

# **IBM SPSS Direct Marketing 24**

**IBM**

**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、27 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM SPSS Statistics バージョン 24 リリース 0 モディフィケーション 0 および新しい版で明記されない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

**原典：** IBM SPSS Direct Marketing 24

**発行：** 日本アイ・ビー・エム株式会社

**担当：** トランスレーション・サービス・センター

---

## 目次

<b>第 1 章 Direct Marketing</b> . . . . .	<b>1</b>	<b>第 5 章 郵便番号応答率</b> . . . . .	<b>17</b>
<b>第 2 章 RFM 分析</b> . . . . .	<b>3</b>	設定 . . . . .	18
トランザクション・データからの RFM スコア . . . . .	3	カテゴリ応答フィールドの作成 . . . . .	19
顧客データからの RFM スコア . . . . .	4	<b>第 6 章 購入の傾向</b> . . . . .	<b>21</b>
RFM ビン . . . . .	5	設定 . . . . .	23
トランザクション・データからの RFM スコアの保存 . . . . .	6	カテゴリ応答フィールドの作成 . . . . .	24
顧客データからの RFM スコアの保存 . . . . .	7	<b>第 7 章 制御パッケージ検定</b> . . . . .	<b>25</b>
RFM の出力 . . . . .	8	<b>特記事項</b> . . . . .	<b>27</b>
<b>第 3 章 クラスタ分析</b> . . . . .	<b>11</b>	商標 . . . . .	28
設定 . . . . .	12	<b>索引</b> . . . . .	<b>31</b>
<b>第 4 章 見込みのプロファイル</b> . . . . .	<b>13</b>		
設定 . . . . .	15		
カテゴリ応答フィールドの作成 . . . . .	15		



---

## 第 1 章 Direct Marketing

Direct Marketing のオプションでは、人口統計的特性、購入の特性、および消費者のさまざまなグループを定義するその他の特性を識別し、正の応答率を最大化するために特定のグループを対象とすることによって、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された一連のツールが用意されています。

**RFM 分析。** この手法によって、新規のオファーに対して応答する可能性が最も高い既存の顧客を識別できます。

**クラスター分析。** これは、データ内の自然なグループ (またはクラスター) 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。例えば、さまざまな人口統計的および購入の特性に基づいて異なる顧客グループを識別できます。

**見込みのプロファイル。** この手法では、過去のキャンペーンまたはテスト・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、連絡先の特定のグループを今後のキャンペーンの対象にすることができます。詳しくは、13 ページの『第 4 章 見込みのプロファイル』のトピックを参照してください。

**郵便番号応答率。** この手法では、過去のキャンペーンの結果を使用して、郵便番号の応答率を計算します。これらの比率を使用して、特定の郵便番号を今後のキャンペーンの対象にすることができます。詳しくは、17 ページの『第 5 章 郵便番号応答率』のトピックを参照してください。

**購入の傾向。** この手法では、テスト・メールまたは過去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、応答する可能性の最も高い連絡先を示します。詳しくは、21 ページの『第 6 章 購入の傾向』のトピックを参照してください。

**制御パッケージ検定。** この手法では、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有意な差があるかどうかを確認します。詳しくは、25 ページの『第 7 章 制御パッケージ検定』のトピックを参照してください。



---

## 第 2 章 RFM 分析

RFM 分析は、新規のオファーに対して応答する可能性が最も高い既存の顧客を識別するために使用する手法です。この手法はダイレクト・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。

- 新しいオファーに反応する可能性が高い顧客を識別するのに最も重要な因子はリーセンサーです。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。
- 次に重要な因子はフリークエンシーです。過去により多くの購入を行った顧客は、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 番目に重要な因子は消費した総額、つまりマネタリーです。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

- 一番最近購入した日付または一番最近購入してからの時間間隔に基づいて、リーセンサー・スコアが顧客に割り当てられます。このスコアは、少数のカテゴリーでリーセンサーの値を単純に順位付けしたものにに基づいています。例えば、5 つのカテゴリーを使用すると、購入日が一番最近である顧客はリーセンサー順位が 5 になり、一番遠い過去に購入した顧客のリーセンサー順位は 1 になります。
- 同様に、顧客にはフリークエンシー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。例えば、5 つのカテゴリーに順位付けとすると、最も頻繁に購入する顧客はフリークエンシー順位が 5 になります。
- 最後に、マネタリー値によって顧客が順位付けされます。最も高いマネタリー値が最も高い順位になります。前述の 5 つのカテゴリーの例では、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。

リーセンサー、フリークエンシー、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 (オファーに反応する可能性が最も高い顧客) とは結合 RFM スコアが最高の顧客です。例えば、5 つのカテゴリーの順位付けでは、合計で 125 種類の結合 RFM スコアがあり、結合 RFM スコアの最高は 555 になります。

データの考慮事項

- データ行がトランザクションを表す場合 (各行が単一トランザクションを表し、各顧客に対し複数のトランザクションがある可能性がある)、トランザクションからの RFM を使用します。詳しくは、『トランザクション・データからの RFM スコア』のトピックを参照してください。
- データ行が顧客とそのすべてのトランザクションの要約情報を表す場合 (消費した総額、総トランザクション数、最新のトランザクション日付の値を表す列を含む)、顧客データからの RFM を使用します。詳しくは、4 ページの『顧客データからの RFM スコア』のトピックを参照してください。

---

### トランザクション・データからの RFM スコア

データの考慮事項

データ・セットは、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。
- 各トランザクションの日付を含む変数。

- 各トランザクションのマネタリー値を含む変数。

トランザクション・データからの RFM スコアの作成

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**」

2. 「**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。
3. 「**トランザクション データ**」を選択し、「**続行**」をクリックします。
4. トランザクション日付を含む変数を選択します。
5. 各トランザクションのマネタリー量を含む変数を選択します。
6. 各顧客のトランザクション量を要約する方法を、合計 (全トランザクションの合計)、平均値、中央値、最大値 (最大トランザクション量) から選択します。
7. 各顧客を識別する固有の変数または変数の組み合わせを選択します。例えば、固有の ID コードまたは姓と名の組み合わせを使用して、ケースを識別できます。

---

## 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。これはリーセンサー・スコアの計算に使用します。
- 総購入回数。これはフリークエンシー・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値の要約。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和 (合計) になりますが、平均、最大値、またはその他の要約統計量にもできません。

RFM スコアを新しいデータ・セットに書き込む場合、アクティブなデータ・セットには、各ケース (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

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**」

2. 「**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。
3. 「**顧客データ**」を選択し、「**続行**」をクリックします。
4. 一番最近のトランザクション日付、または一番最近のトランザクションからの時間間隔を示す数字を含む変数を選択します。
5. 各顧客のトランザクション総数を含む変数を選択します。
6. 各顧客のマネタリー量の要約を含む変数を選択します。
7. RFM スコアを新しいデータ・セットに書き込む場合は、各顧客を識別する固有の変数、または変数の組み合わせを選択します。例えば、固有の ID コードまたは姓と名の組み合わせを使用して、ケースを識別できます。



---

## RFM ビン

大量の数値を少数のカテゴリにグループ化する方法を、**分割**と呼ぶことがあります。RFM 分析では、ピンは順位付けされたカテゴリのことです。「分割」タブを使用して、リーセンサー、フリークエンシー、およびマネタリーの値をそれらのピンに割り当てる方法を変更することができます。

### 「分割方法」

「入れ子」。ネスト分割では、リーセンサー値に単純な順位が割り当てられます。各リーセンサー順位内で顧客にフリークエンシー順位が割り当てられ、各フリークエンシー順位内で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、フリークエンシーやマネタリー順位のスコアの解釈が難しくなるという欠点もあります。例えば、フリークエンシー順位はリーセンサー順位に依存するので、リーセンサー順位が 5 の顧客とリーセンサー順位が 4 の顧客では、どちらもフリークエンシー順位が 5 の場合でも同じことを意味しないこともあります。

「独立」。リーセンサー、フリークエンシー、およびマネタリーの値に単純な順位が割り当てられます。3 つの順位は独立して割り当てられます。そのため、3 つの RFM コンポーネントのそれぞれの解釈は明確です。ある顧客のフリークエンシー・スコアが 5 であるということは、リーセンサー・スコアに関係なく、フリークエンシー・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欠点があります。

### 「ピン数」

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 (ピン) 数。使用できる結合 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。例えば、リーセンサー・ピンが 5、フリークエンシー・ピンが 4、マネタリー・ピンが 3 の場合、合計で 60 個の結合 RFM スコアが作成され、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、この場合に作成される結合 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・コンポーネントのピンの最大数は 9 です。

### 「同順位」

「同順位」とは、リーセンサー、フリークエンシー、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。各ピンにほぼ同数の顧客がいるが理想ですが、同順位の値が多数あると、ピン分布に影響が出る可能性があります。同順位を処理するには 2 つの選択肢があります。

- 「**同順位を同じピンに割り当てる**」。この方法では、ピン分布への影響に関係なく、常に同順位の値を同じピンに割り当てます。これにより、一貫した分割方法が提供されます。2 人の顧客のリーセンサー値が同じ場合、常に同じリーセンサー・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行うことも考えられます。そのため、5 ピンの順位付けでは、理想値の 20% の代わりに、顧客の 50% がリーセンサー・スコアの 5 を取得します。

ネストになっている分割方法では、フリークエンシー・スコアがリーセンサー・スコアのピン内に割り当てられ、マネタリー・スコアがフリークエンシー・スコアのピン内に割り当てられるので、フリークエンシー・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。したがって、同順位の値がどのように処理されるかにかかわらず、同じフリークエンシー値を持つ 2 人の顧客が同じリーセンサー・スコアでない場合は、2 人のフリークエンシー・スコアが同じではない可能性があります。

- 「同順位を無作為に割り当てる」。この方法では、順位付けの前に、非常に小さいランダム分散因子を同順位に割り当てることにより、均等なビン分布になります。つまり、順位付けされたビンに値を割り当てるため、同順位の値はなくなります。このプロセスは、元の値に影響しません。この方法は、同順位のあいまいさをなくすためにのみ使用されます。これによって均等なビン分布が生成されますが（それぞれのビンにはほぼ同数の顧客）、リーセンサー、フリークエンシー、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。

表 1. 同順位を同じビンへ割り当てる場合と、同順位を無作為に割り当てる場合の比較：

ID	一番最近の購入 (リーセンサー)	同順位を同じビンに割り当てる	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ると不均等なビン分布 (5 (10%)、4 (40%)、3 (20%)、2 (20%)、1 (10%)) となります。
- 同順位をランダムに割り当てると、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) 完全にランダム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、同順位の値を持つ特定のケースについて順位付けが変わる可能性があります。例えば、ケース 4 のリーセンサーの順位が 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。

## トランザクション・データからの RFM スコアの保存

トランザクション・データからの RFM 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・セットが常に作成されます。「保存」タブを使用して、保存するスコアおよび他の変数とその保存場所を指定します。

「変数」

各顧客を識別する固有の ID 変数は、自動的に新しいデータ・セットに保存されます。次の追加の変数は新規のデータ・セットに保存することができます。

- 「各顧客について一番最近の取引日付」。
- 「トランザクション数」。各顧客のトランザクション行の総数。
- 「量」。「変数」タブで選択した集計方法に基づいた、各顧客の集計量。

- 「**リーセンサー・スコア**」。一番最近のトランザクションの日付に基づいて、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トランザクション日付を示します。
- 「**フリクエンシー・スコア**」。総トランザクション数に基づいて、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トランザクションを示します。
- 「**マネタリー・スコア**」。選択されたマネタリーの要約統計量に基づいて、各顧客に割り当てられたスコア。より高いスコアの方が、マネタリーの要約統計量について、より高い値を示します。
- 「**RFM スコア**」。3 つの独立したスコアを 1 つの値に結合したもの。 $(\text{リーセンサー} \times 100) + (\text{フリクエンシー} \times 10) + \text{マネタリー}$

デフォルトでは、すべての使用可能な変数は新規データ・セットに含まれます。したがって、含めない変数を選択解除してください。オプションで、独自の変数名を指定することができます。変数名は標準の変数の命名規則に準拠している必要があります。

「場所」

トランザクション・データからの RFM 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・セットが常に作成されます。現在のセッションで新しいデータ・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の命名規則に準拠している必要があります (この制限は、外部データ・ファイル名には適用されません)。

---

## 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、保存するスコア変数とその保存場所を指定します。

「保存された変数の名前」

- 「**固有の名前を自動生成**」。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、新しい変数が固有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・セットに追加する場合に役立ちます。
- 「**カスタム名**」。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の命名規則に準拠している必要があります。

「変数」

保存するスコア変数を選択 (チェック・マークを付ける) します。

- 「**リーセンサー・スコア**」。「変数」タブで選択した「トランザクション日付または区間」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区間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。
- 「**フリクエンシー・スコア**」。「変数」タブで選択した「トランザクション数」変数の値に基づいて、各顧客に割り当てられたスコア。高い値に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」変数の値に基づいて、各顧客に割り当てられたスコア。高い値に高いスコアが割り当てられます。
- 「**RFM スコア**」。3 つの独立したスコアを 1 つの値に結合したもの。 $(\text{リーセンサー} \times 100) + (\text{フリクエンシー} \times 10) + \text{マネタリー}$

「場所」

顧客データでは、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・データ・セット**」。選択した RFM スコア変数は、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) を識別する固有の ID 変数が、現在のセッションで新規データ・セットへ書き込まれます。データ・セット名は標準の変数の命名規則に準拠している必要があります。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) を識別する固有の ID 変数が、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 の出力

### 分割済みデータ

分割済みデータのグラフとテーブルは、計算されたリーセンサー、フリークエンシー、およびマネタリー・スコアに基づいています。

「**リーセンサーおよびフリークエンシー別の平均の金銭的価値のヒート・マップ**」。平均のマネタリー分布のヒート・マップは、リーセンサー・スコアとフリークエンシー・スコアによって定義されるカテゴリーの平均のマネタリー値を表します。より濃い領域はマネタリー値の平均がより高いことを表しています。

「**ビン度数のグラフ**」。ビン度数のグラフには、選択された分割方法のビンの分布が表示されます。各バーは、それぞれの結合 RFM スコアが割り当てられたケース数を表します。

- 通常、すべて (またはほとんどの) バーがほぼ同じ高さになるような均等な分布が求められますが、同順位の値を同じビンに割り当てるデフォルトの分割方法を使用すると、特定の量の分散が予測されます。
- ビン分布の極端な変動がある場合、または空のビンが多い場合、別の分割方法 (ビン数を少なくする、同順位をランダムに割り当てる、あるいはその両方) を実行するか、RFM 分析の適合度を再検討する必要があります。

「**ビン度数のテーブル**」。テーブル形式で表示される以外はビン度数のグラフと同じ情報です。各セルにビン度数が含まれます。

### 分割されていないデータ

分割されていないデータのグラフとテーブルは、リーセンサー、フリークエンシー、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いています。

「**ヒストグラム**」。ヒストグラムは、リーセンサー、フリークエンシー、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、やや非対称の分布を示すのは特別なことではありません。

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。ただしリーセンサーのグラフの解釈は、リーセンサーの測度のタイプ (日付または時間間隔) によって異なります。日付の場合、左のバーが遠い過去の値を表します (古い日付ほど、新しい日付より低い値となります)。時間間隔の場合、左のバーがより最新の値を表します (時間間隔が短いほど、トランザクションが最近であることを表します)。

「変数ペアの散布図」。これらの散布図は、リーセンサー、フリークエンシー、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。

フリークエンシー・スケールの点が明確な線型にグループ化されるのは一般的です。これは、フリークエンシーが多くの場合、個々の値の比較的小さい範囲を表しているためです。例えば、トランザクションの総数が 15 より大きくならない場合、考えられるフリークエンシーの値は 15 個しかありません (少数を数えない場合)。一方で、リーセンサーの値は数百になることもあり、マネタリーの値は数千になることもあります。

リーセンサー軸の解釈は、リーセンサーの測度のタイプ (日付または時間間隔) によって異なります。日付の場合、原点に近い点ほど、遠い過去の日付を表します。時間間隔の場合、原点に近い点ほど、より新しい値を表します。





## 第 3 章 クラスタ分析

クラスタ分析は、データ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。例えば、さまざまな人口統計的および購入の特性に基づいて異なる顧客グループを識別できます。

**例:** 小売と消費者商品の企業は通常、顧客の購入傾向、性別、年齢、収入水準などを説明するクラスタ化手法をデータに適用します。これらの企業は、マーケティングおよび製品開発戦略を顧客グループごとに調整することで、販売を拡大し、ブランド・ロイヤリティを構築します。

### クラスタ分析データの考慮事項

**データ。** このプロシージャは、連続型フィールドとカテゴリー・フィールドの両方に使用できます。各レコード（行）はクラスタ化される顧客を表し、フィールド（変数）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。** レコードの順序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。順序の影響を最小限に抑えるために、レコードの順序をランダムにすることもできます。異なるランダムな順序で並べ替えられたレコードを使用して分析を複数回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、適切な測定レベルを割り当てることが重要です。

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

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。

表2. 測定レベルのアイコン

	数値	文字列	日付	時間
スケール (連続)		該当なし		
順序				
名義				

データ・エディターの「変数ビュー」で測定レベルを変更することができます。または、「変数プロパティの定義」ダイアログで各フィールドに適した測定レベルを指定することができます。

## 測定レベルが不明なフィールド

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

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

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

測定レベルはこのプロシーチャーにおいて重要であるため、すべてのフィールドに測定レベルが定義されるまで、このプロシーチャーを実行するためのダイアログにアクセスすることはできません。

クラスター分析を取得するには

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**」

1. 「**連絡先をクラスターにセグメント化する**」を選択します。
2. セグメント作成に使用するカテゴリー (名義、順序) 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。
3. 「**実行**」をクリックしてプロシーチャーを実行します。

---

## 設定

「**設定**」タブを使用すると、セグメントについて説明するグラフやテーブル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) を識別するデータ・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たりできます。

「**図表と表を表示**」。セグメントについて説明するグラフやテーブルを表示します。

「**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識別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 の命名規則に準拠している必要があります。
- セグメント・メンバーシップのフィールド名は、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プロシーチャーを実行する場合、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**」。セグメント数の決定方法を制御します。
- 「**自動的に判定**」。このプロシーチャーは、指定された最大値を上限として、「最適な」セグメントの個数を自動的に判定します。

「**固定値を指定**」。プロシーチャーは、指定した数のセグメントを作成します。



---

## 第 4 章 見込みのプロファイル

この手法では、過去のキャンペーンまたはテスト・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、連絡先の特定のグループを今後のキャンペーンの対象にすることができます。「応答フィールド」は、過去のキャンペーンまたはテスト・キャンペーンへの応答者を示します。「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含まれます。

**例:** テスト・メールの結果に基づき、企業のダイレクト・マーケティング部門は、人口統計情報を使用して、オファーに応答する可能性が最も高い顧客のタイプに関するプロファイルを生成しようとしています。

### 出力

出力には、各プロファイル・グループについて説明し、応答率 (正の応答のパーセント) および累積応答率を示すテーブルと、累積応答率のグラフが含まれます。対象最小応答率を設定した場合、テーブルには最小累積応答率を満たすプロファイルが色分けして表示され、グラフには指定した最小応答率の値で基準線が表示されます。

### 見込みのプロファイルのデータの考慮事項

「**応答フィールド**」。応答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。文字列または数値を使用できます。このフィールドに購入数または購入金額を示す値が含まれる場合、単一の値がすべての正の応答を示す新しいフィールドを作成する必要があります。詳しくは、15 ページの『**カテゴリ応答フィールドの作成**』のトピックを参照してください。

「**正の応答値**」。正の応答値は、肯定的に (購入したなど) 応答した顧客を示します。欠損値以外の他のすべての応答値は、負の応答を示していると見なされます。応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、名義型、順序型、または連続型 (スケール) です。文字列または数値を使用できます。

**測定レベル**。結果の計算に影響を与えるため、適切な測定レベルを割り当てることが重要です。

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

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。

表 3. 測定レベルのアイコン

	数値	文字列	日付	時間
スケール (連続)		該当なし		
順序				
名義				

データ・エディターの「変数ビュー」で測定レベルを変更することができます。または、「変数プロパティの定義」ダイアログで各フィールドに適した測定レベルを指定することができます。

#### 測定レベルが不明なフィールド

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

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

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

測定レベルはこのプロシージャにおいて重要であるため、すべてのフィールドに測定レベルが定義されるまで、このプロシージャを実行するためのダイアログにアクセスすることはできません。

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。

「ダイレクト マーケティング」 > 「手法を選択」

1. 「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。
2. 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。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。
3. 正の応答を示す値を入力します。値に値ラベルが定義されている場合、ドロップダウン・リストから値ラベルを選択でき、対応する値が表示されます。
4. プロファイルの作成に使用するフィールドを選択します。
5. 「実行」をクリックしてプロシージャを実行します。

#### 関連情報:

詳細な例

---

## 設定

「設定」タブを使用して、最小プロフィール・グループ・サイズを制御し、出力に最小応答率のしきい値を含めることができます。

「**最小プロフィール・グループ・サイズ**」。各プロフィールは、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、最小プロフィール・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が明らかになりますが、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。この値は正整数でなければなりません。

「**結果に最小応答率のしきい値情報を含める**」。結果には、応答率 (正の応答のパーセント) および累積応答率を示すテーブルと、累積応答率のグラフが含まれます。対象最小応答率を入力した場合、テーブルには最小累積応答率を満たすプロフィールが色分けして表示され、グラフには指定した最小応答率の値で基準線が表示されます。この値は 0 より大きくて 100 より小さくする必要があります。

---

## カテゴリー応答フィールドの作成

応答フィールドは、1 つの値がすべての正の応答を表す、カテゴリー型でなければなりません。その他の非欠損値は、負の応答と見なされます。応答フィールドが購入数や購入金額などの連続型の (スケール) 値を表す場合、単一の正の応答値がすべてのゼロ以外の応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。

- 負の応答を 0 (欠損値として処理される空白ではない) として記録する場合、次の数式で計算できます。

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

*NewName* は新しいフィールド名で、*OldName* は元のフィールド名です。これは、0 より大きいすべての非欠損値に 1 の値を、0 以下のすべての非欠損値に 0 を割り当てる論理式です。

- 負の応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応答値には 1 の値が、欠損値であるすべての応答値には 0 の値が割り当てられます。

- 負 (0) の応答値と欠損値の区別が付かない場合、正確な応答値は計算できません。実際の欠損値が比較的少ない場合、計算される応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応答情報が記録された場合など、欠損値が多い場合、計算される応答率は真の応答率より大幅に低くなるため、意味のないものとなります。

カテゴリー応答フィールドを作成するには

- メニューから次の項目を選択します。

「変換」 > 「変数の計算」

- 「ターゲット変数」に、新しいフィールド (変数) 名を入力します。
- 負の応答が 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。
- 負の応答が欠損値 (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。



---

## 第 5 章 郵便番号応答率

この手法では、過去のキャンペーンの結果を使用して、郵便番号の応答率を計算します。これらの比率を使用して、特定の郵便番号を今後のキャンペーンの対象にすることができます。「応答フィールド」は、過去のキャンペーンへの応答者を示します。「郵便番号フィールド」は、郵便番号を含むフィールドを識別します。

**例:** 過去のメーリングの結果に基づいて、企業のダイレクト・マーケティング部門は、郵便番号ごとに応答率を生成します。使用できる応答率の最小値および/またはメーリングに含め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

このプロシージャの出力には、郵便番号ごとの応答率を含む新しいデータ・セットと、十分位の順位 (上位 10%、上位 20% など) で結果を要約したテーブルとグラフが含まれます。テーブルは、ユーザー指定の累積応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。

新しいデータ・セットには、以下のフィールドが含まれています。

- 「**郵便番号**」。郵便番号グループが完全な値の一部のみに基づいている場合、これは郵便番号のそ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ー行ラベルは、元のデータ・セットの郵便番号フィールドの名前です。
- 「**応答率**」。各郵便番号の、正の応答のパーセントです。
- 「**応答**」。各郵便番号の、正の応答の数です。
- 「**連絡先**」。応答フィールドに非欠損値が含まれる、郵便番号ごとの連絡先数の合計です。
- 「**インデックス**」。数式  $N \times P \times (1-P)$  に基づいた重み付けされた応答。 $N$  は連絡先の数で、 $P$  は比率で示される応答率です。
- 「**順位**」。郵便番号の累積応答率を降順の十分位の順位 (上位 10%、上位 20% など) で表したものです。

### 郵便番号応答率データの考慮事項

「**応答フィールド**」。応答フィールドには文字列または数値を使用できます。このフィールドに購入数または購入金額を示す値が含まれる場合、単一の値がすべての正の応答を示す新しいフィールドを作成する必要があります。詳しくは、19 ページの『カテゴリー応答フィールドの作成』のトピックを参照してください。

「**正の応答値**」。正の応答値は、肯定的に (購入したなど) 応答した顧客を示します。欠損値以外の他のすべての応答値は、負の応答を示していると見なされます。応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列または数値を使用できます。

郵便番号応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**」

1. 「上位の応答郵便番号を特定する」を選択します。
2. 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。
3. 正の応答を示す値を入力します。値に値ラベルが定義されている場合、ドロップダウン・リストから値ラベルを選択でき、対応する値が表示されます。
4. 郵便番号を含むフィールドを選択します。
5. 「実行」をクリックしてプロシージャーを実行します。

オプションで下記が可能です。

- 完全な値でなく、郵便番号の最初の  $n$  文字または桁に基づいた応答率を生成する
- 結果を Excel ファイルに自動的に保存する
- 出力表示オプションを制御する

---

## 設定

### 「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応答率を計算するかを指定します。デフォルトでは、完全な郵便番号を使用し、郵便番号が同じすべてのレコードをグループ化して、グループの応答率を計算します。代わりに、完全な郵便番号の一部分 (最初の  $n$  桁または文字で構成) のみに基づいてレコードをグループ化することができます。例えば、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、郵便番号グループごとに 1 つのレコードが含まれます。値を入力する場合は、正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で、完全な値の代わりに、最初の  $n$  桁に基づいて郵便番号をグループ化する場合、元の値の桁数を指定する必要があります。桁数は郵便番号で使用できる最大のものです。例えば、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、先頭のゼロが省略されています。

### 「出力」

郵便番号ごとの応答率を含む新しいデータ・セットに加えて、十分位の順位 (上位 10%、上位 20% など) で結果を要約したテーブルとグラフを表示することができます。テーブルには十分位ごとの応答率、累積応答率、レコード数、累積レコード数が表示されます。グラフには十分位ごとの累積応答率と累積レコード数が表示されます。

「受け入れ可能な応答率の最小値」。対象最小応答率または損益分岐数式を入力した場合、テーブルは最小累積応答率を満たす十分位を色分けして表示し、グラフには指定した最小応答率の値で基準線が表示されません。

- 「目標応答率」。応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応答のパーセント)。この値は 0 より大きくて 100 より小さくする必要があります。
- 「式から損益分岐比率を計算」。数式に基づいて、最小累積応答率を計算します ( $(1$  つのパッケージ送付のコスト/応答ごとの純利益)  $\times 100$ )。どちらの値も正の数値にする必要があります。結果は 0 より大



小さく 100 より小さくなります。例えば、パッケージ送付のコストが \$0.75 で、応答ごとの純利益が \$56 の場合、最小応答率は  $(0.75/56) \times 100 = 1.34\%$  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、テーブルが色分けされて連絡先 (レコード) の累積最大数を超えていない十分位が示され、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。例えば、全体の 50% 以下の連絡先を含む最も高い応答率の十分位を確認する場合があります。この値は 0 より大きくて 100 より小さくする必要があります。
- 「連絡先の数」。最大値を連絡先数で示します。例えば、10,000 個を超えるパッケージを送付しない場合は、値を 10000 に設定できます。値は正整数にする必要があります (グループ化記号はなし)。

使用できる応答率の最小値と連絡先の最大数の両方を指定した場合、テーブルの色分けは、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」

このプロシージャでは、自動的に郵便番号ごとの応答率を含む新しいデータ・セットが作成されます。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。このファイルは Excel 97-2003 形式で保存されます。

---

## カテゴリー応答フィールドの作成

応答フィールドは、1 つの値がすべての正の応答を表す、カテゴリー型でなければなりません。その他の非欠損値は、負の応答と見なされます。応答フィールドが購入数や購入金額などの連続型の (スケール) 値を表す場合、単一の正の応答値がすべてのゼロ以外の応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。

- 負の応答を 0 (欠損値として処理される空白ではない) として記録する場合、次の数式で計算できます。

`NewName=OldName>0`

*NewName* は新しいフィールド名で、*OldName* は元のフィールド名です。これは、0 より大きいすべての非欠損値に 1 の値を、0 以下のすべての非欠損値に 0 を割り当てる論理式です。

- 負の応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応答値には 1 の値が、欠損値であるすべての応答値には 0 の値が割り当てられます。

- 負 (0) の応答値と欠損値の区別が付かない場合、正確な応答値は計算できません。実際の欠損値が比較的少ない場合、計算される応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応答情報が記録された場合など、欠損値が多い場合、計算される応答率は真の応答率より大幅に低くなるため、意味のないものとなります。

カテゴリー応答フィールドを作成するには

1. メニューから次の項目を選択します。

「変換」 > 「変数の計算」

2. 「ターゲット変数」に、新しいフィールド (変数) 名を入力します。

3. 負の応答が 0 として記録される場合、「数式」に  $01dName > 0$  と入力します。*OldName* は元のフィールド名です。
4. 負の応答が欠損値 (空白) として記録される場合、「数式」に  $NOT(MISSING(01dName))$  と入力します。*OldName* は元のフィールド名です。



---

## 第 6 章 購入の傾向

購入の傾向では、テスト・メールまたは過去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、応答する可能性の最も高い連絡先を示します。「応答フィールド」は、テスト・メールまたは過去のキャンペーンへの応答者を示します。「傾向」フィールドは、同様の特性を持つ連絡先が応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 つの基本的なステップがあります。

1. モデルを作成し、モデル・ファイルを保存します。目的 (対象) の結果が判明しているデータ・セットを使用してモデルを作成します。例えば、誰がダイレクト・メール・キャンペーンに応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、誰が応答し、誰が応答しなかったかについての情報を既にあるデータ・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・メールの結果や、過去の類似したキャンペーンに対する応答の情報などです。
2. そのモデルを異なるデータ・セット (目的の結果が不明) 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、テスト・メール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、人口統計上のさまざまな特性に基づいて最も応答および購入の可能性が高い連絡先を特定します。

### 出力

このプロシージャでは自動的に、検定データの傾向スコアを含む新しいフィールドがデータ・セットに作成され、他のデータ・セットのスコアリングに使用できる XML モデル・ファイルも作成されます。オプションの診断出力には、予測した応答と実際の応答を比較する全体のモデル品質グラフと分類テーブルがあります。

### 購入の傾向データの考慮事項

「**応答フィールド**」。応答フィールドには文字列または数値を使用できます。このフィールドに購入数または購入金額を示す値が含まれる場合、単一の値がすべての正の応答を示す新しいフィールドを作成する必要があります。詳しくは、24 ページの『**カテゴリー応答フィールドの作成**』のトピックを参照してください。

「**正の応答値**」。正の応答値は、肯定的に (購入したなど) 応答した顧客を示します。欠損値以外の他のすべての応答値は、負の応答を示していると見なされます。応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**」。傾向の予測に使用されるこのフィールドには、名義型、順序型、または連続型 (スケール) の文字列または数値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レベルを割り当てるのが重要です。












**測定レベル**。結果の計算に影響を与えるため、適切な測定レベルを割り当てるのが重要です。

- *Nominal* (名義データ)。本質的な順位を持たないカテゴリーを表す値である場合 (従業員の勤務先企業での部署など)、変数を名義変数として取り扱うことができます。名義変数の例としては、地域、郵便番号、宗教上の所属などが挙げられます。

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

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。

表 4. 測定レベルのアイコン

	数値	文字列	日付	時間
スケール (連続)		該当なし		
順序				
名義				

データ・エディターの「変数ビュー」で測定レベルを変更することができます。または、「変数プロパティの定義」ダイアログで各フィールドに適した測定レベルを指定することができます。

#### 測定レベルが不明なフィールド

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

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

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

測定レベルはこのプロシージャにおいて重要であるため、すべてのフィールドに測定レベルが定義されるまで、このプロシージャを実行するためのダイアログにアクセスすることはできません。

購入の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。

「ダイレクト マーケティング」 > 「手法を選択」

1. 「購入の可能性が高い連絡先を選択する」を選択します。
2. 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。
3. 正の応答を示す値を入力します。値に値ラベルが定義されている場合、ドロップダウン・リストから値ラベルを選択でき、対応する値が表示されます。
4. 傾向を予測するのに使用するフィールドを選択します。

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

5. 「**モデル情報を XML ファイルにエクスポート**」を選択 (チェック・マークを付ける) します。
6. ディレクトリー・パスと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を選びます。
7. 「**実行**」をクリックしてプロシージャーを実行します。

モデル・ファイルを使用して他のデータ・セットをスコアリングするには

8. スコアリングするデータ・セットを開きます。
9. 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

「モデルの検証」

モデルの検証では、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」セクションで分類テーブルを選択した場合、テーブルは学習セクション (選択) と検定セクション (未選択) に分けられ、比較されます。モデル検証を選択する場合は、必ず分類テーブルも選択してください。スコアは学習サンプルから生成されたモデルに基づいていますが、学習サンプルに含まれるレコードの数は利用可能なレコードの合計より常に少ないです。例えば、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から作成されたモデルは、利用可能なすべてのレコードから作成されたモデルに比べて信頼性に劣る場合があります。

- 「**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パーセントを指定します。応答フィールドが非欠損値の残りのレコードは、検定サンプルに割り当てられます。この値は 0 より大きくて 100 より小さくする必要があります。
- 「**シードを設定して結果を再現**」。レコードがランダムに学習サンプルと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、プロシージャーを実行するたび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。

「**分類テーブル**」。予測された正の応答および負の応答と実際の正の応答および負の応答を比較するテーブルが表示されます。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ある程度示しますが、正しく予測された正の応答のパーセント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、分類テーブルの予測された正の応答カテゴリーに割り当てます。プロシージャーによって生成されたスコアは、連絡先が正の応答をする (たとえば、購入する) 確率を表します。一般的な規則として、対象応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。例えば、少なくとも 5% の応答率を考えている場合、0.05 を指定します。この値は 0 より大きくて 1 より小さくする必要があります。

「再コード化された応答フィールドの名前およびラベル」

このプロシージャーでは、応答フィールドを、1 が正の応答を、0 が負の応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 の命名規則に準拠している必要があります。

「スコアを保存」

傾向スコアを含む新しいフィールドが、自動的に元のデータ・セットに保存されます。スコアは、正の応答の確率を比率で示します。

- フィールド名は、IBM SPSS Statistics の命名規則に準拠している必要があります。
- 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プロシージャーを実行する場合、それぞれ別の名前を指定する必要があります。

---

## カテゴリー応答フィールドの作成

応答フィールドは、1 つの値がすべての正の応答を表す、カテゴリー型でなければなりません。その他の非欠損値は、負の応答と見なされます。応答フィールドが購入数や購入金額などの連続型の (スケール) 値を表す場合、単一の正の応答値がすべてのゼロ以外の応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。

- 負の応答を 0 (欠損値として処理される空白ではない) として記録する場合、次の数式で計算できます。

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

*NewName* は新しいフィールド名で、*OldName* は元のフィールド名です。これは、0 より大きいすべての非欠損値に 1 の値を、0 以下のすべての非欠損値に 0 を割り当てる論理式です。

- 負の応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応答値には 1 の値が、欠損値であるすべての応答値には 0 の値が割り当てられます。

- 負 (0) の応答値と欠損値の区別が付かない場合、正確な応答値は計算できません。実際の欠損値が比較的少ない場合、計算される応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応答情報が記録された場合など、欠損値が多い場合、計算される応答率は真の応答率より大幅に低くなるため、意味のないものとなります。

カテゴリー応答フィールドを作成するには

1. メニューから次の項目を選択します。

「変換」 > 「変数の計算」

2. 「ターゲット変数」に、新しいフィールド (変数) 名を入力します。
3. 負の応答が 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。
4. 負の応答が欠損値 (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。

## 第 7 章 制御パッケージ検定

この手法では、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有意な差があるかどうかを確認します。キャンペーンの有効性は応答によって測定されます。「キャンペーン・フィールド」は、オファー A やオファー B などの異なるキャンペーンを識別します。「応答フィールド」は、連絡先がキャンペーンに応答したかどうかを示します。応答が購入量として記録される場合（例えば、99.99）は、「購入量」を選択します。「はい」または「いいえ」など、連絡先が正の応答または負の応答のどちらを行ったのかを応答で単純に示す場合は「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応答を生み出すかどうかを確認しようとしています。そこで、部門は新しいパッケージ・デザインが非常に高い正の応答率を生み出すかどうかを判断するためのテスト・メールを送付します。テスト・メールは、既存のパッケージを受け取る制御グループと、新しいパッケージ・デザインを受け取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、有意な差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応答および負の応答の度数およびパーセントを示すテーブルと、グループ間で有意な差があるかを示すテーブルが表示されます。

### 制御パッケージ検定データの考慮事項および前提

「キャンペーン・フィールド」。「キャンペーン・フィールド」は、カテゴリ型（名義または順序）でなければなりません。

「有効性応答フィールド」。「有効性フィールド」で「購入量」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）にする必要があります。

負（購入量の場合、0）の応答値と欠損値の区別が付かない場合、正確な応答率は計算できません。実際の欠損値が比較的少ない場合、計算される応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応答情報が記録された場合など、欠損値が多い場合、計算される応答率は真の応答率より大幅に低くなるため、意味のないものとなります。

**仮定:** このプロシージャでは、連絡先が各キャンペーン・グループにランダムに割り当てられていることを前提としています。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、すべての連絡先が任意のグループに割り当てられる確率は等しいものとします。

制御パッケージ検定を取得するには

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」

1. 「キャンペーンの有効性の比較」を選択します。
2. 各連絡先が属するキャンペーン・グループを識別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。
3. 応答の有効性を示すフィールドを選択します。

応答フィールドが購入量を示す場合、フィールドは数値型にする必要があります。



応答フィールドで、連絡先が正の応答または負の応答（「はい」または「いいえ」など）のどちらを行ったのかを単純に示す場合は、「**応答**」を選択し、正の応答を表す値を入力します。値に値ラベルが定義されている場合、ドロップダウン・リストから値ラベルを選択でき、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 が正の応答を、0 が負の応答を示し、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** の命名規則に準拠している必要があります。

4. 「**実行**」をクリックしてプロシーチャーを実行します。

---

## 特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。この資料は、IBM から他の言語でも提供されている可能性があります。ただし、これを入手するには、本製品または当該言語版製品を所有している必要がある場合があります。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

*IBM Director of Licensing*

*IBM Corporation*

*North Castle Drive, MD-NC119*

*Armonk, NY 10504-1785*

*US*

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

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

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者にお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名前はすべて架空のものであり、名前や住所が類似する個人や企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾:

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

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

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

© Copyright IBM Corp. \_年を入れる\_. All rights reserved.

---

## 商標

IBM、IBM ロゴおよび [ibm.com](http://ibm.com) は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。



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

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子会社の米国およびその他の国における商標または登録商標です。

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

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

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

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連会社の米国およびその他の国における商標または登録商標です。



---

## 索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

### [カ行]

クラスター 11  
クラスター分析 11  
クラスター分析 (Direct Marketing のオプション) 11  
購入の傾向 21

### [サ行]

制御パッケージ検定 25

### [マ行]

見込みのプロファイル (Direct Marketing のオプション) 13

### [ヤ行]

郵便番号応答率 17

### [ラ行]

ロジスティック回帰 (Direct Marketing) 21

## R

RFM 3, 6, 7, 8  
顧客データ 4  
トランザクション・データ 3  
分割 5







Printed in Japan