

IBM SPSS Direct Marketing 21



注：この情報とサポートされている製品をご使用になる前に、「注意事項」（p. 115）の一般情報をお読みください。

本版は IBM® SPSS® Statistics 21 , および新版で指示されるまで後続するすべてのリリースおよび変更に対して適用されます。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。

Licensed Materials - Property of IBM

© Copyright IBM Corporation 1989, 2012.

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

はじめに

IBM® SPSS® Statistics は、データ分析の包括的システムです。ダイレクトマーケティング は、このマニュアルで説明されている追加の分析手法を提供するオプションのアドオン モジュールです。ダイレクト マーケティング アドオン モジュールは SPSS Statistics Core システムと組み合わせて使用し、Core システムに完全に統合されます。

IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者がビジネス パフォーマンスを向上させるために信頼する完全で、一貫した正確な情報を提供します。ビジネス インテリジェンス、予測分析、財務実績および戦略管理、および 分析アプリケーションの包括的なポートフォリオを利用することによって、現在の実績を明確、迅速に理解し、将来の結果を予測することができます。豊富な業界のソリューション、実績ある実例、専門サービスと組み合わせ、さまざまな規模の組織が、高い生産性を実現、意思決定を自信を持って自動化し、より良い決定をもたらします。

このポートフォリオの一部として、IBM SPSS Predictive Analytics ソフトウェアを使用する組織は、将来のイベントを予測し、その洞察に基づいて積極的に行動し、より優れた業績を実現することができます。全世界の企業、政府、学術分野のお客様が IBM SPSS の技術を活用し、不正行為を減少させ、リスクを軽減させながら、顧客の獲得、保持、成長において、競争優位を高めることができます。IBM SPSS ソフトウェアを日々の業務に取り入れることによって、組織は業務目標を達成し、大きな競争的優位を獲得することができるよう、意思決定を方向付け、自動化することができるようになります。お問い合わせは、<http://www.ibm.com/spss> を参照してください。

テクニカル サポート

テクニカル サポートのサービスをご利用いただけます。IBM Corp. 製品の使用方法や、対応しているハードウェア環境へのインストールに関して問い合わせることもできます。テクニカル サポートの詳細については、IBM Corp. Web サイト (<http://www.ibm.com/support>) を参照してください。連絡の際は、所属団体名、サポート契約などを確認できるよう、あらかじめ手元にご用意ください。

学生向けテクニカル サポート

IBM SPSS ソフトウェア製品の Student 版、アカデミック版、Grad パック版を使用している学生の場合、学生用の特別オンライン ページ、[Solutions for Education \(http://www.ibm.com/spss/rd/students/\)](http://www.ibm.com/spss/rd/students/) ページを参照してください。大学提供の IBM SPSS ソフトウェアのコピーを使用している場合、大学の IBM SPSS 製品コーディネータにお問い合わせください。

カスタマ サービス

配送やアカウントに関するご質問は、お近くの営業所にお問い合わせください。お問い合わせの際には、シリアル番号をご用意ください。

トレーニング セミナー

IBM Corp. では一般公開およびオンサイトで トレーニング セミナーを実施しています。セミナーでは実践的な講習を行います。セミナーは主要都市で定期的に行われます。セミナーに関する詳細については、<http://www.ibm.com/software/analytics/spss/training> を参照してください。

内容

パート I: ユーザー ガイド

1	ダイレクト マーケティング	1
2	RFM 分析	2
	取引データからの RFM スコア	3
	顧客データからの RFM スコア	5
	RFM ビン	7
	取引データから RFM スコアを保存	10
	顧客データからの RFM スコアの保存	12
	RFM 出力	14
3	クラスタ分析	16
	設定	19
4	見込みのプロファイル	21
	設定	26
	カテゴリ回答フィールドの作成	27
5	郵便番号回答率	28
	設定	32
	カテゴリ回答フィールドの作成	34

6	購入の傾向	36
	設定	41
	カテゴリ回答フィールドの作成	43
7	対照パッケージ検定	44
パート II: 例		
8	トランザクション データからの RFM 分析	49
	取引データ	49
	分析の実行	49
	結果の評価	51
	スコア データと顧客データの結合	53
9	クラスタ分析	57
	分析の実行	57
	出力	59
	クラスタに基づいたレコードの選択	67
	クラスタ モデル ビューアでのフィルタ作成	68
	クラスタ フィールド値に基づいたレコードの選択	70
	要約	73
10	見込みのプロファイル	74
	データの考慮事項	74
	分析の実行	74
	出力	78
	要約	80

11 郵便番号回答率	81
データの考慮事項.....	81
分析の実行.....	81
出力.....	84
要約.....	87
12 購入の傾向	88
データの考慮事項.....	88
予測モデルの作成.....	88
モデルの評価.....	92
モデルの適用.....	93
要約.....	100
13 対照パッケージ検定	101
分析の実行.....	101
出力.....	103
要約.....	103
付録	
A サンプル ファイル	104
B 注意事項	115
索引	118

パート I: ユーザー ガイド

ダイレクト マーケティング

ダイレクト マーケティング オプションでは、人口統計、購買、および消費者の様々なグループを定義するその他の特性を特定し、特定のグループを対象として正の回答率を最大化することによって、ダイレクト マーケティング キャンペーンの結果を向上させるように設計した一連のツールが用意されています。

RFM 分析。 新製品に最も反応しそうな既存の顧客を識別するために使用する技術です。 [詳細は、 2 章 p.2 RFM 分析 を参照してください。](#)

クラスタ分析。 データセット内の自然なグループ（またはクラスタ）を明らかにすることを目的として設計された探索ツールです。たとえば、さまざまな人口統計的および購入の特性に基づいてさまざまな顧客のグループを識別できます。 [詳細は、 3 章 p.16 クラスタ分析 を参照してください。](#)

見込みのプロファイル。 この方法では、以前のまたはテスト キャンペーンの結果を使用して、説明的プロファイルを作成します。プロファイルを使用して、特定グループの連絡先を目標に設定することができます。 [詳細は、 4 章 p.21 見込みのプロファイル を参照してください。](#)

郵便番号回答率。 この方法では、以前のキャンペーンの結果を使用して、郵便番号の回答率を計算します。これらの比率は、今後のキャンペーンの対象の特定郵便番号を目標に設定できます。 [詳細は、 5 章 p.28 郵便番号回答率 を参照してください。](#)

購入の傾向。 この方法では、テスト送付または以前のキャンペーンの結果を使用して、傾向スコアを生成します。スコアは、回答する可能性の最も高い連絡先を示します。 [詳細は、 6 章 p.36 購入の傾向 を参照してください。](#)

パッケージ検定を制御。 この方法では、マーケティング キャンペーンを比較して、さまざまなパッケージまたはオファーの有効性に重要な相違点があるかどうかを確認します。 [詳細は、 7 章 p.44 対照パッケージ検定 を参照してください。](#)

RFM 分析

RFM 分析は、新製品に最も反応しそうな既存の顧客を識別するために使用する技術です。この技術はダイレクト マーケティングで一般的に使用されます。RFM 分析は、次のシンプルな理論に基づいています。

- 新製品に反応しそうな顧客を識別する最も重要な因子は、**最新性** です。最近購入した顧客は、遠い過去に購入した顧客よりも、もう一度購入する可能性が高くなります。
- 2 番目に大事な因子は、**頻度**です。過去により購入が多い顧客は、購入が少ない顧客よりも、反応する可能性が高くなります。
- 3 番目に大事な因子は、全支出金額です。これを**マネタリー**と呼びます。過去により多くの金額を支払った（購入したすべての合計）顧客は、少なく支払った顧客よりも、反応する可能性が高くなります。

RFM 分析はどのように機能するか

- 最近購入した日付または最近購入してからの時間間隔に基づいて、最新性スコアが顧客に割り当てられます。このスコアは、少数のカテゴリへ最新性の値を単純に順位付けすることに基づいています。たとえば、5 つのカテゴリを使用すると、購入日が一番最近である顧客は最新性の順位が 5 になり、一番遠い過去に購入した顧客の最新性の順位は 1 になります。
- 同様に、顧客には頻度順位が割り当てられます。値が高いほど購入の頻度が高いことを示します。たとえば、5 つのカテゴリに順位付けとすると、もっとも頻繁に購入する顧客は頻度順位が 5 になります。
- 最後に、マネタリー値の順位が顧客に付けられます。最も高いマネタリー値をとると最も高い順位になります。5 つのカテゴリの例を続けると、最も多く支払った顧客はマネタリー順位が 5 になります。

最新性、頻度、マネタリー、および結合 RFM スコアの 4 つのスコアが各顧客の結果となります。結合 RFM スコアとは、単純に 3 つのそれぞれのスコアを 1 つの値に連結したものです。「最高」の顧客（製品に最も反応する可能性のある顧客）とは結合 RFM スコアが最高の顧客です。たとえば、5 つのカテゴリの順位付けでは、合計で 125 種類の結合 RFM スコアがあり、結合 RFM スコアの最高点は 555 点になります。

データの考慮事項

- データ行が取引を表す場合は（各行が 1 つの取引を表し、各顧客に複数の取引がある可能性がある）、[取引] から RFM を使用します。詳細は、p.3 取引データからの RFM スコア を参照してください。
- データ行が、すべての取引について要約情報がある（総支出金額、取引の合計数、および一番最近の取引日付などの値を含む列がある）顧客を表す場合は、顧客データから RFM を使用します。詳細は、p.5 顧客データからの RFM スコア を参照してください。

図 2-1
取引と顧客データ

トランザクション行			
ID	性別	日付	金額
1	男性	9/25/2005	21
2	男性	1/15/2006	297
4	男性	2/5/2006	249
4	男性	5/7/2005	172
6	男性	4/16/2005	164
6	男性	4/12/2005	286
7	女性	7/12/2005	403
9	男性		
9	男性		
9	男性		
10	女性		
10	女性		

行が顧客にトランザクションサマリー				
ID	性別	最近	合計金額	購入品番号
1	男性	9/25/2005	21	1
2	男性	1/15/2006	297	1
4	男性	2/5/2006	421	2
6	男性	4/16/2005	450	2

取引データからの RFM スコア

データの考慮事項

データセットには、次の情報を含む変数を含んでいる必要があります。

- 各ケース（顧客）を識別する変数または変数の組み合わせ。
- 各取引日付を持つ変数。
- 各取引のマネタリー値を持つ変数。

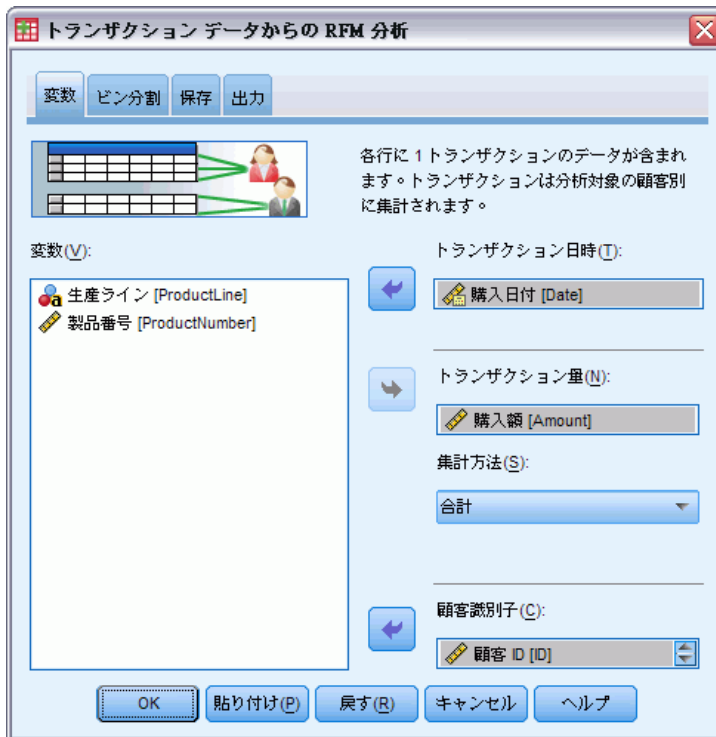
図 2-2
RFM 取引データ

ID	日付	金額
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/13/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

取引データからの RFM スコアの作成

- ▶ メニューから次の項目を選択します。
ダイレクト マーケティング > 手法を選択
- ▶ [最良の連絡先を特定する (RFM) 分析] を選択し、[続行] をクリックします。
- ▶ [取引データ] を選択し、[続行] をクリックします。

図 2-3
取引データ、[変数] タブ



- ▶ 取引日付を含む変数を選択します。
- ▶ 各取引の金銭的価値量を含む変数を選択します。
- ▶ 各顧客の取引金額を集計する方法を、合計（全取引の合計）、平均値、中央値、最大値（最高取引金額）から選択します。
- ▶ 各ケース（顧客）を一意に識別する変数または変数の組み合わせを選択します。たとえば、一意の ID コードまたはラスト ネームとファースト ネームの組み合わせを使用して、ケースを識別できる可能性もあります。

顧客データからの RFM スコア

データの考慮事項

データセットには、次の情報を含む変数を含んでいる必要があります。

- 一番最近購入した日付または一番最近購入した日付からの時間間隔。これはリーセンシ スコアの計算に使用します。

- 購入総数。これはフリクエンシ スコアの計算に使用します。
- すべての購入についてのマネタリーの要約。これはマネタリー スコアの計算に使用します。通常、これはすべての購入の総和（合計）になりますが、平均、最大値、またはその他の集計項目にもできます。

図 2-4
RFM 顧客データ

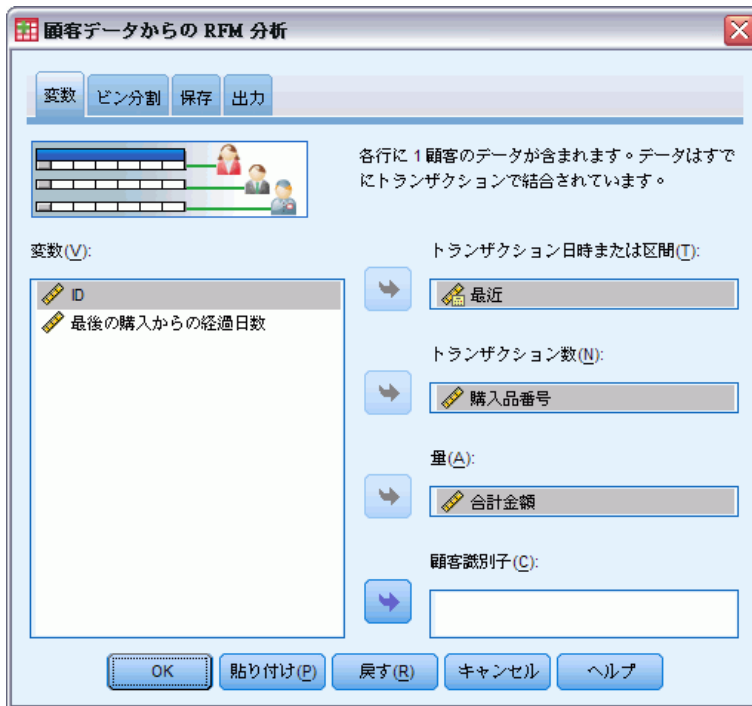
ID	合計金額	最近	購入品番号
1	485.00	09/04/2006	5
2	350.00	11/10/2005	4
3	233.00	06/04/2005	2
4	936.00	08/18/2006	7
5	359.00	07/07/2006	3
6	249.00	07/16/2006	3
7	1089.00	02/15/2006	7
8	423.00	08/21/2006	4
9	689.00	08/31/2006	7
10	325.00	10/13/2005	3

RFM スコアを新しいデータセットに書き込むと、アクティブなデータセットには、ケース（顧客）を識別する変数、または変数の組み合わせも含める必要があります。

顧客データからの RFM スコアの作成

- ▶ メニューから次の項目を選択します。
ダイレクト マーケティング > 手法を選択
- ▶ [最良の連絡先を特定する (RFM) 分析] を選択し、[続行] をクリックします。
- ▶ [顧客データ] を選択し、[続行] をクリックします。

図 2-5
顧客データ、[変数] タブ



- ▶ 一番最近の取引日付、または一番最近の取引からの時間間隔を示す数字を含む変数を選択します。
- ▶ 各顧客の取引総数を含む変数を選択します。
- ▶ 各顧客の金銭的価値量の要約を含む変数を選択します。
- ▶ RFM スコアを新しいデータセットに書き込む場合は、各顧客を一意に識別する変数、または変数の組み合わせを選択します。たとえば、一意の ID コードまたはラスト ネームとファースト ネームの組み合わせを使用して、ケースを識別できる可能性もあります。

RFM ビン

大量の数値を少数のカテゴリにグループ化する処理のことを、**ビン**と呼ぶことがあります。RFM 分析では、**ビン**は順位付けされたカテゴリのことです。[ビン] タブを使用して、リーセンシ、フリクエンシ、およびマネタリーをビンに割り当てる方法を変更することができます。

図 2-6
[RFM ビン] タブ



ビン分割方法

入れ子。入れ子のビンでは、リーセンシの値に簡単な順位が割り当てられます。それぞれのリーセンシ順位には、顧客にフリクエンシ順位が割り当てられ、それぞれのフリクエンシ順位には顧客にマネタリー順位が割り当てられます。これにより、結合 RFM スコアの均等な分布を得られる傾向がありますが、フリクエンシやマネタリーのスコアの解釈が難しくなるという欠点もあります。たとえば、フリクエンシ順位はリーセンシ性順位に依存するので、リーセンシ順位が 5 でフリクエンシ順位が 5 の顧客は、リーセンシ順位が 4 でフリクエンシ順位が 5 の顧客と同じことを意味しないこともあります。

独立。リーセンシ、フリクエンシ、およびマネタリーには単純な順位が割り当てられます。それぞれ 3 つの順位が割り当てられます。そのため、3 つの RFM コンポーネントのそれぞれの解釈ははっきりしています。ある顧客のフリクエンシ スコアが 5 であるということは、リーセンシ スコアにかかわらず、フリクエンシ スコアが 5 である別の顧客と同じであることを意味します。より少ないサンプルでは、結合 RFM スコアが均等でない分布になるという短所があります。

ビン数

各コンポーネントで RFM スコアを作成するために使用するカテゴリ（ビン）数。可能な結合 RFM スコアの総数は、3 つの値を掛け合わせたものになります。たとえば、リーセンシ ビンが 5、フリクエンシ ビンが 4、マネタリー ビンが 3 の場合、合計で 60 個の可能な結合 RFM スコアがあり、値の範囲は 111 から 543 です。

- 各コンポーネントのデフォルトの値は 5 となっており、この場合に可能な結合 RFM スコアは合計で 125 個あり、値の範囲は 111 から 555 になります。
- 各スコア コンポーネントの最大数は 9 です。

同順位

「同順位」とは、リーセンシ、フリクエンシまたはマネタリーの値が2つ以上等しいことをいいます。理想的には、各ビンにほぼ同数の顧客がいるようにしたいのですが、同順位の値が多数あると、ビン分布に影響が出る可能性があります。同順位の処理には、2 つの方法があります。

- **同順位を同じビンへ割り当てる。**この方法では、ビン分布へどのように影響するかにかかわらず、常に同順位の値を同じビンに割り当てます。これにより、一貫したビン方法を提供します。2 つの顧客のリーセンシ値が同じ場合、常に同じリーセンシ スコアが割り当てられます。ただし、極端な例では、1,000 人の顧客のうち 500 人が同日に最新の購入をすることも考えられます。そのため、5 ビンの順位付けでは、目的の 20% の代わりに、顧客の 50% が最新性スコア 5 を受け取ることになります。

入れ子になっているビン方法では、頻度スコアが最新性スコアのビンに割り当てられ、マネタリー スコアが頻度スコアのビンに割り当てられるので、頻度スコアとマネタリー スコアに関して「一貫性」がいくぶん複雑になっていることに注意してください。同順位の値がどのように処理されるかにかかわらず、同じ頻度値を持つ 2 人の顧客が同じ最新性スコアでない場合は、2 人の頻度スコアが同じではない可能性があります。

- **同順位の無作為な割り当て。**これにより、順位付けよりも前に、非常に小さい無作為な分散因子を同順位に割り当てることで均等なビン分布になります。そのため、順位付けられたビンへ値を割り当てるため、同順位の値はありません。この処理は、元の値には影響しません。これは同順位のあいまいさをなくすためだけに使用します。これによって均等なビン分布が生成されますが（それぞれのビンにほぼ同数の顧客）、最新性、頻度、マネタリーの値が類似しているか等しいように思われる顧客について、全く異なるスコア結果が出る可能性もあります。特に、顧客の総数が比較的小さいか同順位の数が比較的多い場合、あるいはその両方が該当する場合に発生します。

テーブル 2-1
同順位を同じビンに割り当てることと、同順位を無作為に割り当てることの比較

ID	一番最近の購入 (最新性)	同順位を同じビンへ割り当てる	同順位を無作為に割り当てる
1	10/29/2006	5	5
2	10/28/2006	4	4
3	10/28/2006	4	4
4	10/28/2006	4	5
5	10/28/2006	4	3
6	9/21/2006	3	3
7	9/21/2006	3	2
8	8/13/2006	2	2
9	8/13/2006	2	1
10	6/20/2006	1	1

- この例では、同じビンに同順位を割り当て不均等なビン分布、5 (10%)、4 (40%)、3 (20%)、2 (20%)、1 (10%) となります。
- 同順位を無作為に割り当てると、各ビンが 20% になりますが、この結果を満たすには、日付値が 10/28/2006 である 4 つのケースを 3 つの異なるビンに割り当て、さらに日付値が 8/13/2006 である 2 つのケースを別々のビンに割り当てます。

同順位を別々のビンに割り当てる方法は、(最終的に各ビン内のケースは同数になるという制約で) 完全に無作為であることに注意してください。同じ方法を使って 2 番目のスコアのセットを計算すると、同じ値のケースについて順位付けが変わる可能性があります。たとえば、ケース 4 の最新性の順位が 5 で、ケース 5 の順位が 3 であっても、2 回目にはその順位が入れ替わる可能性があります。

取引データから RFM スコアを保存

取引データの RFM スコアでは、顧客ごとに行が 1 つ存在する、新しく集計されたデータセットを常に作成します。[保存] タブを使用して、保存したいスコアとそのほかの変数、および保存場所を指定します。

図 2-7
取引データ、[保存] タブ



変数

各顧客を一意に識別する ID 変数は、自動的に新しいデータセットに保存されます。下記の追加変数は、新しいデータセットに保存できます。

- **各顧客について一番最近の取引日付。**
- **取引数。**各顧客の取引行の総数。
- **量。**[変数] タブで選択する集計方法に基づいた各顧客の集計量。
- **リーセンシスコア。**一番最近の取引日付に基づいて、各顧客に割り当てられたスコア。より高いスコアがより最近の取引日付を示します。
- **フリクエンシスコア。**取引総数に基づいて、各顧客に割り当てられたスコア。より高いスコアがより多くの取引を示します。
- **マネタリースコア。**選択したマネタリーの集計項目に基づいて、各顧客に割り当てられたスコア。より高いスコアの方が、マネタリーの集計項目について、より高い値を示します。
- **RFM スコア。**1 つの値に結合された 3 つの各スコア ((リーセンシ x 100) + (フリクエンシ x 10) + マネタリー)。

デフォルトでは、利用可能な変数はすべて、新しいデータセットに含まれます。そのため、含めたくない変数は、選択から外します（チェックを外します）。オプションとして、独自の変数名を指定できます。変数名は、標準の変数名規則に従う必要があります。

位置

取引データの RFM スコアでは、顧客ごとに行が 1 つ存在する、新しく集計されたデータセットを常に作成します。現在のセッションに新しいデータセットを作成したり、RFM スコア データを外部データ ファイルに保存したりできます。データセット名は、標準の変数名規則に従う必要があります。（この制限は、外部データ ファイル名には適用しません）

顧客データからの RFM スコアの保存

顧客データでは、RFM スコア変数をアクティブなデータセットに追加したり、選択したスコア変数を含む新しいデータセットを作成したりできます。[保存] タブを使用して、保存したいスコア変数とその保存場所を指定します。

図 2-8
顧客データ、[保存] タブ

顧客データからの RFM 分析

変数 ビン分割 **保存** 出力

保存変数の名前

一意の名前を自動生成(A)

ユーザー設定(C)

変数(V):

保存	内容	保存変数の名前
<input checked="" type="checkbox"/>	リーセンシスコア	Recency_score
<input checked="" type="checkbox"/>	フリクエンシスコア	Frequency_score
<input checked="" type="checkbox"/>	マネタリースコア	Monetary_score
<input checked="" type="checkbox"/>	RFM スコア	RFM score

位置

アクティブなデータセットに書き込み(A)

新しいデータセットを作成(C) 名前(M):

新しいデータ ファイルを書き込み(W) 参照(B)...

OK 貼り付け(P) 戻す(R) キャンセル ヘルプ

保存変数の名前

- **一意の名前を自動生成します。** スコア変数をアクティブなデータ セットに追加する場合、これにより、新しい変数が一意の名前であることが保証されます。特に、（異なる条件に基づいて）複数の異なる RFM スコアのセットをアクティブなデータセットに追加したい場合に役立ちます。
- **顧客名。** これを使用すると、独自の変数名をスコア変数に割り当てることができます。変数名は、標準の変数命名規則に従う必要があります。

変数

保存するスコア変数を選択（チェックを付ける）します。

- **リーセンシ スコア。** [変数] タブで選択した [取引日付または間隔] 変数の値に基づいて、各顧客に割り当てられたスコア。一番最近の日付または間隔値がより低い値に対して、より高いスコアが割り当てられます。
- **フリクエンシ スコア。** [変数] タブで選択した [取引数] 変数に基づいて、各顧客に割り当てられたスコア。より高い値に対して、高いスコアが割り当てられます。
- **マネタリー スコア。** [変数] タブで選択した [量] 変数に基づいて、各顧客に割り当てられたスコア。より高い値に対して、高いスコアが割り当てられます。
- **RFM スコア。** 1 つの値に結合された 3 つの各スコア（(リーセンシ x 100) + (フリクエンシ x 10) + マネタリー）。

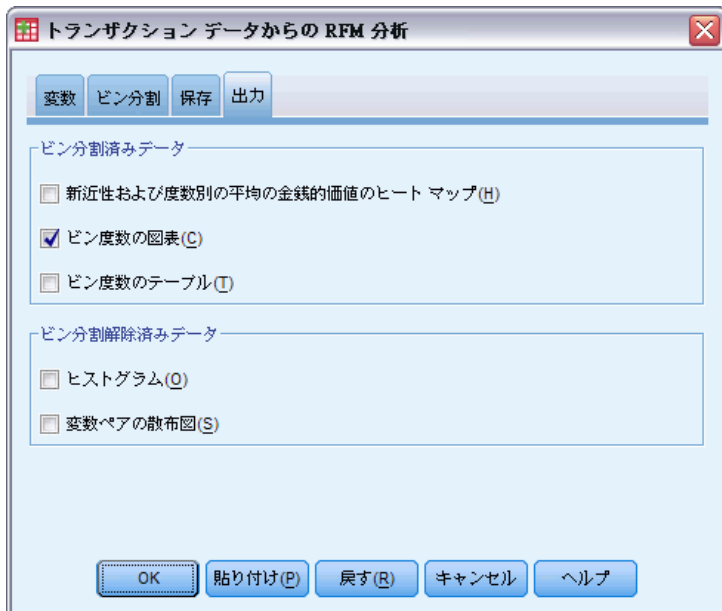
位置

顧客データには、新しい RFM スコアを保存できる場所の選択肢が 3 つあります。

- **アクティブなデータセット。** 選択した RFM スコアは、アクティブなデータセットに追加されます。
- **新規データセット。** 選択した RFM スコア変数と、各顧客（ケース）を一意に識別する ID 変数は、現在のセッションの新規データセットへ上書きされます。データセット名は、標準の変数命名規則に従う必要があります。このオプションは、[変数] タブで 1 つ以上の顧客識別変数を選択する場合にのみ有効です。
- **ファイル。** 選択した RFM スコアと、各顧客（ケース）を一意に識別する ID 変数は、外部データ ファイルに保存されます。このオプションは、[変数] タブで 1 つ以上の顧客識別変数を選択する場合にのみ有効です。

RFM 出力

図 2-9
[RFM 出力] タブ



ビン分割済みデータ

ビン分割済みデータの図表とテーブルは、計算されたリーセンシ、フリクエンシ、およびマネタリー スコアに基づきます。

リーセンシおよびフリクエンシの平均のマネタリーのヒート マップ 平均のマネタリー分布のヒート マップは、リーセンシおよびフリクエンシ スコアで定義されたカテゴリの平均のマネタリーを表します。濃く表示されたエリアは、平均のマネタリーが高いことを表します。

ビン度数の図表。 ビン度数の図法は、選択したビン分割方法でのビン分布を表示します。それぞれの棒が、それぞれの結合 RFM スコアに割り当てられたケース数を表示します。

- すべて（または、ほとんど）の棒の高さがほぼ同じ程度の、規則正しい分布が一般的に求められますが、同順位の値を同じビンに割り当てるデフォルトのビン分割方法を使用した場合は、ある程度の分散は予期する必要があります。
- ビン分布における極端な変動、および/または、空のビンが多い場合は、他のビン分割方法（ビンを減らす、および/または、同順位の値をランダムに割り当てる）を試したり、RFM 分析の適切さを再考する必要があることを示している場合があります。

ビン度数のテーブル。 テーブルの形式での表現以外は、各セルのビン度数を含み、ビン度数の図表と同じ情報。

ビン分割解除済みデータ

ビン分割解除済みデータの図表とテーブルは、リーセンシ、フリクエンシ、およびマネタリー スコアを作成するために使用された元の変数に基づきます。

ヒストグラム。 ヒストグラムは、リーセンシ、フリクエンシ、およびマネタリー スコアを計算するために使用された 3 つの変数の値の相対分布を表示します。これらのヒストグラムが正規分布または対称分布ではなく、やや非対称分布を示すのは一般的ではありません。

それぞれのヒストグラムの横軸は、常に左から右に、低い値から高い値の順になります。ただしリーセンシの図表の解釈は、日付または時間間隔など、リーセンシの測定方法に依存します。日付については、左の棒が遠い過去を表します（古い日付ほど、新しい日付より低い値となります）。時間間隔については、左の棒が最新に近い値を表します（時間間隔が短いほど、取引が最近であることを表します）。

変数ペアの散布図。 散布図は、リーセンシ、フリクエンシ、およびマネタリー スコアを計算するために使用された 3 つの変数の関係を表示します。

フリクエンシ スケールの点が明確な線型にグループ化されるのも一般的です。これは、フリクエンシ「が多くの場合、個々の値の比較的小さい範囲を表しているためです。たとえば、取引の総数が 15 より大きくなならない場合、考えられるフリクエンシの値は 15 個しかありません（少数を数えない場合）。ところが、リーセンシの値が数百件になることもあり、マネタリーの値は数千になることもあります。

リーセンシ軸の解釈は、日付または時間間隔など、リーセンシの測定方法に依存します。日付については、原点に近い点ほど、遠い過去の日付を表します。時間間隔については、原点に近い点ほど、新しい値を表します。

クラスタ分析

クラスタ分析は、データセット内の自然なグループ（またはクラスタ）を明らかにすることを目的として設計された探索ツールです。たとえば、さまざまな人口統計的および購入の特性に基づいてさまざまな顧客のグループを識別できます。

例：小売業者および対消費者企業は定期的に、顧客の購買習慣、性別、年齢、収入レベルなどを説明するデータにクラスタリング手法を適用します。これらの企業は、マーケティングおよび製品開発戦略を各消費者グループごとに作成し、販売を拡大し、ブランドロイヤリティを構築します。

クラスタ分析のデータの考慮事項












データ。この手続きは、連続型フィールドとカテゴリフィールドの両方に使用できます。各レコードはクラスタ化される顧客を表し、フィールド（フィールド）はクラスタ化の基準となる属性を表します。

レコードの順序。レコードの並び順によって結果が異なる可能性があることに注意してください。並び順の影響を最小限に抑えるには、レコードを無作為に並べます。異なる無作為な順序で並べ替えられたレコーを使用していくつかの異なる分析を実行することにより、特定の解の安定性を確認できます。

測定レベル。結果の計算に影響を与えるため、適切な測定レベルの割り当ては重要です。

- **名義データ。**値がランキングを持たないカテゴリを表しているとき、名義（変数）として取り扱うことができます。たとえば、従業員の会社の所属などです。名義変数の例としては、地域やジップコードや所属宗教などがあります。
- **順序データ。**値がランキングをもったカテゴリを表しているとき、変数を順序として取り扱うことができます。たとえば、「かなり不満」から「かなり満足」までのようなサービス満足度のレベルなどです。順序変数の例としては、満足度や信頼度を表す得点や嗜好得点などです。
- **続行。**値が有意な基準を持った順序カテゴリを表しているとき、変数をスケール（連続型）として扱うことができます。値間の距離の比較などに適切です。スケール変数の例としては、年齢や、千ドル単位で表した所得があります。

各フィールドの隣のアイコンは、現在の測定レベルを示します。

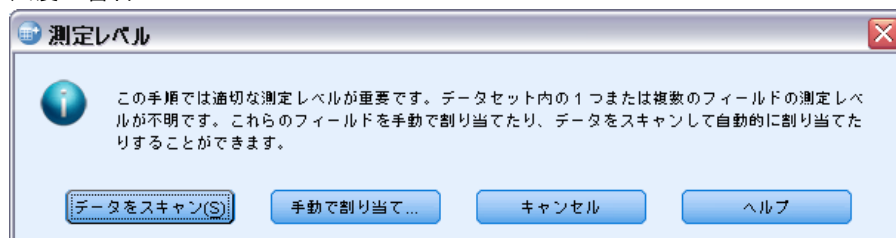
	数値	String	Date	Time
スケール（連続）		利用不可		
順序				
名義				

測定レベルをデータ エディタ の変数ビューで変更することができます。または [変数のプロパティ] ダイアログを使用して各フィールドに適切な測定レベルを提案することができます。

不明な尺度の項目

データセットの 1 つまたは複数の変数（フィールド）の尺度が不明な場合、尺度の警告が表示されます。尺度はこの手順の結果の計算に影響を与えるため、すべての変数に尺度を定義する必要があります。

図 3-1
尺度の警告



- **データをスキャン。** アクティブ データセットのデータを読み込み、デフォルトの尺度を尺度が現在不明なフィールドに割り当てます。データセットが大きい場合は時間がかかります。
- **手動で割り当てる。** 不明な尺度のフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、尺度をこれらのフィールドに割り当てることができます。データ エディタの [変数ビュー] でも、尺度を割り当てることができます。

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

クラスタ分析を行うには

メニューから次の項目を選択します。

Direct Marketing > 手法を選択

- ▶ [連絡先をクラスタにセグメント化する] を選択します。

図 3-2
クラスタ分析の [フィールド] タブ



- ▶ セグメント作成に使用するカテゴリ（名義型、順序型）フィールドおよび連続型（スケール）フィールドを選択します。
- ▶ [実行] をクリックして手続きを実行します。

設定

図 3-3
クラスタ分析の [設定] タブ



[設定] タブを使用すると、セグメントについて説明する図表や表を表示または非表示にしたり、データセット内の各レコードのセグメント（クラスター）を特定するデータセットの新しいフィールドを保存したり、クラスターの解に含めるセグメント数を指定します。

図表と表を表示。 セグメントを説明する表や図表を表示します。

セグメントメンバーシップ。 各レコードが属するセグメントを特定する新しいフィールド（変数）を保存します。

- フィールド名は、IBM® SPSS® Statistics 命名規則に従う必要があります。

- セグメント メンバシップ フィールド名は、データセット内に既に存在するフィールド名と重複することはできません。同じデータセットに複数回この手順を実行する場合、それぞれ別の名前を指定する必要があります。
- **セグメント数。** セグメント数の指定方法を制御します。
- **自動的に判定。** この手続きは、指定された最大値を上限として、「最適な」クラスターの個数を自動的に判定します。

固定値を指定。 この手順では、指定した数のセグメントを作成します。

見込みのプロファイル

この方法では、以前のまたはテスト キャンペーンの結果を使用して、説明的プロファイルを作成します。プロファイルを使用して、特定グループの連絡先を目標に設定することができます。[回答] フィールドは、誰が以前のまたはテスト キャンペーンに回答したかを特定します。[プロファイル] リストには、プロファイルの作成に使用する特性が表示されます。

例€ テスト送付の結果に基づき、企業のダイレクト マーケティングの担当部門は、人口統計情報に基づき、オファーに回答する確率が最も高い顧客の種類に関するプロファイルを生成する場合があります。

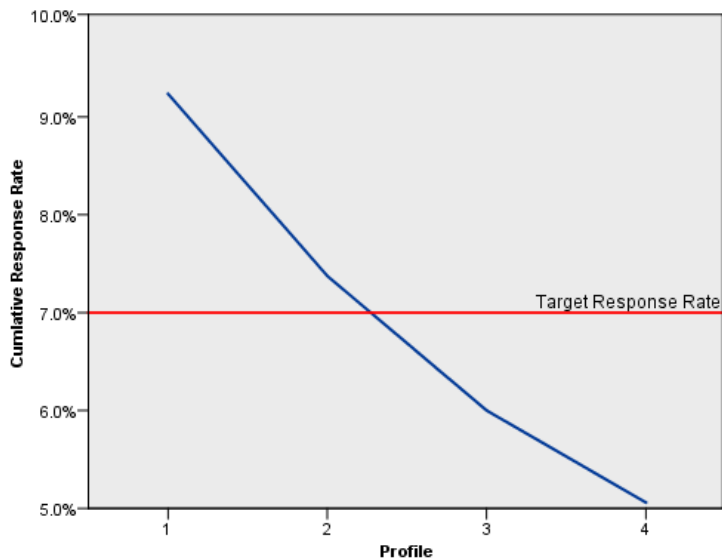
出力

出力には、各プロファイル グループについて説明し、回答率（正の回答の割合）や累積回答率を示す表や、累積回答率のグラフがあります。目標回答率の最小値を指定した場合、表を色分けして累積回答率の最小値に一致するプロファイルを示し、グラフは指定した回答率の最小値に基準線を示します。

図 4-1
回答率表および回答率グラフ

数	回答率			
	説明	グループサイズ	回答率	累積回答率
1	Region = "West", "South", "East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West", "South", "East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West", "South", "East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

緑: 目標回答率を満たす。
赤: 目標回答率を満たしていない。



見込みのプロファイルのデータの考慮事項

回答フィールド。 回答フィールドは、名義型または順序型でなければなりません。この変数は、文字型または数値型のどちらでもかまいません。このフィールドに購入数または金額を示す値を入力する場合、単一の値がすべての正の回答を示す新しいフィールドを作成する必要があります。[詳細は、p. 27 カテゴリ回答フィールドの作成を参照してください。](#)












正の回答値。 正の回答値は、肯定的に（購入した、など）回答した顧客を示します。欠損値以外の他のすべての回答値は、負の回答を示すと想定します。回答フィールドに定義された値ラベルがある場合、これらのラベルは、ドロップダウン リストに表示されます。

次を使用してプロファイルを作成。 これらのフィールドは、名義、順序、または連続型（スケール）のいずれかです。文字型または数値型のどちらでもかまいません。

測定レベル。 結果の計算に影響を与えるため、適切な測定レベルの割り当ては重要です。

- **名義データ。** 値がランキングなどを持たないカテゴリを表しているとき、名義（変数）として取り扱うことができます。たとえば、従業員の会社の所属などです。名義変数の例としては、地域やジップコードや所属宗教などがあります。
- **順序データ。** 値がランキングをもったカテゴリを表しているとき、変数を順序として取り扱うことができます。たとえば、「かなり不満」から「かなり満足」までのようなサービス満足度のレベルなどです。順序変数の例としては、満足度や信頼度を表す得点や嗜好得点などです。
- **続行。** 値が有意な基準を持った順序カテゴリを表しているとき、変数をスケール（連続型）として扱うことができます。値間の距離の比較などに適切です。スケール変数の例としては、年齢や、千ドル単位で表した所得があります。

各フィールドの隣のアイコンは、現在の測定レベルを示します。

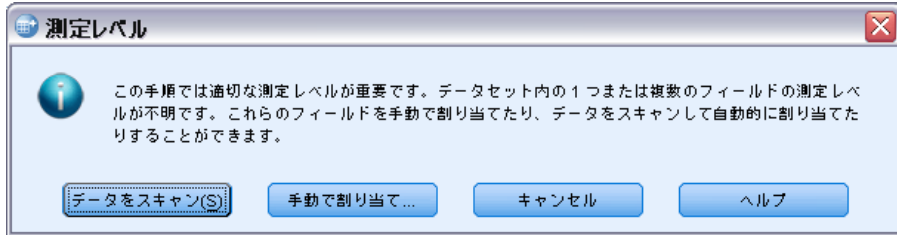
	数値	String	Date	Time
スケール（連続）		利用不可		
順序				
名義				

測定レベルをデータ エディタ の変数ビューで変更することができます。または [変数のプロパティ] ダイアログを使用して各フィールドに適切な測定レベルを提案することができます。

不明な尺度の項目

データセットの 1 つまたは複数の変数（フィールド）の尺度が不明な場合、尺度の警告が表示されます。尺度はこの手順の結果の計算に影響を与えるため、すべての変数に尺度を定義する必要があります。

図 4-2
尺度の警告



- **データをスキャン。** アクティブ データセットのデータを読み込み、デフォルトの尺度を尺度が現在不明なフィールドに割り当てます。データセットが大きい場合は時間がかかります。
- **手動で割り当てる。** 不明な尺度のフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、尺度をこれらのフィールドに割り当てることができます。データ エディタの [変数ビュー] でも、尺度を割り当てることができます。

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

見込みのプロファイルを取得するには

メニューから次の項目を選択します。

Direct Marketing > 手法を選択

- ▶ [オファーに回答した連絡先のプロファイルを生成] を選択します。

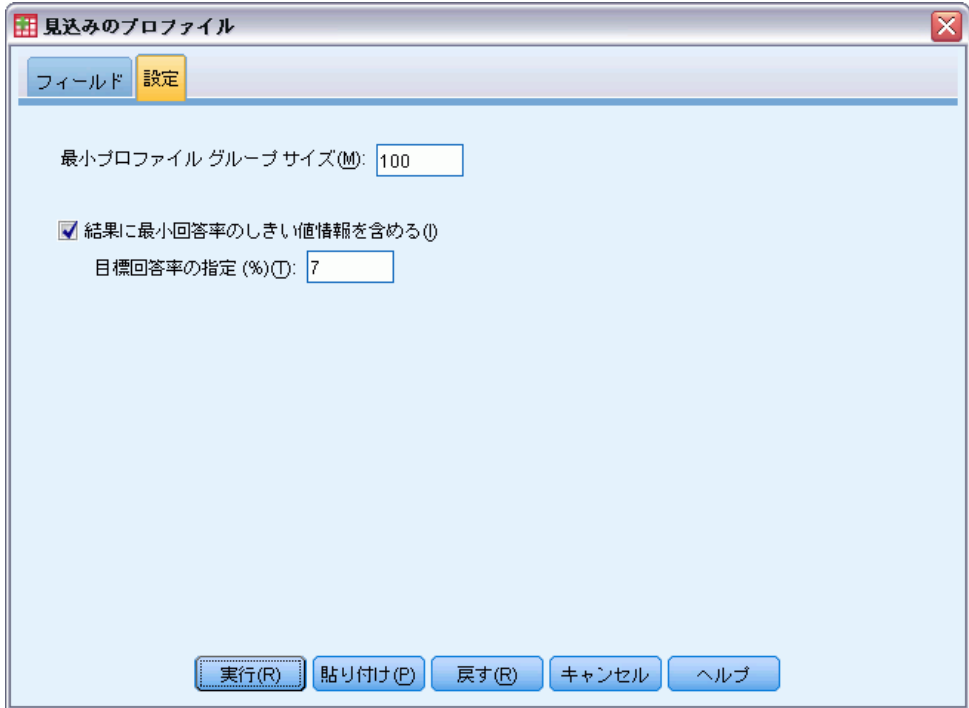
図 4-3
見込みのプロファイルの [フィールド] タブ



- ▶ オファーに回答した連絡先を示すフィールドを選択します。フィールドは、名義型または順序型でなければなりません。
- ▶ 正の回答を示す値を入力してください。値が値ラベルを定義している場合、ドロップダウン リストから値ラベルを選択すると、対応する値が表示されます。
- ▶ プロファイルを作成するのに使用したいフィールドを選択しています。
- ▶ [実行] をクリックして手続きを実行します。

設定

図 4-4
見込みのプロファイルの [設定] タブ



[設定] タブを使用して、最小プロファイル グループ サイズを制御し、出力に最小回答率のしきい値を含めることができます。

最小プロファイル グループ サイズ。 各プロファイルは、データセットの連絡先のグループの共通した特性を示します（西部地区に住む 40 歳未満の女性など）。デフォルトでは、最小プロファイル グループ サイズは 100 です。グループ サイズが小さいほどより多くのグループを明確にし、グループ サイズが大きいほど、結果がより信頼できるものになります。値は正の整数にする必要があります。

結果に最小回答率のしきい値情報を含める。 結果には、回答率（正の回答の割合）や累積回答率を示す表、累積回答率のグラフがあります。目標回答率の最小値を入力した場合、表を色分けして累積回答率の最小値に一致するプロファイルを示し、グラフは指定した回答率の最小値に基準線を示します。値は 0 より大きく 100 未満にする必要があります。

カテゴリ回答フィールドの作成

回答フィールドは、1 つの値がすべての正の回答を示す、カテゴリ型でなければなりません。欠損値以外のその他の回答値は、負の回答を示すと想定します。回答フィールドが購入数や購入金額などの連続型（スケール）値を示す場合、単一正の回答がすべての 0 以外の回答値に割り当てられる新しいフィールドを作成する必要があります。

- 負の回答は 0（空白ではなく、欠損値として処理）として記録されます。これは次の数式で計算できます。

$\text{NewName} = \text{OldName} > 0$

この場合、NewName は新しいフィールドの名前で、OldName は元のフィールドの名前です。これは、欠損値でない 0 より大きいすべての値に 1 の値を、欠損値でない 0 以下のすべての値に 0 を割り当てる論理式です。

- 負の回答に値が記録されない場合、これらの値は欠損値として処理され、数式は少し複雑になります。

$\text{NewName} = \text{NOT}(\text{MISSING}(\text{OldName}))$

この論理式では、欠損値でないすべての回答値には 1 の値が、欠損値であるすべての回答値には 0 の値が割り当てられます。

- 負 (0) の回答値と欠損値の区別が付かない場合、正確な回答値は計算できません。欠損値が比較的少ない場合、計算される回答率にあまり影響はありません。ただし、データセット全体のうち小さな検定サンプルのみの回答情報が記録された場合など、欠損値が多い場合、計算される回答率は真の回答率より大幅に低くなるため、あまり意味のないものとなります。

カテゴリ回答フィールドをするには

- ▶ メニューから次の項目を選択します。
変換(T) > 変数の計算
- ▶ 目標変数に、新しいフィールド（変数）名を入力します。
- ▶ 負の回答が 0 として記録される場合、数式に $\text{OldName} > 0$ と入力します。OldName は元のフィールド名です。
- ▶ 負の回答が欠損値（空白）として記録される場合、数式に $\text{NOT}(\text{MISSING}(\text{OldName}))$ と入力します。OldName は元のフィールド名です。

郵便番号回答率

この方法では、以前のキャンペーンの結果を使用して、郵便番号の回答率を計算します。これらの比率は、今後のキャンペーンの対象の特定郵便番号を目標に設定できます。[回答] フィールドは、誰が以前のキャンペーンに回答したかを特定します。[郵便番号] フィールドは、郵便番号が入力されたフィールドを特定します。

例€ 以前の送付の結果に基づいて、企業のダイレクト マーケティング部門は、郵便番号ごとに回答率を生成します。受け入れ可能な回答率の最小値および/または送付に使用する連絡先の最大数などのさまざまな基準をもとに、特定の郵便番号を対象とすることができます。

出力

この手順からの出力には、郵便番号ごとの回答率を含む新しいデータセット、10 分位ごとの順位（上位 10%、上位 20% など）で結果を要約する表やグラフがあります。表は、ユーザー指定の累積回答率の最小値や連絡先の最大数に基づいて、色分けすることができます。

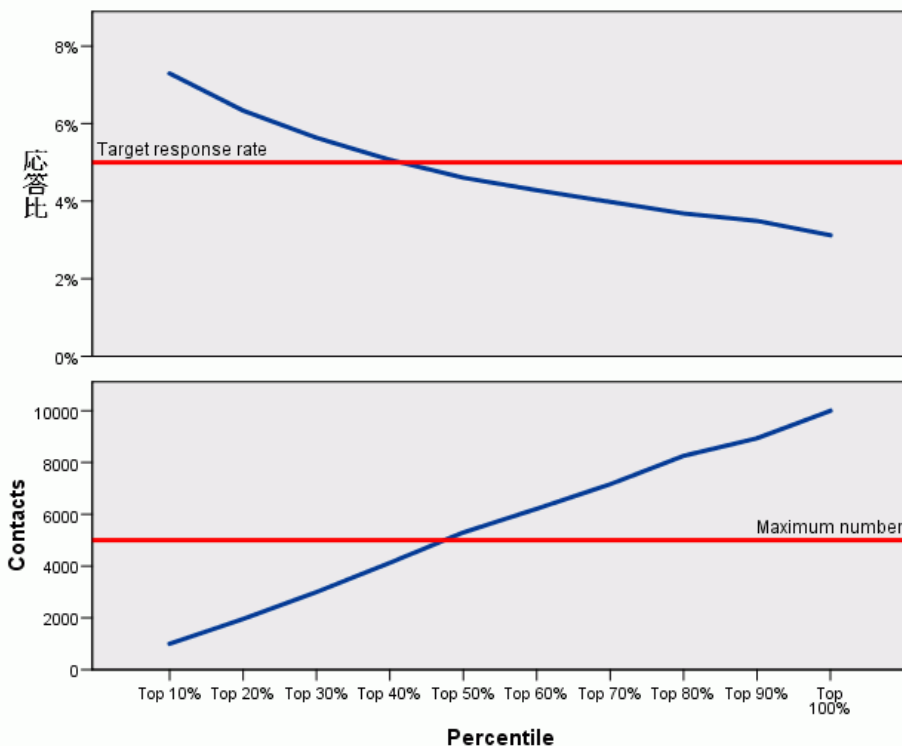
図 5-1
郵便番号ごとの回答率を含むデータセット

	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
1	932	10.0%	4	40	3.6	Top 10%
2	098	8.8%	6	68	5.5	Top 10%
3	740	7.8%	9	116	8.3	Top 10%
4	100	7.7%	7	91	6.5	Top 10%
5	110	7.7%	5	65	4.6	Top 10%
6	954	7.5%	4	53	3.7	Top 10%
7	108	7.3%	6	82	5.6	Top 10%
8	107	7.0%	5	71	4.6	Top 10%
9	090	6.9%	4	58	3.7	Top 10%
10	966	6.9%	4	58	3.7	Top 10%
11	760	6.7%	8	119	7.5	Top 10%
12	113	6.2%	5	80	4.7	Top 10%
13	927	6.0%	3	50	2.8	Top 10%
14	969	6.0%	3	50	2.8	Top 10%

図 5-2
集計表とグラフ

Percentile	応答比	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

緑: 目標回答率を満たす。
赤: 目標回答率を満たしていない。



新しいデータセットには次のフィールドが含まれます。

- **郵便番号**。郵便番号グループが、値の一部のみを基にしている場合、郵便番号の一部を示す値となります。Excel ファイルのこの列のヘッダ行ラベルは、元のデータセットの郵便番号フィールドの名前です。

- **回答率。** 郵便番号ごとの正の回答の割合です。
- **回答。** 郵便番号ごとの正の回答数です。
- **連絡先。** 回答フィールドに欠損値以外の値を投入した、郵便番号ごとの連絡先数の合計です。
- **インデックス。** 式 $N \times P \times (1-P)$ に基づく「重みづけされた」回答です。N は連絡先数を示し、P は、比率で示された回答率です。
- **順位。** 累積郵便番号回答率の10 分位ごとの順位（上位 10%、上位 20% など）で、降順に示します。

郵便番号回答率で一たの考慮事項

回答フィールド。 回答フィールドには、文字または数値を指定できます。このフィールドに購入数または金額を示す値を入力する場合、単一の値がすべての正の回答を示す新しいフィールドを作成する必要があります。[詳細は、p.34 カテゴリ回答フィールドの作成 を参照してください。](#)

正の回答値。 正の回答値は、肯定的に（購入した、など）回答した顧客を示します。欠損値以外の他のすべての回答値は、負の回答を示すと想定します。回答フィールドに定義された値ラベルがある場合、これらのラベルは、ドロップダウン リストに表示されます。

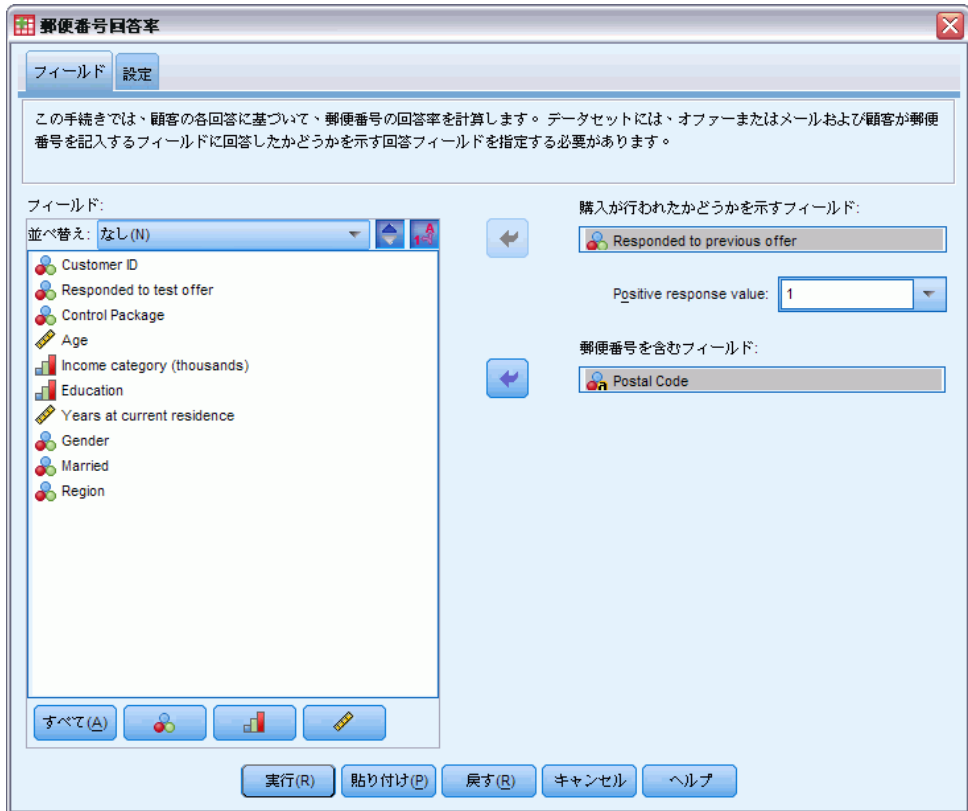
[郵便番号] フィールド [郵便番号] フィールドには、文字または数値を指定できます。

郵便番号回答率を取得するには

メニューから次の項目を選択します。
ダイレクト マーケティング > 手法を選択

- ▶ [上位の回答郵便番号を特定する] を選択します。

図 5-3
郵便番号回答率の [フィールド] タブ



- ▶ オファーに回答した連絡先を示すフィールドを選択します。
- ▶ 正の回答を示す値を入力してください。値が値ラベルを定義している場合、ドロップダウン リストから値ラベルを選択すると、対応する値が表示されます。
- ▶ 郵便番号が入力されたフィールドを選択します。
- ▶ **[実行]** をクリックして手続きを実行します。

オプションとして、次の選択も可能です。

- 完全に値を入力しなくても、郵便番号の最初の n 文字または桁で回答率を生成する。
- 結果は Excel ファイルに自動的に保存する。
- 出力表示を制御する。

設定

図 5-4
郵便番号回答率フィールドの [設定] タブ

基準となる郵便番号のグループ化

レコードをどのようにグループ化して回答率を計算するかを指定します。デフォルトでは、郵便番号全体を使用し、郵便番号がおなじすべてのレコードをグループ化して、グループの回答率を計算します。また、最初の n 桁または文字など、郵便番号の一部を基準にレコードをグループ化することもできます。たとえば、10文字の郵便番号のうち最初の5文字、または5桁の郵便番号のうち最初の3桁を基準にレコードをグループ化できます。出力データセットには、各郵便番号グループに1つつレコードがあります。値を入力する場合、正の整数でなければなりません。

郵便番号の形式

郵便番号フィールドが数値型で、値全体でなく最初の n 桁で郵便番号をグループ化する場合、元の値で桁数を指定する必要があります。桁数は、郵便番号の最大桁数を指定します。たとえば、郵便番号フィールドに 5 桁と 9 桁の郵便番号が混在している場合、桁数として 9 を指定する必要があります。

注：表示形式によっては、一部の 5 桁の郵便番号が 4 桁だけ表示される場合がありますが、先頭に 0 が使用されています。

出力

郵便番号ごとの回答率を含む新しいデータセットのほか、10 分位ごとの順位（上位 10%、上位 20% など）で結果を要約する表やグラフも表示できます。表には各 10 分位ごとの回答率、累積回答率、レコード数、累積レコード数が表示されます。グラフには各 10 分位ごとの累積回答率、累積レコード数が表示されます。

受け入れ可能な回答率の最小値。 目標の回答率の最小値または損益分岐式を入力した場合、表を色分けして累積回答率の最小値に一致する 10 分位を示し、グラフは指定した回答率の最小値に基準線を示します。

- **目標回答率。** 回答率はパーセントで示します（各郵便番号グループの正の回答のパーセント）。値は 0 より大きく 100 未満にする必要があります。
- **式から損益分岐比率を計算。** 式（パッケージ送付のコスト/回答ごとの純利益）x 100 に基づいて、累積回答率の最小値を計算します。2 つの値は正である必要があります。結果は 0 より大きく 100 より小さくなります。たとえば、パッケージ送付のコストが \$0.75 で、回答ごとの純利益が \$56 の場合、最小回答率は $(0.75/56) \times 100 = 1.34\%$ となります。

連絡先の最大数。 連絡先の最大数を指定すると、表を色分けして連絡先（レコード）の累積最大数を超えていない 10 分位を示し、グラフにはその値に基準線が示されます。

- **連絡先のパーセント。** 最大値をパーセントで示します。たとえば、全体の 50% に満たない連絡先を含むもっとも高い回答率の 10 分位を知ることができます。値は 0 より大きく 100 未満にする必要があります。
- **連絡先の数。** 連絡先数をパーセントで示します。たとえば、10,000 パッケージより多く送付したくない場合、値を 10000 に設定できます。値は正の整数でなければなりません（グループ化記号はなし）。

受け入れ可能な回答率の最小値と連絡先の最大数を指定した場合、表の色分けは、最初に条件を満たした方に基づいて行われます。

Excel へエクスポート

この手順では、自動的に郵便番号ごとの回答率を含む新しいデータセットを作成します。データセットの各レコード（行）は、郵便番号を示します。同じ情報を自動的に Excel ファイルに保存できます。ファイルは Excel 97 ~ 2003 形式で保存されます。

カテゴリ回答フィールドの作成

回答フィールドは、1 つの値がすべての正の回答を示す、カテゴリ型でなければなりません。欠損値以外のその他の回答値は、負の回答を示すと想定します。回答フィールドが購入数や購入金額などの連続型（スケール）値を示す場合、単一正の回答がすべての 0 以外の回答値に割り当てられる新しいフィールドを作成する必要があります。

- 負の回答は 0（空白ではなく、欠損値として処理）として記録されません。これは次の数式で計算できます。

`newName=OldName>0`

この場合、NewName は新しいフィールドの名前で、OldName は元のフィールドの名前です。これは、欠損値でない 0 より大きいすべての値に 1 の値を、欠損値でない 0 以下のすべての値に 0 を割り当てる論理式です。

- 負の回答に値が記録されない場合、これらの値は欠損値として処理され、数式は少し複雑になります。

`newName=NOT(MISSING(OldName))`

この論理式では、欠損値でないすべての回答値には 1 の値が、欠損値であるすべての回答値には 0 の値が割り当てられます。

- 負（0）の回答値と欠損値の区別が付かない場合、正確な回答値は計算できません。欠損値が比較的少ない場合、計算される回答率にあまり影響はありません。ただし、データセット全体のうち小さな検定サンプルのみの回答情報が記録された場合など、欠損値が多い場合、計算される回答率は真の回答率より大幅に低くなるため、あまり意味のないものとなります。

カテゴリ回答フィールドをするには

- ▶ メニューから次の項目を選択します。
変換(T) > 変数の計算
- ▶ 目標変数に、新しいフィールド（変数）名を入力します。
- ▶ 負の回答が 0 として記録される場合、数式に `OldName>0` と入力します。
OldName は元のフィールド名です。

- ▶ 負の回答が欠損値（空白）として記録される場合、数式に `NOT(MISSING(OldName))` と入力します。OldName は元のフィールド名です。

購入の傾向

購入の傾向では、テスト送付または以前のキャンペーンの結果を使用して、スコアを生成します。スコアは、回答する可能性の最も高い連絡先を示します。[回答] フィールドは、誰がテスト送付または以前のキャンペーンに回答したかを特定します。[傾向] フィールドは、同様の特性を持つ連絡先が回答する確率を予測するために使用する特性です。

この手法では、予測モデルの作成に 2 項ロジスティック回帰を使用します。予測モデルの作成および適応のプロセスには、次の 2 つの基本ステップがあります。

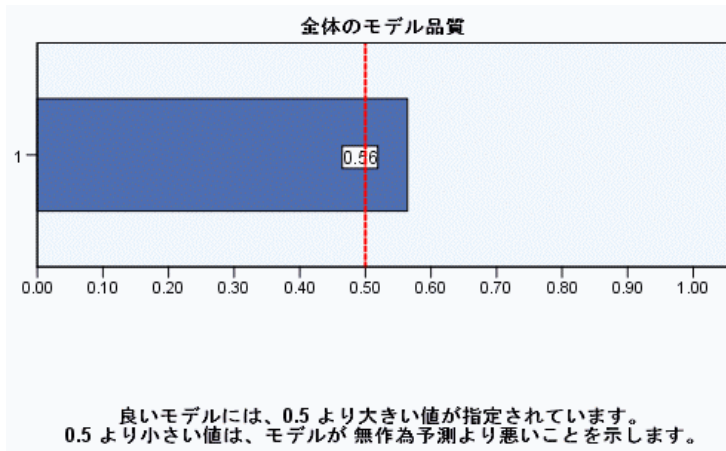
- ▶ モデルを作成し、モデル ファイルを保存します。関心のある出力（目標）が認識されているデータセットを使用してモデルを作成します。たとえば、誰がダイレクト メール キャンペーンに回答する可能性が高いかを予測するモデルを作成する場合、誰が回答し、誰が回答しなかったかについての情報をすぐに含んでいるデータセットから始める必要があります。この場合、小規模な顧客グループへのテスト メールの結果、または過去の類似したキャンペーンに対する回答者の情報となります。
- ▶ そのモデルを異なるデータセット（関心のある結果が認識されていない）に適用し、予測結果を取得します。

例 企業のダイレクト マーケティング部門は、最も回答および購入の傾向が高い連絡先を特定する人口統計上の特性に基づき、さまざまなテスト送付の結果を使用して、連絡先データベースの残りの連絡先に傾向スコアを割り当てます。

出力

この手順では、データセットのスコアリングに使用できる検定データおよび XML モデルファイルの傾向スコアを含むデータセットにファイルを自動的に新規作成します。オプションの診断出力には、予測回答と実際の回答を比較する全体のモデル品質グラフと分類表があります。

図 6-1
全体のモデル品質グラフ



購入の傾向データについての考慮事項

回答フィールド。 回答フィールドには、文字または数値を指定できます。このフィールドに購入数または金額を示す値を入力する場合、単一の値がすべての正の回答を示す新しいフィールドを作成する必要があります。[詳細は、p. 43 カテゴリ回答フィールドの作成](#) を参照してください。

正の回答値。 正の回答値は、肯定的に（購入した、など）回答した顧客を示します。欠損値以外の他のすべての回答値は、負の回答を示すと想定されます。回答フィールドに定義された値ラベルがある場合、これらのラベルは、ドロップダウン リストに表示されます。












次を使用して傾向を予測。 傾向を予測するために使用するフィールドは文字列または数値です。また名義型、順序型、または連続型（尺度型）を指定することができます。ただし、すべての予測フィールドに適切な測定水準を割り当てる必要があります。

測定レベル。 結果の計算に影響を与えるため、適切な測定レベルの割り当ては重要です。

- **名義データ。** 値がランキングなどを持たないカテゴリを表しているとき、名義（変数）として取り扱うことができます。たとえば、従業員の会社の所属などです。名義変数の例としては、地域やジップコードや所属宗教などがあります。

- **順序データ.** 値がランキングをもったカテゴリを表しているとき、変数を順序として取り扱うことができます。たとえば、「かなり不満」から「かなり満足」までのようなサービス満足度のレベルなどです。順序変数の例としては、満足度や信頼度を表す得点や嗜好得点などです。
- **続行.** 値が有意な基準を持った順序カテゴリを表しているとき、変数をスケール（連続型）として扱うことができます。値間の距離の比較などに適切です。スケール変数の例としては、年齢や、千ドル単位で表した所得があります。

各フィールドの隣のアイコンは、現在の測定レベルを示します。

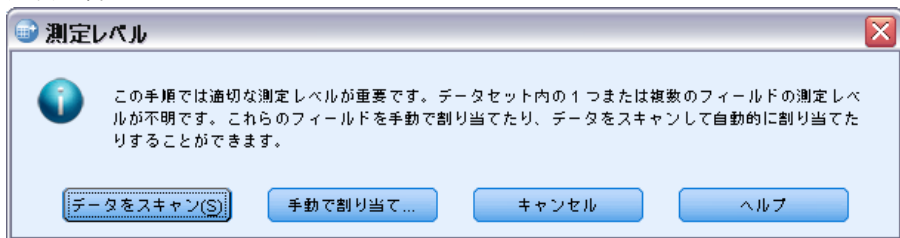
	数値	String	Date	Time
スケール（連続）		利用不可		
順序				
名義				

測定レベルをデータ エディタ の変数ビューで変更することができます。または [変数のプロパティ] ダイアログを使用して各フィールドに適切な測定レベルを提案することができます。

不明な尺度の項目

データセットの 1 つまたは複数の変数（フィールド）の尺度が不明な場合、尺度の警告が表示されます。尺度はこの手順の結果の計算に影響を与えるため、すべての変数に尺度を定義する必要があります。

図 6-2
尺度の警告



- **データをスキャン.** アクティブ データセットのデータを読み込み、デフォルトの尺度を尺度が現在不明なフィールドに割り当てます。データセットが大きい場合は時間がかかります。

- **手動で割り当てる。** 不明な尺度のフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、尺度をこれらのフィールドに割り当てることができます。データ エディタの [変数ビュー] でも、尺度を割り当てることができます。

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

購入傾向スコアを取得するには

メニューから次の項目を選択します。

Direct Marketing > 手法を選択

- ▶ [購入の可能性が高い連絡先を選択する] を選択します。

図 6-3
購入の傾向の [フィールド] タブ



- ▶ オファーに回答した連絡先を示すフィールドを選択します。
- ▶ 正の回答を示す値を入力してください。値が値ラベルを定義している場合、ドロップダウン リストから値ラベルを選択すると、対応する値が表示されます。
- ▶ 傾向を予測するのに使用したいフィールドを選択しています。
他のデータ ファイルをスコアリングするモデル XML ファイルを保存するには
- ▶ [モデル情報を XML ファイルにエクスポート] を選択 (オン) します。
- ▶ ディレクトリ パスまたはファイル名を入力するか、[参照] をクリックして、モデル XML ファイルを保存する場所に移動します。
- ▶ [実行] をクリックして手続きを実行します。
他のデータセットをスコアリングするモデル ファイルを保存するには
- ▶ スコアリングするデータセットを開きます。
- ▶ スコアリング ウィザードを使用して、モデルをデータセットに適用します。メニューから次の項目を選択します。
ユーティリティ > スコアリング ウィザード

設定

図 6-4
購入の傾向の [設定] タブ



購入の傾向

フィールド 設定(N)

モデルの検証

スコア生成に使用するモデルを検証できます。モデルを検証するには、データをデータ区分に分割する必要があります。学習データ区分を使用して、モデルを学習、または作成します。学習データ区分を使用して、モデルを検証します。モデルを検証する場合、この方法ではレコードが自動的にデータ区分に割り当てられます。

モデルを検証(Y)

学習サンプルデータ区分のサイズ(%) (I):

シードを設定して結果を再現(S)

シード(E):

診断出力

全体のモデル品質(Q)

分類表(C)

最小確率(M):

再コード化された回答フィールドの名前およびラベル

この方法では、回答フィールドを、1 が正の回答を、0 が負の回答を示す新しいフィールドに自動的に再コード化します。

新しいフィールド名(N):

新しいフィールドラベル(W):

スコアを保存

この方法では、テスト送付または以前のキャンペーンの結果を使用して、スコアを生成します。スコアは、使用できるよう自動的に保存されます。このタブのその他のコントロールには、保存されたスコアに対する追加のコントロールが用意されています。

スコアの新しいフィールド名(N):

実行(R) 貼り付け(P) 戻す(R) キャンセル ヘルプ

モデルの検証

モデル検証では、診断の目的で、学習グループと検定グループを作成します。[診断出力] セクションで分類表を選択した場合、表は学習セクション（選択）と検定セクション（未選択）に分けられ、比較されます。分類表を選択するまでは、モデル検証を選択しないでください。スコアは学習サンプルから生成されたモデルに基づき、利用可能なレコードの合計より少ないレコード数が常に含まれています。たとえば、デフォルトの学習サンプルサイズは 50% で、利用可能なレコードの半数のみに作成されたモデルは、利用可能なすべてのレコードに作成されたモデルに比べて信用できない場合があります。

- **学習サンプルデータ区分のサイズ(%)**。学習サンプルに割り当てるレコードの割合を指定します。回答フィールドが欠損値でない残りのレコードは、検定サンプルに割り当てられます。値は 0 より大きく 100 未満にする必要があります。
- **シードを設定して結果を再現**。レコードが無作為に学習サンプルや検定サンプルに割り当てられるため、同じ開始乱数シード値を指定していない限り、手順を実行するごとに異なる結果が出る場合があります。

診断出力

全体のモデル品質。全体のモデル品質に関する棒グラフを表示します。値は 0 ~ 1 です。良いモデルには 0.5 より大きい値が指定されます。

分類表。予測された正の回答および負の回答と実際の正の回答および負の回答を比較する表が表示されます。全体の精度を使用して、モデルがどれだけ機能しているかについていくらか示すことができますが、正確に予測された正の回答の割合に、より関心がある場合があります。

- **最小確率**。指定した値より大きいスコア値を持つレコードを、分類表の予測された正の回答カテゴリに割り当てます。手順によって生成されたスコアは、連絡先が正の回答を行う（購入するなど）確率を示します。一般的なルールとして、目標回答率の最小値に近い値を比率で指定する必要があります。たとえば、少なくとも 5% の回答率を考えている場合、0.05 を指定します。値は 0 より大きく 1 未満にする必要があります。

再コード化された回答フィールドの名前およびラベル

この手続きでは、購入フィールドを、1 が正の回答を、0 が負の回答を示す新しいフィールドに自動的に再コード化します。分析が再コード化されたフィールドで実行されます。デフォルトの名前とラベルを上書きして、独自の名前とラベルを指定できます。名前は、IBM® SPSS® Statistics 命名規則に従う必要があります。

スコアを保存

傾向スコアを含む新しいフィールドは、自動的に元のデータセットに保存されます。スコアは、正の回答を確立を比率で示します。

- フィールド名は、SPSS Statistics 命名規則に従う必要があります。
- フィールド名は、データセット内に既に存在するフィールド名と重複することはできません。同じデータセットに複数回この手順を実行する場合、それぞれ別の名前を指定する必要があります。

カテゴリ回答フィールドの作成

回答フィールドは、1 つの値がすべての正の回答を示す、カテゴリ型でなければなりません。欠損値以外のその他の回答値は、負の回答を示すと想定します。回答フィールドが購入数や購入金額などの連続型（スケール）値を示す場合、単一正の回答がすべての 0 以外の回答値に割り当てられる新しいフィールドを作成する必要があります。

- 負の回答は 0（空白ではなく、欠損値として処理）として記録されます。これは次の数式で計算できます。

$\text{NewName} = \text{OldName} > 0$

この場合、NewName は新しいフィールドの名前で、OldName は元のフィールドの名前です。これは、欠損値でない 0 より大きいすべての値に 1 の値を、欠損値でない 0 以下のすべての値に 0 を割り当てる論理式です。

- 負の回答に値が記録されない場合、これらの値は欠損値として処理され、数式は少し複雑になります。

$\text{NewName} = \text{NOT}(\text{MISSING}(\text{OldName}))$

この論理式では、欠損値でないすべての回答値には 1 の値が、欠損値であるすべての回答値には 0 の値が割り当てられます。

- 負 (0) の回答値と欠損値の区別が付かない場合、正確な回答値は計算できません。欠損値が比較的少ない場合、計算される回答率にあまり影響はありません。ただし、データセット全体のうち小さな検定サンプルのみの回答情報が記録された場合など、欠損値が多い場合、計算される回答率は真の回答率より大幅に低くなるため、あまり意味のないものとなります。

カテゴリ回答フィールドをするには

- ▶ メニューから次の項目を選択します。
変換(T) > 変数の計算
- ▶ 目標変数に、新しいフィールド（変数）名を入力します。
- ▶ 負の回答が 0 として記録される場合、数式に $\text{OldName} > 0$ と入力します。OldName は元のフィールド名です。
- ▶ 負の回答が欠損値（空白）として記録される場合、数式に $\text{NOT}(\text{MISSING}(\text{OldName}))$ と入力します。OldName は元のフィールド名です。

対照パッケージ検定

この方法では、マーケティング キャンペーンを比較して、さまざまなパッケージまたはオファーの有効性に重要な相違点があるかどうかを確認します。キャンペーンの有効性は回答によって測定されます。キャンペーンフィールドは、Offer A や Offer B などのさまざまなキャンペーンを特定します。回答フィールドは、連絡先がキャンペーンに回答したかどうかを特定します。回答を購入量（例：“99.99”）として記録する場合、[購入量] を選択します。「はい」または「いいえ」など、回答で連絡先が正の回答または負の回答のどちらを行ったのかを容易に示す場合は [応答] を選択します。

例 企業のダイレクト マーケティング部門が、新しいパッケージ デザインの方が既存のパッケージより多くの正の回答を引き出すかどうかを確認したい場合があります。新しいパッケージ デザインが非常に多くの正の回答率を生み出すかどうかを判断するテスト送付を送信します。テスト送付は、既存のパッケージを受信する対照グループと、新しいパッケージ デザインを受信する検定グループで構成されています。2 つのグループの結果を比較して、大きな違いがあるかどうかを確認します。

出力

出力には、キャンペーン フィールドで定義された各グループの正の回答および負の回答の度数およびパーセントを表示する表、グループ間の差分を示す表があります。

図 7-1
対照パッケージ検定の出力

		対象パッケージ			
		対象		テスト	
		度数	列の N %	度数	列の N %
有効性 (1=Yes 0=No)	0	879	96.6%	984	97.7%
	1	31	3.4%	23	2.3%

対象とテストの間に統計的な有意差はありません。

対照パッケージ検定データの考慮事項および仮定

キャンペーン フィールド。 キャンペーン フィールドは、カテゴリ（名義型または順序型）でなければなりません。

有効性回答フィールド。 有効性フィールドに購入数を選択した場合、フィールドは数値型、測定レベルは連続型（スケール）でなければなりません。

負（購入数の場合、0 の値）の回答値と欠損値の区別が付かない場合、正確な回答値は計算できません。欠損値が比較的少ない場合、計算される回答率にあまり影響はありません。ただし、データセット全体のうち小さな検定サンプルのみの回答情報が記録された場合など、欠損値が多い場合、計算される回答率は真の回答率より大幅に低くなるため、あまり意味のないものとなります。

仮定。 この手順では、連絡先が各キャンペーン グループに無作為に割り当てられていると仮定します。つまり、特定の人口統計、購入履歴、またはその他の特性はグループの割り当てに影響を与えず、すべての連絡先がグループに割り当てられる確率は等しくなります。

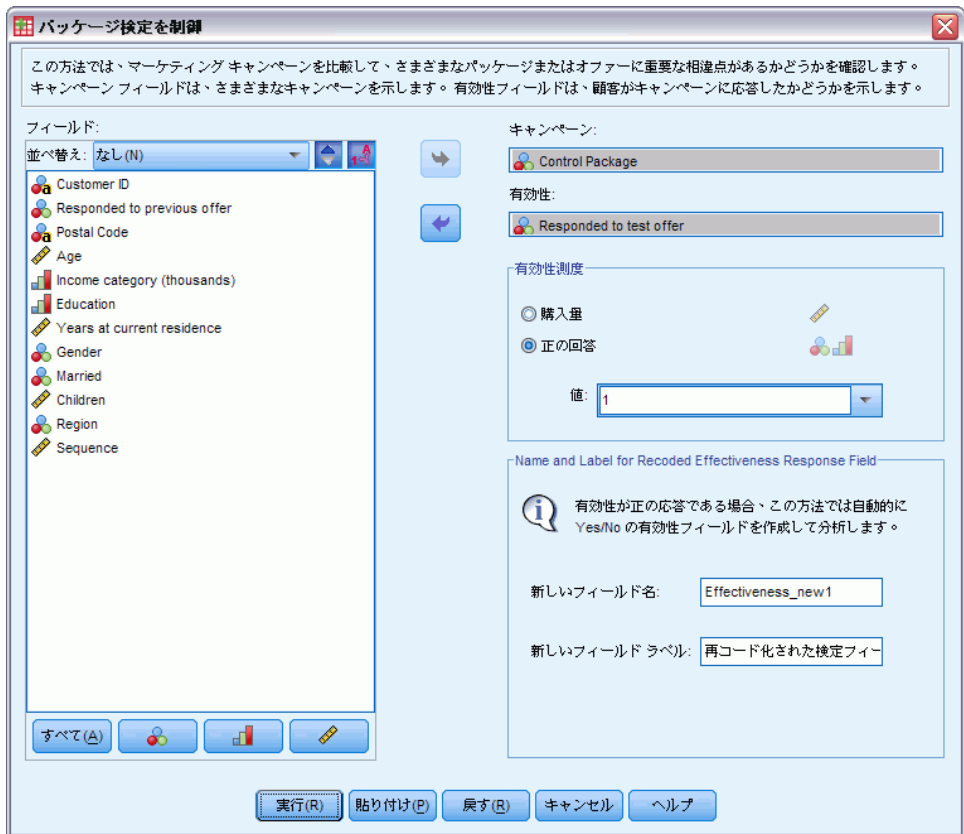
対照パッケージ検定を行うには

メニューから次の項目を選択します。

ダイレクト マーケティング > 手法を選択

- ▶ [キャンペーンの有効性の比較] を選択します。

図 7-2
[対照パッケージ検定] ダイアログ



- ▶ 各連絡先が属するキャンペーン グループを特定するフィールドを選択します (例: オファー A、オファー B など)。このフィールドは名義型または順序型である必要があります。
- ▶ 回答の有効性を示すフィールドを選択します。

回答フィールドが購入数を示す場合、フィールドは数値型でなければなりません。

「はい」または「いいえ」など、回答で連絡先が正の回答または負の回答のどちらを行ったのかを容易に示すのみの場合は [応答] を選択し、正の回答を示す値を入力します。値が値ラベルを定義している場合、ドロップダウン リストから値ラベルを選択すると、対応する値が表示されます。

新しいフィールドが自動的に作成されます。このフィールドでは、1 が正の回答を、0 が負の回答を示し、分析が新しいフィールドで実行されます。デフォルトの名前とラベルを上書きして、独自の名前とラベルを指定できます。名前は、IBM® SPSS® Statistics 命名規則に従う必要があります。

- ▶ [実行] をクリックして手続きを実行します。

パート II: 例

トランザクション データからの RFM 分析

取引データ ファイルでは、データ行は個別の顧客ではなく個別の取引を表し、各顧客に取引の行が複数存在することもあります。この例では、データ ファイル `rfm_transactions.sav` を使用します。詳細は、A 付録 p. 104 サンプル ファイル を参照してください。

取引データ

データセットには、次の情報を含む変数を含んでいる必要があります。

- 各ケース（顧客）を識別する変数または変数の組み合わせ。
- 各取引日付を持つ変数。
- 各取引のマネタリー値を持つ変数。

図 8-1
RFM 取引データ

ID	日付	金額
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/13/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

分析の実行

- ▶ RFM スコアを計算するには、メニューから次の項目を選択します。
Direct Marketing > 手法を選択

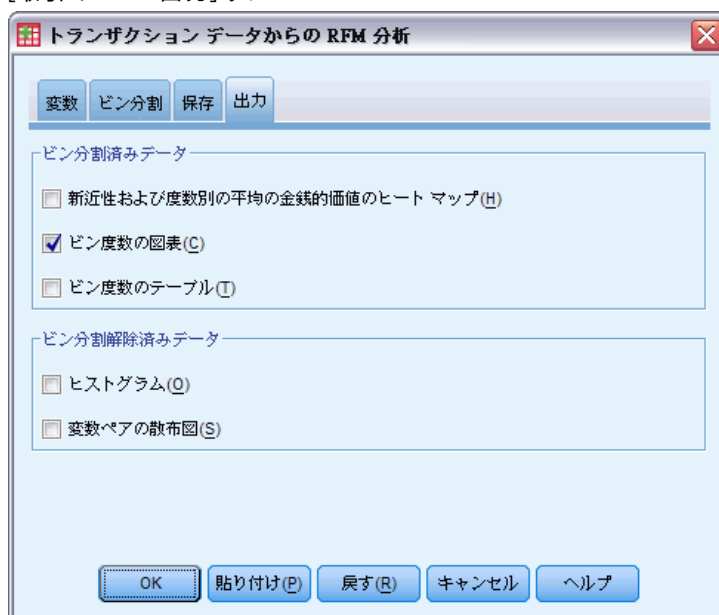
- ▶ [最良の連絡先を特定する (RFM) 分析] を選択し、[続行] をクリックします。
- ▶ [取引データ] をクリックし、[続行] をクリックします。

図 8-2
[取引の RFM: 変数] タブ



- ▶ [戻す] をクリックし、これまでの設定をすべてクリアします。
- ▶ [取引日付] に対し、[購入日付 [日付]] を選択します。
- ▶ [取引量] に対し、[購入量 [量]] を選択します。
- ▶ [集計方法] に対し、[合計] を選択します。
- ▶ [顧客 ID] に対し、[顧客 ID [ID]] を選択します。
- ▶ 次に、[出力] タブをクリックします。

図 8-3
[取引の RFM: 出力] タブ



- ▶ [ビン度数の図表] を選択（チェック）します。
- ▶ [OK] をクリックして手続きを実行します。

結果の評価

取引データの RFM スコアを計算する場合、新しい RFM スコアを含む新しいデータセットが作成されます。

図 8-4
取引データセットの RFM

ID	Date_most_recent	Transaction_count	量	Recency_score	Frequency_score	Monetary_score	RFM_score
1	05/17/2006	10	1313.00	2	3	5	235
2	09/21/2005	11	1230.00	1	5	4	154
3	08/11/2006	13	1194.00	3	5	2	352
4	05/24/2006	9	794.00	2	3	2	232
5	03/13/2005	3	278.00	1	1	2	112
6	07/28/2006	9	922.00	3	2	4	324
7	06/20/2006	11	961.00	2	4	2	242

デフォルトでは、データセットに各顧客に関する次の情報が含まれています。

- 顧客 ID 変数

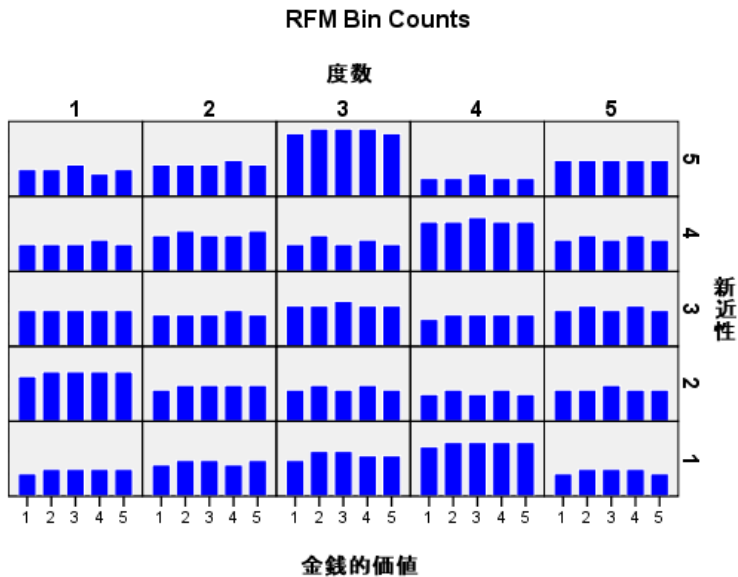
- 最新のトランザクションの日付
- 取引の合計数
- 取引量の集計（デフォルトでは合計）
- リーセンシ、フリクエンシ、マネタリーおよび結合された RFM スコア

新しいデータセットには、各顧客について 1 行（レコード）ずつのみ含まれています。元の取引データは、顧客 ID 変数の値ごとに集計されています。ID 変数は、常に新しいデータセットに含まれています。そうでなければ、RFM スコアを顧客に一致させることができません。

各顧客の結合された RFM スコアは、「(リーセンシ x 100) + (フリクエンシ x 10) + マネタリー」で計算した 3 つのスコアの連結です。

ビューア ウィンドウに表示されるビン度数の図表には、各 RFM カテゴリの顧客数が表示されます。

図 8-5
ビン度数の図表(C)



3 つの RFM コンポーネントそれぞれの 5 つのスコア カテゴリのデフォルトの手法を使用すると、125 の RFM スコア カテゴリが作成されます。図表のそれぞれの棒グラフは、各 RFM カテゴリの顧客数を示します。

理想的には、すべての RFM スコア カテゴリの比較的均等な分布を希望します。実際には、例で表示されたような、ある程度の変動が存在します。空のカテゴリが多く存在する場合、ビン分割方法を変更しようと考えます。

RFM スコアの不均等な分布を扱う方法が多くあります。たとえば、次のとおりです。

- 独立したビン分割方法ではなく、入れ子にされたビン分割方法を使用する。
- スコア カテゴリ（ビン）の数を減らす。
- 大量の同じ値がある場合、異なるカテゴリに同じスコアを持つケースを無作為に割り当てる。

詳細は、2 章 p.7 RFM ビン を参照してください。

スコア データと顧客データの結合

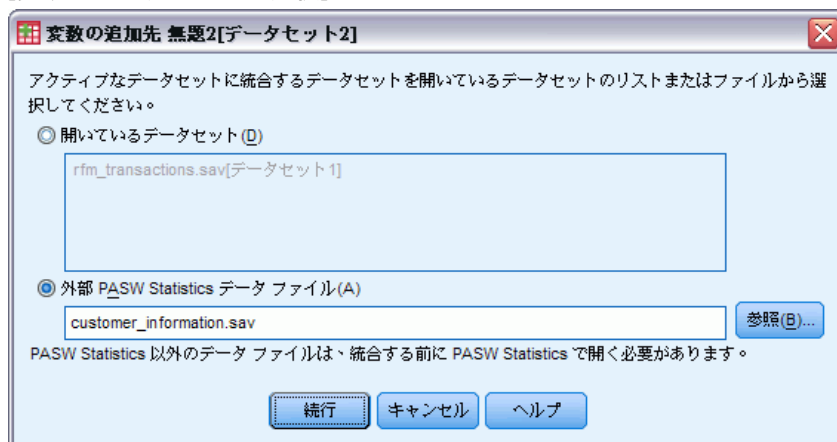
RFM スコアを含むデータセットがあるため、これらのスコアを顧客に一致させる必要があります。スコアを取引データ ファイルに結合することができますが、スコア データを RFM スコア データセット同様、各顧客について 1 行（レコード）ずつ割り当て、顧客の名前およびアドレスのような情報を含むデータ ファイルと結合するのがより一般的です。

図 8-6
[変数ビュー] の RFM スコア データセット

名前	型	幅	小数桁数	ラベル	値
ID	数値	5	0	顧客ID	なし
Date_most_recent	日付	10	0	最新のトランザ...	なし
Transaction_count	数値	7	0	トランザクショ...	なし
量	数値	8	2	量	なし
Recency_score	数値	3	0	新近性スコア	なし
Frequency_score	数値	3	0	度数スコア	なし
Monetary_score	数値	3	0	金銭的価値スコ...	なし
RFM_score	数値	3	0	RFM スコア	なし
名前	文字列	4	0		なし

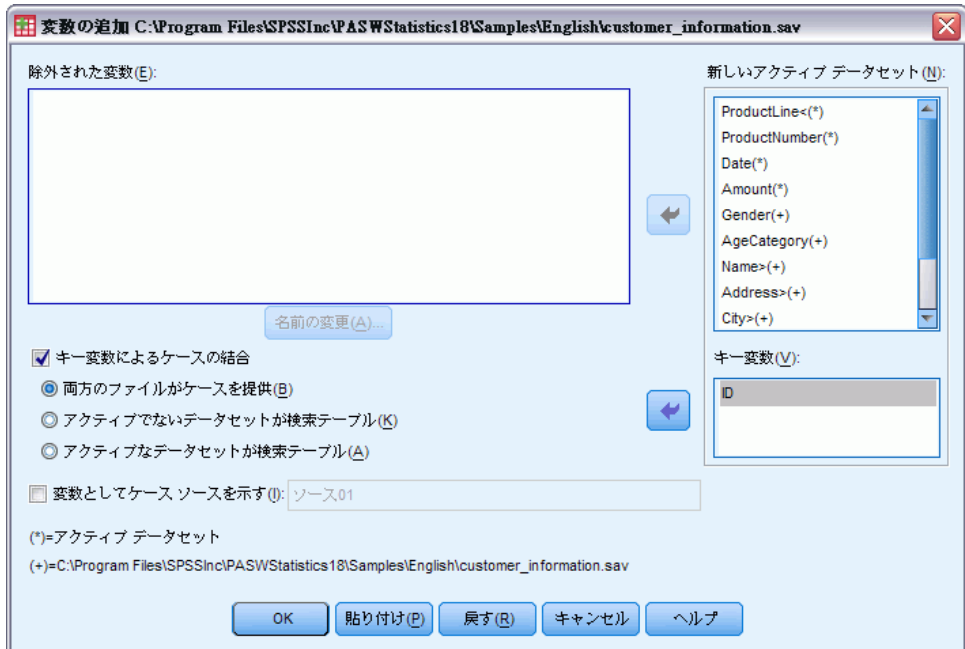
- ▶ RFM スコアを含むデータセットをアクティブなデータセットにします。（データセットを含む [データ エディタ] ウィンドウの任意の場所をクリックします。）
- ▶ メニューから次の項目を選択します。
データ > ファイルの結合 > 変数を追加

図 8-7
[変数の追加、ファイルの選択] ダイアログ



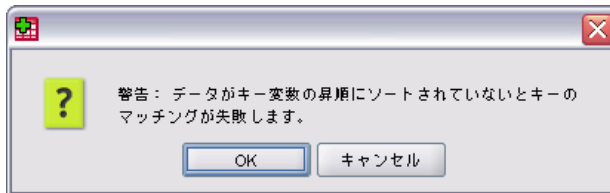
- ▶ [外部データ ファイル] を選択します。
- ▶ [参照] ボタンを使用して、Samples フォルダに移動し、customer_information.sav を選択します。詳細は、A 付録 p.104 サンプル ファイル を参照してください。
- ▶ [続行] をクリックします。

図 8-8
[変数の追加、変数の選択] ダイアログ



- ▶ [キー変数によるケースの結合] チェックボックスをオンにします。
- ▶ [両方のファイルがケースを提供] を選択します。
- ▶ [キー変数] リストの ID を選択します。
- ▶ [OK] をクリックします。

図 8-9
[変数の追加] 警告メッセージ



いずれのファイルもキー変数の昇順でソートするよう警告するメッセージに注意してください。この例では、2つのファイルはすでに、RFMスコアの計算時に選択した顧客ID変数である、キー変数の昇順でソートされています。取引データからRFMスコアを計算する場合、新しいデータセットは顧客ID変数の昇順で自動的にソートされます。スコアデータセットのソート順を変更する場合またはスコアデータセットを結合するデータファイルが昇順にソートされていない場合、まず、2つのファイルを顧客ID変数の昇順でソートする必要があります。

- ▶ [OK] をクリックして、2 つのデータセットを結合します。

RFM スコアを含むデータセットには、各顧客の名前、住所、およびその他の情報が含まれています。

図 8-10
結合されたデータセット

名前	型	幅	小数桁数	ラベル	値
ID	数値	5	0	顧客ID	なし
Date_most_...	日付	10	0	最新のトランザ...	なし
Transaction...	数値	7	0	トランザクショ...	なし
量	数値	8	2	量	なし
Recency_s...	数値	3	0	新近性スコア	なし
Frequency_...	数値	3	0	度数スコア	なし
Monetary_s...	数値	3	0	金銭的価値スコ...	なし
RFM_score	数値	3	0	RFM スコア	なし
名前	文字列	4	0		なし
住所	文字列	7	0		なし
市	文字列	4	0		なし
地方	文字列	14	0		なし
郵便番号	文字列	11	0		なし
国	文字列	7	0		なし
性別	文字列	1	0		{f, 女性}...
年齢カテゴリ	数値	1	0	年齢カテゴリ	{1, <25}...

クラスタ分析

クラスタ分析は、データセット内の自然なグループ（またはクラスタ）を明らかにすることを目的として設計された探索ツールです。たとえば、さまざまな人口統計的および購入の特性に基づいてさまざまな顧客のグループを識別できます。

たとえば、企業のダイレクト マーケティング部門は、顧客データベースの人口統計グループを特定し、マーケティング キャンペーンの戦略を決定し、新しい製品の提供を開発したいと考えています。

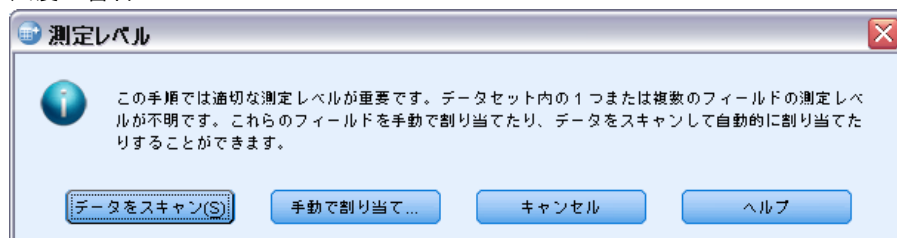
この情報は、dmdata.sav に収集されています。 [詳細は、A 付録 p. 104 サンプル ファイル を参照してください。](#)

分析の実行

- ▶ クラスタ分析を実行するには、メニューから次の項目を選択します。
Direct Marketing > 手法を選択
- ▶ [連絡先をクラスタにセグメント化する] を選択し、[続行] をクリックします。

データセットの 1 つまたは複数の変数（フィールド）の尺度が不明な場合、尺度の警告が表示されます。尺度はこの手順の結果の計算に影響を与えるため、すべての変数に尺度を定義する必要があります。

図 9-1
尺度の警告



- **データをスキャン。** アクティブ データセットのデータを読み込み、デフォルトの尺度を尺度が現在不明なフィールドに割り当てます。データセットが大きい場合は時間がかかります。
- **手動で割り当てる。** 不明な尺度のフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、尺度をこれらのフィールドに割り当てることができます。データ エディタの [変数ビュー] でも、尺度を割り当てることができます。

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

例のファイルでは、尺度が不明なフィールドはなく、すべてのフィールドの尺度が適切であるため、尺度の警告は表示されません。

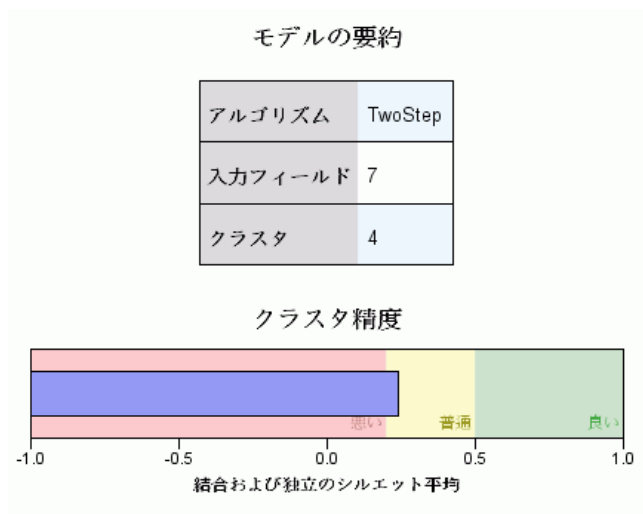
図 9-2
クラスタ分析の [フィールド] タブ



- ▶ セグメントを作成するには、[年齢]、[収入カテゴリ]、[学歴]、[現住所の居住年月]、[性別]、[結婚]、[子供] を選択します。
- ▶ [実行] をクリックして手続きを実行します。

出力

図 9-3
クラスタ モデルの要約

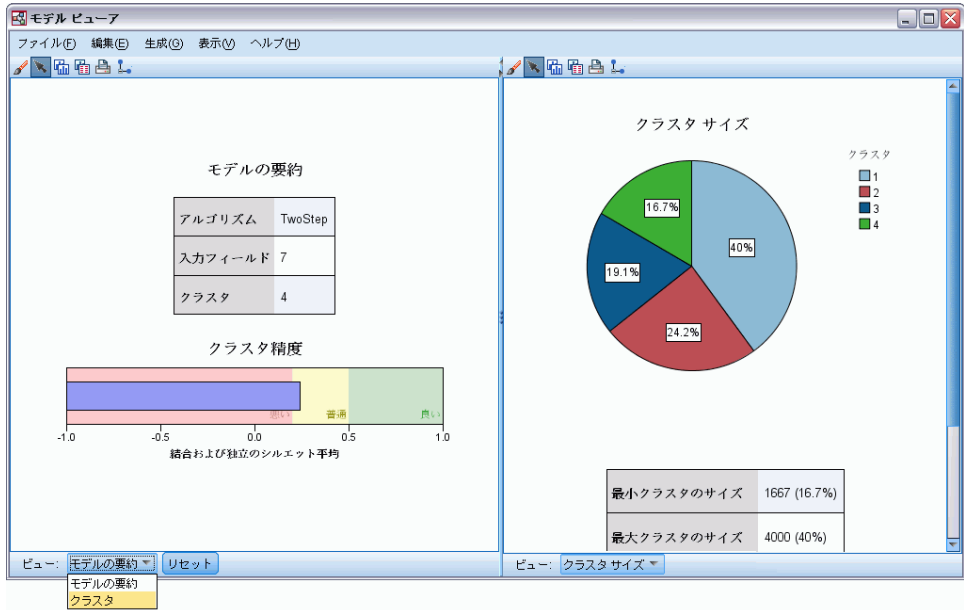


結果はクラスタ モデル ビューアに表示されます。

- モデルの要約では、選択した 7 つの入力フィールドに基づいて、4 つのクラスタが見つかったことを示しています。
- クラスタの品質表は、全体のモデル品質が「普通」の範囲の中間にあることを示しています。





- ▶ クラスタ モデル ビューアの出力をダブル クリックして、モデル ビューアをアクティブにします。

図 9-4
アクティブになったクラスタ モデル ビューア



- ▶ [クラスタ モデル ビューア] ウィンドウの下部にある [ビュー] ドロップダウン リストで [クラスタ] を選択します。

図 9-5
クラスタ ビュー

クラス	1	2	3	4
ラベル				
説明				
サイズ	 40.0% (4000)	 24.2% (2424)	 19.1% (1909)	 16.7% (1667)
機能	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

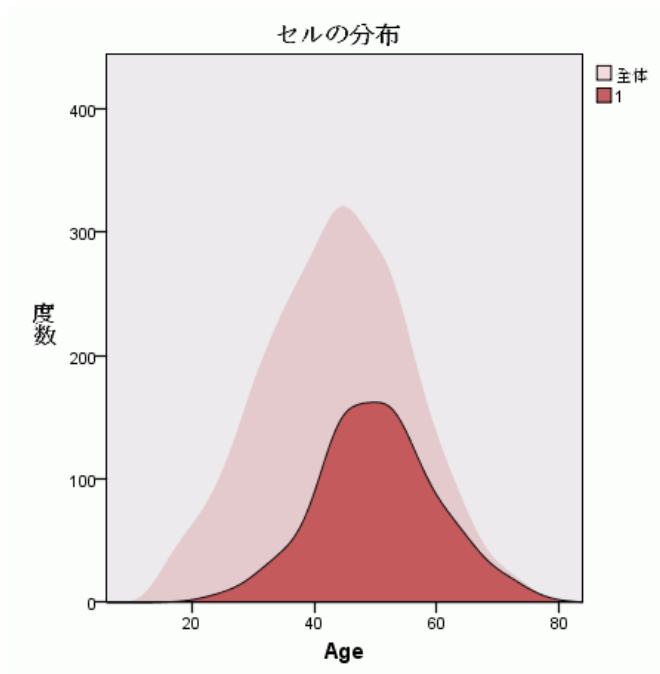
クラスタ ビューには、各クラスタの属性に関する情報が表示されます。

- 連続型（スケール）フィールドの場合、平均値が表示されます。
- カテゴリ（名義型、順序型）フィールドの場合、最頻値が表示されます。最頻値は、最大のレコード数を含むカテゴリです。この例では、各レコードは顧客です。
- デフォルトでは、フィールドはモデルに対する全体の重要度の順に表示されます。この例では、「年齢」の全体の重要度が最も高くなります。また、クラスタ内の重要度やアルファベット順にフィールドを並べ替えることができます。

クラスタ ビューでセルを選択（クリック）すると、クラスタの該当するフィールドの値を集計するグラフを表示できます。

- ▶ たとえば、クラスタ 1 の「年齢」セルをクリックします。

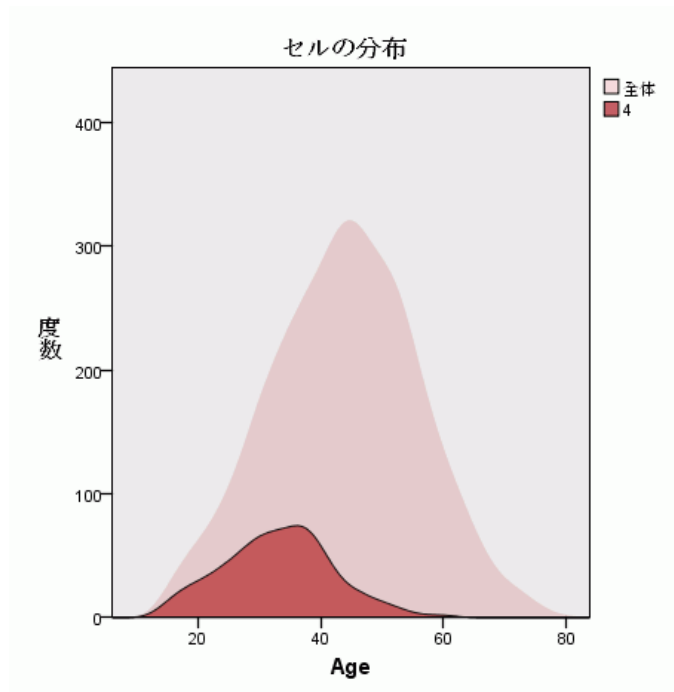
図 9-6
クラスタ 1 の [年齢] ヒストグラム



連続型フィールドの場合、ヒストグラムが表示されます。ヒストグラムには、クラスタ内の値の分布と、フィールドの値の全体の分布が表示されます。ヒストグラムは、クラスタ 1 の顧客の年齢が高いを示します。

- ▶ クラスタ ビューで、クラスタ 4 の「年齢」セルを選択します。

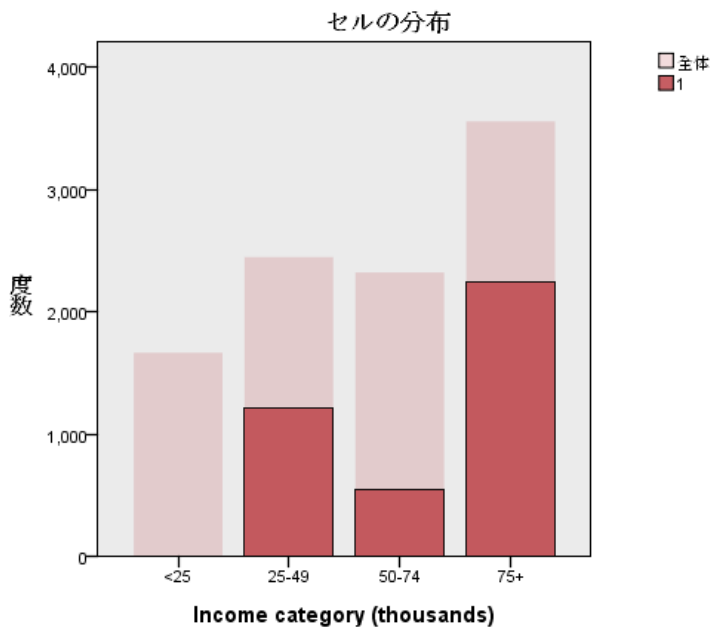
図 9-7
クラスタ 4 の [年齢] ヒストグラム



クラスタ 1 と比較して、クラスタ 4 の顧客は、全体の平均よりも若いことがわかります。

- ▶ クラスタ ビューで、クラスタ 1 の「収入カテゴリ」セルを選択します。

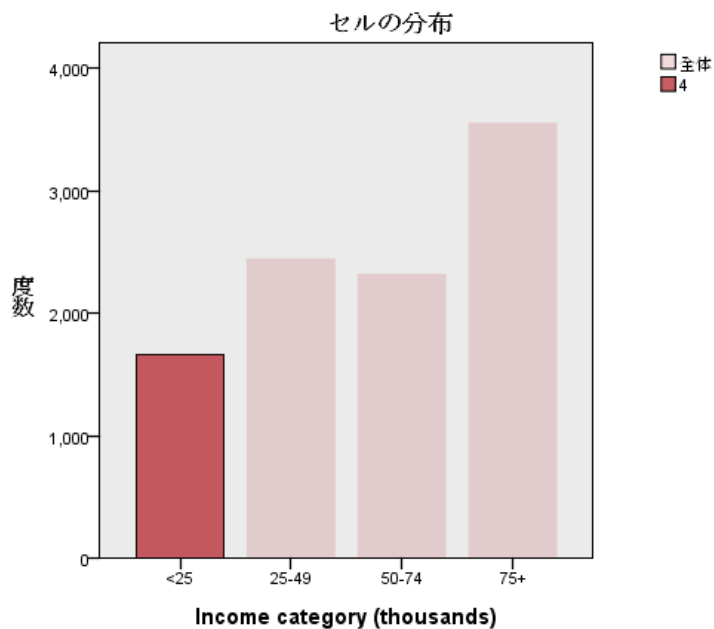
図 9-8
クラスタ 1 の収入カテゴリの棒グラフ



カテゴリ フィールドの場合、棒グラフが表示されます。このクラスタの収入カテゴリ棒グラフの最も注意すべき特徴は、最も低い収入カテゴリに顧客がまったくいないことです。

- ▶ クラスタ ビューで、クラスタ 4 の「収入カテゴリ」セルを選択します。

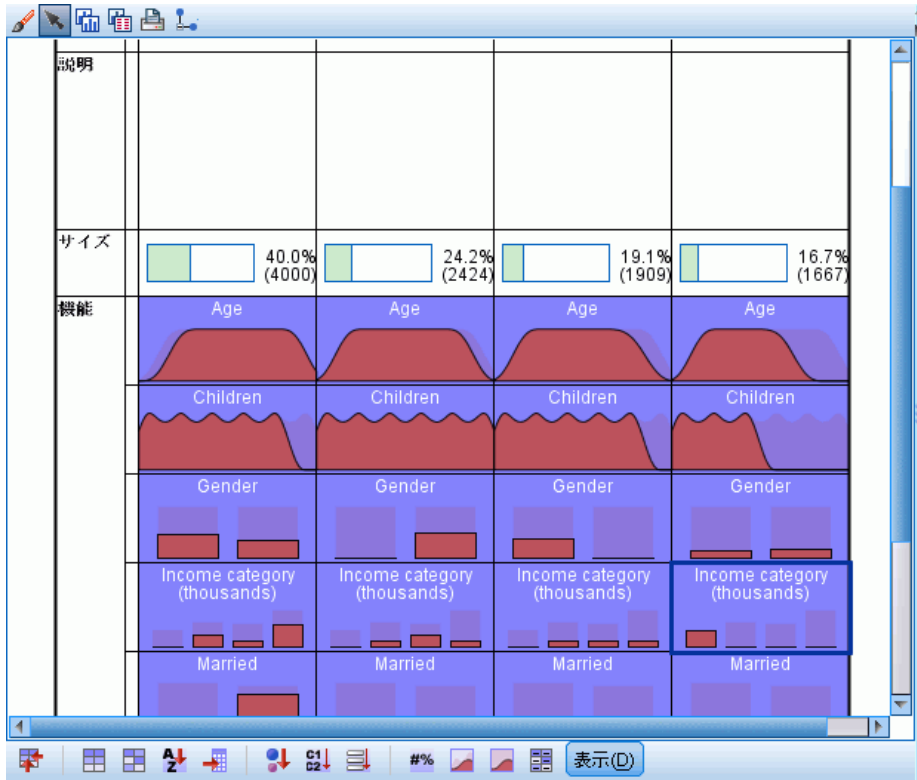
図 9-9
クラスタ 4 の収入カテゴリの棒グラフ



クラスタ 1 と比較して、クラスタ 4 の顧客はすべて、最も低い収入カテゴリであることがわかります。

クラスタ ビューを変更して、セルのグラフを表示することもできます。ビューを変更する、モデル ビューアの下部にあるツールバーを使用して、クラスタ間の値の分布を容易にかつ迅速に比較できるようになります。

図 9-10
クラスタに表示されたグラフ



クラスタ ビューと各セルのグラフに表示された追加情報を見ると、クラスタ間でいくつかの相違点を確認できます。

- クラスタ 1 の顧客は年齢が高く、既婚で子供があり、収入が高い傾向になります。
- クラスタ 2 の顧客はいくらか年齢が高く、シングル マザーで、中程度の収入の傾向になります。
- クラスタ 3 の顧客は若く、子供のいない独身男性の傾向があります。
- クラスタ 4 の顧客は若く、子供のいない独身女性で、低収入の傾向にあります。

クラスタ ビューの [説明] セルは、各クラスタの説明を追加して編集できるテキスト フィールドです。

図 9-11
クラスタの説明を表示したクラスタ ビュー

クラスタ	1	2	3	4
ラベル				
説明	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
サイズ	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
機能	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

クラスタに基づいたレコードの選択

所属クラスタに基づくレコードを、次の 2 とおりの方法で選択できます。

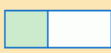
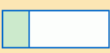


- クラスタ モデル ビューアで、インタラクティブにフィルタ条件を作成します。
- 手続きによって生成されたクラスタ フィールドの値を使用して、フィルタ条件または選択条件を指定します。

クラスタ モデル ビューアでのフィルタ作成

クラスタ モデル ビューアで、特定のクラスタからレコードを選択するフィルタ条件を作成するには

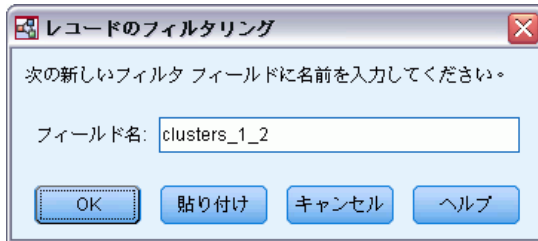
- ▶ クラスタ モデル ビューアをアクティブ化（ダブルクリック）します。
- ▶ [クラスタ モデル ビューア] ウィンドウの下部にある [ビュー] ドロップダウン リストで [クラスタ] を選択します。
- ▶ クラスタ ビューの上部で必要なクラスタのクラスタ番号をクリックします。複数のクラスタを選択する必要がある場合、該当する各追加クラスタ番号を、Ctrl キーを押しながらクリックします。

図 9-12
クラスタ ビューで選択したクラスタ

クラスタ	1	2	3	4
ラベル				
説明	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
サイズ	 40.0% (4000)	 24.2% (2424)	 19.1% (1909)	 16.7% (1667)
機能	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

- ▶ クラスタ モデル ビューアのメニューから、次の項目を選択します。
生成 > レコードのフィルタリング

図 9-13
[レコードのフィルタリング] ダイアログ



- ▶ フィルタ フィールドの名前を入力して、[OK] をクリックします。名前は、IBM® SPSS® Statistics 命名規則に従う必要があります。

図 9-14
データ エディタで除外されたレコード

	ID	Married	Children	Region	ClusterGroup1	clusters_1_2
14	03623	No	0	West	3	.00
15	01353	No	0	West	3	.00
16	07055	No	0	West	3	.00
17	04455	No	0	West	2	1.00
18	07210	No	1	West	2	1.00
19	08054	No	0	West	4	.00
20	06937	No	0	West	4	.00
21	06512	No	0	West	4	.00
22	08315	No	0	West	4	.00
23	09676	No	3	West	2	1.00
24	09636	No	0	West	4	.00
25	08579	No	1	West	2	1.00
26	01480	No	1	West	2	1.00

データセットに新しいフィールドを作成し、フィールドの値に基づいて、データセットのレコードをフィルタリングします。

- フィルタ フィールドの値が 1 のレコードは、今後の分析、グラフ、レポートに含まれます。
- フィルタ フィールドの値が 0 のレコードは除外されます。
- 除外されたレコードは、データセットから削除されません。レコードはフィルタ ステータス インジケータで保持され、データ エディタのレコード番号に斜線として表示されます。

クラスタ フィールド値に基づいたレコードの選択

デフォルトでは、クラスタ分析は、各レコードのクラスタ グループを特定する新しいフィールドを作成します。このフィールドのデフォルト名は ClusterGroupn です。n は、一意のフィールド名にする整数です。

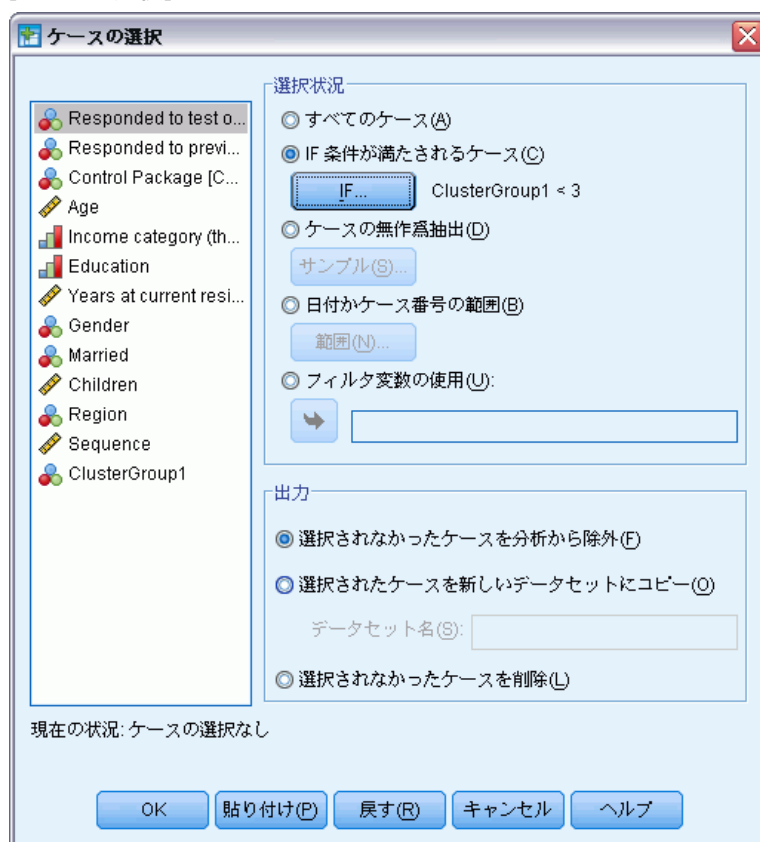
図 9-15
データセットに追加されたクラスタ フィールド

	ID	Gender	Married	Children	Region	ClusterGroup1
1	01359	Female	No	0	West	4
2	06262	Female	No	1	West	2
3	08031	Male	No	0	West	3
4	01971	Male	No	0	West	4
5	09689	Male	No	0	West	3
6	06108	Male	No	1	West	3
7	09853	Male	No	0	West	3
8	06802	Male	No	0	West	4
9	07597	Male	No	0	West	3
10	03692	Male	No	1	West	3
11	00071	Male	No	0	West	4
12	00769	Male	No	0	West	3

クラスタ フィールドの値を使用して特定のクラスタのレコードを選択するには

- ▶ 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

図 9-16
[ケースの選択] ダイアログ



- ▶ [ケースの選択] ダイアログ ボックスで、[IF 条件が満たされるケース] を選択して [IF] をクリックします。

図 9-17
[ケースの選択: IF] ダイアログ



- ▶ 選択条件を入力します。

たとえば、[ClusterGroup1 < 3] を選択すると、クラスタ 1 および 2 のすべてのレコードを選択し、クラスタ 3 以降のレコードを除外します。

- ▶ [続行] をクリックします。

[ケースの選択] ダイアログで、選択したレコードと選択していないレコードを処理するためのオプションがいくつかあります。

選択されなかったケースを分析から除外。 フィルタ条件を指定する新しいフィールドを作成します。除外されたレコードは、データセットから削除されません。レコードはフィルタ ステータス インジケータで保持され、データ エディタのレコード番号に斜線として表示されます。クラスタ モデル ビューアで、インタラクティブにクラスタを選択することと同じです。

選択されたケースを新しいデータセットにコピー。 フィルタ条件を満たすレコードのみを含む新しいデータセットを現在のセッションで作成します。元のデータセットは影響されません。

選択されなかったケースの削除。 選択されなかったレコードは、データセットから削除されます。削除されたレコードを回復するには、変更を保存しないでファイルを終了し、再びファイルを開くしかありません。データファイルに変更を保存すると、ケースは永久に削除されます。

[ケースの選択] ダイアログには、既存の変数をフィルタ変数（フィールド）として使用するオプションもあります。クラスタ モデル ビューアでフィルタ条件をインタラクティブに作成し、データセットと生成されたフィールド フィールドを保存すると、後続のセッションでそのフィルタを使用して、レコードを除外できます。

要約

クラスタ分析は、データセット内の自然なグループ（またはクラスタ）を明らかにできる、有用な探索ツールです。これらのクラスタの情報を使用して、マーケティング キャンペーン戦略を決定し、新しい製品提供を開発します。所属クラスタに基づくレコードを、高度な分析またはターゲット化されたマーケティング キャンペーン向けに選択できます。

見込みのプロファイル

見込みのプロファイルでは、以前のまたはテスト キャンペーンの結果を使用して、説明的プロファイルを作成します。プロファイルを使用して、特定グループの連絡先を目標に設定することができます。たとえば、テスト送付の結果に基づき、企業のダイレクト マーケティングの担当部門は、人口統計情報を基に、特定の種類のオファーに回答する確率が最も高い顧客の種類に関するプロファイルを生成する場合があります。これらの結果に基づいて、同様のオファーに使用する必要のあるメーリング リストの種類を判断することができます。

たとえば、企業のダイレクト マーケティング部門は、顧客データベースのうち約 20% にテスト送付するとします。このテスト送付のきっかけは、年齢、性別、婚姻状況、地理的領域など、各顧客に関する陣頭統計上の特性も含むデータ ファイルに記録されます。結果は、テスト送付に回答した顧客、回答しなかった顧客を示す単純な yes/no の方式で記録されます。

この情報は、dmdata.sav に収集されています。詳細は、A 付録 p. 104 サンプル ファイル を参照してください。

データの考慮事項

回答フィールドは、1 つの値がすべての正の回答を示す、カテゴリ型でなければなりません。欠損値以外のその他の回答値は、負の回答を示すと想定します。回答フィールドが購入数や購入金額などの連続型（スケール）値を示す場合、単一正の回答がすべての 0 以外の回答値に割り当てられる新しいフィールドを作成する必要があります。詳細は、4 章 p. 27 カテゴリ回答フィールドの作成 を参照してください。

分析の実行

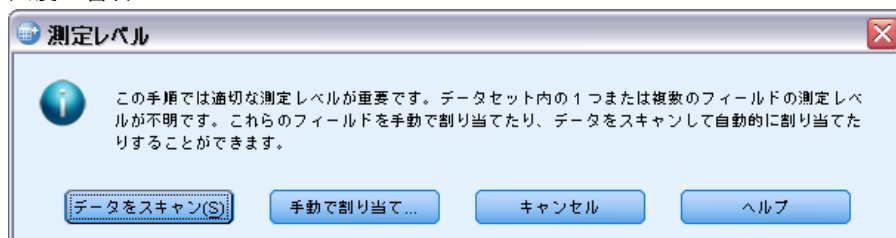
- ▶ 見込みのプロファイル分析を実行するには、メニューから次の項目を選択します。

Direct Marketing > 手法を選択

- ▶ [オファーに回答した連絡先のプロファイルを生成] を選択し、[続行] をクリックします。

データセットの 1 つまたは複数の変数（フィールド）の尺度が不明な場合、尺度の警告が表示されます。尺度はこの手順の結果の計算に影響を与えるため、すべての変数に尺度を定義する必要があります。

図 10-1
尺度の警告



- **データをスキャン。** アクティブ データセットのデータを読み込み、デフォルトの尺度を尺度が現在不明なフィールドに割り当てます。データセットが大きい場合は時間がかかります。
- **手動で割り当てる。** 不明な尺度のフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、尺度をこれらのフィールドに割り当てることができます。データ エディタの [変数ビュー] でも、尺度を割り当てることができます。

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

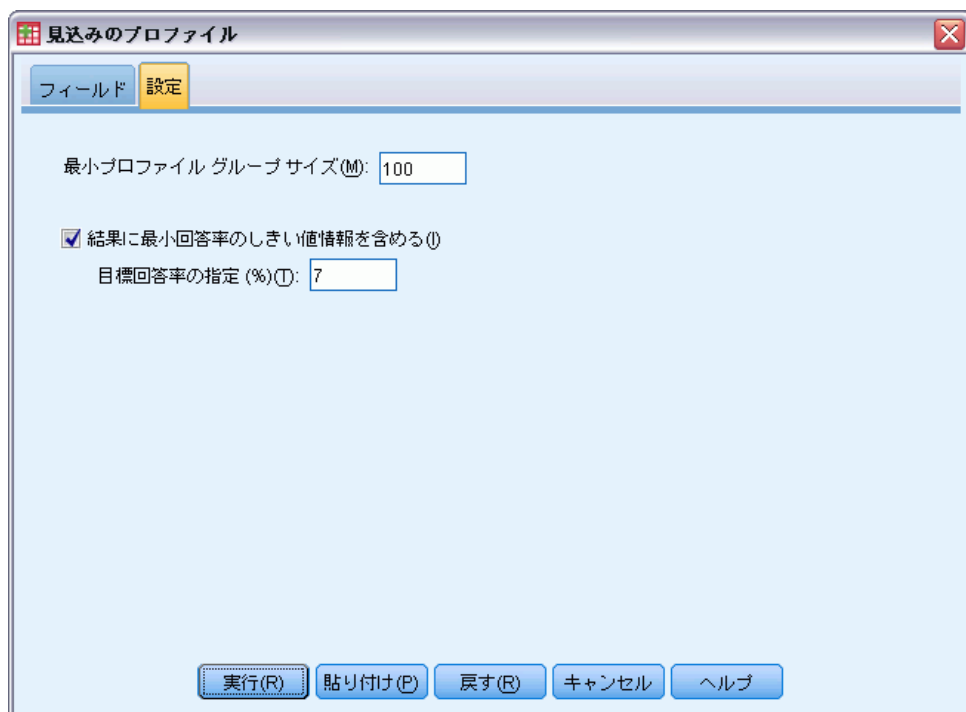
例のファイルでは、尺度が不明なフィールドはなく、すべてのフィールドの尺度が適切であるため、尺度の警告は表示されません。

図 10-2
見込みのプロファイルの [フィールド] タブ



- ▶ 回答フィールドの場合、[テスト オファーに回答済み] を選択します。
- ▶ 正の回答値の場合、ドロップダウン リストから [はい] を選択します。
「はい」は実際、記録された 1 の値に関連する値ラベルであるため、1 の値がテキスト フィールドに表示されます（正の回答値に値ラベルが定義されていない場合、テキスト フィールドに値を入力できます）。
- ▶ [次を使用してプロファイルを作成] の場合、[年齢]、[収入カテゴリ]、[学歴]、[現住所の居住年月]、[性別]、[結婚]、[地域]、[子供] を選択します。
- ▶ [設定] タブをクリックします。

図 10-3
見込みのプロファイルの [設定] タブ



- ▶ [結果に最小回答率のしきい値情報を含める] をオンにします。
- ▶ 目標回答率に、7 と入力します。
- ▶ [実行] をクリックして手続きを実行します。

出力

図 10-4
回答率のテーブル

数	プロフィール			
	説明	グループ サイズ	回答率	累積回答率
1	Region = "West","South","East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West","South","East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West","South","East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

緑: 目標回答率を満たす。
赤: 目標回答率を満たしていない。

回答率の表には、手続きで特定された各プロフィール グループの情報が表示されます。

- プロファイルは降順または回答率の順に表示されます。
- 回答率は、肯定的に（購入した、など）回答した顧客の割合を示します。
- 累積回答率は、現在の行および先行するすべてのプロフィール グループの結合された回答率です。プロフィールは、回答率の高いものから順に並んでいるため、累積回答率は、現在のプロフィール グループと回答率の高いすべてのプロフィール グループとを加算した分の回答率の結合となります。
- プロファイルの説明には、モデルに大きく貢献するフィールドのみの特性が示されます。この例では、地域、性別、婚姻状況がモデルに含まれます。年齢、収入、学歴、現住所の居住年数といった残りのフィールドは、モデルに大きく貢献しないため、モデルには含まれません。
- 表の緑色の領域は、累積回答率が指定した目標回答率以上（この例では7%以上）のプロファイルのセットを示します。
- 表の赤い領域は、累積回答率が指定した目標回答率より小さいプロフィールのセットを示します。
- 表内の最後の行の累積回答率は、全プロフィール グループの回答率であり、テスト送付に含まれるすべての顧客の全体または平均回答率です。

表に示された結果により、西部、南部、東部の女性を対象とする場合、目標回答率より少し上回る回答率を取得することがわかります。

ただし、それらの地域の未婚女性（9.2%）と既婚女性（5.0%）の回答率の間にはかなりの違いがあります。2つのグループの累積回答率は目標回答率を上回りますが、既婚女性の回答率は実際、目標回答率を下回るため、モデルの質を向上させる別の特性を探する必要があります。

スマート出力

図 10-5
スマート出力

回答率はテーブルには、手順によって示される各プロファイルグループの情報が表示されます。プロファイルの説明には、モデルに有意な寄与を提供するこれらのフィールドの特性のみについて記述されています。有意な寄与のないフィールドについては含まれていません。

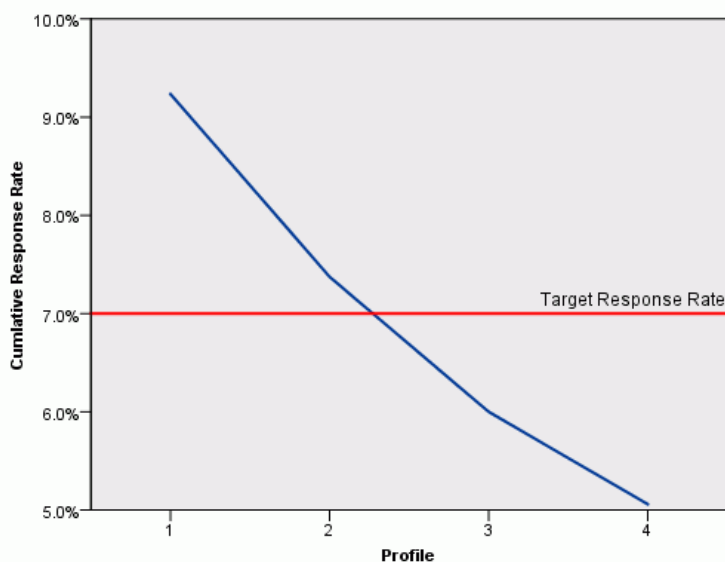
プロファイルは回答率が高い順に表示されます。回答率は、肯定的に反応した（購入した）顧客の割合です。

累積回答率は、現在およびすべての先行プロファイルグループの結合回答率です。プロファイルが回答率の高い順に表示されるため、累積回答率は現在のプロファイルグループと回答率の高いすべてのプロファイルグループとの結合回答率です。です。

指定した目標回答率は7.00%です。緑色の行には7.00%を超える累積回答率が表示され、赤い行には7.00%を下回る累積回答率が表示されます。緑色の領域のプロファイルグループの回答率が7.00%を下回る場合がありますが、累積回答率は7.00%を超えています。

表には、表の解釈方法に関する全般的な情報や、表に記載されている結果の具体的な情報を提供する「スマート出力」が付属しています。

図 10-6
累積回答率図表



累積回答率図表は、表に表示される累積回答率を視覚的に示すものです。プロファイルは回答率の高い順に報告されるため、累積回答率のラインは後続のプロファイルごとに低下していきます。表と同様、グラフは累積回答率がプロファイルグループ2とプロファイルグループ3の間で目標回答率を下回ることを示しています。

要約

この特定のテスト送付について、4 つのプロファイル グループが特定され、結果はオファーに回答した人物かどうかに関連すると思われる重要な人口統計的特性は、性別、地域、婚姻状況のみであることを示します。回答率の高いグループは、南部、東部、西部にすむ未婚女性で構成されています。それに対し、同じ地域の既婚女性の累積回答率は目標回答率を上回りますが、回答率は急速に低下します。

郵便番号回答率

この方法では、以前のキャンペーンの結果を使用して、郵便番号の回答率を計算します。これらの比率は、今後のキャンペーンの対象の特定郵便番号を目標に設定できます。

たとえば、以前の送付の結果に基づいて、企業のダイレクト マーケティング部門は、郵便番号ごとに回答率を生成します。受け入れ可能な回答率の最小値および/または送付に使用する連絡先の最大数などのさまざまな基準をもとに、特定の郵便番号を対象とすることができます。

この情報は、`dmdata.sav` に収集されています。 [詳細は、A 付録 p. 104 サンプル ファイル を参照してください。](#)

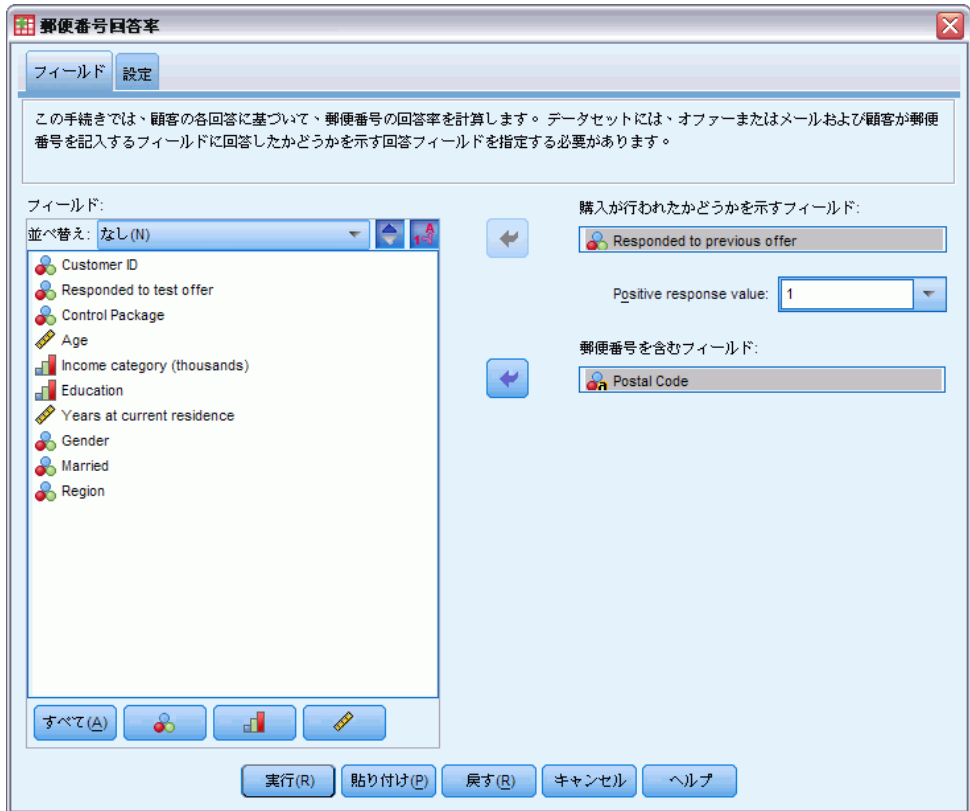
データの考慮事項

回答フィールドは、1 つの値がすべての正の回答を示す、カテゴリ型でなければなりません。欠損値以外のその他の回答値は、負の回答を示すと想定します。回答フィールドが購入数や購入金額などの連続型（スケール）値を示す場合、単一正の回答がすべての 0 以外の回答値に割り当てられる新しいフィールドを作成する必要があります。 [詳細は、5 章 p. 34 カテゴリ回答フィールドの作成 を参照してください。](#)

分析の実行

- ▶ 郵便番号回答率を計算するには、メニューから次の項目を選択します。
Direct Marketing > 手法を選択
- ▶ [上位の回答郵便番号を特定する] を選択し、[続行] をクリックします。

図 11-1
郵便番号回答率の [フィールド] タブ



- ▶ 回答フィールドの場合、[以前のオファーに回答済み] を選択します。
- ▶ 正の回答値の場合、ドロップダウン リストから [はい] を選択します。
「はい」は実際、記録された 1 の値に関連する値ラベルであるため、1 の値がテキスト フィールドに表示されます（正の回答値に値ラベルが定義されていない場合、テキスト フィールドに値を入力できます）。
- ▶ 郵便番号フィールドの場合は、[郵便番号] を選択します。
- ▶ [設定] タブをクリックします。

図 11-2
郵便番号回答率フィールドの [設定] タブ

郵便番号回答率

フィールド 設定

基準となる郵便番号のグループ化

完了値(C)
 最初の 3 桁または 3 文字
 最初の 5 桁または 5 文字
 最初の N 桁または N 文字(F)
 N:

郵便番号の形式

元の郵便番号の記述方法

3 桁(D)
 5 桁(I)
 9 桁
 その他(O)
 桁数(G):

出力

回答率および容量分析(S)

受け入れ可能な回答率の最小値

最小値なし(U)
 目標回答率 (%) (E)
 式から損益分岐比率を計算(K)
 パッケージ送付のコスト(I):
 回答ごとの純利益(P):

連絡先の最大数

すべての連絡先(A)
 連絡先のパーセント(T)
 連絡先の数(M)

Excelへエクスポート

郵便番号回答率を Excel にエクスポート(A)

ファイル名(L):

- ▶ [基準となる郵便番号のグループ化] グループで、[最初の 3 桁または 3 文字] を選択します。最初の 3 桁または 3 文字が同じ郵便番号を持つすべての連絡先の、結合された回答率が計算されます。たとえば、米国の郵便番号の最初の 3 桁は、全 5 桁の郵便番号で定義された領域よりも大きな共通の領域を示します。
- ▶ [出力] グループで、[回答率および容量分析] を選択 (オン) にします。
- ▶ [目標回答率] を選択し、値に 5 を入力します。
- ▶ [連絡先数] を選択し、値として 「5000」と入力します。
- ▶ [実行] をクリックして手続きを実行します。

出力

図 11-3
郵便番号ごとの回答率を含む新しいデータセット

	PostalCode	ResponseRate	Responses	Contacts	Index	Rank	
1	932	10.0%	4	40	3.6	Top 10%	
2	098	8.8%	6	68	5.5	Top 10%	
3	740	7.8%	9	116	8.3	Top 10%	
4	100	7.7%	7	91	6.5	Top 10%	
5	110	7.7%	5	65	4.6	Top 10%	
6	954	7.5%	4	53	3.7	Top 10%	
7	108	7.3%	6	82	5.6	Top 10%	
8	107	7.0%	5	71	4.6	Top 10%	
9	090	6.9%	4	58	3.7	Top 10%	
10	966	6.9%	4	58	3.7	Top 10%	
11	760	6.7%	8	119	7.5	Top 10%	
12	113	6.2%	5	80	4.7	Top 10%	
13	927	6.0%	3	50	2.8	Top 10%	
14	969	6.0%	3	50	2.8	Top 10%	

新しいデータセットが自動的に作成されます。このデータセットには、各郵便番号に 1 つのレコード（行）が含まれています。この例では、各行に最初の 3 桁または 3 文字が同じすべての郵便番号の集計情報が含まれます。

郵便番号が入力されたフィールドのほか、新しいデータセットには次のフィールドが含まれています。

- **回答率。** 郵便番号ごとの正の回答の割合です。レコードは自動的に回答率が高いものから順に並べ替えられます。そのため、回答率が最も高い郵便番号は、データセットの最上位に表示されます。
- **回答。** 郵便番号ごとの正の回答数です。
- **連絡先。** 回答フィールドに欠損値以外の値を投入した、郵便番号ごとの連絡先数の合計です。
- **インデックス。** 式 $N \times P \times (1-P)$ に基づく「重みづけされた」回答です。N は連絡先数を示し、P は、比率で示された回答率です。回答率が同じ 2 つの郵便番号について、この数式は、連絡先数が多い郵便番号に高いインデックス値を割り当てます。
- **順位。** 累積郵便番号回答率の 10 分位ごとの順位（上位 10%、上位 20% など）で、降順に示します。

[郵便番号回答率] ダイアログの [設定] タブで [回答率および容量分析] を選択しているため、回答率の集計表及びグラフがビューアに表示されます。

図 11-4
回答率のテーブル

Percentile	応答比	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

緑: 目標回答率を満たす。
赤: 目標回答率を満たしていない。

この表は、10 分位ごとの順位（上位 10%、上位 20% など）で結果を要約します。

- 累積回答率は、現在の行および先行するすべての行の正の回答の結合された割合です。結果は、回答率の高いものから順に並んでいるため、これは現在の 10 分位回答率の高いとすべての 10 分位の結合された回答率となります。
- 表は、目標回答率および連絡先の最大数に入力した値に基づいて色分けすることができます。累積回答率が 5% 以上で累積連絡先数が 5,000 以下の行は緑色です。色分けは、どの行が最初にしきい値に到達するかによって決定します。この例では、同じ 10 分位で両方のしきい値に達しています。

図 11-5
回答率表のスマート出力

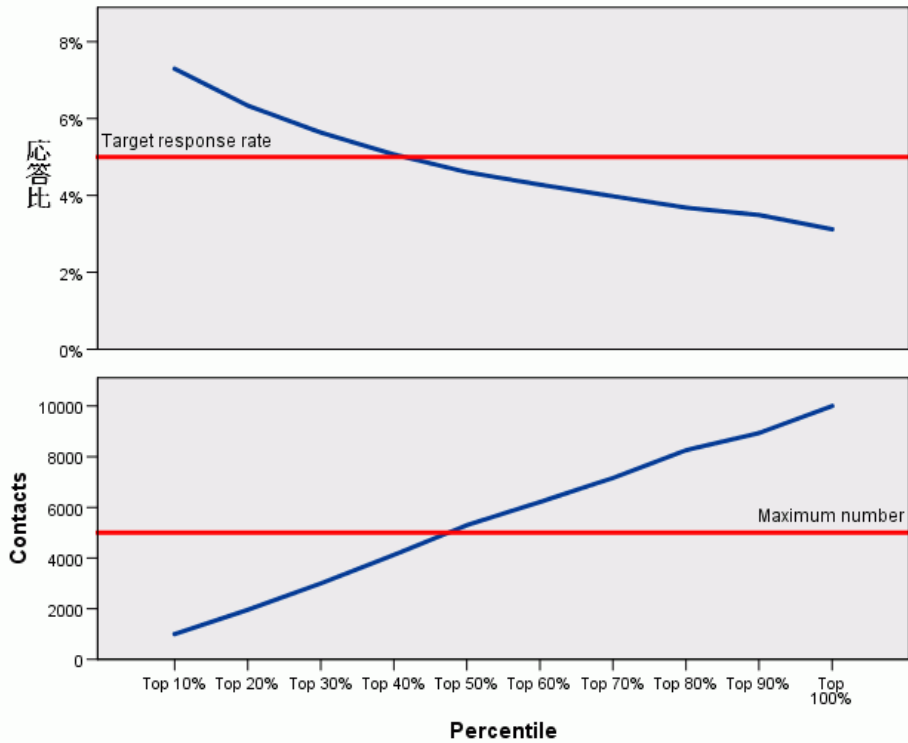
回答率表には、結果が 10 位単位の順位がの高い順 (上位 10%、上位 20% など) 出表示されます。

累積回答率は、現在およびすべての先行行の正の回答の割合の合計です。結果が回答率の高い順に表示されるため、累積回答率は現在の十分位と回答率の高いすべての十分位との回答率の合計です。10 位単位の順位が新しいデータセットにデータセットに含まれているため、特定の累積回答率に合致する郵便番号を容易に特定できます。10 位単位の順位を特定する新しいデータセットのフィールドは、ランクと呼ばれ、1=上位 10%、2=上位 20% となります。

指定された最小回答率は、5.00% です。指定された最大連絡先数は 5000 です。表のカラーコードは、最初に到達するしきい値に基づきます。同じカテゴリの 2 つのしきい値に到達しました。緑色の行の累積回答率は指定した最小回答率以上となり、累積連絡先数は指定した最大連絡先数以下となっています。赤色の行の累積回答率は指定した最小回答率を下回り、累積連絡先数は指定した最大連絡先数を超過しています。

表には、表の読み方について全般的に説明するテキストが記載されています。最小回答率または取引先の最大数を指定している場合、結果が指定のしきい値とどのように関連するかについて説明するセクションも表示されます。

図 11-6
累積回答率図表



累積回答率および 10 分位ごとの累積連絡先数の図表は、回答率表に表示される同じ情報を視覚的に表したものです。累積回答率の最小値と、累積連絡先数の最大値のしきい値には、40 ～ 50 番目の 10 分位で達します。

- 図表は、累積回答率を回答率の 10 分位の高い順に表示するため、累積回答率のラインは後続の 10 分位ごとに低下していきます。
- 連絡先数のラインは累積連絡先数を示すため、ラインは上昇していきます。

表や図表内の情報は、少なくとも 5% の回答率に達する必要があるが、キャンペーンに 5,000 件を超える連絡先が必要ない場合、上位 4 つの 10 分位の郵便番号に焦点を当てる必要があることを示します。10 分位の順

位は新しいデータセットに含まれているため、上位 40% の要件を満たす郵便番号を容易に特定できます。

図 11-7
新しいデータセット

The screenshot shows the PASW Statistics Data Editor window. The title bar reads '*無題15 [Dataset2] - PASW Statistics データ エディタ'. The menu bar includes 'ファイル(F)', '編集(E)', '表示(V)', 'データ(D)', '変換(T)', '分析(A)', 'アプリケーション グラフ(G)', 'ユーティリティ(U)', 'ウィンドウ(W)', and 'ヘルプ(H)'. The main area displays a table with the following data:

	PostalCode	Response...	Responses	Contacts	Index	Rank
49	965	3.57%	2.00	56	1.93	Top 40%
50	618	3.54%	4.00	113	3.86	Top 40%
51	603	3.53%	3.00	85	2.89	Top 40%
52	757	3.48%	4.00	115	3.86	Top 40%
53	948	3.39%	2.00	59	1.93	Top 40%
54	103	3.33%	3.00	90	2.90	Top 40%
55	608	3.33%	3.00	90	2.90	Top 40%
56	612	3.28%	4.00	122	3.87	Top 50%
57	762	3.23%	1.00	31	.97	Top 50%
58	933	3.23%	2.00	62	1.94	Top 50%
59	615	3.19%	3.00	94	2.90	Top 50%

The status bar at the bottom shows 'データ ビュー(D)' and '変数 ビュー(V)'.

注：順位は、1 ～ 10 の整数の値で記録されます。フィールドには値ラベルが定義され、1= 上位 10%、2=上位 20% などと表します。[表示] 設定に応じて、データ エディタのデータ ビューで実際の順位の値または値ラベルを確認できます。

要約

郵便番号回答率の手順では、以前のキャンペーンの結果を使用して、郵便番号の回答率を計算します。これらの比率は、今後のキャンペーンの対象の特定郵便番号を目標に設定できます。この手順では、郵便番号ごとの回答率を含む新しいデータセットを作成します。回答率の表および図表の情報、そして新しいデータセットの 10 分位の順位の情報に基づき、指定した累積回答率の最小値および/または累積連絡先数の最大値を満たす郵便番号のセットを特定できます。

購入の傾向

購入の傾向では、テスト送付または以前のキャンペーンの結果を使用して、傾向スコアを生成します。スコアは、選択したさまざまな特性に基づいて、回答する可能性の最も高い連絡先を示します。

この手法では、予測モデルの作成に 2 項ロジスティック回帰を使用します。予測モデルの作成および適応のプロセスには、次の 2 つの基本ステップがあります。

- ▶ モデルを作成し、モデル ファイルを保存します。関心のある出力（目標）が認識されているデータセットを使用してモデルを作成します。たとえば、誰がダイレクト メール キャンペーンに回答する可能性が高いかを予測するモデルを作成する場合、誰が回答し、誰が回答しなかったかについての情報をすぐに含んでいるデータセットから始める必要があります。この場合、小規模な顧客グループへのテスト メールの結果、または過去の類似したキャンペーンに対する回答者の情報となります。
- ▶ そのモデルを異なるデータセット（関心のある結果が認識されていない）に適用し、予測結果を取得します。

この例では、次の 2 つのデータ ファイルを使用しています。dmdata2.sav はモデルの作成に使用され、そのモデルは dmdata3.sav に適用されます。詳細は、A 付録 p. 104 サンプル ファイル を参照してください。

データの考慮事項

回答フィールド（関心のある目標の結果）は、1 つの値がすべての正の回答を示す、カテゴリ型でなければなりません。欠損値以外のその他の回答値は、負の回答を示すと想定します。回答フィールドが購入数や購入金額などの連続型（スケール）値を示す場合、単一正の回答がすべての 0 以外の回答値に割り当てられる新しいフィールドを作成する必要があります。詳細は、6 章 p. 43 カテゴリ回答フィールドの作成 を参照してください。

予測モデルの作成

- ▶ データ ファイル dmdata2.sav を開きます。

このファイルには、テスト メールを受け取った人々の人口統計上の特徴が含まれ、またそのメールに回答したかしなかったかについての情報も含まれます。この情報は、フィールド（変数）[回答済み]に記録されます。1 の値は、連絡先がメールに回答したことを示し、0 の値は連絡先が回答しなかったことを示します。

図 12-1
データ エディタのデータ ファイルの内容

ID	回答済み	前回	対象パッケージ	郵便番号	年齢	収入	学歴	住所	性別
06262	0	0	1	96600	67	3	5	10	1
03692	0	0	0	95510	53	3	4	9	0
01480	0	0	1	92590	56	3	2	11	1
06118	0	0	1	92670	55	2	3	8	1
07378	0	0	1	92690	56	2	1	6	1
08467	0	0	0	93410	50	1	1	8	1
09621	0	0	1	93480	67	4	2	9	1
03029	0	0	0	93490	47	1	2	10	1
02660	0	0	1	93490	58	4	5	11	1

- ▶ メニューから次の項目を選択します。
ダイレクト マーケティング > 手法を選択
- ▶ [購入の可能性が高い連絡先を選択する] を選択し、[続行] をクリックします。

図 12-2
[購入の傾向] の [フィールド] タブ



- ▶ 回答フィールドの場合、[テスト オファーに回答済み] を選択します。
- ▶ 正の回答値の場合、ドロップダウン リストから [はい] を選択します。
「はい」は実際、記録された 1 の値に関連する値ラベルであるため、1 の値がテキスト フィールドに表示されます（正の回答値に値ラベルが定義されていない場合、テキスト フィールドに値を入力できます）。
- ▶ [次を使用して傾向を予測] の場合、[年齢]、[収入カテゴリ]、[学歴]、[現住所の居住年月]、[性別]、[結婚]、[地域]、[子供] を選択します。
- ▶ [モデル情報を XML ファイルにエクスポート] を選択（オン）します。
- ▶ [参照] をクリックして、ファイルの保存場所に移動し、ファイル名を入力します。

- ▶ [購入の傾向] ダイアログで、[設定] タブをクリックします。

図 12-3
[購入の傾向: 設定] タブ

購入の傾向

フィールド 設定(N)

モデルの検証

スコア生成に使用するモデルを検証できます。モデルを検証するには、データをデータ区分に分割する必要があります。学習データ区分を使用して、モデルを学習、または作成します。学習データ区分を使用して、モデルを検証します。モデルを検証する場合、この方法ではレコードが自動的にデータ区分に割り当てられます。

モデルを検証(Y)

学習サンプルデータ区分のサイズ(%) (T):

シードを設定して結果を再現(S)

シード(E):

診断出力

全体のモデル品質(Q)

分類表(C)

最小確率(M):

再コード化された回答フィールドの名前およびラベル

この方法では、回答フィールドを、1 が正の回答を、0 が負の回答を示す新しいフィールドに自動的に再コード化します。

新しいフィールド名(N):

新しいフィールドラベル(W):

スコアを保存

この方法では、テスト送付または以前のキャンペーンの結果を使用して、スコアを生成します。スコアは、使用できるよう自動的に保存されます。このタブのその他のコントロールには、保存されたスコアに対する追加のコントロールが用意されています。

スコアの新しいフィールド名(N):

- ▶ モデル検証グループで、[モデルの検証] および [シードを設定して結果を複製] をオンにします。
- ▶ 50% のデフォルトの学習サンプル データ区分サイズおよび 2000000 のデフォルト シード値を使用します。
- ▶ [診断出力] グループで、[全体の品質] を選択 (オン) し、[分類表] を選択します。
- ▶ [最小確率] に 0.05 を入力します。一般的なルールとして、目標回答率の最小値に近い値を比率で指定する必要があります。値 0.05 は、回答率 5% を示します。

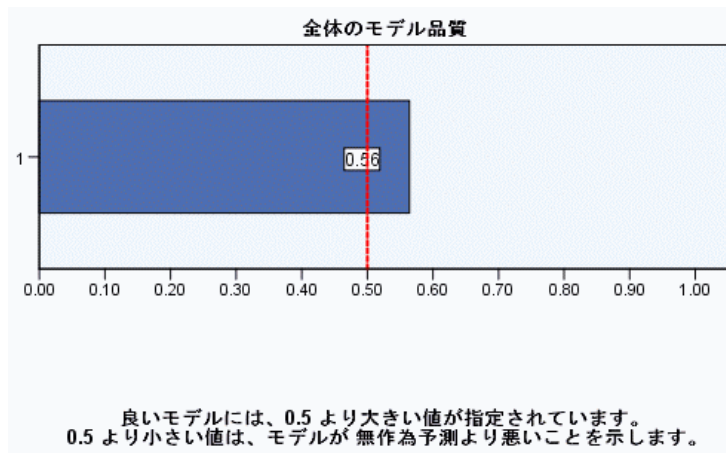
- ▶ **[実行]** をクリックして手続きを実行、モデルを生成します。

モデルの評価

購入の傾向は、モデルの評価に使用できる全体のモデル品質グラフと分類表を作成します。

全体のモデル品質グラフには、モデルの品質に関する視覚的な要約が素早く表示されます。一般的なルールとして、全体のモデル品質は 0.5 より高い必要があります。

図 12-4
全体のモデル品質グラフ



モデルがスコアリングに適していることを確認するには、分類表を検証する必要があります。

図 12-5
分類表

		分類テーブル					
		予測					
		学習サンプル =			検定サンプル		
		再コード化された回答(1=Yes, 0=No)		正解の割合	再コード化された回答(1=Yes, 0=No)		正解の割合
いいえ	はい	いいえ	はい				
再コード化された回答 (1=Yes, 0=No)	いいえ	651	249	72.33	653	267	70.98
	はい	19	20	51.28	36	22	37.93
全体のパーセント		2.84	7.43	71.46	5.22	7.61	69.02

分類表は、目標フィールドの予測値を、目標フィールドの実際の値と比較します。全体の精度比はモデルがどれだけ適切に機能しているかを示しますが、目的が肯定的な回答の確率が指定した最小比率以上となる連絡先のグループを示すモデルを作成することである場合、正しく予測された肯定的な回答の割合により関心がある場合があります。

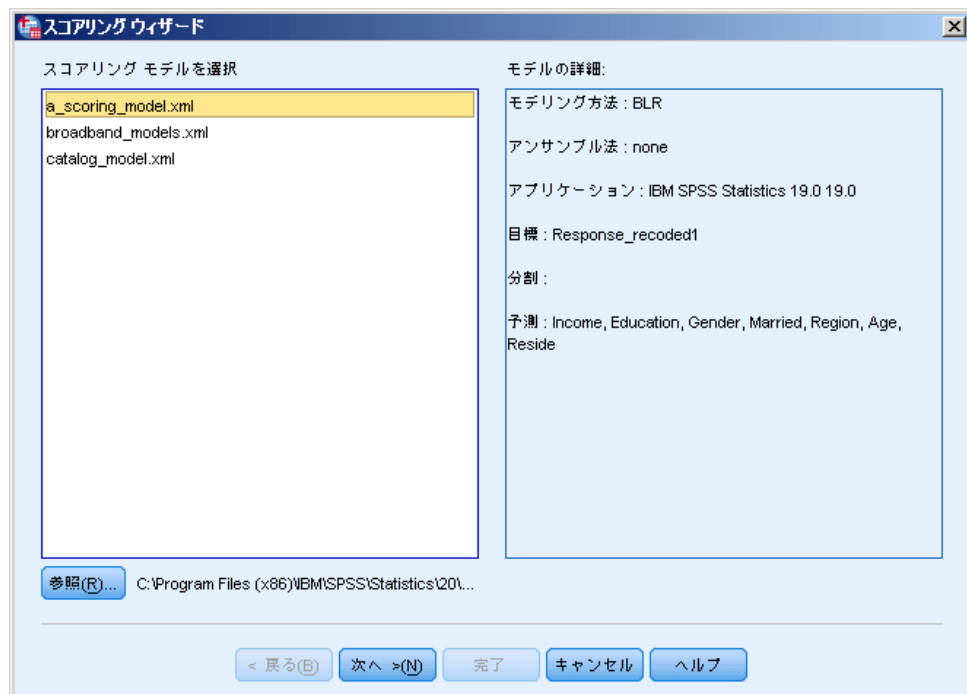
この例では、分類表は**学習サンプル**と**検定サンプル**に分けられます。学習サンプルを使用して、モデルを作成します。その後、モデルを検定サンプルに適用して、モデルがどれだけ適切に機能しているかを確認します。

指定の最小回答率は 0.05 ~ 5% です。分類表では、学習サンプルの肯定的な回答率の正分類比は 7.43%、検定サンプルでは 7.61% であることを示します。検定サンプルの回答率は 5% を超えているため、このモデルは回答率が 5% を超える連絡先のグループをすばやく作成できます。

モデルの適用

- ▶ データ ファイル dmdata3.sav を開きます。このデータ ファイルには、テスト メールに含まれていなかったすべての連絡先に関する人口統計とその他の情報が含まれています。詳細は、[A 付録 p.104 サンプル ファイル](#) を参照してください。
- ▶ スコアリング ウィザードを開きます。スコアリング ウィザードを開くには、メニューから次の項目を選択します。
ユーティリティ > スコアリング ウィザード

図 12-6
スコアリング ウィザード、スコアリング モデルの選択



- ▶ [参照] をクリックし、モデル XML ファイルを保存した場所に移動し、[参照] ダイアログで [選択] をクリックします。

拡張子が .xml または .zip のすべてのファイルはスコアリング ウィザードに表示されます。選択したファイルが有効なモデル ファイルとして認識されると、モデルの説明が表示されます。

- ▶ 作成したモデル XML ファイルを選択し、[次へ] をクリックします。

図 12-7
スコアリング ウィザード:モデル フィールドの一致

The screenshot shows the 'Scoring Wizard' dialog box. It displays the model name 'propensity_scores' and the model type '二項ロジスティック回帰'. Below this, it states 'モデル フィールドをデータセットに一致させる'. A table maps data set fields to model fields, roles, measurement methods, and types. At the bottom, there are radio buttons for '値の置き換えを使用' (selected) and 'システム欠損値を使用'. Navigation buttons include '< 戻る(B)', '次へ >(N)', '完了', 'キャンセル', and 'ヘルプ'.

データセット フィールド	モデル フィールド	役割	測定方法	タイプ(T)
収入カテゴリ	Income	予測変数	名義	数値(N)
学歴	Education	予測変数	名義	数値(N)
性別	Gender	予測変数	名義	数値(N)
結婚	Married	予測変数	名義	数値(N)
地域	Region	予測変数	名義	数値(N)
年齢	Age	予測変数	続行	数値(N)
住所	Reside	予測変数	続行	数値(N)

アクティブ データセットのスコアリングを行うには、データセットにはモデル内のすべての予測値に対応するフィールド (変数) が必要です。モデルに分割フィールドも含まれる場合、データセットにはモデルのすべての分割フィールドに対応するフィールドも必要です。

- デフォルトでは、モデルのフィールドと名前およびタイプが同じアクティブ データセットのフィールドは、自動的に一致します。

- ドロップダウン リストを使用して、データセットのフィールドをモデル フィールドに一致させます。各フィールドのデータ型は、モデルおよびデータセットで同じでなければなりません。
- モデルのすべての予測値（および、ある場合は分割ファイル）がアクティブ データセットのフィールドと一致しない場合、ウィザードを続行することもアクティブ データセットをスコアリングすることもできません。

アクティブ データセットには、[収入] という名前のフィールドが含まれます。モデル フィールド [収入] に該当する [データセット フィールド] 列のセルは、最初は空白です。モデル フィールドに等しいアクティブ データセットのフィールドを選択する必要があります。

- ▶ [収入] モデル フィールドの行内のセルが空白である [データセット フィールド] 列のドロップダウン リストから、[IncomeCategory] を選択します。

注：フィールド名およびデータ型のほか、スコアリングされる実際のデータ値は、モデルの構築に使用されるデータセットのデータ値と同じ方法で記録されます。たとえば、収入が 4 つのカテゴリに分けられた [収入] フィールドでモデルが作成され、アクティブ データセットの IncomeCategory が収入を 6 つのカテゴリまたは 4 つのカテゴリに分ける場合、これらのフィールドはお互いに一致せず、スコアは信頼できません。

[次へ] をクリックして、スコアリング ウィザードの次のステップに進みます。

図 12-8
スコアリング ウィザード:スコアリング関数の選択



スコア関数は、選択したモデルに使用できるスコアの種類の種類です。使用できるスコア関数は、モデルによって異なります。この例で使用されている 2 項ロジスティックモデルの場合、使用できる関数は、予測値、予測値の確率、選択した値の確率、および確信度です。

この例では、メールに肯定的な回答の予測確率に関心があります。そのため、選択した値の確率が必要となります。

- ▶ [選択したカテゴリの確率] をオンにします。
- ▶ [値] 列で、ドロップダウン リストから [1] を選択します。モデルの作成に使用されるデータ ファイルの目標値に基づき、目標の正の値のリストがモデルで定義されます。
- ▶ 他のすべてのスコアリング関数をオフにします。
- ▶ オプションで、アクティブ データセットのスコア値を含む新しいフィールドにより具体的な名前を付けることができます。たとえば、Probability_of_responding のようになります。
- ▶ [完了] をクリックして、モデルをデータセットに適用します。

肯定的な回答の確率を含む新しいフィールドは、データセットの終わりに追加されます。

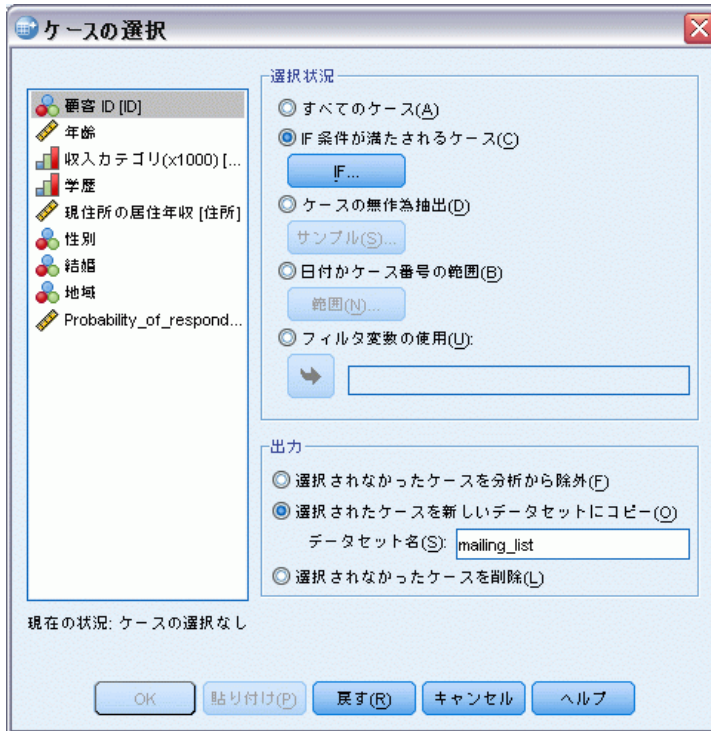
図 12-9
新しい確率フィールドを含むデータセット

住所	性別	結婚	地域	Probability_of_responding
7	1	0	4	.04
9	0	0	4	.03
12	0	0	4	.03
8	0	0	4	.04
13	0	0	4	.07
10	0	0	4	.04
12	0	0	4	.03
15	0	0	4	.05
10	0	0	4	.05
14	0	0	4	.02
5	0	0	4	.12

特定のレベル以上の肯定的な回答をする可能性のある連絡先のサブセットを選択するために、このフィールドを使用できるようになります。たとえば、5%以上の肯定的な回答をする可能性が高くなるケースのサブセットを含む新しいデータセットを作成することができます。

- ▶ 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

図 12-10
[ケースの選択] ダイアログ



- ▶ [ケースの選択] ダイアログ ボックスで、[IF 条件が満たされるケース] を選択して [IF] をクリックします。

図 12-11
[ケースの選択: IF] ダイアログ



- ▶ [ケースの選択: IF] ダイアログに次の式を入力します。

Probability_of_responding >=.05

注： 確率値を含むフィールドに異なる名前を使用した場合、Probability_of_responding の代わりにその名前を入力してください。デフォルト名は、SelectedProbability です。

- ▶ [続行] をクリックします。
- ▶ [ケースの選択] ダイアログで、[選択したケースを新しいデータセットにコピーする] をクリックして、新しいデータセットの名前を入力します。データセット名は、フィールド（変数）命名規則に従う必要があります。
- ▶ [OK] をクリックすると、選択した連絡先を含むデータセットを作成します。

新しいデータセットには、肯定的な回答の予測確率が 5% 以上の連絡先だけが含まれます。

図 12-12
選択した連絡先を含む新しいデータセット

住所	性別	結婚	地域	Probability_of_responding
13	0	0	4	.07
15	0	0	4	.05
10	0	0	4	.05
5	0	0	4	.12
7	0	0	4	.08
10	0	0	4	.10
15	1	0	4	.05
11	1	0	4	.08
9	1	0	4	.08
9	1	0	4	.05

要約

購入の傾向では、テスト送付または以前のキャンペーンの結果を使用して、傾向スコアを生成します。スコアは、選択したさまざまな特性に基づいて、回答する可能性の最も高い連絡先を示します。この方法で、傾向スコアを取得するためにデータセットに適用できる予測モデルを作成します。

対照パッケージ検定

この方法では、マーケティング キャンペーンを比較して、さまざまなパッケージまたはオファーの有効性に重要な相違点があるかどうかを確認します。キャンペーンの有効性は回答によって測定されます。

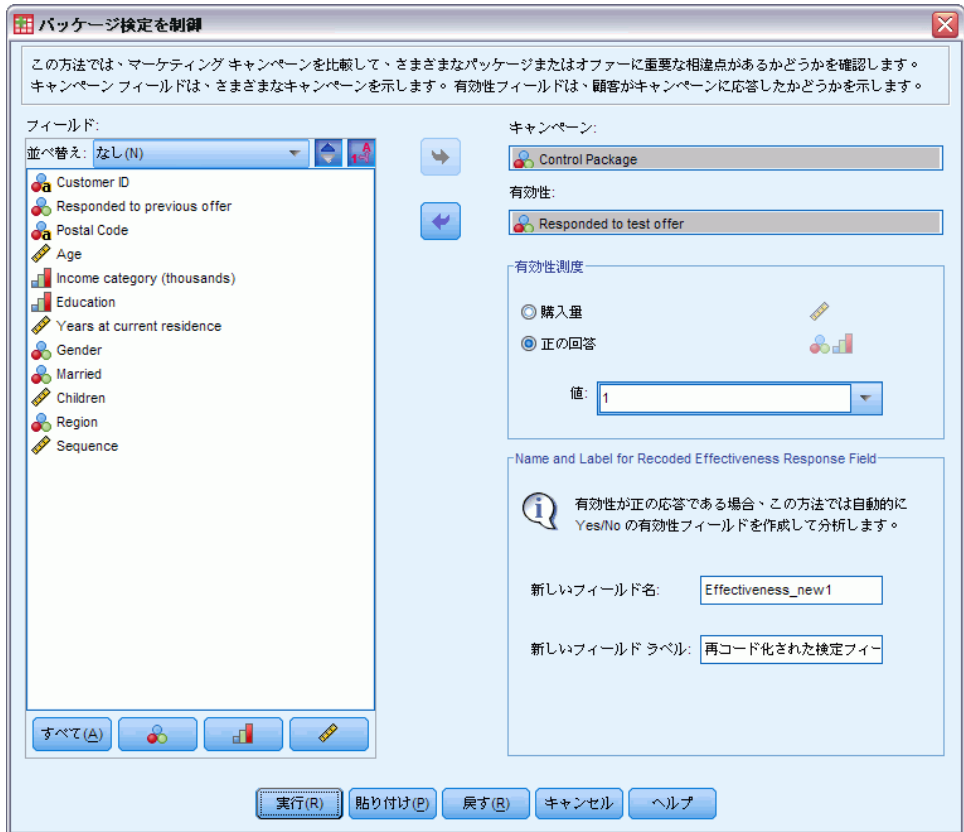
たとえば、企業のダイレクト マーケティング部門が、新しいパッケージ デザインの方が既存のパッケージより多くの正の回答を引き出すかどうかを確認したい場合があります。新しいパッケージ デザインが非常に多くの正の回答率を生み出すかどうかを判断するテスト送付を送信します。テスト送付は、既存のパッケージを受信する対照グループと、新しいパッケージ デザインを受信する検定グループで構成されています。2 つのグループの結果を比較して、大きな違いがあるかどうかを確認します。

この情報は、dmdata.sav に収集されています。 [詳細は、A 付録 p. 104 サンプル ファイル を参照してください。](#)

分析の実行

- ▶ 対照パッケージ検定を実行するには、メニューから次の項目を選択します。
Direct Marketing > 手法を選択
- ▶ [キャンペーンの有効性の比較 (対照パッケージ検定)] を選択し、[続行] をクリックします。

図 13-1
対照パッケージ検定の [フィールド] タブ



- ▶ キャンペーン フィールドについて、[対照パッケージ] を選択します。
- ▶ 有効性回答フィールドについて、[テスト オファーに回答済み] を選択します。
- ▶ [応答] を選択します。
- ▶ 正の回答値の場合、ドロップダウン リストから [はい] を選択します。
「はい」は実際、記録された 1 の値に関連する値ラベルであるため、1 の値がテキスト フィールドに表示されます（正の回答値に値ラベルが定義されていない場合、テキスト フィールドに値を入力できます）。

新しいフィールドが自動的に作成されます。このフィールドでは、1 が正の回答を、0 が負の回答を示し、分析が新しいフィールドで実行されます。デフォルトの名前とラベルを上書きして、独自の名前とラベルを指定できます。この例では、すでに指定されているフィールド名を使用します。

- ▶ [実行] をクリックして手続きを実行します。

出力

図 13-2
対照パッケージ検定の出力

		対象パッケージ			
		対象		テスト	
		度数	列の N %	度数	列の N %
有効性 (1=Yes 0=No)	0	879	96.6%	984	97.7%
	1	31	3.4%	23	2.3%

対象とテストの間に統計的な有意差はありません。

この手順の出力には、キャンペーン フィールドで定義された各グループの正の回答および負の回答の度数およびパーセントを表示する表、グループ間のグループ回答率が大きく異なるかどうかを示す表があります。

- 有効性は、1 が正の回答を、0 が負の回答を示す再コード化したバージョンの回答フィールドです。
- 対照パッケージの正の回答率は 3.8%、検定パッケージの正の回答率は 6.2% です。

テーブル下の簡単なテキストの説明は、グループ間の差が大幅に異なることを示します。つまり、検定パッケージの高い回答率は、無作為確率の結果ではありません。このテキスト表には、分析に含まれるグループの各ペアの比較が記載されています。この例ではグループが 2 つしかないため、比較は 1 つだけです。5 つを超えるグループがある場合、テキストによる説明表は [列部分の比較] テーブルになります。

要約

対照パッケージ検定では、マーケティング キャンペーンを比較して、さまざまなパッケージまたはオファーの有効性に重要な相違点があるかどうかを確認します。この例では、検定パッケージの正の回答率 6.2% は、対照パッケージの正の回答率 3.8% を大きく上回ります。これは、古いパッケージ デザインではなく新しいパッケージ デザインを使用することを推奨していますが、新しいパッケージ デザインに関連する追加コストなど、その他の要素について考慮が必要な場合があります。

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の Samples サブディレクトリにあります。[サンプル] サブディレクトリ内に次の各言語の別のフォルダがあります。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、簡体字中国語、スペイン語、そして繁体中国語です。

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。サンプル ファイルがある言語で使用できない場合、その言語のフォルダには、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

以下は、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です。

- **accidents.sav**。与えられた地域での自動車事故の危険因子を年齢および性別ごとに調べている保険会社に関する架空のデータ ファイルです。各ケースが、年齢カテゴリと性別のクロス分類に対応します。
- **adl.sav**。脳卒中患者に提案される治療の効果を特定するための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。一方のグループは標準的な理学療法を受け、もう一方のグループは感情面の治療も追加で受けました。治療の 3 か月後に、各患者が日常生活の一般的な行動をどの程度とることができるかを、順序変数として得点付けしました。
- **advert.sav**。広告費とその売上成果の関係を調べるための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。この小売業者は、そのために、過去の売上と、それに関する広告費のデータを収集しました。
- **aflatoxin.sav**。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に関して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 つそれぞれの収穫物から 16 のサンプルを受け取って、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**。拒食行動または過食行動の標準的な症状の特定を目指して、調査員 が、摂食障害を持つ大人 55 人の調査を行いました。各患者が 4 年間で 4 回診察を受けたので、観測値は合計で 220 になりました。観測値ごとに、16 種類の症状に関して患者の得点が記録

されました。患者 71 (2 回目)、患者 76 (2 回目)、患者 47 (3 回目) の症状の得点が見つからなかったので、残っている 217 回分の観測値が有効です。

- **bankloan.sav**。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、過去の顧客および見込み客 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 ケースは、以前に貸付を行った顧客です。残りの 150 ケースは見込み顧客で、これらの顧客に関して銀行は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**。過去の顧客 5,000 人に関する財務情報と人口統計情報を含む架空のデータ ファイルです。
- **behavior.sav**。52 人の学生に 15 の状況と 15 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た研究があります。個人間の平均を取ったため、値は非類似度としてみなされます。
- **behavior_ini.sav**。このデータ ファイルには、behavior.sav の 2 次元の解の初期配置が含まれています。
- **brakes.sav**。高性能自動車のディスク ブレーキを生産している工場での品質管理に関する架空のデータ ファイルです。このデータ ファイルには、8 台の機械で生産した 16 個のディスクの直径測定値が含まれています。ブレーキの目標の直径は 322 ミリメートルです。
- **breakfast.sav**。21 人の Wharton School MBA の学生およびその配偶者に、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た研究があります。調査対象者の嗜好は、「すべて」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に基づいて記録されました。
- **breakfast-overall.sav**。このデータ ファイルには、最初のシナリオ (「すべて」) のみの朝食の好みが含まれています。
- **broadband_1.sav**。全国規模のブロードバンド サービスの地域ごとの契約者数を含む架空のデータ ファイルです。このデータ ファイルには、85 地域の月々の契約者数が 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、データが 3 か月分追加されています。
- **car_insurance_claims.sav**。他の場所 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、製造年の線型結合と関連付けることにより、平均請求数はガンマ分布としてモデリングできます。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、定価、仕様が含まれています。定価と仕様はそれぞれ、edmunds.com と製造元のサイトから入手しました。

- **car_sales_upprepared.sav**。変換したバージョンのフィールドを含まない car_sales.sav の修正したバージョンです。
- **carpet.sav**。一般的な例 としては、新しいカーペット専用洗剤を市販することに関心のある企業が消費者の嗜好に関する 5 種類の因子（パッケージのデザイン、ブランド名、価格、サービスシール、料金の払い戻し）の影響について調べたい場合があります。パッケージのデザインには、3 つの因子レベルがあります。それぞれ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、および Bissell）、3 つの価格水準があり、最後の 2 つの因子のそれぞれに対しては 2 つのレベル（「なし」または「あり」）があります。10 人の消費者が、これらの因子により定義された 22 個のプロファイルに順位を付けます。変数「嗜好」には、各プロファイルの平均順位の序列が含まれています。順位が低いほど、嗜好度は高くなります。この変数には、各プロファイルの嗜好測定値がすべて反映されます。
- **carpet_prefs.sav**。このデータ ファイルは carpet.sav 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。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 で定義されている、関連するプロファイルの ID が含まれています。
- **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、架空の月間売上高が含まれています。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、季節性の分解手続き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**。解約率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。解約の傾向スコアは、0 ~ 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の変更を考えている場合があります。
- **ceramics.sav**。新しい上質の合金に標準的な合金より高い耐熱性があるかどうかを特定するための、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 つの合金の別々のテストを表し、軸受けの耐熱温度が記録されます。
- **cereal.sav**。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。各ケースが別々の回答者を表します。
- **clothing_defects.sav**。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、調査員が衣料品のサンプルを取り出し、不良品の数を数えます。

- **coffee.sav.** このデータ ファイルは、6 つのアイスコーヒー ブランド について受けた印象に関連しています。回答者は、アイス コーヒーに対する 23 の各印象属性に対して、その属性が言い表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、および FF で表しています。
- **contacts.sav.** 企業のコンピュータ営業グループの担当者リストに関する架空のデータ ファイルです。各担当者は、所属する会社の部門および会社のランクによって分類されています。また、最新の販売金額、最後の販売以降の経過時間、担当者の会社の規模も記録されています。
- **creditpromo.sav.** 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ための、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。そのうち半分には、今後 3 か月間の買い物に関して利率を下げることをプロモーションする広告を送付しました。残り半分には、通常どおりの定期的な広告を送付しました。
- **customer_dbase.sav.** 自社のデータ ウェアハウスにある情報を使用して、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、顧客の反応が記録されています。
- **customer_information.sav.** 名前や住所など、顧客の連絡先情報を含む架空のデータ ファイルです。
- **customer_subset.sav.** customer_dbase.sav の 80 件のケースのサブセット。
- **debate.sav.** 政治討論の出席者に対して行った調査の、討論の前後それぞれの回答に関する架空のデータ ファイルです。各ケースが別々の回答者に対応します。
- **debate_aggregate.sav.** debate.sav 内の回答を集計する、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.** 月々の特典を送付することを目的とした、購入顧客のデータベースに関する架空のデータ ファイルです。顧客が特典に反応したかどうか、さまざまな人口統計情報と共に記録されています。
- **demo_cs_1.sav.** 調査情報のデータベースをコンパイルするための、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。各ケースが別々の都市に対応し、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.** 調査情報のデータベースをコンパイルするための、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別々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 つの段階からの抽出情報も含まれています。

- **demo_cs.sav**。コンプレックス サンプル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。各ケースが別々の世帯単位に対応し、さまざまな人口統計情報および抽出情報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および購入情報を含む架空のデータです。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を含み、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を含みます。
- **dietstudy.sav**。この架空のデータ ファイルには、“Stillman diet” の研究結果が含まれています。各ケースが別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。
- **dvdplayer.sav**。新しい DVD プレーヤーの開発に関する架空のデータ ファイルです。プロトタイプを使用して、マーケティング チームはフォーカス グループ データを収集しました。各ケースが別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。
- **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 にある “German credit” データセットから取ったものです。
- **grocery_1month.sav**。この架空のデータ ファイルは、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、各ケースが別々の顧客に対応するようにしたものです。その結果、週ごとに変っていた変数の一部が表示されなくなり、買物の総額が、調査を行った 4 週間の買物額の合計になっています。
- **grocery_coupons.sav**。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を含む架空のデータ ファイルです。各顧客を 4 週間に渡って追跡し、各ケースが別々の顧客の週に対応しています。その週に食料品に費やした金額も含め、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 は、予想される社会グループを示す表を作成しました。Guttman は、この表の一部を使用しました。この表では、社会相互作用、グループへの帰属感、メンバとの物理的な近接性、関係の形式化などを表す 5 個の変数が、理論上の 7 つ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（例、フットボールの試合の観戦者）、視聴者（例、映画館または授業の参加者）、公衆（例、新聞やテレビの視聴者）、暴徒（観衆に似ているが、より強い相互作用がある）、第一次集団（親密な関係）、第二次集団（自発的な集団）、および近代コミュニティ（物理的により密接した近接性と特化されたサービスの必要性によるゆるい同盟関係）があります。

- **health_funding.sav**。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）、医療サービス機関への訪問率（人口 10,000 人あたりの人数）のデータを含む、架空のデータ ファイルです。各ケースが別々の都市を表します。
- **hivassay.sav**。HIV 感染を発見する迅速な分析方法を開発するための、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、半分は感染していませんでした。
- **hourlywagedata.sav**。管理職から現場担当まで、またさまざまな経験レベル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**。不正請求の恐れがある、疑いを区別するためにモデルを作成する必要がある保険会社の仮説データ ファイルです。各ケースがそれぞれの請求を表します。
- **insure.sav**。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、請求を行った契約と行わなかった契約のペアを表します。
- **judges.sav**。訓練を受けた審判（および 1 人のファン）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。各行が別々の演技を表し、審判たちは同じ演技を見ました。
- **kinship_dat.sav**。Rosenberg と Kim は、15 種類の親族関係用語（祖父、祖母、父、母、叔父、叔母、兄弟、姉妹、いとこ、息子、娘、甥、姪、孫息子、孫娘）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、1 回目と違う条件に基づいて、2 回目の並べ替えをするように頼みました。このようにして、合計で 6 つの「ソース」が取得できました。各ソースは、15 × 15 の近接行列に対応します。この近接行列のセルの数は、ソースの人数から、ソース内でオブジェクトを分割した回数を引いたものです。
- **kinship_ini.sav**。このデータ ファイルには、kinship_dat.sav の 3 次元の解の初期配置が含まれています。
- **kinship_var.sav**。このデータ ファイルには、kinship_dat.sav の解の次元の解釈に使用できる独立変数である性別、世代、および(ation), and 親等が含まれています。特に、解の空間をこれらの変数の線型結合に制限するために使用できます。
- **marketvalues.sav**。1999 ~ 2000 年の間の、イリノイ州アルゴンキンの新興住宅地での住宅売上に関するデータ ファイルです。これらの売上は、公開レコードの問題となります。

- **nhis2000_subset.sav**。National Health Interview Survey (NHIS) は、米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について対面式で調査が行われます。各世帯のメンバーに関して、人口統計情報、健康に関する行動および状態の観測値が得られます。このデータ ファイルには、2000 年の調査から得ら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。一般使用データおよびドキュメント。ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年にアクセス。
- **ozone.sav**。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 個の気象変数に対する 330 個の観測値が含まれています。それまでの研究者、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。特に興味深いことは、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。各ケースが別々の患者に対応し、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。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **poll_cs.sav**。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。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およびサンプル重み付けが記録されています。ただし、抽出計画では確率比例 (PPS) 法を使用するため、結合選択確率を含むファイル (poll_jointprob.sav) もあります。サンプル抽出後、有権者の人口統計および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**。限られたリソースで資産価値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ます。データ ファイル内の各ケースでは、資産が存在する町、最後に訪問した評価担当者、その評価からの経過時間、当時行われた評価、および資産の売却価値が記録されています。

- **property_assess_cs.sav.** 限られたリソースで資産価値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ます。データ ファイル内の各ケースでは、資産が存在する郡、町、および区域、最後の評価からの経過時間、および当時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 の資産のサンプルが含まれています。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およびサンプル重み付けが記録されています。サンプル抽出後、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.** 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、人口統計情報、最初の犯罪の詳細、初犯から 2 年以内の場合は 2 回目の逮捕までの期間が記録されています。
- **recidivism_cs_sample.sav.** 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 月に最初の逮捕から釈放された元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、2006 年 7 月までの 2 回目の逮捕のデータが記録されています。犯罪者は recidivism_cs.csplan で指定された抽出計画に従って抽出された部門から選択されます。調査では確率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル (recidivism_cs_jointprob.sav) も用意されています。
- **rfm_transactions.sav.** 購入日、購入品目、各取引のマネタリー量など、購買取引データを含む架空のデータ ファイルです。
- **salesperformance.sav.** 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 人の従業員が 3 つのグループに分けられ、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 は技術トレーニングを、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 コースの最後に各従業員がテストを受け、得点が記録されました。データ ファイルの各ケースは別々の訓練生を表し、割り当てられたグループと、テストの得点が記録されています。
- **satisf.sav.** 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 人の顧客を調査し、各ケースは 1 人の顧客からの回答を表します。
- **screws.sav.** このデータ ファイルには、ねじ、ボルト、ナット、鋸 (びょう) の特性に関する情報が含まれています。
- **shampoo_ph.sav.** 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 つの異なる製品が測定され、pH が記録されます。目標範囲は 4.5 ~ 5.5 です。

- **ships.sav.** 他の場所 で表示および分析される、波による貨物船への損害に関するデータセットです。件数は、船舶の種類、建造期間、およびサービス期間によって、ポワゾン率で発生するものとしてモデリングできます。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マ数の集計によって、危険にさらされる確率の値が得られます。
- **site.sav.** 業務拡大に向けて新たな用地を選択するための、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。2 人のコンサルタントを雇って、用地を別々に評価させました。広範囲のレポートに加えて、各用地を「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 から抜粋したものであり、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場合は、まず人口の傾向を反映させてデータを重み付けする必要があります。
- **stocks.sav** このデータ ファイルには、1 年あたりの在庫価格、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、[データの準備] オプションの手続き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が含まれており、データ入力にいくつかエラーがあります。
- **stroke_survival.** この架空のデータ ファイルは、虚血性脳卒中で数回の困難に直面した後リハビリ プログラムを終えた患者の生存時間に関するものです。脳卒中後、心筋梗塞の発生、虚血性脳卒中、または出血性脳卒中が注意され、イベントの時間が記録されます。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のみが含まれるため、サンプルは左側が切り捨てられます。
- **stroke_valid.sav.** この架空のデータ ファイルには、[データの検証] 手続きを使用して確認した後の、医療データベースの状態が含まれています。異常である可能性のあるケースが含まれています。
- **survey_sample.sav.** このデータ ファイルには、人口統計データおよびさまざまな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、いくつかのデータ値が変更され、追加の架空変数がデモの目的で追加されています。
- **telco.sav.** 顧客ベースにおける解約率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。各ケースが別々の顧客に対応し、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 ファイルに似ていますが、「期間」および対数変換された顧客支出の属性が削除され、標準化された対数変換顧客支出の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に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、新しいメニューを追加しようというファースト フード チェーンの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、新メニューはいくつかのランダムに選択した市場にある場所で紹介されます。場所ごとに別々のプロモーションを使用し、最初の 4 週間の新メニューの週間売上高が記録されます。各ケースが場所と週に対応します。
- **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、各ケースが別々の場所に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、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、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_credit.sav.** これは、人口統計および銀行ローン履歴のデータを含む架空のデータ ファイルです。
- **tree_missing_data.sav.** これは、人口統計および銀行ローン履歴のデータと、多数の欠損値を含む架空のデータ ファイルです。
- **tree_score_car.sav.** これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_textdata.sav.** 尺度および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、変数を 2 つだけ含む単純なデータ ファイルです。
- **tv-survey.sav.** 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、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。各行は別々の回答者を表し、各列は別々の条件を表します。
- **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ように計画された調査の情報の一部が含まれています。これは区間調査の良い例であり、他の場所 で表示および分析されています。
- **ulcer_recurrence_recoded.sav.** このファイルでは、ulcer_recurrence.sav の情報が、単に調査終了時のイベント確率ではなく調査の区間ごとのイベント確率をモデリングできるように再編成されています。これは他の場所 で表示および分析されています。
- **verd1985.sav.** このデータ ファイルは調査に関連しています。8 つの変数に対する 15 人の被験者の回答を記録しました。対象となる変数が 3 つのグループに分類されます。グループ 1 には「年齢」と「婚姻」、

グループ 2 には「ペット」と「新聞」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれます。「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。また、その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**。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 (ISP) の取り組みに関する架空のデータ ファイルです。この ISP は、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) パーセンテージを、発見の瞬間から脅威が阻止されるまで追跡しました。
- **wheeze_steubenville.sav**。これは、子供 に対する大気汚染の健康上の影響の長期調査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 スビューベンビル の 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記録が含まれています。
- **workprog.sav**。体の不自由な人をより良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。プログラムの参加者候補のサンプルが追跡されました。その中には、ランダムに選ばれてプログラムに登録された人と、そうでない人がいました。各ケースが別々のプログラム参加者を表します。
- **worldsales.sav** このデータ ファイルには、大陸および製品ごとの販売収益が含まれています。

注意事項

この情報は、世界各国で提供される製品およびサービス向けに作成されています。

IBMはこのドキュメントで説明する製品、サービス、機能は他の国では提供していない場合があります。現在お住まいの地域で利用可能な製品、サービス、および、情報については、お近くの IBM の担当者にお問い合わせください。IBM 製品、プログラム、またはサービスに対する参照は、IBM 製品、プログラム、またはサービスのみが使用することができることを説明したり意味するものではありません。IBM の知的所有権を侵害しない機能的に同等の製品、プログラム、またはサービスを代わりに使用することができます。ただし、IBM 以外の製品、プログラム、またはサービスの動作を評価および確認するのはユーザーの責任によるものです。

IBMは、本ドキュメントに記載されている内容に関し、特許または特許出願中の可能性があります。本ドキュメントの提供によって、これらの特許に関するいかなる権利も使用者に付与するものではありません。ライセンスのお問い合わせは、書面にて、下記住所に送ることができます。

IBM Director of Licensing, IBM Corporation, North Castle Drive,
Armonk, NY 10504-1785, U. S. A.

2 バイト文字セット (DBCS) 情報についてのライセンスに関するお問い合わせは、お住まいの国の IBM Intellectual Property Department に連絡するか、書面にて下記宛先にお送りください。

神奈川県大和市下鶴間1623番14号 日本アイ・ビー・エム株式会社 法務・知的財産 知的財産権ライセンス渉外

以下の条項は、イギリスまたはこのような条項が法律に反する他の国では適用されません。 International Business Machines は、明示的または黙示的に関わらず、第三者の権利の侵害しない、商品性または特定の目的に対する適合性の暗黙の保証を含むがこれに限定されない、いかなる保証なく、本出版物を「そのまま」提供します一部の州では、特定の取引の明示的または暗示的な保証の免責を許可していないため、この文が適用されない場合があります。

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。情報については変更が定期的に行われます。これらの変更は本書の新版に追加されます。IBM は、本書に記載されている製品およびプログラムについて、事前の告知なくいつでも改善および変更を行う場合があります。

IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、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。これらの Web サイトの資料はこの IBM 製品の資料に含まれるものではなく、これらの Web サイトの使用はお客様の責任によるものとします。

IBM はお客様に対する一切の義務を負うことなく、自ら適切と考える方法で、情報を使用また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取得者が (i) 別途作成されたプログラムと他のプログラム（本プログラムを含む）との間の情報交換および (ii) 交換された情報の相互利用を目的とした本プログラムに関する情報の所有を希望する場合、下記住所にお問い合わせください。

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

上記のような情報は、該当する条項および条件に従い、有料で利用できるものとします。

本ドキュメントに記載されている許可されたプログラムおよびそのプログラムに使用できるすべてのライセンス認証された資料は、IBM Customer Agreement、IBM International Program License Agreement、および当社とかわした同等の契約の条件に基づき、IBM によって提供されます。

IBM 以外の製品に関する情報は、それらの製品の供給業者、公開済みの発表、または公開で使用できるソースから取得しています。IBM は、それらの製品のテストは行っておらず、IBM 以外の製品に関連する性能、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給業者に通知する必要があります。

この情報には、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。できる限り詳細に説明するため、例には、個人、企業、ブランド、製品などの名前が使用されています。これらの名称はすべて架空のものであり、実際の企業で使用される名称および住所とは一切関係ありません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーのイラストが表示されない場合があります。

商標

IBM、IBM ロゴ、および [ibm.com](http://www.ibm.com)、SPSS は、世界の多くの国で登録された IBM Corporation の商標です。IBM の商標の現在のリストは、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel、Intel のロゴ、Intel Inside、Intel Inside のロゴ、Intel Centrino、Intel Centrino のロゴ、Celeron、Intel Xeon、Intel SpeedStep、Itanium、および Pentium は、米国およびその他の国の Intel Corporation または関連会社の商標または登録商標です。

Java およびすべての Java ベースの商標およびロゴは、米国およびその他の国の Sun Microsystems, Inc. の商標です。

Linux は、米国およびその他の国における Linus Torvalds の登録商標です。

Microsoft、Windows、Windows NT、および Windows のロゴは、米国およびその他の国における Microsoft 社の商標です。

UNIX は、米国およびその他の国における The Open Group の登録商標です。

この製品は、WinWrap Basic (Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>) を使用します。

その他の製品名およびサービス名等は、IBM または他の会社の商標です。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。



索引

郵便番号回答率, 28, 81
商標, 116

cluster, 16
[RFM], 2, 10, 12, 14, 49
 取引データ, 3, 49
 顧客データ, 5
 ビン, 7

クラスタ分析, 16, 57

サンプル ファイル
 位置, 104

法律に関する注意事項, 115

購入の傾向, 36, 88

対照パッケージ検定, 44, 101

見込みのプロファイル, 21, 74

ロジスティック回帰, 36, 88