IBM SPSS Statistics 28 Brief Guide

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、83 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 28 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。

© Copyright International Business Machines Corporation .

目次

第 1 章概要 1
サンプル・ファイル 1
データ・ファイルを開く 1
分析の実行 2
図表の作成 4

第 2 章データの読み取り 7
IBM SPSS Statistics データ・ファイルの基本構造 7
IBM SPSS Statistics データ・ファイルの読み取り 7
Excel データの読み込み 8
データベースからのデータの読み取り 11
テキスト・ファイルからのデータの読み取り 13

第 3 章データ・エディターの使用 17
数値データの入力 17
文字列データの入力 18
データの定義 19
変数ラベルの追加 19
変数の型と書式の変更 20
値ラベルの追加 20
欠損データの処理 21
数値型変数の欠損値 21
文字列変数の欠損値 22

第 4 章個々の変数に対する要約統計量の検討 23
測定レベル 23
カテゴリー・データの要約統計量 23
カテゴリー・データのグラフ 24
スケール変数の要約統計量 25
スケール変数のヒストグラム 26

第 5 章グラフの作成および編集 27
図表作成の基本 27
図表ビルダー・ギャラリーの使用 27
変数および統計量の定義 28
テキストの追加 29
図表の作成 30

第 6 章出力を使用した作業 33
ビューアーの使用 33
ピボット・テーブル・エディターの使用 34
出力定義へのアクセス 34
ピボット・テーブル 35
層の作成と表示 36
テーブルの編集 37
行および列の非表示 37
データの表示書式の変更 38
テーブルルック 39
定義済み書式の使用 39
テーブルルック・スタイルのカスタマイズ 40

デフォルトのテーブル書式の変更 43
初期表示設定のカスタマイズ 43
変数ラベルおよび値ラベルの表示 44
その他のアプリケーションでの結果の使用 45
Word の表としての結果の貼り付け 46
テキストとしての結果の貼り付け 46
Microsoft Word、PowerPoint、および Excel ファイルへの結果のエクスポート 47
結果の PDF へのエクスポート 52
結果の HTML へのエクスポート 54

第 7 章シンタックスの使用 57
シンタックスを貼り付ける 57
シンタックスを編集する 58
シンタックス・ファイルを開いて実行する 59
ブレークポイントの使用 59

第 8 章データ値の変更 61
スケール変数からカテゴリー変数を作成 61
新しい変数の計算 63
式での関数の使用 64
条件式の使用 65
日付と時刻の使用 65
2 つの日付間の時間の長さの計算 66
日付への期間の加算 67

第 9 章データのソートおよび選択 69
データのソート 69
ファイルの分割処理 69
ファイル分割処理のケースのソート 71
ファイルの分割処理をオンとオフを切り替える 71
ケースのサブセットの選択 71
条件式に基づいたケースの選択 72
無作為抽出の選択 73
時間範囲またはケース範囲の選択 73
未選択のケースの処理 74
ケース選択の状況 74

第 10 章サンプル・ファイル 75

特記事項 83
商標 84

索引 85

第 1 章 概要

このガイドでは、多くの使用可能な機能の使用方法を示します。 このガイドは、ステップバイステップの実践的なガイドとなるように編集されています。例で示されているすべてのファイルは、アプリケーションと一緒にインストールされるため、ここに示す同じ分析を実行し、同じ結果を取得しながら、続けることができます。

さまざまな統計分析手法の詳細な例が必要な場合は、「ヘルプ」メニューから使用可能な、ステップバイステップのケース・スタディーを試してみてください。

サンプル・ファイル

ここで示すほとんどの例では、データ・ファイル *demo.sav* を使用しています。このデータ・ファイルは数千人分の架空の調査で、基本的な人口統計および消費者情報を含んでいます。

Student 版を使用している場合、*demo.sav* のバージョンは、1,500 ケースの制限に対応するように縮小された、元のデータ・ファイルの代表的なサンプルになります。 そのデータ・ファイルを使用して得た結果は、ここに示す結果とは異なります。

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。

データ・ファイルを開く

データ・ファイルを開くには、次の手順を実行します。

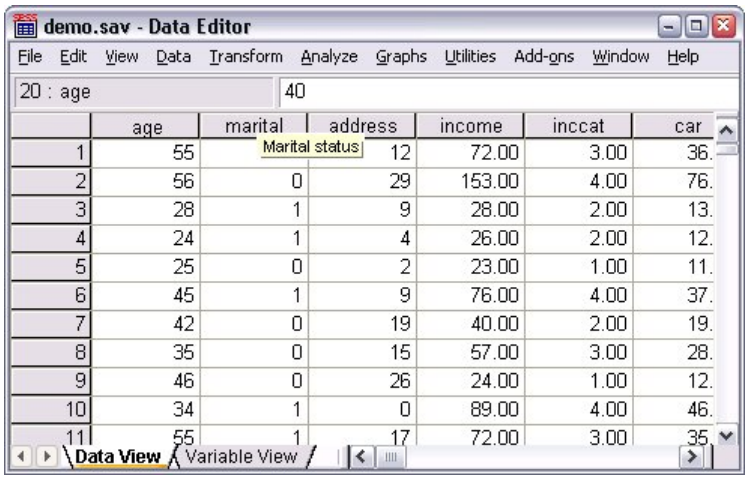
1. メニューから次の項目を選択します。

   「**ファイル**」 > 「**開く**」 > 「**データ...**」

   ファイルを開くダイアログ・ボックスが表示されます。

   デフォルトで、IBM SPSS Statistics データ・ファイル (*.sav* 拡張子) が表示されます。

   この例では、ファイル *demo.sav* を使用しています。



	age	marital	address	income	inccat	car
1	55		12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

図 *1.* データ・エディター内の *demo.sav* ファイル

データ・エディターにデータ・ファイルが表示されます。データ・ビューで、変数名 (列見出し) の上にマウス・カーソルを置くと、よりわかりやすい変数ラベルが表示されます (その変数にラベルが定義されている場合)。

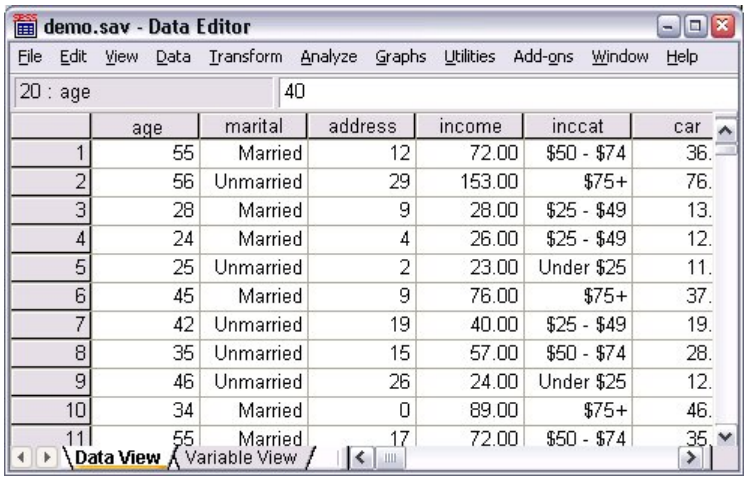
デフォルトでは、実際のデータ値が表示されます。ラベルを表示するには、次の手順を実行します。

2. メニューから次の項目を選択します。

「**表示**」 > 「**値ラベル**」

図 *2.* 「値ラベル」ボタン

または、ツールバーの 「値ラベル」ボタンを使用することもできます。



	age	marital	address	income	inccat	car
1	55	Married	12	72.00	$50 - $74	36.
2	56	Unmarried	29	153.00	$75+	76.
3	28	Married	9	28.00	$25 - $49	13.
4	24	Married	4	26.00	$25 - $49	12.
5	25	Unmarried	2	23.00	Under $25	11.
6	45	Married	9	76.00	$75+	37.
7	42	Unmarried	19	40.00	$25 - $49	19.
8	35	Unmarried	15	57.00	$50 - $74	28.
9	46	Unmarried	26	24.00	Under $25	12.
10	34	Married	0	89.00	$75+	46.
11	55	Married	17	72.00	$50 - $74	35.

図 *3.* 「データ・エディター」に表示される値ラベル

これで、わかりやすい値ラベルが表示されるので、回答が解釈しやすくなります。

分析の実行

アドオン・オプションがある場合、「分析」メニューには、レポートおよび統計分析カテゴリーのリストが表示されます。

簡単な度数分布表 (度数のテーブル) の作成から始めます。この例では Statistics Base オプションが必要です。

1. メニューから次の項目を選択します。

   「**分析**」 > 「**記述統計**」 > 「**度数分布表...**」

   「度数分布表」ダイアログ・ボックスが表示されます。

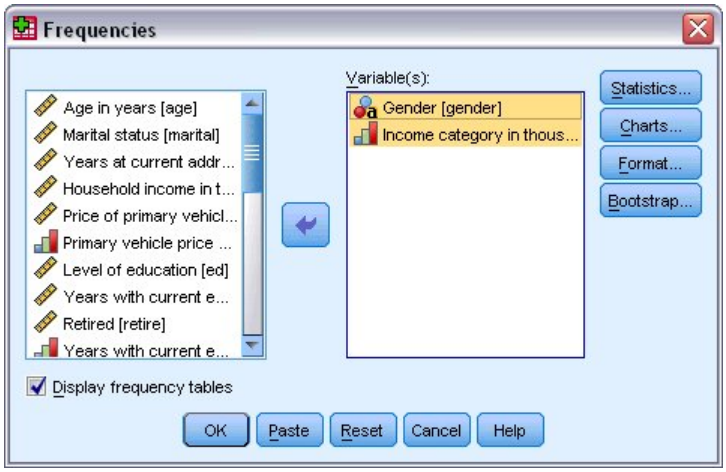


図 *4.* 「度数分布表」ダイアログ・ボックス

各変数の隣にあるアイコンは、データ型と測定のレベルに関する情報を示します。

	数値	**文字列**	**日付**	**時刻**
スケール (連続)		なし		
順序				
名義				

リストで変数ラベルや名前が切り捨てられているように見える場合、その上にカーソルを置くと、完全なラベルまたは名前が表示されます。説明の変数ラベルの後に、大括弧で囲まれた変数名 *inccat* が表示されます。「世帯全体の収入カテゴリー *(*千ドル*)*」が変数ラベルです。変数ラベルがない場合は、変数名だけがリスト・ボックスに表示されます。

ダイアログ・ボックスは、ウィンドウとまったく同じように、外枠や角をクリックしてドラッグするとサイズ変更できます。例えば、ダイアログ・ボックスの幅を大きくすると、変数リストの幅も大きくなります。

ダイアログ・ボックスで、左側のソース・リストから分析する変数を選択して、それらを右側の「変数」リストにドラッグ・アンド・ドロップします。「変数」リストに少なくとも 1 つの変数が配置されるまで、分析を実行する「**OK**」ボタンは無効です。

多くのダイアログで、リストの任意の変数名を右クリックし、ポップアップ・メニューから「**変数情報**」を選択して、追加情報を取得できます。

2. ソース変数リストの「性別 *[*性別*]* 」をクリックし、変数を目標の「変数」リストにドラッグします。
3. ソース・リストの「世帯全体の収入カテゴリー *(*千ドル*) [*収入カテゴリー*]*」をクリックし、目的リストにドラッグします。

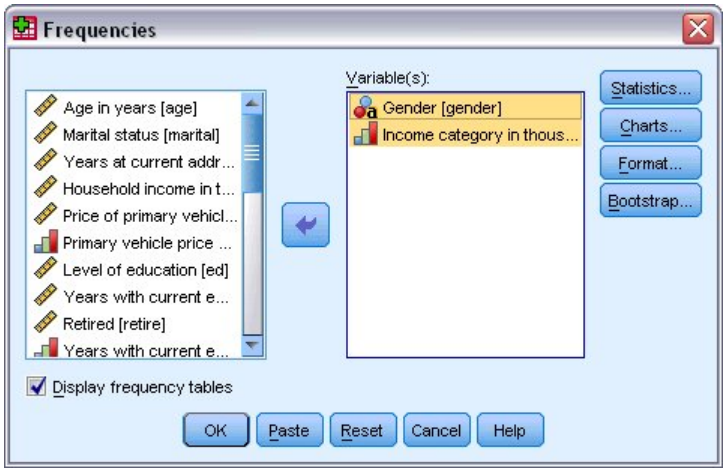


図 5. 分析に選択された変数

4.「**OK**」をクリックし、プロシージャーを実行します。

結果が「ビューアー」ウィンドウに表示されます。

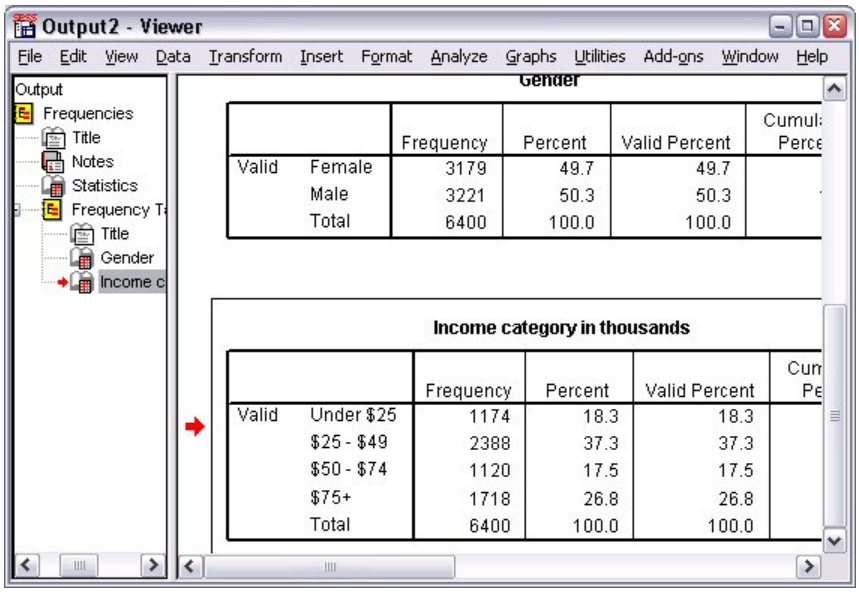


図 6. 収入カテゴリーの度数分布表

図表の作成

一部の統計手続きでは図表を作成できますが、「グラフ」メニューを使用して図表を作成することもできます。

例えば、ワイヤレス電話サービスと PDA (携帯情報端末) の所有の関係を示す図表を作成できます。

1. メニューから次の項目を選択します。

   「**グラフ**」 > 「**図表ビルダー...**」

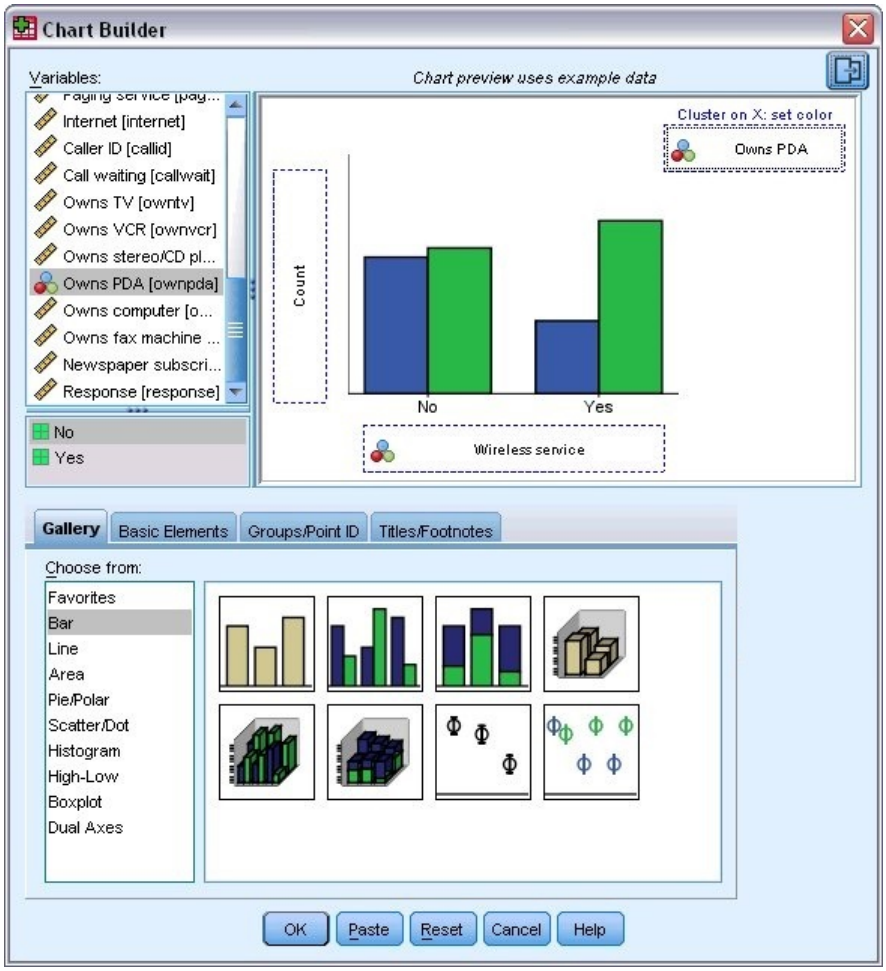


図 7. ドロップ・ゾーンがすべて設定された「図表ビルダー」ダイアログ・ボックス

2. 「**ギャラリー**」タブが選択されていない場合は、クリックします。
3. 「**棒グラフ**」が選択されていない場合は、クリックします。
4. 「クラスター棒グラフ」アイコンをキャンバスにドラッグします。キャンバスとは、ギャラリーの上にある広い領域です。
5. 「変数」リストを下にスクロールし、「ワイヤレス・サービス *[ワイヤレス]*」を右クリックし、測定レベルとして「**名義**」を選択します。
6. 「ワイヤレス・サービス *[ワイヤレス]*」変数を *x* 軸にドラッグします。
7. 「携帯情報端末 *[PDA]*」を右クリックし、測定レベルとして、「**名義**」を選択します。
8. 「携帯情報端末 *[PDA]*」変数をキャンバスの右上隅のクラスター・ドロップ・ゾーンにドラッグします。
9. 「**OK**」 をクリックして、図表を作成します。

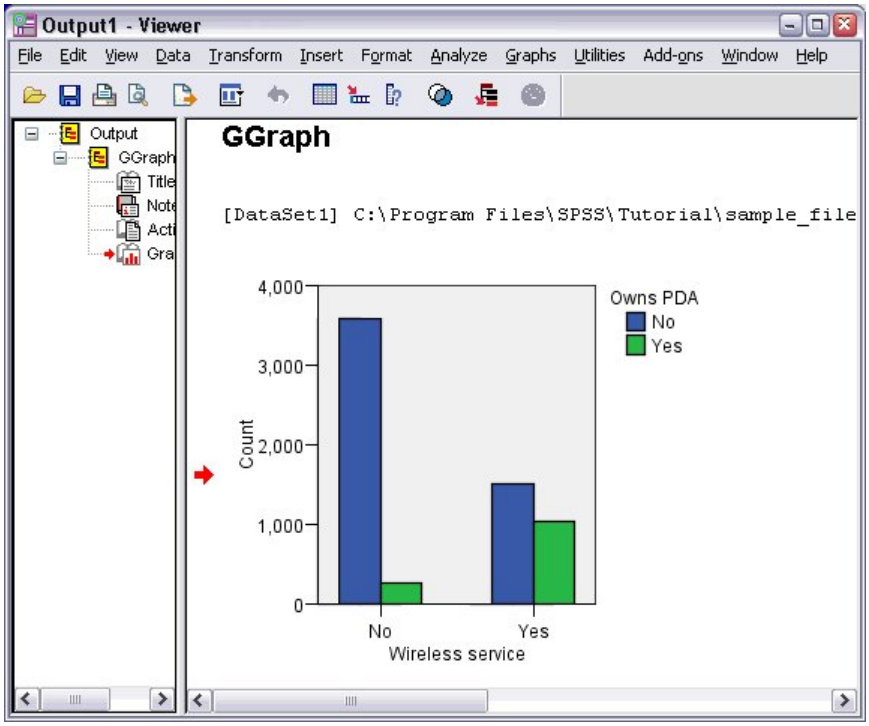


図 *8.* 「ビューアー」ウィンドウに表示された棒グラフ

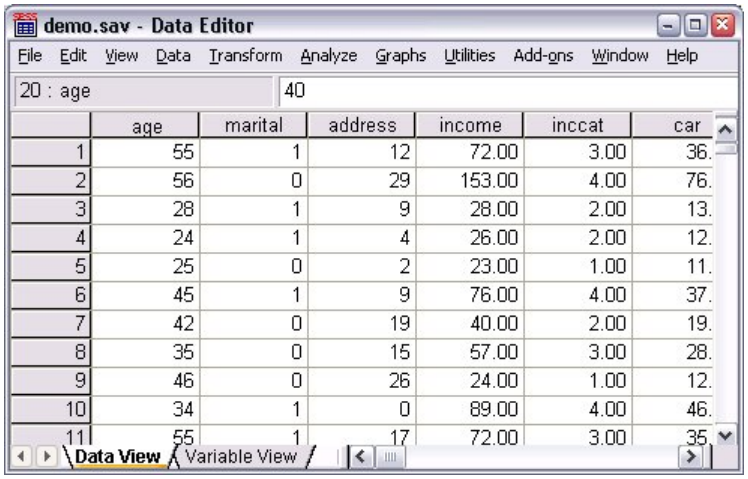
「ビューアー」に棒グラフが表示されます。この図表は、ワイヤレス電話サービスを使用する人は、ワイヤレス・サービスを使用しない人よりも、PDA を持っている可能性がかなり高いことを示しています。

「ビューアー」ウィンドウの内容のウィンドウの図表やテーブルをダブルクリックして、それらを編集したり、結果を他のアプリケーションにコピーして貼り付けることができます。それらのトピックについては、後で説明します。

第 2 章 データの読み取り

データは、直接入力することも、さまざまなソースからインポートすることもできます。この章では、IBM SPSS Statistics データ・ファイル、Microsoft Excel などのスプレッドシート・アプリケーション、Microsoft Access などのデータベース・アプリケーション、およびテキスト・ファイルに保存されたデータを読み取るあらゆる手順を説明します。

IBM SPSS Statistics データ・ファイルの基本構造



	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

図 *9.* データ・エディター

IBM SPSS Statistics データ・ファイルは、ケース (行) と変数 (列) で構成されています。 このデータ・ファイルで、ケースとは調査の個々の回答を指します。 変数は、調査に使用した個々の質問への回答を指します。

IBM SPSS Statistics データ・ファイルの読み取り

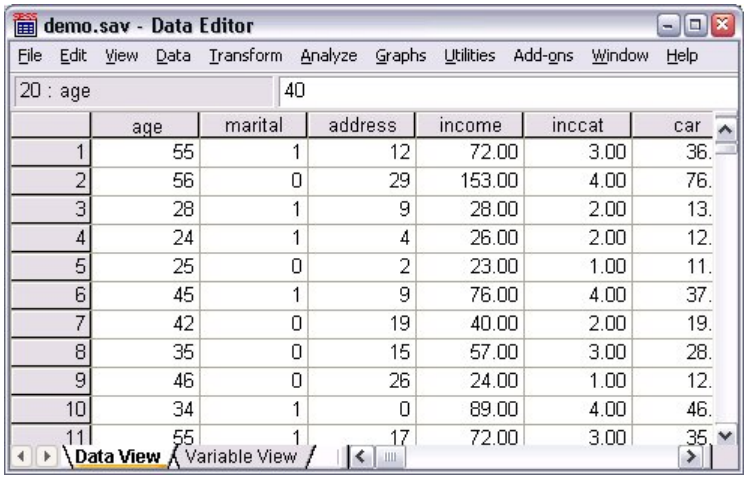
保存したデータ・ファイルは、IBM SPSS Statistics データ・ファイル (拡張子は *.sav*) に格納されます。

1. メニューから次の項目を選択します。

   「**ファイル**」 > 「**開く**」 > 「**データ...**」

2. *demo.sav* を参照して開きます。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

データ・エディターにデータが表示されます。



	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

図 *10.* 開かれたデータ・ファイル

Excel データの読み込み

データ・エディターにすべてのデータを直接入力するのでなく、Microsoft Excel などのアプリケーションからデータを読み取ることができます。列見出しを変数名として読み取ることもできます。

1. メニューから次の項目を選択します。

   「**ファイル**」 > 「**データのインポート**」 > 「**Excel**」

2. `Samples¥English` フォルダーに移動し、`demo.xlsx` を選択します。

   「**Excel ファイルの読み込み**」ダイアログに、このデータ・ファイルのプレビューが表示されます。ファイルの最初のシートの内容が表示されます。ファイルに複数のシートがある場合は、リストからシートを選択することができます。

   Gender の一部の文字列値に、先行スペースがあることを確認できます。*MaritalStatus* の一部の値は、ピリオド (.) として表示されています。

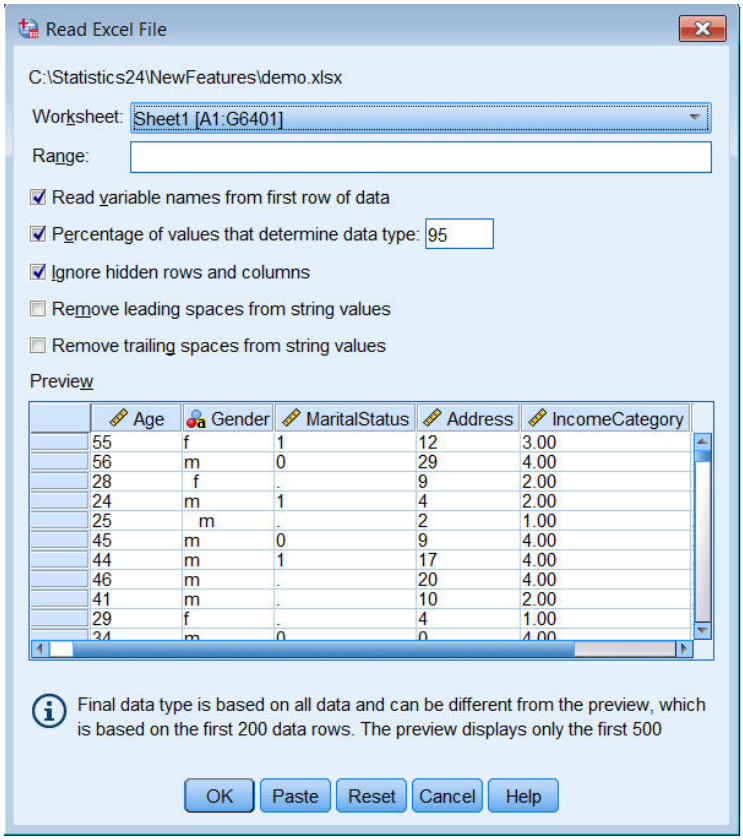
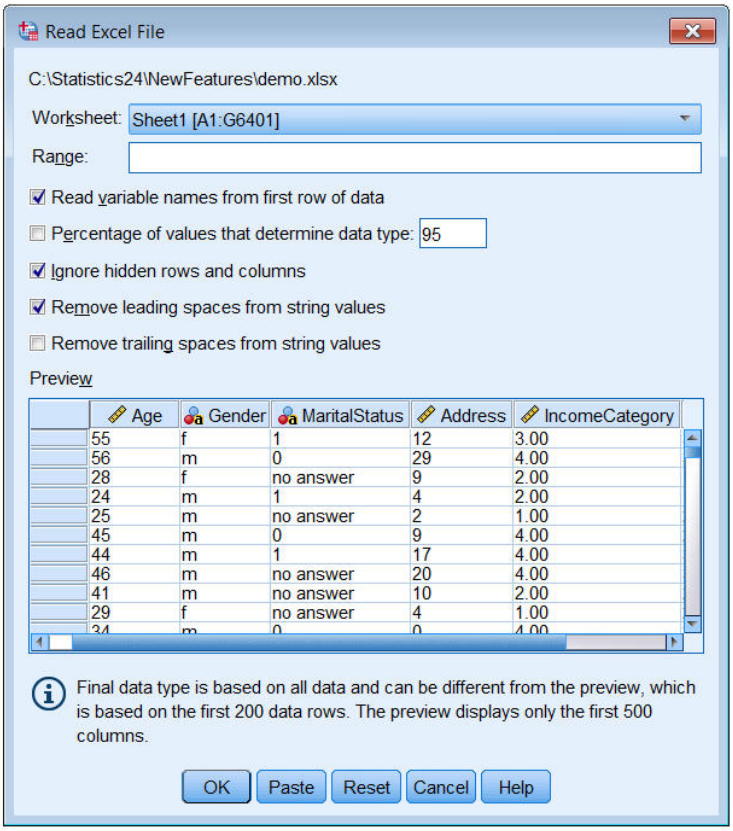


図 *11.* 「*Excel* ファイルの読み込み」ダイアログ

3. 「**データの最初の行から変数名を読み込む**」が選択されていることを確認します。変数名規則に従っていない列見出しは、有効な変数名に変換されます。元の列見出しは、変数ラベルとして保存されます。
4. 「**文字列値から先行スペースを削除**」を選択します。
5. 「**データ型を判定する値のパーセンテージ**」を選択解除します。



システム欠損だった文字列値「no answer」がセルに表示されるようになりました。「値のパーセンテージ」パラメーターがなく、列にデータ・タイプが混在して含まれている場合、変数は文字列データ・タイプとして読み込まれます。すべての値が保持されますが、数値は文字列値として扱われます。

6. *MaritalStatus* を数値変数として扱うには、「**データ型を判定する値のパーセンテージ**」を選択 (チェック) します。
7. 「**OK**」をクリックして、Excel ファイルを読み取ります。

データ・エディターにデータが表示されます。列見出しは変数名として使用されています。 変数名にスペースを含めることはできないので、元の列見出しに含まれていたスペースは取り除かれます。例えば、列見出し「Marital Status」は変数 *MaritalStatus* に変換されます。元の列見出しは、変数ラベルとして保持されます。

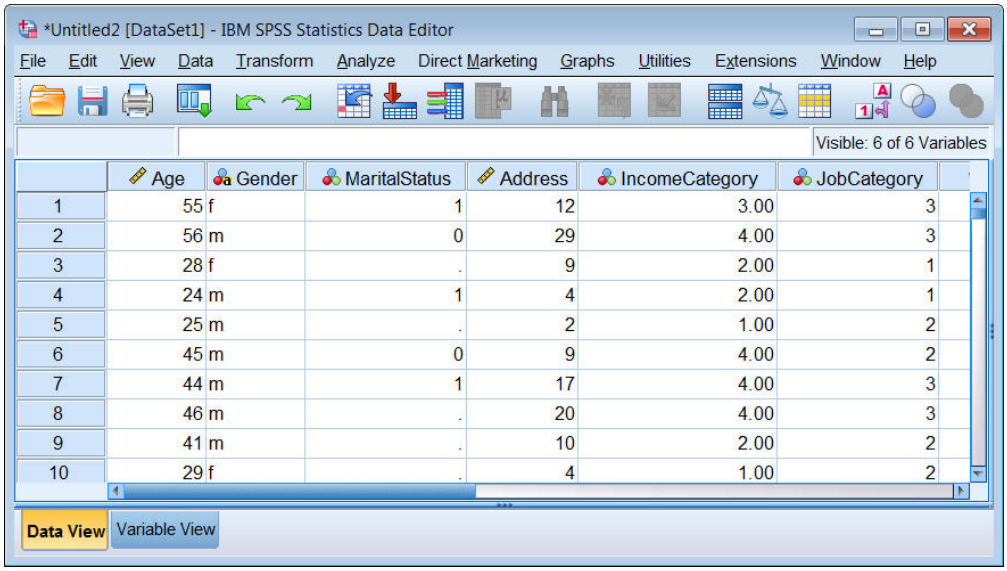


図 *12. インポートした Excel データ*

関連情報

75 ページの『サンプル・ファイル』

データベースからのデータの読み取り

データベース・ウィザードを使用すると、データベース・ソースからデータを簡単にインポートできます。ODBC (Open Database Connectivity) ドライバーを使用するデータベースは、ドライバーをインストールすれば直接読み取ることができます。インストール CD に、多数のデータベース形式に対応した ODBC ドライバーが用意されています。その他のドライバーは、サード・パーティー・ベンダーから入手できます。ここでは、広く普及しているデータベース・アプリケーションの 1 つである Microsoft Access を例に取って説明します。

注: これは Microsoft Windows に固有の例であり、Access 用の ODBC ドライバーが必要です。Microsoft Access ODBC ドライバは、32 ビット版の IBM SPSS Statistics でのみ動作します。 手順は他のプラットフォームの場合と同様ですが、サード・パーティー製の Access 用 ODBC ドライバーが必要な場合があります。

1. メニューから次の項目を選択します。

   「**ファイル**」 > 「**データのインポート**」 > 「**データベース**」 > 「**新規クエリー...**」

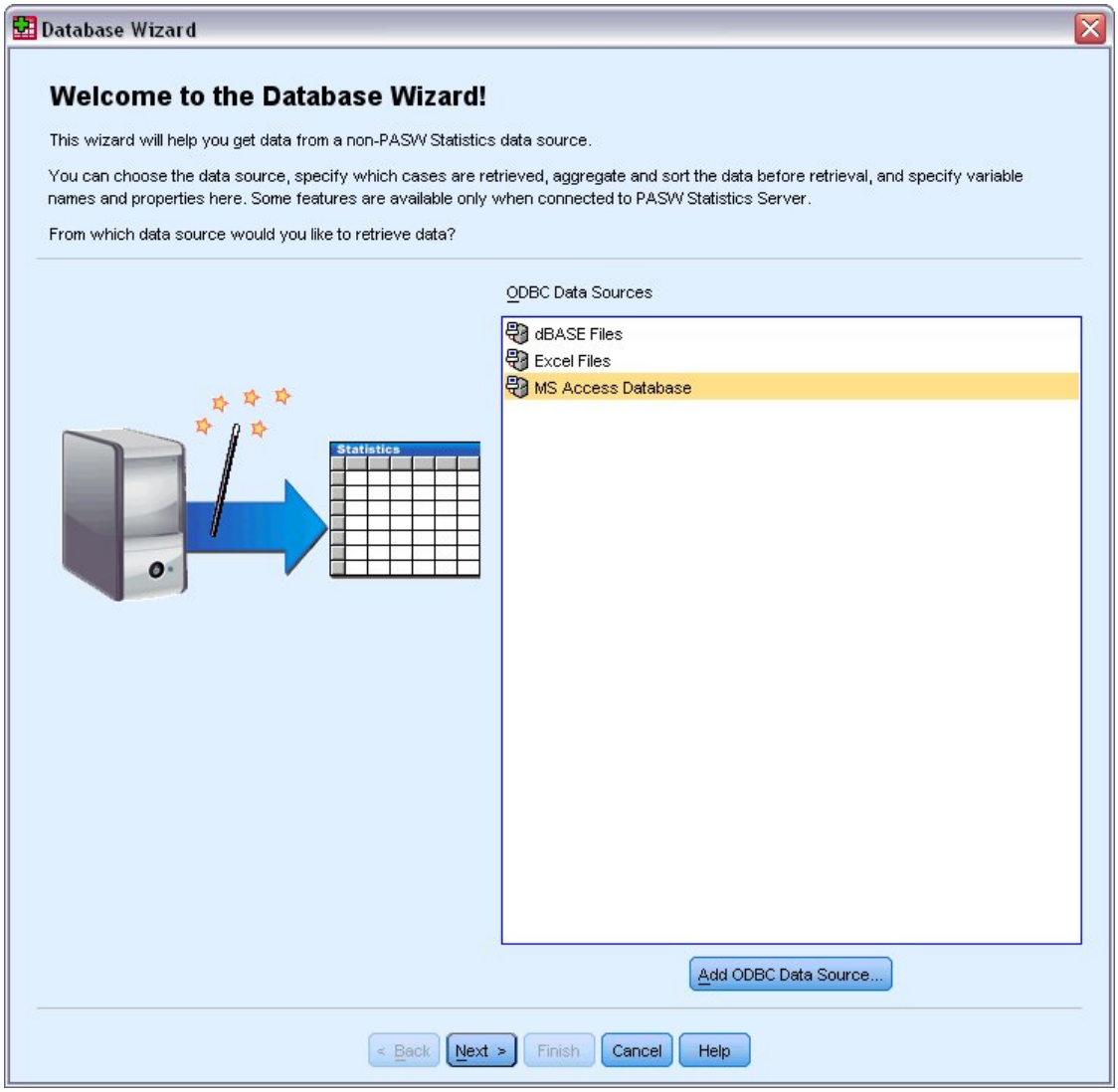


図 *13.* 「データベース・ウィザードにようこそ」ダイアログ・ボックス

2. データ・ソースのリストから「**MS Access Database**」を選択し、「**次へ**」をクリックします。

注: インストールの状況によっては、ウィザードの左側に OLEDB データ・ソースのリストが表示されることもあります (Windows オペレーティング・システムのみ)。ただし、この例では、右側に表示される ODBC データ・ソースのリストを使用します。

3. 「**参照**」をクリックし、開く Access データベース・ファイルに移動します。
4. *demo.mdb* を開きます。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。
5. 「ログイン」ダイアログ・ボックスで「**OK**」をクリックします。

次のステップでは、インポートするテーブルと変数を指定します。

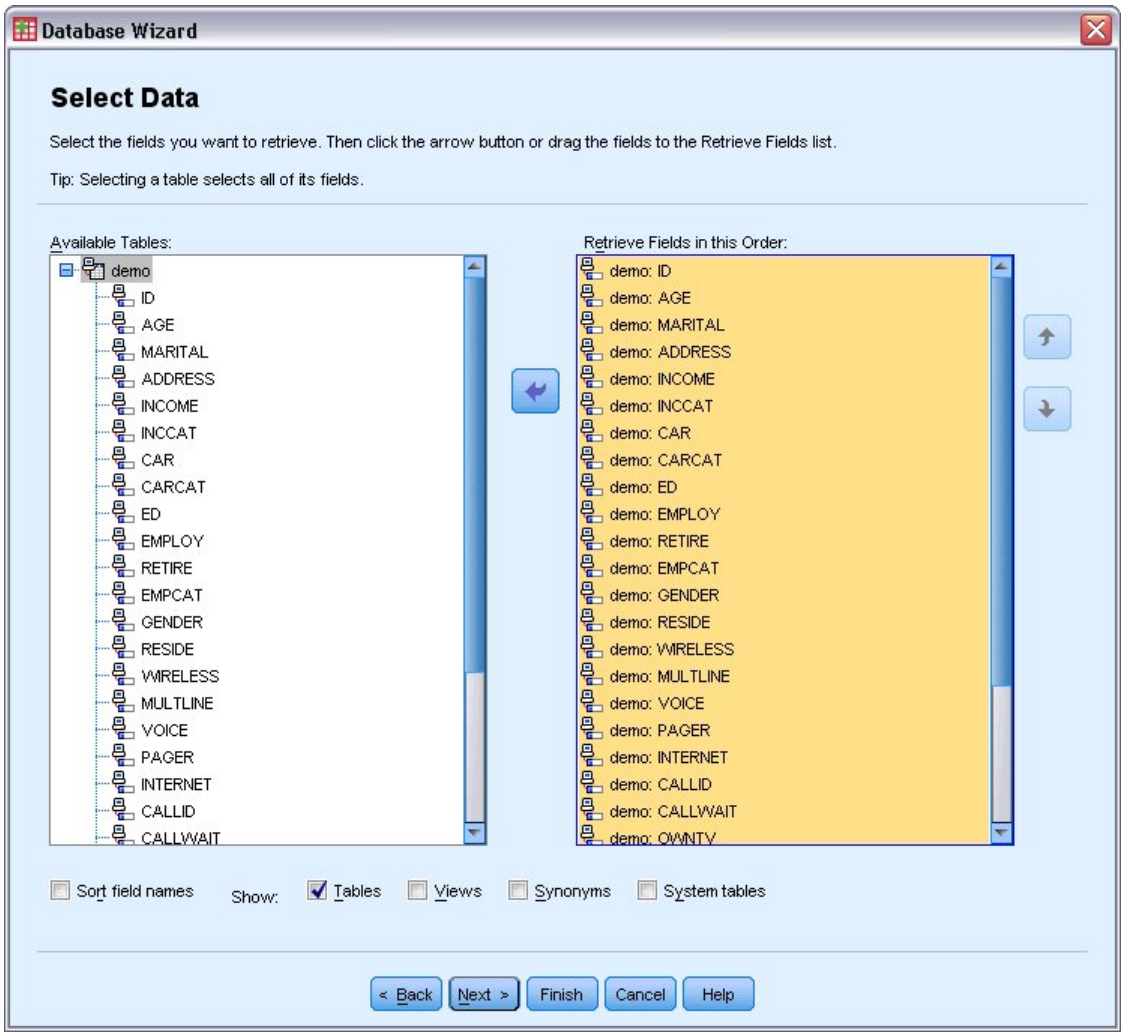


図 *14.* 「データの選択」ステップ

6. 「**demo**」テーブル全体を「読み込むフィールドの順序」リストへドラッグします。
7. 「**次へ**」をクリックします。

次のステップでは、インポートするレコード (ケース) を選択します。

すべてのケースをインポートするわけではない場合は、ケースのサブセット (例えば、31 歳以上の男性) をインポートするか、データ・ソースからケースを無作為に抽出してインポートすることができます。データ・ソースが大きい場合は、ケース数を代表的な小さなサンプルに限定することで、処理時間を短縮することができます。

8. 「**次へ**」をクリックして先へ進みます。

フィールド名を使用して変数名が作成されます。必要に応じて、フィールド名は有効な変数名に変換されます。元のフィールド名は、変数ラベルとして保持されます。データベースをインポートする前に、変数名を変更することもできます。

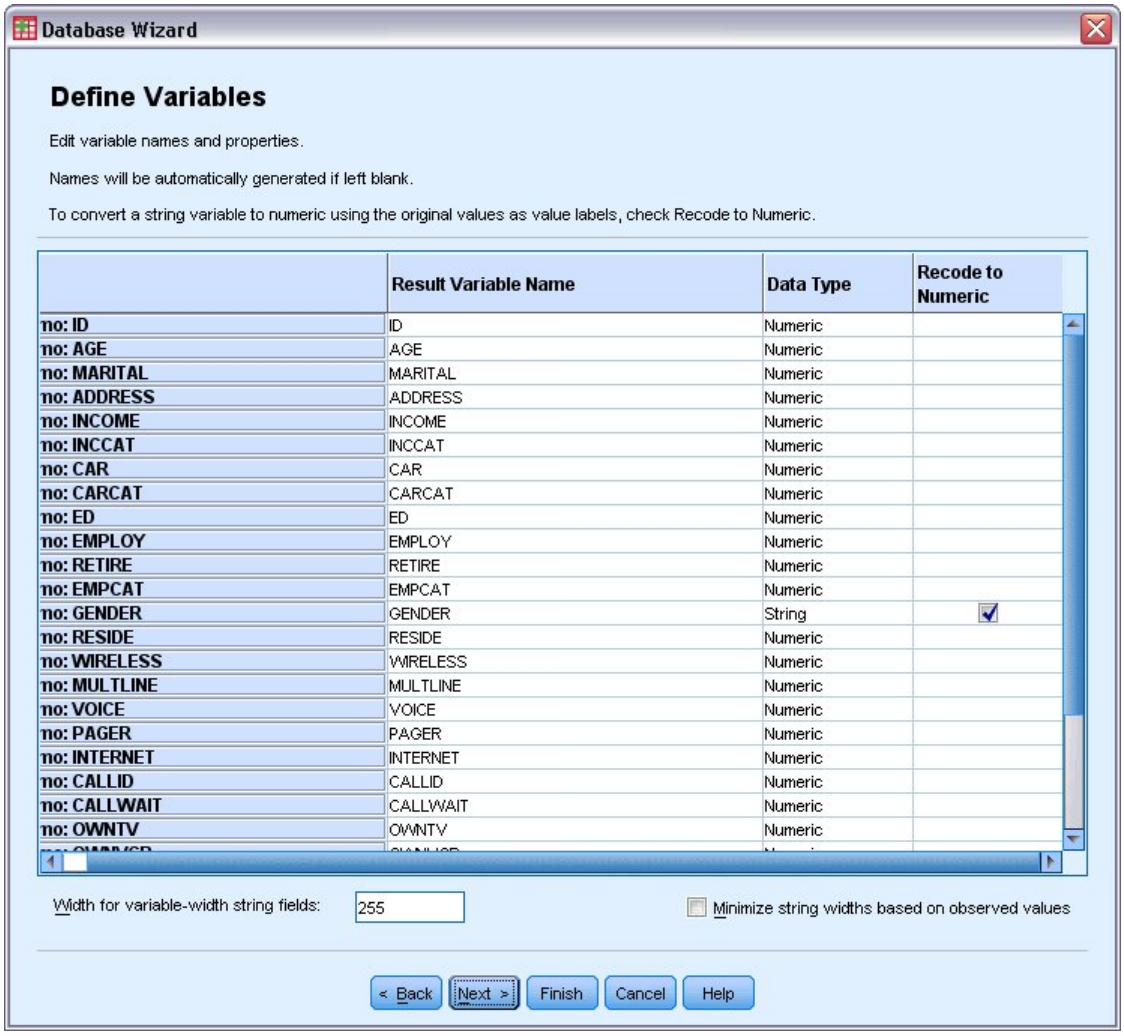


図 *15.* 「変数の定義」ステップ

9. 「性別」フィールドの「**数値への値の再割り当て**」セルをクリックします。このオプションを選択すると、文字列変数が整数型変数に変換され、元の値は新しい変数の値ラベルとして保持されます。
10. 「**次へ**」をクリックして先へ進みます。

    データベース・ウィザードでの選択を基に作成された SQL ステートメントが「結果」ステップに表示されます。このステートメントは、すぐに実行することも、ファイルに保存して後で実行することもできます。
11. 「**完了**」をクリックして、データをインポートします。

これで、インポート用に選択した Access データベース内のデータすべてが、データ・エディターで使用可能になりました。

テキスト・ファイルからのデータの読み取り

テキスト・ファイルは、一般的なデータ・ソースのもう 1 つの代表です。スプレッドシート・プログラムおよびデータベースの多くでは、その内容をさまざまなテキスト・ファイル形式のいずれかで保存することができます。コンマ区切りファイルまたはタブ区切りのファイルとは、各変数がコンマまたはタブで区切られたデータ行を指します。 この例では、データはタブ区切りです。

1. メニューから次の項目を選択します。

   「**ファイル**」 > 「**データのインポート**」 > 「**テキスト・データ**」
2. `Samples¥English` フォルダーに移動し、`demo.txt` を選択します。

テキスト・インポート・ウィザードのステップに従って、指定したテキスト・ファイルの解釈方法を定義します。

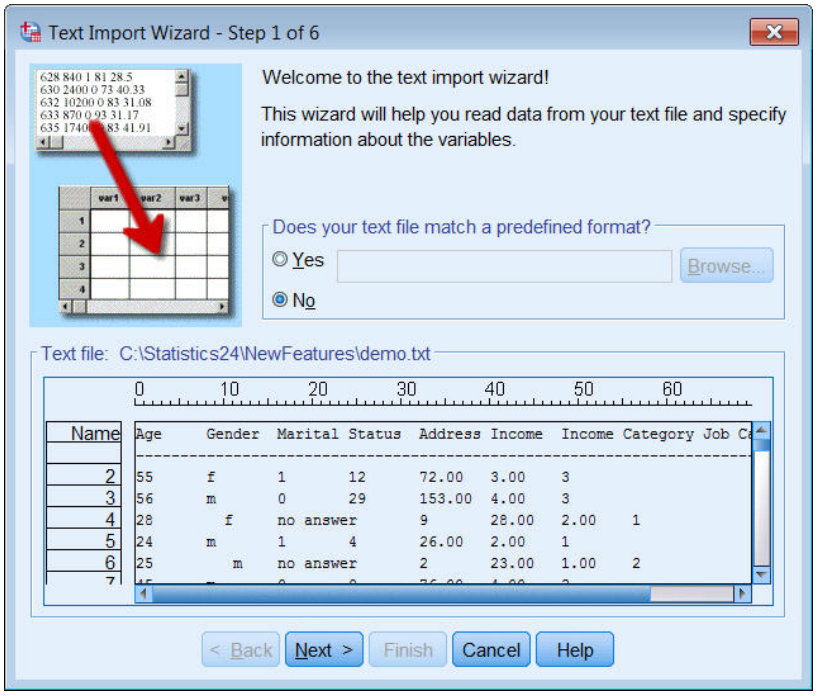


図 *16.* テキスト・インポート・ウィザード*:* ステップ *1/6*

3. ステップ 1 では、定義済みの形式を選択するか、ウィザードで新しい形式を作成できます。「**いいえ**」を選択します。
4. 「**次へ**」をクリックして先へ進みます。

   前述のとおり、このファイルの形式はタブ区切りです。また、変数名はこのファイルの最初の行で定義されています。
5. ウィザードのステップ 2 では、「**自由書式**」を選択して、ファイルが区切り形式の構造になっていることを示します。
6. 「**はい**」を選択して、変数名がファイルの先頭に含まれることを示します。
7. 「**次へ**」をクリックして先へ進みます。
8. ウィザードのステップ 3 では、データの最初のケースが始まる行番号として「2」を入力します (最初の行には変数名があるため)。
9. このステップのその他のオプションの値はデフォルトのままとします。「**次へ**」をクリックして先へ進みます。

   ステップ 4 のデータ・プレビューを見ると、ファイルが正しく読み取られていることを即座に確認できます。
10. 「**タブ**」を選択し、区切りのその他のオプションを選択解除します。ファイルにスペースが含まれているため、デフォルトで「**スペース**」が選択されています。このファイルの場合、スペースはデータ値の一部であり、区切り文字ではありません。ファイルを正しく読み取るには、「**スペース**」の選択を解除する必要があります。
11. 「**文字列値から先行スペースを削除**」を選択します。文字列値の最初にあるスペースは、文字列値が式で評価される方法に影響します。このファイルでは、*Gender* の一部の値に、値の一部ではない先行スペースがあります。これらのスペースを削除しない場合、「f」という値が「f」とは別の値として扱われます。

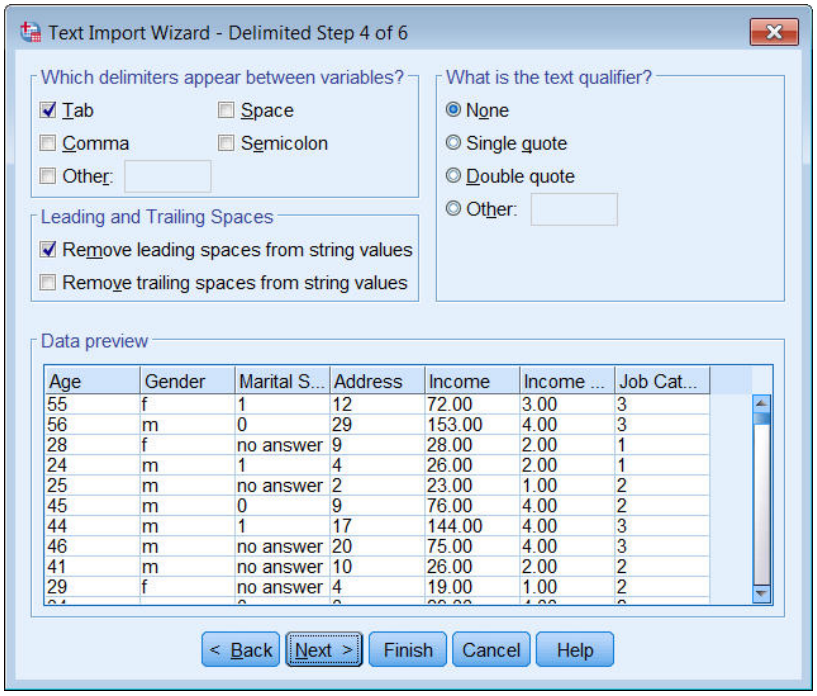


図 *17.* テキスト・インポート・ウィザード*:* ステップ *4/6*

12.「**次へ**」をクリックして先へ進みます。

変数名は命名規則に従うように変更されているため、ステップ 5 では望ましくない名前を編集することができます。

データ型も同様にここで定義できます。例えば、*Income* をドル通貨書式に変更できます。

データ型を変更するには、次の手順に従います。

13.「**データ プレビュー**」で *Income* を選択します。

14.「データ形式」ドロップダウン・リストから「**ドル記号**」を選択します。

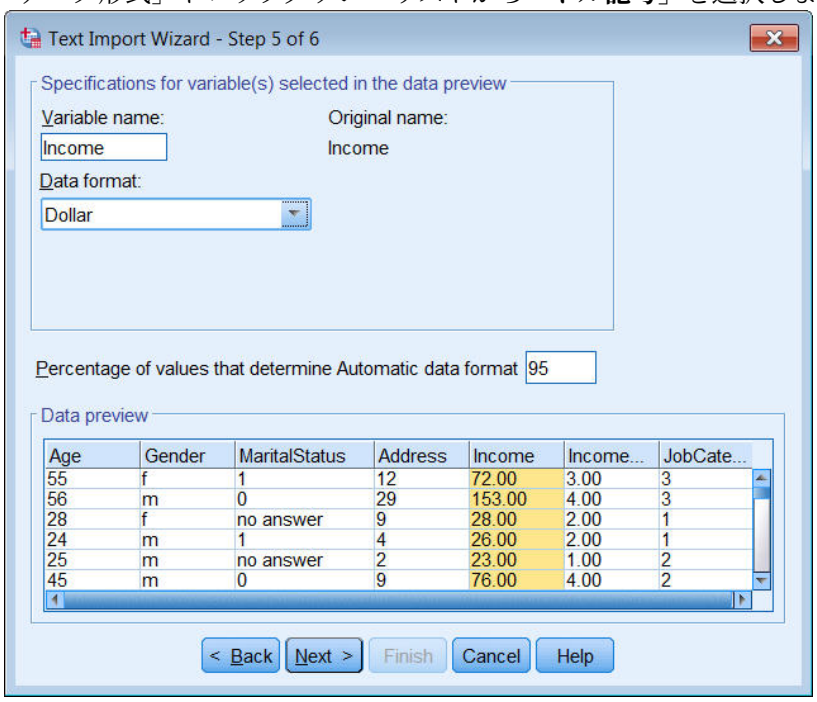


図 *18.* データ型の変更

変数 *MaritalStatus* には、文字列値と数値の両方が含まれます。値の 5 パーセント未満が文字列です。「**自動データ形式を判定する値のパーセンテージ**」のデフォルト設定が 95% であることで、この変数は数値として扱われ、文字列値はシステム欠損として設定されます。データ形式がパーセント値に一致しない場合、この変数は文字列変数として扱われます。設定を 100 に変更した場合、すべての値が保持されますが、数値はすべて文字列として扱われます。

15. 「**次へ**」をクリックして先へ進みます。
16. 最後のステップのオプションはすべてデフォルトの 選択のままとします。「**完了**」をクリックしてデータをインポートします。

第 3 章 データ・エディターの使用

データ・エディターには、アクティブなデータ・ファイルの内容が表示されます。データ・エディター内の情報は、変数とケースで構成されています。

- 「データ・ビュー」では、列が変数を意味し、行がケース (観測値) を表します。
- 「変数ビュー」では、各行が変数であり、各列がその変数に関連する属性です。

変数は、収集したさまざまなデータの種類を表現するために使用されます。アンケート調査がその一般的な例です。アンケート調査のそれぞれの質問に対する回答が変数に相当します。変数には、数値型、文字列型、通貨型、日付型を含むさまざまな型があります。

数値データの入力

データはデータ・エディターに入力することができます。このデータ・エディターは、小さいデータ・ファイルを処理したり、比較的大きなデータ・ファイルの細部を編集するのに便利です。

1. 「データ エディタ」ウィンドウの下部にある「**変数ビュー**」タブをクリックします。

   使用する変数を定義する必要があります。この場合は、年齢、婚姻状況、および 収入 の 3 つの変数のみが必要です。

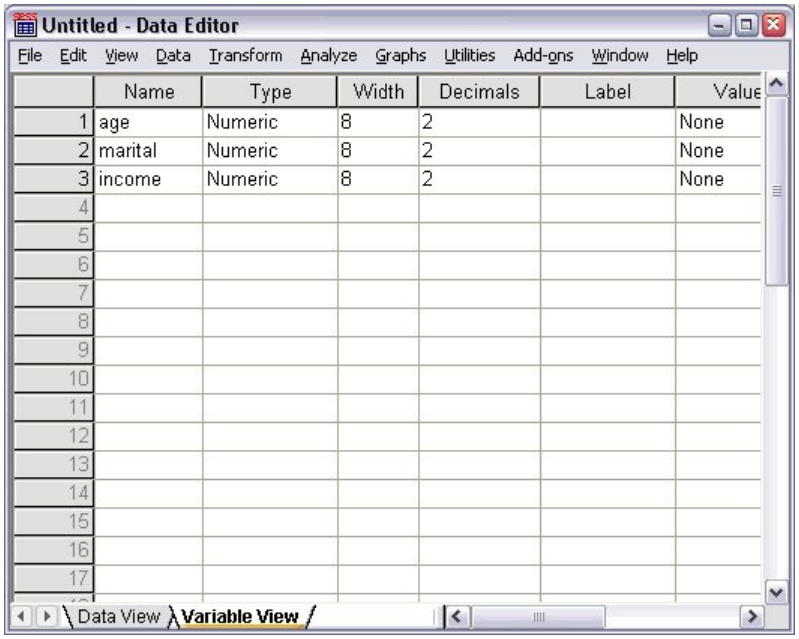


   図 *19.* 「変数ビュー」の変数名

2. 1 列目の 1 行目に、「年齢」と入力します。
3. 2 行目に、「婚姻状況」と入力します。
4. 3 行目に、「収入」と入力します。

   新しい変数は、自動的に数値データ型に指定されます。

   変数名を入力しない場合、一意の名前が自動的に作成されます。ただし、これらの名前は記述的ではないので、データ・ファイルが大きい場合にはお勧めしません。

5. 「**データ ビュー**」タブをクリックして、データを引き続き入力します。

   これで、「変数ビュー」に入力した名前が、「データ・ビュー」の最初の 3 列の見出しになります。

   1 列目の 1 行目からデータを入力していきます。

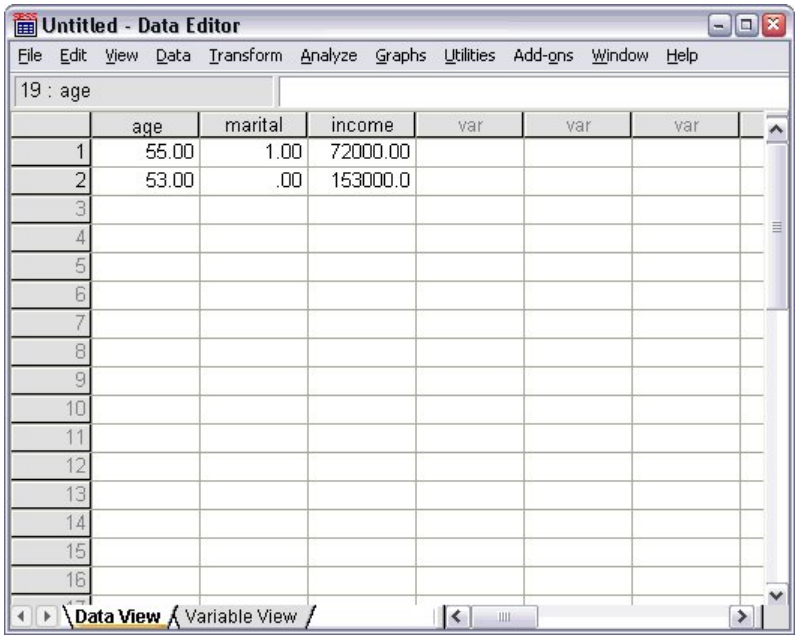


図 *20.* 「データ・ビュー」に入力された値

6. 「年齢」列に、55 と入力します。
7. 「婚姻状況」列に、1 と入力します。
8. 「収入」列に、72000 と入力します。
9. 次の被験者のデータを追加するために、1 列目の 2 行目にカーソルを移動します。
10. 「年齢」列に、53 と入力します。
11. 「婚姻状況」列に、0 と入力します。
12. 「収入」列に、153000 と入力します。

    「年齢」と「婚姻状況」列は、本来の整数となるはずですが、現時点では小数が表示されています。 これらの変数の小数部を非表示にするには、次の手順を実行します。

13. 「データ エディタ」ウィンドウの下部にある「**変数ビュー**」タブをクリックします。
14. 「年齢」行の「小数桁数」列に、「0」と入力して小数部を非表示にします。
15. 「婚姻状況」行の「小数桁数」列に、「0」と入力して小数部を非表示にします。

文字列データの入力

データ・エディターには、テキストの文字列など、非数値データも入力できます。

1. 「データ エディタ」ウィンドウの下部にある「**変数ビュー**」タブをクリックします。
2. 最初の空白行の最初のセルに、変数名として「性別」と入力します。
3. 入力したセルの横にある「型」セルをクリックします。
4. 「型」セルの右側にあるボタンをクリックし、「変数の型」ダイアログ・ボックスを開きます。
5. 「**文字列**」を選択して変数の型を指定します。
6. 「**OK**」をクリックして選択内容を保存し、データ・エディターに戻ります。

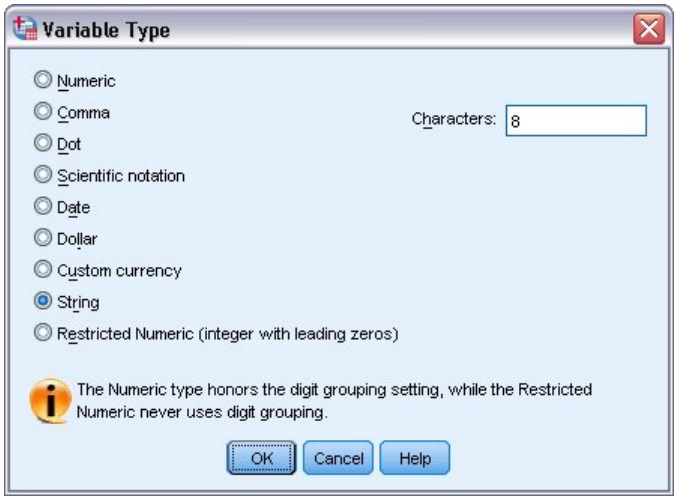


図 *21.* 「変数の型」ダイアログ・ボックス

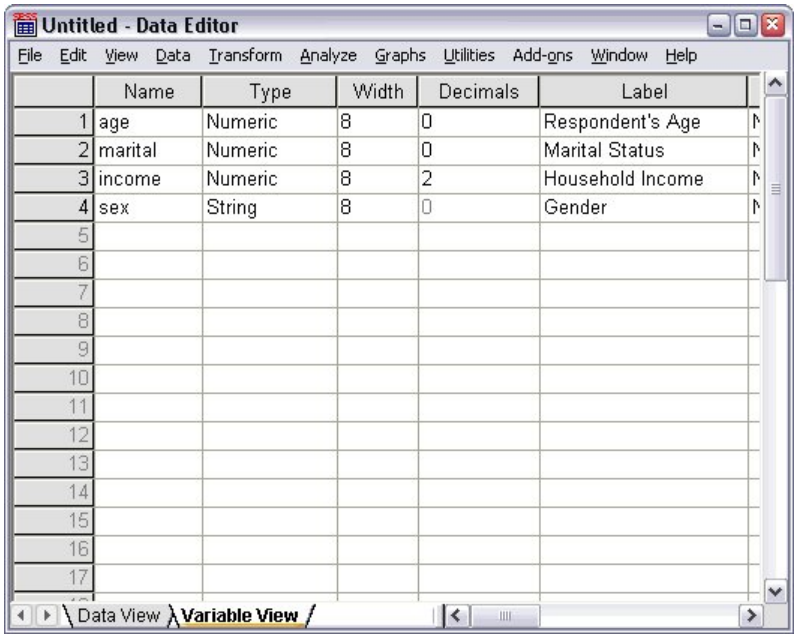
データの定義

データの型を定義することに加えて、変数名とデータ値に対してわかりやすい変数ラベルと値ラベルを定義することもできます。このような記述的ラベルは、統計レポートおよびグラフに使用されます。

変数ラベルの追加

ラベルには、変数の説明が記述されます。この説明は、多くの場合、変数名よりも長くなります。ラベルは最大 255 バイトです。 これらのラベルは、出力の中でさまざまな変数を識別するために使用されます。

1. 「データ エディタ」ウィンドウの下部にある「**変数ビュー**」タブをクリックします。
2. 「年齢」行の「ラベル」列に、「回答者の年齢」と入力します。
3. 「婚姻状況」行の「ラベル」列に、「婚姻状況」と入力します。
4. 「年収」行の「ラベル」列に、「世帯全体の収入」と入力します。
5. 「性別」行の「ラベル」列に、「性別」と入力します。



	Name	Type	Width	Decimals	Label
1	age	Numeric	8	0	Respondent's Age
2	marital	Numeric	8	0	Marital Status
3	income	Numeric	8	2	Household Income
4	sex	String	8	0	Gender

図 *22.* 「変数ビュー」に入力された変数ラベル

変数の型と書式の変更

「型」列には、変数ごとに現在のデータの型が表示されます。最も一般的なデータ型は数値型と文字列型ですが、これ以外にもさまざまな書式がサポートされています。現在のデータ・ファイルでは、「収入」変数が数値型として定義されています。

1. 「収入」行の「型」セルをクリックし、セルの右側にあるボタンをクリックして、「変数の型」ダイアログ・ボックスを開きます。
2. 「**ドル記号**」を選択します。

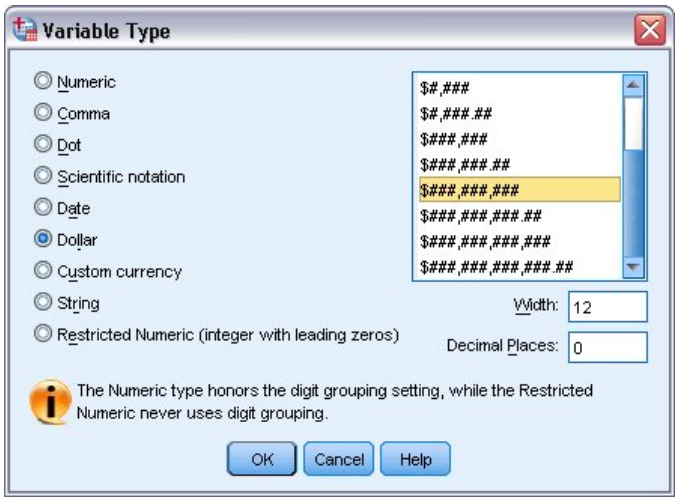


図 *23.* 「変数の型」ダイアログ・ボックス

現在選択されているデータの型のフォーマット・オプションが表示されます。

3. この例では、通貨の書式として、「**$###,###,###**」を選択します。
4. 「**OK**」をクリックして変更内容を保存します。

値ラベルの追加

値ラベルにより、文字列ラベルに変数値をマップすることができるようになります。この例では、「婚姻状況」変数に適切な値が 2 つあります。0 は被験者が未婚であることを意味し、1 は既婚であることを意味します。

1. 「婚姻状況」行の「値」セルをクリックし、セルの右側にあるボタンをクリックして、「値ラベル」ダイアログ・ボックスを開きます。

   「**値**」は実際の数値です。

   「**値ラベル**」は、指定した数値に適用される文字列ラベルです。
2. 「値」フィールドに「0」と入力します。
3. 「ラベル」フィールドに「未婚」と入力します。
4. 「**追加**」をクリックし、このラベルをリストに追加します。

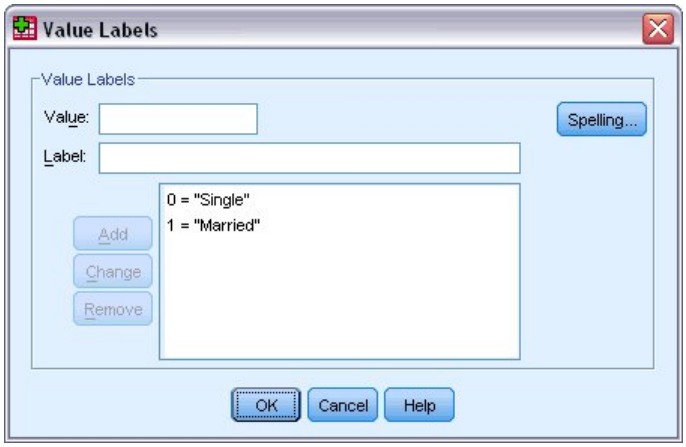


図 *24.* 「値ラベル」ダイアログ・ボックス

5. 「値」フィールドに「1」と入力し、「ラベル」フィールドに「既婚」と入力します。
6. 「**追加**」をクリックし、「**OK**」をクリックして、変更内容を保存し、データ・エディターに戻ります。

   これらのラベルは「データ・ビュー」にも表示できます。これによりデータがさらに読みやすくなります。
7. 「データ エディタ」ウィンドウの下部にある「**データ ビュー**」タブをクリックします。
8. メニューから次の項目を選択します。

「**表示**」 > 「**値ラベル**」

データ・エディターに値を入力すると、ラベルがリストに表示されるようになります。この設定により、有効な回答を提案したり、よりわかりやすい回答が得られるようになります。

「値ラベル」のメニュー項目がすでにアクティブになっている (項目の横にチェック・マークが付いている) 場合、「**値ラベル**」を再度選択すると、値ラベルの表示がオフ になります。

欠損データの処理

欠損データまたは無効なデータは、一般的に、無視できないほど多く見られます。調査回答者は、ある種の質問への回答を拒否したり、答えを知らなかったり、あるいは不適切な形式で回答する場合があります。このようなデータを除外または識別しなければ、分析結果が不正確になる可能性があります。

数値データの場合、空白のデータ・フィールドまたは無効な入力が含まれるデータ・フィールドは、1 つのピリオドによって識別できるシステム欠損フィールドに変換されます。

値が欠損した理由が分析にとって重要となる場合があります。例えば、質問に答えることを拒否した回答者と、質問が該当しなかったため答えなかった回答者とを区別することが有用なことがあります。

数値型変数の欠損値

1. 「データ エディタ」ウィンドウの下部にある「**変数ビュー**」タブをクリックします。
2. 「年齢」行の「欠損値」セルをクリックし、セルの右側にあるボタンをクリックして、「欠損値」ダイアログ・ボックスを開きます。

   このダイアログ・ボックスでは、異なる欠損値を 3 つまで設定すること、または値の範囲に加えて個別の値を 1 つ指定することができます。

図 *25.* 「欠損値」ダイアログ・ボックス

3. 「**個別の欠損値**」を選択します。
4. 最初のテキスト・ボックスに「999」と入力し、他の 2 つのテキスト・ボックスを空のままにします。
5. 「**OK**」をクリックして変更内容を保存し、データ・エディターに戻ります。

   これで欠損データ値が追加されたので、その値にラベルを適用することができます。
6. 「年齢」行の「値」セルをクリックし、セルの右側にあるボタンをクリックして、「値ラベル」ダイアログ・ボックスを開きます。
7. 「値」フィールドに「999」と入力します。
8. 「ラベル」フィールドに「なし」と入力します。
9. 「**追加**」をクリックして、このラベルをデータ・ファイルに追加します。
10. 「**OK**」をクリックして変更内容を保存し、データ・エディターに戻ります。

文字列変数の欠損値

文字列変数の欠損値は、数値型変数の欠損値の場合と同様に処理されます。 ただし、文字列変数では、数値型変数の場合と異なり、空白のフィールドがシステム 欠損として指定されません。その代わり、空白文字列として解釈されます。

1. 「データ エディタ」ウィンドウの下部にある「**変数ビュー**」タブをクリックします。
2. 「性別」行の「欠損値」セルをクリックし、セルの右側にあるボタンをクリックして、「欠損値」ダイアログ・ボックスを開きます。
3. 「**個別の欠損値**」を選択します。
4. 最初のテキスト・ボックスに「NR」と入力します。

   文字列変数の欠損値では、大文字と小文字が区別されます。したがって、*nr* という値は、欠損値として扱われません。
5. 「**OK**」をクリックして変更内容を保存し、データ・エディターに戻ります。

   次に、欠損値にラベルを付けることができます。
6. 「性別」行の「値」セルをクリックし、セルの右側にあるボタンをクリックして、「値ラベル」ダイアログ・ボックスを開きます。
7. 「値」フィールドに「NR」と入力します。
8. 「ラベル」フィールドに「なし」と入力します。
9. 「**追加**」をクリックして、このラベルをプロジェクトに追加します。
10. 「**OK**」をクリックして変更内容を保存し、データ・エディターに戻ります。

第 4 章 個々の変数に対する要約統計量の検討

このセクションでは、単純な要約統計量について、また変数の尺度レベルが必要な統計のタイプにどのように影響するのかについて説明します。データ・ファイル *demo.sav* を使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

測定レベル

異なるデータの種類に対し、適切な要約統計量もさまざまです。次のような測定レベルにより変わります。

「**カテゴリー**」。限られた数の、異なる値またはカテゴリーを持つデータ (例えば、性別や婚姻状況)。これは、**質的データ**とも呼ばれます。カテゴリー変数には、文字列 (英数字) データ、または数字コードを使用してカテゴリーを表す (例えば、0 = 未婚 および 1 = 既婚) 数値型変数があります。カテゴリー・データには 2 つの基本タイプがあります。

- 「**名義**」。カテゴリーに固有の順序が存在しないカテゴリー・データです。例えば、セールス のジョブ・カテゴリーはマーケティング またはリサーチ のジョブ・カテゴリーより上でも下でもありません。
- 「**順序**」。カテゴリーに意味のある順序が存在するカテゴリー・データです。ただし、カテゴリー間には測定可能な距離はありません。例えば、高、中、低 の値には順序が存在しますが、値の間の「距離」は計算できません。

「**スケール**」。間隔、または比率スケールに基づいて測定したデータです。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。これは、**量的データ**または**連続データ**とも呼ばれます。

カテゴリー・データの要約統計量

カテゴリー・データで、最も典型的な要約統計量は、各カテゴリーのケース数またはケースのパーセントです。**最頻値**は、ケース数が最も多いカテゴリーです。順序データでは、カテゴリー数が多い場合、**中央値** (ケースの半分をそれぞれ上下に分ける値)が有効な要約統計量となることもあります。

度数分布表の手続きでは、度数テーブルに、変数の各観測値についてのケース数とケースのパーセントの両方が表示されます。

1. メニューから次の項目を選択します。

   「**分析**」 > 「**記述統計**」 > 「**度数分布表...**」

   注: この機能には、Statistics Base オプションが必要です。

2. 「携帯情報端末 *[PDA]*」と「*TV [TV]*」を選択し、それらを変数リストに移動します。

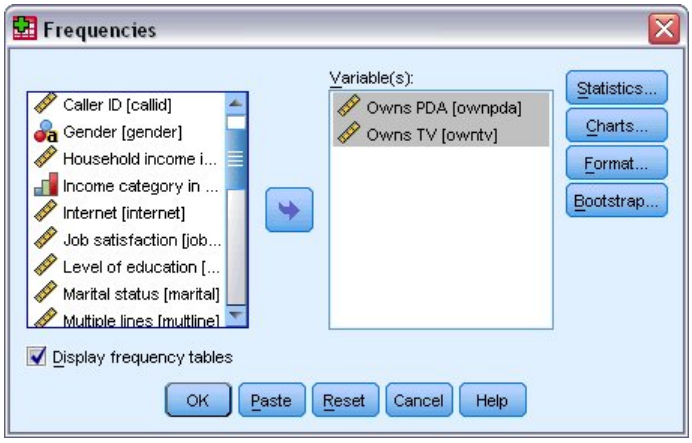
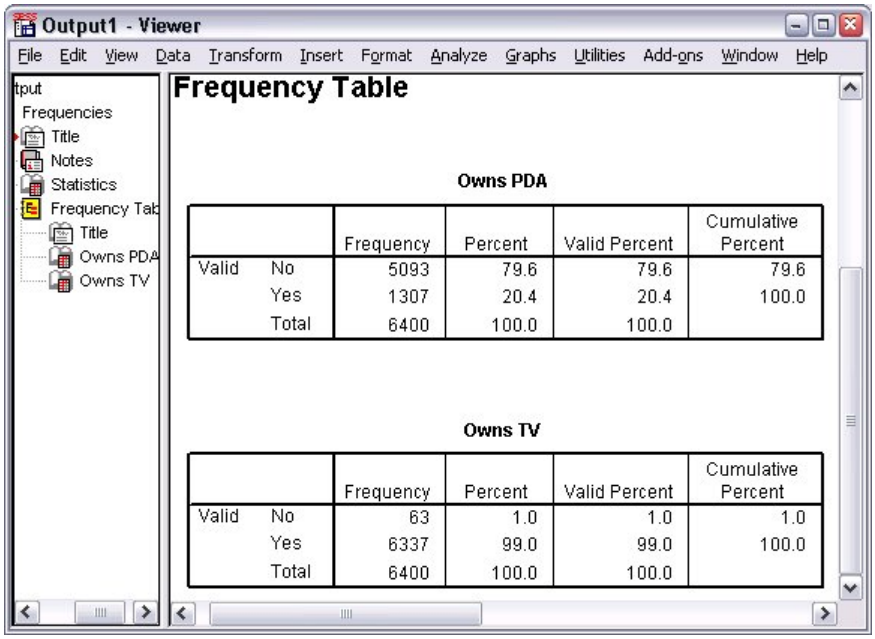


図 *26.* 分析のために選択したカテゴリー変数

3. 「**OK**」をクリックして手続きを実行します。



Owns PDA		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	5093	79.6	79.6	79.6
	Yes	1307	20.4	20.4	100.0
	Total	6400	100.0	100.0	

Owns TV		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	63	1.0	1.0	1.0
	Yes	6337	99.0	99.0	100.0
	Total	6400	100.0	100.0	

図 *27.* 度数分布表

度数分布表は「ビューアー」ウィンドウに表示されます。度数分布表によって、わずか 20.4% の人だけが PDA を所有している一方で、ほとんどすべての人 (99.0%) がテレビを所有していることがわかります。これは、それほど興味深い発見ではないかもしれません。むしろ、テレビを所有していないわずかな人のグループについて、詳しく調べる方が興味深いかもしれません。

カテゴリー・データのグラフ

棒グラフまたは円グラフを使用して、度数分布表に情報を図で表示することができます。

1. 「度数分布表」ダイアログ・ボックスをもう一度開きます。(ここでも変数は 2 つ選択されている必要があります)。

   ツールバーの「ダイアログ リコール」ボタンを使用すると、最近使用した手続きに素早く戻ることができます。

   図 *28.* 「ダイアログ リコール」ボタン

2. 「**図表**」をクリックします。
3. 「**棒グラフ**」を選択し、「**続行**」をクリックします。
4. メインのダイアログ・ボックスで「**OK**」をクリックし、手続きを実行します。

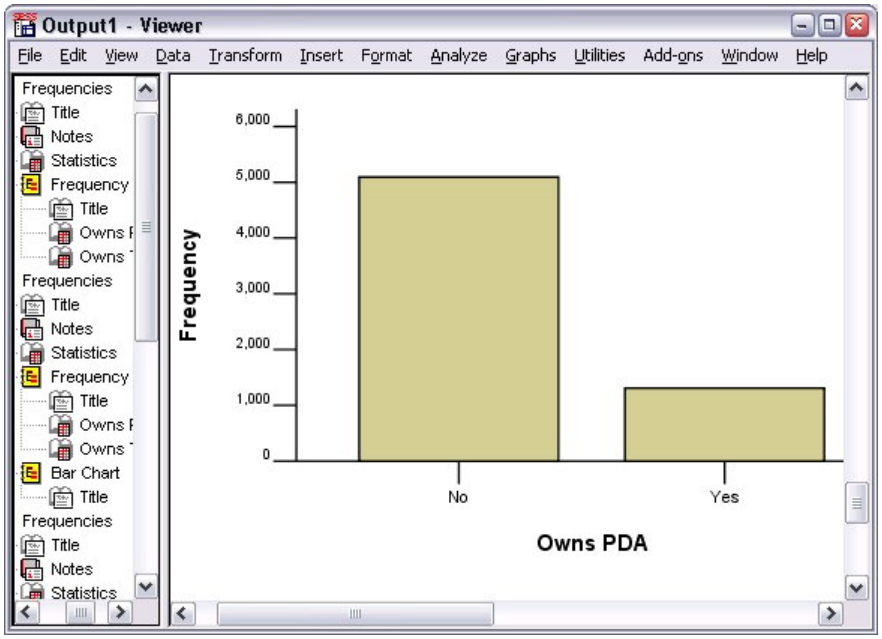


図 *29.* 棒グラフ

度数分布表に加えて、同じ情報が棒グラフによって表示され、ほぼすべての人がテレビを所有しているのに対し、ほとんどの人が PDA を所有していないことがよりわかりやすくなります。

スケール変数の要約統計量

スケール変数には多くの要約統計量を使用できます。例えば、次のものが含まれます。

- **中心傾向の測定方法**。最も一般的な中心傾向の測定方法は、**平均値** (算術平均) および**中央値** (ケースの半分をそれぞれ上下に分ける値) です。
- **分散の測定方法**。データの変化量または広がり量を測定する統計には、標準偏差、最小値、最大値があります。

1. 「度数分布表」ダイアログ・ボックスをもう一度開きます。
2. 「**戻す**」をクリックして、以前の設定をすべてクリアします。
3. 「世帯全体の収入カテゴリ (千ドル)」を選択し、変数リストに移動します。
4. 「**統計**」をクリックします。
5. 「**平均値**」、「**中央値**」、「**標準偏差**」、「**最小値**」、および「**最大値**」を選択します。
6. 「**続行**」をクリックします。
7. メインのダイアログ・ボックスで、「**度数分布表の表示**」を選択解除します。(データ・ファイルにあるケースの数とほぼ同数の異なる数値が存在する場合があるため、通常、度数分布表はスケール変数に対して有用ではありません)。
8. 「**OK**」をクリックして手続きを実行します。

「ビューアー」ウィンドウに度数分布表の統計テーブルが表示されます。

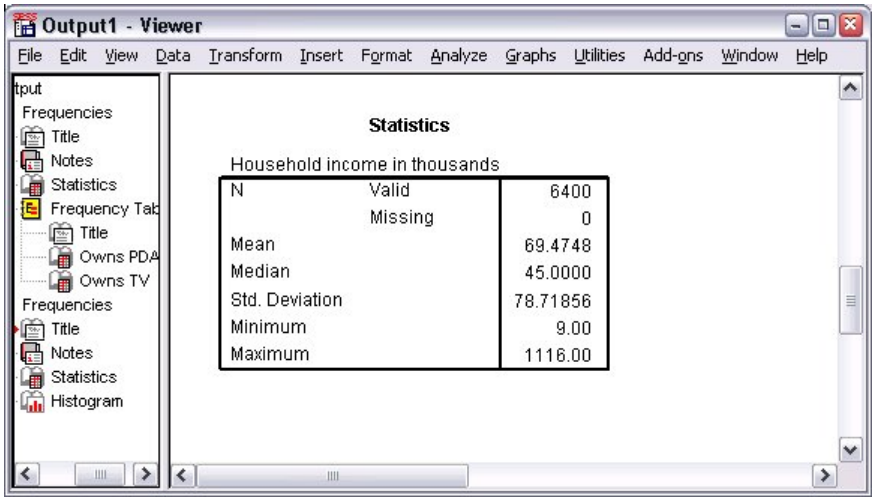


図 *30.* 度数分布表の統計テーブル

この例では、平均値と中央値との間に大きな差があります。平均値は中央値より 25,000 近くも大きく、このことは、値が正規分布ではないことを示しています。ヒストグラムを使用すると、分布を視覚的に確認することができます。

スケール変数のヒストグラム

1. 「度数分布表」ダイアログ・ボックスをもう一度開きます。
2. 「**図表**」をクリックします。
3. 「**ヒストグラム**」を選択し、「**正規曲線 (With normal curve)**」を選択します。
4. メインのダイアログ・ボックスで「**続行**」をクリックし、「**OK**」をクリックして手続きを実行します。

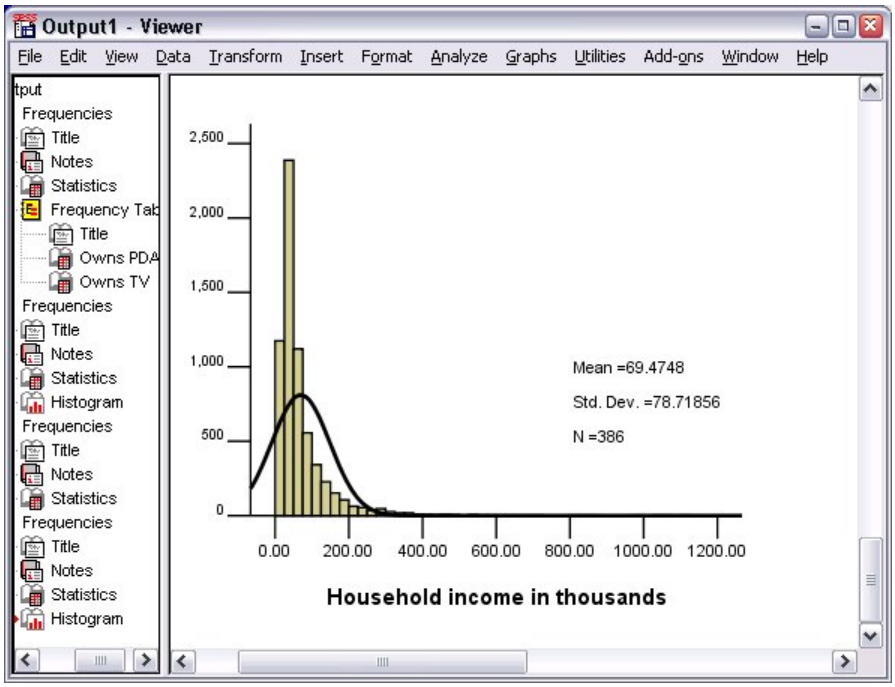


図 *31.* ヒストグラム

ケースの大部分はスケールの下限でクラスター化され、多くは 100,000 未満となります。ただし、500,000 の範囲内、およびその範囲を超えるケースがあります (このケースは少ないため、ヒストグラムに変更を加えないと表示できません)。これら少数のケースに限られる非常に高い値は、平均値には影響を与えますが、中央値にはほぼ、あるいはまったく影響を与えません。そのため、この例の中心傾向の指標には、中央値がより適切となります。

第 5 章 グラフの作成および編集

さまざまな種類の図表を作成および編集することができます。この章では、棒グラフの作成と編集を行います。その他の図表の種類にも、この原則が適用されます。

図表作成の基本

図表作成の基本を説明するため、さまざまなレベルの仕事満足度に対応した平均収入の棒グラフを作成します。この例では、データ・ファイル *demo.sav* を使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

1. メニューから次の項目を選択します。

「**グラフ**」 > 「**図表ビルダー...**」

「図表ビルダー」ダイアログ・ボックスは、図表の作成中にその見え方をプレビューすることができる、インタラクティブなウィンドウです。

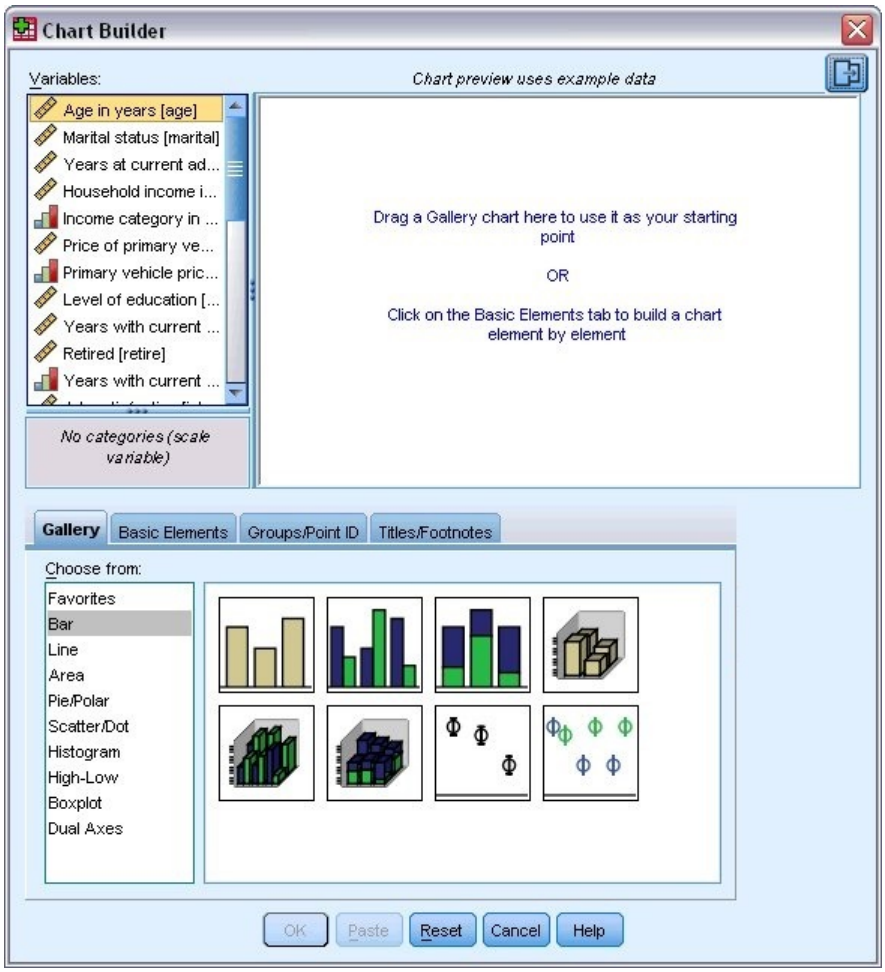


図 *32.* 「図表ビルダー」ダイアログ・ボックス

図表ビルダー・ギャラリーの使用

1. 「**ギャラリー**」タブをクリックします (このタブが選択されていない場合)。

ギャラリーには、さまざまな定義済みの図表が含まれており、図表の種類ごとに整理されています。「基本要素」タブには、初めから図表を作成するための基本要素 (軸やグラフィック要素など) も用意されていますが、「ギャラリー」を使用するほうがより簡単です。

2. 「**棒グラフ**」をクリックします (これが選択されていない場合)。

   ギャラリー内の使用可能な棒グラフを示すアイコンが、ダイアログ・ボックスに表示されます。この図には、特定の図表の種類を指定するのに十分な情報が提供されているはずです。さらに情報が必要な場合、カーソルをアイコンに 合わせると、図表のツールチップの説明を表示することもできます。

3. 単純棒グラフのアイコンを 「キャンバス」にドラッグします。「キャンバス」とは、ギャラリーの上にある広い領域のことです。「図表ビルダー」では、キャンバスに図表のプレビューが表示されます。この図表の描画に使用されるデータは、実際のデータではないことに注意してください。このデータは、サンプルです。

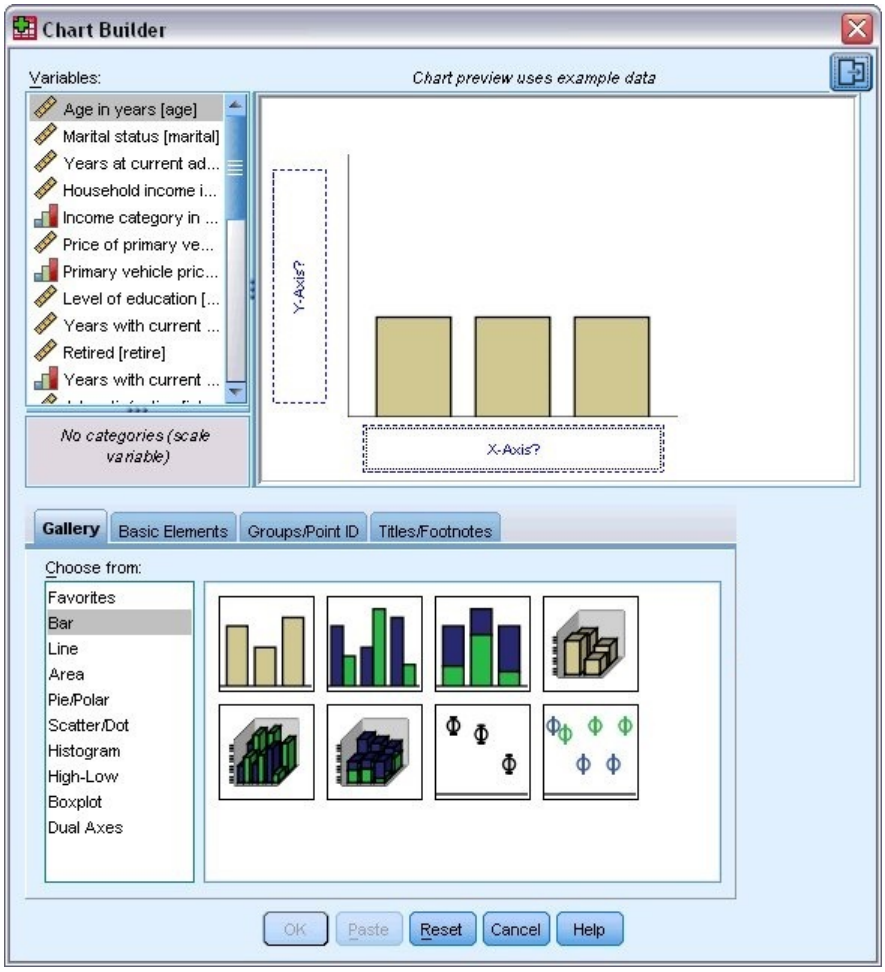


図 *33.* 「図表ビルダー」キャンバスの棒グラフ

変数および統計量の定義

キャンバス上には図表がありますが、棒の高さを制御したり、各棒に対応する変数カテゴリーを指定する変数または統計量がないため、この図表はまだ完全ではありません。変数や統計量のない図表を作成することはできません。キャンバスの左側に配置された「変数」リストからドラッグすることにより、変数を追加します。

図表ビルダーでは、変数の測定レベルが重要です。*X* 軸で「仕事に対する満足度」変数を使用します。ただし、変数の隣にある (ルーラーのような) アイコンは、この変数の測定レベルがスケールとして定義されていることを示しています。正しい図表を作成するには、カテゴリーの測定レベルを使用する必要があります。変数ビューに戻って測定レベルを変更するのではなく、図表ビルダーで一時的に測定レベルを変更することもできます。

1.「変数」リストの「仕事に対する満足度」を右クリックし、「**順序**」を選択します。「仕事に対する満足度」のカテゴリーは満足度のレベルでランク付けされるため、「順序」が適切な測定レベルとなります。測定レベルを変更すると、アイコンが変更されることに注意してください。
2.「仕事に対する満足度」を変数リストから *X* 軸のドロップ・ゾーンにドラッグします。

   Y 軸のドロップ・ゾーンのデフォルトは、度数 の統計量です。別の統計量 (パーセンテージや平均など) を使用する場合は、容易に変更することができます。この例では、どちらの統計量も使用しませんが、統計量を変更する必要があるときのために、後程この手順をレビューします。
3. 図表ビルダーのサイドバーで「**要素のプロパティ**」タブをクリックします。 (サイドバーが表示されていない場合は、図表ビルダーの右上隅にあるボタンをクリックしてサイドバーを表示します。)

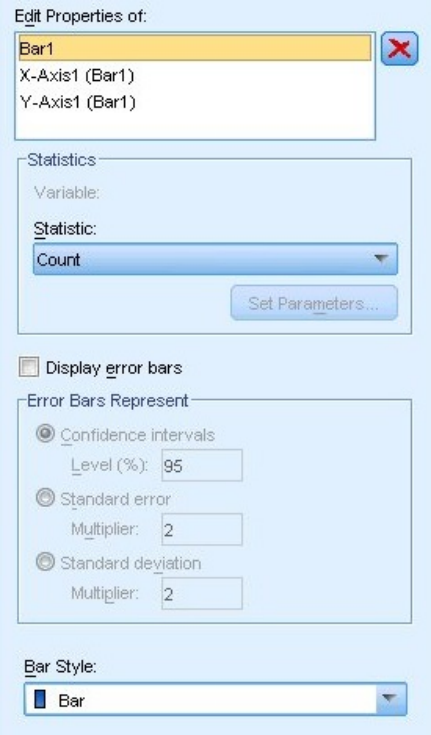


図 *34.* 要素のプロパティ

   「要素のプロパティ」では、さまざまな図表要素のプロパティを変更できます。これらの要素には、グラフィック要素 (棒グラフの棒など) および図表の軸が含まれます。「以下のプロパティーを編集」リストで要素の 1 つを選択し、要素に関連するプロパティーを変更します。リストの右側にある赤色の「*X*」にも注意してください。このボタンで、キャンバスからグラフィック要素が削除されます。「**棒グラフ 1**」が選択されているため、表示されたプロパティーがグラフィック要素、具体的には棒グラフのグラフィック要素に適用されます。

   「統計量」ドロップダウン・リストには、使用可能な特定の統計量が表示されます。通常、同じ統計量をすべての図表の種類に使用できます。統計量では、変数を含む *Y* 軸のドロップ・ゾーンを必要とする場合があります。
4.「世帯全体の収入 *(千ドル)*」を変数リストから *y* 軸のドロップ・ゾーンにドラッグします。*Y* 軸の変数はスカラーで、*X* 軸の変数はカテゴリー (順序はカテゴリーの測定レベルの種類) であるため、*Y* 軸のドロップ・ゾーンのデフォルトは *MEAN* 統計量になります。必要な変数および統計量があるため、要素のプロパティーを変更する必要はありません。

テキストの追加

図表に表題および脚注を追加することもできます。

1.「**表題/脚注**」タブをクリックします。
2.「**表題 1**」を選択します。

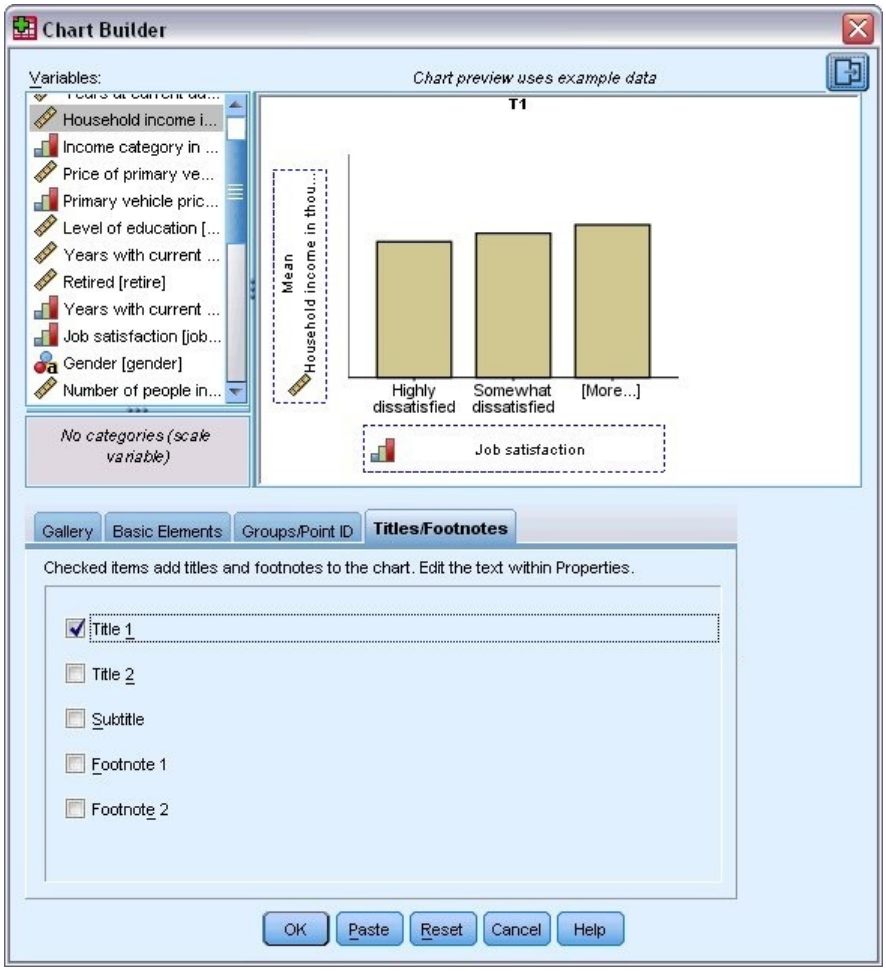


図 *35.* キャンバスに表示された「表題 *1*」

タイトルが、ラベル「**T1**」でキャンバスに表示されます。

3.「**要素のプロパティ**」タブでは、「以下のプロパティを編集」リストで「**表題 1**」を選択します。
4.「内容」テキスト・ボックスに、「`Income by Job Satisfaction`」と入力します。これは、表題として表示されるテキストです。

図表の作成

1.「**OK**」をクリックし、棒グラフを作成します。

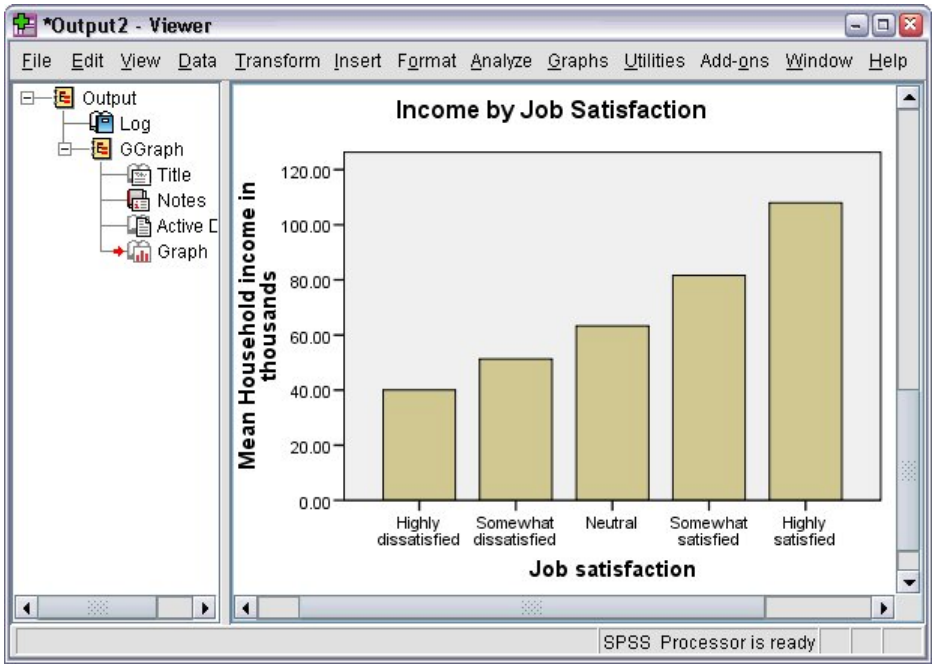


図 *36.* 棒グラフ

この棒グラフにより、仕事に対する満足感の高い回答者ほど、世帯収入も高くなる傾向にあることが明らかになります。

第 6 章 出力を使用した作業

統計手続きの実行結果は「ビューアー」に表示されます。生成された出力は、手続きの実行時に選択した項目に応じて、統計テーブル、図表、グラフ、またはテキストにすることができます。このセクションでは、*viewertut.spv* ファイルおよび *demo.sav* ファイルを使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

ビューアーの使用

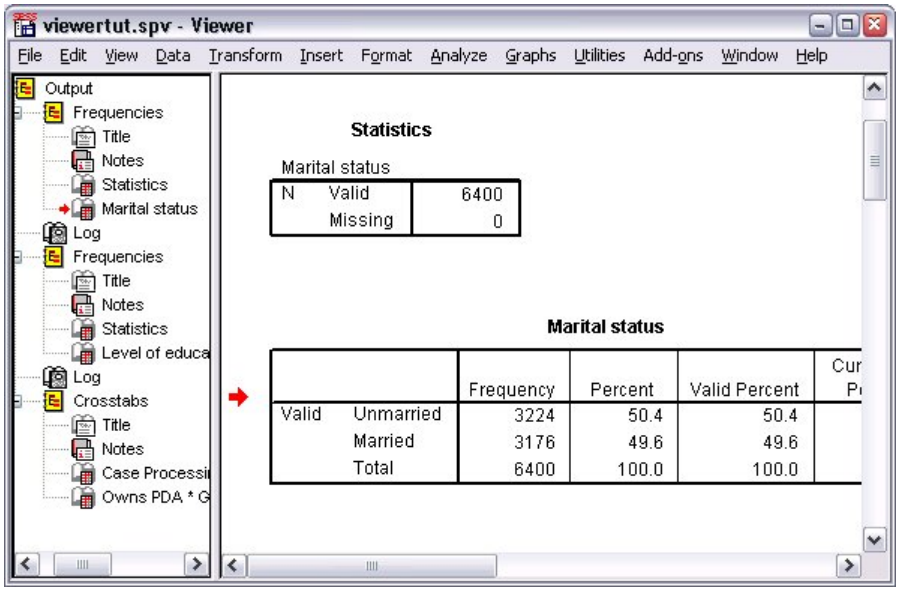


図 *37.* ビューアー

「ビューアー」ウィンドウは 2 つのペインに分かれています。 **アウトライン・ペイン**には、「ビューアー」で保存されたすべての情報の概要が表示されます。**内容のウィンドウ**には、統計テーブル、図表、およびテキスト出力が表示されます。

スクロール・バーを使用して、縦方向および横方向にウィンドウの内容をナビゲートします。簡単にナビゲーションするには、アウトライン・ペインで項目をクリックして、内容のウィンドウにその項目を表示します。

1. アウトライン・ペインの右枠をクリックしてドラッグし、その幅を変更します。

   アウトライン・ペインに表示される開いたブック・アイコンは、それが現在「ビューアー」に表示されていることを示します。ただし、内容のウィンドウの表示部分には表示されていない可能性があります。

2. テーブルまたは図表を非表示にするには、アウトライン・ペインでそのブック・アイコンをダブルクリックします。

   開いたブック・アイコンが閉じたブック・アイコンに変わり、関連付けられている情報が非表示になったことを示します。

3. 非表示の出力を再表示するには、閉じたブック・アイコンをダブルクリックします 。

   特定の統計手続きからのすべての出力、または「ビューアー」内のすべての出力を非表示にすることもできます。

4. 結果を非表示にする手続きの左側にある負符号 (-) のボックスをクリックするか、またはアウトライン・ペインの一番上の項目の横にあるボックスをクリックして、すべての出力を非表示にします。

   アウトラインが省略表示され、これらの結果が非表示になったことがわかります。

出力の表示順序を変更することもできます。

5. アウトライン・ペインで、移動する項目をクリックします。
6. 選択した項目をアウトライン内の新しい場所までドラッグします。

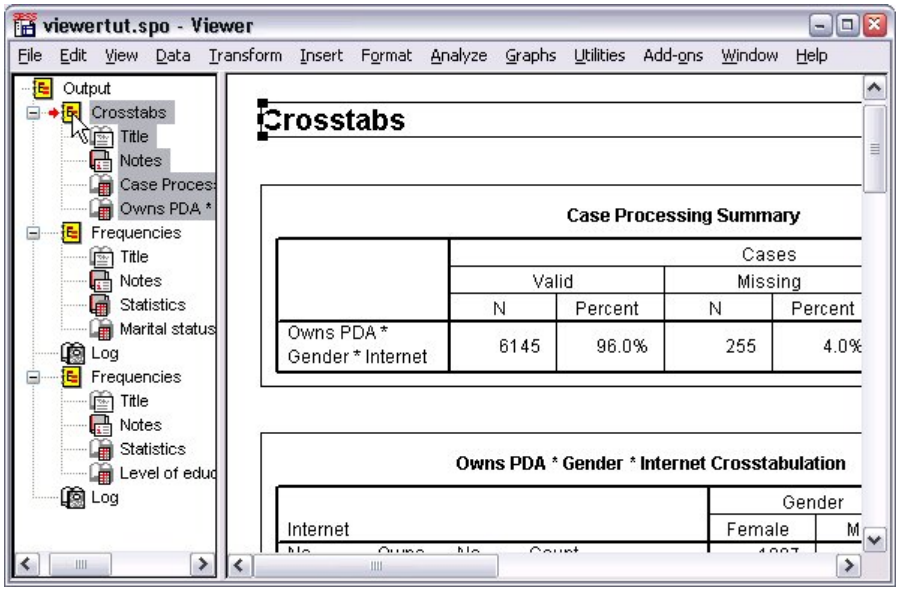


図 *38.* 「ビューアー」で並べ替えられた出力

出力項目を内容のウィンドウでクリックしてドラッグする 方法で、移動することもできます。

ピボット・テーブル・エディターの使用

ほとんどの統計手続きの結果が、**ピボット・テーブル**に表示されます。

出力定義へのアクセス

出力には多数の統計用語が表示されます。これらの用語の定義には、「ビューアー」で直接アクセスできます。

1. 「情報端末と性別とインターネットのクロス表」テーブルをダブルクリックします。
2. 「期待度数」を右クリックし、ポップアップ・メニューから「**用語のヘルプ**」を選択します。

ポップアップ・ウィンドウに定義が表示されます。

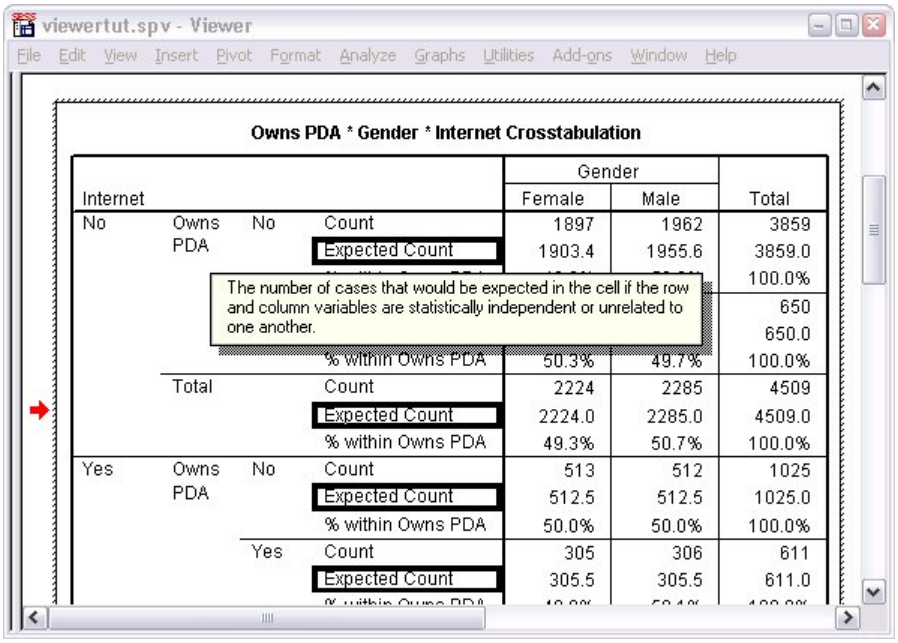


図 *39.* ポップアップ定義

ピボット・テーブル

生成されるデフォルト・テーブルに表示される情報は、簡潔さや明確さがないことがあります。ピボット・テーブルを使用すると、行と列を入れ替える (テーブルを反転させる) ことができ、テーブル内のデータの順序を調整でき、ほかにもさまざまな方法でテーブルを変更できます。例えば、行と列を入れ替えることで、横長のテーブルを縦長のテーブルに変更できます。テーブルのレイアウトを変更しても、結果には影響しません。レイアウト変更によって、情報をより適切な方法で表示できます。

1. 「情報端末と性別とインターネットのクロス表」テーブルがまだアクティブになっていない 場合は、ダブルクリックしてアクティブにします。
2. 「ピボット トレイ」ウィンドウが表示されていない場合は、メニューから次の項目を選択します。

   「**ピボット**」 > 「**ピボット トレイ**」

   ピボット・トレイでは、列、行、および層間でデータを移動できます。

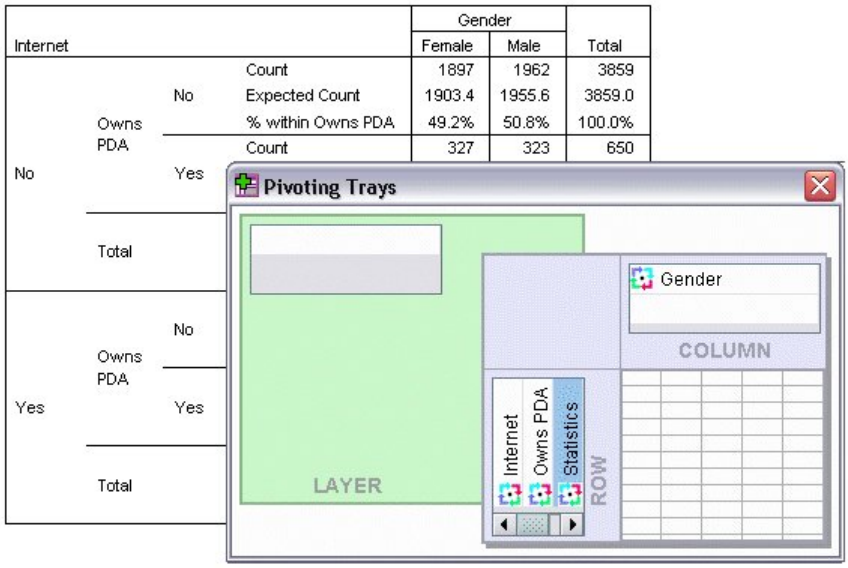


図 *40.* ピボット・トレイ

3. 行ディメンションから列ディメンションの「性別」の下まで「統計量」要素をドラッグします。テーブルは、変更を反映して即時に再構成されます。

   テーブル内の要素の順序が、ピボット・トレイ内の要素の順序に反映されます。

4. 「携帯情報端末」要素を行ディメンションの「インターネット」要素の前にドラッグ・アンド・ドロップして、これらの 2 つの行の順序を反転させます。

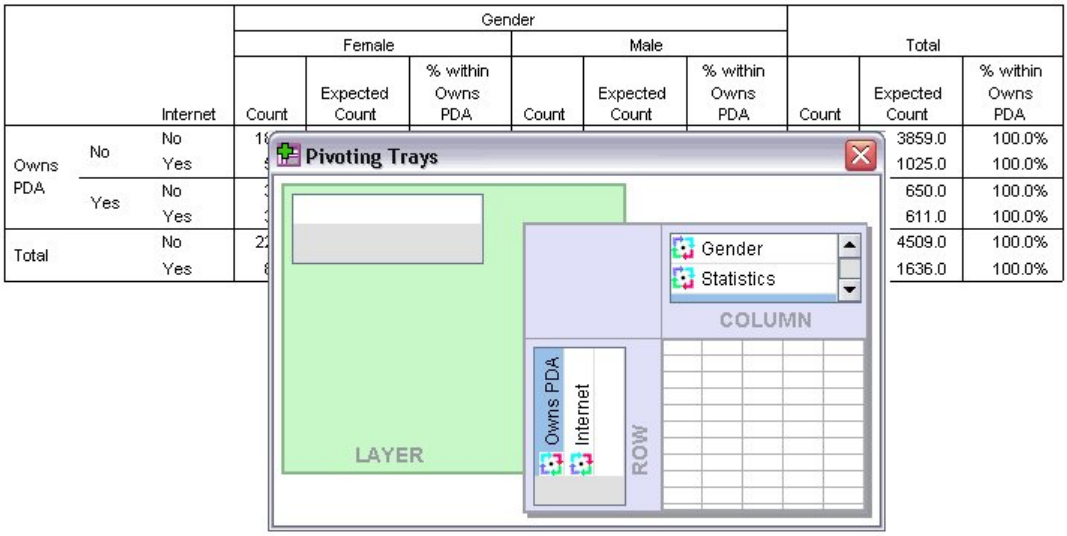


図 *41.* 行の切り替え

層の作成と表示

層は、情報のカテゴリーがネストされている大規模なテーブルの場合に役立ちます。層を作成すると、テーブルの外観が単純になり、読みやすくなります。

1. 列ディメンションからレイヤー・ディメンションまで「性別」要素をドラッグします。

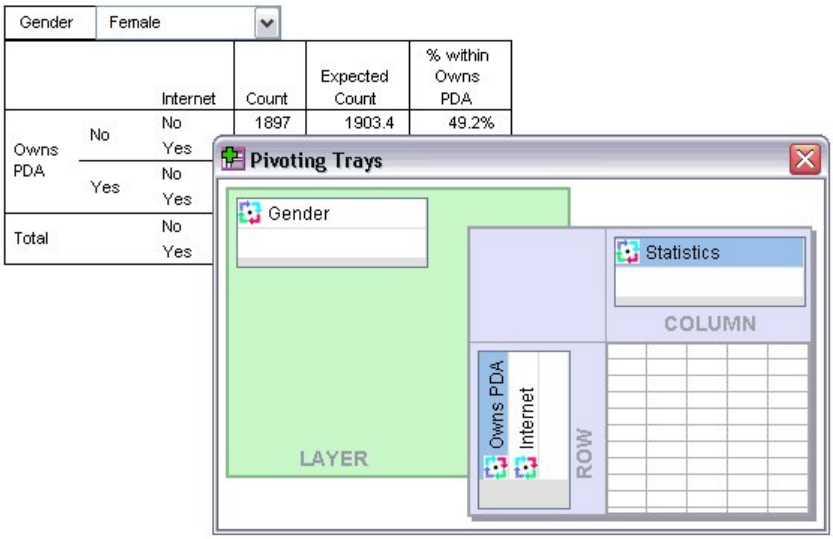


図 *42.* レイヤー・ディメンションの「性別」ピボット・アイコン

別の層を表示するには、テーブルのドロップダウン・リストからカテゴリーを選択します。

テーブルの編集

カスタム・テーブルルックを作成しない限り、ピボット・テーブルは標準の書式設定で作成されます。 テーブル内のテキストの書式設定を変更できます。変更できる書式には、フォント名、フォント・サイズ、フォント・スタイル (太字またはイタリック)、色などがあります。

1. 「教育のレベル」テーブルをダブルクリックします。
2. 「ツールバーの書式設定」が表示されていない場合は、メニューから次の項目を選択します。

   「**表示**」 > 「**ツールバー**」
3. タイトル・テキストの「教育のレベル」をクリックします。
4. ツールバーのフォント・サイズのドロップダウン・リストから「**12**」を選択します。
5. タイトル・テキストの色を変更するには、テキストの色ツールをクリックし、新しい色を選択します。

Level of education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7	21.7
	High school degree	1936	30.3	30.3	52.0
	Some college	1360	21.3	21.3	73.2
	College degree	1355	21.2	21.2	94.4
	Post-undergraduate degree	359	5.6	5.6	100.0
	Total	6400	100.0	100.0	

図 *43.* タイトル・テキストが再書式設定されたピボット・テーブル

テーブルおよびラベルの内容を編集することもできます。 例えば、このテーブルのタイトルを変更できます。

6. タイトルをダブルクリックします。
7. 新しいラベルとして「教育レベル」と入力します。

注: テーブル内の値を変更しても、合計およびその他の統計は再計算されません。

行および列の非表示

テーブルに表示されるデータの一部が役に立たない、あるいはテーブルを必要以上に複雑にしている場合があります。そのような場合でも、データを削除することなく、行全体および列全体を非表示にすることができます。

1. 「教育レベル」テーブルがまだアクティブではない場合は、ダブルクリックしてアクティブにします。
2. 「有効パーセント」列ラベルをクリックして選択します。
3. 「編集」メニューから、または右クリックで表示されるポップアップ・メニューから、次の項目を選択します。

   「**選択**」 > 「**データとラベルのセル**」
4. 「表示」メニューから「**非表示**」を選択するか、または右クリックで表示されるポップアップ・メニューから「**カテゴリを隠す**」を選択します。

   これで、列を削除せずに、非表示になりました。

Education Level

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7
	High school degree	1936	30.3	52.0
	Some college	1360	21.3	73.2
	College degree	1355	21.2	94.4
	Post-undergraduate degree	359	5.6	100.0
	Total	6400	100.0	

図 *44.* テーブルで非表示の「有効パーセント」列

列を再表示するには、次の手順を実行します。

5. メニューから次の項目を選択します。

「**表示**」 > 「**すべて表示**」

行は、列と同じ方法で非表示と表示を切り替えることができます。

データの表示書式の変更

ピボット・テーブルのデータの表示形式は容易に変更できます。

1. 「教育レベル」テーブルがまだアクティブではない場合は、ダブルクリックしてアクティブにします。
2. 「パーセント」列ラベルをクリックして選択します。
3. 「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。

   「**選択**」 > 「**データ セル**」
4. 「書式」メニューから、または右クリックで表示されるポップアップ・メニューから、「**セル プロパティ**」を選択します。
5. 「**書式値**」タブをクリックします。
6. 「小数桁数」フィールドに `0` と入力して、この列の小数点以下をすべて非表示にします。

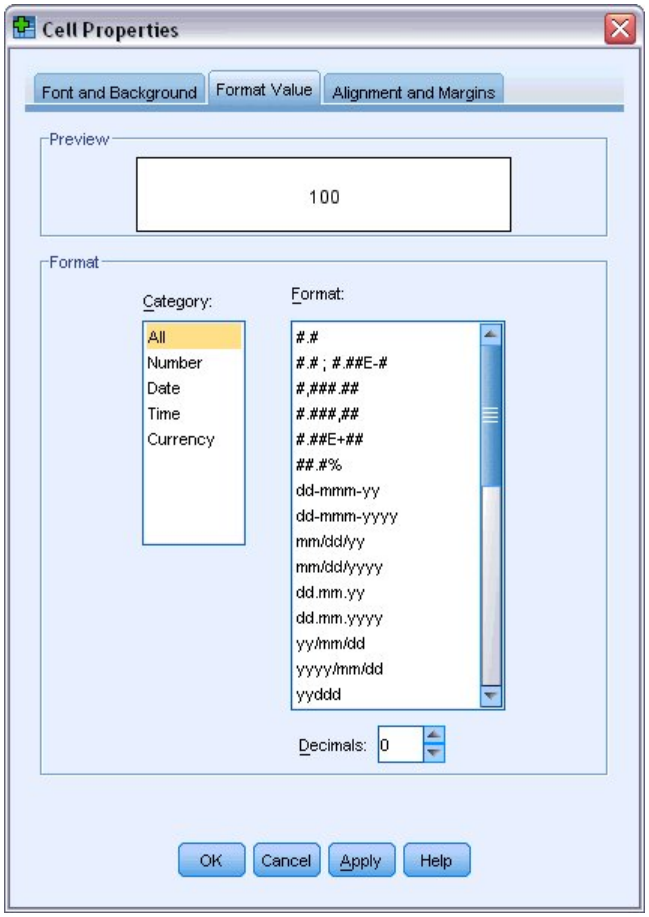


図 *45.* 「セル プロパティ」、「書式値」タブ

このダイアログ・ボックスでは、データ型および書式も変更できます。

7. 適用する型を「カテゴリー」リストから選択した後、その型の書式を「書式」リストで選択します。
8. 「**OK**」または「**適用**」をクリックして、変更を適用します。

Education Level

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	22	21.7
	High school degree	1936	30	52.0
	Some college	1360	21	73.2
	College degree	1355	21	94.4
	Post-undergraduate degree	359	6	100.0
	Total	6400	100	

図 *46.* 小数が非表示になっている「パーセント」列

これで、「パーセント」列の小数が非表示になりました。

テーブルルック

テーブルの書式は、結果の明確さ、簡潔さ、および意義に大きく影響します。読みづらいテーブルの情報は、簡単に理解できません。

定義済み書式の使用

1. 「婚姻状況」テーブルをダブルクリックします。
2. メニューから次の項目を選択します。

「**形式**」 > 「**テーブルルック ...**」

「テーブルルック」ダイアログ・ボックスに、さまざまな定義済みスタイルが表示されます。リストからスタイルを選択して、右側の「サンプル」ウィンドウにそのプレビューを表示します。

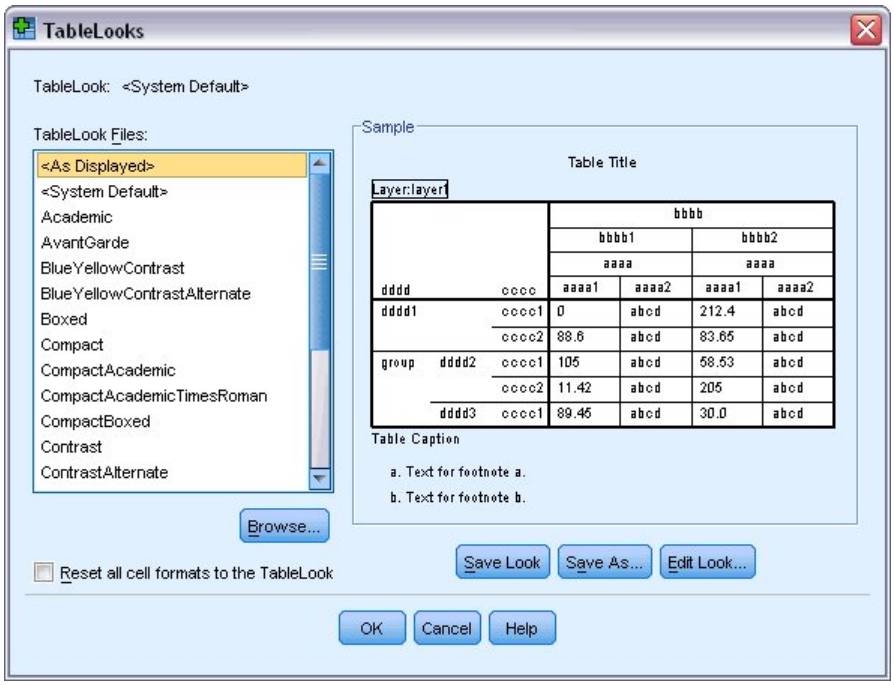


図 *47.* 「テーブルルック」ダイアログ・ボックス

スタイルをそのまま使用することも、既存のスタイルを編集してニーズに合わせることもできます。

3. 既存のスタイルを使用するには、スタイルを選択し、「**OK**」をクリックします。

テーブルルック・スタイルのカスタマイズ

特定のニーズを満たすように、書式をカスタマイズできます。背景の色から枠線のスタイルに至るまで、テーブルのほとんどすべての部分をカスタマイズできます。

1. 「婚姻状況」テーブルをダブルクリックします。
2. メニューから次の項目を選択します。

   「**形式**」 > 「**テーブルルック ...**」
3. 必要な書式にもっとも近いスタイルを選択し、「**テーブル ルックの編集**」をクリックします。
4. 「**セル書式**」タブをクリックして、フォーマット・オプションを表示します。

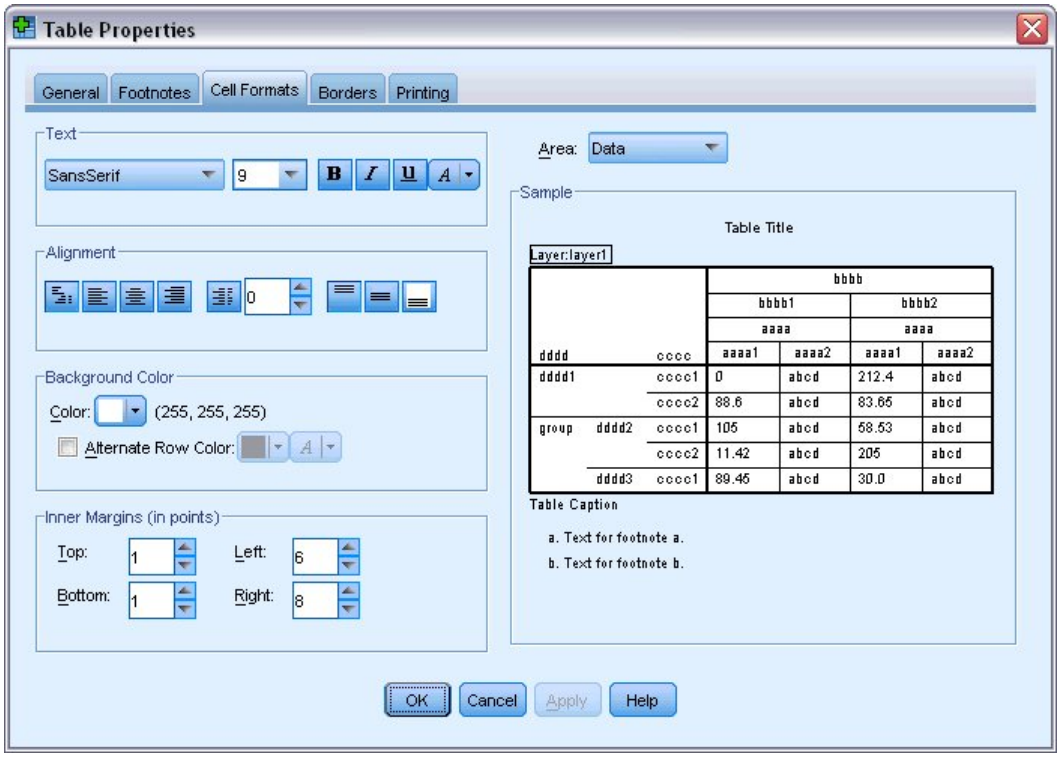


図 *48.* 「テーブル プロパティ」ダイアログ・ボックス

フォーマット・オプションには、フォント名、フォント・サイズ、スタイル、色があります。さらに、位置合わせ、テキストと背景の色、および余白サイズのオプションもあります。

右側の「サンプル」ウィンドウに、書式設定の変更がテーブルに及ぼす影響のプレビューが表示されます。 テーブルの各領域に、異なる書式設定スタイルを適用できます。例えば、タイトルとデータに異なるスタイルを適用できます。編集するテーブルの領域を選択するには、領域の名前を「領域」ドロップダウン・リストで選択するか、または変更する領域を「サンプル」ウィンドウでクリックします。

5. 「領域」ドロップダウン・リストから「**データ**」を選択します。
6. 「背景色」ドロップダウン・パレットから新しい色を選択します。
7. その後、新しいテキストの色を選択します。

「サンプル」ウィンドウに新しいスタイルが表示されます。

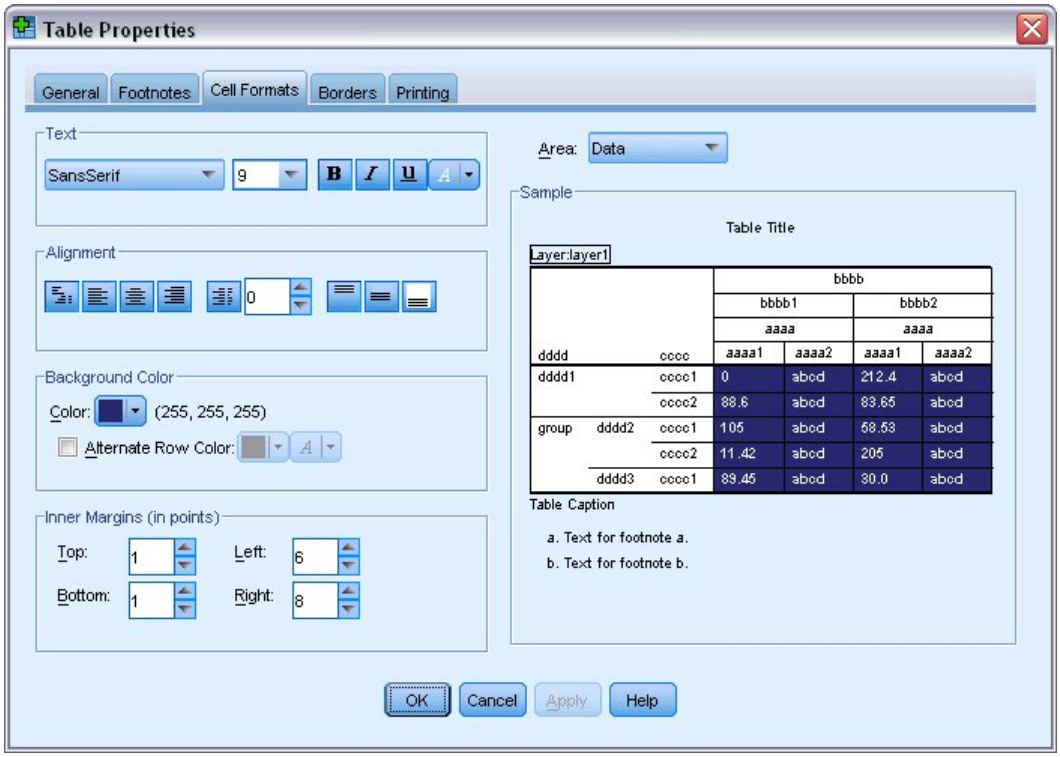


図 *49.* テーブル・セル書式の変更

8. 「**OK**」をクリックして、「テーブルルック」ダイアログ・ボックスに戻ります。

   後から作成するテーブルに簡単に適用できるように、新しいスタイルを保存できます。

9. 「**名前を付けて保存**」をクリックします。
10. ターゲット・ディレクトリーまで移動し、「ファイル名」テキスト・ボックスに新しいスタイルの名前を入力します。
11. 「**保存**」をクリックします。
12. 「**OK**」をクリックして変更を適用し、「ビューアー」に戻ります。

これで、指定したカスタム書式設定がテーブルに適用されました。

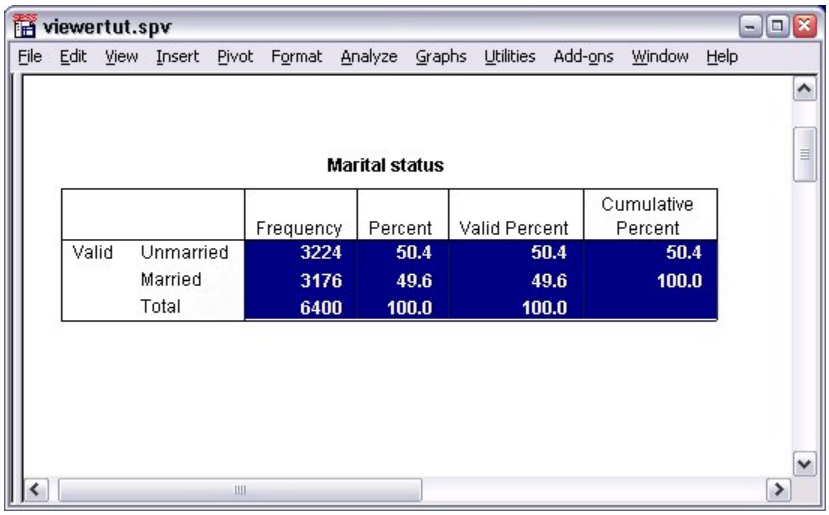


図 *50.* カスタム・テーブルルック

デフォルトのテーブル書式の変更

テーブルの書式は作成後に変更できますが、デフォルトのテーブルルックを変更した方が効率的な場合があります。これにより、テーブルを作成するたびに書式を変更する必要がなくなります。

ピボット・テーブルのデフォルトのテーブルルック・スタイルを変更するには、メニューから次の項目を選択します。

「**編集**」 > 「**オプション...**」

1. 「オプション」ダイアログ・ボックスで「**ピボット テーブル**」タブをクリックします。

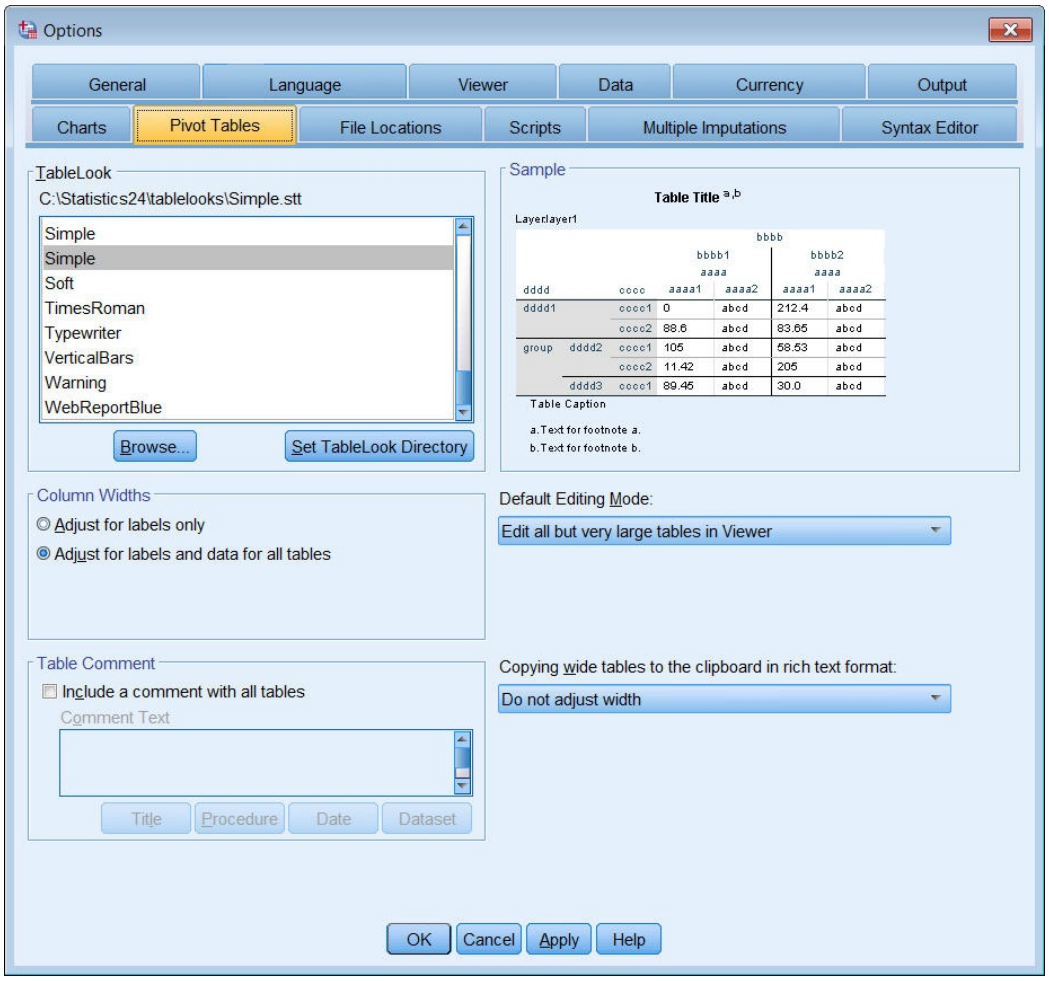


図 *51.* 「オプション」ダイアログ・ボックス

2. すべての新規テーブルに使用する「テーブルルック」スタイルを選択します。

   右側の「サンプル」ウィンドウに、各テーブルルックのプレビューが表示されます。

3. 「**OK**」をクリックして設定を保存し、ダイアログ・ボックスを閉じます。

デフォルトのテーブルルックの変更後に作成したテーブルはすべて、新しい書式設定規則に自動的に従います。

初期表示設定のカスタマイズ

初期表示設定には、「ビューアー」でのオブジェクトの位置合わせ、デフォルトでオブジェクトを表示するか非表示にするか、および「ビューアー」ウィンドウの幅などがあります。これらの設定を変更するには、次の手順を実行します。

1. メニューから次の項目を選択します。

   「**編集**」 > 「**オプション...**」

2.「**ビューアー**」タブをクリックします。

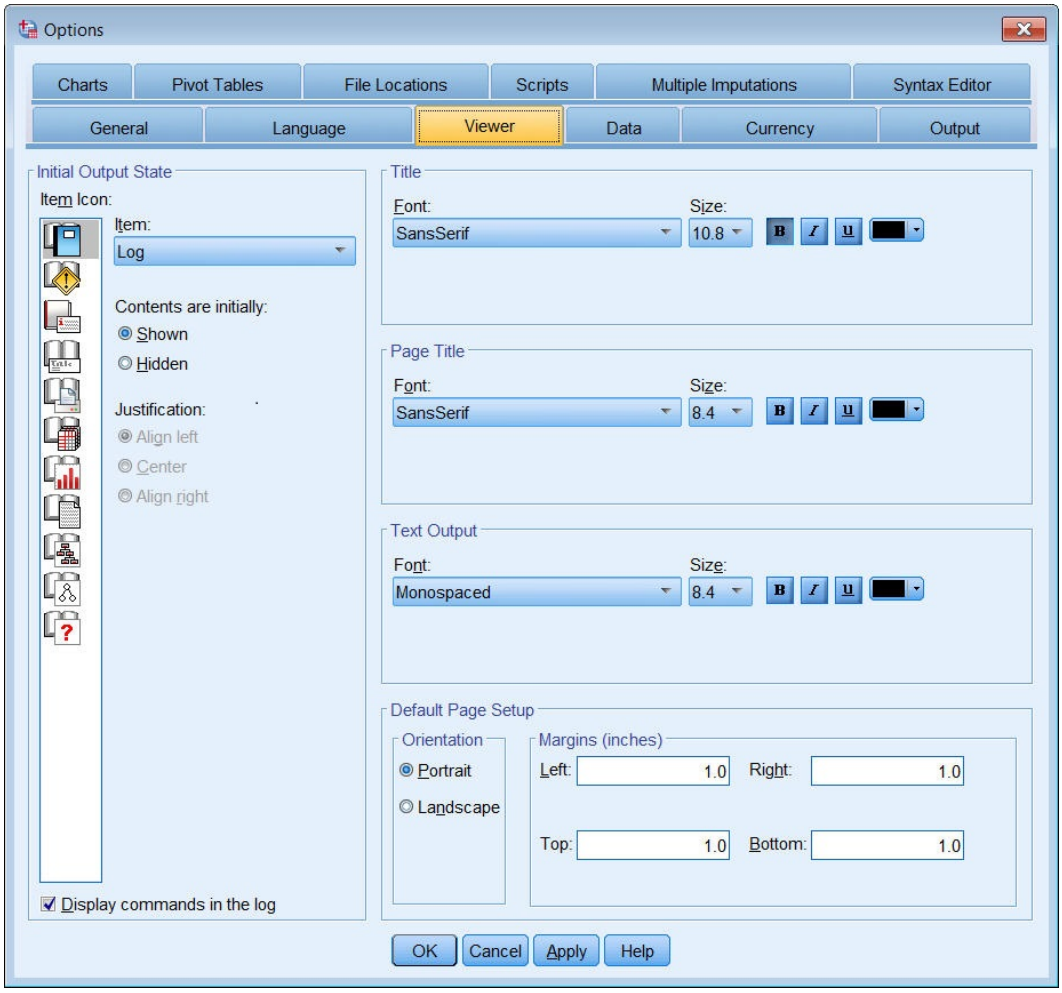


図 *52.* 「ビューアー」オプション

設定はオブジェクト別に適用されます。例えば、テーブルの表示方法を変更せずに、図表の表示方法をカスタマイズできます。カスタマイズするオブジェクトを単に選択し、変更を加えます。

3.「**表題**」アイコンをクリックして、その設定を表示します。

4.「**中央**」をクリックして、「ビューアー」の中央 (水平方向の中央) にすべての表題を表示します。

出力を乱雑にする可能性のあるログや警告メッセージなどの要素を非表示にすることもできます。アイコンをダブルクリックすると、そのオブジェクトの表示プロパティーが自動的に変更されます。

5.「**警告**」アイコンをダブルクリックして、出力内の警告メッセージを非表示にします。

6.「**OK**」をクリックして変更を保存し、ダイアログ・ボックスを閉じます。

変数ラベルおよび値ラベルの表示

ほとんどの場合、変数名または実際のデータ値を表示するより、変数と値のラベルを表示する方が効率的です。ただし、名前とラベルの両方の表示が必要な場合もあります。

1. メニューから次の項目を選択します。

   「**編集**」 > 「**オプション...**」

2.「**出力ラベル**」タブをクリックします。

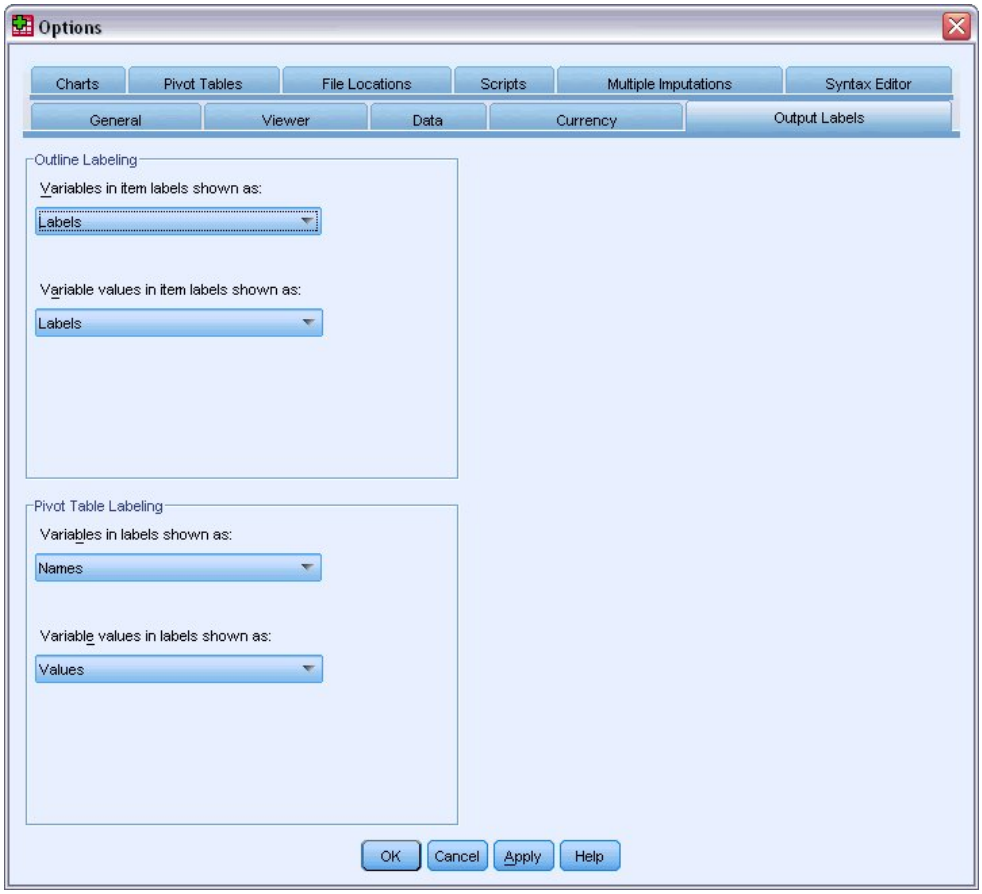


図 *53.* 「ピボット・テーブルのラベル付け」設定

アウトラインのウィンドウと内容のウィンドウにそれぞれ異なる設定を指定できます。例えば、アウトライン・ウィンドウにラベルを表示し、内容のウィンドウに変数名とデータ値を表示するには、次の手順を実行します。

3. 「ピボット・テーブルのラベル付け」グループで、「ラベル中の変数の表示」ドロップダウン・リストから「**名前**」を選択して、ラベルではなく変数名を表示します。
4. その後、「ラベル中の変数値の表示」ドロップダウン・リストから「**値**」を選択して、ラベルではなくデータ値を表示します。

これらの変更は、そのセッション内でこれ以降に生成されるテーブルに反映されます。

marital

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3224	50.4	50.4	50.4
	1	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

図 *54.* 表示される変数の名前および値

その他のアプリケーションでの結果の使用

結果はさまざまなアプリケーションで使用できます。例えば、プレゼンテーションやレポートにテーブルやグラフを組み込むことができます。

以下の例は Microsoft Word に固有の例ですが、他のワード・プロセッシング・アプリケーションでも同様に機能します。

Word の表としての結果の貼り付け

ピボット・テーブルを、ネイティブの Word の表として Word に貼り付けることができます。 すべてのテーブル属性 (フォント・サイズ、色など) が保持されます。 テーブルは Word の表形式で貼り付けられるため、Word で他の表と同様に編集できます。

1. 「ビューアー」でテーブルをクリックして選択します。
2. メニューから次の項目を選択します。

   「**編集**」 > 「**コピー**」
3. ワード・プロセッシング・アプリケーションを開きます。
4. ワード・プロセッサーのメニューから次の項目を選択します。

   「**編集**」 > 「**形式を選択して貼り付け...**」
5. 「形式を選択して貼り付け」ダイアログ・ボックスで「**リッチ・テキスト形式 (RTF)**」を選択します。
6. 「**OK**」をクリックして、結果を現在の文書に貼り付けます。

これで、文書内に表が表示されます。ニーズに応じて、カスタム書式設定を適用したり、データを編集したり、表のサイズを変更したりすることができます。

テキストとしての結果の貼り付け

ピボット・テーブルを、プレーン・テキストとして他のアプリケーションにコピーできます。この方法では、書式設定スタイルは保持されませんが、対象のアプリケーションに貼り付けた後で、テーブルのデータを編集できます。

1. 「ビューアー」でテーブルをクリックして選択します。
2. メニューから次の項目を選択します。

   「**編集**」 > 「**コピー**」
3. ワード・プロセッシング・アプリケーションを開きます。
4. ワード・プロセッサーのメニューから次の項目を選択します。

   「**編集**」 > 「**形式を選択して貼り付け...**」
5. 「形式を選択して貼り付け」ダイアログ・ボックスで「**書式なしテキスト**」を選択します。
6. 「**OK**」をクリックして、結果を現在の文書に貼り付けます。

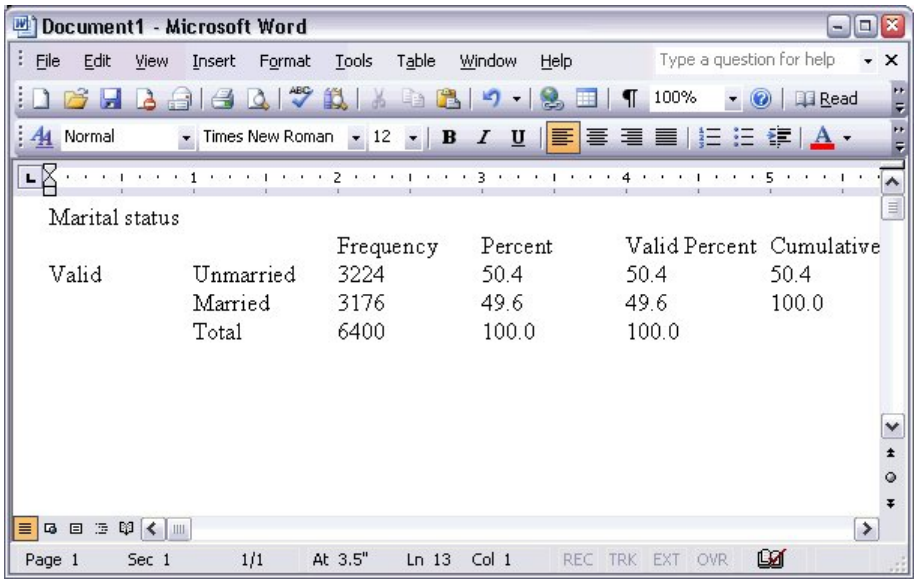


図 *55. Word* で表示されているピボット・テーブル

表の各列はタブで区切られます。列幅を変更するには、ワード・プロセッシング・アプリケーションでタブ・ストップを調整します。

Microsoft Word、PowerPoint、および Excel ファイルへの結果のエクスポート

Microsoft Word、PowerPoint、または Excel ファイルに結果をエクスポートできます。「ビューアー」で選択した項目およびすべての項目をエクスポートできます。このセクションでは、*msouttut.spv* ファイルおよび *demo.sav* ファイルを使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

注: PowerPoint へのエクスポートは、Windows オペレーティング・システムでのみ実行できます。また、Student 版では実行できません。

「ビューアー」のアウトライン・ペインで、エクスポートする特定の項目を選択するか、あるいはすべての項目またはすべての表示項目をエクスポートできます。

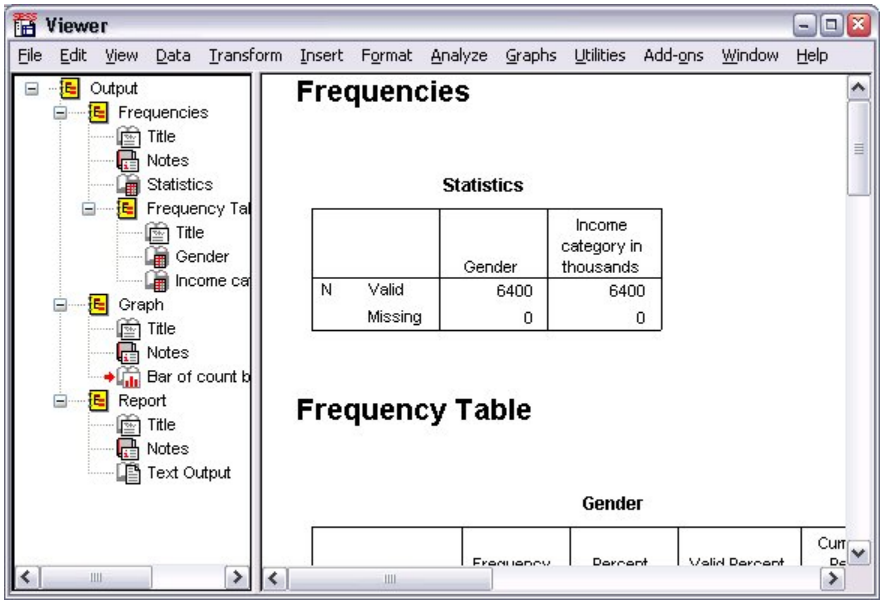


図 *56.* ビューアー

1. 「ビューアー」メニューから次の項目を選択します。

   「**ファイル**」 > 「**エクスポート...**」

   「ビューアー」内のすべてのオブジェクトをエクスポートするのではなく、表示されているオブジェクト (アウトライン・ペインで開いているブック) のみ、またはアウトライン・ペインで選択したオブジェクトのみをエクスポートするように選択できます。アウトライン・ペインで項目を選択しなかった場合は、選択したオブジェクトをエクスポートするためのオプションが 表示されません。

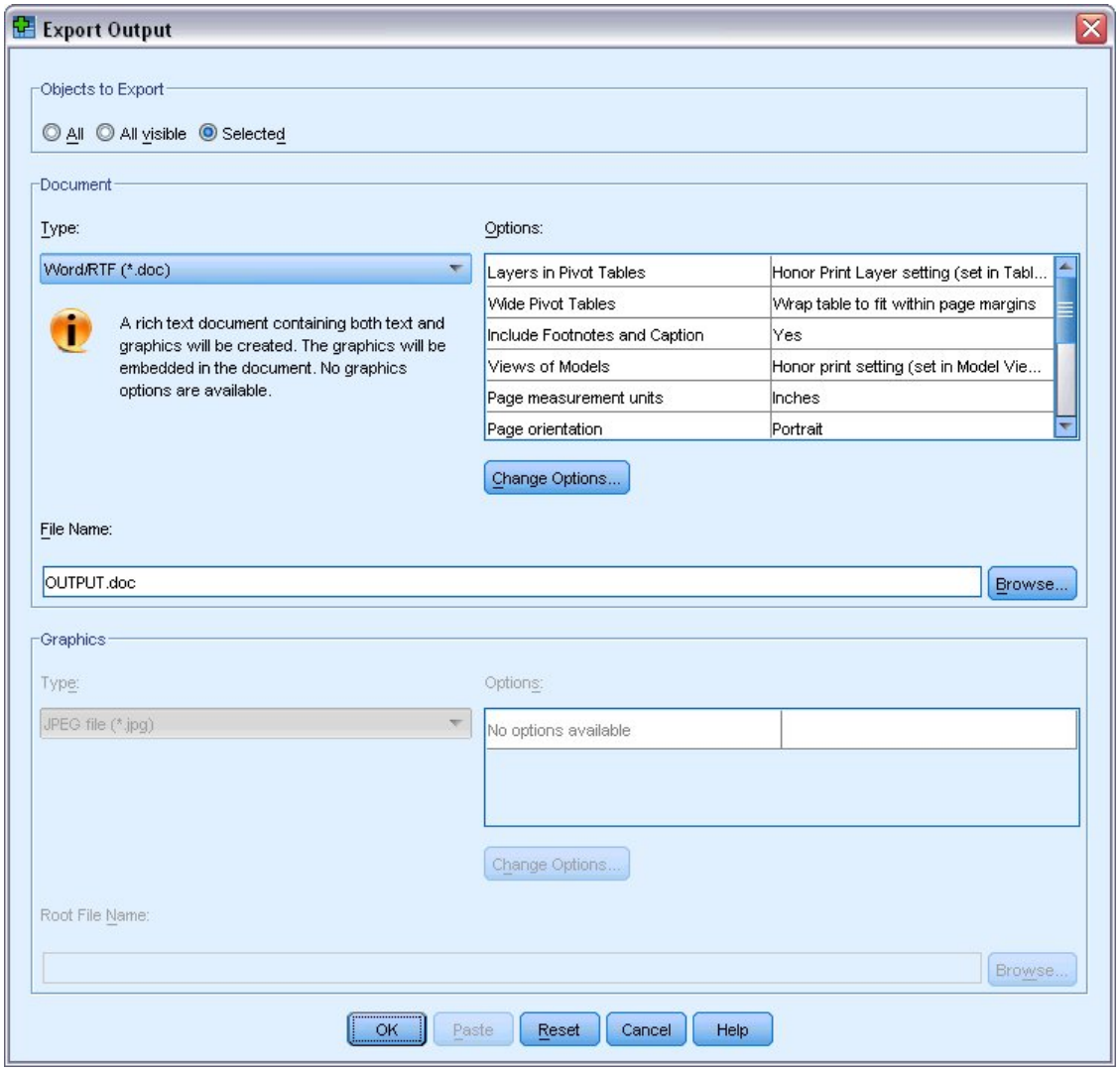
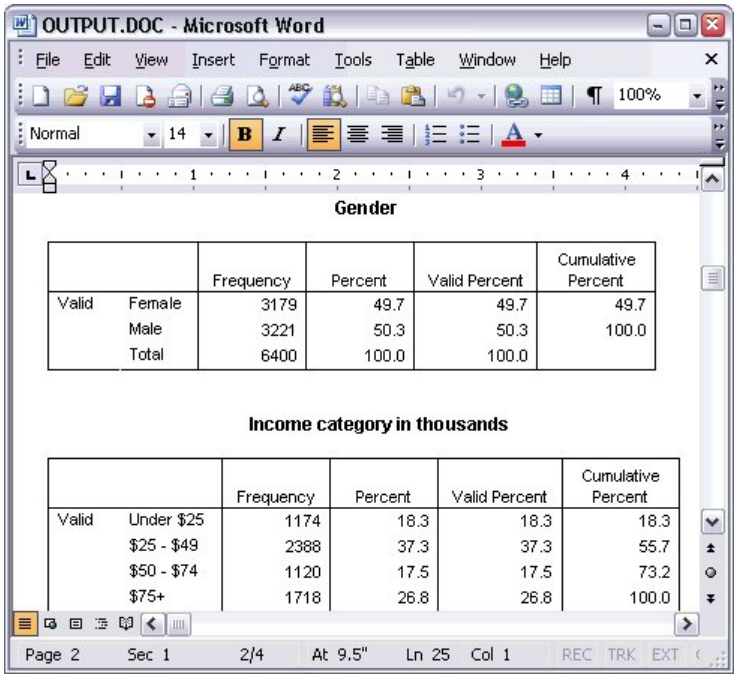


図 *57.* 「出力のエクスポート」ダイアログ・ボックス

2. 「エクスポートするオブジェクト」グループで「**すべて**」を選択します。
3. 「型」ドロップダウン・リストから「**Word/RTF (*.doc)**」を選択します。
4. 「**OK**」をクリックして、Word ファイルを生成します。

生成されたファイルを Word で開くと、結果がどのようにエクスポートされたかを確認できます。すべてのオブジェクトをエクスポートするように選択したため、記録オブジェクト (表示されるオブジェクトではない) が Word に表示されます。

ピボット・テーブルは、元のピボット・テーブルの書式設定 (フォント、色、枠線など) がすべて保持された Word テーブルになります。



OUTPUT.DOC - Microsoft Word

File Edit View Insert Format Tools Table Window Help

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3179	49.7	49.7	49.7
	Male	3221	50.3	50.3	100.0
	Total	6400	100.0	100.0	

Income category in thousands

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under $25	1174	18.3	18.3	18.3
	$25 - $49	2388	37.3	37.3	55.7
	$50 - $74	1120	17.5	17.5	73.2
	$75+	1718	26.8	26.8	100.0

図 *58. Word* で表示されているピボット・テーブル

図表は、Word 文書にグラフィックス・イメージとして組み込まれます。

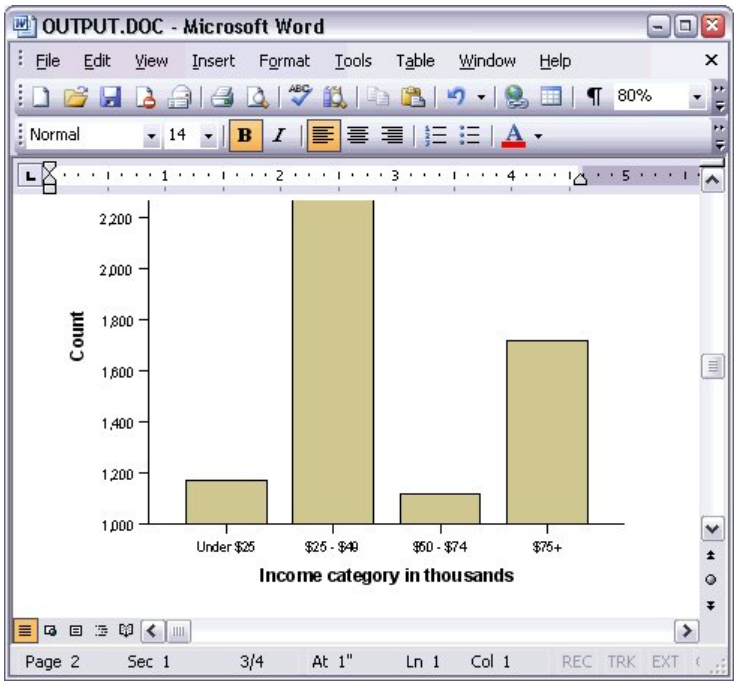


図 *59. Word* で表示されている図表

テキスト出力は、「ビューアー」でテキスト・オブジェクトに使用したフォントと同じフォントで表示されます。位置合わせが適切に行われるように、テキスト出力には固定ピッチ (モノスペース) フォントを使用する必要があります。

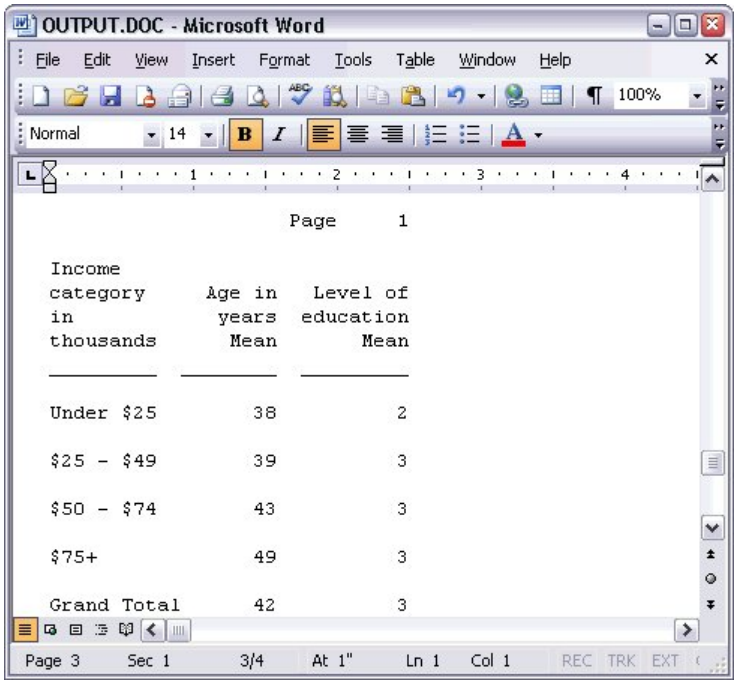


図 *60. Word* で表示されているテキスト出力

PowerPoint ファイルにエクスポートすると、エクスポートした各項目が別々のスライドに配置されます。PowerPoint にエクスポートされたピボット・テーブルは Word の表になりますが、元のピボット・テーブルの書式設定 (フォント、色、枠線など) はすべて保持されます。

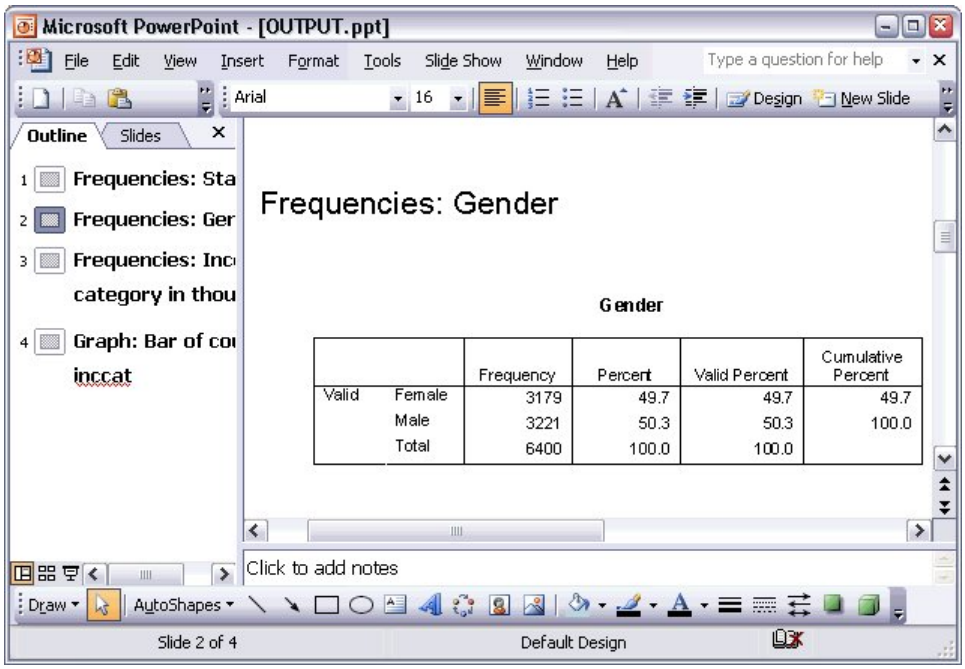


図 *61. PowerPoint* で表示されているピボット・テーブル

PowerPoint にエクスポートするために選択した図表は、PowerPoint ファイルに組み込まれます。

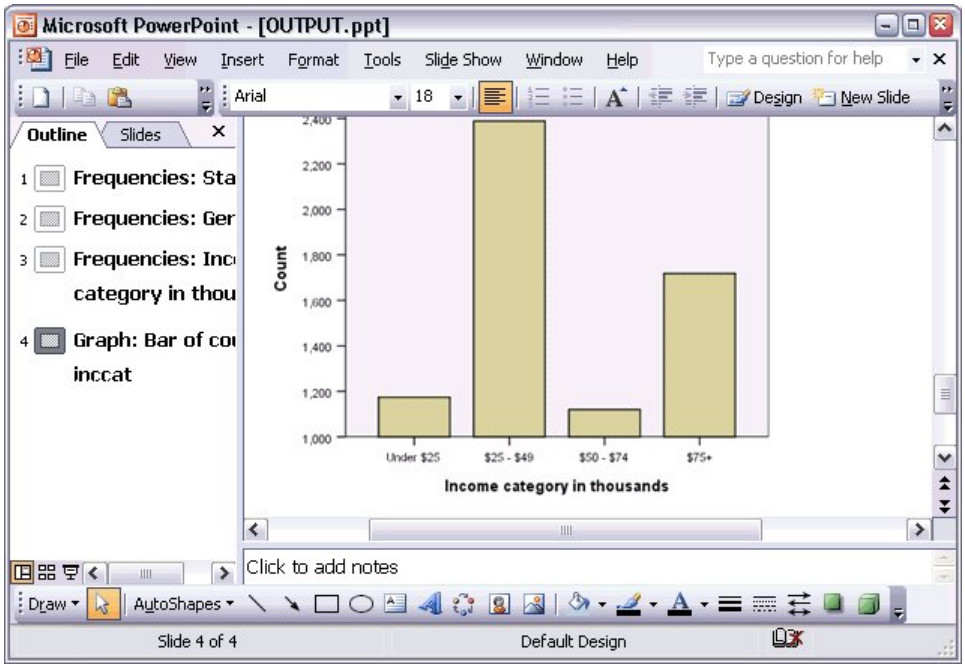
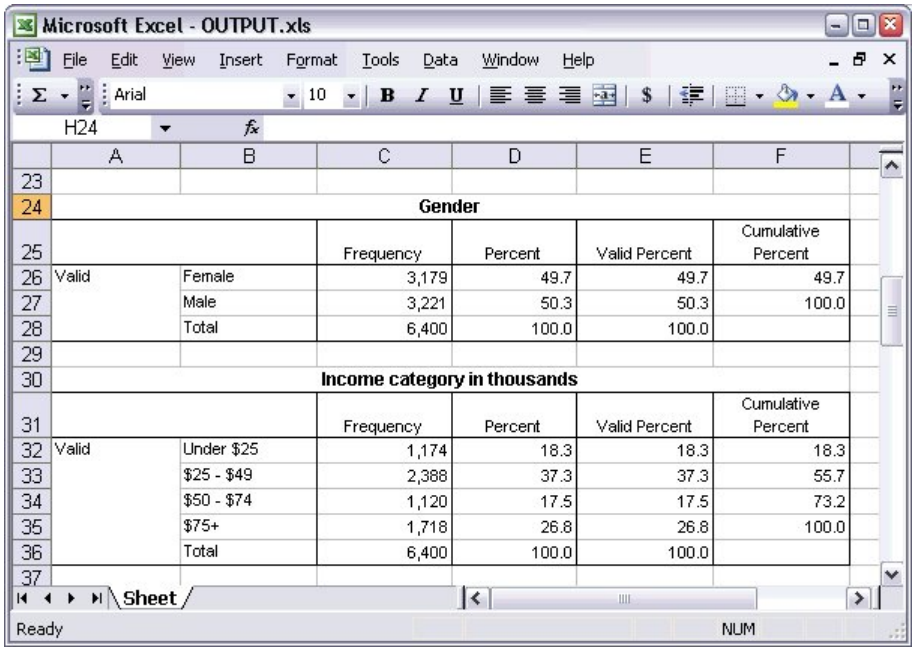


図 *62. PowerPoint* で表示されている図表

注: PowerPoint へのエクスポートは、Windows オペレーティング・システムでのみ実行できます。また、Student 版では実行できません。

Excel ファイルにエクスポートすると、結果が異なる方法でエクスポートされます。



	A	B	C	D	E	F
23						
24	Gender					
25			Frequency	Percent	Valid Percent	Cumulative Percent
26	Valid	Female	3,179	49.7	49.7	49.7
27		Male	3,221	50.3	50.3	100.0
28		Total	6,400	100.0	100.0	
29						
30	Income category in thousands					
31			Frequency	Percent	Valid Percent	Cumulative Percent
32	Valid	Under $25	1,174	18.3	18.3	18.3
33		$25 - $49	2,388	37.3	37.3	55.7
34		$50 - $74	1,120	17.5	17.5	73.2
35		$75+	1,718	26.8	26.8	100.0
36		Total	6,400	100.0	100.0	
37						

図 *63. Excel* で表示されている *OUTPUT.xls*

ピボット・テーブルの行、列、およびセルは、Excel の行、列、およびセルになります。

テキスト出力の各行が Excel ファイルの各 1 行に配置され、各行の内容全体がそれぞれ 1 つのセルに配置されます。

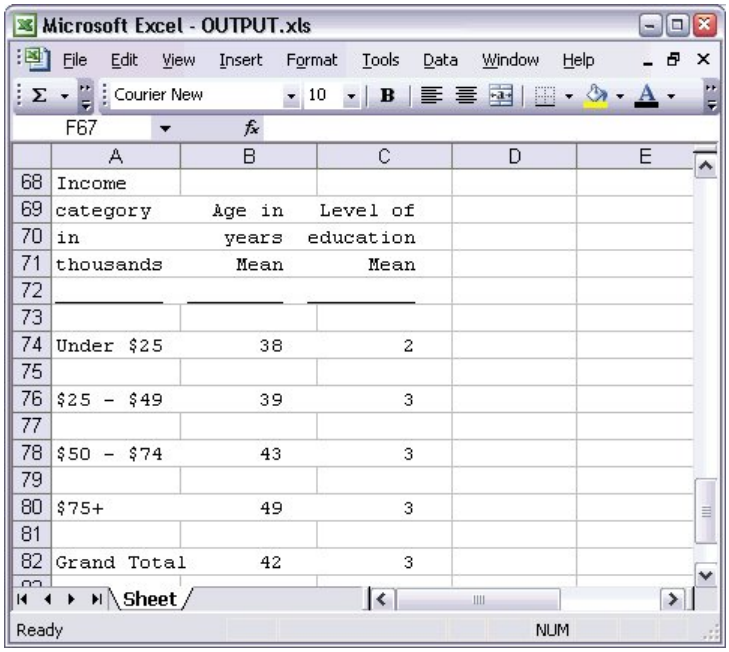


図 *64. Excel* で表示されているテキスト出力

結果の PDF へのエクスポート

ビューアー内のすべてまたは選択された項目を PDF (portable document format) ファイルにエクスポートできます。

1. PDF にエクスポートしたい結果を含む「ビューアー」ウィンドウのメニューから、次の項目を選択します。

   「**ファイル**」 > 「**エクスポート...**」

2. 「出力のエクスポート」ダイアログ・ボックスで、「エクスポート書式のファイルの種類」ドロップダウン・リストから 「**PDF 形式**」を選択します。

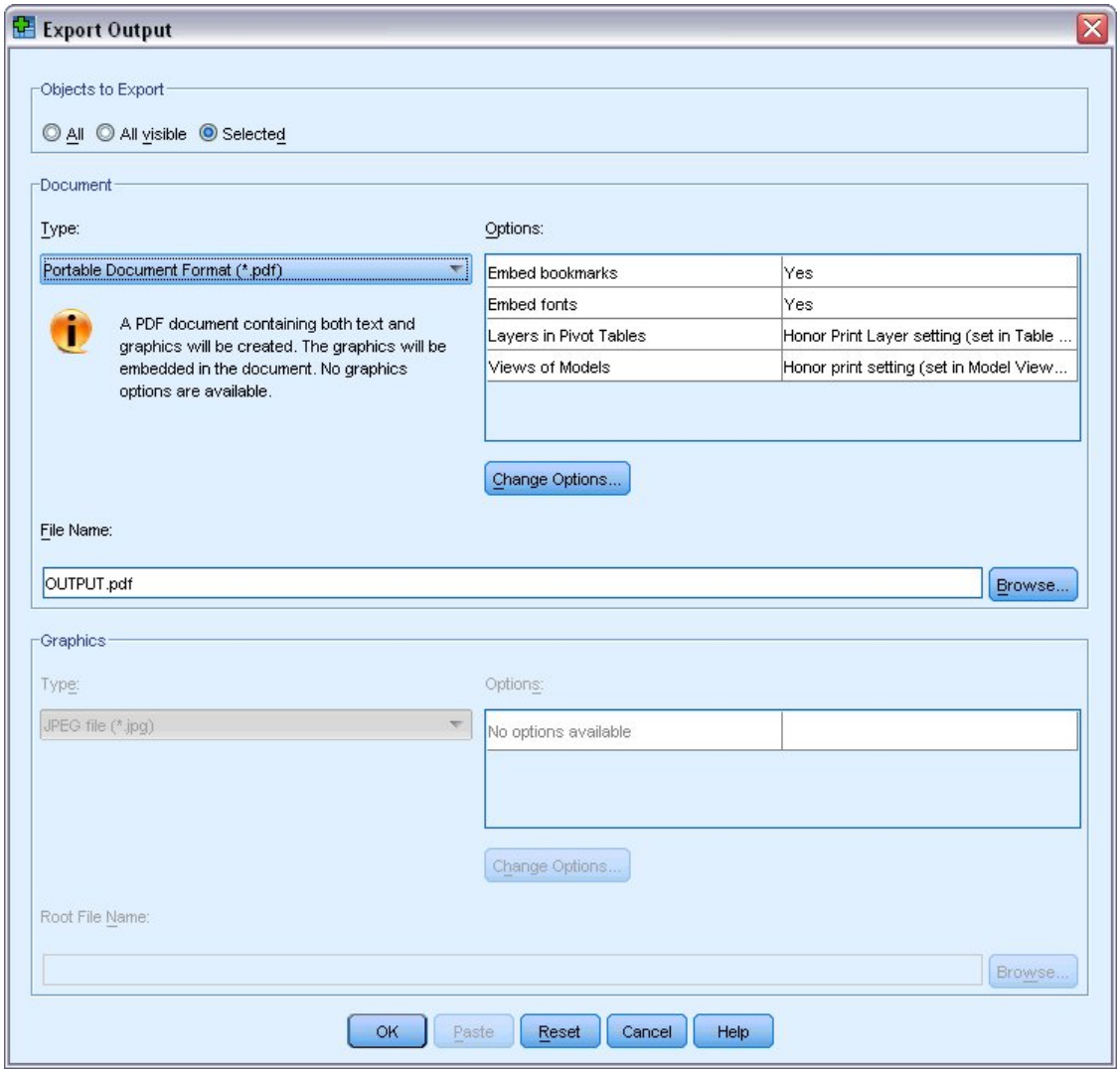


図 *65.* 「出力のエクスポート」ダイアログ・ボックス

- ビューアー文書のアウトライン・ペインは、ナビゲーションが簡単に行えるように、PDF ファイルではブックマークに変換されます。
- PDF 文書のページ・サイズ、方向、余白、ページのヘッダーとフッターの内容と表示、および図表の印刷サイズは、ページ設定オプション (「ビューアー」ウィンドウの「ファイル」メニューの「ページ設定」) で制御します。
- PDF 文書の解像度 (DPI) は、デフォルトまたは現在選択されているプリンタの現在の解像度設定となります (「ページ設定」で変更できます)。最大解像度は 1200 DPI です。プリンタの方が設定が高い場合、PDF 文書の解像度は 1200 DPI になります。注: 高解像度の文書を低解像度プリンタで印刷すると、粗悪な結果になる場合があります。

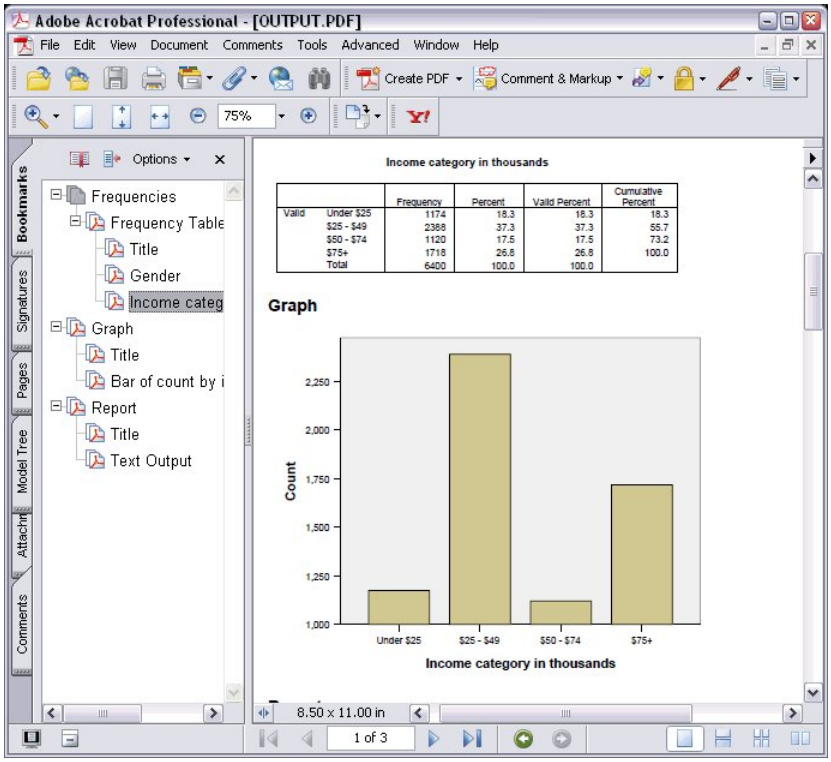


図 *66.* ブックマーク付きの *PDF* ファイル

結果の HTML へのエクスポート

結果は、HTML (ハイパーテキスト・マークアップ言語) にエクスポートすることもできます。HTML として保存すると、グラフィック以外のすべての出力は、単一の HTML ファイルにエクスポートされます。

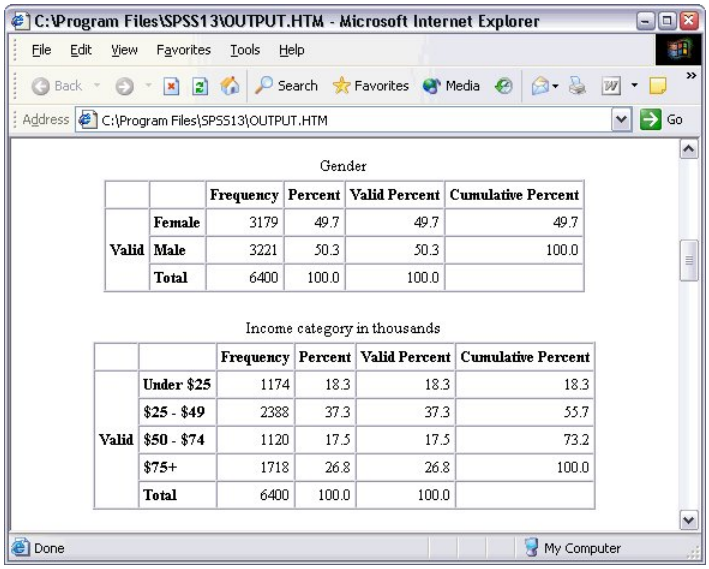


図 *67. Web* ブラウザーの *output.htm*

HTML にエクスポートする場合、図表をエクスポートすることも可能ですが、単一のファイルにエクスポートされるわけではありません。

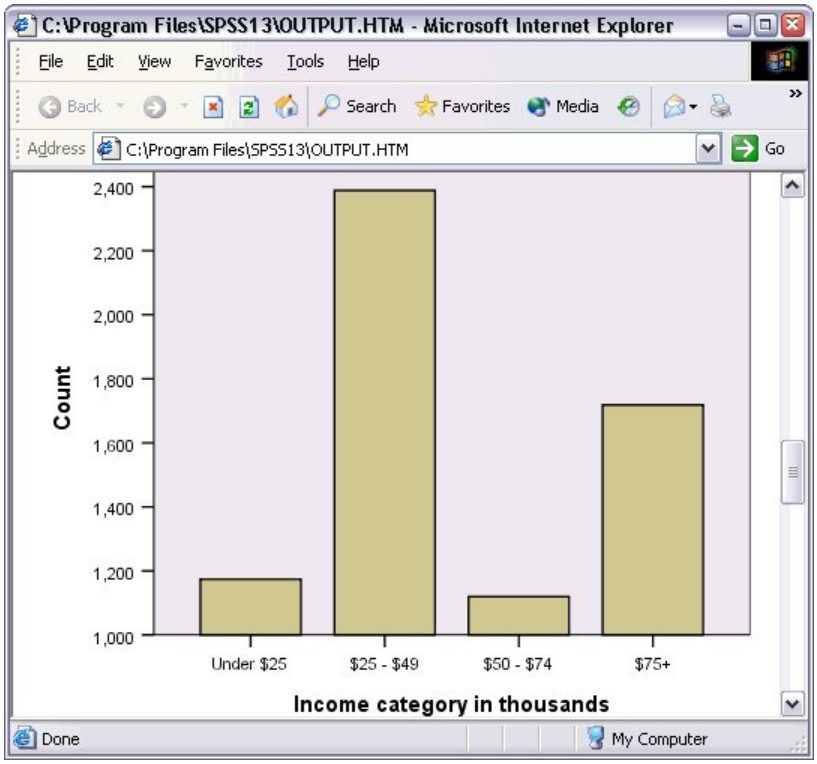


図 *68. HTML* の図表

各図表はユーザーが指定した形式のファイルで保存され、これらのグラフィック・ファイルへの参照が HTML に配置されます。すべての図表 (または選択した図表) を個別のグラフィック・ファイルにエクスポートするオプションもあります。

第 7 章 シンタックスの使用

強力なコマンド言語を使用することにより、多くの一般的なタスクを保存し、自動化できます。コマンド言語を使用すると、メニューやダイアログ・ボックスにはない機能も実行できます。コマンドのほとんどは、メニューやダイアログ・ボックスから利用できます。ただし、コマンド言語を使用しなければ利用できないコマンドやオプションもあります。コマンド言語を使用すると、シンタックス・ファイルにジョブを保存することもできるので、日を改めて分析を繰り返すことができます。

コマンド・シンタックス・ファイルは単純なテキスト・ファイルで、IBM SPSS Statistics シンタックス・コマンドを含んでいます。シンタックス・ウィンドウを開いて直接コマンドを入力することもできますが、多くの場合、作業の一部または全部をダイアログ・ボックスによって行った方が容易です。

この章の例では、データ・ファイル *demo.sav* を使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

注: コマンド・シンタックスは、Student 版では利用できません。

シンタックスを貼り付ける

シンタックスを作成する最も簡単な方法は、ほとんどのダイアログ・ボックスにある「貼り付け」ボタンを使用する方法です。

1. データ・ファイル *demo.sav* を開きます。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。
2. メニューから次の項目を選択します。

   「**分析**」 > 「**記述統計**」 > 「**度数分布表...**」
3. 「婚姻状況 *[婚姻状況]*」を選択し、変数リストに移動します。
4. 「**図表**」をクリックします。
5. 「図表」ダイアログ・ボックスで、「**棒グラフ**」を選択します。
6. 「図表の値」グループで、「**パーセンテージ**」を選択します。
7. 「**続行**」をクリックします。「**貼り付け**」をクリックし、ダイアログ・ボックスで選択した結果作成されたシンタックスをシンタックス・エディターにコピーします。

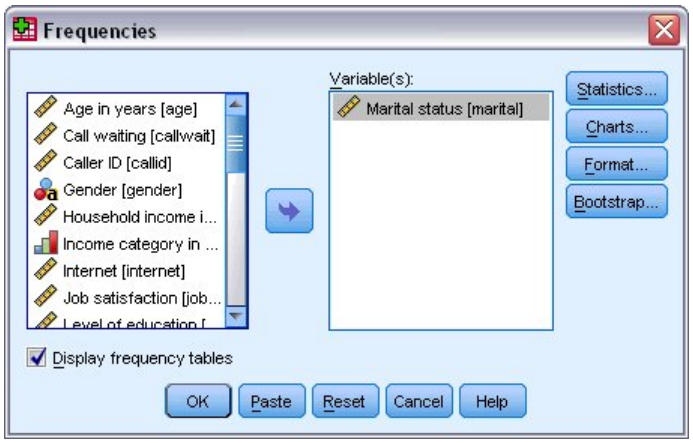


図 *69.* 「度数分布表」ダイアログ・ボックス

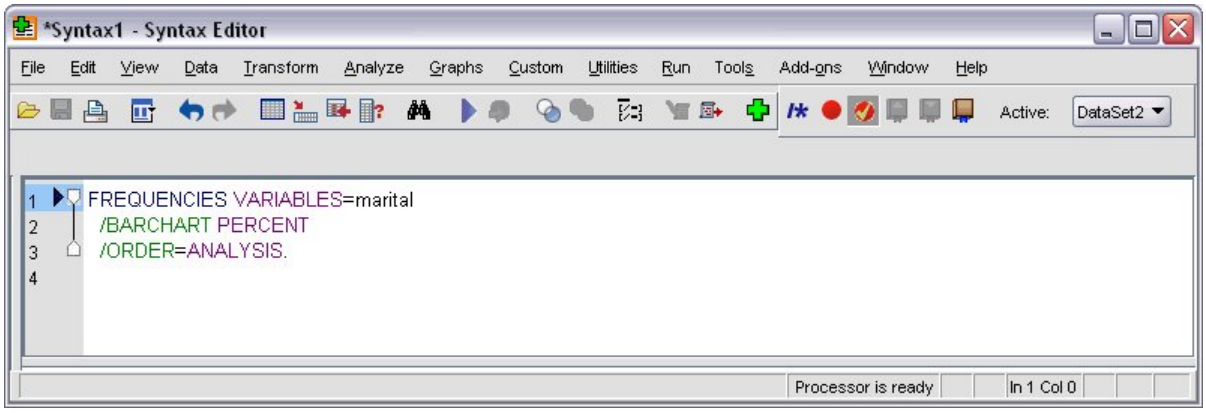


図 *70.* 「度数分布表」シンタックス

8. 現在表示されているシンタックスを実行するには、メニューから次の項目を選択します。

   「**実行**」 > 「**選択**」

シンタックスを編集する

シンタックス・ウィンドウでは、シンタックスを編集できます。例えば、サブコマンド /BARCHART を変更して、パーセンテージではなく度数を表示するようにできます (サブコマンドはスラッシュで示されます)。度数を表示するキーワードがわかっている場合は、キーワードを直接入力できます。キーワードがわからない場合、サブコマンド名の後ろの任意の場所にカーソルを置いて、Ctrl + スペース・バーを押すことで、サブコマンドに利用できるキーワードのリストを取得できます。これで、サブコマンドの自動入力コントロールが表示されます。

BARCHART のサブコマンドから PERCENT キーワードを削除します。

Ctrl + スペースキーを押します。

1. 度数を表す **FREQ** という項目をクリックします。自動入力コントロールで項目をクリックすると、現行カーソル位置にその項目が挿入されます。

   デフォルトでは、自動入力コントロールにより、文字を入力すると使用できる言葉の一覧が表示されます。例えば、棒グラフと同時に円グラフも使用したいとします。円グラフは、別のサブコマンドで指定します。

2. **FREQ** のキーワードの後で Enter キーを押し、サブコマンドの始まりを表すスラッシュを入力します。

シンタックス・エディターにより、現在のコマンドに対するサブコマンドのリストが表示されます。

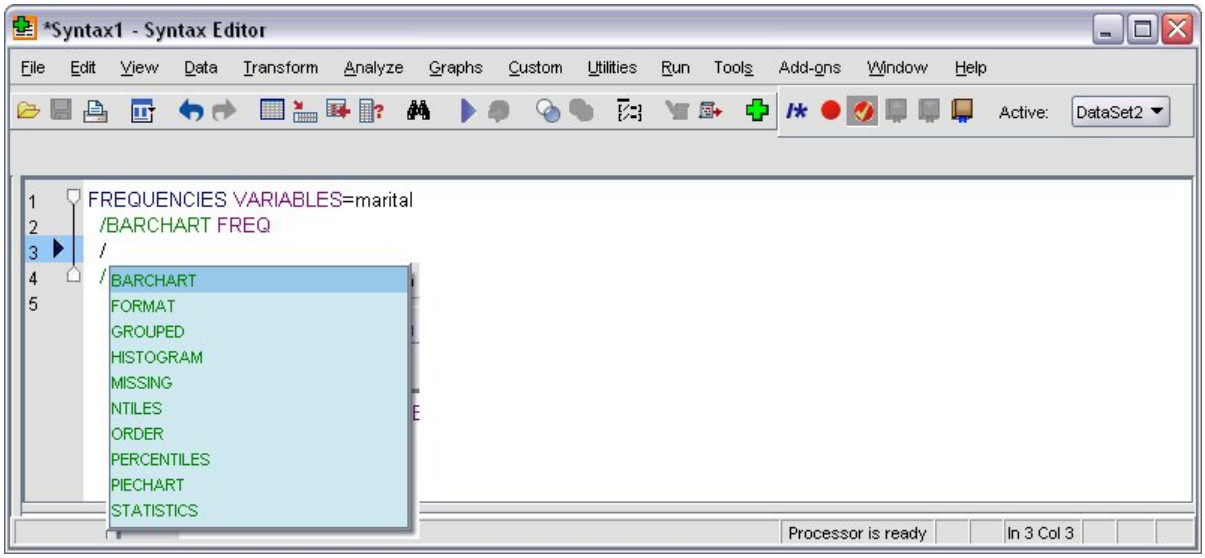


図 *71.* サブコマンドを表示する自動入力コントロール

現在のコマンドのより詳細なヘルプを取得するには、F1 キーを押します。これにより、現在のコマンドの、コマンド・シンタックス・リファレンス情報に直接移動することができます。

シンタックス・ウィンドウに表示されるテキストに、色が付けられています。コードに色を付けることによって、認識された言葉のみに色がつくため、認識されていない言葉を素早く特定できます。例えば、FORMAT サブコマンドのつづりを誤って FRMAT と入力したとします。サブコマンドはデフォルトでは緑色で表示されますが、FRMAT という文字は認識されていないため、色が付きません。

シンタックス・ファイルを開いて実行する

1. 保存済みのシンタックス・ファイルを開くには、メニューから次の項目を選択します。

   「**ファイル**」 > 「**開く**」 > 「**シンタックス...**」

   ファイルを開く際の標準のダイアログ・ボックスが表示されます。
2. シンタックス・ファイルを選択します。シンタックス・ファイルが表示されない場合は、「**シンタックス (*.sps)**」が表示するファイルの種類として選択されていることを確認します。
3. 「**開く**」をクリックします。
4. コマンドを実行するには、シンタックス・ウィンドウの「実行」メニューを使用します。

コマンドを特定のデータ・ファイルに適用する場合は、コマンドを実行する前にそのデータ・ファイルを開いておくか、そのデータ・ファイルを開くコマンドを追加する必要があります。この種類のコマンドは、データ・ファイルを開くダイアログ・ボックスから貼り付けることができます。

ブレークポイントの使用

ブレークポイントを使用すると、コマンド・シンタックスの実行をシンタックスの指定したポイントで停止させ、準備ができた時に実行を続けることができます。これにより、出力またはデータをシンタックス・ジョブの中間で表示することができ、また FREQUENCIES のように、現在のデータの状態についての情報を表示するコマンド・シンタックスを実行することができます。ブレークポイントは、コマンド内の特定の行ではなく、コマンドのレベルに対して設定できます。

コマンドにブレークポイントを挿入するには、次の手順を実行します。

1. コマンドに関連するテキストの左側の領域の任意の場所をクリックします。

   ブレークポイントがコマンド・テキストの左側の領域に赤い円で表示され、クリックした場所に関係なく、コマンド名と同じ行に表示されます。

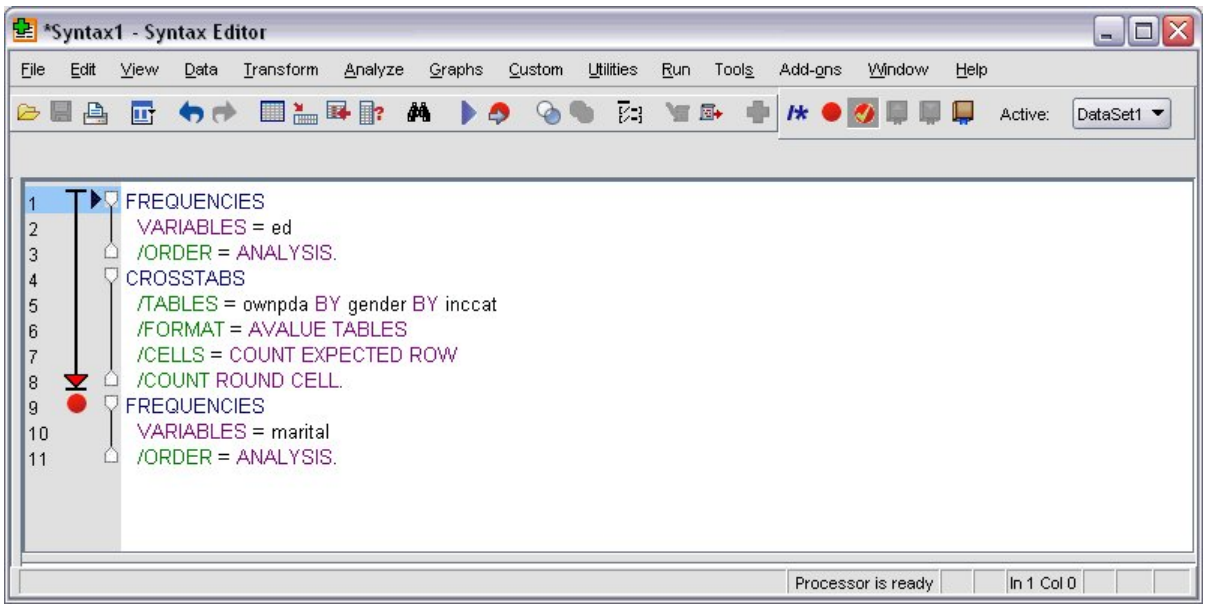


図 72. ブレークポイントで停止した実行

ブレークポイントを含むコマンド・シンタックスを実行する場合、ブレークポイントを含む各コマンドの前で実行は停止します。

コマンド・テキストの左側にある下向き矢印は、シンタックス実行の進行状況を示します。これは実行する最初のコマンドから最後のコマンドの領域に渡って表示され、ブレークポイントを含むコマンド・シンタックスを実行する際に特に便利です。

ブレークポイントの後で実行を再開するには、次の手順を実行します。

2. 「シンタックス・エディター」のメニューから、次の項目を選択します。

   「**実行**」 > 「**続行**」

第 8 章 データ値の変更

開始時のデータが、必要な分析またはレポートに最も有用な方法で構成されているとは限りません。例えば、次のような場合があります。

- スケール変数からカテゴリー変数を作成する。
- いくつかの応答カテゴリーを単一のカテゴリー内に結合する。
- 既存の 2 つの変数の差分を計算した新しい変数を作成する。
- 2 つの日付間の時間の長さを計算する。

この章では、データ・ファイル *demo.sav* を使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

スケール変数からカテゴリー変数を作成

実際は、データ・ファイル *demo.sav* にある複数のカテゴリー変数は、このデータ・ファイル内のスケール変数から派生しています。例えば、変数 *inccat* は、単に 4 つのカテゴリーにグループ化される *income* です。このカテゴリー変数は 1 から 4 の整数値を使用し、次の収入カテゴリー (千単位) ($25 より下、$25 から $49、$50 から $74、および $75 より上) を表現します。

カテゴリー変数 *inccat* を作成するには、以下の手順を実行します。

1. 「データ・エディター」ウィンドウのメニューから、次の項目を選択します。

   「**変換**」 > 「**連続変数のカテゴリー化...**」

   初めの「連続変数のカテゴリー化」ダイアログ・ボックスでは、新しいビン分割変数を作成するスケール変数または順序変数、あるいはその両方を選択します。**ビン分割**とは、2 つ以上の連続値を取得し、同じカテゴリーにグループ化することです。

   連続変数のカテゴリー化では、適切なビン分割の選択を行えるようにデータ・ファイル内の実際の値を使用するため、最初にデータ・ファイルを読み取る必要があります。データ・ファイルに大量のケースが含まれている場合、この処理に時間がかかることがあるため、この初期ダイアログ・ボックスでは、読み取る (「スキャン」する) ケースの数を制限できます。このサンプル・データ・ファイルでは、この処理は必要ありません。ファイルには 6,000 を超えるケースが含まれていますが、この数のケースをスキャンするのにそれほど時間はかかりません。

2. 世帯全体の収入 *(千ドル) [収入]* を「変数」リストから「ビン分割する変数」リストにドラッグ・アンド・ドロップして、「**続行**」をクリックします。

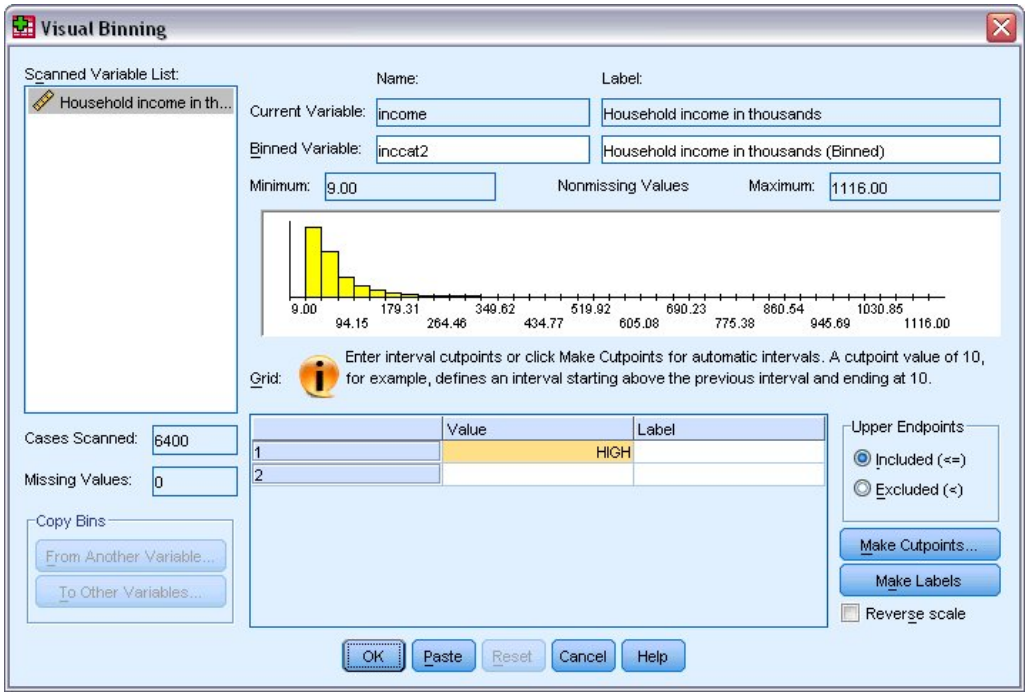


図 *73.* メインの「連続変数のカテゴリー化」ダイアログ・ボックス

3. メインの「連続変数のカテゴリー化」ダイアログ・ボックスで、スキャンされた変数のリストの世帯全体の収入 *(*千ドル*) [*収入*]* を選択します。

   選択した変数の分布がヒストグラムに表示されます (このケースでは、かなりの偏りが見られます)。

4. 新しいビン分割する変数の名前に「`inccat2`」、変数ラベルに「収入カテゴリー ［千ドル]」を入力します。
5. 「**分割点の作成**」をクリックします。
6. 「**等幅の区間**」を選択します。
7. 最初の分割点の位置に「25」、分割点の数に「3」、幅に「25」を入力します。

   ビン分割後のカテゴリーの数は、分割点の数に 1 を加えた数となります。したがってこの例の場合、新しくビン分割された変数には、各カテゴリーにそれぞれ 25 (千ドル) の範囲が含まれる最初の 3 つのカテゴリーと、分割点の最大値である 75 (千ドル) を超えるすべての値が含まれる最後のカテゴリーの、合計 4 つのカテゴリーがあります。

8. 「**適用**」をクリックします。

   グリッドに表示されている値は、定義済みの分割点を表現しており、各カテゴリーの終点の上限になります。また、ヒストグラムの垂直線は、分割点の位置を示します。

   デフォルトでは、これらの分割点の値は、対応するカテゴリーに含まれます。例えば、最初の値の 25 には、25 以下のすべての値が含まれます。ただしこの例では、25 より小さい、25 から 49、50 から 74、および 75 以上に対応するカテゴリーが必要です。

9. 終点の上限グループで、「**除外する (<)**」を選択します。
10. 「**ラベルの作成**」をクリックします。

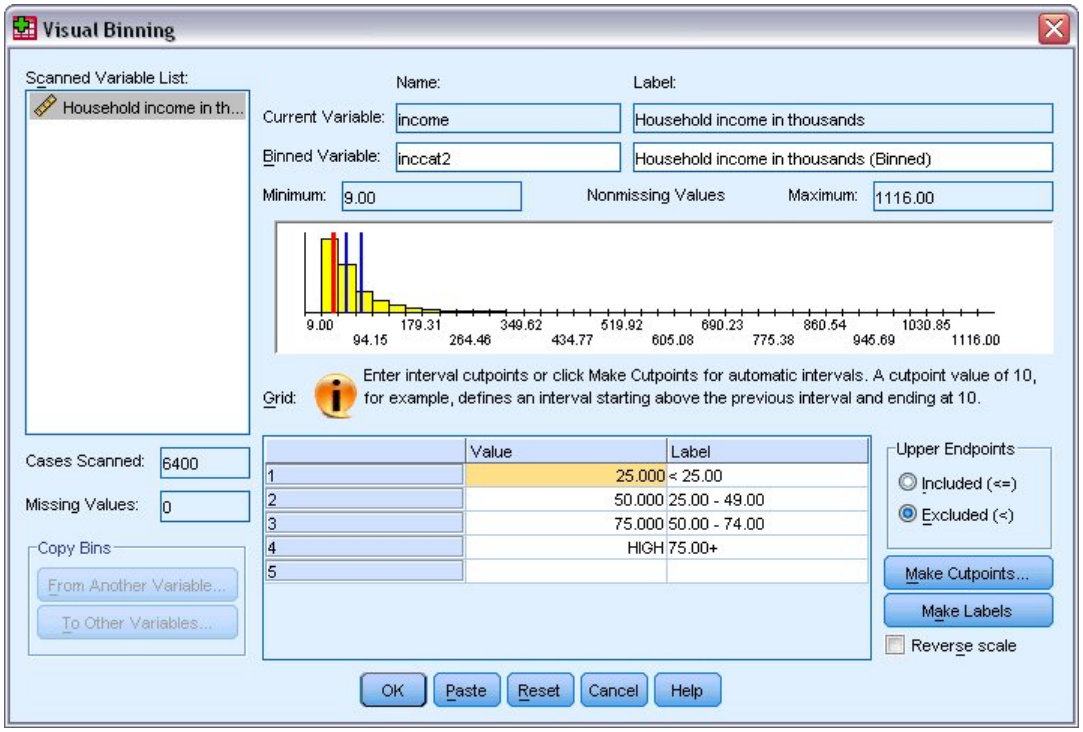


図 *74.* 自動生成された値ラベル

それぞれのカテゴリーに対して、記述的なラベルが自動生成されます。新しくビン分割する変数に割り当てられた実際の値は 1 から始まる連続する整数であるため、値ラベルを非常に有効に使用できます。

グリッド内の分割点およびラベルを手動で入力または変更できます。さらに、ヒストグラムにある分割点の線をドラッグ・アンド・ドロップして分割点の位置を変更したり、ヒストグラムの外に分割点の線をドラッグして分割点を削除することもできます。

11. 「**OK**」をクリックし、新しいビン分割する変数を作成します。

新しい変数が、データ・エディターに表示されます。変数はファイルの最後に追加されるため、「データ・ビュー」の一番右の列、および「変数ビュー」の最終行に表示されます。

新しい変数の計算

さまざまな数学関数を使用すると、高度で複雑な方程式に基づいて新しい変数を計算できます。ただし、この例では、既存の 2 つの変数の値の差分である新しい変数を計算します。

データ・ファイル *demo.sav* には、回答者の現在の年齢の変数と現在の仕事の就業年数の変数が含まれています。ただし、回答者が仕事を始めたときの年齢に対する変数は含まれていません。現在の年齢と現在の仕事の就業年数との差分を計算した新しい変数を作成することができ、これが回答者がその仕事を始めたおおよその年齢になります。

1. 「データ・エディター」ウィンドウのメニューから、次の項目を選択します。

   「**変換**」 > 「**変数の計算...**」

2. 目標変数に「`jobstart`」と入力します。
3. ソース変数で「年齢 *[年齢]*」を選択し、矢印ボタンをクリックして「数式」テキスト・ボックスにコピーします。
4. ダイアログ・ボックスの電卓パッドでマイナス (-) ボタンをクリックします (またはキーボードで「-」キーを押します)。
5. 「勤続年数 *[雇用]*」を選択し、矢印ボタンをクリックして式にコピーします。

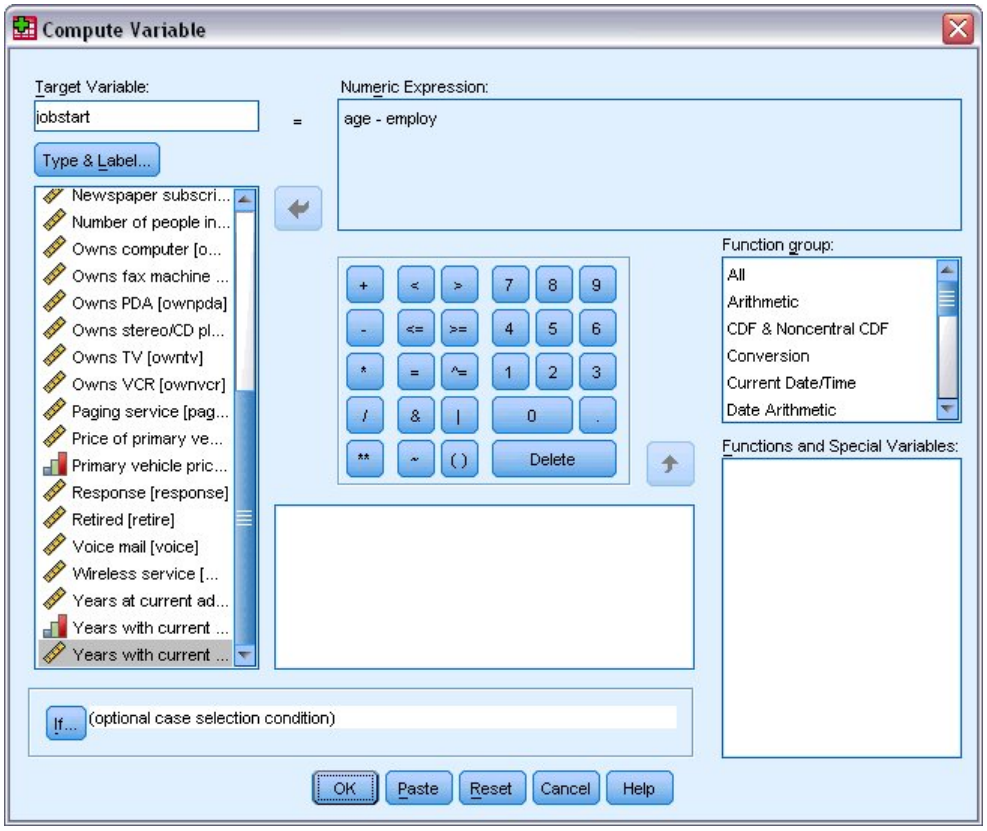


図 *75.* 「変数の計算」ダイアログ・ボックス

注: 正しい雇用変数を選択してください。必要のない 変数に再割り当てされたカテゴリー・バージョンもあります。数式は「年齢 – 勤続年数 *(empcat)*」ではなく、「年齢 – 雇用」である必要があります。

6.「**OK**」をクリックして、新しい変数を計算します。

新しい変数が、データ・エディターに表示されます。変数はファイルの最後に追加されるため、「データ・ビュー」の一番右の列、および「変数ビュー」の最終行に表示されます。

式での関数の使用

式で定義済み関数を使用することもできます。次のような、70 を超える組み込み関数を使用できます。

- 算術関数
- 統計関数
- 分布関数
- 論理関数
- 日付と時刻の集計および因子抽出関数
- 欠損値関数
- ケース間関数
- 文字列関数

関数は、算術演算を行うグループおよび統計的なメトリックを計算する別のグループなど、論理的に異なるグループから編成されています。利便性のために、*$TIME* (現在の日付と時刻) など多くの一般的に使用されるシステム変数も、適切な関数グループに含まれています。

式に関数を貼り付ける

式に関数を貼り付けるには、次の手順を実行します。

1. 式の中で関数を表示させる場所にカーソルを置きます。

2.「関数グループ」リストから適切なグループを選択します。「**すべて**」でラベル付けされたグループでは、すべての使用可能な関数およびシステム変数をリスト表示できます。
3.「関数と特殊変数」リスト内の関数をダブルクリックします (または、関数を選択し、「関数グループ」リストの隣にある矢印をクリックします)。

関数が式に挿入されます。式の一部を強調表示して関数を挿入する場合、式で強調表示した部分が関数の最初の引数として使用されます。

式の関数を編集

関数は、引数を入力するまで完成しません。引数は、貼り付けた関数に疑問符 (?) で表されます。疑問符 (?) の数は、関数を完成させるために必要な引数の最小数を示します。

1. 貼り付けた関数の疑問符 (?) を強調表示します。
2. 引数を入力します。引数が変数名の場合、変数リストから貼り付けることができます。

条件式の使用

条件式 (論理式とも呼ばれます) を使用して、選択したケースのサブセットに変換を適用することができます。条件式は、各ケースに True、False、または欠損値を返します。条件式の結果が True の場合、変換がそのケースに適用されます。結果が False または欠損値の場合、変換はケースには適用されません。

条件式を指定するには、次の手順を実行します。

1.「変数の計算」ダイアログ・ボックスで「**IF**」をクリックします。「ケースの選択」ダイアログ・ボックスが開きます。
2.「**IF 条件を満たしたケースを含む**」を選択します。
3. 条件式を入力します。

次のとおり、ほとんどの条件式に、少なくとも 1 つの関係演算子が含まれています。

```
age>=21
```

または

```
income*3<100
```

最初の例では、年齢 *[年齢]* の値が 21 以上のケースのみが選択されます。2 番目の例では、選択されるケースは、世帯全体の収入 *(千ドル) [収入]* の 3 倍が、100 より小さい必要があります。

また、次に示すように、論理演算子を使用して 2 つ以上の条件式をリンクさせることもできます。

```
age>=21 | ed>=4
```

または

```
income*3<100 & ed=5
```

最初の例では、年齢 *[年齢]* の条件または 教育のレベル *[教育]* の条件のいずれかに一致するケースが選択されます。2 番目の例では、選択されるケースは、世帯全体の収入 *(千ドル) [収入]* および 教育のレベル *[教育]* の両方の条件が一致している必要があります。

日付と時刻の使用

日付と時刻を使用して実行する多くの一般的なタスクは、日付と時刻ウィザードを使用して容易に実行できます。このウィザードを使用して、次の操作を実行できます。

- 日付または時刻を含む文字列変数から日付時刻変数を作成する。
- 日付または時刻の異なる部分が含まれた変数を結合することにより、日付時刻変数を作成する。
- 日付時刻変数に対して値の加算または減算を行う。これには、2 つの日付時刻変数の加算または減算も含まれます。
- 日付変数または時刻変数の一部を抽出する。例えば、mm/dd/yyyy 形式の日付時刻変数から、月初からの日数を抽出します。

このセクションの例では、データ・ファイル *upgrade.sav* を使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

日付と時刻ウィザードを使用するには、次の手順を実行します。

1. メニューから次の項目を選択します。

「**変換**」 > 「**日付と時刻ウィザード...**」

図 *76.* 日付と時刻ウィザードの導入画面

日付と時刻ウィザードの導入画面では、一般的なタスクのグループが表示されます。現行のデータに適用されていないタスクは無効です。例えば、データ・ファイル *upgrade.sav* には文字列変数が含まれていないため、文字列から日付変数を作成するタスクは無効です。

IBM SPSS Statistics の日付と時刻を初めて使用する場合は、「**日付と時刻の表記方法の詳細**」を選択し、「**次へ**」をクリックします。日付時刻変数の簡単な概要が説明された画面が表示され、「ヘルプ」ボタンからはより詳細な情報にリンクされています。

2 つの日付間の時間の長さの計算

日付に関する最も一般的なタスクは、2 つの日付間の時間の長さを計算することです。例として、あるソフトウェア企業が、各顧客が最後にアップグレードを購入してからの年数を特定することにより、アップグレード・ライセンスの購入について分析することに関心を持っているとします。データ・ファイル *upgrade.sav* には、各顧客が最後にアップグレードを購入した日付の変数は含まれていますが、購入してからの年数は含まれていません。最後のアップグレードの日付と次の製品リリースの日付との間の時間の長さを年単位で示す新しい変数を使用することで、この量を測定できます。

2 つの日付間の時間の長さを計算するには、次の手順を実行します。

1. 日付と時刻ウィザードの導入画面で「**日付と時刻で計算**」を選択し、「**次へ**」をクリックします。
2. 「**2 つの日付間の時間単位数の計算**」を選択し、「**次へ**」をクリックします。

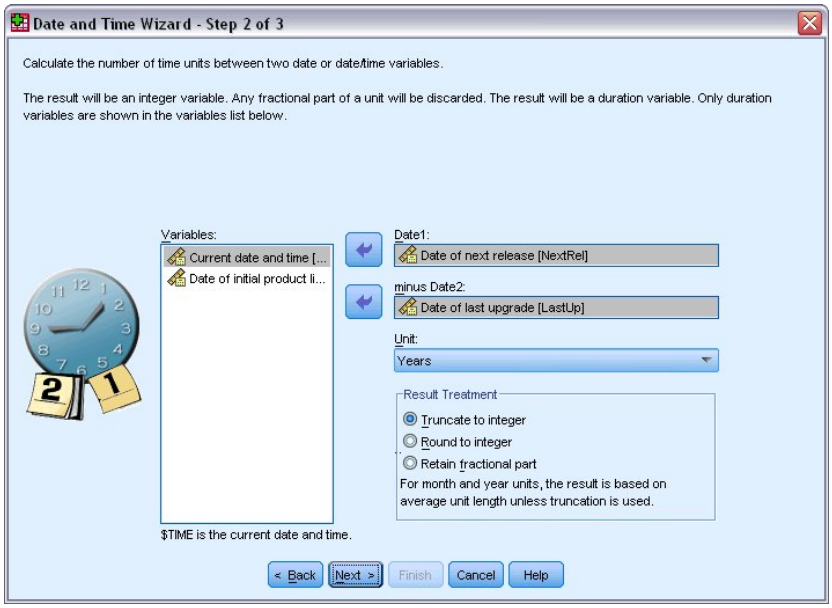


図 *77. 2* つの日付間の時間の長さの計算*:* ステップ *2*

3. ステップ 2 で、「日付 1」に「次のリリースの日付」を選択します。
4. 「引く日付 2」に「最後のアップグレードの日付」を選択します。
5. 「単位」に「**年**」および「結果の処置」に「**整数切り捨て**」を選択します (これらはデフォルトで選択されています)。
6. 「**次へ**」をクリックします。
7. ステップ 3 で、結果変数の名前に「*YearsLastUp*」と入力します。結果変数に既存の変数と同じ名前を付けることはできません。
8. 結果変数のラベルとして「最後のアップグレードからの年数」と入力します。結果変数の変数ラベルはオプションです。
9. デフォルトの「**今すぐ変数を作成する**」を選択したまま、「**完了**」をクリックし、新しい変数を作成します。

データ・エディターに表示された新しい変数「*YearsLastUp*」は、2 つの日付間の年の整数です。年の小数点以下は切り捨てられています。

日付への期間の加算

日付に対して 10 日または 12 カ月などの期間を加算または減算することができます。引き続き、前のセクションで使用したソフトウェア会社の例を使用し、各顧客の初期技術サポート契約が終了する日付の決定について検討します。データ・ファイル *upgrade.sav* には、契約したサポートの年数に対する変数および初期購入日付に対する変数が含まれています。そして、サポート年数を購入日付に加算することにより、初期サポートの終了日付を決定できます。

日付に期間を加算するには、次の手順を実行します。

1. 日付と時刻ウィザードの導入画面で「**日付と時刻で計算**」を選択し、「**次へ**」をクリックします。
2. ステップ 1 で、「**日付と期間の加算または減算**」を選択し、「**次へ**」をクリックします。

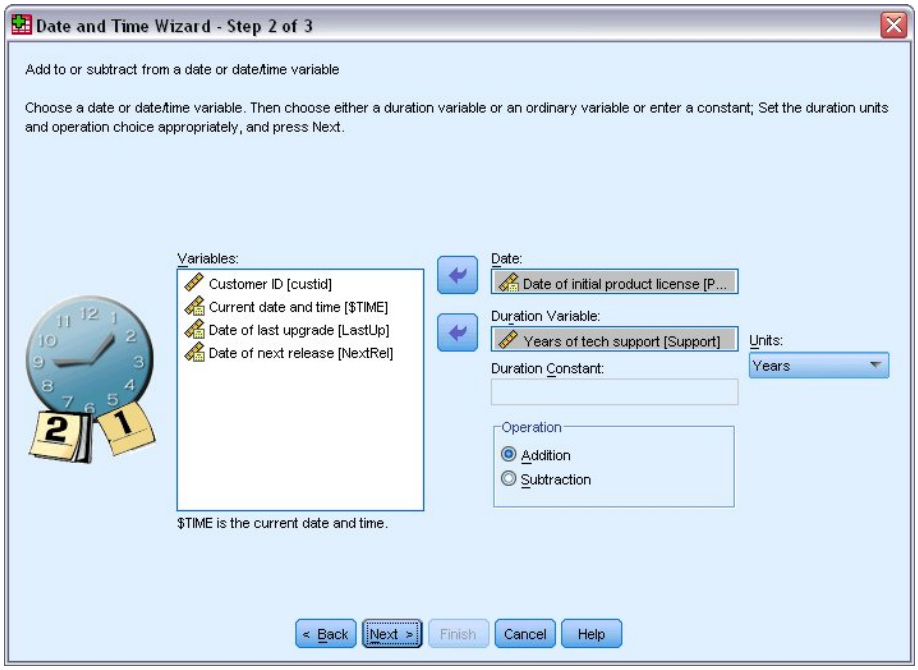


図 *78.* 日付への期間の加算*:* ステップ *2*

3. 日付に 最初の製品ライセンスの日付 を選択します。
4. ステップ 2 で、期間変数に テックサポートの期間 (年) を選択します。

   テックサポートの期間 は単に数値変数であるため、この変数を期間として加算する場合、使用する単位を指定する必要があります。
5. 「単位」ドロップダウン・リストから「**年**」を選択します。
6. 「**次へ**」をクリックします。
7. ステップ 3 で、結果変数の名前に「*SupEndDate*」と入力します。結果変数に既存の変数と同じ名前を付けることはできません。
8. サポート終了日 を結果変数のラベルとして入力します。結果変数の変数ラベルはオプションです。
9. 「**完了**」をクリックして、新しい変数を作成します。

新しい変数が、データ・エディターに表示されます。

第 9 章 データのソートおよび選択

データ・ファイルは必ずしも、特定のニーズに最適な形式で構成されているわけではありません。分析するデータを準備するには、以下の機能を含むさまざまなファイル変換から選択します。

- **データのソート**。1 つ以上の変数に基づいて、ケースをソートできます。
- **ケースのサブセットの選択**。ケースの 1 つのサブセットに限定して分析を行うか、またはさまざまなサブセットを同時に分析できます。

この章の例では、データ・ファイル *demo.sav* を使用します。詳しくは、75 ページの『第 10 章 サンプル・ファイル』のトピックを参照してください。

データのソート

ケースのソート (データ ファイルの行のソート) は、多くの場合役に立ち、特定のタイプの分析では必須となります。

1 つ以上のソート変数の値に基づいてデータ・ファイル内のケースを並べ替えるには、次の手順を実行します。

1. メニューから次の項目を選択します。

   「**データ**」 > 「**ケースの並べ替え...**」

   「ケースの並べ替え」ダイアログ・ボックスが表示されます。

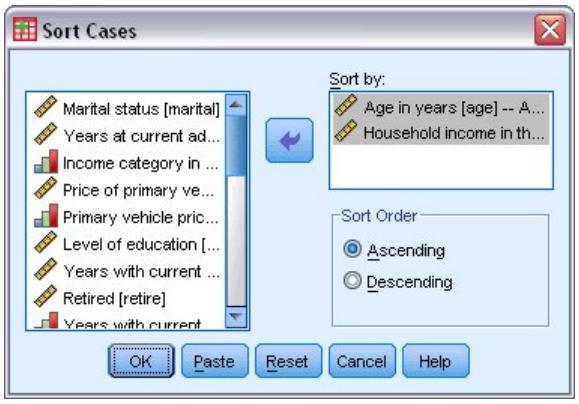


図 *79.* 「ケースの並べ替え」ダイアログ・ボックス

2. 「年齢 *[年齢]*」変数および「世帯全体の収入 *[収入]* 」変数を「並べ替え」リストに追加します。

複数のソート変数を選択した場合は、「並べ替え」リストに表示されている順序によって、ケースのソート順序が決まります。この例では、「並べ替え」リストの項目に基づいて、カテゴリー「年齢 *[年齢]*」内の値「世帯全体の収入 *[収入]*」を基準にケースがソートされます。文字列変数の場合、ソート順序で大文字は小文字より優先されます (例えば、ソート順序で文字列値 *Yes* は *yes* より前になります)。

ファイルの分割処理

分析用にデータ・ファイルを複数のグループに分割するには、次の手順を実行します。

1. メニューから次の項目を選択します。

   「**データ**」 > 「**ファイルの分割...**」

   「ファイルの分割」ダイアログ・ボックスが表示されます。

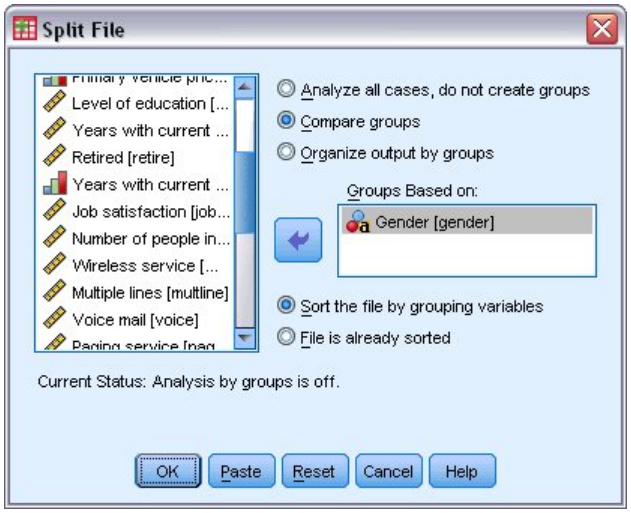


図 *80.* 「ファイルの分割」ダイアログ・ボックス

2. 「**グループの比較**」または「**グループごとの分析**」を選択します。(これらのステップに続く例で、これらの 2 つのオプションの相違点を説明します。)
3. 「性別 *[*性別*]*」を選択して、ファイルをこれらの変数の複数のグループに分割します。

グループ化変数として、数値、短い文字列、および長い文字列を使用できます。グループ化変数によって定義されたサブグループごとに、個別に分析が実行されます。複数のグループ化変数を選択した場合は、「グループ化変数」リストに表示されている順序によって、ケースをグループ化する方法が決まります。

「**グループの比較**」を選択した場合は、すべてのファイルの分割グループの結果が同じテーブルに組み込まれます。これは、「度数」手続きで生成された次の要約統計の表に示されています。

Statistics

Household income in thousands

Female	N	Valid	3179
		Missing	0
	Mean		68.7798
	Median		44.0000
	Std. Deviation		75.73510
Male	N	Valid	3221
		Missing	0
	Mean		70.1608
	Median		45.0000
	Std. Deviation		81.56216

図 *81. 1* つのピボット・テーブルを使用したファイルの分割出力

「**グループごとの分析**」を選択し、「度数」手続きを実行した場合は、2 つのピボット・テーブルが作成されます。1 つは女性用、もう 1 つは男性用です。

Statistics[a]

Household income in thousands

N	Valid	3179
	Missing	0
Mean		68.7798
Median		44.0000
Std. Deviation		75.73510

a. Gender = Female

図 *82.* 女性用ピボット・テーブルを含むファイルの分割出力

Statistics[a]

Household income in thousands

N	Valid	3221
	Missing	0
Mean		70.1608
Median		45.0000
Std. Deviation		81.56216

a. Gender = Male

図 *83.* 男性用ピボット・テーブルを含むファイルの分割出力

ファイル分割処理のケースのソート

「ファイルの分割」手続きにより、いずれかのグループ化変数の異なる値が検出されるたびに、新規サブグループが作成されます。 したがって、ファイルの分割処理を呼び出す前に、グループ化変数の値に基づいてケースをソートすることが重要になります。

デフォルトでは、「ファイルの分割」によって、グループ化変数の値に基づき、データ・ファイルが自動的にソートされます。ファイルが既に適切な順序でソートされている場合は、「**ファイルはすでに並び替え済み**」を選択すると、処理時間を短縮できます。

ファイルの分割処理をオンとオフを切り替える

ファイルの分割処理を呼び出したら、オフにしない限り、そのセッション中で有効なままになります。

- 「**全てのケースを分析**」。このオプションを選択すると、ファイルの分割処理はオフになります。
- 「**グループの比較**」および「**グループごとの分析**」。これらのオプションを選択すると、ファイルの分割処理がオンになります。

ファイルの分割処理が有効である場合は、アプリケーション・ウィンドウの下部のステータス・バーに「**分割ファイル・オン**」というメッセージが表示されます。

ケースのサブセットの選択

変数や複合式などの基準に基づいて、特定のサブグループの分析のみを実行することができます。ケースを無作為に抽出することもできます。サブグループを定義するために使用する基準には、次のものを使用できます。

- 変数値および範囲
- 日時範囲
- ケース (行) 番号
- 演算式
- 論理式
- 関数

分析するケースのサブセットを選択するには、次の手順を実行します。

1. メニューから次の項目を選択します。

   「**データ**」 > 「**ケースの選択...**」

これにより、「ケースの選択」ダイアログ・ボックスが開きます。

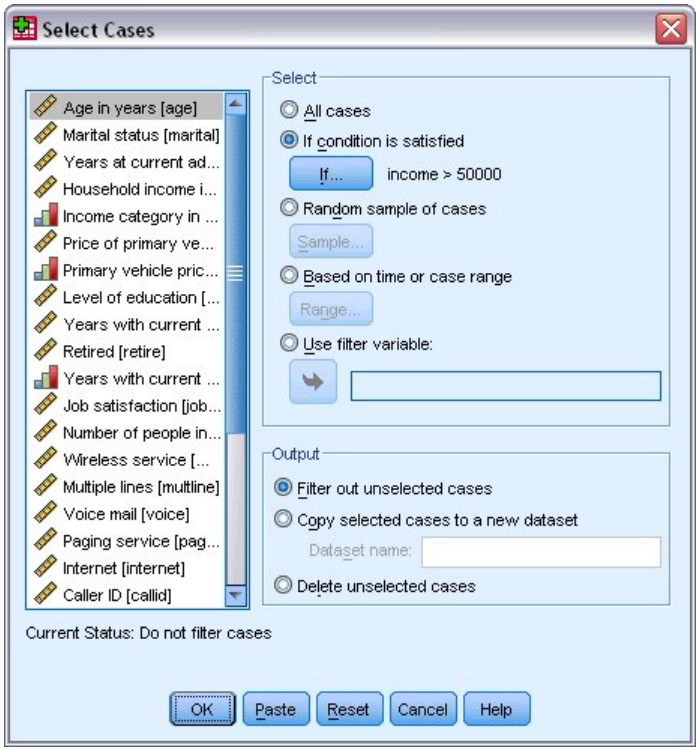


図 *84.* 「ケースの選択」ダイアログ・ボックス

条件式に基づいたケースの選択

条件式に基づいてケースを選択するには、次の手順を実行します。

1. 「**IF 条件が満たされるケース**」を選択し、「ケースの選択」ダイアログ・ボックスで「**If**」をクリックします。

これにより、「ケースの選択: IF 条件の定義」ダイアログ・ボックスが開きます。

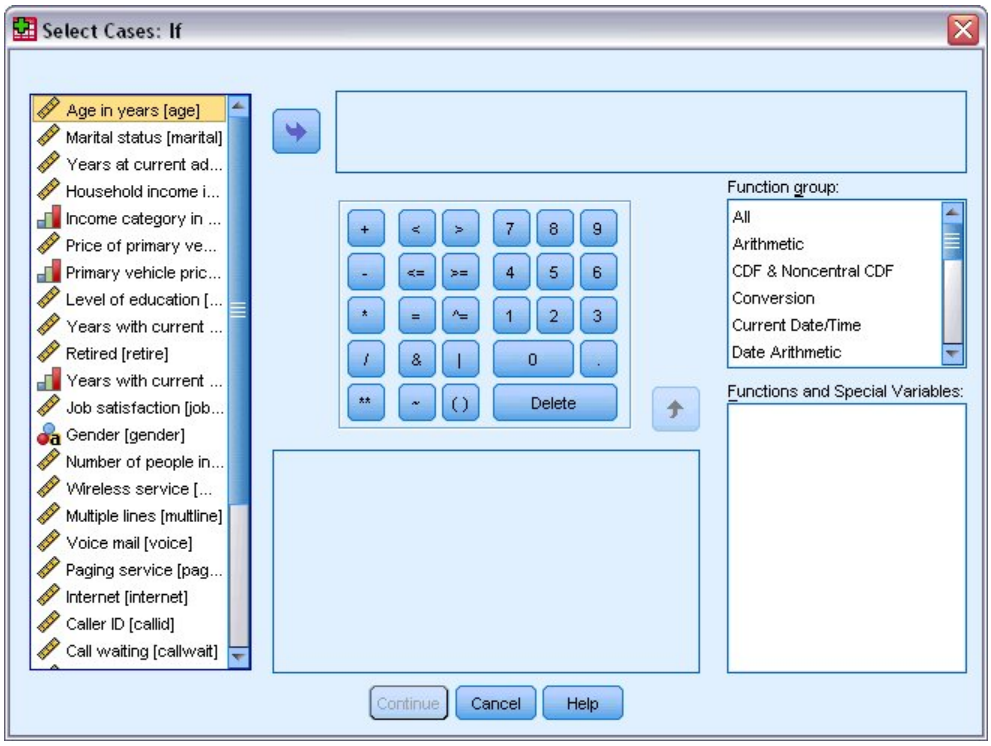


図 *85.* 「ケースの選択: *IF* 条件の定義」ダイアログ・ボックス

条件式では、既存の変数名、定数、算術演算子、論理演算子、関係演算子、および関数を使用できます。出力ウィンドウのテキストと同様に、テキスト・ボックスで式を入力および編集できます。電卓パッド、変数リスト、および関数リストを使用して、要素を式に貼り付けることもできます。詳しくは、65 ページの『条件式の使用』のトピックを参照してください。

無作為抽出の選択

無作為に抽出するには、次のようにします。

1. 「ケースの選択」ダイアログ・ボックスで「**ケースの無作為抽出**」を選択します。
2. 「**サンプル**」をクリックします。

これにより、「ケースの選択: 無作為抽出」ダイアログ・ボックスが開きます。

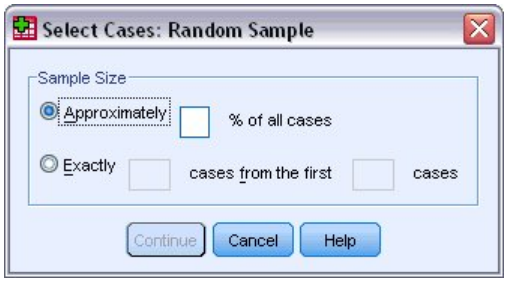


図 *86.* 「ケースの選択*:* 無作為抽出」ダイアログ・ボックス

以下のいずれかのサンプル・サイズを選択できます。

- 「**全ケースから約**」。ユーザー指定のパーセンテージ。このオプションでは、指定したパーセンテージに近いケースが無作為に抽出されます。
- 「**正確に**」。ユーザー指定のケース数。抽出するケースの数も指定する必要があります。この 2 番目の数は、データ・ファイル内のケースの合計数以下である必要があります。 この数がデータ・ファイル内のケースの合計数を超えた場合は、要求された数より少ないケースから抽出が行われます。

時間範囲またはケース範囲の選択

日付、時刻、または観測 (行) 番号に基づいてケースの範囲を選択するには、次の手順を実行します。

1. 「**日付かケース番号の範囲**」を選択し、「ケースの選択」ダイアログ・ボックスで「**範囲**」をクリックします。

   これにより、「ケースの選択: 範囲の定義」ダイアログ・ボックスが開き、観測 (行) 番号の範囲を選択できます。

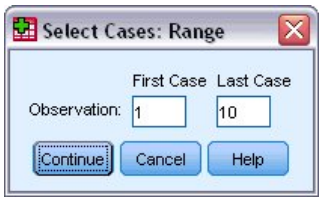


   図 *87.* 「ケースの選択*:* 範囲の定義」ダイアログ・ボックス

   - 「**最初のケース**」。範囲の開始日時値を入力します。日付変数が定義されていない場合は、開始観測番号 (「ファイルの分割」がオンでない場合、データ・エディターの行番号) を入力します。「最後のケース」値を指定しない場合は、開始日時から時系列の終わりまでのすべてのケースが選択されます。
   - 「**最後のケース**」。範囲の終了日時値を入力します。日付変数が定義されていない場合は、終了観測番号 (「ファイルの分割」がオンでない場合、データ・エディターの行番号) を入力します。「最初のケース」値を指定しない場合は、時系列の開始から終了日時までのすべてのケースが選択されます。

   定義済み日付変数を使用した時系列データの場合は、定義済み日付変数に基づいて日時の範囲を選択します。 各ケースは異なる時点の観測を表し、ファイルは日時順にソートされます。

図 *88.* 「ケースの選択*:* 範囲の定義」ダイアログ・ボックス *(*時系列*)*

時系列データの日付変数を生成するには、次の手順を実行します。

2. メニューから次の項目を選択します。

「**データ**」 > 「**日付の定義...**」

未選択のケースの処理

未選択のケースを処理するには、以下のいずれかの方法を選択できます。

- 「**選択されなかったケースを分析から除外**」。未選択のケースは、分析には含まれませんが、データ・セットには残されます。フィルター処理をオフにした場合は、未選択のケースをそのセッション中に後で使用できます。無作為に抽出する場合、または条件式に基づいてケースを選択する場合は、*filter_$* という名前の変数が生成され、選択されたケースの場合は値が 1 に設定され、未選択のケースの場合は値が 0 に設定されます。
- 「**選択されたケースを新しいデータセットにコピー**」。選択されたケースは新しいデータ・セットにコピーされます。元のデータ・セットは影響を受けずに残ります。選択されないケースは、新しいデータ・セットには含まれず、元の状態で元のデータ・セットに残ります。
- 「**選択されなかったケースを削除**」。未選択のケースがデータ・セットから削除されます。削除されたケースを復旧するには、変更を保存せずにファイルを閉じてから、再び開きます。変更をデータファイルに保存した場合は、ケースが永久に削除されます。

注: 未選択のケースを削除し、ファイルを保存した場合は、ケースを復旧できません。

ケース選択の状況

ケースのサブセットを選択したが、未選択のケースを破棄していない場合は、「データ・エディター」で未選択のケースの行番号に斜線が付けられます。

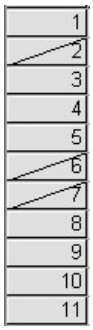


図 *89.* ケース選択の状況

第 10 章 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、年齢カテゴリー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、それに関係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 [1] 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。
- **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。
- **behavior.sav**: 典型的な例 [2] 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す。
- **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。

---

[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.

- **breakfast.sav**: 典型的な調査[3]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。
- **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。
- **car_insurance_claims.sav**: これは、他の場所[4]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[5]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* で説明したものと同じ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別するための、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。
- **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。

---

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

- **coffee.sav**: これは、6 つのアイス・コーヒー・ブランド[6] 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。
- **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。
- **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。
- **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が、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」[7] 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。

---

[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.

[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.

- **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[8] 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、週ごとに変わ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。
- **guttman.sav**: Bell [9] は、予想される社会グループを示す表を作成しました。Guttman [10] 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傍性、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傍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 [11] 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」が取得されたことになります。各ソースは、15×15 の近傍行列に対応しています。この近傍行列のセルの数は、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になります。
- **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。

---

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

[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.

- **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。
- **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/*。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 [12]、[13] 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。
- **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。

---

[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_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、鋲 (びょう) [14] 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 [15] 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、Data Preparation オプション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。
- **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リハビリ・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が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。

[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.

- **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 [16] 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 [17] 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 [18] 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。

---

[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.

- **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[19]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。

[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、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。

© Copyright IBM Corp. 2021. 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 - 2021. All rights reserved.

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]

値の再割り当て 61
値ラベル
　数値変数 20
　「ビューアー」での表示の制御 20
　割り当て 20
新しい変数の計算 63

[カ行]

カテゴリー・データ
　要約統計量 23
区間データ 23
グラフ
　グラフの作成 27
　図表の作成 27
　棒 27
ケース
　選択 71
　ソート 69, 71
ケースのサブセット
　条件式 72
　選択 71
　日時に基づく 73
　未選択のケースの削除 74
　未選択のケースのフィルター処理 74
　無作為抽出 73
　IF 条件が満たされるケース 72
ケースの選択 71
ケースのソート 69
結果のエクスポート
　Excel への 47
　HTML 54
　PowerPoint への 47
　Word への 47
欠損値
　システム欠損 21
　数値型変数の 21
　非数値型変数の 22

[サ行]

サンプル・ファイル
　場所 75
式での関数 64
システム欠損値 21
質的データ 23
順序データ 23
条件式 65
シンタックス 57
シンタックス・ウィンドウ
　コードに色を付ける 58
　コマンドの実行 57, 59
シンタックス・ウィンドウ (続き)
　コマンドの貼り付け 57
　自動入力 58
　ブレークポイント 59
　編集コマンド 58
シンタックス・ファイル
　開く 59
「シンタックス・ヘルプ」ツール 58
シンタックスを貼り付ける
　ダイアログ・ボックスから 57
数値データ 17
スケール・データ 23
スケール変数
　要約統計量 25
図表
　ヒストグラム 26
　棒 24
スプレッドシート・ファイル
　変数名の読み取り 8
　読み取り 8
測定レベル 23

[タ行]

データ入力 17, 18
データベース・ウィザード 11
データベース・ファイル
　読み取り 11
テキスト・インポート・ウィザード 13
テキスト・データ・ファイル
　読み取り 13
度数
　度数のテーブル 23
度数分布表 23

[ナ行]

日時の変数 65

[ハ行]

ヒストグラム 26
日付と時刻ウィザード 65
比データ 23
ピボット・テーブルでの行および列の非表示 37
ピボット・テーブルでの行と列の入れ替え (反転) 35
ピボット・テーブルの編集 37
ファイルの分割処理 69
変数
　データ型 20
　ラベル 19
変数ラベルの作成 19
棒グラフ 24

[マ行]

名義データ 23

[ヤ行]

要約統計量
　カテゴリー・データ 23
　スケール変数 25

[ラ行]

量的データ 23
連続データ 23

A

Access (Microsoft) 11

E

Excel (Microsoft)
　結果のエクスポート 47
Excel ファイル
　読み取り 8

H

HTML
　結果のエクスポート 54

P

PowerPoint (Microsoft)
　結果のエクスポート 47

W

Word (Microsoft)
　結果のエクスポート 47

[特殊文字]

移動
　ピボット・テーブル内の要素 35
　ビューアー内の項目 33
層
　ピボット・テーブルでの作成 36
データ エディタ
　数値データの入力 17
　非数値データの入力 18
データ型
　変数の 20
データの入力
　数値 17
　非数値 18
ピボット・テーブル
　行および列の非表示 37
　行と列の入れ替え 35
　小数点の非表示 38
　書式設定 37
　セル書式 38
ピボット・テーブル (続き)
　セルのデータ型 38
　層 36
　定義へのアクセス 34
　ピボット・トレイ 35
　編集 37
ビューアー
　出力の移動 33
　出力の非表示と表示 33
変数ラベル
　作成 19
文字列データ
　データの入力 18

IBM®