

# **IBM SPSS Custom Tables**

## **26**

The IBM logo, consisting of the letters 'IBM' in a bold, black, sans-serif font. Each letter is composed of horizontal bars, with the 'I' having three bars, the 'B' having six bars, and the 'M' having four bars.

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、19 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 26 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 Custom Tables 26

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

カスタム・テーブル . . . . .	1	特記事項 . . . . .	19
カスタム テーブルのインターフェース . . . . .	1	商標 . . . . .	20
テーブル・ビルダー・インターフェース . . . . .	1		
テーブルの作成 . . . . .	1	索引 . . . . .	23
カスタム テーブル: 「検定統計量」 . . . . .	7		
サンプル・ファイル . . . . .	9		



---

## カスタム・テーブル

以下のカスタム・テーブル機能が、SPSS® Statistics Standard Edition または Custom Tables オプションに含まれています。

---

### カスタム テーブルのインターフェース

#### テーブル・ビルダー・インターフェース

カスタム・テーブルでは、単純なドラッグ・アンド・ドロップ方式によるテーブル・ビルダー・インターフェースを使用して、テーブルをプレビューしながら、変数やオプションを選択できます。また、一般的なダイアログ・ボックスにはないレベルの柔軟性を備えており、例えばウィンドウのサイズやウィンドウ内のペインのサイズを変更することもできます。

#### テーブルの作成

「カスタム テーブル」インターフェースから、テーブルに表示する変数や要約統計量を選択します。

「分析」 > 「テーブル」 > 「カスタム テーブル」

変数リスト。データ・ファイル内の変数はダイアログの左ペインに表示されます。「カスタム・テーブル」では、変数用の 2 つの異なる測定レベルが区別され、測定レベルに応じて、異なる方法で変数が処理されます。

カテゴリ：限られた数の異なる値またはカテゴリを持つデータ (例えば、性別や宗教)。カテゴリ変数には、文字列 (英数字) 変数、または数字コード (例えば、0 = 男性 および 1 = 女性) でカテゴリを表す数値変数があります。これは、質的データとも呼ばれます。カテゴリ変数は、名義変数または順序変数のいずれかです。

- *Nominal* (名義データ)。本質的な順位を持たないカテゴリ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。

カテゴリ変数は、テーブル内のカテゴリ (行、列、および層) を定義します。デフォルトの要約統計量は、度数 (各カテゴリのケース数) です。例えば、性別に関するカテゴリ変数のデフォルト・テーブルには、単純に男性の数と女性の数が表示されます。

スケール：間隔または比率スケールについて測定したデータ。この場合、データ値は、値の順序と値の間の距離の両方を示します。例えば、\$72,195 の給料は \$52,398 の給料より高く、この 2 つの値の間の距離は \$19,797 です。量的データ、連続型データとも呼ばれます。

通常、スケール変数は、カテゴリ変数のカテゴリ内で要約されます。デフォルトの要約統計量は平均値です。例えば、性別のカテゴリに含まれる収入のデフォルト・テーブルには、男性の平均収入と女性の平均収入が表示されます。

カテゴリ変数を使用してグループを定義することなく、スケール変数自体でスケール変数を要約することもできます。これは主に、複数のスケール変数の要約を積み重ねる場合に便利です。

## 多重回答セット

カスタム・テーブルは、多重回答セットという特殊な種類の変数もサポートしています。多重回答セットは、実際には、一般的な意味での変数ではありません。多重回答セットはデータ・エディターには表示されず、他のプロシージャーでも認識されません。多重回答セットは、複数の回答が可能な質問に対する回答を、複数の変数を使用して記録します。多重回答セットは、カテゴリ変数と同様に処理され、カテゴリ変数で実行できるほとんどの処理は、多重回答セットでも実行することができます。

変数リストの各変数の隣にあるアイコンは、変数の型を識別します。

カテゴリ: 変数リストでカテゴリ変数を選択すると、そのカテゴリ変数用に定義されたカテゴリが「変数情報」ペインに表示されます。これらのカテゴリは、テーブルで変数を使用する際に、キャンバス・ペインにも表示されます。変数にカテゴリが定義されていない場合、「変数情報」ペインとキャンバス・ペインには、プレースホルダー・カテゴリとして「カテゴリ 1」と「カテゴリ 2」の 2 つが表示されます。

テーブル・ビルダーに表示される定義済みカテゴリは、値ラベルに基づいています。これは説明的なラベルで、さまざまなデータ値に割り当てられます (例えば、数値 0 と 1 に、値ラベルの「男性」と「女性」が割り当てられるなど)。値ラベルはデータ・エディターの「変数情報」ペインで定義することができます。

キャンバス・ペイン: テーブルを作成するには、キャンバス・ペインの行と列に変数をドラッグ・アンド・ドロップします。キャンバス・ペインには、作成されるテーブルのプレビューが表示されます。キャンバス・ペインのセルには、実際のデータ値は表示されませんが、キャンバス・ペインで、最終的なテーブルのレイアウトをかなり正確に確認することができます。カテゴリ変数の場合、値ラベルが定義されていない固有値がデータ・ファイル内に存在すると、プレビューに表示されているよりも多くのカテゴリが実際のテーブルに含まれることがあります。

## テーブル作成時の基本的な規則と制約事項

- カテゴリ変数の場合、要約統計量は、統計ソース・ディメンション内で最も深いレベルの変数に基づきます。
- カテゴリ変数のデフォルトの統計ソース・ディメンション (行または列) は、変数をキャンバス・ペインにドラッグ・アンド・ドロップする順序に基づきます。例えば、最初に行トレイに変数をドラッグした場合、行ディメンションが、デフォルトの統計ソース・ディメンションになります。
- スケール変数は、行ディメンション内または列ディメンション内で最も深いレベルの変数のカテゴリ内でのみ要約できます (スケール変数は、テーブルの任意のレベルに配置できますが、要約は最も深いレベルで行われます)。
- スケール変数を他のスケール変数内で要約することはできません。ただし、複数のスケール変数の要約を積み重ねたり、カテゴリ変数のカテゴリ内でスケール変数を要約したりすることはできます。あるスケール変数を別のスケール変数内でネストしたり、あるスケール変数を行ディメンション内に配置し、別のスケール変数を列ディメンション内に配置したりすることはできません。
- アクティブ・データ・セット内のいずれかの変数に 12,000 を超える定義済み値ラベルが含まれている場合、テーブル・ビルダーを使用してテーブルを作成することはできません。この制限を超える変数をテーブルに含める必要がなければ、これらの変数を除外する変数セットを定義して適用することができます。12,000 を超える定義済み値ラベルを持つ変数を含める必要がある場合は、CTABLES コマンド構文を使用してテーブルを生成することができます。

## テーブルの作成

1. メニューから次の項目を選択します。

「分析」 > 「テーブル」 > 「カスタム テーブル」

2. 1 つ以上の変数を、キャンバス・ペインの行領域または列領域にドラッグ・アンド・ドロップするか、移動します。
3. 「作成」をクリックしてテーブルを作成します。

### キャンバス・ペインから変数を削除するには

1. キャンバス・ペインで変数を選択 (クリック) します。
2. 右クリックして、ドロップダウン・メニューから「変数を削除」を選択します。

### 変数のネスト

クロス集計表と同様に、ネストでも 2 つのカテゴリ変数の間にある関係を示すことができます。ただし、同じディメンションで別の変数にネストされている変数は除きます。例えば、行ディメンションで「年齢カテゴリ」の中に「性別」をネストして、各年齢カテゴリに属する男性と女性の数を表示できます。

カテゴリ変数の中にスケール変数をネストすることもできます。例えば、「性別」の中に「収入」をネストして、平均収入 (または中央値や他の要約統計量) を男性と女性に分けて表示することができます。

### 変数をネストするには

1. カテゴリ変数をキャンバス・ペインの行領域または列領域にドラッグ・アンド・ドロップします。
2. カテゴリ変数またはスケール変数を、カテゴリ行変数またはカテゴリ列変数の上にドラッグ・アンド・ドロップします。
3. メニューから「すべての変数の上にネスト」、「左にネスト」、または「右にネスト」を選択します。

表 1. ネストしたカテゴリ変数

変数 1	変数 2	要約統計量
カテゴリ 1	カテゴリ 1	12
	カテゴリ 2	34
	カテゴリ 3	56
カテゴリ 2	カテゴリ 1	12
	カテゴリ 2	34
	カテゴリ 3	56

注: カスタム テーブルは、層化分割ファイルの処理は行いません。層化分割ファイルと同じ結果を得るには、分割ファイル変数を、テーブルの最も浅いレベルにネストされた層に配置します。

## 統計の編集

「統計の編集」ペインでは以下の処理を実行できます。

- テーブルの要約統計量の追加と削除を行う。

「統計の編集」ペインで選択可能な統計量 (およびその他のオプション) は、統計量ソース変数の測定レベルに依存します。統計量のソース (統計量の計算に使用する変数) は、以下の基準で決まります。

- 測定レベル: テーブル (または積み重ねテーブルのテーブル・セクション) にスケール変数がある場合、統計量は、そのスケール変数に基づきます。

- 変数の選択順序: カテゴリー変数のデフォルトの統計ソース・ディメンション (行または列) は、変数をキャンバス・ペインにドラッグ・アンド・ドロップする順序に基づきます。例えば、最初に行領域に変数をドラッグした場合、行ディメンションがデフォルトの統計ソース・ディメンションになります。
- ネスト: カテゴリー変数の場合、統計量は、統計ソース・ディメンション内で最も深いレベルの変数に基づきます。

カテゴリー変数の要約統計量: カテゴリー変数で使用できる基本的な統計量は、度数とパーセントです。合計と小計に対して、カスタム要約統計量を指定することもできます。カスタム要約統計量には、一部の順序カテゴリー変数に適した中心傾向 (平均値や中央値など) や散らばり (標準偏差など) の測定値などがあります。

度数: テーブルの各セルに含まれるケースの数、または多重回答セットの回答者の数。重み付けが有効な場合、この値が重み付き度数です。

- 重み付けが有効である場合、この値は重み付き度数です。
- 重み付き度数は、グローバルなデータセット重み付け (「データ」 > 「ケースの重み付け...」) の場合と同じです。

重み付けのない度数: テーブルの各セルに含まれる重み付けのないケースの数。重み付けが有効な場合のみ、度数とは異なって機能します。

調整済み度数 有効ベースの重み付けの計算で使用される調整済み度数。有効ベースの重み付け変数を使用しない場合、調整済み度数は度数と同じです。

行パーセント: 各行でのパーセント。サブテーブルの各行のパーセントを合計すると 100% になります (単純なパーセントの場合)。通常、行パーセントが役立つのは、カテゴリー列 変数が使用されている場合だけです。

列パーセント: 各列でのパーセント。サブテーブルの各列のパーセントを合計すると 100% になります (単純なパーセントの場合)。通常、列パーセントが役立つのは、カテゴリー行 変数が使用されている場合だけです。

サブテーブル・パーセント: 各セルのパーセントは、サブテーブル全体に基づきます。サブテーブル内のすべてのセルのパーセントは同じケース合計数に基づき、サブテーブル内で合計すると 100% になります。ネストされたテーブルでは、最も深いネストレベルの前にある変数によってサブテーブルが定義されます。例えば、「年齢カテゴリー」の中に「性別」があり、その中に「婚姻状況」があるテーブルの場合、「性別」によってサブテーブルが定義されます。

表パーセント: 各セルのパーセントは、テーブル全体に基づきます。すべてのセルのパーセントは同じケース合計数に基づき、テーブル全体で合計すると 100% になります (単純なパーセントの場合)。

#### 信頼区間

- 信頼限界の下限および上限は、度数、パーセント、平均、中央値、パーセンタイル、および合計で使用可能です。
- ラベル内の文字列「&[信頼度レベル]」には、テーブル内の列ラベルの信頼度レベルが含まれます。
- 標準誤差は、度数、パーセント、平均、および合計で使用可能です。
- 信頼区間および標準誤差は、多重回答グループでは使用できません。

水準 信頼区間の信頼水準をパーセントで表したものの。値は 0 より大きく 100 未満であることが必要です。



## 多重回答セット

多重回答セットでは、ケース数、回答数、度数に基づいてパーセントを計算することができます。詳しくは、『多重回答セットの要約統計量』のトピックを参照してください。

パーセントの分母: パーセントの計算には 3 つの方法があります。どの方法で計算されるかは、計算の基準で欠損値をどのように扱うかによって決まります。

単純なパーセント: パーセントはテーブルで使用するケース数に基づいており、すべてのパーセントを合計すると常に 100% になります。カテゴリーをテーブルから除外した場合、そのカテゴリーのケースは分母となる数から除外されます。システム欠損値のあるケースは、常に分母となる数から除外されます。ユーザー欠損値のあるケースは、ユーザー欠損カテゴリーがテーブルから除外される場合 (デフォルト) には除外され、ユーザー欠損カテゴリーがテーブルに含まれる場合には含められます。名前に有効 N または合計 N が含まれないパーセントは、単純なパーセントです。

合計 N パーセント: システム欠損値およびユーザー欠損値があるケースが、単純なパーセントの分母に追加されます。パーセントの合計が、100% にならない場合があります。

有効 N パーセント: ユーザー欠損カテゴリーがテーブルに含まれている場合でも、ユーザーユーザー欠損値があるケースは、単純なパーセントの分母から除外されます。

注: ユーザー欠損カテゴリー以外のカテゴリーを手動で除外した場合、そのカテゴリーのケースは、常に分母となる数から除外されます。

多重回答セットの要約統計量: 多重回答セットには、以下の要約統計量も使用できます。

列/行/層の応答 %: 応答数に基づくパーセントです。

列/行/層の応答 % (基準: 度数): 応答数が分子、合計度数が分母となります。

列/行/層の度数 % (基準: 応答数): 度数が分子、合計応答数が分母となります。

層の列/行の応答 %: サブテーブルでのパーセントです。応答数に基づくパーセントです。

層の列/行の応答 % (基準: 度数): サブテーブルでのパーセントです。応答数が分子、合計度数が分母となります。

層の列/行の応答 % (基準: 応答数): サブテーブルでのパーセントです。度数が分子、合計応答数が分母となります。

応答: 応答数です。

サブテーブル/テーブルの応答 %: 応答数に基づくパーセントです。

サブテーブル/テーブルの応答 % (基準: 度数): 応答数が分子、合計度数が分母となります。

サブテーブル/テーブルの度数 % (基準: 応答数): 度数が分子、合計応答数が分母となります。

スケール変数とカテゴリー・カスタム合計の要約統計量: カテゴリー変数で使用できる度数とパーセントのほかに、スケール変数では以下の要約統計量も使用することができます。また、これらの要約統計量は、カテゴリー変数のカスタム合計や小計要約として使用することもできます。これらの要約統計量は、多重回答セットや文字列 (英数字) 変数では使用できません。

平均値: 算術平均 (合計をケース数で割った値)。

中央値: ケースの中央付近にある値 (50 パーセントイル)。

最頻値: 最も頻繁に出現する値。値の出現頻度が同じである場合、最小の値が表示されます。

最小: 最小 (最低) 値。

最大: 最大 (最高) 値。

欠損値: 欠損値の度数 (ユーザー欠損値とシステム欠損値の両方)。

パーセンタイル: 5、25、75、95、99 のいずれかのパーセンタイル (あるいはこれらすべてのパーセンタイル) を指定できます。

範囲: 最大値と最小値の差分。

標準偏差: 平均値を中心とした散らばりの測定値。正規分布では、68% のケースが平均値  $\pm 1$  標準偏差に含まれ、95% のケースが平均値  $\pm 2$  標準偏差に含まれます。例えば、平均年齢が 45 歳で標準偏差が 10 の場合、正規分布ではケースの 95% が 25 歳から 65 歳までに含まれます (分散の平方根)。

合計: 値の合計。

合計パーセント: 合計に基づくパーセント。行と列 (サブテーブル内)、行全体と列全体 (サブテーブル全体)、層、サブテーブル、テーブル全体で使用することができます。

合計 N: 非欠損値、ユーザー欠損値、およびシステム欠損値の度数。ユーザー欠損カテゴリを除き、手動で除外したカテゴリのケースは含まれません。

調整済み合計 N 有効ベースの重み付けの計算で使用される調整済み合計 N。有効ベースの重み付け変数 (「オプション」タブ) を使用しない場合、調整済み合計 N は合計 N と同じです。この統計量は、多重回答セットでは使用できません。

有効 N: 非欠損値の度数。ユーザー欠損カテゴリを除き、手動で除外したカテゴリのケースは含まれません。

調整済み有効 N 有効ベースの重み付けの計算で使用される調整済み有効 N。有効ベースの重み付け変数 (「オプション」タブ) を使用しない場合、調整済み有効 N は有効 N と同じです。この統計量は、多重回答セットでは使用できません。

分散: 平均値を中心とした散らばりの測定値。平均値からの偏差の平方和を、ケースの数から 1 を引いた値で割った結果に等しくなります。分散は、変数自体の平方を単位として測定されます (標準偏差の平方)。

信頼区間

- 信頼限界の下限および上限は、度数、パーセント、平均、中央値、パーセンタイル、および合計で使用可能です。
- ラベル内の文字列「&[信頼度レベル]」には、テーブル内の列ラベルの信頼度レベルが含まれます。
- 標準誤差は、度数、パーセント、平均、および合計で使用可能です。
- 信頼区間および標準誤差は、多重回答グループでは使用できません。

水準 信頼区間の信頼水準をパーセントで表したものです。値は 0 より大きく 100 未満であることが必要です。

## 積み重ねテーブル

積み重ね変数によって定義されたテーブルの各セクションは個別のテーブルとして扱われ、要約統計量もそれによって計算されます。

## カテゴリーと合計

カスタム・テーブルを使用すると、以下の操作を実行できます。

- カテゴリーの並べ替え。
- 合計の挿入。
- 定義済みの値ラベルがない変数の場合、カテゴリーのソートと合計の挿入のみ、実行することができます。

### カテゴリーと合計のオプションにアクセスするには

1. カテゴリー変数または多重回答セットを、キャンバス・ペインにドラッグ・アンド・ドロップします。
2. キャンバス・ペインで変数を右クリックし、ポップアップ・メニューからカテゴリーまたは合計のオプションから 1 つ選択します。

### カテゴリーをソートするには

1. キャンバス・ペインで変数を右クリックし、ポップアップ・メニューから「カテゴリーを並べ替え」を選択した後、ソート方法を選択します。
  - 値順
  - ラベルごと
  - 出現頻度順
  - 小さい順

### 合計

1. キャンバス・ペインで変数を右クリックし、ポップアップ・メニューから「合計の表示」を選択した後、合計を表示する場所を選択します。
  - カテゴリーの上
  - カテゴリーの下

選択した変数が別の変数内でネストされている場合、各サブテーブルに合計が挿入されます。

。

## カスタム テーブル: 「検定統計量」

「検定統計量」機能では、カスタム テーブルに対する有意差検定を行えます。



カテゴリ ラベルがデフォルトのテーブル ディメンションから移動されたテーブルおよび計算対象カテゴリでは、これらの検定を使用することはできません。

### 列の平均値および列比率の検定

列の平均値検定は、スケール変数で使用可能です。列比率の検定は、カテゴリー変数で使用可能です。

## 列の平均値の比較

列の平均値の等質性の、ペアごとの検定です。テーブルの列にはカテゴリ変数が含まれていなければなりません。また、行の最も内側のレベルとしてスケール変数が必要です。テーブルには、要約統計量として平均値が含まれている必要があります。

通常のカテゴリ変数の場合、分散は、すべてのカテゴリから、または、比較されるカテゴリのみから推定できます。多重回答変数の場合、平均値検定の分散は常に、比較されるカテゴリのみに基づきます。

## 列の比率の比較

列の比率の等質性の、ペアごとの検定です。テーブルの列と行の両方に、少なくとも 1 つのカテゴリ変数が含まれていなければなりません。テーブルには、度数または列パーセントが含まれている必要があります。

## 有意水準

列の平均値および列比率の検定の有意水準。

- この値は 0 より大きくて 1 より小さくする必要があります。
- 有意水準を 2 つ指定する場合、より小さい有意水準以下の有意確率値の識別には大文字が使用されます。より大きい有意水準以下の有意確率値の識別には小文字が使用されます。
- 「**APA** スタイルの添え字を使用」を選択した場合、2 番目の値は無視されます。

## 多重比較のため p 値を調整

**Bonferroni** の修正では、ファミリーごとのエラー率 (FWER) を調整します。

**Benjamini-Hochberg** 法では、偽の発見率 (FDR) を調整します。この方法は **Bonferroni** 修正ほど保守的ではありません。

## 有意差の特定

列の平均値および列比率の検定では、有意差の結果を別個のテーブルまたはメインテーブルに表示できます。

### 個別テーブル

有意差検定の結果は、個別テーブルに表示されます。2 つの値が大きく異なる場合、大きいほうの値に対応するセルには、小さいほうの値を含む列を示すキーが表示されます。

### 有意確率の値を表示

有意確率の値は、セル内の各キー値の後に括弧で囲んで表示されます。このオプションを使用できるのは、有意差の結果が別個のテーブルに表示されている場合のみです。

### メインテーブル

有意差検定の結果は、メインテーブルに表示されます。テーブル内の各列カテゴリは、アルファベットのキーで識別されます。各有意確率ペアについて、列平均または列比率が小さいほうのカテゴリのキーが、列平均または列比率がより大きいほうのカテゴリ内に表示されます。

- ピボット・テーブル内で列ラベルのセルのキーにカーソルを合わせると、その有意度キーを持つテーブル内のセルがすべて強調表示されます。列ディメンションに多重変数を持つテーブルの場合は、そのサブテーブル内のセルだけが強調表示されます。
- テーブル (またはサブテーブル) 内で同じ有意度キーを持つセルをすべて選択するには、その列ラベルのセルを右クリックし、「選択」 > 「この有意度キーですべてのセルを選択」を選択します。

### **APA** スタイルの添え字を使用

添え字を使用した **APA** スタイルの形式で有意差を識別します。2 つの値が大きく

異なる場合、これらの値には別の下付き文字が表示されます。これらの下付き文字は、脚注ではありません。このオプションを有効にすると、現在のテーブルルックで定義されている脚注のスタイルが上書きされ、上付き文字の番号として表示されます。同じ行内で同じ有意度キーを持つセルをすべて選択するには、その有意度キーを持つセルを右クリックし、「有意度が類似したセルを選択」を選択します。

#### 独立性の検定 (カイ 2 乗)

行と列の両方に少なくとも 1 つのカテゴリ変数が存在するテーブルに対する、独立性のカイ 2 乗検定。

#### 小計カテゴリの代わりに小計を使用する

有意差検定の小計のカテゴリに代わって各小計が使用されます。このオプションを選択しなかった場合は、小計カテゴリが非表示になっている小計だけが、小計のカテゴリに代わって検定で使用されます。

#### 検定に多重回答変数を含める

多重回答グループのカテゴリが有意差検定に含まれます。それ以外の場合、多重回答グループは有意差検定に含まれません。

---

## サンプル・ファイル

製品とともにインストールされるサンプル・ファイルは、インストール・ディレクトリーの *Samples* サブディレクトリーにあります。Samples サブディレクトリー内には、英語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ロシア語、中国語 (簡体字)、スペイン語、中国語 (繁体字) の言語ごとに個別のフォルダーがあります。

すべてのサンプル・ファイルが、すべての言語で使用できるわけではありません。あるサンプル・ファイルが特定の言語で使用できない場合、その言語のフォルダーには、サンプル・ファイルの英語バージョンが含まれています。

### 説明

このドキュメントのさまざまな例で使用されているサンプル・ファイルの簡単な説明を以下に示します。

- **accidents.sav:** これは、特定の地域における自動車事故の危険因子を年齢と性別ごとに調査している保険会社に関する架空のデータ・ファイルです。各ケースが、年齢カテゴリと性別のクロス分類に対応しています。
- **adl.sav:** これは、脳卒中患者に対して提案されるタイプの治療の効果を特定する取り組みに関する架空のデータ・ファイルです。医師団は、女性の脳卒中患者たちを、2 つのグループのいずれかにランダムに割り当てました。最初のグループは標準的な理学療法を受け、2 番目のグループはさらに心理療法を追加で受けました。治療の 3 カ月後に、各患者が日常生活における一般的な活動をどの程度行うことができるかを、順序変数として得点付けしました。
- **advert.sav:** これは、広告費とその売上成果との関係を調べることを目的とした、小売業者の取り組みに関する架空のデータ・ファイルです。この目的のために、過去の売上高の数値と、それに関する広告費のデータが収集されています。
- **aflatoxin.sav:** これは、収穫物によって濃度が大きく異なる毒物であるアフラトキシンを、トウモロコシの収穫物で検定することに関する架空のデータ・ファイルです。ある穀物加工業者は、8 つの収穫物から 16 のサンプルを受け取り、10 億分の 1 単位でアフラトキシン・レベルを測定しました。

- **anorectic.sav:** 拒食行動と過食行動の症状の標準化を目的として、調査員<sup>1</sup>が摂食障害を持つ55人の若者について調査を行いました。各患者が4年間で4回診察を受け、観察記録の合計は220件となっています。毎回の観察では、16種類の症状それぞれについて患者にスコアが付けられました。患者71(2回目)、患者76(2回目)、患者47(3回目)の症状のスコアは欠落しているため、有効な観察記録は217件ということになります。
- **anticonvulsants.sav:** 医学研究者は一般化線型混合モデルを使用して、新しい抗けいれん薬により患者がてんかん性発作を起こす割合をどのくらい減らせるかを割り出すことができます。同じ患者からの反復測定は通常正の相関性があるため、いくつかのランダム効果を含む混合モデルが適しています。対象フィールド(発作の回数)は正の整数値を取るため、ポワソン分布および対数リンクを使用した一般化線型混合モデルが適切な場合があります。
- **bankloan.sav:** これは、債務不履行率を低減させるための銀行の取り組みに関する架空のデータ・ファイルです。このファイルには、過去の顧客と見込み客850人に関する財務情報と人口統計情報が含まれています。最初の700件のケースは、以前に貸付を行った顧客です。残りの150件のケースは見込み顧客です。銀行は、これらの顧客に関して信用リスクの良し悪しを分類する必要があります。
- **bankloan\_binning.sav:** これは、過去の5,000人の顧客に関する財務情報と人口統計情報が含まれている架空のデータ・ファイルです。
- **bankloan\_cs.sav:** これは、債務不履行になる可能性がある人物を示す特徴を識別し、それらの特徴を使用して、信用リスクの良し悪しを識別する銀行の取り組みに関する架空のデータ・ファイルです。
- **bankloan\_cs\_noweights.sav:** これは、債務不履行になる可能性がある人物を示す特徴を識別し、それらの特徴を使用して、信用リスクの良し悪しを識別する銀行の取り組みに関する架空のデータ・ファイルです。このファイルには抽出の重み付けは含まれていません。
- **behavior.sav:** 典型的な例<sup>2</sup>として、52人の学生に対し、15種類の状況と15種類の行動の組み合わせについて、0 = 「非常に適切」から9 = 「非常に不適切」までの10段階でランク付けするよう依頼しました。個人間で平均を取り、非類似度として値を取りました。
- **behavior\_ini.sav:** このデータ・ファイルには、*behavior.sav*の2次元の解の初期構成が含まれていません。
- **brakes.sav:** これは、高性能自動車のディスク・ブレーキを生産している工場における品質管理に関する架空のデータ・ファイルです。このデータ・ファイルには、8台の機械で生産した16個のディスクの直径測定値が含まれています。ブレーキの目標の直径は322ミリメートルです。
- **breakfast.sav:** 典型的な調査<sup>3</sup>として、21人のWharton School MBAの学生とその配偶者に対し、15種類の朝食を好みの順に1 = 「最も好き」から15 = 「最も嫌い」までランク付けするよう依頼しました。調査対象者の嗜好は、「全体的な好み」から「スナックとドリンクのみ」まで、6つの異なるシナリオで記録されています。
- **breakfast-overall.sav:** このデータ・ファイルには、最初のシナリオ(「全体的な好み」)における朝食の品目についての好みだけが含まれています。
- **broadband\_1.sav:** これは、全国規模のブロードバンド・サービスの地域ごとの加入者数が含まれている架空のデータ・ファイルです。このデータ・ファイルには、85地域の月々の加入者数が4年間分含まれています。

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

2. Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

3. Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

- **broadband\_2.sav:** このデータ・ファイルは *broadband\_1.sav* と同じですが、3 カ月分のデータが追加されています。
- **cable\_survey.sav:** テレビ、電話、およびインターネット・サービスのケーブル・プロバイダーの経営陣は、潜在的な顧客についての詳細を把握する必要があります。サービス地域で 2000 人の調査を行い、3 つのサービスに関して (1) サービスを利用していないか (2) 他のプロバイダーのサービスに加入しているか、または (3) 当社のサービスを利用しているかについて尋ねました。調査では、性別、年齢カテゴリ (4 レベル)、教育カテゴリ (3 レベル)、収入カテゴリ (4 レベル)、住居タイプ・カテゴリ (4 レベル)、現在の住所での年数カテゴリ (3 レベル)、居住人数など、いくつかの人口統計情報も収集しています。
- **car\_insurance\_claims.sav:** これは、他の場所<sup>4</sup> で表示および分析される、自動車の損害請求に関するデータ・セットです。逆リンク関数を使用して、従属変数の平均値を保険契約者の年齢、車種、製造年の線型結合に関連付けることにより、平均請求数をガンマ分布としてモデル化することができます。申請された請求の数は、尺度重み付けとして使用することができます。
- **car\_sales.sav:** このデータ・ファイルには、自動車のさまざまな車種やモデルの架空の売上推定値、表示価格、物理的仕様が含まれています。表示価格と物理的仕様は、それぞれ *edmunds.com* と製造元のサイトから入手したものです。
- **car\_sales\_uprepared.sav:** これは、フィールドの変換バージョンが含まれていない、*car\_sales.sav* の修正バージョンです。
- **carpet.sav:** 一般的な例<sup>5</sup> として、新しいカーペット専用洗剤の市販を計画している企業が、消費者の嗜好に関する 5 つの因子 (パッケージのデザイン、ブランド名、価格、「*Good Housekeeping*」のシール、返金保証) の影響について調べようとしています。パッケージのデザインには 3 つの因子レベルがあり、それぞれのレベルによって塗布用ブラシの位置が異なります。また、3 つのブランド名 (*K2R*、*Glory*、*Bissell*) と 3 つの価格水準があり、残り 2 つの因子のそれぞれについて 2 つのレベル (「なし」と「あり」) があります。10 人の消費者が、これらの因子によって定義された 22 個のプロファイルに順位を付けます。「嗜好」という変数には、各プロファイルの平均順位のランキングが格納されます。ランキングが低いほど、嗜好度は高くなります。この変数は、各プロファイルの全体的な嗜好測定値を反映しています。
- **carpet\_prefs.sav:** このデータ・ファイルは *carpet.sav* で説明したものと例に基づいていますが、10 人の消費者それぞれから収集した実際のランキングが含まれています。これらの消費者は、22 種類の製品プロファイルを、一番好きなものから一番嫌いなものまで順位付けすることを依頼されています。変数 *PREF1* から *PREF22* には、*carpet\_plan.sav* で定義された、関連するプロファイルの ID が格納されます。
- **catalog.sav:** このデータ・ファイルには、あるカタログ会社が販売した 3 つの製品の、架空の月間売上高の数値が含まれています。また、5 つの予測変数のデータも含まれています。
- **catalog\_seasfac.sav:** このデータ・ファイルは *catalog.sav* と同じですが、季節性の分解プロシージャーとそれに付随する日付変数から計算した一連の季節因子が追加されています。
- **cellular.sav:** これは、チャーン (契約/解約を繰り返す顧客の流動現象) を削減するための携帯電話会社の取り組みに関する架空のデータ・ファイルです。チャーンの傾向スコアは、0 から 100 の範囲でアカウントに適用されます。スコアリングが 50 以上のアカウントは、プロバイダーを変更しようとしていると考えられます。
- **ceramics.sav:** これは、新しい上質の合金は、標準的な合金より高い耐熱性を持っているかどうかを判断するための、ある製造業者の取り組みに関する架空のデータ・ファイルです。各ケースが 1 つの合金の個別のテストを表し、耐熱性の限界温度が記録されます。

4. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

5. Green, P. E., and Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **cereal.sav:** これは、880 人を対象に、朝食の好みについて、年齢、性別、婚姻状況、ライフスタイルが活動的かどうか (週 2 回以上運動するか) を含めて調査した、架空のデータ・ファイルです。各ケースが個別の回答者を表しています。
- **clothing\_defects.sav:** これは、ある衣料品工場での品質管理工程に関する架空のデータ・ファイルです。工場で生産される各ロットから、検査官が衣料品のサンプルを取り出し、不良品の数を数えます。
- **coffee.sav:** これは、6 つのアイス・コーヒー・ブランド<sup>6</sup> の印象に関連するデータ・ファイルです。回答者は、アイス・コーヒーに関する 23 の各印象属性について、その属性が表現していると思われるすべてのブランドを選択しました。機密保持のため、6 つのブランドを AA、BB、CC、DD、EE、FF で表しています。
- **contacts.sav:** これは、企業のコンピューター営業担当者グループの連絡先リストに関する架空のデータ・ファイルです。それぞれの連絡先は、所属する会社の部門と会社のランクによって分類されています。また、最後に販売した金額、最後の販売以降の経過時間、連絡先の会社の規模も記録されています。
- **credit\_card.sav:** クレジット・カードの利用に関する架空の調査では、各対象者のメインのクレジットカードでの毎月の消費を 2 年間追跡調査し、その消費内容を取り引きのタイプ (食糧、小売り、交際費、旅行、その他) に分割しています。データ・セットの各レコードは、特定の月の消費および取り引きのタイプに対応するため、各対象者について収集されるデータは、2 年 × 12 カ月/年 × 5 タイプの取り引き = 120 レコードが必要となります。
- **creditpromo.sav:** これは、最近のクレジットカード・プロモーションの有効性を評価することを目的とした、あるデパートの取り組みに関する架空のデータ・ファイルです。この目的のために、500 人のカード所有者がランダムに選択されました。そのうちの半分に対して、今後 3 カ月間の買い物に関して利率を下げることを知らせる広告を送付しました。残りの半分には、通常どおりの定期的な広告を送付しました。
- **cross\_sell.sav:** ある通信販売の会社には、本のクラブと CD のクラブがあります。毎月、クラブ・メンバーには特別なオファーが提供されます。この会社は、本の購入合計、CD の購入合計、およびクラブ・メンバーに提供されたオファーのタイプに基づいて、月の特別オファー購入合計のモデルを作成したいと考えています。この場合は、2 段階最小二乗回帰が適しています。これは、特別オファーに消費されたお金は、本や CD に使われたお金ではないためです。したがって、応答とこれら 2 つの予測変数の間には、フィードバック・ループが存在します。
- **customer\_dbase.sav:** これは、自社のデータウェアハウスにある情報を使用して、最も反応がありそうな顧客に対して特典を提供するための、ある会社の取り組みに関する架空のデータ・ファイルです。顧客ベースのサブセットをランダムに選択して特典を提供し、それに対する顧客の反応が記録されています。
- **customer\_information.sav:** これは、名前や住所など、顧客の連絡先情報が含まれている架空のデータ・ファイルです。
- **customer\_subset.sav:** これは、*customer\_dbase.sav* の 80 件のケースのサブセットです。
- **debate.sav:** これは、政治討論会の出席者から得た、討論前と討論後の調査に対する回答のペアに関する架空のデータ・ファイルです。各ケースが個々の回答者に対応しています。
- **debate\_aggregate.sav:** これは、*debate.sav* 内の回答が集計されている架空のデータ・ファイルです。各ケースが、討論前後の好みのクロス分類に対応しています。

6. Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.



- **demo.sav:** これは、毎月の特典を送付することを目的とした、購入顧客のデータベースに関する架空のデータ・ファイルです。顧客が特典に反応したかどうか、さまざまな人口統計情報と共に記録されています。
- **demo\_cs\_1.sav:** これは、調査情報のデータベースを作成するための、ある会社の取り組みの最初のステップに関する架空のデータ・ファイルです。各ケースが個別の都市に対応し、地域、地方、地区、および都市の ID が記録されています。
- **demo\_cs\_2.sav:** これは、調査情報のデータベースを作成するための、ある会社の取り組みの第 2 のステップに関する架空のデータ・ファイルです。各ケースが、最初のステップで選択した都市の個別の世帯単位に対応し、地域、地方、地区、都市、区画、および単位の ID が記録されます。計画の最初の 2 段階からのサンプリング情報も含まれています。
- **demo\_cs.sav:** これは、複合サンプリング計画を使用して収集された調査情報を含む架空のデータ・ファイルです。各ケースが個別の世帯単位に対応し、さまざまな人口統計情報と抽出情報が記録されています。
- **diabetes\_costs.sav.** これは、糖尿病を持つ保険契約者に関して保険会社が保持する情報が含まれている架空のデータ・ファイルです。各ケースが個々の保険契約者に対応しています。
- **dietstudy.sav:** この架空のデータ・ファイルには、「Stillman diet」<sup>7</sup> の研究結果が含まれています。各ケースが個々の被験者に対応し、被験者のダイエット前後の体重 (ポンド単位) と、トリグルセリド・レベル (mg/100 ml 単位) が記録されています。
- **dmdata.sav:** これは、ダイレクト・マーケティング企業の人口統計情報と購入情報が含まれている架空のデータ・ファイルです。*dmdata2.sav* には、テスト・メールを受け取った連絡先のサブセットの情報が含まれ、*dmdata3.sav* には、テスト・メールを受け取らなかった残りの連絡先に関する情報が含まれています。
- **dvdplayer.sav:** これは、新しい DVD プレイヤーの開発に関する架空のデータ・ファイルです。マーケティング・チームはプロトタイプを使用して、フォーカス・グループ・データを収集しました。各ケースが個々の調査対象ユーザーに対応し、ユーザーの人口統計情報と、プロトタイプに関する質問への回答が記録されています。
- **Employee data.sav:** これは、従業員固有の情報 (教育レベル、雇用カテゴリー、現在の給与、前職の経験など) が入っている架空データ ファイルです。
- **german\_credit.sav:** このデータ・ファイルは、カリフォルニア大学アーバイン校の Repository of Machine Learning Databases<sup>8</sup> に含まれている「German credit」データ・セットから抜粋したものです。
- **grocery\_1month.sav:** これは、*grocery\_coupons.sav* データ・ファイルの週ごとの購入を「ロールアップ」して、各ケースが個々の顧客に対応するようにした架空のデータ・ファイルです。そのため、週ごとに変っていた変数の一部が表示されなくなり、記録された購入額が、調査を行った 4 週間の購入額の合計になっています。
- **grocery\_coupons.sav:** これは、顧客の購買習慣に関心を持っている食料雑貨店チェーンが収集した調査データが含まれている架空のデータ・ファイルです。各顧客を 4 週間にわたって追跡し、各ケースが個々の顧客の週に対応しています。その週の食料品購入金額など、顧客がいつどこで買物をするかに関する情報が記録されています。

7. Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

- **guttman.sav:** Bell<sup>9</sup> は、予想される社会グループを示す表を作成しました。Guttman<sup>10</sup> は、この表の一部を使用しました。この表では、社会的相互作用、グループへの帰属感、メンバーとの物理的な近接性、関係の形式化などを表す 5 つの変数が、7 つの理論上の社会グループと交差しています。このグループには、観衆 (フットボールの試合の観戦者など)、視聴者 (映画館や授業の参加者など)、公衆 (新聞やテレビの読者や視聴者など)、群衆 (観衆に似ているが、より強い相互の関連がある)、第 1 次集団 (親密な関係)、第 2 次集団 (自発的な集団)、および近代コミュニティー (物理的に密接した近接性と特化されたサービスの必要性による緩い同盟関係) があります。
- **health\_funding.sav:** これは、医療用資金 (人口 100 人あたりの金額)、罹患率 (人口 10,000 人あたりの人数)、医療サービス機関の受診率 (人口 10,000 人あたりの人数) に関するデータが含まれている架空のデータ・ファイルです。各ケースが、個別の都市を表しています。
- **hivassay.sav:** これは、HIV 感染を発見する迅速な分析方法を開発することを目的とした、ある製薬研究所の取り組みに関する架空のデータ・ファイルです。分析の結果は 8 段階の濃さの赤で表現され、色が濃いほど感染の可能性が高くなります。研究所では 2,000 件の血液サンプルに関して試験を行い、その半数が HIV に感染しており、残りの半分は感染していませんでした。
- **hourlywagedata.sav:** これは、さまざまな経験レベルを持つ管理職から現場担当までの看護師の時給に関する架空のデータ・ファイルです。
- **insurance\_claims.sav:** これは、不正請求の恐れがある疑わしい請求にフラグを立てるためのモデルを作成したいと考えているある保険会社に関する架空のデータ・ファイルです。各ケースが個々の請求を表しています。
- **insure.sav:** これは、10 年満期の生命保険契約に対し、顧客が請求を行うかどうかを示す危険因子を調査している保険会社に関する架空のデータ・ファイルです。データ・ファイルの各ケースは、年齢と性別が一致する、請求が記録された契約と記録されなかった契約のペアを表しています。
- **judges.sav:** これは、訓練を受けた審判 (および 1 人の熱心なファン) が、300 件の体操の演技に対して付けた得点に関する架空のデータ・ファイルです。各行が個々の演技を表しています。この審判たちは、同じ演技を見ました。
- **kinship\_dat.sav:** Rosenberg と Kim<sup>11</sup> は、15 種類の親族関係用語 (叔母、兄弟、いとこ、娘、父、孫娘、祖父、祖母、孫息子、母、甥、姪、姉妹、息子、叔父) の分析を行いました。Rosenberg と Kim は、大学生の 4 つのグループ (女性 2 組、男性 2 組) に、類似性に基づいて上記の用語を並べ替えるよう依頼しました。2 つのグループ (女性 1 組、男性 1 組) には、この最初の条件とは異なる条件に基づいて、2 回目の並べ替えをするように依頼しました。したがって、合計で 6 つの「ソース」が取得されたこととなります。各ソースは、15×15 の近接行列に対応しています。この近接行列のセルの数は、ソース内の人数から、ソース内でオブジェクトを分割した回数を引いたものと同じとなります。
- **kinship\_ini.sav:** このデータ・ファイルには、*kinship\_dat.sav* の 3 次元の解の初期布置が含まれています。
- **kinship\_var.sav:** このデータ・ファイルには、*gender* (性別)、*gener* (世代)、*degree* (親等) という独立変数が含まれています。これらの変数を使用して、*kinship\_dat.sav* の解のディメンションを解釈することができます。具体的には、これらの変数を使用して、解の空間をこれらの変数の線型結合に制限することができます。

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

11. Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

- **marketvalues.sav:** これは、1999 年から 2000 年までの、イリノイ州アルゴンキンの新興住宅地開発における住宅売上に関するデータ・ファイルです。これらの売上は、公開レコードの対象となっています。
- **nhis2000\_subset.sav:** National Health Interview Survey (NHIS) は、米国国民を対象とした人口ベースの大規模な調査です。全国の代表的な世帯サンプルで、対面式の調査が行われます。各世帯のメンバーについて、人口統計情報と、健康に関する行動と状態の観測値が収集されます。このデータ・ファイルには、2000 年の調査から取得された情報のサブセットが含まれています (National Center for Health Statistics. National Health Interview Survey, 2000。一般使用のデータ・ファイルとドキュメント。 [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/)。2003 年にアクセス)。
- **ozone.sav:** このデータには、残りの変数からオゾン濃度を予測するための、6 つの気象変数に対する 330 の観測値が含まれています。これまでの研究者<sup>12</sup>、<sup>13</sup> が、他の研究者と共に、これらの変数間に非線型性を確認しています。この場合、標準的な回帰アプローチは使用できません。
- **pain\_medication.sav:** この架空のデータ・ファイルには、慢性関節炎を治療する抗炎症薬の臨床試験の結果が含まれています。特に重要なのは、薬の効果が出るまでの時間と、既存の薬剤との比較です。
- **patient\_los.sav:** この架空のデータ・ファイルには、心筋梗塞 (MI、または「心臓発作」) の疑いで入院した患者の治療記録が含まれています。各ケースは個々の患者に対応し、入院に関連する多くの変数が記録されています。
- **patlos\_sample.sav:** この架空のデータ・ファイルには、心筋梗塞 (MI、または「心臓発作」) の治療中に血栓溶解剤を投薬された患者のサンプルの治療記録が含まれています。各ケースは個々の患者に対応し、入院に関連する多くの変数が記録されています。
- **poll\_cs.sav:** これは、市民の法案支持率を議会開会前に特定するための、世論調査員の取り組みに関する架空のデータ・ファイルです。各ケースは、登録されている有権者に対応しています。ケースごとに、有権者が居住している郡、町、区域が記録されています。
- **poll\_cs\_sample.sav:** この架空のデータ・ファイルには、*poll\_cs.sav* の有権者のサンプルが含まれています。このサンプルは、*poll\_csplan* 計画ファイルで指定されている計画に従って抽出され、このデータ・ファイルには包含確率とサンプル重み付けが記録されています。ただし、サンプリング計画では確率比例 (PPS) 法が使用されるため、結合選択確率が含まれているファイル (*poll\_jointprob.sav*) もあります。サンプルの抽出後、有権者の人口統計と法案に関する意見に対応する追加の変数が収集され、データ・ファイルに追加されました。
- **property\_assess.sav:** これは、限られたリソースで資産価値の評価を最新に保つための、郡の評価担当者の取り組みに関する架空のデータ・ファイルです。各ケースは、前年に郡内で売却された資産に対応しています。データ・ファイル内の各ケースには、その資産が存在する町、最後にその資産を訪問した評価担当者、その評価からの経過時間、そのときに行われた評価、その資産の売却価値が記録されています。
- **property\_assess\_cs.sav:** これは、限られたリソースで資産価値の評価を最新に保つための、州の評価担当者の取り組みに関する架空のデータ・ファイルです。各ケースは、州内の資産に対応しています。データ・ファイル内の各ケースには、その資産が存在する郡、町、区域、最後の評価からの経過時間、そのときに行われた評価が記録されています。
- **property\_assess\_cs\_sample.sav:** この架空のデータ・ファイルには、*property\_assess\_cs.sav* にリストされている資産のサンプルが含まれています。このサンプルは、*property\_assess\_csplan* 計画ファイルで指定されている計画に従って抽出され、このデータ・ファイルには包含確率とサンプル重み付けが記録されています。サンプルの抽出後、現在の価値 変数が収集され、データ・ファイルに追加されました。

12. Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

- **recidivism.sav:** これは、管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ・ファイルです。各ケースは元犯罪者に対応し、この犯罪者の人口統計情報、最初の犯罪の詳細、2 回目の逮捕までの期間 (初犯から 2 年以内の場合) が記録されています。
- **recidivism\_cs\_sample.sav:** これは、管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ・ファイルです。各ケースは、2003 年の 7 月中に最初の逮捕から釈放された元犯罪者に対応し、この犯罪者の人口統計情報、最初の犯罪の詳細、2 回目の逮捕のデータ (2006 年 6 月末日までに逮捕された場合) が記録されています。犯罪者は、*recidivism\_cs.csplan* で指定されたサンプリング計画に従って抽出された部門から選択されています。サンプリング計画では確率比例 (PPS) 法が使用されるため、結合選択確率が含まれているファイル (*recidivism\_cs\_jointprob.sav*) もあります。
- **rfm\_transactions.sav:** これは、購入日、購入品目、各取引の金額など、購買取引データが含まれている架空のデータ・ファイルです。
- **salesperformance.sav:** これは、2 つの新しい販売トレーニング・コースの評価に関する架空のデータ・ファイルです。60 人の従業員を 3 つのグループに分け、全員が標準のトレーニングを受けます。さらに、グループ 2 は技術的なトレーニングを受け、グループ 3 は実践的なチュートリアルを受けます。トレーニング・コースの最後にすべての従業員がテストを受け、そのスコアが記録されました。データ・ファイルの各ケースは個々の受講生を表し、その受講生が割り当てられたグループと、テストのスコアが記録されています。
- **satisf.sav:** これは、ある小売業者が 4 箇所の店舗で行った満足度調査に関する架空のデータ・ファイルです。合計で 582 人の顧客について調査が行われ、各ケースは 1 人の顧客からの回答を表しています。
- **screws.sav:** このデータ・ファイルには、ねじ、ボルト、ナット、鉋 (びょう) <sup>14</sup> の特性に関するデータが含まれています。
- **shampoo\_ph.sav:** これは、あるヘアケア製品工場での品質管理に関する架空のデータ・ファイルです。定期的に、6 つの異なる生産バッチが測定され、その pH が記録されます。目標の範囲は 4.5 から 5.5 です。
- **ships.sav:** これは、他の場所 <sup>15</sup> で表示および分析される、波が貨物船に与える損害に関するデータ・セットです。事故カウントは、船舶の種類、建造期間、サービス期間により、ポワゾン率で発生するものとしてモデル化することができます。各因子のクロス分類によって形成されたテーブルの各セルのサービス月数の集計により、危険にさらされる確率の値が得られます。
- **site.sav:** これは、業務拡大に向けて新たな敷地を選定することを目的とした、ある会社の取り組みに関する架空のデータ・ファイルです。2 人のコンサルタントを雇い、敷地を別々に評価させました。これらのコンサルタントは、詳細なレポートのほかに、各敷地を「良い」、「普通」、「悪い」のいずれかで集計しました。
- **smokers.sav:** このデータ・ファイルは、1998 年の National Household Survey of Drug Abuse から抽出した、アメリカの世帯の確率サンプルです (<http://dx.doi.org/10.3886/ICPSR02934>)。したがって、このデータ・ファイルを分析するには、人口の傾向を反映させるために、最初にデータの重み付けを行う必要があります。
- **stocks.sav:** この架空のデータ・ファイルには、1 年分の在庫の価格と量が含まれています。
- **stroke\_clean.sav:** この架空のデータ・ファイルには、Statistics Base Edition のプロシージャーを使用して整理した後の、医療データベースの状態が含まれています。
- **stroke\_invalid.sav:** この架空のデータ・ファイルには、医療データベースの初期状態と、いくつかのデータ入力エラーが含まれています。

14. Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

15. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

- **stroke\_survival**: これは、虚血性脳卒中後のリハビリ・プログラムを終えた後に数回の困難に直面した患者の生存時間に関する架空のデータ・ファイルです。脳卒中後の心筋梗塞、虚血性脳卒中、または出血性脳卒中の発症と、その発症時刻が記録されています。このサンプルには、脳卒中後に実施されたりハビリ・プログラムの最後まで生存した患者だけが含まれているため、サンプルの左側は切り捨てられています。
- **stroke\_valid.sav**: この架空のデータ・ファイルには、データの検証プロシージャーによる確認後の、医療データベースの状態が含まれています。異常ケースの可能性のあるケースも含まれています。
- **survey\_sample.sav**: このデータ・ファイルには、人口統計データや各種の態度指標などの調査データが含まれています。これは「1998 NORC General Social Survey」の変数のサブセットに基づいていますが、一部のデータ値が変更され、いくつかの架空の変数がデモの目的で追加されています。
- **tcm\_kpi.sav**. これは、ビジネスの週次重要業績評価指標の値が含まれている架空のデータ・ファイルです。同じ期間のいくつかの制御可能メトリックの週次データも含まれます。
- **tcm\_kpi\_upd.sav**. このデータ・ファイルは *tcm\_kpi.sav* と同様ですが、さらに 4 週間分の追加データが含まれています。
- **telco.sav**: これは、顧客ベースにおけるチャーン (契約/解約を繰り返す顧客の流動現象) を削減するための電気通信会社の取り組みに関する架空のデータ・ファイルです。各ケースが個々の顧客に対応し、人口統計やサービス利用状況などのさまざまな情報が記録されています。
- **telco\_extra.sav**: このデータ・ファイルは *telco.sav* データ・ファイルと似ていますが、「期間」変数と対数変換された顧客支出変数が削除され、標準化された対数変換後の顧客支出変数に置き換えられています。
- **telco\_missing.sav**: このデータ・ファイルは *telco.sav* データ・ファイルのサブセットですが、一部の人口統計データ値が欠損値で置き換えられています。
- **testmarket.sav**: この架空のデータ・ファイルは、ファースト・フード・チェーンでの新しいメニューの追加計画に関連しています。新製品をプロモーションするためのキャンペーンには 3 つの候補があるため、新メニューは、ランダムに選択されたマーケットのいくつかの場所で導入されます。場所ごとに別々のプロモーションを展開し、最初の 4 週間について、新メニューの週間売上高が記録されます。各ケースが、個々の場所と週に対応しています。
- **testmarket\_1month.sav**: この架空のデータ・ファイルは、*testmarket.sav* データ・ファイルの週ごとの売上を「ロールアップ」して、各ケースを個々の場所に対応させたものです。そのため、週ごとに変わっていた変数の一部が表示されなくなり、記録された売上高が、調査を行った 4 週間の売上高の合計になっています。
- **tree\_car.sav**: これは、人口統計と自動車購入価格のデータが含まれている架空のデータ・ファイルです。
- **tree\_credit.sav**: これは、人口統計と銀行ローン履歴のデータが含まれている架空のデータ・ファイルです。
- **tree\_missing\_data.sav**: これは、多数の欠損値とともに、人口統計と銀行ローン履歴のデータが含まれている架空のデータ・ファイルです。
- **tree\_score\_car.sav**: これは、人口統計と自動車購入価格のデータが含まれている架空のデータ・ファイルです。
- **tree\_textdata.sav**: これは、測定レベルと値ラベルを割り当てる前の変数のデフォルトの状態を示すことを主な目的とする変数が 2 つだけ含まれている単純なデータ・ファイルです。
- **tv-survey.sav**: これは、テレビ・スタジオで実施された、ヒットした番組の放送期間を延長するかどうかを検討する調査に関する架空のデータ・ファイルです。906 人の回答者に対して、さまざまな条件下でこの番組を視聴するかどうかを質問しました。各行は個々の回答者を表し、各列は個々の条件を表しています。

- **ulcer\_recurrence.sav:** このファイルには、潰瘍の再発を防ぐための 2 つの治療の有効性を比較するための調査から得られた情報の一部が含まれています。これは区間打ち切りデータの好例で、他の場所<sup>16</sup> で表示と分析が行われています。
- **ulcer\_recurrence\_recoded.sav:** このファイルでは、単に調査終了時のイベント確率ではなく調査の間隔ごとのイベント確率をモデル化できるように、*ulcer\_recurrence.sav* の情報が再編成されています。これは、他の場所<sup>17</sup> で表示と分析が行われています。
- **verd1985.sav:** このデータ・ファイルは、調査<sup>18</sup> に関連しています。8 つの変数に対する 15 人の被験者の回答が記録されています。対象となる変数は 3 つのグループに分類されています。グループ 1 には「年齢」と「婚姻」、グループ 2 には「ペット」と「ニュース」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれています。「ペット」は多重名義として尺度化され、「年齢」は順序として尺度化されます。他のすべての変数は、単一名義として尺度化されます。
- **virus.sav:** これは、自社のネットワーク上のウィルスの影響を特定するための、インターネット・サービス・プロバイダー (ISP) の取り組みに関する架空のデータ・ファイルです。ネットワーク上の感染した E メール・トラフィックの (およその) パーセンテージが、検出時点から脅威が阻止されるまでの間、追跡されています。
- **wheeze\_steubenville.sav:** これは、子供に対する大気汚染の健康上の影響に関する長期的な調査<sup>19</sup> から得られたサブセットです。このデータには、オハイオ州スチューベンビルに住む 7 歳、8 歳、9 歳、10 歳の子供を対象に行った喘鳴の状態の反復 2 値測定と、調査の初年に母親が喫煙していたかどうかの固定的な記録が含まれています。
- **workprog.sav:** これは、体の不自由な人をより条件のよい仕事に就かせようとする政府の事業プログラムに関する架空のデータ・ファイルです。このプログラムの参加候補者のサンプルが追跡され、その中には、ランダムに選択されてプログラムに参加した人と、そうでない人がいます。各ケースが、個々のプログラム参加者を表しています。
- **worldsales.sav:** この架空のデータ・ファイルには、大陸別と製品別の販売収益が含まれています。

---

16. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

---

## 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。この資料の他の言語版を IBM から入手できる場合があります。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、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を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。

© IBM 2019. 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。

© Copyright IBM Corp. 1989 - 2019. All rights reserved.

---

## 商標

IBM、IBM ロゴおよび [ibm.com](http://www.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。

## [カ行]

- カスタム・テーブル
  - カテゴリの除外 7
  - カテゴリの並べ替え 7
  - カテゴリ変数 1
  - カテゴリ変数の値ラベル 1
  - 計算対象カテゴリ 7
  - 検定統計量 7
  - 合計 7
  - 小計 7
  - スケール変数 1
  - 測定レベルの変更 1
  - 多重回答セット 1
  - 多重回答セットのパーセント 5
  - テーブルの作成 3
  - パーセント 4, 5
  - 表示形式 3
  - 表示する小数桁数の制御 3
  - ファイルの分割処理 3
  - 要約統計量 4, 5
- カテゴリの削除
  - カスタム・テーブル 7
- カテゴリの除外
  - カスタム・テーブル 7
- カテゴリの並べ替え
  - カスタム・テーブル 7
- 検定統計量
  - カスタム・テーブル 7
- 合計
  - カスタム・テーブル 5, 7

## [サ行]

- 最小
  - カスタム・テーブル 5
- 最大
  - カスタム・テーブル 5
- 最頻値
  - カスタム・テーブル 5
- サンプル・ファイル
  - 場所 9
- 小計
  - カスタム・テーブル 7

- 小数桁数
  - カスタム・テーブルで表示する小数桁数の制御 3
- 測定レベル
  - カスタム・テーブルでの変更 1

## [タ行]

- 多重回答セット
  - パーセント 5
- 中央値
  - カスタム・テーブル 5
- テーブル
  - カスタム・テーブル 1

## [ハ行]

- パーセント
  - カスタム・テーブル 4, 5
  - 多重回答セット 5
- 範囲
  - カスタム・テーブル 5
- 標準偏差
  - カスタム・テーブル 5
- ファイルの分割処理
  - カスタム・テーブル 3
- 分散
  - カスタム・テーブル 5
- 平均値
  - カスタム・テーブル 5

## [ヤ行]

- 有意差検定
  - カスタム・テーブル 7
- 有効 N
  - カスタム・テーブル 5







Printed in Japan