

IBM SPSS Statistics Base 24

The IBM logo, consisting of the letters 'IBM' in a bold, black, sans-serif font. Each letter is composed of eight horizontal bars of equal height and width, stacked vertically to form the characters. The logo is centered horizontally on the page.

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、207 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 Statistics Base 24

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

第 1 章 コードブック	1	T-TEST コマンドの追加機能.	36
コードブックの「出力」タブ	1	第 10 章 一元配置分散分析	39
コードブックの統計タブ	3	一元配置分散分析の対比	39
第 2 章 度数	5	一元配置分散分析のその後の検定	40
度数分布表の統計	5	一元配置分散分析のオプション	41
度数分布表のグラフ	7	ONEWAY コマンドの追加機能	42
度数分布表の書式	7	第 11 章 GLM 1 変量分散分析	43
第 3 章 記述統計	9	GLM モデル	44
記述統計のオプション	9	項の構築	45
DESCRIPTIVES コマンドの追加機能	10	平方和	45
第 4 章 探索	11	GLM の対比	46
探索的分析の統計	12	対比タイプ	46
探索的分析の作図	12	GLM のプロファイル・プロット	47
探索のべき乗変換	12	GLM のオプション	47
探索的分析のオプション	13	UNIANOVA コマンドの追加機能	48
EXAMINE コマンドの追加機能	13	GLM のその後の比較	48
第 5 章 クロス集計表	15	GLM のオプション	50
クロス集計表の層	16	UNIANOVA コマンドの追加機能	50
クロス集計表のクラスター棒グラフ	16	GLM の保存	51
テーブル層に層変数を表示するクロス集計表	16	GLM のオプション	52
クロス集計表の統計	16	UNIANOVA コマンドの追加機能	52
クロス集計表のセル表示	18	第 12 章 2 変量の相関分析	55
クロス集計表の表書式	19	2 変量の相関分析のオプション	55
第 6 章 要約	21	CORRELATIONS コマンドと NONPAR CORR コマ ンドの追加機能	56
ケースの要約のオプション	21	第 13 章 偏相関	57
ケースの要約の統計	22	偏相関のオプション	57
第 7 章 平均値	25	PARTIAL CORR コマンドの追加機能	58
グループの平均のオプション	25	第 14 章 距離	59
第 8 章 OLAP キューブ	29	距離行列の非類似度の測定方法	59
OLAP キューブの統計	29	距離行列の類似度の測定方法	60
OLAP キューブの差分	31	PROXIMITIES コマンドの追加機能	60
OLAP キューブの表題	32	第 15 章 線型モデル	61
第 9 章 t 検定	33	線型モデルの取得方法	61
T 検定	33	目的	61
独立したサンプルの T 検定	33	基本	62
独立したサンプルの T 検定のグループの定義	34	モデル選択	63
独立したサンプルの T 検定のオプション	34	アンサンプル	64
対応のあるサンプルの T 検定	34	詳細設定	64
対応のあるサンプルの t 検定のオプション	35	モデル・オプション	64
T-TEST コマンドの追加機能	35	モデルの要約	64
1 サンプルの T 検定	35	自動データ準備	65
1 サンプルの t 検定のオプション	36	予測値の重要度	65
T-TEST コマンドの追加機能	36	予測対観測	65
		残差	65

外れ値	66
効果	66
係数	66
推定平均	67
モデル構築の要約	67

第 16 章 線型回帰 69

線型回帰の変数選択方法	70
線型回帰の規則の設定	70
線型回帰の作図	70
線型回帰: 新規変数の保存	71
線型回帰の統計	72
線型回帰のオプション	73
REGRESSION コマンドの追加機能	74

第 17 章 順序回帰 75

順序回帰のオプション	76
順序回帰の出力	76
順位回帰の位置モデル	77
項の構築	77
順序回帰の尺度モデル	78
項の構築	78
PLUM コマンドの追加機能	78

第 18 章 曲線推定 79

曲線推定のモデル	80
曲線推定の保存	80

第 19 章 偏最小二乗回帰 81

モデル	83
オプション	83

第 20 章 最近傍分析 85

近傍	87
特徴	88
分割	88
保存	89
出力	89
オプション	90
モデル・ビュー	90
特徴空間	90
変数の重要度	92
同位	92
最近傍の距離	92
四分位分布図	92
特徴選択エラー・ログ	92
k 選択エラー・ログ	93
k および特徴選択エラー・ログ	93
分類テーブル	93
誤差の集計	93

第 21 章 判別分析 95

判別分析: 範囲の定義	96
判別分析: ケースの選択	96
判別分析: 統計	96
判別分析: ステップワイズ法	97

判別分析: 分類	97
判別分析: 保存	98
DISCRIMINANT コマンドの追加機能	98

第 22 章 因子分析 101

因子分析のケースの選択	102
因子分析の記述統計	102
因子分析の因子抽出	102
因子分析の回転	103
因子分析の因子得点	104
因子分析オプション	104
FACTOR コマンドの追加機能	104

**第 23 章 クラスタリングの手続きの選
択 105**

第 24 章 TwoStep クラスタ分析 . . . 107

TwoStep クラスタ分析のオプション	108
TwoStep クラスタ分析の出力	109
クラスタ・ビューアー	110
クラスタ・ビューアー	110
クラスタ・ビューアーのナビゲート	114
レコードのフィルタリング	115

第 25 章 階層クラスタ分析 117

階層クラスタ分析の方法	117
階層クラスタ分析の統計	118
階層クラスタ分析の作図	118
階層クラスタ分析の新規変数の保存	118
CLUSTER コマンド・シンタックスの追加機能	118

**第 26 章 大規模ファイルのクラスタ
分析 121**

大規模ファイルのクラスタ分析の効率	122
大規模ファイルのクラスタ分析の反復	122
大規模ファイルのクラスタ分析の保存	122
大規模ファイルのクラスタ分析のオプション	123
QUICK CLUSTER コマンドの追加機能	123

第 27 章 ノンパラメトリック検定 . . . 125

1 サンプルのノンパラメトリック検定	125
1 サンプルのノンパラメトリック検定を取得する には	125
「フィールド」タブ	125
「設定」タブ	126
NPTESTS コマンドの追加機能	128
独立サンプルのノンパラメトリック検定	128
独立サンプルのノンパラメトリック検定を取得す るには	129
「フィールド」タブ	129
「設定」タブ	129
NPTESTS コマンドの追加機能	131
対応サンプルのノンパラメトリック検定	131
対応サンプルのノンパラメトリック検定を取得す るには	131
「フィールド」タブ	132

「設定」タブ	132
NPTESTS コマンドの追加機能	134
モデル・ビュー	134
モデル・ビュー	134
NPTESTS コマンドの追加機能	139
レガシー ダイアログ	139
カイ 2 乗検定	140
2 項検定	141
ラン検定	142
1 サンプル Kolmogorov-Smirnov 検定	143
2 個の独立サンプルの検定	144
2 個の対応サンプルの検定	146
複数の独立サンプルの検定	147
複数の対応サンプルの検定	148
第 28 章 多重回答分析	151
多重回答分析	151
多重回答グループの定義	151
多重回答の度数表	152
多重回答のクロス集計表	153
多重回答のクロス集計表の範囲の定義	154
多重回答のクロス集計表のオプション	154
MULT RESPONSE コマンドの追加機能	155
第 29 章 結果の報告	157
結果の報告	157
報告書の行の集計	157
集計報告書: 行の集計を取得するには	157
報告書のデータ列/ブレイク列の書式	158
報告書の集計行/最終集計行	158
報告書のブレイク・オプション	158
報告書のオプション	159
報告書のレイアウト	159
報告書の表題	159
報告書の列の集計	159
集計報告書: 列の集計を取得するには	160
データ列の集計関数	160
合計列のデータ列集計	160
報告書の列の書式	161
報告書の列の集計でのブレイク列のオプション	161
報告書の列の集計のオプション	161
報告書の列の集計のレイアウト	161
REPORT コマンドの追加機能	161
第 30 章 信頼性分析	163
信頼性分析の統計	164
RELIABILITY コマンドの追加機能	165
第 31 章 多次元スケーリング法	167
多次元スケーリング法のデータの型	168
多次元スケーリング法の測度の作成	168
多次元スケーリング法モデル	168
多次元スケーリング法のオプション	169
ALSCAL コマンドの追加機能	169

第 32 章 比率統計	171
比率統計	171
第 33 章 ROC 曲線	173
ROC 曲線のオプション	173
第 34 章 シミュレーション	175
モデル・ファイルに基づいてシミュレーションを設計するには	176
カスタム方程式に基づいてシミュレーションを設計するには	176
予測モデルを使用しないシミュレーションを設計するには	177
シミュレーション・プランからシミュレーションを実行するには	178
シミュレーション・ビルダー	178
「モデル」タブ	179
「シミュレーション」タブ	181
「シミュレーションの実行」ダイアログ	190
「シミュレーション」タブ	190
「出力」タブ	191
シミュレーションからのグラフ出力の作業	192
グラフ・オプション	193
第 35 章 地理空間モデリング	195
マップの選択	195
マップの選択	196
地理空間関係	196
座標系の設定	196
投影法の設定	197
投影法と座標系	197
データ ソース	197
データ ソースの追加	198
データとマップのアソシエーション	198
キーの検証	198
地理空間アソシエーション ルール	199
イベント データ フィールドの定義	199
フィールドの選択	199
出力	199
保存	201
規則の作成	201
ビン分割と集計	202
空間的・時間的予測	203
フィールドの選択	203
時間区分	203
集計	204
出力	204
モデル オプション	205
保存	206
詳細設定	206
完了	206
特記事項	207
商標	208
索引	211

第 1 章 コードブック

コードブックは、辞書情報 (変数名、変数ラベル、値ラベル、欠損値など) および、アクティブ・データ・セット内の、すべての変数または指定した変数の要約統計量および多重回答グループの要約統計量を報告します。名義変数、順序変数、および多重回答セットの場合、要約統計量には度数とパーセントが含まれます。スケール変数の場合、要約統計量には、平均値、標準偏差、および 4 分位が含まれます。

注: コードブックは分割ファイル・ステータスを無視します。これはに、欠損値の多重代入 (「欠損値」のアドオン・オプションで使用可能) 用に作成された分割ファイル・グループが含まれます。

コードブックを取得するには

1. メニューから次の項目を選択します。

「分析」 > 「報告書」 > 「コードブック」

2. 「変数」タブをクリックします。
3. 1 つ以上の変数または多重回答セット、あるいはその両方を選択します。

必要に応じて、以下の操作を実行することができます。

- 表示される変数情報を制御する。
- 表示される統計量を制御する (または、すべての要約統計量を除外する)。
- 変数と多重回答セットが表示される順序を制御する。
- ソース・リストの任意の変数の測定レベルを変更し、表示される要約統計量を変更する。詳しくは、3 ページの『コードブックの統計タブ』のトピックを参照してください。

測定レベルの変更

変数の測定レベルを一時的に変更できます。(多重回答セットの測定レベルは変更できません。多重回答セットは常に名義型として扱われます。)

1. ソース・リスト内の変数を右クリックします。
2. ポップアップ・メニューから測定レベルを選択します。

これにより、測定レベルが一時的に変更されます。実際には、これは数値変数にのみ役立ちます。文字列変数の測定レベルは名義変数または順序変数に限られ、どちらもコードブック手続きで同様に処理されます。

コードブックの「出力」タブ

「出力」タブは、それぞれの変数と多重回答グループに含まれる変数情報、変数と多重回答グループが表示される順序、およびオプションのファイル情報テーブルの内容を制御します。

変数情報

変数ごとに表示される辞書情報を制御します。

位置: ファイル順序内の変数の位置を表す整数。多重回答グループでは使用できません。

ラベル: 変数または多重回答グループに関連付けられている説明ラベル。

型: 基本的なデータ型。「数値」、「文字列」、または「多重回答グループ」のいずれかです。

形式: 変数の表示形式 (「A4」、「F8.2」、「DATE11」など)。多重回答グループでは使用できません。

測定レベル: 可能な値は、「名義」、「順序」、「尺度」、「不明」です。表示される値は、辞書に格納されている測定レベルであり、「変数」タブのソース変数リストで測定レベルを変更して指定された一時的な測定レベルのオーバーライドによる影響を受けることはありません。多重回答グループでは使用できません。

注: 数値変数の測定レベルが明示的に設定されていない場合 (外部ソースから読み込まれたデータや新規作成した変数の場合など)、測定レベルは最初のデータ・パスまで「不明」の場合があります。詳しくは、のトピックを参照してください。

役割: 一部のダイアログでは、定義された役割に基づいて、分析する変数を事前に選択することができます。

値ラベル: 特定のデータ値に関連付けられている説明ラベル。

- 「統計」タブで「度数」または「パーセント」が選択されている場合、「値ラベル」をここで選択しなくても、定義済みの値のラベルが出力に含まれます。
- 多重二値セットの場合、「値ラベル」は、セットの定義方法に応じてセット内の基本変数の変数ラベルか、集計値のラベルになります。詳しくは、のトピックを参照してください。

欠損値: ユーザー定義の欠損値。「統計」タブで「度数」または「パーセント」が選択されている場合、「欠損値」をここで選択しなくても、定義済みの値のラベルが出力に含まれます。多重回答グループでは使用できません。

カスタム属性: ユーザー定義のカスタム変数属性。出力には、各変数に関連付けられているすべてのカスタム変数属性の名前と値の両方が含まれます。詳しくは、のトピックを参照してください。多重回答グループでは使用できません。

予約属性: 予約されているシステム変数属性。システム属性は表示はできますが、変更はできません。システム属性名は、ドル記号 (\$) で始まります。名前が「@」または「\$@」で始まる非表示属性は含まれません。出力には、各変数に関連付けられているすべてのシステム属性の名前と値の両方が含まれます。多重回答グループでは使用できません。

ファイル情報

オプションのファイル情報テーブルには、次のファイル属性を含めることができます。

ファイル名: IBM® SPSS® Statistics データ・ファイルの名前。データ・セットが IBM SPSS Statistics 形式で保存されたことがない場合、データ・ファイル名はありません。(「データ・エディター」ウィンドウのタイトル・バーにファイル名が表示されていない場合、アクティブ・データ・セットにはファイル名がありません。)

場所: IBM SPSS Statistics データ・ファイルのディレクトリー (フォルダー) の場所。データ・セットが IBM SPSS Statistics 形式で保存されたことがない場合、場所はありません。

ケースの数: アクティブ・データ・セットのケース数。これはケースの総数です。フィルター条件により要約統計から除外された可能性があるケースもすべて含まれます。

ラベル: FILE LABEL コマンドで定義されたファイル・ラベルです (定義されている場合)。

文書: データ・ファイル文書テキスト。

重み付け状況: 重み付けが有効の場合、重み付け変数の名前が表示されます。詳しくは、のトピックを参照してください。

カスタム属性: ユーザー定義のカスタム・データ・ファイル属性。DATAFILE ATTRIBUTE コマンドを使用して定義されたデータ・ファイル属性。

予約属性: 予約されているシステム・データ・ファイル属性。システム属性は表示はできますが、変更はできません。システム属性名は、ドル記号 (\$) で始まります。名前が「@」または「\$@」で始まる非表示属性は含まれません。出力には、すべてのシステム・データ・ファイル属性の名前と値の両方が含まれます。

変数の表示順

変数と多重回答グループが表示される順序を制御するために、次の選択肢が用意されています。

アルファベット順: 変数名のアルファベット順。

ファイル: データ・セット内に変数が現れる順序 (データ・エディターに変数が表示される順序)。昇順の場合、選択されたすべての変数の後、最後に多重回答グループが表示されます。

測定レベル: 測定レベルでソートします。4 つのソート・グループ (名義、順序、尺度、不明) が作成されます。多重回答グループは名義として扱われます。

注 :数値変数の測定レベルが明示的に設定されていない場合 (外部ソースから読み込まれたデータや新規作成した変数の場合など)、測定レベルは最初のデータ・パスまで「不明」の場合があります。

変数リスト: 「変数」タブの選択された変数のリストに変数と多重回答グループが表示される順序。

カスタム属性名: ソート順序オプションのリストには、ユーザー定義のカスタム変数属性の名前もすべて含まれます。昇順の場合、属性を持たない変数が一番上にソートされ、次に値が定義されていない属性を持つ変数、その次に値が定義された属性を持つ変数が値のアルファベット順にソートされます。

カテゴリーの最大数

出力に各固有値の値のラベル、度数、またはパーセントが含まれている場合、値の数が指定された値を超えていればテーブルにこの情報を表示しないように抑制できます。デフォルトでは、変数の固有値の数が 200 を超えると、この情報は抑制されます。

コードブックの統計タブ

「統計」タブでは、出力に含まれる要約統計量を制御したり、要約統計量の表示を完全に抑制したりすることができます。

度数とパーセント

名義型変数、順序型変数、多重回答セット、およびスケール変数のラベル値の場合、次の統計を使用できません。

度数: 変数のそれぞれの値 (または値の範囲) を持つケースの数。

Percent (パーセント): ある特定の値を持つケースの割合。

中心傾向と分散

スケール変数の場合、次の統計を使用できます。

Mean (平均). 中心傾向の指標。算術平均 (合計をケース数で割った値) です。

Standard Deviation (標準偏差). 平均値の周りの散らばりの指標。正規分布では、平均から 1 標準偏差以内にケースの 68% が含まれ、2 標準偏差以内にケースの 95% が含まれます。例えば平均年齢が 45 で標準偏差が 10 である場合、正規分布ではケースの 95% が 25 と 65 の間に含まれます。

Quartiles (4 分位 (*Kaplan-Meier*)). 25、50、および 75 パーセンタイルに対応する値を表示します。

注: 「変数」タブのソース変数リストで、変数に関連付けられた測定レベルを一時的に変更できます (したがって、その変数に関して表示される要約統計量に変更されます)。

第 2 章 度数

「度数分布表」手続きは、多くのタイプの変数を記述するのに役立つ統計とグラフ表示を生成します。「度数分布表」手続きは、データの調査を開始する場合に適した手続きです。

度数レポートと棒グラフの場合、値を昇順や降順で配置することも、カテゴリーを度数別に順序付けることもできます。変数が多くの異なる値を持っている場合は、度数レポートを抑制することができます。図表には、度数 (デフォルト) またはパーセントのラベルを付けることができます。

例: ある企業の顧客が業種によってどのように分布しているのかを検討するとします。出力から、顧客の 37.5% は政府機関に属し、24.9% は一般企業、28.1% は学術機関、9.4% は医療保険業界に属していることがわかります。販売収益のような連続した量的データの場合は、製品の平均売上高が 3,576 ドルで、標準偏差が 1,078 ドルであることがわかります。

統計と作図: 度数、パーセント、累積パーセント、平均値、中央値、合計、標準偏差、分散、範囲、最小値と最大値、平均値の標準誤差、歪度と尖度 (両方とも標準誤差付き)、4 分位、ユーザー指定のパーセンタイル、棒グラフ、円グラフ、ヒストグラム。

度数分布表データの考慮事項

データ: 数値コードまたは文字列を使用してカテゴリー変数をコード化します (名義尺度または順序尺度)。

仮定: 集計表とパーセントは、いずれの分布のデータについても役立つ説明を提供しますが、順序付けされたカテゴリーまたは順序付けされていないカテゴリーを持つ変数の場合に、特に役立ちます。平均値や標準偏差などのオプションの要約統計のほとんどは、通常理論に基づいており、対称的な分布の量的変数に適しています。中央値、4 分位、パーセンタイルなどの頑健な統計は、正規性の仮定に適合するかどうかにかかわらず、量的変数に適しています。

度数分布表を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「記述統計」 > 「度数分布表...」

2. 1 つ以上のカテゴリー変数または量的変数を選択します。

必要に応じて、以下の操作を実行することができます。

- 「統計」をクリックして、量的変数の記述統計を求める。
- 「図表」をクリックして、棒グラフ、円グラフ、ヒストグラムを作成する。
- 「書式」をクリックして、結果の表示順序を指定する。

度数分布表の統計

パーセンタイル値: 順序付けされたデータをグループに分割する量的変数の値です。データの特定のパーセントがこの値を上回り、残りのパーセントがこの値を下回るように、データが分割されます。4 分位 (25、50、および 75 パーセンタイル) の場合、同じサイズの 4 つのグループに観測値が分割されます。必要な等サイズ・グループの個数が 4 以外の場合は、「等サイズの n グループに分割」を選択します。個別のパーセンタイルを指定することもできます (例えば 95 パーセンタイルは、観測値の 95% が含まれる値です)。

中心傾向: 平均値、中央値、最頻値、すべての値の合計など、分布の位置を記述する統計です。

- *Mean* (平均). 中心傾向の指標。算術平均 (合計をケース数で割った値) です。
- *Median* (中央値). この値より上と下それぞれにケースの半数ずつが該当することになる値。50 パーセントイル。ケース数が偶数の場合の中央値は、昇順または降順にソートしたときに中央に来る 2 つのケースの平均です。中央値は、外れ値に対して敏感でない、中心傾向の指標です。それに対して平均値は、少数の極端に大きいまたは小さい値に影響されることがあります。
- **最頻値:** 最も多く出現する値。複数の値が最高の頻度で出現し、その頻度が同じである場合は、それぞれが最頻値となります。度数分析プロシージャーは、それらの最頻値のうち最小値だけを報告します。
- *Sum* (合計). 欠損値でない値を持つすべてのケースにわたる値の和 (合計)。

散らばり: 標準偏差、分散、範囲、最小値、最大値、平均の標準誤差など、データの変動量または広がり量を測定する統計です。

- **標準偏差:** 平均値の周りの散らばりの指標。正規分布では、平均から 1 標準偏差以内にケースの 68% が含まれ、2 標準偏差以内にケースの 95% が含まれます。例えば平均年齢が 45 で標準偏差が 10 である場合、正規分布ではケースの 95% が 25 と 65 の間に含まれます。
- *Variance* (分散 (信頼性分析)). 平均値の周りの値の散らばりの指標。平均値からの偏差の平方和を、ケース数より 1 少ない値で割ったものに等しくなります。分散の測定単位は、変数自体の単位の 2 乗です。
- *Range* (OK (ファイルオープン時のオプション)). 数値変数の最大値と最小値の差。最大値から最小値を引いた値。
- *Minimum* (最小値). 数値変数の最小値。
- *Maximum* (最大). 数値変数の最大値。
- **標準誤差の平均値:** 同じ分布から抽出したサンプルの間で平均値がどの程度異なるかを示す指標。観測した平均と仮説による値をおおまかに比較するために使用することができます (差と標準誤差の比率が -2 より小さいか +2 より大きい場合は、2 つの値が異なっていると結論付けることができます)。

分布: 尖度と歪度は、分布の形状と対称を示す統計量です。これらの統計量は、その標準誤差とともに表示されます。

- *Skewness* (歪度). 分布の非対称性の指標。正規分布は対称であり、歪度の値は 0 です。歪度が正の大きな値である分布は、右側の裾が長くなります。歪度が負で絶対値が大きい分布は、左側の裾が長くなります。目安として、歪度が標準誤差の 2 倍より大きい場合は、対称分布からずれていると解釈します。
- *Kurtosis* (尖度). 観測値が中心点の周囲に群がっている度合いの指標。正規分布の場合、尖度の統計値は 0 です。尖度が正の場合、正規分布と比較して観測値が分布の中心あたりに多く群がっており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布と比較して厚くなります。尖度が負の場合、正規分布と比較して観測値の群がりが小さくなり、分布の極値までの両裾が厚くなります。急尖的分布の両裾は、正規分布と比較して薄くなります。

値はグループの中間点: データ内の値がグループの中間点にある場合 (例えば、30 代の人すべての年齢が 35 としてコード化されている場合)、グループ化される前の元データの中央値とパーセンタイルを推定するには、このオプションを選択します。

度数分布表のグラフ

図表の種類: 円グラフは、全体に対する部分の割合を表示します。円グラフの各分割は、1つのグループ化変数で定義された1グループに対応します。棒グラフでは、異なる値またはカテゴリーの度数が個別の棒として表示されるため、各カテゴリーを視覚的に比較することができます。ヒストグラムでも棒が表示されますが、等間隔のスケールに沿ってプロットされます。それぞれの棒の高さは、区間内に含まれる量的変数の値の度数に対応しています。ヒストグラムでは、分布の形状、中央、広がりが表示されます。ヒストグラム上に重ね合わせられた正規曲線により、データが正規分布しているかどうかを判断することができます。

図表の値: 棒グラフの場合、スケール軸に度数またはパーセントでラベルを付けることができます。

度数分布表の書式

表示順: 度数分布表は、データ内の実際の値、または値の度数（発生の度数）に従って、昇順と降順のどちらでも配列することができます。ただし、ヒストグラムまたはパーセンタイルを要求した場合は、度数分布表では変数が量的であると仮定されるため、変数の値が昇順で表示されます。

複数の変数: 複数の変数について統計テーブルを作成する場合、1つのテーブルですべての変数を表示することも（「**変数の比較**」）、変数ごとに統計テーブルを分けて表示する（「**変数ごとの分析**」）こともできます。

カテゴリー数の多いテーブルを抑制: 指定された数を超える値を持つ度数分布表を非表示にする場合は、このオプションを選択します。

第 3 章 記述統計

記述統計手続きでは、複数の変数の 1 変量の要約統計量が 1 つの表に表示され、標準化された値 (z 得点) が計算されます。変数は、平均値の大きさの順 (昇順または降順) にすることも、アルファベット順にすることも、変数の選択順 (デフォルト) にすることもできます。

z 得点を保存すると、これらの得点がデータ・エディターのデータに追加されて、図表、データの一覧表示、分析で使用できるようになります。変数が別々の単位で記録されている場合 (例えば、1 人当たりの国内総生産と識字率など)、 z 得点変換を行うことにより、変数の尺度が統一され、視覚的に比較しやすくなります。

例: データの各ケースに、各店員の売り上げの日次合計が数カ月間分収集した形で含まれている場合 (例えば、Bob、Kim、Brian ごとに 1 つのエントリー)、記述統計手続きにより、各店員の一日の平均売り上げを計算して、その結果を平均値が最も大きい値から最も小さい値の順に表示することができます。

統計: サンプル・サイズ、平均値、最小値、最大値、標準偏差、分散、範囲、合計、平均値の標準誤差、および尖度と歪度とそれらの標準誤差。

記述統計データの考慮事項

データ: 数値型変数をグラフを使用してスクリーニングしてから、この数値型変数を使用して、誤差、外れ値、分布の異常を記録します。「記述統計」手続きは、サイズの大きいデータ・ファイル (千件単位のケース) を扱う場合に非常に便利です。

仮定: 使用可能な統計量 (z 得点を含む) のほとんどは、通常の理論に基づいており、対称的な分布の量的変数 (間隔または比率尺度) に適しています。順序付けされていないカテゴリや非対称分布を持つ変数は使用しないでください。 z 得点の分布は元のデータと同じ形状であるため、 z 得点を計算しても、データの問題点が解消されるわけではありません。

記述統計量分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「記述統計」 > 「記述統計...」

2. 1 つ以上の変数を選択します。

必要に応じて、以下の操作を実行することができます。

- 「標準化された値を変数として保存」を選択して z 得点を新しい変数として保存する。
- 「オプション」をクリックしてその他の統計量や表示順を選択する。

記述統計のオプション

平均値および合計: デフォルトでは、平均値 (算術平均) が表示されます。

散らばり: データの広がりまたは偏差を測定する統計には、標準偏差、分散、範囲、最小値、最大値、平均値の標準誤差があります。

- **標準偏差:** 平均値の周りの散らばりの指標。正規分布では、平均から 1 標準偏差以内にケースの 68% が含まれ、2 標準偏差以内にケースの 95% が含まれます。例えば平均年齢が 45 で標準偏差が 10 である場合、正規分布ではケースの 95% が 25 と 65 の間に含まれます。
- **Variance (分散 (信頼性分析)):** 平均値の周りの値の散らばりの指標。平均値からの偏差の平方和を、ケース数より 1 少ない値で割ったものに等しくなります。分散の測定単位は、変数自体の単位の 2 乗です。
- **Range (OK (ファイルオープン時のオプション)):** 数値変数の最大値と最小値の差。最大値から最小値を引いた値。
- **Minimum (最小値):** 数値変数の最小値。
- **Maximum (最大):** 数値変数の最大値。
- **標準誤差の平均値:** 同じ分布から抽出したサンプルの間で平均値がどの程度異なるかを示す指標。観測した平均と仮説による値をおおまかに比較するために使用することができます (差と標準誤差の比率が -2 より小さいか +2 より大きい場合は、2 つの値が異なっていると結論付けることができます)。

分布: 尖度と歪度は、分布の形状や対称の特性を示す統計量です。これらの統計量は、その標準誤差とともに表示されます。

- **Kurtosis (尖度):** 観測値が中心点の周囲に群がっている度合いの指標。正規分布の場合、尖度の統計値は 0 です。尖度が正の場合、正規分布と比較して観測値が分布の中心あたりに多く群がっており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布と比較して厚くなります。尖度が負の場合、正規分布と比較して観測値の群がりが小さくなり、分布の極値までの両裾が厚くなります。急尖的分布の両裾は、正規分布と比較して薄くなります。
- **Skewness (歪度):** 分布の非対称性の指標。正規分布は対称であり、歪度の値は 0 です。歪度が正の大きな値である分布は、右側の裾が長くなります。歪度が負で絶対値が大きい分布は、左側の裾が長くなります。目安として、歪度が標準誤差の 2 倍より大きい場合は、対称分布からずれていると解釈します。

表示順: デフォルトでは、変数は選択された順に表示されます。オプションで、アルファベットの昇順または降順で変数を表示することができます。

DESCRIPTIVES コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 一部の変数について、標準化されたスコア (z スコア) を保存する (VARIABLES サブコマンドを使用)。
- 標準化されたスコアを格納する新しい変数の名前を指定する (VARIABLES サブコマンドを使用)。
- 任意の変数について、値が指定されていない分析ケースから除外する (MISSING サブコマンドを使用)。
- 平均値だけでなく統計値も使用して、変数の表示順を並べ替える (SORT サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 4 章 探索

「探索的分析」手続きは、すべてのケース、またはケースの個別のグループについて、要約統計量と図形表示を作成します。「探索的分析」手続きには、データ・スクリーニング、外れ値の識別、記述統計、仮説の検定、下位母集団 (ケースのグループ) 間の差異の特徴付けなど、多くの用途があります。データ・スクリーニングでは、異常値、極値、データ内のギャップなどの特性が存在するかどうかを示すことができます。データの探索的分析により、データ分析で使用する統計手法が適切なものであるかどうかを判断することができます。探索的分析では、正規分布を必要とする手法を実行する際に、データ変換が必要になる場合があります。あるいは、ノンパラメトリック検定が必要になる場合もあります。

例: 4 種類の強化計画において、ネズミの迷路学習時間の分布を検討するとします。4 つのグループのそれぞれについて、時間の分布が近似正規分布になっているかどうか、4 つの分散が等しいかどうかを確認することができます。また、学習時間の長さについて、上位 5 つのケースと下位 5 つのケースを特定することもできます。箱ひげ図と幹葉図は、グループそれぞれの学習時間の分散を図示して要約します。

統計と作図: 平均値、中央値、5% トリム平均値、標準誤差、分散、標準偏差、最小値、最大値、範囲、4 分位範囲、歪度と尖度およびその標準誤差、平均値の信頼区間 (および指定した信頼度レベル)、パーセントイル、Huber の M 推定量、Andrews のウェーブ推定量、Hampel の M 推定量、Tukey のバイウエイト推定量、5 つの最大値と 5 つの最小値、正規性を検定するための Lilliefors の有意確率と Kolmogorov-Smirnov の統計量、および Shapiro-Wilk の統計量。箱ひげ図、幹葉図、ヒストグラム、正規性プロット、および Levene 検定と変換による水準と広がり の図。

探索的分析データの考慮事項

データ: 「探索的分析」手続きは、量的変数 (区間または比率尺度の測定) で使用することができます。因子変数 (データをケースのグループに分割する場合に使用) は、適切な数の異なる値 (カテゴリ) を持っている必要があります。これらの値は、短い文字型にすることも、数値にすることもできます。箱ひげ図の外れ値を示すために使用されるケースのラベル変数は、短い文字型にすることも、長い文字型 (最初の 15 バイト) にすることも、数値にすることもできます。

仮定: データの分布は、対称分布や正規分布になっている必要はありません。

データの探索的分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「記述統計」 > 「探索的...」

2. 1 つ以上の従属変数を選択します。

必要に応じて、以下の操作を実行することができます。

- 1 つ以上の因子変数を選択する (この変数の値により、ケースのグループが定義されます)。
- ケースにラベルを付けるための識別変数を選択する。
- 「統計」をクリックして、頑健推定量、外れ値、パーセントイル、度数分布表を使用する。
- 「作図」をクリックして、ヒストグラム、正規確率プロットと検定、Levene の統計による水準と広がり の図を使用する。
- 「オプション」をクリックして、欠損値の処理を指定する。

- **平方根の逆数:** 各データ値に対して、平方根の逆数が計算されます。
- **逆数:** 各データ値の逆数が計算されます。
- **平方根:** 各データ値の平方根が計算されます。
- **2 乗:** 各データ値が 2 乗されます。
- **3 乗:** 各データ値が 3 乗されます。

探索的分析のオプション

欠損値: 欠損値の処理を制御します。

- **リストごとに除外:** 従属変数または因子変数に対する欠損値があるケースは、すべての分析から除外されます。これはデフォルトです。
- **ペアごとに除外:** グループ (セル) 内の変数に対する欠損値がないケースが、そのグループの分析に含まれます。これらのケースには、他のグループで使用される変数に対する欠損値を持つケースが含まれている場合もあります。
- **欠損値を出力:** 因子変数の欠損値は、個別のカテゴリとして処理されます。このカテゴリに対して、すべての出力が作成されます。度数分布表には、欠損値のカテゴリが含まれます。因子変数の欠損値は分布には含まれますが、欠損値としてラベル付けされます。

EXAMINE コマンドの追加機能

探索手続きは、EXAMINE コマンド・シンタックスを使用します。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 因子変数によって定義されたグループの出力とプロットのほかに、合計の出力とプロットを要求する (TOTAL サブコマンドを使用)。
- 箱ひげ図のグループに対して共通スケールを指定する (SCALE サブコマンドを使用)。
- 因子変数の交互作用を指定する (VARIABLES サブコマンドを使用)。
- デフォルト以外のパーセンタイルを指定する (PERCENTILES サブコマンドを使用)。
- 5 種類のいずれかの方法に従ってパーセンタイルを計算する (PERCENTILES サブコマンドを使用)。
- 水準と広がり of 図に対して任意のべき乗変換を指定する (PLOT サブコマンドを使用)。
- 表示する極値の数を指定する (STATISTICS サブコマンドを使用)。
- 場所の M 推定量と頑強な推定量のパラメーターを指定する (MESTIMATORS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 5 章 クロス集計表

「クロス集計表」手続きは、2次元表と多次元表を作成し、2次元表のさまざまな検定と連関の測定を提供します。表の構造と、カテゴリーが順序付けされているかどうかにより、使用される検定と測定が決まります。

クロス集計表の統計量と連関の測定は、2次元表についてのみ計算されます。行、列、層の因子 (制御変数) を指定すると、「クロス集計表」手続きは、層因子のそれぞれの値 (または 2 つ以上の制御変数の値の組み合わせ) について、関連する統計量と測定方法から構成された 1 枚のパネルを作成します。例えば、性別が、人生観 (「人生は楽しい」、「日常的」、または「つまらない」と結婚経験 (「はい」または「いいえ」) のクロス表の層因子である場合、女性の 2次元表の結果と男性の 2次元表の結果はそれぞれ別に計算され、交互に並んだパネルとして表示されます。

例: 教育やコンサルティングなどのサービス販売において、中小企業の顧客は、大企業の顧客よりも収益を生む可能性が高いでしょうか。クロス集計により、中小企業 (従業員 500 人未満) の多くが高いサービス収益を生むのに対して、大企業 (従業員 2,500 人以上) の多くのサービス収益は低いことがわかります。

統計量と連関の測定方法: Pearson のカイ 2 乗、尤度比カイ 2 乗、線型と線型による連関検定、Fisher の直接法、Yates の修正カイ 2 乗、Pearson の r 、Spearman のロー、分割係数、ファイ、Cramér の V 、対称および非対称ラムダ、Goodman と Kruskal のタウ、不確定性係数、ガンマ、Somers の d 、Kendall のタウ b 、Kendall のタウ c 、イータ係数、Cohen のカッパ、相対リスク推定値、オッズ比、McNemar 検定、Cochran 統計量と Mantel-Haenszel 統計量、列比率の統計量。

クロス集計表データの考慮事項

データ: 各表変数のカテゴリーを定義するには、数値型変数または短い文字型変数 (8 バイト以下) の値を使用します。例えば性別の場合、1 と 2、または男性 と女性 としてデータをコード化することができます。

仮定: 一部の統計値と測定方法には、統計値に関するセクションで説明したように、順序付けされたカテゴリー (順序データ) または量的な値 (区間データまたは比率データ) を前提とするものがあります。その他の統計値と測定方法は、順序付けされていないカテゴリー (名義データ) を表変数が持っている場合に有効になります。カイ 2 乗に基づく統計値 (ファイ、Cramer の V 、分割係数) の場合、データは多項分布から無作為に抽出されたサンプルでなければなりません。

注: 順序変数は、カテゴリーを表す数値型コード (例: 1 = 低、2 = 中、3 = 高) にすることも、文字列値にすることもできます。ただし、カテゴリーの実際の順序を反映させるため、文字列値のアルファベット順が想定されます。例えば、低、中、高 という値を持つ文字列変数の場合、カテゴリーの順序は高、中、低として解釈されますが、これは正しい順序ではありません。一般に、順序データを表す場合は、数値型コードを使用した方が信頼性が高くなります。

クロス集計を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「記述統計」 > 「クロス集計表...」

2. 1 つ以上の行変数と 1 つ以上の列変数を選択します。

必要に応じて、以下の操作を実行することができます。

- 1 つ以上の制御変数を選択する。
- 「統計」をクリックして、2 次元表または副表の検定と関連の測定方法を選択する。
- 「セル」をクリックして、観測値と期待値、パーセント、残余を指定する。
- 「書式」をクリックして、カテゴリーの順序を制御する。

クロス集計表の層

1 つ以上の層変数を選択すると、それぞれの層変数 (制御変数) のカテゴリーごとに個別のクロス集計が作成されます。例えば、行変数と列変数をそれぞれ 1 つずつ選択し、2 つのカテゴリーを持つ層変数を 1 つ選択した場合、この層変数の各カテゴリーに対して 2 次元表が 1 つずつ作成されます。制御変数の別の層を作成する場合は、「次へ」をクリックします。最初に第 1 の各層変数のカテゴリーの各組み合わせに対してサブテーブルが作成され、次に第 2 の各層変数のカテゴリーの各組み合わせに対してサブテーブルが作成されます (3 番目以降についても同様です)。統計と関連度を要求した場合、2 次元表にだけ統計と関連度が適用されます。

クロス集計表のクラスター棒グラフ

クラスター棒グラフの表示: クラスター棒グラフを使用して、データをケースのグループに要約することができます。「行」の下で指定した変数のそれぞれの値について、棒のクラスターが 1 つ存在します。「列」の下で指定した変数が、各クラスター内で棒を定義する変数になります。この変数のそれぞれの値について、異なる色または異なるパターンの棒のグループが 1 つ存在します。「列」または「行」の下で複数の変数を指定すると、2 つの変数の各組み合わせに対して、クラスター棒グラフが生成されます。

テーブル層に層変数を表示するクロス集計表

テーブル層に層変数を表示する: 層変数 (制御変数) をクロス集計表のテーブル層として表示することができます。これにより、層変数のカテゴリーをドリルダウンできるだけでなく、行変数と列変数の全体的な統計を表示するビューを作成することもできます。

データ・ファイル *demo.sav* (インストール・ディレクトリーの *Samples* ディレクトリーにあります) を使用する場合の例を以下に示します。取得方法は以下のとおりです。

1. 行変数として「世帯全体の収入カテゴリー (千ドル) [収入カテゴリー]」を選択し、列変数として「携帯情報端末 [携帯端末]」を選択し、「教育のレベル [教育]」を層変数として選択します。
2. 「テーブル層に層変数を表示する」を選択します。
3. 「セル表示」サブダイアログで「列」を選択します。
4. クロス集計手続きを実行し、クロス集計表をダブルクリックして、「学歴」ドロップダウン・リストから「大学」を選択します。

クロス集計表の選択されたビューに、大学卒の学歴を持つ回答者の統計が表示されます。

クロス集計表の統計

カイ 2 乗: 2 つの行と 2 つの列から構成される表の場合、「カイ 2 乗」を選択して、Pearson のカイ 2 乗、尤度比カイ 2 乗、Fisher の直接法、Yates の修正カイ 2 乗 (連続性のための修正) を計算します。2 × 2 表の場合、大規模な表の欠損行と欠損列を持たない表で、セルの期待度数が 5 よりも小さいときには、Fisher の直接法が計算されます。他のすべての 2 × 2 表については、Yates の修正カイ 2 乗が計算さ

れます。任意の数の行と列から構成される表の場合、「**カイ 2 乗**」を選択して、Pearson のカイ 2 乗と尤度比カイ 2 乗を計算します。どちらの表変数も量的変数である場合、「**カイ 2 乗**」により、線型と線型による連関検定が得られます。

相関: 順序付けられた値が行と列の両方に含まれている表の場合、「**相関**」により、Spearman の相関係数とロー (数値データのみ) が得られます。Spearman のローは、ランク順位間の関連度です。どちらの表変数 (因子) も量的変数である場合、「**相関**」により、Pearson の相関係数、 r 、変数間の線型による関連度が得られます。

名義: 名義データ (カトリック、プロテスタント、ユダヤなど、固有の順序関係のないデータ) には、「**分割係数**」、「**ファイ (係数) と Cramér の V**」、「**ラムダ**」(対称ラムダと非対称ラムダ、Goodman と Kruskal のタウ)、「**不確定性係数**」を選択することができます。

- **分割係数:** カイ 2 乗に基づく関連度。値の範囲は 0 から 1 までです。値が 0 の場合は行変数と列変数の間に関連がないことを示し、値が 1 に近い場合は変数の間に強い関連があることを示します。取り得る最大値は、テーブルの行と列の数によって決まります。
- **Phi and Cramer's V (標準偏差 (グラフの集計関数)).** ファイは、カイ 2 乗に基づく関連度の指標です。カイ 2 乗統計量を標本サイズで割った結果の平方根を取ります。Cramer の V は、カイ 2 乗に基づく関連度です。
- **Lambda (ラムダ).** 従属変数の値を予測するために独立変数の値を使用するときの、誤差の削減比を反映する関連度。値が 1 の場合は、独立変数によって従属変数が完全に予測されます。値が 0 の場合、独立変数で従属変数を予測することはできません。
- **不確定性係数:** 変数の値を予測するために別の変数の値を使用するときの予測連関指数を示す関連度。例えば値 0.83 は、一方の変数が判明すると、もう一方の変数の値を予測するときに誤差が 83% 低下することを示します。本プログラムは、不確定性係数の対称版と非対称版の両方を計算します。

順序: 順序付けられた値が行と列の両方に含まれている表の場合、「**ガンマ**」(2 次元表の場合は 0 次、3 次元表から 10 次元表の場合は条件付き)、「**Kendall のタウ b**」、および「**Kendall のタウ c**」を選択します。行カテゴリーから列カテゴリーを予測する場合は、「**Somers の d**」を選択します。

- **Gamma (ガンマ).** 2 つの順序変数の間の対称な連関度。範囲は -1 から 1 の間です。値の絶対値が 1 に近い場合は、2 つの変数の間に強い関係があることを示します。値が 0 に近い場合は、関係が弱いかもしれないことを示します。2 次元テーブルの場合は 0 次ガンマを表示します。3 次元テーブルから n 次元テーブルの場合は、条件付きのガンマを表示します。
- **Somers' d (Somers の d).** 2 つの順序変数の間の関連度。範囲は -1 から 1 までです。値の絶対値が 1 に近い場合は、2 つの変数の間に強い関係があることを示します。値が 0 に近い場合は、2 つの変数の間にほとんど (またはまったく) 関係がないことを示します。Somers の d はガンマを非対称に拡張したものであり、独立変数の同順位でないペアの数を含める点だけが異なります。この統計量の対称版も計算します。
- **Kendall's tau-b (Kendall のタウ b).** 同順位を考慮に入れた、順序変数 (ランク付けされた変数) のノンパラメトリックな相関度。係数の符号は関係の方向を示します。絶対値は強さを示し、絶対値が大きいほど関係が強いことを示します。値の範囲は -1 から 1 までですが、-1 または +1 の値が得られるのは平方表の場合のみです。
- **Kendall's tau-c (Kendall のタウ c).** 同順位を無視した、順序変数のノンパラメトリックな関連度。係数の符号は関係の方向を示します。絶対値は強さを示し、絶対値が大きいほど関係が強いことを示します。値の範囲は -1 から 1 までですが、-1 または +1 の値が得られるのは平方表の場合のみです。

間隔尺度の名義: 一方の変数がカテゴリー変数で、もう一方の変数が量的変数の場合は、「**イータ**」を選択します。カテゴリー変数は数値でコード化されている必要があります。

- **イータ:** 関連度の 1 つ。値の範囲は 0 から 1 までです。値が 0 の場合は行変数と列変数の間に関連がないことを示し、値が 1 に近い場合は関連度が高いことを示します。イータは、間隔尺度で測定される従属変数 (収入など) とカテゴリーの少ない独立変数 (性別など) に適しています。イータ値は 2 つ計算されます。1 つは行変数を間隔変数として扱うものであり、もう 1 つは列変数を間隔変数として扱うものです。

Kappa (カッパ (因子分析)). Cohen のカッパは、2 人の評価者が同じ対象を評価するときに、両者による評価の一致度を測定します。値 1 は完全な一致を示します。値 0 は、偶然以外の一致がないことを示します。カッパは、行および列の値が同じスケールを表す平方表に基づきます。1 つの変数に観測値があるがそれ以外の変数に観測値がないセルには 0 の度数を割り当てます。2 つの変数のデータ・ストレージ・タイプ (文字列か数値か) が同じでない場合は、カッパは計算されません。文字列変数の場合は、両方の変数を同じ文字数で定義している必要があります。

リスク: 2 x 2 表の場合の、因子の存在とイベントの発生間の関連の強さの指標。統計量の信頼区間に値 1 が含まれる場合、その因子がそのイベントに関連していると想定することはできません。因子の発生がまれな場合には、オッズ比を推定値または相対リスクとして使用することができます。

McNemar: 関連する 2 つの二分変数を対象とするノンパラメトリック検定。カイ 2 乗分布を使用して応答の変化を検定します。「前後」デザインでの実験的介入による応答の変化を検出する場合に有用です。大きな平方表の場合は、McNemar-Bowker 対称検定が報告されます。

Cochran's and Mantel-Haenszel statistics (Cochran 統計量と Mantel-Haenszel 統計量). Cochran と Mantel-Haenszel の統計量は、1 つ以上の層 (制御) 変数によって定義された共変量パターンを条件として 2 値因子変数と 2 値応答変数の間の独立性を検定するために使用します。他の統計量は層ごとに計算しますが、Cochran と Mantel-Haenszel の統計量はすべての層を対象にして一度に計算します。

クロス集計表のセル表示

ユーザーが有意なカイ 2 乗検定に寄与するパターンをデータ内で検出できるように、「クロス集計表」手続きは、期待度数を表示し、観測度数と期待度数との差を測定する 3 種類の残差 (偏差) を表示します。表の各セルには、度数、パーセンテージ、残差を任意に組み合わせる表示することができます。

度数: 実際の観測ケース数と期待ケース数 (行変数と列変数が互いに独立している場合)。ここで指定した整数よりも小さな度数を非表示にすることができます。非表示の値は <Nとして表示されます。N は、指定する整数です。2 以上の整数を指定する必要があります。値として 0 を指定することもできますが、その場合、非表示になる度数はありません。

列の割合を比較: このオプションは、列比率のペアごとの比較を計算し、有意差のある列のペア (特定の行) を示します。有意差は、下付き文字を使用する APA スタイル形式のクロス集計表に表示され、0.05 の有意確率で計算されます。注: 観測度数または列パーセントを選択せずにこのオプションを指定すると、観測度数がクロス集計表に組み込まれ、列比率の検定結果が APA スタイルの下付き文字で示されます。

- **p 値の調整 (Bonferroni 法):** 列比率のペアごとの比較では、Bonferroni 補正が使用されます。この補正により、観測された有意水準が、複数の比較が実行されたという事実に対して調整されます。

パーセンテージ: パーセンテージにより、行全体または列全体を合計することができます。表 (1 つの層) に表示されるケースの合計数のパーセンテージも選択することができます。注: 「度数」グループで「小さい度数を非表示にする」が選択されている場合、非表示の度数に関連するパーセンテージも非表示になります。

「残差」。標準化されていない未加工の残差により、観測値と期待値との間の差分が示されます。標準化された残差と調整済みの標準化された残差も選択することができます。

- **標準化されていない:** 観測値と期待値の差。ここでの期待値は、2 つの変数の間に関係がないと仮定した場合に期待されるセルのケース数です。残差が正の場合は、行変数と列変数が独立であると仮定した場合のセルのケース数より実際のケース数が多いことを示します。
- **標準化:** 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。平均値は 0 であり、標準偏差は 1 です。
- **調整済みの標準化:** セルの残差 (観測値 - 期待値) をその標準誤差の推定値で割った値。結果として得られる標準化残差は、平均より上または下に標準偏差の何倍離れているかで表されます。

非整数値の重み付け: セル度数は、各セル内のケースの数を表すため、通常は整数値になります。ただし、小数値を持つ重み付け変数 (1.25 など) によってデータ・ファイルに重みが付けられている場合、セル度数も小数値になることがあります。セル度数の計算前または計算後に値の切り捨てや丸めを行ったり、小数値のセル度数を表の表示と統計計算の両方で使用したりすることができます。

- *Round cell counts* (丸めセル度数). ケース重みをそのまま使用しますが、セルの累積重みを丸めてから統計量を計算します。
- *Truncate cell counts* (切り捨てセル度数). ケース重みをそのまま使用しますが、セルの累積重みを切り捨ててから統計量を計算します。
- *Round case weights* (丸めケース重み). 使用前にケースの重みを丸めます。
- *Truncate case weights* (切り捨てケース重み). 使用前にケースの重みを切り捨てます。
- なし: ケースの重みを現状のまま使用し、小数のセル度数を使用します。ただし、正確統計量 (「正確確率検定」オプションでのみ使用可能) が要求された場合は、正確確率検定統計量を計算する前に、セルの累積重みを切り捨てるか丸めます。

クロス集計表の表書式

行は、行変数の値の昇順または降順に整列させることができます。

第 6 章 要約

「ケースの要約」手続きは、1 つ以上のグループ化変数のカテゴリ内の変数に対するサブグループ統計量を計算します。グループ化変数のすべてのレベルがクロス集計されます。統計量の表示順を選択することができます。カテゴリ全体における各変数の要約統計量も表示されます。各カテゴリのデータ値は、リスト表示することも、非表示にすることもできます。サイズの大きなデータ・セットの場合、最初の n 件のケースだけをリスト表示することができます。

例: 地区と顧客業種別の平均製品売上高とはどのようなものでしょうか。例えば、西部地区の企業の顧客からは最高の平均売上高が得られていることから、西部地区の平均売上高は他の地区に比べてわずかに高いことがわかります。

統計: 合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、合計 N のパーセント、グループ化変数内の合計のパーセント、グループ化変数内の N のパーセント、幾何平均、調和平均。

ケースの要約データの考慮事項

データ: グループ化変数は、値が数値型または文字型のカテゴリ変数です。カテゴリの数はある程度少なくする必要があります。他の変数は、ランク付け可能な変数にする必要があります。

仮定: オプションの一部のサブグループ統計量には、平均値や標準偏差などのように、通常理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。頑健な統計量 (中央値や範囲など) は、正規性の仮定を満たしているかどうかにかかわらず、量的変数に適しています。

ケースの要約を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「報告書」 > 「ケースの要約...」

2. 1 つ以上の変数を選択します。

必要に応じて、以下の操作を実行することができます。

- 1 つ以上のグループ化変数を選択して、データをサブグループに分割する。
- 「オプション」をクリックして、出力の表題の変更、出力の下に表示する解説の追加、または欠損値を持つケースの除外を行う。
- 「統計」をクリックして、オプションの統計量を使用する。
- 「ケースの表示」をクリックして、各サブグループのケースをリストする。デフォルトでは、ファイル内の最初の 100 ケースだけが表示されます。「ケースの制限は最初の n 」の値を変更したり、項目の選択を解除してすべてのケースをリストすることができます。

ケースの要約のオプション

「ケースの要約」では、出力の表題の変更や、出力テーブルの下に表示される解説の追加を行うことができます。テキスト内で改行を挿入したい場所に $\$n$ を入力すると、表題や解説の行折り返しを制御することができます。

合計の副見出しの表示と非表示を切り替えたり、分析で使用するいずれかの変数に対して欠損値を持つケースを含めるか除外するかを選択することもできます。通常、欠損値を持つケースを出力する場合、ピリオドやアスタリスクで示すことをお勧めします。その場合、欠損値用に表示する文字、語句、またはコードを入力します。入力しなかった場合は、出力時に欠損値を持つケースに対して特別な処理は行われません。

ケースの要約の統計

各グループ化変数のカテゴリ内の変数に対するサブグループ統計量としては、合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、合計 N のパーセント、グループ化変数内の合計のパーセント、グループ化変数内の N のパーセント、幾何平均、調和平均のうちの 1 つ以上を選択することができます。統計量は、「セル統計量」リストに表示されている順序で、出力に表示されます。カテゴリ全体の各変数についても、要約統計量が表示されます。

First (最初). データ・ファイルで最初に検出されたデータ値を表示します。

Geometric Mean (幾何平均). データ値の積の n 乗根。ここで、 n はケースの数を表します。

Grouped Median (グループ中央値). グループにコード化される、データについて計算した中央値。例えば、年齢データを使用しており、30 代のそれぞれの値を 35 にコード化し、40 代のそれぞれの値を 45 にコード化し、他の年齢層についても同様にコード化する場合、グループ化した中央値は、コード化したデータから計算される中央値です。

Harmonic Mean (調和平均). グループ間で標本サイズが等しくない場合に、グループの平均サイズを推定するために使用します。調和平均は、サンプルの総数を標本サイズの逆数の和で割ったものです。

Kurtosis (尖度). 観測値が中心点の周囲に群がっている度合いの指標。正規分布の場合、尖度の統計値は 0 です。尖度が正の場合、正規分布と比較して観測値が分布の中心あたりに多く群がっており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布と比較して厚くなります。尖度が負の場合、正規分布と比較して観測値の群がりが小さくなり、分布の極値までの両裾が厚くなります。急尖的分布の両裾は、正規分布と比較して薄くなります。

Last (最後). データ・ファイルで最後に検出されたデータ値を表示します。

Maximum (最大). 数値変数の最大値。

Mean (平均). 中心傾向の指標。算術平均 (合計をケース数で割った値) です。

Median (中央値). この値より上と下それぞれにケースの半数ずつが該当することになる値。50 パーセントイル。ケース数が偶数の場合の中央値は、昇順または降順にソートしたときに中央に来る 2 つのケースの平均です。中央値は、外れ値に対して敏感でない、中心傾向の指標です。それに対して平均値は、少数の極端に大きいまたは小さい値に影響されることがあります。

Minimum (最小値). 数値変数の最小値。

N . ケース (観測値またはレコード) の数。

総数のパーセント: 各カテゴリのケースの総数の割合。

総合計のパーセント: 各カテゴリの総合計の割合。

Range (OK (ファイルオープン時のオプション)). 数値変数の最大値と最小値の差。最大値から最小値を引いた値。

Skewness (歪度). 分布の非対称性の指標。正規分布は対称であり、歪度の値は 0 です。歪度が正の大きな値である分布は、右側の裾が長くなります。歪度が負で絶対値が大きい分布は、左側の裾が長くなります。目安として、歪度が標準誤差の 2 倍より大きい場合は、対称分布からずれていると解釈します。

Standard Deviation (標準偏差). 平均値の周りの散らばりの指標。正規分布では、平均から 1 標準偏差以内にケースの 68% が含まれ、2 標準偏差以内にケースの 95% が含まれます。例えば平均年齢が 45 で標準偏差が 10 である場合、正規分布ではケースの 95% が 25 と 65 の間に含まれます。

Standard Error of Kurtosis (尖度の標準誤差). 標準誤差に対する尖度の比率は、正規性の検定として使用することができます (比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。尖度が大きな正の値である場合は、分布の裾が正規分布の裾より長いことを示します。尖度が負の値である場合は、裾が短いことを示します (箱形の一様分布に似た形になります)。

Standard Error of Mean (平均値の標準誤差). 同じ分布から抽出したサンプルの間で平均値がどの程度異なるかを示す指標。観測した平均と仮説による値をおおまかに比較するために使用することができます (差と標準誤差の比率が -2 より小さいか +2 より大きい場合は、2 つの値が異なっていると結論付けることができます)。

Standard Error of Skewness (歪度の標準誤差). 標準誤差に対する歪度の比率は、正規性の検定として使用することができます (比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度が大きな正の値である場合は、右側の裾が長いことを示します。極端な負の値の場合は、左側の裾が長いことを示します。

Sum (合計). 欠損値でない値を持つすべてのケースにわたる値の和 (合計)。

Variance (分散 (信頼性分析)). 平均値の周りの値の散らばりの指標。平均値からの偏差の平方和を、ケース数より 1 少ない値で割ったものに等しくなります。分散の測定単位は、変数自体の単位の 2 乗です。

第 7 章 平均値

「平均値」手続きは、1 つ以上の独立変数のカテゴリ内の従属変数を対象として、サブグループの平均と関連する 1 変量の統計量を計算します。オプションで、一元配置分散分析、イータ、線型性の検定を実行することができます。

例: 3 種類の食用油について、それぞれの平均脂肪吸収量を測定し、一元配置分散分析を実行して、平均値が異なるかどうかを確認します。

統計: 合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、合計 N のパーセント、グループ化変数内の合計のパーセント、グループ化変数内の N のパーセント、幾何平均、調和平均。オプションには、分散分析、イータ、イータ 2 乗、線型性 R および R^2 の検定があります。

平均値データの考慮事項

データ: 従属変数は量的変数で、独立変数はカテゴリ変数です。カテゴリ変数の値は、数値にすることも、文字列にすることもできます。

仮定: オプションの一部のサブグループ統計量には、平均値や標準偏差などのように、通常理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。頑健な統計量 (中央値など) は、正規性の仮定を満たしているかどうかにかかわらず、量的変数に適しています。分散分析は正規性からの逸脱には頑健ですが、各セルのデータは対称でなければなりません。分散分析では、等しい分散を持つ母集団から各グループが取り出されているということも想定されます。この想定を検定するには、「一元配置分散分析」手続きで使用できる Levene の等分散性の検定を使用します。

サブグループの平均を求めるには

1. メニューから次の項目を選択します。

「分析」 > 「平均の比較」 > 「グループの平均...」

2. 1 つ以上の従属変数を選択します。

3. 以下の方法のいずれかを使用して、カテゴリ独立変数を選択します。

- 1 つ以上の独立変数を選択します。独立変数ごとに、結果が個別に表示されます。
- 独立変数の層を 1 つ以上選択します。各層では、サンプルがさらに細分割されます。層 1 と層 2 のそれぞれに 1 つの独立変数がある場合、独立変数ごとに個別の表が作成されるのではなく、1 つのクロス表に結果が表示されます。

4. オプションの統計量、分散分析表、イータ、イータの 2 乗、 R 、 R^2 を使用する場合は、「オプション」をクリックします。

グループの平均のオプション

各グループ化変数の各カテゴリ内の変数に対するサブグループ統計量としては、合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、合計 N のパーセント、グループ化変数内の合計のパーセン

ト、グループ化変数内の N のパーセント、幾何平均、調和平均のうちの 1 つ以上を選択することができます。サブグループの統計量の表示順は、変更することができます。統計量は、「セル統計量」リストに表示されている順序で、出力にも表示されます。カテゴリ全体の各変数についても、要約統計量が表示されます。

First (最初). データ・ファイルで最初に検出されたデータ値を表示します。

Geometric Mean (幾何平均). データ値の積の n 乗根。ここで、 n はケースの数を表します。

Grouped Median (グループ中央値). グループにコード化される、データについて計算した中央値。例えば、年齢データを使用しており、30 代のそれぞれの値を 35 にコード化し、40 代のそれぞれの値を 45 にコード化し、他の年齢層についても同様にコード化する場合、グループ化した中央値は、コード化したデータから計算される中央値です。

Harmonic Mean (調和平均). グループ間で標本サイズが等しくない場合に、グループの平均サイズを推定するために使用します。調和平均は、サンプルの総数を標本サイズの逆数の和で割ったものです。

Kurtosis (尖度). 観測値が中心点の周囲に群がっている度合いの指標。正規分布の場合、尖度の統計値は 0 です。尖度が正の場合、正規分布と比較して観測値が分布の中心あたりに多く群がっており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布と比較して厚くなります。尖度が負の場合、正規分布と比較して観測値の群がりが小さくなり、分布の極値までの両裾が厚くなります。急尖的分布の両裾は、正規分布と比較して薄くなります。

Last (最後). データ・ファイルで最後に検出されたデータ値を表示します。

Maximum (最大). 数値変数の最大値。

Mean (平均). 中心傾向の指標。算術平均 (合計をケース数で割った値) です。

Median (中央値). この値より上と下それぞれにケースの半数ずつが該当することになる値。50 パーセンタイル。ケース数が偶数の場合の中央値は、昇順または降順にソートしたときに中央に来る 2 つのケースの平均です。中央値は、外れ値に対して敏感でない、中心傾向の指標です。それに対して平均値は、少数の極端に大きいまたは小さい値に影響されることがあります。

Minimum (最小値). 数値変数の最小値。

N . ケース (観測値またはレコード) の数。

Percent of total N (総数のパーセント). 各カテゴリのケースの総数の割合。

Percent of total sum (総合計のパーセント). 各カテゴリの総合計の割合。

Range (OK (ファイルオープン時のオプション)). 数値変数の最大値と最小値の差。最大値から最小値を引いた値。

Skewness (歪度). 分布の非対称性の指標。正規分布は対称であり、歪度の値は 0 です。歪度が正の大きな値である分布は、右側の裾が長くなります。歪度が負で絶対値が大きい分布は、左側の裾が長くなります。目安として、歪度が標準誤差の 2 倍より大きい場合は、対称分布からずれていると解釈します。

Standard Deviation (標準偏差). 平均値の周りの散らばりの指標。正規分布では、平均から 1 標準偏差以内にケースの 68% が含まれ、2 標準偏差以内にケースの 95% が含まれます。例えば平均年齢が 45 で標準偏差が 10 である場合、正規分布ではケースの 95% が 25 と 65 の間に含まれます。

Standard Error of Kurtosis (尖度の標準誤差). 標準誤差に対する尖度の比率は、正規性の検定として使用することができます (比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。尖度が大きな正の値である場合は、分布の裾が正規分布の裾より長いことを示します。尖度が負の値である場合は、裾が短いことを示します (箱形の一様分布に似た形になります)。

Standard Error of Mean (平均値の標準誤差). 同じ分布から抽出したサンプルの間で平均値がどの程度異なるかを示す指標。観測した平均と仮説による値をおおまかに比較するために使用することができます (差と標準誤差の比率が -2 より小さいか +2 より大きい場合は、2 つの値が異なっていると結論付けることができます)。

Standard Error of Skewness (歪度の標準誤差). 標準誤差に対する歪度の比率は、正規性の検定として使用することができます (比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度が大きな正の値である場合は、右側の裾が長いことを示します。極端な負の値の場合は、左側の裾が長いことを示します。

Sum (合計). 欠損値でない値を持つすべてのケースにわたる値の和 (合計)。

Variance (分散 (信頼性分析)). 平均値の周りの値の散らばりの指標。平均値からの偏差の平方和を、ケース数より 1 少ない値で割ったものに等しくなります。分散の測定単位は、変数自体の単位の 2 乗です。

第 1 層の統計

分散分析表とイータ: 一元配置分散分析テーブルを表示し、最初の層にある各独立変数のイータおよびイータ 2 乗 (関連度) を計算します。

線型性の検定: 線型成分および非線型成分に関連した平方和、自由度、および平方平均のほか、F 比、R、および R² 乗を計算します。独立変数が短い文字列である場合は、線型性を計算しません。

第 8 章 OLAP キューブ

「OLAP (Online Analytical Processing) キューブ」手続きは、1 つ以上のカテゴリー・グループ化変数のカテゴリー内で、連続型集計変数の合計、平均値、その他の 1 変量の統計量を計算します。テーブル内には、各グループ化変数のカテゴリーごとに個別の層が作成されます。

例: 地域別売上の合計と平均、および地域内の製品別売上の合計と平均。

統計: 合計、ケース数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリーの変数値、グループ化変数の最後のカテゴリーの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、ケース合計のパーセント、集計合計のパーセント、グループ化変数内のケース合計のパーセント、グループ化変数内の集計合計のパーセント、幾何平均、調和平均。

OLAP キューブ・データの考慮事項

データ: 集計変数は量的変数 (区間尺度または比率尺度で測定される連続型変数) で、グループ化変数はカテゴリー変数です。カテゴリー変数の値は、数値にすることも、文字列にすることもできます。

仮定: オプションの一部のサブグループ統計量には、平均値や標準偏差などのように、通常理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。頑健な統計量 (中央値や範囲など) は、正規性の仮定を満たしているかどうかにかかわらず、量的変数に適しています。

OLAP キューブを取得するには

1. メニューから次の項目を選択します。

「分析」 > 「報告書」 > 「OLAP キューブ...」

2. 連続型の集計変数を 1 つ以上選択します。

3. カテゴリー・グループ化変数を 1 つ以上選択します。

必要に応じて、以下の操作を実行することができます。

- 別の要約統計量を選択する (「統計」をクリック)。要約統計量を選択するには、最初に 1 つ以上のグループ化変数を選択する必要があります。
- 変数のペアと、グループ化変数によって定義されるグループのペアの差分を計算する (「差分」をクリック)。
- カスタムのテーブル表題を作成する (「表題」をクリック)。
- 指定した整数よりも小さな度数を非表示にする。非表示の値は <Nとして表示されます。N は、指定する整数です。2 以上の整数を指定する必要があります。

OLAP キューブの統計

各グループ化変数の 1 つのカテゴリー内の集計変数に対して、1 つ以上のサブグループ統計量を指定することができます。指定できるサブグループ統計量には、合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリーの変数値、グループ化変数の最後のカテゴリーの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、ケ

ケース合計のパーセント、集計合計のパーセント、グループ化変数内のケース合計のパーセント、グループ化変数内の集計合計のパーセント、幾何平均、調和平均があります。

サブグループの統計量の表示順は、変更することができます。統計量は、「セル統計量」リストに表示されている順序で、出力にも表示されます。カテゴリー全体の各変数についても、要約統計量が表示されます。

First (最初). データ・ファイルで最初に検出されたデータ値を表示します。

Geometric Mean (幾何平均). データ値の積の n 乗根。ここで、 n はケースの数を表します。

Grouped Median (グループ中央値). グループにコード化される、データについて計算した中央値。例えば、年齢データを使用しており、30 代のそれぞれの値を 35 にコード化し、40 代のそれぞれの値を 45 にコード化し、他の年齢層についても同様にコード化する場合、グループ化した中央値は、コード化したデータから計算される中央値です。

Harmonic Mean (調和平均). グループ間で標本サイズが等しくない場合に、グループの平均サイズを推定するために使用します。調和平均は、サンプルの総数を標本サイズの逆数の和で割ったものです。

Kurtosis (尖度). 観測値が中心点の周囲に群がっている度合いの指標。正規分布の場合、尖度の統計値は 0 です。尖度が正の場合、正規分布と比較して観測値が分布の中心あたりに多く群がっており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布と比較して厚くなります。尖度が負の場合、正規分布と比較して観測値の群がりが小さくなり、分布の極値までの両裾が厚くなります。急尖的分布の両裾は、正規分布と比較して薄くなります。

Last (最後). データ・ファイルで最後に検出されたデータ値を表示します。

Maximum (最大). 数値変数の最大値。

Mean (平均). 中心傾向の指標。算術平均 (合計をケース数で割った値) です。

Median (中央値). この値より上と下それぞれにケースの半数ずつが該当することになる値。50 パーセンタイル。ケース数が偶数の場合の中央値は、昇順または降順にソートしたときに中央に来る 2 つのケースの平均です。中央値は、外れ値に対して敏感でない、中心傾向の指標です。それに対して平均値は、少数の極端に大きいまたは小さい値に影響されることがあります。

Minimum (最小値). 数値変数の最小値。

N. ケース (観測値またはレコード) の数。

Percent of N in (数のパーセント). 他のグループ化変数のカテゴリー内における、指定したグループ化変数に対応するケースの数の割合。グループ化変数が 1 つしかない場合、この値はケースの総数の割合と同じです。

合計のパーセント: 他のグループ化変数のカテゴリー内における、指定したグループ化変数に対応する合計の割合。グループ化変数が 1 つしかない場合、この値はケースの総合計の割合と同じです。

総数のパーセント: 各カテゴリーのケースの総数の割合。

総合計のパーセント: 各カテゴリーの総合計の割合。

Range (OK (ファイルオープン時のオプション)). 数値変数の最大値と最小値の差。最大値から最小値を引いた値。

Skewness (歪度). 分布の非対称性の指標。正規分布は対称であり、歪度の値は 0 です。歪度が正の大きな値である分布は、右側の裾が長くなります。歪度が負で絶対値が大きい分布は、左側の裾が長くなります。目安として、歪度が標準誤差の 2 倍より大きい場合は、対称分布からずれていると解釈します。

Standard Deviation (標準偏差). 平均値の周りの散らばりの指標。正規分布では、平均から 1 標準偏差以内にケースの 68% が含まれ、2 標準偏差以内にケースの 95% が含まれます。例えば平均年齢が 45 で標準偏差が 10 である場合、正規分布ではケースの 95% が 25 と 65 の間に含まれます。

Standard Error of Kurtosis (尖度の標準誤差). 標準誤差に対する尖度の比率は、正規性の検定として使用することができます (比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。尖度が大きな正の値である場合は、分布の裾が正規分布の裾より長いことを示します。尖度が負の値である場合は、裾が短いことを示します (箱形の一様分布に似た形になります)。

Standard Error of Mean (平均値の標準誤差). 同じ分布から抽出したサンプルの間で平均値がどの程度異なるかを示す指標。観測した平均と仮説による値をおおまかに比較するために使用することができます (差と標準誤差の比率が -2 より小さいか +2 より大きい場合は、2 つの値が異なっていると結論付けることができます)。

Standard Error of Skewness (歪度の標準誤差). 標準誤差に対する歪度の比率は、正規性の検定として使用することができます (比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度が大きな正の値である場合は、右側の裾が長いことを示します。極端な負の値の場合は、左側の裾が長いことを示します。

Sum (合計). 欠損値でない値を持つすべてのケースにわたる値の和 (合計)。

Variance (分散 (信頼性分析)). 平均値の周りの値の散らばりの指標。平均値からの偏差の平方和を、ケース数より 1 少ない値で割ったものに等しくなります。分散の測定単位は、変数自体の単位の 2 乗です。

OLAP キューブの差分

このダイアログ・ボックスでは、集計変数間のパーセントの差と算術的な差分を計算することも、グループ化変数によって定義されたグループ間のパーセントの差と算術的な差分を計算することもできます。差分は、「OLAP キューブ: 統計」ダイアログ・ボックスで選択されたすべての測定方法について計算されます。

変数間の差分: 変数のペア間の差分を計算します。各ペアにおいて、2 番目の変数 (「マイナス変数」) の要約統計量の値が、最初の変数の要約統計量の値から引かれます。パーセントの差を求める場合は、「マイナス変数」の集計変数の値が分母として使用されます。変数間の差分を指定するには、最初にメイン・ダイアログ・ボックスで 2 つ以上の集計変数を選択する必要があります。

ケースのグループ間の差分: グループ化変数によって定義されたグループのペア間の差分を計算します。各ペアにおいて、2 番目のカテゴリー (「マイナス・カテゴリー」) の要約統計量の値が、最初のカテゴリーの要約統計量の値から引かれます。パーセントの差を求める場合は、「マイナス・カテゴリー」の要約統計量の値が分母として使用されます。グループ間の差分を指定するには、最初にメイン・ダイアログ・ボックスで 1 つ以上のグループ化変数を選択する必要があります。

OLAP キューブの表題

出力の表題を変更したり、出力テーブルの下に表示される解説を追加したりすることができます。また、テキスト内で改行を挿入する場所に「¥n」と入力することにより、表題や解説の行折り返しを制御することもできます。

第 9 章 t 検定

T 検定

以下の 3 種類の t 検定を使用することができます。

独立したサンプルの t 検定 (2 サンプルの t 検定): 2 つのケースのグループについて 1 つの変数の平均値を比較します。各グループの記述統計量と等分散性の Levene の検定とともに、分散が等しい場合と等しくない場合の両方の t 値と平均値の差の 95% 信頼区間が得られます。

対応のあるサンプルの t 検定 (従属 t 検定): 1 つのグループについて 2 つの変数の平均値を比較します。この検定は、一致しているペアまたはケース・コントロール研究の計画用の検定でもあります。出力には、検定変数の記述統計量、変数間の相関係数、対応間の差の記述統計量、 t 検定、95% 信頼区間が表示されます。

1 サンプルの t 検定: 既知の値または仮説値を 1 つの変数の平均値と比較します。検定値の記述統計量が t 検定とともに表示されます。検定変数の平均値と仮説検定値との差の 95% 信頼区間は、デフォルトの出力として表示されます。

独立したサンプルの T 検定

「独立したサンプルの T 検定」手続きは、ケースの 2 つのグループの平均値を比較します。この検定の場合、被検者を無作為に 2 つのグループに割り当て、応答の差が他の要素によるものでなく、処置 (または処置の欠如) によるものになるようにするのが理想的な方法です。男性と女性の平均収入を比較するような場合は、この方法は当てはまりません。ある個人が男性または女性に無作為に割り当てられることはないためです。このような場合は、他の要素における差が平均値の有意差を隠したり拡大したりすることがないようにする必要があります。平均収入の差は、(性別だけではなく) 教育水準などの因子による影響を受ける可能性があります。

例: 高血圧の患者を偽薬グループと治療グループに無作為に割り当てます。偽薬の被検者には効き目のない錠剤を投与し、治療グループには血圧を下げるのが期待される新薬を投与します。被検者を 2 カ月間治療したあと、2 サンプル t 検定を使用して、偽薬グループと治療グループの平均血圧を比較します。各患者はそれぞれ 1 回の測定を受け、いずれかのグループに所属します。

統計: 各変数: サンプル・サイズ、平均値、標準偏差、平均値の標準誤差。平均値の差: 平均値、標準誤差、信頼区間 (信頼度レベルを指定できます)。検定: 等分散性の Levene 検定、2 つの母平均の差のプールされた分散および等分散でないときの t 検定。

独立したサンプルの T 検定データの考慮事項

データ: 対象となる量的変数の値が、データ・ファイルの 1 つの列に含まれます。この手続きは、2 つの値を持つグループ化変数を使用して、当該ケースを 2 つのグループに分割します。グループ化変数は、数値型変数 (1 と 2、6.25 と 12.5 などの値) にすることも、短い文字型変数 (「はい」と「いいえ」など) にすることもできます。別の方法として、分割点を指定することにより、年齢などの量的変数を使用して、ケースを 2 つのグループに分割することもできます (分割点を 21 にすると、年齢は 21 歳未満のグループと 21 歳以上のグループに分割されます)。

仮定: 等分散の t 検定の場合、観測値は、同じ母集団分散を持つ正規分布から無作為に抽出された独立したサンプルでなければなりません。等分散でない t 検定の場合、観測値は、正規分布から無作為に抽出された独立したサンプルでなければなりません。2 サンプルの t 検定は、正規性からの逸脱に対して非常に頑健です。分布を図示して確認する場合は、分布が対称的で、外れ値がないことを確認します。

独立したサンプルの t 検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「平均の比較」 > 「独立したサンプルの t 検定...

2. 1 つ以上の量的検定変数を選択します。変数ごとに個別の t 検定が計算されます。
3. グループ化変数を 1 つ選択してから「**グループの定義**」をクリックして、比較したいグループの 2 つのコードを指定します。
4. 必要に応じて「**オプション**」をクリックして、欠損データの処理方法と信頼区間のレベルを指定します。

独立したサンプルの T 検定のグループの定義

数値型のグループ化変数の場合、2 つの値または分割点を指定して、 t 検定を行う 2 つのグループを定義します。

- **特定の値を使用:** グループ 1 に値を入力し、グループ 2 に別の値を入力します。他の値を持つケースは分析から除外されます。数値は整数でなくてもかまいません (例えば、6.25 や 12.5 などでもかまいません)。
- **分割点:** グループ化変数の値を 2 つのグループに分割する数値を入力します。分割点未満の値を持つすべてのケースが一方のグループを形成し、分割点以上の値のケースを持つケースがもう一方のグループを形成します。

文字列のグループ化変数の場合、グループ 1 に文字列を入力し、グループ 2 に別の文字列を入力します (「はい」と「いいえ」など)。他の文字列を持つケースは分析から除外されます。

独立したサンプルの T 検定のオプション

信頼区間: デフォルトでは、平均値の差の 95% 信頼区間が表示されます。別の信頼度レベルが必要な場合は、1 から 99 までの値を入力します。

欠損値: 複数の変数を検定する際に、1 つ以上の変数に対してデータが欠損している場合、手続きでどのケースを含めるか (または除外するか) を指定することができます。

- **分析ごとに除外:** 各 t 検定は、検定対象の変数について有効なデータを持っているすべてのケースを使用します。そのため、サンプル・サイズが検定ごとに変化する場合があります。
- **リストごとに除外:** 各 t 検定は、要求された t 検定で使用されるすべての変数に対して有効なデータを持っているケースだけを使用します。そのため、サンプル・サイズはすべての検定で一定になります。

対応のあるサンプルの T 検定

「対応のあるサンプルの T 検定」手続きは、1 つのグループの 2 つの変数の平均を比較します。この手続きは、各ケースの 2 つの変数間の差を計算し、平均が 0 とは異なるかどうかを検定します。

例: 高血圧に関する調査で、調査の開始時にすべての患者を測定し、治療後に再度測定します。したがって、各被験者には、事前 測定値、事後 測定値と呼ばれる 2 つの測定値があります。あるいは、この検定が使用される別の計画として、一致するペアの研究またはケース・コントロール研究があります。この場

合、データ・ファイルの各レコードには、患者とその患者に一致する対照被験者の応答が記録されます。血圧の調査では、患者と対照被験者を、年齢で一致させることもできます (75 歳の患者と 75 歳の対照グループ・メンバー)。

統計: 各変数: 平均値、サンプル・サイズ、標準偏差、平均値の標準誤差。変数の各ペア: 相関関数、平均値の差の平均、 t 検定、平均値の差に対する信頼区間 (信頼度レベルを指定できます)。標準偏差と平均値の差の標準誤差。

対応のあるサンプルの T 検定データの考慮事項

データ: 対応のある検定ごとに、2 つの量的変数を指定します (区間尺度または比例尺度)。一致するペアの研究またはケース・コントロール研究の場合、各検定の被験者とその被験者に一致する対照被験者の応答は、データ・ファイル内の同じケースに含まれている必要があります。

仮定: ペアに対する観測は同じ条件で行う必要があります。また、平均値の差の分布は、正規分布でなければなりません。各変数の分散は、等しくても異なってもかまいません。

対応のあるサンプルの T 検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「平均の比較」 > 「対応のあるサンプルの t 検定...

2. 1 つ以上の変数ペアを選択します。

3. 必要に応じて「オプション」をクリックして、欠損データの処理方法と信頼区間のレベルを指定します。

対応のあるサンプルの t 検定のオプション

信頼区間: デフォルトでは、平均値の差の 95% 信頼区間が表示されます。別の信頼度レベルが必要な場合は、1 から 99 までの値を入力します。

欠損値: 複数の変数を検定する際に、1 つ以上の変数に対してデータが欠損している場合、手続きでどのケースを含めるか (または除外するか) を指定することができます。

- **分析ごとに除外:** 各 t 検定は、検定対象の変数のペアに対して有効なデータを持っているすべてのケースを使用します。そのため、サンプル・サイズが検定ごとに変化する場合があります。
- **リストごとに除外:** 各 t 検定は、検定対象のすべての変数のペアに対して有効なデータを持っているケースだけを使用します。そのため、サンプル・サイズはすべての検定で一定になります。

T-TEST コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 1 サンプルと独立サンプルの両方の t 検定を、1 つのコマンドを実行して生成する。
- リスト上の各変数に対する変数の検定を、対応のある t 検定で実行する (PAIRS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

1 サンプルの T 検定

「1 サンプルの T 検定」手続きは、単一の変数の平均値が、指定された定数と異なっているかどうかを検定します。

例: 研究者は、学生のグループの平均 IQ スコアが 100 とは異なるかどうかを検定することができます。または、シリアル・メーカーは、生産ラインからボックスのサンプルを採集し、95% の信頼度レベルでサンプルの平均重量が 1.3 ポンドと異なるかどうかを確認することができます。

統計: 各検定変数: 平均値、標準偏差、平均値の標準誤差。各データ値と仮説検定値の差の平均、その平均が 0 であることを検定する t 検定、その差に対する信頼区間 (信頼度レベルを指定できます)。

1 サンプルの T 検定データの考慮事項

データ: 仮説検定値に対する量的変数の値を検定するには、量的変数を選択して仮説検定値を入力します。

仮定: この検定は、データが正規分布しているものと仮定しますが、データが正規性から逸脱している場合でも、この検定はかなり頑健です。

1 サンプルの T 検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「平均の比較」 > 「1 サンプルの T 検定...」

2. 同じ仮説値に対して検定する変数を 1 つ以上選択します。
3. 各サンプルの平均値と比較する検定値を数値で入力します。
4. 必要に応じて「オプション」をクリックして、欠損データの処理方法と信頼区間のレベルを指定します。

1 サンプルの t 検定のオプション

信頼区間: デフォルトでは、平均値と仮説検定値との差の 95% 信頼区間が表示されます。別の信頼度レベルが必要な場合は、1 から 99 までの値を入力します。

欠損値: 複数の変数を検定する際に、1 つ以上の変数に対してデータが欠損している場合、手続きでどのケースを含めるか (または除外するか) を指定することができます。

- **分析ごとに除外:** 各 t 検定は、検定対象の変数について有効なデータを持っているすべてのケースを使用します。そのため、サンプル・サイズが検定ごとに変化する場合があります。
- **リストごとに除外:** 各 t 検定は、要求された t 検定で使用されるすべての変数に対して有効なデータを持っているケースだけを使用します。そのため、サンプル・サイズはすべての検定で一定になります。

T-TEST コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 1 サンプルと独立サンプルの両方の t 検定を、1 つのコマンドを実行して生成する。
- リスト上の各変数に対する変数の検定を、対応のある t 検定で実行する (PAIRS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

T-TEST コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 1 サンプルと独立サンプルの両方の t 検定を、1 つのコマンドを実行して生成する。
- リスト上の各変数に対する変数の検定を、対応のある t 検定で実行する (PAIRS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 10 章 一元配置分散分析

「一元配置分散分析」手続きは、単一の因子（独立）変数により、量的従属変数の一元配置分散分析を実行します。分散分析を使用して、一部の平均値は等しいという仮説を検定します。この手法は、2 サンプルの t 検定を拡張したものです。

平均値の間に差があることを判別するだけでなく、どの平均値が異なっているのかを調べたい場合もあります。平均値を比較する検定には、事前対比とその後の検定という 2 種類があります。対比は、実験を実行する前に設定される検定で、その後の検定は、実験が行われた後に実行する検定です。カテゴリ全体の傾向を検定することもできます。

例: ドーナッツは、調理時にかなりの量の脂肪を吸収します。ピーナッツ・オイル、コーン・オイル、ラードという 3 種類の脂肪を使用した実験が設定されています。ピーナッツ・オイルとコーン・オイルは不飽和脂肪で、ラードは飽和脂肪です。使用する脂肪のタイプによって脂肪の吸収量が異なるかどうかを判別するとともに、飽和脂肪と不飽和脂肪とでは脂肪の吸収量が異なるかどうかを判別するための事前対比を設定することもできます。

統計: グループごとのケース数、平均値、標準偏差、平均値の標準誤差、最小値、最大値、平均値の 95% 信頼区間。等分散性の Levene の検定、各従属変数に対する平均値の同等性を検定する分散分析表および頑健な検定、ユーザー指定の事前対比、およびその後の検定と多重比較: Bonferroni、Sidak、Tukey の HSD、Hochberg の GT2、Gabriel、Dunnnett、Ryan-Einot-Gabriel-Welsch の F 検定 (R-E-G-W F)、Ryan-Einot-Gabriel-Welsch の範囲検定 (R-E-G-W Q)、Tamhane の T2、Dunnnett の T3、Games-Howell、Dunnnett の C、Duncan の多重範囲検定、Student-Newman-Keuls (S-N-K)、Tukey の b 、Waller-Duncan、Scheffé、最小有意差。

一元配置分散分析データの考慮事項

データ: 因子変数の値は整数で、従属変数は量的変数（区間尺度）でなければなりません。

仮定: 各グループは、正規母集団から無作為に抽出された、互いに独立したサンプルです。分散分析は正規性からの逸脱に対して頑健ですが、データは対称でなければなりません。グループは、分散の等しい母集団からのグループでなければなりません。この仮定を検定するには、Levene の等分散性の検定を使用します。

一元配置分散分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「平均の比較」 > 「一元配置分散分析...」

2. 1 つ以上の従属変数を選択します。
3. 単一の独立因子変数を選択します。

一元配置分散分析の対比

グループ間平方和をトレンド成分に分割したり、事前対比を指定したりすることができます。

「多項式」。グループ間平方和をトレンド成分に分割します。因子変数の順序付けされた水準全体で、従属変数のトレンドを検定することができます。例えば、最高所得で順位付けられた水準全体で、給与の線型トレンド（増加または減少）を検定することができます。

- **次数:** 1 次、2 次、3 次、4 次、または 5 次の多項式を選択することができます。

係数: t 統計量によって検定されるユーザー指定の事前対比。因子変数の各グループ（カテゴリー）について係数を入力し、入力するたびに「追加」をクリックします。それぞれの新しい値が、係数リストの最後に追加されます。対比のグループをさらに指定するには、「次」をクリックします。「次」と「前」を使用して、対比のグループ間を移動することができます。

係数の順序は、因子変数のカテゴリー値の昇順に対応するため、重要です。リストの最初の係数が因子変数の最小グループ値に対応し、最後の係数が最大グループ値に対応します。例えば、因子変数のカテゴリーが 6 つある場合、係数 $-1, 0, 0, 0, 0.5, 0.5$ は、最初のグループを 5 番目と 6 番目のグループに対比させます。ほとんどのアプリケーションでは、係数の合計が 0 になる必要があります。合計が 0 にならない係数のセットも使用できますが、その場合は警告メッセージが表示されます。

一元配置分散分析のその後の検定

平均値の間に差があることが判明したら、その後の範囲検定とペアごとの多重比較により、どの平均値が異なっているのかを判断することができます。範囲検定は、互いに平均値に差がない等質サブセットを特定します。ペアごとの多重比較は、それぞれのペアごとの平均値の差を検定して、アルファ・レベル 0.05 で有意な差があるグループ平均値をアスタリスクで示す行列を生成します。

等分散が仮定されている

Tukey の HSD 検定、Hochberg の GT2、Gabriel の検定、Scheffé の検定は、多重比較検定および範囲検定です。使用できるその他の範囲検定には、Tukey の b 検定、S-N-K (Student-Newman-Keuls) の検定、Duncan、R-E-G-W F (Ryan-Einot-Gabriel-Welsch F 検定)、R-E-G-W Q (Ryan-Einot-Gabriel-Welsch 範囲検定)、Waller-Duncan の方法があります。使用できる多重比較検定には、Bonferroni、Tukey の HSD 検定、Sidak、Gabriel、Hochberg、Dunnnett の方法、Scheffé、LSD (最小有意差) があります。

- **LSD:** t 検定を使用して、グループ平均間のすべてのペアワイズ比較を実行します。多重比較の場合でも誤差率を調整しません。
- **Bonferroni.** t 検定を使用してグループ平均の間でのペアワイズ比較を行います。ただし、実験ごとの誤差率を総検定数で割った値を各検定の誤差率として設定することによって、全体の誤差率を調整します。したがって、多重比較を実行するとして観測有意水準を調整します。
- **Sidak:** t 統計量に基づくペアごとの多重比較検定。Sidak の方法は、多重比較の有意水準を調整して、Bonferroni の方法より厳しい限界を設定します。
- **Scheffe.** 平均値のすべての可能なペアワイズの組み合わせについて同時結合ペアワイズ比較を実行します。F サンプル分布を使用します。ペアワイズ比較だけでなく、グループ平均のすべての可能な線型結合を調べるために使用することができます。
- **R-E-G-W F :** F 検定に基づく Ryan-Einot-Gabriel-Welsch の多重ステップダウン手続き。
- **R-E-G-W Q :** スチューデント化された範囲に基づく Ryan-Einot-Gabriel-Welsch の多重ステップダウン手続き。
- **S-N-K:** スチューデント化した範囲の分布を使用して、平均値どうしのすべてのペアワイズ比較を実行します。標本サイズが等しい場合は、ステップワイズ法の手続きを使用して等質サブセット内の平均値のペアも比較します。平均値を高い順に順序付け、最初に極値の差を検定します。

- *Tukey*: スチューデント化された範囲の統計量を使用してグループ間のすべてのペアワイズ比較を実行します。実験ごとの誤差率を、すべてのペアワイズ比較の集合に対する誤差率に設定します。
- *Tukey* の *b*: スチューデント化された範囲の分布を使用してグループ間のペアワイズ比較を実行します。臨界値は、*Tukey* の HSD 検定と Student-Newman-Keuls 検定に対応する値の平均です。
- *Duncan*: Student Newman Keuls 検定で使用する回数と同一のステップワイズ比較回数を使用してペアワイズ比較を行います。ただし、個々の検定の誤差率ではなく、検定の集合の誤差率に対する保護レベルを設定します。スチューデント化された範囲統計量を使用します。
- *Hochberg* の *GT2*: スチューデント化された最大法を使用する多重比較と範囲の検定。*Tukey* の HSD 検定に似ています。
- *Gabriel*: スチューデント化された最大法を使用したペアワイズ比較検定。一般に、セルの大きさが等しくないときには *Hochberg* の *GT2* より強力です。セルの大ききのばらつきが大きい場合には、*Gabriel* の検定の方が公平になることがあります。
- *Waller-Duncan*: *t* 統計量に基づく多重比較検定。ベイズの方法を使用します。
- *Dunnett*: 一連の実験群を単一の対照平均と比較するペアワイズの多重比較 *t* 検定。最後のカテゴリが、デフォルトの対照カテゴリになります。代わりに、最初のカテゴリを選択することもできます。「両側」を選択すると、因子の任意のレベル (対照カテゴリを除く) の平均値が対照カテゴリの平均値と等しくないことが検定されます。「< 対照カテゴリ」を選択すると、因子の任意のレベルの平均値が対照カテゴリの平均値よりも小さいかどうかを検定されます。「> 対照カテゴリ」を選択すると、因子の任意のレベルの平均値が対照カテゴリの平均値よりも大きいかどうかを検定されます。

等分散が仮定されていない

等分散を仮定しない多重比較検定は、*Tamhane* の *T2*、*Dunnett* の *T3*、*Games-Howell*、*Dunnett* の *C* です。

- *Tamhane* の *T2*: *t* 検定に基づく保守的なペアワイズ比較検定。この検定は、分散が等しくない場合に適しています。
- *Dunnett* の *T3*: スチューデント化された最大法に基づくペアワイズ比較検定。この検定は、分散が等しくない場合に適しています。
- *Games-Howell*: 公平になる場合があるペアワイズ比較検定。この検定は、分散が等しくない場合に適しています。
- *Dunnett* の *C*: スチューデント化された範囲に基づくペアワイズ比較検定。この検定は、分散が等しくない場合に適しています。

注: 「テーブル プロパティ」ダイアログ・ボックス (ピボット・テーブルをアクティブにして、「書式」メニューから「テーブル プロパティ」を選択) の「空白の行と列を隠す」を選択解除すると、その後の検定からの出力が見やすくなる場合があります。

一元配置分散分析のオプション

統計: 以下のオプションを 1 つ以上を選択します。

- **記述統計**: グループごとに、ケースの数、平均値、標準偏差、平均値の標準誤差、最小値、最大値、各従属変数の 95% 信頼区間を計算します。
- **固定および変量効果**: 固定効果モデルの標準偏差、標準誤差、95% 信頼区間を表示し、変量効果モデルの標準誤差、95% 信頼区間、成分間の推定分散を表示します。

- **等分散性の検定:** Levene 統計量を計算して、グループの分散の等質性を検定します。この検定は、正規性の仮定には依存しません。
- **Brown-Forsythe:** Brown-Forsythe 統計量を計算して、グループ平均値の等質性を検定します。等分散性が仮定できない場合は、 F 統計量よりもこの統計量の方が適しています。
- **Welch:** Welch 統計量を計算して、グループ平均値の等質性を検定します。等分散性が仮定できない場合は、 F 統計量よりもこの統計量の方が適しています。

平均値のプロット: サブグループ平均値 (因子変数の値によって定義されたグループごとの平均値) をプロットするグラフを表示します。

欠損値: 欠損値の処理を制御します。

- **分析ごとに除外:** 特定の分析での従属変数または因子変数のどちらかに欠損値があるケースは、その分析では使用されません。さらに、因子変数に対して指定された範囲外のケースも使用されません。
- **リストごとに除外:** 因子変数またはメイン・ダイアログ・ボックスの従属変数リスト内の従属変数に欠損値があるケースは、すべての分析から除外されます。複数の従属変数を指定していない場合は、このオプションを指定しても効果はありません。

ONEWAY コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 固定効果と変量効果の統計量を取得する。固定効果モデルの場合は、標準偏差、平均値の標準誤差、95% 信頼区間です。変量効果モデルの場合は、標準誤差、95% 信頼区間、成分間の推定分散です (STATISTICS=EFFECTS を使用)。
- 最小有意差、Bonferroni、Duncan、Scheffé の各多重比較検定に対してアルファ・レベルを指定する (RANGES サブコマンドを使用)。
- 平均行列、標準偏差、度数を書き込む。または、平均行列、度数、プールされた分散、プールされた分散の自由度を読み込む。これらの行列を生データの代わりに使用して、一元配置分散分析を実行することができます (MATRIX サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 11 章 GLM 1 変量分散分析

「GLM 1 変量」手続きは、1 つ以上の因子や変数を使用して、1 つの従属変数について回帰分析や分散分析を実行します。因子変数により、母集団がいくつかのグループに分割されます。この一般線型モデル手続きを使用すると、1 つの従属変数のさまざまなグループの平均値に対する他の変数の効果についての帰無仮説を検定することができます。また、因子間の交互作用や因子ごとの効果を調べることができます。これらの因子の一部は、ランダムになっている場合があります。さらに、共変量の効果や、共変量と因子の交互作用を含めることもできます。回帰分析の場合、独立 (予測) 変数は共変量として指定されます。

釣り合い型モデルと不釣り合い型モデルの両方を検定することができます。モデル内の各セルに含まれているケース数が等しい場合、その計画は釣り合っています。「GLM 1 変量」手続きは、仮説の検定のほかに、パラメーターの推定値を生成します。

仮説の検定を実行する場合は、一般的に用いられている事前対比を使用することができます。さらに、全体的な F 検定で有意確率が判明していれば、その後の検定を使用して、特定の平均値間の差分を評価することができます。推定周辺平均から、モデル内のセルの予測平均値が推定されます。これらの平均値のプロファイル・プロット (交互作用プロット) を使用して、一部の関係を簡単に視覚化することができます。

残差、予測値、Cook の距離、てこ比の値は、データ・ファイルに新規変数として保存して、仮定の確認で使用することができます。

測定方法ごとに異なる精度を補正するなどの目的で、WLS 重みを使用して、重み付き最小二乗法 (WLS) 分析用のさまざまな重みを観測値に付けるために使用される変数を指定することができます。

例: シカゴ・マラソンの出場ランナーの個人データが数年分収集されています。各ランナーの完走タイムが従属変数です。その他の因子には、天候 (寒い、快適、暑い)、トレーニングの月数、過去のマラソン出場回数、性別などがあります。年齢は共変量と見なされます。この場合、性別が有意の効果であり、性別と天候の交互作用が有意であることがわかります。

方法: タイプ I、タイプ II、タイプ III、タイプ IV の平方和を使用して、異なる仮説を評価することができます。デフォルトはタイプ III です。

統計: その後の範囲検定と多重比較: 最小有意差、Bonferroni の方法、Sidak の方法、Scheffé の検定、Ryan-Einot-Gabriel-Welsch の多重 F 値、Ryan-Einot-Gabriel-Welsch の多重範囲、Student-Newman-Keuls の検定、Tukey の HSD 検定、Tukey の b 検定、Duncan の方法、Hochberg の GT2、Gabriel の方法、Waller-Duncan の t 検定、Dunnett の方法 (片側と両側)、Tamhane の T2、Dunnett の T3、Games-Howell の方法、Dunnett の C 。記述統計: すべてのセルにおけるすべての従属変数の観測平均値、標準偏差、度数。Levene の等分散性の検定。

作図: 水準と広がり、残差、プロファイル (交互作用)。

GLM 1 変量データの考慮事項

データ: 従属変数は量的変数です。因子はカテゴリ型です。数値または最大 8 文字までの文字列値を持つことができます。共変量は、従属変数に関連する量的変数です。

仮定: データは、正規母集団から無作為に抽出されるサンプルです。この母集団では、すべてセルの分散が同じになります。分散分析は正規性からの逸脱に対して頑健ですが、データは対称でなければなりません。仮定を確認するには、等分散性の検定や水準と広がりの方を使用します。残差と残差プロットについて探索的分析を実行することもできます。

GLM 1 変量分散分析テーブルを作成するには

1. メニューから次の項目を選択します。

「分析」 > 「一般線型モデル」 > 「1 変量...」

2. 従属変数を選択します。

3. 「固定因子」、「変量因子」、「共変量」について、データに対して適切な変数を選択します。

4. オプションで WLS 重みを使用して、重み付き最小二乗法分析用の重み付け変数を指定することができます。重み付け変数の値が 0、負、または欠損値である場合はケースを分析から除外します。モデル内で既に使用されている変数を重み付け変数として使用することはできません。

GLM モデル

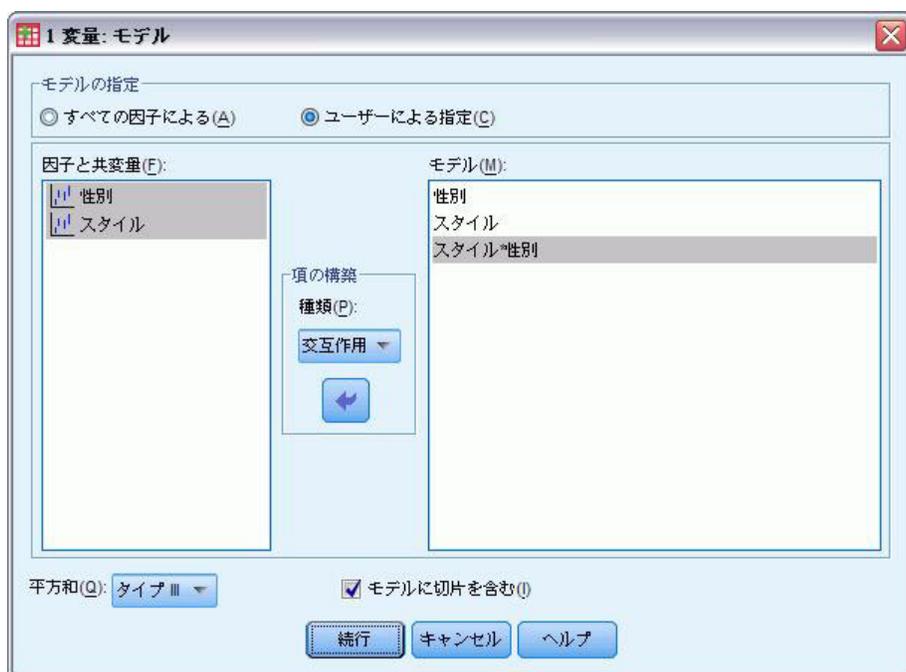


図1. 「1 変量: モデル」ダイアログ・ボックス

「モデルの指定」。すべての因子によるモデルには、すべての因子の主効果、すべての共変量の主効果、すべての因子間の交互作用が含まれます。共変量の交互作用は含まれません。交互作用の一部だけを指定する場合や、因子と共変量の交互作用を指定する場合は、「ユーザーの指定」をクリックします。モデルに含めるすべての項目を指定する必要があります。

因子と共変量: 因子と共変量がリストされます。

モデル: モデルは、使用するデータの性質によって異なります。「ユーザーの指定」を選択すると、分析対象の主効果と交互作用を選択できるようになります。

平方和: 平方和の計算方法。欠損セルがない釣り合い型モデルや不釣り合い型モデルの場合は、タイプ III の平方和の方法が最もよく使用されます。

モデルに切片を含む: 通常、モデルには切片が含まれています。データが原点を通ると想定できる場合は、切片を除外してもかまいません。

項の構築

選択した因子や共変量について、次の項を作成できます。

交互作用: 選択したすべての変数について、最高水準の交互作用項が作成されます。これはデフォルトです。

主効果: 選択した変数ごとに、主効果の項目が作成されます。

2 次まで: 選択した変数で考えられる 2 次の交互作用がすべて作成されます。

3 次まで: 選択した変数で考えられる 3 次の交互作用がすべて作成されます。

4 次まで: 選択した変数で考えられる 4 次の交互作用がすべて作成されます。

5 次まで: 選択した変数で考えられる 5 次の交互作用がすべて作成されます。

平方和

モデルに対して、平方和のタイプを選択することができます。最も一般的に使用されるのはタイプ III です。これがデフォルトです。

タイプ I: この方法は、平方和の階層的分解法とも呼ばれます。各項目は、モデル内で先行する項目に対してのみ調整されます。タイプ I の平方和は、通常、以下のモデルに対して使用します。

- 1 次の交互作用効果の前に主効果が指定され、2 次の交互作用効果の前に 1 次の交互作用効果が指定されている (これ以降も同様)、分散分析の釣り合い型モデル。
- 高次の項目の前に低次の項目が指定されている多項式回帰モデル。
- 最初に指定された効果が 2 番目に指定された効果内でネストされ、2 番目に指定された効果が 3 番目に指定された効果内にネストされている (これ以降も同様)、純粋にネストされているモデル (この形式のネストを指定するには、シンタックスを使用する必要があります)。

タイプ II: この方法では、他のすべての該当する効果に対して調整されたモデルの効果の平方和が計算されます。該当する効果とは、調査対象の効果を含んでいないすべての効果に対応する効果のことです。タイプ II の平方和の方法は、通常、以下のモデルに対して使用します。

- 分散分析の釣り合い型モデル。
- 因子の主効果だけを持つモデル。
- 回帰モデル。
- 純粋にネストされている計画 (この形式のネストを指定するには、シンタックスを使用する必要があります)。

タイプ III: これがデフォルトの方法です。この方法では、計画内の効果の平方和を、その効果を含まない他の効果に対して調整されており、その効果を含む効果 (存在する場合) に直交している平方和として計算します。タイプ III の平方和には、通常の推定形式が一定の状態に保たれている限り、セル度数が変化しないという大きな利点があります。そのため、このタイプの平方和は、多くの場合、欠損セルがない不釣り合

い型モデルで役立ちます。欠損セルのない多因子計画の場合、この方法は Yates の平均値の重み付き 2 乗法に相当します。タイプ III の平方和の方法は、通常、以下のモデルに対して使用します。

- タイプ I とタイプ II に記載されているモデル。
- 空白セルのない釣り合い型モデルまたは不釣り合い型モデル。

タイプ IV: この方法は、欠損セルが存在する場合を目的とした方法です。計画内の効果 F に対して、 F が他のどの効果にも含まれていない場合、タイプ IV = タイプ III = タイプ II となります。 F が他の効果に含まれている場合、タイプ IV は、 F のパラメーター間で行われている対比を、より高いレベルのすべての効果に等しく分配します。タイプ IV の平方和の方法は、通常、以下のモデルに対して使用します。

- タイプ I とタイプ II に記載されているモデル。
- 空白セルがある釣り合い型モデルまたは不釣り合い型モデル。

GLM の対比

1 つの因子の水準の間に差異があるかどうかを検定するには、対比を使用します。対比は、モデルの因子ごと (反復測定モデルでは被験者間因子ごと) に指定することができます。対比は、パラメーターの線型結合を表します。

GLM 1 変量: 仮説の検定は、帰無仮説 $LB = 0$ に基づきます。 L は対比係数行列、 B はパラメーターのベクトルです。対比が指定されると、 L 行列が作成されます。因子に対応する L 行列の列は、対比と一致します。残りの列が調整されて、 L 行列が推定可能な状態になります。

出力には、対比のセットごとの F 統計量が含まれます。対比の差異については、スチューデントの t 分布に基づく Bonferroni の同時信頼区域も表示されます。

使用可能な対比

使用可能な対比には、偏差、単純、差分、Helmert、反復測定、多項式があります。偏差対比と単純対比については、参照カテゴリーを最終カテゴリーと最初のカテゴリーのどちらにするかを選択することができます。

対比タイプ

「**偏差**」。各レベルの平均値 (参照カテゴリーを除く) を、すべてのレベルの平均値 (全平均) と比較します。因子のレベルは、任意の順序にすることができます。

「**単純**」。各レベルの平均値を、指定されたレベルの平均値と比較します。このタイプの対比は、制御グループが存在する場合に便利です。最初のカテゴリーまたは最後のカテゴリーを参照として選択することができます。

「**差分**」。各レベルの平均値 (最初のレベルは除く) を、前のレベルの平均値と比較します (逆 Helmert 対比と呼ばれる場合があります)。

「**Helmert**」。因子の各レベルの平均値 (最後のレベルは除く) を、その後のレベルの平均値と比較します。

「**反復**」。各レベルの平均値 (最後のレベルは除く) を、その後のレベルの平均値と比較します。

「**多項式**」。1 次効果、2 次効果、3 次効果... を比較します。第 1 自由度にはすべてのカテゴリーを通じた 1 次効果が含まれ、第 2 自由度には 2 次効果が含まれます (第 3 自由度以降も同様です)。こうした対比は、多項式のトレンドを推定する場合によく使用されます。

GLM のプロフィール・プロット

プロフィール・プロット (交互作用プロット) は、モデル内の周辺平均を比較する場合に役立ちます。プロフィール・プロットは、1 つの点が、因子の 1 つの水準における従属変数 (共変量を対象に調整) の推定周辺平均を示す線のプロットです。第 2 因子の水準を使用して、個別の線を作成することができます。第 3 因子の各水準を使用して、個別のプロットを作成することができます。固定因子と変量因子がある場合は、それらすべての因子をプロットで使用することができます。多変量分析の場合は、従属変数ごとにプロフィール・プロットが作成されます。反復測定分析では、被験者間因子と被験者内因子の両方をプロフィール・プロットで使用することができます。Advanced Statistics オプションがインストールされている場合のみ、「GLM 多変量」と「GLM 反復測定」を使用することができます。

1 つの因子のプロフィール・プロットでは、推定周辺平均が水準全体で増えているか減っているかが示されます。因子が複数ある場合、平行線は因子間で交互作用がないことを示しています。その場合、1 つの因子に限ってその水準を調べることができます。平行ではない線は、因子間で交互作用があることを示しています。

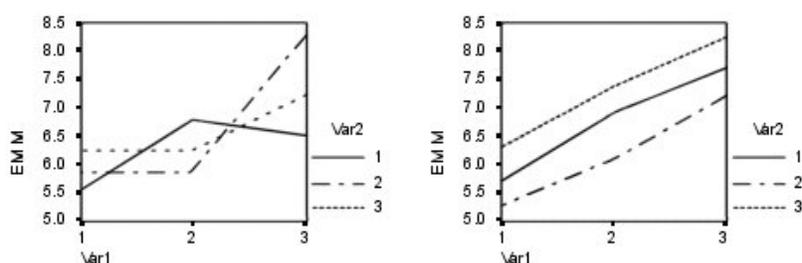


図2. 平行ではないプロット (左) と平行プロット (右)

横軸の因子を選択し、必要に応じて個別の線の因子と個別のプロットの因子を選択することによってプロットを指定したら、そのプロットを「作図」リストに追加する必要があります。

GLM のオプション

このダイアログ・ボックスで、オプションの統計を選択することができます。統計量は、固定効果モデルを使用して計算されます。

推定周辺平均: セルにおける母集団周辺平均について推定したい因子と交互作用を選択します。共変量が存在する場合、これらの平均値は、共変量に対して調整されます。

- **主効果の比較:** 被験者間因子と被験者内因子の両方について、モデル内の主効果に対する推定周辺平均値間で、ペアごとに無補正の比較を行います。この項目を使用できるのは、「平均値の表示」リストで主効果を選択した場合だけです。
- **信頼区間の調整:** 信頼区間と有意性に対して、最小有意差 (LSD) 調整、Bonferroni 調整、または Sidak 調整を選択します。この項目を使用できるのは、「主効果の比較」を選択した場合だけです。

表示: すべてのセルにおけるすべての従属変数の観測平均値、標準偏差、度数を求めるには、「記述統計」を選択します。「効果サイズの推定値」には、それぞれの効果とそれぞれのパラメータ推定値に対する偏相関のイータの 2 乗の値が表示されます。イータの 2 乗統計量は、因子に起因する総変動の比率を記述します。観測値に基づいて対立仮説を立てる際に検定の検出力を取得するには、「観測検定力」を選択します。パラメータ推定値、標準誤差、 t 検定、信頼区間、各検定での観測検定力を表示するには、「パラメータ推定値」を選択します。L 行列を取得するには、「対比係数行列」を選択します。

「等分散性の検定」を選択すると、被験者間因子の場合のみ、被験者間因子のすべての水準の組み合わせによる各従属変数の分散の等質性分析について、Levene の検定が生成されます。水準と広がりとの図と残差プロットのオプションは、データに関する仮定を確認する場合に役立ちます。因子が存在しない場合、この項目は無効になります。それぞれの従属変数について、標準化された予測による観測残差プロットを作成するには、「残差プロット」を選択します。こうしたプロットは、分散が等しいという仮定を調査する場合に役立ちます。モデルによって従属変数と独立変数との関係を正確に記述できるかどうかを確認するには、「不適合度」を選択します。「一般の推定可能関数」を使用すると、一般推定可能関数に基づくカスタム仮説検定を構成することができます。任意の対比係数行列における行は、一般推定可能関数の線型結合です。

有意水準: 信頼区間を構成するために、その後の検定や信頼係数で使用される有意水準を調整したい場合があります。指定された値は、検定の観測検定力の計算にも使用されます。有意水準を指定すると、信頼区間の関連水準がダイアログ・ボックスに表示されます。

UNIANOVA コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 計画に入れ子効果を指定する (DESIGN サブコマンドを使用)。
- 効果の検定に対する効果 (または値) の線型組み合わせを指定する (TEST サブコマンドを使用)。
- 多重対比を指定する (CONTRAST サブコマンドを使用)。
- ユーザー欠損値を含める (MISSING サブコマンドを使用)。
- EPS 基準を指定する (CRITERIA サブコマンドを使用)。
- カスタムの **L** 行列、**M** 行列、または **K** 行列を作成する (LMATRIX サブコマンド、MMATRIX サブコマンド、KMATRIX サブコマンドを使用)。
- 偏差対比と単純対比について、中間参照カテゴリーを指定する (CONTRAST サブコマンドを使用)。
- 多項式対比の計量を指定する (CONTRAST サブコマンドを使用)。
- その後の比較の誤差項を指定する (POSTHOC サブコマンドを使用)。
- 任意の因子について、または因子リスト内の因子間の交互作用について、推定周辺平均値を計算する (EMMEANS サブコマンドを使用)。
- 一時変数の名前を指定する (SAVE サブコマンドを使用)。
- 相関行列のデータ・ファイルを作成する (OUTFILE サブコマンドを使用)。
- 被験者間因子の分散分析表の統計量を格納する行列データ・ファイルを作成する (OUTFILE サブコマンドを使用)。
- 計画行列を新しいデータ・ファイルに保存する (OUTFILE サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

GLM のその後の比較

その後の多重比較検定: 平均値の間に差があることが判明したら、その後の範囲検定とペアごとの多重比較により、どの平均値が異なっているのかを判断することができます。比較は、調整されていない値に基づいて行われます。この検定は、固定被験者間因子に対してのみ使用されます。GLM 反復測定では、この検定は、被験者間因子がない場合は使用できません。また、その後の多重比較検定は、被験者内因子の水準全体の平均に対して実行されます。GLM 多変量の場合、その後の検定は従属変数ごとに実行されます。

Advanced Statistics オプションがインストールされている場合のみ、「GLM 多変量」と「GLM 反復測定」を使用することができます。

多重比較検定では、通常、Bonferroni 検定と Tukey の HSD 検定が使用されます。**Bonferroni の検定**は、スチューデントの t 検定統計量に基づいて、多重比較が行われるという事実に対して有意水準を調整します。**Sidak の t 検定**でも有意水準が調整され、Bonferroni 検定よりも厳しく制限されます。**Tukey の HSD 検定**は、スチューデント化された範囲統計量を使用して、すべてのペアごとの比較をグループ間で行い、実験ごとの誤差率をすべてのペアごとの比較の集合の誤差率に設定します。多数の平均値のペアを検定する場合は、Tukey の HSD 検定の方が Bonferroni 検定よりも有効です。少数のペアの場合は、Bonferroni 検定の方が有効です。

Hochberg の GT2 は Tukey の HSD 検定と類似していますが、スチューデント化された最大法が使用されます。通常は、Tukey の検定の方が有効です。**Gabriel のペアごとの比較検定**も、スチューデント化された最大法を使用しますが、通常、セルのサイズが均等でない場合は、Hochberg の GT2 よりも有効です。セルのサイズのばらつきが大きい場合には、Gabriel の検定の方が公平になることがあります。

Dunnnett のペアごとの多重比較 t 検定は、処理のセットを 1 つの対照平均値と比較します。最後のカテゴリは、デフォルトの対照カテゴリです。代わりに、最初のカテゴリを選択することもできます。両側の検定を選択することも、片側の検定を選択することもできます。因子の任意のレベル (対照カテゴリを除く) の平均値が対照カテゴリの平均値と等しくないことを検定するには、両側の検定を使用します。因子の任意のレベルの平均値が対照カテゴリの平均値よりも小さいかどうかを検定するには、「< 対照カテゴリ」を選択します。同様に、因子の任意のレベルの平均値が対照カテゴリの平均値よりも大きいかどうかを検定するには、「> 対照カテゴリ」を選択します。

Ryan, Einot, Gabriel, Welsch (R-E-G-W) は、2 種類のステップダウン多重範囲検定を開発しました。ステップダウン多重手続きは、最初に、すべての平均値が等しいかどうかを検定します。すべての平均値が等しいわけではない場合は、平均値のサブセットが等しいかどうかを検定します。**R-E-G-W の F 値**は F 検定に基づき、**R-E-G-W の Q 値**はスチューデント化された範囲に基づきます。この検定は、Duncan の多重範囲検定や Student-Newman-Keuls の検定 (これもステップダウン多重手続きです) よりも有効ですが、セルのサイズが等しくない場合はお勧めできません。

分散が等しくない場合は、**Tamhane の T2** (t 検定に基づくペアごとの控えめな比較)、**Dunnnett の T3** (スチューデント化された最大偏差に基づくペアごとの比較検定)、**Games-Howell のペアごとの比較検定** (公平な場合もある)、または **Dunnnett の C** (スチューデント化された範囲に基づくペアごとの比較検定) を使用してください。モデルに複数の因子がある場合、これらのテストは有効ではないため、生成されません。

Duncan の多重範囲検定、Student-Newman-Keuls (**S-N-K**) の検定、**Tukey の b** 検定は、グループ平均を順位付け、範囲の値を計算する範囲検定です。これらの検定は、先に述べた検定ほど頻繁には使用されません。

Waller-Duncan の t 検定は、Bayesian のアプローチを使用しています。この範囲検定は、サンプル・サイズが等しくない場合にサンプル・サイズの調和平均を使用します。

Scheffé の検定の有意水準は、この機能で使用できるペアごとの比較だけでなく、グループ平均で考えられるすべての線型結合を検定できるように設計されています。結果的に、Scheffé の検定は他の検定よりも控えめになってしまうことが多いため、有意確率を求める場合は、平均値間の差が大きくなければなりません。

最小有意差 (**LSD**) のペアごとの多重比較検定は、グループのすべてのペア間の多重 t 検定に相当します。この検定の欠点は、観測された有意水準が多重比較用に調整されないことです。

表示される検定: ペアごとの比較は、LSD、Sidak、Bonferroni、Games-Howell、Tamhane の T2 と T3、Dunnnett の C、Dunnnett の T3 で使用することができます。範囲検定の等質サブグループは、

S-N-K、Tukey の *b*、Duncan、R-E-G-W の *F*、R-E-G-W の *Q*、Waller で使用することができます。Tukey の HSD 検定、Hochberg の GT2、Gabriel の検定、Scheffé の検定は、多重比較検定でもあり、範囲検定でもあります。

GLM のオプション

このダイアログ・ボックスで、オプションの統計を選択することができます。統計量は、固定効果モデルを使用して計算されます。

推定周辺平均: セルにおける母集団周辺平均について推定したい因子と交互作用を選択します。共変量が存在する場合、これらの平均値は、共変量に対して調整されます。

- **主効果の比較:** 被験者間因子と被験者内因子の両方について、モデル内の主効果に対する推定周辺平均値間で、ペアごとに無補正の比較を行います。この項目を使用できるのは、「平均値の表示」リストで主効果を選択した場合だけです。
- **信頼区間の調整:** 信頼区間と有意性に対して、最小有意差 (LSD) 調整、Bonferroni 調整、または Sidak 調整を選択します。この項目を使用できるのは、「主効果の比較」を選択した場合だけです。

表示: すべてのセルにおけるすべての従属変数の観測平均値、標準偏差、度数を求めるには、「記述統計」を選択します。「効果サイズの推定値」には、それぞれの効果とそれぞれのパラメーター推定値に対する偏相関のイータの 2 乗の値が表示されます。イータの 2 乗統計量は、因子に起因する総変動の比率を記述します。観測値に基づいて対立仮説を立てる際に検定の検出力を取得するには、「観測検定力」を選択します。パラメーター推定値、標準誤差、*t* 検定、信頼区間、各検定での観測検定力を表示するには、「パラメーター推定値」を選択します。**L** 行列を取得するには、「対比係数行列」を選択します。

「等分散性の検定」を選択すると、被験者間因子の場合のみ、被験者間因子のすべての水準の組み合わせによる各従属変数の分散の等質性分析について、Levene の検定が生成されます。水準と広がりとの図と残差プロットのオプションは、データに関する仮定を確認する場合に役立ちます。因子が存在しない場合、この項目は無効になります。それぞれの従属変数について、標準化された予測による観測残差プロットを作成するには、「残差プロット」を選択します。こうしたプロットは、分散が等しいという仮定を調査する場合に役立ちます。モデルによって従属変数と独立変数との関係を正確に記述できるかどうかを確認するには、「不適合度」を選択します。「一般の推定可能関数」を使用すると、一般推定可能関数に基づくカスタム仮説検定を構成することができます。任意の対比係数行列における行は、一般推定可能関数の線型結合です。

有意水準: 信頼区間を構成するために、その後の検定や信頼係数で使用される有意水準を調整したい場合があります。指定された値は、検定の観測検定力の計算にも使用されます。有意水準を指定すると、信頼区間の関連水準がダイアログ・ボックスに表示されます。

UNIANOVA コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 計画に入れ子効果を指定する (DESIGN サブコマンドを使用)。
- 効果の検定に対する効果 (または値) の線型組み合わせを指定する (TEST サブコマンドを使用)。
- 多重対比を指定する (CONTRAST サブコマンドを使用)。
- ユーザー欠損値を含める (MISSING サブコマンドを使用)。
- EPS 基準を指定する (CRITERIA サブコマンドを使用)。
- カスタムの **L** 行列、**M** 行列、または **K** 行列を作成する (LMATRIX サブコマンド、MMATRIX サブコマンド、KMATRIX サブコマンドを使用)。
- 偏差対比と単純対比について、中間参照カテゴリーを指定する (CONTRAST サブコマンドを使用)。

- 多項式対比の計量を指定する (CONTRAST サブコマンドを使用)。
- その後の比較の誤差項を指定する (POSTHOC サブコマンドを使用)。
- 任意の因子について、または因子リスト内の因子間の交互作用について、推定周辺平均値を計算する (EMMEANS サブコマンドを使用)。
- 一時変数の名前を指定する (SAVE サブコマンドを使用)。
- 相関行列のデータ・ファイルを作成する (OUTFILE サブコマンドを使用)。
- 被験者間因子の分散分析表の統計量を格納する行列データ・ファイルを作成する (OUTFILE サブコマンドを使用)。
- 計画行列を新しいデータ・ファイルに保存する (OUTFILE サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

GLM の保存

残差、関連測定値、モデルが予測した値は、データ・エディターで新しい変数として保存することができます。これらの変数の多くは、データに関する前提を調べるために使用することができます。値を保存して別の IBM SPSS Statistics セッションで使用するには、現在のデータ・ファイルを保存する必要があります。

「予測値」。モデルがケースごとに予測する値。

- 標準化されていない: モデルが予測する従属変数の値。
- 重み付けあり: 重み付きの、標準化していない予測値。既に WLS 変数を選択している場合に限り使用することができます。
- 標準誤差: 独立変数の値が同じケースを対象とした、従属変数の平均値の標準偏差の推定値。

診断: 独立変数の例外的な値の組み合わせを持つケースと、モデルに大きな影響を与える可能性があるケースを特定するための測定方法。

- *Cook* の距離: 特定のケースを回帰係数の計算から除外した場合にすべてのケースの残差がどの程度変化するかを示す指標。Cook の D が大きいときは、回帰統計量の計算からケースを除外すると係数が大きく変化することを示します。
- てこ比の値: 非中心化てこ比の値。モデルの適合度に対する各観測値の相対的な影響度。

「残差」。標準化されていない残差は、従属変数の実際の値から、モデルが予測した値を引いたものです。標準化された残差、スチューデント化された残差、削除された残差も使用することができます。WLS 変数を選択した場合は、重み付けされた標準化されていない残差を使用することができます。

- 標準化されていない: 観測した値と、モデルによって予測された値との差。
- 重み付けあり: 重み付きの、標準化していない残差。既に WLS 変数を選択している場合に限り使用することができます。
- 標準化: 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。平均値は 0 であり、標準偏差は 1 です。
- スチューデント化: 独立変数の各ケースの値と独立変数の平均値との距離に応じて、ケースごとに異なる標準偏差の推定量で残差を割ったもの。
- 削除済み: ケースを回帰係数の計算から除外した場合のそのケースの残差。従属変数の値と調整済み予測値の差です。

係数の統計量: モデル内のパラメーター推定値の分散共分散行列を、現在のセッションの新しいデータ・セット、または IBM SPSS Statistics の外部データ・ファイルに書き込みます。また、それぞれの従属変数に

対して、パラメーター推定値の行、パラメーター推定値に対応する t 統計量の有意確率値の行、残差自由度の行が作成されます。多変量モデルの場合は、従属変数ごとに同様の行が存在します。行列ファイルを読み込む別の手続きで、この行列ファイルを使用することができます。

GLM のオプション

このダイアログ・ボックスで、オプションの統計を選択することができます。統計量は、固定効果モデルを使用して計算されます。

推定周辺平均: セルにおける母集団周辺平均について推定したい因子と交互作用を選択します。共変量が存在する場合、これらの平均値は、共変量に対して調整されます。

- **主効果の比較:** 被験者間因子と被験者内因子の両方について、モデル内の主効果に対する推定周辺平均値間で、ペアごとに無補正の比較を行います。この項目を使用できるのは、「平均値の表示」リストで主効果を選択した場合だけです。
- **信頼区間の調整:** 信頼区間と有意性に対して、最小有意差 (LSD) 調整、Bonferroni 調整、または Sidak 調整を選択します。この項目を使用できるのは、「主効果の比較」を選択した場合だけです。

表示: すべてのセルにおけるすべての従属変数の観測平均値、標準偏差、度数を求めるには、「記述統計」を選択します。「効果サイズの推定値」には、それぞれの効果とそれぞれのパラメーター推定値に対する偏相関のイータの 2 乗の値が表示されます。イータの 2 乗統計量は、因子に起因する総変動の比率を記述します。観測値に基づいて対立仮説を立てる際に検定の検出力を取得するには、「観測検定力」を選択します。パラメーター推定値、標準誤差、 t 検定、信頼区間、各検定での観測検定力を表示するには、「パラメータ推定値」を選択します。**L** 行列を取得するには、「対比係数行列」を選択します。

「等分散性の検定」を選択すると、被験者間因子の場合のみ、被験者間因子のすべての水準の組み合わせによる各従属変数の分散の等質性分析について、Levene の検定が生成されます。水準と広がりとの図と残差プロットのオプションは、データに関する仮定を確認する場合に役立ちます。因子が存在しない場合、この項目は無効になります。それぞれの従属変数について、標準化された予測による観測残差プロットを作成するには、「残差プロット」を選択します。こうしたプロットは、分散が等しいという仮定を調査する場合に役立ちます。モデルによって従属変数と独立変数との関係を正確に記述できるかどうかを確認するには、「不適合度」を選択します。「一般の推定可能関数」を使用すると、一般推定可能関数に基づくカスタム仮説検定を構成することができます。任意の対比係数行列における行は、一般推定可能関数の線型結合です。

有意水準: 信頼区間を構成するために、その後の検定や信頼係数で使用される有意水準を調整したい場合があります。指定された値は、検定の観測検定力の計算にも使用されます。有意水準を指定すると、信頼区間の関連水準がダイアログ・ボックスに表示されます。

UNIANOVA コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 計画に入れ子効果を指定する (DESIGN サブコマンドを使用)。
- 効果の検定に対する効果 (または値) の線型組み合わせを指定する (TEST サブコマンドを使用)。
- 多重対比を指定する (CONTRAST サブコマンドを使用)。
- ユーザー欠損値を含める (MISSING サブコマンドを使用)。
- EPS 基準を指定する (CRITERIA サブコマンドを使用)。
- カスタムの **L** 行列、**M** 行列、または **K** 行列を作成する (LMATRIX サブコマンド、MMATRIX サブコマンド、KMATRIX サブコマンドを使用)。

- 偏差対比と単純対比について、中間参照カテゴリーを指定する (CONTRAST サブコマンドを使用)。
- 多項式対比の計量を指定する (CONTRAST サブコマンドを使用)。
- その後の比較の誤差項を指定する (POSTHOC サブコマンドを使用)。
- 任意の因子について、または因子リスト内の因子間の交互作用について、推定周辺平均値を計算する (EMMEANS サブコマンドを使用)。
- 一時変数の名前を指定する (SAVE サブコマンドを使用)。
- 相関行列のデータ・ファイルを作成する (OUTFILE サブコマンドを使用)。
- 被験者間因子の分散分析表の統計量を格納する行列データ・ファイルを作成する (OUTFILE サブコマンドを使用)。
- 計画行列を新しいデータ・ファイルに保存する (OUTFILE サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 12 章 2 変量の相関分析

「2 変量の相関分析」手続きは、Pearson の相関係数、Spearman のロー、および Kendall のタウ b をそれぞれの有意水準で計算します。相関は、変数またはランク順の関係を測ります。相関係数を計算する前に、外れ値（これがあると誤った結果を出す可能性があります） および線型関係の証拠があるかどうか、データを調べてください。Pearson の相関係数は、線型の関連性に対する測度です。2 つの変数が完全に関連していても、その関係が直線的でない場合は、Pearson の相関係数はそれらの関連性を測るのに適した統計ではありません。

例: あるバスケットボール・チームのゲーム勝数と、1 ゲームあたりの平均得点には相関関係があるでしょうか。散布図は線型関係があることを示しています。1994 年から 1995 年の NBA シーズンのデータを解析すると、Pearson の相関係数 (0.581) は 0.01 レベルで有意であることがわかります。シーズンあたりの勝数が多いほど、相手チームの得点はより少ないとも考えることができます。これらの変数は負の相関関係 (-0.401) にあり、相関は 0.05 レベルで有意となります。

統計: 各変数: 非欠損値を持つケースの数、平均値、標準偏差。変数の各ペア: Pearson の相関係数、Spearman のロー、Kendall のタウ b 、交差積和、共分散。

2 変量相関分析データの考慮事項

データ: Pearson の相関係数には、対称的な量的変数を使用し、Spearman のローおよび Kendall のタウ b には、量的変数または順序付けされたカテゴリ変数を使用します。

仮定: Pearson の相関係数は、変数のそれぞれの組が 2 変量正規であると仮定します。

2 変量の相関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。

「分析」 > 「相関」 > 「2 変量...」

1. 2 つ以上の数値変数を選択します。

以下のオプションも使用することができます。

- **相関係数:** 正規分布した量的変数には、「**Pearson**」の相関係数を選択します。データが正規分布していない場合、またはデータに順序付けしたカテゴリがある場合は、ランク順の間の関連性を測る「**Kendall のタウ b** 」または「**Spearman**」を選択します。相関係数の値の範囲は、-1 (完全な負の関係) から +1 (完全な正の関係) になります。値 0 は線型関係がないことを示します。結果を解釈するとき、有意相関を理由に因果関係があるという結論を出さないように注意してください。
- **有意差検定:** 両側確率または片側確率を選択できます。関連性の方向が事前に分かっている場合は、「**片側**」を選択します。それ以外の場合は、「**両側**」を選択します。
- **有意な相関係数に星印を付ける:** 5 % 水準で有意な相関係数はアスタリスクが 1 つ、1 % 水準で有意な相関係数はアスタリスクが 2 つ付いた形で識別されます。

2 変量の相関分析のオプション

統計量: Pearson の相関分析には、次のいずれか一方、または両方を選択できます。

- **平均値と標準偏差:** 変数ごとに表示されます。非欠損値を持つケースの数も表示されます。欠損値の設定に関係なく、欠損値は変数ごとに処理されます。
- **交差積和と共分散:** 変数のペアごとに表示されます。交差積和は、平均値を修正した変数の積の和に等しくなります。これは Pearson の相関係数の分子です。共分散は 2 変数間の関係の標準化されていない測度であり、 $N-1$ で割った交差積和と等しくなります。

欠損値: 以下のいずれかを選択することができます。

- **ペアごとに除外:** 相関係数の変数ペアのうち、片方または両方が欠損値であるケースは、分析から除外されます。各係数は、その特定の変数ペアに対して有効なコードを持つすべてのケースに基づくため、使用可能な最大の情報が計算ごとに使用されます。そのため、使用するケースの数が係数ごとに異なる場合があります。
- **リストごとに除外:** 変数の欠損値が 1 つでもあるケースは、すべての相関から除外されます。

CORRELATIONS コマンドと NONPAR CORR コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- Pearson の相関で、相関行列を書き込む (MATRIX サブコマンドを使用)。この相関行列を生データの代わりに使用して、因子分析などの他の分析を実行することができます。
- 2 つの変数リスト間の各変数の相関を取得する (VARIABLES サブコマンドで WITH キーワードを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 13 章 偏相関

「偏相関」手続きは、1 つ以上の追加変数の効果を制御しながら、2 つの変数間の線型関係を記述する偏相関係数を計算します。相関は、線型連関の測度です。2 つの変数を完全に関連付けることができますが、その関係が線型ではない場合、相関係数は連関を測定するための適切な統計ではありません。

例: 医療用資金と罹患率との間に関係はあるでしょうか。このような関係は負の相関になると予測するかもしれませんが、研究によると、医療用資金が増加すると罹患率も増加するという有意な正の相関が報告されています。ただし、医療サービス機関への訪問率を操作すると、観測される正の相関は実質的に排除されます。医療用資金と罹患率は、単に正の関係を持つように見えるだけです。なぜなら、医療用資金が増えるほど、医療機関を受診する人も増え、医者や病院からより多くの病気が報告されることになるためです。

統計: 各変数: 非欠損値を持つケースの数、平均値、標準偏差。自由度と有意確率を持つ偏相関行列と 0 次相関行列。

偏相関データの考慮事項

データ: 対称的な量的変数を使用します。

仮定: 「偏相関」手続きでは、変数の各ペアは 2 変量正規であると仮定されます。

偏相関分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「相関」 > 「偏相関...」

2. 偏相関を計算する 2 つ以上の数値変数を選択します。

3. 1 つ以上の数値型制御変数を選択します。

以下のオプションも使用することができます。

- **有意差検定:** 両側確率または片側確率を選択することができます。関連性の方向が事前に分かっている場合は、「片側」を選択します。それ以外の場合は、「両側」を選択します。
- **有意確率を表示:** デフォルトでは、相関係数ごとに確率と自由度が表示されます。この項目を選択解除すると、5% 水準で有意な係数は 1 つのアスタリスクで表示され、1% 水準で有意な係数は 2 つのアスタリスクで表示され、自由度は非表示になります。この設定は、偏相関行列と 0 次相関行列の両方に影響します。

偏相関のオプション

統計以下のオプションのいずれかだけを選択することも、両方を選択することもできます。

- **平均値と標準偏差:** 変数ごとに表示されます。非欠損値を持つケースの数も表示されます。
- **0 次相関:** 制御変数を含むすべての変数間の単純相関の行列が表示されます。

欠損値: 以下のいずれかのオプションを選択することができます。

- **リストごとに除外:** 制御変数を含め、変数に欠損値のあるケースは、すべての計算から除外されます。
- **ペアごとに除外:** 偏相関の基礎となる 0 次相関の計算では、ペアになった変数の両方または一方に欠損値があるケースは使用されません。ペアごとの削除では、可能な限り多くのデータが使用されます。た

だし、ケースの数は係数全体で異なる場合があります。ペアごとの削除が有効な場合、特定の偏相関の自由度は、すべての 0 次相関の計算で使用されるケースの最小数に基づきます。

PARTIAL CORR コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 0 次相関行列を読み取る、または偏相関行列を書き込む (MATRIX サブコマンドを使用)。
- 2 つの変数リスト間の偏相関を取得する (VARIABLES サブコマンドでキーワード WITH を使用)。
- 複数の分析を実行する (複数の VARIABLES サブコマンドを使用)。
- 2 つの制御変数を使用する場合に、1 次偏相関と 2 次偏相関の両方など、必要な次数値を指定する (VARIABLES サブコマンドを使用)。
- 冗長な係数を非表示にする (FORMAT サブコマンドを使用)。
- 計算できない係数がある場合に、単純相関行列を表示する (STATISTICS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 14 章 距離

この手続きは、変数のペア間またはケースのペア間の類似度または非類似度 (距離) を測定するさまざまな統計量を計算します。これらの類似度または距離測度を、因子分析、クラスター分析、多次元尺度法などの他の手続きで使用して、複雑なデータ・セットを分析することができます。

例: エンジンのサイズ、MPG (ガソリン 1 ガロン当たりの走行距離)、馬力などの特定の特性変数を基に、自動車のペア間の類似度を測定することは可能でしょうか。自動車の類似度を計算することにより、どの自動車が互いに類似していて、どの自動車が類似していないかを知ることができます。より正式な分析を行う場合は、基本的な構造を調べる階層クラスター分析または多次元尺度法を適用することもできます。

統計: 非類似度 (距離) の測度は、区間データの場合には、ユークリッド距離、平方ユークリッド距離、Chebychev、ブロック、Minkowski、またはカスタマイズ、度数データの場合には、カイ 2 乗またはファイ 2 乗、2 値データの場合には、ユークリッド距離、平方ユークリッド距離、サイズの差異、パターンとの差異、分散、形、または Lance と Williams の距離です。類似度の測度は、間隔データの場合には、Pearson の相関係数またはコサイン、2 値データの場合には、Russel と Rao、単純マッチング、Jaccard、Dice、Rogers と Tanimoto、Sokal と Sneath 1、Sokal と Sneath 2、Sokal と Sneath 3、Kulczynski 1、Kulczynski 2、Sokal と Sneath 4、Hamann、ラムダ、Anderberg の D 、Yule の Y 、Yule の Q 、落合、Sokal と Sneath 5、ファイ 4 分点相関、または散らばりです。

距離行列を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「相関」 > 「距離...」

2. ケース間の距離を計算する場合は 1 つ以上の数値型変数を選択し、変数間の距離を計算する場合は 2 つ以上の数値型変数を選択します。
3. 「距離の計算」グループ内の選択肢を選択し、ケース間の近接性と変数間の近接性のどちらを計算するかを指定します。

距離行列の非類似度の測定方法

「測定」グループでデータの種類 (区間、度数、2 値) を選択してから、データの種類に応じた測定方法をドロップダウン・リストから 1 つ選択します。データの種類により、以下の測定方法を選択することができます。

- **区間データ:** ユークリッド距離、平方ユークリッド距離、Chebychev、都市ブロック、Minkowski、またはカスタマイズ。
- **度数データ:** カイ 2 乗測度またはファイ 2 乗測度。
- **2 値データ:** ユークリッド距離、平方ユークリッド距離、サイズの差、パターンの違い、分散、形、または Lance と Williams (「真」と「偽」のそれぞれに値を入力し、対象の 2 値を指定します。他のすべての値は、距離では無視されます)。

「値の変換」グループを使用すると、近接度を計算する前に、ケースまたは変数のデータの値を標準化することができます。これらの変換は、2 値データには適用されません。選択可能な標準化方法は、「 Z 得点」、「-1 から 1 の範囲」、「0 から 1 の範囲」、「最大値を 1」、「平均値を 1」、「標準偏差を 1」です。

「測定方法の変換」グループを使用すると、距離の測定方法によって生成された値を変換することができます。変換された値は、距離測度の計算後に適用されます。選択可能なオプションは、「絶対値」、「符号変換」、「0 ～ 1 の範囲で尺度化」です。

距離行列の類似度の測定方法

「測定」グループでデータの種類 (区間または 2 値) を選択してから、データの種類に応じた測定方法をドロップダウン・リストから 1 つ選択します。データの種類により、以下の測定方法を選択することができます。

- **区間データ:** Pearson の相関係数またはコサイン。
- **2 値データ:** Russell と Rao、単純整合、Jaccard、Dice、Rogers と Tanimoto、Sokal と Sneath 1、Sokal と Sneath 2、Sokal と Sneath 3、Kulczynski 1、Kulczynski 2、Sokal と Sneath 4、Hamann、ラムダ、Anderberg の *D*、Yule の *Y*、Yule の *Q*、落合、Sokal と Sneath 5、ファイ 4 分点相関、または散らばり (「真」と「偽」のそれぞれに値を入力し、対象の 2 値を指定します。他のすべての値は、距離では無視されます)。

「値の変換」グループを使用すると、近接度を計算する前に、ケースまたは変数のデータの値を標準化することができます。これらの変換は、2 値データには適用されません。選択可能な標準化方法は、「z 得点」、「-1 から 1 の範囲」、「0 から 1 の範囲」、「最大値を 1」、「平均値を 1」、「標準偏差を 1」です。

「測定方法の変換」グループを使用すると、距離の測定方法によって生成された値を変換することができます。変換された値は、距離測度の計算後に適用されます。選択可能なオプションは、「絶対値」、「符号変換」、「0 ～ 1 の範囲で尺度化」です。

PROXIMITIES コマンドの追加機能

「距離」手続きは、PROXIMITIES コマンド・シンタックスを使用します。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- Minkowski 距離測度のべき乗として任意の整数を指定する。
- カスタマイズされた距離測度のべき乗および根として任意の整数を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 15 章 線型モデル

線型モデルは、対象と 1 つ以上の予測変数との線型の関係に基づいて連続型対象を予測します。

線型モデルは比較的単純で、スコアリング用に、解釈が容易な数式を提供します。これらのモデルのプロパティについてはよく理解されていて、同じデータ・セットの他のモデル・タイプ (ニューラル・ネットワークまたはディジション・ツリーなど) に比べて、通常、非常に素早く作成できます。

例: 住宅所有者の保険金請求の調査を行うにはリソースが限られている保険会社が、請求のコストを推定するためのモデルを作成したいと考えます。このモデルをサービス・センターに導入することによって、担当者は電話で顧客と対話しながら請求情報を入力し、過去のデータに基づいて「予想」コストを直ちに入手することができます。

フィールド要件: 対象と 1 つ以上の入力が必要です。デフォルトでは、事前定義された役割が「両方」または「なし」のフィールドは使用されません。対象は連続型 (スケール) でなければなりません。予測変数 (入力) には測定レベルの制限はありません。カテゴリ・フィールド (名義型、および順序型) はこのモデルで因子として使用され、連続型フィールドは共変量として使用されます。

注: カテゴリ・フィールドに 1000 を超えるカテゴリがある場合、手続きは実行されず、モデルは作成されません。

線型モデルの取得方法

この機能には Statistics Base オプションが必要です。

メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「自動線型モデル...」

1. 1 つ以上の対象および入力があることを確認してください。
2. 「作成オプション」をクリックし、オプションの作成設定およびモデル設定を指定します。
3. 「モデル オプション」をクリックし、アクティブ・データ・セットにスコアを保存し、外部ファイルにモデルをエクスポートします。
4. 「実行」をクリックして手続きを実行し、「モデル」オブジェクトを作成します。

目的

主な目的は?: 該当する目的を選択します。

- **標準モデルを作成:** この方法では、予測変数を使用して対象を予測する単一モデルが作成されます。一般的に、ブースティング、バギング、または大規模なデータ・セット・アンサンブルと比べ、標準モデルは解釈が容易であり、素早くスコアリングできます。
- **モデル精度を向上 (ブースティング):** この方法では、ブースティングを使用してアンサンブル・モデルが作成されます。これによって、より正確な予測を得るための一連のモデルが生成されます。アンサンブルは、標準モデルと比べて作成とスコアリングに時間がかかる場合があります。

ブースティングによって一連の「コンポーネント・モデル」が生成されます。各コンポーネント・モデルはデータ・セット全体に作成されます。連続する各コンポーネント・モデルを作成する前に、以前の

コンポーネント・モデルの残差に基づきレコードに重みが付けられます。残差が大きいケースには比較的大きな分析の重みが与えられるため、次のコンポーネント・モデルは、これらのレコードの予測に重点を置きます。これらのコンポーネント・モデルがまとまってアンサンプル・モデルを形成します。アンサンプル・モデルは結合ルールを使用して新しいレコードをスコアリングします。使用できる規則は、対象の測定レベルによって異なります。

- **モデル安定性を向上 (バギング):** この方法では、バギング (ブートストラップ集計) を使用してアンサンプル・モデルが作成されます。これによって、より信頼性の高い予測を得るための複数のモデルが生成されます。アンサンプルは、標準モデルと比べて作成とスコアリングに時間がかかる場合があります。

ブートストラップ集計 (バギング) は、元のデータ・セットから置換を行うサンプリングによって、学習データ・セットの複製を作成します。これにより、元のデータ・セットと同じサイズのブートストラップ・サンプルが作成されます。その後、「コンポーネント・モデル」が複製ごとに作成されます。これらのコンポーネント・モデルがまとまってアンサンプル・モデルを形成します。アンサンプル・モデルは結合ルールを使用して新しいレコードをスコアリングします。使用できる規則は、対象の測定レベルによって異なります。

- **非常に大きいデータ・セットのモデルを作成 (IBM SPSS Statistics Server が必要):** この方法では、データ・セットを別々のデータ・ブロックに分割することにより、アンサンプル・モデルが作成されます。上記のモデルのいずれかを作成するにはデータ・セットが大きすぎる場合、または増分モデル作成の場合、このオプションを選択します。このオプションは、作成には時間がかからないものの、標準モデルと比べてスコアリングに時間がかかる場合があります。このオプションには、IBM SPSS Statistics Server 接続が必要です。

ブースティング、バギング、および非常に大きいデータ・セットに関する設定については、64 ページの『アンサンプル』を参照してください。

基本

自動的にデータを準備する: このオプションでは、対象と予測変数を内部で変換してモデルの予測精度を最大化する手続きが可能になります。いずれの変換もモデルに保存され、スコアリングする新しいデータに適用されます。変換フィールドの元のバージョンはモデルから除外されます。デフォルトでは、次の自動データ準備が実行されます。

- **日付および時刻の処理:** 日付の各予測変数は、基準日 (1970-01-01) 以降の経過時間を含む新たな連続型予測値に変換されます。時刻の各予測変数は、基準時刻 (00:00:00) 以降の経過時間を含む連続型予測値に変換されます。
- **測定レベルの調整:** 異なる値が 5 個より少ない連続型予測値は、順序型予測値に変更されます。異なる値が 10 個より多い順序型予測値は、連続型予測値に変更されます。
- **外れ値の処理:** カットオフ値 (平均値からの標準偏差が 3) を超える連続型予測値の値がカットオフ値に設定されます。
- **欠損値の処理:** 名義型予測値の欠損値は、学習データ区分の最頻値と置き換えられます。順序型予測値の欠損値は、学習データ区分の中央値と置き換えられます。連続型予測値の欠損値は、学習データ区分の平均値と置き換えられます。
- **監視結合:** 対象と関連して処理するフィールドの数を減らすことにより、より節約的なモデルを作成します。類似するカテゴリーが、入力と対象との関係に基づいて特定されます。有意差のないカテゴリー (p 値が 0.1 より大きいカテゴリー) は結合されます。すべてのカテゴリーが 1 つに結合される場合、元のバージョンおよび派生したバージョンのフィールドは、予測変数としての値がないため、モデルから除外されます。

信頼度レベル: 係数ビューでモデル係数の区間の推定値を計算するために使用する信頼度のレベルです。0 より大きく 100 より小さい値を指定してください。デフォルトは 95 です。

モデル選択

モデル選択方法: モデル選択方法 (下記参照) のいずれかを選択するか、または主効果のモデル項として使用可能なすべての予測値を単に入力する「すべての予測値を含む」を選択します。デフォルトでは、「変数増加ステップワイズ法」が使用されます。

変数増加ステップワイズ法の選択: モデルの効果がなく状態から、これ以上追加または削除できなくなるまで、ステップワイズ法の基準に従って徐々に効果を追加および削除します。

- **投入または除去の基準:** これは、モデルに効果を加えるかどうか、またはモデルから効果を削除するかどうかを決定するときに使用する統計です。「情報量基準 (AICC)」はモデルを指定された学習セットの尤度に基づき、過度に複雑なモデルにペナルティーを科すよう調整されます。「F 統計量」はモデルのエラーの改善に対する統計検定に基づいています。「調整済み R2 乗」は学習セットの適合度に基づき、過度に複雑なモデルにペナルティーを科すよう調整されます。「オーバーフィット防止基準 (ASE)」は、オーバーフィット防止セットの適合度 (平均平方誤差 (ASE)) に基づきます。オーバーフィット防止セットは、モデルの学習に使用されない元のデータ・セットのおよそ 30% の無作為サブサンプルです。

「F 統計量」以外の基準を選択した場合、各ステップでその基準での最も大きい正の増分に対応する効果がモデルに追加されます。その基準での減少に対応するモデルの効果はいずれも削除されます。

「F 統計量」を基準として選択すると、「次の値より小さい p 値の効果を含む」で指定したしきい値よりも小さい、最小の p 値を持つ効果が、各ステップでモデルに追加されます。デフォルトは 0.05 です。「次の値より大きい p 値の効果削除する」で指定したしきい値より大きい p 値を持つモデルの効果はいずれも削除されます。デフォルトは 0.10 です。

- **最終モデルの最大効果数をカスタマイズする:** デフォルトでは、すべての使用可能な効果をモデルに投入できます。あるいは、ステップワイズ・アルゴリズムが指定した効果の最大数でステップを終了する場合は、アルゴリズムは効果の現在のセットで停止します。
- **ステップの最大数をカスタマイズする:** 特定のステップ数の後、ステップワイズ・アルゴリズムが停止します。デフォルトでは、これは使用可能な効果の数の 3 倍です。あるいは、ステップの最大数を正整数で指定します。

最適サブセットの選択: 「可能なすべての」モデル、または少なくとも変数増加ステップワイズ法より大きい、可能なモデルのサブセットをチェックし、最適サブセットの基準に従って最適サブセットを選択します。「情報量基準 (AICC)」はモデルを指定された学習セットの尤度に基づき、過度に複雑なモデルにペナルティーを科すよう調整されます。「調整済み R2 乗」は学習セットの適合度に基づき、過度に複雑なモデルにペナルティーを科すよう調整されます。「オーバーフィット防止基準 (ASE)」は、オーバーフィット防止セットの適合度 (平均平方誤差 (ASE)) に基づきます。オーバーフィット防止セットは、モデルの学習に使用されない元のデータ・セットのおよそ 30% の無作為サブサンプルです。

最大の基準値を持つモデルが最良のモデルとして選択されます。

注: 最適サブセットによる選択は、変数増加ステップワイズ法による選択に比べてより多くの計算リソースを使用します。最適サブセットが、ブースティング、バギング、または非常に大きいデータ・セットと組み合わせると、変数増加ステップワイズ法の選択を使用して作成された標準モデルよりも大幅に時間がかかる場合があります。

アンサンブル

これらの設定は、「目的」で、ブースティング、バギング、または非常に大きなデータ・セットが要求される場合に起きるアンサンブルの動作を決定します。選択された目的に適用されないオプションは無視されません。

バギングおよび非常に大きなデータ・セット: アンサンブルをスコアリングする際、この規則を使い、基本モデルの予測値を結合して、アンサンブル・スコア値を計算します。

- **連続型対象のデフォルトの結合規則:** 連続型対象に対するアンサンブル予測値は、基本モデルの予測値の平均値または中央値を使用して結合できます。

モデルの精度を上げることが目的である場合、結合規則の選択が無視されることに注意してください。ブースティングは常に、カテゴリ対象のスコアリングには重み付き多数決を使用し、連続型対象のスコアリングには重み付き中央値を使用します。

ブースティングおよびバギング: モデルの精度または安定性の向上が目的である場合、作成する基本モデル数を指定します。バギングの場合、これはブートストラップ・サンプルの数になります。これは正整数である必要があります。

詳細設定

結果の再現: ランダム・シードを設定すると、分析を再現することができます。どのレコードをオーバーフィット防止セットに含むかを選択するには、乱数ジェネレーターを使用します。整数を指定するか、「生成」をクリックします。「生成」をクリックすると、1 から 2147483647 までの整数の疑似乱数が作成されます。デフォルトは 54752075 です。

モデル・オプション

推定値をデータ・セットに保存: デフォルトの変数名は *PredictedValue* です。

モデルのエクスポート: モデルを外部の .zip ファイルに書き込みます。このモデル・ファイルを使用して、スコアリングのために他のデータ・ファイルにモデル情報を適用できます。一意で有効なファイル名を指定します。ファイルの指定が既存ファイルを示す場合、ファイルは上書きされます。

モデルの要約

「モデルの要約」ビューでは、モデルとその適合度についてスナップショット的に一目でわかる要約が表示されます。

テーブル: テーブルは次のようなハイレベルなモデル設定を特定します。

- 「フィールド」タブで指定した対象名。
- 「基本」設定で指定したとおりに、自動データ準備が実行されたか。
- 「モデル選択」設定で指定したモデルの選択方法および選択基準。最終モデルの選択基準の値も表示され、値が小さいほど適切である形式で提示されます。

グラフ: グラフには、最終モデルの精度が表示され、値が大きいほど適切である形式で提示されます。値は、 $100 \times$ 最終モデルの調整済み R^2 乗です。

自動データ準備

このビューには、自動データ準備 (ADP) ステップで、どのフィールドが除外されたか、また変換されたフィールドがどのように派生したかについての情報が表示されます。テーブルには変換または除外されたフィールドごとに、フィールド名、分析での役割、および ADP ステップで実行されたアクションがリストされます。フィールドは、フィールド名のアルファベット順 (昇順) にソートされます。各フィールドに対して行うことのできるアクションには、以下のようなものがあります。

- 「**期間の取得: 月**」では、日付を含むフィールドの値から現在のシステムの日付までの経過時間 (月) が計算されます。
- 「**期間の取得: 時間**」では、時間を含むフィールドの値から現在のシステムの時間までの経過時間 (時間) が計算されます。
- 「**測定の尺度を連続型から順序型に変更**」では、固有値が 5 個未満の連続型フィールドが順序型フィールドに変更されます。
- 「**測定の尺度を順序型から連続型に変更**」では、固有値が 10 個を超える順序型フィールドが連続型フィールドに変更されます。
- 「**外れ値を削除**」では、カットオフ値 (平均値からの標準偏差が 3) を超える連続型予測フィールドの値がカットオフ値に設定されます。
- 「**欠損値の置換**」では、名義型の欠損値が最頻値に、順序型フィールドの欠損値が中央値に、連続型フィールドの欠損値が平均値に置き換えられます。
- 「**カテゴリを結合して目標との関連性を最大化**」では、入力と目標との関係に基づいて「類似する」予測カテゴリーが特定されます。有意差のないカテゴリー (p 値が 0.05 より大きいカテゴリー) が結合されます。
- 「**一定の予測値を/外れ値の処理後/カテゴリの結合後除外する**」では、1 つの値を持つ予測変数が除外されます。これらは、他の ADP アクションが実行された後に行われる可能性があります。

予測値の重要度

通常、モデリングの作業を最も重要な予測フィールドに集中させ、最も重要でない予測フィールドを削除または無視したいと考えます。予測値の重要度グラフを使用すると、モデル推定時に各予測値の相対重要度が示されるので便利です。値が相対的であるため、表示されるすべての予測変数の値の合計は 1.0 となります。予測値の重要度は、モデルの精度に関連しません。予測が正確かどうかに関係なく、予測時における各予測の重要度に関係しません。

予測対観測

縦軸の予測値に対し横軸に観測値を示した分割散布図が表示されます。点は 45 度の線上にあるのが理想です。このビューで、モデルによるレコードの予測にとりわけ問題があるかがわかります。

残差

モデル残差の診断グラフが表示されます。

グラフ・スタイル: さまざまな表示スタイルがあり、「スタイル」ドロップダウン・リストから選択できます。

- **ヒストグラム:** これは、正規分布のオーバーレイが適用された、スチューデント化残差の分割ヒストグラムです。線型モデルでは残差は正規分布になる想定するため、ヒストグラムは理想的にはほぼ滑らかな線になるはずですが。
- **正規 P-P プロット:** これは、スチューデント化残差を正規分布と比較する、分割された確率 - 確率プロットです。作図された点の傾きが通常の線に比べて小さい場合、残差は正規分布より大きい変動を示し、傾きがより大きい場合、残差は正規分布より小さい変動を示します。作図された点が S 字曲線を示す場合、残差の分布は歪んでいます。

外れ値

このテーブルではモデルに悪影響を与えるレコードがリストされ、レコード ID (「フィールド」タブで指定している場合)、対象値、および Cook の距離が表示されます。Cook の距離は、特定のレコードがモデル係数の計算から除外された場合に、すべてのレコードの残差がどのくらい変化するかを示す測度です。Cook の距離が大きい場合、レコードを除外すると係数が大幅に変わることを示しているため、影響があると考えられます。

影響を及ぼすレコードは慎重に検証し、モデルの推定時に重み付けを小さくするか、外れ値を許容可能な大きい値に切り詰めるか、あるいは影響のあるレコードを完全に削除するかを判断する必要があります。

効果

このビューには、モデルの各効果のサイズが表示されます。

スタイル: さまざまな表示スタイルがあり、「スタイル」ドロップダウン・リストから選択できます。

- **ダイアグラム:** この図表では、効果が予測値の重要度の大きい順にソートされます。ダイアグラム内の接続線は、効果の有意性に基づいて重みが付けられます。効果の有意性が大きいほど (p 値が小さいほど) 線が太くなります。接続線の上にマウス・ポインターを置くと、 p 値および効果の重要度を示すツールチップが表示されます。これはデフォルトです。
- **テーブル:** これは、モデル全体の効果および個別のモデル効果を示す分散分析テーブルです。個別の効果は、予測値の重要度が大きい順にソートされます。ただしデフォルトでは、テーブルが省略表示され、モデル全体の結果だけが表示されます。個別のモデル効果の結果を表示するには、テーブル内の「修正モデル」セルをクリックします。

予測値の重要度: 「予測値の重要度」スライダーは、どの予測値がビュー内に表示されるかを制御します。このスライダーを使用してもモデルは変更されませんが、最も重要な予測値にフォーカスすることができます。デフォルトでは、上位 10 件の効果が表示されます。

有意確率: 「有意確率」スライダーは、予測値の重要度に基づく表示のほか、さらにどの効果がビュー内に表示されるかを制御します。有意確率の値がスライダーの値より大きい効果は表示されません。このスライダーを使用してもモデルは変更されませんが、最も重要な効果にフォーカスすることができます。デフォルトでは値が 1.00 になるため、有意確率に基づいてフィルタリングされる効果はありません。

係数

このビューには、モデルの各係数の値が表示されます。因子 (カテゴリー予測値) はモデル内で指標コード化されるため、因子を含む効果には通常、複数の関連する係数があることに注意してください。冗長 (参照) パラメーターに対応するカテゴリー以外は、カテゴリーごとに 1 つの係数があります。

スタイル: さまざまな表示スタイルがあり、「スタイル」ドロップダウン・リストから選択できます。

- **ダイアグラム:** この図表では、まず切片項が表示されてから、すべての効果が予測値の重要度が大きい順にソートされます。因子を含む効果内で、係数はデータ値の昇順でソートされます。ダイアグラム内の接続線は、係数の符号 (図のキーを参照) に基づいて色分けされ、係数の有意性に基づいて重みが付けられます。係数の有意性が大きいほど (p 値が小さいほど) 線が太くなります。接続線上にマウス・ポインターを置くと、係数の値、その p 値、およびパラメーターが関連する効果の重要度を示すツールチップが表示されます。これはデフォルトのスタイルです。
- **テーブル:** 各モデル係数の値、有意差検定、および信頼区間が表示されます。切片項の後、効果は予測値の重要度の降順でソートされます。因子を含む効果内で、係数はデータ値の昇順でソートされます。ただしデフォルトではテーブルが省略表示され、各モデル・パラメーターの係数、有意性、および重要度だけが表示されます。標準誤差、 T 統計量、および信頼区間を表示するには、テーブルの「**係数**」セルをクリックします。テーブル内のモデル・パラメーターの名前上にマウス・ポインターを置くと、パラメーターの名前、パラメーターが関連する効果、および (カテゴリー予測値の場合は) モデル・パラメーターに関連する値のラベルを示すツールチップがテーブルに表示されます。これは特に、自動データ準備がカテゴリー予測値の類似カテゴリーを結合する時に作成された新しいカテゴリーを確認する場合に役立ちます。

予測値の重要度: 「予測値の重要度」スライダーは、どの予測値がビュー内に表示されるかを制御します。このスライダーを使用してもモデルは変更されませんが、最も重要な予測値にフォーカスすることができます。デフォルトでは、上位 10 件の効果が表示されます。

有意確率: 「有意確率」スライダーは、予測値の重要度に基づく表示のほか、さらにどの係数がビュー内に表示されるかを制御します。有意確率の値がスライダーの値より大きい係数は表示されません。このスライダーを使用してもモデルは変更されませんが、最も重要な係数に焦点を当てることができます。デフォルトでは値が 1.00 になるため、有意確率に基づいてフィルタリングされる係数はありません。

推定平均

これは有意な予測変数について表示されるグラフです。横軸の予測変数の各値に対して、縦軸に対象のモデル推定値を表示し、その他のすべての予測変数は一定に保ちます。対象に対する各予測変数の係数の効果が視覚化されるので有用です。

注: 有意な予測変数がない場合、推定平均値は生成されません。

モデル構築の要約

「モデル選択」設定で、「なし」以外のモデル選択アルゴリズムを選択すると、モデル作成プロセスの詳細が一部表示されます。

「**変数増加ステップワイズ法**」: 変数増加ステップワイズ法が選択アルゴリズムである場合、テーブルにはステップワイズ・アルゴリズムの最後の 10 ステップが表示されます。ステップごとに、選択基準の値とそのステップでのモデル内の効果が表示されます。これにより、各ステップのモデルへの寄与度を把握できます。各列で行を並べ替えることができるので、任意のステップでモデル内にどの効果があるかをより簡単に確認できます。

最適サブセット: 最適サブセットが選択アルゴリズムである場合、テーブルには上位 10 件のモデルが表示されます。モデルごとに、選択基準の値とモデル内の効果が表示されます。これにより、上位モデルの安定性が把握できます。相違点が少ない類似した効果が多い傾向があれば、「上位」モデルはかなり信頼できますが、非常に異なる効果を持つ傾向がある場合は、一部の効果が非常に類似している可能性があり、結合

(あるいは 1 つを削除) する必要があります。各列で行を並べ替えることができるので、任意のステップでモデル内にどの効果があるかをより簡単に確認できます。

第 16 章 線型回帰

線型回帰は、線型方程式の係数を推定します。従属変数の値を最適に予測する 1 つ以上の独立変数が使用されます。例えば、営業部員の年間総売上高 (従属変数) を、年齢、教育、経験年数などの独立変数から予測することができます。

例: バスケットボール・チームの 1 シーズン中の勝利ゲーム数は、ゲームごとのチーム得点の平均に関係しているでしょうか。散布図は、これらの変数には線型関係があることを示しています。勝利ゲーム数と相手チームの平均得点にも線型関係があります。これらの変数間には負の相関があります。つまり、勝利ゲーム数が増えると、相手チームの平均得点が減少します。線型回帰では、これらの変数間の関係をモデル化することができます。適切なモデルを使用すると、チームの勝利ゲーム数を予測することができます。

統計: 各変数: 有効ケース数、平均値、標準偏差。各モデル: 回帰係数、相関行列、部分相関および偏相関、多重 R 、 R^2 、調整済み R^2 、 R^2 の変化量、推定値の標準誤差、分散分析表、予測値、残差。また、各回帰係数の 95% 信頼区間、分散共分散行列、変動インフレーション因子、許容度、Durbin-Watson の検定、距離 (Mahalanobis、Cook、てこ比の値)、DfBeta、DfFit、予測区間、ケースごとの診断情報。プロット: 散布図、偏残差の散布図、ヒストグラム、正規確率プロット。

線型回帰データの考慮事項

データ: 従属変数と独立変数は量的でなければなりません。宗教、専攻、居住地区などのカテゴリ変数は、2 値 (ダミー) 変数またはその他の種類の対比変数として再割り当てする必要があります。

仮定: 独立変数の各値に対して、従属変数の分布が正規分布でなければなりません。従属変数の分布の分散は、独立変数のすべての値に対して一定でなければなりません。従属変数と各独立変数の関係は線型で、すべての観測値が独立している必要があります。

線型回帰分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「線型...」

2. 「線型回帰」ダイアログ・ボックスで、数値型の従属変数を選択します。

3. 1 つ以上の数値型の独立変数を選択します。

必要に応じて、以下の操作を実行することができます。

- 独立変数をいくつかのブロックにグループ化し、変数のサブセットごとに異なる投入方法を指定する。
- 選択変数を選択し、この変数に対して特定の値を持つケースのサブセットだけに、分析の対象を制限する。
- プロット上の点を識別するためのケース識別変数を選択する。
- 重み付き最小二乗法分析用に数値型の WLS 重み付け変数を選択する。

WLS. 重み付き最小二乗モデルを使用できます。データ・ポイントに、その分散の逆数で重みを付けます。そのため、分散の大きな観測値は、分散が小さな関連する観測値よりも分析に与える影響が小さくなります。重み付け変数の値が 0、負、または欠損値である場合はケースを分析から除外します。

線型回帰の変数選択方法

方法選択により、独立変数を分析に投入する方法を指定できます。各種の方法を使用して、同じ変数セットからさまざまな回帰モデルを作成できます。

- *Enter (Regression)* (強制投入法 (回帰)). 変数選択手続きの 1 つ。ブロック内のすべての変数を 1 つのステップで投入します。
- *Stepwise* (ステップワイズ法). 各ステップで、式になく、F 値の有意確率が十分に小さい独立変数のうち、その確率が最小の変数を投入します。回帰式に既に含まれている変数の F の確率が著しく高くなった場合は、その変数を除去します。投入対象や除去対象として適格な変数がなくなると、この手法は終了します。
- *Remove* (強制除去法). 変数選択手続きの 1 つ。ブロック内のすべての変数を 1 つのステップで除去します。
- *Backward Elimination* (変数減少法). すべての変数を式に投入してから順番に除去していく変数選択の手順。従属変数との偏相関が最も小さい変数が、最初に除去する候補になります。その変数が除外の基準を満たす場合は、除去されます。最初の変数が除去された後、式に残っている変数のうち、偏相関が最も小さい変数が次の候補になります。除去基準を満たす変数が式に存在しなくなったときに、プロシージャが停止します。
- *Forward Selection* (変数増加法). ステップワイズ変数選択プロシージャの 1 つ。変数を順番にモデルに投入します。最初に式に投入する変数の候補は、従属変数との正または負の相関が最も大きいものです。この変数は、投入基準を満たす場合に限り式に投入されます。最初の変数を投入した場合は、式に含まれていない独立変数のうち、偏相関が最大のものが次の候補になります。投入基準を満たす変数がなくなると、プロシージャが停止します。

出力の有意確率値は、1 つのモデルの当てはめに基づいています。そのため、ステップワイズの方法 (ステップワイズ法、変数増加法、変数減少法) を使用すると、通常は有意確率値が無効になります。

変数を式に投入するには、指定した投入方法にかかわらず、すべての変数が許容基準を満たしている必要があります。デフォルトの許容度は 0.0001 です。また、ある変数を投入すると、モデル内に既に存在する別の変数の許容度が許容基準より下がってしまう場合、その変数は投入されません。

選択されたすべての独立変数が、1 つの回帰モデルに追加されます。ただし、変数の別のサブセットに対して異なる投入方法を指定することもできます。例えば、ある変数ブロックをステップワイズ選択で回帰モデルに投入し、別のブロックを変数増加法で投入することができます。別の変数ブロックを回帰モデルに追加するには、「次へ」をクリックします。

線型回帰の規則の設定

分析では、選択規則によって定義されたケースが使用されます。例えば、ある変数を選択し、その変数に対して「等しい」を選択して、値として 5 を入力すると、選択した変数に対して値「5」を持つケースだけが分析で使用されることになります。値として文字列を使用することもできます。

線型回帰の作図

作図は、正規性、線型性、分散の等質性に対する仮定の妥当性を確認する場合に役立ちます。また、外れ値、異常な観測値、影響の大きなケースを検出する場合も、作図が役立ちます。予測値、残差、その他の診断情報を新しい変数として保存すると、データ・エディターで独立変数を使用して作図する際に、その新しい変数を使用することができます。使用可能な作図は以下のとおりです。

散布図: 従属変数、標準化予測値、標準化残差、削除された残差、調整済み予測値、スチューデント化された残差、スチューデント化された削除済み残差のうち、任意の 2 つを作図することができます。線型性と等分散性を確認するには、標準化予測値に対して標準化残差を作図します。

ソース変数リスト: 従属変数 (DEPENDNT)、予測変数および残差変数 (標準化予測値 (*ZPRED)、標準化残差 (*ZRESID)、削除済み残差 (*DRESID)、調整済み予測値 (*ADJPRED)、スチューデント化された残差 (*SRESID)、スチューデント化された削除済み残差 (*SDRESID)) をリストします。

全ての偏残差の散布図を作成: 独立変数と従属変数の両方が残りの独立変数上で個別に回帰されている場合に、各独立変数の残差と従属変数の残差の散布図を表示します。偏残差プロットを作成するには、2 つ以上の独立変数が式に含まれている必要があります。

標準化残差のプロット: 標準化残差のヒストグラムと正規確率プロットを作成して、標準化残差の分布を正規分布と比較することができます。

いずれかの作図が要求されると、標準化予測値と標準化残差 (*ZPRED および *ZRESID) の要約統計量が表示されます。

線型回帰: 新規変数の保存

診断情報として役立つ予測値、残差、その他の統計量を保存することができます。項目を選択するたびに、1 つ以上の新しい変数がアクティブなデータ・ファイルに追加されます。

「予測値」。各ケースについて回帰モデルが予測する値。

- **標準化されていない:** モデルが予測する従属変数の値。
- **標準化:** 各予測値を標準化形式に変換したものです。すなわち、予測値から平均予測値を引き、その差を予測値の標準偏差で割ったものです。標準化予測値の平均値は 0 であり、標準偏差は 1 です。
- **調整済み:** ケースを回帰係数の計算から除外した場合のそのケースの予測値。
- **平均予測値の標準誤差:** 予測値の標準誤差。独立変数の値が同じケースを対象とした、従属変数の平均値の標準偏差の推定値。

距離: 回帰モデルに大きく影響する可能性がある、独立変数とケースの値の異常な組み合わせを持つケースを特定する測定方法。

- **Mahalanobis:** 独立変数のケースの値と全ケースの平均との差異の程度を示す指標。マハラノビスの距離が大きい場合は、ケースにおいて 1 つ以上の独立変数に極値が存在することを示します。
- **Cook:** 特定のケースを回帰係数の計算から除外した場合にすべてのケースの残差がどの程度変化するかを示す指標。Cook の D が大きいときは、回帰統計量の計算からケースを除外すると係数が大きく変化することを示します。
- **てこ比の値:** 回帰の適合性に対する 1 つのポイントの影響度を測定します。中心化てこ比の範囲は、0 (適合性への影響なし) から (N-1)/N までです。

予測区間: 平均予測区間と個別予測区間の上限と下限。

- **Mean (平均):** 平均予測応答の予測区間の下限と上限 (2 つの変数)。
- **個別:** 1 つのケースに対する従属変数の予測区間の下限と上限 (2 つの変数)。
- **信頼区間:** 2 つの予測区間に対する信頼水準を指定する 1 から 99.99 までの値を入力します。この値を入力する前に、「平均」または「個別」を選択する必要があります。一般的な信頼区間は 90、95、および 99 です。

「残差」。従属変数の実際の値から回帰式で予測された値を引いた値。

- 標準化されていない: 観測した値と、モデルによって予測された値との差。
- 標準化: 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。平均値は 0 であり、標準偏差は 1 です。
- スチューデント化: 独立変数の各ケースの値と独立変数の平均値との距離に応じて、ケースごとに異なる標準偏差の推定量で残差を割ったもの。
- 削除済み: ケースを回帰係数の計算から除外した場合のそのケースの残差。従属変数の値と調整済み予測値の差です。
- スチューデント化された削除: ケースの削除済み残差をその標準誤差で割ったもの。スチューデント化された削除済み残差とその関連するスチューデント化された残差の差は、ケースを削除した場合にその予測値に生ずる差の程度を示します。

影響力の統計: 特定のケースを除外した場合の回帰係数の変化量 (DfBeta) と予測値の変化量 (DfFit)。標準化 DfBeta 値と標準化 DfFit 値も、共分散比とともに使用することができます。

- *DfBeta*. ベータ値の差は、特定のケースを除外した結果として生じる回帰係数の変化です。値は、モデル内の項ごとに (定数項を含む) 計算されます。
- *Standardized DfBeta* (標準化 *DfBeta*). ベータ値の差を標準化したもの。特定のケースを除外することで引き起こされる回帰係数の変化。絶対値が $2/(N$ の平方根) より大きいケースを調べてください (N はケースの数)。値は、モデル内の項ごとに (定数項を含む) 計算されます。
- *DfFit*. 当てはめ値の差は、特定のケースを除外した結果として生じる予測値の変化です。
- *Standardized DfFit* (標準化 *DfFit*). 当てはめ値の差を標準化したもの。特定のケースを除外することで引き起こされる予測値の変化。絶対値が p/N の平方根の 2 倍を超える標準化値を調べてください。ここで、 p はモデル内のパラメーターの数であり、 N はケースの数です。
- 共分散比: 特定のケースを回帰係数の計算から除外した共分散行列の行列式と、すべてのケースを含む共分散行列の行列式の比。この比が 1 に近い場合、そのケースは共分散行列に大きな影響を及ぼしていません。

係数の統計量: 回帰係数がデータ・セットやデータ・ファイルに保存されます。データ・セットは、同じセッション内で引き続き使用できますが、セッションを終了する前に明示的に保存しない限り、ファイルとしては保存されません。データ・セット名は、変数の命名規則に従っている必要があります。

「モデル情報を XML ファイルにエクスポート」。パラメーター推定値とその共分散 (オプション) が、指定したファイルに XML (PMML) 形式でエクスポートされます。このモデル・ファイルを使用して、スコアリングの目的で他のデータ・ファイルにモデル情報を適用できます。

線型回帰の統計

以下の統計を使用することができます。

回帰係数: 「推定値」には、回帰係数 B 、 B の標準誤差、標準化係数ベータ、 B の t 値、 t の両側有意確率が表示されます。「信頼区間」には、回帰係数ごと、または共分散行列に指定された信頼度レベルでの信頼区間が表示されます。「共分散行列」には、回帰係数の共分散行列が、共分散は対角線外に、分散は対角線上に表示されます。また、相関行列も表示されます。

モデルの適合度: モデルに投入された変数とモデルから削除された変数がリストされ、適合度統計量として、多重 R 、 R^2 、調整済み R^2 、推定値の標準誤差、分散分析表が表示されます。

R² 乗の変化量: 独立変数の追加や削除によって生じる R^2 統計量の変化量です。変数に関連する R^2 の変化量が大きければ、その変数は従属変数の適切な予測変数であるということになります。

記述統計: 分析の各変数に対する有効なケース数、平均値、標準偏差を表示します。また、各相関係数に対する片側有意確率とケース数を示す相関行列も表示されます。

Partial Correlation (偏相関): 2 つの変数の間から他の変数との相互関係による相関を取り除いた後にそれら 2 つの変数の間に残る相関。モデル内の他の独立変数の線型効果を従属変数と独立変数の両方から取り除いた後の、従属変数と独立変数の間の相関。

Part Correlation (部分相関 (ピボットテーブル 回帰)): モデル内の他の独立変数の線型効果を独立変数から取り除いた後の、従属変数と独立変数の間の相関。変数を式に追加するときの R^2 乗の変化に関連します。半偏相関とも呼びます。

共線性の診断: 共線性 (または多重共線性) は、独立変数の 1 つが他の独立変数の線型関数であることを示す、望ましくない状況です。尺度化および非中心化された積和行列の固有値、条件指標、分散分解の比率が、変動インフレーション因子 (VIF) と個々の変数の許容度とともに表示されます。

残差: 残差の系列相関に対する Durbin-Watson の検定と、選択基準を満たすケースに対するケースごとの診断 (標準偏差 n 倍以上の外れ値) を表示します。

線型回帰のオプション

使用可能なオプションは次のとおりです。

ステップ法の基準: このオプションは、変数選択の方法として、変数増加法、変数減少法、ステップワイズ法のいずれかが指定されている場合に適用されます。 F 値の有意確率または F 値自身のいずれかに応じて、変数をモデルに投入するか、モデルから除去することができます。

- **Use Probability of F (ステップワイズのための F 値確率):** F 値の有意水準が「投入」の値より小さい場合に変数をモデルに投入し、有意水準が「削除」の値より大きい場合に変数を除去します。「投入」は「削除」より小さくしなければならず、いずれの値も正でなければなりません。さらに多くの変数をモデルに投入するには、「投入」の値を上げてください。さらに多くの変数をモデルから除去するには、「除去」の値を下げてください。
- **Use F Value (ステップワイズのための F 値):** F 値が「投入」の値より大きい場合に変数をモデルに投入し、「削除」の値より小さい場合に変数を除去します。「投入」は「削除」より小さくしなければならず、いずれの値も正でなければなりません。さらに多くの変数をモデルに投入するには、「投入」の値を下げてください。さらに多くの変数をモデルから除去するには、「除去」の値を上げてください。

回帰式に定数項を含む: 回帰モデルには、デフォルトで定数項が含まれます。このオプションを選択解除すると、原点を通る回帰が適用されますが、これが実行されることはほとんどありません。原点を通る回帰の結果は、定数を含む回帰の結果に相当しない場合があります。例えば、 R^2 を通常の方法で解釈することはできません。

欠損値: 以下のいずれかを選択することができます。

- **リストごとに除外:** すべての変数に対して有効な値を持っているケースだけが分析で使用されます。
- **ペアごとに除外:** 相関する変数のペアに対して完全なデータを持つケースを使用して、回帰分析の基礎となる相関係数が計算されます。自由度は、ペア単位の最小数 N が基礎になります。
- **平均値で置換:** 欠損観測値を変数の平均値で置き換えて、すべてのケースが計算に使用されます。

REGRESSION コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 相関行列を書き込むか、生データの代わりに行列を読み込んで、回帰分析を実行する (MATRIX サブコマンドを使用)。
- 許容度を指定する (CRITERIA サブコマンドを使用)。
- 同じ従属変数または異なる従属変数に対して複数のモデルを取得する (METHOD サブコマンドと DEPENDENT サブコマンドを使用)。
- 追加の統計量を取得する (DESCRIPTIVES サブコマンドと STATISTICS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 17 章 順序回帰

順序回帰分析では、一連の予測値における、多分割順序応答の従属性をモデル化することができます。予測値は、因子または共変量となります。順序回帰分析の設計は McCullagh (1980, 1998) の方法に基づいており、シンタックスでは、この手続きのことを PLUM と呼びます。

標準線型回帰分析では、応答 (従属) 変数と、予測 (独立) 変数の重み付き組み合わせとの差を 2 乗した値の合計が最小化されます。推定係数は、予測変数の変化が応答変数にどのように影響するかを表します。応答は、応答のレベルの変化が応答の範囲全体にわたって等しいという意味で、数値であると仮定されます。例えば、身長 150 cm の人と、身長 140 cm の人の身長差は 10 cm です。これは、身長 210 cm の人と身長 200 cm の人の身長差と同じ意味になります。このような関係は、順序変数に対しては必ずしも成り立たず、応答カテゴリーの選択と数は非常に恣意的なものになります。

例: 順序回帰分析を使用して、薬品投与に対する患者の反応を調べることができます。考えられる反応の分類としては、「なし」、「穏やか」、「適度」、「激しい」などがあります。「穏やか」と「適度」の反応の違いは、数値化が困難または不可能なため、知覚に基づいて判別されます。さらに、「穏やか」と「適度」の反応の差は、「適度」と「激しい」の反応の差より大きい場合もあれば、逆に小さい場合もあります。

統計と作図: 観測度数、期待度数、累積度数、度数と累積度数の Pearson 残差、観測確率と期待確率、共変量パターンによる各応答カテゴリーの観測累積確率と期待累積確率、パラメーター推定値の漸近相関行列および漸近分散共分散行列、Pearson のカイ 2 乗と尤度比カイ 2 乗、適合度統計量、反復履歴、平行線の仮定の検定、パラメーター推定値、標準誤差、信頼区間、Cox と Snell の R^2 乗統計量、Negalkerke の R^2 乗統計量、McFadden の R^2 乗統計量。

順序回帰データの考慮事項

データ: 従属変数は順序変数であると仮定され、数値または文字列のどちらかです。順序は、従属変数の値を昇順で並べ替えることによって決定されます。最低値により、最初のカテゴリーが定義されます。因子変数は、カテゴリー型と仮定されます。共変量変数は数値型でなければなりません。複数の連続した共変量を使用すると、非常に大きなセルの確率表が容易に作成されてしまうため、注意してください。

仮定: 使用できる応答変数は 1 つだけで、必ず指定する必要があります。さらに、独立変数間の値の異なるパターンそれぞれについて、応答は独立した多項分布変数であると仮定されます。

関連プロシージャー: 名義ロジスティック回帰分析では、名義従属変数に対して類似したモデルが使用されます。

順序回帰の作成

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「順序...」

2. 従属変数を 1 つ選択します。
3. 「OK」をクリックします。

順序回帰のオプション

「オプション」ダイアログ・ボックスを使用して、反復推定アルゴリズムで使用されるパラメーターの調整、パラメーター推定値の信頼レベルの選択、リンク関数の選択を行うことができます。

反復: 反復アルゴリズムをカスタマイズすることができます。

- **最大反復回数:** 負でない整数を指定してください。0 を指定すると、この手続きは初期推定値を返します。
- **最大段階 2 分:** 正の整数を指定します。
- **対数尤度収束:** このアルゴリズムは、対数尤度の絶対変化または相対変化がこの値よりも小さい場合に停止します。0 を指定した場合、この基準は使用されません。
- **パラメーター収束:** このアルゴリズムは、それぞれのパラメーター推定値の絶対変化または相対変化がこの値よりも小さい場合に停止します。0 を指定した場合、この基準は使用されません。

信頼区間: 0 以上 100 未満の値を指定します。

「デルタ」。0 のセル度数に加算される値。1 未満の負ではない値を指定します。

「特異性許容度」。従属性の高い予測値かどうかを調べる場合に使用されます。オプションのリストから値を選択してください。

リンク関数: リンク関数は、モデルの推定を可能にする累積確率の変換です。以下の 5 つのリンク関数を使用することができます。

- **ロジット:** $f(x)=\log(x/(1-x))$ 。通常、分布が均一なカテゴリで使用されます。
- **補ログ・マイナス・ログ:** $f(x)=\log(-\log(1-x))$ 。通常、上位のカテゴリの確率が高い場合に使用されます。
- **負ログ・マイナス・ログ:** $f(x)=-\log(-\log(x))$ 。通常、下位のカテゴリの確率が高い場合に使用されます。
- **プロビット:** $f(x)=\Phi^{-1}(x)$ 。通常、潜在変数が正規分布する場合に使用されます。
- **コーチット (コーシーの逆関数):** $f(x)=\tan(\pi(x-0.5))$ 。通常、潜在変数に多数の極値が存在する場合に使用されます。

順序回帰の出力

「出力」ダイアログ・ボックスを使用すると、ビューアーで表示するテーブルを作成し、変数を作業ファイルに保存することができます。

表示: 以下に示す各テーブルが生成されます。

- **反復の記述:** 指定された出力反復頻度で、対数尤度値とパラメーター推定値が出力されます。最初の反復と最後の反復は常に出力されます。
- **適合度統計量:** Pearson カイ 2 乗統計量と尤度比カイ 2 乗統計量。これらの統計量は、変数リストで指定されている分類に基づいて計算されます。
- **要約統計量:** Cox と Snell の R^2 乗統計量、Nagelkerke の R^2 乗統計量、McFadden の R^2 乗統計量。
- **パラメーター推定値:** パラメーター推定値、標準誤差、および信頼区間。
- **パラメーター推定値の漸近相関:** パラメーター推定相関の行列。
- **パラメーター推定値の漸近共分散:** パラメーター推定共分散の行列。

- **セル情報:** 観測度数、期待度数、累積度数、度数と累積度数の Pearson 残差、観測確率と期待確率、共変量パターンによる各応答カテゴリーの観測累積確率と期待累積確率。多数の共変量パターンを持つモデル (連続した共変量を持つモデルなど) でこのオプションを使用すると、非常に大きく扱いにくいテーブルが生成される場合があるため、注意してください。
- **平行線の検定:** 位置パラメーターが従属変数のレベル間で等しいという仮説の検定。これは、位置専用モデルでのみ使用することができます。

「保存変数」。以下の各変数が作業ファイルに保存されます。

- **推定応答確率:** 因子/共変量パターンを応答カテゴリーに分類するモデル推定確率。応答カテゴリーの数だけ確率があります。
- 「**予測カテゴリー**」。因子/共変量パターンに対して最大推定確率を持つ応答カテゴリー。
- **予測カテゴリー確率:** 因子/共変量パターンを予測カテゴリーに分類する推定確率。この確率は、因子/共変量パターンの推定確率の最大値でもあります。
- 「**実際のカテゴリー確率**」。因子/共変量パターンを実際のカテゴリーに分類する推定確率。

対数尤度の出力: 対数尤度の表示を制御します。「**多項式定数を含む**」を指定すると、尤度の完全な値が表示されます。定数を含まない積全体で結果を比較するには、定数の除外を選択します。

順位回帰の位置モデル

「位置」ダイアログ・ボックスを使用して、分析用の位置モデルを指定することができます。

モデルの指定: 主効果モデルには、共変量と因子の主効果が含まれますが、交互作用効果は含まれません。カスタム・モデルを作成して、因子または共変量の交互作用のサブセットを指定することができます。

因子/共変量因子と共変量がリストされます。

位置モデル: モデルは、選択した主効果と交互作用効果によって異なります。

項の構築

選択した因子や共変量について、次の項を作成できます。

交互作用: 選択したすべての変数について、最高水準の交互作用項が作成されます。これはデフォルトです。

主効果: 選択した変数ごとに、主効果の項目が作成されます。

2 次まで: 選択した変数で考えられる 2 次の交互作用がすべて作成されます。

3 次まで: 選択した変数で考えられる 3 次の交互作用がすべて作成されます。

4 次まで: 選択した変数で考えられる 4 次の交互作用がすべて作成されます。

5 次まで: 選択した変数で考えられる 5 次の交互作用がすべて作成されます。

順序回帰の尺度モデル

「尺度」ダイアログ・ボックスを使用すると、分析用の尺度モデルを指定することができます。

因子/共変量因子と共変量がリストされます。

尺度モデル: モデルは、選択した主効果と交互作用効果によって異なります。

項の構築

選択した因子や共変量について、次の項を作成できます。

交互作用: 選択したすべての変数について、最高水準の交互作用項が作成されます。これはデフォルトです。

主効果: 選択した変数ごとに、主効果の項目が作成されます。

2 次まで: 選択した変数で考えられる 2 次の交互作用がすべて作成されます。

3 次まで: 選択した変数で考えられる 3 次の交互作用がすべて作成されます。

4 次まで: 選択した変数で考えられる 4 次の交互作用がすべて作成されます。

5 次まで: 選択した変数で考えられる 5 次の交互作用がすべて作成されます。

PLUM コマンドの追加機能

選択内容をシンタックス・ウィンドウに貼り付け、結果として生成された PLUM コマンド・シンタックスを編集すると、順序回帰をカスタマイズすることができます。コマンド・シンタックス言語を使用して、以下の処理を実行することもできます。

- 帰無仮説をパラメーターの線型結合として指定することにより、カスタマイズされた仮説の検定を作成する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 18 章 曲線推定

「曲線推定」手続きは、曲線推定の回帰統計量、および 11 種類の曲線推定の回帰モデルの関連プロットを作成します。従属変数ごとに個別のモデルが作成されます。予測値、残差、および予測区間を新しい変数として保存することもできます。

例: あるインターネット・サービス・プロバイダーが、自社のネットワーク上の E メール・トラフィックで、ウイルスに感染した Eメールの割合を長期的に追跡します。散布図では、その関係が非線型であることが示されます。2 次モデルまたは 3 次モデルをデータに適合して、仮定の妥当性とモデルの適合度を確認できます。

統計: 各モデル: 回帰係数、多重 R 、 R^2 、調整済み R^2 、推定値の標準誤差、分散分析テーブル、予測値、残差、および予測区間。モデル: 線型、対数、逆数、2 次、3 次、べき乗、複合、S 曲線、ロジスティック、成長、および指数。

曲線推定データの考慮事項

データ: 従属変数および独立変数は、量的でなければなりません。アクティブ・データ・セットから、独立変数として (変数の代わりに)「時刻」を選択した場合、「曲線推定」手続きにより、ケース間の時間の長さが一様な時間変数が生成されます。「時刻」が選択された場合、独立変数は時系列測度である必要があります。時系列分析に必要なデータ・ファイルの構成では、各ケース (行) が個別の時刻の一連の観測値を表し、ケース間の時間の長さが一様です。

仮定: データをグラフ表示し、独立変数と従属変数の関係 (線型、指数など) を判断します。よいモデルの残差は、ランダムに分布し、正規分布を示します。線型モデルを使用する場合は、以下の仮定を満たしている必要があります。独立変数の各値に対して、従属変数の分布は正規分布でなければなりません。従属変数の分布の分散は、独立変数のすべての値に対して一定でなければなりません。従属変数と独立変数の関係は線型で、観測値はすべて独立でなければなりません。

曲線推定を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「曲線推定...」

2. 1 つ以上の従属変数を選択します。従属変数ごとに個別のモデルが作成されます。

3. 独立変数を選択します (アクティブ・データ・セットの変数を選択するか、「時刻」を選択します)。

4. 必要に応じて、以下の操作を実行することができます。

- 散布図のケースにラベル付けする変数を 1 つ選択する。散布図の各点ごとに「点の識別」ツールを使用して、「ケースのラベル」変数の値を表示することができます。
- 「保存」をクリックして、予測値、残差、予測区間を新しい変数として保存する。

以下のオプションも使用することができます。

- **回帰式に定数項を含む:** 回帰方程式の定数項を予測します。デフォルトで、定数は含まれます。
- **モデルの作図:** 従属変数の値および選択された各モデルを独立変数に対して作図します。従属変数ごとに個別のグラフが作成されます。
- **分散分析表の表示:** 選択された各モデルに関して、要約された分散分析表を表示します。

曲線推定のモデル

1 つ以上の曲線推定回帰モデルを選択することができます。どのモデルを使用するかは、データをプロットして判断します。変数が線型関係を示している場合は、単純線型回帰モデルを使用します。変数が線型関係を示していない場合は、データを変換してください。データを変換しても線型関係にならない場合は、より複雑なモデルが必要になります。その場合は、データの散布図を表示し、自分が知っている数学関数にプロットが類似していれば、そのタイプのモデルにデータを適合させます。例えば、データが指数関数に類似している場合は、指数モデルを使用します。

線型：式が $Y = b_0 + (b_1 * t)$ であるモデル。系列の値は、時間の線型関数としてモデル化されます。

対数：式が $Y = b_0 + (b_1 * \ln(t))$ であるモデル。

逆数：式が $Y = b_0 + (b_1 / t)$ であるモデル。

2 次：式が $Y = b_0 + (b_1 * t) + (b_2 * t^{**2})$ であるモデル。2 次モデルは、上昇する系列または下降する系列をモデル化するために使用することができます。

3 次：式 $Y = b_0 + (b_1 * t) + (b_2 * t^{**2}) + (b_3 * t^{**3})$ で定義されるモデル。

べき乗：式が $Y = b_0 * (t^{**b_1})$ つまり $\ln(Y) = \ln(b_0) + (b_1 * \ln(t))$ であるモデル。

複合：式が $Y = b_0 * (b_1^{**t})$ または $\ln(Y) = \ln(b_0) + (\ln(b_1) * t)$ であるモデル。

S 曲線：式が $Y = e^{**}(b_0 + (b_1/t))$ つまり $\ln(Y) = b_0 + (b_1/t)$ であるモデル。

ロジスティック：式が $Y = 1 / (1/u + (b_0 * (b_1^{**t})))$ つまり $\ln(1/y-1/u) = \ln(b_0) + (\ln(b_1) * t)$ であるモデル。ここで、u は上限値です。「ロジスティック」を選択した後に、回帰式で使用する上限値を指定します。この値は、最大の従属変数値より大きい正の数値でなければなりません。

成長：式が $Y = e^{**}(b_0 + (b_1 * t))$ つまり $\ln(Y) = b_0 + (b_1 * t)$ であるモデル。

指数：式が $Y = b_0 * (e^{**}(b_1 * t))$ または $\ln(Y) = \ln(b_0) + (b_1 * t)$ であるモデル。

曲線推定の保存

変数の保存: 選択されたモデルごとに、予測値、残差 (従属変数の観測値からモデル予測値を引いた値)、および予測区間 (上限および下限) を保存できます。新しい変数名と説明用のラベルが出力ウィンドウのテーブルに表示されます。

ケースの予測: アクティブ・データ・セットで、独立変数として変数の代わりに「時刻」を選択した場合、時系列の最終点以降の予測期間を指定できます。以下のいずれかのオプションを選択することができます。

- **推定期間を基に最後のケースまでを予測**: 推定期間内のケースに基づき、ファイル内のすべてのケースについて値を予測します。ダイアログ・ボックスの下部に表示される推定期間は、「データ」メニューにある「ケースの選択」オプションの「範囲」サブダイアログ・ボックスで定義されます。推定期間が定義されていない場合は、すべてのケースを使用して値が予測されます。
- **指定による予測**: 推定期間でのケースに基づいて、指定した日付、時刻、または観測数までの値を推定します。この機能は、時系列の最後のケース以降の値を予測するために使用できます。現在定義されている日付変数によって、推定期間の最終点を指定するために使用可能なテキスト・ボックスが決まります。定義された日付変数がない場合は、最終観測 (ケース) 数を指定できます。

日付変数を作成するには、「データ」メニューの「日付の定義」オプションを使用します。

第 19 章 偏最小二乗回帰

「偏相関最小二乗法回帰」手続きは、偏相関最小二乗法 (PLS。潜在的構造投影方法 (projection to latent structure) と呼ばれます) 回帰モデルを推定します。PLS は、通常最小二乗法 (OLS) 回帰、正準相関、構造方程式モデリングに対する代替の予測技術で、予測変数が密接に相関している場合や、予測数がケース数を超えている場合に特に役立ちます。

PLS は、主成分分析と多重回帰の機能を結合したものです。PLS は、従属変数と独立変数の間の共分散を可能な限り説明する潜在的因子のセットを最初に抽出します。その後の回帰ステップで、独立変数の分解を使用して、従属変数の値が予測されます。

テーブル: デフォルトでは、(潜在的因子によって) 説明された分散の比率、潜在的因子の重み、潜在的因子負荷、投影の独立変数の重要度 (VIP)、回帰パラメーター推定値 (従属変数による) のすべてが作成されます。

図表: 投影の変数の重要度 (VIP)、因子得点、最初の 3 つの潜在的因子の因子の重み、モデルへの距離は、すべて「オプション」タブで生成されます。

偏相関最小二乗法回帰データの考慮事項

測定レベル: 従属変数と独立 (予測) 変数は、スケール変数、名義変数、順序変数のいずれかにすることができます。この手続きは、すべての変数に適切な尺度が割り当てられていることを前提とします。ただし、ソース変数リストで変数を右クリックし、ポップアップ・メニューから尺度を選択することにより、変数の尺度を一時的に変更することができます。この手続きでは、カテゴリ (名義または順序) 変数は同等に扱われます。

カテゴリ変数のコード化: この手続きは、手続きの期間について、one-of- c コード化を使用してカテゴリ従属変数を一時的に再割り当てします。変数の c カテゴリが存在する場合、その変数は c ベクトルとして格納され、最初のカテゴリ (1,0,...,0)、次のカテゴリ (0,1,0,...,0) などのように、最後のカテゴリ (0,0,...,0,1) までが示されます。カテゴリ従属変数は、ダミー・コードを使用して表わされます。つまり、参照カテゴリに対応する指標は単純に省略されます。

度数による重み付け: 重み変数は、最も近い整数に丸められてから使用されます。欠損重みを持つケースや重みが 0.5 未満のケースは、分析では使用されません。

欠損値: ユーザー欠損値とシステム欠損値は、無効な値として処理されます。

再調整: カテゴリ変数を表す指示変数を含め、すべてのモデル変数について中心化と標準化が行われます。

偏相関最小二乗法回帰を取得するには

メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「偏相関最小二乗法...」

1. 1 つ以上の従属変数を選択します。
2. 1 つ以上の独立変数を選択します。

必要に応じて、以下の操作を実行することができます。

- カテゴリー (名義または順位) 従属変数の参照カテゴリーを指定する。
- ケースごとの出力と保存されたデータ・セットの固有 ID として使用される変数を指定する。
- 抽出する潜在的因子数の上限を指定する。

前提条件

偏相関最小 2 乗法回帰手続きは、Python 拡張コマンドであり、IBM SPSS Statistics - Essentials for Python が必要です。これは、デフォルトで IBM SPSS Statistics 製品とともにインストールされます。また、無料で入手可能な NumPy および SciPy の各 Python ライブラリーも必要です。

注: ディストリビュート・アナリシス・モードで作業 (IBM SPSS Statistics Server が必要) を行うユーザーについては、NumPy と SciPy がサーバーにインストールされている必要があります。詳しくは、システム管理者に問い合わせてください。

Windows および Mac のユーザー

Windows および Mac の場合は、IBM SPSS Statistics とともにインストールされたバージョンとは別のバージョンの Python 2.7 に、NumPy と SciPy をインストールする必要があります。Python 2.7 の別のバージョンがない場合は、<http://www.python.org> からダウンロードできます。その上で、Python バージョン 2.7 用の NumPy と SciPy をインストールします。インストーラーは <http://www.scipy.org/Download> から入手できます。

NumPy と SciPy を使用できるようにするには、NumPy と SciPy をインストールしたバージョンの Python 2.7 に、Python の場所を設定する必要があります。Python の場所は、「オプション」ダイアログ (「編集」>「オプション」) の「ファイルの場所」タブから設定します。

Linux ユーザー

ソースをダウンロードして NumPy と SciPy をユーザー自身で作成することをお勧めします。ソースは <http://www.scipy.org/Download> から入手できます。NumPy と SciPy は、IBM SPSS Statistics とともにインストールされたバージョンの Python 2.7 にインストールできます。これは、IBM SPSS Statistics のインストール場所にある Python ディレクトリー内にあります。

NumPy と SciPy を、IBM SPSS Statistics とともにインストールされたバージョンとは別のバージョンの Python 2.7 にインストールすることを選択する場合は、Python の場所がそのバージョンを指すように設定する必要があります。Python の場所は、「オプション」ダイアログ (「編集」>「オプション」) の「ファイルの場所」タブから設定します。

Windows および Unix サーバー

NumPy と SciPy は、サーバーで、IBM SPSS Statistics に付属してインストールされるバージョンとは別個のバージョンの Python 2.7 にインストールする必要があります。別個のバージョンの Python 2.7 がサーバーにない場合、<http://www.python.org> からダウンロードできます。Python 2.7 用の NumPy と SciPy は、<http://www.scipy.org/Download> から入手できます。NumPy と SciPy を使用できるようにするには、サーバーに対する Python の場所を、NumPy と SciPy がインストールされているバージョンの Python 2.7 に設定する必要があります。Python の場所は、IBM SPSS Statistics Administration Console から設定します。

モデル

モデル効果の指定: 主効果モデルには、すべての因子と共変量の主効果が含まれます。交互作用を指定するには、「**ユーザー指定**」を選択します。モデルに含めるすべての項目を指定する必要があります。

因子と共変量: 因子と共変量がリストされます。

モデル: モデルは、使用するデータの性質によって異なります。「**ユーザーの指定**」を選択すると、分析対象の主効果と交互作用を選択できるようになります。

項の構築

選択した因子と共変量について、以下の項目を作成することができます。

交互作用: 選択したすべての変数について、最高水準の交互作用項が作成されます。これはデフォルトです。

主効果: 選択した変数ごとに、主効果の項目が作成されます。

2 次まで: 選択した変数で考えられる 2 次の交互作用がすべて作成されます。

3 次まで: 選択した変数で考えられる 3 次の交互作用がすべて作成されます。

4 次まで: 選択した変数で考えられる 4 次の交互作用がすべて作成されます。

5 次まで: 選択した変数で考えられる 5 次の交互作用がすべて作成されます。

オプション

「オプション」タブでは、個々のケース、潜在的因子、予測変数のモデル推定値を保存し、プロットすることができます。

データの各型に対し、データ・セットの名前を指定します。データ・セット名は一意でなければなりません。既存のデータ・セットの名前を指定した場合、そのデータ・セットの内容が置き換えられます。それ以外の場合は、新しいデータ・セットが作成されます。

- **各ケースの推定値を保存:** 予測値、残差、潜在的因子モデルへの距離、潜在的因子スコアのケースごとのモデル推定値を保存します。また、潜在的因子スコアを作図します。
- **潜在的因子の推定値を保存:** 潜在的因子負荷と潜在的因子の重みを保存します。また、潜在的因子の重みを作図します。
- **独立変数の推定値を保存:** 回帰パラメーターの推定値と投影変数の重要度 (VIP) を保存します。また、潜在的因子ごとに VIP を作図します。

第 20 章 最近傍分析

最近傍分析は、他のケースに対する類似性に基づいてケースを分類する方法です。この方法は、マシン学習において、保存されたパターンやケースに完全に一致することを必要とせずにデータのパターンを認識する方法として開発されました。類似したケースは相互に近い位置に存在し、類似していないケースは相互に離れた位置に存在します。そのため、2 つのケース間の距離は、それらのケースの非類似度を示す尺度になります。

相互に近いケースは「近傍」と呼ばれます。新しいケース (ホールドアウト) が存在する場合、モデル内の各ケースからその新しいケースへの距離が計算されます。最も類似したケース (最近傍) の分類が集計され、最近傍が最も多く存在するカテゴリーに新しいケースが配置されます。

調査対象の最近傍数を指定できます。この値は k と呼ばれます。

また、最近傍分析を使用して、連続型対象の値を計算することもできます。この場合、最近傍の平均対象値または中央対象値を使用して、新しいケースの予測値が取得されます。

最近傍分析データの考慮事項

目標および特徴: 目標および特徴は次のとおりです。

- *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。
- *Scale* (スケール データ). 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。

最近傍分析では、名義変数と順序変数が同等に処理されます。この手続きは、各変数に適切な尺度が割り当てられていることを前提とします。ただし、ソース変数リストで変数を右クリックし、ポップアップメニューから尺度を選択することにより、変数の尺度を一時的に変更することができます。

変数リストで各変数の隣にあるアイコンは、以下のような尺度とデータ型を表します。

表 1. 測定レベル・アイコン

	数値	文字列	日付	時間
スケール (連続)		n/a		
順序				
名義				

カテゴリ変数のコード化: この手続きは、手続きの期間について、one-of-c コード化を使用してカテゴリ予測変数と従属変数を一時的に再割り当てします。変数の c カテゴリが存在する場合、その変数は c ベクトルとして格納され、最初のカテゴリ $(1,0,\dots,0)$ 、次のカテゴリ $(0,1,0,\dots,0)$ などのように、最後のカテゴリ $(0,0,\dots,0,1)$ までが示されます。

このコード化方式では、特徴空間の次元数が増加します。具体的には、次元の合計数は、スケール予測変数の数とすべてのカテゴリ予測変数のカテゴリ数の合計になります。そのため、このコード化方式により、学習の速度が遅くなる場合があります。最近傍学習の速度が遅い場合、手続きを実行する前に、類似したカテゴリを結合するか、非常にまれなカテゴリを持つケースを削除して、カテゴリ予測変数のカテゴリ数を削減することをお勧めします。

ホールドアウト・サンプルが定義されている場合でも、one-of-c コード化はすべて学習データに基づきます (88 ページの『分割』を参照)。そのため、学習データには存在しない予測カテゴリを持つケースがホールドアウト・サンプルに含まれている場合、そのケースはスコア化されません。学習データには存在しない従属変数カテゴリを持つケースがホールドアウト・サンプルに含まれている場合、そのケースはスコア化されます。

再調整: スケール機能はデフォルトで標準化されます。ホールドアウト・サンプルが定義されている場合でも、再調整はすべて学習データに基づきます (88 ページの『分割』を参照)。変数を指定して分割を定義する場合、特徴の分布が、学習サンプルとホールドアウト・サンプルで同様になることが重要です。分割全体にわたる分布を調べるには、探索的分析手続きなどを使用してください。

度数による重み付け: 度数による重み付けは、この手続きでは無視されます。

結果の再現: この手続きは、分割と交差検証群の無作為割り当て時に乱数ジェネレーターを使用します。結果を正確に再現するには、同じ手続きの設定を使用するだけでなく、Mersenne Twister のシード (88 ページの『分割』を参照) を設定するか、変数を使用して分割と交差検証群を定義します。

最近傍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。

「分析」 > 「分類」 > 「最近傍法...」

1. 目標がある場合に独立変数または予測変数と見なすことができる 1 つ以上の特徴を指定します。

目標 (オプション): 目標 (従属変数または応答) が指定されていない場合、この手続きは k 最近傍だけを検出します。分類や予測は実行されません。

スケール機能を標準化: 標準化された機能の値は同じ範囲になるため、推定アルゴリズムのパフォーマンスが向上します。調整済み正規化の $[2*(x-\min)/(\max-\min)]-1$ が使用されます。調整済み正規化の値は -1 から 1 までです。

中心ケース識別子 (オプション): 特に関心のあるケースをマークすることができます。例えば、ある研究者が、ある学区の検定スコア (中心ケース) が、類似した学区の検定スコアと比較可能かどうかを確認したいと考えているとします。この研究者は、最近傍分析を使用して、特定の特徴のセットに関して最も類似している学区を検出します。次に、中心となる学区の検定スコアと最近傍の検定スコアを比較します。

中心ケースを臨床研究で使用して、臨床ケースに類似した対照ケースを選択することもできます。中心ケースは、 k 最近傍と距離の表、特徴空間図、同位図、四分位分布図で表示されます。中心ケースについての情報は、「出力」タブで指定されたファイルに保存されます。

指定された変数に正の値を持つケースは、中心ケースとして処理されます。正の値を持たない変数を指定することはできません。

ケースのラベル (オプション): これらの値を使用して、特徴空間図、同位図、四分位分布図のケースにラベルが付けられます。

不明な尺度のフィールド

データ・セット内の 1 つ以上の変数 (フィールド) の測定レベルが不明な場合、測定レベルの警告が表示されます。測定レベルはこの手続きの結果の計算に影響を与えるため、すべての変数について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込み、デフォルトの測定レベルを、測定レベルが現在不明なすべてのフィールドに割り当てます。データ・セットのサイズが大きい場合、この処理には時間がかかります。

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

この手続きでは測定レベルが重要であるため、すべてのフィールドに対して測定レベルが定義されるまで、ダイアログにアクセスしてこの手続きを実行することはできません。

近傍

最近傍数 (k): 最近傍数を指定します。使用する近傍の数が多いほど、より正確なモデルが作成されるとは限らないことに注意してください。

目標が「変数」タブで指定されている場合、代わりに値の範囲を指定すると、その範囲内の「最適な」近傍数が手続きによって選択されます。最近傍数を決定する方法は、特徴選択が「特徴」タブで要求されているかどうかによって異なります。

- 特徴選択が有効な場合、特徴選択は要求された範囲の k の各値について実行され、最も低い誤差率 (または目標がスケールの場合、最も低い平方和の誤差) を持つ k とそれに付随する特徴セットが選択されます。
- 特徴選択が有効ではない場合、 V 群交差検証を使用して、「最適な」近傍数が選択されます。群の割り当ての制御については、「分割」タブを参照してください。

奥行き計算: ケースの類似度の測定に使用される距離基準を指定するための計量です。

- **ユークリッド計量:** x と y の 2 つのケース間の距離は、すべての次元において、それらのケースの値の差の平方和の平方根になります。
- **都市ブロック計量:** 2 つのケースの間の距離は、すべての次元において、それらのケースの値の絶対差の合計になります。これは Manhattan 距離とも呼ばれます。

目標が「変数」タブで指定されている場合、オプションで、距離の計算時に、正規化された重要度によって特徴に重みを付けることができます。予測変数の特徴重要度は、誤差率の比率またはモデルの誤差の平方和と、モデルからモデル全体の誤差率または誤差の平方和に移動された予測変数によって計算されます。正規化された重要度は、合計が 1 になるように特徴重要度の値を再度重み付けすることによって計算されます。

スケール目標の予測: スケール目標が「変数」タブで指定されている場合、予測値が平均値と最近傍の中央値のどちらに基づいて計算されるかを指定します。

特徴

目標が「変数」タブで指定されている場合、「特徴」タブを使用して、特徴選択オプションの要求と指定を行うことができます。デフォルトでは、すべての特徴が特徴選択用に考慮されていますが、オプションで、特徴のサブセットを選択してモデル内に適用することができます。

停止基準: 各ステップでは、モデルに追加した場合に誤差 (カテゴリー目標の誤差率と、スケール目標の誤差の平方和として計算) が最も小さくなる特徴が、モデル・セットに含める特徴であるとみなされます。変数増加法は、指定された条件を満たすまで続行されます。

- **指定される特徴数:** アルゴリズムは、モデル内に強制的に適用される特徴のほかに、ここで指定された数の特徴を追加します。正の整数を指定してください。小さな数値を指定すると、より節約的なモデルが作成されますが、重要な特徴が欠損する可能性があります。大きな数値を指定すると、重要な特徴がすべて取り込まれますが、モデル誤差を増加させる特徴も追加される可能性があります。
- **絶対誤差比の最小変化量:** 絶対誤差比の変化量が、これ以上特徴を追加してもモデルが改善されないことを示した時点で、アルゴリズムが停止します。正の数値を指定してください。最小変化量の値を小さくすると、より多くの特徴が取り込まれるようになりますが、モデルに対して重要な価値を与えない機能も追加される可能性があります。最小変化量の値を大きくすると、より多くの特徴が除外されるようになりますが、モデルにとって重要な特徴が失われる可能性があります。最小変化量の「最適な」値は、データとアプリケーションによって異なります。どの特徴が最も重要であるかを評価するには、出力の特徴選択エラー・ログを確認します。詳しくは、92 ページの『特徴選択エラー・ログ』のトピックを参照してください。

分割

「分割」タブを使用して、データ・セットを学習セットとホールドアウト・セットに分割し、必要に応じてケースを交差検証群に割り当てることができます。

学習およびホールドアウトの分割: このグループは、アクティブなデータ・セットを学習サンプルとホールドアウト・サンプルに分割する方法を指定します。学習サンプルは、最近傍モデルの学習で使用するデータ・レコードから構成されています。モデルを取得するには、データ・セット内のケース何割かを学習サンプルに割り当てる必要があります。ホールドアウト・サンプルは、最終モデルの評価に使用されるデータ・レコードから構成された、独立したセットです。ホールドアウト・ケースはモデルの作成には使用されないため、ホールドアウト・サンプルの誤差は、モデルの予測能力の「公正な」評価を示します。

- **無作為にケースを分割に割り当て:** 学習サンプルに割り当てるケースのパーセントを入力します。残りはホールドアウト・サンプルに割り当てられます。
- **ケースの割り当てに変数を使用:** アクティブなデータ・セット内の各ケースを学習サンプルまたはホールドアウト・サンプルに割り当てる数値型変数を指定します。変数に正の値を持つケースは学習サンプルに割り当てられ、0 の値または負の値を持つケースはホールドアウト・サンプルに割り当てられます。システム欠損値を持つケースは、分析から除外されます。分割変数のユーザー欠損値は、常に有効な値として扱われます。

交差検証群: V 群交差検証を使用して、近傍の「最適な」数が判断されます。パフォーマンス上の理由から、これを特徴選択と組み合わせて使用することはできません。

交差検証では、サンプルが群と呼ばれる複数のサブサンプルに分割されます。その後、各サブサンプルから順にデータを除外して、最近傍モデルが生成されます。最初のモデルは、最初のサンプル群のケースを除くすべてのケースに基づき、2 番目のモデルは、2 番目のサンプル群のケースを除くすべてのケースに基づきます。3 番目以降のモデルについても同様です。各モデルについて、そのモデルの生成時に除外されたサブサンプルに適用することにより、誤差が推定されます。群全体で最も誤差が少ないモデルの数が、最近傍の「最適な」数になります。

- **無作為にケースを群に割り当て:** 交差検証で使用する群の数を指定します。この手続きでは、1 から V (群の数) までの番号が付けられた群に、無作為にケースが割り当てられます。
- **ケースの割り当てに変数を使用:** アクティブなデータ・セット内の各ケースを群に割り当てる数値型変数を指定します。この変数は、1 から V までの値を取る数値型でなければなりません。この範囲内の値が欠損していたり、分割されていたりすると (分割ファイルが有効な場合)、誤差が生じます。

Mersenne Twister のシードを設定: シードを設定すると、分析を複製することができます。このコントロールを使用すると、アクティブ・ジェネレーターとして Mersenne Twister を設定し、「乱数ジェネレーター」ダイアログの固定開始ポイントを指定した場合と同様の結果になりますが、このダイアログでシードを設定することによる大きな違いは、乱数ジェネレーターの現在の状態が保存され、分析の完了後にその状態が復元されることです。

保存

保存する変数の名前: 自動的な名前の生成の場合、すべての作業が保存されます。ユーザー指定の名前の場合、データ・エディターで保存された変数を最初に削除することなく、前回実行された結果を破棄するか置き換えることができます。

保存する変数

- **予測値またはカテゴリー:** スケール目標の予測値またはカテゴリー目標の予測カテゴリーを保存します。
- **予測確率:** カテゴリー目標の予測確率を保存します。最初の n カテゴリーのそれぞれについて、個別の変数が保存されます。 n は、「カテゴリー目標のために保存する最大カテゴリー数」コントロールで指定された値です。
- **学習/ホールドアウトの分割変数:** 「分割」タブで、ケースが学習サンプルとホールドアウト・サンプルに無作為に割り当てられている場合、ケースが割り当てられた分割 (学習またはホールドアウト) の値を保存します。
- **交差検証群変数:** 「分割」タブでケースが群に無作為に割り当てられている場合、ケースが割り当てられた群の値を保存します。

出力

ビューアー出力

- **ケース処理の要約:** ケース処理要約テーブルを表示します。このテーブルには、分析内のケースの数と除外されたケースの数が、全体、学習サンプルごと、ホールドアウト・サンプルごとに要約されて表示されます。
- **図表と表:** 表や図表など、モデルに関連する出力を表示します。モデル・ビューのテーブルには、中心ケースの k 最近傍と距離、カテゴリー応答変数の分類、誤差の集計が表示されます。モデル・ビューのグラフィカル出力には、選択エラー・ログ、特徴重要度図表、特徴空間図表、同位図、4 分位分布図があります。詳しくは、90 ページの『モデル・ビュー』のトピックを参照してください。

ファイル

- **モデルを XML にエクスポート:** このモデル・ファイルを使用して、スコアリングのために他のデータ・ファイルにモデル情報を適用できます。分割ファイルが定義されている場合、このオプションは使用できません。
- **中心ケースと k 最近傍との間の距離をエクスポート:** 中心ケースの場合、中心ケースの (学習サンプルからの) k 最近傍と、対応する k 最短距離について、それぞれ個別の変数が作成されます。

オプション

ユーザー欠損値: カテゴリ変数は、分析に含まれるケースに対して有効な値を取る必要があります。これらのコントロールを使用して、ユーザー欠損値をカテゴリ変数で有効な値として処理するかどうかを判断することができます。

システム欠損値とスケール変数の欠損値は、常に無効な値として処理されます。

モデル・ビュー

「出力」タブで「**図表と表**」を選択すると、この手続きにより、ビューアーに最近傍モデル・オブジェクトが作成されます。このオブジェクトをダブルクリックしてアクティブにすると、モデルの双方向ビューが表示されます。このモデル・ビューには、以下の 2 つのパネルで構成されたウィンドウがあります。

- 最初のパネルはメイン・ビューと呼ばれ、モデルの概要が表示されます。
- 2 番目のパネルには、次の 2 種類のビューのいずれかが表示されます。

モデルの詳細が表示され、モデル自体には焦点を当てていない補助的なモデル・ビュー。

ユーザーがメイン・ビューの一部をドリルダウンした場合に、モデルのいずれかの特徴に関する詳細を表示するリンク・ビュー。

デフォルトでは、最初のパネルに特徴空間が表示され、2 番目のパネルに変数の重要度グラフが表示されます。変数の重要度グラフが使用できない場合 (つまり、「特徴」タブで「**重要度によって重みを付ける**」が選択されていない場合)、「ビュー」ドロップダウンで最初に使用できるビューが表示されます。

ビューに有効な情報がない場合、「ビュー」ドロップダウンの該当する項目テキストが無効になります。

特徴空間

特徴空間図は、特徴空間 (特徴が 4 つ以上ある場合は部分空間) のインタラクティブ・グラフです。それぞれの軸はモデル内の特徴を示し、グラフ上の点の場所が、学習分割とホールドアウト分割のケースに対するこれらの特徴の値を示します。

キー: 特徴の値のほかに、図表内の点がその他の情報を表します。

- 形状は、点が属する分割 (学習またはホールドアウト) を示します。
- 点の色/網掛けは、そのケースの目標の値を示します。それぞれの色の値は、カテゴリ型目標のカテゴリに対応し、網掛けは連続型目標の値の範囲を示します。学習分割で示される値は、観測値です。ホールドアウト分割の場合、この値は予測値になります。目標が指定されていない場合、このキーは表示されません。
- 太い枠線は、そのケースが中心ケースであることを示しています。中心ケースは、その中心ケースの k 最近傍にリンクされて表示されます。

コントロールおよび双方性: 図表内の各種コントロールを使用して、特徴空間を調べることができます。

- 図表内に表示する特徴のサブセットを選択したり、次元上に表示される特徴を変更することができます。
- 「中心ケース」は、特徴空間図で選択された単なる点です。中心ケース変数を指定すると、中心ケースを表す点が最初に選択されます。ただし、どの点を選択しても、その点が一時的に中心ケースになります。点の選択には「通常の」コントロールが適用されます。つまり、点をクリックすると、その点を選択され、他のすべての点を選択解除されます。Ctrl キーを押しながら点をクリックすると、その点が、選択された点のセットに追加されます。同位図などのリンク・ビューは、特徴空間で選択されたケースに基づいて自動的に更新されます。
- 中心ケース用に表示する最近傍 (k) の数は、変更することができます。
- カーソルを図表内の点に移動すると、ケース・ラベルの値 (ケース・ラベルが定義されていない場合はケース数)、観測目標値、予測目標値を示すツールチップが表示されます。
- 特徴空間を元の状態に戻すには、「戻す」ボタンを使用します。

フィールド/変数の追加と削除

新しいフィールド/変数を特徴空間に追加したり、現在表示されているフィールド/変数を削除したりすることができます。

変数パレット

変数の追加や削除を行う前に、変数パレットを表示する必要があります。変数パレットを表示するには、モデル・ビューアーを編集モードにし、ケースを特徴空間で選択する必要があります。

1. モデル・ビューアーを編集モードにするには、メニューから次の項目を選択します。

「表示」 > 「編集モード」

2. 編集モードになったら、特徴空間でいずれかのケースをクリックします。
3. 変数パレットを表示するには、メニューから次の項目を選択します。

「表示」 > 「パレット」 > 「変数」

変数パレットに、特徴空間のすべての変数のリストが表示されます。変数名の隣のアイコンは、変数の測定の尺度を示しています。

4. 変数の測定の尺度を一時的に変更するには、変数パレットの変数を右クリックして、いずれかのオプションを選択します。

変数領域

変数は、特徴空間の「領域」に追加されます。領域を表示するには、変数パレットから変数をドラッグするか、「領域を表示」を選択します。

特徴空間には、 x 軸、 y 軸、 z 軸の各領域があります。

変数を領域に移動するには

変数を領域に移動する場合、以下に示す一般的な規則とヒントを参照してください。

- 変数を領域に移動するには、変数をクリックして変数パレットからドラッグし、領域にドロップします。「領域を表示」を選択した場合、領域を右クリックして、領域に追加する変数を選択することもできます。

- 既に別の変数が存在する領域に変数パレットから変数をドラッグすると、既存の変数が新しい変数に置き換えられます。
- ある領域の変数を、既に別の変数が存在する領域にドラッグすると、変数の位置が入れ替わります。
- 領域内の「X」をクリックすると、その領域から変数が削除されます。
- 視覚化に複数のグラフィック要素がある場合、各グラフィック要素は、それぞれに関連付けられた独自の変数領域を持つことができます。最初に、グラフィック要素を選択してください。

変数の重要度

通常は、最も重要な変数を中心にモデリング作業を行い、あまり重要ではない変数については、削除するか無視したいと考えます。そのために役立つのが、変数の重要度グラフです。このグラフは、モデルの推定において、各変数の相対重要度を示します。変数の値は相対値であるため、表示されるすべての変数の値の合計は 1.0 になります。変数の重要度は、モデルの精度には関係しません。変数の重要度は、予測が正確であるかどうかではなく、予測時の各変数の重要度に関係します。

同位

この図には、各特徴と目標について、中心ケースとその k 最近傍が表示されます。この図は、特徴空間で中心ケースが選択されている場合に使用することができます。

リンク動作: 同位図は、以下の 2 つの方法で特徴空間にリンクされます。

- 同位図には、特徴空間で選択された (中心) ケースがその k 最近傍とともに表示されます。
- 同位図では、特徴空間で選択された k の値が使用されます。

最近傍の距離

このテーブルには、中央ケースの k 最近傍と距離だけが表示されます。このテーブルは、「変数」タブで中心ケースの識別子が指定されている場合に使用でき、この変数によって識別される中心ケースだけが表示されます。

各行の内容は以下のとおりです。

- 「中心ケース」列には、中心ケースのケース・ラベル変数の値が表示されます。ケース・ラベルが定義されていない場合、この列には中心ケースのケース数が表示されます。
- 最近傍グループの i 番目の列には、中心ケースの i 番目の最近傍のケース・ラベル変数の値が表示されます。ケース・ラベルが定義されていない場合、この列には中心ケースの i 番目の最近傍のケース番号が表示されます。
- 最近傍の距離グループの i 番目の列には、中心ケースへの i 番目の最近傍の距離が表示されます。

四分位分布図

この図には、中心ケースとその k 最近傍が、散布図 (または、目標の尺度に応じてドット・プロット) で表示されます。y 軸は目標、x 軸はスケールの特徴を表し、特徴ごとにパネル化されます。この図は、目標が指定されていて、中心ケースが特徴空間で選択されている場合に使用することができます。

- 連続型変数の場合、学習分割の変数の平均値で参照線が描画されます。

特徴選択エラー・ログ

グラフ内の点は、y 軸上ではモデルの誤差 (目標の尺度に応じて、誤差率または平方和の誤差のいずれか) を示し、x 軸上ではリストされた特徴を示します (さらに、x 軸の左側にすべての特徴が表示されます)。このグラフは、特徴選択が有効な場合に使用することができます。

k 選択エラー・ログ

図表内の点は、 y 軸上ではモデルの誤差 (目標の尺度に応じて、誤差率または平方和の誤差のいずれか) を示し、 x 軸上では最近傍 (k) の数を示します。この図表は、目標が指定されていて、 k 選択が有効な場合に使用することができます。

k および特徴選択エラー・ログ

これらの特徴選択図表 (92 ページの『特徴選択エラー・ログ』を参照) は、 k ごとにパネル化されます。この図表は、目標が指定されていて、 k と特徴選択の両方が有効な場合に使用することができます。

分類テーブル

このテーブルには、目標の観測値と予測値のクロス分類が分割ごとに表示されます。目標が指定されていて、その目標がカテゴリ型である場合、このテーブルを使用することができます。

- ホールドアウト分割の「(欠損値)」行には、目標に欠損値が存在するホールドアウト・ケースが表示されます。これらのケースは、「ホールドアウト・サンプル: すべてのパーセント」の値に影響しますが、「正分類パーセント」の値には影響しません。

誤差の集計

このテーブルは、目標変数が存在する場合に使用することができます。このテーブルには、モデルに関連する誤差が表示されます。連続型目標の場合は平方和が表示され、カテゴリ型目標の場合は誤差率 (100% - 全正分類パーセント) が表示されます。

第 21 章 判別分析

判別分析により、所属グループ用の予測モデルが作成されます。このモデルは、各グループを最も適切に判別する予測変数の線型結合に基づいた 1 つの判別関数 (複数のグループの場合は、判別関数のセット) から構成されます。各関数は、所属グループが判明しているケースのサンプルから生成されます。各関数は、予測変数の測定は存在するが所属グループが不明な新規ケースに適用することができます。

注: グループ化変数には、3 つ以上の値を設定することができます。ただし、グループ化変数のコードは整数でなければならず、変数の最小値と最大値を指定する必要があります。この範囲外の値を持つケースは、分析から除外されます。

例: 平均的に、温帯の国に住んでいる人は、熱帯に住んでいる人に比べて 1 日当たりのカロリー消費量が多く、さらに温帯のほうが都市人口の比率が高くなっています。この 2 つのグループの国々をどの程度正確に区別できるかを判断するため、研究者はこれらの情報を 1 つの関数にまとめたいと考えています。さらに、研究者は、人口規模や経済情報も重要であると考えています。この場合は判別分析を使用すると、多重線型回帰式の右側の項に類似した線型判別関数の係数を推定することができます。つまり、係数 a 、 b 、 c 、 d を使用した場合、この関数は以下のようになります。

$$D = a * climate + b * urban + c * population + d * gross\ domestic\ product\ per\ capita$$

これらの変数が 2 つの気候帯の判別に役立つ場合、温帯の国と熱帯の国では D の値が異なることとなります。ステップワイズ変数選択法を使用すると、4 つの変数のすべてを関数に含める必要はないことがわかります。

統計: 各変数: 平均値、標準偏差、1 変量の分散分析。各分析: Box の M 、グループ内相関行列、グループ内共分散行列、グループ別共分散行列、総共分散行列。各正準判別関数: 固有値、分散の割合、正準相関、Wilks のラムダ、カイ 2 乗。各ステップ: 事前確率、Fisher の分類関数の係数、標準化されていない分類関数の係数、各正準関数に対する Wilks のラムダ。

判別分析データの考慮事項

データ: グループ化変数は、整数でコード化された一定数のカテゴリーを持っている必要があります。名義型の独立変数は、ダミー変数または対比変数として再割り当てする必要があります。

仮定: ケースは独立している必要があります。予測変数は多変量正規分布を持っている必要があります。グループ内分散共分散行列はグループ全体で等しくなければなりません。各所属グループは相互に排他的であり (つまり、複数のグループに所属しているケースが存在しない)、グループ全体で包括的である (つまり、すべてのケースがいずれかのグループに所属している) と想定されます。この手続きは、所属グループがカテゴリー変数である場合に最も効果的です。所属グループが連続変数の値 (例えば、高い IQ と低い IQ など) に基づいている場合は、連続変数そのものが提供する詳細な情報を最大限に活用するために、線型回帰を使用することをお勧めします。

判別分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「分類」 > 「判別分析...」

2. 整数値をとるグループ変数を選択し、「範囲の定義」をクリックして目的のカテゴリーを指定します。

3. 独立変数 (つまり、予測変数) を選択します。グループ化変数の値が整数ではない場合、「変換」メニューの「連続数への再割り当て」を選択すると、整数値をとる変数が作成されます。
4. 独立変数を投入する方法を選択します。
 - 同時に独立変数を投入: 許容基準を満たすすべての独立変数を同時に投入します。
 - ステップワイズ法を使用: ステップワイズ法の分析を使用して、変数の投入と削除を制御します。
5. オプションで、ケース選択変数を選択します。

判別分析: 範囲の定義

分析で使用するグループ化変数の最小値と最大値を指定します。この範囲外の値のケースは、判別分析では使用されませんが、分析結果に基づいて既存のグループのいずれかに分類されます。最小値と最大値は整数でなければなりません。

判別分析: ケースの選択

分析用にケースを選択するには、以下の手順を実行します。

1. 「判別分析」ダイアログ・ボックスで、ケース選択変数を選択します。
2. 「値」をクリックし、選択値として使用する整数を入力します。

ここで指定されたケース選択変数の値を持つケースだけを使用して、判別関数が作成されます。統計と分類の結果は、選択されたケースと選択解除されたケースの両方に対して生成されます。このプロセスは、既存のデータに基づいて新しいケースを分類するための手段になります。または、データを学習用と検定用のサブセットに分割し、作成されたモデルの有効性を確認するための手段になります。

判別分析: 統計

記述統計: 使用可能なオプションは、平均 (標準偏差を含む)、1 変量の分散分析、Box の M 検定です。

- *Means* (平均値 (信頼性分析)). 独立変数の合計、グループ平均値、および標準偏差を表示します。
- *Univariate ANOVAs* (1 変量の分散分析 (判別分析)). 一元配置分散分析を実行して、独立変数ごとにグループ平均値の等質性を検定します。
- *Box* の M 検定: グループの共分散行列の等質性を調べる検定。サンプル数が十分に多いが p 値が有意でない場合は、行列が異なるという証拠が不十分であることを意味します。この検定は、多変量正規性からの逸脱に対して敏感です。

関数係数: 使用可能なオプションは、Fisher の分類係数と非標準化係数です。

- *Fisher's* (Fisher の分類関数の係数). 分類に直接使用できる、Fisher の分類関数の係数を表示します。分類関数の一連の係数をグループごとに個別に求め、最大判別得点 (分類関数の値) を持つグループにケースを割り当てます。
- *Unstandardized* (標準化されていない (判別分析)). 標準化していない判別関数の係数を表示します。

行列: 独立変数の使用可能な係数の行列は、グループ内相関行列、グループ内共分散行列、グループ別共分散行列、全共分散行列です。

- **グループ内相関:** 相関を計算する前にすべてのグループの個別の共分散行列を平均化することによって得られるプールされたグループ内相関行列を表示します。
- **グループ内共分散:** プールされたグループ内共分散行列を表示します。全共分散行列とは異なる場合があります。この行列は、すべてのグループの個別の共分散行列を平均化することによって得られます。

- **グループ別共分散:** 各グループの個別の共分散行列を表示します。
- **全共分散:** すべてのケースから得た共分散行列を、1つのサンプルから取り出したかのように表示します。

判別分析: ステップワイズ法

方法: 新しい変数を投入または削除する際に使用される統計方法を選択します。有効な選択肢は、Wilks のラムダ、解明不明の分散、Mahalanobis の距離、最小 F 比率、Rao の V です。Rao の V を使用すると、投入する変数に対して V 単位の最小増分を指定することができます。

- **Wilks のラムダ:** ステップワイズ判別分析における変数選択法の 1 つ。変数が Wilks のラムダを低下させる程度に基づいて式に投入する変数を選択します。各ステップでは、Wilks のラムダが最小になる変数を投入します。
- **解明不明の分散:** 各ステップで、グループ間の説明されない分散の合計を最小化する変数を投入します。
- **Mahalanobis の距離:** 独立変数のケースの値と全ケースの平均との差異の程度を示す指標。マハラノビスの距離が大きい場合は、ケースにおいて 1 つ以上の独立変数に極値が存在することを示します。
- **最小 F 比率:** グループ間のマハラノビスの距離から計算した F 比の最大化に基づく、ステップワイズ分析での変数選択法。
- **Rao の V :** グループ平均値の差の指標。Lawley-Hotelling のトレースとも呼びます。各ステップで、Rao の V における増加を最大化する変数を投入します。このオプションを選択した後、分析に投入する変数が持つべき最小値を入力してください。

「基準」。有効な選択肢は、「ステップワイズのための F 値」と「ステップワイズのための F 値確率」です。変数の投入用の値と削除用の値を入力します。

- **ステップワイズのための F 値:** F 値が「投入」の値より大きい場合に変数をモデルに投入し、「削除」の値より小さい場合に変数を除去します。「投入」は「削除」より大きくなければならず、いずれの値も正でなければなりません。さらに多くの変数をモデルに投入するには、「投入」の値を下げてください。さらに多くの変数をモデルから除去するには、「除去」の値を上げてください。
- **ステップワイズのための F 値確率:** F 値の有意水準が「投入」の値より小さい場合に変数をモデルに投入し、有意水準が「削除」の値より大きい場合に変数を除去します。「投入」は「削除」より小さくなければならず、いずれの値も正でなければなりません。さらに多くの変数をモデルに投入するには、「投入」の値を上げてください。さらに多くの変数をモデルから除去するには、「除去」の値を下げてください。

表示: 「ステップの集計」を選択すると、ステップごとにすべての変数の統計量が表示されます。「ペアごとの Mahalanobis の距離による F 値」を選択すると、グループの各ペアについてペアごとの F 比の行列が表示されます。

判別分析: 分類

事前確率: このオプションは、所属グループの事前の知識に応じて分類係数を調整するかどうかを決定します。

- **すべてのグループが等しい:** すべてのグループについて同じ事前確率が想定されます。係数に対する影響はありません。
- **グループ・サイズから計算:** サンプル内のグループ・サイズの観測結果により、所属グループの事前確率が決定されます。例えば、分析に含まれる観測結果の 50% が最初のグループ、25% が 2 番目のグループ、残りの 25% が 3 番目のグループに分類される場合、分類係数は、他の 2 つのグループに対して最初のグループ内の所属性の尤度を増やすように調整されます。

表示: 使用できる表示オプションは、「ケースごとの結果」、「集計表」、「Leave-one-out 分類」です。

- ケースごとの結果: 実際のグループ、予測グループ、事後確率、および判別得点のコードをケースごとに表示します。
- 集計表: 判別分析に基づいて各グループに正しくまたは誤って割り当てられたケースの数。「混同行列」と呼ぶこともあります。
- Leave-one-out 分類: 分析における各ケースを、そのケース以外のすべてのケースから派生した関数で分類します。「U 手法」とも呼びます。

欠損値を平均値で置換: このオプションを選択すると、分類段階でのみ、欠損値が独立変数の平均値に置き換えられます。

共分散行列の使用: グループ内の共分散行列または個別グループ共分散行列を使用してケースを分類することができます。

- グループ内: プールされたグループ内共分散行列は、ケースの分類に使用します。
- グループ別: グループ別共分散行列は分類に使用します。分類は (元の変数ではなく) 判別関数に基づいて行うため、このオプションは必ずしも 2 次の判別と等価ではありません。

作図: 使用できる作図オプションは、「結合されたグループ」、「グループ別」、「領域マップ」です。

- 結合されたグループ: 最初の 2 つの判別関数の値を使用して全グループ散布図を作成します。関数が 1 つしかない場合は、代わりにヒストグラムが表示されます。
- グループ別: 最初の 2 つの判別関数の値のグループ別散布図を作成します。関数が 1 つしかない場合は、代わりにヒストグラムを表示します。
- 領域マップ: 関数の値に基づいてケースをグループに分類するために使用する境界のプロット。これらの数値は、ケースの分類先グループに対応します。各グループの平均値は、その境界の内側にアスタリスクで示します。判別関数が 1 つしかない場合は、このマップを表示しません。

判別分析: 保存

アクティブなデータ・ファイルに新しい変数を追加することができます。使用できるオプションには、「予測された所属グループ」(単一変数)、「判別得点」(解の判別関数ごとに 1 つの変数)、および判別得点に基づく「所属グループの事後確率」(グループごとに 1 つの変数)があります。

モデル情報を、指定のファイルに XML 形式でエクスポートすることもできます。このモデル・ファイルを使用して、スコアリングのために他のデータ・ファイルにモデル情報を適用できます。

DISCRIMINANT コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 複数の判別分析を実行し (1 つのコマンドを使用)、変数の入力順序を制御する (ANALYSIS サブコマンドを使用)。
- 分類用の事前確率を指定する (PRIORS サブコマンドを使用)。
- 回転後のパターン行列と構造行列を表示する (ROTATE サブコマンドを使用)。
- 抽出する判別関数の数を制限する (FUNCTIONS サブコマンドを使用)。
- 分析用に選択された (または選択されていない) ケースだけに分類を制限する (SELECT サブコマンドを使用)。
- 相関行列の読み取りと分析を行う (MATRIX サブコマンドを使用)。

- 今後の分析用に相関行列を書き込む (MATRIX サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 22 章 因子分析

因子分析は、基礎となる変数（つまり、観測対象の変数セット内における相関パターンを説明する因子）を特定しようとしています。因子分析は、通常、データの分解で使用されます。これにより、多数の顕在変数で観測された分散のほとんどを説明する少数の因子が特定されます。因子分析を使用して、原因のメカニズムに関する仮説を立てたり、以降の分析で使用する変数をスクリーニングしたりすることもできます（例えば、線型回帰分析を実行する前に共線性を特定する場合など）。

「因子分析」手続きには、以下のような高度な柔軟性があります。

- 7 種類の因子抽出方法があります。
- 直接オブリミンや非直交回転のためのプロマックスなど、5 種類の回転方法があります。
- 3 種類の因子得点の計算方法があります。得点を変数として保存して、今後の分析で使用することができます。

例: 政治に関する調査の質問に対して、回答者はどのような考え方で回答するのでしょうか。調査項目間の相関を調べると、税に関する質問や軍事問題に関する質問が互いに相関しているなど、項目のさまざまなサブグループ間に有意な重複があることが明らかになります。因子分析では、基底因子の数を調べることができます。また、多くの場合、因子が表す概念的な意味を特定することもできます。さらに、各回答者の因子得点を計算し、以降の分析で使用することもできます。例えば、因子得点を使用して、投票行動を予測するロジスティック回帰モデルを構築することができます。

統計: 各変数: 有効ケース数、平均値、標準偏差。各因子分析: 有意確率、行列式、逆行列を含む変数の相関行列、反イメージなどの再生相関行列、初期の解（共通性、固有値、および説明された分散のパーセント）、サンプル抽出の適正さの Kaiser-Meyer-Olkin 測定および Bartlett の球面性検定、因子負荷量、共通性、および固有値を含む回転前の解、回転後のパターン行列や変換行列を含む回転後の解。斜交回転: 回転後のパターンと構造行列、因子得点係数行列と因子共分散行列。作図: 固有値のスクリー・プロットと最初の 2 因子または 3 因子の因子負荷プロット。

因子分析データの考慮事項

データ: 変数は、区間 レベルまたは比率 レベルの量的変数でなければなりません。カテゴリー・データ（宗教や出生国など）は、因子分析には適していません。Pearson の相関係数を正確に算出できるデータは、因子分析に適しています。

仮定: データに含まれる変数の各ペアが 2 変量の正規分布で、観測値が独立している必要があります。因子分析モデルは、変数が共通因子（モデルによって推測される因子）と独自因子（観測変数間で重複しない因子）によって決定されることを指定します。つまり、計算された推測値は、すべての独自因子が互いに相関せず、共通因子とも相関しないという前提に基づいているということです。

因子分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「次元分解」 > 「因子分析...」

2. 因子分析のための変数を選択します。

因子分析のケースの選択

分析用にケースを選択するには、以下の手順を実行します。

1. ケース選択変数を選択します。
2. 「値」をクリックし、選択値として使用する整数を入力します。

ケース選択変数の値を持つケースだけが因子分析で使用されます。

因子分析の記述統計

統計: 「1 変量の記述統計」には、変数ごとの平均値、標準偏差、有効ケース数が含まれます。「初期の解」には、初期の共通性と固有値、および説明された分散のパーセントが表示されます。

相関行列: 選択可能なオプションは、係数、有意確率、行列式、KMO と Bartlett の球面性検定、逆行列、再生相関、反イメージです。

- *KMO* と *Bartlett* の球面性検定: Kaiser-Meyer-Olkin 標本妥当性の測度は、変数間の偏相関が小さいかどうかを検定します。*Bartlett* の球面性検定は、相関行列が単位行列であるかどうか (因子モデルが不適切かどうか) を検定します。
- 再生相関: 因子の解から推定した相関行列。残差 (推定した相関と観測した相関の差) も表示します。
- 反イメージ: 反イメージ相関行列は偏相関係数の負の値を含み、反イメージ共分散行列はその偏共分散の負の値を含みます。良好な因子モデルでは、対角要素以外のほとんどの要素が小さい値になります。変数に対する抽出の妥当性の指標は、反イメージ相関行列の対角に示されます。

因子分析の因子抽出

方法: 因子抽出の方法を指定することができます。選択可能な方法は、主成分分析、重み付けのない最小二乗法、一般化した最小二乗法、最尤法、主因子法、アルファ因子法、イメージ因子法です。

- *Principal Components Analysis* (主成分分析 (因子分析)). 観測変数の無相関線型結合を形成するために使用する因子抽出法。第 1 成分が最大の分散を持ちます。次の成分は分散のうちさらに小さな部分を説明し、それ以降の成分も順次さらに小さな部分を説明していきます。各成分はいずれも互いに相関しません。主成分分析は、初期因子解を得るために使用します。相関行列が特異である場合に使用することができます。
- *Unweighted Least-Squares Method* (重み付けのない最小 2 乗法 (因子分析)). 因子抽出法の 1 つ。(対角要素を無視して) 観測相関行列と再生相関行列の差の平方和を最小化します。
- *Generalized Least-Squares Method* (一般化最小 2 乗法 (因子分析)). 因子抽出法の 1 つ。観測相関行列と再生相関行列の差の平方和を最小化します。相関には一意性と逆の重み付けがされるため、一意性の高い変数には、一意性の低い変数より小さい重みが与えられます。
- *Maximum-Likelihood Method* (最尤法 (因子分析)). 因子抽出法の 1 つ。サンプルが多変量正規分布から抽出されている場合に、観測された相関行列を生成した可能性が最も高いパラメーター推定値を生成します。相関には変数の一意性の逆数で重みを付け、反復アルゴリズムを使用します。
- *Principal Axis Factoring* (主因子法 (因子分析)). 共通性の初期推定値として対角上に配置した重相関係数の 2 乗を使用して元の相関行列から因子を抽出する方法。これらの因子負荷量を使用して、対角上にある古い共通性の推定値に置き換わる新しい共通性を推定します。反復間での共通性の変化が抽出の収束基準を満たすまで、反復を続行します。
- *Alpha* (アルファ (因子分析)). 分析中の変数をポテンシャル変数の母集団のサンプルと見なす因子抽出法。この手法は、因子のアルファ信頼性を最大化します。

- *Image Factoring* (イメージ因子法 (因子分析)). Guttman によって開発された、イメージ理論に基づく因子抽出法。変数の共通部分 (偏イメージ) を、仮説的因子の関数としてではなく、残りの変数での線型回帰として定義します。

分析: 相関行列または共分散行列のいずれかを指定することができます。

- **相関行列:** 分析に含まれている変数を異なる尺度で測定する場合に便利です。
- **共分散行列:** 各変数が異なる分散を持つ複数のグループに対して因子分析を適用する場合に便利です。

抽出の基準: 指定された値を超える固有値を持つ因子をすべて保持することも、特定の数の因子だけを保持することもできます。

表示: 回転前の因子解と固有値のスクリー・プロットを要求することができます。

- *Unrotated Factor Solution* (回転のない因子解 (因子分析)). 回転を行わない因子解の因子負荷量 (因子パターン行列)、共通性、および固有値を表示します。
- **スクリー・プロット:** 各因子に関連する分散のプロット。このプロットは、保持する因子の数を決定するために使用します。一般にこのプロットでは、急勾配を示す大きい因子と、緩やかな裾 (スクリー) を示す残りの因子の間に明瞭な区切りが示されます。

収束のための最大反復回数: 解を推定するためにアルゴリズムに含めるステップの最大数を指定することができます。

因子分析の回転

方法: 因子の回転方法を選択することができます。選択可能な方法は、バリマックス、直接オブリミン、コーティマックス、エカマックス、プロマックスです。

- *Varimax Method* (バリマックス法 (因子分析)). 直交回転法の 1 つ。各因子の負荷量が高い変数の個数を最小化します。この方法では、因子の解釈が単純化されます。
- **直接オブリミン法:** 斜交 (非直交) 回転法の 1 つ。デルタが 0 (デフォルト) に等しいとき、解は最も斜交します。デルタが負で絶対値が大きくなるに従って、因子の斜交度は下がります。デフォルトのデルタ 0 をオーバーライドするには、0.8 以下の数値を入力してください。
- *Quartimax Method* (クォーティマックス法 (因子分析)). 回転法の 1 つ。各変数を説明するために必要な因子の数を最小化します。この方法では、観測した変数の解釈が単純化されます。
- *Equamax Method* (エカマックス法 (因子分析)). 回転法の 1 つ。因子を単純化するバリマックス法と、変数を単純化するコーティマックス法を組み合わせたもの。因子に高い負荷を加える変数の数と、変数を説明するために必要な因子の数を最小化します。
- *Promax Rotation* (プロマックス回転 (因子分析)). 斜交回転の 1 つ。因子を相関させることができます。この回転は直接オブリミン回転より高速に計算できるため、大規模なデータ・セットの場合に役立ちます。

表示: 最初の 2 因子または 3 因子の因子負荷プロットと、回転後の解の出力を表示することができます。

- *Rotated Solution* (回転後の解 (因子分析)). 回転後の解を得るためには、回転方法を選択する必要があります。直交回転の場合は、回転するパターン行列と因子変換行列が表示されます。斜交回転の場合は、パターン、構成、および因子の相関行列が表示されます。
- *Factor Loading Plot* (因子負荷プロット). 最初の 3 つの因子による 3 次元の因子負荷プロット。2 因子解の場合は 2 次元のプロットが表示されます。因子を 1 つしか抽出していない場合はプロットは表示されません。回転が要求された場合は、回転した解をプロットに表示します。

収束のための最大反復回数: 回転の実行のためにアルゴリズムに含めるステップの最大数を指定することができます。

因子分析の因子得点

変数として保存: 最終解の各因子に対して新しい変数を 1 つ作成します。

方法: 因子得点の計算方法の選択肢には、回帰、Bartlett 法、Anderson-Rubin 法があります。

- *Regression Method (回帰法 (因子分析))*. 因子得点係数の推定法の 1 つ。得られる得点は平均が 0 になり、分散が、推定因子得点と真の因子の値との重相関の 2 乗に等しくなります。因子が直交する場合であっても、得点が相関することがあります。
- *Bartlett 得点*: 因子得点係数の推定法の 1 つ。得られる得点の平均は 0 になります。変数の範囲内に一意の因子の平方和を最小化します。
- *Anderson-Rubin Method (Anderson-Rubin 法 (因子分析))*. 因子得点係数の推定法の 1 つ。推定する因子の直交性を確保するように Bartlett 法を修正したもの。得られる得点は平均が 0 に、標準偏差が 1 になり、相関しません。

因子得点係数行列を表示: 因子得点を算出するために変数に乗算する係数を表示します。因子得点間の相関も表示します。

因子分析オプション

欠損値: 欠損値の処理方法を指定することができます。使用できる選択項目は、「リストごとに除外」、「ペアごとに除外」、「平均値で置換」です。

係数の表示書式: 出力行列の縦横比を制御することができます。サイズを基準に係数をソートしたり、指定した値未満の絶対値を持つ係数を抑制したりすることができます。

FACTOR コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 抽出時と回転時に、反復の収束基準を指定する。
- 個別の回転後因子のプロットを指定する。
- 保存する因子スコアの数を指定する。
- 主因子法の対角値を指定する。
- 今後の分析用に、相関行列や因子負荷行列をディスクに書き込む。
- 相関行列や因子負荷行列の読み取りと分析を行う。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 23 章 クラスタリングの手続きの選択

クラスター分析は、TwoStep クラスター分析手続き、階層クラスター分析手続き、大規模ファイルのクラスター分析手続きを使用して実行することができます。各手続きは、クラスター作成用の異なるアルゴリズムを採用しており、他の手続きでは使用できないオプションを持っています。

TwoStep クラスター分析: TwoStep クラスター分析手続きは、多くのアプリケーションで選択される分析方法です。この手続きは、以下に示す独自機能を備えています。

- クラスター・モデルから選択するだけでなく、最適な数のクラスターを自動的に選択する。
- カテゴリ変数と連続型変数に基づいて、複数のクラスター・モデルを同時に作成する。
- クラスター・モデルを外部の XML ファイルに保存し、このファイルを読み込んで、新しいデータを使用してクラスター・モデルを更新する。

さらに、TwoStep クラスター分析手続きは、大規模なデータ・ファイルを分析することもできます。

階層クラスター分析: 階層クラスター分析手続きの対象は、小規模なデータ・ファイル (クラスター化されるオブジェクトが数百個のもの) だけに制限されますが、以下に示す独自機能を備えています。

- ケースまたは変数をクラスター化する。
- 考えられる解の範囲を計算し、それぞれの解の所属クラスターを保存する。
- クラスター形成、変数の変換、クラスター間の非類似度の測定を複数の方法で実行する。

すべての変数の型が同じであれば、階層クラスター分析手続きにより、区間 (連続)、度数、2 値の変数を分析することができます。

大規模ファイルのクラスター分析: 大規模ファイルのクラスター分析手続きの対象は連続データだけに制限され、クラスターの数をあらかじめ指定しておく必要がありますが、以下に示す独自機能を備えています。

- 各オブジェクトについて、クラスター中心からの距離を保存する。
- 外部の IBM SPSS Statistics ファイルから初期クラスター中心を読み込み、最後のクラスター中心を外部の IBM SPSS Statistics ファイルに保存する。

さらに、大規模ファイルのクラスター分析手続きは、大規模なデータ・ファイルを分析することもできます。

第 24 章 TwoStep クラスタ分析

「TwoStep クラスタ分析」手続きは、通常は明らかにされることがない、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。この手続きで使用されるアルゴリズムには、従来のクラスタ分析手法とは異なる以下の優れた特徴があります。

- **カテゴリー変数と連続型変数の処理:** 変数が独立していると仮定することにより、カテゴリー変数と連続型変数に対して、多項分布と正規分布を結合したものを配置することができます。
- **クラスタ数の自動選択:** 異なるクラスタ解の間でモデル選択基準の値を比較することにより、この手続きは、最適なクラスタ数を自動的に判定することができます。
- **スケーラビリティ:** レコードを要約するクラスタ機能 (CF) ツリーを作成することにより、TwoStep アルゴリズムで大規模なデータ・ファイルを分析することができます。

例: 小売企業と、消費者製品を扱っている企業は、顧客の購買習慣、性別、年齢、収入レベルなどを記述するデータに対して、定期的にクラスタリング手法を適用します。これらの企業は、販売を拡大し、ブランド・ロイヤリティを構築するために、マーケティング戦略と製品開発戦略を各消費者グループに合わせて調整します。

距離測定: このオプションにより、2 つのクラスタ間の類似度を計算する方法を指定します。

- **対数尤度:** この尤度測定により、変数の確率分布を求めます。連続型変数は正規分布しているものと仮定され、カテゴリー変数は多項分布しているものと仮定されます。すべての変数は独立しているものと仮定されます。
- **ユークリッド:** ユークリッド測定は、2 つのクラスタ間の「直線」距離です。このオプションは、すべての変数が連続型である場合のみ使用できます。

クラスタ数: このオプションにより、クラスタ数の判定方法を指定することができます。

- **自動的に判定:** この手続きは、「クラスタ化の基準」グループで指定された基準を使用して、「最適な」クラスタ数を自動的に判定します。オプションで、手続きで考慮する必要があるクラスタ数の最大値を正の整数で入力することができます。
- **固定値を指定:** 解に含まれるクラスタ数を固定値にすることができます。正整数を入力してください。

連続変数の数: このグループは、「オプション」ダイアログ・ボックスで指定された連続型変数の標準化仕様の集計を示します。詳しくは、108 ページの『TwoStep クラスタ分析のオプション』のトピックを参照してください。

クラスタ化の基準: この選択項目により、自動クラスタリング・アルゴリズムでクラスタ数を判定する方法を決定します。ベイズの情報量基準 (BIC) または赤池情報量基準 (AIC) のどちらかを指定することができます。

TwoStep クラスタ分析データの考慮事項

データ: この手続きは、連続型変数とカテゴリー変数の両方で有効です。ケースはクラスタ化されるオブジェクトを表し、変数はクラスタ化の基準となる属性を表します。

ケースの並び順: クラスタ機能ツリーと最終解は、ケースの並び順によって異なる可能性があることに注意してください。並び順の影響を最小限に抑えるには、ケースを無作為に並べます。得られた解の安定性を

確認するために、異なる無作為の順序でソートしたケースを使用して複数の異なる解を取得することをお勧めします。ファイル・サイズが非常に大きいためにこの方法を実行するのが難しい場合は、異なる無作為の順序でソートされたケースのサンプルを使用して、何回かに分けて実行してください。

仮定: 尤度距離測度は、クラスター・モデル内の変数が独立しているものと仮定します。さらに、各連続型変数には正規 (ガウス) 分布があると仮定し、各カテゴリ変数には多項分布があると仮定します。経験的内部検定は、この手続きが独立の仮定と分布の仮定の両方の違反に対して堅牢であることを示していますが、これらの仮定がどの程度満たされているかについて注意する必要があります。

2 つの連続型変数の独立性を検定するには、「2 変量の相関分析」手続きを使用します。2 つのカテゴリ変数の独立性を検定するには、「クロス集計表」手続きを使用します。連続型変数とカテゴリ変数との間の独立性を検定するには、「グループの平均」手続きを使用します。連続型変数の正規性を検定するには、「探索的分析」手続きを使用します。指定された多項分布がカテゴリ変数にあるかどうかを検定するには、「カイ 2 乗検定」手続きを使用します。

TwoStep クラスター分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「分類」 > 「TwoStep クラスター...」

2. 1 つ以上のカテゴリ変数または連続型変数を選択します。

必要に応じて、以下の操作を実行することができます。

- クラスターの作成基準を調整する。
- ノイズ処理、メモリー割り当て、変数の標準化、クラスター・モデル入力の設定を選択する。
- モデル・ビューアー出力を要求する。
- モデルの結果を作業ファイルまたは外部 XML ファイルに保存する。

TwoStep クラスター分析のオプション

外れ値の処置: クラスター機能 (CF) ツリーがいっぱいの場合、このグループを使用して、クラスター化の実行中に外れ値を特別に処理することができます。CF ツリーが葉ノードにこれ以上ケースを受け入れることができず、葉ノードを分割することができない場合、その CF ツリーはいっぱいになっています。

- ノイズ処理を選択した場合、CF ツリーがいっぱいになると、空きがある葉の中のケースを「ノイズ」葉に配置した後で、CF ツリーが再生されます。葉に含まれているケースの数が最大葉サイズのケースの指定パーセントよりも少ない場合、その葉には空きがあると見なされます。ツリーの再生後、可能であれば、外れ値が CF ツリー内に配置されます。配置できない場合、外れ値は廃棄されます。
- ノイズ処理を選択しなかった場合、CF ツリーがいっぱいになると、さらに大きな距離変更しきい値を使用して CF ツリーが再生されます。最終クラスター化が終了すると、クラスターに割り当てることができなかった値は、外れ値としてラベル付けされます。外れ値クラスターには「-1」という ID 番号が設定され、クラスターの数としてはカウントされません。

メモリー割り当て: このグループを使用して、クラスター・アルゴリズムが使用する最大メモリー量をメガバイト (MB) で指定することができます。この最大値を超えると、手続きはディスクを使用して、メモリー内に納まらない情報を保存します。4 以上の値を指定してください。

- システムで指定できる最大値については、システム管理者に問い合わせてください。
- この値が小さすぎると、アルゴリズムが正しいクラスターの数を検出できない場合や、指定されたクラスターの数を検出できない場合があります。

変数の標準化: クラスタリング・アルゴリズムは、標準化された連続型変数を処理します。標準化されていない連続型変数は、「標準化される変数」リストの変数として残ったままになります。既に標準化されている連続型変数を「標準化されない変数」リストの変数として選択すると、操作にかかる時間と手間を省くことができます。

高度なオプション

CF ツリーの調節基準: 以下のクラスタリング・アルゴリズム設定は、クラスター機能 (CF) ツリーに明示的に適用されるため、変更する場合は注意が必要です。

- **距離の変化の初期のしきい値:** これは、CF ツリーの生成に使用される初期しきい値です。特定のケースが CF ツリーの葉に挿入されたときに、ケース間の距離がこのしきい値よりも小さい場合、その葉は分割されません。ケース間の距離がこのしきい値よりも大きい場合は、葉が分割されます。
- **最大枝数 (葉ノードごと):** 1 つの葉ノードが持つことができる子ノードの最大数。
- **ツリーの最大の深さ:** 1 つの CF ツリーが持つことができるレベルの最大数。
- **最大ノード数:** 関数 $(b^{d+1} - 1) / (b - 1)$ に基づいて、手続きで生成できる CF ツリー・ノードの最大数を示します。 b は最大枝数、 d はツリーの最大の深さです。CF ツリーが大きすぎると、システム・リソースが浪費され、手続きのパフォーマンスに影響する可能性があることに注意してください。各ノードについて、最低 16 バイト必要です。

クラスター・モデルの更新: このグループを使用して、先に実行した分析で生成されたクラスター・モデルをインポートして更新することができます。入力ファイルには、CF ツリーが XML 形式で格納されています。このモデルは、アクティブ・ファイルのデータで更新されます。メイン・ダイアログ・ボックスで、先の分析で指定したときと同じ順序で変数名を選択する必要があります。新しいモデル情報を明示的に同じ名前のファイルに書き込まない限り、XMLファイルは変更されません。詳しくは、『TwoStep クラスタ分析の出力』のトピックを参照してください。

クラスター・モデルの更新を指定すると、元のモデルに対して指定された CF ツリーの生成に関するオプションが使用されます。具体的には、保存されたモデルの距離測度、ノイズ処理、メモリー割り当て、または CF ツリーの調節基準の設定が使用され、ダイアログ・ボックス内のこれらのオプション設定はすべて無視されます。

注: クラスタ・モデルを更新する際に、この手続きは、元のクラスター・モデルの作成で、アクティブなデータ・セット内で選択されているケースは 1 つも使用されなかったものと仮定します。また、この手続きは、モデルの更新で使用されるケースは、元のモデルの作成で使用されたケースと同じ母集団からのものであると仮定します。つまり、連続型変数の平均値と分散、カテゴリー変数のレベルは、両方のケースのセットで同じであると仮定されます。ケースの「新しい」セットと「古い」セットが不均質な母集団からのものからである場合、最良の結果を得るには、ケースを結合したセットで「TwoStep クラスタ分析」手続きを実行する必要があります。

TwoStep クラスタ分析の出力

出力: このグループにより、クラスター化の結果を表示するためのオプションを指定します。

- **ピボット・テーブル:** 結果がピボット・テーブルに表示されます。
- **モデル・ビューアーの図表:** 結果がモデル・ビューアーに表示されます。
- **評価フィールド:** クラスタの作成で使用されなかった変数について、クラスター・データが計算されます。「表示」サブダイアログで選択することにより、評価フィールドをモデル・ビューアーの入力機能とともに表示することができます。欠損値を持つフィールドは無視されます。

作業データ・ファイル: このグループを使用して、変数をアクティブ・データ・セットに保存することができます。

- **クラスターの所属変数を作成:** この変数には、各ケースのクラスター ID 番号が格納されます。この変数の名前は *tsc_n* です。*n* は、特定のセッションでこの手続きが実行したアクティブ・データ・セットの保存操作の順序を示す正の整数です。

XML ファイル: XML 形式でエクスポートできる出力ファイルのタイプには、最終クラスター・モデルと CF ツリーの 2 つがあります。

- **最終モデルをエクスポート:** 最終クラスター・モデルは、指定されたファイルに XML (PMML) 形式でエクスポートされます。このモデル・ファイルを使用して、スコアリングのために他のデータ・ファイルにモデル情報を適用できます。
- **CF ツリーをエクスポート:** このオプションにより、クラスター・ツリーの現在の状態を保存し、後で新しいデータを使用してそれを更新することができます。

クラスター・ビューアー

通常、クラスター・モデルは、調査された変数に基づく類似レコードのグループ (またはクラスター) の検索に使用されます。同じグループのメンバー間の類似度は高く、異なるグループのメンバー間の類似度は低くなります。この結果を使用して、明らかではなかった関連度を特定することができます。例えば、顧客の嗜好、収入レベル、購買習慣のクラスター分析を通して、特定のマーケティング・キャンペーンに回答する確率が高い顧客のタイプを特定できる場合があります。

クラスター表示での結果を解釈する方法は、次の 2 とおりあります。

- クラスターを調べて、そのクラスターに固有の特性を判別します。1 つのクラスターに、高収入な借り手がすべて含まれているかどうか。そのクラスターに、他のクラスターよりも多くのレコードが含まれているかどうか。
- クラスター間でフィールドを調べて、値がクラスター間でどのように分布しているかを判別します。クラスター内の所属は、学歴により決定されているか。信用度の高さにより、クラスターごとの所属が区別されているか。

クラスター・ビューアーのメイン・ビューおよび各種リンク・ビューを使用して、これらの質問に答える上で役立つ洞察力を得ることができます。

クラスター・モデルの情報を表示するには、ビューアーでモデル・ビューアー・オブジェクトをアクティブ化 (ダブルクリック) します。

クラスター・ビューアー

クラスター・ビューアーは 2 つのパネルで構成されており、左側にメイン・ビュー、右側にリンク・ビューまたは補助ビューがあります。メイン・ビューには、以下の 2 つがあります。

- モデルの要約 (デフォルト)。詳しくは、111 ページの『「モデルの要約」ビュー』のトピックを参照してください。
- クラスター。詳しくは、111 ページの『「クラスター」ビュー』のトピックを参照してください。

リンク/補助ビューには、以下の 4 つがあります。

- 予測値の重要度。詳しくは、113 ページの『「クラスター予測値の重要度」ビュー』のトピックを参照してください。

- クラスタ・サイズ (デフォルト)。詳しくは、113 ページの『「クラスタ・サイズ」ビュー』のトピックを参照してください。
- セルの分布。詳しくは、113 ページの『「セルの分布」ビュー』のトピックを参照してください。
- クラスタの比較。詳しくは、113 ページの『「クラスタの比較」ビュー』のトピックを参照してください。

「モデルの要約」ビュー

「モデルの要約」ビューには、陰影によって結果 (悪い、普通、良い) が示される、クラスタの結合および分離のシルエット測定など、クラスタ・モデルのスナップショット、つまり要約が表示されます。このスナップショットを使用すると、品質が悪いかどうかを素早く確認できます。品質が悪い場合は、より良い結果を生成するために、モデル作成ノードに戻ってクラスタ・モデルの設定を修正する決断を下すことができます。

結果 (悪い、普通、良い) は、クラスタ構造の解釈に関する Kaufman と Rousseeuw (1990) の研究に基づいています。「モデルの要約」ビューでは、良い結果は Kaufman と Rousseeuw の評価をクラスタ構造の合理的または強力な証拠として反映するデータに相当し、普通の結果は弱い証拠の評価を反映するデータに相当し、悪い結果は重要な証拠のない評価を反映するデータに相当します。

すべてのレコードに対するシルエット測定平均は $(B-A) / \max(A,B)$ となります。A はレコードのクラスタ中心へのレコードの距離、B はレコードが属さない最近傍のクラスタ中心へのレコードの距離です。シルエット係数 1 は、すべてのケースが自身のクラスタ中心の真上にあることを意味します。値 -1 は、すべてのケースが他のクラスタのクラスタ中心にあることを意味します。平均の 0 の値は、ケースが自身のクラスタ中心と、その他の最近傍のクラスタ中心から等距離にあることを意味します。

要約には、次の情報について示すテーブルも含まれます。

- **アルゴリズム:** 使用されたクラスタリング・アルゴリズム (例えば「TwoStep」)。
- **入力フィールド:** 入力または予測とも呼ばれる、フィールドの数。
- **クラスタ:** 解のクラスタ数。

「クラスタ」ビュー

「クラスタ」ビューには、各クラスタのクラスタ名、サイズ、プロファイルが含まれた、クラスタとフィールドのグリッドがあります。

グリッドの列には、次の情報が含まれています。

- **クラスタ:** アルゴリズムによって作成されたクラスタ番号。
- **ラベル:** 各クラスタに適用されるラベル (デフォルトでは空白です)。セル内をダブルクリックし、クラスタの内容を説明するラベル (例えば「高級車購入者」) を入力します。
- **説明:** クラスタの内容の説明 (デフォルトでは空白です)。セル内をダブルクリックし、クラスタの説明 (例えば「55 歳以上、専門職、収入 \$100,000 以上」) を入力します。
- **サイズ:** クラスタ・サンプル全体のパーセントとしての各クラスタのサイズ。グリッド内の各サイズのセルには、クラスタ内のサイズのパーセントを示す垂直バー、数値形式によるサイズのパーセント、クラスタ・ケース度数が表示されます。
- **フィールド:** 個別の入力または予測。デフォルトでは全体の重要度でソートされています。サイズが等しい列が複数ある場合、それらはクラスタ番号の昇順でソートされて表示されます。

フィールド全体の重要度は、セルの背景の陰影の色で示されます。最も重要なフィールドは濃く、最も重要でないフィールドは陰影なしとなります。テーブルの上のガイドは、各フィールドのセルの色に関連する重要度を示します。

セル上にマウスを移動すると、フィールドの完全名/ラベルとセルの重要度の値が表示されます。ビューおよびフィールドのタイプによっては、より詳細な情報が表示されます。「クラスター中心」ビューの場合、この情報には「平均: 4.32」など、セル統計量やセル値が含まれます。カテゴリ・フィールドの場合、セルは最も頻度の高い (最頻) カテゴリの名前とそのパーセントを示します。

「クラスター」ビュー内では、以下のさまざまな方法を選択して、クラスター情報を表示できます。

- クラスターとフィールドの入れ替え。詳しくは、『クラスターとフィールドの入れ替え』のトピックを参照してください。
- フィールドのソート。詳しくは、『フィールドのソート』のトピックを参照してください。
- クラスターのソート。詳しくは、『クラスターのソート』のトピックを参照してください。
- セル内容の選択。詳しくは、『セル内容』のトピックを参照してください。

クラスターとフィールドの入れ替え: デフォルトでは、クラスターは列として、フィールドは行として表示されます。この表示を逆にするには、「フィールドのソート基準」ボタンの左にある「クラスターとフィールドを入れ替え」ボタンをクリックします。例えば、表示されるクラスターが多い場合にこれを行うと、データを表示するために必要となる横方向のスクロール量を減らすことができます。

フィールドのソート: 「フィールドのソート基準」ボタンを使用すると、どのようにフィールド・セルが表示されるかを選択できます。

- **全体の重要度:** これは、デフォルトのソート順序です。フィールドは全体の重要度の降順にソートされ、ソート順序はクラスター間で同じです。同じ重要度の値を持つフィールドがある場合、それらのフィールドは、フィールド名の昇順でソートされてリストされます。
- **クラスター内重要度:** フィールドは、各クラスターのフィールドの重要度に応じてソートされます。同じ重要度の値を持つフィールドがある場合、それらのフィールドは、フィールド名の昇順でソートされてリストされます。このオプションを選択すると、通常、ソート順序はクラスターによって異なります。
- 「名前」。フィールドは、名前のアルファベット順にソートされます。
- **データ順:** フィールドは、データ・セット内のフィールドの順序でソートされます。

クラスターのソート: デフォルトでは、クラスターはサイズの降順でソートされています。「クラスターのソート基準」ボタンを使用すると、名前のアルファベット順にクラスターをソートできます。または固有のラベルを作成してある場合は、代わりにラベルの英数字順にソートできます。

同じラベルを持つフィールドは、クラスター名でソートされます。クラスターがラベルでソートされている場合にクラスターのラベルを編集すると、ソート順序は自動的に更新されます。

セル内容: 「セル」ボタンを使用すると、フィールドおよび評価フィールドのセル内容の表示を変更できます。

- **クラスター中心:** デフォルトでは、セルには、フィールド名/ラベルと各クラスター/フィールドの組み合わせの中心傾向が表示されます。連続型フィールドの場合は平均値が表示され、カテゴリ・フィールドの場合は最頻値 (最も頻繁に発生するカテゴリ) がカテゴリ・パーセントとともに表示されます。

- **絶対分布:** フィールド名/ラベルと各クラスター内のフィールドの絶対分布が表示されます。カテゴリ・フィールドの場合、データ値が昇順に並んでいるカテゴリを重ね合わせた棒グラフが表示されます。連続型フィールドの場合、各クラスターに対して同じエンドポイントと区間を使用する平滑密度プロットが表示されます。

濃い赤はクラスター分布を示し、薄い赤は全体のデータを表します。

- **相対分布:** フィールド名/ラベルと相対分布がセルに表示されます。一般的に、相対分布が代わりに表示されるという点を除いて、絶対分布の表示と類似しています。

濃い赤はクラスター分布を示し、薄い赤は全体のデータを表します。

- **基本ビュー:** クラスターが多いと、スクロールせずにすべての詳細を確認するのは難しい場合があります。スクロールの量を減らすために、このビューを選択して、よりコンパクトなバージョンのテーブルに表示を変更します。

「クラスター予測値の重要度」ビュー

「予測値の重要度」ビューには、モデルの推定における各フィールドの相対重要度が表示されます。

「クラスター・サイズ」ビュー

「クラスター・サイズ」ビューには、各クラスターが含まれた円グラフが表示されます。各クラスターのサイズのパーセントが各スライスに表示されます。各スライス上にマウスを移動すると、そのスライスに度数が表示されます。

グラフの下のテーブルには、以下のサイズ情報がリストされます。

- 最小クラスターのサイズ (度数と全体の割合)
- 最大クラスターのサイズ (度数と全体の割合)
- 最大クラスターの最小クラスターに対するサイズの比率

「セルの分布」ビュー

「セルの分布」ビューには、「クラスター」メイン・パネルのテーブルで選択したフィールド・セルのデータの分布に関する、拡張された、より詳細なプロットが示されます。

「クラスターの比較」ビュー

「クラスターの比較」ビューは、グリッド・スタイルのレイアウトで構成され、フィールドは行に、選択したクラスターは列に表示されます。このビューは、クラスターを構成する要素をより良く理解する上で役立ちます。また、全体のデータと比較するだけでなく、クラスター同士で比較して、クラスター間の差分を表示することもできます。

表示するクラスターを選択するには、「クラスター」メイン・パネルでクラスター列の一番上をクリックします。Ctrl キーまたは Shift キーを押しながらクリックすることで、比較対象として複数のクラスターを選択または選択解除できます。

注: 表示するクラスターは、5 個まで選択できます。

クラスターは選択した順に表示されます。一方でフィールドの順序は、「フィールドのソート基準」で決まります。「クラスター内重要度」を選択した場合、フィールドは常に全体の重要度でソートされます。

背景のプロットには、各フィールドの全体の分布が表示されます。

- カテゴリ・フィールドはドット・プロットとして表示されます。ドットのサイズは、フィールドごとの各クラスターの最も頻度の高い (最頻) カテゴリを示します。

- 連続型フィールドは箱ひげ図として表示され、全体の中央値と 4 分位範囲を示します。

これらの背景ビューに、選択したクラスターの箱ひげ図が重ね合わせて表示されます。

- 連続型フィールドの場合、四角形のポイント・マーカと水平線は、それぞれ各クラスターの中央値と 4 分位範囲を示します。
- 各クラスターは異なる色で表され、ビューの最上位に表示されます。

クラスター・ビューアーのナビゲート

クラスター・ビューアーはインタラクティブ表示です。以下を行うことができます。

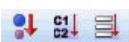
- フィールドまたはクラスターを選択して詳細を表示する。
- クラスターを比較して関心のある項目を選択する。
- 表示を変更する。
- 軸を入れ替える。

ツールバーの使用

ツールバー・オプションを使用して、左右のパネルに表示される情報を制御します。ツールバー・コントロールを使用して、表示方向 (上から下、左から右、右から左) を変更できます。また、ビューアーをデフォルト設定にリセットし、ダイアログ・ボックスを開いて、メイン・パネルの「クラスター」ビューの内容を指定することもできます。

「フィールドのソート基準」、「クラスターのソート基準」、「セル」、「表示」の各オプションは、メイン・パネルで「クラスター」ビューを選択した場合にのみ使用できます。詳しくは、111 ページの『「クラスター」ビュー』のトピックを参照してください。

表2. ツールバーのアイコン:

アイコン	トピック
	『クラスターとフィールドの入れ替え』を参照
	『フィールドのソート基準』を参照
	『クラスターのソート基準』を参照
	『セル』を参照

「クラスター」ビュー表示の制御

メイン・パネルの「クラスター」ビューの表示内容を制御するには、「表示」ボタンをクリックします。これにより、「表示」ダイアログが開きます。

フィールド: デフォルトで選択されています。すべての入力フィールドを非表示にするには、チェック・ボックスを選択解除します。

評価フィールド: 表示する評価フィールド (クラスター・モデルの作成には使用されないが、クラスターの評価のためにモデル・ビューアーに送信されるフィールド) を選択します。デフォルトでは表示される評価フィールドはありません。注: 評価フィールドは、複数の値が含まれた文字列でなければなりません。使用可能な評価フィールドがない場合、このチェック・ボックスは選択できません。

クラスターの説明: デフォルトで選択されています。すべてのクラスターの説明のセルを非表示にするには、チェック・ボックスを選択解除します。

クラスター・サイズ: デフォルトで選択されています。すべてのクラスター・サイズのセルを非表示にするには、チェック・ボックスを選択解除します。

カテゴリーの最大数: カテゴリー・フィールドのグラフに表示するカテゴリーの最大数を指定します。デフォルトは 20 です。

レコードのフィルタリング

特定のクラスターまたはクラスターのグループのケースについて詳細を確認したい場合は、選択したクラスターに基づいてレコードのサブセットを選択して、さらに分析を行うことができます。

1. クラスター・ビューアーのクラスター・ビューでクラスターを選択します。複数のクラスターを選択するには、Ctrl キーを押しながらクリックします。
2. 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...」

3. フィルター変数名を入力します。選択したクラスターのレコードには、このフィールドに対して値 1 が返されます。その他のレコードにはすべて値 0 が返され、フィルターの状態を変更するまで、以降の分析から除外されます。
4. 「OK」をクリックします。

第 25 章 階層クラスター分析

この手続きでは、選択された特性に基づいて、ケース（または変数）が相対的に等質なグループの識別が試みられます。このとき使用されるアルゴリズムは、個々のクラスターの各ケース（または変数）から始めて、クラスターが 1 つになるまで複数のクラスターを結合します。未調整の変数を分析するか、さまざまな標準化変換の中から選択することができます。距離測度または類似度は、「近接」手続きで生成されます。統計量が段階ごとに表示されるので、最良の解を選択するのに役立ちます。

例: 同じような視聴者層を持つテレビ番組のグループを識別することはできるでしょうか？ 階層クラスター分析では、視聴者の特性に基づいて、テレビ番組（ケース）を同種のグループにクラスター化できます。これを利用して、マーケティング用にセグメントを特定できます。また、比較可能な都市を選択して、さまざまなマーケティング戦略を検定できるように、都市（ケース）を等質グループにクラスター化することができます。

統計: クラスター凝集経過工程、距離（または類似度）行列、および単一の解または解の範囲の所属クラスター。**プロット:** デンドログラムおよびつららプロット。

階層クラスター分析データの考慮事項

データ: 変数には、量的データ、2 値データ、または度数データを使用できます。スケーリングの違いはクラスター解（複数可）に影響する場合がありますので、変数のスケーリングは重要な問題です。変数に大きなスケーリングの差がある場合（ある変数はドルで測定され、別の変数は年で測定されるなど）、変数の標準化を検討する必要があります（「階層クラスター分析」手続きでは、この標準化を自動的に実行できます）。

ケースの並び順: 同順位の距離または類似度が入力データ内に存在するか、結合中に更新されたクラスター内で発生する場合は、導き出されるクラスター解は、ファイル内のケースの順序に依存している可能性があります。得られた解の安定性を確認するために、異なる無作為の順序でソートしたケースを使用して複数の異なる解を取得することをお勧めします。

仮定: 使用する距離測度、または類似度は、分析するデータに適していなければなりません（距離測度と類似度の選択について詳しくは、「近接」手続きを参照してください）。また、関連する変数をすべて分析に含める必要があります。影響を与える変数を除外すると、誤った解が導かれる可能性があります。階層クラスター分析は予備的な方法なので、結果は独立サンプルで確認するまで仮として扱うべきです。

階層クラスター分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「分類」 > 「階層クラスター...」

2. ケースをクラスター化する場合は、1 つ以上の数値変数を選択します。変数をクラスター化する場合は、3 つ以上の数値変数を選択します。

オプションとして、ケースのラベル付けをする識別変数を選択できます。

階層クラスター分析の方法

クラスター化の方法: グループ間平均連結法、グループ内平均連結法、最近傍法、最遠隣法、重心法、メディアン法、および Ward 法が選択可能です。

測定: クラスタ化に使用する距離測度や類似度を指定できます。データの種類、および適切な距離測度または類似度を指定します。

- **区間:** ユークリッド距離、平方ユークリッド距離、コサイン、Pearson の相関、Chebychev、都市ブロック、Minkowski、およびカスタマイズが選択可能です。
- **度数:** カイ 2 乗測度およびファイ 2 乗測度が選択可能です。
- **2 値:** ユークリッド距離、平方ユークリッド距離、サイズの差、パターンの違い、分散、散らばり、形、単純一致、ファイ 4 点相関、ラムダ、Anderberg の *D*、Dice、Hamann、Jaccard、Kulczynski 1、Kulczynski 2、Lance-Williams、Ochiai、Rogers-Tanimoto、Russel-Rao、Sokal-Sneath 1、Sokal-Sneath 2、Sokal-Sneath 3、Sokal-Sneath 4、Sokal-Sneath 5、Yule の *Y*、および Yule の *Q* が選択可能です。

値の変換: 近接度を計算する前に、ケースまたは値のデータ値を標準化できます (2 値データには使用できません)。標準化の方法として、*z* スコア、-1 から 1 の範囲、0 から 1 の範囲、最大値を 1、平均値を 1、および標準偏差を 1 が選択可能です。

測定方法の変換: 距離測度によって生成された値を変換できます。変換された値は、距離測度の計算後に適用されます。絶対値、符号変換、および 0 から 1 の範囲で再スケールが選択可能です。

階層クラスタ分析の統計

クラスタ凝集経過工程: 各段階で結合されたケースまたはクラスタ、結合中のケース間またはクラスタ間の距離、およびケース (または変数) がクラスタに結合した最後のクラスタ・レベルを表示します。

距離行列: 項目間の距離または類似度が得られます。

所属クラスタ: クラスタの結合の 1 つ以上の段階で、各ケースが割り当てられるクラスタを表示します。単一の解および解の範囲が使用可能です。

階層クラスタ分析の作図

デンドログラム: デンドログラムが表示されます。デンドログラムは、形成されたクラスタの結合性の評価に使用でき、保持すべき適切なクラスタ数に関する情報を提供します。

つらら: すべてのクラスタまたは指定された範囲のクラスタを含むつららプロットが表示されます。つららプロットは、分析の反復ごとに、どのようにケースがクラスタへ結合されるかという情報を表示します。「方向」で、垂直プロットまたは水平プロットを選択できます。

階層クラスタ分析の新規変数の保存

所属クラスタ: 単一の解または解の範囲の所属クラスタを保存できます。保存された変数は、後続の分析でグループ間のその他の差の探索に使用できます。

CLUSTER コマンド・シンタックスの追加機能

階層クラスタ手続きは、CLUSTER コマンド・シンタックスを使用します。コマンド・シンタックス言語を使用して、以下の処理を実行することもできます。

- 1 つの分析で複数のクラスタリング方式を使用する。
- 近接行列の読み取りと分析を行う。

- 今後の分析用に、近接行列をディスクに書き込む。
- カスタマイズされた (べき乗の) 距離測度におけるべき乗と根の値を指定する。
- 保存された変数の名前を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 26 章 大規模ファイルのクラスター分析

この手続きは、大量のケースを処理できるアルゴリズムを使用して、選択された特性に基づくケース内で相対的に等質なグループを特定しようとしています。ただし、このアルゴリズムを使用するには、クラスターの数を指定する必要があります。あらかじめわかっている場合は、初期クラスター中心を指定することができます。ケースを分類する場合、クラスター中心を反復して更新することも、分類だけを行うこともできます。所属クラスター、距離情報、最終クラスター中心を保存することができます。オプションで、変数を指定して、その変数の値をケースごとの出力のラベル付けに使用することができます。また、分散分析の F 統計量を要求することもできます。この統計は便宜的なものですが（この手続きは異なるグループを形成しようとするため）、統計量の相対的なサイズにより、グループの分離に対する各変数の寄与率の情報が提供されません。

例: 類似した視聴者層の興味を引くテレビ番組を特定できるグループとは、どのようなものでしょうか。大規模ファイルのクラスター分析では、視聴者の特性を基に、テレビ番組（ケース）を k 個の等質グループにクラスター化することができます。この処理を使用して、マーケティングのセグメントを特定することができます。また、比較可能な都市を選択して、さまざまなマーケティング戦略を検定できるように、都市（ケース）を等質グループにクラスター化することができます。

統計: 完全な解: 初期クラスター中心、分散分析表。各ケース: クラスター情報、クラスター中心からの距離。

大規模ファイルのクラスター分析データの考慮事項

データ: 変数は、区間レベルまたは比率レベルの量的変数でなければなりません。変数が 2 値または度数の場合は、階層クラスター分析手続きを使用します。

ケースと初期クラスター中心の順序: 初期クラスター中心を選択するためのデフォルトのアルゴリズムは、ケースの順序によって異なります。「反復」ダイアログ・ボックスの「**移動平均を使用**」オプションを使用すると、初期クラスター中心の選択方法に関係なく、結果として得られた解が潜在的にケースの順序に依存するようになります。これらの方法のいずれかを使用すると、異なる無作為な順序でソートされたケースを使用していくつかの異なる解を得ることにより、特定の解の安定性を確認することができます。初期クラスター中心を指定し、「**移動平均を使用**」オプションを選択しなければ、ケースの順序に関連する問題を回避することができます。ただし、ケースからクラスター中心への固定された距離が存在する場合、初期クラスター中心の順序が解に影響することがあります。特定の解の安定性を評価するには、初期中心値の異なる転置の分析から得られた結果を比較します。

仮定: 距離は、単純なユークリッド距離を使用して計算されます。距離または類似度の別の測定方法を使用したい場合は、「階層クラスター分析」手続きを使用します。変数の尺度は、重要な考慮事項です。変数を異なる尺度で測定した場合（例えば、ある変数はドル単位、別の変数は年単位で表現されている場合）、誤解を招く結果になる可能性があります。こうした場合には、大規模ファイルのクラスター分析を実行する前に、変数を標準化する必要があります（これは、「記述統計」手続きで実行することができます）。この手続きは、適切な数のクラスターが選択されていて、関連する変数がすべて含まれているものと想定します。選択されているクラスターの数が適切でなかったり、重要な変数が省略されていたりすると、誤解を招く適切な結果になる可能性があります。

大規模ファイルのクラスター分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「分類」 > 「大規模ファイルのクラスター...」

2. クラスター分析で使用する変数を選択します。
3. クラスターの数を指定します (クラスターの最小数は 2 で、データ・ファイル内のケース数を超える数を指定することはできません)。
4. 「反復と分類」または「分類のみ」を選択します。
5. オプションで、ケースにラベルを付けるための識別変数を選択することができます。

大規模ファイルのクラスター分析の効率

大規模ファイルのクラスター分析コマンドは、基本的には効率的です。これは、階層クラスター分析コマンドで使用されるアルゴリズムなどの多くのクラスタリング・アルゴリズムとは異なり、大規模ファイルのクラスター分析コマンドの場合は、ケースのペア間における距離の計算で、一部のペアだけを計算対象とするためです。

効率を最大限に高めるには、サンプルのケースを使用し、「反復と分類」方法を選択してクラスター中心を決定します。次に、「最終値の書き込み」を選択します。次に、データ・ファイル全体を復元してから「分類のみ」を方法として選択し、「初期値の読み込み」を選択して、サンプルから推定された中心を使用してファイル全体を分類します。ファイルやデータ・セットに対して、データの読み取りと書き込みを行うことができます。データ・セットは、同じセッション内で引き続き使用できますが、セッションを終了する前に明示的に保存しない限り、ファイルとしては保存されません。データ・セット名は、変数の命名規則に従う必要があります。詳しくは、のトピックを参照してください。

大規模ファイルのクラスター分析の反復

注: 以下のオプションを使用できるのは、「大規模ファイルのクラスター分析」ダイアログ・ボックスで「反復と分類」方法を選択した場合だけです。

「最大反復回数」。大規模ファイルのクラスター分析のアルゴリズムでの反復回数を制限します。収束基準が満たされていない場合でも、この回数の反復が実行されると、反復が停止します。この数値は、1 から 999 の範囲内で指定する必要があります。

バージョン 5.0 以前の Quick Cluster コマンドで使用するアルゴリズムを再現するには、「最大反復回数」を 1 に設定します。

収束基準: 反復を停止するタイミングを決定します。この値は、初期クラスター中心間の最小距離の比率を表すため、0 より大きく 1 以下の値を指定する必要があります。例えば、基準が 0.02 の場合、完全な反復によって、いずれかの初期クラスター中心間の最小距離の 2% を超える距離だけクラスター中心が移動されなければ、反復が停止します。

移動平均を使用: 各ケースの割り当て後にクラスター中心を更新するように要求することができます。このオプションを選択しなかった場合、新しいクラスター中心の計算は、すべてのケースの割り当ての終了後に実行されます。

大規模ファイルのクラスター分析の保存

解に関する情報を新しい変数として保存して、その後の分析で使用することができます。

所属クラスター: ケースごとの最終所属クラスターを示す新しい変数が作成されます。新しい変数の値の範囲は、1 からクラスターの個数までです。

クラスター中心からの距離: 各ケースとその分類の中心との間のユークリッド距離を示す新しい変数が作成されます。

大規模ファイルのクラスター分析のオプション

統計: 初期クラスター中心、分散分析表、ケースに対するクラスター情報のいずれかを選択することができます。

- 初期クラスター中心: 各クラスターの変数の平均値の最初の推定値。デフォルトでは、クラスターと同数の、十分に間隔の空いたケースがデータから選択されます。初期クラスター中心は分類の 1 巡目を使用され、その後更新されます。
- ANOVA table (分散分析表 (カテゴリ/カテゴリの回帰)). クラスター変数ごとに、1 変量 F 検定を含む分散分析テーブルを表示します。F 検定は記述に過ぎず、結果として得られる確率を解釈すべきではありません。すべてのケースが 1 つのクラスターに割り当てられる場合、分散分析テーブルは表示されません。
- ケースに対するクラスター情報: それぞれのケースについて、最終クラスター割り当てと、ケースの分類に使用するクラスター中心とケースの間のユークリッド距離を表示します。また、最終クラスター中心間のユークリッド距離も表示します。

欠損値: 選択可能なオプションは、「リストごとに除外」または「ペアごとに除外」です。

- リストごとに除外: クラスター変数に対する欠損値を持つケースを分析から除外します。
- ペアごとに除外: 非欠損値を持つすべての変数から計算された距離に基づいて、ケースがクラスターに割り当てられます。

QUICK CLUSTER コマンドの追加機能

大規模ファイルのクラスター手続きは、QUICK CLUSTER コマンド・シンタックスを使用します。コマンド・シンタックス言語を使用して、以下の処理を実行することもできます。

- 最初の k ケースを初期クラスター中心として受け入れることにより、通常はこれらのケースの推定に使用されるデータ・パスを禁止する。
- コマンド・シンタックスの一部として初期クラスター中心を指定する。
- 保存された変数の名前を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 27 章 ノンパラメトリック検定

ノンパラメトリック検定は、データの基本的な分布について最小限の仮定を行います。これらのダイアログで選択可能な検定は、データがどのように編成されるかによって以下の 3 つの大きなカテゴリーにグループ化されます。

- 1 サンプルの検定で、1 つのフィールドを分析する。
- 対応サンプルの検定で、ケースの同じセットの 2 つ以上のフィールドを比較する。
- 独立サンプル検定で、別のフィールドのカテゴリーでグループ化された 1 つのフィールドを分析する。

1 サンプルのノンパラメトリック検定

1 サンプルのノンパラメトリック検定は、1 つ以上のノンパラメトリック検定を使用して、単一フィールドの差分を識別します。ノンパラメトリック検定では、データが正規分布になると仮定しません。

目的は?: 目的により、一般的に使用される互いに異なる検定の設定を迅速に指定できます。

- **観測データを仮説と自動的に比較する:** カテゴリー数が 2 つだけのカテゴリー・フィールドには 2 項検定が、その他すべてのカテゴリー・フィールドにはカイ 2 乗検定が、連続型フィールドには Kolmogorov-Smirnov 検定が適用されます。
- **ランダム性の順序を検定する:** ラン検定を使用して、ランダム性についてデータ値の観測された順序を検定します。
- **カスタム分析:** 「設定」タブで検定の設定を手動で修正する場合、このオプションを選択します。後で「設定」タブのオプションに変更を行い、その変更が現在選択されている目的と互換性がない場合、この設定が自動的に選択されることに注意してください。

1 サンプルのノンパラメトリック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。

「分析」 > 「ノンパラメトリック検定」 > 「1 サンプル...」

1. 「実行」をクリックします。

必要に応じて、以下の操作を実行することができます。

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

「フィールド」タブ

「フィールド」タブでは、検定するフィールドを指定します。

事前定義された役割を使用: このオプションでは、既存のフィールド情報が使用されます。事前定義された役割が「入力」、「対象」または「両方」に指定されているフィールドはすべて、検定フィールドとして使用されます。1 つ以上の検定フィールドが必要です。

カスタム・フィールド割り当ての使用: このオプションを使用すると、フィールドの役割をオーバーライドできます。このオプションを選択した後、以下のフィールドを指定してください。

- **検定フィールド:** 1 つ以上のフィールドを選択してください。

「設定」タブ

「設定」タブは、アルゴリズムがデータをどのように処理するかを調整するために変更できる、複数グループの設定で構成されています。現在選択されている目的と互換性のない変更をデフォルト設定に対して行うと、「目的」タブが自動的に更新され、「分析のカスタマイズ」オプションを選択します。

検定の選択

これらの設定は、「フィールド」タブで指定したフィールドに実行する検定を指定します。

データに基づいて検定を自動的に選択します: 有効な (非欠損) カテゴリが 2 つしかないカテゴリ・フィールドには 2 項検定が、その他すべてのカテゴリ・フィールドにはカイ 2 乗検定が、連続型フィールドには Kolmogorov-Smirnov 検定が適用されます。

検定のカスタマイズ: この設定では、実行する特定の検定を選択できます。

- **観測された 2 値の確率を仮説と比較する (2 項検定):** 2 項検定をすべてのフィールドに適用できます。これにより、フラグ型フィールド (カテゴリ数が 2 つだけのカテゴリ・フィールド) の観測された分布が、指定した 2 項分布から予想されるものと同じかどうかを検定する 1 サンプル検定が作成されます。さらに、信頼区間を要求することができます。検定の設定について詳しくは、『2 項検定のオプション』を参照してください。
- **観測された確率を仮説と比較する (カイ 2 乗検定):** カイ 2 乗検定が、名義型フィールドおよび順序型フィールドに適用されます。これにより、フィールドのカテゴリの観測度数と期待度数の差に基づいてカイ 2 乗統計量を計算する、1 サンプル検定が作成されます。検定の設定について詳しくは、127 ページの『カイ 2 乗検定のオプション』を参照してください。
- **観測された分布を仮説と比較する (Kolmogorov-Smirnov 検定):** コルモゴロフ・スミルノフ検定が、連続型フィールドと順序型フィールドに適用されます。これにより、1 つのフィールドのサンプル累積分布関数を、一様分布、正規分布、ポワソン分布、または指数分布と等質であるかどうかを検定する 1 サンプル検定が作成されます。検定の設定について詳しくは、127 ページの『Kolmogorov-Smirnov 検定のオプション』を参照してください。
- **中央値を仮説と比較する (Wilcoxon の符号付き順位検定):** Wilcoxon の符号付き順位検定が、連続型フィールドと順序型フィールドに適用されます。これにより、フィールドの中央値に対する 1 サンプル検定が作成されます。仮説の中央値として数値を指定してください。
- **ランダム性の順序をテストする (ラン検定):** ラン検定が、すべてのフィールドに適用されます。これにより、2 分されたフィールドの値の順序が無作為かどうかについての 1 サンプル検定が作成されます。検定の設定について詳しくは、127 ページの『ラン検定のオプション』を参照してください。

2 項検定のオプション: 2 項検定はフラグ型フィールド (カテゴリが 2 つだけのカテゴリ・フィールド) を対象としていますが、「成功」を定義するルールを使用することにより、すべてのフィールドに適用されます。

仮説の比率: 「成功」として定義されるレコードの期待比率 (p) を指定します。0 より大きく、1 より小さい値を指定してください。デフォルトは 0.5 です。

信頼区間: 2 項データの信頼区間を計算するには、以下の方法が使用できます。

- **Clopper-Pearson (正確):** 累積 2 項分布に基づいた正確な区間。
- **Jeffreys:** Jeffreys を事前に使用し、 p の事後分布に基づくベイズ区間。
- **尤度比:** p の尤度関数に基づいた区間。

カテゴリ・フィールドの成功を定義する:「成功」(仮説の比率に対して検定されたデータ値 (複数可)) を、カテゴリ・フィールドに対してどのように定義するかを指定します。

- 「**データ内で最初に検出されたカテゴリを使用する**」を指定すると、サンプル内で最初に検出された値を使用して 2 項検定を実行し、「成功」を定義します。このオプションは、値が 2 つだけの名義型フィールドまたは順序型フィールドにのみ適用できます。このオプションが使用されている、「フィールド」タブで指定したその他すべてのカテゴリ・フィールドは検定されません。これはデフォルトです。
- 「**成功の値を指定する**」を指定すると、指定した値のリストを使用して 2 項検定を実行し、「成功」を定義します。文字列値または数値のリストを指定してください。リストの値は、サンプルに存在していなくても構いません。

連続型フィールドの成功を定義する:「成功」(検定値に対して検定されたデータ値 (複数可)) を、連続型フィールドに対してどのように定義するかを指定します。成功は、分割点以下の値として定義されます。

- 「**サンプルの中点**」を指定すると、最小値と最大値の平均に分割点が設定されます。
- 「**カスタム分割点**」を指定すると、分割点の値を指定できます。

カイ 2 乗検定のオプション: すべてのカテゴリの確率が等しい: サンプルのすべてのカテゴリで同等の度数を作成します。これはデフォルトです。

期待確率をカスタマイズする: カテゴリの指定したリストに不等な度数を指定できます。文字列値または数値のリストを指定してください。リストの値は、サンプルに存在していなくても構いません。「**カテゴリ**」列でカテゴリ値を指定します。「**相対度数**」列で、各カテゴリに 0 より大きい値を指定します。カスタム度数は比率として処理されるので、例えば、1、2、および 3 の度数を指定することは 10、20、および 30 の度数を指定することと同じであり、両者ともレコードの 1/6 が最初のカテゴリに、1/3 が 2 番目のカテゴリに、1/2 が 3 番目のカテゴリに当てはまるよう指定しています。カスタムの期待確率を指定する場合、カスタムのカテゴリ値には、データ内のすべてのフィールド値が含まれる必要があります。そうでない場合は、そのフィールドの検定は実行されません。

Kolmogorov-Smirnov 検定のオプション: このダイアログは、検定する分布、および仮説の分布のパラメータを指定します。

正規: 「**サンプル データを使用**」では、観測された平均値および標準偏差が使用されます。「**カスタム**」では、値を指定することができます。

一様: 「**サンプル データを使用**」では、観測された最小値および最大値が使用されます。「**カスタム**」では、値を指定できます。

指数: 「**サンプル平均**」では、観測平均値が使用されます。「**カスタム**」では、値を指定することができます。

ポワソン: 「**サンプル平均**」では、観測平均値が使用されます。「**カスタム**」では、値を指定することができます。

ラン検定のオプション: ラン検定はフラグ型フィールド (カテゴリが 2 つだけのカテゴリ・フィールド) を対象としていますが、グループを定義するルールを使用することにより、すべてのフィールドに適用できます。

カテゴリ・フィールドのグループを定義する: 使用可能なオプションは次のとおりです。

- 「**サンプル内のカテゴリは 2 つだけです**」を指定すると、サンプルで検出された値を使用してラン検定を実行し、グループを定義します。このオプションは、値が 2 つだけの名義型フィールドまたは順序

型フィールドにのみ適用できます。このオプションが使用されている、「フィールド」タブで指定したその他すべてのカテゴリー・フィールドは検定されません。

- 「データを 2 つのカテゴリに再コード化する」を指定すると、指定した値のリストを使用してラン検定を実行し、グループの 1 つを定義します。サンプルのその他の値はすべて、もう 1 つのグループを定義します。リストの値がすべてサンプルに存在する必要はありませんが、少なくとも 1 つのレコードが各グループに必要です。

連続型フィールドの分割点を定義する: 連続型フィールドのグループのを定義方法を指定します。最初のグループは分割点以下の値として定義されます。

- 「サンプル中央値」は、分割点をサンプル中央値に設定します。
- 「サンプル平均値」は、分割点をサンプル平均値に設定します。
- 「カスタム」を指定すると、分割点の値を指定できます。

検定オプション

有意水準: すべての検定の有意水準 (アルファ) を指定します。0 から 1 までの数値を指定してください。デフォルトは 0.05 です。

信頼区間 (%): 作成されたすべての信頼区間の信頼度レベルを指定します。0 から 100 までの数値を指定してください。デフォルトは 95 です。

除外されたケース: 検定のケース基準の決定方法を指定します。

- 「リストごとに除外」を指定すると、「フィールド」タブで指定したフィールドに対して欠損値を持つレコードは、すべての分析から除外されます。
- 「検定ごとに除外」を指定すると、特定の検定に使用されるフィールドに対して欠損値を持つレコードは、その検定から除外されます。分析に複数の検定を指定すると、各検定は個別に評価されます。

ユーザー欠損値

カテゴリー・フィールドのユーザー欠損値: レコードを分析に含めるには、カテゴリー・フィールドに有効値が必要です。これらのコントロールを使用すると、カテゴリー・フィールドでユーザー欠損値を有効値として扱うかどうかを決定できます。システム欠損値および連続型フィールドの欠損値は常に無効として処理されます。

NPTESTS コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 手続きを 1 回実行して、1 サンプル検定、独立サンプル検定、対応サンプル検定を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

独立サンプルのノンパラメトリック検定

独立サンプルのノンパラメトリック検定は、1 つ以上のノンパラメトリック検定を使用して 2 つ以上のグループ間の差分を特定します。ノンパラメトリック検定では、データが正規分布になると仮定しません。

目的は?: 目的により、一般的に使用される互いに異なる検定の設定を迅速に指定できます。

- **自動的にグループ間の分布を比較する:** この目的では、2 つのグループのデータには Mann-Whitney の U 検定が、または k 個のグループのデータには Kruskal-Wallis の一元配置分散分析が適用されます。

- **グループ間の中央値を比較する:** メディアン検定を使用して、グループ間で観測された中央値が比較されます。
- **カスタム分析:** 「設定」タブで検定の設定を手動で修正する場合、このオプションを選択します。後で「設定」タブのオプションに変更を行い、その変更が現在選択されている目的と互換性がない場合、この設定が自動的に選択されることに注意してください。

独立サンプルのノンパラメトリック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。

「分析」 > 「ノンパラメトリック検定」 > 「独立サンプル...」

1. 「実行」をクリックします。

必要に応じて、以下の操作を実行することができます。

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

「フィールド」タブ

「フィールド」タブでは、検定するフィールドおよびグループを定義するために使用するフィールドを指定します。

事前定義された役割を使用: このオプションでは、既存のフィールド情報が使用されます。事前定義された役割が「目標」または「両方」に指定されているすべての連続型フィールドと順序型フィールドは、検定フィールドとして使用されます。事前定義された役割が「入力」である単一のカテゴリー・フィールドは、グループ化フィールドとして使用されます。そうでない場合、デフォルトではグループ化フィールドは使用されないため、カスタム・フィールドの割り当てを使用する必要があります。1 つ以上の検定フィールドおよびグループ化フィールドが必要です。

カスタム・フィールド割り当ての使用: このオプションを使用すると、フィールドの役割をオーバーライドできます。このオプションを選択した後、以下のフィールドを指定します。

- **検定フィールド:** 1 つ以上の連続型フィールドまたは順序型フィールドを選択します。
- **グループ:** カテゴリー・フィールドを選択します。

「設定」タブ

「設定」タブは、アルゴリズムがデータをどのように処理するかを調整するために変更できる、複数グループの設定で構成されています。現在選択されている目的と互換性のない変更をデフォルト設定に対して行うと、「目的」タブが自動的に更新され、「分析のカスタマイズ」オプションを選択します。

検定の選択

これらの設定は、「フィールド」タブで指定したフィールドに実行する検定を指定します。

データに基づいて検定を自動的に選択します: この設定では、2 グループのデータには Mann-Whitney の U 検定が、 k 個のグループのデータには Kruskal-Wallis 一元配置分散分析が適用されます。

検定のカスタマイズ: この設定では、実行する特定の検定を選択できます。

- **グループ間の分布を比較する:** サンプルが同じ母集団から抽出されているかどうかに関する独立サンプル検定を作成します。

「**Mann-Whitney の U (2 サンプル)**」では、各ケースの順位を使用して、グループが同じ母集団から抽出されているかどうかを検定します。昇順に並んだグループ化フィールドの最初の値が最初のグループを定義し、2 番目の値が 2 番目のグループを定義します。グループ化フィールドに 2 つを超える値がある場合、この検定は作成されません。

「**Kolmogorov-Smirnov (2 サンプル)**」は、2 つの分布間の中央値、散らばり、歪度などの差分に対して敏感です。グループ化フィールドに 2 つを超える値がある場合、この検定は作成されません。

「**ランダム性の順序をテストする (2 サンプルの Wald-Wolfowitz)**」は、基準として所属グループを指定してラン検定を作成します。グループ化フィールドに 2 つを超える値がある場合、この検定は作成されません。

「**Kruskal-Wallis (k サンプル)**」は、Mann-Whitney の U 検定の拡張であり、一元分散分析のノンパラメトリック版です。オプションで、「すべてのペアごと」の多重比較、または「ステップワイズのステップダウン」の比較のいずれかで、 k 個のサンプルの多重比較を要求できます。

「**順序付けのサンプルをテストする (k サンプルの Jonckheere-Terpstra)**」は、 k 個のサンプルが自然な順序の場合、Kruskal-Wallis に比べてより強力な方法です。例えば、 k 個の母集団が k 段階の温度上昇を表すとします。異なる温度でも同じ応答分布になるという仮説が、温度が上昇するにつれて応答の大きさが増えていくという対立仮説に対して検定されます。この場合、対立仮説が順序付けされるため、使用する検定としては Jonckheere-Terpstra が最も適しています。「**最小から最大**」では、1 番目のグループの位置パラメータは 2 番目のグループのもの以下であり、その 2 番目のグループのものは 3 番目のグループのもの以下であり、以降についても同様となる、という対立仮説が指定されます。「**最大から最小**」では、1 番目のグループの位置パラメータは 2 番目のグループのもの以上であり、その 2 番目のグループのものは 3 番目のグループのもの以上であり、以降についても同様となる、という対立仮説が指定されます。また、どちらのオプションの対立仮説でも、各位置がすべて等しくなることはないものと想定されます。オプションで、「すべてのペアごと」の多重比較、または「ステップワイズのステップダウン」の比較のいずれかで、 k 個のサンプルの多重比較を要求できます。

- **グループ間の範囲を比較する:** サンプルの範囲が同じかどうかに関する独立サンプル検定を作成します。「**Moses の外れ値反応検定 (2 サンプル)**」は、比較グループに対して制御グループを検定します。昇順に並んだグループ化フィールドの最初の値が制御グループを定義し、2 番目の値が比較グループを定義します。グループ化フィールドに 2 つを超える値がある場合、この検定は作成されません。
- **グループ間の中央値を比較する:** サンプルの中央値が同じかどうかに関する独立サンプル検定を作成します。「**メディアン検定 (k サンプル)**」では、プールされたサンプル中央値 (データ・セットのすべてのレコードで計算) またはカスタム値を仮説の中央値として使用できます。オプションで、「すべてのペアごと」の多重比較、または「ステップワイズのステップダウン」の比較のいずれかで、 k 個のサンプルの多重比較を要求できます。
- **グループ間の信頼区間を推定する:** 「**Hodges-Lehman 推定 (2 サンプル)**」を指定すると、2 つのグループの中央値の差分に対する独立サンプル推定および信頼区間が作成されます。グループ化フィールドに 2 つを超える値がある場合、この検定は作成されません。

検定オプション

有意水準: すべての検定の有意水準 (アルファ) を指定します。0 から 1 までの数値を指定してください。デフォルトは 0.05 です。

信頼区間 (%): 作成されたすべての信頼区間の信頼度レベルを指定します。0 から 100 までの数値を指定してください。デフォルトは 95 です。

除外されたケース: 検定のケース基準の決定方法を指定します。「リストごとに除外」を指定すると、サブコマンドで指定したフィールドに対して欠損値を持つレコードは、すべての分析から除外されます。「**検定ごとに除外**」を指定すると、特定の検定に使用されるフィールドに対して欠損値を持つレコードは、その検定から除外されます。分析に複数の検定を指定すると、各検定は個別に評価されます。

ユーザー欠損値

カテゴリー・フィールドのユーザー欠損値: レコードを分析に含めるには、カテゴリー・フィールドに有効値が必要です。これらのコントロールを使用すると、カテゴリー・フィールドでユーザー欠損値を有効値として扱うかどうかを決定できます。システム欠損値および連続型フィールドの欠損値は常に無効として処理されます。

NPTESTS コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 手続きを 1 回実行して、1 サンプル検定、独立サンプル検定、対応サンプル検定を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

対応サンプルのノンパラメトリック検定

1 つ以上のノンパラメトリック検定を使用して 2 つ以上の関連したフィールド間の差分を識別します。ノンパラメトリック検定では、データが正規分布になると仮定しません。

データの考慮事項: 各レコードは、データ・セットの各フィールドに 2 つ以上の関連する測定が格納されている特定の被験者に対応しています。例えば、ダイエット計画の効果に関する研究で、定期的に測定された各被験者の体重が「**ダイエット前の体重**」、「**中間の体重**」、および「**ダイエット後の体重**」などのフィールドに格納される場合は、対応サンプルのノンパラメトリック検定を使用して分析できます。これらの値は「**関連**」しています。

目的は?: 目的により、一般的に使用される互いに異なる検定の設定を迅速に指定できます。

- **観測データを自動的に仮説データと比較する:** この目的では、2 つのフィールドが指定された場合のカテゴリー・データに McNemar の検定が、3 つ以上のフィールドが指定された場合のカテゴリー・データに Cochran の Q が、2 つのフィールドが指定された場合の連続型データに Wilcoxon の一致するペアの符号付き順位検定が、そして 3 つ以上のフィールドが指定された場合の連続型データに順位付けによる Friedman の二次元配置分散分析が適用されます。
- **カスタム分析:** 「設定」タブで検定の設定を手動で修正する場合、このオプションを選択します。後で「設定」タブのオプションに変更を行い、その変更が現在選択されている目的と互換性がない場合、この設定が自動的に選択されることに注意してください。

異なる測定レベルのフィールドが指定されている場合、まず測定レベルごとに区分され、その後、適切な検定が各グループに適用されます。例えば、「**観測データを仮説データと自動的に比較する**」を目的に選択し、3 つの連続型フィールドおよび 2 つの名義型フィールドを指定した場合、Friedman の検定が連続型フィールドに適用され、McNemar の検定が名義型フィールドに適用されます。

対応サンプルのノンパラメトリック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。

「分析」 > 「ノンパラメトリック検定」 > 「対応サンプル...」

1. 「実行」をクリックします。

必要に応じて、以下の操作を実行することができます。

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

「フィールド」タブ

「フィールド」タブでは、検定するフィールドを指定します。

事前定義された役割を使用: このオプションでは、既存のフィールド情報が使用されます。事前定義された役割が「対象」または「両方」に指定されているフィールドはすべて、検定フィールドとして使用されません。2 つ以上の検定フィールドが必要です。

カスタム・フィールド割り当ての使用: このオプションを使用すると、フィールドの役割をオーバーライドできます。このオプションを選択した後、以下のフィールドを指定してください。

- **検定フィールド:** 2 つ以上のフィールドを選択します。各フィールドは個別の対応サンプルに対応しています。

「設定」タブ

「設定」タブは、手続きがデータをどのように処理するかを調整するために変更できる、複数グループの設定で構成されています。他の目的と互換性のない変更をデフォルト設定に対して行くと、「目的」タブが自動的に更新され、「分析のカスタマイズ」オプションを選択します。

検定の選択

これらの設定は、「フィールド」タブで指定したフィールドに実行する検定を指定します。

データに基づいて検定を自動的に選択します: この設定では、2 つのフィールドが指定された場合のカテゴリー・データに McNemar の検定を、3 つ以上のフィールドが指定された場合のカテゴリー・データに Cochran の Q が、2 つのフィールドが指定された場合の連続型データに Wilcoxon の一致するペアの符号付き順位検定が、そして 3 つ以上のフィールドが指定された場合の連続型データに順位付けによる Friedman の二次元配置分散分析が適用されます。

検定のカスタマイズ: この設定では、実行する特定の検定を選択できます。

- **2 値のデータの変更の検定:** 「McNemar の検定 (2 サンプル)」をカテゴリー・フィールドに適用できます。2 つのフラグ型フィールド (値が 2 つのみのカテゴリー・フィールド) 間の値の組み合わせの確率が等しいかどうかに関する対応サンプルの検定を作成します。「フィールド」タブで 3 つ以上のフィールドが指定されている場合、この検定は実行されません。検定の設定について詳しくは、133 ページの『McNemar 検定: 成功の定義』を参照してください。「Cochran の Q (k サンプル)」はカテゴリー・フィールドに適用できます。これは、 k 個のフラグ型フィールド (値が 2 つのみのカテゴリー・フィールド) 間の値の組み合わせの確率が等しいかどうかに関する対応サンプルの検定を作成します。オプションで、「すべてのペアごと」の多重比較、または「ステップワイズのステップダウン」の比較のいずれかで、 k 個のサンプルの多重比較を要求できます。検定の設定について詳しくは、133 ページの『Cochran の Q: 成功の定義』を参照してください。
- **多項データの変更の検定:** 「周辺等質性検定 (2 サンプル)」は、ペアになっている 2 つの順序型フィールド間の値の組み合わせの確率が等しいかどうかに関する対応サンプル検定を作成します。周辺等質性検定は通常、反復測定で使用されます。この検定は、McNemar の検定を 2 値回答から多重回答に拡張したものです。「フィールド」タブで 3 つ以上のフィールドが指定されている場合、この検定は実行されません。

- **中央値の差分を仮説と比較:** これらの検定は、2 つのフィールド間の中央値の差分が 0 とは異なるかどうかに関する対応サンプルの検定を作成します。検定は連続型フィールドと順序型フィールドに適用されます。「フィールド」タブで 3 つ以上のフィールドが指定されている場合、これらの検定は実行されません。
- **信頼区間を推定:** 2 つのペアになっているフィールド間の中央値差分の対応サンプル推定および信頼区間を作成します。検定は連続型フィールドと順序型フィールドに適用されます。「フィールド」タブで 3 つ以上のフィールドが指定されている場合、この検定は実行されません。
- **関連性を数量化:** 「Kendall の一致係数 (k サンプル)」は、審判または評価者間の一致の測定方法を作成します。ここでの各レコードは、複数の項目 (フィールド) に対する 1 人の審判の評価です。オプションで、「すべてのペアごと」の多重比較、または「ステップワイズのステップダウン」の比較のいずれかで、 k 個のサンプルの多重比較を要求できます。
- **分布の比較:** 「Friedman (k サンプル)」は、同じ母集団から k 個の対応サンプルが抽出されたかどうかについての対応サンプルの検定を作成します。オプションで、「すべてのペアごと」の多重比較、または「ステップワイズのステップダウン」の比較のいずれかで、 k 個のサンプルの多重比較を要求できます。

McNemar 検定: 成功の定義: McNemar 検定はフラグ型フィールド (カテゴリーが 2 つだけのカテゴリー・フィールド) を対象としていますが、「成功」を定義するルールを使用することにより、すべてのカテゴリー・フィールドに適用されます。

カテゴリー・フィールドの成功を定義する: カテゴリー・フィールドの「成功」をどのように定義するのかを指定します。

- 「データ内で最初に検出されたカテゴリを使用する」を指定すると、サンプル内で最初に検出された値を使用して検定を実行し、「成功」を定義します。このオプションは、値が 2 つだけの名義型フィールドまたは順序型フィールドにのみ適用できます。このオプションが使用されている、「フィールド」タブで指定したその他すべてのカテゴリー・フィールドは検定されません。これはデフォルトです。
- 「成功の値を指定する」を指定すると、指定された値のリストを使用して検定を実行し、「成功」を定義します。文字列値または数値のリストを指定してください。リストの値は、サンプルに存在していなくても構いません。

Cochran の Q: 成功の定義: Cochran の Q 検定はフラグ型フィールド (カテゴリーが 2 つだけのカテゴリー・フィールド) を対象としていますが、「成功」を定義するルールを使用することにより、すべてのカテゴリー・フィールドに適用されます。

カテゴリー・フィールドの成功を定義する: カテゴリー・フィールドの「成功」をどのように定義するのかを指定します。

- 「データ内で最初に検出されたカテゴリを使用する」を指定すると、サンプル内で最初に検出された値を使用して検定を実行し、「成功」を定義します。このオプションは、値が 2 つだけの名義型フィールドまたは順序型フィールドにのみ適用できます。このオプションが使用されている、「フィールド」タブで指定したその他すべてのカテゴリー・フィールドは検定されません。これはデフォルトです。
- 「成功の値を指定する」を指定すると、指定された値のリストを使用して検定を実行し、「成功」を定義します。文字列値または数値のリストを指定してください。リストの値は、サンプルに存在していなくても構いません。

検定のオプション

有意水準: すべての検定の有意水準 (アルファ) を指定します。0 から 1 までの数値を指定してください。デフォルトは 0.05 です。

信頼区間 (%): 作成されたすべての信頼区間の信頼度レベルを指定します。0 から 100 までの数値を指定してください。デフォルトは 95 です。

除外されたケース: 検定のケース基準の決定方法を指定します。

- 「**リストごとに除外**」を指定すると、サブコマンドで指定したフィールドに対して欠損値を持つレコードは、すべての分析から除外されます。
- 「**検定ごとに除外**」を指定すると、特定の検定に使用されるフィールドに対して欠損値を持つレコードは、その検定から除外されます。分析に複数の検定を指定すると、各検定は個別に評価されます。

ユーザー欠損値

カテゴリー・フィールドのユーザー欠損値: レコードを分析に含めるには、カテゴリー・フィールドに有効値が必要です。これらのコントロールを使用すると、カテゴリー・フィールドでユーザー欠損値を有効値として扱うかどうかを決定できます。システム欠損値および連続型フィールドの欠損値は常に無効として処理されます。

NPTESTS コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 手続きを 1 回実行して、1 サンプル検定、独立サンプル検定、対応サンプル検定を指定する。

シンタックスについて詳しくは、「**コマンド・シンタックス・リファレンス**」を参照してください。

モデル・ビュー

モデル・ビュー

手続きによって、ビューアーに「モデル・ビューアー」オブジェクトが作成されます。このオブジェクトをダブルクリックしてアクティブにすると、モデルの双方向ビューが表示されます。モデル・ビューのウィンドウは 2 つのパネルで構成されています。左側のパネルはメイン・ビューで、右側のパネルはリンク・ビュー (補助ビュー) です。

メイン・ビューには、以下の 2 つがあります。

- 仮説の要約。これがデフォルト・ビューです。詳しくは、135 ページの『仮説の要約』のトピックを参照してください。
- 信頼区間の要約。詳しくは、135 ページの『信頼区間の要約』のトピックを参照してください。

リンク/補助ビューには 7 種類あります。

- 1 サンプルの検定。1 サンプルの検定が要求された場合は、これがデフォルト・ビューになります。詳しくは、135 ページの『1 サンプルの検定』のトピックを参照してください。
- 対応サンプル検定。対応サンプル検定が要求され、1 サンプルの検定が要求されなかった場合は、これがデフォルト・ビューになります。詳しくは、136 ページの『対応サンプル検定』のトピックを参照してください。
- 独立サンプルの検定。対応サンプル検定も 1 サンプルの検定も要求されていない場合は、これがデフォルト・ビューになります。詳しくは、137 ページの『独立サンプルの検定』のトピックを参照してください。
- カテゴリー・フィールド情報。詳しくは、138 ページの『カテゴリー・フィールド情報』のトピックを参照してください。

- 連続型フィールド情報。詳しくは、138 ページの『連続型フィールド情報』のトピックを参照してください。
- ペアごとの比較。詳しくは、138 ページの『ペアごとの比較』のトピックを参照してください。
- 等質サブセット。詳しくは、139 ページの『等分散性サブセット』のトピックを参照してください。

仮説の要約

「モデルの要約」ビューでは、ノンパラメトリック検定についてスナップショット的に一目でわかる要約が表示されます。帰無仮説と決定を強調し、有意な p 値に注意を引き付けます。

- 各行は、それぞれ別の検定に対応します。行をクリックすると、検定に関する追加情報がリンク・ビューに表示されます。
- 列の見出しをクリックすると、その列の値によって行がソートされます。
- 「戻す」ボタンを使用すると、「モデル・ビューアー」を元の状態に戻すことができます。
- 「フィールド フィルター」ドロップダウン・リストを使用すると、選択したフィールドを使用する検定のみを表示できます。

信頼区間の要約

「信頼区間の要約」には、ノンパラメトリック検定で生成された信頼区間が表示されます。

- 各行は、各信頼区間に対応します。
- 列の見出しをクリックすると、その列の値によって行がソートされます。

1 サンプルの検定

「1 サンプルの検定」ビューには、要求された 1 サンプルのノンパラメトリック検定に関連する詳細が表示されます。表示される情報は、選択された検定によって異なります。

- 「検定」ドロップダウン・リストを使用すると、任意の種類の 1 サンプルの検定を選択できます。
- 「フィールド」ドロップダウン・リストを使用すると、「検定」ドロップダウン・リストで選択した検定を使用して検定されたフィールドを選択できます。

2 項検定

「2 項検定」には、積み上げ棒グラフと検定テーブルが表示されます。

- 積み上げ棒グラフには、検定フィールドの「成功」カテゴリおよび「失敗」カテゴリの観測度数および仮説度数が表示されます。「失敗」カテゴリは「成功」カテゴリの上に積み上げられます。棒グラフの棒の上にマウス・ポインターを置くと、ツールチップにカテゴリの割合が表示されます。棒に目に見える差がある場合、検定フィールドに仮説の 2 項分布がない可能性を示します。
- テーブルには検定の詳細が表示されます。

カイ 2 乗検定

「カイ 2 乗検定」ビューには、クラスター棒グラフと検定テーブルが表示されます。

- クラスター棒グラフには、検定フィールドの各カテゴリの観測度数および仮説度数が表示されます。クラスター棒グラフの棒の上にマウス・ポインターを置くと、ツールチップに観測度数と仮説度数およびそれらの差 (残差) が表示されます。観測と仮説の棒に目に見える差がある場合は、検定フィールドに仮説の分布がない可能性を示します。
- テーブルには検定の詳細が表示されます。

Wilcoxon の符号付き順位

「Wilcoxon の符号付き順位検定」ビューには、ヒストグラムと検定テーブルが表示されます。

- ヒストグラムには、観測された中央値と仮説の中央値を示す縦線が表示されます。
- テーブルには検定の詳細が表示されます。

ラン検定

「ラン検定」ビューには、グラフと検定テーブルが表示されます。

- グラフには正規分布が示され、ランの観測数が縦線でマークされます。ただし、正確確率検定が実行された場合、検定は正規分布に基づきません。
- テーブルには検定の詳細が表示されます。

Kolmogorov-Smirnov 検定

「Kolmogorov-Smirnov 検定」ビューには、ヒストグラムと検定テーブルが表示されます。

- ヒストグラムには、仮説の一樣分布、正規分布、ポアソン分布、または指数分布の確率密度関数のオーバーレイが含まれます。この検定は累積分布に基づくことに注意してください。テーブルで報告される「最遠距離差」は、累積分布に関して解釈する必要があります。
- テーブルには検定の詳細が表示されます。

対応サンプル検定

「1 サンプルの検定」ビューには、要求された 1 サンプルのノンパラメトリック検定に関連する詳細が表示されます。表示される情報は、選択された検定によって異なります。

- 「**検定**」ドロップダウン・リストを使用すると、任意の種類の 1 サンプルの検定を選択できます。
- 「**フィールド**」ドロップダウン・リストを使用すると、「**検定**」ドロップダウン・リストで選択した検定を使用して検定されたフィールドを選択できます。

McNemar 検定

「McNemar 検定」ビューには、クラスター棒グラフと検定テーブルが表示されます。

- クラスター棒グラフには、検定フィールドで定義された、2x2 のテーブルの非対角のセルの観測度数および仮説度数が表示されます。
- テーブルには検定の詳細が表示されます。

符号検定

「符号検定」ビューには、積み上げヒストグラムと検定テーブルが表示されます。

- 積み上げヒストグラムには、差分の符号を積み上げフィールドとして使用して、フィールド間の差分が表示されます。
- テーブルには検定の詳細が表示されます。

Wilcoxon の符号付き順位検定

「Wilcoxon の符号付き順位検定」ビューには、積み上げヒストグラムと検定テーブルが表示されます。

- 積み上げヒストグラムには、差分の符号を積み上げフィールドとして使用して、フィールド間の差分が表示されます。
- テーブルには検定の詳細が表示されます。

周辺等質性検定

「周辺等質性検定」ビューには、クラスター棒グラフと検定テーブルが表示されます。

- クラスター棒グラフには、検定フィールドで定義されたテーブルの非対角のセルの観測度数が表示されます。
- テーブルには検定の詳細が表示されます。

Cochran の Q 検定

「Cochran の Q 検定」ビューには、積み上げ棒グラフと検定テーブルが表示されます。

- 積み上げ棒グラフには、検定フィールドの「成功」カテゴリおよび「失敗」カテゴリの観測度数が表示されます。「失敗」カテゴリは「成功」カテゴリの上に積み上げられます。棒グラフの棒の上にマウス・ポインターを置くと、ツールチップにカテゴリの割合が表示されます。
- テーブルには検定の詳細が表示されます。

Friedman の順位付けによる変数の双方向分析

「Friedman の順位付けによる変数の双方向分析」ビューには、パネル化されたヒストグラムと検定テーブルが表示されます。

- ヒストグラムには、順位の観測分布が検定フィールドごとにパネル化されて表示されます。
- テーブルには検定の詳細が表示されます。

Kendall の一致係数

「Kendall の一致係数」ビューには、パネル化されたヒストグラムと検定テーブルが表示されます。

- ヒストグラムには、順位の観測分布が検定フィールドごとにパネル化されて表示されます。
- テーブルには検定の詳細が表示されます。

独立サンプルの検定

「独立サンプルの検定」ビューには、要求された独立サンプルのノンパラメトリック検定に関する詳細が表示されます。表示される情報は、選択された検定によって異なります。

- 「**検定**」ドロップダウン・リストを使用すると、任意の種類の独立サンプルの検定を選択できます。
- 「**フィールド**」ドロップダウン・リストを使用すると、「**検定**」ドロップダウン・リストで選択した検定を使用して検定された検定とグループ化フィールドの組み合わせを選択できます。

Mann-Whitney 検定

「Mann-Whitney 検定」ビューには、人口ピラミッド・グラフと検定テーブルが表示されます。

- 人口ピラミッド・グラフには、グループ化フィールドのカテゴリごとの背中合わせのヒストグラムが表示され、各グループのレコード数とグループの平均順位が注記されます。
- テーブルには検定の詳細が表示されます。

Kolmogorov-Smirnov 検定

「Kolmogorov-Smirnov 検定」ビューには、人口ピラミッド・グラフと検定テーブルが表示されます。

- 人口ピラミッド・グラフには、グループ化フィールドのカテゴリごとの背中合わせのヒストグラムが表示され、各グループのレコード数が示されます。観測された累積分布の線は、「**累積**」ボタンをクリックすることによって、表示または非表示にすることができます。
- テーブルには検定の詳細が表示されます。

Wald-Wolfowitz のラン検定

「Wald-Wolfowitz のラン検定」ビューには、積み上げ棒グラフと検定テーブルが表示されます。

- 人口ピラミッド・グラフには、グループ化フィールドのカテゴリごとの背中合わせのヒストグラムが表示され、各グループのレコード数が示されます。
- テーブルには検定の詳細が表示されます。

Kruskal-Wallis 検定

「Kruskal-Wallis の検定」ビューには、箱ひげ図と検定テーブルが表示されます。

- グループ化フィールドのカテゴリごとに、別個の箱ひげ図が表示されます。ボックスの上にマウス・ポインターを置くと、ツールチップに平均順位が表示されます。
- テーブルには検定の詳細が表示されます。

Jonckheere-Terpstra 検定

「Jonckheere-Terpstra 検定」ビューには、箱ひげ図と検定テーブルが表示されます。

- グループ化フィールドのカテゴリごとに、別個の箱ひげ図が表示されます。
- テーブルには検定の詳細が表示されます。

外れ値対応の Moses の検定

「外れ値対応の Moses の検定」ビューには、箱ひげ図と検定テーブルが表示されます。

- グループ化フィールドのカテゴリごとに、別個の箱ひげ図が表示されます。ポイント・ラベルは、「レコード ID」ボタンをクリックすることによって、表示または非表示にすることができます。
- テーブルには検定の詳細が表示されます。

メディアン検定

「メディアン検定」ビューには、箱ひげ図と検定テーブルが表示されます。

- グループ化フィールドのカテゴリごとに、別個の箱ひげ図が表示されます。
- テーブルには検定の詳細が表示されます。

カテゴリー・フィールド情報

「カテゴリー・フィールド情報」ビューには、「フィールド」ドロップダウン・リストで選択したカテゴリー・フィールドの棒グラフが表示されます。使用できるフィールドのリストは、「仮説の要約」ビューで現在選択されている検定で使用されているカテゴリー・フィールドに制限されます。

- グラフの棒の上にマウス・ポインターを移動すると、カテゴリーの割合がツールチップに表示されます。

連続型フィールド情報

「連続型フィールド情報」ビューには、「フィールド」ドロップダウン・リストで選択した連続型フィールドのヒストグラムが表示されます。使用できるフィールドのリストは、「仮説の要約」ビューで現在選択されている検定で使用されている連続型フィールドに制限されます。

ペアごとの比較

「ペアごとの比較」ビューでは、ペアワイズの多重比較が要求された場合に、 k 個のサンプルのノンパラメトリック検定で作成された距離のネットワーク図と比較表が表示されます。

- 距離のネットワーク図は、ネットワークのノード間の距離がサンプル間の差分に対応する比較表のグラフィカル表現です。黄色い線は統計的に有意な差分に対応し、黒い線は統計的に有意でない差分に対応します。ネットワークの線の上にマウス・ポインターを置くと、線で接続されたノード間の差分の調整済み有意確率を示したツールチップが表示されます。
- 比較表には、すべてのペアごとの比較の数値型結果が表示されます。各行は個別のペアごとの比較に対応しています。列の見出しをクリックすると、その列の値によって行がソートされます。

等分散性サブセット

「等分散性サブセット」ビューでは、ステップワイズ法のステップダウン法による多重比較が要求された場合の、 k 個のサンプルのノンパラメトリック検定で作成された比較表が表示されます。

- サンプル・グループの各行は、各関連サンプルに対応します (データでは別個のフィールドとして表現)。統計的に有意差がないサンプルは同じ色のサブセットにグループ化されます。識別されたサブセットごとに個別の列があります。すべてのサンプルが統計的に有意差がある場合は、サンプルごとに別個のサブセットが存在します。統計的に有意差があるサンプルがない場合は、サブセットは 1 つになります。
- 検定統計量、有意確率値、および調整済みの有意確率値は、複数のサンプルを含むサブセットごとに計算されます。

NPTESTS コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 手続きを 1 回実行して、1 サンプル検定、独立サンプル検定、対応サンプル検定を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

レガシー ダイアログ

「レガシー」ダイアログの多くは、ノンパラメトリック検定も実行します。これらのダイアログは、「正確確率検定」オプションで提供される機能をサポートします。

カイ 2 乗検定: 変数をカテゴリーに集計し、観測度数と期待度数の差に基づいてカイ 2 乗統計量を計算します。

2 項検定: 2 分変数の各カテゴリーの観測度数を、2 項分布の期待度数と比較します。

ラン検定: 1 つの変数の 2 つの値の発生順序がランダムかどうかを検定します。

1 サンプル Kolmogorov-Smirnov 検定: 変数の観測累積分布関数を、指定した理論分布 (正規分布、一様分布、指数分布、またはポワソン分布) と比較します。

2 個の独立サンプルの検定: 1 つの変数に関して 2 つのケース・グループを比較します。Mann-Whitney の U 検定、2 サンプル Kolmogorov-Smirnov 検定、Moses の極値反応検定、および Wald-Wolfowitz のラン検定が使用可能です。

2 個の対応サンプルの検定: 2 つの変数の分布を比較します。Wilcoxon の符号付き順位検定、符号検定、および McNemar 検定が使用可能です。

複数の独立サンプルの検定: 1 つの変数に関して 2 つ以上のケース・グループを比較します。Kruskal-Wallis 検定、メディアン検定、および Jonckheere-Terpstra 検定が使用可能です。

複数の対応サンプルの検定: 2 つ以上の変数の分布を比較します。Friedman の検定、Kendall の W 、および Cochran の Q が使用可能です。

上記すべての検定に、4 分位ならびに平均値、標準偏差、最小値、最大値、および非欠損ケース数が使用可能です。

カイ 2 乗検定

「カイ 2 乗検定」手続きは、変数をカテゴリ別に集計してカイ 2 乗統計量を計算します。この適合度検定は、各カテゴリ内の観測度数と期待度数を比較し、すべてのカテゴリに同じ比率で値が含まれているか、または各カテゴリにユーザー指定の比率で値が含まれているかを検定します。

例: カイ 2 乗検定を使用して、飴玉を入れた袋の中に、青、茶、緑、オレンジ、赤、黄色の飴玉がそれぞれ同じ比率で入っているかどうかを判断することができます。また、青 5%、茶 30%、緑 10%、オレンジ 20%、赤 15%、黄色 15% の比率で飴玉が入っているかどうかを検定して調べることもできます。

統計: 平均値、標準偏差、最小値、最大値、4 分位。非欠損ケースと欠損ケースの数とパーセント、各カテゴリの観測ケース数と期待ケース数、残差、カイ 2 乗統計量。

カイ 2 乗検定データの考慮事項

データ: 順序付けられている数値型カテゴリ変数、または順序付けられていない数値型カテゴリ変数を使用します (順序または名義の尺度レベル)。文字型変数を数値型変数に変換するには、「変換」メニューの「連続数への再割り当て」手続きを使用します。

仮定: ノンパラメトリック検定では、基本的な分布形状についての仮定は必要ありません。データは、無作為のサンプルとして仮定されます。各カテゴリの期待度数は 1 以上です。期待度数が 5 未満となるカテゴリが 20% を超えてはいけません。

カイ 2 乗検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「カイ 2 乗...」

2. 1 つ以上の検定変数を選択します。各変数が、個別の検定を作成します。
3. 必要に応じて「オプション」をクリックし、記述統計量、4 分位、欠損データの処理を指定します。

カイ 2 乗検定の期待範囲と期待値

期待範囲: デフォルトでは、変数の各値がカテゴリとして定義されます。特定の範囲内でカテゴリを定義するには、「指定された範囲を使用」を選択して、整数値の下限と上限を入力します。カテゴリは、範囲内の各整数値に対して定義され、範囲外の値を持つケースは除外されます。例えば、「下限」に 1 を指定し、「上限」に 4 を指定した場合、1 から 4 までの整数値だけがカイ 2 乗検定で使用されます。

期待値: デフォルトでは、すべてのカテゴリが同じ期待値を持ちます。カテゴリは、ユーザーが指定した期待比率を取ることができます。「値」を選択し、検定変数の各カテゴリに 0 より大きな値を入力して「追加」をクリックします。追加した値は、値リストの一番下に表示されます。この順序は、検定変数のカテゴリ値の昇順に対応するため、重要です。リストの最初の値が検定変数のグループ値の最低値に対応し、最後の値が最高値に対応します。値リストの各要素が合計され、それぞれの値がこの合計値で除算されて、対応するカテゴリの期待ケースの比率が計算されます。例えば、3、4、5、4 という値リストの場合、期待比率は、3/16、4/16、5/16、4/16 となります。

カイ 2 乗検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- 記述統計: 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- 4 分位: 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- 検定ごとに除外: 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- リストごとに除外: いずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (カイ 2 乗検定)

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変数ごとに異なる最小値と最大値、または期待度数を指定する (CHISQUARE サブコマンドを使用)。
- 1 つの変数を、異なる期待度数や異なる範囲で検定する (EXPECTED サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

2 項検定

「2 項検定」手続きは、2 分変数の 2 つのカテゴリの観測度数を、指定された確率パラメーターの 2 項分布での期待度数と比較します。デフォルトでは、両方のグループの確率パラメーターは 0.5 です。この確率を変更するには、最初のグループの検定比率を入力します。2 番目のグループの確率は、最初のグループで指定した比率を 1 から減算した値になります。

例: コインを投げた場合、表が出る確率は 2 分の 1 です。この仮説に基づいて、コインを 40 回投げ、結果 (表または裏) を記録します。2 項検定により、投げた回数の 3/4 が表であり、観測された有意水準が小さい (0.0027) ことがわかります。この結果は、表の出る確率が 1/2 ではないことを示しています。この場合、コインが歪んでいる可能性があります。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。

2 項検定データの考慮事項

データ: 検定対象の変数は、数値型の 2 分変数でなければなりません。文字型変数を数値型変数に変換するには、「変換」メニューの「連続数への再割り当て」手続きを使用します。2 分変数とは、「はい」または「いいえ」、「true」または「false」、0 または 1 など、有効な値が 2 つしかない変数のことです。データ・セット内で検出された最初の値が最初のグループを定義し、もう一方の値が 2 番目のグループを定義します。変数が 2 分変数でない場合は、分割点を指定する必要があります。分割点により、分割点以下の値を持つケースが最初のグループに割り当てられ、残りのケースが 2 番目のグループに割り当てられません。

仮定: ノンパラメトリック検定では、基本的な分布形状についての仮定は必要ありません。データは、無作為のサンプルとして仮定されます。

2 項検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「2 項...」

2. 1 つ以上の数値型検定変数を選択します。

3. 必要に応じて「オプション」をクリックし、記述統計量、4 分位、欠損データの処理を指定します。

2 項検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- 記述統計: 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- 4 分位: 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- 検定ごとに除外: 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- リストごとに除外: 検定対象のいずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPART TESTS コマンドの追加機能 (2 項検定)

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変数が 3 つ以上のカテゴリーを持っている場合に、特定のグループを選択して他のグループを除外する (BINOMIAL サブコマンドを使用)。
- 変数ごとに分割点または確率を指定する (BINOMIAL サブコマンドを使用)。
- 同じ変数を異なる分割点または確率に対して検定する (EXPECTED サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

ラン検定

「ラン検定」手続きは、変数の 2 つの値が出現する順序がランダムであるかどうかを検定します。ランとは、類似した一連の観測値のことです。ランが非常に多いサンプルや非常に少ないサンプルは、そのサンプルがランダムではないことを示しています。

例: 20 人を対象に調査を行い、ある製品を購入するかどうかを調べるとします。20 人の対象者のすべてが同じ性別である場合は、前提となるサンプルのランダム性に大きな問題があることとなります。こうした場合にラン検定を使用すると、サンプルが無作為に抽出されたかどうかを判断することができます。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。

ラン検定データの考慮事項

データ: 変数は数値型でなければなりません。文字型変数を数値型変数に変換するには、「変換」メニューの「連続数への再割り当て」手続きを使用します。

仮定: ノンパラメトリック検定では、基本的な分布形状についての仮定は必要ありません。連続確率分布のサンプルを使用します。

ラン検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「ラン...」

2. 1 つ以上の数値型検定変数を選択します。
3. 必要に応じて「オプション」をクリックし、記述統計量、4 分位、欠損データの処理を指定します。

ラン検定の分割点

分割点: 選択した変数を 2 分するための分割点を指定します。観測された平均値、中央値、最頻値を使用することも、分割点として指定された値を使用することもできます。分割点未満の値を持つケースが一方のグループに割り当てられ、分割点以上の値を持つケースがもう一方のグループに割り当てられます。選択された分割点ごとに検定が 1 回実行されます。

ラン検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- **記述統計:** 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- **4 分位:** 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- **検定ごとに除外:** 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- **リストごとに除外:** いずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (ラン検定)

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変数ごとに異なる分割点を指定する (RUNS サブコマンドを使用)。
- 1 つの変数を複数の異なるカスタム分割点に対して検定する (RUNS サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

1 サンプル Kolmogorov-Smirnov 検定

「1 サンプルによる Kolmogorov-Smirnov 検定」手続きは、1 つの変数に対して観測された累積分布関数を、指定された理論分布 (正規分布、一様分布、ポアソン分布、または指数分布) と比較します。

Kolmogorov-Smirnov の Z は、観測された累積分布関数と理論的な累積分布関数との間の最大差 (絶対値) から計算されます。この適合度検定は、観測値が指定した分布から適切に取得されているかどうかを検定します。

例: パラメトリック検定では、通常、正規分布の変数が必要です。1 サンプルによる Kolmogorov-Smirnov 検定を使用して、収入 などの変数が正規分布しているかどうかを検定することができます。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。

1 サンプルによる Kolmogorov-Smirnov 検定データの考慮事項

データ: 量的変数を使用します (測度の区間または比率レベル)。

仮定: Kolmogorov-Smirnov 検定は、検定分布のパラメーターが事前に指定されていると想定します。この手続きは、サンプルからパラメーターを推定します。サンプルの平均値と標準偏差は正規分布のパラメーターで、サンプルの最小値と最大値は一様分布の範囲を定義し、サンプルの平均値はポアソン分布と指数分布のパラメーターです。この検定では、仮説分布からの逸脱に関する検出性が大きく低下する可能性があります。推定パラメーターを持つ正規分布を検定する場合は、調整済みの K-S Lilliefors の検定 (「探索的分析」手続きで使用可能) を使用することをお勧めします。

1 サンプルによる Kolmogorov-Smirnov 検定を行うには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「1 サンプルによる K-S 検定...」

2. 1 つ以上の数値型検定変数を選択します。各変数が、個別の検定を作成します。
3. 必要に応じて「オプション」をクリックし、記述統計量、4 分位、欠損データの処理を指定します。

1 サンプルによる Kolmogorov-Smirnov 検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- 記述統計: 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- 4 分位: 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- 検定ごとに除外: 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- リストごとに除外: いずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPART TESTS コマンド追加機能 (1 サンプルによる Kolmogorov-Smirnov 検定)

コマンド・シンタックス言語では、検定分布のパラメーターを指定することもできます (K-S サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

2 個の独立サンプルの検定

「2 個の独立サンプルの検定」手続きは、1 つの変数について 2 つのグループのケースを比較します。

例: 装着感と外観の改善、矯正期間の短縮などを目的とした歯列矯正器が新たに開発されました。新しい歯列矯正器の装着期間が従来の矯正器よりも短いかどうかを調査するため、従来の矯正器を装着する 10 名の子供と、新しい矯正器を装着する 10 名の子供をそれぞれ無作為に選択しました。Mann-Whitney の U の検定により、従来の矯正器を装着した子供に比べ、新しい矯正器を装着した子供のほうが装着期間が短かったことがわかります。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。検定: Mann-Whitney の U 、Moses の外れ値反応、Kolmogorov-Smirnov の Z 、Wald-Wolfowitz のラン。

2 個の独立サンプルの検定データの考慮事項

データ: 順序付け可能な数値変数を使用します。

仮定: 無作為に抽出された独立サンプルを使用します。Mann-Whitney の U の検定は、2 つの分布の同等性を検定します。この検定を使用して 2 つの分布間の位置の相違を検定するには、それらの分布が同じ形状であると仮定する必要があります。

2 個の独立サンプルの検定を行うには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「2 個の独立サンプルの検定...」

2. 1 つ以上の数値変数を選択します。

3. グループ化変数を選択し、「グループの定義」をクリックして、ファイルを 2 つのグループまたはサンプルに分割します。

2 個の独立サンプルの検定の種類

検定の種類: 2 個の独立サンプル (グループ) が同じ母集団から抽出されたものかどうかを検定する場合、4 つの検定を使用することができます。

Mann-Whitney の U の検定は、最も一般的な 2 個の独立サンプルの検定です。この検定は、2 個のグループについては、Wilcoxon の順位和検定と Kruskal-Wallis の検定に相当します。Mann-Whitney は、2 個の母集団のサンプルが位置的に同じかどうかを検定します。両方のグループからの観測値を組み合わせて順位を付け、観測値が同順位の場合は、平均順位を割り当てます。同順位の数、観測値の総数に対して小さい数値でなければなりません。2 個の母集団が位置的に同じである場合、それらの順位は 2 個のサンプル間で無作為に混合されます。この検定は、グループ 1 の得点がグループ 2 の得点より多い場合の回数と、グループ 2 の得点がグループ 1 の得点より多い場合の回数を計算します。Mann-Whitney の U 統計量は、これら 2 つの数の小さいほうの値になります。Wilcoxon 順位和 W 統計量も表示されます。グループの平均順位が同じでない限り、 W は小さいほうの平均順位を持つグループの順位和になります。これは、「2 個の独立サンプルの検定: グループの定義」ダイアログ・ボックスで最後に名前が表示されているグループの順位和です。

Kolmogorov-Smirnov の Z 検定と **Wald-Wolfowitz のラン検定**は、分布の位置と形状の両方における差を検出する、一般的な検定です。Kolmogorov-Smirnov の検定は、両方のサンプルの観測累積分布関数間の最大絶対差に基づいています。最大絶対差が著しく大きい場合、2 個の分布は異なるものと見なされます。Wald-Wolfowitz のラン検定は、両方のグループの観測値を組み合わせて順位付けを行います。2 個のサンプルが同じ母集団から抽出されたものである場合、2 個のグループは順位全体にわたって無作為に散らばります。

Moses の極値反応の検定は、実験変数が一部の被験者に対しては一方向に影響し、他の被験者に対しては反対方向に影響することを前提としています。この検定は、制御グループと比較した極値反応数を検定します。制御グループのスパンに注目するこの検定は、実験グループの極値が制御グループと結合された場合に、その極値がスパンにどの程度の影響を与えるかを示す尺度になります。制御グループは、「2 個の独立サンプルの検定: グループの定義」ダイアログ・ボックスのグループ 1 の値によって定義されます。両方のグループからの観測値を組み合わせて順位が付けられます。制御グループのスパンは、制御グループの最大値と最小値の順位間の差に 1 を加算して計算されます。偶然の外れ値はスパンの範囲を容易にゆがめる可能性があるため、制御ケースの 5% は、制御グループの上端と下端から取り除かれます。

2 個の独立サンプルの検定のグループの定義

ファイルを 2 つのグループまたはサンプルに分割するには、整数値をグループ 1 に入力し、別の整数値をグループ 2 に入力します。他の値を持つケースは分析から除外されます。

2 個の独立サンプルの検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- **記述統計:** 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- **4 分位:** 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- **検定ごとに除外:** 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- **リストごとに除外:** いずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPART TESTS コマンドの追加機能 (2 個の独立サンプルの検定)

コマンド・シンタックス言語を使用すると、Moses 検定から除外するケースの数を指定することもできます (MOSES サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

2 個の対応サンプルの検定

「2 個の対応サンプルの検定」手続きは、2 つの変数の分布を比較します。

例: 一般的に、家族が自分たちの住宅を売りに出す場合、提示価格で売れるでしょうか。Wilcoxon の符号付き順位検定を 10 戸の家のデータに適用すると、7 戸が提示価格よりも安値で、1 戸が高値で、2 戸が提示価格で売れていることがわかります。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。検定: Wilcoxon の符号付き順位検定、符号検定、McNemar 検定。「正確確率検定」オプション (Windows オペレーティング・システム上でのみ使用可能) がインストールされている場合は、周辺等質性検定も使用することができます。

2 個の対応サンプルの検定データの考慮事項

データ: 順序付け可能な数値変数を使用します。

仮定: 2 つの変数に対して特定の分布は仮定されませんが、対応のある差の母集団の分布は対称であると仮定されます。

2 個の対応サンプルの検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「2 個の対応サンプルの検定 ...」

2. 1 つ以上の変数ペアを選択します。

2 個の対応サンプルの検定の種類

このセクションの各検定は、2 つの対応する変数の分布を比較します。どの検定を使用するかは、データの種類に応じて異なります。

連続的なデータの場合は、符号検定または Wilcoxon の符号付き順位検定を使用します。**符号検定**は、すべてのケースについて 2 つの変数間の差を計算し、その差を、正、負、同じのいずれかに分類します。2 つの変数の分布が類似している場合は、正と負の差の数に有意な差はありません。**Wilcoxon の符号付き順位検定**は、ペア間の差の符号と大きさの両方に関する情報を考慮します。Wilcoxon の符号付き順位検定は、データに関する情報をより多く使用するため、符号検定よりも強力です。

2 値データの場合は、**McNemar 検定**を使用します。通常、この検定は、指定されたイベントの発生前と発後に 1 回ずつ使用されます。そのため、各被験者から回答を得るのは 2 回ということになります。

McNemar 検定は、最初の回答比 (イベントが発生する前) と最後の回答比 (イベントが発生した後) が等しいかどうかを判別します。この検定は、計画の前後で実験的介入によって発生する回答の変化を検出する場合に役立ちます。

カテゴリー・データの場合は、**周辺等質性検定**を使用します。この検定は、McNemar の検定を 2 値回答から多重回答に拡張したものです。この検定は、回答の変化を (カイ 2 乗分布を使用して) 検定し、計画

の前後で実験的介入によって発生する回答の変化を検出する場合に役立ちます。正確確率検定がインストールされている場合のみ、周辺等質性検定を使用することができます。

2 個の対応サンプルの検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- 記述統計: 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- 4 分位: 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- 検定ごとに除外: 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- リストごとに除外: いずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPART TESTS コマンドの追加機能 (2 個の関連サンプル)

コマンド・シンタックス言語を使用すると、リスト上の各変数について、変数の検定を行うこともできます。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

複数の独立サンプルの検定

「複数の独立サンプルの検定」手続きは、1 つの変数で複数のケースのグループを比較します。

例: 100 ワット電球が寿命に達するまでの平均時間に、3 種類の製品の間で差があるのでしょうか。

Kruskal-Wallis 一元配置分散分析により、3 種類の製品の平均寿命は異なることがわかります。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。検定: Kruskal-Wallis の H 検定、メディアン検定。

複数の独立サンプルの検定データの考慮事項

データ: 順序付け可能な数値変数を使用します。

仮定: 無作為に抽出された独立サンプルを使用します。Kruskal-Wallis の H 検定では、検定対象のサンプルの形状が類似している必要があります。

複数の独立サンプルの検定を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「 K 個の独立サンプルの検定 ...」

2. 1 つ以上の数値変数を選択します。
3. グループ化変数を選択し、「範囲の定義」をクリックして、グループ化変数の最小値と最大値として整数値を指定します。

複数の独立サンプルの検定の種類

複数の独立サンプルが同じ母集団から抽出されたものかどうかを判断する場合、3 つの検定を使用することができます。Kruskal-Wallis の H 検定、メディアン検定、Jonckheere-Terpstra の検定はいずれも、複数の独立サンプルが同じ母集団から抽出されたものかどうかを検定します。

Mann-Whitney U 検定の拡張である **Kruskal-Wallis の H 検定**は、一元配置分散分析のノンパラメトリック版で、分布の位置の差を検出します。より汎用的な (ただし、それほど強力ではない) **メディアン検定**は、分布の位置と形状の違いを検出します。Kruskal-Wallis の H 検定とメディアン検定では、サンプルが抽出される k 個の母集団に事前の順位付けがないものと仮定されます。

k 個の母集団に自然な事前の順位付け (昇順または降順) がある場合は、**Jonckheere-Terpstra 検定**のほうが適しています。例えば、 k 個の母集団が k 段階の温度上昇を表すとします。異なる温度でも同じ応答分布になるという仮説が、温度が上昇するにつれて応答の大きさが増えていくという対立仮説に対して検定されます。この場合、対立仮説が順序付けされるため、使用する検定としては Jonckheere-Terpstra が最も適しています。Exact Tests アドオン・モジュールがインストールされている場合のみ、Jonckheere-Terpstra 検定を使用することができます。

複数の独立サンプルの検定範囲の定義

範囲を定義するには、「**最小値**」と「**最大値**」に、グループ化変数の最小カテゴリと最大カテゴリに対応する最小値と最大値を整数で入力します。この範囲外の値を持つケースは除外されます。例えば、最小値に 1、最大値に 3 を指定すると、1 から 3 までの整数値だけが使用されます。最小値は最大値よりも小さくしなければなりません。また、必ず両方の値を指定する必要があります。

複数の独立サンプルの検定のオプション

統計: 以下の要約統計のいずれかだけを選択することも、両方を選択することもできます。

- **記述統計:** 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- **4 分位:** 25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値: 欠損値の処理を制御します。

- **検定ごとに除外:** 複数の検定が指定されている場合、欠損値がないかどうか、各検定が個別に評価されます。
- **リストごとに除外:** いずれかの変数に対する欠損値を持つケースは、すべての分析から除外されます。

NPART TESTS コマンドの追加機能 (K 個の独立サンプルの検定)

コマンド・シンタックス言語では、メディアン検定で観測された中央値以外の値を指定することもできます (MEDIAN サブコマンドを使用)。

シンタックスについて詳しくは、「**コマンド・シンタックス・リファレンス**」を参照してください。

複数の対応サンプルの検定

「複数の対応サンプルの検定」手続きは、複数の変数の分布を比較します。

例: 一般の人が、医者、法律家、警察官、教師という職業に対して感じる敬意の程度に差はあるでしょうか。10 人に依頼して、この 4 つの職業に対して感じる敬意の程度で順序を付けてもらいました。Friedman の検定では、一般の人がこの 4 つの職業に対して感じる敬意の程度に差があることがわかります。

統計: 平均値、標準偏差、最小値、最大値、非欠損ケースの数、4 分位。検定: Friedman の検定、Kendall の W 検定、Cochran の Q 検定。

複数の対応サンプルの検定データの考慮事項

データ: 順序付け可能な数値変数を使用します。

仮定: ノンパラメトリック検定では、基本的な分布形状についての仮定は必要ありません。無作為に抽出された依存関係のあるサンプルを使用します。

複数の対応サンプルの検定を行うには

1. メニューから次の項目を選択します。

「分析」 > 「ノンパラメトリック検定」 > 「レガシー ダイアログ」 > 「**K** 個の対応サンプルの検定 ...」

2. 2 つ以上の数値型検定変数を選択します。

複数の対応サンプルの検定の種類

関連する複数の変数の分布を比較する場合、3 つの検定を使用することができます。

Friedman の検定は、1 サンプルの反復測定計画や、セルごとに 1 つの観測値を持つ双方向分散分析に相当するノンパラメトリック検定です。Friedman の検定は、 k 個の関連する変数は同じ母集団から発生するという帰無仮説を検定します。ケースごとに、 k 個の変数が 1 から k までランク付けされます。検定の統計量は、これらのランクに基づきます。

Kendall の W は、Friedman の統計を正規化したものです。Kendall の W は、評価者の一致度の尺度である一致係数として解釈することができます。各ケースは審判または評価者で、各変数は評価対象の項目または個人です。変数ごとに、ランクの合計が計算されます。Kendall の W は、0 (まったく一致していない) から 1 (完全に一致している) までの値を取ります。

Cochran の Q は、Friedman の検定と同じですが、すべての応答が 2 値の場合に使用することができます。この検定は、McNemar 検定を k 個のサンプルの状況に拡張したものです。Cochran の Q は、複数の関連する 2 分変数では平均値が同じになるという仮説を検定します。変数は、同一の個体または一致する個体上で測定されます。

複数の対応サンプルの検定の統計

以下の統計を選択することができます。

- **記述統計:** 平均値、標準偏差、最小値、最大値、および非欠損ケースの数を表示します。
- **4 分位:** 25、50、および 75 パーセンタイルに対応する値を表示します。

NPART TESTS コマンドの追加機能 (K 個の対応サンプルの検定)

シンタックスの詳細については、「*Command Syntax Reference*」を参照してください。

第 28 章 多重回答分析

多重回答分析

多重 2 分グループと多重カテゴリー・グループを分析する場合、2 つの手続きを使用することができます。「多重回答の度数表」手続きは、度数分布表を表示します。「多重回答のクロス集計表」手続きは、2 次元と 3 次元のクロス集計表を表示します。どちらの手続きも、使用する前に多重回答グループを定義する必要があります。

例: この例では、市場調査で多重回答項目を使用する場合について説明します。データは架空のもので、実際のデータではありません。ある航空会社が、自社と競合する航空会社を評価するために、特定の路線を利用する旅客を調査するとします。この例では、アメリカン航空が、シカゴとニューヨーク間で自社の旅客が他の航空会社をどの程度利用しているか、航空会社の選択時にスケジュールとサービスが相対的にどの程度重視されるかについて調査します。客室乗務員が、旅客の搭乗時に簡単な調査票を手渡します。最初の質問は、「アメリカン航空、ユナイテッド航空、トランスワールド航空、US エア、その他の航空会社のうち、最近 6 カ月間にこの路線で 1 回以上利用したすべての航空会社に丸を付けてください」というものです。これは、旅客が複数の回答に丸を付けることができる質問であるため、多重回答質問になります。ただし、変数には各ケースについて 1 つの値しか設定できないため、この質問を直接コード化することはできません。回答をそれぞれの質問にマップするには、複数の変数を使用する必要があります。これには 2 つの方法があります。その 1 つは、それぞれの選択肢 (例: アメリカン航空、ユナイテッド航空、トランスワールド航空、US エア、その他) に対応する変数を定義する方法です。乗客がユナイテッド航空に丸をつけた場合、変数 *united* にコード 1 が割り当てられ、丸を付けなかった場合には 0 が割り当てられます。これは、変数のマッピングの**多重 2 分法**です。回答をマップするもう 1 つの方法は、**多重カテゴリー法**です。この方法では、質問に対して考えられる最大回答数を予測して、それと同じ数の変数を設定し、コードを使用して利用した航空会社を指定します。調査票のサンプルを詳細に調べると、最近 6 カ月間にこの路線で利用した航空会社は、どの旅客も 3 社以下であることがわかります。さらに、航空会社への規制緩和により、その他のカテゴリーに 10 社の航空会社名が記載されていることがわかります。多重回答法を使用する場合、3 つの変数を定義して、1 = アメリカン航空、2 = ユナイテッド航空、3 = トランス・ワールド航空、4 = US エア、5 = デルタ航空 などのように、各変数をコード化することになります。ある旅客がアメリカン航空とトランス・ワールド航空に丸を付けた場合、最初の変数にコード 1、2 番目の変数にコード 3 が割り当てられ、3 番目の変数は欠損値コードとなります。別の旅客は、アメリカン航空とデルタ航空に丸を付けたとします。その場合は、最初の変数にコード 1、2 番目の変数にコード 5 が割り当てられ、3 番目の変数は欠損値コードとなります。一方、多重 2 分法を使用する場合は、14 個の異なる変数を使用することになります。この調査ではどちらのマッピング方法も使用できますが、回答の分布に応じて選択する方法を決定する必要があります。

多重回答グループの定義

「多重回答グループを定義」手続きは、基本変数を多重 2 分グループと多重カテゴリー・グループに分類します。これにより、度数分布表とクロス集計表を作成することができます。多重回答グループは、最大 20 個まで定義することができます。各グループには一意の名前を付ける必要があります。グループを削除するには、多重回答グループのリスト上でそのグループを強調表示して「**削除**」をクリックします。グループを変更するには、リスト上でそのグループを強調表示し、グループの定義特性を変更してから「**変更**」をクリックします。

基本変数は、2 分変数としてコード化することも、カテゴリ変数としてコード化することもできます。2 分変数を使用する場合は、「**2 分**」を選択して多重 2 分グループを作成します。集計値には整数値を入力してください。集計値が 1 回以上発生する各変数は、多重 2 分グループのカテゴリになります。成分変数と同じ範囲の値を持つ多重カテゴリ・グループを作成するには、「**カテゴリ**」を選択します。多重カテゴリ・グループのカテゴリの範囲には、最小値と最大値を整数値で入力してください。この手続きは、成分変数全体で範囲内に含まれる各整数値を合計します。空のカテゴリについては、表は作成されません。

各多重回答グループには、7 文字までの一意の名前を割り当てる必要があります。この手続きは、ユーザーが割り当てた名前の前にドル記号 (\$) を付加します。*casenum*、*sysmis*、*jdate*、*date*、*time*、*length*、*width* の各予約名は使用できません。多重回答グループの名前は、多重回答手続き専用です。多重回答グループの名前を他の手続きで参照することはできません。オプションとして、多重回答グループに対して記述的な変数ラベルを入力することができます。ラベルの最大文字数は、40 文字です。

多重回答グループを定義するには

1. メニューから次の項目を選択します。

「分析」 > 「多重回答」 > 「変数グループの定義...」

2. 2 つ以上の変数を選択します。
3. 変数が 2 分変数としてコード化されている場合は、カウントしたい値を指定します。変数がカテゴリとしてコード化されている場合は、カテゴリの範囲を定義します。
4. 各多重回答グループに対して、一意の名前を入力します。
5. 「追加」をクリックして、多重回答グループを定義済みグループのリストに追加します。

多重回答の度数表

「多重回答の度数表」手続きは、多重回答グループの度数分布表を作成します。最初に、1 つ以上の多重回答グループを定義する必要があります (『多重回答グループの定義』を参照)。

多重 2 分グループの場合、出力に示されるカテゴリ名は、そのグループ内の基本変数に対して定義された変数ラベルから取得されます。変数ラベルが定義されていない場合、変数名がラベルとして使用されます。多重カテゴリ・グループの場合、カテゴリ・ラベルは、グループ内の最初の変数の値ラベルから取得されます。最初の変数で欠損しているカテゴリが、グループ内の他の変数に存在する場合には、その欠損カテゴリに対して値ラベルを定義してください。

欠損値: 欠損値が存在するケースは、テーブル単位で除外されます。以下のオプションのいずれかだけを選択することも、両方を選択することもできます。

- **2 分グループをリストごとに除外:** いずれかの変数に対する欠損値を持つケースを多重 2 分グループの作表から除外します。このオプションは、2 分グループとして定義された多重回答グループに対してのみ適用されます。デフォルトでは、ケースのどの成分変数にも集計値が含まれていない場合、そのケースは多重 2 分グループに対して欠損とみなされます。すべての変数ではなく、一部の变数で欠損値があるケースは、1 つ以上の変数に集計値が含まれている場合、そのグループの作表に含められます。
- **カテゴリ・グループをリストごとに除外:** いずれかの変数に対する欠損値を持つケースを多重カテゴリ・グループの作表から除外します。このオプションは、カテゴリ・グループとして定義されている多重回答グループに対してのみ適用されます。デフォルトでは、定義された範囲内の有効な値がケースのどの成分変数にも含まれていない場合のみ、そのケースは多重カテゴリ・グループに対して欠損とみなされます。

例: 調査の質問から作成された各変数は、基本変数です。多重回答項目を分析するには、各変数を 2 種類の多重回答グループ (多重 2 分グループと多重カテゴリー・グループ) のどちらか 1 つに結合する必要があります。例えば、航空会社の調査で、3 つの航空会社 (アメリカン航空、ユナイテッド航空、トランス・ワールド航空) のうち最近 6 カ月間に利用したのはどの会社かという質問をする場合、2 分変数を使用して**多重 2 分グループ**を定義すると、グループ内の 3 つの変数はそれぞれグループ化変数のカテゴリーになります。3 つの航空会社の度数とパーセントは 1 つの度数分布表に表示されます。回答者が 3 つ以上の航空会社について回答しないことがわかっている場合は、各航空会社に 1 つずつ、計 3 つのコードを持つ変数を 2 つ作成することができます。**多重カテゴリー・グループ**を定義すると、同じコードをまとめて基本変数に追加することによって値が集計されます。結果として生成される値のグループは、各基本変数の値のグループと同じになります。例えば、ユナイテッド航空を挙げた 30 の回答は、航空会社 1 に対してユナイテッド航空を挙げた 5 つの回答と、航空会社 2 に対してユナイテッド航空を挙げた 25 の回答の合計です。3 つの航空会社の度数とパーセントは、1 つの度数分布表に表示されます。

統計: 度数、回答のパーセント、ケースのパーセント、有効なケース数、欠損ケース数を表示する度数分布表。

多重回答の度数分布表データの考慮事項

データ: 多重回答グループを使用します。

仮定: 度数とパーセントは、任意の分布から取得したデータの有用な記述を提供します。

関連プロシージャー: 「多重回答グループを定義」手続きを使用して、多重回答グループを定義することができます。

多重回答の度数分布表を作成するには

1. メニューから次の項目を選択します。

「分析」 > 「多重回答」 > 「度数分布表...」

2. 1 つ以上の多重回答グループを選択します。

多重回答のクロス集計表

「多重回答のクロス集計表」手続きは、多重回答グループ、基本変数、または結合をクロス集計します。ケースまたは回答に基づいたセル・パーセントの取得、欠損値の処理の変更、対応するクロス集計表の作成を行うこともできます。最初に、1 つ以上の多重回答グループを定義する必要があります (『多重回答グループを定義するには』を参照)。

多重 2 分グループの場合、出力に示されるカテゴリー名は、そのグループ内の基本変数に対して定義された変数ラベルから取得されます。変数ラベルが定義されていない場合、変数名がラベルとして使用されます。多重カテゴリー・グループの場合、カテゴリー・ラベルは、グループ内の最初の変数の値ラベルから取得されます。最初の変数で欠損しているカテゴリーが、グループ内の他の変数に存在する場合には、その欠損カテゴリーに対して値ラベルを定義してください。この手続きは、3 つの行にある各列でカテゴリー・ラベルを表示します。各行の最大文字数は 8 文字です。単語が分割されないようにするには、行と列の項目を逆にするか、ラベルを再定義します。

例: 多重 2 分グループと多重カテゴリー・グループは、どちらもこの手続きで他の変数とクロス集計することができます。ある航空会社の調査で、旅客に次のように尋ねました。次の航空会社のうち、最近 6 カ月間に 1 回以上利用したすべての航空会社に丸を付けてください (アメリカン航空、ユナイテッド航空、トランス・ワールド航空)。航空会社を選択する場合、スケジュールとサービスのどちらを重視しますか。

どちらか 1 つだけ選択してください。この質問について、2 分変数または多重カテゴリーとしてデータを入力し、そのデータを 1 つのセットに結合すると、航空会社の選択肢とサービスまたはスケジュールに関連する質問をクロス集計できるようになります。

統計: セル、行、列、合計度数を持つクロス集計表と、セル、行、列、合計パーセントを持つクロス集計表。セルのパーセントは、ケースに基づいて計算することも、回答に基づいて計算することもできます。

多重回答のクロス集計表データの考慮事項

データ: 多重回答グループまたは数値カテゴリー変数を使用します。

仮定: 度数とパーセントは、任意の分布から取得したデータの有用な記述を提供します。

関連プロシージャ: 「多重回答グループを定義」手続きを使用して、多重回答グループを定義することができます。

多重回答のクロス集計表を作成するには

1. メニューから次の項目を選択します。

「分析」 > 「多重回答」 > 「クロス集計表...」

2. クロス集計表の各次元について、1 つ以上の数値変数または多重回答グループを選択します。

3. 各基本変数の範囲を定義します。

オプションで、制御変数または多重回答グループの各カテゴリーについて、2 元配置のクロス集計表を作成することができます。「層」リストに対して、1 つ以上の項目を選択してください。

多重回答のクロス集計表の範囲の定義

値の範囲は、クロス集計表のすべての基本変数について定義する必要があります。集計したい最小カテゴリー値と最大カテゴリー値を整数で入力してください。この範囲外のカテゴリーは、分析から除外されます。範囲内の値は整数であると想定されます (整数以外は切り捨てられます)。

多重回答のクロス集計表のオプション

セルのパーセント: セル度数は常に表示されます。行パーセント、列パーセント、2 次元表 (全体) パーセントについては、表示するかどうかを選択することができます。

パーセンテージ計算の分母: セルのパーセントの分母は、ケース数 (または回答者数) にすることができます。多重カテゴリー・グループ全体で変数を一致させることを選択した場合、このオプションは使用できません。セルのパーセントの分母は、回答数にすることもできます。多重 2 分グループの場合、回答数は、ケース全体でカウントされる値の数に等しくなります。多重カテゴリー・グループの場合、回答数は、定義された範囲内の値の数になります。

欠損値: 以下のオプションのいずれかだけを選択することも、両方を選択することもできます。

- **2 分グループをリストごとに除外:** いずれかの変数に対する欠損値を持つケースを多重 2 分グループの作表から除外します。このオプションは、2 分グループとして定義された多重回答グループに対してのみ適用されます。デフォルトでは、ケースのどの成分変数にも集計値が含まれていない場合、そのケースは多重 2 分グループに対して欠損とみなされます。すべての変数ではなく、一部の変数で欠損値があるケースは、1 つ以上の変数に集計値が含まれている場合、そのグループの作表に含められます。
- **カテゴリー・グループをリストごとに除外:** いずれかの変数に対する欠損値を持つケースを多重カテゴリー・グループの作表から除外します。このオプションは、カテゴリー・グループとして定義されてい

る多重回答グループに対してのみ適用されます。デフォルトでは、定義された範囲内の有効な値がケースのどの成分変数にも含まれていない場合のみ、そのケースは多重カテゴリー・グループに対して欠損とみなされます。

デフォルトでは、2 つの多重カテゴリー・グループを集計する場合、この手続きは、最初のグループ内の各変数を 2 番目のグループ内の各変数と集計して、各セルの度数を合計します。そのため、一部の回答が 1 つの表で複数回表示される場合があります。次のオプションを選択できます。

グループ間で変数を順に整合: 最初のグループ内の最初の変数を 2 番目のグループ内の最初の変数とペアにし、これ以降も同じように処理します。このオプションを選択すると、セル・パーセントの分母は、回答者数ではなく回答数になります。ペア化は、多重 2 分グループや基本変数では使用できません。

MULT RESPONSE コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 5 次元までのクロス集計表を作成する (BY サブコマンドを使用)。
- 値ラベルの抑制など、出力書式オプションを変更する (FORMAT サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 29 章 結果の報告

結果の報告

ケースのリストと記述統計量は、データの調査と表示を行うための基本的なツールです。ケースのリストを取得するには、データ・エディターまたは「ケースの要約」手続きを使用します。度数と記述統計量を取得するには「度数分布表」手続きを使用し、部分母集団の統計量を取得するには「グループの平均」手続きを使用します。各手続きは、情報を明確にするために設計された書式を使用します。別の書式で情報を表示したい場合は、「報告書: 行の集計」と「報告書: 列の集計」で、データの表示を制御することができます。

報告書の行の集計

「報告書: 行の集計」手続きは、各種の要約統計量が行に割り当てられる報告書を作成します。要約統計量の有無にかかわらず、ケースのリストを表示することもできます。

例: ある小売店のチェーンを展開している会社が、給与、勤務年数、各従業員が勤務している店舗と部門などの従業員情報を記録しているとします。この場合、店舗と部門別 (ブレイク変数) に分類した個別の従業員情報 (リスト) を作成して、各店舗、各部門、各店舗内の部門に関する要約統計量 (平均給与など) を表示することができます。

データ列: 必要なケースのリストまたは要約統計量に対するレポート変数をリストし、データ列の表示書式を制御します。

ブレイク列: 報告書を複数のグループに分割する任意のブレイク変数をリストし、要約統計量とブレイク列の表示書式を制御します。ブレイク変数が複数ある場合は、リスト内の先行するブレイク変数のカテゴリ内にある各ブレイク変数の各カテゴリに対して、個別のグループが作成されます。ブレイク変数は、ケースを一定数の有意なカテゴリに分割する個別のカテゴリ変数でなければなりません。各ブレイク変数の個々の値は、データ列の左側の個別の列に、ソートされて表示されます。

報告書: 全体の要約統計量、欠損値の表示、ページ番号付け、表題など、報告書全体の特性を制御します。

ケースの表示: それぞれのケースについて、データ列変数の実際の値 (または値ラベル) を表示します。これにより、集計報告書よりも大幅に長いリスト報告書が作成される場合があります。

プレビュー: 報告書の最初のページだけを表示します。このオプションは、報告書全体を処理することなく報告書の書式をプレビューする場合に役立ちます。

データは並べ替え済み: ブレイク変数を使用する報告書の場合、報告書を作成する前に、データ・ファイルをブレイク変数値でソートする必要があります。データ・ファイルがブレイク変数で既にソートされている場合、このオプションを選択すると処理時間を短縮することができます。このオプションは、プレビュー報告書の実行後に使用すると、特に便利です。

集計報告書: 行の集計を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「報告書」 > 「報告書: 行の集計...」

2. 「データ列」に対して 1 つ以上の変数を選択します。選択された各変数に対して、報告書内に 1 つの列が生成されます。
3. 報告書をサブグループでソートして表示するには、「ブレイク列」に対して 1 つ以上の変数を選択します。
4. ブレイク変数で定義されたサブグループの要約統計量を表示する報告書の場合、「ブレイク列変数」リスト内のブレイク変数を選択し、「ブレイク列」グループ内の「集計」をクリックして集計項目を指定します。
5. 全体的な要約統計量を表示する報告書の場合、「集計」をクリックして集計項目を指定します。

報告書のデータ列/ブレイク列の書式

「書式」ダイアログ・ボックスは、列の表題、列幅、テキストの位置合わせ、データ値または値ラベルの表示を制御します。データ列の「書式」は、報告書ページの右側のデータ列の書式を制御します。ブレイク列の「書式」は、左側のブレイク列の書式を制御します。

列の表題: 選択された変数について、列の表題を制御します。長い表題は、列内で自動的に折り返されます。表題を折り返す場所に手で改行を挿入するには、Enter キーを使用します。

値の表示位置: 選択された変数について、列内のデータ値または値ラベルの位置合わせを制御します。値またはラベルの位置合わせは、列見出しの位置合わせには影響しません。指定した文字数分だけ列の内容をインデントすることも、内容を中央に位置合わせすることもできます。

表示内容: 選択された変数について、データ値と定義済み値ラベルのどちらを表示するかを制御します。値ラベルが定義されていない値については、常にデータ値が表示されます (列の集計報告書のデータ列では使用できません)。

報告書の集計行/最終集計行

2 つの「集計行」ダイアログ・ボックスは、ブレイク・グループと報告書全体について、要約統計量の表示を制御します。「集計行」ダイアログ・ボックスは、ブレイク変数で定義された各カテゴリーのサブグループ統計量を制御します。「最終集計行」ダイアログ・ボックスは、報告書の終わりに表示される全体の統計量を制御します。

使用できる要約統計量は、合計、平均値、最小値、最大値、ケースの数、指定した値の上下でのケースのパーセント、指定した範囲値内のケースのパーセント、標準偏差、尖度、分散、歪度です。

報告書のブレイク・オプション

「ブレイク・オプション」は、ブレイク・カテゴリー情報のスペースとページ編集を制御します。

ページの制御: 選択されたブレイク変数のカテゴリーに対してスペースとページ編集を制御します。ブレイク・カテゴリー間のブランク行の数を指定することも、各ブレイク・カテゴリーを新規ページで開始することもできます。

集計の前の空白行: ブレイク・カテゴリー・ラベル間、またはデータと要約統計量間のブランク行の数を制御します。このオプションは、報告書を結合してブレイク・カテゴリーに個々のケース・リストと要約統計量の両方を表示する場合に特に便利です。このような報告書では、ケース・リストと要約統計量の間にスペースを挿入することができます。

報告書のオプション

「報告書: オプション」は、欠損値の処理と表示、報告書のページ番号付けを制御します。

欠損値のあるケースをリストごとに除外: レポート変数のいずれかに対する欠損値を持つケースを (報告書から) 除外します。

欠損値の表示: データ・ファイル内の欠損値を表す記号を指定することができます。この記号は 1 文字に限られ、システム欠損値 とユーザー欠損値 の両方を表すために使用されます。

開始ページ番号: 報告書の最初のページのページ番号を指定することができます。

報告書のレイアウト

「報告書: レイアウト」ダイアログ・ボックスは、各報告書ページの幅と長さ、ページ上での報告書の配置、ブランク行とラベルの挿入を制御します。

ページ的设计: 行数 (上端と下端) と文字数 (左と右) で表されるページ余白を制御し、余白内での報告書の位置合わせを制御します。

ページ表題と脚注: ページ表題とフッターを報告書の本文から何行分離するかを制御します。

ブレイク列: ブレイク列の表示を制御します。複数のブレイク変数を指定する場合は、各変数を個別の列に配置することも、最初の列に配置することもできます。最初の列にすべてのブレイク変数を配置すると、幅の狭い報告書が作成されます。

列の表題: 表題に付ける下線、表題と報告書本文との間のスペース、列表題の縦の位置合わせなど、列の表題の表示を制御します。

データ列の行とブレイク列のラベル: 各ブレイク・カテゴリーの最初のブレイク・ラベルを基準として、データ列情報 (データ値や要約統計量) の配置を制御します。データ列情報の最初の行は、ブレイク・カテゴリー・ラベルと同じ行から開始することも、指定した行数をブレイク・カテゴリー・ラベル行に加算した行から開始することもできます (列の集計報告書では使用できません)。

報告書の表題

「報告書: 表題」は、報告書の表題とフッターについて、内容と配置を制御します。10 行までのページ表題と 10 行までのページ・フッターを指定して、各行の内容を、左揃え、中央、または右揃えにすることができます。

変数を表題またはフッターに挿入すると、現在の値ラベルまたは変数の値が表題またはフッターに表示されます。表題の場合、変数の値に対応する値ラベルがページの先頭に表示されます。フッターの場合は、変数の値に対応する値ラベルがページの終わりに表示されます。値ラベルがない場合は、実際の値が表示されません。

特殊変数: 特殊変数の *DATE* と *PAGE* を使用して、現在の日付またはページ番号を報告書のヘッダーまたはフッターの任意の行に挿入することができます。データ・ファイルに *DATE* または *PAGE* という変数が含まれている場合、これらの変数を報告書の表題またはフッターで使用することはできません。

報告書の列の集計

「報告書: 列の集計」手続きは、各要約統計量が個別の列に表示される要約統計量を作成します。

例: ある小売店のチェーンを展開している会社が、給与、勤務年数、各従業員が勤務している部門などの従業員情報を記録しているとします。この場合、部門ごとに給与の要約統計量（平均値、最小値、最大値など）を示す報告書を作成することができます。

データ列: 必要な要約統計量に対するレポート変数をリストし、各変数の表示書式と、各変数に表示される要約統計量を制御します。

ブレイク列: 報告書を複数のグループに分割するオプションのブレイク変数をリストし、ブレイク列の表示書式を制御します。ブレイク変数が複数ある場合は、リスト内の先行するブレイク変数のカテゴリ内にある各ブレイク変数の各カテゴリに対して、個別のグループが作成されます。ブレイク変数は、ケースを一定数の有意なカテゴリに分割する個別のカテゴリ変数でなければなりません。

報告書: 欠損値の表示、ページ番号付け、表題など、報告書全体の特性を制御します。

プレビュー: 報告書の最初のページだけを表示します。このオプションは、報告書全体を処理することなく報告書の書式をプレビューする場合に役立ちます。

データは並べ替え済み: ブレイク変数を使用する報告書の場合、報告書を作成する前に、データ・ファイルをブレイク変数値でソートする必要があります。データ・ファイルがブレイク変数で既にソートされている場合、このオプションを選択すると処理時間を短縮することができます。このオプションは、プレビュー報告書の実行後に使用すると、特に便利です。

集計報告書: 列の集計を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「報告書」 > 「報告書: 列の集計...」

2. 「データ列」に対して 1 つ以上の変数を選択します。選択された各変数に対して、報告書内に 1 つの列が生成されます。

3. 変数の集計項目を変更するには、「データ列変数」リストで変数を選択して「集計」をクリックします。

4. 1 つの変数について複数の集計項目を取得するには、必要な集計項目ごとに、ソース・リストで変数を選択して「データ列変数」ボックスに移動します。

5. 既存の列の合計値、平均値、比率、またはその他の関数を含む列を表示するには、「合計変数の挿入」をクリックします。これにより、合計 という変数が「データ列」リストに挿入されます。

6. 報告書をサブグループでソートして表示するには、「ブレイク列」に対して 1 つ以上の変数を選択します。

データ列の集計関数

「集計行」は、選択されたデータ列変数で表示する要約統計量を制御します。

使用できる要約統計量は、合計、平均値、最小値、最大値、ケースの数、指定した値の上下でのケースのパーセント、指定した範囲値内のケースのパーセント、標準偏差、尖度、分散、歪度です。

合計列のデータ列集計

「集計列」は、複数のデータ列を集計する合計の要約統計量を制御します。

選択できる合計の要約統計量は、列の合計、列の平均値、最小値、最大値、2 つの列間の値の差、1 つの列の値を別の列の値で割った商、列の値を乗算した積です。

列の合計: 合計 の列は、「集計列」リスト内の列の合計になります。

列の平均: 合計 の列は、「集計列」リスト内の列の平均値になります。

列の最小: 合計 の列は、「集計列」リスト内の列の最小値になります。

列の最大: 合計 の列は、「集計列」リスト内の列の最大値になります。

1 列目 - 2 列目: 合計 の列は、「集計列」リスト内の列の差になります。「集計列」リストには 2 つの列だけが含まれている必要があります。

1 列目/2 列目: 合計 の列は、「集計列」リスト内の列の商になります。「集計列」リストには 2 つの列だけが含まれている必要があります。

1 列目/2 列目の %: 合計 の列は、「集計列」リスト内の 2 番目の列に対する 1 番目の列のパーセントになります。「集計列」リストには 2 つの列だけが含まれている必要があります。

列の積: 合計 の列は、「集計列」リスト内の列の積になります。

報告書の列の書式

「報告書: 列の集計」のデータ列とブレイク列の書式オプションは、「報告書: 行の集計」で説明しているオプションと同じです。

報告書の列の集計でのブレイク列のオプション

ブレイク列のオプションにより、ブレイク・カテゴリーの小計表示、スペース、ページ編集を制御します。

小計: ブレイク・カテゴリーの小計表示を制御します。

ページの制御: 選択されたブレイク変数のカテゴリーに対してスペースとページ編集を制御します。ブレイク・カテゴリー間のブランク行の数を指定することも、各ブレイク・カテゴリーを新規ページで開始することもできます。

小計の前の空白行: ブレイク・カテゴリー・データと小計の間のブランク行の数を制御します。

報告書の列の集計のオプション

以下のオプションにより、列の集計報告書内の総計の表示、欠損値の表示、ページ編集を制御します。

総計: 各列の総計の表示とラベル付けを行います (総計は、列の下端に表示されます)。

欠損値: 報告書から欠損値を除外することも、報告書内で欠損値を示す単一文字を選択することもできます。

報告書の列の集計のレイアウト

「報告書: 列の集計」の「報告書: レイアウト」オプションは、「報告書: 行の集計」で説明しているオプションと同じです。

REPORT コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 1 つの集計行の列に別の集計関数を表示する。

- データ列変数以外の変数に対して、または集計関数の各種結合 (複合関数) に対して、集計行をデータ列に挿入する。
- 中央値、最頻値、度数、パーセントを集計関数として使用する。
- 要約統計量の表示形式をさらに正確に制御する。
- レポート内のさまざまな箇所に空白行を挿入する。
- レポートのリスト表示で、 n 番目のケースごとに空白行を挿入する。

REPORT シンタックスは複雑なため、シンタックスを使用して新しいレポートを作成する際に、ダイアログ・ボックスを使用して生成されたものに近いレポートを作成するには、対応するシンタックスをコピーして貼り付け、そのシンタックスを修正して必要なレポートを作成すると便利です。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 30 章 信頼性分析

信頼性分析では、測定尺度の特性と、測定尺度を構成する項目の特性を調査することができます。「信頼性分析」手続きは、一般に使用される多数の測定尺度の信頼性を計算するとともに、測定尺度の各項目間の関係についての情報を提供します。級内相関係数を使用して、評価者間の信頼性推定値を計算することができます。

例: 自分で作成した調査票が、顧客の満足度の測定に役立つかどうかを調べることができます。信頼性分析を使用して、調査票の項目が互いにどれだけ関連しているのかを判別し、反復性の全体的な指標や尺度全体の内部的な一貫性を求めることができます。また、尺度から削除する必要がある、問題のある項目を識別することもできます。

統計: 各変数と尺度の記述統計、項目全体の要約統計、項目間の相関と共分散、信頼性推定値、分散分析表、級内相関係数、Hotelling の T^2 、Tukey の加法性の検定。

モデル: 以下の信頼性のモデルを使用することができます。

- **アルファ (Cronbach):** これは、項目間の平均相関に基づく内部一貫性のモデルです。
- **折半法:** このモデルは、尺度を 2 つに分割し、分割した部分間の相関を調べます。
- **Guttman:** このモデルは、真の信頼性の Guttman の下限を計算します。
- **平行モデル:** このモデルは、反復全体において、すべての項目の分散と誤差分散が等しいと仮定します。
- **厳密平行モデル:** このモデルは、平行モデルの仮定だけでなく、すべての項目で平均値が等しいと仮定します。

信頼性分析データの考慮事項

データ: 2 分データ、順序データ、間隔データのいずれかを使用できますが、数値でコード化されたデータでなければなりません。

仮定: 観測値は独立している必要があります。また、誤差は項目間で相関してはなりません。項目の各組は 2 変数正規分布でなければなりません。尺度は加法的でなければなりません。したがって、各項目は合計得点に対して線型になります。

関連プロシージャ: 尺度項目の次元数を調べる場合は (項目得点のパターンを説明するために複数の構成が必要かどうかを調べる場合)、因子分析または多次元尺度法を使用します。等質な変数グループを特定するには、階層クラスター分析を使用して変数をクラスター化します。

信頼性分析を実行するには

1. メニューから次の項目を選択します。
「分析」 > 「スケール」 > 「信頼性分析...」
2. 加法尺度の潜在的成分として、複数の変数を選択します。
3. 「モデル」ドロップダウン・リストからモデルを選択します。

信頼性分析の統計

尺度と項目を記述するさまざまな統計を選択することができます。デフォルトで報告される統計は、ケース数、項目数、および以下の信頼性推定値です。

- **アルファ・モデル:** アルファ係数。2 分変数の場合、これは Kuder-Richardson 20 (KR20) 係数に相当します。
- **折半法のモデル:** 形式間の相関、Guttman の折半法の信頼性、Spearman-Brown の信頼性 (等長と不等長)、2 つの部分ごとのアルファ係数。
- **Guttman のモデル:** 信頼係数ラムダ 1 からラムダ 6 まで。
- **平行モデルと厳密平行モデル:** モデルの適合度の検定、誤差分散の推定値、共通分散の推定値、真の分散の推定値、共通項目間相関の推定値、信頼性の推定値、信頼性の不偏推定値。

記述統計: すべてのケースについて、尺度または項目の記述統計量を生成します。

- **項目:** すべてのケースについて、項目の記述統計量を生成します。
- **尺度:** 尺度の記述統計量を生成します。
- **項目を削除したときの尺度:** 各項目を、他の項目から構成される尺度と比較する要約統計量を表示します。この統計量には、項目が尺度から削除された場合の尺度の平均と分散、項目と他の項目から構成される尺度との相関、項目が尺度から削除された場合の Cronbach のアルファが含まれます。

要約: 尺度のすべての項目にわたる項目分布の記述統計量を求めます。

- **Means (平均値 (信頼性分析)):** 項目平均値の要約統計量。最小項目平均値、最大項目平均値、平均項目平均値、平均値の範囲と分散、および最大平均値と最小平均値の比を表示します。
- **分散:** 項目分散の要約統計量。最小項目分散、最大項目分散、平均項目分散、項目分散の範囲と分散、および最大項目分散と最小項目分散の比を表示します。
- **共分散:** 項目間共分散の要約統計量。最小項目間共分散、最大項目間共分散、平均項目間共分散、項目間共分散の範囲と分散、および最大項目間共分散と最小項目間共分散の比を表示します。
- **相関:** 項目間相関の要約統計量。最小項目間相関、最大項目間相関、平均項目間相関、項目間相関の範囲と分散、および最大項目間相関と最小項目間相関の比を表示します。

項目間: 項目間の相関行列または共分散行列を作成します。

分散分析表: 平均が等しい検定を行います。

- **F 検定:** 反復測定の分散分析テーブルを表示します。
- **Friedman カイ 2 乗:** Friedman のカイ 2 乗および Kendall の一致係数を表示します。このオプションは、順位の形式になっているデータに適しています。カイ 2 乗検定は、分散分析テーブルの通常の F 検定に代わるものです。
- **Cochran カイ 2 乗:** Cochran の Q を表示します。このオプションは 2 分されるデータに適しています。Q 統計量は、分散分析テーブルの通常の F 統計量に代わるものです。

Hotelling の T2: スケール上のすべての項目が同じ平均を持つという帰無仮説の多変量の検定を行います。

Tukey の加法性の検定: 項目間に相乗的交互作用がないという仮説の検定を行います。

級内相関係数: ケース内の値の一貫性または一致を測定します。

- **モデル:** 級内相関係数を計算するモデルを選択します。選択できるモデルは、二元配置混合、二元配置変量、一元配置変量です。人的効果の変量で項目効果が固定である場合は「二元配置混合」を、人的効果と項目効果の両方が変量である場合は「二元配置変量」を、人的効果の変量である場合は「一元配置変量」を選択します。
- **タイプ:** 指標の種類を選択します。選択できる種類は、一貫性と絶対一致です。
- **信頼区間:** 信頼区間のレベルを指定します。デフォルトは 95% です。
- **検定値:** 仮説検定の係数の仮説値を指定します。この値が、観測値と比較されます。デフォルト値は 0 です。

RELIABILITY コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 相関行列の読み取りと分析を行う。
- 今後の分析用に、相関行列を書き込む。
- 折半法について、等分割以外の分割方法を指定する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 31 章 多次元スケーリング法

多次元スケーリング法は、オブジェクトまたはケース間の距離測度のセットに構造を見いだそうとする方法です。この作業は、観測値を概念的な空間 (通常は 2 次元または 3 次元) の中の特定の場所に割り当てることによって達成されます。この時、空間内の点の間の距離が任意の非類似度にできるだけ一致するように割り当てられます。多くの場合、データをさらに理解するために、この概念的な空間の次元を解釈し、使用することができます。

変数が客観的に測定された場合は、データ分解のテクニックとして多次元スケーリング法を使用できます (必要に応じて、「多次元スケーリング法」手続きは多変量データから距離を計算します)。多次元スケーリング法は、オブジェクトまたは概念間の非類似度を主観的に評価する場合にも適用できます。さらに、「多次元スケーリング法」手続きでは、複数の評価者や複数のアンケート回答者がいる場合など、複数のソースからの非類似度のデータを処理することができます。

例: 人はどのようにして異なる車同士の関連性を認識するでしょうか? 異なる車種の間類似度を示す回答者のデータがある場合は、多次元スケーリング法を使用して、消費者の認識を示す次元を特定できます。例えば、車の価格とサイズが、回答者が報告した類似度を説明する 2 次元空間を定義するかもしれません。

統計: 各モデル: データ行列、最適スケール化データ行列、S- ストレス (Young)、ストレス (Kruskal)、RSQ、刺激座標、平均ストレス、および各刺激の RSQ (RMDS モデル)。個別差分 (INDSCAL) モデル: 被験者の重み付け、および各被験者の奇異性のインデックス。反復された多次元スケーリング法モデルの各行列: 各刺激のストレスおよび RSQ。作図: 刺激座標 (2 次元または 3 次元)、差異対距離の散布図。

多次元スケーリング法データの考慮事項

データ: データが非類似度データの場合、すべての非類似度は量的でなければならず、さらに同じ測定基準で測定される必要があります。データが多変量データの場合、変数は量的データ、2 値データ、または度数データです。スケーリングの違いは解に影響する場合があるので、変数のスケーリングは重要な問題です。変数に大きなスケーリングの違いがある場合 (ある変数はドルで測定され、別の変数は年で測定されるなど)、変数の標準化を検討する必要があります (「多次元スケーリング法」手続きでは、標準化は自動的に実行できます)。

仮定: 「多次元スケーリング法」手続きは比較的、分布の仮定を必要としません。結果を正しく計算するには、多次元スケーリング法のオプション・ダイアログ・ボックスで、必ず適切な測定レベル (順序、区間、または比率) を選択してください。

関連プロシージャ: データ分解を目的としていて、特に変数が量的な場合は、代替方法として因子分析を検討します。類似ケースのグループを識別したい場合は、多次元スケーリング分析を、階層クラスター分析または *k*-Means クラスター分析で補足することを検討してください。

多次元スケーリング分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「スケール」 > 「多次元スケーリング法...」

2. 分析に最低 4 つの数値変数を選択します。
3. 「距離」グループで、「データが距離行列」または「データから距離行列を作成」を選択します。

4. 「データから距離行列を作成」を選択すると、個別の行列にグループ化変数を選択することもできます。グループ化変数は、数値と文字列のどちらでもかまいません。

オプションで、以下の操作も実行することができます。

- データが距離の場合、距離行列の形状を指定する。
- データから距離を作成する場合、使用する距離測度を指定する。

多次元スケーリング法のデータの型

アクティブ・データ・セットが 1 組のオブジェクト内の距離、または 2 組のオブジェクト間の距離を表している場合、正しい結果を得るには、データ行列の型を指定する必要があります。

注: 「モデル」ダイアログ・ボックスで行の条件付けを指定している場合、「対称形」は選択できません。

多次元スケーリング法の測度の作成

多次元スケーリング法では、非類似度データを使用してスケーリング解を作成します。データが多変量データ (測定された複数の変数値) の場合は、多次元スケーリング法の解を計算するために非類似度データを作成しなければなりません。データから非類似度の測定方法の作成を詳細に指定できます。

測定: 分析で使用する非類似度を指定できます。データの種類に応じて「測定」グループから選択肢を 1 つ選び、その後、測定の種類に応じてドロップダウン・リストから測定方法を 1 つ選択します。選択肢は次のとおりです。

- **区間:** ユークリッド距離、平方ユークリッド距離、Chebychev、都市ブロック、Minkowski、またはカスタマイズ。
- **度数:** カイ 2 乗測度、またはファイ 2 乗測度。
- **2 値:** ユークリッド距離、平方ユークリッド距離、サイズの差、パターンの違い、分散、または Lance-Williams。

距離行列の作成: 分析の単位を選択できます。「変数間」または「ケース間」が選択可能です。

値の変換: 特定のケース (変数が非常に異なるスケールで測定される場合など) では、値を標準化してから近接を計算することをお勧めします (2 値データには適用できません)。「標準化」ドロップダウン・リストから標準化の方法を選択してください。標準化が不要な場合は、「なし」を選択します。

多次元スケーリング法モデル

多次元スケーリング法モデルを正確に推定できるかどうかは、データとモデル自体の要素によって決まります。

測定レベル: データのレベルを指定できます。「順序」、「区間」または「比率」が選択可能です。変数が順序型の場合は、「同一順序の観測を同一としない」を選択すると、同順位のデータ (ケースが違うが同じ値を持つデータ) が最適な形で解決されるように、変数が連続変数として扱われます。

条件付け: どの比較が重要かを指定できます。「行列」、「行」、または「無条件」が選択可能です。

次元: スケーリング解 (複数可) の次元を指定できます。範囲内の各数値に対して、解が 1 つ計算されます。1 から 6 までの整数を指定してください。スケーリング・モデルとして「ユークリッド距離」を選択した場合のみ、最小値 1 を指定できます。解を 1 つだけ求めるには、最小値と最大値に同じ数値を指定します。

スケーリング・モデル: スケーリングを実行するための仮定を指定できます。「ユークリッド距離」または「個別差分ユークリッド距離」(INDSCAL と呼ばれます) が選択可能です。個別差分ユークリッド距離モデルでは、データに対して適切な場合は、「負の被験者の重み付けを許可」を選択できます。

多次元スケーリング法のオプション

多次元スケーリング法分析のオプションを指定できます。

表示: 各種の出力形式を選択できます。「グループ・プロット」、「被験者ごとのプロット」、「データ行列」、および「モデルとオプションの要約」が選択可能なオプションです。

「基準」。反復を停止する基準を指定できます。デフォルトを変更するには、「S-ストレス収束値」、「最小S-ストレス値」、および「最大反復回数」に値を入力します。

距離が n 以下は欠損値として扱う: この値より小さい距離は分析から除外されます。

ALSCAL コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 多次元スケーリング法の文献で ASCAL、AINDS、GEMSCAL と呼ばれる 3 つの追加モデルを使用する。
- 区間データと比率データに対して多項式変換を実行する。
- 順序データを使用して、(距離ではなく) 類似度を分析する。
- 名義データを分析する。
- さまざまな座標や重み付け行列をファイルに保存し、それらを分析用に読み込む。
- 多次元展開を制限する。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

第 32 章 比率統計

「比率統計量」手続きは、2 つのスケール変数間の比率を記述する要約統計量の総合的なリストを作成します。

グループ化変数値を使用して、出力を昇順または降順にソートすることができます。比率統計量報告書は出力内で非表示にすることができ、結果は外部ファイルに保存することができます。

例: 家屋の査定額と販売額との間の比率は、5 つの郡においてどの程度異なるのでしょうか。出力結果から、比率の分布は郡によって大きく異なることがわかります。

統計: ユーザー指定の範囲または中央値比率に含まれたパーセンテージで計算された中央値、平均値、重み付き平均値、信頼区間、散らばり係数 (COD)、中央値中心化の変動係数、平均値中心化の変動係数、価格関連格差 (PRD)、標準偏差、平均絶対偏差 (AAD)、範囲、最小値と最大値、濃度インデックス。

比率統計データの考慮事項

データ: 数値コードまたは文字列を使用してグループ化変数をコード化します (名義尺度または順序尺度)。

仮定: 比率の分子と分母を定義する変数は、正の値を取るスケール変数でなければなりません。

比率統計を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「記述統計」 > 「比率...」

2. 分子変数を選択します。

3. 分母変数を選択します。

必要に応じて、以下の操作を実行することができます。

- グループ化変数を選択し、結果でのグループの順序を指定する。
- ビューアーに結果を表示するかどうかを選択する。
- 後で使用できるように結果を外部ファイルに保存するかどうかを選択し、結果を保存するファイルの名前を指定する。

比率統計

中心傾向: 中心傾向の測度は、比率の分布を示す統計量です。

- **中央値:** この値よりも小さい比率の数と、この値よりも大きい比率の数が同じになります。
- **平均値:** 比率の合計を比率の総数で割った値です。
- **重み付き平均値:** 分子の平均値を分母の平均値で割った値です。重み付き平均値は、分母で重みを付けた比率の平均値にもなります。
- **信頼区間:** 平均値、中央値、重み付き平均値の信頼区間を表示します (要求された場合)。信頼水準として、0 以上 100 未満の値を指定してください。

散らばり: 以下の各統計は、観測値での変動量 (散布度) を測定します。

- **AAD:** 平均絶対偏差は、中央値の比率の絶対偏差合計を比率の総数で割った値です。

- **COD:** 散らばり係数は、平均絶対偏差を中央値のパーセントで表した値です。
- **PRD:** 価格関連格差 (回帰指数とも呼ばれます) は、平均値を加重平均で割った値です。
- **中央値中心化の COV:** 中央値中心化の変動係数は、中央値からの偏差の平方平方根を中央値のパーセントで表した値です。
- **平均値中心化の COV:** 平均値中心化の変動係数は、標準偏差を平均値のパーセントで表した値です。
- **標準偏差:** 標準偏差は、平均値の比率の平方偏差合計を、比率の総数から 1 を引いた数で割り、正の平方根を取った値です。
- **範囲:** 範囲とは、最大比から最小比を引いた値のことです。
- **最小値:** 最小値とは、最小比のことです。
- **最大値:** 最大値とは、最大比のことです。

濃度インデックス: 集中係数は、区間に含まれる比率のパーセントを表します。この値は、以下の 2 つの方法で計算することができます。

- **範囲間の比率:** 区間の最小値と最大値を指定することにより、区間が明示的に定義されます。最小値の比率と最大値の比率を入力し、「追加」をクリックして区間を取得します。
- **範囲内の比率:** 中央値のパーセントを指定することにより、区間が暗黙的に定義されます。0 から 100 までの値を入力して、「追加」をクリックします。区間の下限は「 $(1 - 0.01 \times \text{値}) \times \text{中央値}$ 」で算出され、上限は「 $(1 + 0.01 \times \text{値}) \times \text{中央値}$ 」で算出されます。

第 33 章 ROC 曲線

2 つのカテゴリを持つ 1 つの変数で被験者を分類する場合、その分類方法のパフォーマンスを評価するには、この手続きを使用すると便利です。

例: 銀行は、ローン返済の履行と不履行について顧客を正しく分類したいと考えています。そのための意思決定を行う特殊な方法が開発されています。ROC 曲線を使用して、こうした方法のパフォーマンスを評価することができます。

統計: 信頼区間を持つ ROC 曲線の下領域と ROC 曲線の座標点。作図: ROC 曲線。

方法: ROC 曲線の下領域の面積の推定値は、ノンパラメトリックに計算することも、2 負指数モデルを使用してパラメトリックに計算することもできます。

ROC 曲線データの考慮事項

データ: 検定変数は量的変数です。多くの場合、検定変数は、判別分析またはロジスティック回帰からの確率か、任意のスケール上の得点 (被験者をカテゴリに分類する評価者の「確信の強さ」を示す) によって構成されます。状態変数は、任意のタイプにすることができます。この変数は、被験者が属する真のカテゴリを示します。状態変数の値は、どのカテゴリを正とみなすかべきかを示します。

仮定: 評価者のスケール上の数値が大きくなるほど、被験者が一方のカテゴリに属するという評価者の確信が強くなり、このスケール上の数値が小さくなるほど、被験者が他方のカテゴリに属するという評価者の確信が強くなると想定されます。ユーザーは、どの方向が正であるかを指定する必要があります。さらに、各被験者が属している真のカテゴリがわかっているものと想定されます。

ROC 曲線を作成するには

1. メニューから次の項目を選択します。

「分析」 > 「ROC 曲線...」

2. 1 つ以上の確率検定変数を選択します。

3. 1 つの状態変数を選択します。

4. 状態変数の正の値を特定します。

ROC 曲線のオプション

ROC 分析の場合、以下のオプションを指定することができます。

分類: 正の分類を行う際に、分割値を含めるか除外するかを指定することができます。この設定は、現在、出力には影響しません。

検定方向: 正のカテゴリに対するスケールの方向を指定することができます。

面積の標準誤差のパラメーター: 曲線の下領域の面積の標準誤差を推定する方法を指定することができます。使用できる方法は、ノンパラメトリックと 2 負指数です。さらに、信頼区間の水準を設定することもできます。有効な範囲は 50.1% から 99.9% です。

欠損値: 欠損値の処理方法を指定することができます。

第 34 章 シミュレーション

線型回帰などの予測モデルでは、結果や目標値を予測するための既知の入力セットが必要になります。しかし、多くの場合、実世界のアプリケーションの入力値は不確定なものです。シミュレーションにより、予測モデルに対する入力の不確定要素を考慮し、その不確定要素を前提としてモデルのさまざまな出力結果の尤度を評価することができます。例えば、材料のコストを入力として持つ収益モデルを使用する場合、そのコストには、市場の変動による不確定性が存在します。シミュレーションを使用すると、その不確定性をモデル化し、その不確定性が収益に与える影響を判断することができます。

IBM SPSS Statistics のシミュレーションでは、モンテカルロ法が使用されます。不確定性の入力、三角分布などの確率分布によってモデル化され、これらの分布から描画することにより、不確定性の入力に対するシミュレーションされた値が生成されます。既知の値を持つ入力は、その既知の値で固定されます。不確定性の入力ごとにシミュレーションされた値と、既知の入力の固定値を使用して、予測モデルが評価され、モデルの目標が計算されます。このプロセスを何度も繰り返し (通常は数万回から数十万回)、その結果として得られた目標値の分布を使用して、確率的な性質の質問に回答することができます。IBM SPSS Statistics のコンテキストでは、プロセスを繰り返すごとに、不確定性の入力に対してシミュレーションされた値、固定入力の値、およびモデルの予測目標のセットで構成されたデータの個別のケース (レコード) が生成されます。

シミュレーションする変数の確率分布を指定することで、予測モデルを使用せずにデータをシミュレーションすることもできます。生成されるデータの各ケースは、指定された変数のシミュレーションされた値のセットで構成されます。

シミュレーションを実行するには、予測モデル、不確定性の入力の確率分布、これらの入力と固定入力の値との相関などの詳細情報を指定する必要があります。シミュレーションのすべての詳細情報を指定したら、シミュレーションを実行することができます。また、必要に応じて、その指定内容をシミュレーション・プラン・ファイルに保存することができます。シミュレーション・プランは、他のユーザーと共有することができます。プランを共有しているユーザーは、それがどのように作成されたかを知らなくても、シミュレーションを実行することができます。

2 つのインターフェースを使用して、シミュレーションを操作することができます。シミュレーション・ビルダーは、シミュレーションを設計して実行するユーザー向けの高度なインターフェースです。このインターフェースには、シミュレーションの設計、シミュレーション・プラン・ファイルへの仕様の保存、出力の指定、シミュレーションの起動を行うための包括的な機能が用意されています。シミュレーションは、IBM SPSS のモデル・ファイル、またはシミュレーション・ビルダーで定義した一連のカスタム方程式に基づいて作成することができます。既存のシミュレーション・プランをシミュレーション・ビルダーに読み込み、設定を変更してシミュレーションを実行することもできます。必要に応じて、更新したプランを保存することもできます。シミュレーション・プランを持っていて、主にシミュレーションを実行するユーザーについては、よりシンプルなインターフェースが用意されています。このインターフェースを使用して設定を変更し、さまざまな条件下でシミュレーションを実行することができますが、シミュレーションを設計するための、シミュレーション・ビルダーのすべての機能が用意されているわけではありません。

モデル・ファイルに基づいてシミュレーションを設計するには

1. メニューから次の項目を選択します。

「分析」 > 「シミュレーション...」

2. 「SPSS モデル ファイルを選択」をクリックし、「続行」をクリックします。
3. モデル・ファイルを開きます。

モデル・ファイルは、IBM SPSS Statistics または IBM SPSS Modeler から作成されたモデル PMML を格納する XML ファイルです。詳しくは、179 ページの『「モデル」タブ』のトピックを参照してください。

4. シミュレーション・ビルダーの「シミュレーション」タブで、シミュレーションされた入力の確率分布と固定入力の値を指定します。アクティブ・データ・セットにシミュレーションされた入力の履歴データが含まれている場合は、「すべて適合」をクリックします。これにより、各入力のデータに最も適合する分布が自動的に判別され、入力間の相関が判別されます。履歴データに適合しないシミュレーションされた各入力については、分布の種類を選択して必須パラメーターを入力することにより、明示的に分布を指定する必要があります。
5. 「実行」をクリックしてシミュレーションを実行します。デフォルトでは、シミュレーションの詳細を指定するシミュレーション・プランは、「保存」設定で指定された場所に保存されます。

使用可能なオプションは次のとおりです。

- 保存されたシミュレーション計画の場所を変更する。
- シミュレーションされた入力間で既知の相関を指定する。
- カテゴリー入力間の連関の分割表を自動的に計算し、それらの入力のデータが生成されるときにそれらの連関を使用する。
- 固定入力値のばらつきの影響、またはシミュレーションされた入力の分布パラメーターのばらつきの影響を調査するための感度分析を指定する。
- 生成する最大ケース数や裾サンプルの要求など、詳細なオプションを指定する。
- 出力をカスタマイズする。
- シミュレーションされたデータをデータ・ファイルに保存する。

カスタム方程式に基づいてシミュレーションを設計するには

1. メニューから次の項目を選択します。

「分析」 > 「シミュレーション...」

2. 「等式の入力」をクリックし、「続行」をクリックします。
3. シミュレーション・ビルダーの「モデル」タブで「新しい等式」をクリックし、予測モデルの各方程式を定義します。
4. 「シミュレーション」タブをクリックし、シミュレーションされた入力の確率分布と固定入力の値を指定します。アクティブ・データ・セットにシミュレーションされた入力の履歴データが含まれている場合は、「すべて適合」をクリックします。これにより、各入力のデータに最も適合する分布が自動的に判別され、入力間の相関が判別されます。履歴データに適合しないシミュレーションされた各入力については、分布の種類を選択して必須パラメーターを入力することにより、明示的に分布を指定する必要があります。

5. 「実行」をクリックしてシミュレーションを実行します。デフォルトでは、シミュレーションの詳細を指定するシミュレーション・プランは、「保存」設定で指定された場所に保存されます。

使用可能なオプションは次のとおりです。

- 保存されたシミュレーション計画の場所を変更する。
- シミュレーションされた入力間で既知の相関を指定する。
- カテゴリー入力間の連関の分割表を自動的に計算し、それらの入力のデータが生成されるときにそれらの連関を使用する。
- 固定入力値のばらつきの影響、またはシミュレーションされた入力の分布パラメーターのばらつきの影響を調査するための感度分析を指定する。
- 生成する最大ケース数や裾サンプルの要求など、詳細なオプションを指定する。
- 出力をカスタマイズする。
- シミュレーションされたデータをデータ・ファイルに保存する。

予測モデルを使用しないシミュレーションを設計するには

1. メニューから次の項目を選択します。

「分析」 > 「シミュレーション...」

2. 「シミュレーション データの作成」をクリックし、「続行」をクリックします。
3. シミュレーション・ビルダーの「モデル」タブで、シミュレーションするフィールドを選択します。アクティブ・データ・セットのフィールドを選択することも、「新規」をクリックして新しいフィールドを定義することもできます。
4. 「シミュレーション」タブをクリックし、シミュレーションするフィールドの確率分布を指定します。アクティブ・データ・セットにそれらのフィールドのうちいずれかの履歴データが含まれている場合は、「すべて適合」をクリックします。これにより、データに最も適合する分布が自動的に判別され、フィールド間の相関が判別されます。履歴データに適合しないフィールドについては、分布の種類を選択して必須パラメーターを入力することにより、明示的に分布を指定する必要があります。
5. 「実行」をクリックしてシミュレーションを実行します。デフォルトでは、シミュレーションされたデータは、「保存」設定で指定された新規データ・セットに保存されます。さらに、シミュレーションの詳細を指定するシミュレーション・プランも、「保存」設定で指定された場所に保存されます。

使用可能なオプションは次のとおりです。

- シミュレーションされたデータまたは保存されたシミュレーション・プランの場所を変更する。
- シミュレーションするフィールド間の既知の相関を指定する。
- カテゴリー・フィールド間の連関の分割表を自動的に計算し、それらのフィールドのデータが生成されるときにそれらの連関を使用する。
- シミュレーションするフィールドの分布パラメーターのばらつきの影響を調査するための感度分析を指定する。
- 生成するケース数の設定などの詳細なオプションを指定する。

シミュレーション・プランからシミュレーションを実行するには

シミュレーション・プランからシミュレーションを実行する場合、2つの方法があります。1つは「シミュレーションの実行」ダイアログ（主にシミュレーション・プランから実行するために設計されたダイアログ）を使用する方法で、もう1つはシミュレーション・ビルダーを使用する方法です。

「シミュレーションの実行」ダイアログを使用するには

1. メニューから次の項目を選択します。

「分析」 > 「シミュレーション...」

2. 「既存のシミュレーション プランを開く」をクリックします。
3. 「シミュレーション ビルダーで開く」チェック・ボックスにチェック・マークが付いていないことを確認してから、「続行」をクリックします。
4. シミュレーション・プランを開きます。
5. 「シミュレーションの実行」ダイアログで、「実行」をクリックします。

シミュレーション・ビルダーからシミュレーションを実行するには

1. メニューから次の項目を選択します。

「分析」 > 「シミュレーション...」

2. 「既存のシミュレーション プランを開く」をクリックします。
3. 「シミュレーション ビルダーで開く」チェック・ボックスを選択して、「続行」をクリックします。
4. シミュレーション・プランを開きます。
5. 「シミュレーション」タブで、必要に応じて設定を変更します。
6. 「実行」をクリックしてシミュレーションを実行します。

必要に応じて、以下の操作を行うことができます。

- 感度分析をセットアップまたは変更して、固定入力の値の変更による影響や、シミュレーションされた入力の分布パラメーターの変更による影響を調べる。
- シミュレーションされた入力の分布と相関を新しいデータに再適合させる。
- シミュレーションされた入力の分布を変更する。
- 出力をカスタマイズする。
- シミュレーションされたデータをデータ・ファイルに保存する。

シミュレーション・ビルダー

シミュレーション・ビルダーには、シミュレーションを設計して実行するための包括的な機能が用意されています。これを使用して、次の一般的なタスクを実行することができます。

- PMML モデル・ファイルに定義された IBM SPSS モデルのシミュレーションを設計して実行する。
- ユーザー指定の一連のカスタム方程式で定義された予測モデルのシミュレーションを設計して実行する。
- 予測モデルを使用せずにデータを生成するシミュレーションを設計して実行する。
- 既存のシミュレーション・プランに基づいてシミュレーションを実行し、必要に応じてプランの設定を変更する。

「モデル」タブ

予測モデルに基づくシミュレーションの場合は、「モデル」タブでモデルのソースを指定します。予測モデルを含まないシミュレーションの場合は、「モデル」タブでシミュレーション対象のフィールドを指定します。

SPSS モデル・ファイルを選択: このオプションを選択すると、予測モデルが IBM SPSS モデル・ファイル内で定義されます。IBM SPSS のモデル・ファイルは、IBM SPSS Statistics または IBM SPSS Modeler から作成されたモデル PMML を含む XML ファイルまたは圧縮ファイル・アーカイブ (.zip ファイル) です。IBM SPSS Statistics の線型回帰やデシジョン・ツリーなど、予測モデルを手順によって作成し、モデル・ファイルにエクスポートすることができます。別のモデル・ファイルを使用するには、「参照」をクリックして、必要なファイルにナビゲートします。

シミュレーションでサポートされている PMML モデル

- 線型回帰
- 自動線型モデル
- 一般化線型モデル
- 一般化線型混合モデル
- 一般線型モデル
- 2 項ロジスティック回帰
- 多項ロジスティック回帰
- 順序多項回帰
- Cox 回帰
- ツリー
- ブースティング・ツリー (C5)
- 判別
- 2 段階クラスター
- 大規模ファイルのクラスター
- ニューラル・ネット
- ルール・セット (デシジョン・リスト)

注:

- シミュレーションでは、複数の目標フィールド (変数) または区分が存在する PMML モデルを使用することはできません。
- 二項ロジスティック回帰モデルへの文字列入力の値は 8 バイトに制限されています。このような文字列入力をアクティブ・データセットに対して適合させる場合、データ内の値の長さが 8 バイトを超えないようにしてください。8 バイトを超えるデータ値は、入力に関連付けられているカテゴリー分布から除外され、「不整合カテゴリー」出力テーブルに不一致として表示されます。

等式の入力: このオプションは、ユーザーが作成した 1 つ以上の方程式によって予測モデルを構成することを指定します。方程式を作成するには、「新しい等式」をクリックします。この操作により、方程式エディターが開きます。既存の方程式を変更したり、コピーして新しい方程式のテンプレートとして使用することができます。また、既存の方程式の順序を変更したり、削除したりすることもできます。

- シミュレーション・ビルダーは、連立方程式のシステムや、目標変数内で線型ではない方程式のシステムはサポートしていません。

- カスタム方程式は、指定された順序で評価されます。特定の目標の方程式が他の目標に依存する場合、先行する方程式によって他の目標を定義する必要があります。

例えば、以下の 3 つの方程式がある場合、利益 (*profit*) の方程式は収益 (*revenue*) と費用 (*expenses*) の値によって異なるため、収益 と費用 の方程式が利益 の方程式よりも先行している必要があります。

$revenue = price * volume$

$expenses = fixed + volume * (unit_cost_materials + unit_cost_labor)$

$profit = revenue - expenses$

モデルなしでシミュレーション データを作成: このオプションは、予測モデルを使用せずにデータをシミュレーションする場合に選択します。シミュレーション対象のフィールドを指定するには、アクティブ・データ・セットからフィールドを選択するか、「新規」をクリックして新規フィールドを定義します。

方程式エディター

方程式エディターを使用して、予測モデルのカスタム方程式の作成や変更を行うことができます。

- 方程式の数式には、アクティブ・データ・セットのフィールドや、方程式エディターで定義した新しい入力フィールドを含めることができます。
 - 尺度や値ラベル、目標に対して出力を生成するかどうかなど、目標のプロパティを指定することができます。
 - 以前に定義した方程式の目標を現在の方程式に対する入力として使用して、結合方程式を作成することができます。
 - 記述的なコメントを方程式に追加することができます。コメントは、方程式とともに「モデル」タブに表示されます。
1. 目標の名前を入力します。オプションで、「目標」テキスト・ボックスの下にある「編集」をクリックして「定義された入力」ダイアログを開き、目標のデフォルト・プロパティを変更することができます。
 2. 式を作成するには、「数式」フィールドに成分を貼り付けるか、「数式」フィールドに直接入力します。
- アクティブ・データ・セットのフィールドを使用して式を作成することも、「新規」ボタンをクリックして新しい入力を定義することもできます。これにより、「入力の定義」ダイアログが開きます。
 - 関数を貼り付けるには、「関数グループ」リストからグループを選択し、「関数」リストで関数をダブルクリックします (または、関数を選択し、「関数グループ」リストの隣にある矢印をクリックします)。疑問符で示されたパラメーターを入力します。「すべて」というラベルの付いた関数グループには、使用可能なすべての関数がリストされます。現在選択されている関数の短い説明が、ダイアログ・ボックスの予約領域に表示されます。
 - 文字定数は、引用符で囲む必要があります。
 - 値に小数が含まれている場合、小数点にはピリオド (.) を使用する必要があります。

注: シミュレーションでは、文字列の目標を持つカスタム方程式はサポートされていません。

定義された入力: 「定義された入力」ダイアログでは、新しい入力を定義し、目標のプロパティを設定することができます。

- 式で使用する入力がアクティブ・データ・セットに存在しない場合は、その入力を定義してから、式でその入力を使用する必要があります。

- 予測モデルを使用せずにデータのシミュレーションを行う場合は、アクティブ・データ・セットに存在しないすべてのシミュレーション対象入力を定義する必要があります。

「名前」。目標または入力の名前を指定します。

目標: 目標の尺度を指定できます。デフォルトの尺度は連続型です。この目標に対して出力を作成するかどうかを指定することもできます。例えば、一連の結合方程式で、最終方程式での目標の出力のみ必要な場合は、他の目標の出力を非表示にすることができます。

シミュレーションする入力: このオプションを選択すると、指定された確率分布に従って入力の値がシミュレーションされます (確率分布は、「シミュレーション」タブで指定します)。入力のデータに最も適合する分布を検出する (「シミュレーション」タブで「適合」または「すべて適合」をクリック) 際に考慮される分布のデフォルト・セットは、尺度によって決まります。例えば、連続型の尺度の場合は、(連続型データに適した) 正規分布が考慮されますが、2 項分布は考慮されません。

注: 文字列入力の場合は、尺度として「文字列型」を選択します。文字列入力をシミュレーションできるのは、カテゴリー分布に制限されます。

固定値入力: 入力の値が既知で、その既知の値で入力固定されることを指定します。固定入力は、数値にすることも、文字列にすることもできます。固定入力の値を指定してください。文字列値は、引用符で囲まないでください。

値ラベル: 目標、シミュレーションされた入力、固定入力に対して、値ラベルを指定することができます。値ラベルは、出力グラフと出力テーブルで使用されます。

「シミュレーション」タブ

「シミュレーション」タブでは、予測モデル以外のシミュレーションのすべてのプロパティを指定します。「シミュレーション」タブでは、次の一般的なタスクを実行することができます。

- シミュレーションされた入力の確率分布と固定入力の値を指定する。
- シミュレーションされた入力間の相関を指定する。カテゴリー入力の場合は、それらの入力のデータを生成するときに、アクティブ・データ・セット内のそれらの入力間に存在する連関を使用するように指定できます。
- 裾のサンプリングや、分布を過去のデータに適合させる基準など、拡張オプションを指定する。
- 出力をカスタマイズする。
- シミュレーション・プランの保存先を指定し、必要に応じて、シミュレーションされたデータを保存する。

シミュレーションしたフィールド

シミュレーションを実行するには、各入力フィールドを固定入力またはシミュレーションする入力として指定する必要があります。シミュレーションされた入力は値が不確定で、指定された確率分布から抽出することによって生成されます。シミュレーションする入力に履歴データを使用できる場合、そのデータに最も適合する分布が、これらの入力間の相関とともに自動的に特定されます。履歴データを使用できない、または特定の分布か相関が必要な場合には、手動で分布や相関を指定することもできます。

固定入力は、値が確定されている入力で、シミュレーションで生成された各ケースに対して変化することはありません。例えば、価格などのいくつかの入力の関数として売り上げの線型回帰モデルがあり、現在の市場価格に価格を固定したいとします。この場合、価格を固定入力として指定します。

予測モデルに基づくシミュレーションの場合は、モデル内の各予測変数がシミュレーション対象の入力フィールドとなります。予測モデルを使用しないシミュレーションの場合は、「モデル」タブで指定されたフィールドがシミュレーション対象の入力となります。

シミュレーションされる入力の分布を自動的に適合して相関を計算する: シミュレーションする入力の履歴データがアクティブ・データ・セットに含まれる場合、これらの入力間の相関を特定できるだけでなく、これらの入力のデータに最も適合する分布を自動的に見つけることができます。手順は以下のとおりです。

1. シミュレーション対象の各入力、アクティブ・データ・セットの適切なフィールドに一致していることを確認します。入力は「入力」列にリストされ、「当てはめ」列にはアクティブ・データ・セット内の一致するフィールドが表示されます。「適合」ドロップダウン・リストから異なる項目を選択することにより、アクティブ・データ・セットの異なるフィールドに入力を一致させることができます。

「適合」列の「-なし-」の値は、入力をアクティブ・データ・セットのフィールドに自動的に一致させることができなかったことを示します。デフォルトでは、入力とデータ・セット・フィールドの一致は、名前、尺度、および型 (数値型または文字列型) に基づいて行われます。アクティブ・データ・セットに入力の履歴データが含まれない場合は、下記の説明に従って、入力の分布を手動で指定するか、固定入力として入力を指定します。

2. 「すべて適合」をクリックします。

最も適合する分布とその関連パラメーターは「分布」列に表示され、履歴データのヒストグラム (棒グラフ) には分布が重なってプロットされます。シミュレーションされた入力間の相関は「相関」設定に表示されます。入力の行を選択し、「適合の詳細」をクリックすることにより、適合の結果を調べ、特定の入力の自動分布適合をカスタマイズできます。詳しくは、184 ページの『適合の詳細』のトピックを参照してください。

入力の行を選択し、「適合」をクリックすることにより、特定の入力の自動分布適合を実行できます。アクティブ・データ・セットのフィールドに一致するすべてのシミュレーションされた入力の相関の計算も自動的に行われます。

注:

- シミュレーション入力の欠損値のあるケースは、分布の適合、相関の計算、およびオプションの分割表の計算から除外されます (カテゴリ分布での入力の場合)。オプションで、カテゴリ分布の入力のユーザー欠損値を有効として処理するかどうかを指定することができます。デフォルトでは、欠損値として扱われます。詳細については、186 ページの『高度なオプション』のトピックを参照してください。
- 連続型入力と順序型入力において、検定済みのどの分布でも適切な適合度が見つからない場合、経験的分布が最適な適合度とみなされます。連続型入力の場合、経験的分布が履歴データの累積分布関数になります。順序型入力の場合、経験的分布が履歴データのカテゴリ分布になります。

手動による分布の指定: 「タイプ」ドロップダウン・リストから分布を選択し、「パラメーター」グリッドに分布パラメーターを入力することにより、シミュレーションされた入力の確率分布を手動で指定することができます。分布のパラメーターを入力すると、指定されたパラメーターに基づき、分布のサンプル・プロットが「パラメーター」グリッドの隣に表示されます。特定の分布に関するいくつかの注意事項を以下に示します。

- **カテゴリ:** カテゴリ分布は、カテゴリと呼ばれる一定数の値を持つ入力フィールドを記述します。各カテゴリには確率が関連付けられ、すべてのカテゴリの確率の合計は 1 となります。カテゴリを入力するには、「パラメーター」グリッド内の左側の列をクリックし、カテゴリ値を指定します。右側の列には、カテゴリに関連付けられる確率を入力します。

注: PMML モデルのカテゴリ入力には、モデルによって決定されたカテゴリがあるため、これを変更することはできません。

- **負の 2 項 - 失敗:** 指定した数の成功が確認される前の、一連の試行回数の中の失敗数の分布を示します。パラメーター *thresh* は指定された成功数で、パラメーター *prob* は、指定された試行回数の中の成功する確率です。
- **負の 2 項 - 試行回数:** 指定した数の成功が確認される前に必要な試行回数の分布を示します。パラメーター *thresh* は指定された成功数で、パラメーター *prob* は、指定された試行回数の中の成功する確率です。
- **範囲:** この分布は一連の区間で構成されます。各区間には確率が割り当てられ、すべての区間の確率の合計は 1 となります。特定の区間内の値は、その区間に定義された一様分布から抽出されます。区間の指定は、最小値、最大値、関連付けられた確率を入力することにより行います。

例えば、原料のコストが単位あたり \$10 から \$15 の範囲に収まる確率が 40%、\$15 から \$20 の範囲に収まる確率が 60% であるとします。この場合、「10 - 15」と「15 - 20」の 2 つの区間で構成される範囲の分布を使用してコストをモデル化し、最初の区間に関連付けられる確率を 0.4、2 番目の区間に関連付けられる確率を 0.6 に設定します。区間は連続させる必要はなく、重複しても構いません。例えば、区間として \$10 から \$15 と \$20 から \$25 や、\$10 から \$15 と \$13 から \$16 など指定することができます。

- **Weibull:** パラメーター *c* は、分布の原点の位置を指定するオプションのロケーション・パラメーターです。

ベルヌーイ分布、ベータ、2 項分布、指数分布、ガンマ、対数正規分布、負の 2 項分布 (試行回数と失敗数)、正規分布、ポワソン、一様分布のパラメーターは、「変数の計算」ダイアログ・ボックスで使用できる関連する確率変数関数と同じ意味を持っています。

固定入力の指定: 「分布」列の「タイプ」ドロップダウン・リストから「固定」を選択して固定値を入力することにより、固定入力を指定します。入力が数値か文字列かによって、値は数値または文字列となります。文字列値は、引用符で囲まないでください。

シミュレーション値の境界の指定: 多くの分布では、シミュレーション値の上限および下限を指定できます。「最小」テキスト・ボックスに値を入力して下限を指定し、「最大」テキスト・ボックスに値を入力して上限を指定することができます。

入力のロック: 列内のロック・アイコン付きチェック・ボックスを選択して入力をロックすると、その入力が自動分布適合から除外されます。これは、分布または固定値を手動で指定し、自動分布適合の影響を受けないようにする場合に最も役立ちます。また、「シミュレーションの実行」ダイアログでシミュレーション・プランを実行するユーザーとシミュレーション・プランを共有し、特定の入力に変更されないようにする場合にも、ロックが役立ちます。このため、ロックされた入力の指定を「シミュレーションの実行」ダイアログで変更することはできません。

感度分析: 感度分析を使用して、シミュレーションされたケースの独立したセット (実質的には別のシミュレーション) を指定された値ごとに生成することにより、固定入力における系統的な変化の影響や、シミュレーションされた入力のパラメーター分布における系統的な変化の影響を調べることができます。感度分析を指定するには、固定入力またはシミュレーションされた入力を選択して「感度分析」をクリックします。感度分析は、単一の固定入力、またはシミュレーションされた入力の単一のパラメーター分布だけに制限されます。詳しくは、185 ページの『感度分析』のトピックを参照してください。

適合の状況を示すアイコン

「適合」列のアイコンは、各入力フィールドの適合の状況を示します。

表 3. 状況アイコン:

アイコン	説明
	入力に分布は指定されておらず、入力は固定入力として指定されていません。シミュレーションを実行するには、この入力の分布を指定するか、入力を固定入力となるように定義して固定値を指定する必要があります。
	入力はアクティブ・データ・セットに存在しないフィールドに適合していました。アクティブ・データ・セットに入力の分布を再適合させる場合を除き、操作は必要ありません。
	最も適合する分布は、「適合の詳細」ダイアログのその他の分布で置き換えられています。
	入力は最も適合する分布に設定されています。
	この入力には、分布が手動で指定されているか、感度分析の反復が指定されています。

適合の詳細: 「適合の詳細」ダイアログには、特定の入力に関する自動分布適合の結果が表示されます。分布は適合度順に配置され、最も適合した分布が最初にリストされます。「使用」列の該当する分布のラジオ・ボタンを選択することにより、最も適合する分布を上書きすることができます。「使用」列のラジオ・ボタンを選択すると、その入力の過去のデータのヒストグラム (または棒グラフ) に重ね合わせた分布の散布図も表示されます。

適合度統計量: デフォルトの場合と連続型フィールドの場合、Anderson-Darling 検定を使用して適合度が決定されます。また、連続型フィールドの場合のみ、適合度に対して Kolmogorov-Smirnoff 検定を指定することができます。これを指定するには、「高度なオプション」設定で該当する項目を選択します。連続型入力の場合、両方の検定の結果が「適合度統計量」列 (Anderson-Darling 法の場合は A 列、Kolmogorov-Smirnoff 法の場合は K 列) に表示され、選択された検定を使用して分布が並べ替えられます。順序型入力と名義型入力の場合は、カイ 2 乗検定が使用されます。検定に関連する p 値も表示されます。

パラメーター: 適合する各分布に関連する分布パラメーターが「パラメーター」列に表示されます。ベルヌーイ分布、ベータ、2 項分布、指数分布、ガンマ、対数正規分布、負の 2 項分布 (試行回数と失敗数)、正規分布、ポワソン、一様分布のパラメーターは、「変数の計算」ダイアログ・ボックスで使用できる関連する確率変数関数と同じ意味を持っています。詳しくは、のトピックを参照してください。カテゴリ型分布の場合、カテゴリがパラメーター名になり、関連する確率がパラメーター値になります。

カスタマイズした分布設定で再適合: デフォルトでは、入力の尺度を使用して、自動分布適合で考慮される分布の設定が決定されます。例えば、連続型入力の適合時には、対数正規やガンマなどの連続型分布が考慮されますが、ポワソンや二項分布などの離散型分布は考慮されません。「再適合」列で分布を選択することにより、デフォルトの分布のサブセットを選択することができます。「処理 (尺度)」ドロップダウン・リストから異なる尺度を選択し、「再適合」列で分布を選択することにより、分布のデフォルト設定を上書きすることもできます。「再適合を実行」をクリックして、ユーザー指定の分布設定を使用して再適合を行います。

注:

- シミュレーション入力の欠損値のあるケースは、分布の適合、相関の計算、およびオプションの分割表の計算から除外されます (カテゴリ分布での入力の場合)。オプションで、カテゴリ分布の入力のユーザ欠損値を有効として処理するかどうかを指定することができます。デフォルトでは、欠損値として扱われます。詳細については、186 ページの『高度なオプション』のトピックを参照してください。
- 連続型入力と順序型入力において、検定済みのどの分布でも適切な適合度が見つからない場合、経験的分布が最適な適合度とみなされます。連続型入力の場合、経験的分布が履歴データの累積分布関数になります。順序型入力の場合は、経験的分布が履歴データのカテゴリ分布になります。

感度分析: 感度分析では、固定入力を変えた場合、またはシミュレーションされた入力の分布パラメータを変えた場合に、指定された一連の値にどのような効果があるのかを調べることができます。指定した値ごとに、シミュレーションされたケースの独立したセット (実質的に、個別のシミュレーション) が生成されるため、入力の変化による効果を調べることができます。シミュレーションされたケースの各セットは、**反復**と呼ばれます。

反復: このオプションを使用して、入力を変化させる値のセットを指定することができます。

- 分布パラメータの値を変える場合は、ドロップダウン・リストから該当するパラメータを選択します。反復グリッドごとに、パラメータ値のセットを入力してください。「**続行**」をクリックすると、値の反復数を指定するインデックスとともに、指定された値が関連する入力の「パラメータ」グリッドに追加されます。
- カテゴリ型分布と範囲型分布の場合、カテゴリと区間の確率はそれぞれ変化させることができますが、カテゴリの値と区間のエンドポイントを変化させることはできません。カテゴリまたは区間をドロップダウン・リストから選択し、反復グリッドごとに「パラメータ値」に確率のセットを指定してください。その他のカテゴリまたは区間の確率は、それぞれ自動的に調整されます。

反復なし: 入力の反復をキャンセルするには、このオプションを使用します。「**続行**」をクリックすると、反復が削除されます。

相関

シミュレーション対象の入力フィールドは、多くの場合、身長と体重などのように、相関することがわかっています。シミュレーション値でこれらの相関を保持するには、シミュレーションする入力間の相関を示す必要があります。

適合時に相関を再計算する: 「シミュレーションされるフィールド」設定の「**すべて適合**」操作または「**適合**」操作によって分布をアクティブ・データ・セットに適合させる際に、シミュレーションされる入力間の相関を自動的に計算する場合は、このオプションを選択します。

適合時に相関を再計算しない: 手で相関を指定し、分布をアクティブ・データ・セットに自動的に適合させる際にこの相関が上書きされないようにするには、このオプションを選択します。相関グリッドに入力する値は、-1 から 1 の範囲でなければなりません。値として 0 を入力すると、関連する入力のペア間に相関がないことを示します。

リセット: すべての相関を 0 にリセットします。

カテゴリ分布の入力に適合多元分割表を使用: カテゴリ分布の入力について、それらの入力間の連関を記載した多元分割表を、アクティブ・データ・セットから自動的に計算できます。この分割表は、その後、それらの入力のデータが生成されるときに使用されます。シミュレーション・プランを保存することを選択した場合、分割表は計画ファイル内に保存され、計画の実行時に使用されます。

- **アクティブ データセットから分割表を計算:** 分割表を含む既存のシミュレーション・プランを操作する場合は、アクティブ・データ・セットから分割表を再計算できます。このアクションにより、読み込まれた計画ファイルの分割表がオーバーライドされます。
- **読み込まれたシミュレーション計画の分割表を使用:** デフォルトでは、分割表を含むシミュレーション・プランを読み込むと、そのプランの表が使用されます。「**アクティブ データセットから分割表を計算**」を選択すると、アクティブ・データ・セットから分割表を再計算できます。

高度なオプション

ケースの最大数: 生成するシミュレーション・データの最大ケース数と、関連する目標値を指定します。感度分析を指定した場合、この値が各反復の最大ケース数になります。

停止基準の目標: 予測モデルに複数の目標が含まれている場合、停止基準の適用先となる目標を選択することができます。

停止基準: 以下の選択項目は、許可されている最大数のケースが生成される前に、シミュレーションを停止するための基準を指定します。

- **最大数に達成するまで続行する:** ケースの最大数に達するまで、シミュレーションされたケースが生成されます。
- **裾がサンプリングされた場合に停止する:** 指定された目標分布のいずれかの裾を適切にサンプリングするには、このオプションを使用します。指定された裾のサンプリングが完了するかケースの最大数に達するまで、シミュレーションされたケースが生成されます。予測モデルに複数の目標が含まれている場合、「**停止基準の目標**」ドロップダウン・リストから、この基準の適用先となる目標を選択します。

タイプ: 10,000,000 などの目標値や 99 パーセンタイルなどのパーセンタイルを指定することにより、裾領域の境界を定義することができます。「**タイプ**」ドロップダウン・リストで「**値**」を選択した場合は、「**値**」テキスト・ボックスに境界の値を入力し、「**端**」ドロップダウン・リストを使用して、これが左右どちらの裾領域の境界であるかを指定します。「**タイプ**」ドロップダウン・リストで「**パーセンタイル**」を選択した場合は、「**パーセンタイル**」テキスト・ボックスに値を入力します。

度数: 裾が適切にサンプリングされるように、裾領域内の目標値の数を指定します。この数に達すると、ケースの生成が停止します。

- **平均の信頼区間が指定されたしきい値内にある場合に停止する:** 特定の目標の平均を指定された精度で認識する場合は、このオプションを使用します。指定された精度に達するかケースの最大数に達するまで、シミュレーションされたケースが生成されます。このオプションを使用するには、信頼度レベルとしきい値を指定します。指定したレベルが関連付けられた信頼区間がしきい値の範囲内になるまで、シミュレートされたケースが生成されます。例えば、このオプションを使用して、95% の信頼度レベルでの平均の信頼区間が平均値の 5% 以内になるまでケースを生成するよう指定することができます。予測モデルに複数の目標が含まれている場合、「**停止基準の目標**」ドロップダウン・リストから、この基準の適用先となる目標を選択します。

しきい値のタイプ: しきい値は、数値として指定することも、平均のパーセントとして指定することもできます。「**しきい値のタイプ**」ドロップダウン・リストで「**値**」を選択した場合は、「**値のしきい値**」テキスト・ボックスにしきい値を入力します。「**しきい値のタイプ**」ドロップダウン・リストで「**パーセント**」を選択した場合は、「**パーセントのしきい値**」テキスト・ボックスに値を入力します。

サンプリングするケースの数: シミュレーションされた入力の分布を自動的にアクティブ・データ・セットへ適合するとき使用するケースの数を指定します。データ・セットが非常に大きい場合は、分布の適合に使用されるケースの数を制限することをお勧めします。「**ケース数を N に制限**」を選択すると、最初の N 件のケースだけが使用されます。

適合度の基準 (連続型): 連続型入力の場合、適合度の Anderson-Darling 検定または Kolmogorov-Smirnoff 検定を使用して、シミュレーションされた入力の分布をアクティブ・データ・セットに適合させる際に、分布の順位付けを行うことができます。Anderson-Darling 検定がデフォルトで選択されます。裾領域での最適な適合度を取得する場合は、特にこの検定をお勧めします。

経験的分布: 連続型入力の場合、経験的分布が履歴データの累積分布関数になります。連続入力の経験的分布の計算に使用されるビンの数を指定することができます。デフォルトは 100 で、最大値は 1000 です。

結果の再現: ランダム・シードを設定すると、シミュレーションを再現することができます。整数を指定するか、「生成」をクリックします。「生成」をクリックすると、1 から 2147483647 までの整数の疑似乱数が作成されます。デフォルトは 629111597 です。

注: 特定のランダム シードについては、スレッド数を変更しない限り、結果が再現されます。特定のコンピュータでは、SET THREADS コマンド構文を実行することでスレッド数を変更しない限り、数は一定です。異なるコンピュータでシミュレーションを実行すると、スレッド数が変わる場合があります。各コンピュータでのスレッド数の決定には内部アルゴリズムが使用されるためです。

カテゴリ分布の入力のユーザー欠損値: これらのコントロールは、カテゴリ分布における入力のユーザー欠損値が有効な値として処理されるかどうかを指定します。他のすべての型の入力では、システム欠損値とユーザー欠損値が常に無効な値として処理されます。分布適合、相関の計算、およびオプションである分割表の計算の対象にケースを含めるには、すべての入力に有効な値が指定されている必要があります。

密度関数

以下の設定を使用して、連続型目標の確率密度関数と累積分布関数の出力や、カテゴリ型目標の予測値の棒グラフをカスタマイズすることができます。

確率密度関数 (PDF): 確率密度関数は、目標値の分布を表示します。連続型目標の場合、この関数を使用して、目標が特定の領域内に含まれる確率を判断することができます。カテゴリ型目標 (尺度が名義型または順序型の目標) の場合は、目標の各カテゴリに該当するケースのパーセントを表示する棒グラフが生成されます。PMML モデルのカテゴリ型目標では、カテゴリ値の追加オプションを使用して、以下に示す設定を報告することができます。

2 ステップのクラスター・モデルと大規模ファイルのクラスター・モデルの場合、所属クラスターの棒グラフが生成されます。

累積分布関数 (CDF): 累積分布関数は、目標の値が指定の値以下になる確率を表示します。この関数は、連続型目標でのみ使用することができます。

スライダーの位置: PDF グラフおよび CDF グラフ上の移動可能な基準線の初期位置を指定できます。下限の線および上限の線に指定する値は、パーセンタイルではなく水平軸上の位置を基準とします。下限の線を除去するには「負の無限大」を、上限の線を除去するには「無限」を選択します。デフォルトでは、これらの線はそれぞれ 5 番目と 95 番目のパーセンタイルに配置されます。複数の分布関数が単一のグラフ上に表示されている場合 (感度分析が繰り返されたことにより、複数の目標や結果が生成されたため)、デフォルトは最初の反復または最初の目標の分布を基準とします。

基準線 (連続型): 連続型目標の確率密度関数と累積分布関数に垂直方向のさまざまな基準線を追加するように要求することができます。

- **シグマ:** 対象の平均に基づき、標準偏差から指定数値分プラスおよびマイナスの位置に基準線を追加することができます。

- **パーセンタイル:** 「下部」テキスト・ボックスと「上部」テキスト・ボックスに値を入力することにより、対象の分布の 1 つまたは 2 つのパーセンタイル値の位置に基準線を追加することができます。例えば、「上部」テキスト・ボックスに「95」を入力すると、95 パーセンタイルという意味になり、観測の 95% が該当する値となります。同様に、「下部」テキスト・ボックスに「5」を入力すると、5 パーセンタイルという意味になり、観測の 5% が該当する値となります。
- **ユーザー指定の基準線:** 指定した目標値で基準線を追加することができます。

注: 複数の分布関数が単一のグラフ上に表示されている場合 (感度分析が繰り返されたことにより、複数の目標や結果が生成されたため)、基準線は最初の反復または最初の目標の分布にのみ適用されます。他の分布に基準線を追加するには、「グラフ」オプション・ダイアログを使用します。このダイアログには、PDF グラフまたは CDF グラフからアクセスします。

個別の連続型目標から結果をオーバーレイする: 複数の連続型目標が存在する場合、そのすべての目標の分布関数を単一のグラフに表示するかどうかを指定します (一方のグラフで確率密度関数を、もう一方のグラフで累積分布関数を表示)。このオプションを選択しなかった場合、各目標の結果は個別のグラフに表示されます。

レポートするカテゴリ値: カテゴリ型目標を持つ PMML モデルの場合、モデルの結果は、カテゴリごとの予測確率のセットになります。各カテゴリに、目標値が含まれます。最も高い確率を持つカテゴリが予測カテゴリとなり、このカテゴリを使用して、上記の「確率密度関数」設定で説明した棒グラフが生成されます。「予測カテゴリ」を選択すると、この棒グラフが生成されます。「予測確率」を選択すると、目標のカテゴリごとに予測確率の分布を示すヒストグラムが生成されます。

感度分析のグループ化: 感度分析を含むシミュレーションでは、分析によって定義された反復ごとに、予測目標値の独立したセットが生成されます (変化する入力値ごとに 1 つの反復)。反復が存在する場合、カテゴリ目標の予測カテゴリの棒グラフは、すべての反復の結果が含まれるクラスター棒グラフとして表示されます。カテゴリをグループ化することも、反復をグループ化することもできます。

出力

トルネード図: トルネード図は、さまざまな指標を使用して、目標とシミュレーションされた入力との間の関係を表示する棒グラフです。

- **目標と入力の相関:** このオプションを選択すると、指定された目標とそのシミュレーションされた各入力との間の相関係数のトルネード図が作成されます。この種類のトルネード図は、名義型または順序型の尺度を持つ目標値も、カテゴリ分布のシミュレーションされた入力もサポートしていません。
- **分散への貢献度:** このオプションを選択すると、シミュレーションされた各入力からの目標の分散への貢献度を示すトルネード図が作成されます。このトルネード図により、各入力がある目標の全体的な不確実性にどの程度貢献しているかを評価することができます。この種類のトルネード図は、順序型または名義型の尺度を持つ目標値や、カテゴリ型分布、ベルヌーイ分布、2 項分布、ポアソン分布、負の 2 項分布のシミュレーションされた入力をサポートしていません。
- **変更する目標の感度:** このオプションを選択すると、入力に関連付けられている分布の標準偏差の指定数を加算または減算することにより、シミュレーションされた各入力の変調の目標に対する影響を表示するトルネード図が作成されます。この種類のトルネード図は、順序型または名義型の尺度を持つ目標値や、カテゴリ型分布、ベルヌーイ分布、2 項分布、ポアソン分布、負の 2 項分布のシミュレーションされた入力をサポートしていません。

目標分布の箱ひげ図: 箱ひげ図は、連続型目標の場合のみ使用することができます。予測モデルに複数の連続型目標があり、1 つのチャート上にすべての目標に関する箱ひげ図を表示する場合は、「個別目標の結果をオーバーレイ」を選択します。

目標対入力の散布図: 目標をシミュレーションされた入力と比較する散布図は、連続型目標とカテゴリ型目標の両方で使用することができます。この図には、連続型入力とカテゴリ型入力の両方を持つ目標の分布が表示されます。カテゴリ対象またはカテゴリ型入力を含む散布図は、ヒート・マップとして表示されます。

パーセンタイル値のテーブルを作成: 連続型目標の場合、目標の分布の指定されたパーセンタイルの表を作成することができます。4 分位 (25、50、および 75 パーセンタイル) の場合、同じサイズの 4 つのグループに観測値が分割されます。必要な等サイズ・グループの数が 4 以外の場合は、「区間」を選択してその数を指定します。個別のパーセンタイル (99 パーセンタイルなど) を指定するには、「ユーザー指定のパーセンタイル」を選択します。

目標分布の記述統計量: このオプションを選択すると、連続型目標、カテゴリ型目標、および連続型入力に対して、記述統計の表が作成されます。連続型目標の場合、この表には、平均値、標準偏差、中央値、最小値、最大値、指定されたレベルでの平均値の信頼区間、目標分布の 5 パーセンタイルと 95 パーセンタイルが表示されます。カテゴリ型目標の場合、この表には、目標の各カテゴリに該当するケースの割合 (パーセント) が表示されます。PMML モデルのカテゴリ型目標の場合、この表には、目標の各カテゴリの平均確率が表示されます。連続型入力の場合、この表には、平均値、標準偏差、最小値、最大値が表示されます。

入力の相関および分割表: このオプションでは、シミュレーションした入力間の相関係数の表が表示されます。カテゴリ分布の入力が分割表から生成される場合、それらの入力用に生成されるデータの分割表も表示されます。

出力に含めるシミュレーションされた入力: デフォルトでは、シミュレーションされたすべての入力が出力に含まれます。シミュレーションされた入力を選択して、出力から除外することができます。これにより、トルネード図、散布図、テーブル形式の出力から、シミュレーションされた入力が除外されます。

連続型対象の制限範囲: 1 つ以上の連続型対象について、有効値の範囲を指定できます。指定された範囲外の値は、対象に関連するすべての出力および分析から除外されます。下限を設定するには、「限度」列で「下限」を選択し、「最小」列に値を入力します。上限を設定するには、「限度」列で「上限」を選択し、「最大」列に値を入力します。下限と上限の両方を設定するには、「限度」列で「両方」を選択し、「最小」列と「最大」列に値を入力します。

表示形式: 目標と入力 (固定入力とシミュレーションされた入力の両方) の値を表示する場合の形式を設定することができます。

保存

このシミュレーションのプランを保存: シミュレーションの現在の仕様をシミュレーション・プラン・ファイルに保存することができます。シミュレーション・プラン・ファイルの拡張子は `.splan` です。シミュレーション・ビルダーでプランを再度開き、必要に応じて変更を加えて、シミュレーションを実行することができます。シミュレーション・プランは、他のユーザーと共有することができます。プランを共有しているユーザーは、「シミュレーションの実行」ダイアログでそのシミュレーションを実行することができます。シミュレーション・プランには、密度関数の設定、グラフとテーブルの出力設定、適合、経験的分布、ランダム・シードの高度なオプション設定を除くすべての仕様が含まれています。

シミュレーション・データを新規データ・ファイルとして保存: シミュレーションされた入力、固定入力、予測目標値を、SPSS Statistics のデータ・ファイル、現在のセッションの新規データ・セット、Excel ファイルに保存することができます。データ・ファイルの各ケース (または各行) は、予測目標値と、目標値を生成するシミュレーションされた入力および固定入力から構成されます。感度分析を指定した場合、反復回数がラベル付けされたケースの連続セットが、反復のたびに発生します。

「シミュレーションの実行」ダイアログ

「シミュレーションの実行」ダイアログは、シミュレーション・プランを持っていて、主に既存のシミュレーションを実行するユーザーを対象に設計されています。このインターフェースには、さまざまな条件下でシミュレーションを実行するために必要な機能も用意されています。これを使用して、次の一般的なタスクを実行することができます。

- 感度分析をセットアップまたは変更して、固定入力の変更に伴う影響や、シミュレーションされた入力の分布パラメーターの変更に伴う影響を調べる。
- 不確実な入力の確率分布（およびこれらの入力間の相関）を新しいデータに再適合させる。
- シミュレーションされた入力の分布を変更する。
- 出力をカスタマイズする。
- シミュレーションを実行する。

「シミュレーション」タブ

「シミュレーション」タブでは、感度分析を指定し、シミュレーションされた入力の確率分布とシミュレーションされた入力間の相関を新しいデータに再適合させ、シミュレーションされた入力に関連する確率分布を変更することができます。

「シミュレーションした入力」グリッドには、シミュレーション・プランで定義された各入力フィールドのエントリーが含まれます。各エントリーには、入力の名前とその入力に関連する確率分布のタイプが、関連する分布曲線のサンプル・プロットとともに表示されます。各入力には、状況アイコン（チェック・マークが付いた色付きの円）も表示されます。このアイコンは、分布を新しいデータに再適合する際に便利です。さらに、入力にはロック・アイコンが表示される場合もあります。このアイコンは、入力がロックされているため、「シミュレーションの実行」ダイアログで入力を変更したり、新しいデータへの再適合を行ったりできないことを示します。ロックされた入力を変更するには、シミュレーション・ビルダーでシミュレーション・プランを開く必要があります。

各入力は、シミュレーションされた入力または固定入力のいずれかです。シミュレーションされた入力は値が不確定で、指定された確率分布から抽出することによって生成されます。固定入力は、値が確定されている入力であり、シミュレーションで生成された各ケースに対して変化することはありません。特定の入力を処理するには、シミュレーションされた入力グリッドで入力のエントリーを選択します。

感度分析を指定するには

感度分析を使用して、シミュレーションされたケースの独立したセット（実質的には別のシミュレーション）を指定された値ごとに生成することにより、固定入力における系統的变化の影響や、シミュレーションされた入力のパラメーター分布における系統的变化の影響を調べることができます。感度分析を指定するには、固定入力またはシミュレーションされた入力を選択して「感度分析」をクリックします。感度分析は、単一の固定入力、またはシミュレーションされた入力の単一のパラメーター分布だけに制限されます。詳しくは、185 ページの『感度分析』のトピックを参照してください。

分布を新しいデータに再適合させるには

シミュレーションされた入力の確率分布（およびシミュレーションされた入力間の相関）をアクティブ・データ・セットのデータに自動的に再適合させるには、以下の手順を実行します。

1. シミュレーションする各モデル入力が、アクティブ・データ・セット内の適切なフィールドに一致していることを確認します。シミュレーションされた各入力は、その入力に関連する「フィールド」ドロ

プダウン・リストで指定されたアクティブ・データ・セットのフィールドに適合されます。以下のような、疑問符付きのチェック・マークを示す状況アイコンが表示されている入力を探すことにより、一致しない入力を簡単に特定することができます。



2. 必要なフィールドの対応を変更するには、「データセットのフィールドに当てはめ」をクリックして、リストから該当するフィールドを選択します。
3. 「すべて適合」をクリックします。

適合した各入力について、最もデータに適合する分布が表示され、過去のデータのヒストグラム (または棒グラフ) に重ね合わせてその分布がプロットされます。適切な適合が見つからない場合は、経験的分布が使用されます。経験的分布に適合する入力の場合、経験的分布は実際にはヒストグラムによって表されるため、過去のデータのヒストグラムだけが表示されます。

注: 状況アイコンの完全なリストについては、181 ページの『シミュレーションしたフィールド』のトピックを参照してください。

確率分布を変更するには

シミュレーションされた入力の確率分布を変更し、必要に応じて、シミュレーションされた入力を固定入力に変更することができます (固定入力をシミュレーションされた入力に変更することもできます)。

1. 該当する入力を選択し、「分布を手動で設定」を選択します。
2. 該当する分布の種類を選択し、分布パラメーターを指定します。シミュレーションされた入力を固定入力に変更するには、「種類」ドロップダウン・リストで「固定」を選択します。

分布のパラメーターを入力すると、(入力のエントリーに表示された) 分布のサンプル・プロットが更新され、変更内容が反映されます。確率分布を手動で指定する方法については、181 ページの『シミュレーションしたフィールド』のトピックを参照してください。

適合時にカテゴリー入力のユーザー欠損値を含める: これは、アクティブ・データ・セット内のデータへの再適合を行うときに、カテゴリー分布における入力のユーザー欠損値が有効な値として処理されるかどうかを指定します。他のすべての型の入力では、システム欠損値とユーザー欠損値が常に無効な値として処理されます。分布適合および相関の計算の対象にケースを含めるには、すべての入力に有効な値が指定されている必要があります。

「出力」タブ

「出力」タブを使用して、シミュレーションによって生成された出力をカスタマイズすることができます。

密度関数: 密度関数は、シミュレーションによって生成された一連の結果を調べるための主要な手段です。

- **確率密度関数:** 確率密度関数は、目標値の分布を表示します。この表示により、目標が特定の領域内に含まれる確率を判断することができます。固定された結果のセット (「品質の低いサービス」、「品質が中程度のサービス」、「品質が高いサービス」など) を持つ目標の場合、目標の各カテゴリーに該当するケースのパーセントを表示する棒グラフが生成されます。
- **累積分布関数:** 累積分布関数は、目標の値が指定の値以下になる確率を表示します。

トルネード図: トルネード図は、さまざまな指標を使用して、目標とシミュレーションされた入力との間の関係を表示する棒グラフです。

- **目標と入力の間:** このオプションを選択すると、指定された目標とそのシミュレーションされた各入力との間の相関係数のトルネード図が作成されます。
- **分散への貢献度:** このオプションを選択すると、シミュレーションされた各入力からの目標の分散への貢献度を示すトルネード図が作成されます。このトルネード図により、各入力がある目標の全体的な不確実性にどの程度貢献しているかを評価することができます。
- **変更する目標の感度:** このオプションを選択すると、入力に関連する分布の標準偏差に 1 を加算または減算することにより、シミュレーションされた各入力の変調の目標に対する効果を表示するトルネード図が作成されます。

目標対入力の散布図: このオプションを選択すると、シミュレーションされた入力と目標値を比較する散布図が生成されます。

目標分布の箱ひげ図: このオプションを選択すると、目標分布の箱ひげ図が生成されます。

4 分位表: このオプションを選択すると、目標分布の 4 分位表が生成されます。分布の 4 分位は、25、50、75 のパーセンタイルです。この場合、同じサイズの 4 つのグループに観測値が分割されます。

入力の相関および分割表: このオプションでは、シミュレーションした入力間の相関係数の表が表示されます。カテゴリ分布の入力間の連関の分割表は、カテゴリ・データを分割表から生成することがシミュレーション・プランで指定されている場合に表示されます。

個別目標の結果をオーバーレイ: シミュレーション対象の予測モデルに複数のターゲットが含まれている場合は、個別の目標の結果を 1 つのグラフに表示するかどうかを指定することができます。この設定は、確率密度関数のグラフ、累積分布関数のグラフ、箱ひげ図に適用されます。例えば、このオプションを選択すると、すべての目標の確率密度関数が 1 つのグラフに表示されます。

このシミュレーションのプランを保存: シミュレーション・プラン・ファイルにシミュレーションの変更内容を保存することができます。シミュレーション・プラン・ファイルの拡張子は *.splan* です。「シミュレーションの実行」ダイアログまたはシミュレーション・ビルダーでプランを再度開くことができます。シミュレーション・プランには、出力の設定を除くすべての仕様が含まれます。

シミュレーション・データを新規データ・ファイルとして保存: シミュレーションされた入力、固定入力、予測目標値を、SPSS Statistics のデータ・ファイル、現在のセッションの新規データ・セット、Excel ファイルに保存することができます。データ・ファイルの各ケース (または各行) は、予測目標値と、目標値を生成するシミュレーションされた入力および固定入力から構成されます。感度分析を指定した場合、反復回数がラベル付けされたケースの連続セットが、反復のたびに発生します。

ここでは使用できない機能を使用して出力をさらにカスタマイズしたい場合は、シミュレーション・ビルダーからシミュレーションを実行することをお勧めします。詳しくは、178 ページの『シミュレーション・プランからシミュレーションを実行するには』のトピックを参照してください。

シミュレーションからのグラフ出力の作業

シミュレーションから生成された各種グラフには、表示をカスタマイズするためのインタラクティブ機能が組み込まれています。インタラクティブ機能を使用するには、出力ビューアーでグラフ・オブジェクトをアクティブ化 (ダブルクリック) します。すべてのシミュレーション・グラフが、グラフボードでの視覚化に対応しています。

連続型目標の確率密度関数グラフ: このグラフには 2 つのスライドする縦の基準線があり、これらの基準線によってグラフが個別の領域に分割されます。グラフの下のテーブルには、目標が各領域内に含まれる確

率が表示されます。複数の密度関数が同じグラフに表示される場合、このテーブルには各密度関数に関連する確率ごとに行が割り当てられます。各基準線には、線を簡単に移動するためのスライダー（逆三角形）があります。グラフで「**グラフ**」オプションボタンをクリックすると、各種の追加機能が使用可能になります。具体的には、スライダーの位置を明示的に設定したり、基準線を追加したり、グラフの表示を連続曲線とヒストグラムの間で切り替えたりすることができます。詳しくは、『**グラフ・オプション**』のトピックを参照してください。

連続型目標の累積分布関数グラフ: このグラフにも、上記の確率密度関数で説明した 2 つの移動可能な縦の基準線と関連するテーブルがあります。また、「**グラフ**」オプション・ダイアログにアクセスすることもできます。このダイアログで、スライダーの位置を明示的に設定したり、固定基準線を追加したり、累積分布関数を増加関数（デフォルト）と減少関数のどちらで表示するかを指定したりすることができます。詳しくは、『**グラフ・オプション**』のトピックを参照してください。

感度分析の反復があるカテゴリー型目標の棒グラフ: 感度分析の反復があるカテゴリー型目標の場合、予測目標カテゴリーの結果は、すべての反復の結果が含まれるクラスター化された棒グラフとして表示されます。このグラフには、ドロップダウン・リストがあります。このドロップダウン・リストを使用して、カテゴリーと反復のどちらでクラスター化するかを選択することができます。2 ステップのクラスター・モデルと大規模ファイルのクラスター・モデルの場合、クラスター数と反復のいずれかでクラスター化することができます。

感度分析の反復がある複数の目標の箱ひげ図: 複数の連続型目標と感度分析の反復がある予測モデルの場合、すべての目標の箱ひげ図を 1 つのグラフで表示することを選択すると、クラスター化された箱ひげ図が作成されます。このグラフには、ドロップダウン・リストがあります。このドロップダウン・リストを使用して、目標と反復のどちらでクラスター化するかを選択することができます。

グラフ・オプション

「**グラフ**」オプション・ダイアログでは、シミュレーションで生成された確率密度関数と累積分布関数のアクティブなグラフの表示をカスタマイズすることができます。

表示: 「**表示**」ドロップダウン・リストは、確率密度関数のグラフにのみ適用されます。これを使用して、グラフの表示を連続曲線からヒストグラムに切り替えることができます。複数の密度関数を同じグラフで表示する場合、この機能を使用することはできません。その場合、密度関数は連続曲線としてのみ表示することができます。

順序: 「**順序**」ドロップダウン・リストは、累積分布関数のグラフにのみ適用されます。このオプションは、累積密度関数を増加関数（デフォルト）と減少関数のどちらで表示するかを指定します。減少関数で表示する場合、横軸の特定のポイントにある関数の値は、そのポイントの右側に目標がある確率を示します。

スライダーの位置: 「**上限**」と「**下限**」テキスト・ボックスに値を入力することにより、スライドする基準線の位置を明示的に設定することができます。左側の線を削除するには、「**-無限大**」を選択します。この操作により、線の位置が実質的に負の無限大に設定されます。右側の線を削除するには、「**無限大**」を選択して、線の位置を実質的に無限大に設定します。

基準線: 確率密度関数と累積分布関数に、垂直方向のさまざまな固定基準線を追加することができます。複数の関数が単一のグラフ上に表示されている場合（感度分析が繰り返されたことにより、複数の目標や結果が生成されたため）、基準線を適用する特定の関数を指定できます。

- **シグマ:** 対象の平均に基づき、標準偏差から指定数値分プラスおよびマイナスの位置に基準線を追加することができます。

- **パーセンタイル:** 「下部」テキスト・ボックスと「上部」テキスト・ボックスに値を入力することにより、対象の分布の 1 つまたは 2 つのパーセンタイル値の位置に基準線を追加することができます。例えば、「上部」テキスト・ボックスに「95」を入力すると、95 パーセンタイルという意味になり、観測の 95% が該当する値となります。同様に、「下部」テキスト・ボックスに「5」を入力すると、5 パーセンタイルという意味になり、観測の 5% が該当する値となります。
- **ユーザー指定の位置:** 横軸に沿って、指定した位置に基準線を追加することができます。

基準線にラベルを付ける (Label reference lines): このオプションは、選択した基準線にラベルを適用するかどうかを制御します。

基準線を削除するには、「グラフ」オプション・ダイアログで関連する選択項目をクリアし、「**続行**」をクリックします。

第 35 章 地理空間モデリング

地理空間モデリング手法は、地理空間 (マップ) 成分を含む、データ内のパターンを検出するために設計されています。地理空間モデリング ウィザードは、時間成分が含まれている地理空間データと時間成分が含まれていない地理空間データを分析する方法を提供します。

イベントおよび地理空間データ (地理空間アソシエーション ルール) に基づいてアソシエーションを検索
地理空間アソシエーション ルールを使用して、空間プロパティと非空間プロパティの両方に基づいて、データ内のパターンを検出できます。例えば、場所と人口統計の属性によって犯罪データ内のパターンを特定できます。検出したパターンから、特定のタイプの犯罪が発生する可能性が高い場所を予測するルールを作成できます。

時系列と地理空間データ (空間的時系列的予測) を使用して予測を作成 (Make predictions using time series and geospatial data (Spatio-Temporal Prediction))

空間的時系列的予測では、位置データ、予測の入力フィールド (予測変数)、1 つ以上の時間フィールド、および対象フィールドが含まれるデータを使用します。各位置のデータには、各時間間隔における各予測変数の値と対象の値を表す、多数の行が含まれています。

地理空間モデリング ウィザードの使用

1. メニューから次の項目を選択します。

「分析」 > 「空間および時間モデリング」 > 「空間モデリング」

2. ウィザードの手順に従います。

例

ヘルプ システムに詳細な例があります。

- 地理空間アソシエーション ルール: 「ヘルプ」 > 「トピック」 > 「ケース スタディ」 > 「Statistics Base」 > 「空間的アソシエーション ルール」
- 空間的時系列的予測: 「ヘルプ」 > 「トピック」 > 「ケース スタディ」 > 「Statistics Base」 > 「空間的時系列的予測」

マップの選択

地理空間モデリングでは、1 つ以上のマップ データ ソースを使用できます。マップ データ ソースには、地理的領域および他の地理的フィーチャ (道路や川など) を定義する情報が含まれます。多くのマップ ソースには、人口統計や他の記述データおよびイベント データ (犯罪レポートや失業率) も含まれます。以前に定義したマップ仕様ファイルを使用するか、ここでマップ仕様を定義して、今後を使用するためにその仕様を保存できます。

マップ仕様のロード(L)

以前に定義したマップ仕様 (.mplan) ファイルをロードします。ここで定義したマップ データ ソースは、マップ仕様ファイルに保存できます。空間的時系列的予測では、複数のマップが指定されたマップ仕様ファイルを選択した場合、ファイルの中の 1 つのマップを選択するように求められます。

マップ ファイルの追加(A)

ESRI シェープ (.shp) ファイルまたは ESRI シェープファイルが含まれている .zip アーカイブを追加します。

- .shp ファイルと同じ場所に、対応する .dbf ファイルがなければなりません。また、そのファイルのルート名は、.shp ファイルのものと同じでなければなりません。
- ファイルが .zip アーカイブである場合、.shp および .dbf ファイルのルート名は、.zip アーカイブと同じものでなければなりません。
- 対応する投影法 (.prj) ファイルがない場合は、投影システムを選択するように求められます。

関係 地理空間アソシエーション ルールの場合、この列で、イベントとマップ内のフィーチャとの関係を定義します。この設定は、空間的・時間的予測では使用できません。

上へ移動、下へ移動

マップ要素のレイヤーの順序は、リスト内の順序によって決定されます。リスト内の最初のマップが一番下のレイヤーになります。

マップの選択

空間的・時間的予測では、複数のマップが指定されたマップ仕様ファイルを選択した場合、ファイルの中の 1 つのマップを選択するように求められます。空間的・時間的予測では、複数のマップはサポートされません。

地理空間関係

地理空間アソシエーション ルールの場合、「地理空間関係」ダイアログで、イベントとマップ内のフィーチャとの関係を定義します。

- この設定は、地理空間アソシエーション ルールにのみ適用されます。
- この設定による影響を受けるのは、データ ソースを選択するためのステップで、コンテキスト データとして指定されたマップに関連付けられているデータ ソースのみです。

関係

近い イベントの発生場所が、マップ上の指定ポイントまたは領域に近い場所です。

内部 イベントの発生場所が、マップ上の指定領域内にあります。

含む イベントの領域に、マップのコンテキスト オブジェクトが含まれています。

交差 異なるマップの線または領域が相互に交わっている場所。

クロス 複数のマップの場合に、(道路、川、鉄道の) 異なる線の線が相互に交わっている場所。

北、南、東、西

イベントの発生場所が、マップ上の指定ポイントの北、南、東、または西の領域内にあります。

座標系の設定

マップの投影法 (.prj) ファイルがない場合、またはデータ ソースの 2 つのフィールドを座標セットとして定義する場合、座標系を設定する必要があります。

デフォルトの地理 (経度と緯度)(G)

座標系は、経度と緯度です。

単純なデカルト (X と Y)

座標系は単純な X と Y 座標です。

ウェルノウン ID を使用 (WKID)

一般的な投影法の「ウェルノウン ID」。

座標系の名前を使用

座標系は、名前を指定した投影法に基づきます。名前は、括弧で囲みます。

投影法の設定

マップで提供されている情報からは投影システムを決定できない場合、投影システムを指定する必要があります。この状態が発生する最も一般的な原因は、マップに関連付けられている投影 (.prj) ファイルがないこと、または、使用できない投影ファイルにあります。

- 都市、地域、または国 (メルカトル図法)
- 大きな国、複数の国、または大陸 (ヴィンケル図法)
- 赤道に非常に近い領域 (メルカトル図法)
- 北極または南極に近い領域 (心射図法)

メルカトル投影法は、多くのマップで使用される一般的な投影法です。この投影法は、地球を、平面に広げられた円柱として扱います。メルカトル投影法では、大きな対象のサイズと形状にゆがみが生じます。このゆがみは、赤道から遠くなるほど、そして北極または南極に近づくほど増大します。ヴィンケル図法と心射図法の投影法においては、マップは 2 次元で表示される 3 次元の球形の一部を表すという事実に関する調整が行われます。

投影法と座標系

複数のマップを選択し、各マップで異なる投影法と座標系を使用している場合、使用する投影システムが含まれたマップを選択する必要があります。すべてのマップで、マップを出力で組み合わせた場合に、その投影システムが使用されます。

データ ソース

データ ソースは、シェープ ファイルとともに提供された dBase ファイル、IBM SPSS Statistics データ ファイル、または現行セッション内の開いているデータ セットにすることができます。

コンテキスト データ。コンテキスト データは、マップ上のフィーチャを識別します。コンテキスト データには、モデルの入力として使用できるフィールドを含めることもできます。マップ シェープ (.shp) ファイルに関連したコンテキスト dBase (.dbf) ファイルを使用するには、そのコンテキスト dBase ファイルがシェープファイルと同じ場所になければならず、またルート名が同じでなければなりません。例えば、シェープファイルが geodata.shp の場合、dBase ファイルの名前は geodata.dbf でなければなりません。

イベント データ。イベント データには、発生したイベント (犯罪や事故など) の情報が含まれます。このオプションは、地理空間アソシエーション ルールの場合にのみ使用可能です。

点の密度。カーネル密度の推定の時間間隔および座標データ。このオプションは、空間時間的予測の場合にのみ使用可能です。

追加。データ ソースを追加するためのダイアログが開きます。データ ソースは、シェープ ファイルとともに提供された dBase ファイル、IBM SPSS Statistics データ ファイル、または現行セッション内の開いているデータ セットにすることができます。

関連付け。データをマップに関連付けるために使用される識別子 (座標またはキー) を指定するためのダイアログが開きます。各データ ソースには、データをマップに関連付ける 1 つ以上の識別子が含まれていな

ければなりません。シェープファイルに付属の dBase ファイルには通常、デフォルト識別子として自動的に使用されるフィールドが含まれています。他のデータ ソースの場合、識別子として使用するフィールドを指定する必要があります。

キーの検証: マップとデータ ソース間で一致するキーを検証するためのダイアログが開きます。

地理空間アソシエーション ルール

- 少なくとも 1 つのデータ ソースがイベント データ ソースでなければなりません。
- すべてのイベント データ ソースで、同じ形式のマップ アソシエーション識別子 (座標またはキー値) を使用しなければなりません。
- イベント データ ソースがキー値が含まれたマップに関連付けられている場合、すべてのイベント ソースで同じマップ フィーチャ タイプ (例えば、多角形、点、線) を使用する必要があります。

空間的時期的予測

- コンテキスト データ ソースが存在する必要があります。
- データ ソースが 1 つのみの場合 (関連マップがないデータ ファイル)、座標値を含める必要があります。
- 2 つのデータ ソースがある場合、一方のデータ ソースはコンテキスト データでなければならず、他方のデータ ソースは点の密度データでなければなりません。
- データ ソースを 2 つより多く組み込むことはできません。

データ ソースの追加

データ ソースは、シェープ ファイルおよびコンテキスト ファイルとともに提供された dBase ファイル、IBM SPSS Statistics データ ファイル、または現行セッション内の開いているデータ セットにすることができます。

同じデータ ソースを複数回追加できます (それぞれで別の空間的アソシエーションを使用する場合)。

データとマップのアソシエーション

各データ ソースには、データをマップに関連付ける 1 つ以上の識別子が含まれていなければなりません。

座標 データ ソースに、デカルト座標を表すフィールドが含まれます。X 座標と Y 座標を表すフィールドを選択します。地理空間アソシエーション ルールの場合、Z 座標が存在する場合があります。

キー値 データ ソース内のフィールドのキー値が、選択したマップ キーに対応します。例えば、地域マップに、各地域をラベル付ける名前識別子 (マップ キー) を含めることができます。この識別子は、地域の名前も含まれたデータ内のフィールド (データ キー) に対応します。フィールドは、2 つのリストに表示されている順序に基づいて、マップ キーに突き合わされます。

キーの検証

「キーの検証」ダイアログでは、選択した識別子キーに基づいて、マップとデータ ソース間で一致するレコードの要約が示されます。一致しないデータ キー値がある場合は、手動でその値をマップ キー値に一致させることができます。

地理空間アソシエーション ルール

地理空間アソシエーション ルールの場合、マップおよびデータ ソースを定義した後のウィザードでの残りのステップは以下のとおりです。

- 複数のイベント データ ソースがある場合は、イベント データ ソースを結合する方法を定義します。
- 分析で条件および予測として使用するフィールドを選択します。

オプションで、以下の操作も実行することができます。

- 異なる出力オプションを選択します。
- スコアリング モデル ファイルを保存します。
- モデルで使用するデータ ソースに予測値と規則の新規フィールドを作成します。
- アソシエーション ルールを構築するための設定をカスタマイズします。
- ビン分割と集計の設定をカスタマイズします。

イベント データ フィールドの定義

地理空間アソシエーション ルールでは、複数のイベント データ ソースがある場合、イベント データ ソースは結合されます。

- デフォルトでは、すべてのイベント データ ソースに共通のフィールドのみが含まれます。
- 共通フィールド、特定のデータ ソースのフィールド、またはすべてのデータ ソースのフィールドのリストを表示し、含めるフィールドを選択できます。
- 共通フィールドの場合、「タイプ」および「尺度」はすべてのデータ ソースで同じでなければなりません。競合がある場合は、共通フィールドごとに使用するタイプおよび尺度を指定できます。

フィールドの選択

使用可能なフィールドのリストには、イベント データ ソースのフィールドおよびコンテキスト データ ソースのフィールドが含まれています。

- 「データ ソース」リストからデータ ソースを選択することで、表示されるフィールドのリストを制御できます。
- 少なくとも 2 つのフィールドを選択する必要があります。少なくとも 1 つが条件でなければならず、また少なくとも 1 つが予測でなければなりません。「両方 (条件および予測)」リストに 2 つのフィールドを選択するなど、この要件を満たすための複数の方法があります。
- アソシエーション ルールは、条件フィールドの値に基づいて、予測フィールドの値を予測します。例えば、規則「If x=1 and y=2, then z=3」では、x と y の値が条件であり、z の値が予測です。

出力

規則表 各規則表には、確信度、ルール サポート、リフト、条件サポート、およびデプロイアビリティの上位の規則と値が表示されます。各テーブルは、選択した基準の値でソートされます。選択した基準に基づいて、すべての規則または上位「数」位までの規則を表示できます。

ソート可能なワード クラウド

選択した基準の値に基づいた、上位の規則のリスト。テキストのサイズは、規則の相対重要度を示します。インタラクティブ出力オブジェクトには、確信度、ルール サポート、リフト、条件サポート、およびデプロイアビリティの上位の規則が含まれます。選択した基準により、デフォルトで表示される規則のリストが決定されます。出力では、別の基準をインタラクティブに選択できます。「表示する規則の最大数」により、出力に表示するルールの数を決定します。

マップ 選択した基準に基づいた上位の規則のインタラクティブ棒グラフおよびマップ。各インタラクティブ出力オブジェクトには、確信度、ルール サポート、リフト、条件サポート、およびデプロイアビリティの上位の規則が含まれます。選択した基準により、デフォルトで表示される規則のリストが決定されます。出力では、別の基準をインタラクティブに選択できます。「表示する規則の最大数」により、出力に表示するルールの数を決定します。

モデル情報テーブル

フィールド変換。

分析で使用されているフィールドに適用される変換について説明します。

レコード要約 (Record Summary)。

含まれているレコードと除外されたレコードの数とパーセント。

規則統計。

条件サポート、確信度、ルール サポート、リフト、およびデプロイアビリティの要約統計量。統計量には、平均、最小、最大、および標準偏差が含まれます。

最頻出項目。

最頻出の項目。項目は、規則の条件または予測に組み込まれています。例えば、age < 18 や gender=female。

最頻出フィールド。

規則内で最頻出のフィールド。

除外された入力。

分析から除外されたフィールドと、各フィールドが除外された理由。

規則表、ワード クラウド、およびマップの基準

信頼度。

正しい規則の予測のパーセント。

ルール サポート。

規則が真であるケースのパーセント。例えば、規則が「If x=1 and y=2, then z=3」の場合、ルールサポートは、データ内で x=1、y=2、かつ z=3 であるケースの実際のパーセントです。

リフト。

リフトは、規則により、無作為確率と比較して予測が改善される度合いを示す指標です。これは、予測値の出現数全体に対する正しい予測の比率です。値は、1 より大きくなくてはなりません。例えば、予測値が 20% の確率で発生し、予測の確信度が 80% の場合、リフトの値は 4 になります。

条件サポート。

規則条件が存在しているケースのパーセント。例えば、規則が「If x=1 and y=2, then z=3」の場合、条件サポートは、データ内で x=1 かつ y=2 であるケースの比率です。

デプロイアビリティ。

条件が true だった場合の、誤った予測のパーセンテージ。デプロイアビリティは、条件サポートの値または条件サポートからルール サポートを引いた値によって、「1 から信頼度を引いた値」が乗算された値に相当します。

保存

マップおよびコンテキスト データをマップ仕様として保存する(M)

マップ仕様を外部ファイル (.mplan) に保存します。このマップ仕様ファイルをその後の分析のためにウィザードにロードできます。また、マップ仕様ファイルは SPATIAL ASSOCIATION RULES コマンドでも使用できます。

すべてのマップおよびデータ ファイルを仕様にコピー

マップ シェープファイル、外部データ ファイル、およびマップ仕様で使用されるデータセットのデータが、マップ仕様ファイルに保存されます。

スコアリング

指定したデータ ソースに新規フィールドとして、最適な規則値、規則の確信度値、および規則の数値 ID 値を保存します。

スコアを作成するデータ ソース

新規フィールドを作成するデータ ソース。データ ソースは、現行セッションで開かれていない場合、現行セッションで開かれます。新規フィールドを保存するには、変更したファイルを明示的に保存する必要があります。

目標値 選択した対象 (予測) フィールドの新規フィールドを作成します。

- 対象フィールドごとに 2 つの新規フィールド (予測値および確信度値) が作成されます。
- 連続型 (スケール) 対象フィールドの場合、予測値は、値範囲を記述する文字列になります。「(value1, value2)」という形式の値は、「value1 より大きく、value2 以下」という意味です。

最適な規則の数

指定した最適な規則の数の新規フィールドを作成します。規則ごとに 3 つの新規フィールド (規則値、確信度値、規則の数値 ID 値) が作成されます。

名前の接頭辞

新規フィールド名に使用する接頭辞。

規則の作成

規則の作成パラメータは、生成されるアソシエーション ルールの基準を設定します。

規則ごとの項目

規則の条件および予測に含めることができるフィールド値の数。項目の総数は、10 を超えてはなりません。例えば、規則「If x=1 and y=2, then z=3」の場合、2 つの条件項目と 1 つの予測項目があります。

最大予測値。

規則の予測に含めることができるフィールド値の最大数。

最大条件。

規則の条件に含めることができるフィールド値の最大数。

ペアの除外

指定したフィールドのペアが同じ規則に含まれないように除外します。

規則基準

信頼度。

出力に含めるために必要な規則の最小確信度。確信度は、正しい予測のパーセントです。

ルール サポート。

出力に含めるために必要な規則の最小ルール サポート。値は、観測データで規則が真であるケースのパーセントを表します。例えば、規則が「If x=1 and y=2, then z=3」の場合、ルール サポートは、データ内で x=1、y=2、かつ z=3 であるケースの実際のパーセントです。

条件サポート。

出力に含めるために必要な規則の最小条件サポート。値は、条件が存在するケースのパーセントを表します。例えば、規則が「If x=1 and y=2, then z=3」の場合、条件サポートは、データ内で x=1 かつ y=2 であるケースのパーセントです。

リフト。

出力に含めるために必要な規則の最小リフト。リフトとは、規則によって無作為確率の場合よりも予測が改善される度合いを示す指標のことです。これは、予測値の出現数全体に対する正しい予測の比率です。例えば、予測値が 20% の確率で発生し、予測の確信度が 80% の場合、リフトの値は 4 になります。

同じものとして扱う

同じフィールドとして扱う必要がある、フィールドのペアを指定します。

ビン分割と集計

- マップ内のフィーチャより多くのレコードがデータ内に存在する場合、集計が必要になります。例えば、個別の国のデータ レコードがあるが、州のマップがある場合です。
- 連続型フィールドおよび順序フィールドの集計要約統計量の方法を指定できます。名義型フィールドは、最頻値に基づいて集計されます。

連続 連続型 (スケール) フィールドの場合、要約統計量は平均、中央値、または合計にすることができます。

順序 順序フィールドの場合、要約統計量は、中央値、最頻値、最大値、または最小値にすることができます。

ビン数 連続型 (スケール) フィールドのビンの最大数を設定します。連続型フィールドは常に、値の範囲にグループ化 (「ビン分割」) されます。例えば、5 以下、5 より大きく 10 以下、10 より大のようになります。

マップをグループ集計(M)

集計をデータとマップの両方に適用します。

特定のフィールドのユーザー指定の設定

特定のフィールドについて、デフォルトの要約統計量およびビン数をオーバーライドできます。

- アイコンをクリックして「**フィールド ピッカー**」ダイアログを開き、リストに追加するフィールドを選択します。
- 「**集計**」列で、要約統計量を選択します。
- 連続型フィールドの場合、「**ビン**」列のボタンをクリックして、「**ビン**」ダイアログのフィールドのカスタム ビン数を指定します。

空間的時期的予測

空間的時期的予測の場合、マップおよびデータ ソースを定義した後のウィザードでの残りのステップは以下のとおりです。

- 対象フィールド、時間フィールド、およびオプションの予測変数を指定します。
- 時間フィールドの時間間隔または周期的な期間を定義します。

オプションで、以下の操作も実行することができます。

- 異なる出力オプションを選択します。
- モデルの構築パラメータをカスタマイズします。
- 集計設定をカスタマイズします。
- 予測値を現行セッションのデータ セットまたは IBM SPSS Statistics 形式のデータ ファイルに保存します。

フィールドの選択

使用可能なフィールドのリストには、選択したデータ ソースのフィールドが含まれています。「データ ソース」リストからデータ ソースを選択することで、表示されるフィールドのリストを制御できます。

対象 対象フィールドが必要です。対象は、値を予測するフィールドです。

- 対象フィールドは、連続型 (スケール) の数値フィールドでなければなりません。
- 2 つのデータ ソースがある場合、対象はカーネル密度の推定になり、「密度」が対象名として表示されます。この選択を変更することはできません。

予測値 1 つ以上の予測フィールドを指定できます。この設定はオプションです。

時刻フィールド

期間を表す 1 つ以上のフィールドを選択するか、「**周期的な期間**」を選択する必要があります。

- 2 つのデータ ソースがある場合、両方のデータ ソースから時間フィールドを選択する必要があります。両方の時間フィールドは、同じ間隔を表している必要があります。
- 周期的な期間の場合、ウィザードの「時間区分」パネルで周期を定義するフィールドを指定する必要があります。

時間区分

このパネル内のオプションは、フィールドを選択するステップにおける「時刻フィールド」または「周期的な期間」の選択に基づきます。

時刻フィールド

選択された時間フィールド。 フィールドを選択するステップで 1 つ以上の時間フィールドを選択した場合、そのフィールドがこのリストに表示されます。

時間区分: リストから該当する時間間隔を選択します。時間間隔に応じて、観測の間隔 (増分) や開始値などの他の設定も指定できます。この時間間隔は、選択したすべての時間フィールドで使用されます。

- このプロシージャでは、すべてのケース (レコード) が等間隔を表しているものと想定されます。
- 選択した時間間隔に基づいて、プロシージャは、欠損している観測値、および集計する必要がある、同じ時間間隔内の複数の観測値を検出できます。例えば、時間間隔が日で、日付 2014-10-27 の後に 2014-10-29 が続いている場合、2014-10-28 の観測値が欠損しています。時間間隔が月の場合、同じ月の複数の日付は集計されます。

- 一部の時間間隔では、追加設定により、通常の等間隔内にブレイクを定義できます。例えば、時間間隔が日であるが、平日のみが有効な場合、1 週間に 5 日あり、週が月曜日から始まるということを指定できます。
- 選択された時間フィールドが日付形式のフィールドでも時刻形式のフィールドでもない場合、時間間隔は自動的に「期間」に設定され、変更できません。

サイクル フィールド

フィールドを選択するステップで「**周期的な期間**」を選択した場合、周期的な期間を定義するフィールドを指定する必要があります。周期的な期間は、1 年の月数や 1 週間の日数などの、反復周期の変動を示します。

- 周期的な期間を定義するフィールドを最大で 3 つ指定できます。
- 最初のサイクル フィールドは、サイクルの最高レベルを表します。例えば、年、四半期、および月で周期的な変動がある場合、年を表すフィールドが最初のサイクル フィールドです。
- 最初のサイクル フィールドと 2 番目のサイクル フィールドのサイクルの長さは、後続レベルにおける周期になります。例えば、サイクル フィールドが年、四半期、および月の場合、最初のサイクルの長さは 4 であり、2 番目のサイクルの長さは 3 です。
- 2 番目のサイクル フィールドおよび 3 番目のサイクル フィールドの開始値は、該当するそれぞれの周期的な期間の最初の値です。
- サイクルの長さおよび開始値は、正整数でなければなりません。

集計

- フィールドを選択するステップで「**予測変数**」を選択した場合、予測変数の集計要約方法を選択できます。
- 定義された時間間隔内に複数のレコードがある場合、集計が必要になります。例えば、時間間隔が月の場合、同じ月の複数の日付は集計されます。
- 連続型フィールドおよび順序フィールドの集計要約統計量の方法を指定できます。名義型フィールドは、最頻値に基づいて集計されます。

連続 連続型 (スケール) フィールドの場合、要約統計量は平均、中央値、または合計にすることができます。

順序 順序フィールドの場合、要約統計量は、中央値、最頻値、最大値、または最小値にすることができます。

特定のフィールドのユーザー指定の設定

特定の予測変数のデフォルト集計要約統計量をオーバーライドできます。

- アイコンをクリックして「**フィールド ピッカー**」ダイアログを開き、リストに追加するフィールドを選択します。
- 「**集計**」列で、要約統計量を選択します。

出力

マップ

目標値。

選択した対象フィールドの値のマップ。

相関 相関のマップ。

クラスタ

相互に似ているクラスタの場所を強調表示するマップ。クラスタのマップは、経験的モデルでのみ提供されます。

位置類似度のしきい値。

クラスタの作成に必要な類似度。値は、ゼロより大きく 1 より小さい数値にする必要があります。

最大クラスタ数の指定。

表示するクラスタの最大数。

モデル評価テーブル

モデル指定。

対象、入力、場所の各フィールドなど、分析の実行に使用する指定の要約。

時間的情報の要約。

モデルで使用される時間フィールドおよび時間間隔を示します。

平均値構造における効果の検定。

出力に、モデルおよび各効果の検定統計値、自由度、および有意水準が含まれます。

モデル係数の平均値構造。

出力に、各モデル項の係数値、標準誤差、検定統計値、有意水準、および信頼区間が含まれます。

自己回帰係数。

出力に、各ラグの係数値、標準誤差、検定統計値、有意水準、および信頼区間が含まれます。

空間共分散のテスト。

バリオグラムに基づくパラメトリック モデルの場合、空間共分散構造の適合度検定の結果を表示します。この検定結果により、空間共分散構造をパラメトリックにモデル化するか、それともノンパラメトリック モデルを使用するかを判別することができます。

パラメトリック空間共分散。

バリオグラムに基づくパラメトリック モデルの場合、パラメトリック空間共分散のパラメータ推定値を表示します。

モデル オプション

モデルの設定

定数項を自動的に含める(A)

モデルに定数項を含めます。

自己相関誤差の回帰の最大ラグ (Maximum autoregression lag)

自己相関誤差の回帰の最大ラグ。この値は、1 から 5 までの整数にする必要があります。

空間共分散

空間共分散の推定方法を指定します。

パラメトリック(P)

推定方法をパラメトリックにします。方法は、「ガウス」、「指数」、または「べき乗指数」のいずれかにすることができます。「べき乗指数」では、「べき乗」値を指定できます。

ノンパラメトリック(E)

推定方法をノンパラメトリックにします。

保存

マップおよびコンテキスト データをマップ仕様として保存する(M)

マップ仕様を外部ファイル (.mplan) に保存します。このマップ仕様ファイルをその後の分析のためにウィザードにロードできます。また、マップ仕様ファイルは SPATIAL TEMPORAL PREDICTION コマンドでも使用できます。

すべてのマップおよびデータ ファイルを仕様にコピー

マップ シェープ ファイル、外部データ ファイル、およびマップ仕様で使用されるデータ セットのデータが、マップ仕様ファイルに保存されます。

スコアリング

選択したデータ ファイルに対象フィールドの予測値、分散、および信頼境界の上限と下限を保存します。

- 現行セッションの開いているデータ セットまたは IBM SPSS Statistics 形式のデータ ファイルに予測値を保存できます。
- データ ファイルをモデルで使用されているデータ ソースにすることはできません。
- データ ファイルには、モデルで使用されているすべての時間フィールドおよび予測変数が含まれている必要があります。
- これらの時間値は、モデルで使用されている時間値よりも大きくなければなりません。

詳細設定

欠損値のある最大ケース数 (%)

欠損値のあるケースの最大パーセント。

有意水準

バリオグラムに基づくパラメトリック モデルが適切かどうかを判別するための有意水準。この値は、0 よりも大きく 1 未満でなければなりません。デフォルト値は 0.05 です。有意水準は、空間共分散構造の適合度検定で使用されます。適合度統計は、パラメトリック モデルとノンパラメトリック モデルのどちらを使用するかの判別に使用されます。

不確実性因子 (%)

不確実性因子とは、将来の予測に関する不確実性の増加を表すパーセント値です。将来に向けたステップごとに、指定したパーセント分だけ、予測の不確実性の上限と下限が増加します。

完了

地理空間モデリング ウィザードの最終ステップで、モデルを実行するか、生成されたコマンド シンタックスをシンタックス ウィンドウに貼り付けることができます。今後の使用のために、生成されたシンタックスを変更して保存できます。

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

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

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

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

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

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

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

索引

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

[ア行]

赤池情報量基準

線型モデル 63

アルファ因子法 102

アルファ係数

信頼性分析 163, 164

アンサンブル

線型モデル 64

イータ

グループの平均 25

クロス集計表 16

イータの 2 乗

グループの平均 25

GLM 1 変数 47, 50, 52

一元配置分散分析 39

因子変数 39

オプション 41

欠損値 41

コマンドの追加機能 42

その後の検定 40

対比 39

多項式の対比 39

多重比較 40

統計 41

位置モデル

順序回帰 77

一致するペアの研究

対応のあるサンプルの T 検定 34

一般化最小二乗法

因子分析 102

イメージ因子法 102

因子得点 104

因子分析 101

因子得点 104

回転法 103

概要 101

記述統計 102

ケースの選択 102

係数の表示書式 104

欠損値 104

コマンドの追加機能 104

収束 102, 103

抽出法 102

統計 101, 102

因子分析 (続き)

プロットのロード 103

例 101

エカマックス回転

因子分析 103

円グラフ

度数分布表 7

オーバーフィット防止基準

線型モデル 63

重み付き最小二乗法

線型回帰 69

重み付き平均値

比率統計 171

重み付き予測値

GLM 51

重み付けのない最小二乗法

因子分析 102

折半法の信頼性

信頼性分析 163, 164

[カ行]

カイ 2 乗 140

オプション 141

期待値 140

期待範囲 140

クロス集計表 16

欠損値 141

線型と線型による連関 16

統計 141

独立性 16

尤度比 16

1 サンプルの検定 140

Fisher の直接法 16

Pearson 16

Yates の連続修正 16

カイ 2 乗距離

距離 59

カイ 2 乗検定

1 サンプルのノンパラメトリック検定
126, 127

回帰

作図 70

線型回帰 69

複数の回帰 69

回帰係数

線型回帰 72

階層クラスター分析 117

値の変換 117

距離行列 118

距離測度 117

階層クラスター分析 (続き)

クラスター化の方法 117

クラスター凝集経過工程 118

クラスター変数 117

ケースのクラスタリング 117

コマンドの追加機能 118

所属クラスター 118

新変数の保存 118

測定方法の変換 117

つららプロット 118

デンドログラム 118

統計 117, 118

プロット方向 118

類似度の測定方法 117

例 117

階層的分解 45

外れ値

線型回帰 70

探索的 12

TwoStep クラスター分析 108

価格関連格差 (PRD)

比率統計 171

学習サンプル

最近傍分析 88

確率密度関数

シミュレーション 187

カスタム・モデル

GLM 44

仮説の要約

ノンパラメトリック検定 135

カッパ

クロス集計表 16

カテゴリ・フィールド情報

ノンパラメトリック検定 138

観測度数

クロス集計表 18

順序回帰 76

観測平均値

GLM 1 変数 47, 50, 52

感度分析

シミュレーション 185

ガンマ

クロス集計表 16

幾何平均

グループの平均 25

ケースの要約 22

OLAP キューブ 29

記述統計 9

統計 9

表示順 9

z 得点の保存 9

記述統計量	空間モデリング 195	クロス集計表 (続き)
記述統計 9	クォーティマックス回転	制御変数 16
ケースの要約 22	因子分析 103	セル表示 18
探索的 12	クラスターの度数	層 16
度数分布表 5	TwoStep クラスタ分析 109	多重回答 153
比率統計 171	クラスター分析	統計 16
GLM 1 変量 47, 50, 52	階層クラスター分析 117	表の抑制 15
TwoStep クラスタ分析 109	効率 122	ケースごとの診断情報
期待度数	大規模ファイルのクラスター分析 121	線型回帰 72
クロス集計表 18	クラスター・ビューアー	ケースの数
順序回帰 76	概要 110	グループの平均 25
逆 Helmert 対比	基本ビュー 112	ケースの要約 22
GLM 46	「クラスター」ビュー 111	OLAP キューブ 29
逆数モデル	「クラスター中心」ビュー 111	ケースのリスト 21
曲線推定 80	クラスターとフィールドの入れ替え	ケース・コントロール研究
級内相関係数 (ICC)	112	対応のあるサンプルの T 検定 34
信頼性分析 164	クラスターとフィールドの反転 112	傾向化除去正規プロット
共線性の診断情報	クラスターのサイズ 113	探索的 12
線型回帰 72	クラスターのソート 112	結合ルール
行パーセント	クラスターの比較 113	線型モデル 64
クロス集計表 18	「クラスターの比較」ビュー 113	欠損値
共分散行列	クラスター表示のソート 112	一元配置分散分析 41
順序回帰 76	「クラスター予測値の重要度」ビュー	因子分析 104
線型回帰 72	113	カイ 2 乗検定 141
判別分析 96, 97	「クラスター・サイズ」ビュー 113	最近傍分析 90
GLM 51	クラスター・モデルについて 110	線型回帰 73
共分散比	使用 114	対応のあるサンプルの t 検定 35
線型回帰 71	セル内容のソート 112	多重回答のクロス集計表 154
曲線推定 79	セル内容の表示 112	多重回答の度数表 152
残差の保存 80	セルの分布 113	探索的 13
定数を含む 79	「セルの分布」ビュー 113	独立したサンプルの t 検定 34
分散分析 79	フィールドのソート 112	複数の独立サンプルの検定 148
モデル 80	フィールド表示のソート 112	偏相関 57
予測 80	モデルの要約 111	報告書の行の集計 159
予測値の保存 80	要約ビュー 111	ラン検定 143
予測区間の保存 80	予測値の重要度 113	列の集計報告書 161
極値	レコードのフィルタリング 115	1 サンプルによる Kolmogorov-Smirnov
探索的 12	クラスタリング 110	検定 144
許容度	クラスターの表示 110	1 サンプルの t 検定 36
線型回帰 72	全体表示 110	2 項検定 142
距離 59	手続きの選択 105	2 個の対応サンプルの検定 147
値の変換 59, 60	グラフ	2 個の独立サンプルの検定 145
ケース間の距離の計算 59	ケースのラベル 79	2 変量の相関分析 55
コマンドの追加機能 60	グループ間の差分	ROC 曲線 173
測定方法の変換 59, 60	OLAP キューブ 31	厳密平行モデル
統計 59	グループ中央値	信頼性分析 163, 164
非類似度の測定方法 59	グループの平均 25	コードブック 1
変数間の距離の計算 59	ケースの要約 22	出力 1
類似度の測定方法 60	OLAP キューブ 29	統計 3
例 59	グループの比較	効果サイズの推定値
距離測定	OLAP キューブ 31	GLM 1 変量 47, 50, 52
階層クラスター分析 117	グループ平均 25, 29	合計
距離 59	クロス集計表 15	記述統計 9
最近傍分析 87	クラスター棒グラフ 16	グループの平均 25
近接	クロス集計表 15	ケースの要約 22
階層クラスター分析 117	書式 19	度数分布表 5

合計 (続き)
 OLAP キューブ 29
 合計パーセント
 クロス集計表 18
 合計列
 報告書 160
 交互作用項 45, 77, 78
 項目の構築 45, 77, 78
 誤差の集計
 最近傍分析 93
 固有値
 因子分析 102
 線型回帰 72

[サ行]

最近傍の距離
 最近傍分析 92
 最近傍分析 85
 オプション 90
 近傍 87
 出力 89
 特徴選択 88
 分割 88
 変数の保存 89
 モデル・ビュー 90
 最後
 グループの平均 25
 ケースの要約 22
 OLAP キューブ 29
 最初
 グループの平均 25
 ケースの要約 22
 OLAP キューブ 29
 最小値
 記述統計 9
 グループの平均 25
 ケースの要約 22
 探索的 12
 度数分布表 5
 比率統計 171
 報告書の列の比較 160
 OLAP キューブ 29
 最小有意差
 一元配置分散分析 40
 GLM 48
 サイズの差異測定
 距離 59
 最大値
 記述統計 9
 グループの平均 25
 ケースの要約 22
 探索的 12
 度数分布表 5
 比率統計 171
 報告書の列の比較 160

最大値 (続き)
 OLAP キューブ 29
 最大枝数
 TwoStep クラスタ分析 108
 最適サブセット
 線型モデル 63
 最頻値
 度数分布表 5
 最尤法
 因子分析 102
 削除された残差
 線型回帰 71
 GLM 51
 サブグループ平均 25, 29
 残差
 曲線推定で保存 80
 クロス集計表 18
 線型回帰での保存 71
 残差プロット
 GLM 1 変数 47, 50, 52
 参照カテゴリー
 GLM 46
 散布図
 シミュレーション 188
 線型回帰 70
 視覚化
 クラスタリング・モデル 110
 時系列分析
 ケースの予測 80
 予測 80
 辞書
 コードブック 1
 指数モデル
 曲線推定 80
 自動データ準備
 線型モデル 65
 自動分布適合
 シミュレーション 181
 四分位分布図
 最近傍分析 92
 シミュレーション 175
 新しい入力の作成 180
 インタラクティブ・グラフ 192
 確率密度関数 187
 感度分析 185
 グラフ・オプション 193
 サポートされるモデル 179
 散布図 188
 シミュレーションされたデータの保存
 189
 シミュレーション・ビルダー 178
 シミュレーション・プランの作成
 176, 177
 シミュレーション・プランの実行
 178, 190

シミュレーション (続き)
 シミュレーション・プランの保存 189
 出力 187, 188
 裾サンプル 186
 停止基準 186
 トルネード図 188
 入力間の相関 185
 箱ひげ図 188
 分布適合のカスタマイズ 184
 分布適合の結果 184
 分布の新しいデータへの再適合 190
 分布の適合 181
 方程式エディター 180
 目標と入力の表示形式 188
 目標分布のパーセンタイル 188
 モデル指定 179
 累積分布関数 187
 what-if 分析 185
 シミュレーション・ビルダー 178
 尺度モデル
 順序回帰 78
 主因子法 102
 収束
 因子分析 102, 103
 大規模ファイルのクラスタ分析 122
 従属 t 検定
 対応のあるサンプルの T 検定 34
 周辺等質性検定
 対応サンプルのノンパラメトリック検
 定 132
 2 個の対応サンプルの検定 146
 主成分分析 101, 102
 順位相関係数
 2 変量の相関分析 55
 順序回帰 75
 位置モデル 77
 オプション 76
 コマンドの追加機能 78
 尺度モデル 78
 統計 75
 リンク 76
 小計
 列の集計報告書 161
 乗算
 列の報告書全体の乗算 160
 情報量基準
 線型モデル 63
 初期しきい値
 TwoStep クラスタ分析 108
 除算
 列の報告書全体の除算 160
 書式設定
 報告書の列 158
 信頼区間
 一元配置分散分析 41
 線型回帰 72

- 信頼区間 (続き)
 - 線型回帰での保存 71
 - 対応のあるサンプルの t 検定 35
 - 探索的 12
 - 独立したサンプルの t 検定 34
 - 1 サンプルの t 検定 36
 - GLM 46, 47, 50, 52
 - ROC 曲線 173
 - 信頼区間の要約
 - ノンパラメトリック検定 135, 136
 - 信頼性分析 163
 - 記述統計 164
 - 級内相関係数 164
 - 項目間の相関係数および共分散 164
 - コマンドの追加機能 165
 - 統計 163, 164
 - 分散分析表 164
 - 例 163
 - Hotelling の T2 164
 - Kuder-Richardson 20 164
 - Tukey の加法性の検定(K) 164
 - 水準と広がり の図
 - 探索的 12
 - GLM 1 変量 47, 50, 52
 - 推定周辺平均
 - GLM 1 変量 47, 50, 52
 - スケール
 - 信頼性分析 163
 - 多次元スケーリング法 167
 - スチューデント t 検定 33
 - スチューデント化された残差
 - 線型回帰 71
 - ステップワイズ選択
 - 線型回帰 70
 - ストレス
 - 多次元スケーリング法 167
 - 図表
 - ROC 曲線 173
 - すべての因子によるモデル
 - GLM 44
 - 正規確率プロット
 - 線型回帰 70
 - 探索的 12
 - 正規性検定
 - 探索的 12
 - 制御変数
 - クロス集計表 16
 - 成長モデル
 - 曲線推定 80
 - 線型回帰 69
 - 重み 69
 - 欠損値 73
 - コマンドの追加機能 74
 - 作図 70
 - 残差 71
 - 新変数の保存 71
 - 線型回帰 (続き)
 - 選択変数 70
 - 統計 72
 - ブロック 69
 - 変数選択方法 70, 73
 - モデル情報のエクスポート 71
 - 線型性の検定
 - グループの平均 25
 - 線型と線型による連関
 - クロス集計表 16
 - 線型モデル 61
 - アンサンプル 64
 - 外れ値 66
 - 曲線推定 80
 - 係数 66
 - 結果の再現 64
 - 結合ルール 64
 - 残差 65
 - 自動データ準備 62, 65
 - 情報量基準 64
 - 信頼度レベル 62
 - 推定平均 67
 - 分散分析表 66
 - 目的 61
 - モデル構築の要約 67
 - モデル選択 63
 - モデルの要約 64
 - モデル・オプション 64
 - 予測値の重要度 65
 - 予測対観測 65
 - R 2 乗統計量 64
 - 選択変数
 - 線型回帰 70
 - 尖度
 - 記述統計 9
 - グループの平均 25
 - ケースの要約 22
 - 探索的 12
 - 度数分布表 5
 - 報告書の行の集計 158
 - 報告書の列の集計 160
 - OLAP キューブ 29
 - 尖度の標準誤差
 - グループの平均 25
 - ケースの要約 22
 - OLAP キューブ 29
 - 層
 - クロス集計表 16
 - 相関
 - クロス集計表 16
 - シミュレーション 185
 - 偏相関 57
 - 0 次 57
 - 2 変量の相関分析 55
 - 相関行列
 - 因子分析 101, 102
 - 相関行列 (続き)
 - 順序回帰 76
 - 判別分析 96
 - 総計
 - 列の集計報告書 161
 - 相対リスク
 - クロス集計表 16
 - その後の多重比較 40
- ## [夕行]
- 対応サンプル 146, 148
 - 対応サンプルのノンパラメトリック検定 131
 - フィールド 132
 - Cochran の Q 検定 133
 - McNemar 検定 133
 - 対応のあるサンプルの T 検定 34
 - オプション 35
 - 欠損値 35
 - 対応のある変数の選択 34
 - 大規模ファイルのクラスター分析
 - 概要 121
 - クラスター距離 122
 - クラスター情報の保存 122
 - 欠損値 123
 - 効率 122
 - コマンドの追加機能 123
 - 収束基準 122
 - 所属クラスター 122
 - 統計 121, 123
 - 反復 122
 - 方法 121
 - 例 121
 - 対数モデル
 - 曲線推定 80
 - 対比
 - 一元配置分散分析 39
 - GLM 46
 - タウ b
 - クロス集計表 16
 - タウ c
 - クロス集計表 16
 - 多項式の対比
 - 一元配置分散分析 39
 - GLM 46
 - 多次元スケーリング法 167
 - 値の変換 168
 - 基準 169
 - 距離行列の作成 168
 - 距離測度 168
 - コマンドの追加機能 169
 - 次元 168
 - 条件付け 168
 - スケーリング・モデル 168
 - 測定レベル 168

- 多次元スケーリング法 (続き)
 - データの型の定義 168
 - 統計 167
 - 表示オプション 169
 - 例 167
 - 多重 R
 - 線型回帰 72
 - 多重回答
 - コマンドの追加機能 155
 - 多重回答グループを定義 151
 - カテゴリー 151
 - 名前の設定 151
 - ラベルの設定 151
 - 2 分変数 151
 - 多重回答セット
 - コードブック 1
 - 多重回答のクロス集計表 153
 - 値の範囲の定義 154
 - 回答数に基づくパーセント 154
 - グループ間で変数を順に整合 154
 - ケース数に基づくパーセント 154
 - 欠損値 154
 - セルのパーセント 154
 - 多重回答の度数表 152
 - 欠損値 152
 - 多重回答分析
 - クロス集計表 153
 - 多重回答のクロス集計表 153
 - 多重回答の度数表 152
 - 度数分布表 152
 - 多重比較
 - 一元配置分散分析 40
 - 探索 11
 - オプション 13
 - 欠損値 13
 - コマンドの追加機能 13
 - 作図 12
 - 統計 12
 - べき乗変換 12
 - 単純対比
 - GLM 46
 - 中央値
 - グループの平均 25
 - ケースの要約 22
 - 探索的 12
 - 度数分布表 5
 - 比率統計 171
 - OLAP キューブ 29
 - 中心傾向の測定
 - 探索的 12
 - 度数分布表 5
 - 比率統計 171
 - 調整済み R 2 乗
 - 線型回帰 72
 - 線型モデル 63
 - 調和平均
 - グループの平均 25
 - ケースの要約 22
 - OLAP キューブ 29
 - 直接オプティミム回転
 - 因子分析 103
 - 散らばり係数 (COD)
 - 比率統計 171
 - 散らばりの測定
 - 記述統計 9
 - 探索的 12
 - 度数分布表 5
 - 比率統計 171
 - 地理空間モデリング 195, 196, 197, 198, 199, 201, 202, 203, 204, 205, 206
 - つららプロット
 - 階層クラスター分析 118
 - ツリーの深さ
 - TwoStep クラスター分析 108
 - 適合度
 - 順序回帰 76
 - てこ比の値
 - 線型回帰 71
 - GLM 51
 - デンドログラム
 - 階層クラスター分析 118
 - 同位
 - 最近傍分析 92
 - 等分散性サブセット
 - ノンパラメトリック検定 139
 - 等分散性の検定
 - 一元配置分散分析 41
 - GLM 1 変数 47, 50, 52
 - 特徴空間図
 - 最近傍分析 90
 - 特徴選択
 - 最近傍分析 92
 - 独立サンプルの検定
 - ノンパラメトリック検定 137
 - 独立サンプルのノンパラメトリック検定 128
 - 「フィールド」タブ 129
 - 独立したサンプルの T 検定 33
 - オプション 34
 - グループ化変数 34
 - グループの定義 34
 - 欠損値 34
 - 信頼区間 34
 - 文字型変数 34
 - 独立性の検定
 - カイ 2 乗 16
 - 都市ブロック距離
 - 最近傍分析 87
 - 度数 5
 - グラフ 7
 - 書式 7
 - 度数 (続き)
 - 統計 5
 - 表示順 7
 - 表の抑制 7
 - 度数分布表
 - 探索的 12
 - 度数分布表 5
 - トルネード図
 - シミュレーション 188
- ## [ナ行]
- ノイズ処理
 - TwoStep クラスター分析 108
 - 濃度インデックス
 - 比率統計 171
 - ノンパラメトリック検定
 - カイ 2 乗 140
 - 複数の対応サンプルの検定 148
 - 複数の独立サンプルの検定 147
 - モデル・ビュー 134
 - ラン検定 142
 - 1 サンプル Kolmogorov-Smirnov 検定 143
 - 2 個の対応サンプルの検定 146
 - 2 個の独立サンプルの検定 144
- ## [ハ行]
- パーセンタイル
 - シミュレーション 188
 - 探索的 12
 - 度数分布表 5
 - パーセンテージ
 - クロス集計表 18
 - バギング
 - 線型モデル 61
 - 箱ひげ図
 - 因子の水準の比較 12
 - シミュレーション 188
 - 探索的 12
 - 変数の比較 12
 - パターン行列
 - 因子分析 101
 - パターンの差異測定
 - 距離 59
 - パラメーター推定値
 - 順序回帰 76
 - GLM 1 変数 47, 50, 52
 - バリマックス回転
 - 因子分析 103
 - 範囲
 - 記述統計 9
 - グループの平均 25
 - ケースの要約 22

範囲 (続き)	標準偏差	分散 (続き)
度数分布表 5	記述統計 9	探索的 12
比率統計 171	グループの平均 25	度数分布表 5
OLAP キューブ 29	ケースの要約 22	報告書の行の集計 158
反復	探索的 12	報告書の列の集計 160
因子分析 102, 103	度数分布表 5	OLAP キューブ 29
大規模ファイルのクラスター分析 122	比率統計 171	分散拡大係数
反復測定	報告書の行の集計 158	線型回帰 72
GLM 46	報告書の列の集計 160	分散分析
反復履歴	GLM 1 変量 47, 50, 52	一元配置分散分析 39
順序回帰 76	OLAP キューブ 29	曲線推定 79
判別分析 95	表題	グループの平均 25
関数係数 96	OLAP キューブ 32	線型回帰 72
記述統計量 96	比率統計 171	線型モデル 66
基準 97	統計 171	モデル 44
共分散行列 97	プースティング	GLM 1 変量 43
行列 96	線型モデル 61	分布の測定
グループ化変数 95	ファイ	記述統計 9
ケースの選択 96	クロス集計表 16	度数分布表 5
欠損値 97	ファイ 2 乗距離	分布の適合
コマンドの追加機能 98	距離 59	シミュレーション 181
作図 97	不確定性係数	分類
事前確率 97	クロス集計表 16	ROC 曲線 173
ステップワイズ法 95	複合モデル	分類テーブル
統計 95, 96	曲線推定 80	最近傍分析 93
独立変数 95	複数の回帰	ページの制御
範囲の定義 96	線型回帰 69	行の集計報告書 159
判別分析法 97	複数の対応サンプルの検定 148	列の集計報告書 161
表示オプション 97	検定の種類 149	ページの番号付け
分類変数の保存 98	コマンドの追加機能 149	行の集計報告書 159
モデル情報のエクスポート 98	統計 149	列の集計報告書 161
例 95	複数の独立サンプルの検定 147	ペアごとの比較
Mahalanobis の距離 97	オプション 148	ノンパラメトリック検定 138
Rao の V 97	グループ化変数 148	平均
Wilks のラムダ 97	欠損値 148	一元配置分散分析 41
ヒストグラム	検定の種類 147	記述統計 9
線型回帰 70	コマンドの追加機能 148	グループの平均 25
探索的 12	統計 148	ケースの要約 22
度数分布表 7	範囲の定義 148	サブグループ 25, 29
標準化	符号検定	探索的 12
TwoStep クラスター分析 108	対応サンプルのノンパラメトリック検	度数分布表 5
標準回帰係数	定 132	比率統計 171
線型回帰 72	2 個の対応サンプルの検定 146	複数の列の報告書 160
標準化された値	ブロック距離	報告書の行の集計 158
記述統計 9	距離 59	報告書の列の集計 160
標準化されていない残差	プロットのロード	OLAP キューブ 29
GLM 51	因子分析 103	平均値 25
標準化残差	プロファイル・プロット	オプション 25
線型回帰 71	GLM 47	統計 25
GLM 51	分割係数	平均値の標準誤差
標準誤差	クロス集計表 16	グループの平均 25
記述統計 9	分割表 15	ケースの要約 22
探索的 12	分散	OLAP キューブ 29
度数分布表 5	記述統計 9	平均絶対偏差 (AAD)
GLM 47, 50, 51, 52	グループの平均 25	比率統計 171
ROC 曲線 173	ケースの要約 22	

平行線の検定
順序回帰 76
平行モデル
信頼性分析 163, 164
平方ユークリッド距離
距離 59
平方和 45
GLM 44
べき乗推定値
GLM 1 変数 47, 50, 52
べき乗モデル
曲線推定 80
変換行列
因子分析 101
偏最小二乗回帰 81
変数のエクスポート 83
モデル 83
偏差対比
GLM 46
変数間の差分
OLAP キューブ 31
変数減少法
線型回帰 70
変数増加ステップワイズ法
線型モデル 63
変数増加法
最近傍分析 88
線型回帰 70
変数の重要度
最近傍分析 92
変数の比較
OLAP キューブ 31
偏相関 57
オプション 57
欠損値 57
コマンドの追加機能 58
線型回帰 72
統計 57
0 次相関 57
偏相関プロット
線型回帰 70
変動係数 (COV)
比率統計 171
ホールドアウト・サンプル
最近傍分析 88
棒グラフ
度数分布表 7
報告書
行の集計報告書 157
合計の列 160
合計列 160
列の値の乗算 160
列の値の除算 160
列の集計報告書 159
列の比較 160
報告書の行の集計 157

報告書の行の集計 (続き)
欠損値 159
コマンドの追加機能 161
ソート順 157
データ列 157
表題 159
表題の変数 159
フッター 159
ブレイク間隔 158
ブレイク列 157
ページの制御 158
ページ的设计 159
ページの番号付け 159
列の書式 158
報告書の列の集計 159
欠損値 161
合計列 160
コマンドの追加機能 161
小計 161
総計 161
ページの制御 161
ページ的设计 159
ページの番号付け 161
列の書式 158

[マ行]

幹葉図
探索的 12
メディアン検定
2 個の独立サンプルの検定 147
メモリー割り当て
TwoStep クラスタ分析 108
モデル・ビュー
最近傍分析 90
ノンパラメトリック検定 134
モンテカルロ・シミュレーション 175

[ヤ行]

ユークリッド距離
距離 59
最近傍分析 87
尤度比カイ 2 乗
クロス集計表 16
順序回帰 76
尤度比区間
1 サンプルのノンパラメトリック検定
126
要約 21
オプション 21
統計 22
予測
曲線推定 80

予測値
曲線推定で保存 80
線型回帰での保存 71
予測値の重要度
線型モデル 65
予測区間
曲線推定で保存 80
線型回帰での保存 71

[ラ行]

ラムダ
クロス集計表 16
ラン検定
オプション 143
欠損値 143
コマンドの追加機能 143
統計 143
分割点 142, 143
1 サンプルのノンパラメトリック検定
126, 127
リスク
クロス集計表 16
リンク
順序回帰 76
類似度の測定方法
階層クラスタ分析 117
距離 60
累積度数
順序回帰 76
累積分布関数
シミュレーション 187
列の集計報告書 159
列パーセント
クロス集計表 18
列比率の統計量
クロス集計表 18
連続型フィールド情報
ノンパラメトリック検定 138
ロー
クロス集計表 16
2 変量の相関分析 55
ロジスティック・モデル
曲線推定 80
[ワ行]
歪度
記述統計 9
グループの平均 25
ケースの要約 22
探索的 12
度数分布表 5
報告書の行の集計 158
報告書の列の集計 160

歪度 (続き)

OLAP キューブ 29

歪度の標準誤差

グループの平均 25

ケースの要約 22

OLAP キューブ 29

[数字]

0 次相関

偏相関 57

1 サンプル Kolmogorov-Smirnov 検定
143

オプション 144

欠損値 144

検定分布 143

コマンドの追加機能 144

統計 144

1 サンプルの T 検定 35

オプション 36

欠損値 36

コマンドの追加機能 35, 36

信頼区間 36

1 サンプルのノンパラメトリック検定
125

カイ 2 乗検定 127

フィールド 125

ラン検定 127

2 項検定 126

Kolmogorov-Smirnov 検定 127

2 項検定 141

オプション 142

欠損値 142

コマンドの追加機能 142

統計 142

1 サンプルのノンパラメトリック検定
126

2 分変数 141

2 個の対応サンプルの検定 146

オプション 147

欠損値 147

検定の種類 146

コマンドの追加機能 147

統計 147

2 個の独立サンプルの検定 144

オプション 145

グループ化変数 145

グループの定義 145

欠損値 145

検定の種類 145

コマンドの追加機能 146

統計 145

2 サンプル t 検定

独立したサンプルの t 検定 33

2 次モデル

曲線推定 80

2 変量の相関分析

オプション 55

欠損値 55

コマンドの追加機能 56

相関係数 55

統計 55

有意水準 55

3 次モデル

曲線推定 80

4 分位

度数分布表 5

5% トリム平均

探索的 12

A

Anderson-Rubin 因子得点 104

Andrews のウェーブ推定量

探索的 12

B

Bartlett 因子得点 104

Bartlett の球面性の検定

因子分析 102

Bonferroni

一元配置分散分析 40

GLM 48

Box の M 検定

判別分析 96

Brown-Forsythe 統計量

一元配置分散分析 41

C

Chebychev の距離

距離 59

Clopper-Pearson 区間

1 サンプルのノンパラメトリック検定
126

Cochran 統計量

クロス集計表 16

Cochran の Q

複数の対応サンプルの検定 149

Cochran の Q 検定

対応サンプルのノンパラメトリック検
定 132, 133

Cohen のカッパ

クロス集計表 16

Cook の距離

線型回帰 71

GLM 51

Cox と Snell の R² 乗

順序回帰 76

Cramér の V

クロス集計表 16

Cronbach のアルファ

信頼性分析 163, 164

D

d

クロス集計表 16

Descriptives

コマンドの追加機能 10

DfBeta

線型回帰 71

DfFit(F)

線型回帰 71

Duncan の多重範囲検定

一元配置分散分析 40

GLM 48

Dunnett の C(U)

一元配置分散分析 40

GLM 48

Dunnett の t 検定

一元配置分散分析 40

GLM 48

Dunnett の T3(3)

一元配置分散分析 40

GLM 48

Durbin-Watson の統計量

線型回帰 72

F

F 統計量

線型モデル 63

Fisher の LSD

GLM 48

Fisher の直接法

クロス集計表 16

Friedman の検定

対応サンプルのノンパラメトリック検
定 132

複数の対応サンプルの検定 149

G

Gabriel のペアごとの比較検定

一元配置分散分析 40

GLM 48

Games と Howell のペアごとの比較検定

一元配置分散分析 40

GLM 48

GLM

行列の保存 51

その後の検定 48

プロファイル・プロット 47

GLM (続き)
平方和 44
変数の保存 51
モデル 44
GLM 1 変量 43, 48, 50, 52
オプション 47, 50, 52
診断情報 47, 50, 52
推定周辺平均 47, 50, 52
対比 46
表示 47, 50, 52
Goodman と Kruskal のガンマ
クロス集計表 16
Goodman と Kruskal のタウ
クロス集計表 16
Goodman と Kruskal のラムダ
クロス集計表 16
Guttman のモデル
信頼性分析 163, 164

H

Hampel の M 推定量
探索的 12
Helmert 対比
GLM 46
Hochberg の GT2(H)
一元配置分散分析 40
GLM 48
Hodge-Lehman の推定
対応サンプルのノンパラメトリック検
定 132
Hotelling の T2
信頼性分析 163, 164
Huber の M 推定量
探索的 12

I

ICC. 級内相関係数を参照 164

J

Jeffreys 区間
1 サンプルのノンパラメトリック検定
126

K

k および特徴選択
最近傍分析 93
k 選択
最近傍分析 93
Kendall の W(K)
複数の対応サンプルの検定 149

Kendall の一致係数 (W)
対応サンプルのノンパラメトリック検
定 132
Kendall のタウ b
クロス集計表 16
2 変量の相関分析 55
Kendall のタウ c 16
クロス集計表 16
Kolmogorov-Smirnov 検定
1 サンプルのノンパラメトリック検定
126, 127
Kolmogorov-Smirnov の Z(K)
1 サンプルによる Kolmogorov-Smirnov
検定 143
2 個の独立サンプルの検定 145
KR20
信頼性分析 164
Kruskal のタウ
クロス集計表 16
Kruskal-Wallis の H(K)
2 個の独立サンプルの検定 147
Kuder-Richardson 20 (KR20)
信頼性分析 164

L

Lance と Williams の非類似度測定方法
59
距離 59
Levene の検定
一元配置分散分析 41
探索的 12
GLM 1 変量 47, 50, 52
Lilliefors の検定
探索的 12

M

M 推定量
探索的 12
Mahalanobis の距離
線型回帰 71
判別分析 97
Manhattan 距離
最近傍分析 87
Mann-Whitney の U(M)
2 個の独立サンプルの検定 145
Mantel-Haenszel 統計量
クロス集計表 16
McFadden の R2 乗
順序回帰 76
McNemar 検定
クロス集計表 16
対応サンプルのノンパラメトリック検
定 132, 133

McNemar 検定 (続き)
2 個の対応サンプルの検定 146
Minkowski の距離
距離 59
Moses の極値反応の検定
2 個の独立サンプルの検定 145

N

Nagelkerke の R2 乗
順序回帰 76
Newman-Keuls
GLM 48

O

OLAP キューブ 29
統計 29
表題 32

P

Pearson カイ 2 乗
クロス集計表 16
順序回帰 76
Pearson 残差
順序回帰 76
Pearson の相関
クロス集計表 16
2 変量の相関分析 55
PLUM
順序回帰 75

R

R 2
グループの平均 25
線型回帰 72
R 2 乗の変化量 72
R 2 乗
線型モデル 64
r 相関係数
クロス集計表 16
2 変量の相関分析 55
R 統計量
グループの平均 25
線型回帰 72
Rao の V
判別分析 97
ROC 曲線 173
統計と作図 173
Ryan-Einot-Gabriel-Welsch 多重 F
一元配置分散分析 40
GLM 48

Ryan-Einot-Gabriel-Welsch 多重範囲
一元配置分散分析 40
GLM 48
R-E-G-W F(R)
一元配置分散分析 40
GLM 48
R-E-G-W Q(Q)
一元配置分散分析 40
GLM 48

S

S モデル
曲線推定 80
Scheffé の検定
一元配置分散分析 40
GLM 48
Shapiro-Wilk の検定
探索的 12
Sidak の t 検定
一元配置分散分析 40
GLM 48
Somers の d
クロス集計表 16
Spearman の相関係数
クロス集計表 16
2 変量の相関分析 55
Spearman-Brown の信頼性
信頼性分析 164
Student-Newman-Keuls の検定
一元配置分散分析 40
GLM 48
S- ストレス
多次元スケーリング法 167

T

t 検定
対応のあるサンプルの T 検定 34
独立したサンプルの t 検定 33
1 サンプルの T 検定 35
GLM 1 変量 47, 50, 52
Tamhane の T2(M)
一元配置分散分析 40
GLM 48
Tukey の b 検定
一元配置分散分析 40
GLM 48
Tukey の HSD 検定
一元配置分散分析 40
GLM 48
Tukey の加法性の検定(K)
信頼性分析 163, 164
Tukey のバイウエイ推定量
探索的 12

TwoStep クラスター分析 107
オプション 108
外部ファイルへの保存 109
作業ファイルへの保存 109
統計 109

V

V
クロス集計表 16

W

Wald-Wolfowitz のラン(W)
2 個の独立サンプルの検定 145
Waller-Duncan の t 検定
一元配置分散分析 40
GLM 48
Welch 統計量
一元配置分散分析 41
what-if 分析
シミュレーション 185
Wilcoxon の符号付き順位検定
対応サンプルのノンパラメトリック検定 132
1 サンプルのノンパラメトリック検定 126
2 個の対応サンプルの検定 146
Wilks のラムダ
判別分析 97

Y

Yates の連続修正
クロス集計表 16

Z

z 得点
記述統計 9
変数として保存 9



Printed in Japan