IBM SPSS Data Preparation 24

注記

本書および本書で紹介する製品をご使用になる前に、 33 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 24 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 Data Preparation 24

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 Data Preparation の概要

演算システムの処理能力の向上に伴い、それに比例して情報に対する需要も増大するため、より多くのデータ収集され、ケースの個数、変数の個数、およびデータ入力エラーの件数も増加します。これらのエラーは、データウェアハウジングの究極の目標である予測モデルの予測における問題となるため、データを「クリーン」に保つ必要があります。ただし、増大するウェアハウス格納データの量は、ケースを手動で確認する能力を遥かに超えているため、データ検証の自動プロセスを実装することが不可欠です。

Data Preparation アドオン・モジュールを使用すると、アクティブなデータ・セットの中にある異常なケースや、無効なケース、変数、およびデータ値を特定し、モデル作成のデータを準備できます。

Data Preparation のプロシージャーの使用

Data Preparation のプロシージャーの使用方法は、目的に応じて異なります。データのロード後の標準的な処理の順序は次のようになります。

- **メタデータの準備。** データ・ファイル内の変数を確認し、有効な値、ラベル、および測定レベルを決定します。使用不可でありながら誤ってコード化されることの多い変数値の組み合わせを特定します。この情報に基づいて検証規則を定義します。これは時間のかかる作業ですが、類似した属性を持つデータ・ファイルを定期的に検証する必要がある場合は、実施する価値があります。
- **データの検証。** 基本チェックを実行し、無効なケース、変数、およびデータ値を特定するために定義された検証規則に基づくチェックを実行します。無効なデータが見つかったら、原因を調べ、修正します。これには、メタデータの準備において別の手順が必要になることがあります。
- **モデルの準備。** 自動データ準備を使用して、モデル作成を向上させる元のフィールドの変換を取得します。多くの予測モデルで問題を引き起こす可能性がある潜在的な統計量の外れ値を特定します。一部の外れ値は、特定されていない無効な変数値が原因で発生します。これには、メタデータの準備において別の手順が必要になることがあります。

データ・ファイルが「クリーン」になったら、他のアドオン・モジュールからモデルを作成できます。

第 2 章 検証規則

規則は、ケースが有効かどうかを決定するために使用されます。検証規則には次の 2 種類があります。

- **単一変数規則:** 単一変数規則は、範囲外の値のチェックなど、1 つの変数に適用されるチェック一定の一連のチェックによって構成されます。単一変数規則では、有効な値は、値の範囲または許容可能な値のリストとして表現できます。
- **クロス変数規則:** クロス変数規則は 1 つの変数または変数の組み合わせに対して適用できるユーザー定義の規則です。クロス変数規則は、無効な値にフラグを立てる論理式で定義されます。

検証規則は、データ・ファイルのデータ・ディクショナリーに保存されます。これにより、規則を指定したら、その後その規則を再利用できます。

事前定義の検証規則の読み込み(L)

インストール済み環境に含まれている外部データ・ファイルから事前定義の規則を読み込むことによって、すぐ使用できる一連の検証規則を取得できます。

事前定義の検証規則を読み込むには

1. メニューから次の項目を選択します。

   データ > 検証 > 事前定義の規則を読み込み...

あるいは、「データ・プロパティーのコピー」ウィザードを使用して、任意のデータ・ファイルから規則を読み込むこともできます。

検証規則の定義

「検証規則の定義」ダイアログ・ボックスでは、単一変数検証規則およびクロス変数検証規則を作成および表示できます。

検証規則を作成および表示するには

1. メニューから次の項目を選択します。

   データ > 検証 > 規則の定義...

   このダイアログ・ボックスには、データ・ディクショナリーから読み取られた単一変数検証規則およびクロス変数検証規則が取り込まれます。規則がないときは、各自の目的に合わせて変更できる新規プレースホルダー規則が自動的に作成されます。
2. 「単一変数規則」タブと「クロス変数規則」タブで個々の規則を選択し、そのプロパティーを表示および変更します。

単一変数規則の定義

「単一変数規則」タブでは、単一変数検証規則を作成、表示、および変更することができます。

規則: このリストは、単一変数検証規則を名前順で表示し、規則を適用できる変数のタイプを表示します。このダイアログ・ボックスが開くと、データ・ディクショナリーで定義されている規則が表示されます。定

義されている規則がない場合は、「単一変数規則 1」という名前のプレースホルダー規則が表示されます。「規則」リストの下に以下のボタンが表示されます。

- **新規:** 「規則」リストの最後に新規項目を追加します。この規則が選択され、「SingleVarRule *n*」という名前が割り当てられます。この *n* は整数であるため、新しい規則の名前が単一変数規則とクロス変数規則において一意になります。
- **複製:** 選択されている規則のコピーを「規則」リストの最後に追加します。規則の名前は、単一変数規則とクロス変数規則の中で一意となるように調整されます。例えば、「SingleVarRule 1」を複製すると、最初の複製規則の名前は「SingleVarRule 1 のコピー」となり、2 番目は「SingleVarRule 1 のコピー (2)」となります。
- **削除:** 選択されている規則を削除します。

規則定義: 以下のコントロールを使用して、選択されている規則のプロパティーを表示および設定できます。

- **名前**: 規則の名前は、単一変数規則とクロス変数規則の中で一意でなければなりません。
- **型:** 規則を適用することができる変数の型です。「**数値**」、「**文字列**」、および「**日付**」のいずれかを選択します。
- **形式:** 日付変数に適用できる規則の日付形式を選択できます。
- **有効値:** 有効値は、範囲または値のリストのいずれかで指定できます。

  範囲の定義

  範囲定義コントロールを使用して、有効な範囲を指定できます。範囲外の値は、無効としてフラグが立てられます。

  範囲を指定するには、最小値と最大値のいずれかまたは両方を指定します。チェック・ボックス・コントロールを使用すると、範囲内でラベルのない値および非整数値にフラグを立てることができます。

  リストの定義

  リスト定義コントロールでは、有効な値のリストを定義できます。リストに含まれない値は、無効としてフラグが立てられます。

  グリッドにリスト値を入力します。チェック・ボックスは、許容値のリストに対して文字列データ値がチェックされるときに大文字と小文字を区別するかどうかを指定します。
- **ユーザー欠損値を許可する:** ユーザー欠損値に無効としてフラグを立てるかどうかを制御します。
- **システム欠損値を許可する:** システム欠損値に無効としてフラグを立てるかどうかを制御します。文字列型の規則には適用されません。
- **空白値を許可する:** 空白 (完全に空) の文字列値を無効として区別するかどうかを制御します。非文字列型の規則には適用されません。

クロス変数規則の定義

「クロス変数規則」タブでは、クロス変数検証規則を作成、表示、および変更することができます。

規則: このリストには、クロス変数検証規則が名前順に表示されます。ダイアログ・ボックスが開くと、「CrossVarRule 1」という名前のプレースホルダー規則が表示されます。「規則」リストの下に以下のボタンが表示されます。

- **新規:** 「規則」リストの最後に新規項目を追加します。この規則が選択され、「CrossVarRule *n*」という名前が割り当てられます。この *n* は整数であるため、単一変数規則とクロス変数規則において新規規則の名前が一意になります。
- **複製:** 選択されている規則のコピーを「規則」リストの最後に追加します。規則の名前は、単一変数規則とクロス変数規則の中で一意となるように調整されます。例えば「CrossVarRule 1」を複製すると、最初の複製規則の名前は「CrossVarRule 1 のコピー」となり、2 番目の名前は「CrossVarRule 1 のコピー (2)」となります。
- **削除:** 選択されている規則を削除します。

規則定義: 以下のコントロールを使用して、選択されている規則のプロパティーを表示および設定できます。

- **名前**: 規則の名前は、単一変数規則とクロス変数規則の中で一意でなければなりません。
- **論理式:** これは、実質的に規則の定義です。無効なケースが 1 に評価されるように式を記述してください。

式の作成

1. 式を作成するには、「式」フィールドに構成要素を貼り付けるか、直接入力します。

- 関数またはよく使用されるシステム変数を貼り付けるには、「関数グループ」リストからグループを選択し、「関数と特殊変数」リストで関数または変数をダブルクリックします (または、関数あるいは変数を選択し、「**挿入**」をクリックします)。疑問符が付いているすべてのパラメーターに値を入力します (これは関数にのみ適用されます)。「**すべて**」というラベルが付いている関数グループには、使用可能な関数とシステム変数がすべて含まれています。現在選択されている関数または変数の簡単な説明が、ダイアログ・ボックスの予約領域に表示されます。
- ストリング定数は、引用符またはアポストロフィで囲みます。
- 値に小数が含まれる場合、小数点として必ずピリオド (.) を使用してください。

第 3 章 データの検証

「データの検証」ダイアログ・ボックスでは、アクティブなデータ・セットの中にある疑わしいケースまたは無効なケース、変数、およびデータ値を特定することができます。

例: データ・アナリストが月次の顧客満足度レポートを依頼者に提供する必要があるとします。依頼者が毎月受け取るデータは、不完全な顧客 ID、範囲外の変数値、および間違って入力されることの多い変数値の組み合わせがないかどうか品質チェックを行う必要があります。「データの検証」ダイアログ・ボックスでは、分析者は、顧客を一意に特定する変数を指定したり、有効な変数の範囲に対する単一変数規則を定義したり、不可能な組み合わせを捕捉するためのクロス変数規則を定義したりすることができます。このプロシージャーは、問題のあるケースと変数のレポートを返します。さらに、このデータには毎月同じデータ要素が含まれるため、分析者は翌月新しいデータ・ファイルに規則を適用できます。

統計: このプロシージャーは、さまざまなチェックを通らない変数、ケース、およびデータ値、単一変数規則およびクロス変数規則の違反数、および分析変数の簡単な記述要約のリストを作成します。

重み付け: このプロシージャーは、重み付け変数の指定を無視し、代わりにこの変数を一般の分析変数として扱います。

データを検証するには

1. メニューから次の項目を選択します。

   「データ」 > 「検証」 > 「データの検証...」
2. 基本変数チェックまたは単一変数検証規則による検証のための分析変数を 1 つ以上選択します。

   または、次を行うことができます。
3. **「クロス変数規則」**タブをクリックし、1 つ以上のクロス変数規則を適用する。

オプションで以下の操作を実行できます。

- 重複した ID や不完全な ID がないかチェックするためのケース識別変数を 1 つ以上選択します。ケース ID 変数は、ケースごとの出力にラベルを付けるためにも使用されます。2 つ以上のケース ID 変数が指定された場合は、それらの値の組み合わせがケース識別子として扱われます。

測定レベルが不明なフィールド

データ・セットの 1 つ以上の変数 (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。測定レベルはこの手順の結果の計算に影響を与えるため、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: アクティブなデータ・セットのデータを読み込み、測定レベルが現在不明なフィールドに対し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ります。

手動で割り当て: 不明な測定レベルのフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、測定レベルをこれらのフィールドに割り当てることができます。データ・エディターの「変数ビュー」でも、測定レベルを割り当てることができます。

測定レベルがこの手順で重要であるため、すべてのフィールドに測定レベルが定義されるまで、ダイアログにアクセスしてこの手順を実行することはできません。

「データの検証」の「基本チェック」

「基本チェック」タブでは、分析変数、ケース識別子、およびケース全体の基本チェックを選択できます。

分析変数: 「変数」タブで分析変数を選択した場合、以下の有効性のチェックを選択できます。チェック・ボックスを使用して、チェックをオンまたはオフにできます。

- **欠損値の最大パーセント:** 欠損値の割合 (パーセント) が指定された値より大きい分析変数を報告します。指定する値は、100 以下の正数である必要があります。
- **1 つのカテゴリーのケースの最大パーセント:** カテゴリー型の分析変数がある場合、このオプションにより、1 つの欠損していないカテゴリーを表すケースの割合 (パーセント) が、指定された値より大きいカテゴリー分析変数が報告されます。指定する値は、100 以下の正数である必要があります。パーセントは、変数の非欠損値を持つケースに基づきます。
- **度数が 1 のカテゴリーのケースの最大パーセント:** カテゴリー型の分析変数がある場合、このオプションにより、ケースを 1 つだけ含む変数のカテゴリーの割合 (パーセント) が、指定された値より大きいカテゴリー分析変数が報告されます。指定する値は、100 以下の正数である必要があります。
- **最小変動係数:** スケール型の分析変数がある場合、このオプションにより、変動係数の絶対値が指定された値より小さいスケール分析変数が報告されます。このオプションは、平均値がゼロではない変数に対してのみ適用されます。指定する値は、負でない数値であることが必要です。0 を指定すると、変動係数のチェックがオフになります。
- **最小標準偏差:** スケール型の分析変数がある場合、このオプションにより、標準偏差が指定された値より小さいスケール分析変数が報告されます。指定する値は、負でない数値であることが必要です。0 を指定すると、標準偏差のチェックがオフになります。

ケース識別子: 「変数」タブでケース識別子変数を選択した場合、以下の有効性のチェックを選択できます。

- **不完全な ID をチェックする:** このオプションにより、ケース識別子が不完全なケースが報告されます。ある 1 つのケースで ID 変数が空または欠落している場合、その識別子は不完全と見なされます。
- **重複した ID をチェックする:** このオプションにより、ケース識別子が重複したケースが報告されます。不完全な識別子は重複している可能性のある値のグループから除外されます。

空のケースをチェックする: このオプションは、すべての変数が空または空白であるケースを報告します。空のケースを特定するために、ファイル内のすべての変数 (ID 変数を除く) を使用するか、または「変数」タブに定義された分析変数のみを使用することができます。

「データの検証」の「単一変数規則」

「単一変数規則」タブでは、使用可能な単一変数検証規則が表示され、それらの規則を分析変数に適用できます。追加の単一変数規則を定義するには、「**規則の定義**」をクリックします。詳しくは、 3 ページの『単一変数規則の定義』を参照してください。

分析変数: このリストは、分析変数を表示し、それらの分布を要約し、各変数に適用された規則の数を表示します。ユーザー欠損値とシステム欠損値が要約に含まれないことに注意してください。「表示」ドロップダウン・リストは、どの変数を表示するかを制御します。「**すべての変数**」、「**数値型変数**」、「**文字型変数**」、および「**日付変数**」のいずれかを選択できます。

規則: 分析変数に規則を適用するには、1 つ以上の変数を選択し、「規則」リストで適用するすべての規則にチェック・マークを付けます。「規則」リストには、選択された分析変数に対して適切な規則のみが表示

されます。例えば、数値型の分析変数が選択されている場合は数値規則のみが表示され、文字列変数が選択されている場合は文字列規則のみが表示されます。分析変数が選択されていないか、またはデータ型が混在している場合、規則は表示されません。

変数の分布: 「分析変数」リストに表示されている分布の要約は、すべてのケースを基にするか、最初の *n* 個 (「ケース」テキスト・ボックスに指定) のケースのスキャンを基にすることができます。「**再スキャン**」をクリックすると、分布の要約が更新されます。

「データの検証」の「クロス変数規則」

「クロス変数規則」タブでは、使用可能なクロス変数規則が表示され、それらの規則をデータに適用することができます。追加のクロス変数規則を定義するには、「**規則の定義**」をクリックします。詳しくは、4 ページの『クロス変数規則の定義』を参照してください。

「データの検証」の「出力」

ケースごとのレポート: 単一変数検証規則またはクロス変数検証規則を適用した場合、個々のケースについて検証規則違反がリストされたレポートを要求できます。

- **違反の最小数:** このオプションは、ケースをレポートに含めるために必要な規則違反の最小数を指定します。正整数を指定してください。
- **最大ケース数:** このオプションは、ケースのレポートに含まれるケースの最大数を指定します。1000 以下の正整数を指定してください。

単一変数検証規則: 単一変数検証規則を適用した場合、結果を表示するかどうか、およびどのように表示するかを選択できます。

- **分析変数ごとに違反を要約する:** このオプションは、それぞれの分析変数について、違反したすべての単一変数検証規則と、それぞれの規則に違反した値の数を表示します。また、変数ごとに単一変数規則違反の総数を報告します。
- **規則ごとに違反を要約する:** このオプションは、それぞれの単一変数検証規則について、違反した規則と、変数あたりの無効な値の数を報告します。また、すべての変数について、規則ごとの違反した値の総数を報告します。

分析変数に対する記述統計を表示する: このオプションを使用すると、分析変数の記述統計量を要求できます。カテゴリー変数ごとに度数分布表が生成されます。スケール変数に対して、平均値、標準偏差、最小値、最大値を含む要約統計量の表が生成されます。

検証規則違反のあるケースをアクティブ・データ・セットの先頭に移動: このオプションは、単一変数規則またはクロス変数規則に違反するケースを、容易に確認できるようにするために、アクティブなデータ・セットの先頭に移動します。

「データの検証」の「保存」

「保存」タブでは、規則違反を記録する変数をアクティブなデータ・セットに保存できます。

集計変数: 保存できる個々の変数です。保存する変数のチェック・ボックスをオンにします。変数のデフォルト名が入力されますが、この名前は編集できます。

- **空のケース指示変数:** 空のケースには値 1 が割り当てられます。他のすべてのケースは 0 にコード化されます。変数の値には、「基本チェック」タブで指定した範囲が反映されます。

- **重複 ID のグループ:** 同じケース識別子を持つケース (不完全な識別子を持つケースを除く) には同じグループ番号を割り当てられます。一意または不完全な識別子を持つケースは 0 にコード化されます。
- **不完全な ID 指示変数:** ケース識別子が空または不完全なケースには値 1 が割り当てられます。その他すべてのケースは 0 にコード化されます。
- **検証規則違反:** これは、ケースごとの単一変数検証規則違反とクロス変数検証規則違反の合計数です。

既存の集計変数を置き換える: データ・ファイルに保存される変数の名前が一意ではない場合、同じ名前の変数を置き換えます。

指示変数を保存する: このオプションを使用すると、検証規則違反の完全な記録を保存できます。それぞれの変数は検証規則の適用に対応しており、ケースが規則に違反している場合は値が 1 になり、違反していない場合は値が 0 になります。

第 4 章 自動データ準備

分析に向けたデータの準備は、どのプロジェクトでも最も重要なステップの 1 つであり、最も時間がかかるプロセスの 1 つです。自動データ準備 (ADP) は、この作業をユーザーに代わって行います。データ分析および修正の特定、問題のあるフィールドまたは有用でないと考えられるフィールドの除外、必要に応じた新しい属性の派生、高度なスクリーニング手法によるパフォーマンスの改善が実行されます。完全な**自動化**方式でアルゴリズムを使用できます。この場合、アルゴリズムが修正を選択して適用します。またはアルゴリズムを**対話**式に使用できます。この場合、変更を行う前にその変更内容をプレビューし、必要に応じて変更を受け入れるか拒否します。

ADP を使用すると、関連する統計の概念を事前に理解する必要なく、モデル作成のためのデータを迅速かつ容易に準備できます。モデルの作成とスコアリングがより迅速に行われます。また、ADP を使用すると、自動モデリング・プロセスの頑強性が強化されます。

注: ADP は、分析用にフィールドを準備する際に、古いフィールドの既存の値およびプロパティーを置き換えるのではなく、調整または変換を含む新しいフィールドを作成します。古いフィールドは以降の分析では使用されません。役割は「なし」に設定されます。 また、ユーザー欠損値情報はこれらの新たに作成されるフィールドには転送されません。新たに作成されたフィールドの欠損値はすべてシステム欠損値となります。

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

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

目的は? 自動データ準備では、他のアルゴリズムがモデルを作成する速度に影響し、それらのモデルの予測精度を改善するデータ準備ステップが推奨されます。これには、フィールドの変換、構築、および選択が含まれます。目標も変換することができます。データ準備プロセスで重点を置く必要があるモデル作成の優先順位を指定できます。

- **速度および精度のバランス:** このオプションでは、モデル作成アルゴリズムによるデータ処理の速度と、予測精度の両方を同等に優先するように、データを準備します。
- **速度の最適化:** このオプションでは、モデル作成アルゴリズムによるデータ処理の速度を優先するよう、データを準備します。非常に大きいデータ・セットを処理する場合、または迅速な回答を求めている場合は、このオプションを選択します。
- **精度の最適化:** このオプションでは、モデル作成アルゴリズムによって生成される予測の精度を優先するよう、データを準備します。
- **カスタム分析:** 「設定値」タブでアルゴリズムを手動で変更する場合は、このオプションを選択します。これ以降に「設定値」タブで行うオプションの変更がその他の目的のいずれかと矛盾する場合、この設定が自動的に選択されることに注意してください。

自動データ準備を取得するには

メニューから次の項目を選択します。

1. メニューから次の項目を選択します。

   「変換」 > 「モデル作成のデータ準備」 > 「自動...」

2. **「実行」**をクリックします。

オプションで以下の操作を実行できます。

- 「目的」タブで目的を指定します。
- 「フィールド」タブでフィールド割り当てを指定します。
- 「設定」タブでエキスパート設定を指定します。

インタラクティブなデータ準備を取得するには

1. メニューから次の項目を選択します。

   「変換」 > 「モデル作成のデータ準備」 > 「インタラクティブ...」

2. ダイアログ上部のツールバーにある**「分析」**をクリックします。
3. 「分析」タブをクリックし、推奨されるデータ準備ステップを確認します。
4. 問題がなければ、**「実行」**をクリックします。そうでない場合は**「分析をクリア」**をクリックし、必要な設定の変更を行い、**「分析」**をクリックします。

オプションで以下の操作を実行できます。

- 「目的」タブで目的を指定します。
- 「フィールド」タブでフィールド割り当てを指定します。
- 「設定」タブでエキスパート設定を指定します。
- **「XML の保存」**をクリックして、推奨されるデータ準備ステップを XML ファイルに保存します。

「フィールド」タブ

「フィールド」タブでは、以降の分析のために準備する必要があるフィールドを指定します。

事前定義された役割を使用: このオプションは、既存のフィールド情報を使用します。役割が「目標」の単一フィールドがある場合、そのフィールドが目標として使用されます。それ以外の場合、目標はありません。事前定義された役割が「入力」のフィールドはすべて、入力として使用されます。入力フィールドは、少なくとも 1 つ必要です。

カスタム・フィールド割り当てを使用: デフォルトのリストからフィールドを移動してフィールドの役割をオーバーライドする場合、ダイアログは自動的にこのオプションに切り替わります。カスタム・フィールドの割り当てを行う場合、次のフィールドを指定します。

- **目標 (オプション):** 目標を必要とするモデルを構築する予定の場合は、目標フィールドを選択します。これは、フィールドの役割を「目標」に設定する場合と類似しています。
- **入力:** 1 つ以上の入力フィールドを選択します。これは、フィールドの役割を「入力」に設定する場合と類似しています。

「設定」タブ

「設定」タブは、アルゴリズムによるデータの処理方法を調整するために変更できる、複数の設定値グループで構成されています。他の目的と互換性のないデフォルトの設定値を変更すると、「目的」タブが自動的に更新され、「**分析のカスタマイズ**」オプションが選択されます。

日付と時刻の準備

多くのモデル作成アルゴリズムは、日付や時刻の詳細を直接処理することはできません。以下の設定により、既存データの日付および時刻から、モデル入力として使用できる新しい期間データを取得できます。日付と時刻を含むフィールドは、日付また時刻のストレージ・タイプを使用して事前に定義しておく必要があります。元の日付および時刻フィールドは、自動データ準備に従うモデル入力としては推奨されません。

モデリング用の日付と時刻を準備: このオプションを選択解除すると、選択内容を維持した状態で「日付と時刻の準備」のその他のすべてのコントロールが無効になります。

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

- **基準日:** 入力データの日付情報に関して、計算する期間の開始日を指定します。「**今日の日付**」を選択すると、ADP の実行時には常に現在のシステム日付が使用されます。特定の日付を使用するには、「**固定日付**」を選択して必要な日付を入力します。
- **期間 (日数) の単位:** 期間の単位を ADP により自動的に決定するか、または「**固定単位**」（「年」、「月」、または「日」）から選択するかを指定します。

基準時刻からの経過時間を計算: 時刻を含む各変数の基準時刻以降の時間/分/秒の数を生成します。

- **基準時刻:** 入力データの時刻情報に関して、計算する期間の開始時刻を指定します。「**現在の時刻**」を選択すると、ADP の実行時には常に現在システム時刻が使用されます。特定の時刻を使用するには、「**固定時刻**」を選択して必要な詳細情報を入力します。
- **期間 (時間数) の単位:** 期間の単位を ADP により自動的に決定するか、または「**固定単位**」（「時間」、「分」、または「秒」）から選択するかを指定します。

サイクル時間要素を抽出: これらの設定を使用して、1 つの日付または時刻フィールドを 1 つ以上のフィールドに分割します。例えば 3 つの日付チェック・ボックスをすべて選択すると、入力日付フィールド「1954-05-23」は 3 つのフィールド (1954、5、23) に分割され、各フィールドには「**フィールド名**」パネルで定義した接尾辞が使用され、元の日付フィールドは無視されます。

- **日付から取得:** 日付入力について、年、月、日付またはそれらの組み合わせを抽出するかどうかを指定します。
- **時刻から取得:** 時刻入力について、時間、分、秒またはそれらの組み合わせを抽出するかどうかを指定します。

フィールドの除外

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

低品質の入力フィールドを除外: このオプションを選択解除すると、選択内容を維持した状態で「フィールドの除外」のその他のすべてのコントロールが無効になります。

欠損値の多いフィールドの除外: 欠損値の割合が指定されている割合を超えているフィールドが、以降の分析から除外されます。 0 以上 100 以下の値を指定します (0 はこのオプションの選択解除を示します)。ただし、値がすべて欠損値であるフィールドは自動的に除外されます。デフォルトは 50 です。

一意のカテゴリーが多すぎる名義型フィールドを除外: カテゴリー数が指定された数を超えている名義型フィールドが、以降の分析から除外されます。正整数を指定してください。デフォルトは 100 です。ID、住所、名前など、レコードごとに固有の情報を含むフィールドをモデル作成から自動的に除外する場合に役立ちます。

単一カテゴリーの値が多いカテゴリー・フィールドの除外: 1 つのカテゴリーに指定されている割合を超えるレコードが含まれている順序型フィールドおよび名義型フィールドが、以降の分析から除外されます。0 以上 100 以下の値を指定します (0 はオプションの選択解除を示します)。ただし、定数フィールドは自動的に除外されます。デフォルトは 95 です。

測定の調整

測定レベルの調整: このオプションを選択解除すると、選択内容を維持した状態で「測定の調整」のその他のコントロールが無効になります。

測定レベル: 値が「少なすぎる」連続型フィールドの測定レベルを順序型に調整できるようにするかどうか、および値が「多すぎる」順序型フィールドを連続型に調整できるかどうかを指定します。

- **>順序フィールドの値の最大数:** カテゴリーの数が指定の数を超えている順序型フィールドが、連続型フィールドとして再計算されます。正整数を指定してください。デフォルトは 10 です。この値は、連続型フィールドの値の最小数以上でなければなりません。
- **連続型フィールドの値の最小数:** 固有値の数がこの指定の数を下回る連続型フィールドが、順序型フィールドとして再計算されます。正整数を指定してください。デフォルトは 5 です。この値は、順序型フィールドの値の最大数以下でなければなりません。

データ品質の向上

データ品質向上のためにフィールドを準備: このオプションを選択解除すると、選択内容を維持した状態で、「データ品質の向上」のその他のコントロールがすべて無効になります。

外れ値の処理: 入力と目標の外れ値を置き換えるかどうかを指定します。置き換える場合、標準偏差で測定した外れ値のカットオフ基準と、外れ値を置き換える方法を指定します。外れ値を置き換えるには、トリミング (カットオフ値を設定) するか、または外れ値を欠損値として設定します。欠損値として設定した外れ値は、次のように選択された欠損値処理の設定にしたがって処理されます。

欠損値の置換: 連続型フィールド、名義型フィールド、または順序型フィールドの欠損値を置き換えるかどうかを指定します。

名義フィールドの並べ替え: これを選択すると、最小のカテゴリー (発生する頻度が最も少ない) が最初に、最大カテゴリー (発生する頻度が最も多い) が最後になるように名義型 (セット型) フィールドの値の順序を並べ替えます。新しいフィールド値は、0 (頻度が最も少ないカテゴリー) から始まります。元のフィールドが文字列型である場合でも、新しいフィールドは数値型になることに注意してください。例えば、名義型フィールドのデータ値が「A」、「A」、「A」、「B」、「C」、「C」の場合、自動データ準備は「B」を 0 に、「C」を 1 に、「A」を 2 に再割り当てします。

フィールドの尺度設定

フィールドの尺度設定: このオプションを選択解除すると、選択内容を維持した状態で、「フィールドの尺度設定」のその他のすべてのコントロールが無効になります。

分析の重み付け: この変数には、分析 (回帰またはサンプリング) の重みが含まれています。分析の重み付けを使用して、対象フィールドのレベル間の分散における差異を処理します。連続型フィールドを選択します。

連続型入力フィールド: 「**z-スコア変換**」または「**min/max 変換**」を使用して、連続型入力フィールドを正規化します。入力のスケール変更は、「選択および構築」設定で「**フィールド構築の実行**」を選択した場合に特に便利です。

- **z-スコア変換:** 観測された平均と標準偏差を母集団パラメーター推定値として使用して、フィールドが標準化され、*z* スコアが指定された**最終平均値**と**最終標準偏差**を使用して正規分布の対応する値にマップされます。「**最終平均値**」に数値を指定し、「**最終標準偏差**」に正数を指定します。標準化された再スケールに応じて、デフォルトはそれぞれ 0 および 1 となります。
- **min/max 変換:** 観測された平均と標準偏差を母集団パラメーター推定値として使用し、フィールドが**最小値**と**最大値**が指定された一様分布の対応する値にマップされます。「**最大値**」には「**最小値**」よりも大きい数値を指定してください。

連続型対象: Box-Cox 変換を使用して、連続型対象を、指定された**最終平均値**と**最終標準偏差**を持つ近似正規分布のフィールドに変換します。「**最終平均値**」に数値を指定し、「**最終標準偏差**」に正数を指定します。デフォルトはそれぞれ 0 および 1 となります。

注: 対象が ADP によって変換されている場合、変換された対象を使用して作成された後続のモデルは、変換された単位をスコアリングします。結果を解釈して使用するために、予測値を元のスケールに変換する必要があります。詳しくは、を参照してください。詳しくは、 24 ページの『スコアの後方変換』を参照してください。

フィールドの変換

データの予測精度を向上させるために、入力フィールドを変換することができます。

モデル作成にフィールドを変換: このオプションを選択解除すると、選択内容を維持した状態で「フィールドの変換」のその他のすべてのコントロールが無効になります。

カテゴリー入力フィールド: 使用可能なオプションは次のとおりです。

- **まばらなカテゴリーを結合して目標との関連性を最大化:** 目標と関連して処理するフィールドの数を減らして、より節約的なモデルを作成します。類似するカテゴリーが、入力と目標の間の関係に基づいて特定されます。大きく相違しないカテゴリー、つまり *p* 値が指定された値より大きいカテゴリーが、結合されます。 0 より大きく、1 以下の値を指定します。すべてのカテゴリーが 1 つのカテゴリーに結合されると、元のバージョンのフィールドおよび派生バージョンのフィールドは、予測値がないため、以降の分析からは除外されます。
- **目標がない場合、度数に基づいてまばらなカテゴリーを結合する:** データ・セットに目標がない場合、順序型フィールドおよび名義型フィールドのまばらなカテゴリーを結合できます。等度数法を使用して、レコード数合計の割合が指定された最小値よりも少ないカテゴリーは結合されます。0 以上 100 以下の値を指定します。デフォルトは 10 です。ケース数が指定された最小パーセントに満たないカテゴリーがない場合、または残っているカテゴリーが 2 つだけの場合、結合が停止します。

連続型入力フィールド: データ・セットにカテゴリー型目標が含まれている場合、強い関連を持つ連続型入力フィールドを分割して、処理のパフォーマンスを向上させることができます。ビンが「等質なサブセット」のプロパティーに基づいて作成され、指定した *p* 値を等質なサブセットを決定する基準値のアルファとして使用する Scheffe 手法で特定されます。0 より大きく、1 以下の値を指定します。デフォルトは 0.05 です。分割操作によって特定フィールドに単一ビンが生成される場合、予測値としての値がないため、元のバージョンのフィールドおよび分割されたフィールドは除外されます。

注: ADP の分割化は最適カテゴリー化とは異なります。最適カテゴリー化では、エントロピー情報を使用して、連続型フィールドをカテゴリー・フィールドに変換します。これには、データをソートし、メモリー内にすべて保存する必要があります。ADP では、等質サブセットを使用して連続型フィールドを分割します。つまり ADP 分割では、データをソートしてメモリー内にすべて保存する必要はありません。等質サブセット方法を使用して連続型フィールドを分割すると、分割後のカテゴリー数は、常に目標のカテゴリー数以下になります。

選択および構築

データの予測精度を向上させるために、既存のフィールドに基づいて新しいフィールドを構築できます。

フィールド選択を実行: 目標との相関の *p* 値が、指定された *p* 値よりも大きい場合、分析から連続型入力が除外されます。

フィールド構築の実行: 複数の既存フィールドの組み合わせから新しいフィールドを作成するには、このオプションを選択します。古いフィールドは、以降の分析には使用されません。このオプションは、目標が連続型である連続型入力フィールド、または目標がない連続型入力フィールドにのみ適用されます。

フィールド名

新しいフィールドや変換されたフィールドを容易に特定できるようにするために、ADP は新しい基本名、接頭辞または接尾辞を作成し、適用します。各自のニーズおよびデータとの関連性を高めるため、これらの名前を修正することができます。

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

また、「選択および構築」設定を使用して、構築されるフィールドに適用する接頭辞名を指定します。番号の接尾辞をこの接頭辞のルート名に追加して、新しい名前が作成されます。番号の形式は、作成する新しいフィールドの数によって異なります。次に例を示します。

- 構築フィールド数が 1 から 9 の場合、feature1 から feature9 となります。
- 構築フィールド数が 10 から 99 の場合、feature01 から feature99 となります。
- 構築フィールド数が 100 から 999 の場合、feature001 から feature999 となります。

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

日時から計算した期間: 日付および時刻の両方から計算された期間に適用する名前の拡張子を指定します。

日時から抽出したサイクル要素: 日付および時刻の両方から抽出したサイクル要素に適用する名前の拡張子を指定します。

変換の適用と保存

インタラクティブ・データ準備または自動データ準備のどちらのダイアログを使用しているかによって、変換の適用および保存の設定が多少異なります。

インタラクティブ・データ準備での変換の適用の設定

変換されたデータ: この設定は、変換されたデータの保存先を指定します。

- **新しいフィールドをアクティブなデータ・セットに追加:** 自動データ準備によって作成されるフィールドはすべて、新しいフィールドとしてアクティブなデータ・セットに追加されます。「**分析済みフィールドの役割を更新**」を選択すると、自動データ準備により以降の分析から除外されるすべてのフィールドの役割が「なし」に設定されます。
- **変換されたデータを含む新しいデータ・セットまたはファイルを作成:** 自動データ準備により推奨されるフィールドが、新しいデータ・セットまたはファイルに追加されます。「**分析されていないフィールドを追加**」を選択すると、「フィールド」タブで指定されていない元のデータ・セットのフィールドが、新しいデータ・セットに追加されます。ID、住所、名前など、モデル作成で使用されない情報を含むフィールドを新しいデータ・セットに転送する場合に役立ちます。

自動データ準備の適用および保存の設定

「変換されたデータ」グループは、インタラクティブ・データ準備の場合と同じです。自動データ準備では、次の追加オプションを使用できます。

変換の適用: 「自動データ準備」ダイアログでこのオプションを選択解除すると、選択内容を維持した状態で、その他の適用および保存のコントロールがすべて無効になります。

変換をシンタックスとして保存: 推奨される変換がコマンド・シンタックスとして外部ファイルに保存されます。「インタラクティブ・データ準備」ダイアログでは、「**貼り付け**」をクリックすると変換がコマンド・シンタックスとしてシンタックス・ウィンドウに貼り付けられるため、このコントロールはありません。

変換を XML として保存: 推奨される変換が XML として外部ファイルに保存されます。これは、TMS MERGE を使用してモデル PMML に結合するか、または TMS IMPORT を使用して別のデータ・セットに適用できます。「インタラクティブ・データ準備」ダイアログでは、ダイアログ上部のツールバーにある「**XML の保存**」をクリックすると変換が XML として保存されるため、このコントロールはありません。

「分析」タブ

注: 「分析」タブは「インタラクティブ・データ準備」ダイアログで使用され、ユーザーは推奨される変換を確認できます。「自動データ準備」ダイアログにはこのステップは含まれていません。

1. ADP 設定値 (「目的」、「フィールド」、および「設定値」タブで行われたすべての変更を含む) を確認したら、「**データの分析**」をクリックします。アルゴリズムにより設定値がデータ入力に適用され、その結果が「分析」タブに表示されます。

「分析」タブには、データの処理の概要を示す表形式の出力とグラフィカル出力が表示され、スコアリング用にデータを変更または改善する方法に関する推奨事項が表示されます。これらの推奨事項を確認し、承認または拒否することができます。

「分析」タブは 2 つのパネルで構成されています。左側はメイン・ビュー、右側はリンク・ビューまたは補助ビューです。以下の 3 つのメイン・ビューがあります。

- フィールド処理の要約 (デフォルト)。詳しくは、『フィールド処理の要約』を参照してください。
- フィールド。詳しくは、 19 ページの『フィールド』を参照してください。
- アクションの要約。詳しくは、 20 ページの『アクションの要約』を参照してください。

以下の 4 つのリンク/補助ビューがあります。

- 予測精度 (デフォルト)。詳しくは、 20 ページの『予測精度』を参照してください。
- フィールド・テーブル。詳しくは、 20 ページの『フィールド・テーブル』を参照してください。
- フィールドの詳細。詳しくは、 21 ページの『フィールドの詳細』を参照してください。
- アクションの詳細。詳しくは、 22 ページの『アクションの詳細』を参照してください。

ビュー間のリンク

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

ビューのリセット

元の分析に関する推奨事項を再度表示し、「分析」ビューで行った変更をすべて取り消す場合、メイン・ビュー・パネルの下部にある「**戻す**」をクリックします。

フィールド処理の要約

「フィールド処理の要約」テーブルは、フィールドの状態および作成されるフィールドの数の変更など、予測される処理による全体的な影響のスナップショットを示します。

モデルは実際には作成されていないため、データ準備の前後の全体的な予測精度の変化の測定またはグラフはありません。代わりに、推奨される各予測値の予測精度のグラフを表示できます。

このテーブルには次の情報が表示されます。

- 目標フィールドの数。
- 元の (入力) 予測値の数。
- 分析およびモデリングでの使用が推奨される予測値。これには、推奨されるフィールドの合計数、推奨される元の未変換フィールドの数、推奨される変換後のフィールドの数 (中間バージョンのフィールド、日付/時刻の予測値から派生したフィールド、および構築された予測値を除く)、推奨される日付/時刻から派生したフィールドの数、推奨される構築済み予測値の数が含まれます。
- 元の形式、派生したフィールド、または構築済み予測値への入力など、いかなる形式でも使用が推奨されない入力予測値の数。

「**フィールド**」の情報に下線が付いている場合は、クリックするとリンク・ビューに詳細情報が表示されます。「フィールド・テーブル」リンク・ビューには、「**目標**」、「**入力フィールド**」、および「**未使用の入力フィールド**」の詳細が表示されます。詳しくは、 20 ページの『フィールド・テーブル』を参照してください。「**分析での使用が推奨されるフィールド**」は、「予測精度」リンク・ビューに表示されます。詳しくは、 20 ページの『予測精度』を参照してください。

フィールド

「フィールド」メイン・ビューには、処理済みフィールドと、ADP が下流モデルでそれらのフィールドの使用を推奨するかどうかが表示されます。任意のフィールドの推奨を上書きできます。例えば、作成されたフィールドを除外するか、または ADP が除外を推奨するフィールドを含める場合などです。フィールドが変換されている場合、推奨された変換を受け入れるか、または元のバージョンを使用するかを決定できます。

「フィールド」ビューは 2 つのテーブルで構成されています。1 つは目標フィールドのテーブル、もう 1 つは処理または作成された予測値のテーブルです。

「目標」テーブル

「**目標**」テーブルは、データで目標が定義されている場合にのみ表示されます。

このテーブルには 2 つの列があります。

- **名前:** 目標フィールドの名前またはラベルです。フィールドが変換されている場合でも、元の名前が常に使用されます。
- **測定レベル:** 測定レベルを示すアイコンが表示されます。マウス・ポインターをアイコンの上に移動すると、データを説明するラベル (連続、順序、名義など) が表示されます。

  目標が変換されている場合は、「**測定の尺度**」列には最終変換バージョンが反映されます。*注:* 目標の変換をオフにすることはできません。

「予測変数」テーブル

「**予測変数**」テーブルは常に表示されます。テーブルの各行はフィールドを示します。デフォルトでは、行は予測精度の降順でソートされています。

通常のフィールドの場合、元の名前が常に行の名前として使用されます。元のバージョンおよび派生バージョンの日付/時刻フィールドがテーブルの個別の行に表示されます。また、表には構築された予測変数も含まれています。

テーブルに表示される変換されたバージョンのフィールドは、常に最終バージョンを表している点に注意してください。

デフォルトでは、推奨されたフィールドのみが「予測変数」テーブルに表示されます。その他のフィールドを表示するには、表の上にある「**テーブルに非推奨フィールドを追加する**」ボックスを選択します。テーブルの下部にこれらのフィールドが表示されます。

このテーブルには次の列があります。

- **使用バージョン:** フィールドを下流で使用するかどうか、および推奨される変換を使用するかどうかを制御するドロップダウン・リストが表示されます。デフォルトでは、ドロップダウン・リストには推奨が反映されます。

  変換された通常の予測値の場合、ドロップダウン・リストには「**変換**」、「**元のデータ**」、および「**使用しない**」の 3 つの選択項目が表示されます。

  未変換の通常の予測値の場合は、「**元のデータ**」と「**使用しない**」が表示されます。

  派生した日付/時刻フィールドおよび構築された予測値の場合、選択項目は「**変換**」と「**使用しない**」です。

元の日付フィールドの場合、ドロップダウン・リストは使用不可であり、「**使用しない**」に設定されます。

注: 元のバージョンと変換されたバージョンの両方がある予測値の場合、**オリジナル**バージョンと**変換された**バージョンの間で変更すると、これらのフィールドの「**測定の尺度**」と「**予測精度**」の設定が自動的に更新されます。

- **名前:** 各フィールドの名前はリンクになっています。名前をクリックすると、そのフィールドに関する詳細情報がリンク・ビューに表示されます。詳しくは、 21 ページの『フィールドの詳細』を参照してください。
- **測定レベル:** データ型を示すアイコンが表示されます。マウス・ポインターをアイコンの上に移動すると、データを説明するラベル (連続、順序、名義など) が表示されます。
- **予測精度:** ADP が推奨するフィールドに対してのみ予測精度が表示されます。この列は、目標が定義されていない場合には表示されません。予測精度の範囲は 0 から 1 であり、値が大きいほど予測精度が高くなります。一般に、予測精度は ADP 分析の予測を比較するときに役立ちますが、予測精度の値を分析間で比較しないでください。

アクションの要約

自動データ準備で実行された各アクションについて、入力予測値は変換および/または除外されます。アクションを通過したフィールドは、次のアクションで使用されます。最後のステップまで通過したフィールドが、モデル作成での使用を推奨されます。変換された予測値および構築された予測値への入力は除外されます。

「アクションの要約」は、ADP で実行された処理アクションをリストした単純なテーブルです。**アクション**に下線が付いている場合は、クリックすると実行したアクションに関する詳細情報がリンク・ビューに表示されます。詳しくは、 22 ページの『アクションの詳細』を参照してください。

注: 各フィールドの元のバージョンと最終変換バージョンのみが表示されます。分析実行中に使用された中間バージョンは表示されません。

予測精度

分析を初めて実行する場合、または「フィールド処理の要約」メイン・ビューで「**分析での使用が推奨される予測変数**」を選択した場合にデフォルトで表示されます。グラフには、推奨される予測値の予測精度が表示されます。フィールドは予測精度でソートされ、値が最も大きいフィールドが先頭になります。

変換されたバージョンの通常の予測値の場合、「設定値」タブの「フィールド名」パネルで選択した接尾辞がフィールド名に反映されます (例: *_transformed*)。

各フィールド名の後に、測定レベルを示すアイコンが表示されます。

推奨される各予測値の予測精度は、目標が連続型またはカテゴリー型のいずれであるかに基づいて、線形回帰または Naive Bayes モデルから算出されます。

フィールド・テーブル

「フィールド処理の要約」メイン・ビューの「**目標**」、「**予測値**」、または「**未使用の予測変数**」をクリックすると表示されます。「フィールド・テーブル」ビューには、該当するフィールドをリストした単純なテーブルが表示されます。

このテーブルには 2 つの列があります。

- **名前:** 予測値の名前。

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

  変換されたバージョンの通常の予測値の場合、「設定値」タブの「フィールド名」パネルで選択した接尾辞が名前に反映されます (例: *_transformed*)。

  日付と時刻から派生したフィールドの場合、最終変換バージョンの名前が使用されます (例: *bdate_years*)。

  構築された予測値の場合、構築された予測値の名前が使用されます (例: *Predictor1*)。
- **測定レベル:** データ型を示すアイコンが表示されます。

  目標の場合、「**測定の尺度**」には常に、変換されたバージョンが反映されます (目標が変換されている場合)。例えば、順序型 (順序セット) から連続型 (範囲、スケール) への変換またはこの逆の変更などです。

フィールドの詳細

「フィールド」メイン・ビューで「**名前**」をクリックすると表示されます。「フィールドの詳細」ビューには、選択されたフィールドの分布、欠損値、予測精度グラフ (該当する場合) が表示されます。また、フィールドの処理履歴と変換後のフィールドの名前も表示されます (該当する場合)。

各グラフ・セットについて、2 つのバージョンが並んで表示され、変換が適用されたフィールドと変換が適用されていないフィールドを比較できます。変換されたバージョンのフィールドが存在していない場合は、元のバージョンのグラフのみが表示されます。派生した日付または時刻フィールドと構築された予測値については、新しい予測値のグラフのみが表示されます。

注: カテゴリーが多すぎるためにフィールドが除外されている場合は、処理履歴のみが表示されます。

分布図

連続型フィールドの分布は、正規曲線が重なり、平均値を示す垂直基準線を使用したヒストグラムとして表示されます。カテゴリー・フィールドは棒グラフとして表示されます。

ヒストグラムには、標準偏差と歪度を示すラベルが付いています。ただし値の個数が 2 以下の場合、または元のフィールドの差異が 10 〜 20 より小さい場合、歪度は表示されません。

グラフの上にマウス・ポインターを移動すると、ヒストグラムの平均値、または棒グラフのカテゴリーのレコード合計数の度数およびパーセントが表示されます。

欠損値のグラフ

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

ADP が欠損値の処理を実行した場合、変換後の円グラフには置換値、つまり欠損値の代わりに使用される値がラベルとして表示されます。

このグラフにマウス・ポインターを移動すると、欠損値の数と全レコード数に対する割合が表示されます。

予測精度グラフ

推奨フィールドの変換前と変換後の予測精度が棒グラフに表示されます。目標が変換されている場合、変換後の目標に基づいて予測精度が計算されます。

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

マウス・ポインターをこのグラフに移動すると、予測精度値が表示されます。

処理履歴テーブル

このテーブルには、変換されたバージョンのフィールドがどのように派生したかが示されます。 ADP によって行われた操作が、実行順に表示されます。ただし特定のステップで、特定のフィールドに対して複数の操作が実行されている場合があります。

注: 未変換のフィールドの場合、このテーブルは表示されません。

テーブル内の情報は 2 列または 3 列で表示されます。

- **アクション:** アクションの名前。例えば「連続型予測値」などです。詳しくは、『アクションの詳細』を参照してください。
- **詳細:** 実行された処理のリストです。例えば、標準単位への変換などです。
- **関数:** 構築された予測の場合にのみ表示されます。これは、入力フィールドの線型結合を示します (例: .06*age + 1.21*height)。

アクションの詳細

「アクションの要約」メイン・ビューで下線が付いている「**アクション**」を選択すると、「アクションの詳細」リンク・ビューが表示されます。「アクションの詳細」リンク・ビューには、実行された各処理ステップのアクション固有の情報と共通の情報の両方が表示されます。アクション固有の詳細情報が最初に表示されます。

各アクションの説明が、リンク・ビュー上部にタイトルとして表示されます。アクション固有の詳細情報がタイトルの下に表示されます。この情報には、派生した予測値の数、フィールドの再計算、目標の変換、結合または並べ替えられたカテゴリー、構築または除外された予測値などが含まれます。

アクションが処理されるたびに、予測値の除外または結合などに伴って、処理で使用されている予測値の数が変化することがあります。

注: アクションがオフになっている場合、または目標が指定されていない場合は、「アクションの要約」メイン・ビューでアクションをクリックすると、アクションの詳細情報の代わりにエラー・メッセージが表示されます。

アクションの数は最大 9 つですが、すべての分析ですべてのアクションがアクティブになる訳ではありません。

「テキスト・フィールド」テーブル

このテーブルには、次の数値が表示されます。

- 分析から除外された予測値。

「日時および時刻の予測値」テーブル

このテーブルには、次の数値が表示されます。

- 日付および時刻の予測値から算出した期間。
- 日付および時刻の要素。
- 派生した日付および時刻の予測値の合計。

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

「予測値のスクリーニング」テーブル

このテーブルには、処理から除外された以下の予測値の数が表示されます。
- 定数。
- 欠損値が非常に多い予測値。
- 1 つのカテゴリーのケース数が非常に多い予測値。
- カテゴリー数が非常に多い名義型フィールド (セット)。
- 除外された予測値の合計。

「測定レベルの確認」テーブル

このテーブルには、再計算されたフィールドの数が以下のように分類して表示されます。
- 連続型フィールドとして再計算された順序型フィールド (順序セット)。
- 順序型フィールドとして再計算された連続型フィールド。
- 再計算の合計。

連続型または順序型である入力フィールド (目標または予測値) がない場合は、脚注に表示されます。

「外れ値」テーブル

このテーブルには外れ値の処理方法の数が表示されます。
- 「設定値」タブの「入力と対象の準備」パネルの設定に応じて、外れ値が検出され削除された連続型フィールドの数、または外れ値が検出され、欠損として設定された連続型フィールドの数。
- 外れ値の処理後に定数となったために除外された連続型フィールドの数。

脚注の 1 つには外れ値のカットオフ値が表示され、連続型入力フィールド (対象または予測値) がない場合は別の脚注が表示されます。

「欠損値」テーブル

このテーブルには、欠損値が置換されたフィールドの数が、以下のように分類して表示されます。
- 目標。目標が指定されていない場合はこの行は表示されません。
- 予測値。名義型 (セット)、順序型 (順序セット)、および連続型に分類して表示されます。
- 置換された欠損値の合計。

「目標」テーブル

このテーブルには、対象が変換されたかどうかが次のように表示されます。
- 正規性への Box-Cox 変換。指定されている基準 (平均および標準偏差) とラムダを示す列にさらに分割されます。
- 安定性を向上させるために並べ替えられた対象カテゴリー。

「カテゴリー予測値」テーブル

このテーブルには、以下に該当するカテゴリー予測値の数が表示されます。

- 安定性を向上させるためにカテゴリーが最小から最大の順に並べ替えられている。
- 目標との関連性を最大化するためにカテゴリーが結合されている。
- まばらなカテゴリーを処理するためにカテゴリーが結合されている。
- 目標との関連性が低いために除外されている。
- 結合後に定数となったために除外されている。

カテゴリー予測値がない場合は脚注が表示されます。

「連続型予測値」テーブル

2 つのテーブルがあります。1 番目のテーブル表には、次のいずれかの変換の数が表示されます。

- 標準単位に変換された予測値。また、変換された予測値の数、指定された平均値、標準偏差が表示されます。
- 共通範囲にマップされた予測値。また、指定された最小値および最大値と、min-max 変換を使用して変換された予測値の数も表示されます。
- 分割された予測値と分割された予測値の数。

2 番目のテーブルには、予測値領域の構築の詳細が、以下のような予測値の数として表示されます。

- 構築済み。
- 目標との関連性の低さのために除外されている。
- 分割後に定数となったために除外されている。
- 構築後に定数となったために除外されている。

入力された連続型予測値がない場合は脚注が表示されます。

スコアの後方変換

目標が ADP によって変換されている場合、変換された目標を使用して作成された後続のモデルは、変換された単位をスコアリングします。結果を解釈して使用するために、予測値を元のスケールに変換する必要があります。

1. スコアを後方変換するには、メニューから次の項目を選択します。

   「変換」 > 「モデル作成のデータ準備」 > 「スコアの後方変換...」

2. 後方変換するフィールドを選択します。このフィールドには、変換された目標のモデル予測値が入力されている必要があります。
3. 新しいフィールドの接尾部を指定します。この新しいフィールドには、未変換の目標の元のスケールでモデル予測値が入力されている必要があります。
4. ADP 変換が含まれている XML ファイルの場所を指定します。これは、「インタラクティブ・データ準備」ダイアログまたは「自動データ準備」ダイアログで保存したファイルでなければなりません。詳しくは、 17 ページの『変換の適用と保存』を参照してください。

第 5 章 例外ケースの特定

異常値検出プロシージャーは、クラスター・グループの基準値からの偏差に基づいて、例外的なケースを検索します。このプロシージャーは、任意の推論的データ分析に先立つ予備的なデータ分析ステップで、データ監査の目的で例外的なケースを素早く検出するために設計されています。このアルゴリズムは一般的な異常値検出のために設計されています。つまり、異常ケースの定義は、異常値の定義が適切にできる、医療保険業界での普通でない支払いパターンの検出や金融業界での不正資金浄化 (マネー・ロンダリング) の検出などのような特定の用途に固有のものではありません。

例: 脳卒中の治療結果に関する予測モデルは、異常な観測値の影響を受けやすいため、予測モデルを作成するデータ・アナリストはデータの品質に注意します。こうした異常な観測値の中には、非常に特異なケースを表しているため予測に使用するのは適当でないものがあります。また、技術的には「正しい」値であっても、誤って入力されたために、データ検証のプロシージャーでは検出できない観測値もあります。「例外ケースの特定」プロシージャーは、分析者が外れ値の取り扱いを決めることができるように、それらの外れ値を見つけて報告します。

統計: このプロシージャーは、ピア・グループ、連続型変数とカテゴリー変数のピア・グループ・ノルム、ピア・グループ・ノルムの偏差に基づく異常値指標、および異常と見なされるケースに最も寄与している変数の変数影響値を作成します。

データの考慮事項

データ: このプロシージャーは、連続型変数およびカテゴリー変数の両方に使用できます。それぞれの行は異なる観測値を表し、それぞれの列はピア・グループの基となる異なる変数を表します。データ・ファイルでは出力をマークするためにケース識別変数を使用できますが、分析では使用されません。欠損値は使用できます。重み付け変数が指定されている場合、重み付け変数は無視されます。

検出モデルは、新しい検定データ・ファイルに適用できます。検定データの要素は、学習データの要素と同じである必要があります。また、アルゴリズム設定によっては、モデルの作成に使用される欠損値の処理が、スコアリングの前に検定データ・ファイルに適用される場合があります。

ケースの並び順: ケースの並び順によって解が異なる可能性があることに注意してください。並び順の影響を最小限に抑えるには、ケースを無作為に並べます。特定の解の安定性を確認するには、異なる無作為な順序でソートされたケースを使用していくつかの異なる解を取得します。ファイル・サイズが非常に大きい場合は、異なる無作為な順序でソートされたケースのサンプルを使用し、複数回に分けて実行することができます。

仮定: このアルゴリズムは、すべての変数が一定でなく独立していること、およびすべての入力変数について欠損値を持つケースがないことを仮定します。各連続型変数は正規（ガウス）分布であると仮定し、各カテゴリー変数は多項分布であると仮定します。経験的内部検定は、このプロシージャーが独立仮定および分布仮定の両方の違反に対して堅牢であることを示していますが、これらの仮定がどの程度満たされているか把握するようにしてください。

例外ケースを特定するには

1. メニューから次の項目を選択します。

   「データ」 > 「例外ケースの特定...」

2. 1 つ以上の分析変数を選択します。
3. オプションで、出力のラベル付けに使用するケース識別子変数を選択します。

測定レベルが不明なフィールド

データ・セットの 1 つ以上の変数 (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。測定レベルはこの手順の結果の計算に影響を与えるため、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: アクティブなデータ・セットのデータを読み込み、測定レベルが現在不明なフィールドに対し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ります。

手動で割り当て: 不明な測定レベルのフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、測定レベルをこれらのフィールドに割り当てることができます。データ・エディターの「変数ビュー」でも、測定レベルを割り当てることができます。

測定レベルがこの手順で重要であるため、すべてのフィールドに測定レベルが定義されるまで、ダイアログにアクセスしてこの手順を実行することはできません。

「例外ケースの特定」の「出力」

異常なケースとそれらが異常と見なされる理由のリスト: このオプションでは 3 つの表が作成されます。

- 異常ケースの指数リストは、異常と見なされたケースとその異常値指標の値を表示します。
- 異常ケースのピア ID リストは、例外ケースと、それに対応するピア・グループに関する情報を表示します。
- 異常理由リストは、ケース番号、理由変数、変数影響値、変数の値、および理由ごとの変数のノルムを表示します。

すべての表は、異常値指標の降順でソートされます。さらに、「変数」タブでケース識別子変数が指定されている場合は、ケース識別子が表示されます。

集計: このグループのコントロールは分布の要約を作成します。

- **ピア・グループのノルム:** このオプションを選択すると、「連続型変数ノルム」表 (分析で連続型変数が使用されている場合) または「カテゴリー変数ノルム」表 (分析でカテゴリー変数が使用されている場合) が表示されます。「連続型変数ノルム」表には、ピア・グループごとに、各連続型変数の平均および標準偏差が表示されます。また「カテゴリー変数ノルム」表には、ピア・グループごとに、各カテゴリー変数の最頻値 (度数が最も大きいカテゴリー)、度数、および度数パーセントが表示されます。連続型変数の平均とカテゴリー変数の最頻値は、分析のノルム値として使用されます。
- **異常値指標:** 異常値指標の要約には、異常度が最も高いと特定されたケースの異常値指標の記述統計量が表示されます。
- **各分析変数の理由度数** それぞれの理由に対し、各変数が理由として出現する頻度およびその割合 (パーセント) がこの表に表示されます。また、この表は、それぞれの変数の影響の記述統計量を報告します。「オプション」タブで理由の最大数が 0 に設定されている場合、このオプションは使用できません。
- **処理されたケース:** ケース処理要約には、アクティブなデータ・セットにおけるすべてのケースの度数とその度数のパーセント、分析に組み込まれたケースと除外されたケース、および各ピア・グループのケースが表示されます。

「例外ケースの特定」の「保存」

変数の保存: このグループのコントロールにより、モデル変数をアクティブ・データ・セットに保存できます。また、保存する変数と名前が競合する既存の変数を置き換えることもできます。

- **異常値指標:** 各ケースの異常値指標を指定された名前の変数に保存します。
- **ピア・グループ:** ケースごとに、ピア・グループの ID、ケース度数、および割合 (パーセント) で表されたサイズを、指定されたルート名の変数に保存します。例えば、ルート名として *Peer* が指定されると、変数 *Peerid*、*PeerSize*、および*PeerPctSize* が生成されます。*Peerid* はケースのピア・グループ ID、*PeerSize* はそのグループのサイズ、および *PeerPctSize* はグループのサイズ (パーセント) です。
- **理由:** 理由変数のセットを指定されたルート名で保存します。理由変数のセットは、理由となる変数の名前、変数の影響測度、変数の値、およびノルム値で構成されます。セット数は、「オプション」タブで要求された理由の数に応じて変わります。例えばルート名 *Reason* が指定されている場合に生成される変数は *ReasonVar_k*、*ReasonMeasure_k*、*ReasonValue_k*、および *ReasonNorm_k* です (*k* は 理由の順序 (*k* 番目) です)。理由の数が 0 に設定されている場合は、このオプションを使用できません。

モデル・ファイルのエクスポート: モデルを XML 形式で保存できます。

「例外ケースの特定」の「欠損値」

「欠損値」タブでは、ユーザー欠損値とシステム欠損値の処理方法を制御します。

- **分析から欠損値を除外する:** 欠損値を持つケースが分析から除外されます。
- **分析に欠損値を含める:** 連続型変数の欠損値は対応する全平均に置換され、カテゴリー変数の欠損カテゴリーはグループ化され、有効なカテゴリーとして扱われます。その後、処理された変数が分析で使用されます。必要であれば、ケースごとの欠損変数の比率を表す追加の変数の作成を要求し、その変数を分析で使用することもできます。

「例外ケースの特定」の「オプション」

例外ケースを特定する基準: 以下の選択項目により、異常値リストに含めるケースの数が決まります。

- **異常値指標値が最高であるケースのパーセント:** 100 以下の正数を指定します。
- **異常値指標値が最高であるケースの固定数:** 分析に使用されるアクティブなデータ・セット内のケースの総数以下の正整数を指定します。
- **異常値指標値が最小値以上のケースのみを特定する:** 負ではない数値を指定します。ケースの異常値指標の値が指定されたカットオフ点以上の場合、そのケースは異常と見なされます。このオプションを使用する場合は、「**ケースのパーセント**」と「**ケースの固定数**」オプションを指定してください。例えば、ケースの固定数として 50 を指定し、カットオフ値として 2 を指定した場合、異常値リストには最大で 50 個のケースが含まれ、各ケースは 2 以上の異常値指標値を持ちます。

ピア・グループの数: プロシージャーは、指定された最小値から最大値までの範囲内で最適な数のピア・グループを検索します。これらの値は正整数である必要があり、最小値は最大値以下の値である必要があります。指定された値が等しい場合、プロシージャーは固定数のピア・グループを仮定します。

注: データ内の変動の量によっては、データがサポートできるピア・グループの数が、指定された最小値より小さくなる場合もあります。そのような状況では、プロシージャーが作成するピア・グループが少なくなる場合があります。

理由の最大数: 理由は、変数の影響測度、この理由の変数名、変数の値、および対応するピア・グループの値で構成されます。負ではない整数を指定してください。この値が、分析で使用される処理済み変数の数以上である場合、すべての変数が表示されます。

DETECTANOMALY コマンドの追加機能

このコマンド・シンタックス言語により、以下の操作が可能です。

- すべての分析変数を明示的に指定せずに、アクティブなデータ・セット内のいくつかの変数を除外する (EXCEPT サブコマンドを使用)。
- 連続型変数とカテゴリー変数の影響を均衡させるための調整値を指定する (CRITERIA サブコマンドで MLWEIGHT キーワードを使用)。

シンタックスの詳細については、「コマンド・シンタックス・リファレンス 」を参照してください。

第 6 章 最適カテゴリー化

「最適カテゴリー化」プロシージャーは、1 つ以上のスケール変数 (以下 **分割入力変数**と呼びます) を離散化するために、各スケール変数の値をビンに分配します。ビンの構成は、分割プロセスを「監視」するカテゴリー・ガイド変数に基づいて最適化されます。以降の分析では、元のデータ値の代わりにビンを使用できます。

例: 次に示すように、1 つの変数が取る値の個数を減らすことには、有用な点が数多くあります。

- 他のプロシージャーに必要なデータ要件を満たすことができます。離散化された変数は、カテゴリー型として扱うことができるため、カテゴリー変数を必要とするプロシージャーに使用できます。例えば「クロス集計表」プロシージャーでは、すべての変数がカテゴリー型であることが必要です。
- データ・プライバシーを保護できます。実際の値の代わりに分割された値をレポートすることで、データ・ソースのプライバシーを保護できます。「最適カテゴリー化」プロシージャーでは、ビンの選択に関するガイドが利用できます。
- パフォーマンスが向上します。プロシージャーの中には、値の個数を減らすことでより効率的に処理できるものもあります。例えば多項ロジスティック回帰は、離散化された変数を使用することにより、処理速度を向上させることができます。
- データの完全な区切りまたは準完全な区切りが明確になります。

最適カテゴリー化と連続変数のカテゴリー化: 「連続変数のカテゴリー化」ダイアログ・ボックスでは、いくつかの方法で、ガイド変数を使用せずにビンを自動作成できます。これら「監視なし」の規則は、度数分布表などの記述統計量を生成する際には有効ですが、最終的に予測モデルを作成することが目的である場合は、最適カテゴリー化がより適しています。

出力: このプロシージャーを使用すると、ビンの分割点および各分割入力変数の記述統計量の表を作成できます。この他にも、分割入力変数の分割された値を含む新しい変数をアクティブなデータ・セットに保存したり、離散化する新しいデータで使用できるように、分割規則をコマンド・シンタックスとして保存したりできます。

最適カテゴリー化データの考慮事項

データ: このプロシージャーでは、分割入力変数が数値型スケール変数であることを前提としています。またガイド変数は、カテゴリー型にする必要があり、文字列または数値にできます。

最適カテゴリー化を行うには

1. メニューから次の項目を選択します。

   「変換」 > 「最適カテゴリー化...」
2. 分割入力変数を 1 つ以上選択します。
3. ガイド変数を選択します。

分割されたデータ値を含む変数は、デフォルトでは生成されません。これらの変数を保存するには、「保存」タブを使用してください。

最適カテゴリー化の出力

「出力」タブにより結果の表示が制御されます。

- **ビンの終点:** 各分割入力変数の一連の終点を表示します。
- **分割される変数の記述統計量:** 各分割入力変数について、有効値を持つケースの数、欠損値を持つケースの数、異なる有効値の数、および最小値/最大値が表示されます。ガイド変数の場合は、関連する各分割入力変数のクラス分布が表示されます。
- **分割される変数のモデル・エントロピー:** 各分割入力変数について、ガイド変数についての当該変数の予測精度の測度が表示されます。

「最適カテゴリー化」の「保存」

アクティブ・データ・セットへの変数の保存: 以降の分析で、分割されたデータ値が含まれている変数を元の変数の代わりに使用できます。

分割規則をシンタックスとして保存: 他のデータ・セットを分割するために使用できるコマンド・シンタックスが生成されます。再割り当て規則は、分割アルゴリズムによって決定される分割点に基づきます。

最適カテゴリー化の欠損値

「欠損値」タブでは、欠損値の処理にリストワイズ除去とペアワイズ除去のいずれを使用するかを指定します。ユーザー欠損値は、常に無効として処理されます。元の変数値を新しい変数に再割り当てする場合、ユーザー欠損値はシステム欠損値に変換されます。

- **ペアごと:** このオプションは、ガイド変数と分割入力変数の各ペアに対して適用されます。プロシージャーでは、ガイド変数および分割入力変数に非欠損値が含まれているすべてのケースが使用されます。
- **リストごと:** このオプションは、「変数」タブで指定されたすべての変数に適用されます。ケースで変数が欠損している場合、そのケース全体が除外されます。

最適カテゴリー化のオプション

前処理: 分割入力変数を多数の異なる値に「事前分割」することにより、最終的なビンの品質を大きく損なうことなく、処理時間を短縮できます。作成されるビンの数の上限は、ビンの最大数によって決まります。つまり、最大数を 1000 と指定した場合、分割入力変数が持つ個別値の個数が 1000 未満であれば、その分割入力変数に対して作成される前処理済みのビンの数は、分割入力変数が持つ個別値の個数に等しくなります。

使用頻度の少ないビン: 場合によっては、プロシージャーによって作成されるビンのケース数が非常に少ないことがあります。このような疑似分割点は、次の方法により削除されます。

ある変数に対し、アルゴリズムによって n_{final} 個の分割点が検出された (つまり $n_{final}+1$ 個のビンが検出された) とします。ビン $i = 2, ..., n_{final}$ (値が 2 番目に小さなビンから、値が 2 番目に大きなビンまで) に対して、次の計算を実行します。

$$\frac{sizeof(b_i)}{\min(sizeof(b_{i-1}), sizeof(b_{i+1}))}$$

sizeof(b) はビンのケースの数です。

この値が指定されている結合しきい値よりも小さい場合、b_i は使用頻度が少ないものとしてみなされ、b_{i-1} と b_{i+1} のうち、クラス情報エントロピーが小さい方と結合されます。

このプロシージャーでは、すべてのビンについて上記の一連の処理が行われます。

ビンの終点: このオプションは、区間の下限の定義方法を指定します。分割点の値はプロシージャーによって自動的に決定されるため、このオプションは、必要に応じて使用してください。

最初の (最小の) ビン / 最後の (最大の) ビン: これらのオプションは、各分割入力変数の最小分割点と最大分割点の定義方法を指定します。プロシージャーでは通常、分割入力変数は実数直線上の任意の値を取ることができると想定されますが、理論上または実用上の理由から範囲を制限する場合は、最小値/最大値によってその範囲を定めることができます。

OPTIMAL BINNING コマンドの追加機能

このコマンド・シンタックス言語により、以下の操作が可能です。

- 等度数法による監視なしの分割の実行 (CRITERIA サブコマンドを使用)。

シンタックスの詳細については、「コマンド・シンタックス・リファレンス 」を参照してください。

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。この資料は、IBM から他の言語でも提供されている可能性があります。ただし、これを入手するには、本製品または当該言語版製品を所有している必要がある場合があります。

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

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 Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

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。

記載されている性能データとお客様事例は、例として示す目的でのみ提供されています。実際の結果は特定の構成や稼働条件によって異なります。

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。これらの名前はすべて架空のものであり、名前や住所が類似する個人や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］

［ハ行］

［マ行］

［ラ行］

B

M

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