新規コメント」

- ストリームの背景を右クリックし、ポップアップ・メニューで「新規コメント」をクリックします。
  - ツールバーにある「新規コメント」ボタンをクリックします。
3. コメント テキストを入力します。
  4. ストリームの別のノードをクリックして、コメントを保存します。

または、次を行うことができます。

5. 独立したコメントの挿入 (前項を参照)
6. 次のいずれかを実行します。
  - コメントを選択して F2 キーを押し、ノードまたはナゲットを選択します。
  - ノードまたはナゲットを選択して F2 キーを押し、コメントを選択します。
  - (3 ボタンのマウスのみ) マウス・ポインタをコメントに移動し、中央のボタンを押したままマウス・ポインタをノードまたはナゲットに移動して、マウス・ボタンを離します。

コメントを追加のノードまたはナゲットに接続するには

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

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

既存のコメントを編集するには

1. 次のいずれかを実行します。
  - コメント テキスト・ボックスをダブルクリックします。
  - テキスト・ボックスを選択して Enter キーを押します。
  - テキスト・ボックスを右クリックして、メニューを表示し、「編集」をクリックします。
2. コメント テキストを編集します。編集時、Ctrl + C キーを使用してテキストをコピーするなど、標準的な Windows ショートカット・キーを使用できます。編集中にそのほかのオプションが、コメントのポップアップ・メニューに表示されます。

3. テキスト・ボックスの外部を 1 回クリックするとサイズ変更のコントロールが表示され、もう 1 回クリックするとコメントが完了します。

コメント テキスト・ボックスをサイズ変更するには

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

既存のコメントを移動するには

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

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

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

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

コメントとノードまたはナゲットの接続を解除するには

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

コメントを削除するには

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

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

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

ストリームのコメントの表示または非表示

1. 次のいずれかを実行します。
  - メイン・メニューで次の各項目をクリックします。

「表示」 > 「コメント」

- ツールバーにある「コメントの表示/非表示」 ボタンをクリックします。

**ストリームのコメントの一覧表示:** 特定のストリームまたはスーパーノードに作成されたすべてのコメントのリストを表示できます。

このリスト上で、次のことができます。

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

コメントの一覧表示

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

- メイン・メニューで次の各項目をクリックします。

「ツール」 > 「ストリームのプロパティ」 > 「コメント」

- マネージャー・パネルでストリームを右クリックし、「ストリームのプロパティ」 を選択して 「コメント」 をクリックします。
- ストリーム領域上でストリームの背景を右クリックし、「ストリームのプロパティ」 を選択して 「コメント」 をクリックします。

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

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

**配置ボタン :** リスト内の選択したコメントを上下に移動させます。

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

**注釈のコメントへの変換:** ストリームまたはスーパーノードに作成された注釈は、コメントに変換できません。

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

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

ストリームの注釈をコメントに変換するには

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

スーパーノードの注釈をコメントに変換するには

1. 領域の「スーパーノード」アイコンをダブルクリックします。
2. 「注釈」タブをクリックします。
3. 「注釈をコメントとして表示」 チェック・ボックスを選択します。
4. 「OK」をクリックします。

## 注釈

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

### 注釈の追加

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

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

**名前。** ストリーム領域に表示されるノードについて、自動生成された名前を調整するか、固有の名前を作成するには、「カスタム」を選択します。

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

**キーワード:** プロジェクト・レポート内で使用したりストリーム内のノードを検索する場合に、リポジトリに保存されているオブジェクトの検索または追跡時に使用したりするキーワードを指定します (121 ページの『IBM SPSS Collaboration and Deployment Services Repository について』 を参照)。複数のキーワードをセミコロンで区切って指定できます income; crop type; claim value各キーワードの前後にある空白文字は削除されます。例えば、income ; crop type は、income;crop type と同じ結果になります。(ただし、キーワード内の空白文字は削除されません。例えば、スペースが 1 個の crop type とスペースが 2 個の crop type は同一ではありません。)

メイン・テキスト領域には、ノードの操作やノード作成時の決定事項などの操作に関する注釈を記入します。例えば、ストリームを共有、再利用する場合、「空白が多いフィールドをフィルター・ノードを使用して破棄する。」のような情報を記載しておけば、大変役に立ちます。ノードに注釈を付けると、ノードに情報が保存されます。これらの注釈を、プロジェクト・パネルを使用して作成したプロジェクト・レポートに表示するように設定することもできます。詳しくは、トピック 145 ページの『プロジェクトの概要』を参照してください。

**注釈をコメントとして表示 :** (ストリームおよびスーパーノードの注釈についてのみ) このボックスをチェックすると、注釈をストリーム領域上に表示される独立したコメントに変換します。詳しくは、トピック 58 ページの『ノードおよびストリームへのコメントおよび注釈の追加』を参照してください。

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

## データ・ストリームの保存

ストリームを作成したら、今後再利用するためにストリームを保存できます。

ストリームを保存するには

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

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

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

## ステートの保存

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

ステートを保存するには

1. 「ファイル」メニューで次の各項目をクリックします。  
「ステート」 > 「ステートの保存」または「名前を付けてステートを保存」
2. 「保存」ダイアログ・ボックスで、ステートを保存するフォルダーを指定します。

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

## ノードの保存

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

## 複数のストリーム・オブジェクトの保存

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

1. 保存するオブジェクトに対応するチェック・ボックスを選択してください。
2. 各オブジェクトを指定した場所に保存するには、「OK」をクリックします。

各オブジェクトに対応する標準の「保存」ダイアログ・ボックスが表示されます。保存が完了すると、アプリケーションが終了します。

## 出力の保存

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

1. 保存したい出力を表示するときは、出力ウィンドウのメニューで、次をクリックします。

「ファイル」 > 「保存」

2. 出力ファイルの名前と場所を指定します。
3. 必要に応じて、現在のプロジェクトにファイルを追加するには、「保存」ダイアログ・ボックスにある「ファイルをプロジェクトに追加」を選択します。詳しくは、トピック 145 ページの『プロジェクトの概要』を参照してください。

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

## 暗号化および復号化の情報

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

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

項目を暗号化するには

1. 「保存」ダイアログ・ボックスで、暗号化するアイテムについて、「オプション」をクリックします。「暗号化オプション」ダイアログ・ボックスが開きます。
2. 「このファイルを暗号化」を選択します。
3. オプションで、さらにセキュリティーを高めるために、「キーワードのマスク」を選択します。こうすることで、入力したものがすべてドットの列として表示されます。
4. パスワードを入力します。警告 :パスワードを忘れた場合は、ファイルまたはモデルは開くことができなくなります。
5. 「キーワードのマスク」を選択した場合は、パスワードを再入力して、正しくパスワードを入力したことを確認します。
6. 「OK」をクリックして、「保存」ダイアログ・ボックスに戻ります。

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

## ファイルのロード

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

- ストリーム (.str)

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

新しくファイルを開く

ストリームは、「ファイル」メニューから直接ロードすることができます。

1. 「ファイル」メニューの「ストリームを開く」をクリックします。

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

「モデル」 > 「モデルを開く」または「モデル・パレットをロード」

最近使用したファイルを開く

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

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

## データ・ストリームのマッピング

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

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

データのマッピングを行うには、次の2種類の方法があります。

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

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

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

## データのテンプレートへのマッピング

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

**ステップ 1 :**元の入力ノードの必須フィールドを指定します。ストリーム操作が適切に行われるためには、必須フィールドを指定する必要があります。詳しくは、トピック 68 ページの『必須フィールドの指定』を参照してください。

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

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

**ステップ 4 :**マップされたフィールドを確認します。表示されているダイアログ・ボックスで、置換データ・ソースからストリームへ、フィールドが正しくマッピングされているかどうかを確認します。マップされていない必須フィールドは赤色で表示されています。これらのフィールドはストリーム操作で使用されるので、下流の操作が正しく行われるようにするためには、新しいデータ・ソース中の適切なフィールドと置き換える必要があります。詳しくは、トピック 68 ページの『マップされたフィールドの調査』を参照してください。

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

## ストリーム間のマッピング

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

ストリーム間でデータをマップするには

1. 新規ストリームへの接続に使用するノードを右クリックします。
2. メニューで次の各項目をクリックします。

「データ・マッピング」 > 「マップ先」

3. カーソルを使用して、対象ストリーム上の目的のノードを選択します。

4. 表示されたダイアログ・ボックスで、フィールドが正しく対応しているかどうかを確認して、「OK」をクリックします。

## 必須フィールドの指定

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

必須フィールドを設定するには

1. 置換するテンプレート ストリームの入力ノードを右クリックします。
2. メニューで次の各項目をクリックします。

「データ・マッピング」 > 「必須フィールドの指定」

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

## マップされたフィールドの調査

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

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

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

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

---

## ヒントとショートカット

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

- **ストリームを素早く構築するにはダブルクリックを使用する:** パレット上のノードを単にダブルクリックすれば、現在のストリーム上にそのノードが追加、接続されます。

- キーの組み合わせを使用して下流のノードを選択する：Ctrl+Q キーと Ctrl+W キーを使用して、下流のすべてのノード選択を切り替えます。
- ショートカット・キーを使用してノードの接続、接続解除を行う：ストリーム領域上のノードを選択している場合、F2 キーを押した後、Tab キーを使用して接続先ノードを選択し、次に Shift キーを押しながらスペース・バーを押すことにより、元のノードと接続先ノードを接続することができます。選択したノードのすべての入力接続および出力接続を解除するには、F3 キーを押します。
- 「ノード・パレット」タブによく使うノードを設定する：「ツール」メニューの「パレットの管理」をクリックすると、ダイアログ・ボックスが開きます。ここで、ノード・パレット上に表示されるノードを追加、削除、または移動します。
- ノードの名前の変更と追加に関するヒント：各ノードのダイアログ・ボックスには「注釈」タブが用意されています。このタブでは、ノードのストリーム領域上での独自の名前を設定したり、ストリームを構築するために役立つツールヒントを追加することができます。また、進捗状況の追跡を行う目的で、または処理の説明、必要な作業、達成すべきビジネス上の目的などを記録するために、長い注釈を記入できる欄も用意されています。
- CLEM 式に値を自動的に入力する：さまざまなダイアログ・ボックス（フィールド作成ノードや置換ノードのダイアログ・ボックスなど）からアクセスできる式ビルダーを使用すると、CLEM 式にフィールド値を自動的に挿入できます。式ビルダーの値ボタンをクリックして、既存のフィールド値から選択してください。



図 13. 「値」ボタン

- ファイルを素早く参照する：「開く」ダイアログ・ボックスでファイルを参照するときに、「ファイル」リストを使用して（ダイアログ・ボックスの上部で「検索対象」フィールドの横にある黄色いひし形のボタンをクリックします）、前に使用したディレクトリーや IBM SPSS Modeler のデフォルト・ディレクトリーにアクセスします。進むボタンおよび戻るボタンを使用して、過去に利用したディレクトリーを参照することができます。
- 不要な出力ウィンドウを最小化する：出力ウィンドウの右上角にある赤い X ボタンを使えば、出力ウィンドウを素早く閉じて削除することができます。こうすることにより、マネージャー・パネルの「出力」タブに、注目する結果や興味のある結果だけを保持することができます。

このソフトウェアでは、さまざまなキーボード・ショートカットを利用することができます。詳しくは、トピック 171 ページの『キーボード・アクセシビリティ』を参照してください。

ご存じですか

- マウスをドラッグして、ストリーム領域上の複数のノード グループを選択することができます。
- あるストリームから別のストリームにノードをコピーして貼り付けることができます。
- ダイアログ・ボックスや出力ウィンドウからヘルプを参照することができます。
- CRISP-DM (Cross Industry Standard Process for Data Mining) に関するヘルプを参照できます。（「ヘルプ」メニューの「CRISP-DM ヘルプ」をクリックします。）



---

## 第 6 章 欠損値の処理

---

### 欠損値の概要

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

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

IBM SPSS Modeler の欠損値には、次の 2 種類があります。

- **ヌル値またはシステム欠損値**：これらの値は、データベースまたはソース・ファイルに空白のまま残された文字列以外の値であり、入力ノードまたはデータ型ノードで特に「欠損値」として定義されていません。システム欠損値は `$null$` 値として表示されます。空の文字列が特定のデータベースでヌルとして処理される場合でも、IBM 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。例えば、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**：これらは、入力ノードまたはデータ型ノード内で欠損値として明示的に定義されている `unknown`、`99`、`-1` などの値です。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。例えば、`@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。

**混在データの読み込み**：数値ストレージ（整数、実数、時間、タイムスタンプ、または日付のいずれか）をフィールドに読み込む場合は、数値以外のすべての値がヌル またはシステム欠損値 に設定されることに注意してください。これは、他のアプリケーションと異なり、フィールド内にストレージの混在を許さないためです。これを回避するためには、必要に応じて入力ノードまたは外部アプリケーション内の記憶域タイプを変更し、混在データを含むフィールドを文字列として読み込む必要があります。

**Oracle からの空の文字列の読み取り**：Oracle データベースとの間で読み書きするときには、Oracle が IBM SPSS Modeler やその他のほとんどのデータベースとは異なり、空の文字列値をヌル値と同様に処理および格納することに注意してください。つまり、Oracle データベースから抽出されたデータは、同じデータがファイルやその他のデータベースから抽出された場合とは異なって動作し、また異なる結果が返ることがあります。

---

## 欠損値の処理

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

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

一般的に、アプローチは次の 2 通りあります。

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

これら 2 つのアプローチは、データ検査ノードを使用し、大半を自動化することができます。例えば、フィルター・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。

## 欠損値を含むレコードの処理

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

- 条件抽出ノードを使用して、スタッフ レコードを削除できます。
- データ・セットが大きい場合、空白を含むレコードを破棄することができます。

## 欠損値を含むフィールドの処理

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

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

フィールドの尺度

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

**数値型フィールド:** 連続型 などの数値型フィールドの場合、多くのモデルは数値型フィールドに空白が含まれていると機能しないため、常に数値以外のすべての値を削除してからモデルを作成する必要があります。

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

スクリーニングまたは削除フィールド

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

- データ検査ノードを使用し、品質に基づいてフィールドをフィルタリングすることができます。
- フィールド選択ノードを使用して、指定した欠損値のパーセンテージ以上のフィールドをスクリーニングし、指定した目標値を基準にした重要度に基づいてフィールドをランク付けします。
- フィールドを削除する代わりに、データ型ノードを使用してフィールドの役割をなしに設定することもできます。これにより、フィールドはデータ・セット内に維持されますが、モデル作成プロセスからは除外されます。

---

## 欠損値の代入または置換

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

次の方法で、欠損値の代入ができます。

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

**無作為：** 正常または均一分布に基づいたランダム値で置き換えます。

**式：** ユーザー設定の式を指定することができます。例えば、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。フィルター・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。

代わりに、特定のフィールドに対して強制するには、データ型ノードを使用して、フィールドのデータ型が正しい値だけを利用するために、空白値の置換が必要なフィールドの 「検査」 列を 「強制」 に設定します。

---

## 欠損値用 CLEM 関数

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

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

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

次に示すように、フィールドのリスト全般にわたってヌルをカウントできます。

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

フィールドのリストを入力として受け入れる関数を使用する場合は、特殊関数 @FIELDS\_BETWEEN および @FIELDS\_MATCHING を使用できます。以下に例を示します。

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

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

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

このステートメントでは、範囲内にない値を、**\$null\$** として表示されるシステム欠損値で置き換えます。not() 関数を使うと、負の数値も含めた他のすべての数値を指定することもできます。詳しくは、トピック 118 ページの『空白値とヌル値処理関数』を参照してください。

#### レコードの破棄に関する注意

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

```
Drug = 'drugC' and not(@NULL(Drug))
```

前のバージョンでは、このような状況の場合ヌル値を除外していました。

---

## 第 7 章 CLEM 式の作成

---

### CLEM について

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

CLEM は、IBM SPSS Modeler で次の目的のために使用されます。

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

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

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

また、より複雑な CLEM 式を作成することも可能です。例えば、条件ルールのセットに基づいて、新規フィールドを派生させることができます。例えば、以下の式を使用して、新規の値カテゴリーを作成します。  
If: CardID = @OFFSET(CardID,1), Then: @OFFSET(ValueCategory,1), Else: 'exclude'

この例では @OFFSET 関数が使用され、「あるレコードの CardID フィールドの値が前のレコードの値と同じ場合は、前のレコードの ValueCategory と呼ばれるフィールドの値を返す」ことを指示しています。それ以外の場合は、文字列「exclude」を割り当てます。つまり、隣接するレコードの CardID が同じ場合、それらには同じ値カテゴリーが割り当てられます（exclude 文字列があるレコードは、後に、条件抽出ノードを使用して抜き出すことができます）。

---

### CLEM の例

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

簡単な式

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

$(After - Before) / Before * 100.0$

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

同様に、次の式は単にフィールド *salary* の各値の対数を返します。

```
log(salary)
```

複雑な式

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

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

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

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

次の式も同様です。

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

文字は常に単一逆引用符で囲むことに注意してください。

式中での関数の組み合わせ

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

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

この式は、次のようにさらに短い形式で記述することもできます。

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

よく使われる関数の組み合わせの例を、もう一つ次に示します。

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

この式は、フィールド *web\_page* 中の値を、値の最後の文字から先頭方向に向かって、文字 *`n`* を検索していきます。この式で値の長さに 7 のような特定の数字を使用すると、その値 (7 文字) 未満の値に対して式が不正になってしまいます。そこで、この式では *length* 関数と組み合わせることによって、動的に現在の値の長さを計算して使用します。

特殊関数

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

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

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

```
@BLANK(@FIELD)-> T
```

CLEM マニュアルには、ほかにもさまざまな例が記載されています。詳しくは、トピック 89 ページの『CLEM リファレンス概要』を参照してください。

---

## 値とデータ型

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

```
3
1.79
'banana'
```

次に、有効なフィールド名の例を示します。

```
Product_ID
'$P-NextField'
```

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

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

次の要素も使用できます。

- 文字コード: ``a`` や 3 など
- 項目のリスト: [1 2 3]、['Type 1' 'Type 2'] など

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

### 引用規則

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

- **文字列:** 文字列を指定する場合は、常に二重引用符を使用します ("Type 2" や "value")。単一引用符を使用することもできますが、引用符で囲まれたフィールドと誤解される危険性があります。
- **文字:** 常に ``` のような単一逆引用符を使用します。例えば、関数 `stripchar(`d`,`drugA`)` 内の文字 `d` を参照してください。これに関する唯一の例外は、文字列中の特定の文字を参照するために整数を使用する場合です。例えば、関数 `lowertoupper("druga"(5))` → "A" 内の文字 5 を参照してください。注: 英国および米国の標準キーボードでは、逆引用符文字 (アクサングラフ、Unicode 0060) のキーは、Esc キーのすぐ下にあります。
- **フィールド:** CLEM 式内で使用される場合、通常、フィールドは引用符で囲まれません (`subscr(2,arrayID)`) → CHAR)。スペースや他の特殊文字を囲む必要がある場合には、単一引用符を使用します ('Order Number')。データ・セット中に単一引用符で囲まれているのに未定義のフィールドがあると、それは文字列として読み込まれてしまいます。
- **パラメーター:** 常に単一引用符を使用します ('\$P-threshold')。

---

## 式と条件

CLEM 式が返す結果の例を次に示します (新しい値を作成する場合に使用)。

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

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

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

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

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

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

- `abs(Signal)` が計算され、その結果に対して `sqrt` が適用されます。
- `max(T1, T2)` が計算されます。
- 2 つの結果を乗算します。x は + より優先されます。
- 最後に、上の結果に `Baseline` が加算されます。

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

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

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

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

---

## ストリーム、セッション、およびスーパーノード・パラメーター

パラメーターは、CLEM 式とスクリプトで使用するために定義できます。実際のところ、パラメーターはユーザー定義の変数であり、保存されて、現在のストリーム、セッション、またはスーパーノードで存続します。さらに、スクリプトを使用する場合と同様に、ユーザー・インターフェースからもアクセスできます。例えば、ストリームを保存すると、そのストリームに設定されているパラメーターも保存されます (これは、ローカル・スクリプト変数と異なる点です。ローカル・スクリプト変数は、宣言されたスクリプト内でのみ使用できます)。多くの場合、パラメーターはスクリプトで使用され、スクリプト内でハードコーディングする必要のないフィールドや値の情報を提供することにより、スクリプトの動作を制御します。

パラメーターの有効範囲は、それがどこで設定されたかによって異なります。

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

パラメーターは、スーパーノード用にも設定できます。この場合、スーパーノード内にカプセル化されたノードでだけ表示できます。

### CLEM 式でのパラメーターの使用

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

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

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

CLEM 式に挿入するパラメーターをリストから選択することができます。詳しくは、トピック 84 ページの『フィールド、パラメーター、およびグローバル変数の選択』を参照してください。

---

## 文字列の処理

文字列に対しては、次のような操作を行うことができます。

- 大文字または小文字への文字列の変換 — `uppertolower(CHAR)`。
- 文字列変数からの ``ID_`` や ``$`` などの指定された文字の削除 — `stripchar(CHAR,STRING)`。
- 文字列変数の長さ（文字数）の判別 — `length(STRING)`。
- 文字列値のアルファベット順のチェック — `alphabefore(STRING1, STRING2)`。
- 値からの先頭または末尾の空白文字の削除 — `trim(STRING)`、`trim_start(STRING)`、または `trimend(STRING)`。
- 文字列からの最初または最後の `n` 文字の抽出 — `startstring(LENGTH, STRING)` または `endstring(LENGTH, STRING)`。例えば、製品名と 4 桁の ID コードを結合する (ACME CAMERA-D109) フィールド名 `item` があるとします。4 桁のコードのみを含む新規フィールドを作成するには、フィールド作成コードで次の式を指定します。

```
endstring(4, item)
```

- 特定のパターンの一致 — `STRING matches PATTERN`。例えば、職名に「market」が付いている人を選択するには、条件抽出ノードで次のように指定します。

```
job_title matches "*market*"
```

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

の下で以下の条件を指定します。 `replace('|',';',@FIELD)`

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

```
hassubstring(museums,"museum_of_design")
```

詳しくは、トピック 103 ページの『文字列関数』を参照してください。

---

## 空白および欠損値の処理

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

- @BLANK(FIELD) は、Age などの特定のフィールドの値が空白であるレコードを判別するために使用できます。
- @NULL(FIELD) は、指定されたフィールドの値がシステム欠損値であるレコードを判別するために使用できます。IBM SPSS Modeler では、システム欠損値は \$null\$ 値として表示されます。

詳しくは、トピック 118 ページの『空白値とヌル値処理関数』を参照してください。

---

## 数値の処理

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

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

詳しくは、トピック 100 ページの『数値関数』を参照してください。

---

## 時間と日付の処理

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

### 経過時間の算出

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

```
date_in_months(Date)
```

### 日付/時間値の比較

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

```
date_before(Date1, Date2)
```

### 差分の計算

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

```
date_weeks_difference(Date1, Date2)
```

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

今日の日付

関数 @TODAY を使用すれば、データ・セットに現在の日付を追加することができます。現在の日付は、「ストリームのプロパティ」ダイアログ・ボックスで指定されている日付の形式を使用して、指定されたフィールドまたは新しいフィールドに文字列として追加されます。詳しくは、トピック 108 ページの『日付および時刻の関数』を参照してください。

---

## 複数フィールドの要約

CLEM 言語には、複数フィールドにわたって要約統計量を返すさまざまな関数が含まれます。これらの関数は調査データの分析に特に役立つことがあり、その場合、1 つの質問に対する複数の回答を複数フィールドに保存することもできます。詳しくは、トピック 82 ページの『複数回答データの処理』を参照してください。

比較関数

複数フィールドにわたって、min\_n 関数と max\_n 関数を使用して値を比較できます。以下に例を示します。

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

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

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

同じフィールド・セットにわたってヌル値をカウントするには次のようにします。

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

この例でカウントされているのは、カードを保持する人たちの数ではなく、保持されているカードの数であることに注意してください。詳しくは、トピック 97 ページの『比較関数』を参照してください。

指定された値が複数のフィールドで出現する回数をカウントするには、count\_equal 関数を使用できます。以下の例では、値 Y を含むリスト内のフィールド数をカウントします。

```
count_equal("Y",[Answer1, Answer2, Answer3])
```

リスト内のフィールドが以下の値の場合、示されたように関数は値 ■Y の結果を返します。

表 4. 関数値:

Answer1	Answer2	Answer3	カウント
Y	N	Y	2
Y	N	N	1

数値関数

複数フィールドにわたって、sum\_n 関数、mean\_n 関数、および sdev\_n 関数を使用して統計を取得できます。以下に例を示します。

```
sum_n(['card1ba1' 'card2ba1' 'card3ba1'])
mean_n(['card1ba1' 'card2ba1' 'card3ba1'])
```

詳しくは、トピック 100 ページの『数値関数』を参照してください。

フィールドのリストを生成

フィールドのリストを入力として受け入れる関数を使用する場合は、特殊関数 @FIELDS\_BETWEEN(start, end) および @FIELDS\_MATCHING(pattern) を入力として使用できます。例えば、フィールドの順序が上記の sum\_n の例に示すようなものと仮定すると、以下は同等になります。

```
sum_n(@FIELDS_BETWEEN(card1ba1, card3ba1))
```

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

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

詳しくは、トピック 119 ページの『特殊フィールド』を参照してください。

---

## 複数回答データの処理

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

- value\_at
- first\_index / last\_index
- first\_non\_null / last\_non\_null
- first\_non\_null\_index / last\_non\_null\_index
- min\_index / max\_index

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

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

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

表 5. 車のランク付けの例:

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

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

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

詳しくは、トピック 97 ページの『比較関数』を参照してください。

複数回答セットの参照

特別な @MULTI\_RESPONSE\_SET 関数を使用して、複数回答セットのすべてのフィールドを参照することができます。例えば、上記の 3 つの *car* フィールドが *car\_rankings* という名前の複数回答セットにある場合、次の関数が同じ結果を返します。

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

---

## Clem 式ビルダー

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

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

## Clem 式ビルダーへのアクセス

Clem 式ビルダーは、条件抽出、バランス、フィールド作成、置換、精度分析、レポート、およびテーブル・ノードの式などの CLEM 式を使うノードのすべてで使用できます。式フィールドのすぐ右にある計算機ボタンをクリックして、式ビルダーを開くことができます。

## CLEM 式の作成

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

Clem 式ビルダーを使用して CLEM 式を作成する手順

1. Clem 式フィールドとフィールドのリストを参照しながら、ウィンドウに入力します。

または

2. スクロール・リストから適切なフィールドと関数を選択します。
3. 黄色の矢印ボタンをダブルクリックまたはクリックして、CLEM 式フィールドにフィールドや関数を追加します。
4. オペランドを式に挿入するには、ダイアログ・ボックスの中央にあるオペランド ボタンを使用します。

## 関数の選択

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

これらのカテゴリーの多くは、CLEM の言語記述の参照セクションに説明されています。詳しくは、トピック 94 ページの『関数のリファレンス』を参照してください。

他のカテゴリーは次の通りです。

- **一般的な関数:** 最も使用される関数のうちいくつかの選択が表示されます。

- **最近使用した演算子:** 現在のセッション内で使われた CLEM 関数が表示されます。
- **@ 関数:** 名前の前に「@」記号が付く、すべての特別な関数のリストが表示されます。
- **データベース関数:** ストリームにデータベース接続が含まれている場合 (データベース入力ノードによって)、この選択にはユーザー定義関数 (UDF) など、そのデータベースから使用できる関数が表示されます。
- **データベース集計:** ストリームに (データベース・ソース・ノードによって) データベース接続が含まれている場合、これを選択すると、そのデータベースで使用できる集計オプションがリストされます。
- **データベース・ウィンドウ集計:** ストリームに (データベース・ソース・ノードによって) データベース接続が含まれている場合、これを選択すると、そのデータベースで使用できるウィンドウ集計オプションがリストされます。

これらのオプションは、「集計」という名前が付いていますが、レコード集計ノードで使用するものではなく、フィールド作成または選択などのノードで使用するのに適しています。これは、これらのオプションの出力が実際のレコード集計ではなくスカラーであるためです。したがって、出力に示されるデータの量をレコード集計ノードと同じ方法で削減することはありません。例えば、この種の集計を使用してデータの行の移動平均 (「現在行とそれより前のすべての行の平均」など) を求めることができます。

- **組み込み集計:** 使用できる集計モードのリストを表示します。
- **演算子:** 式を作成するときに使用できるすべての演算子が表示されます。演算子は、ダイアログ・ボックスの中央にあるボタンからも使用できます。
- **すべての関数:** 利用できるすべての CLEM 関数が表示されます。

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

## フィールド、パラメーター、およびグローバル変数の選択

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

詳しくは、トピック 78 ページの『ストリーム、セッション、およびスーパーノード・パラメーター』を参照してください。

フィールドの他に、次の項目から選択することもできます。

**複数回答セット:** 詳しくは、「*IBM SPSS Modeler* ソース、プロセス、および出力ノード」を参照してください。

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

**パラメーター:** 詳しくは、トピック 78 ページの『ストリーム、セッション、およびスーパーノード・パラメーター』を参照してください。

**グローバル値:** 詳しくは、「*IBM SPSS Modeler* ソース、プロセス、および出力ノード」を参照してください。

## 値の表示または選択

フィールドの値は、式ビルダー、データ検査レポートを含めたシステム内の多くの場所からと、時間区分ノード内で将来の値を編集しているときに、表示できます。この機能を使用するには、データがソースまたはデータ型ノード内で完全にインスタンス化されていることが必要であり、そのようにすると、ストレージ、データ型、および値が明らかになります。

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



図 14. 値ピッカーボタン

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

## CLEM 式の検査

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

検査する項目を次に示します。

- 値やフィールド名の引用符が正しいかどうか
- パラメーターやグローバル変数の使用法が正しいかどうか
- 演算子が適切に使用されているかどうか
- 参照されているフィールドが存在しているかどうか
- 参照されているグローバル値が存在しているか、また定義が正しいかどうか

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

---

## 検索と置換

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

1. テキスト領域内にカーソルを置いて、Ctrl + F キーを押して「検索/置換」ダイアログ・ボックスにアクセスします。
2. 検索するテキストを入力するか、最近検索した項目のドロップダウン・リストから選択します。
3. 置換テキストがある場合は、入力します。
4. 「次を検索」 をクリックして、検索を開始します。

- 「置換」をクリックして現在の選択内容を置換するか、「すべて置換」をクリックしてすべてまたは選択したインスタンスを更新します。
- 各操作が終了すると、ダイアログ・ボックスが閉じます。テキスト領域で F3 を押すと最後の検索操作が繰り返され、または Ctrl + F キーを押すとダイアログに再度アクセスします。

#### 検索オプション

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

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

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

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

#### 正規表現シンタックス

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

表 6. 文字の一致：

文字	一致
x	文字 x
¥	円記号
¥0n	8 進法の値を持つ文字 0n (0 ≤ n ≤ 7)
¥0nn	8 進法の値を持つ文字 0nn (0 ≤ n ≤ 7)
¥0mnn	8 進法の値を持つ文字 0mnn (0 ≤ m ≤ 3, 0 ≤ n ≤ 7)
¥xhh	16 進法の値を持つ文字 0xhh
¥uhhhh	16 進法の値を持つ文字 0xhhhh
¥t	タブ文字 (¥u0009)
¥n	改行文字 (¥u000A)
¥r	復帰文字 (¥u000D)
¥f	改ページ文字 (¥u000C)
¥a	アラート (バル) 文字 (¥u0007)
¥e	エスケープ文字 (¥u001B)
¥cx	xに対応する制御文字

表 7. 文字クラス的一致：

文字クラス	一致
[abc]	a、b、または c (単純クラス)
[^abc]	a、b、または c 以外の文字 (減法)
[a-zA-Z]	a から z または A から Z の各文字 (範囲)

表7. 文字クラスの一致 (続き):

文字クラス	一致
「a-d[m-p]」	a から d、または m から p (結合)。また、これを [a-dm-p] と指定することもできます。
「a-z&&[def]」	a から z、および d、e、または f (交差)
「a-z&&[^bc]」	a から z、ただし b と c 以外 (減法)。また、これを [ad-z] と指定することもできます。
「a-z&&[^m-p]」	a から z、ただし m から p は含まない (減法)。また、これを [a-lq-z] と指定することもできます。

表8. 事前設定された文字クラス:

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

表9. 境界の一致:

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



---

## 第 8 章 CLEM 言語リファレンス

---

### CLEM リファレンス概要

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

CLEM 式は、値、フィールド名、演算子、および関数で構成されます。正しい構文を使用して、さまざまなデータ操作を作成することができます。

---

### CLEMデータ型

CLEM データ型は次のいずれかで構成できます。

- 整数値
- 実数
- 文字
- 文字列
- リスト
- フィールド
- 日付/時刻

#### 引用符の使用規則

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

- 文字列: 文字列を指定する場合は、常に二重引用符を使用します (例: "Type 2")。単一引用符を使用することもできますが、引用符で囲まれたフィールドと誤解される危険性があります。
- フィールド: スペースや他の特殊文字を囲む必要がある場合にのみ、単一引用符を使用します (例: 'Order Number')。データ・セット中に単一引用符で囲まれているのに未定義のフィールドがあると、それは文字列として読み込まれてしまいます。
- パラメーター: パラメーターを使用する場合は、常に単一引用符を使用します (例: '\$P-threshold')。
- 文字: 常に単一逆引用符 (^) を使用します (例: stripchar(`d`, "drugA"))。

これらの規則は、以降の項目で詳細に説明しています。

### 整数値

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

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

## 実数

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

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

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

```
Na > 0.6
```

次のようには、指定できません。

```
Na > 0,6
```

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

## 文字

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

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

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

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

## 文字列

基本的に、文字列は二重引用符で囲んでください。文字列の例は "c35product2" や "referrerID" などです。文字列内で特殊文字を指定するには、"¥\$65443" のように円記号を使用します。(円記号文字を指定するには、¥ のように円記号を 2 つ使用します)。文字列を単一引用符で囲むこともできますが、その場合引用符で囲まれたフィールド ('referrerID') と区別できない可能性があります。詳しくは、トピック 103 ページの『文字列関数』を参照してください。

## リスト

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

## フィールド

CLEM 式内で、関数名以外の名前はフィールド名とみなされます。これらは単純に Power、val27、state\_flag のように記述できますが、名前が数字から始まる場合またはスペースなどアルファベット以外の文字（アンダースコアを除く）を含む場合は、'Power Increase'、'2nd answer'、'#101'、'\$P-NextField' のように名前を単一引用符で囲みます。

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

## 日付

日付の計算は、基準日に基づいて行われます。基準日は、「ストリームのプロパティ」ダイアログ・ボックスで指定します。デフォルトの基準日は、1900 年 1 月 1 日です。

CLEM 言語は、次の日付の形式をサポートします：

表 10. CLEM 言語の日付の形式

形式	例
DDMMYY	150163
MMDDYY	011563
YYMMDD	630115
YYYYMMDD	19630115
YYYYDDD	4 桁の年、およびそれに続く年単位の日を表す 3 桁の数字。例えば、2000032 は、2000 年の 32 番目の日付、つまり 2000 年 2 月 1 日を表します。
DAY	現在のロケールの曜日。例えば、英語の場合、Monday、Tuesday などです。
MONTH	現在のロケールの月。例えば、January、February などです。
DD/MM/YY	15/01/63
DD/MM/YYYY	15/01/1963
MM/DD/YY	01/15/63
MM/DD/YYYY	01/15/1963
DD-MM-YY	15-01-63
DD-MM-YYYY	15-01-1963
MM-DD-YY	01-15-63
MM-DD-YYYY	01-15-1963
DD.MM.YY	15.01.63
DD.MM.YYYY	15.01.1963
MM.DD.YY	01.15.63
MM.DD.YYYY	01.15.1963
DD-MON-YY	15-JAN-63、15-jan-63、15-Jan-63
DD/MON/YY	15/JAN/63、15/jan/63、15/Jan/63
DD.MON.YY	15.JAN.63、15.jan.63、15.Jan.63
DD-MON-YYYY	15-JAN-1963、15-jan-1963、15-Jan-1963
DD/MON/YYYY	15/JAN/1963、15/jan/1963、15/Jan/1963

表 10. CLEM 言語の日付の形式 (続き)

形式	例
DD.MON.YYYY	15.JAN.1963、15.jan.1963、15.Jan.1963
MON YYYY	Jan 2004
q Q YYYY	四半期を表す 1 桁の数字 (1 から 4)、およびそれに続く文字 <i>Q</i> と 4 桁の年。例えば、2004 年 12 月 25 日は、4 Q 2004 とし て表されます。
ww WK YYYY	1 年の内での週を表す 2 桁の数で、次に文字 <i>WK</i> と 4 桁の年 が続きます。年内の週は、週の最初の日が月曜日で、また、少な くとも最初の週に 1 日以上あるという仮定の下に計算されま す。

## 時間

CLEM 言語は、次の時間の形式をサポートします:

表 11. CLEM 言語の時間の形式

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

## CLEM 演算子

次の演算子が利用できます。

表 12. CLEM 言語の演算子:

演算	コメント	優先順位 (次項参照)
または	2 つの CLEM 式間で使用されます。どちらかが真 (true) の場合、または両方が真 (true) の場合に、値を返します。	10
and	2 つの CLEM 式間で使用されます。両方が真 (true) の場合に、値を返します。	9

表 12. CLEM 言語の演算子 (続き) :

演算	コメント	優先順位 (次項参照)
=	任意の比較可能な 2 つの項目間で使用されます。 ITEM1 が ITEM2 と等しい場合に true を返します。	7
==	= と同等	7
/=	任意の比較可能な 2 つの項目間で使用されます。 ITEM1 が ITEM2 と等しくない場合に真が返されます。	7
/==	/= と同等	7
>	任意の比較可能な 2 つの項目間で使用されます。 ITEM1 が ITEM2 より大きい場合に true を返します。	6
>=	任意の比較可能な 2 つの項目間で使用されます。 ITEM1 が ITEM2 以上の場合に true を返します。	6
<	任意の比較可能な 2 つの項目間で使用されます。 ITEM1 が厳密に ITEM2 より小さい場合に真を返します。	6
<=	任意の比較可能な 2 つの項目間で使用されます。 ITEM1 が ITEM2 以下の場合に true を返します。	6
&&=_0	2 つの整数間に用いられます。 ブール式の $INT1 \ \&\& \ INT2 = 0$ と同じになります。	6
&&/=_0	2 つの整数間に用いられます。 ブール式の $INT1 \ \&\& \ INT2 = 0$ と同じになります。	6
+	2 つの数値で加算します。 $NUM1 + NUM2$ 。	5
>>	2 つの文字列を次のように連結します。 $STRING1 \ \>\> \ STRING2$ 。	5
-	1 つの数値をもう 1 つの数値から減算します。 $NUM1 - NUM2$ 。1 つの数値の前にも使用できます。 $- \ NUM$ 。	5
*	2 つの数値を乗算するのに使用されます。 $NUM1 * NUM2$ 。	4
&&	2 つの整数間に用いられます。 結果は、整数 $INT1$ と $INT2$ のビット単位の「論理積」になります。	4
&&~	2 つの整数間に用いられます。 結果は、 $INT1$ と、 $INT2$ のビット単位の補数との、ビット単位の「論理積」になります。	4
	2 つの整数間に用いられます。 結果は、 $INT1$ と $INT2$ のビット単位の「包含論理和」になります。	4
~	整数の前に用いられます。 $INT$ のビット単位の補数を生成します。	4

表 12. CLEM 言語の演算子 (続き) :

演算	コメント	優先順位 (次項参照)
/&	2 つの整数間に用いられます。結果は、INT1 と INT2 のビット単位の「排他的論理和」になります。	4
INT1 << N	2 つの整数間に用いられます。N の数だけ位置を左にシフトした INT のビット・パターンを生成します。	4
INT1 >> N	2 つの整数間に用いられます。N の数だけ位置を右にシフトした INT のビット・パターンを生成します。	4
/	1 つの数値をもう 1 つの数値で除算するのに使われます: NUM1 / NUM2。NUM1 / NUM2。	4
**	2 つの数字の間に用いられます: BASE ** POWER。BASE の POWER 乗を返します。	3
rem	2 つの整数間に用いられます。剰余 INT1 - (INT1 div INT2) * INT2 を返します。	2
div	2 つの整数間に用いられます。整数の除算を実行します。	2

### 演算子の優先順位

優先順位は、複数の 2 項演算子を使ったカッコで囲まれていない式などの、複雑な式の解析方法を定めるものです。例を次に示します。

3 + 4 \* 5

相対的な優先順位に従って \* は + より前に解析されるため、(3 + 4) \* 5 ではなく 3 + (4 \* 5) として解析されます。CLEM 言語中のすべての演算子には、それに対応した優先順位があります。この値が小さいほど、その演算子は処理リスト上で重要な意味を持ち、他の演算子よりも先に処理されます。

## 関数のリファレンス

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

表 13. IBM SPSS Modeler データで使用される CLEM 関数

関数の種類	説明
情報	フィールド値を詳しく調べる場合に用いられます。例えば、関数 is_string は、データ型が文字列型のすべてのレコードに対して真を返します。
変換	新しいフィールドの作成や、ストレージ・タイプの変換に用いられます。例えば、関数 to_timestamp は選択されているフィールドをタイムスタンプに変換します。
比較	フィールドの値を互いに比較したり、指定した文字列と比較する場合に用いられます。例えば、<= は、あるフィールドの値がもう 1 つのフィールドの値以下かどうかを比較します。

表 13. IBM SPSS Modeler データで使用される CLEM 関数 (続き)

関数の種類	説明
論理	if、then、else などの論理演算を行うために用いられます。
数値	フィールド値の自然ログ数の算出など、数値計算に用いられます。
三角関数	指定された角度のアーコサインの算出など、三角関数の計算に用いられます。
確率	学生からの $t$ 分布値が特定値に満たなくなる確率など、さまざまな分布を基準にして確率を返します。
ビット単位	整数をビット・パターンとして操作する場合に用いられます。
Random	無作為に項目を選択したり、無作為な 数字を生成するために用いられます。
ストリング	指定した文字を削除する stripchar のように、文字列に関するさまざまな操作を行うために用いられます。
SoundEx	正しいスペルが分からない場合に、特定文字の発音方法についての音声的な仮定を基準にして、文字列を検索するために用いられます。
日時	日付、時間、タイムスタンプ・フィールドに対してさまざまな操作を行うために用いられます。
シーケンス	データ・セットのレコード・シーケンスの詳細を調べたり、そのシーケンスに基づいた操作を行うために用いられます。
グローバル	グローバル・ノードが作成したグローバル値にアクセスするために用いられます。例えば、@MEAN は、データ・セット全体のフィールドのすべての値の平均を参照するために用いられます。
空白とヌル	アクセス、フラグ設定、およびユーザーが指定した空白やシステム欠損値を埋めるために用いられます。例えば、@BLANK(FIELD) は、空白があるレコードに真のフラグを設定するために用いられます。
特殊フィールド	調査対象の特定のフィールドを表すために用いられます。例えば、@FIELD は複数のフィールドを作成する場合に用いられます。

## 関数の表記方法について

このガイドでは、次の規約を関数中のアイテムを参照するために使用します。

表 14. 関数の表記方法について：

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

表 14. 関数の表記方法について (続き):

表記方法	説明
<i>TIME</i>	<i>power_flux</i> のような時刻フィールド。ここで、値の形式は HHMMSS のようになります。

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

表 15. *rem* 関数の説明:

関数	結果	説明
<i>INT1 rem INT2</i>	<i>Number</i>	<i>INT1</i> を <i>INT2</i> で除算した剰余を返します。つまり、 $INT1 - (INT1 \text{ div } INT2) * INT2$ です。

項目をリストにする方法や、関数内で文字を指定する方法などの使用方法の詳細は、別の場所で説明されています。詳しくは、トピック 89 ページの『CLEMデータ型』を参照してください。

## 情報関数

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

表 16. CLEM 情報関数:

関数	結果	説明
@BLANK(FIELD)	<i>Boolean</i>	上流のデータ型ノードまたは入力ノードで設定された空白処理規則 (「データ型」タブ) にしたがって、値が空白のレコードに対して真を返します。
@NULL(ITEM)	<i>Boolean</i>	値が未定義のすべてのレコードに対して真を返します。未定義の値はシステムのヌル値で、IBM SPSS Modeler では \$null\$ として表されます。
<i>is_date</i> (ITEM)	<i>Boolean</i>	データ型が日付のすべてのレコードに対して真 (true) を返します。
<i>is_datetime</i> (ITEM)	<i>Boolean</i>	データ型が日付、時間またはタイムスタンプのすべてのレコードに対して真 (true) を返します。
<i>is_integer</i> (ITEM)	<i>Boolean</i>	データ型が整数のすべてのレコードに対して真 (true) を返します。
<i>is_number</i> (ITEM)	<i>Boolean</i>	データ型が数値のすべてのレコードに対して真 (true) を返します。
<i>is_real</i> (ITEM)	<i>Boolean</i>	データ型が実数のすべてのレコードに対して真 (true) を返します。
<i>is_string</i> (ITEM)	<i>Boolean</i>	データ型が文字列のすべてのレコードに対して真 (true) を返します。
<i>is_time</i> (ITEM)	<i>Boolean</i>	データ型が時間のすべてのレコードに対して真 (true) を返します。
<i>is_timestamp</i> (ITEM)	<i>Boolean</i>	データ型がタイムスタンプのすべてのレコードに対して真 (true) を返します。

## 変換関数

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

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形式に依存します。例えば、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形式として一致する「**MON YYYY**」を選択します。

表 17. CLEM 変換関数

関数	結果	説明
<code>ITEM1 &gt;&gt; ITEM2</code>	ストリング	2 つのフィールドの値を連結し、結果の文字列を <code>ITEM1ITEM2</code> の形式で返します。
<code>to_integer(ITEM)</code>	<i>Integer</i>	指定されたフィールドのストレージを整数に変換します。
<code>to_real(ITEM)</code>	実数	指定されたフィールドのストレージを実数に変換します。
<code>to_number(ITEM)</code>	<i>Number</i>	指定されたフィールドのストレージを数値に変換します。
<code>to_string(ITEM)</code>	ストリング	指定されたフィールドのストレージを文字列に変換します。
<code>to_time(ITEM)</code>	時間	指定されたフィールドのストレージを時間に変換します。
<code>to_date(ITEM)</code>	<i>Date</i>	指定されたフィールドのストレージを日付に変換します。
<code>to_timestamp(ITEM)</code>	タイム・スタンプ	指定されたフィールドのストレージをタイムスタンプに変換します。
<code>to_datetime(ITEM)</code>	<i>Datetime</i>	指定されたフィールドのストレージを日付、時間またはタイムスタンプ値に変換します。
<code>datetime_date(ITEM)</code>	<i>Date</i>	数値、文字列またはタイムスタンプの日付値を返します数値 (秒単位) を日付へ変換しなおすことができるのは、この関数だけです。ITEM が文字列の場合は、現在のデータ形式で文字列を解析することにより日付を作成します。この関数が正常に機能するためには、「ストリームのプロパティ」ダイアログ・ボックスの「日付の形式」に、正しい値が指定されていなければなりません。ITEM が数値の場合は、基準日 (または紀元) からの秒数として解釈します。日付の端数は切り捨てられます。ITEM がタイムスタンプの場合は、日付をタイムスタンプの一部として返します。ITEM が日付の場合は、変更せずに返します。

## 比較関数

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

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

表 18. CLEM 比較関数

関数	結果	説明
count_equal(ITEM1, LIST)	<i>Integer</i>	フィールドの LIST から ITEM1 と等しい数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_greater_than(ITEM1, LIST)	<i>Integer</i>	フィールドの LIST から ITEM1 より大きい数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_less_than(ITEM1, LIST)	<i>Integer</i>	フィールドの LIST から ITEM1 より小さい数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_not_equal(ITEM1, LIST)	<i>Integer</i>	フィールドの LIST から ITEM1 と等しくない数値を返します。ITEM1 が NULL の場合は、NULL を返します。
count_nulls(LIST)	<i>Integer</i>	フィールドの LIST から NULL 値の数を返します。
count_non_nulls(LIST)	<i>Integer</i>	フィールドの LIST から NULL 値以外の数を返します。
date_before(DATE1, DATE2)	<i>Boolean</i>	日付値の順序の確認に用いられます。DATE1 が DATE2 より前の場合に真 (true) を返します。
first_index(ITEM, LIST)	<i>Integer</i>	フィールドの LIST から ITEM を含む最初のフィールドの索引、または値が見つからないなら 0 を返します。サポート対象は、文字列、整数、実数型のみです。
first_non_null(LIST)	<i>Any</i>	提供されるフィールド・リストの最初の非ヌル値を返します。ストレージ・タイプはすべてサポート対象です。
first_non_null_index(LIST)	<i>Integer</i>	非ヌル値を含む特定の LIST の最初のフィールドの索引、またはすべての値がヌル 0 を返します。ストレージ・タイプはすべてサポート対象です。
ITEM1 = ITEM2	<i>Boolean</i>	ITEM1 が ITEM2 と等しい場合に真 (true) を返します。
ITEM1 /= ITEM2	<i>Boolean</i>	2 つの文字列が異なるか、同じでも 0 の場合に真 (true) を返します。
ITEM1 < ITEM2	<i>Boolean</i>	ITEM1 が ITEM2 より小さい場合に真 (true) を返します。
ITEM1 <= ITEM2	<i>Boolean</i>	ITEM1 が ITEM2 以下の場合に真 (true) を返します。
ITEM1 > ITEM2	<i>Boolean</i>	ITEM1 が ITEM2 より大きい場合に真 (true) を返します。
ITEM1 >= ITEM2	<i>Boolean</i>	ITEM1 が ITEM2 以上の場合に真 (true) を返します。
last_index(ITEM, LIST)	<i>Integer</i>	フィールドの LIST から ITEM を含む最新のフィールドの索引、または値が見つからないなら 0 を返します。サポート対象は、文字列、整数、実数型のみです。
last_non_null(LIST)	<i>Any</i>	提供されるフィールド・リストの最後の非ヌル値を返します。ストレージ・タイプはすべてサポート対象です。
last_non_null_index(LIST)	<i>Integer</i>	非ヌル値を含む特定の LIST の最後のフィールドの索引、またはすべての値がヌル 0 を返します。ストレージ・タイプはすべてサポート対象です。
max(ITEM1, ITEM2)	<i>Any</i>	ITEM1 または ITEM2 のどちらか大きい方を返します。
max_index(LIST)	<i>Integer</i>	数値フィールドの LIST から最大値を含むフィールドの索引、またはすべての値がヌルなら 0 を返します。例えば、3番目にリストされたフィールドに最大値がある場合は、インデックス値 3 を返します。複数のフィールドに最大値がある場合は、最初にリストされたもの (左端) を返します。
max_n(LIST)	<i>Number</i>	数値フィールドの LIST から最大値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。

表 18. CLEM 比較関数 (続き)

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

## 論理関数

CLEM 式を使用して論理演算を行うことができます。

表 19. CLEM 論理関数

関数	結果	説明
COND1 and COND2	Boolean	この演算は論理積で、COND1 と COND2 の両方が真 (true) の場合に真 (true) の値を返します。COND1 が偽 (false) の場合、COND2 は評価されません。こうすることによって、COND2 の演算が正しいかどうかを COND1 で判断することができます。例えば、length(Label) >=6 and Label(6) = 'x' です。
COND1 or COND2	Boolean	この演算は、(包括的) 論理和で、COND1 と COND2 のどちらかが真 (true) の場合、または両方とも真 (true) の場合に、真 (true) の値を返します。COND1 が真 (true) の場合、COND2 は評価されません。
not(COND)	Boolean	この演算は論理否定で、COND が偽 (false) の場合に真 (true) の値を返します。それ以外の場合、この演算は 0 の値を返します。
if COND then EXPR1 else EXPR2 endif	Any	この演算は条件評価です。COND が真 (true) の場合、この演算は EXPR1 の結果を返します。それ以外の場合は、EXPR2 を評価した結果を返します。
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	Any	この演算は複数の分岐を持つ条件評価です。COND1 が真 (true) の場合、この演算は EXPR1 の結果を返します。それ以外の場合で、COND2 が真ならば、この演算は EXPR2 を評価した結果を返します。それ以外の場合は、EXPR_N を評価した結果を返します。

## 数値関数

CLEM には、一般的に使われるさまざまな数値関数が用意されています。

表 20. CLEM 数値関数：

関数	結果	説明
-NUM	Number	NUM を否定する場合に用いられます。対応する数値の符号を逆にした値を返します。
NUM1 + NUM2	Number	NUM1 と NUM2 を合計した値を返します。
NUM1 - NUM2	Number	NUM1 から NUM2 を減算した値を返します。
NUM1 * NUM2	Number	NUM1 を NUM2 で乗算した値を返します。
NUM1 / NUM2	Number	NUM1 を NUM2 で除算した値を返します。
INT1 div INT2	Number	整数の除算を行うために用いられます。INT1 を INT2 で除算された値を返します。
INT1 rem INT2	Number	INT1 を INT2 で除算した剰余を返します。つまり、 $INT1 - (INT1 \text{ div } INT2) * INT2$ です。
INT1 mod INT2	Number	この関数は廃止されました。代わりに rem 関数を使用します。
BASE ** POWER	Number	POWER 乗までべき乗した BASE を返します。BASE と POWER はどちらも任意の数値です (ただし、POWER が整数の 0 以外のいずれかのデータ型のゼロの場合、BASE はゼロ以外である必要があります)。POWER が整数の場合は、BASE のべき乗を順次掛けていくことによって計算されます。したがって、BASE が整数の場合、結果は整数になります。POWER が整数の 0 の場合、結果は常に BASE と同じデータ型の 1 になります。POWER が整数ではない場合、結果は $\exp(\text{POWER} * \log(\text{BASE}))$ のように計算されます。
abs(NUM)	Number	NUM の絶対値を返します。この値は常に、同じデータ型の数値になります。
exp(NUM)	実数	NUM 乗までべき乗した e を返します。この e は自然対数の底です。
fracof(NUM)	実数	$\text{NUM} - \text{intof}(\text{NUM})$ として定義される、NUM の小数部を返します。
intof(NUM)	Integer	引数を切り捨てて整数にします。NUM と同じ符号で、 $\text{abs}(\text{INT}) \leq \text{abs}(\text{NUM})$ となる最大の絶対値を持つ整数を返します。
log(NUM)	実数	NUM の自然対数 (底 e) を返します。NUM は、ゼロ以外でなければなりません。
log10(NUM)	実数	NUM の常用対数を返します。この関数は、 $\log(\text{NUM}) / \log(10)$ として定義されます。
negate(NUM)	Number	NUM を否定する場合に用いられます。対応する数値の符号を逆にした値を返します。
round(NUM)	Integer	NUM が正の数の場合は $\text{intof}(\text{NUM} + 0.5)$ 、NUM が負の数の場合は $\text{intof}(\text{NUM} - 0.5)$ を使用して、NUM を整数に丸めるために用いられます。
sign(NUM)	Number	NUM の符号を判断するために用いられます。NUM が整数の場合、この演算は -1, 0, または 1 を返します。NUM が実数の場合、NUM が負、0, または正の値かによって、-1.0, 0.0, または 1.0 を返します。
sqrt(NUM)	実数	NUM の平方根を返します。NUM は正でなければなりません。

表 20. CLEM 数値関数 (続き) :

関数	結果	説明
sum_n(LIST)	Number	数値フィールドの LIST から合計値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。
mean_n(LIST)	Number	数値フィールドの LIST から平均値を返します。フィールドのすべての値が NULL の場合は、NULL を返します。
sdev_n(LIST)	Number	数値フィールドの LIST から標準偏差を返します。フィールドのすべての値が NULL の場合は、NULL を返します。

## 三角関数

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

表 21. CLEM 三角関数

関数	結果	説明
arccos(NUM)	実数	指定した角度のアーコサインを計算します。
arccosh(NUM)	実数	指定した角度の双曲線アーコサインを計算します。
arcsin(NUM)	実数	指定した角度のアークサインを計算します。
arcsinh(NUM)	実数	指定した角度の双曲線アークサインを計算します。
arctan(NUM)	実数	指定した角度のアークタンジェントを計算します。
arctan2(NUM_Y, NUM_X)	実数	NUM_Y / NUM_X のアークタンジェントを計算し、2 つの数値の符号を使用して、象限情報を派生させます。結果は、 $-\pi < \text{ANGLE} \leq \pi$ (radians) - $180 < \text{ANGLE} \leq 180$ (degrees) の範囲内の実数です。
arctanh(NUM)	実数	指定した角度の双曲線アークタンジェントを計算します。
cos(NUM)	実数	指定した角度のコサインを計算します。
cosh(NUM)	実数	指定した角度の双曲線コサインを計算します。
pi	実数	この定数は、パイに最も近い値の実数です。
sin(NUM)	実数	指定した角度のサインを計算します。
sinh(NUM)	実数	指定した角度の双曲線サインを計算します。
tan(NUM)	実数	指定した角度のタンジェントを計算します。
tanh(NUM)	実数	指定した角度の双曲線タンジェントを計算します。

## 確率関数

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

表 22. CLEM 確率関数

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

表 22. CLEM 確率関数 (続き)

関数	結果	説明
<code>cdf_t(NUM, DF)</code>	実数	指定した自由度の $t$ 分布からの値が特定の数字より小さくなる確率を返します。

## ビット単位の整数演算

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

表 23. CLEM のビット単位の整数演算：

関数	結果	説明
<code>~~ INT1</code>	<i>Integer</i>	$INT1$ のビット単位の補数を生成します。つまり、 $INT1$ で 0 である各ビットが 1 になります。 <code>~~ INT = -(INT + 1)</code> は常に真です。
<code>INT1    INT2</code>	<i>Integer</i>	この演算の結果は、 $INT1$ と $INT2$ のビット単位の「包括的論理和」または です。つまり、 $INT1$ と $INT2$ のどちらかまたは両方に 1 がある各ビットが 1 になります。
<code>INT1   /&amp; INT2</code>	<i>Integer</i>	この演算の結果は、 $INT1$ と $INT2$ のビット単位の「排他的論理和」または です。つまり、 $INT1$ と $INT2$ のどちらかにだけ (両方ではない) 1 がある各ビットが 1 になります。
<code>INT1 &amp;&amp; INT2</code>	<i>Integer</i>	$INT1$ と $INT2$ のビット単位の「論理積」を生成します。つまり、 $INT1$ と $INT2$ の両方で 1 である各ビットが 1 になります。
<code>INT1 &amp;&amp;~~ INT2</code>	<i>Integer</i>	この演算の結果は、 $INT1$ と、 $INT2$ のビット単位の補数との、ビット単位の「論理積」です。つまり、 $INT1$ のあるビット位置の値が 1 で、 $INT2$ の同じビット位置の値が 0 の場合に、結果は 1 になります。これは $INT1 \&\& (~~INT2)$ と同じであり、 $INT2$ 内で設定された $INT1$ のビットを消去する場合に役に立ちます。
<code>INT &lt;&lt; N</code>	<i>Integer</i>	$N$ の数だけ位置を左にシフトした $INT1$ のビット・パターンを生成します。 $N$ の値が負の場合は、右にシフトします。
<code>INT &gt;&gt; N</code>	<i>Integer</i>	$N$ の数だけ位置を右にシフトした $INT1$ のビット・パターンを生成します。 $N$ の値が負の場合は、左にシフトします。
<code>INT1 &amp;&amp;=_0 INT2</code>	<i>Boolean</i>	ブール式 $INT1 \&\& INT2 \neq 0$ と同じですが、より効率的です。
<code>INT1 &amp;&amp;/=_0 INT2</code>	<i>Boolean</i>	ブール式 $INT1 \&\& INT2 == 0$ と同じですが、より効率的です。
<code>integer_bitcount(INT)</code>	<i>Integer</i>	$INT$ の 2 の補数表現の、1 または 0 のビットの数をカウントします。 $INT$ が負ではない場合、 $N$ は 1 のビット数になります。 $INT$ が負の場合、 $N$ は 0 のビット数になります。符号の拡張のため、負ではない整数には無限大の数の 0 のビットがあります。また、負の整数には無限大の数の 1 のビットがあります。 <code>integer_bitcount(INT) = integer_bitcount(-(INT+1))</code> は常に真です。

表 23. CLEM のビット単位の整数演算 (続き):

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

## 乱数関数

次の関数は、無作為に項目を選択したり、乱数を生成する場合に用いられます。

表 24. CLEM 乱数関数:

関数	結果	説明
oneof(LIST)	Any	無作為 (ランダム) に選択された <i>LIST</i> の要素を返します。リスト項目は、[ITEM1,ITEM2,...,ITEM_N] のように入力する必要があります。フィールド名のリストを指定することもできます。
random(NUM)	Number	1~ <i>NUM</i> の範囲の、同じデータ型 ( <i>INT</i> または <i>REAL</i> ) の一様に分布した乱数を返します。整数を使用する場合、整数だけが返されます。実数 (10 進数) を使用する場合、実数値が返されます (精度はストリーム・オプションによって決まります)。この関数で返される可能性がある最大の乱数は、 <i>NUM</i> になります。
random0(NUM)	Number	random( <i>NUM</i> ) と同じ性質を持ちますが、値の範囲が 0 から始まります。この関数で返される最大の乱数が <i>NUM</i> になることはありません。

## 文字列関数

CLEM では、文字列に対して次の操作を行うことができます。

- 文字列の比較
- 文字列の生成
- 文字へのアクセス

CLEM では、文字列とは、一組の二重引用符で囲まれた一連の任意の文字のことです (例: "string quotes")。任意の単一の英数字が、文字 (CHAR) になります。これらは、`<文字>` の形式で、単一逆引用符を使用して CLEM 式内で宣言されます。例えば、`z`、`A`、または `2` です。範囲外の文字、または文字列に対する逆索引文字は、未定義の動作が生じます。

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

表 25. CLEM 文字列関数 :

関数	結果	説明
allbutfirst(N, STRING)	String	STRING の先頭 N 文字を削除した文字列を返します。
allbutlast(N, STRING)	String	STRING の最後の文字を削除した文字列を返します。
alphabefore(STRING1, STRING2)	Boolean	文字列のアルファベット順を確認するために用いられます。STRING1 が STRING2 より前にある場合に真を返します。
endstring(LENGTH, STRING)	String	指定した文字列から最後の N 文字を抽出します。文字列の長さが指定した長さよりも短いか、またはそれに等しい場合は、変更されません。
hasendstring(STRING, SUBSTRING)	Integer	この関数は、isendstring(SUBSTRING, STRING) と同じです。
hasmidstring(STRING, SUBSTRING)	Integer	この関数は、ismidstring(SUBSTRING, STRING) (埋め込みサブ文字列) と同じです。
hasstartstring(STRING, SUBSTRING)	Integer	この関数は、isstartstring(SUBSTRING, STRING) と同じです。
hassubstring(STRING, N, SUBSTRING)	Integer	この関数は、issubstring(SUBSTRING, N, STRING) と同じです。N のデフォルトは 1 です。
count_substring(STRING, SUBSTRING)	Integer	指定したサブ文字列が文字列内に発生する回数を返します。例を次に示します。 count_substring("foooo.txt", "oo") は 3 を返します。
hassubstring(STRING, SUBSTRING)	Integer	この関数は、issubstring(SUBSTRING, 1, STRING) と同じです。N のデフォルトは 1 です。
isalphacode(CHAR)	Boolean	CHAR が、文字コードが文字である指定された文字列 (通常フィールド名) 中の文字の場合に真を返します。それ以外の場合、この関数は 0 の値を返します。例えば、isalphacode(produce_num(1)) です。
isendstring(SUBSTRING, STRING)	Integer	文字列 STRING がサブ文字列 SUBSTRING で終わる場合、この関数は、STRING 内の SUBSTRING の整数の添字を返します。それ以外の場合、この関数は 0 の値を返します。
islowercode(CHAR)	Boolean	CHAR が指定された文字列 (通常フィールド名) の小文字の場合に、真 (true) の値を返します。それ以外の場合、この関数は 0 の値を返します。例えば、islowercode(``) と islowercode(country_name(2)) は両方とも有効な式です。
ismidstring(SUBSTRING, STRING)	Integer	SUBSTRING が STRING の部分文字列で、STRING の初めの文字から始まっていないか、または最後の文字で終わっていない場合、この関数は部分文字列が始まる位置の添字を返します。それ以外の場合、この関数は 0 の値を返します。

表 25. CLEM 文字列関数 (続き):

関数	結果	説明
isnumbercode(CHAR)	Boolean	指定された文字列 (通常フィールド名) の CHAR が、文字コードが数字である文字の場合に真を返します。それ以外の場合、この関数は 0 の値を返します。例えば、isnumbercode(product_id(2)) です。
isstartstring(SUBSTRING, STRING)	Integer	文字列 STRING がサブ文字列 SUBSTRING から始まる場合、この関数は添字 1 を返します。そうでない場合は、この関数は 0 の値を返します。
issubstring(SUBSTRING, N, STRING)	Integer	この関数は、文字列 STRING の N 番目の文字から始めて、文字列 SUBSTRING と等しいサブ文字列を検索します。文字列が見つかった場合、一致する部分文字列が始まる位置の添字 M (整数) を返します。それ以外の場合、この関数は 0 の値を返します。N が与えられていない場合、この関数はデフォルトで 1 になります。
issubstring(SUBSTRING, STRING)	Integer	この関数は、文字列 STRING の N 番目の文字から始めて、文字列 SUBSTRING と等しいサブ文字列を検索します。文字列が見つかった場合、一致する部分文字列が始まる位置の添字 M (整数) を返します。それ以外の場合、この関数は 0 の値を返します。N が与えられていない場合、この関数はデフォルトで 1 になります。
issubstring_count(SUBSTRING, N, STRING):	Integer	指定した STRING 内で N 番目に発生した SUBSTRING のインデックスを返します。N 番目に発生する SUBSTRING よりも少ない場合、0 を返します。
issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)	Integer	この関数は issubstring と同じですが、添字 STARTLIM から、またはその前から始まり、添字 ENDLIM で、またはその前で終わるように、マッチングが制限されます。STARTLIM 制約または ENDLIM 制約は、どちらかの引数に偽 (false) の値を指定することによって無効にできます。例えば、issubstring_lim(SUBSTRING, N, false, false, STRING) は issubstring と同じです。
isuppercode(CHAR)	Boolean	この関数は CHAR が大文字の場合に、真 (true) の値を返します。それ以外の場合、この関数は 0 の値を返します。例えば、isuppercode(``) と isuppercode(country_name(2)) は両方とも有効な式です。
last(CHAR)	String	この関数は、STRING の最後の文字 CHAR を返します (少なくとも 1 文字以上の長さがなければなりません)。
length(STRING)	Integer	文字列 STRING の長さ (つまり文字列内の半角文字数) を返します。

表 25. CLEM 文字列関数 (続き):

関数	結果	説明
locchar(CHAR, N, STRING)	Integer	シンボル値フィールド中の文字の位置を識別するために用いられます。この関数は、文字列 <i>STRING</i> 内の文字 <i>CHAR</i> の検索を、 <i>STRING</i> の <i>N</i> 番目の文字から開始します。この関数は、文字が見つかった ( <i>N</i> から始まる) 位置を示す値を返します。文字が見つからない場合は 0 を返します。関数のオフセット ( <i>N</i> ) が無効な場合 (例えばオフセットが文字列の長さを超えているなど)、この関数は \$null\$ を返します。 例えば、locchar(`n`, 2, web_page) は、フィールド <i>web_page</i> 内の `n` 文字の検索を、フィールド値の 2 番目の文字から開始します。 注: 指定する文字を、忘れずに、単一逆引用符で囲むようにしてください。
locchar_back(CHAR, N, STRING)	Integer	locchar に似ていますが、 <i>N</i> 番目の文字から前方向に検索される点が異なります。例えば、locchar_back(`n`, 9, web_page) と指定すると、フィールド <i>web_page</i> の 9 番目の文字から、文字列の先頭方向に向かって検索が開始されます。関数のオフセットが無効な場合 (例えばオフセットが文字列の長さを超えているなど)、この関数は \$null\$ を返します。できる限り、locchar_back を関数 length(<field>) とともに使用して、フィールドの現在の値の長さを動的に使用することをお勧めします。例えば、locchar_back(`n`, (length(web_page)), web_page) です。
lowertoupper(CHAR) lowertoupper (STRING)	CHAR または String	文字または文字列を入力にすることができ、同じデータ型の新しい項目を返すために用いられます。その際小文字はすべて同じ文字の大文字に変換されます。例えば、lowertoupper(`a`), lowertoupper("My string"), および lowertoupper(field_name(2)) はすべて有効な式です。
matches	Boolean	文字列が指定したパターンに一致する場合、真を返します。パターンは文字列リテラルにする必要があります。パターンを含むフィールド名にはなりません。クエスチョン・マーク (?) をパターンに含めて正確に 1 つの文字に一致させることができます。アスタリスク (*) は 0 かそれ以上の文字数に一致します。リテラル・クエスチョン・マークまたはアスタリスクを (むしろ、ワイルドカードとして使用しないで) 一致させるために、バックスラッシュをエスケープ文字として使用することができます。
replace(SUBSTRING, NEWSUBSTRING, STRING)	String	指定した <i>STRING</i> 内で、 <i>SUBSTRING</i> のすべてのインスタンスを <i>NEWSUBSTRING</i> を使用して置き換えます。

表 25. CLEM 文字列関数 (続き):

関数	結果	説明
replicate(COUNT, STRING)	String	指定した回数だけコピーされた元の文字列を含む文字列を返します。
stripchar(Char, STRING)	String	文字列またはフィールドから、指定した文字を削除します。この関数を利用すれば、データから通貨表記などの余分な記号を削除して、単純な数字または名前を取得できます。例えば、シntax スtripchar(`\$, 'Cost') を使用すると、すべての値からドル記号を削除した新しいフィールドが返されます。 注：指定する文字を、忘れずに、単一逆引用符で囲むようにしてください。
skipchar(Char, N, STRING)	Integer	文字列 <i>STRING</i> の <i>N</i> 文字目から、 <i>CHAR</i> 以外の文字を検索します。この関数は、見つかった文字の位置を示す整数サブ文字列を返します。 <i>N</i> 番目以降のすべての文字が <i>CHAR</i> の場合は、0 を返します。関数のオフセットが無効な場合 (例えばオフセットが文字列の長さを超えているなど)、この関数は \$null\$ を返します。 locchar は、よく関数 skipchar と一緒に、 <i>N</i> (文字列の検索開始点) の値を判断するために用いられます。例えば、skipchar(`s`, (locchar(`s`, 1, "MyString")), "MyString") です。
skipchar_back(Char, N, STRING)	Integer	skipchar に似ていますが、 <i>N</i> 番目の文字から前に戻る方向に検索される点が異なります。
startstring(LENGTH, STRING)	String	指定した文字列から最初の <i>N</i> 文字を抽出します。文字列の長さが指定した長さよりも短いか、またはそれに等しい場合は、変更されません。
strmember(Char, STRING)	Integer	locchar(Char, 1, STRING) と同じです。 <i>CHAR</i> が最初に出現する位置を示す整数サブ文字列、または 0 を返します。関数のオフセットが無効な場合 (オフセットが文字列の長さを超えているなど)、この関数は \$null\$ を返します。
subscrs(N, STRING)	CHAR	入力文字列 <i>STRING</i> の <i>N</i> 番目の文字 <i>CHAR</i> を返します。この関数は、STRING( <i>N</i> ) という短い形式で記述することもできます。例えば、lowertoupper("name"(1)) は有効な式です。
substring(N, LEN, STRING)	String	文字列 <i>SUBSTRING</i> を返します。この文字列は、文字列 <i>STRING</i> の添字 <i>N</i> から <i>LEN</i> 文字分の文字列で構成されています。
substring_between(N1, N2, STRING)	String	添字 <i>N1</i> から始まり、添字 <i>N2</i> で終わる <i>STRING</i> のサブ文字列を返します。
trim(STRING)	String	指定した文字列から、文字列の前後の空白文字を削除します。
trim_start(STRING)	String	指定した文字列から、文字列の前の空白文字を削除します。

表 25. CLEM 文字列関数 (続き):

関数	結果	説明
trimend(String)	String	指定した文字列から、文字列の後の空白文字を削除します。
unicode_char(NUM)	CHAR	NUM の Unicode 値を返します。
unicode_value(CHAR)	NUM	CHAR の Unicode 値を返します。
uppertolower(CHAR) uppertolower (STRING)	CHAR または String	文字または文字列を入力にすることができ、同じデータ型の新しい項目を返すために用いられます。その際、大文字はすべて同じ文字の小文字に変換されます。 注：文字列は二重引用符で、文字は単一の逆引用符で忘れずに指定するようにしてください。単純なフィールド名の場合は、引用符は使用しません。

## SoundEx 関数

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

表 26. CLEM soundex 関数

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

## 日付および時刻の関数

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

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

表 27. CLEM 日付と時刻の関数：

関数	結果	説明
@TODAY	String	「ストリームのプロパティ」ダイアログ・ボックスで「日/分をロールオーバー」を選択している場合、この関数は現在の日付形式を使用して、現在の日付を文字列として返します。2桁の日付形式を使用しており、「日/分をロールオーバー」を選択していない場合は、現在のサーバーの \$null\$ を返します。
to_time(ITEM)	Time	指定されたフィールドのストレージを時間に変換します。
to_date(ITEM)	Date	指定されたフィールドのストレージを日付に変換します。
to_timestamp(ITEM)	タイム・スタンプ	指定されたフィールドのストレージをタイムスタンプに変換します。
to_datetime(ITEM)	Datetime	指定されたフィールドのストレージを日付、時間またはタイムスタンプ値に変換します。
datetime_date(ITEM)	Date	数値、文字列またはタイムスタンプの日付値を返します数値（秒単位）を日付へ変換しなおすことができるのは、この関数だけです。ITEM が文字列の場合は、現在のデータ形式で文字列を解析することにより日付を作成します。この関数が正常に機能するためには、「ストリームのプロパティ」ダイアログ・ボックスの「日付の形式」に、正しい値が指定されていなければなりません。ITEM が数値の場合は、基準日（または紀元）からの秒数として解釈します。日付の端数は切り捨てられます。ITEM がタイムスタンプの場合は、日付をタイムスタンプの一部として返します。ITEM が日付の場合は、変更せずに返します。
date_before(DATE1, DATE2)	Boolean	DATE1 が DATE2 より前の日付またはタイムスタンプの場合に true を返します。それ以外の場合、この関数は 0 の値を返します。
date_days_difference(DATE1, DATE2)	Integer	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの日数を整数で返します。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。
date_in_days(DATE)	Integer	基準日から DATE で表される日付またはタイムスタンプまでの日数を整数で返します。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。例えば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_in_months(DATE)	実数	基準日から DATE で表される日付またはタイムスタンプまでの月数を実数で返します。これは、1 か月を 30.4375 日と仮定した近似値になります。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。例えば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。

表 27. CLEM 日付と時刻の関数 (続き) :

関数	結果	説明
date_in_weeks (DATE)	実数	基準日から DATE で表される日付またはタイムスタンプまでの週数を実数で返します。ここでは、1 週間を 7.0 日と仮定しています。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適正に行うには、有効な日付を指定する必要があります。例えば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_in_years (DATE)	実数	基準日から DATE で表される日付またはタイムスタンプまでの年数を実数で返します。これは、1 年 365.25 日を基準とした近似値になります。DATE が基準日より前の場合、この関数は負の数値を返します。計算を適切に行うには、有効な日付を指定する必要があります。例えば、日付として 2001 年 2 月 29 日を指定することはできません。2001 年は閏年ではないので、この日付は存在しません。
date_months_difference (DATE1, DATE2)	実数	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの月数を実数で返します。これは、1 か月を 30.4375 日と仮定した近似値になります。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。
datetime_date (YEAR, MONTH, DAY)	Date	YEAR、MONTH、および DAY の日付値を作成します。引数は整数でなければなりません。
datetime_day (DATE)	Integer	指定された DATE またはタイムスタンプから、日付を返します。結果は 1 から 31 の範囲の整数になります。
datetime_day_name (DAY)	String	指定された DAY のフルネームを返します。引数は、1 (日曜) から 7 (土曜) の範囲の整数でなければなりません。
datetime_hour (TIME)	Integer	TIME またはタイムスタンプから時間を返します。結果は 0 から 23 の範囲の整数になります。
datetime_in_seconds (TIME)	実数	TIME に保存された秒の部分の部分を返します。
datetime_in_seconds (DATE)、 datetime_in_seconds (DATETIME)	実数	現在の DATE または DATETIME と基準日の間の差 (1900-01-01) から集計した数値を秒に変換して返します。
datetime_minute (TIME)	Integer	TIME またはタイムスタンプから分を返します。結果は 0 から 59 の範囲の整数になります。
datetime_month (DATE)	Integer	DATE またはタイムスタンプから月を返します。結果は 1 から 12 の範囲の整数になります。
datetime_month_name (MONTH)	String	指定された DAY のフルネームを返します。引数は、1 から 12 の範囲の整数でなければなりません。
datetime_now	タイム・スタンプ	現在の時刻をタイムスタンプとして返します。
datetime_second (TIME)	Integer	TIME またはタイムスタンプから秒を返します。結果は 0 から 59 の範囲の整数になります。
datetime_day_short_name (DAY)	String	DAY の名前を省略形で返します。引数は、1 (日曜) から 7 (土曜) の範囲の整数でなければなりません。
datetime_month_short_name (MONTH)	String	MONTH の名前を省略形で返します。引数は、1 から 12 の範囲の整数でなければなりません。
datetime_time (HOUR, MINUTE, SECOND)	Time	指定された HOUR、MINUTE、および SECOND の時間値を返します。引数は整数でなければなりません。

表 27. CLEM 日付と時刻の関数 (続き):

関数	結果	説明
datetime_time(ITEM)	<i>Time</i>	ITEM の時間値を返します。
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	タイム・スタンプ	与えられた YEAR、MONTH、DAY、HOUR、MINUTE、および SECOND のタイムスタンプ値を返します。
datetime_timestamp(DATE, TIME)	タイム・スタンプ	DATE および TIME のタイムスタンプ値を返します。
datetime_timestamp (NUMBER)	タイム・スタンプ	与えられた秒数のタイムスタンプ値を返します。
datetime_weekday(DATE)	<i>Integer</i>	指定されたDATE またはタイムスタンプから、曜日を返します。
datetime_year(DATE)	<i>Integer</i>	DATE またはタイムスタンプから年を返します。結果は 2002 のような整数になります。
date_weeks_difference (DATE1, DATE2)	実数	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの週数を実数で返します。ここでは、1 週間を 7.0 日と仮定しています。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。
date_years_difference (DATE1, DATE2)	実数	日付またはタイムスタンプ DATE1 から日付またはタイムスタンプ DATE2 までの年数を実数で返します。これは、1 年 365.25 日を基準とした近似値になります。DATE2 が DATE1 よりも前の場合、この関数は負の数値を返します。
time_before(TIME1, TIME2)	<i>Boolean</i>	TIME1 が TIME2 より前の時間またはタイムスタンプの場合に真を返します。それ以外の場合、この関数は 0 の値を返します。
time_hours_difference (TIME1, TIME2)	実数	時間またはタイムスタンプ TIME1 と TIME2 間の時間差 (時間) を実数で返します。「ストリームのプロパティ」ダイアログ・ボックスで「日/分をロールオーバー」を選択している場合、TIME1 の値の方が大きいと、その値は前の日付を参照します。ロールオーバー・オプションをオンにしていなかった場合、TIME1 の値の方が大きいと、返される値は負になります。
time_in_hours(TIME)	実数	TIME で表される時間を実数で返します。例えば、時間の形式 HHMM では、式 time_in_hours('0130') は 1.5 として評価されます。TIME は時間またはタイムスタンプを示します。
time_in_mins(TIME)	実数	TIME で表される分を実数で返します。TIME は時間またはタイムスタンプを示します。
time_in_secs(TIME)	<i>Integer</i>	TIME で表される秒を整数で返します。TIME は時間またはタイムスタンプを示します。
time_mins_difference(TIME1, TIME2)	実数	時間またはタイムスタンプ TIME1 と TIME2 間の時間差 (分) を実数で返します。「ストリームのプロパティ」ダイアログ・ボックスで「日/分をロールオーバー」を選択している場合、TIME1 の値の方が大きいと、その値は前の日 (または、現在の形式で分と秒だけが指定されている場合は前の時間) を参照します。ロールオーバー・オプションをオンにしていなかった場合、TIME1 の値の方が大きいと、返される値は負になります。

表 27. CLEM 日付と時刻の関数 (続き) :

関数	結果	説明
time_secs_difference(TIME1, TIME2)	Integer	TIME1と TIME2 の時間またはタイムスタンプの差異を秒で返します。「ストリームのプロパティ」ダイアログ・ボックスで「日/分をロールオーバー」を選択している場合、TIME1 の値の方が大きいと、その値は前の日 (または、現在の形式で分と秒だけが指定されている場合は前の時間) を参照します。ロールオーバー・オプションをオンにしていない場合、TIME1 の値の方が大きいと、返される値は負になります。

## 日付と時刻の値の変換:

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

to\_date(DATE)

この返還を行うには、ストリームのデフォルト日付形式として、一致する日付形式 **MON YYYY** を選択します。

フィルター・ノードを使用して文字列値を日付へ変換する例については、streams サブフォルダー内の ¥Demos フォルダにインストールされている、ストリーム broadband\_create\_models.str を参照してください。

**数値として保存される日付:** 上記の例の DATE がフィールド名であるのに対して、to\_date は CLEM 関数であることに注意してください。数値として保存された日付がある場合は、数値が基準日 (または紀元) からの秒数として解釈される datetime\_date 関数を用いることでそれらを変換できます。

datetime\_date(DATE)

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

datetime\_date((date\_in\_days(DATE)-7)\*60\*60\*24)

## シーケンス関数

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

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

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

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

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

シーケンスおよび特殊関数は、次の特徴によってすぐに判別できます。

- 関数名の最初に @ が付いている。
- 関数名が大文字である。

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

例

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

```
@SINCE(Income > Outgoings)
```

この関数は、指定した条件が真 (true) であった最後のレコードのオフセットを返します。つまり、指定した条件が真 (true) であった最後のレコード以前のレコード数を返します。指定した条件が一度も真 (true) にならなかった場合、@SINCE は @INDEX + 1 を返します。

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

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

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

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

この場合、@SINCE は現在のレコードに対して条件を評価しません。前のレコードと現在のレコードに対して条件を評価する場合は、同様な関数 @SINCE0 を使用します。現在のレコードで条件が真 (true) の場合は、@SINCE0 は 0 を返します。

表 28. CLEM シーケンス関数：

関数	結果	説明
MEAN(FIELD)	実数	指定された <i>FIELD</i> または <i>FIELDS</i> に対して、値の平均値を返します。

表 28. CLEM シーケンス関数 (続き):

関数	結果	説明
@MEAN(FIELD, EXPR)	実数	現在のレコードを含めて、現在のノードが受け取った最後の <i>EXPR</i> レコードまでの、 <i>FIELD</i> の値の平均値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> には、0 より大きい整数として評価される任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が現在までに受け取ったレコード数を超過している場合は、現在までに受け取ったすべてのレコードの平均が返されます。
@MEAN(FIELD, EXPR, INT)	実数	現在のレコードを含めて、現在のノードが受け取った最後の <i>EXPR</i> レコードまでの、 <i>FIELD</i> の値の平均値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> には、0 より大きい整数として評価される任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が現在までに受け取ったレコード数を超過している場合は、現在までに受け取ったすべてのレコードの平均が返されます。 <i>INT</i> には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@DIFF1(FIELD)	実数	<i>FIELD1</i> の最初の差分を返します。したがって、1 つの引数を指定する形式では、単純にフィールドの現在値と前の値の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@DIFF1(FIELD1, FIELD2)	実数	2 つの引数を指定する形式では、 <i>FIELD2</i> に関する <i>FIELD1</i> の最初の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@DIFF2(FIELD)	実数	<i>FIELD1</i> の 2 番目の差分を返します。したがって、1 つの引数を指定する形式では、単純にフィールドの現在値と前の値の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@DIFF2(FIELD1, FIELD2)	実数	2 つの引数を指定する形式では、 <i>FIELD2</i> に関する <i>FIELD1</i> の最初の差分を返します。前に関連するレコードが存在しない場合は 0 を返します。
@INDEX	Integer	現在のレコードのインデックスを返します。インデックスは、レコードが現在のノードに到達したときにレコードに対して割り振られます。最初のレコードにはインデックス 1 が与えられます。インデックスは、その後の各レコードに対して 1 ずつ増やされます。
@LAST_NON_BLANK(FIELD)	Any	上流の入力ノードまたはデータ型ノードで定義されるように、空白でない <i>FIELD</i> の最後の値を返します。それまでに読み込んだレコードの <i>FIELD</i> の値がすべて空白である場合は、\$null\$ を返します。ユーザー欠損値とも呼ばれる空白値は、各フィールドに個別に定義することができることに注意してください。
@MAX(FIELD)	Number	指定された <i>FIELD</i> の最大値を返します。

表 28. CLEM シーケンス関数 (続き):

関数	結果	説明
@MAX(FIELD, EXPR)	Number	現在のレコードを含めて、現在までに受け取った過去 <i>EXPR</i> レコードの <i>FIELD</i> の最大値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> は、0 より大きい整数を評価する任意の式を使用できます。
@MAX(FIELD, EXPR, INT)	Number	現在のレコードを含めて、現在までに受け取った過去 <i>EXPR</i> レコードの <i>FIELD</i> の最大値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> は、0 より大きい整数を評価する任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの最大値が返されます。 <i>INT</i> には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@MIN(FIELD)	Number	指定された <i>FIELD</i> の最小値を返します。
@MIN(FIELD, EXPR)	Number	現在のレコードを含めて、現在までに受け取った過去 <i>EXPR</i> レコードの <i>FIELD</i> の最小値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> は、0 より大きい整数を評価する任意の式を使用できます。
@MIN(FIELD, EXPR, INT)	Number	現在のレコードを含めて、現在までに受け取った過去 <i>EXPR</i> レコードの <i>FIELD</i> の最小値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> は、0 より大きい整数を評価する任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの最小値が返されます。 <i>INT</i> には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。

表 28. CLEM シーケンス関数 (続き):

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

表 28. CLEM シーケンス関数 (続き):

関数	結果	説明
@SDEV(FIELD, EXPR, INT)	実数	現在のレコードを含めて、現在のノードが受け取った最後の <i>EXPR</i> レコードまでの、 <i>FIELD</i> の値の標準偏差を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> は、0 より大きい整数を評価する任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が、現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの標準偏差を返します。 <i>INT</i> には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@SINCE(EXPR)	Any	任意の CLEM 式が真 (true) の場合に、 <i>EXPR</i> から過ぎたレコード数を返します。
@SINCE(EXPR, INT)	Any	2 番目の引数 <i>INT</i> には、前方参照するレコードの最大数を指定します。 <i>EXPR</i> が一度も真 (true) にならなかった場合、 <i>INT</i> は @INDEX+1 です。
@SINCE0(EXPR)	Any	現在のレコードも考慮します。一方、@SINCE は現在のレコードは考慮しません。@SINCE0 は、現在のレコードについて <i>EXPR</i> が真 (true) の場合に 0 を返します。
@SINCE0(EXPR, INT)	Any	2 番目の引数 <i>INT</i> には、前方参照するレコードの最大数を指定します。
@SUM(FIELD)	Number	指定された <i>FIELD</i> または <i>FIELDS</i> に対して、値の合計値を返します。
@SUM(FIELD, EXPR)	Number	現在のレコードを含めて、現在のノードが受け取った最後の <i>EXPR</i> レコードまでの、 <i>FIELD</i> の値の合計値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> には、0 より大きい整数として評価される任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの合計が返されます。
@SUM(FIELD, EXPR, INT)	Number	現在のレコードを含めて、現在のノードが受け取った最後の <i>EXPR</i> レコードまでの、 <i>FIELD</i> の値の合計値を返します。 <i>FIELD</i> は数値型フィールドの名前でなければなりません。 <i>EXPR</i> には、0 より大きい整数として評価される任意の式を使用できます。 <i>EXPR</i> を省略した場合、または <i>EXPR</i> の値が現在までに受け取ったレコード数を超えている場合は、現在までに受け取ったすべてのレコードの合計が返されます。 <i>INT</i> には、参照する値の最大数を指定します。この方法は、2 つの引数だけを使用するよりも効率的です。
@THIS(FIELD)	Any	現在のレコードの <i>FIELD</i> で指定された名前のフィールドの値を返します。@SINCE 式でのみ使用されます。

## グローバル関数

関数 @MEAN、@SUM、@MIN、@MAX、および @SDEV は、最も広範囲の場合、現在までに読み取られたすべてのレコード (現在のレコードを含む) で機能します。しかし、現在のレコードの値とデータセット全体での値

とを比較できると便利な場合もあります。グローバル・ノードを使用してデータ・セット全体の値を生成したら、CLEM 式でグローバル関数を使用してこれらの値にアクセスすることができます。

例えば、

@GLOBAL\_MAX(Age)

は、データセット内で最も大きい Age の値を返します。一方、

$(\text{Value} - @GLOBAL\_MEAN(\text{Value})) / @GLOBAL\_SDEV(\text{Value})$

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

表 29. CLEM グローバル関数：

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

## 空白値とヌル値処理関数

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

表 30. CLEM の空白値関数とヌル値関数：

関数	結果	説明
@BLANK(FIELD)	Boolean	上流のデータ型ノードまたは入力ノードで設定された空白処理規則（「データ型」タブ）にしたがって、値が空白のレコードに対して真を返します。

表 30. CLEM の空白値関数とヌル値関数 (続き) :

関数	結果	説明
@LAST_NON_BLANK(FIELD)	Any	上流の入力ノードまたはデータ型ノードで定義されるように、空白でない FIELD の最後の値を返します。それまでに読み込んだレコードの FIELD の値がすべて空白である場合は、\$null\$ を返します。ユーザー欠損値とも呼ばれる空白値は、各フィールドに個別に定義することができることに注意してください。
@NULL(FIELD)	Boolean	FIELD の値がシステム欠損値 \$null\$ の場合に true を返します。ユーザー定義の空白値など、他のすべての値については false を返します。両方を確認する場合は、@BLANK(FIELD) および @NULL(FIELD) を使用します。
undef	Any	一般的に、CLEM で \$null\$ 値を入力するために用いられます。例えば、置換ノードで空白値にヌルを挿入するために用いられます。

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

## 特殊フィールド

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

表 31. CLEM 特殊フィールド :

関数	結果	説明
@FIELD	Any	式のコンテキスト中に指定されているすべてのフィールドに対して処理を行います。
@TARGET	Any	ユーザー定義の分析関数で CLEM 式を使用する場合、@TARGET は対象のフィールドを表すか、または分析される対象と予測のペアに対する「正しい値」を表します。通常この関数は、精度分析ノードで使用されます。
@PREDICTED	Any	ユーザー定義の分析関数で CLEM 式を使用する場合、@PREDICTED は、分析される対象と予測のペアに対して、予測される値を表します。通常この関数は、精度分析ノードで使用されます。
@PARTITION_FIELD	Any	現在のデータ区分フィールドの名前を置き換えます。

表 31. CLEM 特殊フィールド (続き):

関数	結果	説明
@TRAINING_PARTITION	Any	現在の学習用データ区分の値を返します。例えば、条件抽出ノードを使用して学習レコードを選択するには、CLEM 式 (@PARTITION_FIELD = @TRAINING_PARTITION) を使用します。この方法は、データ内の各データ区分を表すためにどの値が使用されているかに関係なく、常に条件抽出ノードが正しく動作することを保証します。
@TESTING_PARTITION	Any	現在のテスト用データ区分の値を返します。
@VALIDATION_PARTITION	Any	現在の検証用データ区分の値を返します。
@FIELDS_BETWEEN(start, end)	Any	データ中のフィールドの普通の順序 (つまり、挿入) にもとづく、指定された開始フィールドと最終フィールドの間 (開始、最終フィールドを含む) のフィールド名のリストを返します。
@FIELDS_MATCHING(pattern)	Any	指定したパターンに一致するフィールド名のリストを返します。クエスチョン・マーク (?) をパターンに含めて正確に 1 つの文字に一致させることができます。アスタリスク (*) は 0 かそれ以上の文字数に一致します。リテラル・クエスチョン・マークまたはアスタリスクを (むしろ、ワイルドカードとして使用しないで) 一致させるために、バックスラッシュをエスケープ文字として使用することができます。
@MULTI_RESPONSE_SET	Any	名前の付いた複数回答セットでフィールドのリストを返します。

---

## 第 9 章 リポジトリでの IBM SPSS Modeler の使用

---

### IBM SPSS Collaboration and Deployment Services Repository について

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

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

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

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

アダプターのインストールに関する詳細は、「*IBM SPSS Modeler 展開インストール・ガイド*」を参照してください。この資料は IBM SPSS Modeler 展開 DVD に収録されています。IBM SPSS Collaboration and Deployment Services Deployment Manager から IBM SPSS Modeler リポジトリ・オブジェクトにアクセスする方法については、「*IBM SPSS Modeler 展開ガイド*」を参照してください。

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

#### 拡張バージョン管理と検索のサポート

リポジトリには、包括的なオブジェクトのバージョン管理および検索機能があります。例えば、ストリームを作成して、そのストリームを リポジトリに格納して、他の部門の研究者と共有する場合を考えます。後で、IBM SPSS Modeler のストリームを更新する場合、以前のバージョンを上書きしないで、更新されたバージョンをリポジトリに追加できます。すべてのバージョンが、アクセス可能な状態で維持され、名前、ラベル、使用されているフィールドまたはその他の属性で検索できます。例えば、純収益を入力フィールドと使用しているすべてのモデルのバージョンや、特定の作成者により作成されたすべてのモデルを検索できます(この処理を従来のファイル・システムで行うと、各バージョンごとに異なるファイル名で保存する必要がありました。またバージョン間の関連がソフトウェアから認識できない場合があります)。

#### シングル・サインオン

シングル・サインオン機能を使用すると、接続するたびにユーザー名やパスワードを入力することなくリポジトリに接続することができます。ユーザーの既存のローカル・ネットワーク・ログインの詳細では、IBM SPSS Collaboration and Deployment Services への必要な認証が設定されています。この機能は、以下によって異なります。

- IBM SPSS Collaboration and Deployment Services を、シングル・サインオン・プロバイダを使用するよう設定する必要がある
- プロバイダと互換性のあるホストにログインする必要がある

詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。

---

## リポジトリ・オブジェクトの保存と展開

IBM SPSS Modeler で作成されるストリームは、拡張子 `.str` の付いたファイルとして、そのままリポジトリに保存できます。このように、全社規模の複数のユーザーが単一のストリームにアクセスできます。詳しくは、トピック 124 ページの『リポジトリのオブジェクトの保存』を参照してください。

リポジトリでストリームを展開することもできます。展開ストリームは、追加メタデータのあるファイルとして保存されます。展開ストリームは、自動化スコアリングやモデル更新など、IBM SPSS Collaboration and Deployment Services のエンタープライズ レベルの機能を十分に活用できます。例えば、新規データが利用可能になると、モデルを定期的なスケジュールで自動的に更新することができます。またあるいは一連のストリームを Champion-Challenger 分析を目的として展開することができ、ストリームを比較して最も効果的な世族モデルを含むシナリオを決定します。

ストリームは、ストリームとして (拡張子 `.str`)、またはシナリオとして (拡張子 `.scn`) 展開できます。ストリームとして展開すると、ストリームをシンクライアント・アプリケーション IBM SPSS Modeler Advantage によって使用できます。詳しくは、トピック 141 ページの『IBM SPSS Modeler Advantage でストリームを開く』を参照してください。シナリオとして展開すると、ストリームを Predictive Applications バージョン 5、IBM SPSS Modeler Advantage の先行アプリケーションによって使用することができます。

詳しくは、136 ページの『ストリーム展開のオプション』を参照してください。

シナリオとして展開されたストリームの要件

- 企業データに一貫してアクセスするには、IBM SPSS Collaboration and Deployment Services の Enterprise View コンポーネントを介してシナリオとして展開されたストリームにアクセスする必要があります。つまり、IBM SPSS Modeler には、ストリームの定義されたスコアリングまたはモデル作成ブランチ内に少なくとも 1 つの Enterprise View 入力ノードが存在する必要があります。
- Enterprise View ノードを使用するには、まず IBM SPSS Collaboration and Deployment Services を、すでに Enterprise View ビュー、アプリケーション・ビューおよびデータ・プロバイダ定義 (DPD) が定義されたサイトにインストールし、設定する必要があります。詳細は、各サイトの管理者に問い合わせるか、当社 Web サイト <http://www.ibm.com/software/analytics/spss/products/deployment/cds/> を参照してください。
- DPD は特定の ODBC データ・ソースに対して定義されます。IBM SPSS Modeler で DPD を使用する場合、ODBC を同じ名前を持つ IBM SPSS Modeler サーバー・ホスト上で定義し、DPD で参照されるデータ・ソースとして同じデータ・ストアに接続する必要があります。
- また、IBM SPSS Collaboration and Deployment Services Enterprise View Driver をストリームの変更または実行に使用する各コンピューターにインストールする必要があります。Windows の場合、SPSS Modeler または SPSS Modeler Server がインストールされているコンピューターにドライバーをインストールするだけで、ドライバー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。IBM SPSS Collaboration and Deployment Services Enterprise View Driver のインストールの詳細は、各サイトの管理者に問い合わせてください。

その他の展開のオプション

IBM SPSS Collaboration and Deployment Services は、企業コンテンツ管理に次のような広範の機能を提供し、ストリームを展開またはエクスポートするための多くのメカニズムを利用することができます。

- 後で IBM SPSS Modeler Solution Publisher ランタイムで利用するために、ストリームとモデルをエクスポートします。
- 1 つ以上のモデルを、モデル情報をエンコードするための XML 形式の PMML としてエクスポートします。詳しくは、トピック 142 ページの『PMML としてのモデルのインポートおよびエクスポート』を参照してください。

---

## リポジトリへの接続

1. リポジトリに接続するには、IBM SPSS Modeler メニューから次の項目をクリックします。

「ツール」 > 「リポジトリ」 > 「オプション...」

2. 「リポジトリ URL」フィールドで、アクセスしたいリポジトリ・インストール環境のディレクトリー・パス (つまり、URL) を入力または選択します。一度に接続できるレポジトリは 1 つだけです。

設定は、各サイトまたはインストール環境によって変わってきます。特定のログイン詳細については、各サイトのシステム管理者にお問い合わせください。

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

「資格情報の設定」。このボックスのチェック・マークを外した状態にして、**シングル・サインオン**機能を有効にします。これにより、ローカル・コンピューターのユーザー名およびパスワードの詳細を使用してログインを試みます。シングル・サインオンを使用できない場合、またはこのボックスをチェックしてシングル・サインオンを無効にした場合 (例えば、管理者アカウントにログインした場合)、資格情報を入力するための詳細画面が表示されます。

## リポジトリの資格情報の入力

設定によっては、「リポジトリ: 資格情報」ダイアログ・ボックス内で以下のフィールドを指定する必要があります。

「**ユーザー ID とパスワード**」。ログオンに使用する有効なユーザー名とパスワードを指定します。必要があれば、詳細についてローカル・システムの管理者にお問い合わせください。

「**プロバイダー**」。認証用のセキュリティー・プロバイダーを選択します。異なるセキュリティー・プロバイダーを使用するよう、リポジトリを設定することができます。必要な場合、詳しい情報を各サイトの管理者にお問い合わせください。

「**リポジトリおよびユーザー ID を記憶**」。毎回、接続時に同じ情報を再入力しなくても済むように、現在の設定をデフォルトとして保存します。

---

## リポジトリ・コンテンツを参照

リポジトリにより、Windows Explorer のように格納した内容を参照することができます。また格納された各オブジェクトのバージョンも参照することができます。

1. IBM SPSS Collaboration and Deployment Services Repository ウィンドウを開くには、SPSS Modeler のメニューで次の各項目をクリックします。

「ツール」 > 「リポジトリ」 > 「探索...」

1. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者にお問い合わせください。

エクスプローラ ウィンドウには、最初にフォルダー階層のツリー・ビューが表示されます。フォルダー名をクリックしてコンテンツを表示します。

現在の選択または検索基準に一致するオブジェクトが右側の領域にリストされ、選択されているバージョンの詳細な情報が右下の領域に表示されます。表示される属性値は、最も新しいバージョンに適用されます。

---

## リポジトリのオブジェクトの保存

リポジトリにストリーム、ノード、モデル、モデル・バレット、プロジェクト、出力オブジェクトを格納できます。他のユーザーやアプリケーションはそれらにアクセスすることができます。

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

ストリーム出力をリポジトリに、その他のユーザーが IBM SPSS Collaboration and Deployment Services Deployment Portal を使用してインターネットで表示できる形式で公開することもできます。

## オブジェクトの特性の設定

オブジェクトを保存する場合、「リポジトリ: 保存」ダイアログボックスが表示され、オブジェクトのさまざまなプロパティの値を設定できます。次の操作が可能です。

- オブジェクトを保存する名前およびリポジトリ・フォルダーを選択する
- バージョン・ラベルやその他の検索可能なプロパティ例など、オブジェクトに関する情報を追加する
- 1 つまたは複数の分類トピックをオブジェクトに割り当てる
- オブジェクトのセキュリティー・オプションを設定する

次の項では、設定できるプロパティについて説明します。

## オブジェクトを保存する場所の選択

「リポジトリ: 保存」ダイアログ・ボックスで、以下を入力します。

**ファイルの保存場所:** オブジェクトが保存される場所、現在のフォルダーが表示されます。リストでフォルダー名をダブルクリックすると、そのフォルダーが現在のフォルダーに設定されます。親フォルダーを選択するには、「上位フォルダー」ボタンを使用します。現在のレベルにフォルダーを作成するには、「新規フォルダー」ボタンを使用します。

**ファイル名。** オブジェクトを保存する名前。

**保存：** オブジェクトを現在の場所に保存します。

## 保存するオブジェクトに関する情報の追加

「リポジトリ: 保存」ダイアログ・ボックスの「情報」タブ上のフィールドは、いずれもオプションです。

「作成者」。リポジトリのオブジェクトを作成したユーザーのユーザー名。デフォルトでは、リポジトリ一接続に使用されたユーザー名が表示されますが、ここでこの名前を変更できます。

「バージョン・ラベル」。オブジェクトのバージョンを示すためにリストからラベルを選択するか、「追加」をクリックして、新しいラベルを作成します。ラベルには「[]」文字を使用しないでください。このオブジェクト・バージョンにラベルを割り当てない場合は、チェック・ボックスがオンになっていないことを確認してください。詳しくは、トピック 133 ページの『オブジェクトのプロパティの表示および編集』を参照してください。

**説明。** オブジェクトの説明です。説明によって、オブジェクトを検索できます (注意事項を参照)。

**キーワード:** オブジェクトに関連し、検索に使用できる 1 つまたは複数のキーワード (注記を参照)。

**有効期限 :** その日以降、オブジェクトが一般ユーザーに表示されなくなる日付。ただし、この日以降も所有者やリポジトリ管理者には表示されます。有効期限の日付を設定するには、「日付」オプションを選択して日付を入力するか、カレンダーのボタンを使用して日付を選択します。

**保存 :** オブジェクトを現在の場所に保存します。

注: 「説明」フィールドおよび「キーワード」フィールドの情報は、オブジェクトの「注釈」タブの SPSS Modeler に入力したものと異なるものとして扱われます。説明またはキーワードによるリポジトリ検索では、「注釈」タブからの情報は返されません。詳しくは、トピック 130 ページの『リポジトリのオブジェクトの検索』を参照してください。

## 保存したオブジェクトへのトピックの割り当て

トピックは、リポジトリに保存されているコンテンツの階層分類システムです。オブジェクトを保存するときに使用可能なトピックから選択でき、またユーザーはトピックによってオブジェクトを検索できます。使用できるトピックのリストは、適切な権限を持つリポジトリ・ユーザーによって設定されます (詳細は、『Deployment Manager ユーザー・ガイド』を参照してください)。

「リポジトリ: 保存」ダイアログ・ボックスの「トピック」タブで、トピックをオブジェクトに割り当てる手順は、次のとおりです。

1. 「追加」ボタンをクリックします。
2. 使用できるトピックのリストのトピック名をクリックします。
3. 「OK」をクリックします。

トピックの割り当てを削除する手順は、次のとおりです。

4. 割り当てたトピックのリストでトピックを選択します。
5. 「削除」をクリックします。

## 保存したオブジェクトのセキュリティー・オプションを設定

「リポジトリ: 保存」ダイアログ・ボックスの「セキュリティー」タブで、保存したオブジェクトのさまざまなセキュリティー・オプションを設定または変更できます。1 つまたは複数の主成分 (ユーザーまたはユーザーのグループ) について、次のことを実行できます。

- オブジェクトへのアクセス権限を割り当てる
- オブジェクトへのアクセス権限を変更する
- オブジェクトへのアクセス権限を削除する

**主成分 :** このオブジェクトにアクセス権限があるユーザーまたはグループのリポジトリ・ユーザー名。

**アクセス権。** このユーザーまたはグループがオブジェクトに対して持つアクセス権限。

**追加：** 1 つまたは複数のユーザーまたはグループをこのオブジェクトのアクセス権限を持つユーザーまたはグループのリストに追加できます。詳しくは、トピック『アクセス権リストへのユーザーの追加』を参照してください。

**変更：** このオブジェクトに対する選択したユーザーまたはグループのアクセス権限を変更することができます。デフォルトでは、読み取りアクセス権が付与されています。所有者、書き込み、読み取り、削除、変更という追加のアクセス権限を与えることができます。

**削除：** このオブジェクトに対するアクセス権リストから選択したユーザーまたはグループを削除します。

**アクセス権リストへのユーザーの追加：** 「リポジトリ: 保存」ダイアログ・ボックスの「セキュリティ」タブで、「追加」を選択すると、以下のフィールドを使用できます。

**プロバイダーの選択：** 認証用のセキュリティ・プロバイダーを選択します。異なるセキュリティ・プロバイダーを使用するよう、リポジトリを設定することができます。必要な場合、詳しい情報を各サイトの管理者に問い合わせてください。

**検索：** 追加するユーザーまたはグループのリポジトリ・ユーザー名を入力し、「検索」を検索するとユーザー・リストに名前が表示されます。一度に複数のユーザー名を追加するには、このフィールドを空白にし、「検索」をクリックすると、すべてのリポジトリ・ユーザー名のリストが表示されます。

**ユーザー・リスト：** リストから 1 つまたは複数のユーザー名を選択し、「OK」をクリックすると、そのユーザー名がアクセス権リストに追加されます。

**オブジェクトに対するアクセス権限の変更：** 「リポジトリ: 保存」ダイアログ・ボックスの「セキュリティ」タブで、「変更」を選択すると、以下のフィールドを使用できます。

**所有者：** このユーザーまたはグループに、オブジェクトに対する所有者のアクセス権限を与えます。所有者は、削除および変更のアクセス権限など、オブジェクトに対して完全なコントロール権を持ちます。

**読み取り：** デフォルトでは、オブジェクトの所有者ではないユーザーまたはグループには、オブジェクトに対する読み取りアクセス権限のみが与えられます。該当するチェック・ボックスをオンにすると、このユーザーまたはグループの書き込み、削除および権限の変更のアクセス権限が追加されます。

## ストリームの保存

リポジトリにストリームを `.str` ファイルとして保存できます。リポジトリから他のユーザーはそれらにアクセスすることができます。

注：リポジトリ機能をさらに利用するための、ストリーム展開の詳細については、135 ページの『ストリームの展開』を参照してください。

現在のストリームを保存するには

1. メイン・メニューで次の各項目をクリックします。

「ファイル」 > 「保存」 > 「ストリームとして保存...」

2. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。

- Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

## プロジェクトの格納

リポジトリに完全な IBM SPSS Modeler プロジェクトを *.cpj* ファイルとして保存できます。他のユーザーもプロジェクトにアクセスすることができます。

プロジェクト・ファイルがその他の IBM SPSS Modeler オブジェクトのコンテナであるため、リポジトリのプロジェクトのオブジェクトを格納することを IBM SPSS Modeler に知らせる必要があります。この操作では、「プロジェクト・プロパティ」ダイアログ・ボックスでの設定を使用します。詳しくは、トピック 148 ページの『プロジェクトのプロパティの設定』を参照してください。

プロジェクトをリポジトリに格納するよう設定すると、新規のオブジェクトをそのプロジェクトに追加するたびに、IBM SPSS Modeler はそのオブジェクトを格納するよう要求するメッセージを自動的に出力します。

IBM SPSS Modeler とのセッションを終了するときに、プロジェクト・ファイルの新しいバージョンを格納して、その追加を記憶する必要があります。プロジェクト・ファイルは、自動的にこのオブジェクトの最新バージョンを含めます（そして取得します）。IBM SPSS Modeler とのセッション中にオブジェクトを追加しなかった場合は、このプロジェクト・ファイルを復元してはいけません。ただし、変更したプロジェクトのオブジェクト（ストリーム、出力など）の新しいバージョンを格納する必要があります。

プロジェクトを格納する手順は、次のとおりです。

- IBM SPSS Modeler のマネージャ パレットの「CRISP-DM」タブまたは「クラス」タブでプロジェクトを選択し、メニューから次の項目をクリックします。

「ファイル」 > 「プロジェクト」 > 「プロジェクトを格納...」

- 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

## ノードの格納

リポジトリに現在のストリームの各ノード定義を *.nod* ファイルとして保存できます。リポジトリから他のユーザーはそのファイルにアクセスすることができます。

ノードを格納する手順は、次のとおりです。

- ストリーム キャンバスにあるノードを右クリックして、「ノードを格納」をクリックします。
- 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
- Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

## 出力オブジェクトの格納

リポジトリに現在のストリームの出力オブジェクトを *.nod* ファイルとして保存できます。リポジトリから他のユーザーはそのファイルにアクセスすることができます。

出力オブジェクトを格納する手順は、次のとおりです。

1. SPSS Modeler のマネージャー・パネルの「出力」タブでオブジェクトを選択し、メイン・メニューで次の項目をクリックします。

「ファイル」 > 「出力」 > 「出力を格納...」

2. または、「出力」タブにあるオブジェクトを右クリックして、「格納」をクリックします。
3. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
4. Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

## モデルおよびモデル・パレットの格納

リポジトリに各モデルを *.gm* ファイルとして保存できます。リポジトリから他のユーザーはそれらにアクセスすることができます。また、モデル・パレットのすべての内容を *.gen* ファイルとしてリポジトリに保存することもできます。

モデルの格納。

1. SPSS Modeler の「モデル」パレットのオブジェクトをクリックし、メイン・メニューで次の項目をクリックします。

「ファイル」 > 「モデル」 > 「モデルの格納...」

2. または、モデル・パレットにあるオブジェクトを右クリックして、「格納」をクリックします。
3. 「格納手順の完了」から先に進みます。

モデル・パレットの格納。

1. モデル・パレットの背景を右クリックします。
2. ポップアップ・メニューで「パレットを格納」をクリックします。
3. 「格納手順の完了」から先に進みます。

格納手順の完了。

1. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
2. Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

---

## リポジトリからのオブジェクトの取得

リポジトリに格納されたストリーム、モデル、モデル・パレット、ノード、プロジェクト、出力オブジェクトを取得できます。

注：ここで説明したようにメニュー・オプションを使用するほか、SPSS Modeler ウィンドウの右上にあるマネージャー・パネルの該当するタブを右クリックして、ストリーム、出力オブジェクト、モデルおよびモデル・パレットを取得することができます。

1. ストリームを取得するには、IBM SPSS Modeler のメインメニューの次の項目をクリックしてください。

「ファイル」 > 「ストリームを取得...」

2. モデル、モデル・パレット、プロジェクト、または出力オブジェクトを取得するには、IBM SPSS Modeler のメイン・メニューで次の項目をクリックします。

「ファイル」 > 「モデル」 > 「モデルを取得...」

または

「ファイル」 > 「モデル」 > 「モデル・パレットを取得...」

または

「ファイル」 > 「プロジェクト」 > 「プロジェクトを取得...」

または

「ファイル」 > 「出力」 > 「出力を取得...」

3. または、マネージャ またはプロジェクト・パネル内で右クリックし、ポップアップ・メニューから「取得」をクリックします。
4. ノードを取得するには、IBM SPSS Modeler のメインメニューの次の項目をクリックしてください。

「挿入」 > 「リポジトリからノード (またはスーパーノード)...」

- a. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
5. 「リポジトリ: 取得」ダイアログで、オブジェクトを参照し、オブジェクトを選択して「取得」ボタンをクリックします。詳しくは、トピックを参照してください。

## 取得するオブジェクトの選択

「リポジトリ: 読み込み/検索」ダイアログ・ボックスでは、以下のフィールドを使用できます。

**検索対象:** 現在のフォルダーのフォルダー階層が表示されます。異なるフォルダーに移動するには、このリストから選択して直接移動するか、このフィールドの下にあるオブジェクト・リストを使用して移動します。

「上位フォルダー」ボタン: 階層内の現在のフォルダーの 1 つ上のレベルに移動します。

「新規フォルダー」ボタン: 階層内の現在のフォルダーに新しいフォルダーを作成します。

**ファイル名。** 選択したオブジェクトのリポジトリ・ファイル名。そのオブジェクトを取得するには、「取得」をクリックします。

**ファイルの種類：** 取得するオブジェクトの種類。フォルダーと共に、この種類のオブジェクトのみが、オブジェクト・リストに表示されます。取得する異なる種類のオブジェクトを表示するには、リストからオブジェクトの種類を選択します。

「**ロック状態で開く**」。デフォルトでは、オブジェクトを取得すると、リポジトリ内でロックされ、他のユーザーは更新できません。取得時にオブジェクトをロックしたくない場合、このボックスをオフにします。

**説明、キーワード：** オブジェクトを保存したとき、オブジェクトに関する詳細が定義されている場合、これらの詳細情報がここに表示されます。詳しくは、トピック 124 ページの『保存するオブジェクトに関する情報の追加』を参照してください。

「**バージョン**」。最新版以外のオブジェクトのバージョンを取得するには、このボタンをクリックします。すべてのバージョンについての詳細な情報が表示され、必要なバージョンを選択できます。

## オブジェクトのバージョンの選択

「リポジトリ: バージョンの選択」ダイアログ・ボックスで特定のバージョンのリポジトリ・オブジェクトを選択する手順は次のとおりです。

1. (オプション) 該当する列の見出しをクリックして、バージョン、ラベル、サイズ、作成日、作成ユーザーによってリストをソートします。
2. 使用するオブジェクトのバージョンを選択します。
3. 「続行」をクリックします。

---

## リポジトリのオブジェクトの検索

オブジェクトを、名前、フォルダー、タイプ、ラベル、日付、またはその他の基準で検索できます。

名前で検索

名前でオブジェクトを検索する手順は、次のとおりです。

1. IBM SPSS Modeler のメイン・メニューで次の各項目をクリックします。

「ツール」 > 「リポジトリ」 > 「探索...」

- a. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
2. 「検索」タブをクリックします。
  3. 「検索するオブジェクトの名前」フィールドで、検索したいオブジェクトの名前を指定します。

名前でオブジェクトを検索する場合、ワイルドカード文字としてアスタリスク (\*) を使用すると、任意の文字列、疑問符 (?) を使用すると、1 文字に相当します。例えば、*\*cluster\** は、*cluster* という文字列を名前の一部に持っているすべてのオブジェクトに一致します。文字列 *m0?\_\** を検索すると、*M01\_cluster.str* と *M02\_cluster.str* は一致しますが、*M01a\_cluster.str* には一致しません。検索する場合は大文字と小文字の区別はされません (*cluster* は *Cluster* と *CLUSTER* にも一致します)。

注：オブジェクトが多数ある場合は、検索に時間がかかる場合があります。

## 他の基準による検索

タイトル、ラベル、日付、作成者、キーワード、インデックス付きコンテンツ、または説明に基づいて検索を実行できます。すべての指定された検索基準に一致するオブジェクトだけが検索されます。例えば、特定のラベルが付けられ、特定の日付以降に変更された、1 つ以上のクラスタリング・モデルを含むすべてのストリームを検索できます。

「**オブジェクト・タイプ**」。モデル、ストリーム、出力、ノード、スーパーノード、プロジェクト、モデル・パレット、シナリオまたは他の種類のオブジェクトに検索を制限することができます。

- **モデル**。カテゴリー (分類、近似、クラスタリング、その他)、または Kohonen などの特定のモデル作成アルゴリズムによりモデルを検索できます。

使用されているフィールド名で検索することもできます。例えば、入力フィールドまたは出力 (対象) フィールドとして、*income* という名前のフィールドを使用しているすべてのモデルを検索できます。

- **ストリーム** : ストリームについて、使用されるフィールド名またはストリームに含まれるモデル タイプ (カテゴリーまたはアルゴリズム) で検索を制限できます。

**トピック** : 適切な権限を持つリポジトリー・ユーザーによって設定されたリストから、特定のトピックに関連するモデルを検索することができます (詳細は「*Deployment Manager ユーザーズ・ガイド*」を参照)。リストを取得するには、このボックスをオンにし、表示される「トピックを追加」ボタンをクリックして、リストから 1 つまたは複数のトピックを選択し、「OK」をクリックします。

**ラベル**。特定のオブジェクト・バージョンのラベルに検索を制限します。

「**日付**」。作成日または修正日を指定して、以前、以降、または指定された日付範囲の間にあるオブジェクトを検索できます。

「**作成者**」。特定のユーザーが作成したオブジェクトに検索を制限します。

**キーワード** : 指定されたキーワードの検索。IBM SPSS Modeler では、キーワードは、ストリーム、モデル、または出力オブジェクトの「注釈」タブで指定します。

**説明**。説明フィールドにある特定の用語を検索します。IBM SPSS Modeler では、説明は、ストリーム、モデル、または出力オブジェクトの「注釈」タブで指定します。検索するフレーズが複数なら、セミコロンで区切ります (例: *income; crop type; claim value*)。 (検索するフレーズでは空白が重要です。例えば、スペースが 1 個の *crop type* とスペースが 2 個の *crop type* は同一ではありません。)

---

## リポジトリーのオブジェクトを変更

SPSS Modeler から直接、リポジトリーの既存のオブジェクトを変更できます。次の操作が可能です。

- フォルダーの作成、名前の変更または削除
- オブジェクトのロックまたはロック解除
- オブジェクトの削除

## フォルダーの編集、名前の変更、削除

1. リポジトリーでフォルダーの操作を実行するには、SPSS Modeler のメニューで次の項目をクリックします。

「ツール」 > 「リポジトリー」 > 「探索...」

- a. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者にお問い合わせください。
2. 「**フォルダー**」 タブがアクティブであることを確認してください。
3. 新しいフォルダーを作成するには、親フォルダーを右クリックして、「**新規フォルダー**」 をクリックします。
4. フォルダー名を変更するには、そのフォルダーを右クリックして「**フォルダー名の変更**」 をクリックします。
5. フォルダーを削除するには、そのフォルダーを右クリックして「**フォルダーの削除**」 をクリックします。

## リポジトリ・オブジェクトのロックおよびロック解除

オブジェクトをロックして、ほかのユーザーが既存のバージョンを更新したり新しいバージョンを作成しないようにすることができます。ロックされたオブジェクトは、オブジェクト・アイコン上に南京錠のシンボルで示されます。



図 15. ロックされたオブジェクト

オブジェクトをロックするには

1. リポジトリ・ウィンドウで目的のオブジェクトを右クリックします。
2. 「**ロック**」 をクリックします。

オブジェクトをロックを解除するには

1. リポジトリ・ウィンドウで目的のオブジェクトを右クリックします。
2. 「**ロック解除**」 をクリックします。

## リポジトリのオブジェクトを削除

オブジェクトをリポジトリから削除する前に、全部のバージョンを削除するか、またはあるオブジェクトの特定のバージョンを削除するのかを決める必要があります。

オブジェクトの全部のバージョンを削除するには

1. リポジトリ・ウィンドウで目的のオブジェクトを右クリックします。
2. 「**オブジェクトの削除**」 をクリックします。

オブジェクトの最新バージョンを削除するには

1. リポジトリ・ウィンドウで目的のオブジェクトを右クリックします。
2. 「**削除**」 をクリックします。

オブジェクトの前回のバージョンを削除するには

1. リポジトリ・ウィンドウで目的のオブジェクトを右クリックします。
2. 「**バージョンの削除**」 をクリックします。
3. 削除するバージョンを選択して「**OK**」 をクリックします。

---

## リポジトリ・オブジェクトのプロパティの管理

SPSS Modeler からさまざまなオブジェクトのプロパティを制御できます。次の操作が可能です。

- フォルダーのプロパティを表示する
- オブジェクトのプロパティを表示および編集する
- オブジェクトのバージョン・ラベルを作成、適用、および削除する

### フォルダーのプロパティの表示

リポジトリ・ウィンドウにあるフォルダーのプロパティを表示するには、目的のフォルダーを右クリックします。「**フォルダーのプロパティ**」をクリックします。

「全般」タブ

このタブには、フォルダー名、作成日時、および修正日時が表示されます。

「権限」タブ

このタブでは、フォルダーの読み取り権限と書き込み権限を指定します。親フォルダーへのアクセスを持つすべてのユーザーとグループがリストされます。権限は階層によって異なります。例えば、読み取り権限を持っていない場合、書き込み権限を持つことができません。書き込み権限を持っていない場合、削除権限を持つことはできません。

**ユーザーとグループ**：このフォルダーに対して少なくとも読み取りアクセス権を持っているリポジトリ・ユーザーおよびグループが表示されます。「書き込み」チェック・ボックスおよび「削除」チェック・ボックスをオンにすると、このフォルダーに対するこれらのアクセス権が、特定のユーザーまたはグループに割り当てられます。別のユーザーまたはグループにアクセスを割り当てるには、「権限」タブの右側にある「**ユーザーとグループの追加**」アイコンをクリックします。利用可能なユーザーとグループのリストは、管理者により管理されています。

**権限のカスケード**：子フォルダーがある場合、現在のフォルダーに対する変更が、子フォルダーにどのように適用されるかを制御できます。

- **すべての権限をカスケードする**：現在のフォルダーからすべての子、および子孫のフォルダーから権限設定をカスケードします。この方法は、複数のフォルダーの権限を一度に素速く設定できます。親フォルダーで必要な権限を設定し、それを必要に応じてカスケードします。
- **変更のみカスケードする**：前回の変更が適用されてから行われた変更のみをカスケードします。例えば、新しいグループが追加されたので、そのグループに「販売」ツリーにある全てのフォルダーへのアクセスを与えたい場合、グループ・アクセスをルートの「販売」フォルダーに与えて、その変更をすべてのサブフォルダーにカスケードできます。既存のサブフォルダーに対するすべてのそれ以外の権限は、以前のまま変更されません。
- **カスケードしない**：変更は、現在のフォルダーにのみ適用され、子フォルダーにカスケードされません。

### オブジェクトのプロパティの表示および編集

「オブジェクト・プロパティ」ダイアログ・ボックスで、プロパティを表示し、編集することができます。プロパティには変更できないものもありますが、いつでも新しいバージョンを追加することで更新できます。

1. リポジトリ・ウィンドウで目的のオブジェクトを右クリックします。
2. 「**オブジェクトのプロパティ**」をクリックします。

## 「全般」タブ

**名前。** リポジトリで表示したときのオブジェクトの名前です。

**作成日：** オブジェクト (バージョンでない) が作成された日付です。

**最終更新日：** 最新のバージョンが更新された日時です。

「**作成者**」。ユーザーのログイン名です。

**説明。** デフォルトでは、SPSS Modeler でオブジェクトの「注釈」タブで指定された説明が含まれています。

**リンクされたトピック：** リポジトリを使用すると、モデルおよび関連付けられたオブジェクトを、必要に応じてトピックごとに整理できます。使用できるトピックのリストは、適切な権限を持つリポジトリ・ユーザーによって設定されます (詳細は、「Deployment Manager ユーザーズ・ガイド」を参照)。

**キーワード：** ストリーム、モデル、または出力オブジェクトの「注釈」タブでキーワードを指定します。複数のキーワードは、スペースで区切る必要があります。また、最大で 255 文字 (半角) までです。(キーワードにスペースが含まれる場合は、キーワード間の区切りには疑問符を使用してください。)

## 「バージョン」タブ

リポジトリに格納されたオブジェクトには、複数のバージョンがあります。「バージョン」タブは、各バージョンについての情報を表示します。

格納されているオブジェクトの指定されたバージョンで、次のプロパティを指定または変更できます。

「**バージョン**」。バージョンの一意の識別子で、そのバージョンが格納された時間に基づいて生成されます。

**ラベル。** バージョンの現在のラベルです。バージョン識別子と異なり、ラベルはオブジェクトのあるバージョンから別のバージョンに移動できます。

各バージョンで、ファイル・サイズ、作成日時、および作成者も表示されます。

**ラベルの編集：** 「バージョン」タブの右上にある「ラベルの編集」アイコンを使用して、格納されたオブジェクトのラベルを定義、適用または削除します。詳しくは、トピック 135 ページの『オブジェクトのバージョン・ラベルの管理』を参照してください。

## 「権限」タブ

「権限」タブで、オブジェクトに対する読み取りと書き込みの権限を設定できます。現在のオブジェクトへのアクセスを持つすべてのユーザーとグループがリストされます。権限は階層によって異なります。例えば、読み取り権限を持っていない場合、書き込み権限を持つことができません。書き込み権限を持っていない場合、削除権限を持つことはできません。

**ユーザーとグループ：** このオブジェクトに対して少なくとも読み取りアクセス権を持っているリポジトリ・ユーザーおよびグループが表示されます。「書き込み」チェック・ボックスおよび「削除」チェック・ボックスをオンにすると、このオブジェクトに対するこれらのアクセス権が、特定のユーザーまたはグループに割り当てられます。別のユーザーまたはグループにアクセスを割り当てるには、「権限」タブの右側にある「ユーザーとグループの追加」アイコンをクリックします。利用可能なユーザーとグループのリストは、管理者により管理されています。

## オブジェクトのバージョン・ラベルの管理

「バージョン・ラベルの編集」ダイアログ・ボックスで以下のことができます。

- 選択したオブジェクトへのラベルの適用
- 選択したオブジェクトからのラベルの削除
- 新しいラベルの定義およびオブジェクトへの適用

ラベルをオブジェクトに適用するには

1. 「**使用できるラベル**」 リストで 1 つ以上のラベルを選択します。
2. 右矢印のボタンをクリックして、選択したラベルを 「**適用されたラベル**」 リストに移動します。
3. 「**OK**」 をクリックします。

オブジェクトからラベルを削除するには

1. 「**適用されたラベル**」 リストで 1 つ以上のラベルを選択します。
2. 左矢印のボタンをクリックして、選択したラベルを 「**使用できるラベル**」 リストに移動します。
3. 「**OK**」 をクリックします。

新しいラベルを定義しオブジェクトに適用するには

1. 「**新規ラベル**」 フィールドでラベル名を入力します。
2. 右矢印のボタンをクリックして、新しいラベルを 「**適用されたラベル**」 リストに移動します。
3. 「**OK**」 をクリックします。

---

## ストリームの展開

ストリームをシナプ리케이션の IBM SPSS Modeler Advantage で使用できるようにするには、リポジトリ内でストリーム (.str ファイル) として展開する必要があります。

ストリームをストリーム (.str ファイル) またはシナリオ (.scn ファイル) のどちらで展開しても、オブジェクトは IBM SPSS Collaboration and Deployment Services のエンタープライズレベルの機能を詳細に活用できます。詳しくは、トピック 122 ページの『リポジトリ・オブジェクトの保存と展開』を参照してください。

現在のストリームを展開するには (「ファイル」メニューを使用した方法)

1. メイン・メニューで次の各項目をクリックします。  
  
「ファイル」 > 「保存」 > 「展開」
2. 展開の種類を選択し、必要に応じてダイアログ・ボックスの残りを入力します。
3. 「**ストリームとして展開**」 をクリックすると、IBM SPSS Modeler Advantage または IBM SPSS Collaboration and Deployment Services と使用するようストリームを展開します。「**シナリオとして展開**」 をクリックすると、IBM SPSS Collaboration and Deployment Services または Predictive Applications バージョン 5 と使用するようストリームを展開します。
4. 「**保存**」 をクリックします。詳細は、「**ヘルプ**」 をクリックしてください。
5. 「**展開プロセスの完了**」 から先に進みます。

現在のストリームを展開するには (「ツール」メニューを使用した方法)

1. メイン・メニューで次の各項目をクリックします。

「ツール」 > 「ストリームのプロパティ」 > 「展開」

2. 展開の種類を選択し、必要に応じて「展開」タブの残りを入力し、「保存」をクリックします。詳しくは、トピック『ストリーム展開のオプション』を参照してください。

展開プロセスの完了

1. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。
2. Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

## ストリーム展開のオプション

「ストリーム・オプション」ダイアログ・ボックスの「展開」タブを使用すると、ストリーム展開のオプションを指定できます。ストリームまたはシナリオとして展開できます。

ストリームとして展開すると、シンクライアント・アプリケーション IBM SPSS Modeler Advantage でストリームを開いたり変更したりできます。ストリームは、拡張子 *.str* のファイルとしてリポジトリに保存されます。

シナリオとして展開すると、リポジトリのストリームは拡張子 *.scn* のファイルとして保存されます。ストリームは Predictive Applications バージョン 5 でも使用できるようになります。

シナリオとして展開しても、ストリームとして展開しても、マルチユーザー・アクセス、自動スコアリング、モデル・リフレッシュ、Champion-Challenger 分析など、IBM SPSS Collaboration and Deployment Services で使用できる追加機能を利用できます。

「展開」タブで、IBM SPSS Modeler がストリームに作成するストリームの説明をプレビューすることもできます。詳しくは、トピック 55 ページの『ストリームの説明』を参照してください。

注：エンタープライズ・データへの一貫したアクセスを実現するために、シナリオとして展開されたストリームは IBM SPSS Collaboration and Deployment Services Enterprise View を使用してソース・データにアクセスする必要があります。そのため、この場合は、指定された各スコアリングまたはモデル作成ブランチ内に Enterprise View ビュー入力ノードが少なくとも 1 つ必要です。

**展開タイプ**：ストリームの展開方法を選択します。すべてのストリームは、展開のために指定されたスコアリング・ノードが必要です。追加の要件およびオプションは展開の種類によって異なります。

- **<none>**。ストリームはリポジトリに展開されません。「ストリームの説明をプレビュー」を除くすべてのオプションが無効になります。
- **スコアリングのみ**：「保存」ボタンをクリックすると、ストリームがリポジトリに展開されます。「スコアリング・ノード」フィールドで指定したノードを使用して、データをスコアリングできます。
- **モデル・リフレッシュ**：「スコアリングのみ」と同じですが、「モデル作成ノード」フィールドおよび「モデル・ナゲット」フィールドで指定したオブジェクトを使用して、モデルをリポジトリで更新できます。注：IBM SPSS Collaboration and Deployment Services では、自動モデル・リフレッシュはデフォルトではサポートされないため、リポジトリからストリームを実行するときこの機能を使用する場合には、この展開タイプを選択する必要があります。詳しくは、トピック 139 ページの『モデル・リフレッシュ』を参照してください。

**スコアリング・ノード** : グラフ、出力、エクスポート、またはエクスポート・ノードを選択して、データのスコアリングに使用するストリームブランチを識別します。ストリームは実際いくつもの有効なブランチ、モデル、ターミナル・ノードを含んでいますが、唯一のスコアリングブランチが展開のために定義されている必要があります。ストリームを展開する最も基本的な要件です。

**スコアリング・パラメーター** : パラメーターを指定することができ、スコアリングブランチが実行される際にそのパラメーターを変更することができます。詳しくは、トピック 138 ページの『スコアリングおよびモデル作成パラメーター』を参照してください。

**モデル作成ノード** : モデル・リフレッシュについて、リポジトリのモデルを再生成または更新するために使用されるモデル作成ノードを指定します。「**モデル・ナゲット**」に指定されたモデル・ナゲットと同じ種類のモデル作成ノードである必要があります。

**モデル構築のパラメーター** : パラメーターを指定することができ、モデル作成ノードが実行される際にそのパラメーターを変更することができます。詳しくは、トピック 138 ページの『スコアリングおよびモデル作成パラメーター』を参照してください。

**モデル・ナゲット** : モデル・リフレッシュについて、リポジトリでストリームが更新されるたびに更新または再生成されるモデル・ナゲットを指定します (通常はスケジュール化されたジョブの一環として)。モデルはスコアリングブランチにある必要があります。複数のモデルがスコアリングブランチに存在する場合がありますが、定義されるのは 1 つだけです。ストリームが最初に作成される場合、新規データが利用可能になると更新または再生成されるプレースホルダー・モデルとなる場合があります。

**ストリームとして展開**: IBM SPSS Modeler Advantage または IBM SPSS Collaboration and Deployment Services でストリームを使用する場合に、このオプションをクリックします (以下の『注』を参照)。

**シナリオとして展開**: IBM SPSS Collaboration and Deployment Services または Predictive Applications バージョン 5 でストリームを使用する場合に、このオプションをクリックします (以下の『注』を参照)。

**検査** : 展開するために有効なストリームかどうかを確認します。展開する前にすべてのストリームが指定されたスコアリング・ノードが必要です。シナリオとして展開している場合、ストリームは、有効なエンタープライズ・ビュー ソース・ノードを含める必要があります。これらの条件が満たされない場合、エラーメッセージが表示されます。

**保存** : 有効な場合ストリームを展開します。有効でない場合、エラー・メッセージが表示されます。「**修正**」 ボタンをクリックしてエラーを修正し、もう一度実行してください。

**ストリームの説明をプレビュー**: IBM SPSS Modeler がストリームについて作成するストリームの説明の内容を表示できます。詳しくは、トピック 55 ページの『ストリームの説明』を参照してください。

**注** : (シナリオまたはシナリオとしての展開)複数の Enterprise View ノードをモデリングブランチ内で使用することができます。その場合、ブランチ内の全 エンタープライズ・ノードに対する単一のデータ接続の使用が多くの場合推奨され、Champion-Challenger 分析に必要となります。

- Champion-Challenger サポートを必要としない場合、データ・プロバイダ 定義 (DPD) によってのみ接続が異なる限り、様々な Enterprise View 接続を同じブランチ内で使用することができます。
- これらの制限は、指定されたブランチ内でのみ適用されます。スコアリングとモデル作成ブランチの間では、様々な Enterprise View 接続をそうした制限なしに使用することができます。

## スコアリングおよびモデル作成パラメーター

ストリームを IBM SPSS Collaboration and Deployment Services に展開する場合、モデルが更新またはスコアリングされるごとに、どのパラメーターを表示または編集するかを選択できます。例えば、最大値および最小値、またはジョブが実行されるごとに変更する必要があるその他の値を指定することができます。

1. ストリームの展開後に表示または編集できるようパラメーターを表示可能にするには、「スコアリング・パラメーター」ダイアログ・ボックスのリストから、そのパラメーターを選択します。

使用可能なパラメーターのリストは、「ストリームのプロパティ」ダイアログ・ボックスの「パラメーター」タブで定義されます。詳しくは、トピック 51 ページの『ストリームとセッション・パラメーターの設定』を参照してください。

## スコアリングブランチ

ストリームを展開している場合、ストリームのいずれかのブランチを**スコアリングブランチ** (スコアリング・ノードを含むブランチ) として指定する必要があります。ブランチをスコアリングブランチとして指定すると、そのブランチはストリーム領域で強調表示され、スコアリングブランチのナゲットへのモデル・リンクとなります。この視覚的表示は、スコアリングブランチがすぐに明らかになっていない、複数のブランチによる複雑なストリームで特に役立ちます。

注：スコアリングブランチとして指定できるストリームブランチは 1 つだけです。

ストリームでスコアリングブランチが既に定義されている場合、新しくブランチを指定すると、それがスコアリングブランチとして置き換えられます。「カスタム カラー」オプションを使用して、スコアリングブランチを示す色を設定できます。詳しくは、トピック 156 ページの『表示オプションの設定』を参照してください。

「ストリーム・マークアップを表示/非表示」ツールバー・ボタンを使用して、スコアリングブランチの表示を表示または非表示にできます。



図 16. 「ストリーム・マークアップを表示/非表示」ツールバー・ボタン

## 展開するスコアリングブランチの特定

ターミナル・ノードのポップアップ・メニューまたは「ツール」メニューを使用して、スコアリングブランチを指定できます。ポップアップ・メニューを使用すると、スコアリング・ノードがストリーム・プロパティの「展開」タブで自動的に設定されます。

ブランチをスコアリングブランチとして指定するには (ポップアップ・メニュー)

1. モデル・ナゲットをターミナル・ノード (ナゲットから下流の処理ノードまたは出力ノード) に接続します。
2. ターミナル・ノードを右クリックします。
3. メニューで、「**スコアリングブランチとして使用**」を選択します。

ブランチをスコアリングブランチとして指定するには (「ツール」メニュー)

1. モデル・ナゲットをターミナル・ノード (ナゲットから下流の処理ノードまたは出力ノード) に接続します。
2. メイン・メニューで次の各項目をクリックします。

「ツール」 > 「ストリームのプロパティ」 > 「展開」

3. 「展開タイプ」 リストで、必要に応じて「スコアリングのみ」または「モデル・リフレッシュ」をクリックします。詳しくは、トピック 136 ページの『ストリーム展開のオプション』を参照してください。
4. 「スコアリング・ノード」 フィールドをクリックして、リストからターミナル・ノードを選択します。
5. 「OK」をクリックします。

## モデル・リフレッシュ

モデル・リフレッシュは、新しいデータを使用するストリームの既存のモデルを再構築するプロセスです。リポジトリのストリーム自体は変わりません。例えば、アルゴリズムやストリーム固有の設定は同じままですが、モデルは新しいデータに保持され、新しいバージョンのモデルが古いモデルより優れている場合、更新されます。

リフレッシュに設定できるストリームのモデル・ナゲットは 1 つだけです。このモデルは、**リフレッシュモデル**と呼ばれます。ストリームのプロパティの「展開」タブで「**モデル・リフレッシュ**」オプションをクリックすると (136 ページの『ストリーム展開のオプション』参照)、そのとき指定したモデル・ナゲットがリフレッシュ モデルとなります。モデル・ナゲットのポップアップ・メニューからモデルをリフレッシュ モデルに指定することもできます。指定するには、ナゲットがあらかじめスコアリングブランチにある必要があります。

ナゲットの「リフレッシュ モデル」状態をオフにすると、ストリームの展開タイプを「スコアリングのみ」に設定することとなり、ストリームのプロパティのダイアログの「展開」タブがそれに従って更新されます。現在のスコアリングブランチにあるナゲットのポップアップ・メニューで「**リフレッシュ モデルとして使用**」オプションを使用して、この状態をオンにしたりオフにしたりすることができます。

スコアリングブランチのナゲットのモデル・リンクを削除すると、ナゲットの「リフレッシュ モデル」状態も削除されます。「編集」メニューまたはツールバーを使用して、モデル・リンクの削除を取り消すことができます。取り消した場合、ナゲットの「リフレッシュ モデル」の状態も元に戻ります。

## リフレッシュ モデルの選択方法

スコアリングブランチのほか、リフレッシュ モデルへのリンクもストリーム内で強調表示されます。リフレッシュ モデルに選択されるモデル・ナゲット、つまり強調表示されるリンクは、ストリーム内のナゲットの数によって異なります。

ストリームで 1 つのモデル

リンクしたモデルがスコアリングに 1 つあり、そのように特定されている場合、そのナゲットがストリームのリフレッシュ モデルとなります。

ストリームに複数のモデル

ストリームにリンクしたナゲットが複数ある場合、次のようにしてリフレッシュ モデルを選択します。

「ストリームのプロパティ」ダイアログ・ボックスの「展開」タブでモデル・ナゲットが定義されており、それがストリーム内にある場合、そのナゲットがリフレッシュ モデルになります。

「展開」タブでナゲットが定義されていない場合、またはナゲットが定義されているがスコアリングブランチにない場合、ターミナル・ノードに最も近いナゲットがリフレッシュ モデルとなります。

リフレッシュ リンクであるすべてのモデル・リンクの選択を解除すると、スコアリングブランチのみが強調表示され、リンクは強調表示されなくなります。展開タイプは「スコアリングのみ」に設定されます。

注：一方のリンクを置換え状態に設定できますが、もう一方は置き換えることができません。この場合、リフレッシュ モデルに選択されているモデル・ナゲットは、スコアリングブランチが指定されているときにリフレッシュ リンクがあり、ターミナル・ノードに最も近いモデル・ナゲットです。

ストリームにモデルがない

ストリームにモデルがない場合、またはモデル・リンクのないモデルだけがある場合、展開タイプは「スコアリングのみ」に設定されます。

## エラーのあるスコアリングブランチのチェック

スコアリングブランチを指定すると、シナリオとして展開しているときストリーム内に Enterprise View ノードがないなど、エラーの有無がチェックされます。

エラーが見つかり、スコアリングブランチがスコアリングブランチエラーの色で強調表示され、エラー・メッセージが表示されます。「カスタム カラー」オプションを使用して、エラーの色を設定できます。詳しくは、トピック 156 ページの『表示オプションの設定』を参照してください。

エラーが見つかった場合、次の手順を実行してください。

1. エラー・メッセージの内容に従って、エラーを修正します。
2. 次のいずれかを実行します。
  - ターミナル・ノードを右クリックし、ポップアップ・メニューで「シナリオを確認」をクリックします。
  - メイン・メニューで次の各項目をクリックします。

「ツール」 > 「ストリームのプロパティ」 > 「展開」

「チェック」をクリックします。

3. 必要に応じて、エラーがなくなるまでこのプロセスを繰り返します。

---

## 第 10 章 外部アプリケーションへのエクスポート

---

### 外部アプリケーションへのエクスポートについて

IBM SPSS Modeler では、データマイニング・プロセス全体を、外部のアプリケーションにエクスポートするためのいくつかの機構を備えており、データの準備作業を行ってから、IBM SPSS Modeler の外部で、ユーザー独自のアプリケーションを使用してモデルを構築できます。

前の項では、マルチユーザー・アクセス、ジョブ・スケジュール、その他の機能を利用するための IBM SPSS Collaboration and Deployment Services リポジトリへのストリームの展開方法について説明しました。同じように、IBM SPSS Modeler ストリームも次のアプリケーションと組み合わせて使用できます。

- IBM SPSS Modeler Advantage
- PMML 形式のファイルをインポートおよびエクスポートできるアプリケーション

IBM SPSS Modeler Advantage でのストリーム使用の詳細は、『IBM SPSS Modeler Advantage でストリームを開く』を参照してください。

PMML 形式をサポートするアプリケーションとモデルを共有するための、PMML ファイルとしてのモデルのエクスポートおよびインポートの詳細は、142 ページの『PMML としてのモデルのインポートおよびエクスポート』を参照してください。

---

### IBM SPSS Modeler Advantage でストリームを開く

IBM SPSS Modeler ストリームは、シンクライアント・アプリケーション IBM SPSS Modeler Advantage と組み合わせて使用できます。IBM SPSS Modeler Advantage 全体でカスタマイズされたアプリケーションを作成できますが、アプリケーション ワークフローの基本として、IBM SPSS Modeler で作成されたストリームを使用することもできます。

IBM SPSS Modeler Advantage でストリームを開く手順は次のとおりです。

1. 「ストリームとして展開」 オプションをクリックして、IBM SPSS Collaboration and Deployment Services リポジトリでストリームを展開します。詳しくは、トピック 135 ページの『ストリームの展開』を参照してください。
2. 「IBM SPSS Modeler Advantage で開く」 ツールバー・ボタンをクリックするか、メイン・メニューで次の項目をクリックします。

「ファイル」 > 「IBM SPSS Modeler Advantage で開く」

1. 必要に応じて、リポジトリへの接続の設定を行います。詳しくは、トピック 123 ページの『リポジトリへの接続』を参照してください。特定のポート、パスワード、およびその他詳細については、各サイトのシステム管理者に問い合わせてください。

注：リポジトリ・サーバーにも、IBM SPSS Modeler Advantage ソフトウェアをインストールする必要があります。

1. Repository の「保存」ダイアログ・ボックスで、オブジェクトを保存したいフォルダーを選択し、記録したいその他の情報を指定して「保存」ボタンをクリックします。詳しくは、トピック 124 ページの『オブジェクトの特性の設定』を参照してください。

インストールすると、ストリームがすでに開いている IBM SPSS Modeler Advantage を起動します。IBM SPSS Modeler のストリームは閉じます。

---

## PMML としてのモデルのインポートおよびエクスポート

PMML (Predictive Model Markup Language) は、モデルへの入力、データ・マイニングのデータの準備に使用する返還、モデル自体を定義するパラメーターなど、データ・マイニングおよび統計モデルを説明する XML 形式です。IBM SPSS Modeler は PMML をインポートおよびエクスポートし、IBM SPSS Statistics など、この形式をサポートする他のアプリケーションとモデルを共有できるようにします。

PMML の詳細は、データ・マイニング・グループの Web サイト (<http://www.dmg.org>) を参照してください。

モデルをエクスポートするには

PMML エクスポートでは、IBM SPSS Modeler 内で生成されたほとんどの種類のモデルがサポートされます。詳しくは、トピック 143 ページの『PMML をサポートするモデルの種類』を参照してください。

1. モデル・パレットのモデル・ナゲットを右クリックします(または、キャンバス上のモデル・ナゲットをダブルクリックして、「ファイル」メニューを選択します)。
2. メニューで、「**PMML のエクスポート**」をクリックします。
3. 「エクスポート」(または「保存」) ダイアログ・ボックスで、対象ディレクトリーとモデルの一意の名前を指定します。

注：「ユーザー・オプション」ダイアログ・ボックスで、PMML エクスポートのオプションを変更できます。メイン・メニューで次の各項目をクリックします。

「ツール」 > 「オプション」 > 「ユーザー・オプション」

そこで「PMML」タブをクリックします。

詳しくは、トピック 157 ページの『PMML エクスポート・オプション の設定』を参照してください。

PMML として保存されたモデルをインポートするには

IBM SPSS Modeler または別のアプリケーションから PMML としてエクスポートされたモデルは、生成済みモデル・パレットへインポートできます。詳しくは、トピック 143 ページの『PMML をサポートするモデルの種類』を参照してください。

1. モデル・パレット内で、パレットを右クリックし、メニューから「**PMML のインポート**」を選択します。
2. インポートするファイルを選択し、必要に応じて、変数のラベルに関するオプションを指定します。
3. 「**オープン**」をクリックします。

**モデル内に存在すれば変数ラベルを使用**：PMML が、データ・ディクショナリー内の変数に対して、変数名と変数ラベル (*RefID* に対する *Referrer ID* など) の両方を指定している場合があります。元のエクスポートされた PMML に変数ラベルが存在するときに変数ラベルを使用するには、このオプションを選択します。

変数ラベル・オプションを選択したにもかかわらず、PMML 内に変数ラベルがない場合、変数名は通常のように使用されます。

## PMML をサポートするモデルの種類

### PMML のエクスポート

**IBM SPSS Modeler モデル:** IBM SPSS Modeler で作成された次のモデルは、PMML 4.0 としてエクスポートされます。

- C&R Tree
- QUEST
- CHAID
- 線型回帰
- ニューラル・ネット
- C5.0
- ロジスティック回帰
- 一般化線型
- SVM
- Apriori
- Carma
- K-Means
- Kohonen
- TwoStep
- GLMM (固定効果専用 GLMM モデルの場合にのみサポートされます)
- ディジション・リスト
- Cox
- シーケンス (Sequence PMML モデルのスコアリングはサポートされません)
- Statistics モデル

**データベース固有モデル:** データベース固有のアルゴリズムを使用して作成されたモデルの場合、PMML エクスポートは、IBM InfoSphere Warehouse モデルのみ利用可能です。Microsoft または Oracle Data Miner の Analysis Services を使用して作成されたモデルをエクスポートすることはできません。また、PMML としてエクスポートされた IBM モデルを、IBM SPSS Modeler に再度インポートすることはできません。

### PMML のインポート

IBM SPSS Modeler では、すべての IBM SPSS Statistics 製品の現在のバージョンで作成された PMML モデルをインポートおよびスコアリングできます。IBM SPSS Statistics 17.0 で生成されたモデルまたは変換 PMML と同様に、IBM SPSS Modeler からエクスポートされたモデルもインポートおよびスコアリングできます。基本的には、次の例外を除いて、スコアリング エンジンはいずれの PMML をスコアリングできます。

- Apriori、CARMA、異常値検出、およびシーケンス・モデルをインポートするここはできません。
- スコアリングに使用できる場合でも、IBM SPSS Modeler へのインポート後に PMML をブラウズすることはできません。(これには、初めに IBM SPSS Modeler からエクスポートされたモデルは含まれません。この制限を回避するには、PMML ではなく生成されたモデル・ファイル [*\*.gm*] としてモデルをエクスポートします。)

- PMML としてエクスポートされた IBM InfoSphere Warehouse モデルを、インポートすることはできません。
- インポート時には制限つき検証が行われますが、モデルのスコアリング試行時には完全検証が実行されます。そのため、インポートは正常に行われますが、スコアリングは失敗したり不正な結果が生成されます。

---

## 第 11 章 プロジェクトとレポート

---

### プロジェクトの概要

プロジェクトは、データ・マイニング作業に関連したファイルのグループです。プロジェクトには、データ・ストリーム、グラフ、生成されたモデル、レポート、およびその他 IBM SPSS Modeler で作成したすべてが含まれます。IBM SPSS Modeler プロジェクトの第一印象は、単に出力を編成するための手段のように思われますが、それ以外にもさまざまな用途に利用することができます。プロジェクトを使用して、次のような処理ができます。

- プロジェクト・ファイル内の各オブジェクトに注釈を付ける。
- CRISP-DM 手法を使用してデータ・マイニング作業のガイド役をする。プロジェクトには、詳細な説明と CRISP-DM によるデータ・マイニング作業の実例を含んだ、CRISP-DM ヘルプ・システムも用意されています。
- データ・マイニングの目標をプレゼンテーションするための PowerPoint スライド・ショーや、使用するアルゴリズムに関する白書などの、IBM SPSS Modeler 以外のオブジェクトを追加する。
- 注釈に基づいて総合レポート、および簡単な更新レポートの両方を作成する。これらのレポートは、社内イントラネットなどに公開しやすい HTML 形式で生成することができます。

注：プロジェクト・パネルが IBM SPSS Modeler ウィンドウに表示されていない場合は、「表示」メニューで「プロジェクト」をクリックします。

プロジェクトに追加したオブジェクトを表示するには、**クラス・ビュー**と **CRISP-DM ビュー**の 2 つの方法を使用できます。プロジェクトに追加したものはすべて両方のビューに追加されます。2 つのビューを切り替えながら作業を行って、最適な編成を作成できます。

### CRISP-DM ビュー

CRISP-DM (Cross-Industry Standard Process for Data Mining) をサポートすることで、IBM SPSS Modeler プロジェクトには、業界で実証された標準のデータ・マイニング作業の編成手段が用意されます。

CRISP-DM では、作業の開始 (ビジネス要件の収集) から完了 (結果の展開) までのプロセスを大きく 6 つのフェーズに分けて記述します。通常、一部のフェーズは Clementine での作業とは関係ありませんが、IBM SPSS Modeler のプロジェクト・パネルでは、プロジェクトに関連するすべての資料やリソースを集中管理、保存、および追跡できるように、6 つのフェーズがすべて用意されています。例えば、一般的に業務の理解フェーズでは、要件の収集と各作業担当者との打ち合わせにより、データ・マイニングの最終目標を決定します。IBM SPSS Modeler でデータ処理などの作業を行うわけではありません。プロジェクト・パネルを利用すれば、このような打ち合わせ内容などの資料や記録を「業務の理解」フォルダーに保存して、将来の作業の参考にしたり、レポートに記載したりすることができます。

プロジェクト・パネルの CRISP-DM ビューには、データ・マイニング作業のライフ サイクルを説明する独自のヘルプ・システムも用意されています。IBM SPSS Modeler から、このヘルプには「ヘルプ」メニューの「**CRISP-DM ヘルプ**」をクリックしてアクセスします。

注：プロジェクト・パネルがウィンドウに表示されていない場合は、「表示」メニューで「プロジェクト」をクリックします。

## デフォルトのプロジェクト・フェーズの設定

プロジェクトに追加されたオブジェクトは、CRISP-DM のデフォルト・フェーズに追加されます。このため、オブジェクトを使用するデータ・マイニング・フェーズに応じて、手作業でオブジェクトを編成し直す必要があります。現在作業中のフェーズを、デフォルト・フォルダーに設定することが賢明です。

デフォルト・フォルダーとして使用するフェーズを選択する手順は、次のとおりです。

1. CRISP-DM ビューで、デフォルトとして設定するフェーズに対応するフォルダーを右クリックします。
2. メニューから、「デフォルト値に設定」 をクリックします。

デフォルト・フォルダーは太字で表示されます。

## クラス・ビュー

プロジェクト・パネルのクラス・ビューでは、IBM SPSS Modeler で行った作業が作成したオブジェクトの種類ごとに分類、編成されます。保存したオブジェクトは、次の任意のカテゴリに追加することができます。

- ストリーム
- ノード
- モデル
- テーブル、グラフ、およびレポート
- その他 (スライド・ショーやデータ・マイニング作業の関連資料などの非 IBM SPSS Modeler ファイル)

オブジェクトをクラス・ビューに追加すると、そのオブジェクトは CRISP-DM ビューのデフォルト・フェーズ・フォルダーにも追加されます。

注：プロジェクト・パネルがウィンドウに表示されていない場合は、「表示」メニューで「プロジェクト」をクリックします。

---

## プロジェクトの作成

基本的にプロジェクトは、プロジェクトに関連付けたすべてのファイルへの参照を含んだファイルです。つまり、プロジェクト内の各項目は個別に保存されるとともに、プロジェクト・ファイル (.cpj) 内に参照として保存されます。このような参照構造を採用しているため、次の点に注意してください。

- プロジェクトに追加する各項目は、プロジェクトに追加する前に、個別に保存しておく必要があります。追加する項目がまだ保存されていない場合は、プロジェクトに追加する前に保存するよう要求するメッセージが表示されます。
- ストリームのように個別に更新されるオブジェクトは、プロジェクト・ファイル内でも更新されます。
- ファイル・システムから手作業でオブジェクト (ストリーム、ノード、出力オブジェクトなど) を移動または削除すると、プロジェクト・ファイル内のリンクが無効になります。

## 新規プロジェクトの作成

新しいプロジェクトは、IBM SPSS Modeler ウィンドウから簡単に作成できます。まだプロジェクトを開いていない場合は、プロジェクトの作成を開始することができます。または、すでにプロジェクトを開いている場合、それを閉じて、新しいプロジェクトを作成することもできます。

メイン・メニューで次の各項目をクリックします。

「ファイル」 > 「プロジェクト」 > 「新規プロジェクト...」

## プロジェクトへの追加

プロジェクトを作成するか、または既存のプロジェクトを開いた後に、データ・ストリーム、ノード、およびレポートなどのオブジェクトを追加できます。オブジェクトを追加するために、さまざまな方法を利用できます。

### マネージャーからのオブジェクトの追加

IBM SPSS Modeler ウィンドウの右上にあるマネージャーを使用して、ストリームまたは出力を追加することができます。

1. 適切なマネージャ タブから、テーブルやストリームなどのオブジェクトを選択します。
2. 右クリックして、「プロジェクトに追加」を選択します。

オブジェクトがすでに保存されたことがある場合は、それが適切なオブジェクト・フォルダー (クラス・ビュー) またはデフォルトのフェーズ・フォルダー (CRISP-DM ビュー) に自動的に追加されます。

3. 代わりに、オブジェクトをマネージャーからプロジェクト・パネルにドラッグ・アンド・ドロップすることもできます。

注：先にオブジェクトの保存を要求するメッセージが表示されることもあります。オブジェクトを保存する際には、「保存」ダイアログ・ボックスで「ファイルをプロジェクトに追加」を忘れずに選択してください。このオプションを選択すると、オブジェクトを保存した後、そのオブジェクトが自動的にプロジェクトに追加されます。

### ストリーム領域からのノードの追加

「保存」ダイアログ・ボックスを使用して、ストリーム領域から個別のノードを追加できます。

1. キャンバスでノードを選択します。
2. ノードを右クリックして、「ノードの保存」をクリックします。または、メイン・メニューで次の各項目をクリックします。

「編集」 > 「ノード」 > 「ノードの保存...」

3. 「保存」ダイアログ・ボックスで、「ファイルをプロジェクトに追加」を選択します。
4. ノード名を指定して、「保存」をクリックします。

ファイルが保存され、プロジェクトにノードが追加されます。ノードは、クラス・ビューの「ノード」フォルダーと、CRISP-DM ビューのデフォルトのフェーズ・フォルダーに追加されます。

### 外部ファイルの追加

IBM SPSS Modeler 以外のさまざまなオブジェクトをプロジェクトに追加することができます。これは、IBM SPSS Modeler 内でデータ・マイニング・プロセス全体を管理する場合に役立ちます。例えば、データへのリンク、注釈、プレゼンテーション、およびグラフィックなどをプロジェクトに保存できます。CRISP-DM ビューでは、外部ファイルを保存するフォルダーを選択できます。クラス・ビューの場合、外部ファイルは「その他」フォルダーにしか保存できません。

プロジェクトに外部ファイルを追加する手順は、次のとおりです。

1. デスクトップからプロジェクトにファイルをドラッグ・アンド・ドロップします。

または

2. CRISP-DM ビューまたはクラス・ビュー内の対象フォルダーを右クリックします。
3. メニューで、「フォルダーに追加」 をクリックします。
4. ダイアログ・ボックス内でファイルを選択して、「開く」 をクリックします。

IBM SPSS Modeler プロジェクト内の選択したオブジェクトへの参照が追加されます。

## プロジェクト IBM SPSS Collaboration and Deployment Services Repository への転送

すべてのコンポーネント・ファイルも含めたプロジェクト全体を、IBM SPSS Collaboration and Deployment Services Repository へ一度に転送できます。すでに目的の場所にあるオブジェクトは、移動しません。これは逆にも機能します。プロジェクト全体を IBM SPSS Collaboration and Deployment Services Repository からローカル・ファイル・システムに移行できます。

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

### プロジェクトの転送

転送するプロジェクトがプロジェクト・パネル内で開かれていることを確認します。

プロジェクトを転送する手順は、次のとおりです。

1. ルート プロジェクト・フォルダーを右クリックして、「プロジェクトの転送」 をクリックします。
2. メッセージが表示されたら、IBM SPSS Collaboration and Deployment Services Repository へログインします。
3. プロジェクトの新しい場所を指定して、「OK」 をクリックします。

## プロジェクトのプロパティの設定

プロジェクトのプロパティのダイアログ・ボックスを使用して、プロジェクトの内容やドキュメントをカスタマイズできます。プロジェクトのプロパティにアクセスする手順は、次のとおりです。

1. プロジェクト・パネル内のオブジェクトまたはフォルダーを右クリックして、「プロジェクトのプロパティ」 をクリックします。
2. プロジェクトの基本情報を指定するには、「プロジェクト」 タブをクリックします。

作成日：プロジェクトの作成日を表示します (編集不可)。

「要約」。プロジェクト・レポートに表示される、データ・マイニング・プロジェクトの要約を入力できません。

内容：このプロジェクト・ファイルに参照されるコンポーネントの種類と数を一覧表示します (編集不可)。

未保存のオブジェクトを名前を付けて保存：未保存のオブジェクトをローカル・ファイル・システムに保存するか、またはリポジトリに格納するかを指定します。詳しくは、トピック 121 ページの『IBM SPSS Collaboration and Deployment Services Repository について』を参照してください。

プロジェクトのロード中にオブジェクト参照を更新：プロジェクトのコンポーネントへの参照を更新する場合に、このオプションを選択します。注：プロジェクトに追加したファイルは、そのプロジェクト・フ

ファイル内に保存されるわけではありません。プロジェクトには、各ファイルへの参照が保存されます。つまり、元のファイルを移動または削除すると、その参照もプロジェクトから削除されます。

## プロジェクトの注釈

プロジェクト・パネルには、データ・マイニング作業に関する注釈を記述するための、さまざまな方法が用意されています。プロジェクト・レベルの注釈は、しばしば「大局的な」目標や決定などの追跡に使用され、フォルダーやノード レベルの注釈は、詳細な説明を伝えるために使用されます。「注釈」タブには、取得できない欠損データの除外や、データ探索の過程で確立された有望な仮説などの、プロジェクト・レベルの詳細を記録するために十分な領域が用意されています。

プロジェクトに注釈を付ける手順は、次のとおりです。

1. CRISP-DM ビューまたはクラス・ビューで、プロジェクト・フォルダーを選択します。
2. フォルダーを右クリックして、「プロジェクトのプロパティ」 をクリックします。
3. 「注釈」タブをクリックします。
4. プロジェクトを記述するキーワードとテキストを入力します。

## フォルダーのプロパティと注釈

個別のプロジェクト・フォルダー (CRISP-DM ビューおよびクラス・ビューの両方の中) に注釈を付けることができます。CRISP-DM ビューでは、注釈機能がデータ・マイニングの各フェーズに対する組織目標を記録するための、最も効果的な手段になります。例えば、「ビジネスの理解」 フォルダーで注釈ツールを使用して、「今回の調査におけるビジネス目標は、高い価値を持つ顧客の顧客離れを減らすことにある。」のような情報を記入することができます。ここに記入したテキストは、「レポートに入れる」 オプションを選択すると、プロジェクト・レポートに自動的に記載されます。

フォルダーに注釈を付ける手順は、次のとおりです。

1. プロジェクト・パネル内でフォルダーを選択します。
2. フォルダーを右クリックして、「フォルダーのプロパティ」 をクリックします。

CRISP-DM ビューのフォルダーには、各フェーズの目的の概略や、関連するデータ・マイニング作業を行うためのガイドなどの注釈が付けられます。これらの注釈は、必要に応じて削除したり編集したりすることができます。

**名前。** この領域には、選択したフィールドの名前が表示されます。

**ツールヒント テキスト** : マウス・ポインタをプロジェクト・フォルダー上に移動した時に表示される、カスタムのツールヒントを作成します。この機能は、CRISP-DM ビュー内で各フェーズの目標を記載したり、「作業中」や「完了」のようにフェーズのステータスを表示したりする場合に役立ちます。

**「注釈」フィールド** : このフィールドは、プロジェクト・レポートに記載できるより長い注釈を記載するために利用されます。CRISP-DM ビューの注釈には、データ・マイニングの各フェーズの注釈が記載されていますが、プロジェクトの内容に応じて、自由に変更することができます。

**レポートに入れる** : この注釈をレポートに記載するには、「レポートに入れる」 を選択します。

## オブジェクトのプロパティ

オブジェクトのプロパティを表示したり、プロジェクト・レポートに個別のオブジェクトを記載するかどうかを選択できます。オブジェクトのプロパティにアクセスする手順は、次のとおりです。

1. プロジェクト・パネルで、オブジェクトを右クリックします。

2. メニューで「**オブジェクトのプロパティ**」をクリックします。

**名前。**この領域には、保存したオブジェクトの名前が一覧表示されます。

**パス：**この領域には、オブジェクトの保存場所が一覧表示されます。

**レポートに入れる：**生成するレポートにオブジェクトの詳細を記載する場合に選択します。

## プロジェクトを閉じる

IBM SPSS Modeler を終了したり、新しいプロジェクトを開いたりすると、既存のプロジェクト・ファイル (.cpj) が閉じます。

製品と関連するいくつかのファイル (ストリーム、ノード、またはグラフなど) は開いたままとなる場合があります。これらのファイルを開いたままにする場合、「... これらのファイルを保存して閉じますか?」というメッセージに「いいえ」と答えてください。

プロジェクトを閉じた後に関連するファイルを変更して保存した場合、それらの変更内容は、次にプロジェクトを開いた時に反映されます。変更内容を反映しないようにするには、プロジェクトからそのファイルを削除するか、それを別のファイル名で保存してください。

---

## レポートの生成

プロジェクトのもっとも役立つ機能の 1 つとして、プロジェクト内の各項目や注釈に基づいてレポートを生成できる機能があります。この機能は、「CRISP-DM 手法」で説明しているように、効果的なデータ・マイニングを行うために必要不可欠です。レポートは、いくつかのファイル タイプの 1 つに直接生成することも、ただちに確認できるように画面上のウィンドウに出力することもできます。この出力から、印刷や保存、または Web ブラウザー内でのレポートの表示を行うことができます。保存されたレポートは、組織内の他の部門に配布できます。

レポートは、データ・マイニング・プロセス中にプロジェクト・ファイルから何度も生成され、プロジェクトの関係者に配布されます。レポートには、プロジェクト・ファイルで参照されるオブジェクト情報や、作成した注釈などから、必要な情報や重要な情報を記載します。レポートは、クラス・ビューまたは CRISP-DM ビューのどちらかに基づいて作成できます。

レポートを生成する手順は次のとおりです。

1. CRISP-DM ビューまたはクラス・ビューで、プロジェクト・フォルダーを選択します。
2. フォルダーを右クリックして、「**プロジェクトのレポート**」をクリックします。
3. レポート・オプションを指定して、「**レポートの生成**」 ボタンをクリックします。

「レポート」ダイアログ・ボックスのオプションには、目的に応じたさまざまなレポートを生成するいくつかの方法が用意されています。

**出力名：**レポートの出力を画面に送ることを選択した場合は、出力ウィンドウの名前を指定します。独自の名前を指定することも、IBM SPSS Modeler に自動的にウィンドウの命名をさせることもできます。

**画面に出力：**レポートを生成し、出力ウィンドウに表示する場合に選択します。出力ウィンドウからさまざまな種類のファイルへレポートをエクスポートするオプションもあります。

**ファイルに出力：**レポートを生成して、「ファイルの種類」リストで指定された種類のファイルとして保存する場合に選択します。

**ファイル名** :生成されたレポートのファイル名を指定します。ファイルは、デフォルトで IBM SPSS Modeler #bin ディレクトリーへ保存されます。別の場所を指定するには、反復記号の「...」 ボタンを使用します。

**ファイルの種類** :指定できるファイルの種類を次に示します。

- **HTML ドキュメント** : レポートは、単一の HTML ファイルとして保存されます。レポートにグラフが含まれている場合は PNG ファイルとして保存され、HTML ファイルに参照されます。レポートをインターネットで公開する場合は、HTML ファイルと、そのファイルが参照する画像の両方を必ずアップロードしてください。
- **テキスト・ドキュメント** : レポートは、単一のテキスト・ファイルとして保存されます。レポートにグラフが含まれている場合は、ファイル名とパスの参照のみがレポートに含まれます。
- **Microsoft Word ドキュメント** : レポートは、単一のドキュメントとして保存され、グラフはドキュメント内に直接埋め込まれます。
- **Microsoft Excel ドキュメント** : レポートは、単一のスプレッドシートとして保存され、グラフはスプレッドシート内に直接埋め込まれます。
- **Microsoft PowerPoint ドキュメント** : 各フェーズが新しいスライドに表示されます。どのようなグラフであっても、PowerPoint スライドに直接埋め込まれます。
- **出力オブジェクト**: IBM SPSS Modeler 内で開くと、このファイル (.cou ) は 「レポート形式」 の 「画面に出力」 と同じになります。

注 : Microsoft Office ファイルをエクスポートするには、対応するアプリケーションがインストールされている必要があります。

**表題**。 レポートのタイトルを指定します。

**レポート構造**: 「CRISP-DM」 または 「クラス」 のいずれかを選択します。CRISP-DM ビューは、「大局的な」 観点からのステータス・レポートと、データ・マイニングの各フェーズの詳細情報を提供します。クラス・ビューはオブジェクトを基準にしたビューで、データやストリームの内部的な追跡を行う場合に適しています。

「作成者」。 デフォルトのユーザー名が表示されますが、変更できます。

**レポートに表示する項目** : オブジェクトをレポートに含める方法を選択します。プロジェクト・ファイルに追加されたすべての項目を記載するには、「すべてのフォルダーとオブジェクト」 を選択します。オブジェクトのプロパティーで 「レポートに入れる」 が選択されているかどうかを基準にして、項目を記載することもできます。代わりに、レポートに表示しない項目を確認するために、除外することを選択した項目だけを記載することもできます ( 「レポートに入れる」 の選択を解除する)。

**選択** : レポート内で 「最近使用した項目」 のみを選択することで、プロジェクトの更新内容を提供できるようにします。代わりに 「古い項目」 のパラメーターを設定して、古い項目や未解決の問題を追跡することもできます。レポートで時間的なことを考慮しない場合は、「すべての項目」 を選択します。

**表示順** : フォルダー内のオブジェクトの順序を決めるために、次のオブジェクト特性の組み合わせを選択できます。

- **データ型**: データ型を基準にオブジェクトをグループ化します。
- **名前**. オブジェクトをアルファベット順に並べます。
- **追加日** : プロジェクトに追加された日付を使用して、オブジェクトをソートします。

## 生成されたレポートの保存とエクスポート

画面に生成されたレポートは、新規出力ウィンドウに表示されます。レポートに含まれるグラフは、インラインの画像として表示されます。

### レポート用語集

各ストリームのノードの合計数がレポート内に表示されます。次の見出しに数が表示され、CRISP-DM 用語集ではなく IBM SPSS Modeler 用語集が使用されます。

- **データ・リーダー**：入力ノード。
- **データ・ライター**：エクスポート・ノード。
- **モデル・ビルダー**：構築ノードまたはモデル作成ノード。
- **モデル・アプライヤー**：ナゲットとも呼ばれる、生成されたモデル。
- **出力ビルダー**：グラフ・ノードまたは出力ノード。
- **その他**：プロジェクトに関連するその他のノード。例えば、ノード・パレットの「フィールド設定」タブまたは「レコード設定」タブで使用できるノード。

レポートを保存する手順は次のとおりです。

1. 「ファイル」メニューで「保存」をクリックします。
2. ファイル名を指定します。

レポートは、出力オブジェクトとして保存されます。

レポートをエクスポートする手順は次のとおりです。

3. 「ファイル」メニューで「エクスポート」をクリックし、エクスポートするファイルの種類を選択します。
4. ファイル名を指定します。

レポートは、選択した形式で保存されます。

次の種類のファイルへエクスポートできます。

- HTML ドキュメント(H)
- テキスト
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

注：Microsoft Office ファイルをエクスポートするには、対応するアプリケーションがインストールされている必要があります。

ウィンドウの上部にあるボタンを使用して、次の作業を実行します。

- レポートを印刷します。
- 外部の Web ブラウザー内で HTML としてレポートを表示します。

---

## 第 12 章 IBM SPSS Modeler のカスタマイズ

---

### IBM SPSS Modeler オプションのカスタマイズ

必要に応じて IBM SPSS Modeler をカスタマイズするためには、さまざまな操作があります。このカスタマイズは、主にメモリーの割り当て、デフォルトのディレクトリー、音声および色の使用など、特定のユーザー・オプションの設定で構成されています。また、IBM SPSS Modeler ウィンドウの下部にあるノード・パレットをカスタマイズすることもできます。

---

### IBM SPSS Modeler オプションの設定

IBM SPSS Modeler のオプションをカスタマイズしたり、設定するには、さまざまな方法があります。

- メモリー使用量やロケールなどのシステム・オプションを設定するには、「ツール > オプション」メニューの「システム・オプション」をクリックします。
- 表示フォントや色などのユーザー・オプションを設定するには、「ツール > オプション」メニューの「ユーザー・オプション」をクリックします。
- IBM SPSS Modeler と連携するアプリケーションの場所を指定するには、「ツール > オプション」メニューの「ヘルパー・アプリケーション」をクリックします。
- IBM SPSS Modeler が使用するデフォルト・ディレクトリーを指定するには、「ファイル」メニューの「ディレクトリーの設定」または「サーバー・ディレクトリーの設定」をクリックします。

また、ストリームの一部またはすべてに適用するオプションを設定することができます。詳しくは、トピック 44 ページの『ストリームのオプションの設定』を参照してください。

### システム・オプション

IBM SPSS Modeler の使用言語またはロケールを指定するには、「ツール > オプション」メニューの「システム・オプション」をクリックします。ここでは、IBM SPSS Modeler の最大メモリー使用量を設定することもできます。このダイアログ・ボックスで行った変更内容は、IBM SPSS Modeler を再起動しないと有効にはならないことに注意してください。

**最大メモリー:** IBM SPSS Modeler のメモリー使用量を MB 単位で制限する場合に選択します。一部のプラットフォームでは、リソースに制約のあるコンピューターや負荷の高いコンピューターの負担を軽減するために、IBM SPSS Modeler はプロセスのサイズを制限しています。そのため、大量のデータを扱うと、「メモリー不足」のエラーが発生することがあります。ここに新しい値を指定して、メモリーの負荷を調整することができます。

**システム ロケールを使用 :** このオプションはデフォルトで選択され、「日本語」に設定されています。他の言語を指定する場合は、このオプションの選択を解除して、リストから適切な言語やロケールを選択してください。

### メモリーの管理

「システム・オプション」ダイアログ・ボックスの「最大メモリー」の設定以外にも、メモリーの使用を最適化するためのさまざまな方法があります。

- 「ストリームのプロパティ」ダイアログ・ボックスの「名義型ノードの最大メンバー数」オプションを調整する。このオプションには、名義型フィールドの最大メンバー数を指定します。これ以降のフィールドの測定は、データ型不明になります。詳しくは、トピック 44 ページの『ストリームの一般的なオプションの設定』を参照してください。
- IBM SPSS Modeler のウィンドウの右下に表示されている、IBM SPSS Modeler が使用中のメモリおよびメモリの総割り当て量 (xxMB/xxMB) の領域をクリックして、メモリを解放する。この領域をクリックすると、いったん領域が暗くなって押されたような状態になります。その後領域が元の色に戻ると、IBM SPSS Modeler で解放できるメモリがすべて解放されます。

## デフォルト・ディレクトリーの設定

ファイルの参照や出力に使われるデフォルト・ディレクトリーを指定するには、「ファイル」メニューの「ディレクトリーの設定」または「サーバー・ディレクトリーの設定」を選択します。

- **ディレクトリーの設定** :作業ディレクトリーを設定します。デフォルトの作業ディレクトリーは、IBM SPSS Modeler のインストール場所、または IBM SPSS Modeler を起動するために用いられたコマンド・ラインのパスをベースにしています。ローカル・モードの場合、作業ディレクトリーはクライアント側の操作や出力ファイルで使われたパスになります (相対パスで参照された場合)。
- **サーバー・ディレクトリーの設定** : リモート・サーバーへの接続がある場合、「ファイル」メニューの「サーバー・ディレクトリーの設定」オプションは常に有効になっています。このオプションを使用して、入力や出力に使われるすべてのサーバー・ファイルやデータ・ファイルのデフォルト・ディレクトリーを指定します。デフォルトのサーバー・ディレクトリーは、`$CLEO/data` になります。ここで、`$CLEO` は IBM SPSS Modeler のサーバー・バージョンがインストールされたディレクトリーを表します。コマンド・ラインで `-server_directory` フラグと `modelerclient` コマンド・ライン引数を使用して、このデフォルトを上書きすることもできます。

## 「ユーザー・オプション」ダイアログ・ボックスの設定

IBM SPSS Modeler の一般オプションを設定するには、「ツール > オプション」メニューの「ユーザー・オプション」を選択します。これらのオプションは、IBM SPSS Modeler で使用するすべてのストリームに適用されます。

ここでは、次の種類のオプションを設定することができます。これらのオプションを設定するには、該当するタブをクリックしてください。

- 「通知」オプション : モデルの上書きやエラー・メッセージなどのオプションを設定します。
- 「表示」オプション : グラフや背景の色を設定します。
- シンタックスの色表示オプション。
- モデルを Predictive Model Markup Language (PMML) へエクスポートするときに使った PMML エクスポート・オプション。
- 名前、イニシャル、メール・アドレスなどのユーザまたは作成者の情報。この情報を、作成したノードまたはその他のオブジェクトの「注釈」タブに表示することができます。

数字区切り文字、時刻や日付の形式、最適化、ストリーム・レイアウト、およびストリーム・スクリプトなどのストリーム固有のオプションの設定は、「ストリームのプロパティ」ダイアログ・ボックスから行います。このダイアログ・ボックスは、「ファイル」メニューと「ツール」メニューから利用できます。

## 通知オプションの設定

「ユーザー・オプション」ダイアログ・ボックスの「通知」タブでは、IBM SPSS Modeler の警告および確認ウィンドウの表示や種類に関係するさまざまな設定を指定できます。また、新しい出力やモデルが生成された時の、マネージャー・パネルの「出力」タブおよび「モデル」タブの動作も設定することができます。

**ストリーム実行フィードバック・ダイアログの表示** : このオプションを選択すると、ストリームが 3 秒間実行されている場合、進捗指標のインジケータが表示されます。ダイアログ・ボックスには、ストリームで作成された出力オブジェクトの詳細も表示されています。

- **完了時にダイアログを閉じる** : デフォルトでは、ストリームが実行を完了すると、このダイアログ・ボックスが閉じられます。ストリームの実行が完了した後もダイアログ・ボックスを表示したままにする場合は、このボックスの選択を解除 (チェックを外す) します。

**ノードがファイルを上書きする時に警告** : このオプションを選択すると、ノード操作により既存のファイルが上書きされる場合に、警告のエラー・メッセージが表示されます。

**ノードがデータベース・テーブルを上書きする時に警告** : このオプションを選択すると、ノード操作により既存のデータベース・テーブルが上書きされる場合に警告のエラー・メッセージが表示されます。

### 音声通知

下にあるリストを使用して、イベントやエラーが発生した時に通知するサウンドを指定します。さまざまな音声やサウンドを利用できます。選択したサウンドを再生する場合は、「再生」ボタン (スピーカー) を使用します。サウンド・ファイルを探して選択する場合は、「...」ボタンを使用します。

注 : IBM SPSS Modeler のサウンドを作成するために使われる .wav ファイルは、インストール・ディレクトリー中の /media/sounds ディレクトリーにあります。

- **すべての音をミュート** : イベントのサウンド通知を行わない場合に選択します。

### 視覚通知

このグループのオプションを使用して、新規項目生成時のマネージャー・パネル右上の「出力」タブおよび「モデル」タブの動作を指定します。リストから「**新規モデル**」または「**新規出力**」を選択して、対応するタブの動作を設定してください。

**新規モデル**については、次のモデルを使用できます。

**モデルをストリームに追加** : デフォルトでは、「モデル」タブのほか、モデルが作成されるとすぐに新しいモデルがストリームに追加されます。ストリームでは、モデルが作成されたモデル作成ノードへリンクした状態で、モデルが表示されます。このボックスをオフにすると、モデルは「モデル」タブにのみ追加されます。

**前のモデルを置換** : デフォルトでは、「モデル」タブおよびストリーム領域のこのストリームの既存モデルを上書きします。このボックスをオフにすると、「モデル」タブおよび領域の既存モデルにモデルが追加されます。この設定は、モデルリンクのモデル置換設定によって上書きされます。

**新規出力**については、次のモデルを使用できます。

**出力超過時に警告 [n]** : 「出力」タブの項目数が事前に指定された数を超えた場合に警告を表示するかどうかを選択します。デフォルトの数は 20 です。必要に応じて、この数を変更することができます。

すべてのケースで次のオプションを利用できます。

**タブを選択**：ストリームの実行時にオブジェクトが生成された場合に、「出力」タブまたは「モデル」タブに切り替えるかどうかを指定します。

- マネージャー・パネルの該当するタブに切り替える場合は、「常時」を選択します。
- 現在領域に表示されているストリームによってオブジェクトが生成された時にだけタブに切り替える場合は、「現在のストリームが生成した場合」を選択します。
- 出力やモデルが生成されたことを、対応するタブに切り替えて知らせない場合は、「しない」を選択します。

**タブを点滅**：新しい出力やモデルが生成された場合に、マネージャー・パネルにある「出力」タブまたは「モデル」タブを点滅させるかどうかを選択します。

- マネージャー・パネル中の対応するタブを点滅させる場合は (タブが選択されていない場合)、「選択されていない場合」を選択します。
- オブジェクトが生成されたことを、対応するタブを点滅させて知らせない場合は、「しない」を選択します。

**パレットを見えるようにスクロール (新規モデルの場合)**：最新のモデルが見えるように、マネージャー・パネルの「モデル」タブを自動的にスクロールするかどうかを選択します。

- スクロールする場合は、「常時」を選択します。
- 現在領域に表示されているストリームによってオブジェクトが生成された時にだけスクロールする場合は、「現在のストリームが生成した場合」を選択します。
- 「モデル」タブの自動スクロールを行わない場合は、「しない」を選択します。

**ウィンドウを開く (新規出力の場合)**：生成時に出力ウィンドウを自動的に開くかどうかを選択します。

- 常に新しく出力ウィンドウを開く場合は、「常時」を選択します。
- 現在領域に表示されているストリームによって出力が生成された時にだけ新しいウィンドウを開く場合は、「現在のストリームが生成した場合」を選択します。
- 生成された出力に対して自動的に新しいウィンドウを開かない場合は、「しない」を選択します。

このタブのデフォルト設定に戻す場合は、「デフォルト値」を選択します。

## 表示オプションの設定

「ユーザー・オプション」ダイアログ・ボックスの「表示」タブを使用して、IBM SPSS Modeler のフォントや色の表示を設定することができます。

**ようこそダイアログが起動時に表示**：ようこそダイアログ・ボックスが起動時に表示されます。ようこそダイアログには、アプリケーションの例チュートリアルを起動するオプション、デモ ストリームまたは既存のストリームやプロジェクトを開くオプション、または新規ストリームを作成するオプションがあります。

**ストリームおよびスーパーノードのマークアップを表示**：デフォルトでは、ストリームおよびスーパーノードのマークアップがあれば表示されます。マークアップはストリームのコメント、モデル・リンク、およびスコアリングブランチの強調表示で構成されています。

**標準のフォントと色 (再始動後に有効)**：このコントロール・ボックス内のオプションを使用して、IBM SPSS Modeler の画面デザイン、カラー・スキーム、および表示されるフォントのサイズを指定します。ここで指定した内容は、いったん IBM SPSS Modeler を終了してから再起動しないと有効にはなりません。

- 「ルック・アンド・フィール」。標準のカラー・スキームおよび画面デザインを選択できます。次の項目を選択できます。

**SPSS 標準** (デフォルト) : IBM SPSS 製品共通のデザイン。

**SPSS クラシック** : 以前のバージョンの IBM SPSS Modeler のユーザーにお馴染みのデザイン。

**Windows** : ストリーム領域やパレットのコントラストを強くする場合などに役立つ Windows のデザイン。

- **ノードのデフォルトのフォント・サイズ**: ノード・パレット内、およびストリーム領域にノードを表示する際に使用するフォント・サイズを指定します。

注 : 「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」タブの「レイアウト」ペインでストリームのノード・アイコンのサイズを設定できます。メイン・メニューから次の各項目を選択します。

「ツール」 > 「ストリームのプロパティ」 > 「オプション」 > 「レイアウト」

**カスタム カラー**: このテーブルには、さまざまな表示項目に使われる色の、現在の設定が表示されています。テーブルに表示された項目について、「色」列の該当する行をダブルクリックし、リストから色を選択して、現在の色を変更できます。カスタム カラーを指定するには、リストの最後にスクロールして、「色...」 エントリーをクリックします。

**グラフ カテゴリの色順序** : このテーブルには、新しく作成したグラフを表示する場合に使われる色の、現在の設定が表示されています。表示されている色の順序が、グラフで使われる順番を表しています。例えば、色のオーバーレイに使われている名義型フィールドに 4 種類の固有値がある場合、ここに記載されている先頭から 4 つの色が使われます。テーブルに表示された項目について、「色」列の該当する行をダブルクリックし、リストから色を選択して、現在の色を変更できます。カスタム カラーを指定するには、リストの最後にスクロールして、「色...」 エントリーをクリックします。ここで行った変更は、それまでに作成したグラフには適用されません。

このタブのデフォルト設定に戻す場合は、「デフォルト値」を選択します。

## シンタックス表示オプションの設定

「ユーザー・オプション」ダイアログ・ボックスの「シンタックス」タブを使用して、IBM SPSS Modeler で作成するスクリプトのフォント属性や表示色のオプションを設定できます。

**シンタックスの強調表示**。このテーブルには、さまざまなシンタックス項目 (フォントやそれが表示されるウィンドウの両方を含む) に使われる色の、現在の設定が表示されています。テーブルに表示された各項目について、行の該当するドロップダウン・リストをクリックし、リストから色を選択して色を変更できます。また、フォント項目については、太字や斜体の強調を選択できます。

**プレビュー**: このテーブルには、「シンタックスの強調表示」テーブルで選択した色やフォント属性が使用されたサンプルのシンタックス表示が示されます。このプレビューは、ユーザーが選択を行うとすぐに更新されます。

このタブのデフォルト設定に戻す場合は、「デフォルト値」を選択します。

## PMML エクスポート・オプション の設定

「PMML」タブで、IBM SPSS Modeler から Predictive Model Markup Language (PMML) へモデルをエクスポートする方法を制御できます。詳しくは、トピック 142 ページの『PMML としてのモデルのインポートおよびエクスポート』を参照してください。

**PMML をエクスポート**：モデルを PMML (Predictive Model Markup Language) としてエクスポートします。PMML を対象アプリケーションと最高の状態で動作させるための各種設定をここでを行います。

- 「**拡張機能**」を選択し、標準 PMML 相当する機能のない特殊なケースに PMML 拡張機能を使用できるようにします。ほとんどのケースで、標準 PMML と同じ結果が生成されます。
- 「**標準 PMML として**」を選択し、PMML 標準にできる限り準拠するように PMML をエクスポートします。

**標準 PMML オプション**：「**標準 PMML として...**」オプションを選択した場合には、線型回帰モデルとロジスティック回帰モデルをエクスポートするのに有効な方法を、次の 2 つから選択できます。

- **PMML <GeneralRegression> モデルとして**
- **PMML <Regression> モデルとして**

PMML の詳細は、データ・マイニング・グループの Web サイト (<http://www.dmg.org>) を参照してください。

## ユーザー情報の設定

**ユーザーと作成者の情報**：ここで入力した情報は、作成したノードおよびその他のオブジェクトの「注釈」タブに表示されます。

---

## ノード・パレットのカスタマイズ

ストリームは、ノードを使用して構築されます。IBM SPSS Modeler ウィンドウの下部にあるノード・パレットには、ストリームの構築に使われるすべてのノードが用意されています。詳しくは、トピック 17 ページの『ノード・パレット(N)』を参照してください。

ノード・パレットは、次の 2 つの方法で再編成することができます。

- パレット・マネージャーをカスタマイズする。詳しくは、トピック『パレット・マネージャのカスタマイズ』を参照してください。
- サブパレットを含むパレットがノード・パレットにどのように表示されるかを変更する。詳しくは、トピック 160 ページの『サブパレットの作成』を参照してください。

## パレット・マネージャのカスタマイズ

パレット・マネージャーを、IBM SPSS Modeler の使用方法に合わせてカスタマイズすることができます。例えば、データベースの時系列データを頻繁に分析するような場合、データベース入力ノード、時間区分ノード、時系列ノード、時系列グラフ・ノードがすべて一意のパレットのタブで使用できるようにしておく便利です。パレット・マネージャーを使用すると、これらのノードをノード・パレットのカスタム・パレット・タブを作成して容易に調整することができます。

パレット・マネージャーを使用して、次のようなさまざまなタスクを実行することができます。

- ストリーム領域の下のノード・パレットに表示されるパレット・タブを制御する。
- パレット・タブがノード・パレットに表示される順序を変更する。
- パレット・タブおよび関連するサブパレットを作成および編集する。
- 「お気に入り」タブのデフォルトのノード選択を編集する。

パレット・マネージャーにアクセスするには、「ツール」メニューの「**パレットの管理**」をクリックします。

**パレット名:** ノード・パレット上に表示されているかどうかに関係なく、使用できるパレット・タブが表示されます。ここでは、作成したパレット・タブが表示されます。詳しくは、トピック『パレット・タブの作成』を参照してください。

**ノード数:** 各パレット・タブに表示されるノード数。数が多いほど、タブのノードを分割するサブパレットを作成しやすくなります。詳しくは、トピック 160 ページの『サブパレットの作成』を参照してください。

**表示 ?:** ノード・パレット上にパレット・タブを表示します。詳しくは、トピック『ノード・パレットのパレット・タブの表示』を参照してください。

**サブパレット:** パレット・タブに表示するサブパレットを選択するには、該当する「パレット名」強調表示し、このボタンをクリックすると「サブパレット」ダイアログ・ボックスが表示されます。詳しくは、トピック 160 ページの『サブパレットの作成』を参照してください。

**復元デフォルト。** このボタンをクリックすると、パレットおよびサブパレットに行ったすべての変更および追加を完全に削除し、デフォルトのパレット設定に戻します。

## パレット・タブの作成

カスタム・パレット・タブを作成する手順は、次のとおりです。

1. 「ツール」メニューからパレット・マネージャーを起動します。
2. 「表示 ?」列の右側にある「パレットを追加」ボタンをクリックすると、「パレットの作成/編集」ダイアログ・ボックスが表示されます。
3. 一意の「パレット名」に入力します。
4. 「使用可能なノード」領域で、パレット・タブに追加されるノードを選択します。
5. 「ノードを追加」の右矢印ボタンをクリックして、選択強調表示されているノードを「選択されたノード」領域に移動します。該当するすべてのノードを追加するまで、この手順を繰り返します。

該当するノードをすべて追加した後、次のようにしてパレット・タブに表示される順序を変更することができます。

6. 単純な矢印ボタンをクリックすると、ノードが 1 行上または下に移動します。
7. 線の付いた矢印ボタンを使用すると、ノードがリストの最上位または最下位に移動します。
8. ノードをパレットから削除するには、「選択されたノード」領域の右側にある「削除」ボタンをクリックします。

## ノード・パレットのパレット・タブの表示

IBM SPSS Modeler 内に使用しないオプションがある場合があります。この場合、パレット・マネージャーを使用して、これらのノードを含むタブを非表示にすることができます。

ノード・パレットに表示されるタブを選択する手順は、次のとおりです。

1. 「ツール」メニューからパレット・マネージャーを起動します。
2. 「表示 ?」列のチェック・ボックスを使用して、各パレット・タブを表示するか非表示にするかを選択します。

ノード・パレットからパレット・タブを完全に削除するには、ノードを強調表示して、「表示 ?」列の右側にある「削除」ボタンをクリックします。削除したパレット・タブを復元することはできません。

注：「お気に入り」タブを除き、IBM SPSS Modeler で提供されたデフォルトのパレット・タブを削除することはできません。

#### ノード・パレット上の表示順の変更

表示するパレット・タブを選択した後、次のようにしてノード・パレットに表示される順序を変更することができます。

1. 単純な矢印ボタンをクリックすると、パレット・タブが 1 行上または下に移動します。上に移動するとノード・パレットの左側に移動し、下に移動するとノード・パレットの右側に移動します。
2. 線の付いた矢印ボタンを使用すると、パレット・タブがリストの最上位または最下位に移動します。リストの最上位にあるタブ パレットは、ノード・パレットの左側に表示されます。

### 「パレット」タブのサブパレットの表示

ノード・パレットに表示されるパレット・タブを制御するのと同じように、親パレット・タブで使用可能なサブ パレットを制御することができます。

パレット・タブで表示するサブ パレットを選択する手順は、次のとおりです。

1. 「ツール」メニューからパレット・マネージャーを起動します。
2. 必要なパレットを選択します。
3. 「サブ パレット」ボタンをクリックすると、「サブ パレット」ダイアログ・ボックスが表示されます。
4. 「表示？」列のチェック・ボックスを使用して、各サブパレットを「パレット」タブに追加するかどうかを選択します。「すべて」のサブ パレットは常に表示され、削除することはできません。
5. パレット・タブからサブ パレットを完全に削除するには、サブ パレットを強調表示して、「表示？」列の右側にある「削除」ボタンをクリックします。

注：「モデル作成パレット」タブで提供されているデフォルトのサブ パレットを削除することはできません。

#### パレット・タブ上の表示順の変更

表示するサブパレットを選択した後、次のようにして親パレットタブに表示される順序を変更することができます。

1. 単純な矢印ボタンをクリックすると、サブパレットが 1 行上または下に移動します。
2. 線の付いた矢印ボタンを使用すると、サブパレットがリストの最上位または最下位に移動します。

親パレット・タブを選択すると、作成したサブ パレットがノード・パレットに表示されます。詳しくは、トピック 161 ページの『パレット・タブの表示の変更』を参照してください。

### サブパレットの作成

既存のノードを作成するカスタム・パレット・タブに追加することができるため、スクロールせずに容易に表示できないほど多くのノードを選択することができます。スクロールしなくてもすむようにするために、パレット・タブに選択するノードを配置するサブパレットを作成することができます。例えば、ストリーム作成に最も頻繁に使用するノードを含むパレット・タブを作成する場合、入力ノード、フィールド設定、モデル作成および出力で選択を分割するサブパレットを作成することができます。

注：親パレット・タブに追加されたサブパレット・ノードからのみ、選択することができます。

サブパレットを作成するには

1. 「ツール」メニューからパレット・マネージャーを起動します。
2. サブパレットを追加するパレットを選択します。
3. 「サブ パレット」ボタンをクリックすると、「サブ パレット」ダイアログ・ボックスが表示されます。
4. 「表示 ?」列の右側にある「サブ パレットを追加」ボタンをクリックすると、「サブ パレットの作成/編集」ダイアログ・ボックスが表示されます。
5. 一意の「サブ パレット名」に入力します。
6. 「使用可能なノード」領域で、サブパレットに追加されるノードを選択します。
7. 「ノードを追加」の右の矢印のボタンをクリックして、選択されたノードを「選択したノード」に移動します。
8. 該当するノードを追加したら、「OK」をクリックして「サブ パレット」ダイアログ・ボックスに戻ります。

親パレット・タブを選択すると、作成したサブ パレットがノード・パレットに表示されます。詳しくは、トピック『パレット・タブの表示の変更』を参照してください。

## パレット・タブの表示の変更

IBM SPSS Modeler で使用できるノード数が多すぎるため、ノード・パレットの左側または右側へのスクロールが必要ない小さい画面に表示することができません。これは、「モデル作成パレット」タブで特に注意する必要があります。スクロールを少なくするために、サブパレットに含まれるノードのみを表示することができます (該当する場合)。詳しくは、トピック 160 ページの『サブパレットの作成』を参照してください。

パレット・タブに表示されるノードを変更するには、パレット・タブを選択し、左側のメニューですべてのノードを表示するか、特定のサブパレットのノードのみを表示するかを選択します。

---

## CEMI ノードの管理

CEMI は廃止され、より柔軟で使いやすい機能セットを提供する CLEF と置き換えられました。詳しくは、このリリースで提供されている「*IBM SPSS Modeler 16 CLEF 開発者ガイド*」を参照してください。



---

## 第 13 章 ストリームおよびノードのパフォーマンスの考察

ストリームは、最も効果的な構成でノードを配置したり、必要に応じてノード・キャッシュを有効にしたり、このセクションで説明するその他の検討事項に注意を払うことによって、パフォーマンスが最大になるように設計できます。

ここで説明される検討事項のほかに、データベースの効果的な使用、特に SQL 最適化を利用することによって、一般的に、より実質的な追加のパフォーマンス改善が実現します。

---

### ノードの順序

SQL 最適化を使用していなくても、ストリーム内のノードの順序がパフォーマンスに影響を与えることがあります。一般的な目的は下流の処理を最小化することです、そのため、データの量を削減するノードがある場合、ストリームの初めあたりにそれらを配置します。IBM SPSS Modeler Server では、コンパイル時に自動的に順序変更ルールを適用し、安全であると判断した場合に特定のノードを進めることができます（この機能は、デフォルトで有効です。お使いのコンピューターでこの機能が有効であることをシステム管理者に確認してください）。

SQL 最適化を使用する場合は、可用性と効率性を最大限にするように設定します。データベースで実行できない操作がストリームに含まれているときに最適化が停止するので、ストリームの先頭に SQL 最適化の操作をグループ化するのが、もっともよい方法です。このストラテジによりデータベース内での処理が増えるので、IBM SPSS Modeler へ持ち込まれるデータが少なくなります。

ほとんどのデータベース内で、次の操作を実行できます。ストリームの「先頭」に次の操作をグループ化するようにしてください。

- キーによる結合
- 条件抽出
- レコード集計
- ソート
- サンプリング
- レコード追加
- すべてのフィールドが選択される *include* (含む) モードでの、識別の操作
- 置換操作
- 標準的な算術式または文字列操作を使用した、基本的な算出操作（どの操作がデータベースにサポートされるかに依存）
- フラグ設定

以下の操作は、ほとんどのデータベース内で実行できません。上記リストの操作の「後」に、次の操作をストリームに配置する必要があります。

- フラット・ファイルのような、非データベースのデータに対する操作
- 順序による結合
- バランス

- *discard* モードでの識別の操作、またはフィールドのサブセットのみが異種として選択された場合の識別の操作
- 処理対象以外のレコードからのデータへのアクセスが必要な操作
- ステート型および度数型フィールドの導出
- 時系列ノードの操作
- "@" (時系列) 関数に関連する操作
- データ型チェック・モードの「警告」と「中止」
- モデル作成、アプリケーション、および分析

注：ディジション・ツリー、ルールセット、線型回帰、および因子生成のモデルで、SQL を生成し、データベースへプッシュバックできます。

- データを処理中の同じデータベース以外の任意の場所へのデータ出力

---

## ノードのキャッシュ

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

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

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

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

キャッシュを有効にするには

1. ストリーム領域でノードを右クリックし、メニューの「**キャッシュ**」をクリックします。
2. キャッシュ設定用のサブメニューから、「**使用する**」を選択します。
3. キャッシュを解除するには、ノードを右クリックして、キャッシュ設定用のサブメニューから「**使用しない**」を選択します。

データベース内でのノードのキャッシュ

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

55文字より長い文字列でデータベース・キャッシュを使用している場合、どちらかの上流のキャッシュノードから、フィールドの値が読み込まれているそのタイプのノードがあるようにするか、または *options.cfg*

ファイルの `default_sql_string_length` パラメーターを用いて文字列の長さを設定します。それにより、一時テーブルの該当する列を適切な幅に設定し、文字列を調整します。

データベースのキャッシングを利用するには、SQL 最適化とデータベース・キャッシングの両方を有効にする必要があります。Client の設定は、Server の最適化設定によって無効になるので注意してください。詳しくは、トピック 47 ページの『ストリームの最適化オプションの設定』を参照してください。

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

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

---

## パフォーマンス : プロセス・ノード

**ソート :** ソート・ノードでは、ソートする前に入力データ・セット全体を読み込む必要があります。データは、一定の限度までメモリーに格納され、超過分はディスクへ書き出されます。ソート・アルゴリズムは組合せのアルゴリズムです。データは上限までメモリーに読み込まれ、高速で複合的なクイックソート・アルゴリズムを使用してソートされます。すべてのデータがメモリーに収まれば、ソートはこれで完了です。そうでない場合は、マージソートのアルゴリズムが適用されます。ソート済みのデータがファイルへ書き込まれ、次のデータの塊りがメモリーに読み込まれ、ソートされ、ディスクへ書き込まれます。この過程はすべてのデータが読み込まれるまで繰り返され、その後、ソート済みのデータの塊りが併合されます。この併合も、ディスクに格納されたデータに対して繰り返して行うことが必要な場合があります。処理のピーク時には、ソート・ノードが、ソート済みと未ソートの完璧な 2 つのデータ・セットをディスク上に持つこととなります。

このアルゴリズムの全体実行時間は約  $N \cdot \log(N)$  です。ここでの  $N$  はレコード数です。メモリー内でのソートはディスクからの併合よりも高速なので、ソートにメモリーを多く割り当てるほど、実際の実行時間を減らすことができます。このアルゴリズムは、IBM SPSS Modeler Server の構成オプションである「メモリー使用量の乗数」で制御される、物理 RAM の一部をアルゴリズム自体に割り当てます。ソートに使用されるメモリーを増やすには、物理 RAM を増やすか、この値を増やします。使用されるメモリー部分がプロセスの作業セットより大きくなり、メモリーの一部がディスクへページングされると、内部ソート・アルゴリズムのメモリーアクセス・パターンがランダムになり、過度のページングの原因となるので、パフォーマンスが減衰することを憶えておいてください。ソート・アルゴリズムはソート・ノード以外の複数のノードでも使用されますが、同じパフォーマンス・ルールが適用されます。

**データ分割 :** データ分割ノードは、ビンの境界を計算するために、レコードのビンへの割り当て前に入力データ・セット全体を読み込みます。データ・セットは、境界の計算中はキャッシュへ入れられ、その後、割り当て用に再スキャンされます。データ分割の手法が「固定幅」または「平均 + 標準偏差」の場合、データ・セットはディスクへ直接キャッシュされます。これらの手法には直線的に増加する実行時間が必要であり、また、データ・セット全体を格納するのに十分なディスク領域が必要です。データ分割手法が「ランク」または「分位」の場合、データ・セットは先に説明したソート・アルゴリズムを使用してソートされ、ソート済みのデータ・セットがキャッシュとして使用されます。これらの手法のソートの実行時間は  $M \cdot N \cdot \log(N)$  です。ここでの  $M$  はビンに分割されるフィールド数、 $N$  はレコード数です。また、実行にはデータ・セットのサイズの 2 倍のディスク領域が必要です。

生成されたビンに基づいてフィールド作成ノードを生成しておく、以後のパスのパフォーマンスが向上します。フィールド作成の速度は、データ分割に比べてかなり速くなります。

**キーによる結合:** 結合の手法がキー (データベース結合と同じ) であるレコード結合ノードでは、各入力データ・セットがキー・フィールドでソートされます。プロシージャのこの部分の実行時間は  $M*N*\log(N)$  です。 $M$  は入力の数、 $N$  は最大の入力に含まれているレコードの数です。すべての入力データ・セットのほかに、最大のデータ・セットの 2 番目のコピーを格納できるだけの十分なディスク・スペースが必要です。結合自体の実行時間そのものは、出力データ・セットのサイズに比例し、キーが一致する頻度に左右されます。最悪の場合は出力が入力のデカルト積であり、実行時間が  $NM$  に近づくことがあります。これは稀な例ですが、ほとんどの結合では一致するキーが少ない場合があります。1 つのデータ・セットがほかに比べて比較的大きい場合、または、入力データがすでにキー・フィールドでソートされている場合は、「最適化」タブを使用して、このノードのパフォーマンスを改善できます。

**集計:** 「連続キー」オプションが設定されていないと、このノードでは、集計済みの出力を作成する前に入力データ・セット全体を読みます (ただし、格納しません)。さらに極端な状況、つまり、集計されたデータのサイズが制限 (IBM SPSS Modeler Server 構成オプションの「メモリー使用量の乗数」で決定される) に達すると、データセットの残りの部分は、「連続キー」のオプションが設定されたようにしてソートされ、処理されます。このオプションが設定されていると、集計済みの出力レコードが入力データが読み込まれるごとに作成されるので、データは格納されません。

**重複レコード:** 重複レコード・ノードでは、入力データセット内の一意のキー・フィールドのすべてが格納されます。すべてのフィールドがキー・フィールドですべてのレコードが一意であるような場合、データセット全体が格納されます。デフォルトでは、重複レコード・ノードはキー・フィールドのデータをソートし、各グループの最初の重複レコードを選択 (または破棄) します。重複レコード・キーが少ない小さなデータセットの場合、または事前にソートされているデータセットの場合、処理のスピードおよび効率を改善するオプションを選択できます。

**データ型:** このキャッシュは、下流の処理で使用されます。キャッシュでは、データ・セット全体を格納しても処理速度を上げるために、十分なディスク領域が必要です。

**評価:** 評価ノードでは、分位を計算するために入力データをソートする必要があります。スコアと結果のレコード順が各ケースで異なるので、評価済みの各モデルに対してソートが繰り返されます。実行時間は  $M*N*\log(N)$  です。ここでの  $M$  はモデル数、 $N$  はレコード数です。

---

## パフォーマンス :モデル作成ノード

**ニューラルおよび Kohonen:** ニューラル・ネットワークの学習アルゴリズム (Kohonen アルゴリズムも含む) により、学習データに多くの経路が作成されます。データは、一定の限度までメモリーに格納され、超過分はディスクへ書き出されます。アクセス方法がランダムで、その結果ディスク活動が過度になるため、ディスクから学習データへアクセスすることはコスト的に高くなります。ノードのダイアログ・ボックスの「モデル」タブで「速度優先で最適化」オプションを選択することですべてのデータがメモリーに格納されるようにし、これらのアルゴリズムのディスク領域の使用を無効化することができます。データを格納するのに必要なメモリー量がサーバー・プロセスの作業セットより大きい場合、メモリーの一部がディスクへページングされて、それによってパフォーマンスが落ちることになります。

「メモリー優先で最適化」が有効にされると、IBM SPSS Modeler Server 構成オプションの「モデル作成メモリー制限の割合」の値に従って、物理 RAM の指定されたパーセンテージがアルゴリズムに割り当てられます。学習ニューラル・ネットワークにさらに多くのメモリーを使用するには、RAM をさらに増やすか、このオプションの値を増やすかですが、設定値が高すぎると、ページングの原因になります。

ニューラル・ネットワーク・アルゴリズムの実行時間は、精度のレベルに左右されます。実行時間は、ノードのダイアログ・ボックスで停止条件を設定することで制御できます。

**K-Means** : K-Means クラスタリング・アルゴリズムには、ニューラル・ネットワーク・アルゴリズムと同様の、メモリー使用量を制御するオプションがあります。データへのアクセスがシーケンシャルであるため、ディスクに格納されたデータへのパフォーマンスのほうが良好です。

---

## パフォーマンス :CLEM 式

データ・ストリーム内を参照する CLEM シーケンス関数 (「@ 関数」) では、最長の参照を可能にするのに十分なデータを格納する必要があります。参照の程度に制限がない操作では、フィールドのすべての値が格納される必要があります。制限がない操作とは、例えば @OFFSET(Sales, Month) のように、オフセット値がリテラルの整数ではない操作です。オフセット値がフィールド名 *Month* なので、その値は実行されるまで不明です。正確な結果を保証するために、サーバーは *Sales* フィールドのすべての値を保存しておく必要があります。上限値がわかっている場合は、例えば @OFFSET(Sales, Month, 12) のように、上限値を追加の引数として指定する必要があります。この指定により、サーバーは *Sales* の 12 の最新値以降を保存しないようになります。制限付きか付きでないかにかかわらず、シーケンス関数によって、ほとんど常に SQL 生成が抑制されます。



---

## 第 14 章 IBM SPSS Modeler のアクセシビリティ

---

### IBM SPSS Modeler のアクセシビリティの概要

IBM SPSS Modeler では、視覚障害やその他の機能障害のあるユーザーに対する特定のサポートの他に、すべてのユーザーにアクセシビリティ・サポートが提供されます。このセクションでは、画面読み上げソフトウェアやキーボード・ショートカットなどのアクセシビリティの概要、およびこれらの機能を使った作業方法を説明していきます。

---

### アクセシビリティサポートの種類

視覚障害をお持ちの方や、キーボードを使った操作しかできないような方でも利用できるように、さまざまな代替手段が用意されています。例えば、ストリームの構築、オプションの設定、および出力の読み込みなどの作業を、すべてマウスを使わないで行うことができます。利用できるキーボード・ショートカットについては、後述します。また、IBM SPSS Modeler は、JAWS for Windows のような画面読み上げソフトウェアもサポートしています。さらに、コントラストを強くするために、カラー・スキーマを最適化することもできます。これらのサポート内容の詳細は、以降のトピックで説明していきます。

### 視力の弱いユーザーのためのアクセシビリティ

ソフトウェアを使いやすくするために、IBM SPSS Modeler にはさまざまなプロパティが用意されています。

#### 表示オプション

グラフの表示色を選択することができます。また、ソフトウェア自体に対して特定の Windows 設定を利用することもできます。これは、コントラストを強くするような場合に役立ちます。

1. 表示オプションを設定するには、「ツール」メニューの「ユーザー・オプション」をクリックします。
2. 「表示」タブをクリックします。このタブには、ソフトウェアの表示色、グラフの色、およびノードのフォント・サイズなどを設定できます。

#### サウンドを通知に使用

サウンドをオンまたはオフにして、ソフトウェアの特定の操作をサウンドで知らせるかどうかを制御できます。例えば、ノードの作成、削除、または新規出力やモデルの生成時に、サウンドを鳴らすことができます。

1. 通知オプションを設定するには、「ツール」メニューの「ユーザー・オプション」をクリックします。
2. 「通知」タブをクリックします。

#### 新規ウィンドウの自動起動

「ユーザー・オプション」ダイアログ・ボックスの「通知」タブでは、テーブルやグラフなどの新しく生成された出力を、新しいウィンドウに表示するかどうかを指定することもできます。このオプションを無効にして、必要に応じて出力ウィンドウを表示することもできます。

1. これらのオプションを設定するには、「ツール」メニューの「ユーザー・オプション」をクリックします。
2. 「通知」タブをクリックします。
3. ダイアログ・ボックスで、「ビジュアル通知」のリストから「新規出力」を選択します。
4. 「ウィンドウを開く」で、「開かない」を選択します。

#### ノード・サイズ

ノードは、標準のサイズまたは小さいサイズで表示することができます。必要に応じて、大きさを調整してください。

1. ノード・サイズ・オプションを設定するには、「ファイル」メニューの「ストリームのプロパティ」をクリックします。
2. 「レイアウト」タブをクリックします。
3. 「アイコン・サイズ」リストで、「標準」を選択します。

## 視覚障害ユーザーのためのアクセシビリティ

視覚障害者向けのサポート機能は、JAWS for Windows のような画面読み上げソフトウェアを使用しているかどうか大きく依存しています。IBM SPSS Modeler と一緒に使用する画面読み上げソフトウェアを最大限に活用できるように、多くの設定を指定できます。

#### 表示オプション

スクリーン・リーダー (画面読み上げソフトウェア) は、画面のコントラストが強い方が、より正確に動作する傾向があります。Windows をすでに高コントラストの設定にしている場合は、それらの Windows 設定をソフトウェアで使用することもできます。

1. 表示オプションを設定するには、「ツール」メニューの「ユーザー・オプション」をクリックします。
2. 「表示」タブをクリックします。

#### サウンドを通知に使用

サウンドをオンまたはオフにして、ソフトウェアの特定の操作をサウンドで知らせることができます。例えば、ノードの作成、削除、または新規出力やモデルの生成時に、サウンドを鳴らすことができます。

1. 通知オプションを設定するには、「ツール」メニューの「ユーザー・オプション」をクリックします。
2. 「通知」タブをクリックします。

#### 新規ウィンドウの自動起動

「ユーザー・オプション」ダイアログ・ボックスの「通知」タブでは、新しく生成された出力を、新しいウィンドウに表示するかどうかを指定することもできます。このオプションを無効にして、必要に応じて出力ウィンドウを表示することもできます。

1. これらのオプションを設定するには、「ツール」メニューの「ユーザー・オプション」をクリックします。
2. 「通知」タブをクリックします。
3. ダイアログ・ボックスで、「ビジュアル通知」のリストから「新規出力」を選択します。
4. 「ウィンドウを開く」で、「開かない」を選択します。

## キーボード・アクセシビリティ

製品の機能にキーボードを使用してアクセスできます。基本的には、Alt キーと他の適切なキーを同時に押してメニュー項目を選択したり (例: Alt + F キーで「ファイル」メニューを選択)、Tab キーを使用してダイアログ・ボックス中のコントロール間を移動することができます。しかし、それ以外に製品のメイン・ウィンドウのそれぞれに関連した固有な問題があり、ダイアログ・ボックス内での操作についての便利なヒントもあります。

このセクションでは、ストリームを開く、ノード・ダイアログ・ボックスの使用、および出力作業などを行うために利用できる、キーボードによるアクセシビリティを説明していきます。また、効率的な移動を行うために、キーボード・ショートカットの一覧も記載されています。

### メイン・ウィンドウ内を移動するためのショートカット

大半の作業は IBM SPSS Modeler のメイン・ウィンドウで行われます。メイン領域はストリーム領域と呼ばれ、データ・ストリームの作成と実行に用いられます。ウィンドウの下部にあるノード・パレットには、利用できるノードが用意されています。パレットは、データ・マイニング操作の種類に応じたタブにわかれており、それぞれに適切なノードが配置されています。例えば、IBM SPSS Modeler にデータを取り込むために使われるノードは、「入力」タブに配置されています。また、フィールドの作成、フィルタリング、およびデータ型の設定に使われるノードは、「フィールド設定」タブに配置されています。

ウィンドウの右側には、ストリーム、出力、およびプロジェクトを管理するためのさまざまなツールが用意されています。ウィンドウ右側の上半分には、ストリーム、出力、および生成されたモデルを管理するために用いられるマネージャがあります。このマネージャには、3 つのタブがあります。ストリーム、出力、生成されたモデルなどのオブジェクトにアクセスするには、適切なタブを選択して、リストからオブジェクトを選択します。メイン・ウィンドウ右側の下半分には、プロジェクトの作業を行うためのプロジェクト・パネルがあります。このエリアには、プロジェクトに 2 の異なるビューが存在することを反映して 2 つのタブがあります。クラス・ビューは、プロジェクト・オブジェクトをタイプごとにソートしますが、CRISP-DM ビューは、データ準備およびモデル作成などの関連するデータ・マイニング・フェーズごとにオブジェクトをソートします。これらの IBM SPSS Modeler ウィンドウの各部の詳細は、ヘルプやマニュアルを参照してください。

IBM SPSS Modeler のメイン・ウィンドウで、およびストリームの構築に使用できるショートカットの一覧を次の表に示します。ダイアログ・ボックスや出力で用いられるショートカットは後述します。これらのショートカット・キーは、メイン・ウィンドウでしか利用できないことに注意してください。

表 32. メイン・ウィンドウのショートカット

ショートカット・キー	関数
Ctrl+F5	フォーカスをノード・パレットに移動します。
Ctrl+F6	フォーカスをストリーム領域に移動します。
Ctrl+F7	フォーカスをマネージャ・パネルに移動します。
Ctrl+F8	フォーカスをプロジェクト・パネルに移動します。

表 33. ノードとストリームのショートカット

ショートカット・キー	関数
Ctrl+N	ストリーム領域に新しい空のストリームを作成します。
Ctrl+O	既存のストリームを選択して開くことのできる「開く」ダイアログボックスを表示します。

表 33. ノードとストリームのショートカット (続き)

ショートカット・キー	関数
Ctrl+数字キー	フォーカスをウィンドウまたはペインの対応するタブに移動します。例えば、タブのある領域やウィンドウで、Ctrl-1 を押すと左から 1 番目のタブが、Ctrl-2 を押すと 2 番目のタブが選択されます。
Ctrl+下矢印キー	ノード・パレットでフォーカスをパレット・タブからそのタブの最初のノードまで移動するのに使用します。
Ctrl+上矢印キー	ノード・パレットでフォーカスをノードからそのパレット・タブまで移動するのに使用します。
強制投入法 :	ノード・パレット (「生成されたモデル」パレットの調整済みモデルも含めて) でノードが選択されている場合に、このキーを押すとノードがストリーム領域に追加されます。キャンバスでノードがすでに選択されている場合に <b>Enter</b> を押すと、そのノードのダイアログ・ボックスが開きます。
Ctrl+Enter	パレットでノードが選択されている場合に、ノードをストリーム領域に追加します。ただし、追加したノードは選択されず、フォーカスはノード・パレットのまま変わりません。
Alt+Enter	パレットでノードが選択されている場合に、ノードをストリーム領域に追加して選択します。ただしフォーカスはノード・パレットのまま変わりません。
Shift+スペースキー	ノードまたはコンポーネントがパレットにフォーカスを置く場合、ノードまたはコンポーネントの選択と非選択を切り替えます。ほかのノードまたはコンポーネントも選択されている場合、これらの選択は解除されます。
Ctrl+Shift+スペースキー	ノードまたはコメントがストリームにフォーカスを置く場合、またはノードまたはコメントがパレットにフォーカスを置く場合、ノードまたはコメントの選択と非選択を切り替えます。そのほかの選択されたノードまたはコメントに影響はありません。
左右の矢印キー	ストリーム領域にフォーカスがある場合、ストリーム全体を画面上で平行に移動します。パレット・タブにフォーカスがある場合、タブの間を順次移動します。パレット・ノードにフォーカスがある場合、パレットのノード間を移動します。
上下の矢印キー	ストリーム領域にフォーカスがある場合、ストリーム全体を画面上で平行に移動します。パレット・ノードにフォーカスがある場合、パレットのノード間を移動します。サブパレットにフォーカスがある場合、このパレット・タブに対して他のサブパレット間を移動します。
Alt+左右の矢印キー	ストリーム領域上で選択されているノードまたはコメントを、矢印キーの方向に平行に移動します。
Alt+上下の矢印キー	ストリーム領域上で選択されているノードまたはコメントを、矢印キーの方向で垂直方向に移動します。
Ctrl+A	ストリーム内のすべてのノードを選択します。
Ctrl+Q	ノードにフォーカスがある場合に、そのノードと下流にあるすべてのノードを選択し、上流にあるすべてのノードを非選択にします。
Ctrl+W	ノードにフォーカスがある場合に、そのノードと下流にあるすべての選択したノードを非選択にします。
Ctrl+Alt+D	選択したノードを複製します。

表 33. ノードとストリームのショートカット (続き)

ショートカット・キー	関数
Ctrl+Alt+L	ストリームでモデル・ナゲットが選択されている場合、「挿入」ダイアログを開いて .nod ファイルからストリームへ保存したモデルをロードできるようにします。
Ctrl+Alt+R	選択したノードの「注釈」タブを表示し、ノードの名前が変更できるようになります。
Ctrl+Alt+U	ユーザー入力ソース・ノードを生成します。
Ctrl+Alt+C	ノードのキャッシュのオンとオフを切り替えます。
Ctrl+Alt+F	ノードのキャッシュを消去します。
タブ	ストリーム領域で、現在のストリームのすべての入力ノードおよびコメントを順次移動します。ノード・パレットで、パレットのノード間を移動します。選択したサブパレットで、サブパレットの最初のノードに移動します。
Shift+Tab	Tab と同じ操作を実行しますが、順序が逆になります。
Ctrl+Tab	マネージャー・パネルまたはプロジェクト・パネルにフォーカスを置き、ストリーム領域にフォーカスを移動します。ノード・パレットにフォーカスを置き、ノードとパレット・タブ間でフォーカスを移動します。
アルファベット・キー	現在のストリームのノードにフォーカスがある状態から、名前が押したキーの文字で始まる次のノードにフォーカスを移動します。
[F1] キー	ヘルプのフォーカスに関連したトピックを開きます。
[F2] キー	ストリーム領域上で選択されているノードの接続処理を開始します。Tab キーを使用して接続先のノードを選択し、Shift + スペース・バーを押すと接続が行われます。
[F3] キー	ストリーム領域上で選択されているノードに対するすべての接続を削除します。
[F6] キー	マネージャー・パネル、プロジェクト・パネルおよびノード・パレット間でフォーカスを移動します。
F10	「ファイル」メニューを表示します。
Shift+F10	ノードまたはストリームのポップアップ・メニューを表示
Delete	キャンバスから選択したノードを削除します。
Esc	ポップアップ・メニューまたはダイアログ・ボックスを閉じます。
Ctrl+Alt+X	スーパーノードを展開します。
Ctrl+Alt+Z	スーパーノードをズーム・インします。
Ctrl+Alt+Shift+Z	スーパーノードからズーム・アウトします。
Ctrl+E	ストリーム領域上にフォーカスを置きながら、現在のストリームを実行します。

ほかにも、コピーの場合は Ctrl-C キーなど、IBM SPSS Modeler ではさまざまな標準のショートカットキーを利用することができます。詳しくは、トピック 23 ページの『ショートカット・キーの使用』を参照してください。

## ダイアログ・ボックスや表のショートカット・キー

ダイアログ・ボックス、表、およびダイアログ・ボックス中の表に関する作業を行う際に役立つさまざまなショートカットキーや画像読み上げソフトウェア用キーが用意されています。これらのキーの一覧を次に示します。

表 34. ダイアログ・ボックスと Clem 式ビルダーのショートカット

ショートカット・キー	関数
Alt+4	開かれているすべてのダイアログ・ボックスまたは出力ウィンドウを閉じるために用いられます。出力は、マネージャー・パネルの「出力」タブから取得することができます。
[Ctrl] + [End]	Clem 式ビルダー内のコントロールにフォーカスがある場合に、挿入ポイントを式の最後に移動します。
Ctrl+1	Clem 式ビルダーで、フォーカスを CLEM 式編集コントロールに移動します。
Ctrl+2	Clem 式ビルダーで、フォーカスを関数リストに移動します。
Ctrl+3	Clem 式ビルダーで、フォーカスをフィールド・リストに移動します。

### テーブルのショートカット

テーブルのショートカットは、出力テーブルや、データ型、フィルター、およびレコード結合などのノードのダイアログ・ボックス中にあるテーブル・コントロールで用いられます。一般的には、テーブル中のセル間の移動には Tab キーを、テーブル・コントロールを終了するには Ctrl-Tab キーを使用します。注：場合によっては、画面読み上げソフトウェアがセルの内容をすぐに読み上げ始めないこともあります。矢印キーを 1 回または 2 回押すと、ソフトウェアがリセットされて、読み上げが開始されます。

表 35. テーブルのショートカット

ショートカット・キー	関数
Ctrl+W	テーブルで選択されている行 ( <b>row</b> ) の簡単な説明を読み上げます。例えば、「選択されている行 2 の値は、性別、フラグ型、男/女…」のように読み上げます。
Ctrl+Alt+W	テーブルで選択されている行 ( <b>row</b> ) の長い説明を読み上げます。例えば、「選択されている行 2 の値は、フィールド = 性別、データ型 = フラグ型、性別 = 男/女…」のように読み上げます。
Ctrl+D	テーブルで選択されている領域の、簡単な説明 ( <b>description</b> ) を読み上げます。例えば、「列 6 の 1 行目が選択されています。」のように読み上げます。
Ctrl+Alt+D	テーブルで選択されている領域の、長い説明 ( <b>description</b> ) を提供します。例えば、「1 行 6 列が選択されています。選択されている列は、フィールド、データ型、欠損値です。選択された行は 1 です。」のようになります。
Ctrl+T	テーブルで選択されている領域の、簡単な説明を提供します。例えば、「フィールド、データ型、欠損値」のようになります。
Ctrl+Alt+T	テーブルで選択されている領域の、長い説明を提供します。例えば、「選択されている列は、フィールド、データ型、欠損値です。」のようになります。
Ctrl+R	テーブルで、テーブル中のレコード数 ( <b>records</b> ) を提供します。

表 35. テーブルのショートカット (続き)

ショートカット・キー	関数
Ctrl+Alt+R	テーブルで、テーブル中のレコード数 ( <b>records</b> )、および列名を提供します。
Ctrl+I	テーブルで、フォーカスがあるセルのセル情報 ( <b>information</b> ) または内容を読み上げます。
Ctrl+Alt+I	テーブルで、フォーカスがあるセルのセル情報 ( <b>information</b> ) の長い説明 (列名とセルの内容) を読み上げます。
Ctrl+G	テーブルで、簡単な一般選択情報 ( <b>general</b> ) を提供します。
Ctrl+Alt+G	テーブルで、長い一般選択情報 ( <b>general</b> ) を提供します。
Ctrl+Q	テーブルで、テーブル・セルを素早く ( <b>quick</b> ) 切り替えられるようにします。Ctrl-Q を押すと、テーブル内を矢印キーを使用して移動すれば、その詳細な説明が「性別 = 女性」のように読み上げられます。もう一度 Ctrl-Q を選択すると、簡単な説明に切り替わります (セルの内容)。

## コメントのショートカット

画面上のコメントを処理している場合、次のショートカットを使用できます。

表 36. コメントのショートカット

ショートカット・キー	関数
Alt+C	コメントの表示/非表示機能を切り替えます。
Alt+M	コメントが現在表示されている場合は新しいコメントを挿入し、非表示になっている場合はコメントを表示します。
タブ	ストリーム領域で、現在のストリームのすべての入力ノードおよびコメントを順次移動します。
強制投入法 :	コメントにフォーカスがある場合、編集の開始を指示します。
Alt+Enter または Ctrl+Tab	編集を終了し、編集の変更を保存します。
Esc	編集を中止します。編集集中に行われた変更は失われます。
Alt+Shift+上矢印キー	グリッドによる配置の設定がオン (またはオフ) の場合、1 グリッド・セル (または 1 ピクセル) ごとにテキスト領域の高さを縮小します。
Alt+Shift+下矢印キー	グリッドによる配置の設定がオン (またはオフ) の場合、1 グリッド・セル (または 1 ピクセル) ごとにテキスト領域の高さを拡大します。
Alt+Shift+左矢印キー	グリッドによる配置の設定がオン (またはオフ) の場合、1 グリッド・セル (または 1 ピクセル) ごとにテキスト領域の幅を縮小します。
Alt+Shift+右矢印キー	グリッドによる配置の設定がオン (またはオフ) の場合、1 グリッド・セル (または 1 ピクセル) ごとにテキスト領域の幅を拡大します。

## クラスター・ビューアーおよびモデル・ビューアーのショートカット

ショートカット・キーは、クラスター・ビューアーおよびモデル・ビューアー・ウィンドウを移動するために使用できます。

表 37. 汎用ショートカット - クラスター・ビューアーおよびモデル・ビューアー

ショートカット・キー	関数
タブ	フォーカスを次の画集コントロールに移動します。

表 37. 汎用ショートカット - クラスター・ビューアーおよびモデル・ビューアー (続き)

ショートカット・キー	関数
Shift+Tab	フォーカスを前の画面コントロールに移動します。
[↓]	ドロップダウン・リストにフォーカスがある場合、リストを開くか、リストの次の項目に移動します。 メニューにフォーカスがある場合、メニューの次の項目に移動します。 サムネイル・グラフにフォーカスがある場合、セット内の次のサムネイル (または最後のサムネイルにフォーカスがある場合は最初のサムネイル) に移動します。
[↑]	ドロップダウン・リストにフォーカスがある場合、リストの前の項目に移動します。 メニューにフォーカスがある場合、メニューの前の項目に移動します。 サムネイル・グラフにフォーカスがある場合、セット内の前のサムネイル (または最初のサムネイルにフォーカスがある場合は最後のサムネイル) に移動します。
強制投入法 :	使用中のドロップダウン・リストを閉じるか、または使用中のメニューで選択を行います。
[F6] キー	ウィンドウの左側および右側のパネルでフォーカスを切り替えます。
左右の矢印キー	タブにフォーカスがある場合、メニューの前後のタブに移動します。 メニューにフォーカスがある場合、前後のメニューに移動します。
Alt+文字	その名前が下線の付いた文字を含むボタンまたはメニューを選択します。
Esc	使用中のメニューまたはドロップダウン・リストを閉じます。

#### クラスター・ビューアーのみ

クラスター・ビューアーには、クラスター特徴グリッドを含むクラスター・ビューがあります。

モデルの要約ビューの代わりにクラスター・ビューを選択する手順は、次のとおりです。

1. 「表示」 ボタンが選択されるまで、Tab キーを繰り返し押します。
2. 下向き矢印を 2 回押して、「クラスター」 を選択します。

ここから、グリッド内の個別のセルを選択できます。

3. 視覚化ツールバーの最後のアイコンが選択されるまで、Tab キーを繰り返し押します。



図 17. 「視覚化ツリーを表示」アイコン

4. Tab キーをもう一度押し、その後スペース、そして矢印キーを押します。

次のキーボード・ショートカットを使用できます。

表 38. クラスター・ビューアーのショートカット

ショートカット・キー	関数
矢印キー	グリッド内の各セル間でフォーカスを移動します。右側のパネルのセルの分布表示が、フォーカスが移動するごとに変わります。

表 38. クラスター・ビューアーのショートカット (続き)

ショートカット・キー	関数
Ctrl+コンマ	セルにフォーカスがあるグリッドの列全体を選択または選択解除します。 列を選択部分に追加するには、矢印キーを使用して、その列のセルに移動し、もう一度 Ctrl キーを押しながらコンマを押します。
タブ	フォーカスをグリッドから次の画面コントロールに移動します。
Shift+Tab	フォーカスをグリッドから前の画面コントロールに移動します。
[F2] キー	編集モード (ラベルおよびセル名のセルのみ) に入ります。
強制投入法 :	編集した変更を保存し、編集モード (ラベルおよびセル名のセルのみ) を終了します。
Esc	変更を保存せずに編集モード (ラベルおよびセル名のセルのみ) を終了します。

### ショートカット・キーの例 :ストリームの構築

キーボードや画面読み上げソフトウェアを使用した、ユーザーのストリームの構築作業を理解するために、マウスを使わないでストリームを構築する例を取り上げます。この例では、可変長ファイルノード、フィールド作成ノード、およびヒストグラム・ノードがあるストリームを、次の手順にしたがって作成していきます。

1. **IBM SPSS Modeler**を始動します。IBM SPSS Modeler が最初に始動される際、フォーカスはノード・パレットの「お気に入り」タブにあります。
2. **Ctrl+下矢印キー**: フォーカスをタブ自体から、タブの内容へと移動します。
3. **右矢印キー**: フォーカスを可変長ファイル・ノードに移動します。
4. **スペース・バー**: 可変長ファイル・ノードを選択します。
5. **Ctrl+Enter**: 可変長ファイル・ノードをストリーム領域に追加します。このキーの組み合わせでは、可変長ファイル・ノードが選択されたままになるため、次のノードを追加すると可変長ノードに接続されます。
6. **Tab**: フォーカスをノード・パレットに移動します。
7. **右矢印キーを 4 回**: フィールド作成ノードを移動します。
8. **スペース・バー**: フィールド作成ノードを選択します。
9. **Alt+Enter**: フィールド作成ノードをストリーム領域に追加して、フィールド作成ノードを選択します。これで、このノードは次に追加されるノードと接続されることとなります。
10. **Tab**: フォーカスをノード・パレットに移動します。
11. **右矢印キーを 5 回**: フォーカスがパレット中のヒストグラム・ノードに移動します。
12. **スペース・バー**: ヒストグラム・ノードを選択します。
13. **Enter**: ノードをストリームに追加し、フォーカスをストリーム領域に移動します。

次の例に進むか、後で次の例を試す場合はストリームを保存します。

### ショートカット・キーの例 :ノードの編集

この例では、前の例で作成したストリームを使用します。このストリームは、可変長ファイル・ノード、フィールド作成ノード、およびヒストグラム・ノードから成り立っています。この例では、ストリーム中の 3 番目にあるヒストグラム・ノードにフォーカスがある所から始まります。

1. **Ctrl+ 左矢印キーを 2 回**: フォーカスを可変長ファイル・ノードに移動します。

2. **Enter** : 「可変長ファイル」ダイアログ・ボックスを開きます。「ファイル」フィールドに Tab キーで移動し、テキスト・ファイルのパスおよび名前を入力して、そのファイルを選択します。Ctrl+Tab キーでダイアログの下部分に移動し、Tab キーで「OK」ボタンに移動して Enter キーを押すとダイアログ・ボックスが閉じます。
3. **Ctrl+右矢印キー**: フォーカスを 2 番目のフィールド作成ノードに移動します。
4. **Enter** : 「フィールド作成ノード」ダイアログ・ボックスが開きます。Tab キーを何回か押してフィールドを選択し、フィールド作成条件を指定します。Ctrl+Tab キーで「OK」ボタンまで移動し、Enter キーを押すとダイアログ・ボックスが閉じます。
5. **Ctrl+右矢印キー**: フォーカスを 3 番目のヒストグラム・ノードに移動します。
6. **Enter** : 「ヒストグラム・ノード」ダイアログ・ボックスが表示されます。Tab キーを何回か押してフィールドを選択し、グラフ・オプションを指定します。ドロップダウン・リストでは、下矢印キーを押してリストを開きリスト項目を強調表示してから、Enter キーを押してリスト項目を選択します。Tab キーで「OK」ボタンまで移動し、Enter キーを押すとダイアログ・ボックスが閉じます。

この時点で、他のノードを追加したり、現在のストリームを実行することができます。ストリームを構築するときには、忘れずに次のヒントに従ってください。

- 手作業でノードを接続する場合、F2 キーを使用して接続の始点を作成し、Tab キーを同時に使用して終点まで移動し、Shift キーとスペース・バーを使用して接続を完了します。
- ストリーム領域で選択されているノードのすべての接続を解除するには、F3 キーを使用します。
- ストリームを作成した後、Ctrl + E キーを押すと現在のストリームを実行します。

ショートカット・キーの完全な一覧が用意されています。詳しくは、トピック 171 ページの『メイン・ウィンドウ内を移動するためのショートカット』を参照してください。

## 画面読み上げソフトウェアの使用

多数の画面読み上げソフトウェアが市場で販売されています。IBM SPSS Modeler は IBM SPSS Modeler と一緒にインストールされる Java Access Bridge を使用して Windows JAWSに対応するよう設定されています。JAWS がインストール済みであれば、IBM SPSS Modeler を起動する前に JAWS を起動するだけで利用できます。

IBM SPSS Modeler 独自のデータ・マイニング・プロセスのグラフィック表示のため、グラフは視覚的に使用するために最適化されています。ただし、画面読み上げソフトウェアを使用して「文字的に」表示できる出力やモデルを理解しながら、作業を行うことができます。

注：64 ビットのクライアント・コンピューターでは、支援技術機能が機能しません。これは Java Access Bridge が 64 ビットの操作向けに設計されていないためです。

### IBM SPSS Modeler デクシオナリー・ファイルの使用

IBM SPSS Modeler デクシオナリー・ファイル (*Awt.JDF*) は、JAWS に含めて利用することができます。このファイルを使用するには

1. IBM SPSS Modeler のインストール・ディレクトリー中の *laccessibility* サブディレクトリーに移動して、デクシオナリー・ファイル (*Awt.JDF*) をコピーします。
2. コピーしたファイルを、JAWS スクリプトのあるディレクトリーに貼り付けます。

他の JAVA アプリケーションを稼働させている場合は、*Awt.JDF* という名前のファイルがすでに存在していることもあります。この場合、手作業でデクシオナリー・ファイルを編集しないと、このファイルを使用できないこともあります。

## HTML 出力での画面読み上げソフトウェアの使用

IBM SPSS Modeler で画面読み上げソフトウェアを使用して HTML 出力を表示する場合、いくつかの問題が発生する可能性があります。次のように、多数の出力タイプが影響を受けます。

- 線形回帰ノード、ロジスティック回帰ノード、および因子分析ノードの「詳細」タブに表示する出力。
- レポート・ノードの出力

これらの各ウィンドウまたはダイアログ・ボックスには、出力をデフォルトのブラウザに表示するためのツールがツールバーに用意されています。このデフォルトのブラウザは、標準の画面読み上げソフトウェアをサポートしています。ブラウザに表示した出力情報を、画面読み上げソフトウェアを使用して調査します。

### 「インタラクティブ・ツリー」ウィンドウのアクセス

「インタラクティブ・ツリー」ウィンドウ内でのディビジョン・ツリーの標準的な表示方法は、画面読み上げソフトウェアを使用する際に問題になることがあります。アクセス可能なバージョンにアクセスするには、インタラクティブ・ツリーのメニューから次のメニュー項目を選択してください。

「表示」 > 「アクセス可能なウィンドウ」

標準のツリー・マップと似ていますが、JAWS が正しく読み上げることができるビューを表示します。標準的な矢印キーを使用すると、上下左右に移動できます。アクセス可能なウィンドウ内を移動すると、それに応じて「インタラクティブ・ツリー」ウィンドウ内のフォーカスも移動します。選択を変更するにはスペース・バーを使用します。また、現在の選択を拡張するには、Ctrl+スペースを使用します。

---

## 使用上のヒント

IBM SPSS Modeler 環境をより使いやすくするための、さまざまなコツがあります。IBM SPSS Modeler で作業を行う際の一般的なヒントを次に示します。

- **拡張テキスト・ボックスの終了**：拡張テキスト・ボックスを終了するには Ctrl+Tab を使用します。注：Ctrl-Tab は、テーブル・コントロールを終了する際にも用いられます。
- **矢印キーよりも Tab キーを利用する**：ダイアログ・ボックスのオプションを選択する場合、オプション・ボタン間の移動には Tab キーを使用します。ここでは、矢印キーを移動に使用することはできません。
- **ドロップダウン・リスト**：ダイアログ・ボックスのドロップダウン・リストでは、Esc キーまたはスペース・バーを使用して項目を選択し、リストを閉じることができます。他のコントロールにタブで移動しており、閉じていないドロップダウン・リストを閉じるには、Esc キーを使用することもできます。
- **実行ステータス**：大きいデータベースでストリームを実行している場合、JAWS ではストリーム・ステータスの読み上げを遅くすることができます。ステータス・レポートを更新するには、Ctrl キーを定期的に押してください。
- **ノード・パレットの使用**：ノード・パレットのタブを初めて参照する場合、JAWS がノード名の代わりに「グループボックス」と読み上げることがあります。この場合、Ctrl-右矢印キーまたは Ctrl-左矢印キーを使用して、画面読み上げソフトウェアをリセットしてから、ノード名を読み上げてください。
- **メニューの読み上げ**：最初にメニューを開く際に、JAWS が最初のメニュー項目を読み上げないことがあります。この問題が発生している可能性がある場合は、下矢印キーを押した後に上矢印キーを押して、最初のメニュー項目を確認してください。
- **カスケード・メニュー**：JAWS は、カスケード・メニューのメイン（第 1 レベル）メニューは読み上げません。メニュー内を移動している際に、何も読み上げられないことがあった場合は、右矢印キーを押して子メニューの項目を読み上げてください。

また、IBM SPSS Modeler Text Analytics をインストールした場合、次のヒントでインタラクティブ・ワークベンチ・インターフェースがより使いやすくなります。

- **ダイアログ・ボックスの入力**：ダイアログ・ボックスに入力直後、Tab キーを押して最初のコントロールに焦点を当てることが必要な場合があります。
- **拡張テキスト・ボックスの終了**：Ctrl+Tab キーを押して、拡張テキスト・ボックスを終了し、次のコントロールに移動します。注：Ctrl-Tab は、テーブル・コントロールを終了する際にも用いられます。
- **最初の文字を入力してツリー・リストの要素を検索**：カテゴリ領域、抽出された結果の領域、またはライブラリー・ツリーで要素を検索する場合、領域にフォーカスがある際に要素の最初の文字を入力することができます。これにより、入力した文字で始まる要素の次の出現を選択します。
- **ドロップダウン・リスト**：ダイアログ・ボックスのドロップダウン・リストでは、スペース・バーを使用して項目を選択し、リストを閉じることができます。

使用上のその他のヒントは、以降のトピックで詳細に説明します。

## 他のソフトウェアとの相性

弊社開発チームが IBM SPSS Modeler で JAWS などの画面読み上げソフトウェアをテストした際に、SMS (Systems Management Server) を使用していると、JAWS による IBM SPSS Modeler などの Java ベース・アプリケーションの読み上げが正常に行われられない可能性があることがわかりました。SMS を無効化するとこの問題を解決できます。SMS の詳細は、Microsoft の Web サイトを参照してください。

## JAWS および Java

JAWS のバージョンによって、Java ベースのソフトウェア・アプリケーションのサポート・レベルも異なります。IBM SPSS Modeler は最近のバージョンのすべての JAWS で利用できますが、バージョンによっては Java ベースのシステムで使用する際に小さな問題が発生する場合があります。詳細は、JAWS for Windows の Web サイト <http://www.FreedomScientific.com> を参照してください。

## IBM SPSS Modeler でのグラフの使用

ヒストグラム、評価グラフ、線グラフ、および散布図のような、視覚的に表示された情報を、画面読み上げソフトウェアで確認することは困難です。ただし、Web グラフや棒グラフの場合は、出力ウィンドウで利用できるテキスト要約情報を参照することができます。

---

## 第 15 章 Unicode のサポート

---

### IBM SPSS Modeler 内での Unicode のサポート

IBM SPSS Modeler は、IBM SPSS Modeler と IBM SPSS Modeler Server の両方で、Unicode を完全にサポートしています。これにより、多言語データベースなどの Unicode をサポートする他のアプリケーションと、ロケール固有のコード化スキーム間で変換を行ったときに生じる情報の欠落なしに、データの交換が可能になります。

- IBM SPSS Modeler は、内部でデータを Unicode で保存しており、データベースに Unicode で保存されている多言語データを、情報の失うことなしに読み取ったり、書き込んだりすることができます。
- IBM SPSS Modeler は、UTF-8 エンコード・テキスト・ファイルを読み書きできます。テキスト・ファイルのインポートとエクスポートのデフォルトは、ロケールのエンコードになっていますが、それ以外に UTF-8 もサポートされています。この設定は、ファイル・インポートおよびエクスポート・ノードで指定できます。また、「ストリームのプロパティ」ダイアログ・ボックスで、デフォルトのエンコードを変更できます。詳しくは、トピック 44 ページの『ストリームの一般的なオプションの設定』を参照してください。
- ロケールのエンコードで保存されている Statistics、SAS、およびテキスト・データ・ファイルは、インポート時に UTF-8 に変換され、エクスポート時に元に戻されます。任意のファイルに書き込む場合、そのロケール文字セットに対応する文字が存在しない Unicode 文字がある場合、別の文字で置換され、また、警告が表示されます。この問題は、データが、Unicode をサポートするデータ・ソース (データベースまたは UTF-8 テキスト・ファイル) からインポートされ、さらに、そのデータにロケールが異なる文字や、複数のロケールまたは文字セットからの文字が含まれている場合についてのみ発生します。
- IBM SPSS Modeler Solution Publisher のイメージは、UTF-8 でエンコードされており、プラットフォームおよびロケール間で完全に移植可能です。

#### Unicode について

Unicode 規格の目標は、多言語テキストの一貫性のあるエンコード法を提供することです。これにより、国、地域、およびアプリケーションを越えた共有が容易になります。Unicode 規格は現在、バージョン 4.0.1 ですが、世界中で一般的に使用されているすべての文字セットのスーパーセットの文字セットを定義しており、それぞれの文字に一意的な名前とコード ポイントを割り当てています。文字とそのコード ポイントは、ISO-10646 により定義された Universal Character Set (UCS) のものと同一です。詳細は、Unicode ホームページを参照してください。



---

## 特記事項

本情報は全世界で提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21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

以下の保証は、国または地域の法律に沿わない場合は、適用されません。IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

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。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、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。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしく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保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Software Group

ATTN: Licensing

200 W. Madison St.

Chicago, IL; 60606

U.S.A.

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

この文書に含まれるいかなるパフォーマンス・データも、管理環境下で決定されたものです。そのため、他の操作環境で得られた結果は、異なる可能性があります。一部の測定が、開発レベルのシステムで行われた可能性があります。その測定値が、一般に利用可能なシステムのものと同じである保証はありません。さらに、一部の測定値が、推定値である可能性があります。実際の結果は、異なる可能性があります。お客様は、お客様の特定の環境に適したデータを確かめる必要があります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、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者をお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

本書には、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。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーの図表は表示されない場合があります。

---

## 商標

IBM、IBM ロゴおよび [ibm.com](http://ibm.com) は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子会社の米国およびその他の国における商標または登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

UNIX は The Open Group の米国およびその他の国における登録商標です。

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連会社の米国およびその他の国における商標または登録商標です。

他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。

# 索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

## [ア行]

アイコン  
オプションの設定 22, 49  
アクセシビリティ 169, 180  
例 177  
IBM SPSS Modeler の機能 169  
IBM SPSS Modeler のヒント 179  
値  
データ検査から表示 85  
CLEM 式への追加 85  
アプリケーション 27  
アプリケーションの例 3  
一時ディレクトリー 14  
色  
設定 156  
印刷 24  
ストリーム 22, 40  
因子 179  
「インタラクティブ・ツリー」ウィンドウ  
アクセシビリティ 179  
インポート  
PMML 142, 143  
エクスポート  
ストリームの説明 57  
PMML 142, 143  
エラー・メッセージ 50  
エンコード 44, 181  
演算子  
文字列の結合 97  
CLEM 式内 83  
演算子の優先順位 92  
オートメーション 75  
オブジェクト  
プロパティ 149  
オブジェクトのプロパティ、IBM SPSS  
Collaboration and Deployment Services  
Repository 133  
オプション 153  
構文 157  
ストリームのプロパティ 44, 46,  
47, 48, 49, 50  
ユーザー 154  
display 156  
IBM SPSS Modeler の 153

オプション (続き)  
PMML 157

## [カ行]

カイ 2 乗分布  
確率関数 101  
概要 89  
IBM SPSS Modeler 11, 153  
科学的表記法  
表示書式 46  
確率関数 101  
カスタム・パレットの作成 159  
サブパレットの作成 160  
画面読み上げソフトウェア 171, 174,  
178, 179  
例 177  
関数 91, 92, 95, 96, 112  
欠損値の処理 73  
データベース 83  
ユーザー定義関数 (UDF) 83  
例 75  
CLEM 式内 83  
@BLANK 73  
@FIELD 83, 119  
@GLOBAL\_MAX 117  
@GLOBAL\_MEAN 117  
@GLOBAL\_MIN 117  
@GLOBAL\_SDEV 117  
@GLOBAL\_SUM 117  
@PARTITION 119  
@PREDICTED 83, 119  
@TARGET 83, 119  
カンマ 44  
キーボード・ショートカット 171, 174,  
175  
キーワード  
ノードの注釈設定 63  
起動ダイアログ・ボックス 156  
キャッシュ  
キャッシュの設定 40  
使用可能化 153  
取り消し 41, 44  
保存 41  
キャッシュ・ファイル・ノード  
ロード 65  
切り取り 20  
空白 72, 80  
空白の処理  
CLEM 関数 118

空白文字  
文字列から削除する 79, 103  
クライアント  
デフォルト・ディレクトリー 154  
クラス 19, 145, 146  
グラフ  
出力の保存 65  
プロジェクトへの追加 146  
グループ化記号  
数字の表示形式 44  
グローバル値  
CLEM 式内 84  
グローバル関数 117  
警告 50  
オプションの設定 155  
欠損値 72, 80  
レコード 72  
CLEM 式 73  
欠損値検査ノード  
欠損値 72  
言語  
オプション 153  
検索  
ストリーム内でのノード 54  
コマンド・ライン  
IBM SPSS Modeler の開始 12  
コメント  
キーボード・ショートカット 175  
ストリームのすべてを一覧表示 62  
ノードとストリームの 58

## [サ行]

サーバー  
サーバーの COP の検索 14  
接続の追加 13  
デフォルト・ディレクトリー 154  
ログイン 12  
最小化 21  
サイズの変更 21  
サブパレット  
作成 160  
パレット上の表示 160  
パレット・タブからの削除 160  
三角関数 101  
シーケンス関数 112  
時間関数 91, 92  
time\_before 97, 108  
time\_hours\_difference 108  
time\_in\_hours 108  
time\_in\_mins 108

- 時間関数 (続き)
    - time\_in\_secs 108
    - time\_mins\_difference 108
    - time\_secs\_difference 108
  - 時間と日付の関数 91, 92
  - 時間の形式 46, 91, 92
  - 時間フィールド
    - 変換 112
  - 式 89
  - 式ビルダー
    - アクセス 83
    - 概説 83
    - 使用 83
  - 次数
    - 測定単位 46
  - 指数関数 100
  - システム
    - オプション 153
  - 実行時間、表示 50
  - 実行を中止 20
  - 実数 89, 90
  - シナリオ 135
    - 展開オプション 136
    - defined 122
  - シナリオの展開 136
  - 出力オブジェクト
    - IBM SPSS Collaboration and Deployment Services Repository での格納 128
  - 出力ノード 38
  - 出力ファイル
    - 保存 65
  - ショートカット
    - 一般的な使用法 68
    - キーボード 23, 171, 174, 175
  - 条件 78
  - 小数桁数
    - 表示形式 46
  - 小数点記号
    - 数字の表示形式 44
  - 情報関数 96
  - 新機能 7
  - シングル・サインオン 12
  - シングル・サインオン、IBM SPSS Collaboration and Deployment Services Repository 121, 123
  - ズーム 20
  - 数字の表示形式 46
  - 数値関数 100
  - スクリプト 25, 75
  - スクリプトの色
    - 設定 157
  - スコアリング
    - ブランチ 57, 138, 140
  - ステート
    - 保存 64
  - ステート (続き)
    - ロード 65
  - ストリーム 11
    - オプション 44, 46, 47, 48, 49
    - 構築 37
    - コメントの追加 58
    - 作成 37
    - 実行 57
    - 実行時間を表示 50
    - 注釈 58, 63
    - 展開オプション 136
    - 名前変更 55, 63
    - ノードの迂回 39
    - ノードの接続 38
    - ノードの追加 38, 40
    - ノードの無効化 40
    - バックアップ・ファイル 64
    - ビューへの調整 22
    - プロジェクトへの追加 146, 147
    - 保存 64
    - ロード 65
    - IBM SPSS Collaboration and Deployment Services Repository での格納 126
    - ストリームの実行 57
    - ストリームのすべてのコメントを一覧表示 62
    - ストリームの説明 55, 57
    - ストリームのデフォルトの文字コード 44
    - ストリームのビューへの調整 22
    - ストリームのプロパティ
      - Analytic Server 49
    - ストリーム名 63
    - ストリーム領域
      - 設定 49
    - ストリーム・パラメーター 52
    - ストリーム・ファイルのバックアップ
      - 復元 64
    - ストリーム■ 16
  - 図表
    - 出力の保存 65
  - 正規分布
    - 確率関数 101
  - 整数 89
  - 生成されたモデル・パレット 18
  - セッション・パラメーター 52
  - 接続
    - サーバー・クラスター (server cluster) 14
    - IBM SPSS Collaboration and Deployment Services Repository への 123
    - IBM SPSS Modeler Server への 12, 13, 14
  - 接続の COP の検索 14
  - セット 44
  - 線型回帰
    - PMML としてのエクスポート 157
  - ソース・ノード 38
  - ソート・ノード
    - パフォーマンス 165
  - 属性 (attribute) 27
- ## [タ行]
- ターミナル・ノード 38
  - 置換ノード
    - 欠損値 73
  - 知識の発見 27
  - 注釈
    - コメントへの変換 62
    - ストリーム 58, 63
    - ノード 58, 63
    - プロジェクト 149
    - folder 149
  - 重複レコード・ノード
    - パフォーマンス 165
  - ツールチップ
    - ノードの注釈設定 63
  - ツールバー 20
  - 追加
    - プロジェクトに 146
  - 通貨の表示形式 46
  - 通知
    - オプションの設定 155
  - ツリー・ベースの分析
    - 典型的なアプリケーション 27
  - データ
    - preview 43
  - データ型ノード
    - 欠損値 73
    - パフォーマンス 165
  - データ監査ノード
    - 探索に使用 27
    - データ・マイニングでの使用 28
  - データのマッピング 68
  - データ分割ノード
    - パフォーマンス 165
  - データベース
    - 関数 83
  - データ・ストリーム
    - 構築 37
    - 作成 37
  - データ・タイプ 77
    - パラメーター 52
  - データ・マイニング 27
    - アプリケーションの例 36
    - 戦略 29
  - データ・マイニングの用途 28
  - データ・マッピング・ツール 66, 67
  - テーブル
    - プロジェクトへの追加 146

- ディクショナリー・ファイル 178
- ディレクトリ
  - デフォルト 154
- テキストの文字コード 44
- テキスト・データ・ファイル
  - エンコード 181
- デシジョン・ツリー
  - アクセシビリティ 179
- デフォルト
  - プロジェクト・フェーズ 146
- 展開 122
- 展開オプション
  - シナリオ 136
- 展開タイプ 136
- 典型的なアプリケーション 27
- テンプレート 67
- テンプレート・フィールド 68
- 等号演算子 97
- 特殊関数 119
- 特殊文字
  - 文字列から削除する 79
- ドメイン名 (Windows)
  - IBM SPSS Modeler Server 12

## [ナ行]

- ナゲット 57
  - defined 18
- ナビゲート
  - キーボード・ショートカット 171
- 名前変更
  - ストリーム 55
  - ノード 63
- 日をロールオーバー 46
- ニューラル・ネットワーク・ノード
  - ラージ・セット 44
- ニューラル・ノード
  - パフォーマンス 166
- 入力ノード
  - 更新 44
  - データのマッピング 67
- ヌル 80
- ノード 11
  - オプションの設定 40
    - 概要 38
  - カスタム・サブパレットの作成 160
  - カスタム・パレットの作成 159
  - 検索 54
  - コメントを追加 58
  - 削除 38
  - 実行時間 50
  - 順序 163
  - 使用可能化 40
  - 使用不可化 40
  - ストリーム内の迂回 39
  - ストリーム内の接続 38

- ノード (続き)
  - ストリーム内の無効化 40
  - 接続の削除 40
  - 注釈 58, 63
  - 重複 40
  - 追加 38, 40
  - データのプレビュー 43
  - データ・プレビュー 43
  - パフォーマンス 165, 166
  - パレット上の表示 159
  - パレットからの削除 159
  - パレット・タブのカスタマイズ 161
  - プロジェクトへの追加 146, 147
  - 編集 40
  - 保存 64
  - ロード 65
  - ロック 43
- IBM SPSS Collaboration and Deployment Services Repository での格納 127
- ノードとストリームに名前をつける 63
- ノードの無効化 40
- ノードの有効化 40
- ノードのロック 43
- ノード名 63
- ノード・パレットの選択 159
- ノイズを含んだデータ 28

## [ハ行]

- バージョン・ラベル、IBM SPSS Collaboration and Deployment Services Repository オブジェクト 135
- パフォーマンス
  - プロセス・ノード 165
  - モデル作成ノード 166
  - CLEM 式 167
- パラメーター
  - シナリオでの使用 138
  - スコアリング 138
  - ストリーム■ 52
  - タイプ 52
  - モデル構築 138
  - ランタイム・プロパティ 52
  - CLEM 式内 84
  - session 52
- 貼り付け 20
- パレット 16
  - カスタマイズ 158
- パレット・タブのカスタマイズ 161
- 比較関数 97
- ビジュアル・プログラミング 15
- 日付
  - 操作 112
  - 変換 112
- 日付関数 91, 92

- 日付関数 (続き)
  - date\_before 97, 108
  - date\_days\_difference 108
  - date\_in\_days 108
  - date\_in\_months 108
  - date\_in\_weeks 108
  - date\_in\_years 108
  - date\_months\_difference 108
  - date\_weeks\_difference 108
  - date\_years\_difference 108
  - @TODAY 関数 108
- 日付と時間関数
  - datetime\_date 108
  - datetime\_day 108
  - datetime\_day\_name 108
  - datetime\_day\_short\_name 108
  - datetime\_hour 108
  - datetime\_in\_seconds 108
  - datetime\_minute 108
  - datetime\_month 108
  - datetime\_month\_name 108
  - datetime\_month\_short\_name 108
  - datetime\_now datetime\_second 108
  - datetime\_time 108
  - datetime\_timestamp 108
  - datetime\_weekday 108
  - datetime\_year 108
- 日付の書式 46, 91, 92
- 日付/時刻の値 80
- 必須フィールド 66, 68
- ビット単位関数 102
- 非等号演算子 97
- 表 174
  - 出力の保存 65
- 評価ノード
  - パフォーマンス 165
- 表記方法 95
- 表示形式
  - グループ化記号 46
  - 小数桁数 46
  - currency 46
  - numbers 46
  - scientific 46
- 開く
  - ステート 65
  - ストリーム 65
  - ノード 65
  - プロジェクト 146
  - モデル 65
  - output 65
- ピリオド 44
- ヒント
  - アクセシビリティ 179
    - 一般的な使用法 68
- フィールド 27, 89, 91
  - 値の表示 85

フィールド (続き)  
 CLEM 式内 84  
 フィールド選択ノード  
 欠損値 72  
 フィールドの関連付けの解除 66  
 フィールドのマッピング 66  
 フォルダー、IBM SPSS Collaboration and  
 Deployment Services Repository 131,  
 133  
 フォント 156, 157  
 複数回答セット  
 CLEM 式内 82, 84  
 複数カテゴリ設定  
 CLEM 式内 82  
 複数二分法設定  
 CLEM 式内 82  
 複数の IBM SPSS Modeler セッション  
 15  
 ブランチ、モデル作成およびスコアリング  
 57, 138, 140  
 プロジェクト 19, 145  
 オブジェクトの追加 147  
 オブジェクトのプロパティ 149  
 クラス・ビュー 146  
 作成 146  
 新規作成 146  
 注釈 149  
 デフォルト・フォルダーの設定 146  
 閉じる 150  
 フォルダーのプロパティ 149  
 プロパティの設定 148  
 レポートの生成 150  
 CRISP-DM ビュー 145  
 IBM SPSS Collaboration and  
 Deployment Services Repository での  
 148  
 IBM SPSS Collaboration and  
 Deployment Services Repository での  
 格納 127  
 プロセス・ノード 38  
 パフォーマンス 165  
 プロパティ  
 データ・ストリーム 44  
 プロジェクト・フォルダー 149  
 レポート・フェーズ 150  
 文書 3  
 分布関数 101  
 変換関数 97  
 変数 27  
 ポート番号  
 IBM SPSS Modeler Server 12, 13  
 ホスト名 (host name)  
 IBM SPSS Modeler Server 12, 13  
 保存  
 出力オブジェクト 65  
 ステート 64

保存 (続き)  
 ストリーム 64  
 ノード 64  
 複数のオブジェクト 64  
 ホット・キー 23

## [マ行]

マウス  
 IBM SPSS Modeler での使用 23, 38  
 マウスの中央ボタン  
 シミュレート 23, 38  
 マシン学習 27  
 メイン・ウィンドウ 16  
 メッセージ  
 生成された SQL の表示 48  
 メモリー  
 管理 153  
 文字 89, 90  
 文字列 89, 90  
 一致 79  
 置換 79  
 CLEM 式内の操作 79  
 文字列関数 103  
 文字列の連結 97  
 モデル 57  
 エクスポート 157  
 更新 139  
 置換 155  
 ブランチ 57  
 プロジェクトへの追加 146  
 IBM SPSS Collaboration and  
 Deployment Services Repository での  
 格納 128  
 「モデル」パレット 128  
 モデル作成ノード 38  
 パフォーマンス 166  
 モデル作成パレット・タブのカスタマ  
 イズ 161  
 モデルの置換 155  
 モデルのリフレッシュ 139  
 モデル・ナゲット 57  
 モデル・リフレッシュ 136  
 元に戻す 20

## [ヤ行]

ユーザー ID  
 IBM SPSS Modeler Server 12  
 ユーザー定義関数 (UDF) 83  
 ユーザー・オプション 154  
 優先順位 92  
 ようこそダイアログ・ボックス 156  
 より小演算子 97  
 より大演算子 97

## [ラ行]

ラジアン  
 測定単位 46  
 ラベル、IBM SPSS Collaboration and  
 Deployment Services Repository オブジェ  
 クト 135  
 ランタイム・プロパティ 52  
 リスト 89, 90  
 領域 16  
 ルール・セット  
 評価 44  
 ルール・ノードの構築  
 ロード 65  
 例  
 アプリケーション ガイド 3  
 概説 5  
 レコード 27  
 欠損値 72  
 レコード結合ノード  
 パフォーマンス 165  
 レポート  
 出力の保存 65  
 生成 150  
 プロジェクトへの追加 146  
 プロパティの設定 150  
 ロード  
 ステート 65  
 ノード 65  
 ログ・ファイル  
 生成された SQL の表示 48  
 ロケール  
 オプション 153  
 ロジスティック回帰 179  
 PMML としてのエクスポート 157  
 論理関数 99

## [数字]

508 準拠 169

## A

abs 関数 100  
 Aggregate ノード  
 パフォーマンス 165  
 allbutfirst 関数 103  
 allbutlast 関数 103  
 alphabefore 関数 103  
 AND 演算子 99  
 arccos 関数 101  
 arccosh 関数 101  
 arcsin 関数 101  
 arcsinh 関数 101  
 arctan 関数 101  
 arctan2 関数 101

arctanh 関数 101

## C

case 27  
cdf\_chisq 関数 101  
cdf\_f 関数 101  
cdf\_normal 関数 101  
cdf\_t 関数 101  
Champion Challenger 分析 122, 136  
CLEM 83  
    概要 25, 75  
    関数 83  
    式 78, 89  
    式の検査 85  
    データ型 89, 90, 91  
    例 75  
    CLEM 式の作成 83  
    language 89  
CLEM 関数  
    空白とヌル 118  
    グローバル 117  
    欠損値 73  
    三角関数 101  
    使用可能なもののリスト 94  
    情報 96  
    数値 100  
    特殊関数 119  
    日時 108  
    比較 97  
    ビット単位 102  
    変換 97  
    ランダム 103  
    論理 99  
    probability 101  
    sequence 112  
    string 103  
CLEM 式  
    パフォーマンス 167  
CLEM 式内の円記号 90  
CLEM 式の検査 85  
Clem 式ビルダー 174  
Coordinator of Processes 14  
COP 14  
copy 20  
cos 関数 101  
cosh 関数 101  
count\_equal 関数 81, 97  
count\_greater\_than 関数 81, 97  
count\_less\_than 関数 81, 97  
count\_non\_nulls 関数 97  
count\_not\_equal 関数 81, 97  
count\_nulls 関数 73, 81, 97  
count\_substring 関数 103  
CRISP-DM 19, 145  
    プロジェクト・ビュー 145

CRISP-DM プロセス・モデル 29, 30

## D

datetime\_date 関数 97  
date\_before 関数 97  
DIFF 関数 112  
div 関数 100  
DTD 142

## E

endstring 関数 103  
Enterprise View ノード 136

## F

f 分布  
    確率関数 101  
first\_index 関数 82, 97  
first\_non\_null 関数 82, 97  
first\_non\_null\_index 関数 82, 97  
fracof 関数 100

## H

hasendstring 関数 103  
hasmidstring 関数 103  
hasstartstring 関数 103  
hassubstring 関数 103  
HTML 出力  
    スクリーン・リーダー 179

## I

IBM InfoSphere Warehouse (ISW)  
    PMML のエクスポート 143  
IBM SPSS Collaboration and Deployment Services 122  
IBM SPSS Collaboration and Deployment Services Enterprise View 122, 136  
IBM SPSS Collaboration and Deployment Services Repository 121, 122  
    オブジェクトおよびバージョンの削除 132  
    オブジェクトの格納 124  
    オブジェクトの取得 129  
    オブジェクトのプロパティ 133  
    オブジェクトのロックおよびロック解除 132  
    検索 130  
    参照 123  
    シングル・サインオン 123  
    接続 123

IBM SPSS Collaboration and Deployment Services Repository (続き)  
    フォルダー 131, 133  
    プロジェクトの転送 148

IBM SPSS Collaboration and Deployment Services Repository オブジェクトのロック 132

IBM SPSS Collaboration and Deployment Services Repository オブジェクトのロック解除 132

IBM SPSS Collaboration and Deployment Services Repository からのオブジェクトの取得 129

IBM SPSS Collaboration and Deployment Services Repository でのオブジェクトの格納 124

IBM SPSS Collaboration and Deployment Services Repository のオブジェクトの検索 130

IBM SPSS Modeler 1, 15  
    アクセスビリティ 169  
    オプション 153  
    概説 11, 153  
    コマンド・ラインからの実行 12  
    はじめに 11  
    ヒントとショートカット 68  
    文書 3

IBM SPSS Modeler Advantage 122, 135

IBM SPSS Modeler Server 2  
    ドメイン名 (Windows) 12  
    ポート番号 12, 13  
    ホスト名 (host name) 12, 13  
    ユーザー ID 12  
    password 12

IBM SPSS Modeler Server 接続の追加 13, 14

IBM SPSS Modeler Server へのログイン 12

if, then, else 関数 99

INDEX 関数 112

integer\_bitcount 関数 102

integer\_leastbit 関数 102

integer\_length 関数 102

intof 関数 100

isalphacode 関数 103

isendstring 関数 103

islowercode 関数 103

ismidstring 関数 103

isnumbercode 関数 103

isstartstring 関数 103

issubstring 関数 103

issubstring\_count 関数 103

issubstring\_lim 関数 103

isuppercode 関数 103

is\_date 関数 96

is\_datetime 関数 96

is\_integer 関数 96  
is\_number 関数 96  
is\_real 関数 96  
is\_string 関数 96  
is\_time 関数 96  
is\_timestamp 関数 96

## J

Java 180  
JAWS 169, 178, 180

## K

Kohonen ノード  
パフォーマンス 166  
ラージ・セット 44  
K-Means ノード  
パフォーマンス 166  
ラージ・セット 44

## L

labels  
表示 44  
value 142  
variable 142  
last\_index 関数 82, 97  
LAST\_NON\_BLANK 関数 112  
last\_non\_null 関数 82, 97  
last\_non\_null\_index 関数 82, 97  
length 関数 103  
locchar 関数 103  
locchar\_back 関数 103  
log 関数 100  
log10 関数 100  
lowertoupper 関数 103

## M

managers 18  
matches 関数 103  
MAX 関数 112  
max 関数 97  
max\_index 関数 82, 97  
max\_n 関数 81, 97  
MEAN 関数 112  
mean\_n 関数 81, 100  
member 関数 97  
MIN 関数 112  
min 関数 97  
min\_index 関数 82, 97  
min\_n 関数 81, 97  
mod 関数 100

## N

negate 関数 100  
NOT 演算子 99  
numbers 80, 89, 90

## O

OFFSET 関数 112  
oneof 関数 103  
OR 演算子 99  
output 18

## P

password  
IBM SPSS Modeler Server 12  
pi 関数 101  
PMML  
エクスポート・オプション 157  
モデルのインポート 142, 143  
モデルのエクスポート 142, 143  
PMML モデル  
線型回帰 157  
ロジスティック回帰 157  
PowerPoint ファイル 146  
power (指数) 関数 100  
Predictive Applications 136  
preview  
ノード・データ 43

## R

random 関数 103  
random0 関数 103  
refresh  
ソース・ノード 44  
regression 179  
rem 関数 100  
replace 関数 103  
replicate 関数 103  
round 関数 100

## S

SAS ファイル  
エンコード 181  
scrolling  
オプションの設定 49  
SDEV 関数 112  
sdev\_n 関数 81, 100  
sign 関数 100  
sin 関数 101  
SINCE 関数 112  
sinh 関数 101

skipchar 関数 103  
skipchar\_back 関数 103  
soundex 関数 108  
soundex\_difference 関数 108  
spaces  
文字列から削除する 79, 103

SQL 生成

プレビュー 48  
ログイン 48

sqrt 関数 100  
startstring 関数 103  
Statistics ファイル  
エンコード 181  
stripchar 関数 103  
strmember 関数 103  
subscrs 関数 103  
substring 関数 103  
substring\_between 関数 103  
SUM 関数 112  
sum\_n 関数 81, 100

## T

t 分布  
確率関数 101  
tan 関数 101  
tanh 関数 101  
testbit 関数 102  
THIS 関数 112  
time\_before 関数 97  
to\_date 関数 97, 108  
to\_dateline 関数 108  
to\_datetime 関数 97  
to\_integer 関数 97  
to\_number 関数 97  
to\_real 関数 97  
to\_string 関数 97  
to\_time 関数 97, 108  
to\_timestamp 関数 97, 108  
trim 関数 103  
trimend 関数 103  
trim\_start 関数 103

## U

undef 関数 118  
Unicode のサポート 181  
unicode\_char 関数 103  
unicode\_value 関数 103  
uppertolower 関数 103  
UTF-8 エンコード 44, 181

## V

values 77

value\_at 関数 82, 97

### [特殊文字]

@BLANK 関数 73, 96, 118

@DIFF 関数 112

@FIELD 関数 73, 119

@FIELDS\_BETWEEN 関数 73, 81, 119

@FIELDS\_MATCHING 関数 73, 81, 119

@INDEX 関数 112

@LAST\_NON\_BLANK 関数 112, 118

@MAX 関数 112

@MEAN 関数 112

@MIN 関数 112

@MULTI\_RESPONSE\_SET 関数 82, 119

@NULL 関数 73, 96, 118

@OFFSET 関数 112

パフォーマンスに関する考慮事項 167

@PARTITION\_FIELD 関数 119

@PREDICTED 関数 119

@SDEV 関数 112

@SINCE 関数 112

@SUM 関数 112

@TARGET 関数 119

@TESTING\_PARTITION 関数 119

@THIS 関数 112

@TODAY 関数 108

@TRAINING\_PARTITION 関数 119

@VALIDATION\_PARTITION 関数 119







Printed in Japan