IBM SPSS Decision Trees 23

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、 25 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 SPSS Statistics バージョン 23 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 Decision Trees 23

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 デシジョン・ツリーの作成

Decision Tree のプロシージャーは、ツリー・ベースの分類モデルを作成します。ケースをグループに分類したり、独立 (予測) 変数の値に基づいて従属 (ターゲット) 変数の値を予測します。このプロシージャーには、探索的および確証的な分類分析のための検証ツールが用意されています。

このプロシージャーは、以下の目的に使用できます。

セグメンテーション。特定のグループのメンバーだと考えられる人物を識別します。

層化。高リスク、中リスク、低リスクのグループなど、複数のカテゴリーの 1 つにケースを割り当てます。

予測。規則を作成し、それを使用して、誰かが債務不履行になる可能性や、自動車や家の潜在的再販価値など、将来的な事象を予測します。

データの分解と変数のスクリーニング。変数の大きなセットから、形式的なパラメトリック・モデルを構築するために使用する予測値の有用なサブセットを選択します。

相互作用の識別。特定のサブグループにのみ関連する関係を識別し、それらを形式的なパラメトリック・モデル内で指定します。

カテゴリーの結合と連続型変数の離散化。グループ予測カテゴリーと連続型変数を、情報の損失を最小限に抑えながら再コード化します。

例: ある銀行が、貸し出し申込者をそれぞれの信用リスクが適切であるかどうかに基づいて分類しようと考えています。過去の顧客に関する既知の信用格付けなど、さまざまな要因に基づいて、将来の顧客が債務不履行になりそうかどうかを予測するモデルを構築できます。

ツリー・ベースの分析には、以下のような優れた機能が用意されています。

- 高リスクまたは低リスクの等質なグループを識別できます。
- 個々のケースについて予測を行うための規則を簡単に構成できます。

データの考慮事項

データ: 従属変数および独立変数として、以下を使用できます。

- *Nominal (名義データ).* 本質的な順位を持たないカテゴリーを表す値である場合 (従業員の勤務先企業での部署など)、変数を名義変数として取り扱うことができます。名義変数の例としては、地域、郵便番号、宗教上の所属などが挙げられます。
- *Ordinal (順序データ).* 値が本質的な順位を持つカテゴリーを表す場合 (例えば、サービス満足度のレベルを「非常に不満」から「非常に満足」までの順位で評価する場合) は、変数を順序変数として扱うことができます。順序変数の例としては、満足度や信頼度を表す得点や嗜好得点などが挙げられます。
- *Scale (スケール データ).* 意味のある測定基準を持つ順序カテゴリーを値が表しており、値の間の距離の比較が可能である場合は、変数をスケール (連続型) として扱うことができます。スケール変数の例としては、年齢や、千ドル単位で表した所得が挙げられます。

度数による重み付け。重み付けが有効な場合、小数表記の重み付けは最も近い整数に丸められます。したがって、重みの値が 0.5 未満のケースには重み 0 が割り当てられ、それにより分析から除外されます。

仮定: このプロシージャーでは、すべての分析変数に適切な測定レベルが割り当てられていることが前提です。また、一部の機能では、分析に含まれる従属変数のすべての値に値ラベルが定義されていることが前提です。

- **測定レベル**。測定レベルはツリーの計算に影響するため、すべての変数に適切な測定レベルを割り当てる必要があります。デフォルトでは、数値変数がスケール変数、文字列変数が名義変数という前提であり、これは実際の測定レベルを反映していない場合があります。変数リストで各変数の隣に表示されているアイコンは、変数の型を識別します。

表 1. 測定レベル・アイコン :

アイコン	**測定レベル**
	スケール
	名義
	順序

ソース変数リストの変数を右クリックし、ポップアップ・メニューから測定レベルを選択することにより、その変数の測定レベルを一時的に変更できます。

- **値ラベル**。このプロシージャーのダイアログ・ボックス・インターフェースでは、カテゴリー (名義、順序) 従属変数のすべての非欠損値に値ラベルが定義されているか、いずれの非欠損値にも定義されていないことが前提です。一部の機能は、カテゴリー従属変数の少なくとも 2 つの非欠損値に値ラベルが定義されている場合にのみ使用できます。少なくとも 2 つの非欠損値に値ラベルが定義されている場合、値ラベルのない他の値を含むすべてのケースは分析から除外されます。

デシジョン・ツリーを取得するには

1. メニューから次の項目を選択します。

   「分析」 > **「分類」** > **「ツリー...」**
2. 従属変数を選択します。
3. 1 つ以上の独立変数を選択します。
4. 成長手法を選択します。

オプションで、以下を行うこともできます。

- ソース・リスト内の変数について、測定レベルを変更する。
- 独立変数リストの最初の変数を最初の分割変数としてモデルに適用する。
- ツリーの成長プロセスにケースが与える影響を定義する影響度変数を選択する。影響度の値が小さいケースは影響が少なく、値が大きいと影響も多くなります。影響度変数の値は、正でなければなりません。
- ツリーの検証を行う。
- ツリーの成長基準をカスタマイズする。
- ターミナル・ノード番号、予測値、および予測確率を変数として保存する。

- モデルを XML (PMML) 形式で保存する。

測定レベルが不明なフィールド

データ・セット内の 1 つまたは複数の変数 (フィールド) の測定レベルが不明な場合、測定レベルの警告が表示されます。測定レベルはこのプロシージャーの結果の計算に影響するため、すべての変数に測定レベルを定義する必要があります。

「**データをスキャン**」。アクティブ・データ・セット内のデータを読み取り、測定レベルが現在不明なすべてのフィールドにデフォルトの測定レベルを割り当てます。データ・セットが大きい場合は時間がかかることがあります。

「**手動で割り当て**」。測定レベルが不明なすべてのフィールドをリスト表示するダイアログを開きます。このダイアログを使用して、これらのフィールドに測定レベルを割り当てることができます。また、測定レベルはデータ・エディターの「変数ビュー」で割り当てることもできます。

測定レベルはこのプロシージャーにとって重要なので、すべてのフィールドに測定レベルが定義されるまで、このプロシージャーを実行するためのダイアログにアクセスすることはできません。

測定レベルの変更

1. ソース・リストで変数を右クリックします。
2. ポップアップ・メニューから測定レベルを選択します。

これにより、Decision Tree のプロシージャーで使用する測定レベルが一時的に変更されます。

成長手法

使用可能な成長手法は、以下のとおりです。

CHAID. カイ 2 乗自動反復検出。各ステップにおいて、CHAID は、従属変数と最も強い交互作用を持つ独立 (予測) 変数を選択します。各予測のカテゴリーは、従属変数に関して有意な差がない場合に統合されます。

Exhaustive CHAID. 各予測について可能なすべての分割を調べる CHAID の修正版。

CRT. 分類ツリーと回帰ツリー。CRT は、従属変数に関して可能なかぎり等質なセグメントにデータを分割します。すべてのケースの従属変数が同じ値であるターミナル・ノードは、等質な「純粋」ノードです。

QUEST. Quick, Unbiased, Efficient Statistical Tree。多くのカテゴリーを持つ予測変数を考慮に入れ、他の方式での偏りを回避する高速な手法。QUEST は、従属変数が名義変数である場合に限って指定することができます。

それぞれの成長手法には、以下のような利点と制約があります。

表 2. 成長手法の機能:

機能	CHAID*	CRT	QUEST
カイ 2 乗を基にする**	X		
独立 (予測) 変数の代理変数		X	X
ツリーの剪定		X	X
ノードの多重分割	X		
ノードの 2 分割		X	X

表 2. 成長手法の機能 (続き):

機能	CHAID*	CRT	QUEST
影響度変数	X	X	
事前確率		X	X
誤分類コスト	X	X	X
高速計算	X		X

*Exhaustive CHAID を含みます。

**QUEST では名義独立変数に対してもカイ 2 乗測度が使用されます。

カテゴリーの選択

カテゴリー (名義、順序) 従属変数では、以下のことが可能です。

- 分析の対象になるカテゴリーを制御する。
- 目的の対象カテゴリーを識別する。

カテゴリーの包含または除外

分析を従属変数の特定のカテゴリーに制限することができます。

- 従属変数の値が「除外」リストに含まれているケースは、分析に含まれません。
- 名義従属変数については、ユーザー欠損カテゴリーを分析に含めることもできます (デフォルトでは、ユーザー欠損カテゴリーは「除外」リストに表示されます)。

対象カテゴリー

選択されている (チェック・マークが付けられている) カテゴリーは、その分析で最も関心のあるカテゴリーとして扱われます。例えば、債務不履行になる可能性が高い個人を識別することに主な関心がある場合は、信用格付けが「悪い」というカテゴリーを対象カテゴリーとして選択することが考えられます。

- デフォルトの対象カテゴリーはありません。カテゴリーが選択されていない場合、分類規則のオプションの一部とゲイン関連の出力は利用できません。
- 複数のカテゴリーが選択されている場合は、対象カテゴリーごとに個別のゲイン・テーブルおよびグラフが作成されます。
- 対象カテゴリーとして 1 つ以上のカテゴリーを指定しても、ツリー・モデル、リスク推定値、誤分類の結果には影響しません。

カテゴリーと値ラベル

このダイアログ・ボックスには、従属変数に対する定義済みの値ラベルが必要です。カテゴリー従属変数のうち少なくとも 2 つの値に値ラベルが定義されている場合にのみ使用できます。

カテゴリーを包含または除外し、対象カテゴリーを選択するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、少なくとも 2 つの値ラベルが定義されているカテゴリー (名義、順序) 従属変数を選択します。
2. 「**カテゴリー**」をクリックします。

検証

検証を使用することにより、大きな母集団に対してツリー構造がどの程度一般化できるかを評価できます。交差検証と分割サンプル検証の 2 つの検証方式があります。

交差検証

交差検証では、サンプルをいくつかのサブサンプル (**群**) に分割します。その後、各サブサンプルのデータを順番に除外しながらツリー・モデルを生成します。すなわち、最初のツリーは最初のサンプル群以外のすべてのケースに基づいて、2 番目のツリーは 2 番目のサンプル群以外のすべてのケースに基づいて、というように処理されます。ツリーごとに、生成時に除外したサブサンプルにツリーを適用することにより、誤分類のリスクを推定します。

- 最大 25 個のサンプル群を指定できます。値を大きくすると、各ツリー・モデルで除外されるケースの数が少なくなります。
- 交差検証では、単一の最終的なツリー・モデルが作成されます。交差検証における最終的なツリーに対するリスク推定値は、すべてのツリーのリスクの平均として計算します。

分割サンプル検証

分割サンプル検証では、モデルは学習サンプルを使用して生成され、ホールドアウト・サンプルで検定されます。

- 学習サンプルのサイズは、合計サンプル・サイズに対するパーセント、またはサンプルを学習サンプルと検定サンプルに分割する変数によって指定できます。
- 変数を使用して学習サンプルと検証サンプルを定義する場合は、変数の値が 1 のケースが学習サンプルに割り当てられ、それ以外のすべてのケースが検証サンプルに割り当てられます。この変数は、従属変数、重み付け変数、影響度変数、または強制独立変数にすることはできません。
- 学習用と検定用の両方のサンプルの結果を表示するか、検定サンプルのみの結果を表示することができます。
- 分割サンプル検証を小さいデータ・ファイル (ケース数が少ないデータ・ファイル) に対して使用する場合は、注意が必要です。学習サンプルのサイズが小さいと、カテゴリーによってはツリーを適切に成長させるだけのケースがない場合もあるため、作成されるモデルが不十分なものになる可能性があります。

デシジョン・ツリーを検証するには

1. メインの「デシジョン・ツリー」ダイアログで、「**検証**」をクリックします。
2. 「**交差検証**」または「**分割サンプル検証**」を選択します。

注: どちらの検証方式でも、ケースはサンプル・グループに対してランダムに割り当てられます。後続の分析で完全に同じ結果が得られるようにするには、最初に分析を実行する前に (「変換」メニューの「乱数ジェネレーター」で) 乱数シードを設定し、2 回目以降の分析ではシードをその値に再設定する必要があります。

ツリーの成長基準

使用可能な成長基準は、成長手法、従属変数の測定レベル、またはこの 2 つの組み合わせによって異なる場合があります。

成長の制限

「成長の制限」タブでは、ツリー内のレベル数を制限し、親ノードおよび子ノードのケースの最小数を制御できます。

「**ツリーの最大の深さ**」。ルート・ノードの下に成長できるレベル数の最大値を制御します。「**自動**」設定は、ツリーのルート・ノードの下のレベルを CHAID 手法および Exhaustive CHAID 手法では 3 つまでに、CRT 手法および QUEST 手法では 5 つまでに制限します。

「**ケースの最小数**」。ノードのケースの最小数を制御します。この基準を満たさないノードは、分割されません。

- 最小値を大きくすると、ノード数の少ないツリーが作成されます。
- 最小値を小さくすると、ノード数の多いツリーが作成されます。

ケース数が少ないデータ・ファイルの場合、親ノードで 100、子ノードで 50 というデフォルト値のケース数が原因で、ルート・ノードの下にノードを持たないツリーが作成されることがあります。このような場合は、最小値を小さくするとよりよい結果が得られる可能性があります。

成長の制限を指定するには

1. メインの「デシジョン・ツリー」ダイアログで、「**基準**」をクリックします。
2. 「**成長の制限**」タブをクリックします。

CHAID 基準

CHAID 手法および Exhaustive CHAID 手法の場合、以下を制御できます。

「**有意レベル**」。ノード分割およびカテゴリー結合に使用する有意確率値を制御できます。どちらの基準でも、デフォルトの有意レベルは 0.05 です。

- ノード分割では、値は 0 よりも大きく、1 未満でなければなりません。値を小さくすると、ノード数の少ないツリーが作成されます。
- カテゴリー結合の場合、値は 0 よりも大きく、1 以下でなければなりません。カテゴリーが結合されないようにするには、値として 1 を指定します。スケール独立変数の場合、これは、最終的なツリーにおける変数のカテゴリーの数が、指定された区間数 (デフォルトは 10) であることを意味します。詳しくは、7 ページの『CHAID 分析のスケールの区間』のトピックを参照してください。

「**カイ 2 乗統計**」。順序の従属変数の場合、ノード分割およびカテゴリー結合を決定するカイ 2 乗は、尤度比手法を使用して計算されます。名義従属変数では、以下の手法を選択できます。

- 「**Pearson**」。この手法は、計算は速くなりますが、サンプルが小さい場合には注意して使用する必要があります。これはデフォルトの手法です。
- 「**尤度比**」。この手法は Pearson よりも堅固ですが、計算に時間がかかります。小さいサンプルの場合は、こちらの手法が推奨されます。

「**モデルの推定**」。名義従属変数または順序従属変数の場合は、以下を指定できます。

- 「**最大反復回数**」。デフォルトは 100 です。最大反復回数に達したためにツリーの成長が止まる場合は、この最大数を大きくするか、ツリーの成長を制御する他の基準を 1 つ以上変更することが考えられます。
- 「**期待されるセル度数の最小変化量**」。この値は、0 よりも大きく 1 未満でなければなりません。デフォルトは 0.05 です。値を小さくすると、ノード数の少ないツリーが作成されます。

「**Bonferroni 法を使用した有意確率値の調整**」。多重比較のときに、結合基準および分割基準の有意確率値の調整に Bonferroni 法を使用します。これはデフォルトです。

「**ノード内で結合したカテゴリーの再分割を許可**」。カテゴリー結合を明示的に回避しない限り、モデルを記述するツリーが最も単純になるように独立 (予測) 変数のカテゴリーどうしが結合されます。このオプションを選択すると、結合されたカテゴリーを再分割するとよりよい解を得られる場合、再分割されます。

CHAID 基準を指定するには

1. 「デシジョン・ツリー」ダイアログで、成長手法として「**CHAID**」または「**Exhaustive CHAID**」を選択します。
2. 「**基準**」をクリックします。
3. 「**CHAID**」タブをクリックします。

CHAID 分析のスケールの区間

CHAID 分析では、スケール独立 (予測) 変数は分析の前に常に離散的グループ (0 から 10、11 から 20、21 から 30 など) にまとめられます。グループ数の初期数および最大数は、制御することができます (ただし、最初の分割の後で連続するグループが結合されることがあります)。

- 「**固定数**」。すべてのスケール独立変数は、最初に同じ数のグループにまとめられます。デフォルトは 10 です。
- 「**カスタム**」。最初に、各スケール独立変数はその変数に対して指定された数のグループにまとめられます。

スケール独立変数の区間を指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、1 つ以上のスケール独立変数を選択します。
2. 成長手法として、「**CHAID**」または「**Exhaustive CHAID**」を選択します。
3. 「**基準**」をクリックします。
4. 「**区間**」タブをクリックします。

CRT 分析および QUEST 分析では、分割は必ず 2 分割であり、スケール独立変数と順序独立変数は同じように扱われます。したがって、スケール独立変数に対して区間の数を指定することはできません。

CRT 基準

CRT 成長手法は、ノード内の等質性を最大化しようとします。ノードがケースの等質なサブセットでない度合いは、**不純度**を示します。例えば、すべてのケースで従属変数が同じ値を持つターミナル・ノードは、「純粋」なので等質なノードであり、それ以上の分割は不要です。

不純度の測定に使用される手法、およびノードの分割に必要な最小限の不純度の減少量を選択できます。

「**不純度の測定**」。スケール従属変数には、最小二乗偏差 (LSD) が不純度の測定方法として使用されます。これは、度数による重み付けまたは影響値を考慮して調整された、ノード内分散として計算されます。

カテゴリー (名義、順序) 従属変数には、以下の不純度測定方法を選択できます。

- 「**Gini**」。従属変数の値に関して子ノードの等質性を最大化する分割を見つけます。Gini は、従属変数の各カテゴリーのメンバーシップの確率の 2 乗に基づいて計算されます。この値は、ノード内のすべてのケースが 1 つのカテゴリーに含まれるときに最小 (ゼロ) になります。これがデフォルトの測定方法です。

- 「**Twoing**」。従属変数のカテゴリーは、2 つのサブクラスにグループ化されます。2 つのグループを最適に分ける分割を見つけます。
- 「**順序測度による Twoing**」。Twoing に似ていますが、隣接するカテゴリーのみがグループ化可能である点が異なります。この測定方法は、順序の従属変数でのみ使用できます。

「**改善度の最小変化量**」。ノードの分割に必要な最小限の不純度の減少量。デフォルトは 0.0001 です。値を大きくすると、ノード数の少ないツリーが作成されます。

CRT 基準を指定するには

1. 成長手法として、「**CRT**」を選択します。
2. 「**基準**」をクリックします。
3. 「**CRT**」タブをクリックします。

QUEST 基準

QUEST 手法では、ノードを分割する際の有意レベルを指定できます。有意レベルが指定された値以下でない限り、独立変数を使用してノードを分割することはできません。この値は、0 よりも大きく 1 未満でなければなりません。デフォルトは 0.05 です。値を小さくすると、最終的なモデルから除外される独立変数が多くなります。

QUEST 基準を指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、名義独立変数を選択します。
2. 成長手法として、「**QUEST**」を選択します。
3. 「**基準**」をクリックします。
4. 「**QUEST**」タブをクリックします。

ツリーの剪定

CRT 手法および QUEST 手法では、ツリーを**剪定**することにより、モデルのオーバーフィッティングを回避できます。ツリーは停止基準を満たすまで成長し、その後、リスクに対して指定された最大差に基づいて最小のサブツリーまで自動的にトリム化されます。リスク値は、標準誤差によって表されます。デフォルトは 1 です。この値は、0 または正でなければなりません。最小リスクのサブツリーを取得するには、0 を指定します。

ツリーを剪定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、成長手法として「**CRT**」または「**QUEST**」を選択します。
2. 「**基準**」をクリックします。
3. 「**剪定**」タブをクリックします。

剪定とノードの非表示の対比

剪定されたツリーを作成すると、ツリーから剪定されたノードは最終的なツリーでは使用できなくなります。最終的なツリーでは、子ノードを対話式に選択して非表示にしたり表示したりできますが、ツリーの作成プロセスで剪定したノードを表示することはできません。詳しくは、19 ページの『第 2 章 ツリー・エディター』のトピックを参照してください。

代理変数

CRT および QUEST では、独立 (予測) 変数に**代理変数**が使用されます。変数の値が欠損しているケースでは、元の変数との関連度が高い他の独立変数が分類に使用されます。このような代替予測値を代理変数と呼びます。モデル内で使用する代理変数は、最大数を指定できます。

- デフォルトでは、代理変数の最大数は独立変数の数から 1 を引いた値です。言い換えると、各独立変数について、他のすべての独立変数を代理変数として使用可能です。
- モデルで代理変数を使用しない場合は、代理変数の数として 0 を指定します。

代理変数を指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、成長手法として「**CRT**」または「**QUEST**」を選択します。
2. 「**基準**」をクリックします。
3. 「**代理変数**」タブをクリックします。

オプション

使用可能なオプションは、成長手法、従属変数の測定レベル、従属変数の値に対して定義された値ラベルの有無によって異なります。

誤分類コスト

カテゴリー (名義、順序) 従属変数では、誤分類コストにより、分類の誤りに関連する相対ペナルティーについての情報を含めることができます。以下に例を示します。

- 信用力のある顧客に対して貸付を拒否することのコストは、後から債務不履行を起こす顧客の信用枠を拡大するコストとは異なると考えられます。
- 心臓疾患のリスクが高い人を低リスクと誤分類するコストは、低リスクの人を高リスクと誤分類するコストよりもずっと高いはずです。
- 大量のメールを、応答しそうにない人に送るコストは、かなり低いと考えられますが、応答しそうな人にメール送らないことのコストは、比較的高い (収益の逸失という意味で) と思われます。

誤分類コストと値ラベル

このダイアログ・ボックスは、カテゴリー従属変数のうち少なくとも 2 つの値に値ラベルが定義されている場合にのみ使用できます。

誤分類コストを指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、少なくとも 2 つの値ラベルが定義されているカテゴリー (名義、順序) 従属変数を選択します。
2. 「**オプション**」をクリックします。
3. 「**誤分類コスト**」タブをクリックします。
4. 「**カスタム**」をクリックします。
5. グリッドに 1 つ以上の誤分類コストを入力します。値は、負でない値である必要があります (正しい分類は対角線上に示され、常に 0 です)。

「**行列の追加**」。多くの場合、コストは対称にすること、すなわち A を B と誤分類するコストが B を A と誤分類するコストと同じになることが望まれます。以下のコントロールを使用すると、対称的なコスト行列を指定することが簡単になります。

- 「**下段の三角形の複製**」。行列の下段の三角形 (対角線の下) の値を、対応する上段の三角形のセルにコピーします。
- 「**上段の三角形の複製**」。行列の上段の三角形 (対角線の上) の値を、対応する下段の三角形のセルにコピーします。
- 「**セルの平均値を使用**」。行列のそれぞれの半分に含まれる各セルに対しては、2 つの値 (上段の三角形と下段の三角形) の平均が計算され、その平均によって両方の値が置き換えられます。例えば、A を B に誤分類するコストが 1 で、B を A に誤分類するコストが 3 の場合、このコントロールを使用するとどちらの値も平均値 (1+3)/2 = 2 に置き換えられます。

利益

カテゴリー従属変数では、収益と費用の値を従属変数のレベルに割り当てることができます。

- 利益は、収益から費用を引いて計算されます。
- 利益の値は、ゲイン・テーブルの平均利益および ROI (投資収益率) の値に影響します。ツリー・モデルの基本構造には影響しません。
- 収益と費用の値は数値である必要があり、また、グリッドに表示される従属変数のすべてのカテゴリーについて指定する必要があります。

利益と値ラベル

このダイアログ・ボックスには、従属変数に対する定義済みの値ラベルが必要です。カテゴリー従属変数のうち少なくとも 2 つの値に値ラベルが定義されている場合にのみ使用できます。

利益を指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、少なくとも 2 つの値ラベルが定義されているカテゴリー (名義、順序) 従属変数を選択します。
2. 「**オプション**」をクリックします。
3. 「**利益**」タブをクリックします。
4. 「**カスタム**」をクリックします。
5. グリッドにリストされているすべての従属変数カテゴリーに対して、収益と費用の値を入力します。

事前確率

カテゴリー従属変数を含む CRT ツリーおよび QUEST ツリーでは、グループ・メンバーシップの事前確率を指定できます。**事前確率**は、独立 (予測) 変数の値について何も情報がない時点での、従属変数の各カテゴリーの全体的な相対頻度の推定値です。事前確率の使用は、母集団全体を代表していないサンプルのデータに基づいたツリーの成長を修正するのに役立ちます。

「**学習サンプルから取得 (経験的事前確率)**」。この設定は、データ・ファイル内の従属変数の値の分布が母集団の分布を代表している場合に使用します。分割サンプル検証を使用している場合は、学習サンプル内のケースの分布が使用されます。

注: 分割サンプル検証では、ケースは学習サンプルに対してランダムに割り当てられるので、学習サンプル内でのケースの実際の分布を事前に知ることはできません。詳しくは、5 ページの『検証』のトピックを参照してください。

「**カテゴリー間で同じ**」。この設定は、従属変数のカテゴリーが母集団の均等な代表の場合に使用します。例えば、4 つのカテゴリーがある場合、各カテゴリーには約 25% のケースが含まれます。

「**カスタム**」。グリッドにリストされる従属変数の各カテゴリーに対して、負でない値を入力します。値は、カテゴリーにまたがる値の分布を表す比率、パーセント、度数などの値です。

「**誤分類コストを使用して事前確率を調整**」。カスタムの誤分類コストを定義した場合は、そのコストに基づいて事前確率を調整できます。詳しくは、9 ページの『誤分類コスト』のトピックを参照してください。

利益と値ラベル

このダイアログ・ボックスには、従属変数に対する定義済みの値ラベルが必要です。カテゴリー従属変数のうち少なくとも 2 つの値に値ラベルが定義されている場合にのみ使用できます。

事前確率を指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、少なくとも 2 つの値ラベルが定義されているカテゴリー (名義、順序) 従属変数を選択します。
2. 成長手法として、「**CRT**」または「**QUEST**」を選択します。
3. 「**オプション**」をクリックします。
4. 「**事前確率**」タブをクリックします。

スコア

順序従属変数を含む CHAID および Exhaustive CHAID では、従属変数の各カテゴリーにカスタムのスコアを割り当てることができます。スコアは、従属変数のカテゴリー間の順序および距離を定義します。スコアを使用することにより、順序の値どうしの相対距離を増減させたり、値の順序を変更することができます。

- 「**各カテゴリーに一般的な順序を使用**」。従属変数の最も低いカテゴリーにスコア 1 が割り当てられ、次に低いカテゴリーにスコア 2 が割り当てられ、以下同様に続きます。これはデフォルトです。
- 「**カスタム**」。グリッドにリストされる従属変数の各カテゴリーに対して、スコアを数値で入力します。

例

表 3. カスタム・スコア値:

値のラベル	元の値	スコア
非熟練者	1	1
熟練肉体労働者	2	4
事務職	3	4.5
専門職	4	7
管理職	5	6

- このスコアでは、*非熟練者* と *熟練肉体労働者* の間の相対距離が増え、*熟練肉体労働者* と *事務職* の間の相対距離は減ります。
- このスコアでは、*管理職* と *専門職* の順序が逆転しています。

スコアと値ラベル

このダイアログ・ボックスには、従属変数に対する定義済みの値ラベルが必要です。カテゴリー従属変数のうち少なくとも 2 つの値に値ラベルが定義されている場合にのみ使用できます。

スコアを指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、少なくとも 2 つの値ラベルが定義されている順序従属変数を選択します。
2. 成長手法として、「**CHAID**」または「**Exhaustive CHAID**」を選択します。
3. 「**オプション**」をクリックします。
4. 「**スコア**」タブをクリックします。

欠損値

「欠損値」タブでは、名義独立 (予測) 変数のユーザー欠損値の処理を制御します。

- 順序およびスケール独立変数のユーザー欠損値の処理は、成長手法によって異なります。
- 名義従属変数の処理は、「カテゴリー」ダイアログ・ボックスで指定します。詳しくは、 4 ページの『カテゴリーの選択』のトピックを参照してください。
- 順序およびスケール従属変数の場合、従属変数のシステム欠損値またはユーザー欠損値があるケースは常に除外されます。

「**欠損値として扱う**」。ユーザー欠損値は、システム欠損値と同じように扱います。システム欠損値の処理は、成長手法によって異なります。

「**有効値として扱う**」。名義独立変数のユーザー欠損値は、ツリーの成長と分類で通常の値として扱われます。

手法に応じた規則

全体ではなく一部の独立変数にシステム欠損値またはユーザー欠損値がある場合

- CHAID および Exhaustive CHAID では、独立変数のシステム欠損値とユーザー欠損値は、1 つの結合したカテゴリーとして分析に含まれます。スケールおよび順序独立変数では、アルゴリズムによってまず有効な値を使用してカテゴリーが生成されてから、欠損しているカテゴリーをそれと最も類似している (有効な) カテゴリーと結合するか、別のカテゴリーとして保持するかが決定されます。
- CRT および QUEST では、独立変数に欠損値のあるケースはツリーの成長プロセスから除外されますが、手法に代理変数が含まれている場合は代理変数を使用して分類されます。名義ユーザー欠損値が欠損として扱われる場合も、この方法で処理されます。詳しくは、 9 ページの『代理変数』のトピックを参照してください。

名義独立変数のユーザー欠損値の扱いを指定するには

1. メインの「デシジョン・ツリー」ダイアログ・ボックスで、少なくとも 1 つの名義独立変数を選択します。
2. 「**オプション**」をクリックします。
3. 「**欠損値**」タブをクリックします。

モデル情報の保存

モデルの情報を変数として作業中のデータ・ファイルに保存できます。また、モデル全体を XML (PMML) 形式で外部ファイルに保存することもできます。

「保存された変数」

「**ターミナル・ノード番号**」。各ケースが割り当てられるターミナル・ノード。値はツリー・ノード番号です。

「**予測値**」。モデルによって予測される従属変数のクラス (グループ) または値。

「**予測確率**」。モデルの予測に関連付けられた確率。従属変数のカテゴリーごとに 1 つの変数が保存されます。スケール従属変数では使用できません。

「**サンプルの割り当て (学習/検定)**。分割サンプル検証で、この変数はケースが学習または検定サンプルで使用されたかどうかを示します。学習サンプルの場合の値は 1、検定サンプルの場合の値は 0 です。分割サンプル検証を選択していない場合は使用できません。詳しくは、 5 ページの『検証』のトピックを参照してください。

「XML としてツリー・モデルをエクスポート」

ツリー・モデル全体を XML (PMML) 形式で保存できます。このモデル・ファイルを使用することにより、スコアリングのために他のデータ・ファイルにモデル情報を適用できます。

「**学習サンプル**」。指定されたファイルにモデルを書き込みます。分割サンプル検証が行われるツリーでは、これが学習サンプル用のモデルになります。

「**検定サンプル**」。指定されたファイルに検定サンプル用のモデルを書き込みます。分割サンプル検証を選択していない場合は使用できません。

出力

使用可能な出力オプションは、成長手法、従属変数の測定レベル、その他の設定によって異なります。

ツリー表示

ツリーの初期の外観を制御したり、ツリーの表示を完全に抑止することができます。

「**ツリー**」。デフォルトでは、ツリー図はビューアーに表示される出力に含まれます。このオプションを選択解除すると、出力からツリー図が除外されます。

「**表示**」。これらのオプションにより、ビューアーにおけるツリー図の初期の外観を制御します。これらの属性はすべて、生成されるツリーを編集することによって変更することもできます。

- 「**方向**」。ツリーは、ルート・ノードを最上として上から下に、左から右に、または右から左に表示できます。
- 「**ノードの内容**」。ノードでは、テーブル、グラフ、またはその両方を表示できます。カテゴリー従属変数の場合、テーブルには度数とパーセントが表示され、グラフは棒グラフが表示されます。スケール従属変数の場合、テーブルには平均、標準偏差、ケース数、予測値が表示され、グラフはヒストグラムが表示されます。
- 「**スケール**」。デフォルトでは、大きいツリーはページに収まるように自動的に縮小表示されます。カスタムのスケール・パーセントは、200% まで指定できます。
- 「**独立変数の統計**」。CHAID および Exhaustive CHAID の場合、統計には有意確率値および自由度の他に、*F* 値 (スケール従属変数用) またはカイ 2 乗値 (カテゴリー従属変数用) が含まれます。CRT の場合、改善値が示されます。QUEST の場合、スケールおよび順序独立変数に対しては *F*、有意確率値、および自由度が示され、名義独立変数に対してはカイ 2 乗、有意確率値、および自由度が示されます。
- 「**ノード定義**」。ノード定義には、各ノード分割で使用される独立変数の値が表示されます。

「**表形式のツリー**」。親ノードの番号、独立変数の統計、ノードの独立変数の値、スケール従属変数の平均と標準偏差、カテゴリー従属変数の数とパーセントなど、ツリーの各ノードの要約情報です。

ツリーの初期表示を制御するには

1. メインの「デシジョン・ツリー」ダイアログで、**「出力」**をクリックします。
2. **「ツリー」**タブをクリックします。

統計

使用できる統計テーブルは、従属変数の測定レベル、成長手法、その他の設定に応じて異なります。

「モデル」

「要約」。要約には、使用されている手法、モデルに含まれている変数、指定されていてもモデルには含まれていない変数が含まれます。

「リスク」。リスク推定値およびその標準誤差。ツリーの予測精度の指標です。

- カテゴリー従属変数の場合のリスク推定値は、事前確率と誤分類コストを調整した後で誤って分類されたケースの比率です。
- スケール従属変数の場合のリスク推定値は、ノード内の分散です。

「分類テーブル」。カテゴリー (名義、順序) 従属変数の場合、このテーブルには、正しく分類されたケースと誤って分類されたケースの数が、従属変数のカテゴリーごとに示されます。スケール従属変数では使用できません。

「コスト、事前確率、スコア、および利益の値」。カテゴリー従属変数の場合、このテーブルには、分析で使用されるコスト、事前確率、スコア、および利益の値が示されます。スケール従属変数では使用できません。

「独立変数」

「モデルに対する重要度」。CRT 成長手法の場合に、モデルにとっての重要度に応じて各独立 (予測) 変数に順位を付けます。QUEST 手法および CHAID 手法では使用できません。

「分割による代理」。CRT 成長手法および QUEST 成長手法でモデルに代理変数が含まれている場合に、ツリーの分割ごとに代理変数をリストします。CHAID 手法では使用できません。詳しくは、 9 ページの『代理変数』のトピックを参照してください。

「ノード・パフォーマンス」

「要約」。スケール従属変数の場合、このテーブルには、ノード番号、ケース数、および従属変数の平均値が示されます。利益が定義されているカテゴリー従属変数の場合、このテーブルには、ノード番号、ケース数、平均利益、および ROI (投資収益率) の値が示されます。利益が定義されていないカテゴリー従属変数では使用できません。詳しくは、 10 ページの『利益』のトピックを参照してください。

「対象カテゴリー順」。対象カテゴリーが定義されているカテゴリー従属変数の場合、このテーブルには、パーセント・ゲイン、応答数のパーセント、およびインデックスのパーセント (リフト) がノードまたはパーセンタイルのグループごとに示されます。対象カテゴリーごとに別々のテーブルが作成されます。対象カテゴリーが定義されていないスケール従属変数またはカテゴリー従属変数では使用できません。詳しくは、 4 ページの『カテゴリーの選択』のトピックを参照してください。

「行」。ノードのパフォーマンス・テーブルは、ターミナル・ノード、パーセンタイル、またはその両方で結果を表示できます。両方を選択すると、対象カテゴリーごとに 2 つずつテーブルが作成されます。パーセンタイル・テーブルには、各パーセンタイルの累積値がソート順序に基づいて表示されます。

「**パーセンタイルの増分**」。パーセンタイル・テーブルに対しては、パーセンタイルの増分を 1、2、5、10、20、または 25 から選択できます。

「**累積統計の表示**」。ターミナル・ノードのテーブルの場合に、累積結果を示す追加の列を各テーブルに表示します。

統計出力を選択するには

1. メインの「デシジョン・ツリー」ダイアログで、「**出力**」をクリックします。
2. 「**統計**」タブをクリックします。

グラフ

使用できるグラフは、従属変数の測定レベル、成長手法、その他の設定に応じて異なります。

「**モデルに対する独立変数の重要度**」。独立変数 (予測) ごとのモデルの重要度を表す棒グラフ。成長手法が CRT の場合にのみ使用できます。

「ノード・パフォーマンス」

「**ゲイン**」。ゲインは、各ノードにおける対象カテゴリー内のすべてのケースのパーセントで、(ノード対象数 *n* / 総対象数 *n*) x 100 で計算されます。ゲイン・グラフは、ゲインの累積パーセンタイルを表す折れ線グラフで、(累積パーセンタイル対象数 *n* / 総対象数 *n*) x 100 で計算されます。対象カテゴリーごとに別々の折れ線グラフが作成されます。対象カテゴリーが定義されたカテゴリー従属変数でのみ使用できます。詳しくは、 4 ページの『カテゴリーの選択』のトピックを参照してください。

ゲイン・グラフは、パーセンタイルのゲインのテーブルにあるゲインのパーセント 列と同じ値をプロットするもので、累積値もレポートします。

Index (インデックス). インデックスは、目標カテゴリーのノード応答率と、サンプル全体に対する目標カテゴリー全体の応答率との比です。インデックス・グラフは、累積パーセンタイルのインデックス値の折れ線グラフです。カテゴリー従属変数でのみ使用できます。累積パーセンタイル・インデックスは、(累積パーセンタイル応答パーセント / 総応答パーセント) x 100 で計算されます。対象カテゴリーごとに別々のグラフが作成され、対象カテゴリーを定義する必要があります。

インデックス・グラフは、パーセンタイルのゲインのテーブルにあるインデックス 列と同じ値をプロットします。

Response (応答). 指定した対象カテゴリーのノードにおけるケースの割合。応答グラフは、応答の累積パーセンタイルを表す折れ線グラフで、(累積パーセンタイル対象数 *n* / 累積パーセンタイル総数 *n*) x 100 で計算されます。対象カテゴリーが定義されたカテゴリー従属変数でのみ使用できます。

応答グラフは、パーセンタイルのゲインのテーブルにある応答 列と同じ値をプロットします。

「**平均**」。従属変数の平均値の累積パーセンタイルを表す折れ線グラフ。スケール従属変数でのみ使用できます。

「**平均利益**」。平均利益の累積を表す折れ線グラフ。利益が定義されたカテゴリー従属変数でのみ使用できます。詳しくは、 10 ページの『利益』のトピックを参照してください。

平均利益グラフは、パーセンタイルのゲイン要約のテーブルにある*利益* 列と同じ値をプロットします。

「**投資収益率 (ROI)**」。累積 ROI (投資収益率) の折れ線グラフ。ROI は、費用に対する利益の比率で計算されます。利益が定義されたカテゴリー従属変数でのみ使用できます。

ROI グラフは、パーセンタイルのゲイン要約のテーブルにある *ROI* 列と同じ値をプロットします。

「**パーセンタイルの増分**」。この設定により、すべてのパーセンタイル・グラフに関してグラフに表示されるパーセンタイルの増分を 1、2、5、10、20、または 25 に制御できます。

グラフ出力を選択するには

1. メインの「デシジョン・ツリー」ダイアログで、「**出力**」をクリックします。
2. 「**作図**」タブをクリックします。

選択規則とスコアリング規則

「規則」タブでは、選択規則や分類/予測規則をコマンド・シンタックス、SQL、または単純な (平易な英語の) テキストの形式で生成できます。これらの規則は、ビューアーで表示したり外部ファイルに保存したりすることができます。

「**シンタックス**」。ビューアーに表示される出力と外部ファイルに保存される選択規則の両方について、選択規則の形式を制御します。

- **IBM® SPSS® Statistics.** コマンド・シンタックス言語。規則は、ケースのサブセットを選択するために使用できるフィルター条件を定義するコマンドのセットとして、またはケースのスコアリングのために使用できる `COMPUTE` ステートメントとして表されます。
- 「**SQL**」。データベースからレコードを選択または抽出したり、それらのレコードに値を割り当てる標準的な SQL 規則を生成します。生成された SQL 規則には、テーブル名などのデータ・ソース情報は一切含まれません。
- 「**単純なテキスト**」。平易な英語による疑似コード。規則は、ノードごとにモデルの分類や予測を記述する「if...then」論理ステートメントのセットとして表されます。この形式の規則では、定義済みの変数ラベルと値ラベル、または変数名とデータ値を使用できます。

「**タイプ**」。IBM SPSS Statistics および SQL 規則で、生成される規則のタイプを選択規則にするかスコアリング規則にするかを制御します。

- 「**ケースに値を割り当てる**」。規則を使用して、ノードのメンバーシップ基準を満たすケースにモデルの予測を割り当てることができます。ノードのメンバーシップ基準を満たすノードごとに、それぞれ別の規則が生成されます。
- 「**ケースの選択**」。規則を使用することにより、ノードのメンバーシップ基準を満たすケースを選択できます。IBM SPSS Statistics および SQL 規則では、選択基準を満たすすべてのケースを選択する 1 つの規則が生成されます。

「**IBM SPSS Statistics および SQL 規則に代理変数を含める**」。CRT および QUEST の場合は、モデルで使用されている代理予測変数を規則に含めることができます。代理変数を含む規則は、非常に複雑になることがあります。一般的に、ツリーに関する概念的な情報を取得することだけが目的の場合は、代理変数は除外します。一部のケースに不完全な独立変数 (予測) データが含まれているときに、ツリーを模倣する規則が必要な場合は、代理変数を含めます。詳しくは、 9 ページの『代理変数』のトピックを参照してください。

「**ノード**」。生成される規則の有効範囲を制御します。有効範囲内のノードごとに、それぞれ別の規則が生成されます。

- 「**すべてのターミナル ノード**」。ターミナル・ノードごとに規則を生成します。

- 「**最適なターミナル・ノード数**」。インデックス値に基づいて、上位 *n* 個のターミナル・ノードに対して規則を生成します。この数値がツリー内のターミナル・ノードの数よりも大きい場合は、すべてのターミナル・ノードに対して規則が生成されます (下の注を参照)。
- 「**ケースの指定のパーセントまでの最適なターミナル・ノード数**」。インデックス値に基づいて、上位 *n* パーセントのケースのターミナル・ノードに対して規則を生成します (下の注を参照)。
- 「**インデックス値がカットオフ値と一致するかその値を超えるターミナル・ノード数**」。インデックス値が指定された値以上であるすべてのターミナル・ノードに対して規則を生成します。インデックス値が 100 を超えている場合、そのノードの対象カテゴリー内のケースのパーセントがルート・ノードのパーセントを超えています (下の注を参照)。
- 「**すべてのノード**」。すべてのノードに対して規則を生成します。

注 1: インデックス値に基づくノード選択は、対象カテゴリーが定義されたカテゴリー従属変数でのみ使用できます。複数の対象カテゴリーを指定している場合、対象カテゴリーごとに個別の規則のセットが生成されます。

注 2: ケースを選択するための IBM SPSS Statistics および SQL 規則 (値を割り当てるための規則でなく) では、「**すべてのノード**」および「**すべてのターミナル ノード**」を使用すると、分析で使用されるすべてのケースを選択する規則が実際には生成されます。

「**規則をファイルにエクスポート**」。規則を外部テキスト・ファイルに保存します。

最終的なツリー・モデルで選択されているノードに基づいて、選択規則やスコアリング規則を対話式に生成して保存することもできます。詳しくは、 22 ページの『ケースの選択規則とスコアリング規則』のトピックを参照してください。

注: 規則をコマンド・シンタックスの形式で他のデータ・ファイルに適用する場合、そのデータ・ファイルに含まれる変数は、最終的なモデルに含まれる独立変数と同じ名前を持ち、同じ測定基準で測定され、同じユーザー定義欠損値 (存在する場合) を持つ必要があります。

選択規則またはスコアリング規則を指定するには

1. メインの「デシジョン・ツリー」ダイアログで、「**出力**」をクリックします。
2. 「**規則**」タブをクリックします。

第 2 章 ツリー・エディター

ツリー・エディターを使用すると、以下を行うことができます。
- 選択したツリーの枝を表示したり非表示にしたりする。
- ノードの内容、ノード分割で表示される統計、およびその他の情報の表示を制御する。
- ノード、背景、枠線、グラフ、およびフォントの色を変更する。
- フォントのスタイルおよびサイズを変更する。
- ツリーの位置合わせを変更する。
- 選択したノードに基づいてさらに分析を行うために、ケースのサブセットを選択する。
- 選択したノードに基づいてケースの選択やスコアリングを行う規則を作成し、保存する。

ツリー・モデルを編集するには
1. ビューアーのウィンドウでツリー・モデルをダブルクリックします。

   または
2. 「編集」メニュー、または右クリックによるポップアップ・メニューから、次の項目を選択します。

   「内容編集」 > 「別ウィンドウ」

ノードの非表示と表示

1 つの親ノードの下にある 1 つの枝のすべての子ノードを非表示にする (省略する) には
1. 親ノードの右下隅にある小さなボックス内の負符号 (-) をクリックします。

   その枝で親ノードよりも下にあるすべてのノードが非表示になります。

   1 つの親ノードの下にある 1 つの枝のすべての子ノードを表示する (展開する) には
2. 親ノードの右下隅にある小さなボックス内の正符号 (+) をクリックします。

注: 枝の子ノードを非表示にすることは、ツリーを剪定することとは異なります。剪定されたツリーが必要な場合は、ツリーを作成する前に剪定を要求する必要があります。また、剪定された枝は最終的なツリーには含まれません。詳しくは、8 ページの『ツリーの剪定』のトピックを参照してください。

複数のノードの選択

現在選択しているノードに基づいて、ケースの選択、スコアリング規則と選択規則の生成などの操作を実行できます。複数のノードを選択するには、以下の操作を行います。
1. 選択するノードをクリックします。
2. Ctrl キーを押しながら、選択する他のノードをクリックします。

複数選択できるのは、1 つの枝の兄弟ノードと親ノードの両方またはいずれかと、別の枝の子ノードです。ただし、同じノードの枝の親ノードと子以下に対しては、複数選択を実行できません。

大きなツリーの処理

ツリー・モデルに含まれるノードや枝の数が非常に多いためにフルサイズでツリー全体を表示するのが困難または不可能なことがあります。大きなツリーを処理する場合に役立つ機能が、以下のようにいくつか用意されています。

- **ツリー・マップ**。通常のツリーよりも大幅に小さい、ツリーの簡易版であるツリー・マップを使用して、ツリー内を移動し、ノードを選択できます。詳しくは、『ツリー・マップ』のトピックを参照してください。
- **スケーリング**。ツリー表示のスケール率を変更することにより、ズームアウトおよびズームインできます。詳しくは、『ツリー表示のスケーリング』のトピックを参照してください。
- **ノードと枝の表示**。ノード内のテーブルだけ、またはグラフだけを表示したり、ノード・ラベルや独立変数情報の表示を抑止することにより、ツリーをさらにコンパクトにすることができます。詳しくは、21 ページの『ツリーに表示される情報の制御』のトピックを参照してください。

ツリー・マップ

ツリー・マップではツリーがコンパクトに簡易化されて表示され、これを使用することによってツリー内を移動したりノードを選択することができます。

ツリー・マップ・ウィンドウを使用するには

1. ツリー・エディターのメニューから次の項目を選択します。

   「**表示**」 > 「**ツリー マップ**」

- 現在選択されているノードは、ツリー・モデル・エディターとツリー・マップ・ウィンドウの両方で強調表示されます。
- ツリーのうち現在ツリー・モデル・エディターの表示領域に表示されている部分が、ツリー・マップでは赤色の長方形で示されます。表示領域に表示されるツリーのセクションを変更するには、この長方形を右クリックしてドラッグします。
- 現在ツリー・エディターの表示領域に表示されていないノードをツリー・マップ内で選択すると、選択したノードを含むように表示が移動します。
- ツリー・マップでもツリー・エディターと同じように、Ctrl キーを押しながらクリックすることによって複数のノードを選択できます。同じノードの枝の親ノードと子以下に対しては、複数選択を実行できません。

ツリー表示のスケーリング

デフォルトでは、ツリーはビューアーのウィンドウに収まるように自動的にスケールが変更されます。そのため、一部のツリーは初期状態では読みにくくなる可能性があります。事前設定されたスケールを選択するか、5 〜 200% のカスタムのスケール値を独自に入力できます。

ツリーのスケールを変更するには

1. ツールバーのドロップダウン・リストからスケールのパーセントを選択するか、カスタムのパーセント値を入力します。

   または

2. ツリー・エディターのメニューから次の項目を選択します。

   「**表示**」 > 「**スケール...**」

ツリー・モデルを作成する前にスケール値を指定することもできます。詳しくは、 13 ページの『出力』のトピックを参照してください。

ノードの要約ウィンドウ

ノードの要約ウィンドウには、選択したノードが大きく表示されます。また、要約ウィンドウを使用すると、選択したノードに基づいて選択規則またはスコアリング規則を表示、適用、保存できます。

- 要約表、グラフ、規則の間で表示を切り替えるには、ノードの要約ウィンドウの「表示」メニューを使用します。
- 表示する規則のタイプを選択するには、ノードの要約ウィンドウの「規則」メニューを使用します。詳しくは、 22 ページの『ケースの選択規則とスコアリング規則』のトピックを参照してください。
- ノードの要約ウィンドウのすべてのビューには、選択したすべてのノードの要約が結合されて反映されます。

ノードの要約ウィンドウを使用するには

1. ツリー・エディターでノードを選択します。複数のノードを選択するには、Ctrl キーを押しながらクリックします。
2. メニューから次の項目を選択します。

   「表示」 > 「要約」

ツリーに表示される情報の制御

ツリー・エディターの「オプション」メニューでは、ノードの内容、独立変数 (予測値) の名前と統計、ノードの定義、その他の表示設定を制御できます。これらの設定の多くは、ツールバーからも制御できます。

ツリーの色とテキスト・フォントの変更

ツリー内の以下の色を変更できます。

- ノードの枠線、背景、テキストの色
- 枝の色および枝のテキストの色
- ツリーの背景色
- 予測カテゴリーの強調表示色 (カテゴリー従属変数)
- ノード・グラフの色

また、ツリー内のすべてのテキストについて、活字フォント、スタイル、サイズを変更できます。

注: 個別のノードや枝について、色やフォント属性を変更することはできません。色の変更は同じ種類のすべての要素に適用され、フォントの変更 (色以外) はすべてのグラフ要素に適用されます。

色とテキストのフォント属性を変更するには

1. ツールバーを使用して、ツリー全体のフォント属性、または各種のツリー要素の色を変更します (ツールバーの各コントロールの上にマウス・カーソルを置くと、コントロールの説明がツールチップに表示されます)。

   または

2. ツリー・エディター内の任意の場所をダブルクリックして「プロパティー」ウィンドウを開くか、メニューから次の項目を選択します。

「**表示**」 > 「**プロパティー**」

3. 枠線、枝、ノードの背景、予測カテゴリー、およびツリーの背景に関しては、「**色**」タブをクリックします。
4. フォントの色と属性に関しては、「**テキスト**」タブをクリックします。
5. ノード・グラフの色に関しては、「**ノードの図表**」タブをクリックします。

ケースの選択規則とスコアリング規則

ツリー・エディターを使用することにより、以下を行うことができます。

- 選択したノードに基づいて、ケースのサブセットを選択する。詳しくは、『ケースのフィルタリング』のトピックを参照してください。
- IBM SPSS Statistics コマンド・シンタックスまたは SQL 形式で、ケースの選択規則またはスコアリング規則を生成する。詳しくは、『選択規則とスコアリング規則の保存』のトピックを参照してください。

また、Decision Tree のプロシージャーを実行してツリー・モデルを作成するときに、さまざまな条件に基づいて規則を自動的に保存することもできます。詳しくは、 16 ページの『選択規則とスコアリング規則』のトピックを参照してください。

ケースのフィルタリング

特定のノードまたはノードのグループ内のケースについて詳細を知るためには、選択したノードに基づいて、さらに分析を行うケースのサブセットを選択できます。

1. ツリー・エディターでノードを選択します。複数のノードを選択するには、Ctrl キーを押しながらクリックします。
2. メニューから次の項目を選択します。

「**規則**」 > 「**ケースのフィルタリング...**」

3. フィルター変数名を入力します。選択したノードのケースには、この変数の値として 1 が返されます。それ以外のケースには値として 0 が返され、これらのケースはフィルターの状況を変更するまで以後の分析から除外されます。
4. 「**OK**」をクリックします。

選択規則とスコアリング規則の保存

ケースの選択またはスコアリングの規則を外部ファイルに保存し、その規則を別のデータ・ソースに適用することができます。これらの規則は、ツリー・エディターで選択したノードに基づきます。

「**シンタックス**」。ビューアーに表示される出力と外部ファイルに保存される選択規則の両方について、選択規則の形式を制御します。

- **IBM SPSS Statistics.** コマンド・シンタックス言語。規則は、ケースのサブセットを選択するために使用できるフィルター条件を定義するコマンドのセットとして、またはケースのスコアリングのために使用できる `COMPUTE` ステートメントとして表されます。
- 「**SQL**」。データベースからレコードを選択/抽出したり、それらのレコードに値を割り当てる標準的な SQL 規則を生成します。生成された SQL 規則には、テーブル名などのデータ・ソース情報は一切含まれません。

「**タイプ**」。選択規則またはスコアリング規則を生成できます。

- 「**ケースの選択**」。規則を使用することにより、ノードのメンバーシップ基準を満たすケースを選択できます。IBM SPSS Statistics および SQL 規則では、選択基準を満たすすべてのケースを選択する 1 つの規則が生成されます。
- 「**ケースに値を割り当てる**」。規則を使用して、ノードのメンバーシップ基準を満たすケースにモデルの予測を割り当てることができます。ノードのメンバーシップ基準を満たすノードごとに、それぞれ別の規則が生成されます。

「**代理変数を含める**」。CRT および QUEST の場合は、モデルで使用されている代理予測変数を規則に含めることができます。代理変数を含む規則は、非常に複雑になることがあります。一般的に、ツリーに関する概念的な情報を取得することだけが目的の場合は、代理変数は除外します。一部のケースに不完全な独立変数 (予測) データが含まれているときに、ツリーを模倣する規則が必要な場合は、代理変数を含めます。詳しくは、 9 ページの『代理変数』のトピックを参照してください。

ケースの選択規則またはスコアリング規則を保存するには

1. ツリー・エディターでノードを選択します。複数のノードを選択するには、Ctrl キーを押しながらクリックします。
2. メニューから次の項目を選択します。

   「**規則**」 > 「**エクスポート...**」
3. 目的の規則のタイプを選択し、ファイル名を入力します。

注: 規則をコマンド・シンタックスの形式で他のデータ・ファイルに適用する場合、そのデータ・ファイルに含まれる変数は、最終的なモデルに含まれる独立変数と同じ名前を持ち、同じ測定基準で測定され、同じユーザー定義欠損値 (存在する場合) を持つ必要があります。

特記事項

本書は米国 IBM が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で提供し、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が、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証できません。 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にお願いします。

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、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。

著作権使用許諾:

本書には、様々なオペレーティング・プラットフォームでのプログラミング手法を例示するサンプル・アプリケーション・プログラムがソース言語で掲載されています。お客様は、サンプル・プログラムが書かれているオペレーティング・プラットフォームのアプリケーション・プログラミング・インターフェースに準拠したアプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。このサンプル・プログラムは、あらゆる条件下における完全なテストを経ていません。従って IBM は、これらのサンプル・プログラムについて信頼性、利便性もしくは機能性があることをほのめかしたり、保証することはできません。これらのサンプル・プログラムは特定物として現存するままの状態で提供されるものであり、いかなる保証も提供されません。 IBM は、お客様の当該サンプル・プログラムの使用から生ずるいかなる損害に対しても一切の責任を負いません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

© (お客様の会社名) (西暦年). このコードの一部は、IBM Corp. のサンプル・プログラムから取られています。

商標

IBM、IBM ロゴおよび 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。

Adobe、Adobe ロゴ、PostScript、PostScript ロゴは、Adobe Systems Incorporated の米国およびその他の国における登録商標または商標です。

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。

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]

[カ行]

[サ行]

[タ行]

[ナ行]

[ハ行]

[ヤ行]

[ラ行]

C

G

Q

S

T

Printed in Japan

日本アイ・ビー・エム株式会社
〒103-8510 東京都中央区日本橋箱崎町19-21