IBM SPSS Statistics Base 19

IBM®

Note: Before using this information and the product it supports, read the general information under Notices p.339 .

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright SPSS Inc. 1989, 2010.

はじめに

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

SPSS Inc., an IBM Company について

SPSS Inc., an IBM Company は、余禄分析ソフトウェアおよびソリューションの世界的なリーディング カンパニーです。当社のデータ収集、統計、モデリング、展開という製品の包括的なポートフォリオによりお客様の考えや意見を収集、見込み客との対話の結果を予測、分析を業務プロセスに組み込むことによりこれらの見解に判断を下すことができます。SPSS Inc. のソリューションにより、分析、IT アーキテクチャ、業務プロセスの収束に焦点を当て、組織全体の相互接続した経営目標に取り組みます。世界中の民間、政府、学術分野のお客様が SPSS Inc. のテクノロジを包括的に利用しています。お客様の関心を呼び、拡大する一方、不正やリスクを軽減、緩和します。2009 年 10 月、SPSS Inc. は IBM 社に買収されました。詳細は http://www.spss.com をご覧ください。

テクニカル サポート

テクニカル サポートのサービスをご利用いただけます。SPSS Inc.製品の使用方法や、対応しているハードウェア環境へのインストールに関して問い合わせることもできます。テクニカル サポートにご連絡するには、http://support.spss.com の SPSS Inc. Web サイトを参照いただくか、http://support.spss.com/default.asp?refpage=contactus.asp の Web サイトでお近くの営業所にお問い合わせください。連絡の際は、所属団体名、サポート契約などを確認できるよう、あらかじめ手元にご用意ください。

カスタマ サービス

製品の発送やお支払いに関してご質問がある場合は、SPSS 社までお問い合わせください (SPSS Japan のホームページは http://www.spss.co.jp です)。お問い合せの際には、シリアル番号をご用意ください。

© Copyright SPSS Inc. 1989, 2010

トレーニング セミナー

SPSS Inc. では一般公開およびオンサイトで トレーニング セミナーを実施しています。セミナーでは実践的な講習を行います。セミナーは主要都市で定期的に開催されます。セミナーの詳細については、SPSS 社までお問い合わせください (SPSS Japan のホームページは http://www.spss.co.jp です)。

追加の出版物

Marija Noruš による『SPSS Statistics: Guide to Data Analysis』、『SPSS Statistics: Statistical Procedures Companion』、『SPSS Statistics: Advanced Statistical Procedures Companion』が Prentice Hall から出版されました。補助的な資料としてご利用いただけます。これらの出版物には、SPSS Statistics Base モジュール、Advanced Statistics モジュール、Regression モジュールの統計的手続きについて記載されています。初めてデータ分析を行う場合、高度なアプリケーションを使用する場合に応じて、この本は IBM® SPSS® Statistics が提供している機能を効率よく使用するための手助けとなります。出版物の内容、サンプルの図表などの詳細は、作者の Web　サイトを参照してください。 http://www.norusis.com

内容

1 コードブック 1

[Codebook Output (コードブック出力)] タブ 3
[Codebook Statistics (コードブック統計)] タブ 6

2 度数 8

度数分布表の統計 10
度数分布表の図表 12
度数分布表の書式 12

3 記述統計 14

記述統計のオプション 16
DESCRIPTIVES コマンドの追加機能 17

4 探索的 18

探索的分析の統計 20
探索的分析の作図 21
探索的分析のべき乗変換 22
探索的分析のオプション. 22
EXAMINE コマンドの追加機能 23

5 クロス集計表 24

クロス集計表の層 26
クロス集計表クラスタ棒グラフ 26
テーブル層に層変数を表示するクロス集計表. 26

クロス集計表の統計 27
クロス集計表のセル表示の設定 30
クロス集計表の表書式. 32

6 要約 33

ケースの要約のオプション 35
ケースの要約の統計 36

7 平均値 38

グループの平均のオプション. 40

8 OLAP キューブ 43

OLAP キューブの統計 45
OLAP キューブの差分 47
OLAP キューブの表題 48

9 t 検定 49

独立したサンプルの t 検定. 49
独立したサンプルの t 検定のグループの定義. 51
独立したサンプルの t 検定のオプション 52
対応のあるサンプルの t 検定 52
対応のあるサンプルの t 検定のオプション 53
1 サンプルの t 検定 54
1 サンプルの t 検定のオプション 55
t 検定コマンドの追加機能. 56

10 一元配置分散分析 57

一元配置分散分析の対比 58

一元配置分散分析のその後の検定 59
一元配置分散分析のオプション 62
ONEWAY コマンドの追加機能 63

11 GLM – 1 変量分散分析 64

GLM モデル 66
項の構築 ロジット モデル 67
平方和 67
GLM の対比 68
対比の種類 69
GLM のプロファイル プロット 70
GLM のその後の比較 71
GLM の保存 73
GLM のオプション 75
UNIANOVA コマンドの追加機能 76

12 2 変量の相関分析 78

2 変量の相関分析のオプション 80
CORRELATIONS および NONPAR CORR コマンドの追加機能 80

13 偏相関分析 82

偏相関のオプション 84
PARTIAL CORR コマンドの追加機能 84

14 距離行列 86

距離行列の非類似度の測定方法 88
距離行列の類似度の測定方法 89
PROXIMITIESL コマンドの追加機能 90

15 線型モデル 91

線型モデルを構成するには 92
目的 93
基本 94
モデルの選択 95
アンサンブル 97
アドバンス 98
モデル 選択 98
モデルの要約 99
自動データ準備 100
予測値の重要度 101
予測対観測 102
残差 103
外れ値 104
効果 105
係数 107
推定平均値 109
モデル構築の要約 110

16 線型回帰 111

線型回帰の変数選択方法 113
線型回帰の規則の設定 114
線型回帰の作図 114
線型回帰: 新変数の保存 116
線型回帰の統計 119
線型回帰のオプション 120
REGRESSION コマンドの追加機能 121

17 順序回帰 122

順序回帰分析のオプション 123
順序回帰分析の出力 125

順序回帰分析の位置モデル 126
項の構築 ロジット モデル 128
順序回帰分析の尺度モデル 128
項の構築 ロジット モデル 128
PLUM コマンドの追加機能 129

18 曲線推定 130

曲線推定のモデル 132
曲線推定の保存 133

19 偏相関最小 2 乗法回帰 134

モデル 136
オプション 137

20 最近隣分析 139

近隣 144
特徴 146
分割 147
保存 149
出力 150
オプション 152
モデル ビュー 153
特徴空間 154
変数の重要度 158
同位 159
最近隣の距離 159
四文位分布図 160
特徴空間エラー ログ 161
k 選択エラー ログ 162
k および特徴選択エラー ログ 163
分類テーブル 164
誤差の集計 164

21 判別分析 165

判別分析: 範囲の定義 167
判別分析: ケースの選択 167
判別分析: 統計 168
判別分析: ステップワイズ法 169
判別分析: 分類 170
判別分析: 保存 172
DISCRIMINANT コマンドの追加機能 172

22 因子分析 173

因子分析のケースの選択 174
因子分析の記述統計 175
因子分析の因子抽出 176
因子分析の回転 178
因子分析の因子得点 179
因子分析オプション 180
FACTOR コマンドの追加機能 180

23 クラスタリングの手続きの選択 181

24 TwoStep クラスタ分析 183

TwoStep クラスタ分析のオプション 186
TwoStep クラスタ分析の出力 188
クラスタ ビューア 189
クラスタ ビューア 190
クラスタ ビューアの操作方法 200
レコードのフィルタリング 202

25 階層クラスタ分析 203

階層クラスタ分析の方法 205

階層クラスタ分析の統計 206
階層クラスタ分析の作図 207
階層クラスタ分析の新変数の保存 207
CLUSTER コマンド シンタックスの追加機能 208

26 大規模ファイルのクラスタ分析 209

大規模ファイルのクラスタ分析の効率 211
大規模ファイルのクラスタ分析の反復 211
大規模ファイルのクラスタ分析の保存 212
大規模ファイルのクラスタ分析のオプション 213
QUICK CLUSTER コマンドの追加機能 213

27 ノンパラメトリック検定 215

1 サンプルのノンパラメトリック検定 215
1 サンプルのノンパラメトリック検定を行うには 216
[フィールド] タブ 217
[設定] タブ 217
独立サンプルのノンパラメトリック検定 224
独立サンプルのノンパラメトリック検定を行うには 224
[フィールド] タブ 225
[設定] タブ 226
対応サンプルのノンパラメトリック検定 229
対応サンプルのノンパラメトリック検定を行うには 230
[フィールド] タブ 231
[設定] タブ 231
モデル ビュー 236
仮説の要約 237
信頼区間の要約 239
1 サンプル検定 239
対応サンプル検定 245
独立サンプル検定 251
カテゴリ フィールド情報 259
連続型フィールド情報 260
ペアごとの比較 261
等質サブセット 262
NPTESTS コマンドの追加機能 263

レガシー ダイアログ 263
カイ 2 乗検定 264
2 項検定 283
ラン検定 285
1サンプルによる Kolmogorov-Smirnov 検定 287
2 個の独立サンプルの検定 289
2 個の対応サンプルの検定 292
複数の独立サンプルの検定 295
複数の対応サンプルの検定 297
2 項検定 283
ラン検定 285
1サンプルによる Kolmogorov-Smirnov 検定 287
2 個の独立サンプルの検定 289
2 個の対応サンプルの検定 292
複数の独立サンプルの検定 295
複数の対応サンプルの検定 297

28 多重回答の分析 300

多重回答グループを定義 301
多重回答の度数表 302
多重回答のクロス集計表 304
多重回答のクロス集計表の範囲の定義 306
多重回答のクロス集計表のオプション 306
MULT RESPONSE コマンドの追加機能 307

29 結果の報告 308

報告書の行の集計 308
集計報告書: 行の集計を取得するには 309
報告書のデータ列/ブレーク列の書式 310
報告書の集計行/最終集計行 311
報告書のブレーク オプション 311
報告書のオプション 312
報告書のレイアウト 313
報告書の表題 314

報告書の列の集計 315
集計報告書: 列の集計を取得するには 315
データ列の集計関数 316
合計列のデータ列集計 317
報告書の列の書式 318
報告書の列の集計でのブレーク列のオプション 318
報告書の列の集計のオプション 319
報告書の列の集計のレイアウト 319
REPORT コマンドの追加機能 319

30 信頼性分析 321

信頼性分析の統計 323
RELIABILITY コマンドの追加機能 325

31 多次元尺度法 326

多次元尺度法のデータの形式 328
多次元尺度法の尺度の作成 328
多次元尺度法のモデル 329
多次元尺度法のオプション 330
ALSCAL コマンドの追加機能 331

32 比率統計量 332

比率統計量 334

33 ROC 曲線 336

ROC 曲線のオプション 338

付録

A　Notices　339

索引　342

章

1

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）と、アクティブ データセット内のすべての、または指定した変数と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、および多重回答グループの場合、要約統計量に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：コードブックはファイルの分割の状態を無視します。これは、欠損値の多重代入用に作成されたファイルの分割も含みます (Missing Values アドオン オプションで利用可能)。

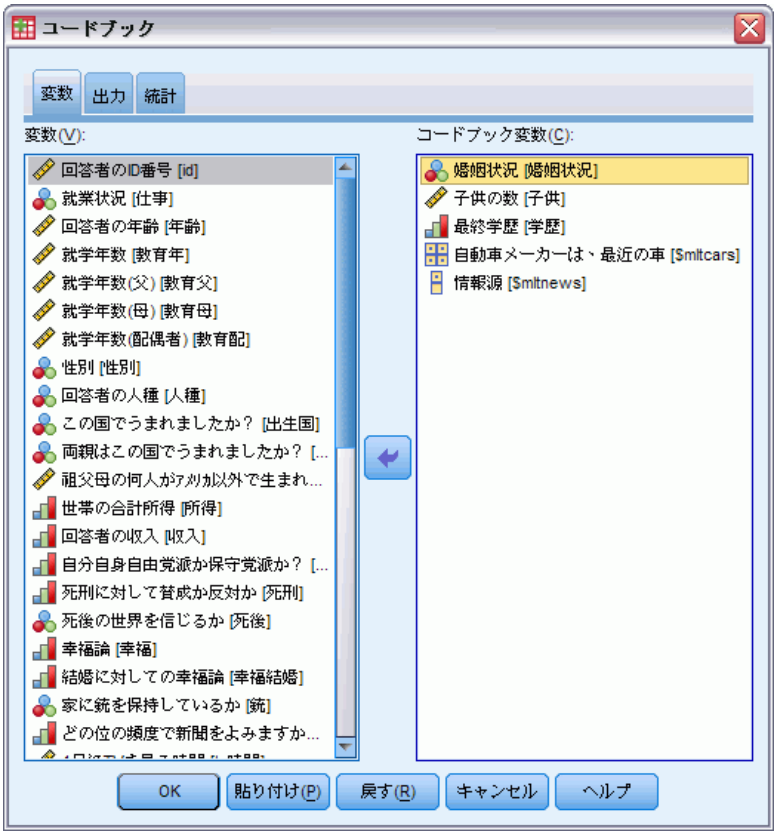
コードブックを取得するには

▶ メニューから次の項目を選択します。
分析(A) > 報告書 > コードブック

▶ ［変数］タブをクリックします。

© Copyright SPSS Inc. 1989, 2010

図 1-1
[Codebook (コードブック)] ダイアログ、[変数] タブ



▶ 1 つ以上の変数または多重回答グループ、あるいはその両方を選択します。

オプションとして、次の選択が可能です。

- 表示される変数情報を制御します。
- 表示される統計量を制御します (またはすべての要約統計量を除外します)。
- 変数と多重回答グループが表示される順序を制御します。
- 表示される要約統計量を変更するために、ソース リストの変数の尺度を変更します。詳細は、 p. 6 [Codebook Statistics (コードブック統計)] タブ を参照してください。

尺度の変更

変数の尺度を一時的に変更できます(文字型変数または多重回答グループの尺度は変更できません。これらは常に名義変数として扱われます)。

▶ ソース リスト内の変数を右クリックします。

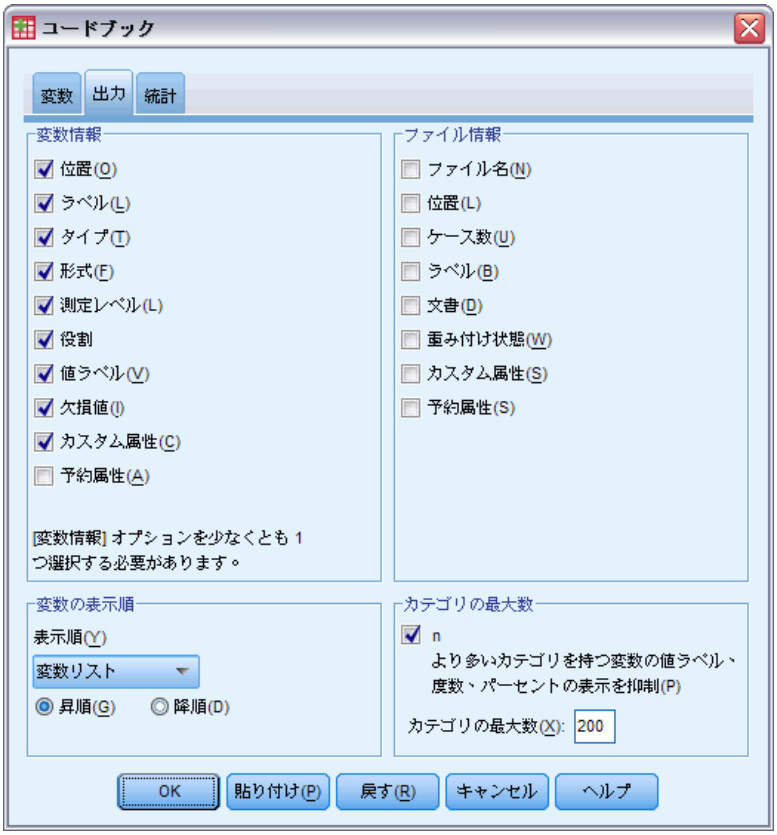
▶ ポップアップ コンテキスト メニューから尺度を選択します。

これにより、尺度が一時的に変更されます。実際、これは数値型変数にのみ役立ちます。文字列変数の尺度は名義変数または順序変数に限られ、いずれもコードブックの手続きで同じように処理されます。

[Codebook Output (コードブック出力)] タブ

[出力] タブは、変数と多重回答グループごと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。

図 1-2
[Codebook (コードブック)] ダイアログ、[出力] タブ



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する記述的なラベル。

型 基本的なデータ型。数値型、文字型、多重回答グループのいずれかになります。

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

測定レベル。使用できる値は、整数、順序、スケール、および不明です。表示される値は、辞書に格納されている尺度であり、[変数] タブのソース変数リストで尺度を変更して一時的な尺度の上書きを指定しても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 :数値型変数の尺度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)、尺度は最初のデータパスまで「不明」の場合があります。

役割。一部のダイアログは、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する記述的なラベル。

- [統計] タブで、[度数] または [パーセント] が選択されていると、[値ラベル] をここで選択しなくても、定義された値レベルが出力に含まれます。
- 多重 2 分変数グループの場合、[値ラベル] は、グループの定義に応じてグループ内の基本変数の変数ラベルか、カウントされた値のラベルになります。

欠損値 ユーザー指定の欠損値。[統計] タブで、[度数] または [パーセント] が選択されていると、ここで [欠損値] を選択しなくても、定義された値レ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 ユーザー指定の変数属性。出力には、各変数に関連するカスタム変数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。

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 ($) で開始します。「@」または「$@」で開始する名前を持つ非表示属性は含まれません。出力には、各変数に関連する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 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] 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、アクティブなデータセットにはファイル名がありません)。

位置SPSS Statistics データ ファイルのディレクトリ (フォルダ) の場所。データセットが 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 アクティブなデータセット内のケースの数。これはケースの総数です。フィルタ条件により要約統計量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 (ある場合) です。

文書 データ ファイル文書のテキスト。

重み付けの状態 重み付けがオンの場合、重み付け変数の名前が表示されます。

カスタム属性 ユーザー指定のカスタム データ ファイル属性。DATAFILE ATTRIBUTE コマンドで定義されるデータ ファイル属性です。

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 ($) で開始します。「@」または「$@」で開始する名前を持つ非表示属性は含まれません。出力には、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、次のいずれかの順序を選択できます。

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。

尺度。尺度順に表示されます。名義型、順序型、スケール型、不明の 4 つのソート グループを作成します。多重回答グループは名義として扱われます。

注 :数値型変数の尺度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)、尺度は最初のデータパスまで「不明」の場合があります。

変数リスト順。[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 並べ替え順序のリストには、ユーザー指定のカスタム変数属性の名前も含まれています。昇順の場合、属性を持たない変数が最初に表示され、次に値の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表示されます。

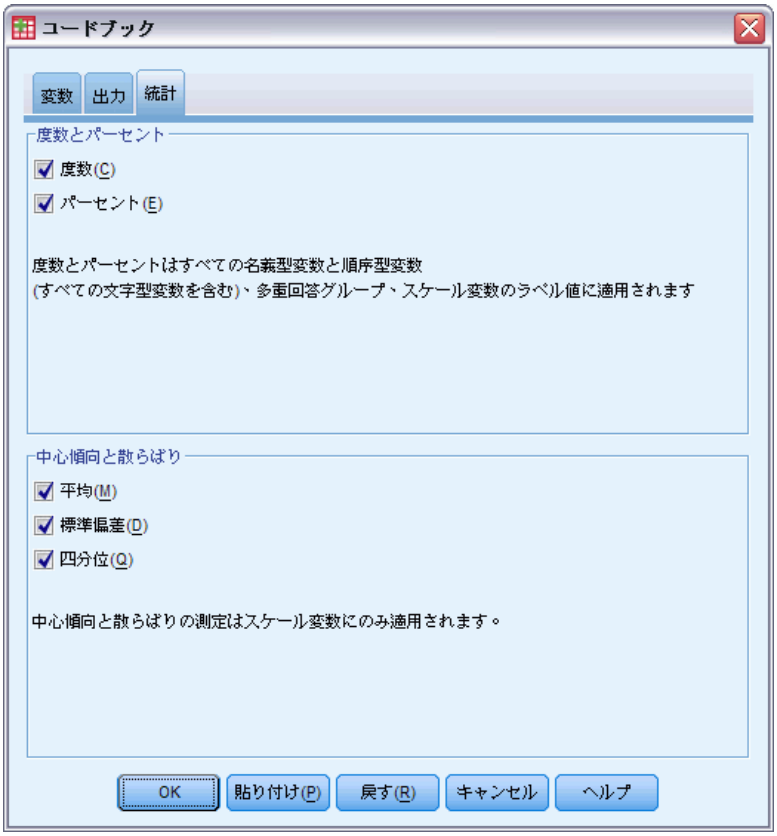
カテゴリの最大数

出力に一意の値ごとの値ラベル、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から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、変数の一意の値の数が 200 を超えるとこの情報の表示は抑制されます。

[Codebook Statistics (コードブック統計)] 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。

図 1-3
[Codebook (コードブック)] ダイアログ、[統計] タブ



度数とパーセント

名義型変数、順序型変数、多重回答グループ、およびスケール変数のラベル付きの値については、次の統計を使用できます。

カウント. 変数のそれぞれの変数値（あるいは変域）を持つケースの数です。

パーセント. 特定の値を持つケースのパーセントです。

中心傾向と散らばり

スケール変数の場合、次の統計を使用できます。

平均. 中心傾向の測定値。観測値の合計をケース数で割った算術平均。

標準偏差. 平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。

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、変数に関連する尺度を一時的に変更できます（したがって、その変数に対して表示される要約統計量が変更されます）。

章
2

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有効な統計と図形を表示します。度数分布表手続きは、データをざっと見るには最適なスタート位置です。

度数レポートと棒グラフでは、昇順または降順で値を配置することもでき、さらにカテゴリを度数別に順序付けて表示することもできます。度数分布表は、変数に多くのカテゴリがあるとき抑制することができます。図表には度数分析（デフォルト）またはパーセントでラベル表示することができます。

例:€ ある企業の顧客は業種によりどのように分布しているでしょうか?出力から、次のことがわかります。顧客の 37.5% は公の省庁関係者で、24.9% は一般企業、28.1% は学術機関、9.4% は医療機関の関係者です。販売収益のような連続した量的データの場合は、平均製品売上高は 3,576 ドルで、標準偏差が 1,078 ドルということがわかります。

統計量と作図。度数、パーセント、累積パーセント、平均値、中央値、最頻値、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。

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

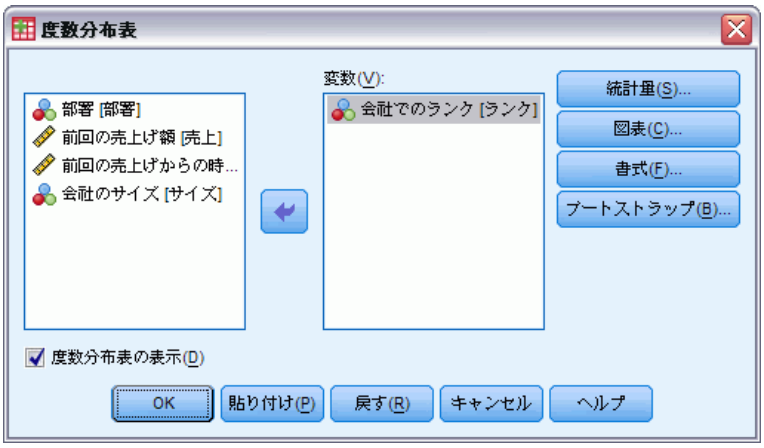
仮定。集計表とパーセントは、どの分布から取り出すデータにも有効ですが、特に順序付けしたカテゴリまたは順序付けしていないカテゴリには、役に立つ統計です。平均値と標準偏差のようなオプションの要約統計のほとんどは、正規理論に基づいていて、分布が対称な量的変数に適しています。中央値、4 分位、およびパーセンタイルのような頑健な統計は正規性の仮定に適合する量的変数にも、適合しない量的変数にも適しています。

度数分布表を取得するには

▶ メニューから次の項目を選択します。

分析(A) > 記述統計 > 度数分布表...

図 2-1
[度数分布表] メーン ダイアログ ボックス



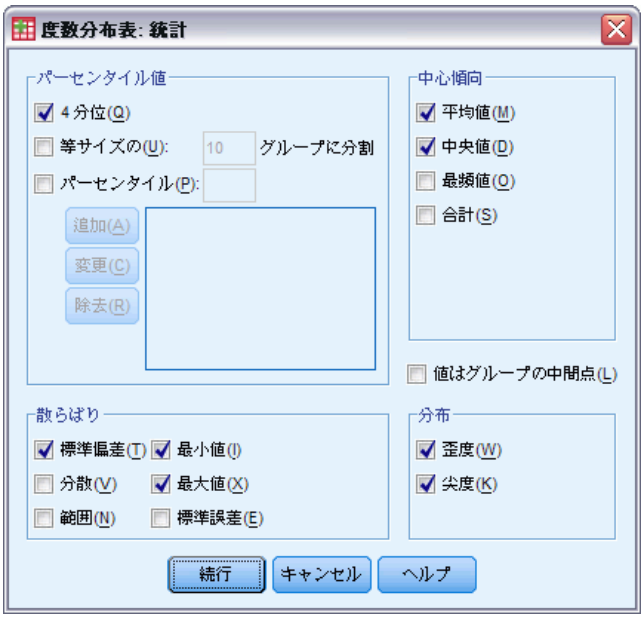
▶ 1 つ以上のカテゴリ変数または量的変数を選択します。

オプションとして、次の選択が可能です。

- 量的変数の記述統計を求めるには [**統計**] をクリックします。
- 棒グラフ、円グラフ、およびヒストグラムを作成するには [**図表**] をクリックします。
- 結果が表示される順序を指定するには [**書式**] をクリックします。

度数分布表の統計

図 2-2
[度数分布表: 統計] ダイアログ ボックス



パーセンタイル値。順序付けしたデータをグループに分割する量的変数の値で、指定されたパーセンテージのケースがその値を上回り、残りのパーセンテージのケースがその値を下回るような値。4 分位（25、50、75 のパーセンタイル）は、観測をサイズの同じ 4 つのグループに分割します。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（たとえば、95 パーセンタイル、観測の 95% が入る値）を指定することもできます。

中心傾向。分布の位置を記述する統計としては、平均値、中央値、最頻値およびすべての値の合計があります。

- **平均.** 中心傾向の測定値。観測値の合計をケース数で割った算術平均。
- **中央値.** ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。
- **最頻値(O).** 最も多く出現する値。複数の値が最高の頻度で出現する場合は、それぞれが最頻値となります。度数分析手続きは、それらのうちの最小の値だけを最頻値として報告します。
- **合計.** 欠損値のないすべてのケースに対する変数の値の合計または全体。

散らばり。データの変動量または広がり量を測定する統計としては、標準偏差、分散、範囲、最小値、最大値、平均の標準誤差があります。

- **標準偏差**. 平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **分散 (信頼性分析)**. 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **範囲**. 数値型変数の最大値と最小値の差。
- **最小値**. 数値型変数がとる最も小さい値。
- **最大値**. 数値型変数の最大値。
- **平均値の標準誤差**. 同一の分布から取り出したサンプル間で平均値がどの程度ばらついているかを測ったもの。観測された平均と仮説された値を比較するために使うことができます (すなわち、差と標準誤差の比率が -2 より小さいか +2 より大きい場合に、2 つの値は異なっていると結論付けることができます)。

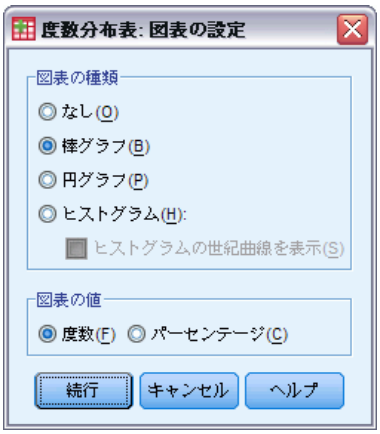
分布。尖度と歪度は、分布の形状や対称を示す統計量です。この統計量は標準誤差とともに表示されます。

- **歪度**. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。
- **尖度**. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、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、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。

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 (たとえば、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。

度数分布表の図表

図 2-3
[度数分布表: 図表の設定] ダイアログ ボックス

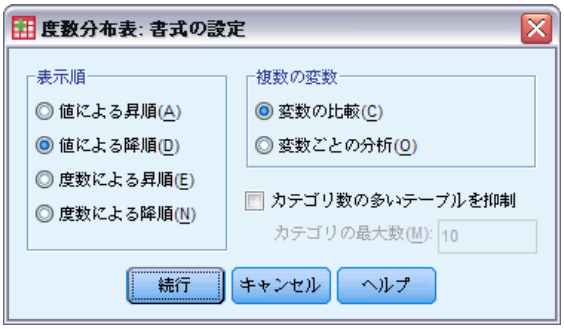


グラフの種類。円グラフは、全体に対する部分の割合を表示します。円グラフの各分割は、1 つのグループ化変数で定義されたグループに対応します。棒グラフは異なる値またはカテゴリの度数を別個の棒として表示するので、カテゴリを視覚的に比較することができます。ヒストグラムにも棒がありますが、こちらは等間隔のスケールに沿ってプロットされます。それぞれの棒の高さは区間内に入る量的変数の値の度数です。ヒストグラムは、分布の行列の形、中央、および広がりを示します。ヒストグラム上に重ね合わせた正規曲線を使用すると、データが正規に分布されているかどうかを判断することができます。

図表の値。棒グラフでは、スケール軸のラベルに度数またはパーセントを使用できます。

度数分布表の書式

図 2-4
[度数分布表: 書式の設定] ダイアログ ボックス



表示順。度数分析は、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、昇順または降順のいずれでも配置することができます。しかし、ヒストグラムまたはパーセンタイルを要求すると、変数が量的であるとみなしその値を昇順で表示します。

複数の変数。複数の変数の統計テーブルを作成する場合、1 つのテーブルにすべての変数を表示すること（[**変数の比較**]）も、変数ごとに統計テーブルを分けて表示すること（[**変数ごとの分析**]）もできます。

カテゴリ数の多いテーブルを抑制。このオプションは、指定した数以上の値がある度数分布表を表示しないようにします。

章
3

記述統計

記述統計手続きにより、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、その平均値の大きさか（平均値の昇順または降順）、アルファベット順、または変数リスト順（デフォルト）に配列することができます。

保存した z 得点はデータ エディタのデータに追加され、図表、データの一覧表表示または分析に使用することができます。変数が別々の単位で記録されている場合（1 人当たりの国内総生産や識字率など）、z 得点変換を行えば、変数の尺度が統一され、変数を一目で比較できるようになります。

例:€ データの各ケースに各店員の 1 日ごとの売上の合計が、毎日の売上を数か月間分集計した形で含まれている場合（たとえば、井上、加藤、山田にそれぞれ 1 ずつ入力）、記述統計手続きにより、各店員の一日の平均売上が計算され、平均値が最も大きい店員を先頭に最も小さい店員まで表示されます。

統計量。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。

データ。誤差、外れ値、および分布の異常を見つけるためにデータをグラフ表示して、修正した数値型変数を使用します。記述統計手続きは、大量のデータ ファイル（千や万単位のケース）を扱う場合に非常に便利です。

仮定。利用可能な統計量（z スコアなど）はそのほとんどが通常の理論に基づいており、対称型の分布を持つ数量変数（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または非対称分布変数は避けます。z 得点の分布は元データと同じ形をしているため、z 得点の計算を行っても、データの問題点が解消されるわけではありません。

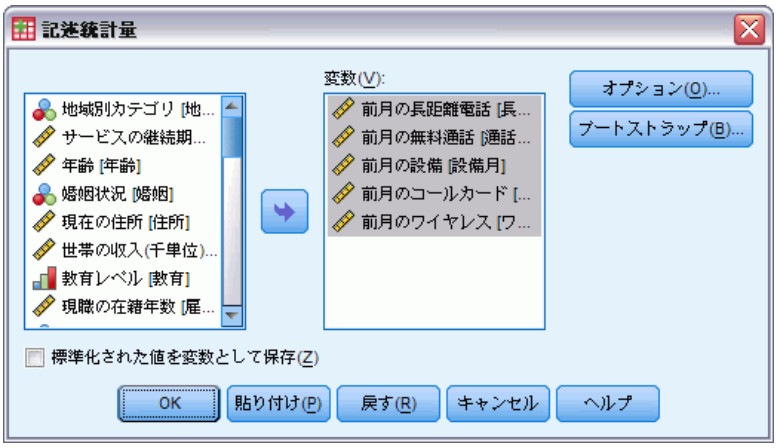
記述統計を行うには

▶ メニューから次の項目を選択します。

分析(A) > 記述統計 > 記述統計...

© Copyright SPSS Inc. 1989, 2010

図 3-1
[記述統計] ダイアログ ボックス



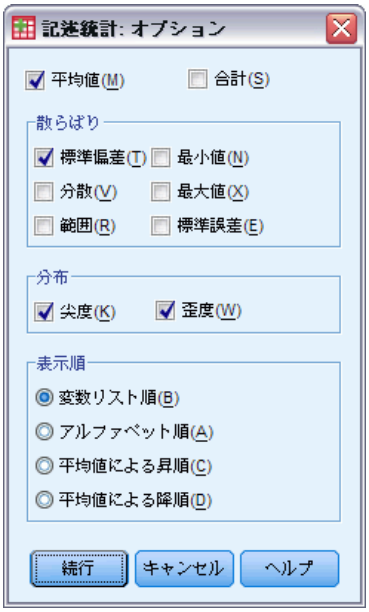
▶ 1 つ以上の変数を選択します。

オプションとして、次の選択が可能です。

- z 得点を新しい変数として保存するには、[**標準化された値を変数として保存**] チェック ボックスをオンにします。
- その他の統計量や表示順を選択する場合は、[**オプション**] をクリックします。

記述統計のオプション

図 3-2
[記述統計: オプション] ダイアログ ボックス



[平均値] と [合計]。 デフォルトでは、平均値、つまり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、および平均値の標準誤差があります。

- **標準偏差.** 平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **分散 (信頼性分析).** 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **範囲.** 数値型変数の最大値と最小値の差。
- **最小値.** 数値型変数がとる最も小さい値。
- **最大値.** 数値型変数の最大値。
- **標準誤差(E).** 同一の分布から取り出したサンプル間で平均値がどの程度ばらついているかを測ったもの。観測された平均と仮説された値を比較するために使うことができます (すなわち、差と標準誤差の比率が -2 より小さいか +2 より大きい場合に、2 つの値は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、分布の形状や対称を表す統計量です。この統計量は標準誤差とともに表示されます。

- **尖度**. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、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、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。
- **歪度**. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

表示順。デフォルトでは、変数は選択した変数リストの順に表示されます。[オプション] 機能で表示順を、アルファベット順、平均値による昇順または降順から選択することができます。

DESCRIPTIVES コマンドの追加機能

コマンド シンタックス言語を使用して、次のことも実行できます。

- 一部の変数に対して（すべての変数ではありません）、標準化された得点（z 得点）を保存（VARIABLES サブコマンドを使用）。
- 標準化された得点を含める新しい変数の名前を指定（VARIABLES サブコマンドを使用）。
- 変数の欠損値のあるケースを分析から除外（MISSING サブコマンドを使用）。
- 平均値だけでなく統計値も加えた順番に変数の表示を並べ替える（SORT サブコマンドを使用）。

複雑なシンタックス情報については、「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。

章
4

探索的

探索的分析手続きは、ケースのすべて、またはケースのグループごとについて、要約統計量と図形表示を作成します。探索的分析手続きを使用するには多くの理由があり、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、および下位母集団（ケースのグループ）間での相違点の特徴付けといったことができます。データ スクリーニングでは、データが異常値、極値、データ内のギャップ、またはその他の特性を持っているかどうかがわかります。データに対して探索的分析を行うと、データ分析に対して考えてい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できます。探索的分析によっては、正規分布を前提とする手法を行う際にデータ変換が必要であることが示されることもあります。または、ノンパラメトリック検定が必要であると判断することもあります。

例:€ 4 種類の強化計画に基づいてネズミ用の迷路学習時間の分布を見てみましょう。4 つのグループそれぞれに対して、時間の分布が近似的に正規分布しているかどうか、かつ 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、学習時間の最大の 5 つのケースおよび最小の 5 つのケースを識別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で示して要約します。

統計量と作図。 平均値、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係数）、パーセンタイル、Huber の M 推定量、Andrews のウェイブ推定量、Hampel の M 推定量、Tukey のバイウェイト推定量、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、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。

データ。 探索的分析手続きは、量的変数（区間または比尺度の測定）に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解するときに使用する）には、妥当な異なった値（カテゴリ）がなければなりません。これらの値は、短い文字型または数値にすることができます。箱ひげ図の外れ値を示すために使用するケースのラベルの変数は、短い文字型、長い文字型（最初の 15 バイト）、または数値にすることができます。

仮定。 データの分布は、対称または正規である必要はありません。

© Copyright SPSS Inc. 1989, 2010

データの探索的分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > 記述統計 > 探索的...

図 4-1
[探索的分析] ダイアログ ボックス

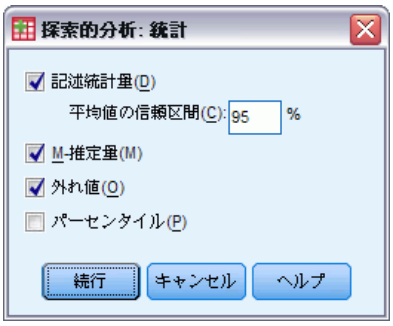
▶ 従属変数を 1 つ以上選択します。

オプションとして、次の選択が可能です。

- 値がケースのグループを定義する 1 つ以上の因子変数を選択できます。
- ケースにラベルを付けるための識別変数を選択できます。
- [**統計**] をクリックすると、M-推定量、外れ値、パーセンタイル、および記述統計量を使用できます。
- [**作図**] をクリックすると、ヒストグラム、正規性の検定とプロット、Levene の統計による水準と広がりの図を使用できます。
- [**オプション**] をクリックすると、欠損値の処理を行えます。

探索的分析の統計

図 4-2
[探索的分析: 統計] ダイアログ ボックス



記述統計。中心傾向と散らばりの測度は、デフォルトで表示されます。中心傾向の測度は分布の位置を表していて、それには平均値、中央値、および 5% トリム平均値が含まれます。散らばりの測度は値の非類似性を表し、標準誤差、分散、標準偏差、最小値、最大値、範囲、および 4 分位範囲が含まれます。記述統計には、分布の形状の測度も含まれ、歪度、および尖度はその標準誤差とともに表示されます。平均値の 95% 水準の信頼区間も表示されますし、任意の信頼水準を指定することもできます。

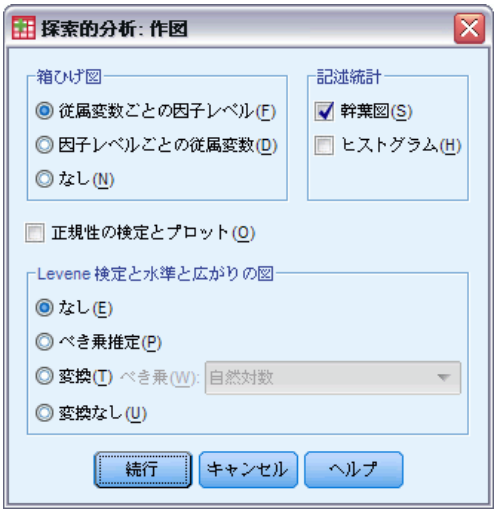
M-推定量。位置を推定するためのサンプル平均値と中央値の頑健な推定量。それぞれの推定量は、ケースに適用する重みで違いがあります。Huber の M-推定量、Andrews のウェイブ推定量、Hampel の M-推定量、および Tukey のバイウェイト推定量が表示されます。

外れ値。ケース ラベルにより 5 つの最大値と 5 つの最小値を表示します。

パーセンタイル。5、10、25、50、75、90、95 番目のパーセンタイルに値を表示します。

探索的分析の作図

図 4-3
[探索的分析: 作図] ダイアログ ボックス



箱ひげ図。複数の従属変数があるときには箱ひげ図の表示を制御します。[**従属変数ごとの因子レベル**] では、従属変数ごとに個別の図表が生成されます。1 つの図表内で、因子変数によって定義されたグループのそれぞれに箱ひげ図が作成されます。[**因子レベルごとの従属変数**] では、因子変数によって定義されたグループごとに個別の図表が生成されます。1 つの図表内で、各従属変数の箱ひげ図を並べて表示します。異なる時に測定した異なる変数が特定の特性を表す場合に、特に便利です。

記述統計量。[記述統計] グループを使用すると、幹葉図とヒストグラムを選択できます。

正規性の検定とプロット。正規確率と傾向化除去正規確率プロットを表示します。正規性を検定するための Lilliefors の有意確率と Kolmogorov-Smirnov の統計量も表示されます。整数以外の重みが指定されると、重みづけされたサンプル サイズが 3 ～ 50 の場合、Shapiro-Wilk 統計量が計算されます。重みがない場合または整数の重みの場合、重みづけされたサンプル サイズが 3 ～ 5,000 であれば、統計量が算出されます。

Levene 検定と水準と広がりの図。水準と広がりの図のデータ変換を制御します。水準と広がりの図のすべてに、回帰直線および等分散性の Levene の頑健な検定が表示されます。変換を選択すると、Levene の検定が変換データに基づいて実行されます。因子変数を選択しないと、水準と広がりの図は作成されません。[**べき乗推定**] では、セル内の分散が等しくなるようにべき乗変換の推定を行うだけではなく、すべてのセルの中央値の自然対数と 4 分位範囲の自然対数のプロットを作成します。水準と広がりの図を使うと、グループ全体で分散を安定させるため（より等しくするため）の変

換のべき乗を決定することができます。[**変換**] では、べき乗推定からの推奨に従って、べき乗の選択肢の 1 つを選択し、さらに変換データのプロットを作成することができます。4 分位範囲および変換データの中央値がプロットされます。[**変換なし**] では、生データのプロットが作成されます。これは、1 乗による変換と等しくなります。

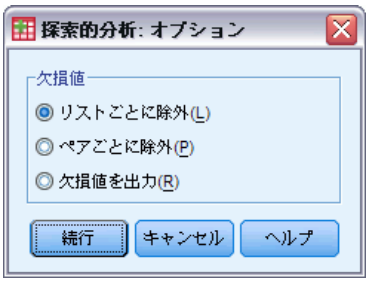
探索的分析のべき乗変換

これは、水準と広がりの図用のべき乗変換です。データを変換するには、変換用のべき乗を選択しなければなりません。次のオプションのどちらかを選択できます。

- **自然対数。**自然対数変換。これはデフォルトです。
- **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。
- **平方。**各データ値が 2 乗されます。
- **立方。**各データ値が 3 乗されます。

探索的分析のオプション

図 4-4
[探索的分析: オプション] ダイアログ ボックス



欠損値。欠損値の処理を管理します。

- **リストごとに除外。**従属または因子変数に対し欠損値のあるケースは、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）内の変数に欠損値がないケースは、そのグループの分析に含まれます。ケースは、他のグループで使用する変数に欠損値がある場合も含みます。
- **欠損値を出力。**因子変数の欠損値は、別のカテゴリとして処理されます。すべての出力は、この追加カテゴリに作成されます。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、欠損として表示されます。

EXAMINE コマンドの追加機能

探索的分析手続きでは、EXAMINE コマンド シンタックスを使用します。コマンド シンタックスを使用すると、次の作業も実行できます。

- 因子変数で定義したグループの出力とプロットに加えて、合計の出力とプロットを要求（TOTAL サブコマンドを使用）。
- 箱ひげ図のグループ用に共通尺度を指定（SCALE サブコマンドを使用）。
- 因子変数の交互作用を指定（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（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（PERCENTILES サブコマンドを使用）。
- 水準と広がりの図に使用する任意のべき乗変換を指定（PLOT サブコマンドを使用）。
- 表示する極値の数を指定（STATISTICS サブコマンドを使用）。
- 位置の M-推定量および頑健推定量のパラメータを指定（MESTIMATORS サブコマンドを使用）。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
5

クロス集計表

クロス集計表手続きでは、2 元表および多次元表を作成し、さらに 2 元表の連関のさまざまな検定および測定を行うことができます。表の構造およびカテゴリが順序付けされているかどうかにより、使用される検定および測定が決定します。

クロス集計の統計および連関の測定が計算される対象は、2 次元表に限られます。行、列、および層（制御変数）を指定すると、[クロス集計表] 手続きは、層の各値（または 2 個以上の制御変数の値の組み合わせ）に対して 1 パネルの連関の統計量および測定方法を作成します。たとえば、性別が 人生観（人生は楽しいか、日常的か、またはつまらないか）と 結婚経験（はい、いいえ）のクロス表の層である場合、女性についての 2 次元表の結果は男性についてのものとは別々に計算され、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業務において、中小企業の顧客は、大企業より収益を生む可能性が高いでしょうか。クロス集計より、この分野では、大企業（従業員 2,500 人以上）の収益が低く、中小企業（従業員 500 人未満）の収益が高いことを示しています。

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、Cramer 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、および Cochran 統計量と Mantel-Haenszel 統計量、および列比率の統計量。

データ。 各表変数のカテゴリを定義するには、数値型変数または短い文字型変数（8 バイト以下）の値を使います。たとえば、性別の場合、1 と 2 または 男性と 女性のようにデータをコード化できます。

仮定。 統計値および測定方法のなかには、順序付けされたカテゴリ（順位データ）または量的な値（区間データまたは比率データ）であることを前提とするものがあります（詳細は統計値についての章を参照）。また、表変数に順序付けされたカテゴリでないもの（名義データ）があるときに有効なものもあります。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となります。

© Copyright SPSS Inc. 1989, 2010

注：順序変数は、カテゴリを表す数値型コード（たとえば、1 = 低、2 = 中、3 = 高）または文字型値です。ただし、カテゴリの本当の順序を反映させるため、文字型値のアルファベット順を仮定しています。たとえば、値が 低、中、高である文字型変数では、カテゴリの順序は 高、低、中と解釈されますが、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には、数値型コードを使用した方が信頼性が高いといえます。

クロス集計を行うには

▶ メニューから次の項目を選択します。

分析(A) > 記述統計 > クロス集計表...

図 5-1
[クロス集計表] ダイアログ ボックス

▶ 1 つ以上の行変数および列変数を選択します。

オプションとして、次の選択が可能です。

- 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の統計値および連関の測定方法を選択する。
- [**セル**] をクリックして、[観測] および [期待] 値、[パーセンテージ]、[残差] を選択する。
- [**書式**] をクリックして、カテゴリの表示順を制御する。

クロス集計表の層

1 つ以上の層変数を選択すると、各層変数（制御変数）のカテゴリごとに別々のクロス集計が作成されます。たとえば、行変数が 1 つ、列変数が 1 つで、2 つのカテゴリを持つ層変数が 1 つある場合、この層変数の各カテゴリに対してそれぞれ2 次元表が 1 つ得られます。別の制御変数の層を作成するには、[次] をクリックします。第 1 層変数のカテゴリと第 2 層変数との組み合わせに対してそれぞれ表が作成されます。統計量および連関の測定が要求されている場合は、2 次元表にだけ適用されます。

クロス集計表クラスタ棒グラフ

クラスタ棒グラフの表示。クラスタ棒グラフを使用して、データをケースのグループに要約できます。［行］の一覧で指定した変数の各値を表す棒のクラスタが 1 つあります。各クラスタ内で棒を定義する変数は、［列］の一覧で指定した変数です。この変数の各値に対してそれぞれ違う色が付けられたり、または違うパターンの棒のグループが 1 つあります。列または行で 1 つ以上の変数を指定した場合は、行と列の変数のそれぞれの組み合わせに対してクラスタ棒グラフが作成されます。

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。層変数（対照変数）をクロス集計表のテーブル層として表示することを選択できます。層変数のカテゴリについてドリルダウンを可能にするほか、行変数および列変数の全体の統計を表示するビューを作成できます。

データ ファイル demo.sav () を使用する例を次に示します。

- 行変数として［世帯全体の収入カテゴリ（千ドル）［収入カテゴリ]]、列変数として［携帯情報端末［携帯端末]]、そして層変数として［教育のレベル［教育]] を選択します。
- [テーブル層に層変数を表示] を選択します。
- ［セル表示］サブダイアログで [列] を選択します。
- 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。

図 5-2
テーブル層に層変数を表示するクロス集計表

世帯全体の収入カテゴリ (千ドル) と 携帯情報端末 と 教育のレベル のクロス表

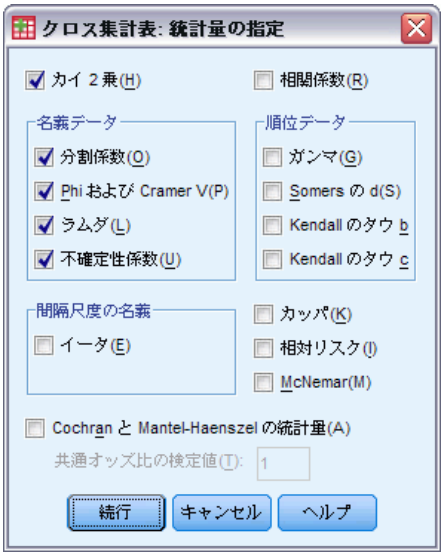
教育のレベル 大学

	統計量		携帯情報端末 持っていない	携帯情報端末 持っている	合計
世帯全体の収入カテゴリ (千ドル)	25未満	度数	146	50	196
		携帯情報端末 の %	15.8%	11.6%	14.5%
	25 - 49	度数	335	155	490
		携帯情報端末 の %	36.3%	35.9%	36.2%
	50 - 74	度数	187	72	259
		携帯情報端末 の %	20.3%	16.7%	19.1%
	75以上	度数	255	155	410
		携帯情報端末 の %	27.6%	35.9%	30.3%
合計		度数	923	432	1355
		携帯情報端末 の %	100.0%	100.0%	100.0%

クロス集計表の選択したビューには、学歴が大学卒の回答者の統計が表示されます。

クロス集計表の統計

図 5-3
[クロス集計表: 統計量の指定] ダイアログ ボックス



カイ 2 乗。 2 つの行と 2 つの列を持つ表の場合は、[カイ 2 乗] チェック ボックスをオンにして、Pearson のカイ 2 乗、尤度比カイ 2 乗、Fisher 直接法、および 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または欠損列を持たない表で、セル

の期待度数が 5 より少ないときには、Fisher の直接法が計算されます。他のすべての 2 × 2 表に対して、Yates の補正によるカイ 2 乗が計算されます。任意の数の行と列で構成される表の場合は、[**カイ 2 乗**] チェックボックスをオンにして、Pearson のカイ 2 乗および尤度比のカイ 2 乗を計算します。表変数がどちらも量的変数である場合は、**カイ 2 乗**は線型と線型による連関検定になります。

相関係数。行および列の両方に順位の値が含まれている表の場合は、[**相関係数**] により Spearman の相関係数、ロー（数値データのみ）が得られます。Spearman のローは、ランク順間の関連度です。表変数（因子）がどちらも量的変数である場合は、[**相関**] により Pearson の相関係数、r、変数間の線型による連関の測定が得られます。

名義。名義データ（カトリック、プロテスタント、ユダヤなどの非順序尺度）の場合には、[**ファイ**]（係数）と [**Cramer の V**]、[**分割係数**]、[**ラムダ**]（対称ならびに非対称ラムダおよび Goodman と Kruskal のタウ）、[**不確定性係数**] を選択できます。

- **分割係数(O).** カイ 2 乗に基づく連関度。値は 0 から 1 の範囲になります。値 0 は行変数と列変数の間に関連がないことを示し、1 に近い値は変数間に強い関連があることを示します。可能な最大値は、表の行と列の数によって決まります。
- **標準偏差 (グラフの集計関数).** ファイは、カイ 2 乗に基づく関連度で、カイ 2 乗統計量をサンプル サイズで割り、その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。
- **ラムダ.** 独立変数の値が従属変数の値を予測しようとするときの、誤差の減少を反映した予測連関指数です。値 1 は、独立変数が従属変数を完全に予測することを意味します。値 0 は、独立変数が従属変数の予測に役立たないことを意味します。
- **不確定性係数(U).** 1 つの変数値がその他の変数値の予測に使われるとき、誤差内の予測連関指数を示す連関度。たとえば、値 0.83 は一方の変数がもう一方の変数の値を予測する際に、誤差を 83% 減らすという情報を示します。プログラムは、不確実性係数の対称版と非対称版の両方を計算します。

順序。行および列の両方に順序の値が含まれている場合は、[**ガンマ**]（2 次元表には 0 次で、3 次元～10 次元表には条件付）、[**Kendall のタウ b**]、および [**Kendall のタウ c**] を選択します。行カテゴリから列カテゴリを予測する場合は、[**Somers の d**] を選択します。

- **ガンマ.** 2 つの順序変数間の対称な連関度で、-1 から 1 の範囲を取ります。絶対値 1 に近い値は、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、関係が弱いかまったくないことを示します。2 次元表では、0 次ガンマが表示されます。3 次元表から n 次元表では、条件付きのガンマが表示されます。

- **Somers の d.** 2 つの順序変数間の関連度で、-1 から 1 の範囲を取ります。絶対値 1 に近い値は、2 つの変数間に強い関係があることを示します。0 に近い値は、変数間の関係が弱いかまったくないことを示します。Somers の d は、独立変数の同順位でないペアの数の差を取ることによって、ガンマ係数を非対称に拡張したものです。この統計量の対称版も計算されます。
- **Kendall のタウ b.** 同順位を考慮する順序変数の、ノンパラメトリックな相関度。係数の符号はその関係の方向を示し、その絶対値は、より強い関係を示すより大きな絶対値によって強度を示します。-1 から 1 までの値を取りますが、-1 または +1 が得られるのは平方表からだけです。
- **Kendall のタウ c.** 同順位を無視する順序変数の、ノンパラメトリックな関連度。係数の符号はその関係の方向を示し、その絶対値は、より強い関係を示すより大きな絶対値によって強度を示します。-1 から 1 までの値を取りますが、-1 または +1 が得られるのは平方表からだけです。

間隔尺度の名義。一方の変数がカテゴリ変数で、他の一方が量的変数であるとき、[**イータ**] を選択します。カテゴリ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(E).** 0 から 1 までの範囲の連関度です。0 は行変数および列変数の間に連関がないことを示し、1 に近い値は連関の度合いが高いことを示します。イータは、間隔尺度で測定された従属変数（収入など）とカテゴリの少ない独立変数（性別など）の分析に適しています。2 つのイータ値が計算されます。1 つは、行変数を間隔変数として扱うもので、もう 1 つは、列変数を間隔変数として扱うものです。

カッパ（因子分析）. Cohen のカッパは、2 つの評価者が同じ対象を評価するときに、それらの評価の一致の度合を測定します。値 1 は完全な一致を表します。値 0 は、偶然以外の一致がないことを表します。カッパは、両方の変数が同じカテゴリ値を持ち、かつ両方の変数が同じ数のカテゴリを持つ表でだけ利用できます。

リスク(R). 2 x 2 表の場合は、ある因子の存在とあるイベントの発生の関連の強さの測定値になります。統計量の信頼区間に 1 が含まれる場合は、その因子がそのイベントに関連していると仮定することはできません。因子の発生がまれなときには、オッズ比を推定値または相対リスクとして使うことができます。

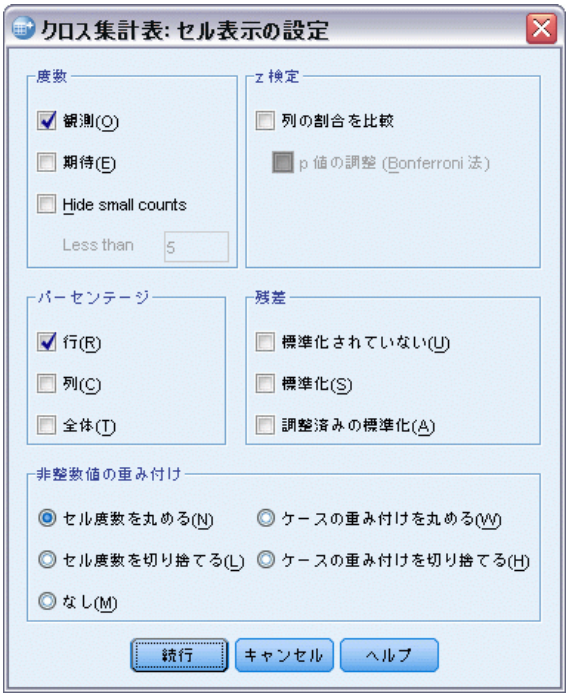
McNemar(M). 関連する 2 つの 2 分変数に対するノンパラメトリック検定。カイ 2 乗分布を使って応答の変化を検定します。設計の前後での実験的介入により、反応の変化の検出に便利です。大きな平方表では、McNemar-Bowker 対称検定が報告されます。

Cochran 統計量と Mantel-Haenszel 統計量. Cochran と Mantel-Haenszel 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われま

す。他の統計量は層ごとに計算されますが、Cochran と Mantel-Haenszel 統計量は、すべての層に対して一度に計算されます。

クロス集計表のセル表示の設定

図 5-4
[クロス集計表: セル表示の設定] ダイアログ ボックス



カイ 2 乗検定で有意になるデータのパターンを見つけやすくするために、[クロス集計表] 手続きにより、期待度数および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) が表示されます。クロス表の各セルは、度数、パーセント、および残差から任意に選択して組み合わせることができます。

度数。行変数および列変数が相互に独立している場合は、実際の観測ケース数および期待ケース数。

列比率の比較。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 (指定した行) を示します。有意差は、APA スタイル形式がサブスクリプト文字を使用しているクロス集計表で表示され、0.05 の有意確率で計算されます。

- **p 値の調整 (Bonferroni 法)。**列比率のペアごとの比較で Bonferroni 補正を使用し、複数の比較が行われる事実に対する観測された有意確立を調整します。

パーセンテージ。パーセンテージは、行全体または列全体を合計できます。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使えます。

残差。標準化されていない残差により、観測値および期待値の間の差分が示されます。標準化された残差および調整済みの標準化された残差も選択できます。

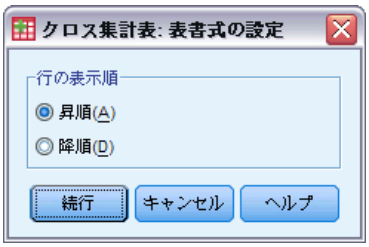
- **標準化されていない(N).** 観測値と期待値の差。期待値は、2 つの変数の間に関係がないと想定した場合に期待されるセルのケース数です。正の残差は、行変数と列変数が独立であると想定される場合に、セルの期待されたケース数より実際のケース数が多いことを示します。
- **標準化(A).** 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼ばれ、平均は 0 で、標準偏差 1 になります。
- **調整済みの標準化(A).** セルの残差（観測度数 - 期待度数）をその標準誤差の推定値で割った値。結果の標準化残差は、標準偏差を単位として平均より上または下で表されます。

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ので、通常は整数値になります。ただし、小数値変数を含む重み付け変数（たとえば 1.25）によって現在データ ファイルが重み付けられている場合、セル度数は小数値になります。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テーブル表示と統計計算の両方に使用することができます。

- **丸めセル度数.** ケース重みはそのままで使用されますが、セルの累積重みはすべての統計量が計算される前に丸められます。
- **切り捨てセル度数.** ケース重みはそのままで使用されますが、セルの累積重みはすべての統計量が計算される前に切り捨てられます。
- **丸めケース重み.** ケース重みは、使用前に丸められます。
- **切り捨てケース重み.** ケース重みは、使用前に切り捨てられます。
- **なし(M).** ケースの重み付けがそのまま使用され、小数値のセル度数が使用されます。ただし、正確検定統計量（[正確確率検定] オプションが利用できるときのみ）が要求された場合、セルの累積重みは、正確確率検定統計量が計算される前に、丸められるか、切り捨てられます。

クロス集計表の表書式

図 5-5
[クロス集計表: 表書式の設定] ダイアログ ボックス



行は、行変数の値の昇順または降順に整列できます。

章
6

要約

ケースの要約手続きで、1 つ以上のグループ化変数から成る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できます。カテゴリ全体での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をリスト表示したり、または表示を抑制できます。大きいデータセットでは、最初の n ケースだけを一覧表示できます。

例:€ 地区および顧客業種別の平均製品売上高はどのようなものでしょうか。他地区に比べて西部地区の平均売上高はわずかに高いことがわかります。西部地区の企業顧客からは最高の平均売上高が得られているからです。

統計量 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、総数のパーセント、グループ変数での合計のパーセント、グループ変数でのケース数のパーセント、幾何平均、調和平均。

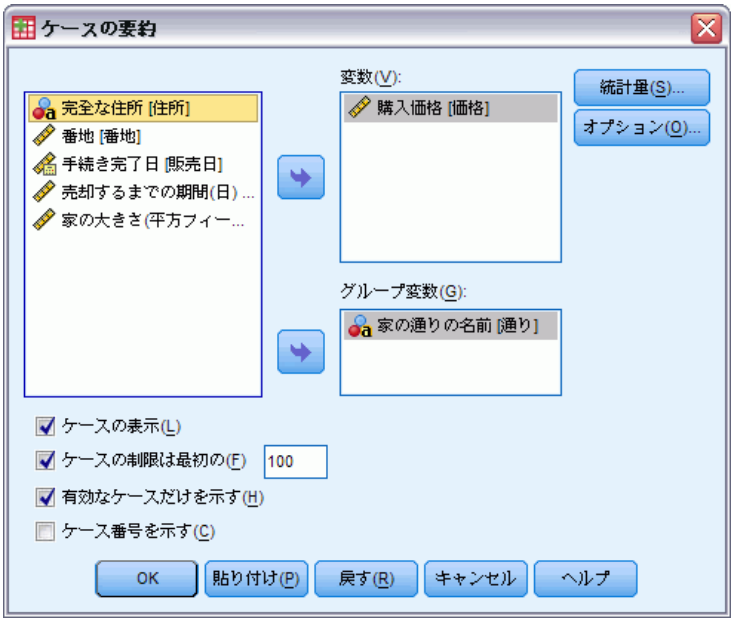
データ。 グループ化変数は、値が数値型または文字型のカテゴリ変数です。カテゴリの数はかなり少ないものでなくてはなりません。他の変数は、ランク付けできるものでなくてはなりません。

仮定。 オプションのサブグループ統計量の中には、平均値や標準偏差などのように、通常の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。中央値や範囲などの頑健な統計は、正規性の仮定に合う場合と合わない場合のある量的変数に適しています。

ケースの要約を行うには

▶ メニューから次の項目を選択します。
分析(A) > 報告書 > ケースの要約...

図 6-1
[ケースの要約] ダイアログ ボックス



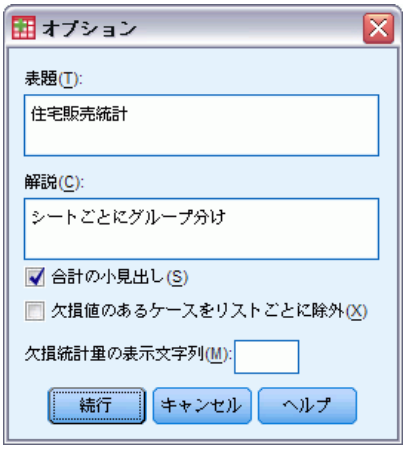
▶ 1 つ以上の変数を選択します。

オプションとして、次の選択が可能です。

- 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、出力表題を変更、解説を追加、または欠損値を持つケースを除外する。
- [**統計**] をクリックして、オプションの統計量にアクセスする。
- [**ケースの表示**] を選択して、各サブグループのケースを一覧で表示する。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**] の値を増やすか減らしたり、その項目を選択解除してすべてのケースを表示したりできます。

ケースの要約のオプション

図 6-2
[オプション] ダイアログ ボックス

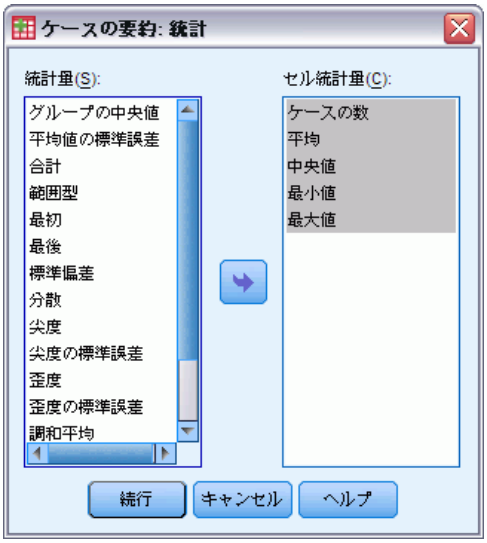


要約を使用して、出力の表題を変更したり、出力テーブルの下に表示される解説を追加することができます。また、表題や解説で改行を指定することもできます。テキスト中の任意の位置で ¥n と入力すると、その位置で改行されます。

欠損値のケースは、ピリオドまたはアスタリスクを付けて出力することが望ましい場合がしばしばあります。欠損値の発生時に表示させたい文字、語句、またはコードを入力します。それ以外の場合は、出力時に欠損値の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

図 6-3
[ケースの要約: 統計] ダイアログ ボックス



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、合計のパーセント、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、幾何平均および調和平均を 1 つまたは複数選択できます。統計量が [セル統計量] リストに現れる順序は、出力に表示される順序です。カテゴリ全体で各変数に対する要約統計量も表示されます。

最初. データ ファイルで発生した最初のデータ値を表示します。

幾何平均. データの値の積の n 乗根です。n はケースの数を表します。

グループ中央値. グループに対してコード化されたデータについて計算された中央値。たとえば、年齢データで、30 代の各値が 35 にコード化され、40 代の各値が 45 にコード化されるとすると、グループ中央値はコード化されたデータから計算された中央値になります。

調和平均. グループ間でサンプルの大きさが等しくないときに、グループの大きさの平均を予測するために使われます。調和平均は、サンプルの合計をサンプルの大きさの逆数の和で割ったものです。

尖度. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、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、正規分布に対して厚くなります。

負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。

最後. データ ファイルで発生した最後のデータ値を表示します。

最大値. 数値型変数の最大値。

平均. 中心傾向の測定値。観測値の合計をケース数で割った算術平均。

中央値. ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最小値. 数値型変数がとる最も小さい値。

ケースの数. ケース（観測値やレコード）の数。

総数のパーセント. 各カテゴリのケースの総数のパーセントです。

総合計のパーセント. 各カテゴリの総和のパーセントです。

範囲. 数値型変数の最大値と最小値の差。

歪度. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

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うことができます（すなわち、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）。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します（箱型の一様分布の裾のようになります）。

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うことができます（すなわち、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）。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

合計. 欠損値のないすべてのケースに対する変数の値の合計または全体。

分散（信頼性分析）. 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

章
7

平均値

［平均の比較］手続きでは、1 つ以上の独立変数のカテゴリ内の従属変数を対象に、サブグループの平均と関連した 1 変量の統計量を計算します。オプションとして、一元配置分散分析、イータ、線型性の検定などを得ることができます。

例:€ 異なる 3 種類の食用油について、その平均脂肪吸収量を測定し、一元配置分散分析によって平均値が異なるかどうかを確かめます。

統計量。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、総数のパーセント、グループ変数での合計のパーセント、グループ変数でのケース数のパーセント、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。

データ。従属変数は量的で、独立変数はカテゴリです。カテゴリ変数の値は、数値または文字型です。

仮定。オプションのサブグループ統計量の中には、平均値や標準偏差などのように、通常の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、正規性の仮定に適合する量的変数にも、適合しない量的変数にも適しています。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。また各グループは、分散が等しい母集団からのものであると仮定します。この仮定を検定するには、［一元配置分散分析］手続きで利用できる 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

▶ メニューから次の項目を選択します。
分析(A) > 平均の比較 > グループの平均...

図 7-1
[平均値] ダイアログ ボックス

▶ 従属変数を 1 つ以上選択します。

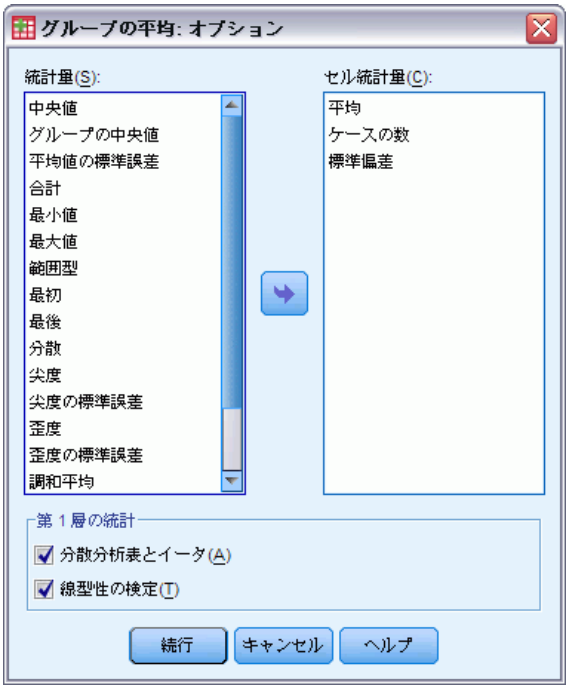
▶ 以下の方法の 1 つを使って、カテゴリ独立変数を選択します。

- 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。各層は、サンプルをさらに細分割します。層 1 と層 2 にある独立変数が 1 つの場合、その結果は 1 つのクロス表に表示され、独立変数ごとに別の表が作成されることはありません。

▶ 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

図 7-2
[グループの平均: オプション] ダイアログ ボックス



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、合計のパーセント、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、幾何平均および調和平均を 1 つまたは複数選択できます。サブグループの統計量は、表示順を変更できます。[セル統計量] リストに表示される統計量の表示順は、出力の表示順です。カテゴリ全体で各変数に対する要約統計量も表示されます。

最初. データ ファイルで発生した最初のデータ値を表示します。

幾何平均. データの値の積の n 乗根です。n はケースの数を表します。

グループ中央値. グループに対してコード化されたデータについて計算された中央値。たとえば、年齢データで、30 代の各値が 35 にコード化され、40 代の各値が 45 にコード化されるとすると、グループ中央値はコード化されたデータから計算された中央値になります。

調和平均. グループ間でサンプルの大きさが等しくないときに、グループの大きさの平均を予測するために使われます。調和平均は、サンプルの合計をサンプルの大きさの逆数の和で割ったものです。

尖度. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、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、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。

最後. データ ファイルで発生した最後のデータ値を表示します。

最大値. 数値型変数の最大値。

平均. 中心傾向の測定値。観測値の合計をケース数で割った算術平均。

中央値. ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最小値. 数値型変数がとる最も小さい値。

ケースの数. ケース（観測値やレコード）の数。

総数のパーセント. 各カテゴリのケースの総数のパーセントです。

総和のパーセント. 各カテゴリの総和のパーセントです。

範囲. 数値型変数の最大値と最小値の差。

歪度. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

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うことができます（すなわち、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）。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します（箱型の一様分布の裾のようになります）。

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うことができます（すなわち、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）。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

合計. 欠損値のないすべてのケースに対する変数の値の合計または全体。

分散 (信頼性分析). 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

第 1 層の統計

分散分析表とイータ(A). 一元配置分散分析表を表示して、最初の層内にある各独立変数にイータおよびイータ 2 乗（連関度）を表示します。

線型性の検定(T). F 比、R および R2 乗以外に線型および非線型成分と関連する平方和、自由度および平均平方を計算します。独立変数が短い文字型の場合、線型性の検定は計算されません。

章

8

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 グループ化変数のカテゴリ内で、連続集計変数の合計、平均値、その他の 1 変量の統計量を計算します。各グループ化変数のカテゴリごとに別々の層がテーブルに作成されます。

例:€ 地域別売上げの合計と平均、地域内の製品群など。

統計量 選択対象の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、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、幾何平均と調和平均があります。

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変数）で、グループ化変数はカテゴリ変数です。カテゴリ変数の値は、数値または文字型です。

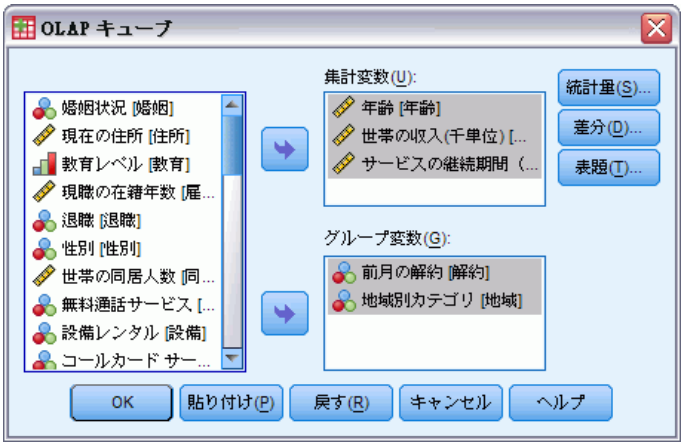
仮定。 オプションのサブグループ統計量の中には、平均値や標準偏差などのように、通常の理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。頑健な統計量（中央値や範囲など）は、正規性の仮定に適合する量的変数にも、適合しない量的変数にも適しています。

OLAP キューブを行うには

▶ メニューから次の項目を選択します。
分析(A) > 報告書 > OLAP キューブ...

© Copyright SPSS Inc. 1989, 2010

図 8-1
[OLAP キューブ] ダイアログ ボックス



▶ 連続型の集計変数を 1 つ以上選択します。

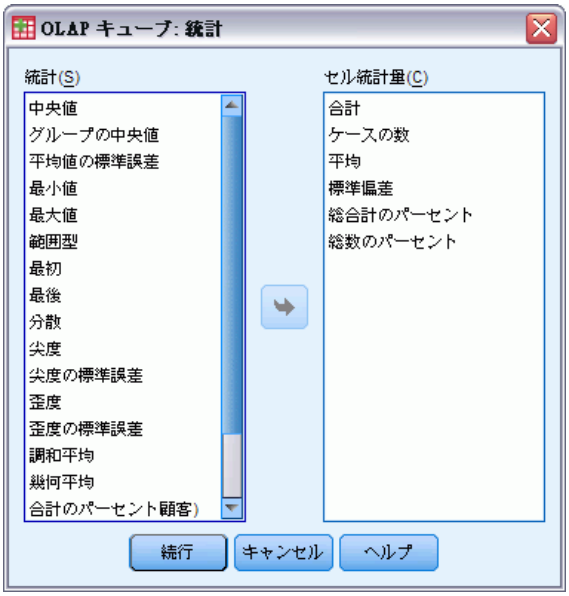
▶ カテゴリ型のグループ化変数を 1 つ以上選択します。

次のオプションが選択できます。

- 別の要約統計量を選択する（[**統計**] をクリックします）。要約統計量を選択する前に、1 つ以上のグループ変数を選択します。
- 変数のペア、および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**] をクリックします）。
- カスタム テーブルの表題を作成する（[**表題**] をクリックします）。

OLAP キューブの統計

図 8-2
[OLAP キューブ: 統計] ダイアログ ボックス



各グループ化変数のカテゴリ内の集計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、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、幾何平均と調和平均を 1 つまたは複数選択できます。

サブグループの統計量は、表示順を変更できます。[セル統計量] リストに表示される統計量の表示順は、出力の表示順です。カテゴリ全体で各変数に対する要約統計量も表示されます。

最初. データ ファイルで発生した最初のデータ値を表示します。

幾何平均. データの値の積の n 乗根です。n はケースの数を表します。

グループ中央値. グループに対してコード化されたデータについて計算された中央値。たとえば、年齢データで、30 代の各値が 35 にコード化され、40 代の各値が 45 にコード化されるとすると、グループ中央値はコード化されたデータから計算された中央値になります。

調和平均. グループ間でサンプルの大きさが等しくないときに、グループの大きさの平均を予測するために使われます。調和平均は、サンプルの合計をサンプルの大きさの逆数の和で割ったものです。

尖度. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、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、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。

最後. データ ファイルで発生した最後のデータ値を表示します。

最大値. 数値型変数の最大値。

平均. 中心傾向の測定値。観測値の合計をケース数で割った算術平均。

中央値. ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最小値. 数値型変数がとる最も小さい値。

ケースの数. ケース（観測値やレコード）の数。

グループ変数でのケース数のパーセント. 他のグループ化変数のカテゴリ内における、指定されたグループ化変数のケース数のパーセント。グループ化変数が 1 つしかない場合は、ケースの総数のパーセントと同じになります。

合計のパーセント. 他のグループ化変数のカテゴリ内における、指定されたグループ化変数の合計のパーセント。グループ化変数が 1 つしかない場合は、総和のパーセントと同じになります。

総数のパーセント. 各カテゴリのケースの総数のパーセントです。

総合計のパーセント. 各カテゴリの総和のパーセントです。

範囲. 数値型変数の最大値と最小値の差。

歪度. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

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うことができます（すなわち、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）。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します（箱型の一様分布の裾のようになります）。

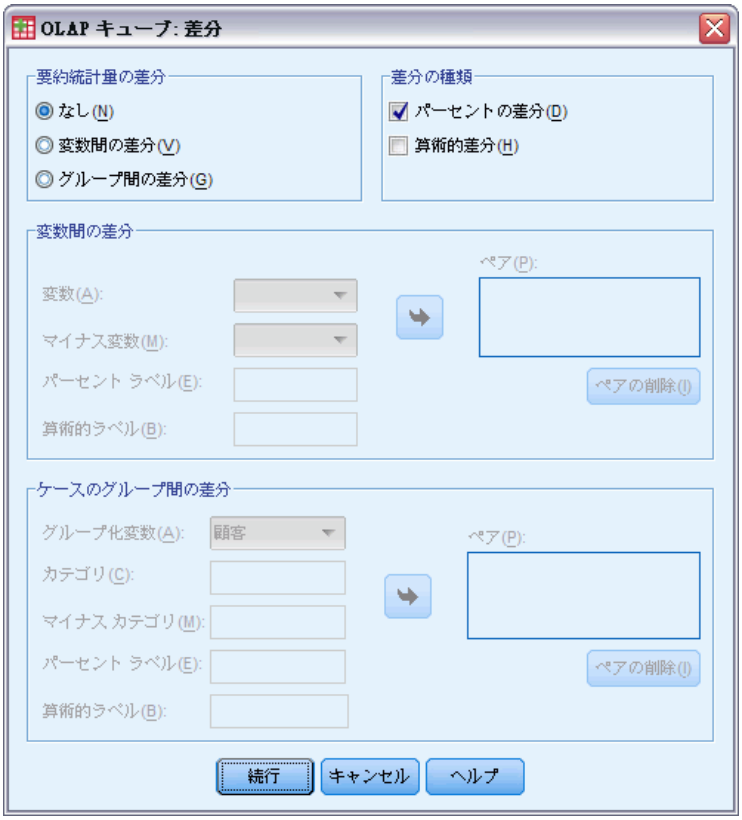
歪度の標準誤差. 標準誤差に対する歪度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

合計. 欠損値のないすべてのケースに対する変数の値の合計または全体。

分散 (信頼性分析). 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

OLAP キューブの差分

図 8-3
[OLAP キューブ: 差分] ダイアログ ボックス



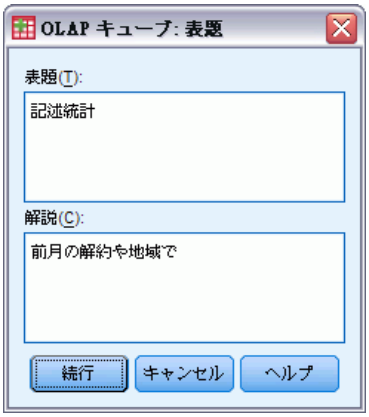
このダイアログ ボックスでは、集計変数間、またはグループ化変数によって定義されるグループ間の算術的な差分や割合を計算できます。差分は、[OLAP キューブ: 統計] ダイアログ ボックスで選択したすべての測定方法で計算されます。

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 1 番目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前に、メイン ダイアログ ボックスで少なくとも 2 つ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（[マイナス カテゴリ]）の要約統計量の値が 1 番目のカテゴリ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] 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前に、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

図 8-4
[OLAP キューブ: 表題] ダイアログ ボックス



出力の表題を変更したり、出力テーブルの下に解説を加えたりできます。また、表題や解説の改行を制御することもできます。テキスト内の必要な位置で「¥n」と入力すると、その位置で改行されます。

章

9

t 検定

t 検定には、次の 3 種類があります。

[独立したサンプルの t 検定] (2 サンプル t 検定)。 2 グループのケースについて 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 の検定の他に、分散が等しい場合および分散が等しくない場合の 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 t 検定] (独立 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。この検定はまた、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のための検定です。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、および 95% 信頼区間が含まれています。

1 サンプルの t 検定。 既知の値または仮説値を 1 変数の平均値と比較します。検定変数の記述統計量は、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の差の 95% 信頼区間は、デフォルト出力に含まれています。

独立したサンプルの t 検定

[独立サンプルの t 検定] 手続きでは、2 つのグループのケースによる平均値を比較します。理想的には、この検定の場合、被検者を 2 個のグループに対して無作為に割り当て、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) によるものとします。男性および女性の平均収入を比較する場合、この検定は当てはまりません。ある被検者が男性または女性に無作為に割り当てられていません。そのような状況においては、他の要素における差が平均値の有意差を隠したり、または大きくしないようにします。平均収入の差は、教育水準などの要素により影響を受けるかもしれません (性別だけによる影響は受けないでしょう)。

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指定します。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効力があると考えられる新薬を投与します。被検者の 2 か月間の治療後に、2 サンプルの t 検定を使用して、偽薬グループおよび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、1 つのグループに所属します。

© Copyright SPSS Inc. 1989, 2010

統計量 各変数に対して: サンプル サイズ、平均値、標準偏差、および平均値の標準誤差。平均値の差に対して: 平均値、標準誤差、および信頼区間 (信頼水準を指定できます)。検定: 等分散性の Levene 検定、および 2 つの母平均の差のプールされた分散ならびに等分散でないときの t 検定。

データ。 任意の量的変数の値がデータ ファイルの 1 つの列にあります。この手続きでは、2 つの値を持ったグループ化変数を使用して当該ケースを 2 つのグループに分けます。グループ化変数は、数値型変数 (1 や 2、または 6.25 や 12.5) または短い文字型変数 (はいと いいえなど) のどちらかです。また別の方法として、量的変数 (年齢など) を使用して、分割値を指定することにより、ケースを 2 つのグループに分割することもできます (分割値を 21 にすると 年齢は 21 未満のグループと 21 以上のグループに分割されます)。

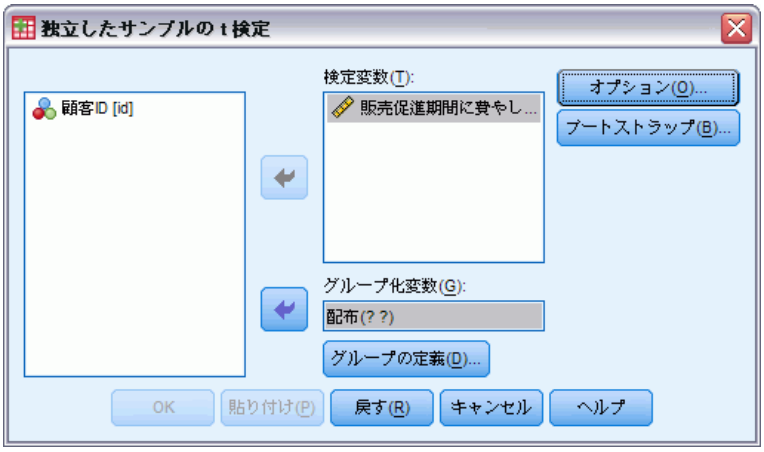
仮定。 等分散の t 検定の場合、観測値は、同じ母集団分散を持つ正規分布からの独立した無作為サンプルでなくてはなりません。等分散でない t 検定の場合、観測値は、正規分布からの独立した無作為サンプルでなくてはなりません。2 サンプルの t 検定は、正規性からの逸脱に対して非常に頑健です。分布をグラフとしてチェックするときには、分布が対称的であり、さらに外れ値が皆無であることを確認します。

独立したサンプルの t 検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 平均の比較 > 独立したサンプルの t 検定...

図 9-1
[独立したサンプルの t 検定] ダイアログ ボックス



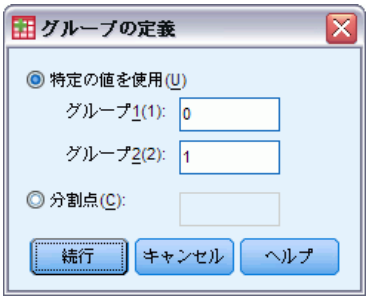
▶ 1 つ以上の量的検定変数を選択します。t 検定が変数ごとに行われます。

▶ 変数のリストから 1 つのグループ化変数を選択して [グループ化変数] ボックスに移動し、[**グループの定義**] をクリックして比較するグループの 2 つのコードを指定します。

▶ オプションとして、[**オプション**] をクリックして、欠損データの扱いと信頼区間の水準を指定します。

独立したサンプルの t 検定のグループの定義

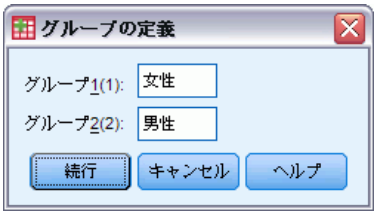
図 9-2
数値型変数の [グループの定義] ダイアログ ボックス



数値型のグループ化変数では、2 つの値または分割値を指定して、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 に値を入力して、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。数字は整数でなくてもかまいません（たとえば、6.25 や 12.5 でも有効です）。
- **分割値。**グループ化変数の値を 2 つのグループに分割する数字を入力します。分割値未満の値のケースが一方のグループを形成し、分割値以上の値のケースは他の一方のグループを形成します。

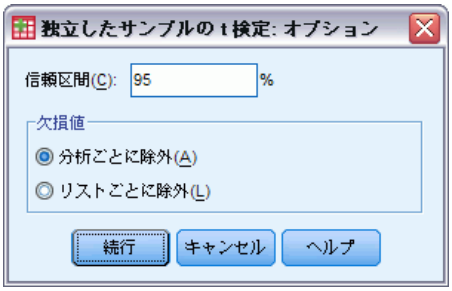
図 9-3
文字型変数の [グループの定義] ダイアログ ボックス



短い文字型のグループ化変数では、[グループ 1] と [グループ 2] に 1 つずつ文字列（はいと いいえなど）を入力します。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

図 9-4
[独立したサンプルの t 検定: オプション] ダイアログ ボックス



信頼区間。デフォルトでは、平均値の差の 95% 信頼区間を表示します。1 から 99 の範囲の数値を入力して、別の信頼水準を表示することもできます。

欠損値。複数の変数を検定する場合で、1 つ以上の変数に対して欠損している場合、どのケースを含める（または除外する）かを指示できます。

- **分析ごとに除外。**各 t 検定では、検定する変数のデータが有効なケースをすべて使用します。したがって、サンプル サイズが検定ごとに変化します。
- **リストごとに除外。**各 t 検定では、要求された t 検定において使用されるすべての変数に対して有効なデータをもつケースのみを使用します。したがって、サンプル サイズが検定を通じて一定になります。

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] 手続きでは、1 つのグループの 2 つの変数の平均を比較します。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 と異なるかどうかを検定します。

例:€ 高血圧に関する調査で、すべての患者を調査の開始時に測定し、治療後に再度測定します。このように、各被験者には 2 つの測定値があり、多くの場合、測定前と測定後と呼ばれます。この検定が使用されるもう 1 つの計画は、一致するペアの研究またはケース コントロール研究です。ここで、データ ファイルの各記録には患者および一致する対照被験者の回答が含まれます。血圧の調査では患者と対照被験者を、年齢で (75 歳の患者と 75 歳の対照グループメンバーを) 一致させることもできます。

統計量 各変数に対して: 平均値、サンプル サイズ、標準偏差、および平均値の標準誤差。変数の各ペアについて: 相関係数、平均値の差の平均、t 検定、平均値の差に対する信頼区間 (信頼水準は指定可能)。標準誤差と平均値の差の標準誤差。

データ。各対応のある検定で 2 つの数量変数（区間尺度または比例尺度）を指定します。一致するペアの研究またはケース コントロール研究に対して、各検定の被験者と対になる対照被験者に対する応答は、データ ファイル内の同じケースにある必要があります。

仮定。ペアに対する観測は同じ条件で行われる必要があります。また、平均値の差は正規分布していなくてはなりません。各変数の分散は、等しい場合と異なる場合があります。

対応のあるサンプルの t 検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 平均の比較 > 対応のあるサンプルの t 検定...

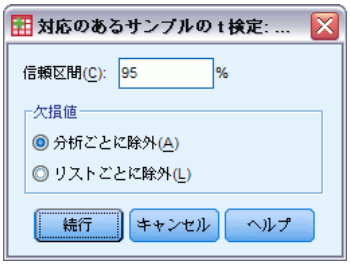
図 9-5
[対応のあるサンプルの t 検定] ダイアログ ボックス

▶ 1 つ以上の変数ペアを選択

▶ オプションとして、[**オプション**] をクリックして、欠損データの扱いと信頼区間の水準を指定します。

対応のあるサンプルの t 検定のオプション

図 9-6
[対応のあるサンプルの t 検定: オプション] ダイアログ ボックス



信頼区間。デフォルトでは、平均値の差の 95% 信頼区間を表示します。1 から 99 の範囲の数値を入力して、別の信頼水準を表示することもできます。

欠損値。複数の変数を検定する場合で、1 つ以上の変数に対して欠損している場合、どのケースを含める（または除外する）かを指示できます。

- **分析ごとに除外。**各 t 検定では、検定するペアの変数のデータが有効なケースをすべて使用します。したがって、サンプル サイズが検定ごとに変化します。
- **リストごとに除外。**各 t 検定では、検定する変数のすべてのペアに対して有効なデータを持つケースだけを使用します。したがって、サンプル サイズが検定を通じて一定になります。

1 サンプルの t 検定

[1 サンプルの t 検定] 手続きでは、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調べます。

例。ある研究者が、学生グループの平均の IQ スコアが 100 と異なるかどうかを検定したい場合があります。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% の確信度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チェックすることができます。

統計量。各検定変数に対して: 平均値、標準偏差、および平均値の標準誤差。各データ値と仮説検定値の差の平均、それが 0 であることを検定する t 検定とその信頼区間 (信頼水準を指定できます)。

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、量的変数と仮説検定値を選択します。

仮定。この検定では、データが正規分布しているものと仮定していますが、データが正規性から逸脱している場合にも、かなり頑健です。

1 サンプルの t 検定を行うには

▶ メニューから次の項目を選択します。

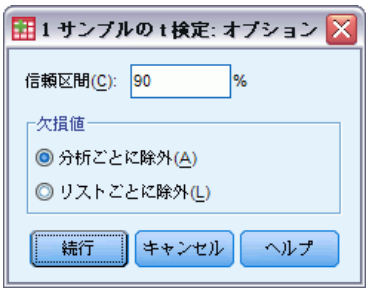
分析(A) > 平均の比較 > 1 サンプルの t 検定...

図 9-7
[1 サンプルの t 検定] ダイアログ ボックス

- 仮説値に対して検定する 1 つ以上の変数を選択します。
- 各サンプルの平均値と比較する数値を［検定値］ボックスに入力します。
- オプションとして、[**オプション**] をクリックして、欠損データの扱いと信頼区間の水準を指定します。

1 サンプルの t 検定のオプション

図 9-8
[1 サンプルの t 検定オプション] ダイアログ ボックス



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% 信頼区間を表示します。1 から 99 の範囲の数値を入力して、別の信頼水準を表示することもできます。

欠損値。複数の変数を検定する場合で、1 つ以上の変数に対して欠損している場合、どのケースを含める（または除外する）かを指示できます。

- **分析ごとに除外。**各 t 検定では、検定する変数のデータが有効なケースをすべて使用します。したがって、サンプル サイズが検定ごとに変化します。
- **リストごとに除外。**各 t 検定では、要求された t 検定において使用されるすべての変数に対して有効なデータをもつケースのみを使用します。したがって、サンプル サイズが検定を通じて一定になります。

t 検定コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 1 サンプルと独立サンプルの両 t 検定を、単一コマンドで実行することができます。
- リストに記載されている各変数に対する変数の検定を、対応のある t 検定で行なうことができます (PAIRS サブコマンド使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
10

一元配置分散分析

［一元配置分散分析］手続きは、量的従属変数に対して一元配置分散分析を一因子（独立）変数別に作成します。分散分析は、いくつかの平均値は等しいという仮説を検定するときに使用します。この手法は、2 サンプルの t 検定の拡張です。

平均値間に差があることを判断する以外に、どの平均値が違うかを知ることもできます。平均値を比較する検定には、事前対比、その後の範囲検定の 2 種類があります。対比は実験を実行する前に設定された検定であり、その後の検定は実験が実行された後で実行される検定です。カテゴリ全体の傾向を検定することもできます。

例:€ ドーナツは料理をするとかなりの量で脂肪を吸収します。実験では、ピーナツ オイル、コーン オイルおよびラードの 3 種類の脂肪を使用しています。ピーナッツ油とコーン油は不飽和の脂肪で、ラードは飽和脂肪です。吸収された脂肪の量が使用した脂肪のタイプに関係しているどうかを判断することができ、さらに事前対比を設定すると脂肪の吸収量が飽和脂肪と不飽和脂肪では違うかどうかを判断することもできます。

統計量 各グループの、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、および平均値の 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、その後の範囲検定と多重比較: Bonferroni、Sidak、Tukey の HSD、Hochberg の GT2、Gabriel、Dunnett、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、Dunnett の T3、Games-Howell、Dunnett の C、Duncan の多重範囲検定、Student-Newman-Keuls (S-N-K)、Tukey の b、Waller-Duncan、Scheffé、および最小有意差。

データ。因子変数の値は整数でなければならず、さらに従属変数は量的変数（区間尺度の測定）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。データは対称であるべきですが、分散分析は正規性からの逸脱に対し頑健です。グループは分散の等しい母集団から発生してい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。

© Copyright SPSS Inc. 1989, 2010

一元配置分散分析を行うには

▶ メニューから次の項目を選択します。
分析(A) > 平均の比較 > 一元配置分散分析...

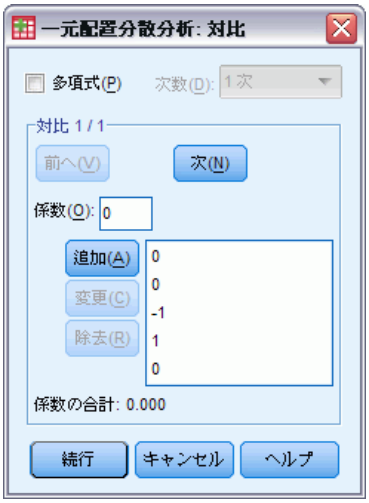
図 10-1
[一元配置分散分析] ダイアログ ボックス

▶ 従属変数を 1 つ以上選択します。

▶ 変数のリストから 1 つの独立因子変数を選択して、[因子] ボックスに移動します。

一元配置分散分析の対比

図 10-2
[一元配置分散分析: 対比] ダイアログ ボックス



グループ間平方和をトレンド成分に分割したり、事前対比を指定できます。

多項式。グループ間平方和を傾向成分に分割します。因子変数の順序付けされた水準全体で従属変数の傾向を検定することができます。たとえば、得た最高の順位の水準で給与の線形傾向（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 次の直交多項式を選択できます。

係数。t 統計量で検定す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）について係数を入力し、入力するごとに [**追加**] をクリックします。それぞれの新しい値は係数リストの下部に加えられます。対比の追加グループを指定するには、[**次**] をクリックします。[**次**] と [**前**] を使うと、対比の設定の間を移動できます。

係数の次数は、因子変数のカテゴリ値の昇順に対応するので重要です。リスト上の最初の係数は因子変数の最も小さいグループの値に対応し、最後の係数は最も大きい値に対応します。たとえば、因子変数のカテゴリが 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 番目のグループと対比させます。ほとんどの場合、係数は合計して 0 になる必要があります。合計が 0 にならないセットも使用できますが、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

図 10-3
[一元配置分散分析: その後の多重比較] ダイアログ ボックス

平均値の間に差があることが判明した後は、その後の範囲検定とペアごとの多重比較により、どの平均値が相異しているのかを決めることができます。範囲検定は、互いに平均値に差がない等質サブグループを識別しま

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、5% 水準で有意な差があるグループの平均値には星印を付けます。

等分散が仮定されている

Tukey の HSD 検定、Hochberg の GT2、Gabriel の検定、および Scheffé の検定は、多重比較検定と範囲検定の両方です。利用できるその他の範囲検定としては、Tukey の b 検定、S-N-K (Student-Newman-Keuls)、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)、および Waller-Duncan の方法があります。利用できる多重比較検定は、Bonferroni、Tukey の HSD 検定、Sidak、Gabriel、Hochberg、Dunnett の方法、Scheffé、および LSD (最小有意差) があります。

- **最小有意差(L).** t 検定を使って、グループ平均間のすべてのペアごとの比較を実行します。多重比較の誤差率は調整されません。
- **Bonferroni の方法.** t 検定を使ってグループ平均のペアごとの比較を行いますが、実験ごとの誤差率を総検定数で割った値に各検定の誤差率を設定することによって、全体の誤差率を制御します。したがって、有意確率は、多重比較がなされているとして調整されます。
- **Sidak(D).** t 検定に基づいたペアごとの多重比較検定。Sidak の方法は、多重比較の有意確率を調整して、Bonferroni の方法より厳しい限界を設定します。
- **Scheffe の検定.** 平均値の可能なペアごと組み合せに対して、ペアごとの同時比較を実行します。F 分布を使います。ペアごとの比較だけでなく、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うこともできます。
- **R-E-G-W の F(R).** F 検定に基づいた Ryan-Einot-Gabriel-Welsch の多重ステップダウン手続き。
- **R-E-G-W の Q(Q).** スチューデント化された範囲に基づいた Ryan-Einot-Gabriel-Welsch の多重ステップダウン手続き。
- **Student-Newman-Keuls(S).** スチューデント化された範囲の分布を使って、平均値間のすべてのペアごとの比較を行います。サンプル サイズが等しい場合は、ステップワイズ法手続きを使って等質サブセット内の平均値のペアも比較します。平均値は最も高いものから最も低いものへと順序付けられ、極値の差が最初に検定されます。
- **Tukey(T).** スチューデント化された範囲の統計量を使ってグループ間のすべてのペアごとの比較を行います。すべてのペアごとの比較の集合に対する誤差率に、実験ごとの誤差率を設定します。
- **Tukey の b.** スチューデント化された範囲の分布を使ってグループ間のペアごとに比較を行います。臨界値は、Tukey の HSD 検定と Student-Newman-Klaus 検定に対応する値の平均です。

- **Duncan(D).** Student Newman Keuls 検定で使用される次数と同一の、ステップごとの比較次数を使ってペアごとに比較しますが、個々の検定に対する誤差率ではなく、検定の集合の誤差率に対する保護水準を設定します。スチューデント化された範囲の統計量を使います。
- **Hochberg の GT2(H).** スチューデント化された最大絶対値を使う多重比較と範囲検定。Tukey の HSD 検定に似ています。
- **Gabriel(G).** スチューデント化された最大法を使用したペアごとの比較検定。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。セルの大きさ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(W).** T 統計量に基づいた多重比較検定。ベイズの方法を使用します。
- **Dunnett.** 1 組の実験群を単一の対照（コントロール）と比較するペアごとの多重比較 T 検定。 最後のカテゴリがデフォルトの対照カテゴリとなります。代わりに、最初のカテゴリを選択できます。**[両側]** は、因子の任意のレベルの平均値（対照カテゴリを除く）が対照カテゴリの平均値と等しくないことを検定します。**[<対照カテゴリ]** は、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ことを検定します。**>[対照カテゴリ]** は、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ことを検定します。

等分散が仮定されていない

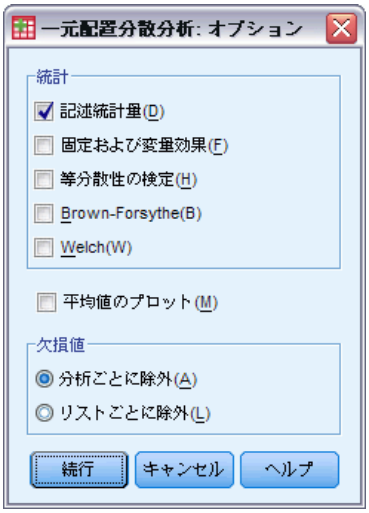
等分散を仮定しない多重比較検定は、Tamhane の T2、Dunnett の T3、Games-Howell、および Dunnett の C です。

- **Tamhane の T2(M).** t 検定に基づいた保守的なペアごとの比較検定。この検定は、分散が等しくないときに適しています。
- **Dunnett の T3(3).** スチューデント化された最大法に基づいたペアごとの比較検定。この検定は、分散が等しくないときに適しています。
- **Games-Howell(A).** 時々公平なペアごとの比較検定。この検定は、分散が等しくないときに適しています。
- **Dunnett の C(U).** スチューデント化された範囲に基づいたペアごとの比較検定。この検定は、分散が等しくないときに適しています。

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] メニューの **[テーブル プロパティ]** を選択します）の [全般] タブで **[空白の行と列を隠す]** のチェックを外すと、その後の検定からの出力を解釈し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

図 10-4
[一元配置分散分析: オプション] ダイアログ ボックス



統計量 次の 1 つ以上を選択します。

- **記述統計量。**各グループに対してケースの数、平均値、標準偏差、平均値の標準誤差、最小値、最大値、および各従属変数の 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、および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。**グループの分散の等質性を検定するため Levene の統計を計算します。この検定は、正規性の仮定に依存しません。
- **Brown-Forsythe。**グループ平均値の等質性を検定するため Brown-Forsythe の統計を計算します。等分散仮定が適用できない場合は、この統計量が F 統計量よりも適しています。
- **Welch。**グループ平均値の等質性を検定するため Welch の統計を計算します。等分散仮定が適用できない場合は、この統計量が F 統計量よりも適しています。

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。欠損値の処理を管理します。

- **分析ごとに除外。**ある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、その分析で使用されません。さらに、因子変数で指定し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に取り込まれた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、影響はありません。

ONEWAY コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 固定効果および変量効果の統計を取得します。固定効果モデルの標準偏差、標準誤差、95% 信頼区間。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（STATISTICS=EFFECTS 使用）。
- 最小有意差、Bonferroni、Duncan、Scheffé の各多重比較検定に対するアルファ レベルを指定します（RANGES サブコマンド使用）。
- 平均行列、標準偏差と度数の書き込み、または、平均行列、度数、プールされた分散とプールされた分散の自由度の読み込みを行ないます。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取得します（MATRIX サブコマンド使用）。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
11

GLM - 1 変量分散分析

[GLM - 1 変量分散分析] 手続きは、1 つの従属変数に対する回帰分析や分散分析を 1 つ以上の因子や変数を使用して行う手続きです。因子変数により、母集団をいくつかのグループに分けます。この [一般的な線型モデル] 手続きを使用すると、1 つの従属変数をさまざまに分けたグループの平均値に対する他の変数の効果について、帰無仮説を検定できます。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、その一部は任意にすることもできます。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ができます。回帰分析では、独立 (予測) 変数は共変量として指定します。

検定は、釣り合い型モデルと不釣り合い型モデルの両方に対して実行できます。モデル内の各セルに含まれているケース数が等しければ、その計画は釣り合っています。[GLM - 1 変量分散分析] 手続きでは、仮説の検定の他に、パラメータの推定値を生成します。

仮説を検定する際は、一般的に用いられている事前対比を利用できます。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、特定平均値間の差分を評価できます。推定周辺平均から、モデルに含まれるセルの予測平均値を推定できるとともに、こうした平均値のプロファイル プロット (交互作用プロット) を使用して関係の一部を簡単に視覚化できます。

残差、予測値、Cook の距離、てこ比の値は、データ ファイルに新変数として保存し、仮定の確認に利用できます。

WLS 重みで、測定方法ごとに異なる精度を補正するなどの目的で、観測値に重み付き最小 2 乗法 (WLS) 分析のためのさまざまな重みを付けるのに使用する変数を指定できます。

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集められています。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、または暑い)、トレーニング月数、過去のマラソン出場回数、性別などが含まれています。年齢は共変量と見なします。ここでは、性別が有意の効果であること、そして性別と天候の交互作用が有意であるということが判明するかもしれません。

方法。異なる仮説を評価する場合は、タイプ I、タイプ II、タイプ III、タイプ IV 平方和を使用できます。デフォルトは タイプ III です。

© Copyright SPSS Inc. 1989, 2010

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、Scheff 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、および 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、および度数。Levene の等分散性の検定。

作図。水準と広がりの図、残差、およびプロファイル（交互作用）。

データ。従属変数は量的変数です。因子も同じくカテゴリ型です。数値または最高 8 文字までの文字値を持つことできます。共変量は、従属変数に関連する量的変数です。

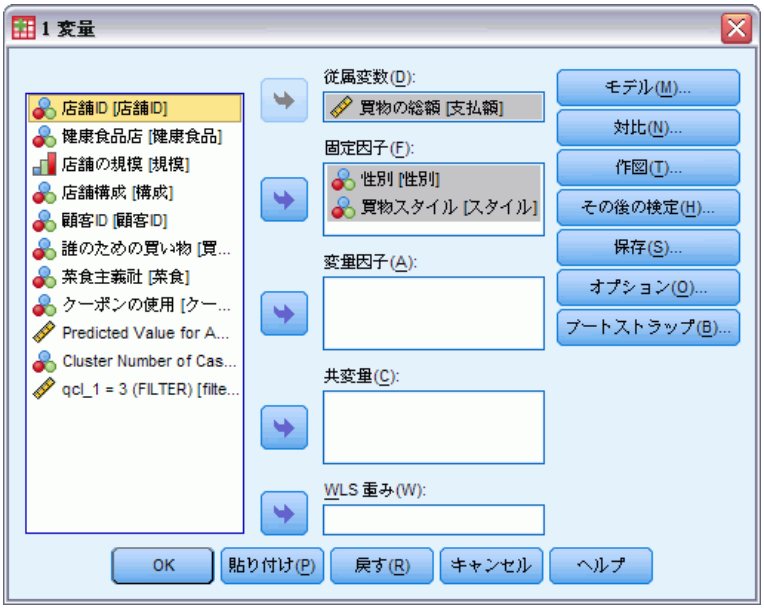
仮定。データは正規母集団からの任意のサンプルで、この母集団では、すべてセルの分散が同じです。データは対称であるべきですが、分散分析は正規性からの逸脱に対し頑健です。仮定の確認には、等分散性の検定や水準と広がりの図を使用できます。残差と残差プロットについて探索的分析を行うこともできます。

GLM – 1 変量分散分析テーブルを作成するには

▶ メニューから次の項目を選択します。

分析(A) > 一般線型モデル > 1 変量...

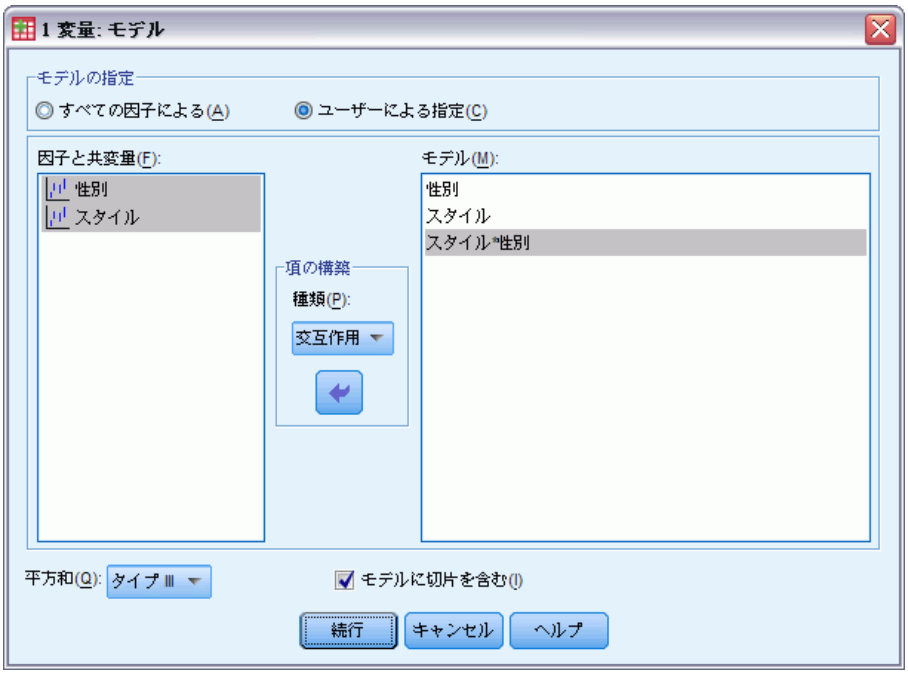
図 11-1
[GLM 1 変量] ダイアログ ボックス



- ［従属変数］ボックスに従属変数を選択します。
- データに合わせて、［固定因子］、［変量因子］、［共変量］の変数を選択します。
- オプションで、［WLS 重み］を使用して、重み付き最小 2 乗法分析のための重み付け変数を指定できます。重み付け変数の値がゼロ、負、または欠損の場合、ケースは分析から除外されます。モデル内ですでに使用された変数は、重み付け変数として使うことはできません。

GLM モデル

図 11-2
［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、因子の主効果、共変量の主効果、因子間の交互作用のすべてが含まれます。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。モデルに含める項目はすべて指示する必要があります。

因子と共変量。要素および共変量はリストされます。

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**］をクリックすれば、分析対象の主効果と交互作用を選択できます。

平方和。平方和の計算方法。釣り合い型モデルと不釣り合い型モデルで欠損セルがない場合は、タイプ III の平方和の方法が最もよく利用されます。

モデルに切片を含む。通常、モデルには定数項が含まれます。データが原点を通ると仮定できる場合は、切片を除外できます。

項の構築 ロジット モデル

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。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

平方和

モデルには、平方和のタイプを選択できます。最も一般的に使われ、デフォルトでもあるのがタイプ III です。

タイプ I。この方法は、平方和の階層的分割法という名でも知られています。各項は、モデル内で先行する項目に対してだけ調整されます。タイプ I の平方和は、一般に次のような場合に使用します。

- 1 次の交互作用の効果より先になんらかの主効果が指定され、2 次の交互作用の効果より先になんらかの 1 次の交互作用の効果が指定されているといった具合の分散分析の釣り合い型モデル。
- 高次の項より先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の中に入れ子になり、2 番目に指定された効果が 3 番目に指定された効果の中に入れ子になり、という具合に続く純粋な入れ子モデル。(入れ子の形式は、シンタックスを使用しなければ指定できません)。

タイプ II。この方法では、他のすべての「適当な」効果に対して調整されるモデルの効果の平方和が計算されます。適当な効果とは、調査中の効果を含んでいないすべての効果に対応するものです。タイプ II の平方和の方法は、一般に次のような場合に使用します。

- 分散分析の釣り合い型モデル。
- 因子の主効果しか持たないモデル。
- 回帰モデル。
- 純粋に入れ子になっている計画。(入れ子の形式は、シンタックスを使用して指定できます)

タイプ III。デフォルトです。この方法では、計画内にある効果の平方和を、その計画を含まないその他の効果、またはそうした効果に対して直交的な効果を対象に調整した平方和として計算します。タイプ III の平方和には、一般的な推定形式が一定の状態に保たれている限り、セル度数が変わらないという大きな利点があります。したがって、このタイプの平方和は、欠損セルがない不釣り合い型モデルに有用だと考えられます。欠損セルのない多因子計画でこの方法に相当するのが、Yates の平均値の重み付き 2 乗法です。タイプ III の平方和の方法が一般的に使用されるのは次の場合です。

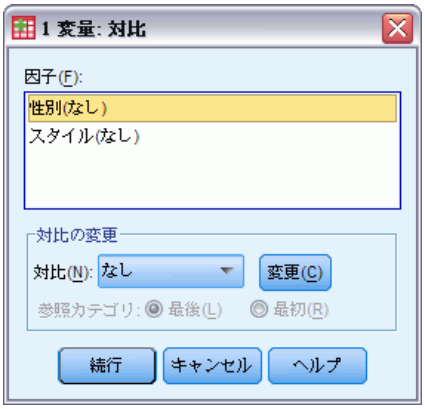
- タイプ I とタイプ II に記載されているモデル。
- 空白セルのない釣り合い型モデルまたは不釣り合い型モデル。

タイプ IV。この方法は、欠損セルが存在する状況を目的としたものです。計画における効果 F に対して、F が他のどの効果にも含まれていないとき、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とき、タイプ IV は、F におけるパラメータ間で行われている対比を、より高いレベルの効果のすべてに等しく分配します。タイプ IV の平方和の方法は、一般に次のような場合に使用します。

- タイプ I とタイプ II に記載されているモデル。
- 空白セルを伴う釣り合い型モデルまたは不釣り合い型モデル。

GLM の対比

図 11-3
[1 変量: 対比] ダイアログ ボックス



1 つの因子の水準同士に差異があるかどうかを検定するときに用いられるのが対比です。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）に指定できます。対比は、パラメータの線型結合を表します。

仮説の検定は、LB = 0 という帰無仮説に基づきます。この場合、L は対比係数行列で、B はパラメータのベクトルです。対比が指定されると、L 行列が作成されます。要素に対応する L 行列の列は、対比と一致します。残りの列は、L 行列が推定できるように調整されます。

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 の同時信頼区間も表示されます。

利用できる対比

利用できる対比には、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、参照カテゴリの最初か最終を選択できます。

対比の種類

偏差。各レベルの平均値（参照カテゴリを除く）をすべてのレベルの平均値（偏差）と比較します。因子の水準は任意の順序になります。

単純。各レベルの平均値を特定のレベルの平均値と比較します。この対比の型はコントロール群があるときに有効です。最初または最後のカテゴリを参照として選択できます。

差分。各レベル（最初は除く）の平均値を前のレベルの平均値と比較します（「逆 Helmert 対比」と呼ばれることもあります）。

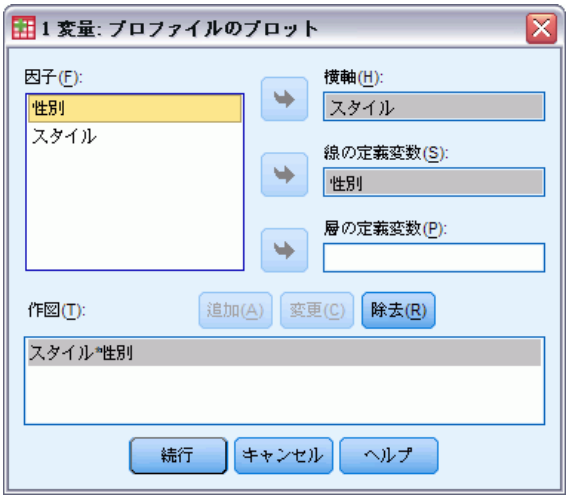
Helmert。因子の各水準（最終を除く）の平均とその後の水準の平均とを比較します。

反復測定。因子の各水準（最終を除く）の平均とその後の水準の平均とを比較します。

多項式。1 次効果、2 次効果、3 次効果等を比較します。第 1 自由度にはすべてのカテゴリを通じての 1 次効果が含まれ、第 2 自由度には 2 次効果が含まれるなど、それぞれが含まれます。このような対比は、多項式のトレンドを推定する場合に頻繁に使用します。

GLM のプロファイル プロット

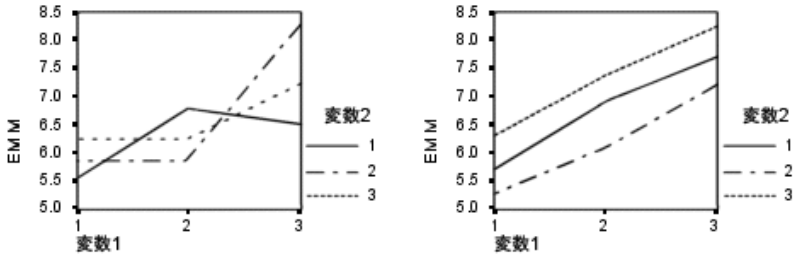
図 11-4
[1 変量: プロファイルのプロット] ダイアログ ボックス



プロファイル プロット（交互作用プロット）は、モデル中の周辺平均を 比較するときに役立ちます。プロファイル プロットは、点を使って、因子の 1 つの水準で従属変数（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 因子の水準を使って、個別の線を作成できます。第 3 因子の各水準を使用して、層別のプロットを作成できます。固定因子と変量因子がある場合には、そのすべてをプロットに使用できます。多変量分析では、従属変数ごとにプロファイル プロットが作成されます。反復測定分析では、プロファイル プロットに被験者間因子と被験者内因子の両方を使用できます。Advanced Statistics オプションをインストールしている場合にだけ「GLM - 多変量」と「GLM - 反復測定」を使用できます。

1 つの因子についてのプロファイル プロットから、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わかります。2 つ以上の因子では、因子間で交互作用がないと平行線で示されるので、1 つの因子に限ってそのレベルを調べることができます。平行ではない線は、交互作用を示しています。

図 11-5
平行ではないプロット（左）と平行プロット（右）



水平軸の因子を選択してプロットを指定し、オプションとして線の定義変数と層の定義変数を選択したら、そのプロットは［作図］ボックスの一覧に追加しなければなりません。

GLM のその後の比較

図 11-6
［その後の検定］ダイアログ ボックス

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後は、その後の範囲検定とペアごとの多重比較により、どの平均値が相異しているのかを決めることができます。比較は調整済みでない値に基づいて行われます。このような検定は、固定被験者間因子の場合にだけ使用します。GLM - 反復測定では、被験者間因子がなければこの検定は利用できません。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します。GLM- 多変量では、その後の検定は、それぞれの測定ごとに従属変数の被験者内での平均に対して実行します。Advanced Statistics オプションをインストールしている場合にだけ「GLM - 多変量」と「GLM - 反復測定」を使用できます。

多重比較検定では、通常、Bonferroni 検定と Tukey の HSD 検定を使用します。**Bonferroni の検定**では、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が調整されます。**Sidak の t 検定**でも有意水準が調整され、Bonferroni の方法よりも厳しく限定されます。**Tukey の HSD 検定**では、スチューデント化された範囲統計量を使っ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。大量の平均値のペアを検定する場合は、Tukey の HSD 検定の方が Bonferroni 検定より有効です。少量のペアの場合は Bonferroni の方法の方が有効です。

Hochberg の GT2 検定は Tukey の HSD 検定と類似していますが、ここで使用するのはスチューデント化された最大法です。一般的には、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、一般的に、セルの大きさが均等ではない場合は、Hochberg の GT2 より有効です。セルの大きさ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。

Dunnett のペアごとの多重比較 t 検定では、処理のグループを単一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できます。両側または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の平均値（対照カテゴリを除く）が対照カテゴリのそれと等しくないことを検定するには、両側の検定を使います。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ことを検定するには、[< **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大きいことを検定するには、> [**対照カテゴリ**] を選択します。

Ryan と 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では、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くなければ、サブグループの平均値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し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）より有効ですが、セルの大きさ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）、**Dunnett の T3**（スチューデント化し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 **Dunnett の C**（スチューデント化した範囲に基づくペアごとの比較検定）を使用します。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く、生成されません。

Duncan の多重範囲検定、Student-Newman-Keuls（**S-N-K**）の方法、および **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われていません。

Waller-Duncan の t 検定では、Bayesian のアプローチが使われています。この範囲検定では、サンプル サイズが等しくない場合にサンプル サイズの調和平均が使われます。

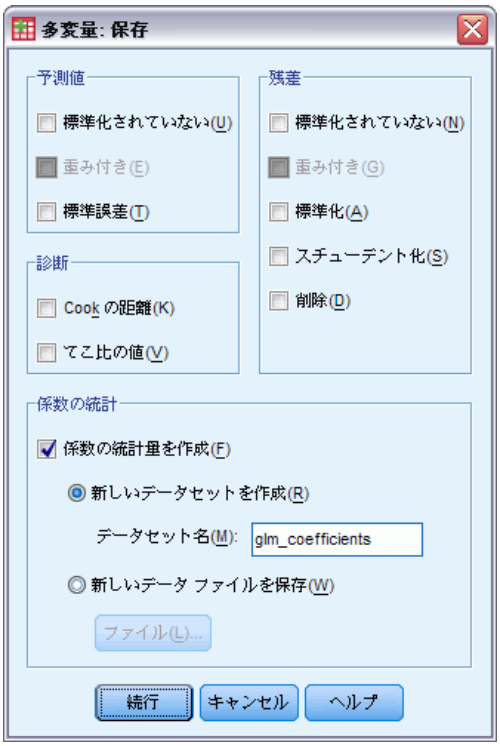
Scheffé の検定の有意水準は、この機能で利用できるペアごとの比較だけにとどまらず、グループ平均に可能なすべての線型結合を検定できる設計になっています。結果的に、Scheffé の検定は他の検定より控えめになってしまうことが多いため、有意確率を求める場合は、平均値間の差が大きい必要があります。

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ごとの多重 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を多重比較向けに調整する試みが一切行われないことです。

表示される検定。 ペアごとの比較は、LSD、Sidak の方法、Bonferroni の方法、Games と Howell の方法、Tamhane の T2 と T3、Dunnett の C と Dunnett の T3 で使用できます。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 および Waller の方法で使用できます。Tukey の HSD 検定、Hochberg の GT2、Gabriel の検定、および Scheffé の検定は、多重比較検定と範囲検定の両方です。

GLM の保存

図 11-7
[保存] ダイアログ ボックス



モデルが予測した値、残差、関連測定値は、データ エディタで新変数として保存できます。これらの変数の多くは、データの仮定を調べるために利用できます。値を保存して別の IBM® SPSS® Statisticsセッションで利用するためには、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 (判別分析).** 従属変数を予測するモデルの値。

- **重み付き**. 重み付きの標準化されていない予測値。WLS 変数がすでに選択されている場合にのみ使用することができます。
- **標準誤差**. 独立変数の同じ値を持つケースに対する従属変数の平均値の標準偏差の推定値。

診断。独立変数の一般的でない値の組み合せのケースとモデルに大きな影響を及ぼすケースを識別するための測定方法。

- **Cook の距離**. 特定のケースが回帰係数の計算から除外された場合に、すべてのケースの残差がどのくらい変化するかを示す測定量。Cook の距離が大きいときは、回帰統計量の計算からケースを除外したことが係数を実質的に変化させたことを示しています。
- **てこ比の値**. 非心てこ比の値。モデルの適合度に関する各観測の相対的な影響度。

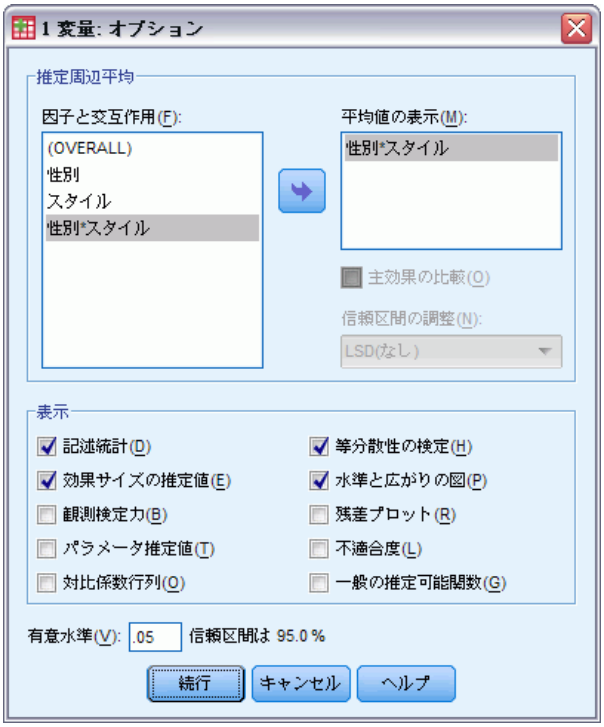
残差。標準化されていない残差は、従属変数の実際の値からモデルにより予測される値を引いたものです。標準化された残差、スチューデント化された残差、削除された残差も使用できます。WLS 変数を選択すると、重み付けした標準化されていない残差が使用できます。

- **標準化されていない (判別分析)**. 観測値およびモデルで予測された値との差。
- **重み付き**. 重み付きの標準化されていない残差。WLS 変数がすでに選択されている場合にのみ使用することができます。
- **標準化残差**. 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼ばれ、平均は 0 で、標準偏差 1 になります。
- **スチューデント化された残差**. 残差を、独立変数の平均値からの独立変数の各ケース値の距離に依存して、ケースごとに違う標準偏差の推定量で割ったもの。
- **削除**. あるケースが回帰係数の計算から除外されたときのケースの残差。従属変数と調整済み予測値の間の差です。

係数統計量。モデルにおけるパラメータ推定値の分散共分散行列を、現在のセッションにある新しいデータセット、または SPSS Statistics 形式の外部データ ファイルに出力します。また、それぞれの従属変数に対しても、パラメータ推定値の行、パラメータ推定値に対応する t 統計量の有意確率の行、および残差自由度の行が存在します。多変量モデルの場合は、各従属変数に同様の行があります。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できます。

GLM のオプション

図 11-8
[オプション] ダイアログ ボックス



このダイアログ ボックスでオプションの統計を利用できます。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。セルにおける母周辺平均値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。この項目は、[平均値の表示] リストで主効果を選択した場合にだけ選択できます。
- **信頼区間の調整。**最小有意差 (LSD)、Bonferroni の方法、Sidak の方法の信頼区間と有意水準に対する調整を選択します。この項目は、[**主効果の比較**] を選択している場合にだけ選択できます。

表示。[**記述統計量**] を選択すると、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、および度数を求めることができます。[**効果サイズの推定値**] は、偏相関のイータの 2 乗の値をそれぞれの効果とそれぞれのパラメータ推定値に与えるものです。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するものです。観測した値に基づいて対

立仮説を立てるときに検定の検出力を得るには、[**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、およびそれぞれの検定での観測検定力を創出するには、[**パラメータ推定値**] を選択します。L 行列を得るには、[**対比係数行列**] を選択します。

等分散性の検定は、被験者間因子の場合にだけ、各従属変数の分散の等質性分析について、被験者間因子のすべての水準の組み合せを通じて Levene の検定を創出するものです。水準と広がりの図および残差プロットのオプションは、データに関する仮定のチェックに役立ちます。この項目は、因子が存在しないときは無効になります。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。このようなプロットは、分散が等しいという仮定の検証に役立ちます。[**不適合度**] を選択して、モデルにより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チェックします。**一般の推定可能関数** を使うと、一般推定可能関数に基づいてユーザー指定の仮説の検定を構築できます。任意の対比係数行列における行は、一般推定可能関数の一次結合です。

有意水準。信頼区間を構築するために、その後の検定や信頼係数で使わ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使用できます。有意水準を指定するときは、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

コマンド シンタックスを使用すると、次の作業も実行できます。

- 計画中の入れ子効果の指定 (DESIGN サブコマンドを使用)。
- 効果と効果や値の線型組み合せ検定の指定 (TEST サブコマンドを使用)。
- 多重対比の指定 (CONTRAST サブコマンドを使用)。
- ユーザー欠損値の包含 (MISSING サブコマンドを使用)。
- EPS 基準の指定 (CRITERIA サブコマンドを使用)。
- ユーザー定義の L 行列、M 行列、または K 行列の作成 (LMATRIX、MMATRIX、および KMATRIX サブコマンドを使用)。
- 全平均対比または単純対比での、中間参照カテゴリの指定 (CONTRAST サブコマンドを使用)。
- 多項式対比の計量の指定 (CONTRAST サブコマンドを使用)。
- その後の比較の誤差項の指定 (POSTHOC サブコマンドを使用)。
- 因子または因子の一覧にある因子間の交互作用の推定周辺平均値の計算 (EMMEANS サブコマンドを使用)。
- 仮の変数の名前の指定 (SAVE サブコマンドを使用)。
- 相関行列のデータ ファイルの作成 (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含む行列データ ファイルの作成 (OUTFILE サブコマンドを使用)。
- 計画行列の新しいデータ ファイルへの保存 (OUTFILE サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
12

2 変量の相関分析

有意水準のある Pearson の相関係数、Spearman のローおよび Kendall のタウ b を計算するには、[2 変量の相関分析] 手続きを使用します。変数またはランク順の関係を計算するには、相関分析を使用します。ただし、相関係数を計算する前に、2 変数間に外れ値 (これがあると誤った結果を出す可能性があります) および線型関係があるかどうかを確かめるためにグラフを描いてみてください。Pearson の相関係数は、線型連関の尺度になります。2 変数に強い関係があっても、その関係が直線的でない場合は、Pearson の相関係数は関連を測るのには適していません。

例。あるバスケットボール チームの勝数と 1 ゲームごとの平均得点の関係は相関しているでしょうか。散布図は直線関係があることを示しています。1994 年 ～ 1995 年の NBA シーズンにおけるデータ解析によると、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、相手チームの得点が少ない (1 ゲームあたりの防御点が高い) ように見えます。この場合、変数は負 (-0.401) に相関し、その相関は 0.05 レベルで有意となります。

統計量 各変数に対して、非欠損値を持つケースの数、平均値および標準偏差。変数のペアごとに、有意水準のある Pearson の相関係数、Spearman のロー、Kendall のタウ b、交差積和、共分散。

データ。Pearson の相関係数には、正規分布している集団から抽出された量的変数を使用し、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、それぞれの変数の組が 2 変量正規であると仮定します。

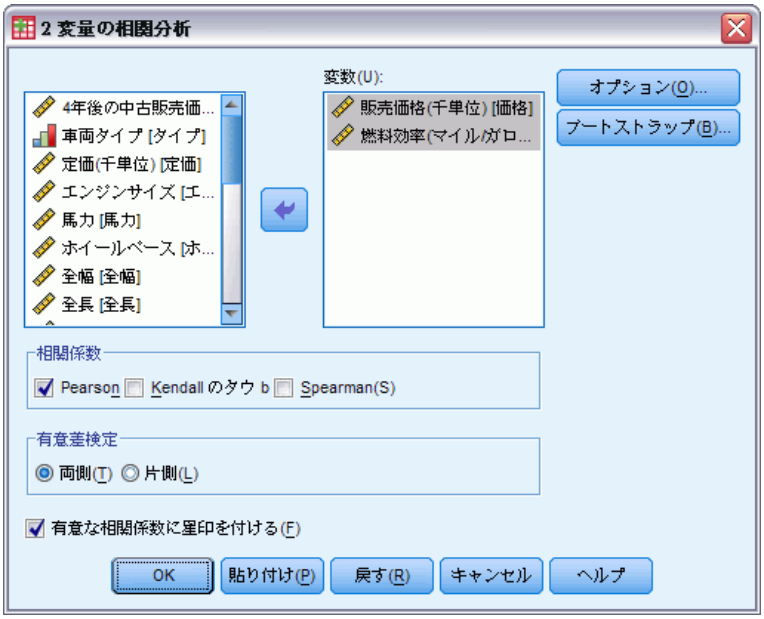
2 変量の相関分析を行うには

メニューから次の項目を選択します。

分析(A) > 相関 > 2 変量...

© Copyright SPSS Inc. 1989, 2010

図 12-1
[2 変量の相関分析] ダイアログ ボックス



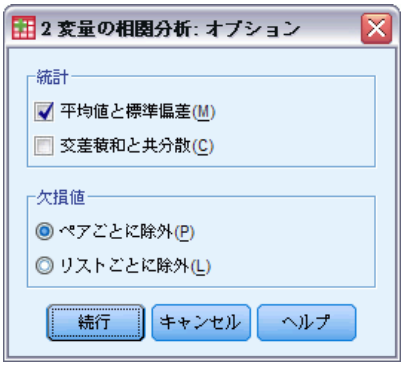
▶ 2 つ以上の数値型変数を選択します。

次のオプションも選択できます。

- **相関係数。**正規分布している集団から抽出された量的変数には、[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か、または順序付けされたカテゴリを持っている場合、[Kendall **のタウ** b] または [Spearman] を選択して、ランク順の間の連関を測ります。相関係数の範囲は、-1（完全な負の相関）から +1（完全な正の相関）までの値をとります。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、相関が有意であるからといって、因果関係があるという結論を出さないように注意する必要があります。
- **有意差検定。**両側確率または片側確率を選択できます。連関の方向がわかっている場合は、[**片側**] を選択します。不明の場合は、[**両側**] を選択します。
- **有意な相関係数に星印を付ける。**相関係数が 5% 水準で有意な場合は星印が 1 つ、1% 水準で有意な場合は星印が 2 つ付いた形で表示されます。

2 変量の相関分析のオプション

図 12-2
[2 変量の相関分析: オプション] ダイアログ ボックス



統計。 Pearson の相関分析については、次のどちら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。また欠損値を除いた全ケース数も表示されます。欠損値の除外の仕方に関係なく、欠損値は変数単位で扱われます。
- **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。これが、Pearson の相関係数の算出式の分子になります。共分散は 2 変数間の関係の標準化されていない尺度であり、交差積和を N-1 で割った値と等しくなります。

欠損値。 次のオプションのうち 1 つを選択できます。

- **ペアごとに除外。** 相関係数の変数の組み合せのうち、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、その組み合せの変数が両方とも欠損値でないすべてのケースから計算されるため、計算時には、利用できる最大の情報が使用されます。そのため、係数ごとに使用したケースの数が異なる場合があります。
- **リストごとに除外。** 変数の欠損値のあるケースは、すべての相関から除外されます。

CORRELATIONS および NONPAR CORR コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- Pearson の相関係数では、相関行列を書き出して、生データの代わりに因子分析などの他の分析を実行 (MATRIX サブコマンドを使用)。
- 2 つの変数リスト間の各変数の相関を得る (VARIABLES サブコマンドのキーワード WITH を使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
13

偏相関分析

偏相関手続きは、1 つ以上の追加変数の効果をコントロールしながら 2 つの変数間の線形関係を表す偏相関係数を計算します。相関は、線形連関の測度です。2 つの変数は完全に関連していますが、その関係が線形ではない場合、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?このような関係が負の相関となると予測していても、研究では、医療用資金が増加すると罹患率も増加するように、正の相関を報告します。しかし、医療サービス機関への訪問率を操作すると、観測された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、資金が増えるほど、人々は医療を受けることができ、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からです。

統計量 各変数に対して、非欠損値を持つケースの数、平均値および標準偏差。自由度と有意確率を伴った偏相関行列および 0 次相関行列。

データ。対称的な量的変数を使用します。

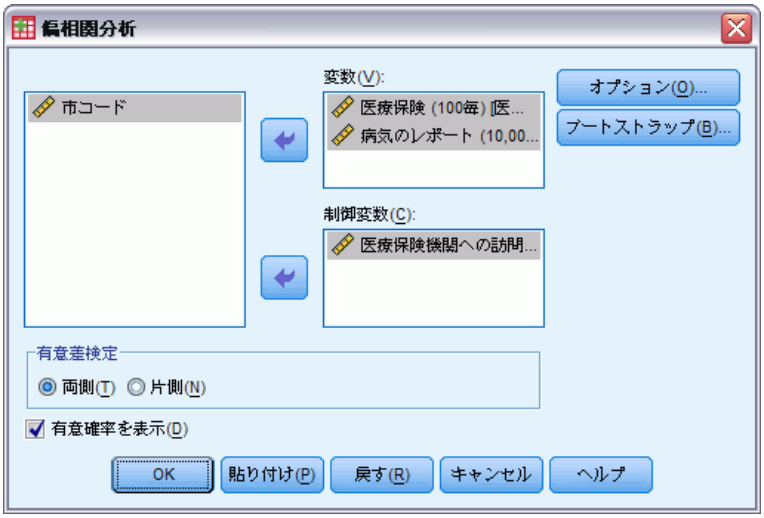
仮定。偏相関手続きは、各ペアの変数は 2 変量正規であると仮定しています。

偏相関を行うには

▶ メニューから次の項目を選択します。

分析(A) > 相関 > 偏相関...

図 13-1
[偏相関分析] ダイアログ ボックス



▶ 偏相関を計算する 2 つ以上の数値変数を選択します。

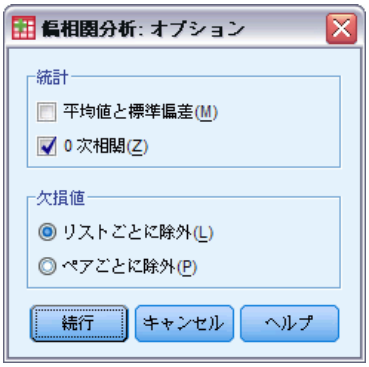
▶ 1 つ以上の数値型制御変数を選択します。

次のオプションも選択できます。

- **有意差検定。**両側確率または片側確率を選択できます。連関の方向がわかっている場合は、[**片側**] を選択します。不明の場合は、[**両側**] を選択します。
- **有意確率を表示。**デフォルトでは、確率と自由度は各相関係数に表示されます。この項目の選択を解除すると、5% 水準で有意な係数は 1 つの星印で識別され、1% 水準で有意な係数は 2 つの星印で識別され、自由度は抑制されます。この設定は、偏相関と 0 次相関両方に影響します。

偏相関のオプション

図 13-2
[偏相関分析: オプション] ダイアログ ボックス



統計。次の中から 1 つまたは両方を選択することができます。

- **平均値と標準偏差。**変数ごとに表示されます。また欠損値を除いた全ケース数も表示されます。
- **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。次のオプションのどちらかを選択できます。

- **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。**偏相関が基になっている 0 次相関の計算には、ペアになった変数の両方または 1 つ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ります。ペアごとの削除が有効なとき、特定の部分係数の自由度は、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いています。

PARTIAL CORR コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 0 次相関行列の読み込み、または偏相関行列の書き出し (MATRIX サブコマンド使用)。
- 2 つの変数リスト間の偏相関の取得 (VARIABLES サブコマンドで WITH キーワード使用)。
- 複数分析の取得 (複数の VARIABLES サブコマンド使用)。
- 制御変数が 2 つある場合の、要求する次数値の指定 (たとえば、1 次偏相関と 2 次偏相関) (VARIABLES サブコマンド使用)。

- 冗長な係数の非表示（FORMAT サブコマンド使用）。
- 計算できない係数がある場合における、単純相関行列の表示（STATISTICS サブコマンド使用）。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
14

距離行列

この手続きでは、変数のペア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することができます。これらの類似度または距離は、因子分析、クラスタ分析、または多次元尺度法などの他の手続きで使用でき、複雑なデータセットの分析に役立ちます。

例:€ エンジンのサイズ、ガソリン 1 ガロン当たりの走行距離、馬力などの任意の特性変数をもとに、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って、自動車が互いに類似しているか、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には、基本的構造を調べる階層クラスタ分析または多次元尺度法を使用することもできます。

統計量 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、都市ブロック、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、パターンの差異、分散、形、または Lance と Williams の距離です。間隔データの類似度の測度は、Pearson の相関係数またはコサインです。また 2 値データの類似度の測度は、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、4 分点相関ファイ、または散らばりです。

距離行列を取得するには

▶ メニューから次の項目を選択します。

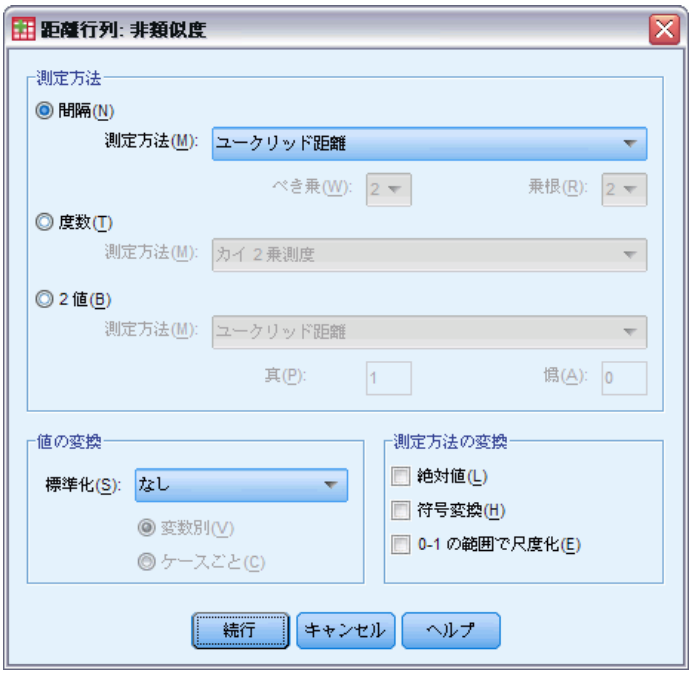
分析(A) > 相関 > 距離...

図 14-1
[距離行列] ダイアログ ボックス

- ケース間の距離を計算するには、[変数] ボックスに少なくとも 1 つの数値型変数を選択し、変数間の距離を計算するには、少なくとも 2 つの数値型変数を選択します。
- [距離の計算] グループの [ケース間] または [変数間] をクリックして、ケース間または変数間のどちらの距離を計算するか指定します。

距離行列の非類似度の測定方法

図 14-2
[距離行列: 非類似度] ダイアログ ボックス



[測定] グループでデータの種類 ([区間]、[度数]、または [2 値]) を選択します。次に、ドロップダウン リストからデータの種類に応じて測定方法を 1 つ選択します。データの種類別に使用できる測定方法は次のとおりです。

- **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、サイズの差異、パターンの差異、分散、形、Lance と Williams。([真] と [偽] の各ボックスに値を入力し、対象の 2 値を指定します。他の値は無視されます)。

[値の変換] グループでは、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。2 値データの場合、[値の変換] は使用できません。使用できる標準化方法は、[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]、または [標準偏差を 1] です。

[測定方法の変換] グループでは、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。変換した値は、距離を計算した後で適用されます。使用できるオプションは、[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

図 14-3
[距離行列: 類似度] ダイアログ ボックス

[測定] グループでデータの種類 ([区間] または [2 値]) を選択します。次に、ドロップダウン リストからデータの種類に応じて測定方法を 1 つ選択します。データの種類別に使用できる測定方法は次のとおりです。

- **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、4 分点相関ファイ、散らばり。([真] と [偽] の各ボックスに値を入力し、対象の 2 値を指定します。他の値は無視されます)。

[値の変換] では、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。2 値データの場合、[値の変換] は使用できません。使用できる標準化方法は、[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]、または [標準偏差を 1] です。

[測定方法の変換] グループでは、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。変換した値は、距離を計算した後で適用されます。使用できるオプションは、[絶対値]、[符号変換]、[0 ～ 1 の範囲で尺度化] です。

PROXIMITIESL コマンドの追加機能

距離行列手続きでは、PROXIMITIES コマンド シンタックスが使用されます。コマンド シンタックスを使用すると、次の作業も実行できます。

- Minkowski の距離のべき乗として任意の整数を指定。
- ユーザー指定の距離のべき乗および根として任意の整数を指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
15

線型モデル

線型モデルは、対象と 1 つまたは複数の予測値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、スコアリングを行うための解釈しやすい数学式を提供しています。これらのモデルのプロパティについてはよく理解され、通常は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シジョン ツリーなど）に比べてすぐに構築できます。

例:世帯主の保険請求を調査するためのリソースが制限されている保険会社が、請求コストを推定するためのモデルを作成したいと考えています。このモデルをサービス センターに提供することによって、担当者は顧客との電話中に請求情報を入力し、過去のデータに基づいて「予測される」請求のコストをすぐに計算できます。

© Copyright SPSS Inc. 1989, 2010

図 15-1
[フィールド] タブ

フィールドの要件:対象フィールドと、少なくとも 1 つの入力フィールドが必要です。デフォルトでは、定義済みの役割が [両方] または [なし] のフィールドは使用されません。目標は、連続型 (スケール) でなければなりません。予測フィールド (入力) に尺度の制限はありません。

線型モデルを構成するには

この機能は Statistics Base オプションが必要です。

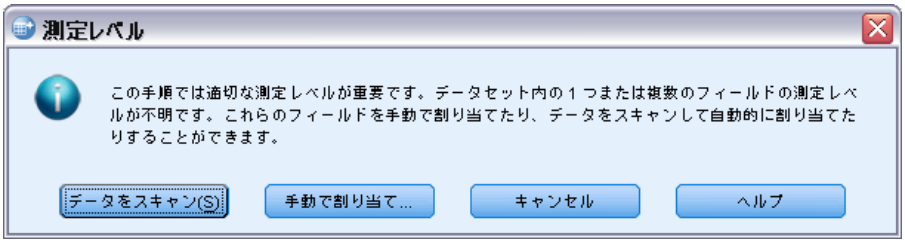
メニューから次の項目を選択します。

分析(A) > Regression > 自動線型モデル...

- ▶ 対象フィールドと入力フィールドが少なくとも 1 つずつ必要です。
- ▶ [**作成オプション**] をクリックして、オプションの作成設定およびモデル設定を指定します。
- ▶ [**モデル オプション**] をクリックし、スコアをアクティブなデータセットに保存したり、モデルを外部ファイルにエクスポートできます。
- ▶ [**実行**] をクリックして手続きを実行してモデル オブジェクトを作成します。

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

図 15-2
尺度の警告



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

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

目的

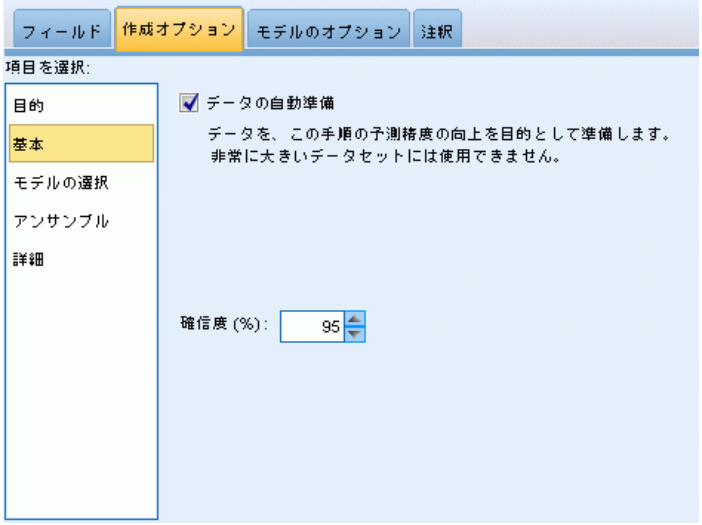
主な目的は？

- **標準モデルを作成します。**この方式により、予測値を使用して目標を予測する単一モデルを構築します。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、素早くスコア付けできます。
- **拡張モデルの精度（ブースティング）。**この方式により、ブースティングを使用してアンサンブル モデルを構築し、より正確な予測を得るための一連のモデルを生成します。アンサンブルは、標準モデルと比べて構築とスコア付けに時間がかかる場合があります。
- **拡張モデルの安定性（バギング）。**この方式により、バギング（ブートストラップ集計）を使用してアンサンブル モデルを構築し、より信頼性の高い予測を得るための複数のモデルを生成します。アンサンブルは、標準モデルと比べて構築とスコア付けに時間がかかる場合があります。
- **非常に大きなデータセットのモデルを構築（IBM® SPSS® Statistics Server が必要）。**この方式により、データセットを個別のデータ ブロックに分割してアンサンブル モデルを構築します。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、または増分モデル構築の場合、こ

のオプションを選択します。このオプションは、標準モデルと比べて構築には時間がかからないものの、スコア付けに時間がかかる場合があります。このオプションには、SPSS Statistics Server 接続が必要です。

基本

図 15-3
[基本] の設定



自動的にデータを準備する: 内部的にに対象フィールドおよび予測フィールドを変換し、モデルの予測精度を最大化できます。モデルによって変換が保存され、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。

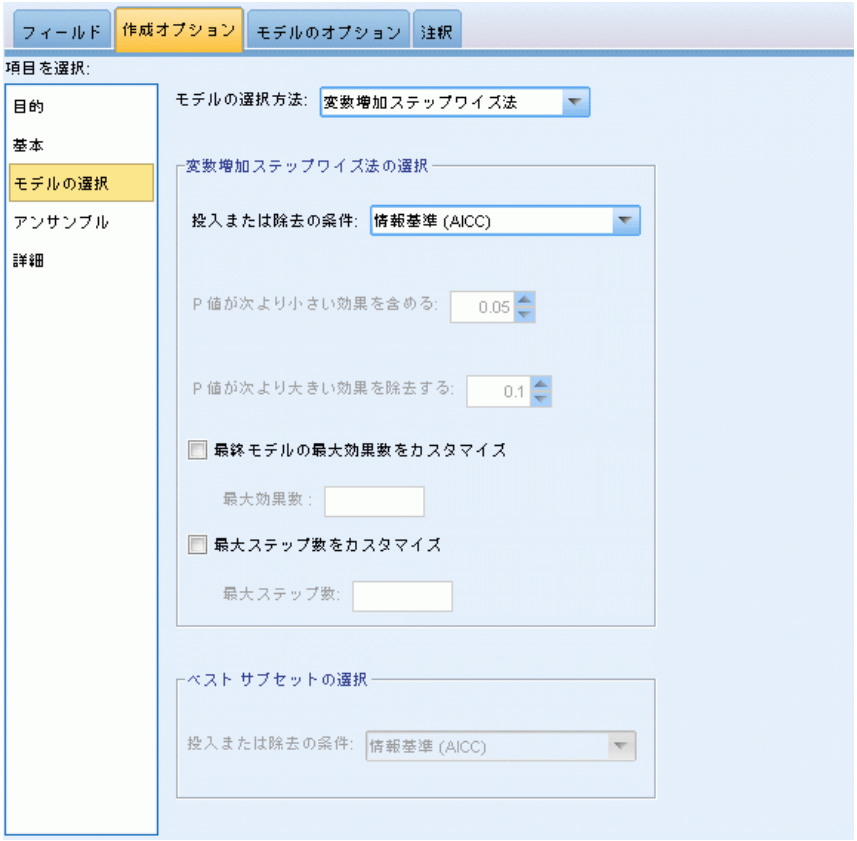
- **日付および時間の処理**各日付の予測フィールドは、基準日 (1970-01-01) 以降の経過時間を含む連続型予測フィールドに変換されます。各時間の予測フィールドは、基準時刻 (00:00:00) 以降の経過時間を含む連続型予測フィールドに変換されます。
- **測定レベルの調整**: 値が 5 個より少ない連続型予測フィールドは順序型予測フィールドに変更されます。10 個より多くの値を持つ順序型予測フィールドは連続型予測フィールドに変更されます。
- **外れ値の処理**:分割値を超える連続型予測フィールドの値 (平均値からの標準偏差が 3) は分割値に設定されます。

- **欠損値の処理**: 名義型予測フィールドの欠損値は、学習データ区分の最頻値と置き換えられます。順序型予測フィールドの欠損値は、学習データ区分の中央値と置き換えられます。連続型予測フィールドの欠損値は、学習データ区分の平均値と置き換えられます。
- **監視結合**: 目標と関連して処理するフィールドの数を減らして、より節約的なモデルを作成します。同様のカテゴリが、入力フィールドと目標フィールド間の関係に基づいて特定されます。それほど重要でないカテゴリ、つまり p が 0.1 より大きいカテゴリは、結合されます。すべてのカテゴリが 1 つのカテゴリに結合される場合、予測値としての値がないため、元のバージョンのフィールドおよび派生した化されたフィールドはモデルから除外されます。

確信度レベル。係数ビューでモデル係数の間隔の推定値を計算するために使用する確信度のレベルです。0 より大きく、100 より小さいの値を指定します。デフォルトは 95 です。

モデルの選択

図 15-4
[モデルの選択] の設定



モデルの選択方法：モデルの選択方法（下記参照）のいずれかまたは [なし] を選択します。[なし] では使用できるすべての予測値を主効果のモデル項として入力します。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：モデルの効果がない状態から始まり、ステップワイズ法の基準に従ってこれ以上追加または削除できなくなるまで一度に 1 ステップずつ効果を追加および削除します。

- **投入または除去の基準**：効果がモデルに加えられるまたは取り除かれるかどうかを決定するときに使用される統計です。情報基準 (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を課すよう調整します。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を課すよう調整します。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、または 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。基準の減少に対応するモデルの効果は削除されます。

  基準として [F 統計量] が選択されると、各ステップで最も小さい p 値が指定されたしきい値より小さい効果がモデルに追加されます（[次の値より小さい p 値の効果を含む]）。デフォルトは 0.05 です。p 値が指定されたしきい値より大きいモデルの効果は削除されます（[次の値より大きい p 値の効果を削除する] は削除されます。デフォルトは 0.10 です。

- **最終モデルの効果の最大数をカスタマイズする**：デフォルトでは、すべての効果をモデルに投入できます。また、ステップワイズ アルゴリズムがステップを指定した効果の最大数で終了する場合、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 アルゴリズムが終了します。デフォルトでは、使用できる効果数の 3 倍です。または、ステップの最大数を正の整数で指定します。

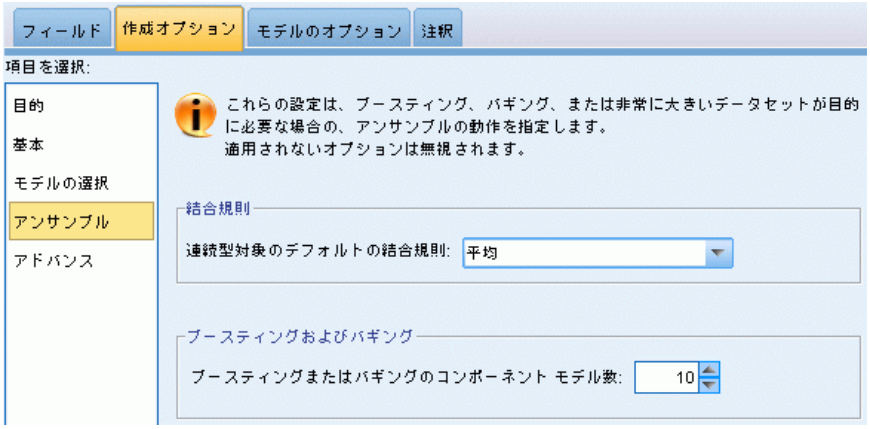
最適サブセットの選択：「使用できるすべての」モデル、または少なくとも変数増加ステップワイズ法より大きいサブセットの使用できるモデルをチェックし、最適サブセットの基準に従って最適サブセットを選択します。情報基準 (AICC) はモデルを指定された学習セットの尤度に基づき、過度に複雑なモデルにペナルティを課すよう調整します。調整済み R2 乗は学習セットの適合度に基づき、過度に複雑なモデルにペナルティを課すよう調整します。オーバーフィット防止基準 (ASE) は、オーバーフィット防止セットの適合度（平均平方誤差、または ASE）に基づきます。オーバーフィット防止セットは、モデルの学習に使用されない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。

基準の値が最大であるモデルが最適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、変数増加法による選択に比べて計算の効率が高くなります。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て実行されると、変数増加法による選択を使用して作成された標準モデルに比べ、大幅に時間がかかる場合があります。

アンサンブル

図 15-5
[アンサンブル] の設定



これらの設定では、目的でブースティング、バギング、および非常に大きなデータセットが必要となる場合に発生するアンサンブルの動作を指定します。選択された目的に当てはまらないオプションが無視されます。

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にスコアを付ける際、このルールを適用して基本モデルに予測値を結合して、アンサンブルのスコア値を計算します。

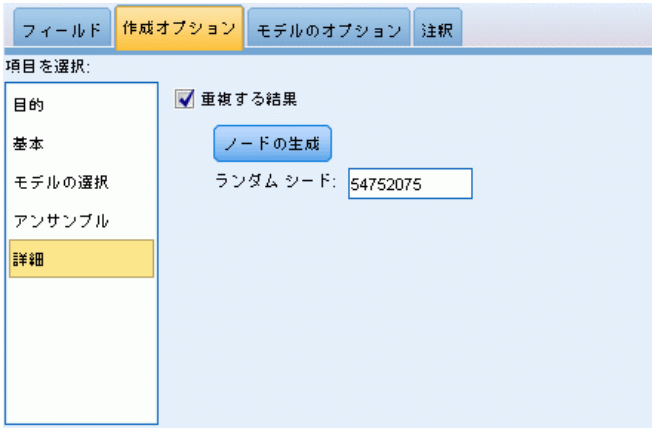
- **連続目標のデフォルトの結合ルールです。**連続目標に対するアンサンブル予測値は、基本モデルから予測値の平均値や中央値を使用して結合できます。

モデルの精度を上げること目的である場合、結合ルールの選択は無視されます。ブースティングは常に、カテゴリ的な目標のスコア付けには重み付けされた大多数の票を使用し、連続目標のスコア付けには重み付けされた中央値を使用します。

ブースティングとバギングモデルの精度と安定性を向上することが目標である場合、構築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のサンプル数となります。値は正の整数にする必要があります。

アドバンス

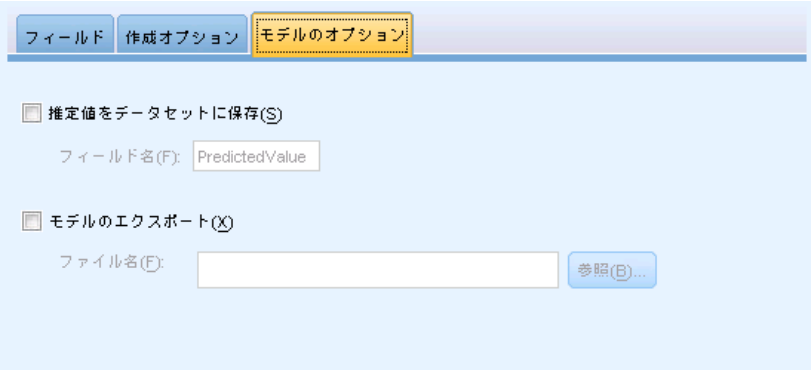
図 15-6
[アドバンス] の設定



結果を複製：ランダム シードを設定すると分析を複製できます。ランダム数値の生成機能は、オーバーフィット防止セットに含まれるレコードの選択に使用します。整数を指定、または [**生成**] をクリックすると、1 ～ 2147483647 の擬似無作為の整数を作成します。デフォルトは 54752075 です。

モデル 選択

図 15-7
[モデル オプション] タブ

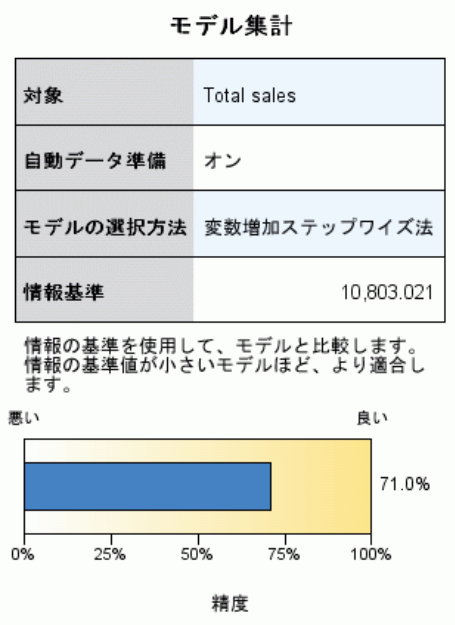


予測値をデータセットに保存します。デフォルトの変数名は、PredictedValue です。

モデルのエクスポート 外部の .zip ファイルにモデルを書き込みます。このモデル ファイルを使用して、得点付け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てください。指定されたファイルが存在する場合、ファイルが上書きされます。

モデルの要約

図 15-8
[モデルの要約] ビュー



[モデル要約] ビューはスナップショットで、モデルと適合度について一目でわかる要約です。

テーブル :次のように、高レベルなモデル設定を示します。

- The name of the target 目標の名前。
- 自動データ準備が基本設定に指定された通りに実行されたかどうか。
- モデル選択設定で指定されたモデル選択方法および選択基準。最終モデルの選択基準の値が表示され、小さく表示されているものがより適切な形式であることを示します。

グラフ: グラフには、最終モデルの精度が表示され、大きく表示されているものがより適切な形式であることを示します。値は、100 × 最終モデルの調整済み R^2 乗です。

自動データ準備

図 15-9
[自動データ準備] ビュー

自動データ準備

対象: Total sales

フィールド	役割	行われるアクション
Age category	予測値	カテゴリを結合して対象との関連を最大化
Primary keyword set	予測値	カテゴリを結合して対象との関連を最大化
Promotion	予測値	測定レベルを連続型から順序型に変更
Secondary keyword set	予測値	カテゴリを結合して対象との関連を最大化

元のフィールド名が X の場合、変換されたフィールド名は X_transformed となります。 元のフィールドが分析から除外され、代わりに変換されたフィールドが分析に追加されます。

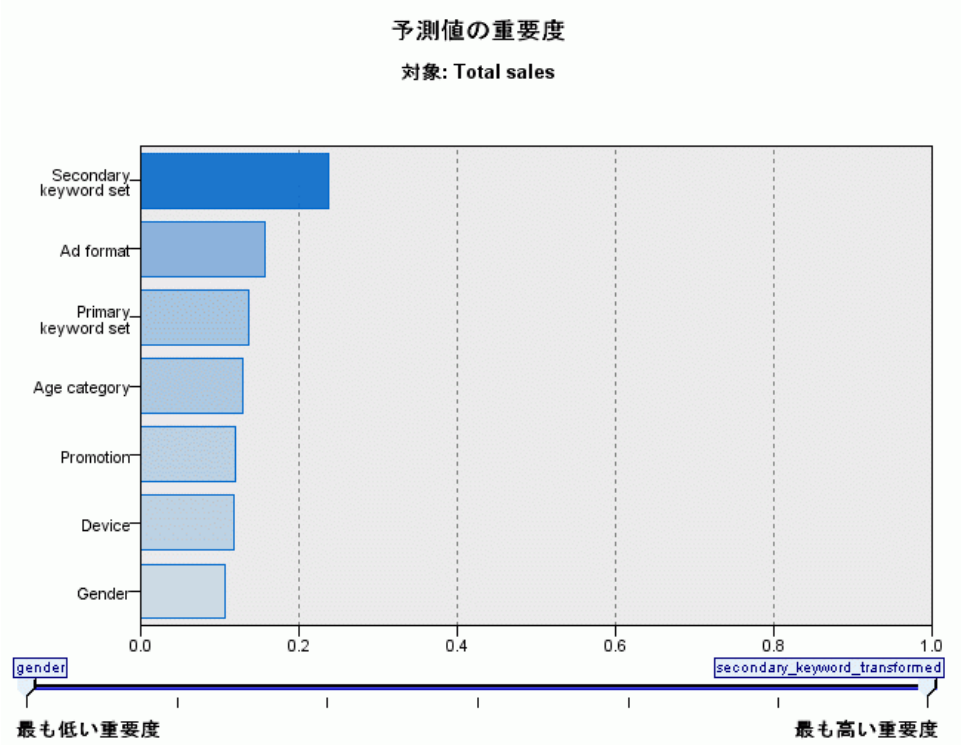
このビューには除外されたフィールドについて、また自動データ準備 (ADP) ステップで変換されたフィールドがどのように派生したかについての情報が表示されます。変換または除外されたフィールドごとに、フィールド名、分析内の役割、ADP ステップで実行されたアクションについて表示されます。フィールドは、フィールド名のアルファベット順 (昇順) に並べ替えられます。各フィールドに行われるアクションには、次のようなものがあります。

- [**期間の計算: 月**] 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期間 (月) を計算します。
- [**期間の計算: 時刻**] は、時刻を含むフィールドの値から現在のシステムの時刻までの期間 (時間) を計算します。
- [**尺度を連続型から順序型に変更**] を選択すると、一意の値が 5 つより少ない連続型フィールドを順序型フィールドに変更します。
- [**尺度を順序型から連続型に変更**] を選択すると、一意の値が 10 つより少ない順序型フィールドを連続型フィールドに変更します。
- [**外れ値を除外**] は、分割値を超える連続型予測フィールドの値 (平均値からの標準偏差が 3) は分割値に設定されます。
- [**欠損値を置換**] では、名義型の欠損値を最頻値に、順序型フィールドの欠損値は中央値に、連続型フィールドの欠損値は平均値に置き換えます。

- ［まばらなカテゴリを結合して目標との関連性を最大化］では、同様の予測カテゴリが、入力フィールドと目標フィールド間の関係に基づいて特定されます。それほど重要でないカテゴリ、つまり p 値が 0.05 より大きいカテゴリは、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、ADP アクションが実行された後、値が 1 つの予測値を削除します。

予測値の重要度

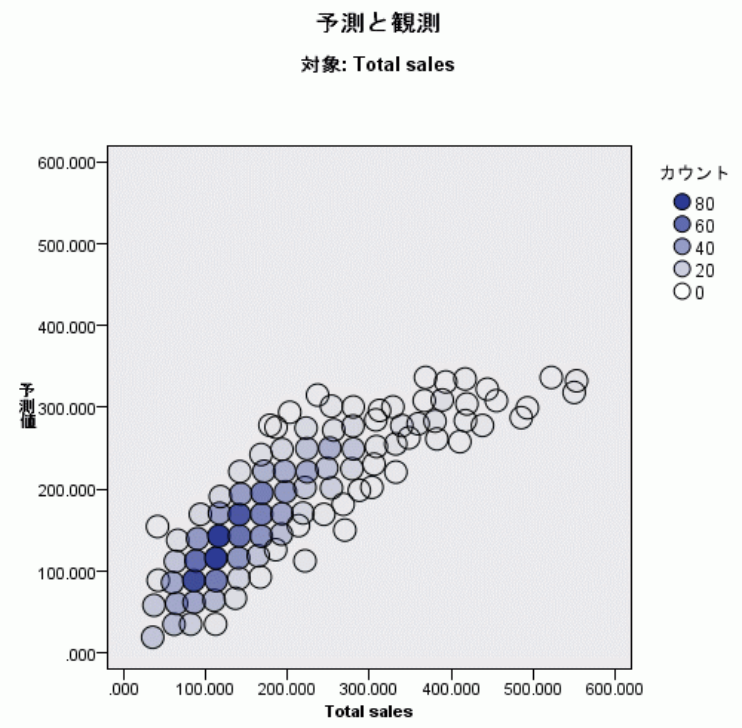
図 15-10
［予測値の重要度］ビュー



通常、モデリングの効果を最も重要な予測値フィールドに集中させ、最も重要でない変数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を示して、これを実現できます。値が相対的であるため、表示されるすべての予測値の値の合計は 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ていません。予測が正確かどうかに関係なく、予測時の各予測値の重要度にのみ関連します。

予測対観測

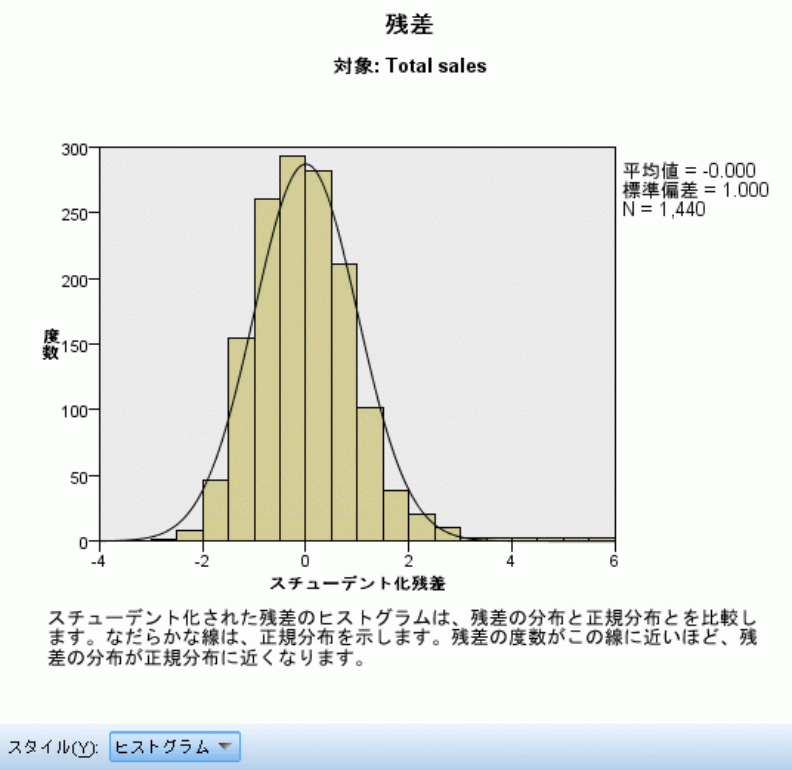
図 15-11
[予測対観測] ビュー



縦軸に予測値を、横軸に観測値を示した分割散布図を表示します。点は 45 度の線にあるのが理想です。このビューはレコードがモデルによって特に不正に予測されているかどうかを示します。

残差

図 15-12
[残差] ビュー、ヒストグラム スタイル



モデル残差の診断グラフを表示します。

グラフのスタイル: 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指定できます。

- **ヒストグラム**: スチューデント化残差の分割ヒストグラムを、正規分布のオーバーレイによって表示します。線型モデルは残差に正規分布があると想定するため、ヒストグラムがほぼ滑らかな線になります。
- **P-P プロット**:スチューデント化残差を正規分布と比較する分割確率-確率プロットです。作図された点の傾斜が通常の線に比べて勾配が小さい場合、残差は正規分布より大きな変動を示し、勾配が大きい場合、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 型カーブを示す場合、残差の分布は歪んでいます。

外れ値

図 15-13
[外れ値] ビュー

外れ値

対象: Total sales

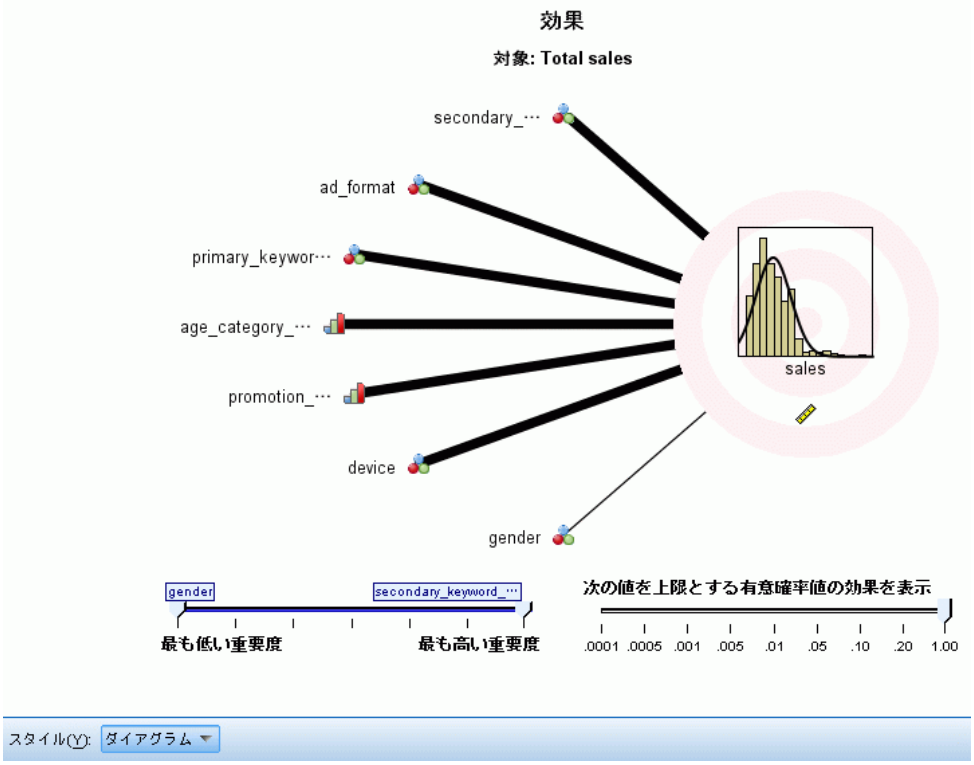
Total sales	クックの距離
560.040	0.026
566.440	0.025
548.990	0.018
539.630	0.018
485.430	0.014
543.240	0.014

このテーブルにはモデルに悪影響を与えるレコードを一覧表示し、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 の距離を示します。Cook の距離は、特定のケースがモデル係数の計算から除外された場合に、すべてのケースの残差がどのくらい変化するかを示す測定量です。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ため、影響力が大きいと考えられます。

影響力が大きいレコードを慎重に検証し、モデルの推定時に重みを小さくするか、外れ値を受け入れ可能なしきい値に変換するか、影響力の大きい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

図 15-14
[効果] ビュー、ダイアグラム スタイル



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 リストから指定できます。

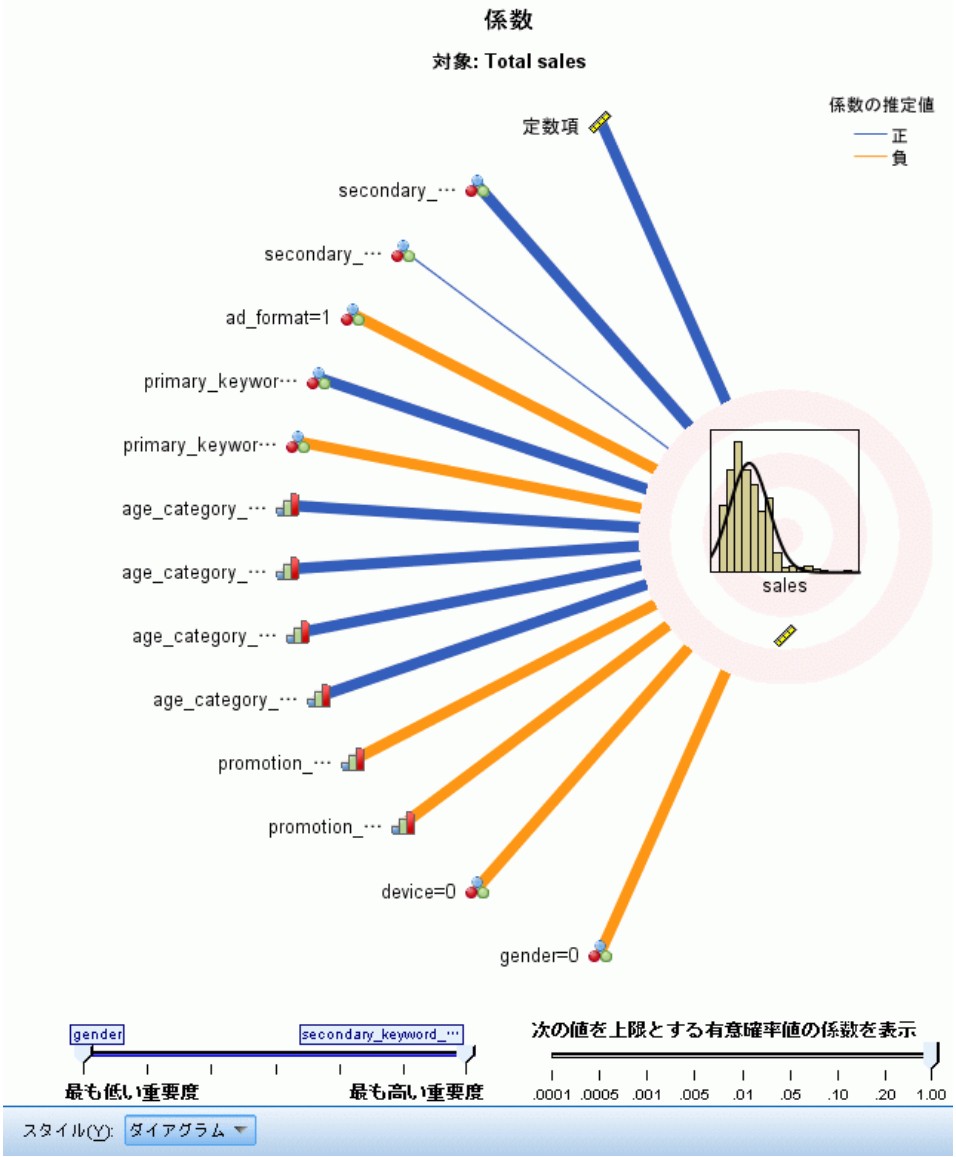
- **ダイアグラム**: 効果について、予測値の重要度が大きいものから順に上から下に並べ替えて表示するグラフです。ダイアグラムで繋がった線は、効果の有意確率に基づいて重みがつけられます。線の太いほど効果の有意確率は大きくなります (p 値は小さくなります)。繋がった線にマウスを置くと、p 値と効果の重要度を示すツールヒントが表示されます。これはデフォルトです。
- **テーブル。** モデル全体および個々のモデルの効果を示す ANOVA テーブルです。各効果は、予測値の重要度が大きいものから順に上から下に並べ替えられます。デフォルトでは、テーブルが折りたたまれ、モデル全体の結果だけが表示されます。個々のモデル効果の結果を表示するには、テーブルの [**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］スライダーで、ビュー内に表示された予測値を制御します。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焦点を当てることができます。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］スライダーで、予測値の重要度に基づいて表示される効果のほか、ビュー内に表示された効果を詳細に制御します。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処理される効果はありません。

係数

図 15-15
[係数] ビュー、ダイアグラム スタイル



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（カテゴリ型予測フィールド）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**があります。一方は冗長（参照）パラメータに対応するカテゴリを除いたものとなります。

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 リストから指定できます。

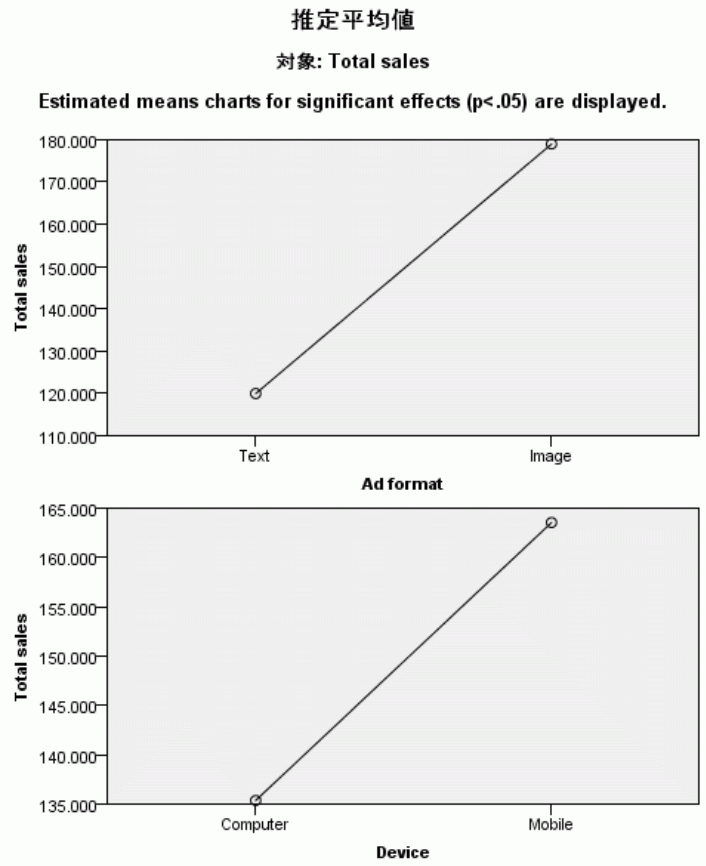
- **ダイアグラム**:まず定数項を示し、効果について予測値の重要度が大きいものから順に上から下に並べ替えて表示するグラフです。因子を含む効果内で、係数はデータ値が小さい順に並べ替えられます。ダイアグラムで繋がった線は、係数の符号（ダイアグラムのキーを参照）に基づいて色分けされ、係数の有意確率に基づいて重みがつけられます。線の太いほど係数の有意確率は大きくなります（p 値は小さくなります)。繋がった線にマウスを置くと、係数の値、p 値、パラメータが関連する効果の重要度を示すツールヒントが表示されます。これはデフォルトのスタイルです。
- **テーブル。**各モデル係数の値、有意性検定、信頼区間が表示されます。定数項の後、予測値の重要度が大きいものから順に上から下に並べ替えられます。因子を含む効果内で、係数はデータ値が小さい順に並べ替えられます。デフォルトでは、テーブルが折りたたまれ、各モデル パラメータの係数、有意度、重要度だけが表示されます。標準誤差、t 統計量、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。テーブルのモデル パラメータの名前にマウス ポインタを置くと、パラメータの名前、パラメータが関連する効果、そしてカテゴリ型予測値の場合は、モデル パラメータに関連する値のラベルを示すツールヒントが表示されます。自動データ準備がカテゴリ型予測値の同様のカテゴリを結合する時に作成された新しいカテゴリ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] スライダーで、ビュー内に表示された予測値を制御します。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焦点を当てることができます。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] スライダーで、予測値の重要度に基づいて表示される係数のほか、ビュー内に表示された係数を詳細に制御します。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処理される係数はありません。

推定平均値

図 15-16
[推定平均値] ビュー



有意な予測フィールドについて表示するグラフです。縦軸に対象フィールドのモデル推定値を、横軸に予測フィールドの各値を示し、ほかのすべての予測フィールドを一定にして表示します。対象フィールドに対する各予測フィールドの係数の効果について、役立つ視覚化を提供します。

注：予測値が重要でない場合、推定平均値は生成されません。

モデル構築の要約

図 15-17
[モデル構築の要約] ビュー、変数増加ステップワイズ法アルゴリズム

モデル構築の集計

対象: Total sales

		ステップ						
		1	2	3	4	5	6	7
情報基準		11,949.413	11,597.758	11,347.000	11,118.878	10,965.287	10,816.338	10,803.021
効果	secondary_keyword_transformed	✓	✓	✓	✓	✓	✓	✓
	ad_format		✓	✓	✓	✓	✓	✓
	primary_keyword_transformed			✓	✓	✓	✓	✓
	age_category_transformed				✓	✓	✓	✓
	promotion_transformed					✓	✓	✓
	device						✓	✓
	gender							✓

モデル作成方法は、情報基準を使用する変数増加ステップワイズ法です。
チェックマークは、効果がこのステップのモデルにあることを意味します。

[モデルの選択] 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、モデル構築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に選択されると、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、モデルの選択基準の値と効果が表示されます。モデルに対する各ステップの寄与度を表します。各列で行を並べ替え、指定したステップのモデルの効果をより用意に確認できます。

最適サブセット: 最適サブセットが選択アルゴリズムに選択されると、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、モデルの選択基準の値と効果が表示されます。上位モデルの安定性について表示されます。相違点が少ない類似した効果が多い場合、「上位の」モデルは信頼できます。非常に異なる効果がある場合、一部の効果は非常に類似している場合があり、結合するか一方を削除する必要があります。各列で行を並べ替え、指定したステップのモデルの効果をより用意に確認できます。

章
16

線型回帰

線型回帰では、線型方程式の係数を推定します。線型方程式には 1 つ以上の独立変数があり、従属変数の値を最適に予測します。たとえば、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、勝利ゲームが増加すると、相手チームの平均得点が減少します。線型回帰では、変数間の関係をモデリングできます。適当なモデルを使用すると、チームの勝利ゲーム数を予測できます。

統計量。各変数: 有効ケース数、平均値および標準偏差。各モデル: 回帰係数、相関行列、部分相関および偏相関、多重回答、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、および残差。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。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。

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。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。従属変数と各独立変数の関係は線型であり、すべての観測が独立している必要があります。

線型回帰分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > 回帰 > 線型...

© Copyright SPSS Inc. 1989, 2010

図 16-1
[線型回帰] ダイアログ ボックス

▶ [線型回帰] ダイアログ ボックスの変数のリストから数値型の従属変数を選び、[従属変数] ボックスに移動します。

▶ 変数のリストから 1 つ以上の数値型の独立変数を選び、[独立変数] ボックスに移動します。

オプションとして、次の選択が可能です。

- 独立変数を幾つかのブロックに分け、変数の組み合せを変えて投入方法を指定する。
- ケース選択変数を選んで、その変数の特定の値を持つケースだけに限定して分析を行う。
- ケースを識別する変数を選んで、プロット上の点を識別する。
- [WLS 重み] に数値型変数を選択して、重み付き最小 2 乗法分析を行う。

WLS. 重み付き最小 2 乗法モデルを利用できます。データ点には、その変数の逆数で重みが付けられます。これは、分散の大きな観測は小さな分散に関連する観測値より分析に与える影響が小さいことを意味しています。重み付け変数の値がゼロ、負、または欠損の場合、ケースは分析から除外されます。

線型回帰の変数選択方法

[方法] ボックスでの選択により、独立変数を分析に投入する方法を指定できます。変数は同じ組み合せのまま方法を変えることで、さまざまな種類の回帰モデルを求めることができます。

- **強制投入法 (回帰).** 変数選択のための手順で、ブロック内のすべての変数は、1 つのステップで投入されます。
- **ステップワイズ法.** 各ステップにおいて、回帰式にない独立変数の F 値の有意確率が十分小さければ、この変数は入力されます。すでに回帰式に含まれている独立変数でも、F 値確率が十分に大きくなると除去されます。入力や除去の対象の変数がなくなると、この方法は終わります。
- **強制除去法.** ブロックの中の全ての変数は、変数の選択法によって、ある 1 つのステップで、強制除去されます。
- **変数減少法.** すべての変数を等式に入力してから順番に除去していく変数選択の手順。従属変数と最も小さい偏相関を持つ変数が、最初に除去する対象となります。その変数が除去するための基準を満たす場合は除去されます。最初の変数が削除されると、等式内に残っている変数のうち、最も小さい偏相関を持つ変数が次の対象となります。等式内に除去基準を満たす変数がなくなると、手順きは終了します。
- **変数増加法.** 変数を順番にモデルに入力していく、ステップごとの変数選択の手順。方程式に最初に入力する変数は、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、入力基準を満たしている場合に限って方程式に入力されます。最初の変数が入力されると、次は方程式に含まれていない独立変数のうち、最も大きい偏相関を持つものが対象となります。入力基準を満たす変数がないと、この手続きは中止します。

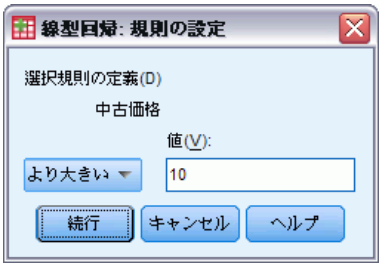
出力内の有意値は、1 つのモデルの適合に基づいたものです。したがって、ステップワイズの方法 (ステップワイズ法、変数増加法、変数減少法) を使用したときには、一般に、有意値は無効になります。

指定した投入方法に関係なく、変数を回帰式に投入するには、その変数が許容基準を満たしている必要があります。デフォルトの許容水準は 0.0001 です。また、ある変数を投入すると、すでにモデルに投入されてい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。しかし、変数の組み合せを変えて、変数の投入方法を指定することもできます。たとえば、1 番目の変数のブロックをステップワイズ法で回帰モデルに投入し、2 番目のブロックを変数増加法で投入することもできます。2 番目の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

図 16-2
[線型回帰: 規則の設定] ダイアログ ボックス



選択規則により定義されたケースが分析に含まれます。たとえば、変数に対して [**等しい**] を選択し、[値] ボックスに 5 を入力すると、選択された変数の値に 5 を持つケースだけを分析で使用します。文字列も指定できます。

線型回帰の作図

図 16-3
[線型回帰: 作図] ダイアログ ボックス

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のに役立ちます。また、外れ値、異常な観測値、影響力の大きいケースを発見する上でも有用です。予測値、残差、その他の診断を新変数として保

存すると、それらの変数をデータ エディタで使用できるようになり、独立変数を使用して作図できます。次の作図が利用できます。

[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、またはスチューデント化され削除された残差から 2 つプロットできます。標準化予測値に対して標準化残差を作図すると、線型性と等分散性を確認できます。

ソース変数リスト. 従属変数（DEPENDNT）、および予測変数および残差変数（標準化予測値（*ZPRED)、標準化残差（*ZRESID)、削除ケース残差（*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)）を表示します。

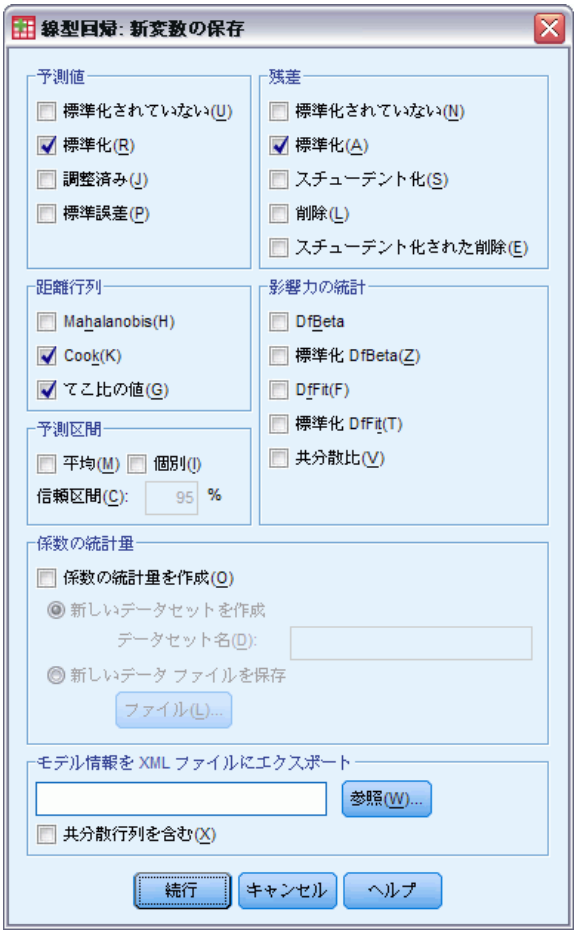
全ての偏残差の散布図を作成。独立変数と従属変数の両変数を残りの独立変数から離れて回帰させて、各独立変数の残差と従属変数の残差の散布図を表示します。偏残差プロットを作成するには、2 つ以上の独立変数をモデル内に投入する必要があります。

[標準化残差のプロット]標準化残差の分布を正規分布と比較するために、ヒストグラムと正規確率プロットを作成できます。

どの作図を選んでも、標準化予測値と標準化残差（*ZPRED と *ZRESID）の要約統計量が表示されます。

線型回帰: 新変数の保存

図 16-4
[線型回帰: 新変数の保存] ダイアログ ボックス



予測値、残差、および診断に役立つその他の統計量を保存します。各統計量の選択により、1 つ以上の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。各ケースに対して回帰モデルが予測する値です。

- **標準化されていない(N).** 従属変数を予測するモデルの値。
- **標準化(A).** 各予測値を標準化された形式に変換したもの。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は 0 で、標準偏差は 1 です。
- **調整済み(J).** 回帰係数の計算に含まれないケースの予測値。
- **標準誤差(P).** 予測値の標準誤差。独立変数の同じ値を持つケースに対する従属変数の平均値の標準偏差の推定値。

距離。回帰モデルに大きな影響をもたらす可能性がある、独立変数とケースの値の異常な組み合せを伴うケースを識別する測定です。

- **Mahalanobis(H).** 独立変数のケースの値がケースの平均からどの程度離れているかを測るもの。大きい Mahalanobis の距離は、1 つ以上の独立変数に極値を持つケースを特定します。
- **Cook(K).** 特定のケースが回帰係数の計算から除外された場合に、すべてのケースの残差がどのくらい変化するかを示す測定量。Cook の距離が大きいときは、回帰統計量の計算からケースを除外したことが係数を実質的に変化させたことを示しています。
- **てこ比の値(G).** 回帰の適合性に対する 1 つの点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両方の上限と下限です。

- **平均.** 平均予測応答の予測区間に対する下限と上限 (2 つの変数)。
- **個別(I).** 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 つの予測区間の値に対する信頼水準を指定するために、1 から 99.99 の値を入力します。この値を入力する前に、平均値または個別値が選択されている必要があります。一般的な信頼区間値は 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です。

- **標準化されていない(N).** 観測値およびモデルで予測された値との差。
- **標準化(A).** 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼ばれ、平均は 0 で、標準偏差 1 になります。
- **スチューデント化(S).** 残差を、独立変数の平均値からの独立変数の各ケース値の距離に依存して、ケースごとに違う標準偏差の推定量で割ったもの。
- **削除(L).** あるケースが回帰係数の計算から除外されたときのケースの残差。従属変数と調整済み予測値の間の差です。
- **スチューデント化された削除(E).** 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間の差は、それ自身の予測によって、ケースの削除によりどの程度の差が生じるかを示します。

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) を使用できます。標準化 DfBeta 値および標準化 DfFit 値も共分散比とともに使用できます。

- **DfBeta(s).** ベータ値の差は、特定のケースの除外から発生する回帰係数の変化です。値は、モデル内の各項に対して定数項を含めて計算されます。

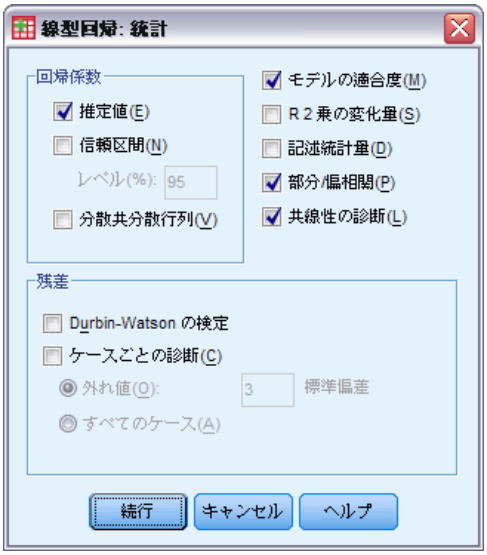
- **標準化 DfBeta.** ベータ値の標準化された差。特定のケースの除去によって発生する回帰係数の変化。絶対値が 2 より大きいケースを N の平方根で割って調べることができます (N はケースの数)。値は、モデル内の各項に対して定数項を含めて計算されます。
- **DfFit.** 当てはめ値の差は、特定ケースの除外から発生する予測値の変化です。
- **標準化 DfFit.** 当てはめ値の標準化された差。特定ケースの除去によって発生する予測値の変化。2 を超える絶対値に p/N の平方根を掛けた値である標準化値を調べることができます (p はモデル内のパラメータの数、N はケースの数です)。
- **共分散比(V).** すべてのケースの分散共分散行列の行列式に対する、回帰係数の計算からあるケースが除外された場合の分散共分散行列の行列式の比。この比率が 1 に近い場合、そのケースは分散共分散行列に大きな影響を及ぼしていません。

係数統計量。回帰係数をデータセットやデータ ファイルに保存します。データセットは、同じセッションの今後で利用可能ですが、セッション終了前に明示的に保存しない限り、保存されません。データセット名は、変数命名規則に従う必要があります。

モデル情報を XML ファイルにエクスポート。パラメータ推定とその共分散（オプション）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、得点付け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。

線型回帰の統計

図 16-5
[統計] ダイアログ ボックス



次の統計を使用できます。

回帰係数。[**推定値**] は、回帰係数 B、B の標準誤差、標準化係数ベータ、B の t 値、および t の両側有意確率を表示します。[**信頼区間**] には、各回帰係数または分散協分散行列の指定された信頼係数で信頼区間を表示します。[**分散共分散行列**] では、回帰係数の分散共分散行列を、共分散は対角線外に、分散は対角線上に表示します。相関行列も表示します。

モデルの適合度。モデルから投入および削除された変数が表示され、多重回答、R^2、および調整済み R^2、推定値の標準誤差、および分散分析表の適合度統計量が表示されます。

[R 2 乗の変化量] 独立変数を加えたり取り除いたりすることによって生じる R^2 統計量の変化。ある独立変数についてこの R^2 統計量の変化が大きければ、その独立変数は従属変数をよく説明する予測変数であると見なすことができます。

記述統計。分析での各変数に対する有効ケース数、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とともに、相関行列が表示されます。

偏相関. 2 つの変数の間から他の変数の関係を取り除いた後で、それら 2 つの変数間に残った相関。従属変数と独立変数の両方からモデル内の他の独立変数の線型効果が取り除かれた後の、独立変数と従属変数の間の相関。

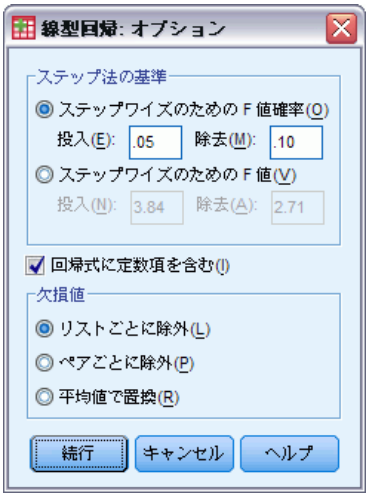
部分相関 (ピボットテーブル 回帰). 独立変数からモデル内の他の独立変数の線型効果が取り除かれた後の、独立変数と従属変数の間の相関。変数が方程式に追加されるときに 2 乗された R の変化と関連します。セミパーシャル相関とも呼ばれます。

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、選択基準に合う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

図 16-6
[線型回帰: オプション] ダイアログ ボックス



次のオプションを使用できます。

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、または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できます。F 値 の有意確率または F 値自身のいずれかを、モデルへの変数の投入や除去に使用します。

- **ステップワイズのための F 値確率.** F 値の有意確率が [投入] の値よりも小さい場合、変数はモデルに入力され、[除去] の値よりも大きい場合除去されます。[投入] の値は [除去] の値より小さい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモ

デルに投入するには、[投入] 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] 値を下げてください。

- **ステップワイズのための F 値.** 変数は、その F 値が [投入] 値よりも大きい場合にモデルに投入され、[除去] よりも小さい場合に除去されます。[投入] の値は [除去] の値より大きい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 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] 値を上げてください。

回帰式に定数項を含む。デフォルトでは、回帰モデルに定数項が含まれます。このチェック ボックスをオフにすると、通常は使用しない原点を通る回帰になります。原点を通るようにした回帰と、定数を含むようにした回帰の結果は異なります。たとえば、R^2 を通常の方法で解釈できなくなります。

欠損値。次のオプションのうち 1 つを選択できます。

- **リストごとに除外。**分析で使うすべての変数が有効な値であるケースだけを使用します。
- **ペアごとに除外。**相関している変数のペアが両方とも完全なデータであるケースを使用して、回帰分析の基礎となる相関係数を計算します。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、すべてのケースを計算に使用します。

REGRESSION コマンドの追加機能

コマンド シンタックス言語を使用して、次のことも実行できます。

- 相関行列の書き出しや、生データの代替として行列を読み込むことによる回帰分析の取得 (MATRIX サブコマンド使用)。
- 許容度の指定 (CRITERIA サブコマンド使用)。
- 同じ従属変数や、異なる従属変数に対する複数モデルの取得 (METHOD と DEPENDENT サブコマンド使用)。
- 追加の統計量の取得 (DESCRIPTIVES と STATISTICS サブコマンド使用)。

複雑なシンタックス情報については、「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。

章
17

順序回帰

順序回帰分析では、一連の予測変数における、多分割順序応答の従属性をモデリングできます。予測変数は因子または共変量となります。順序回帰分析のデザインは、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 と呼んでいます。

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せとの差を 2 乗したものの合計を最小化します。推定された係数は、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、応答の水準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。たとえば、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 の人の身長差と意味が同じです。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、応答カテゴリの選択と数はきわめて恣意的です。

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。反応は、「なし」、「穏やか」、「適度」、「激しい」に分類できます。「穏やか」と「適度」の反応の違いは、数値化が困難、また不可能なため、知覚に基づいて判別されます。さらに言えば、「穏やか」と「適度」の応答の差は、「適度」と「激しい」の応答の差よりも大きい場合もあり、または逆に小さい場合もあります。

統計量と作図。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の観測累積確率と期待累積確率、パラメータ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、反復の記述、平行線の仮定の検定、パラメータ推定値、標準誤差、信頼区間、Cox と Snell の R2 乗 統計量、Negalkerke の R2 乗 統計量、McFadden の R2 乗 統計量 。

データ。従属変数は、順序変数であると仮定され、数値または文字型のどちらかです。順序は、従属変数の値を昇順で並べ替えることにより決定されます。最低値によって、最初のカテゴリが定義されます。因子変数は、カテゴリ型と仮定されます。共変量変数は数値型でなければなりません。複数の連続共変量があると、非常に大きなセル確率テーブルが容易に作成されてしまうことに注意してください。

© Copyright SPSS Inc. 1989, 2010

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、応答は独立多項分布変数であると仮定されます。

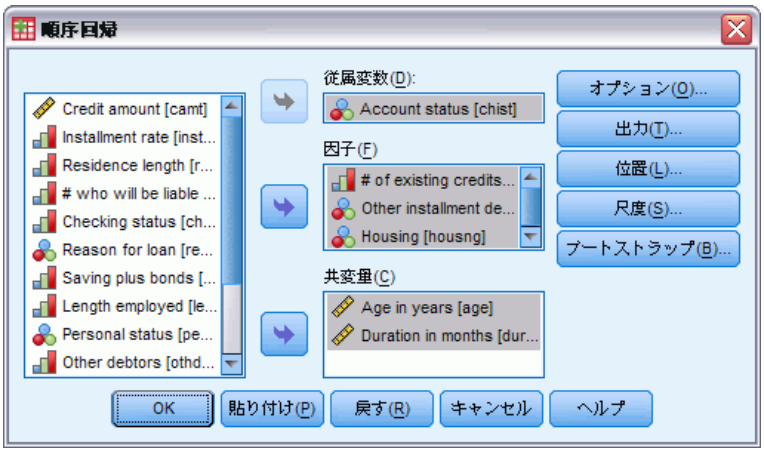
関連手続き。名義ロジスティック回帰分析では、名義従属変数に対して似たモデルを使用します。

順序回帰の作成

▶ メニューから次の項目を選択します。

分析(A) > 回帰 > 順序...

図 17-1
[順序回帰] ダイアログ ボックス



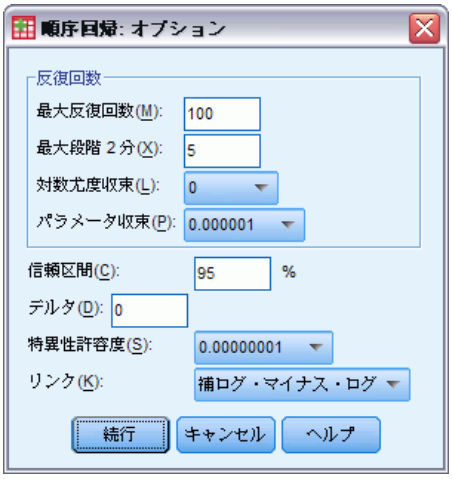
▶ 従属変数を 1 つ選択します。

▶ [OK] をクリックします。

順序回帰分析のオプション

[オプション] ダイアログ ボックスでは、反復推定アルゴリズムで使用されるパラメータの調整、パラメータ推定値の信頼水準の選択、リンク関数の選択を行うことができます。

図 17-2
[順序回帰: オプション] ダイアログ ボックス



反復回数。反復アルゴリズムをカスタマイズできます。

- **最大反復回数。**負でない整数を指定してください。0 を指定した場合、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ければ停止します。0 を指定した場合、この収束基準は使用されません。
- **パラメータ収束。**このアルゴリズムは、それぞれのパラメータ推定値の絶対変化または相対変化がこの値よりも小さければ停止します。0 を指定した場合、この収束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 未満の負でない値を指定します。

特異性許容度。従属性の高い予測変数かどうかを調べる場合に使用します。オプションのリストから値を選択します。

リンク関数。リンク関数とは、モデルの推定を行えるように累積確率を変換するためのものです。利用できるリンク関数は次の表に示す 5 種類です。

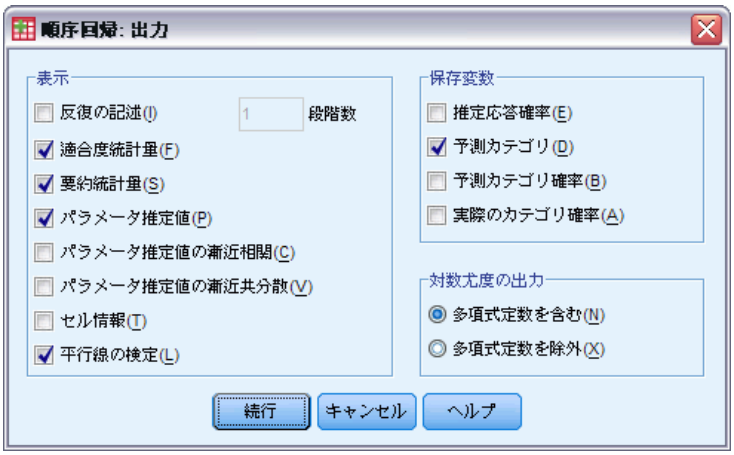
関数（スクリプト ウィンドウ、新規手続き）	形式	代表的な適用例
ロジット	$\log(\xi / (1-\xi))$	分布が均一なカテゴリ
補ログ・マイナス・ログ	$\log(-\log(1-\xi))$	順序の高いカテゴリがより確率が高い
負ログ・マイナス・ログ	$-\log(-\log(\xi))$	順序の低いカテゴリがより確率が高い

関数（スクリプト ウィンドウ、新規手続き）	形式	代表的な適用例
プロビット	$\Phi^{-1}(\xi)$	潜在的変数が正規分布する
コーチット（コーシーの逆関数）	$\tan(\pi(\xi-0.5))$	潜在的変数に多数の外れ値が存在する

順序回帰分析の出力

[出力] ダイアログ ボックスでは、ビューアに表示するテーブルを作成し、変数を作業ファイルに保存できます。

図 17-3
[順序回帰: 出力] ダイアログ ボックス



表示。次のテーブルが作成されます。

- **反復の記述。**指定された出力反復頻度で、度数対数尤度およびパラメータ推定値を出力します。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 および尤度比カイ 2 乗統計量。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 R2 乗 統計量、Nagelkerke の R2 乗 統計量、および McFadden の R2 乗 統計量 。
- **パラメータ推定値。**パラメータ推定値、標準誤差、および信頼区間。
- **パラメータ推定値の漸近相関。**パラメータ推定相関係数の行列。
- **パラメータ推定値の漸近共分散。**パラメータ推定共分散の行列。
- **セル情報。**観測度数、期待度数、累積度数、度数と累積度数に対する Pearson 残差、観測確率、期待確率、および共変量パターンによる各応答カテゴリの観測累積度数と期待累積度数。多数の共変量パターンを含むモデル（たとえば、連続共変量を含むモデル）では、このオプ

ションにより、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ことに注意してください。

- **平行線の検定。**位置パラメータが従属変数の水準にわたって等価であるという仮説の検定。これは、位置だけのモデルでしか使用できません。

保存変数。次の変数を作業ファイルに保存します。

- **推定応答確率。**因子/共変量パターンを応答カテゴリに分類するモデル推定確率。応答カテゴリの数だけ確率があります。
- **予測カテゴリ。**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。
- **予測カテゴリ確率。**因子/共変量パターンを予測カテゴリ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。
- **実カテゴリ確率。**因子/共変量パターンを実際のカテゴリに分類する推定確率。

対数尤度を出力。対数尤度の表示を制御します。[**多項式定数を含む**] により、尤度の完全な値が求められます。その定数を含まない積について結果を比較するには、除外を選択します。

順序回帰分析の位置モデル

[位置] ダイアログ ボックスでは、分析する位置モデルを指定できます。

図 17-4
[順序回帰: 位置] ダイアログ ボックス

モデルの指定。主効果モデルには、共変量および因子主効果が含まれますが、交互作用効果は含まれません。ユーザーの指定によるモデルを作成して、因子交互作用または共変量交互作用のサブセットを指定できます。

因子/共変量。要素および共変量はリストされます。

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 ロジット モデル

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。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

順序回帰分析の尺度モデル

［尺度］ダイアログ ボックスでは、分析する尺度モデルを指定できます。

図 17-5
［順序回帰: 尺度］ダイアログ ボックス

因子/共変量。要素および共変量はリストされます。

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 ロジット モデル

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。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

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、PLUM コマンド シンタックスを編集することで、順序回帰分析をカスタマイズできます。コマンド シンタックスを使用すると、次の作業も実行できます。

- 帰無仮説をパラメータの線型結合として指定することによって、カスタマイズした仮説の検定を作成。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
18

曲線推定

［曲線推定］手続きは、曲線推定の回帰統計量と 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。個々のモデルは、従属変数ごとに作成されます。また、予測値、残差、および予測区間を新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 サービス会社が、自社のネットワーク上を流れるウイルスに感染した電子メールの割合を長期にわたり追跡しています。散布図では、その関係が非線形であることが示されています。2 次モデルまたは 3 次モデルをデータに当てはめ、仮定の有効性とモデルの適合度を確認します。

統計量。各モデル: 回帰係数、多重回答、R^2、調整済み R^2、推定値の標準誤差、分散分析表、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、複合成長、S 曲線、ロジスティック、成長、および指数。

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。独立変数としてアクティブなデータセットから（変数ではなく）**[時間]** を選択すると、［曲線推定］手続きにより、ケース間の時間の長さが一定の時間変数が生成されます。**[時間]** を選択した場合は、従属変数が時系列でなければなりません。時系列分析に使用するデータ ファイルでは、各ケース（行）が個別の時間の一連の観測値を表し、さらにケース間の時間の長さが一定でなければなりません。

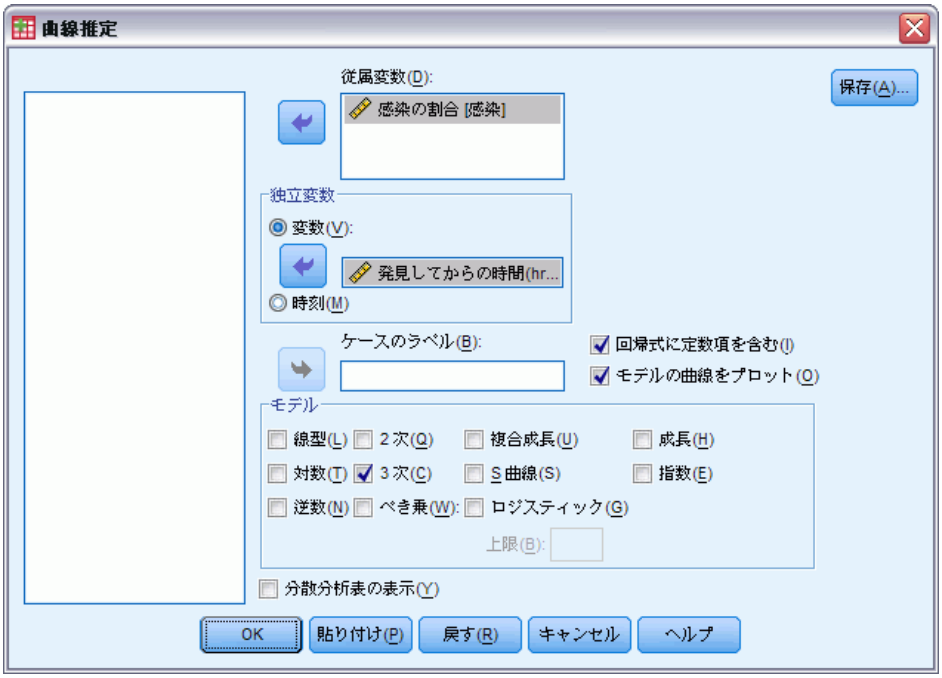
仮定。データをグラフ表示し、独立変数と従属変数の関係を判断します（線型、指数など）。理想的なモデルの残差は、ランダムに分布し、正規分布を示します。線型モデルを使用する場合は、次の 3 つの仮定を満たしている必要があります。すなわち、独立変数の各値に対して、従属変数の分布は正規分布でなければなりません。従属変数の分布の分散は、独立変数のすべての値に対して一定でなければなりません。従属変数と独立変数の関係は線型で、観測値はすべて独立でなければなりません。

曲線推定を行うには

▶ メニューから次の項目を選択します。
分析 > 回帰 > 曲線推定

© Copyright SPSS Inc. 1989, 2010

図 18-1
[曲線推定] ダイアログ ボックス



▶ 従属変数を 1 つ以上選択します。個々のモデルは、従属変数ごとに作成されます。

▶ 独立変数として、アクティブなデータセットの変数を 1 つ選択するか、[時間] を選択します。

▶ 次のオプションが選択できます。

- ［ケースのラベル］ボックスに、散布図のケースにラベル付けする変数を 1 つ選択する。散布図で［点の識別］ツールを使用すると、各点に［ケースのラベル］ボックスの変数の値を表示できます。
- [保存] をクリックして、予測値、残差、および予測区間を新変数として保存する。

次のオプションも選択できます。

- **回帰式に定数項を含む。**回帰式の定数項を推定します。デフォルトでは、このチェック ボックスはオンになっています。
- **モデルの曲線をプロット。**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図表が作成されます。
- **分散分析表の表示。**選択された各モデルに、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 つ以上の曲線推定の回帰モデルを選択できます。選択するモデルを決めるには、まずデータをプロットします。変数に線型関係があれば、単純な線型回帰モデルを使用します。線型関係がなければ、データを変換します。変換できない場合は、さらに複雑なモデルが必要となります。データの散布図を調べ、そのプロットがどれかの関数と似ていれば、データをそのモデルの型に当てはめます。たとえば、データが指数関数に似ている場合は、指数モデルを選択します。

線型 (ピボットテーブル GLM). 方程式が Y = b0 + (b1 * 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. Y = b0 + (b1 * ln(t)) のような方程式をもつモデルです。

逆数. Y = b0 + (b1 / t) のような方程式をもつモデルです。

2 次 (ピボットテーブル GLM). 方程式が Y = b0 + (b1 * t) + (b2 * t**2) であるモデル。2 次モデルは、上昇する系列か下降する系列をモデル化するために使うことができます。

3 次 (ピボットテーブル GLM). Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3) のような方程式で定義されるモデルです。

べき乗. Y = b0 * (t**b1) または ln(Y) = ln(b0) + (b1 * ln(t)) のような方程式をもつモデルです。

複合成長. Y = b0 * (b1**t) または ln(Y) = ln(b0) + (ln(b1) * t) のような方程式で表わされるモデルです。

S 曲線. Y = e**(b0 + (b1/t)) または ln(Y) = b0 + (b1/t) のような方程式をもつモデルです。

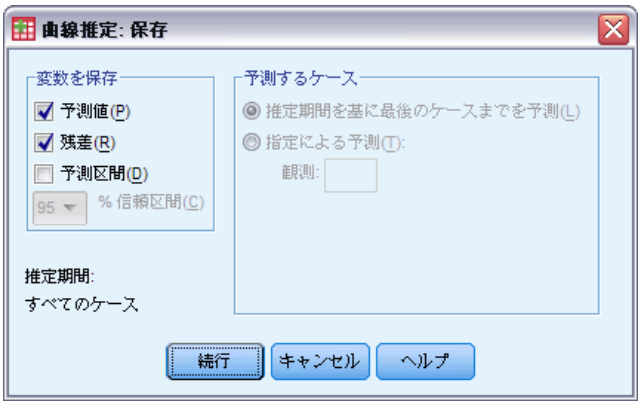
ロジスティック (正規確率プロット 検定分布). Y = 1 / (1/u + (b0 * (b1**t))) または ln(1/y-1/u) = ln (b0) + (ln(b1) * t) のような方程式をもつモデルです。u は、上限の値です。[ロジスティック] を選択したら、回帰式で使う上限の値を指定してください。この値は従属変数の最大値より大きい正の数である必要があります。

成長. Y = e**(b0 + (b1 * t)) または ln(Y) = b0 + (b1 * t) のような方程式をもつモデル。

指数. Y = b0 * (e**(b1 * t)) または ln(Y) = ln(b0) + (b1 * t) のような方程式をもつモデルです。

曲線推定の保存

図 18-2
[曲線推定: 保存] ダイアログ ボックス



変数を保存。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。新しい変数名と記述的なラベルが出力ウィンドウのテーブルに表示されます。

予測するケース。アクティブなデータセット内で、独立変数として変数の代わりに [時間] を選択した場合は、時系列の最終点以降の予測期間を指定できます。次のオプションのどちらかを選択できます。

- **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いてファイル内のすべてのケースに値を予測します。ダイアログ ボックスの下に表示される推定期間は、［データ］メニューの［ケースの選択］の［ケースの選択: 範囲の定義］ダイアログ ボックスで定義されたものです。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の予測が行われます。
- **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うことができます。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ボックスが異なります。日付変数が定義されていない場合は、［観測］ボックスで最終観測（ケース）数を指定します。

日付変数の作成には、［データ］メニューの［日付の定義］を使用します。

章
19

偏相関最小 2 乗法回帰

偏 2 段階最小 2 乗回帰手順は、偏 2 段階最小 2 乗（PLS とは「潜在的構造投影方法（projection to latent structure）のことです」）回帰モデルを推定します。PLS は、通常最小 2 乗法（OLS）回帰、正準相関、構造方程式モデリングに対する代わりの予測技術であり、予測変数が相関しているとき、あるいは予測数がケース数を超えているときは特に役立ちます。

PLS は、主成分分析および多重回帰を組み合わせた機能です。最初に従属変数と独立変数の間の共分散をできるだけたくさん説明する潜在的要素を展開します。そして、回帰手順は独立変数の分解を使用して従属変数の値を予測します。

使用方法。PLS は拡張コマンドです。PLS を実行する予定のシステム上にインストールされるべき IBM® SPSS® Statistics - Integration Plug-In for Python が必要です。PLS 拡張モジュールは別にインストールする必要があります。下記の URL からダウンロードしてください。
http://www.spss.com/devcentral

注 :PLS 拡張モジュールは Python ソフトウェアによって異なります。SPSS Inc. は Python ソフトウェアの所有者でもライセンス付与者でもありません。Python ユーザーは、Python Web サイトに記載されている、Python ライセンス許諾契約書の条件に同意する必要があります。SPSS Inc. は、Python プログラムの品質についていかなる説明も行いません。SPSS Inc. は、Python プログラムの使用にかかわる責任を一切負わないものとします。

テーブル。（潜在的因子により）説明された分散の比率、潜在的因子の重み、因子負荷、投影の独立変数の重要度（VIP）、回帰パラメータ推定値（従属変数による）はすべてデフォルトにより作成されます。

図。投影の変数の重要度（VIP）、因子得点、3 つの潜在的因子の最初の因子の重み、モデルへの距離はすべて［オプション］タブから作成されます。

測定レベル。従属変数および独立変数（予測）は尺度は、名義、順序、またはスケールのいずれかです。適切な尺度がすべてのジェン数に割り当てられたということを前提にする手順もありますが、ソース変数リストの変数を右クリックしてコンテキスト メニューから尺度を選択することで、その変数の尺度を一時的に変更できます。カテゴリ（名義または順序）変数は手順により等しく扱われます。

© Copyright SPSS Inc. 1989, 2010

カテゴリ変数のコード化。手順は、その手順の間 c コードの 1 つを使用してカテゴリ独立変数を記録します。変数の中に、c カテゴリがあるなら、変数は c ベクトルとして保存され、最初のカテゴリは (1, 0, . . . , 0) で示され、次のカテゴリは (0, 1, 0, . . . , 0)、というようになり最後のカテゴリ (0, 0, . . . , 0, 1) と示されます。カテゴリ従属変数は参照カテゴリに対応する指標を単に省略した形式のダミー コードを使用して表示されます。

度数による重み付け。重み変数は使用前にもっとも近い全数に丸められます。欠損重みまたは、重みが 0.5 以下のケースは分析に使用されません。

欠損値。ユーザーおよびシステム欠損値は無効として取り扱われます。

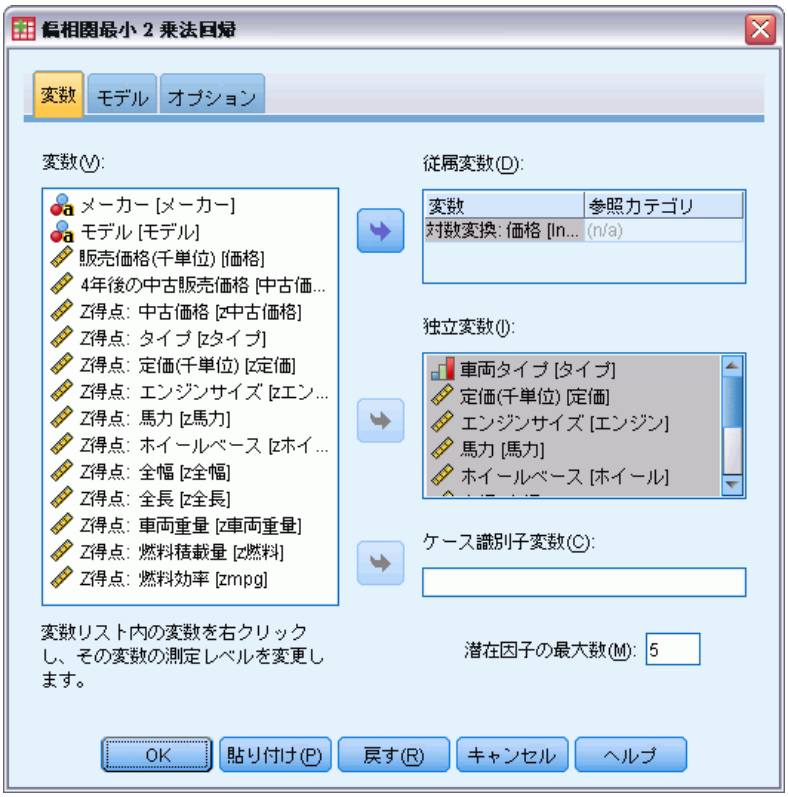
再調整。すべてのモデル変数は、カテゴリ変数を表す指示変数を含んで標準化および中心化されます。

偏 2 段階最小 2 乗回帰を取得するには次のようにします。

メニューから次の項目を選択します。

分析(A) > Regression > 偏相関最小 2 乗法...

図 19-1
[偏 2 段階最小 2 乗回帰変数] タブ



▶ 最低 1 つの従属変数を選択します。

▶ 最低 1 つの独立変数を選択します。

オプションとして、次の選択が可能です。

- カテゴリ（名義、順位1）従属変数の参照カテゴリを指定します。
- ケース主体の出力の一意の識別子として使用された、データセットを保存した変数を指定します。
- 表示される潜在的要素数の上限を指定します。

モデル

図 19-2
[偏 2 段階最小 2 乗回帰モデル] タブ

モデル効果を指定。主効果モデルには、共変量および因子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。モデルに含める項目はすべて指示する必要があります。

因子と共変量。要素および共変量はリストされます。

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**] をクリックすれば、分析対象の主効果と交互作用を選択できます。

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。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項目を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

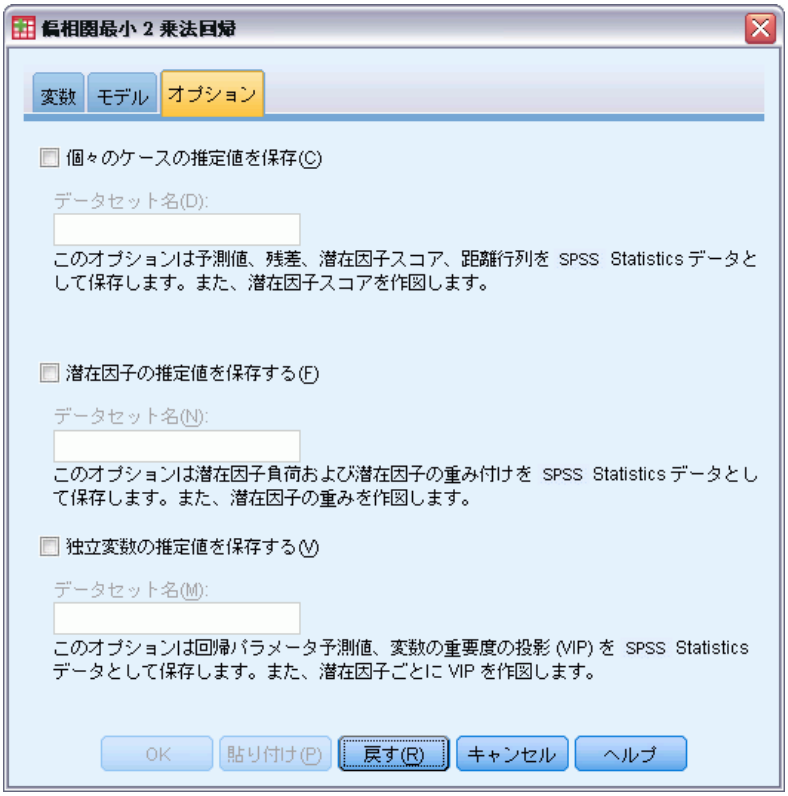
3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

オプション

図 19-3
[偏 2 段階最小 2 乗回帰オプション] タブ



[オプション] タブを押すことにより、ユーザーが各ケース、潜在的因子、および予測変数を保存し、プロットできるようになります。

データの各型は、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くてはいけません。既存のデータセットの名前を指定すると、内容は置き換えられ、新規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、潜在因子モデルへの距離、潜在因子スコアのケースごとのモデル推定値を保存します。潜在的因子得点をプロットします。
- **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。潜在的因子の重みをプロットします。
- **独立変数の推定値を保存。**回帰パラメータ推定値と投影変数の重要度 (VIP) を保存。潜在的因子により VIP をプロットします。

章
20

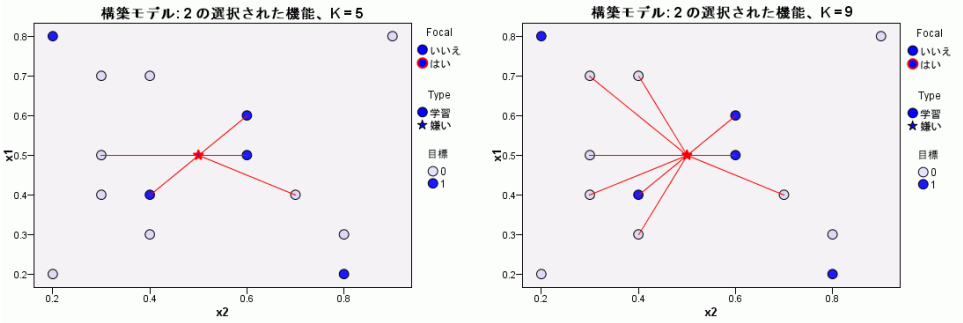
最近隣分析

最近隣分析は、他のケースに対する類似度に基づいてケースを分類する方法です。機械学習において、この方法は保存されたパターン、またはケースに対する正確な一致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。類似したケースはお互いに近く、類似していないケースはお互いに離れています。そのため、2 つのケース間の距離は、非類似度を示す尺度です。

お互いに近いケースは、「近隣」と呼ばれます。新しいケース（ホールドアウト）が示されると、モデルの各ケースからの距離が計算されます。最も類似したケース「最近隣」の分類が集計され、新しいケースは最大数の最近隣を含むカテゴリに投入されます。

ユーザーは、検証する最近隣の数を指定できます。値は k です。図では、新しいケースが 2 つの異なる値の k を使用してどのように分類するかを示します。k = 5 の場合、最近隣の大多数がカテゴリ 1 に属するため、新しいケースはカテゴリ 1 に投入されます。ただし、k = 9 の場合、最近隣の大多数がカテゴリ 0 に属するため、新しいケースはカテゴリ 0 に投入されます。

図 20-1
分類で k を変更した場合の効果



際近隣分析を使用して、連続型目標の値を計算することもできます。この場合、最近隣の平均または中央目標値を使用して、新しいケースの予測値を取得します。

目標および特徴。目標および特徴は次のとおりです。

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

  名義変数および順序変数は、最近隣分析によって同等に処理されます。適切な尺度が各変数に割り当てられたということを前提にする手順もありますが、ソース変数リストの変数を右クリックしてコンテキスト メニューから尺度を選択することで、変数の尺度を一時的に変更できます。

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

測定レベル	データの型			
	数値	文字列	日付	時刻
スケール（連続）		利用不可		
順序				
名義				

カテゴリ変数のコード化。この手順では、手順の期間に対する one-of-c コード化を使用してカテゴリ予測変数および従属変数を一時的に記録します。変数の c カテゴリが存在する場合、変数は最初のカテゴリ（1, 0, . . . , 0）、次のカテゴリ（0, 1, 0, . . . , 0）、. . . そして最後のカテゴリ（0, 0, . . . , 0, 1）が表示され、c ベクトルとして格納されます。

このコード化方式によって、特徴空間の次元数が増加します。特に、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値のカテゴリ数を合計したものです。結果として、このコード化方式によって、学習が遅くなる場合があります。最近隣学習の速度が遅い場合、学習を実行する前に類似したカテゴリを結合するか極端にまれなカテゴリをも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いています (分割 を参照)。そのため、ホールドアウト サンプルに学習データにはない予測変数カテゴリを持つケースがある場合、それらのケースはスコア化されません。ホールドアウト サンプルに学習データにはない従属変数カテゴリを持つケースがある場合、それら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いて行われます (「分割 」 (p. 147) を参照)。変数を指定して分割を定義する場合、特徴に学習サンプル、検定サンプル、ホールドアウト サンプル全体の類似した分布が含まれていることが重要です。 たとえば、[探索的分析] 手続きを使用して、分割全体の分布を検証します。

度数による重み付け。度数による重み付けは、この手続きによって無視されます。

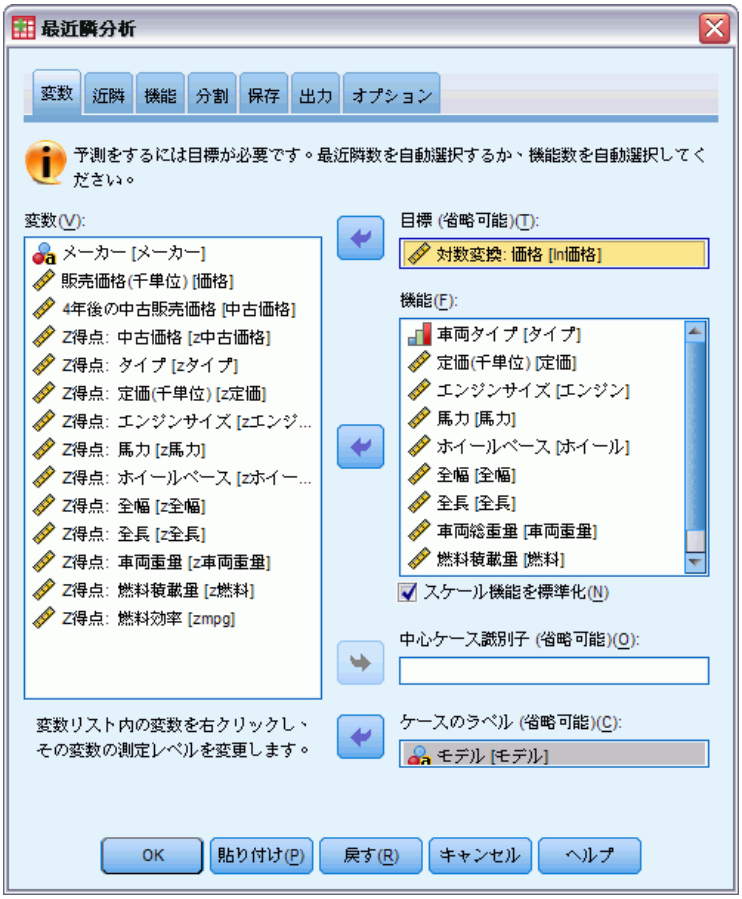
結果の再現 この手続きでは、分割の無作為割り当て時に乱数ジェネレータを使用します。結果を正確に複製する場合、同じ手続きの設定を使用するほか、Mersenne Twister のシード (「分割 」 (p. 147) 参照) を設定、または変数を使用して分割および交差検証群を定義します。

最近隣分析を取得するには

メニューから次の項目を選択します。

分析(A) > 分類 > 最近隣法(N)...

図 20-2
[最近隣分析: 変数] タブ



▶ 1 つまたは複数の特徴を指定し、目標がある場合独立変数または予測変数について考えられるようにします。

目標 (省略可能)。目標が指定されていない場合 (従属変数または応答)、手続きでは k 最近隣のみを検出します。分類または予測は実行されません。

スケール機能を標準化(N) 標準化された機能には同じ範囲の値があり、推定アルゴリズムのパフォーマンスを向上させます。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 ～ 1 です。

中心ケース識別子 (省略可能)(O) 特に重要なケースをマークすることができます。たとえば、研究者がある学区の検定スコア、中心ケースが同じような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ます。彼は最近隣分析を使用して、与えられたセットの特徴に関して最も近

い学区を検出します。そして、焦点となる学区の検定スコアと最近隣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象ケースを選択することもできます。中心ケースは、k 最近隣および距離の表、特徴空間図表、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。

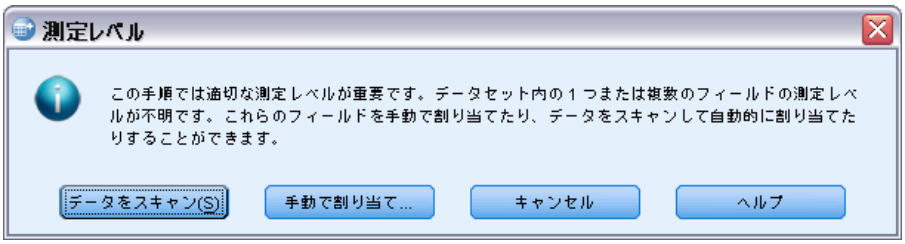
指定された変数の正の値を持つケースは、中心ケースとして処理されます。正の値を持たない変数を指定することはできません。

ケースのラベル (省略可能)(C) ケースは、特徴空間図、同位図、四分位分布図のこれらの値を使用して表示されます。

不明な尺度の項目

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

図 20-3
尺度の警告

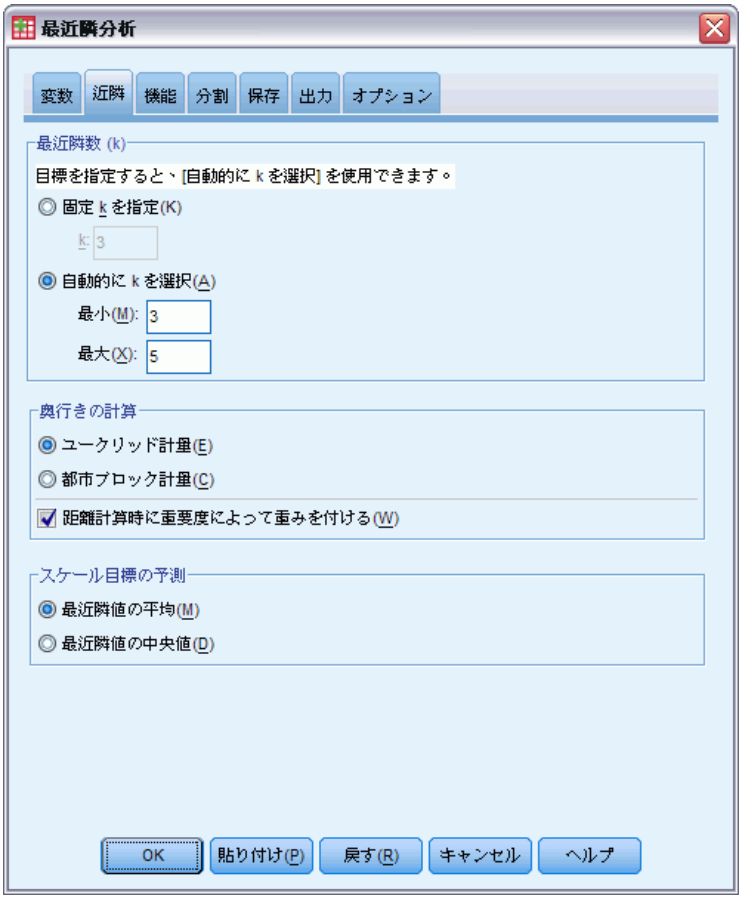


- **データをスキャン。** アクティブ データセットのデータを読み込み、デフォルトの尺度を尺度が現在不明なフィールドに割り当てます。データセットが大きい場合は時間がかかります。
- **手動で割り当てる。** 不明な尺度のフィールドをすべて表示するダイアログが開きます。このダイアログを使用して、尺度をこれらのフィールドに割り当てることができます。データ エディタの [変数ビュー] でも、尺度を割り当てることができます。

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

近隣

図 20-4
[最近隣分析: 近隣] タブ



最近隣数 (k) 最近隣数を指定します。より大きな数の近隣を使用すると、必ずしも正確なモデルが作成されるとは限りません。

目標が［変数］タブで指定されている場合、値の範囲を指定し、手続きで範囲内の「最適な」近隣数を選択することができます。最近隣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。

- 特徴選択が有効である場合、特徴選択は要求された範囲の k の各値に実行され、 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）の k および付随する特徴セットが選択されます。
- 特徴選択が有効でない場合、V 群交差検証を使用して、「最適な」近隣数を選択します。群の割り当てについては、コントロールの［データ区分］タブを参照してください。

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

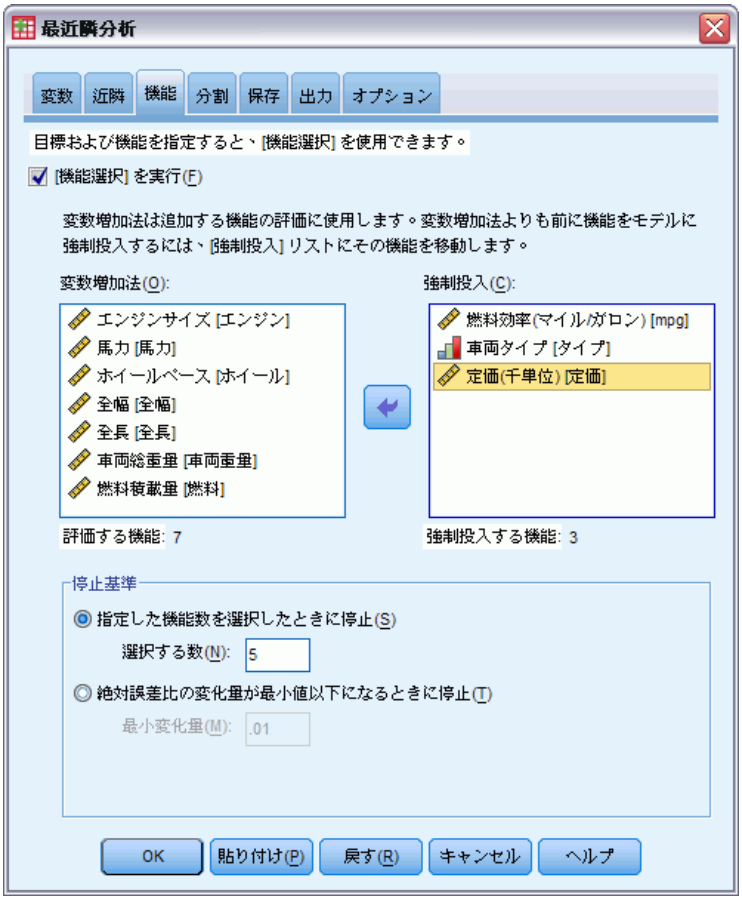
- **ユークリッド計量(E)** x および y の 2 つのケース間の距離は、すべての次元においてケースの値の間の差異を平方の合計の平方根です。
- **都市ブロック計量(C)** 2 つのケースの間の距離は、すべての次元の、それらのケースの値の絶対差の合計になります。Manhattan 距離とも呼ばれます。

オプションで、目標が［変数］タブで指定されている場合、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つけることができます。予測変数の特徴重要度は、予測変数をモデルからすべてのモデルの誤差率または誤差の平方和に移動して、誤差率の比率またはモデルの誤差の平方和によって計算されます。正規化された重要度は、合計が 1 となるよう、特徴重要度の値を再度重み付けして計算します。

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、予測値が平均値または最近隣の中央地のどちらに基づいて計算されるかを指定します。

特徴

図 20-5
[最近隣分析: 特徴] タブ



[特徴] タブを使用すると、目標が [変数] タブで指定されている場合に、特徴選択のオプションを要求および指定することができます。デフォルトでは、特徴選択にすべての特徴が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に強制することができます。

停止基準 各ステップで、モデルへの追加により誤差が最も小さくなる（カテゴリ目標の誤差率およびスケール目標の誤差の平方和として計算）特徴がモデル セットに選択すると見なされます。変数増加法は、指定された条件を満たすまで続行します。

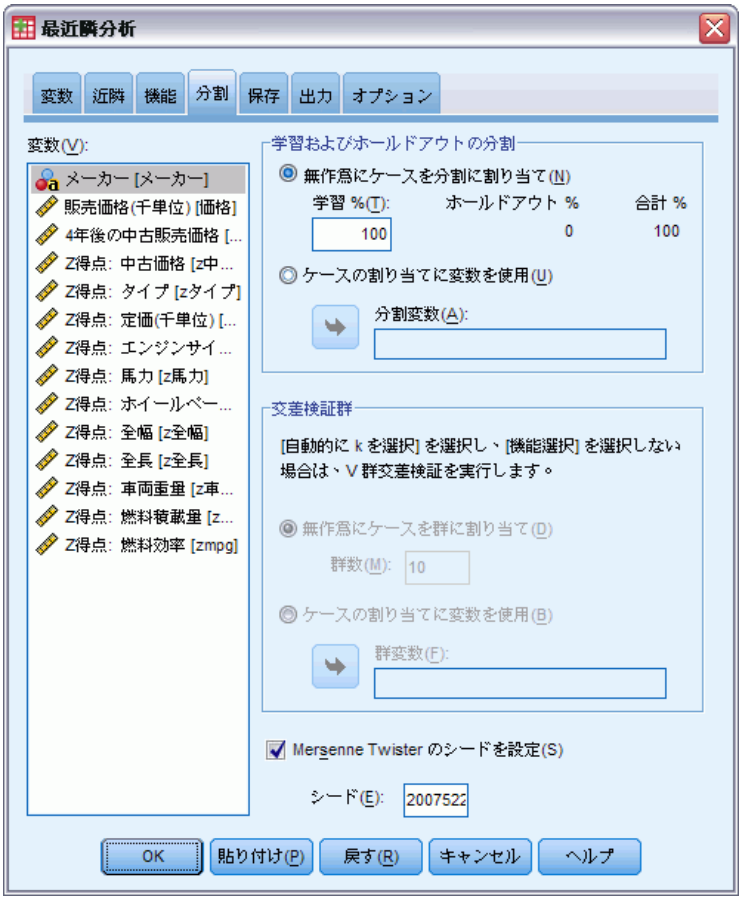
- **指定される特徴数** アルゴリズムでは、モデルに強制的に投入された特徴に加え、固定された特徴数を追加します。正の整数を指定します。選択する数値を減らすと、より節約的なモデルが作成され、重要な特徴が欠損するというリスクがあります。選択する数値を増やすと、す

べての重要な特徴を取得しますが、モデル誤差が増加する特徴を追加するというリスクがあります。

- **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す場合、アルゴリズムは停止します。正の数を指定します。最小変化量の値を小さくすると、より多くの特徴を選択しますが、モデルに多くの値を追加しない特徴を選択するというリスクがあります。最小変化量の値を大きくすると、より多くの特徴を除外しますが、モデルに重要な特徴を失うというリスクがあります。最小変化量の「最適な」値は、データおよび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か評価する方法については、出力の特徴選択エラー ログを参照してください。 詳細は、p. 161 特徴空間エラー ログ を参照してください。

分割

図 20-6
[最近隣分析: 分割] タブ



［データ区分］タブを使用して、データセットを学習セットおよびホールドアウト 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分割する方法をサンプルの学習およびホールドアウトに指定します。**学習サンプル**では、最近隣モデルを学習するために使用するデータ レコードを判断します。データセット内のケースのいくらかの割合は、モデルを取得するために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、最終のモデルを評価するために使用するデータ レコードの独立セットです。ホールドアウト ケースを使用してモデルを構築できなかったため、ホールドアウト サンプルの誤差によってそのモデルの予測能力を「公正に」評価します。

- **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割合を指定します。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の各ケースを学習サンプル、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、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、常に有効なものとして扱われます。

交差検証群V 群交差検証は、近隣の最適な数を決定するために使用されます。パフォーマンス上の理由で、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、サンプルを群と呼ばれる複数のサブサンプルに分割します。分割の後、最近隣モデルが生成されますが、各サブサンプルのデータは除外されます。つまり、最初のモデルは最初のサブサンプル以外のすべてのケースを基に生成され、2 番目のモデルは 2 番目のサブサンプル以外のすべてのケースを基に生成されます。それぞれのモデルを、そのモデルの生成時に除外したサブサンプルに適用し、誤差を推定します。最近隣の「最適な」数は、群全体で最も誤差が少ない数です。

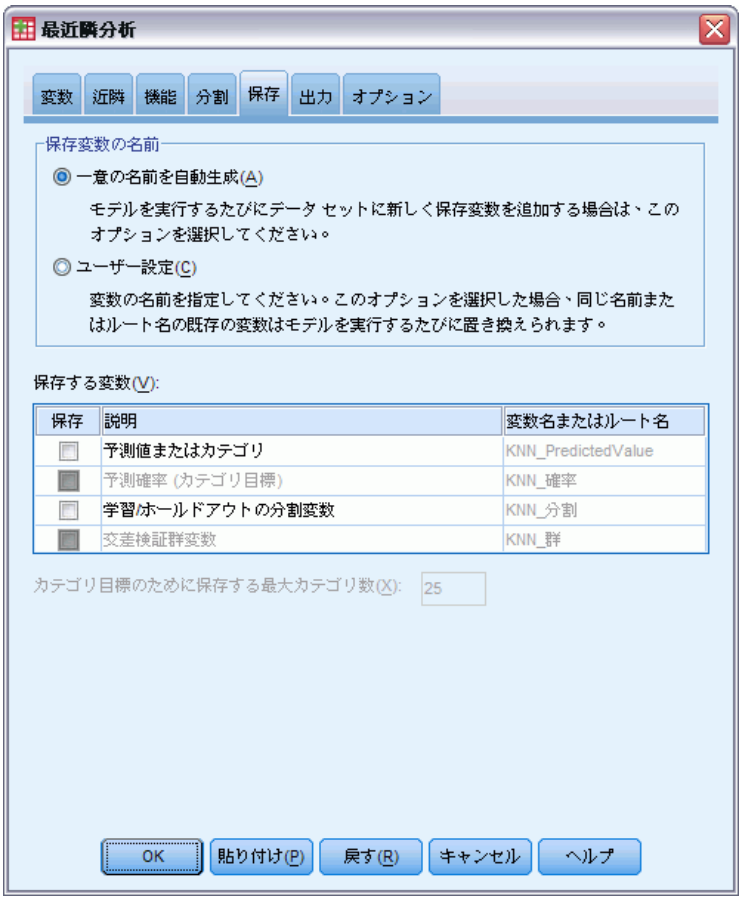
- **無作為にケースを群に割り当て** 交差健勝に使用する必要のある群の数を指定します。この手続きでは、1 から V（群の数）まで、ケースを群に割り当てます。
- **ケースの割り当てに変数を使用** アクティブなデータセットの各ケースを群に割り当てる数値型変数を指定します。変数は数値型で、1 ～ V までの値である必要があります。この範囲内の値が欠損している場合、分割に対して分割ファイルが有効である場合、誤差が生じます。

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 ジェネレータとして Mersenne Twister を設定し、［乱数ジェネレータ］ダイアログの固定開

始ポイントを指定することと同様の設定ができますが、このダイアログでシードを設定すると、乱数ジェネレータの現在の状態を保持し、分析が完了した後、その状態を復元します。

保存

図 20-7
[最近隣分析: 保存] タブ



保存する変数の名前 自動的な名前の生成によって、すべての作業を保存することができます。ユーザー指定の名前によって、Data Editor 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または置き換えることができます。

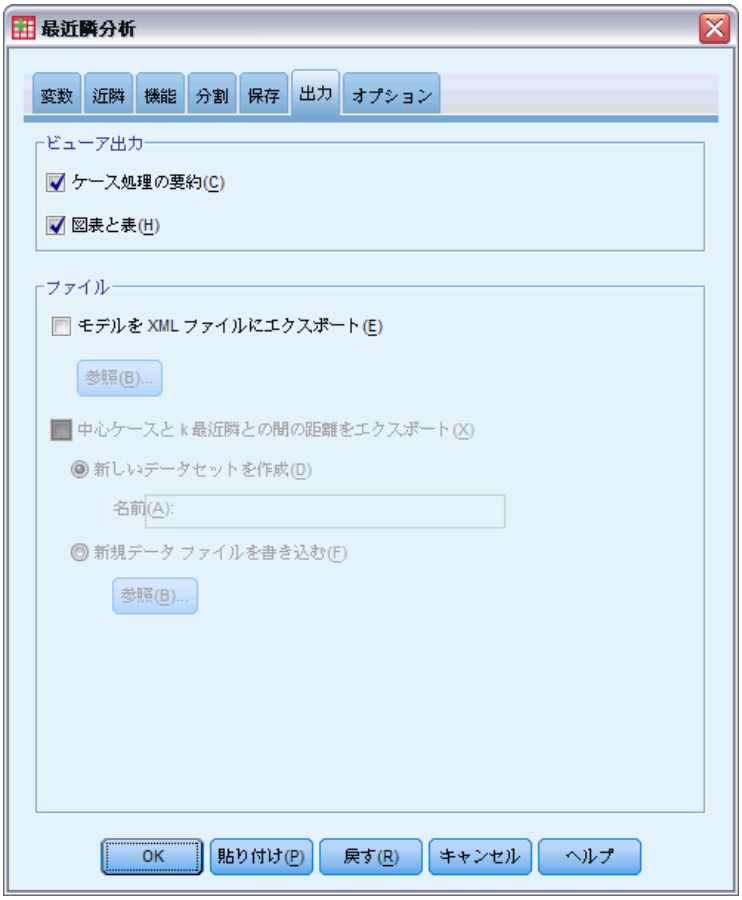
保存する変数

- **予測値またはカテゴリ** これにより、スケール目標に予測された値を保存し、カテゴリ目標に予測カテゴリを保存します。

- **予測確率** カテゴリ目標の予測確率を保存します。各変数は、それぞれの最初の n カテゴリに対して保存されます。この場合、 n は [**カテゴリ目標のために保存する最大カテゴリ数**] コントロールで指定されます。
- **学習/ホールドアウトの分割変数** ケースが [データ区分] タブで学習サンプルおよびホールド 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の値（学習またはホールドアウト）を保存します。
- **交差検証群変数** ケースが [データ区分] タブで交差検証群に割り当てられた場合、ケースが割り当てられた群の値を保存します。

出力

図 20-8
[最近隣分析: 出力] タブ



ビューア出力

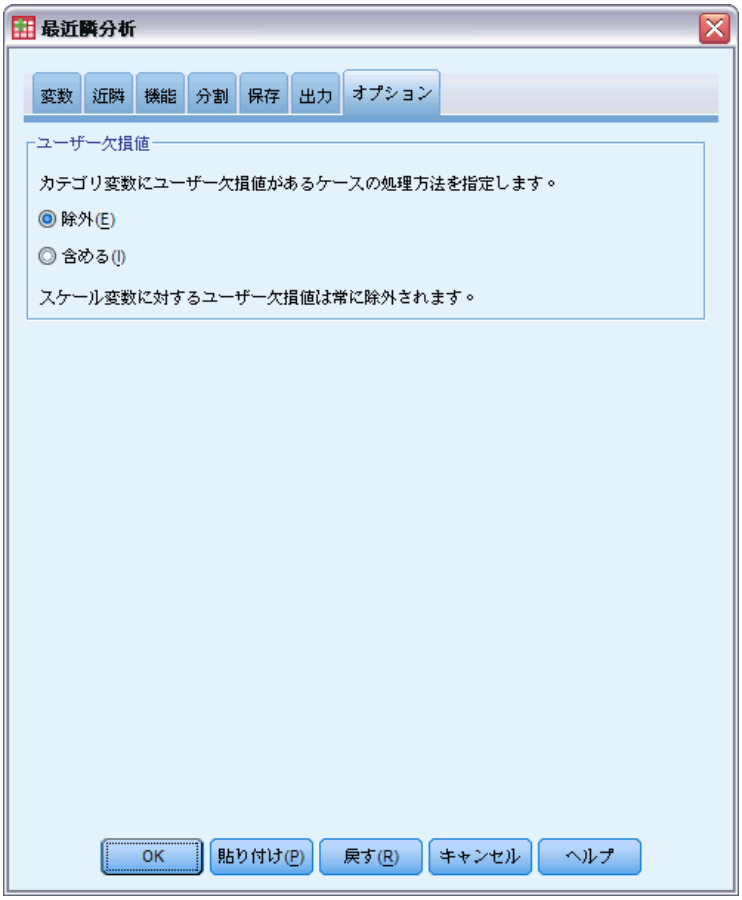
- **ケース処理の要約** 分析に含まれたケースおよび除外されたケースの数を全体、学習サンプルおよびホールドアウト サンプルごとに要約するケース処理要約テーブルを表示します。
- **図表と表** 表および図表など、モデルに関連する出力を表示します。モデル ビューの表には中心ケースの k 最近隣および距離、カテゴリ応答変数の分類、誤差の集計が表示されています。モデル ビューのグラフィカル出力には、選択エラー ログ、特徴重要度図表、特徴空間図表、同位図、四分位分布図があります。 詳細は、 p. 153 モデル ビュー を参照してください。

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、得点付け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 最近隣との間の距離をエクスポート** 中心ケースでは、各中心ケースの k 最近隣（学習サンプル）および対応する k 最短距離にそれぞれ値が作成されます。

オプション

図 20-9
[最近隣分析: オプション] タブ

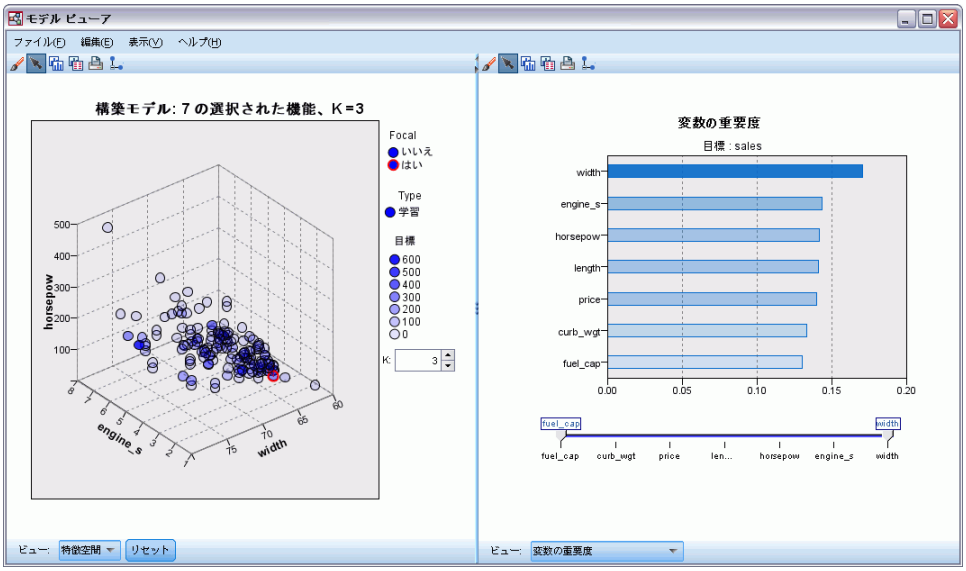


ユーザー欠損値。カテゴリ変数は、分析の対象となるケースに対して有効な値を取る必要があります。このオプションを使用すると、ユーザー欠損値をカテゴリ変数で有効な値として扱うかどうかを決定できます。

システム欠損値およびスケール変数の欠損値は常に無効なものとして処理されます。

モデル ビュー

図 20-10
最近隣分析のモデル ビュー



[出力] タブで [**図表および表**] を選択すると、手続きでは、ビューアに最近隣モデル オブジェクトが作成されます。このオブジェクトを有効化 (ダブルクリック) すると、モデルの双方向ビューを取得します。モデル ビューには、2 つのパネルのウィンドウがあります。

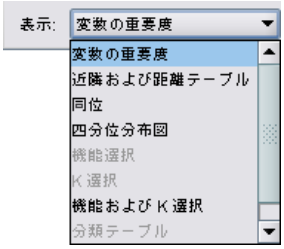
- 1 つめのパネルはメイン ビューと呼ばれ、モデルの概要が表示されます。
- 2 つめのパネルには、次の 2 種類のビューのいずれかが表示されます。

  モデルの詳細を表示するが、モデル自体に焦点を当てていない補助的モデル ビュー。

  ユーザーがメイン ビューの一部について掘り下げた場合、モデルのある特徴についての詳細を示すリンク ビュー。

デフォルトでは、1 つめのパネルで特徴空間を示し、2 つめのパネルで変数の重要度グラフを表示します。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。

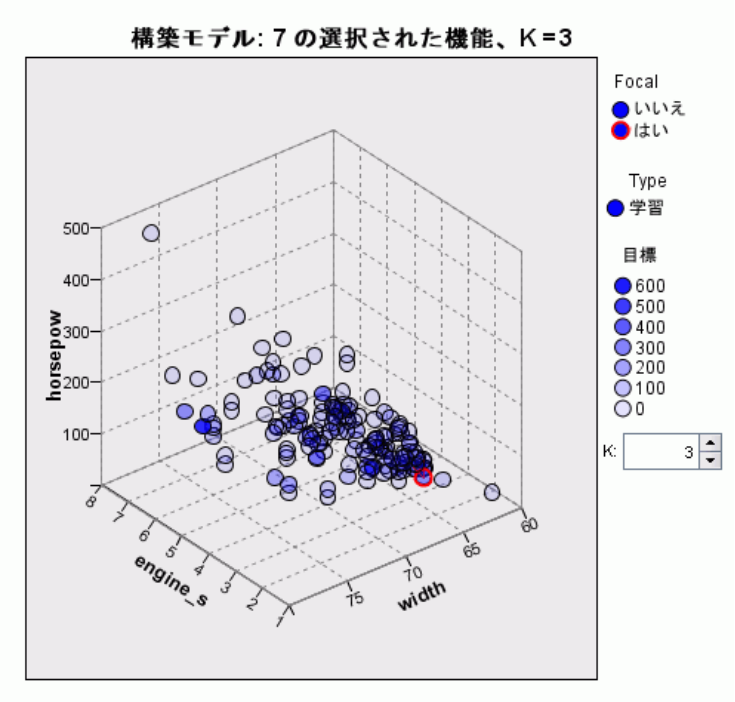
図 20-11
最近隣分析の [モデル ビュー] ドロップダウン



ビューに有効な情報がない場合、[ビュー] ドロップダウンの項目テキストが無効になります。

特徴空間

図 20-12
特徴空間



特徴空間図は、特徴空間（または、3 件を上回る特徴がある場合、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の特徴を示し、グラフの点の場所は、学習およびホールドアウト分割のケースにおけるこれらの特徴の値を示します。

キー。特徴の値のほか、図表内の点はその他の情報を示します。

- 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。それぞれの色でカテゴリ目標のカテゴリを示し、網掛けは連続型目標の値の範囲を示します。学習分割に示された値は観測値で、ホールドアウト分割は、予測値となります。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、ケースが中心ケースであることを示します。中心ケースは、k 最近隣へのリンクを示します。

コントロールおよび双方向性 図表内の多くの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。

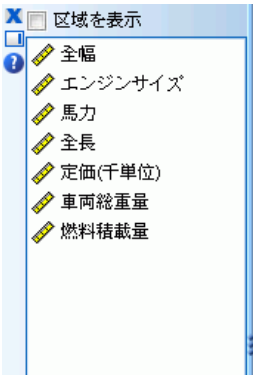
- 図表内に表示する特徴のサブセットを選択でき、また次元で表示される特徴を変更できます。
- 「中心ケース」は特徴空間図に選択された点です。中心ケース変数を指定すると、中心ケースを示す点が最初に選択されます。ただし、いかなる点を選択しても、一時的に中心ケースとなります。ポイント選択の「通常の」コントロールが適用されます。点をクリックすると、その点が選択され、その他の点はすべて選択解除されます。点でコントロールをクリックすると、選択された点のセットに追加されます。同位図などのリンク ビューは、特徴空間で選択されたケースに基づいて自動的に更新されます。
- 最近隣の数（k）を変更して中心ケースで表示することができます。
- カーソルを図内の点に移動すると、ケース ラベルの値を含む tooltip、またはケース ラベルが定義されていない場合はケース数、そして観測目標値および予測目標値が表示されます。
- ［リセット］ボタンを使用して、特徴空間を元の状態に戻すことができます。

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/変数を削除できます。

変数パレット

図 20-13
変数パレット



変数を追加および削除する前に、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。

▶ モデル ビューア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

▶ 編集モードにしたら、特徴空間でケースをクリックします。

▶ 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

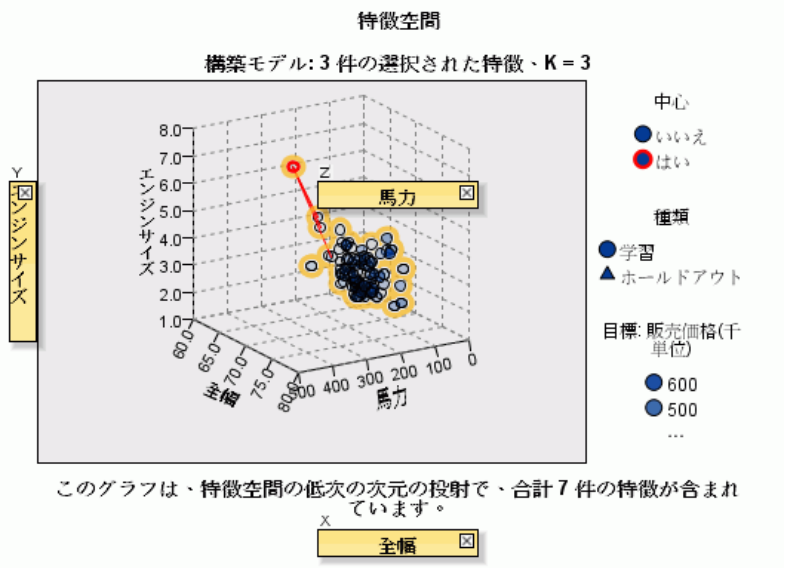
変数パレットには、特徴空間のすべての変数が表示されます。変数名の隣のアイコンは、変数の測定レベルを示します。

▶ 変数の測定レベル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、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。

図 20-14
変数領域



特徴空間には、x 軸、y 軸、z 軸の領域があります。

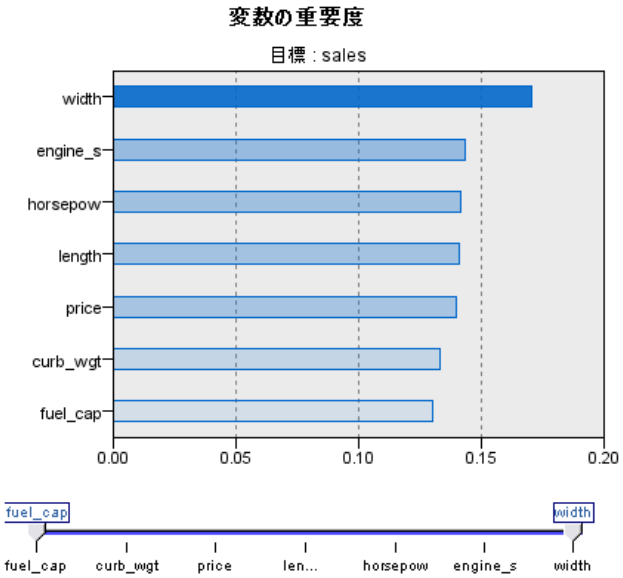
変数を領域に移動

変数を領域に移動するいくつかの規則およびヒントがあります。

- 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、領域に追加したい変数を選択することもできます。
- 変数を変数パレットから別の変数がある領域にドラッグすると、古い変数が新しい変数に置き換えられます。
- 変数をある領域から別の変数がある領域にドラッグすると、変数の位置が入れ替えられます。
- 領域の X をクリックすると、その領域の変数が削除されます。
- 視覚化の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には、それぞれの関連する変数ゾーンがあります。初めに、該当するグラフィック要素を選択します。

変数の重要度

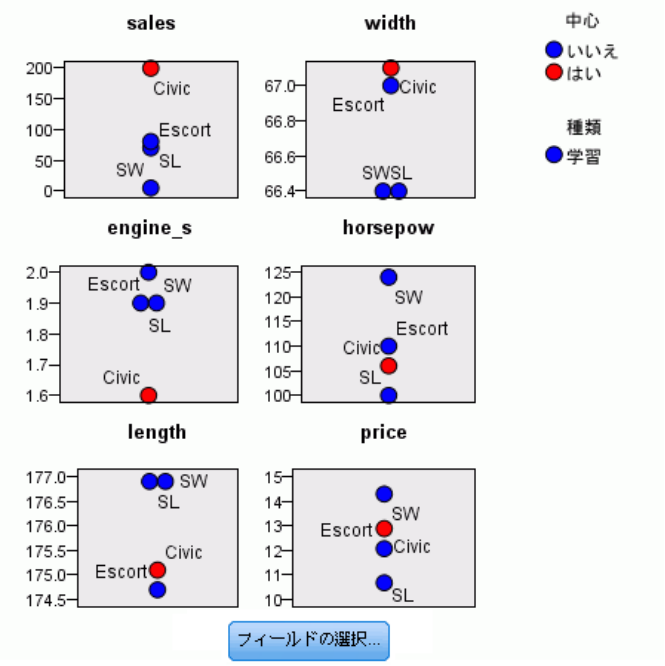
図 20-15
変数の重要度



通常、モデリングの効果を最も重要な変数に集中させ、最も重要でない変数を削除または無視したいと考えます。変数の重要度グラフを使用すると、モデル推定時に各変数の相対重要度を示して、これを実現できます。値が相対的であるため、表示されるすべての変数の値の合計は 1.0 となります。変数の重要度は、モデルの精度に関連していません。予測が正確かどうかに関係なく、予測時の各変数の重要度にのみ関連します。

同位

図 20-16
同位図



この図は、各特徴および目標の中心ケースおよび k 最近隣を表示します。中心ケースが特徴空間で選択されている場合に使用できます。

リンク同位図は、2 通りの方法で特徴空間にリンクしています。

- 特徴空間で選択された（中心）ケースは、k 最近隣とともに同位図に表示されます。
- 特徴空間で選択された k の値は、同位図で使用されます。

最近隣の距離

図 20-17
最近隣の距離

焦点ケース	最寄りの隣人			最寄りの距離		
	1	2	3	1	2	3
シビック	SL	Escort	SW	0.053674	0.0599112	0.0643422

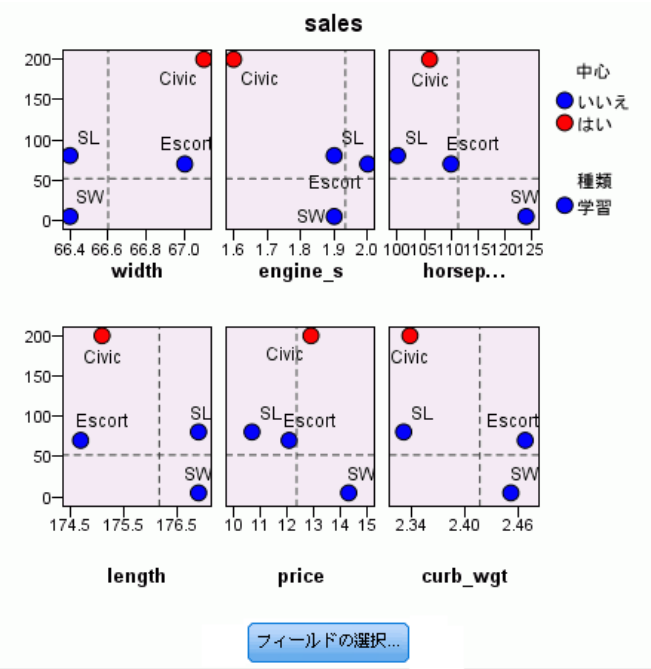
この表には、中心ケースのみの k 最近隣と距離が表示されます。[変数] タブで中心ケースの識別子が指定されている場合に使用でき、この変数で識別される中心ケースのみが表示されます。

次の列の各行に値が表示されます。

- [**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。ケースの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。
- 最近隣グループの i 番目の列には、中心ケースの i 番目の最近隣のケース 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番目の最近隣のケース数が表示されます。
- 最近隣の距離の i 番目の列には、中心ケースへの i 番目の最近隣の距離が表示されます。

四文位分布図

図 20-18
四分位分布図

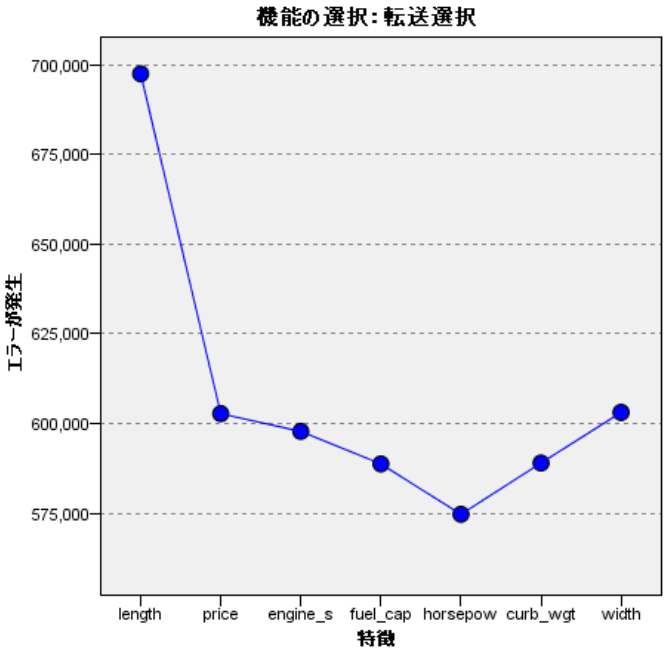


この表には、中心ケースと k 最近隣が散布図（または、目標の尺度に応じてドット プロット）で表示されます。y 軸には目標、x 軸には特徴を表示し、特徴ごとにパネル表示します。目標があり、中心ケースが特徴空間で選択されている場合に使用できます。

■ 連続変数について参照線が、学習分割の変数の平均値で描画されます。

特徴空間エラー ログ

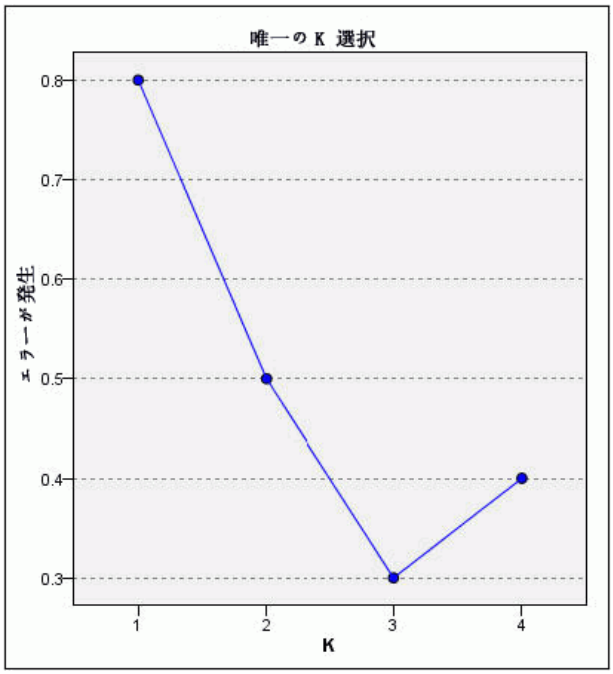
図 20-19
フィールド選択



図内の点は、モデルの y 軸に誤差（目標の尺度に応じて誤差率または誤差の平方和）を示し、x 軸は特徴を示します（x の左側にすべての特徴が表示されます）。目標があり、特徴選択が有効である場合、この図を使用することができます。

k 選択エラー ログ

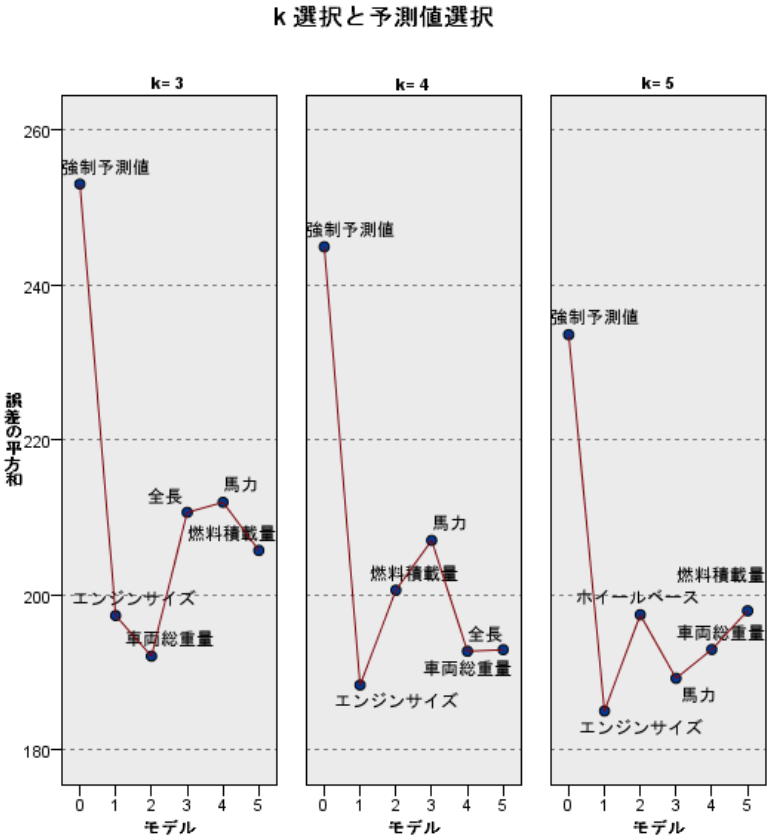
図 20-20
k の選択



図内の点は、モデルの y 軸に誤差（目標の尺度に応じて誤差率または誤差の平方和）を示し、x 軸は最近隣（k）の数を示します。目標があり、 k の選択が有効である場合、この図を使用することができます。

k および特徴選択エラー ログ

図 20-21
k および特徴選択



これらは、特徴選択図表（「特徴空間エラー ログ」（ p. 161 ）参照）で、k ごとにパネル表示されます。目標があり、 k および特徴選択が有効である場合、この図を使用することができます。

分類テーブル

図 20-22
分類テーブル

パーティション		予言		
		0	1	%正しい
訓練	0	111	1	99.11%
	1	7	33	82.50%
	全体の%	77.64%	22.37%	94.74%

この表には、目標の観測値と予測値のクロス分類を分割後とに表示します。目標があり、カテゴリ変数である場合に使用できます。

- ホールドアウト分割の [(欠損値)] 行には、目標に欠損値を持つホールドアウト 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] の値に対応しており、[正分類パーセント] には対応していません。

誤差の集計

図 20-23
誤差の集計

パーティション	二乗和のエラーが発生
訓練	622043

この表は、目標変数がある場合に使用できます。モデルに関連する誤差、連続型目標の場合は平方和、カテゴリ方目標の場合は誤差率(すべての正分類パーセントは 100% −) を表示します。

章
21

判別分析

判別分析は所属グループの予測モデルを構築します。予測モデルは、グループ間で最良の判別を行う予測変数の線型結合をもとに、判別関数（または、複数のグループの場合、判別関数グループ）で構成されます。関数は所属グループがわかっているケースのサンプルから生成され、その関数は、所属グループがわからない予測変数の測定を含む新しい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 つ以上の値を設定できます。しかし、グループ化変数のコードは、整数でなければならず、さらに最小値と最大値を指定しなければなりません。この範囲外の値のケースは、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摂取が多く、さらに温帯の方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判別できるかを確かめるため、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、人口や経済情報も重要であると考えています。判別分析を使って、多重線型回帰式の右側の項に似ている線型判別関数の係数を予測することができます。すなわち、係数 a、b、c、および d を使用した場合、式は次のようになります。

D = a * 気候 + b * 都市人口 + c * 人口 + d * 1 人当たりの国内総生産

これらの変数が 2 つの気候地帯の判別に有効な場合、温帯の国と熱帯の国では、D の値が異なります。ステップワイズ変数選択法を使用すると、4 つの変数すべてを関数に取り入れる必要はないことがわかります。

統計量。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、グループ内分散共分散行列、グループ別分散共分散行列、全分散共分散行列。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。

データ。グループ化変数には、整数でコード化された一定数のカテゴリがなければなりません。名義尺度の独立変数は、ダミー変数または対比変数として再割り当てしなければなりません。

© Copyright SPSS Inc. 1989, 2010

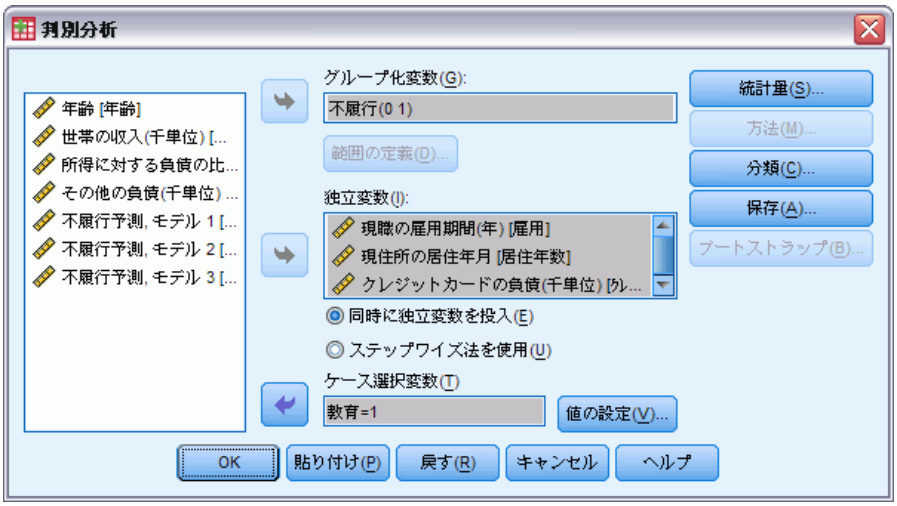
仮定。ケースは独立していなければなりません。予測変数は多変量正規分布を示していなければならず、さらにグループ内分散共分散行列はグループ全体で等しくなければなりません。各所属グループは重複せず（1 つのケースが複数のグループに所属していない）、グループ全体ですべてのケースを網羅している（全ケースがグループに所属している）と仮定します。この手続きは、所属グループがカテゴリ変数で構成されている場合に最も有効です。所属グループが連続変数の値（たとえば、高い IQ 対低い IQ）に基づいて構成されている場合、連続変数そのものの豊富な情報を最大限に活用するために、線型回帰を使用するとよいでしょう。

判別分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > 分類 > 判別分析...

図 21-1
[判別分析] ダイアログ ボックス



▶ 整数値をとるグループ化変数を選択し、[**範囲の定義**] をクリックして、目的のカテゴリを指定します。

▶ 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ない場合、[変換] メニューの [値の再割り当て] で値が整数の変数を作成します）。

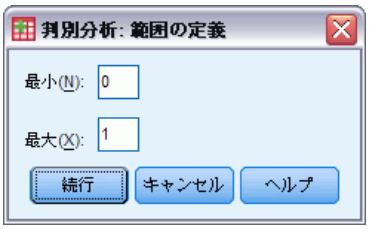
▶ 独立変数を投入する方法を選択します。

- **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が同時に投入されます。
- **ステップワイズ法を使用。**ステップワイズ法を使って変数の投入と削除を制御します。

▶ オプションとして、[選択] をクリックし、[ケース選択変数] ボックスに変数を選択することもできます。

判別分析: 範囲の定義

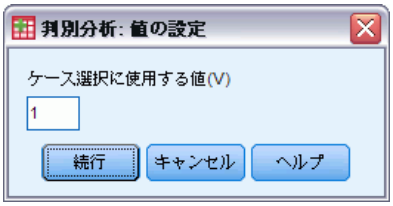
図 21-2
[判別分析: 範囲の定義] ダイアログ ボックス



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、判別分析に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 1 つに分類されます。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

図 21-3
[判別分析: 値の設定] ダイアログ ボックス



分析用にケースを選択するには、次のようにします。

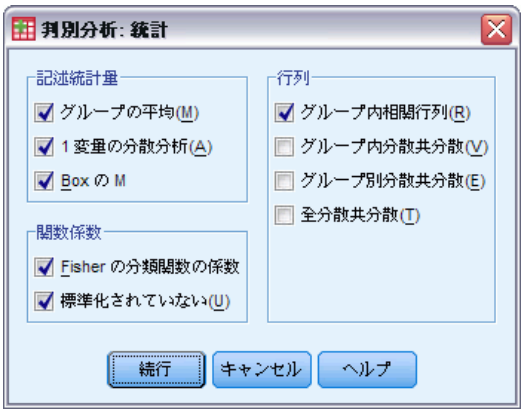
▶ [判別分析] ダイアログ ボックスで、ケース選択変数を選択します。

▶ [**値の設定**] をクリックし、ケース選択に使用する整数の値を入力します。

判別関数の作成に使用されるのは、このボックスで指定したケース選択変数の値を持つケースのみです。統計および分類の結果は、選択されたケースと選択されていないケースの両方に対して生成されます。これを利用し、既存のデータに基づいて新しいケースを分類したり、データを学習用および検定用サブグループに分割し、作成されたモデルの有効性を確認したりできます。

判別分析: 統計

図 21-4
[判別分析: 統計] ダイアログ ボックス



記述統計。使用できるオプションは、[平均値] (標準偏差を含む)、[1 変量の分散分析]、および [Box の M] です。

- **平均値 (信頼性分析).** 独立変数に対して合計とグループ平均値、標準偏差を表示します。
- **1 変量の分散分析 (判別分析).** 各独立変数に対してグループ平均値の同等性を検定するため一元配置分散分析を実行します。
- **Box の M.** グループの分散共分散行列の同等性を調べる検定。サンプルが十分に大きい場合、有意でない p 値は、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。使用できるオプションは、[Fisher の分類関数の係数]、および [標準化されていない] です。

- **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。各グループについて独立した分類関数の係数のセットが求められ、ケースは最大判別得点 (分類関数の値) を持つグループに割り当てられます。
- **標準化されていない (判別分析).** 標準化されていない判別関数の係数を表示します。

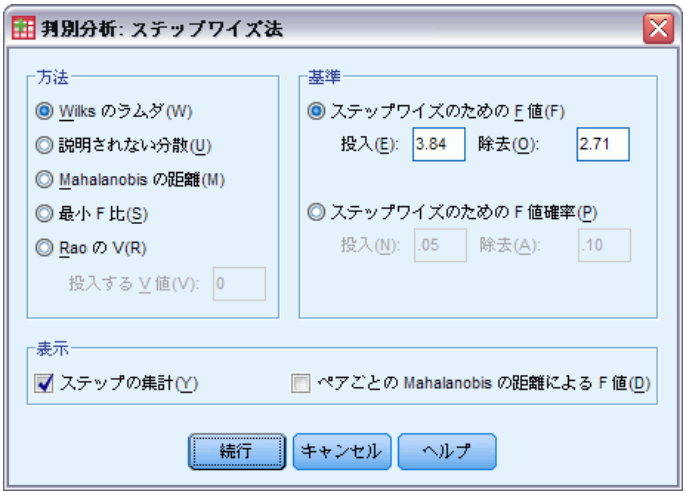
行列。独立変数に使用できる係数行列は、[グループ内相関行列]、[グループ内分散共分散]、[グループ別分散共分散]、および [全分散共分散] です。

- **グループ内相関行列.** 相関を計算する前に、グループすべての個別の分散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。

- **グループ内分散共分散.** 全分散共分散行列とは異なる場合もある、プールされたグループ内分散共分散行列を表示します。この行列は、すべてのグループの個別の分散共分散行列を平均化することによって得られます。
- **グループ別分散共分散.** 各グループに個別の分散共分散行列を表示します。
- **全分散共分散.** 1 つのサンプルから取り出したかのように、すべてのケースからの分散共分散行列を表示します。

判別分析: ステップワイズ法

図 21-5
[判別分析: ステップワイズ法] ダイアログ ボックス



方法。新しい変数を投入または除去するために使用する統計量を選択します。[Wilks のラムダ]、[説明されない分散]、[Mahalanobis の距離]、[最小 F 比]、および [Rao の V] を選択できます。[Rao の V] では、投入する変数に対して V の増加の最小値を指定できます。

- **Wilks のラムダ.** 変数が Wilks のラムダを低下させる量を基に分析に入力する変数を選ぶ、ステップごとの判別分析における変数選択法。各ステップで、Wilks のラムダを最小化させる変数が入力されます。
- **説明されない分散.** 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にする変数が投入されます。
- **Mahalanobis の距離.** 独立変数のケースの値がケースの平均からどの程度離れているかを測るもの。大きい Mahalanobis の距離は、1 つ以上の独立変数に極値を持つケースを特定します。

- **最小 F 比.** グループ間の Mahalanobis の距離から計算した F 比の最大化に基づく、ステップワイズによる変数選択法。
- **Rao の V.** グループ平均間の差の測定値。Lawley-Hotelling のトレースとも呼ばれます。各ステップで、Rao の V における増加を最大化する値が入力されます。このオプションを選択したら、変数が分析に入るために必要な最小の値を入力してください。

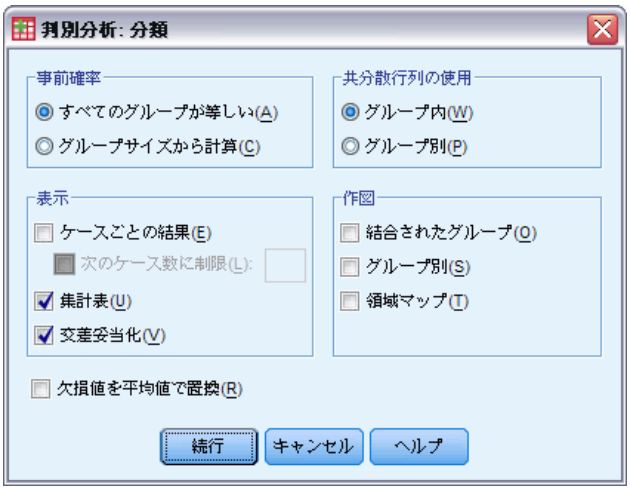
基準。利用できる選択肢は [**スワップワイズのための F 値**] と [**スワップワイズのための F 値確率**] です。変数の入力や削除に必要な値を指定します。

- **ステップワイズのための F 値.** 変数は、その F 値が [投入] 値よりも大きい場合にモデルに投入され、[除去] よりも小さい場合に除去されます。[投入] の値は [除去] の値より大きい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 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] 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 値の有意確率が [投入] の値よりも小さい場合、変数はモデルに入力され、[除去] の値よりも大きい場合除去されます。[投入] の値は [除去] の値より小さい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 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] 値を下げてください。

表示。[**ステップの集計**] をオンにすると、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**] をオンにすると、グループの各ペアに対してペアごとの F 比の行列が表示されます。

判別分析: 分類

図 21-6
[判別分析: 分類] ダイアログ ボックス



事前確率。このオプションは、分類係数を所属グループの事前知識に対して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等しい事前確率が仮定されます。係数には影響がありません。
- **グループサイズから計算。**サンプルで観測されたグループサイズから、所属グループの事前確率を決定します。たとえば、分析に含まれる観測の 50% が 1 番目のグループに属し、25% が 2 番目に属し、25% が 3 番目に属す場合、他の 2 つに対して 1 番目のグループに所属する尤度が大きくなるように、分類係数が調整されます。

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]、および [交差妥当化] です。

- **ケースごとの結果(E).** 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は、各ケースに表示されます。
- **集計表(U).** 判別分析に基づいて各グループに正しくまたは誤って分けられたケースの数。「コンフュージョン行列 (Confusion Matrix)」と呼ばれることもあります。
- **交差妥当化(V).** 分析における各ケースは、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されます。これは、「U-方法」とも呼ばれます。

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] または [グループ別] のどちらかをクリックして、ケース分類に使用する共分散行列を選びます。

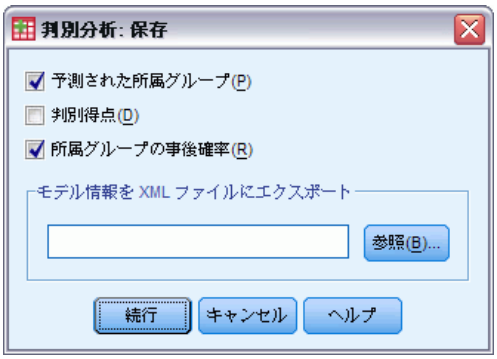
- **グループ内.** プールされたグループ内分散共分散行列は、ケースを分類するときに使用します。
- **グループ別.** グループ別分散共分散行列は、分類のために使われます。分類は元の変数ではなく判別関数に基づいているため、このオプションは 2 次の判別と等しくないときもあります。

作図。使用できるオプションは、[結合されたグループ]、[グループ別]、および [領域マップ] です。

- **結合されたグループ(O).** 最初の 2 つの判別関数値のすべての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(S).** 最初の 2 つの判別関数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が表示されます。
- **領域マップ(T).** 関数の値に基づいて、グループのケースを分類するための境界の作図です。これらの数字は、ケースが分類される先のグループに対応します。各グループの平均は、その境界内に星印で示されます。判別関数が 1 個の場合、このマップは表示されません。

判別分析: 保存

図 21-7
[判別分析: 保存] ダイアログ ボックス



作業中のデータ ファイルに新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 [所属グループの事後確率] (グループごとに 1 つの変数) です。

モデル情報を指定されたファイルに XML (PMML) 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、得点付け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

コマンド シンタックス言語を使用して、次のことも実行できます。

- 複数の判別分析を実行し (1 つのコマンドを使用)、変数の投入順序を制御する (ANALYSIS サブコマンドを使用)。
- 分類に使用する事前確率を指定する (PRIORS 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) ケースに対して分類を制限する (SELECT サブコマンドを使用)。
- 相関行列を読み込み、分析する (MATRIX サブコマンドを使用)。
- その後の分析で使用する相関行列を書き込む (MATRIX サブコマンドを使用)。

複雑なシンタックス情報については、「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。

章
22

因子分析

因子分析は、基礎となる変数、すなわち観測変数グループ内部の相関パターンを説明する **因子**を特定しようとする試みです。因子分析は、通常、データの分解の際に、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、少数の因子を識別するために使用します。因子分析は、原因のメカニズムに関する仮説を立てる場合、または次の分析に必要な変数を選別する（線型回帰分析を実行する前に共線性を識別するなど）場合にも使われます。

因子分析手続きは次のような柔軟性に富んだ機能を備えています。

- 7 種類の因子抽出方法があります。
- 非直交回転のためのプロマックスや直接オブリミン法など、5 種類の回転方法があります。
- 因子得点の計算方法は 3 種類あり、得点は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。

例:€ 政治についての調査質問に対する人々の回答内容は、どのような基本的姿勢によるものなのでしょうか。調査項目間の相関を調べると、税に関する質問が互いに相関していたり、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、また多くの場合、因子が表す概念的な意味を識別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、次の分析に利用することもできます。たとえば、因子得点をもとに、投票を予測するロジスティック回帰モデルを構築することもできます。

統計量 各変数: 有効ケース数、平均値および標準偏差。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があります。オブリミン回転: 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があります。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です。

© Copyright SPSS Inc. 1989, 2010

データ。変数は**区間**または**比例**尺度の量的変数でなければなりません。カテゴリ データ（宗教、出生国など）は因子分析には適しません。Pearson の相関係数が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 データ中の変数の各ペアは 2 変量の正規分布に従い、観測値は独立していることが必要です。因子分析モデルは、変数が共通因子（モデルにより推測される因子）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、すなわち計算された推測値は、すべての独自因子が互いに、また共通因子と相関しないという仮定に基づいていることを条件とします。

因子分析を行うには

▶ メニューから次の項目を選択します。

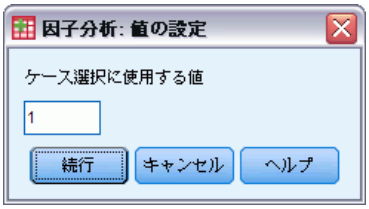
分析(A) > 次元分解 > 因子分析...

▶ 因子分析のための変数を選択します。

図 22-1
[因子分析] ダイアログ ボックス

因子分析のケースの選択

図 22-2
[因子分析: 値の設定] ダイアログ ボックス



分析用にケースを選択するには、次のようにします。

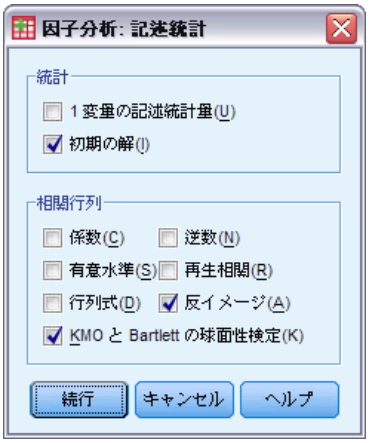
▶ 選択変数を選択します。

▶ [**値の設定**] をクリックし、ケース選択に使用する整数の値を入力します。

[ケース選択変数] に指定した値を持つケースのみが因子分析に使われます。

因子分析の記述統計

図 22-3
[因子分析: 記述統計] ダイアログ ボックス



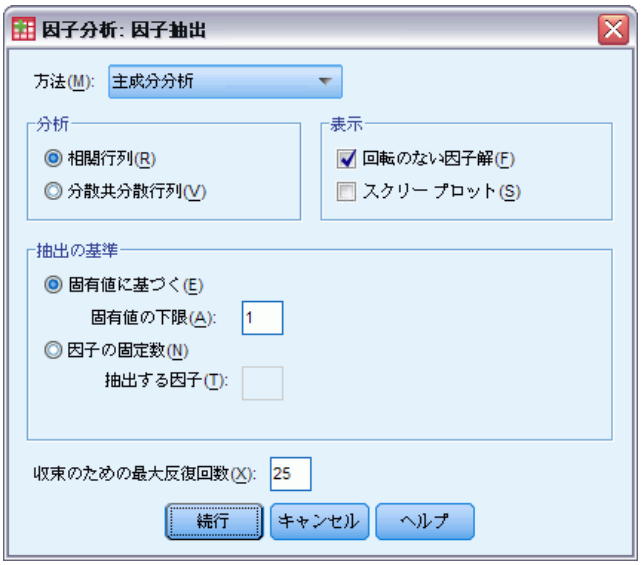
統計。[**1 変量の記述統計**] をクリックすると、変数ごとの平均値、標準偏差、有効ケース数が表示されます。[**初期の解**] をクリックすると、初期の共通性と固有値、説明された分散のパーセントが表示されます。

相関行列。使用できるオプションは、係数、有意確率、行列式、KMO と Barlett の球面性検定、逆行列、再生相関、および反イメージです。

- **KMO と Bartlett の球面性検定.** 変数間の偏相関が小さいかどうかを調べるサンプル適正検定の Kaiser-Meyer-Olkin 量。Bartlett の球面性の検定は、相関行列が単位行列かどうかを検定し、因子モデルが不適切かどうかを示します。
- **再生相関(R).** 因子の解からの推定相関行列。残差（推定された相関と観測された相関の差）も表示されます。
- **反イメージ(A).** 反イメージ相関行列は偏相関係数の負の数を含み、反イメージ分散共分散行列はその偏共分散の負の数を含みます。よい因子モデルでは、対角線上にない要素のほとんどは小さい値となります。変数に対する抽出の妥当性は、反イメージ相関行列の対角線上に表示されます。

因子分析の因子抽出

図 22-4
[因子分析: 因子抽出] ダイアログ ボックス



[方法] 因子抽出の方法を指定できます。使用できる方法は、主成分分析、重み付けのない最小 2 乗法、一般化最小 2 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。

- **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われる因子抽出方法。第 1 主成分が最大の分散を持ちます。以降の成分は、その分散のより小さな部分を徐々に説明し、それらはすべて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われます。相関行列が特異であるときに使うことができます。
- **重み付けのない最小 2 乗法 (因子分析).** 対角成分を無視して、観測相関行列と再生相関行列間の差の 2 乗を最小化する因子抽出法です。
- **一般化最小 2 乗法 (因子分析).** 観測相関行列と再生相関行列の間の差の平方和を最小化する因子抽出法。相関は、一意性と逆に重み付けをするため、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。
- **最尤法 (因子分析).** サンプルが多変量正規分布から抽出されている場合に、観測相関行列を生成した可能性が最も高いパラメータ推定値を生成する因子抽出法。相関には、変数の一意性の逆数が重み付けされ、反復アルゴリズムが用いられます。
- **主因子法 (因子分析).** 共通性の初期推定値の対角線上に配置された重相関係数の 2 乗によって、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は、対角線上にある古い共通性の推定値に置き換わる新し

い共通性を推定するために使われます。反復による変化が抽出のための収束基準を満たすまで、反復は続きます。

- **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。因子のアルファ信頼性を最大化します。
- **イメージ因子法 (因子分析).** Guttman によってイメージ理論を基に開発された因子抽出法。偏イメージと呼ばれる変数の共通部分は、仮説的因子の関数としてではなく、残りの変数に関する線型回帰として定義されます。

分析 相関行列または共分散行列のどちらかを指定できます。

- **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。
- **分散共分散行列** 各変数が異なる分散を持つ複数のグループに因子分析を適用する場合に便利です。

抽出の基準 固有値が指定した値を超えるすべての因子を保持するか、または保持する因子数を指定できます。

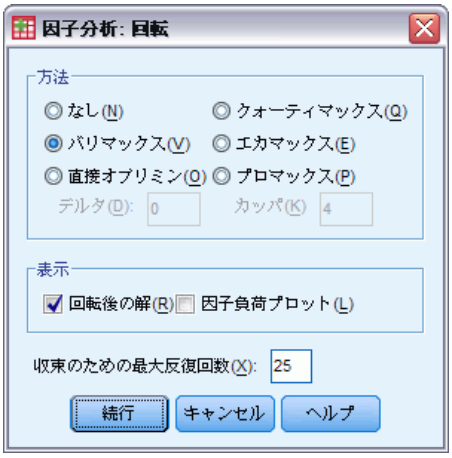
表示。回転前の因子解および固有値のスクリー プロットを指定できます。

- **回転のない因子解 (因子分析).** 因子解を求めるときに、回転を行わず求めた、因子負荷量、共通性、固有値を表示します。
- **スクリー プロット(S).** 各因子と関連する分散のプロット。このプロットで、保持する因子の数を決めるために使われます。一般に、この作図は、大きい因子の急勾配と残りの因子の緩勾配の間に明瞭な区切りを示します。

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られるステップの最大数を指定できます。

因子分析の回転

図 22-5
[因子分析: 回転] ダイアログ ボックス



方法。因子回転の方法を選択することができます。使用できる方法は、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、またはプロマックスです。

- **バリマックス法 (因子分析).** 各因子に高い負荷量を持つ変数の数を最小化する直交回転法。この方法は、因子の解釈を単純化します。
- **直接オブリミン法.** 斜交 (非直交) 回転の方法。デルタが 0 (デフォルト) のとき、解は最も斜交します。デルタが負になるに従って、因子の斜交度は下ります。デフォルト値の 0 を無効にするには、0.8 以下の数を入力してください。
- **クォーティマックス法 (因子分析).** 各変数を説明するために必要な因子数を最小化する回転法。この方法は、観測変数の解釈を単純化します。
- **エカマックス法 (因子分析).** 因子を単純化するバリマックス法と、変数を単純化するクォーティマックス法を組み合せた回転法。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が最小化されます。
- **プロマックス回転 (因子分析).** 斜交回転であり、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も早く計算されるため、大きなデータセットを使用する場合に役立ちます。

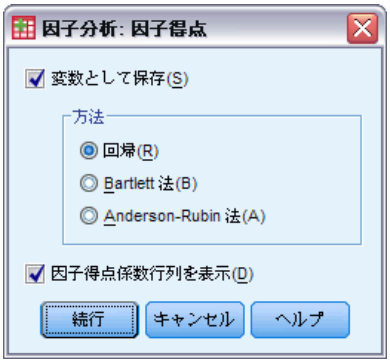
表示。最初の 2 因子または 3 因子の因子負荷プロットおよび回転後の解の出力を表示できます。

- **回転後の解 (因子分析).** 回転後の解を得るためには、回転方法が選択されている必要があります。直交回転では、回転されるパターン行列と因子変換行列が表示されます。斜交回転では、パターン、構成、および因子相関行列が表示されます。
- **因子負荷プロット.** 最初の 3 つの因子の 3 次元の因子負荷プロット。2 因子解に対しては、2 次元の作図が表示されます。1 つの因子だけが抽出された場合、作図は表示されません。因子の回転が要求されると、回転した解を表示します。

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られるステップの最大数を指定できます。

因子分析の因子得点

図 22-6
[因子分析: 因子得点] ダイアログ ボックス



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。

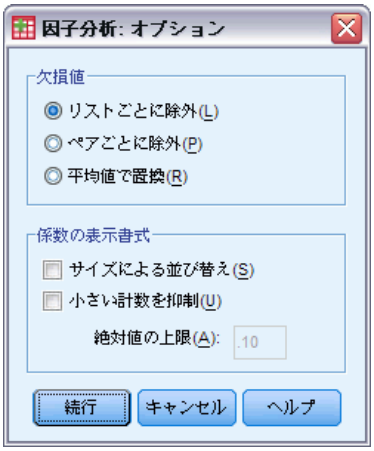
方法。 因子得点を計算するための代わりの方法は、回帰、Bartlett 法または Anderson-Rubin 法です。

- **回帰法 (因子分析).** 因子得点係数を推定する方法。得られる得点では、平均値が 0 で、分散は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。得点は、因子が直交であっても相関することがあります。
- **Bartlett 得点.** 因子得点係数を推定する方法。因子得点の平均値は 0 です。変数の範囲内にある独自の因子の平方和は最小化されます。
- **Anderson-Rubin 法 (因子分析).** 因子得点係数の推定法。推定された因子の直交化を保証する Bartlett 法の修正版です。得られる得点では、平均が 0 で標準偏差が 1 となり、相関しません。

因子得点係数行列を表示 因子得点を計算するために変数に乗じ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

図 22-7
[因子分析: オプション] ダイアログ ボックス



欠損値。欠損値の扱い方を指定できます。使用できる方法は、[**リストごとに**除外]、[**ペアごとに**除外]、または [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できます。係数をサイズにより分類したり、絶対値が指定した値未満の係数を抑制できます。

FACTOR コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 抽出と回転における、反復の収束基準を指定します。
- 個別の回転因子のプロットを指定します。
- 保存する因子得点の個数を指定します。
- 主因子法の対角線要素の値を指定します。
- 相関行列や因子負荷行列を、今後の分析のためにディスクに書き込みます。
- 相関行列や因子負荷行列を読み込み、分析します。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
23

クラスタリングの手続きの選択

クラスタ分析は、TwoStep、階層、または大規模ファイルのクラスタ分析の各手続きを使用して実行できます。各手続きは、クラスタの作成に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備えています。

TwoStep クラスタ分析。[2 ステップ クラスタ分析] 手続きは、多くのアプリケーションで選択される方法です。この手続きには、次に示す独自の機能があります。

- クラスタ モデルを選択するだけではなく、最適なクラスタ数を自動的に選択します。
- カテゴリ変数や連続変数に基づいてクラスタ モデルを同時に作成できます。
- クラスタ モデルを外部の XML ファイルに保存したり、このファイルを読み込んで新しいデータでクラスタ モデルを更新できます。

さらに、[TwoStep クラスタ分析] 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。

階層クラスタ分析。[階層クラスタ分析] 手続きは、小規模なデータ ファイル (クラスタ化されるオブジェクトが数百個のもの) に限定されますが、次に示す独自の機能があります。

- ケースまたは変数をクラスタ化する機能。
- 可能な解の範囲を計算し、それらの解の所属クラスタを保存する機能。
- 複数の方法でクラスタの書式化、変数の変換、およびクラスタ間の非類似度の測定を実行。

すべての変数が同じ種類である場合は、[階層クラスタ分析] 手続きによって、間隔 (連続)、度数、または 2 値変数を分析できます。

大規模ファイルのクラスタ分析。[大規模ファイルのクラスタ分析] 手続きは、連続データに限られ、クラスタの個数を事前に指定しておく必要がありますが、次に示す独自の機能があります。

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

© Copyright SPSS Inc. 1989, 2010

さらに、[大規模ファイルのクラスタ分析] 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。

章
24

TwoStep クラスタ分析

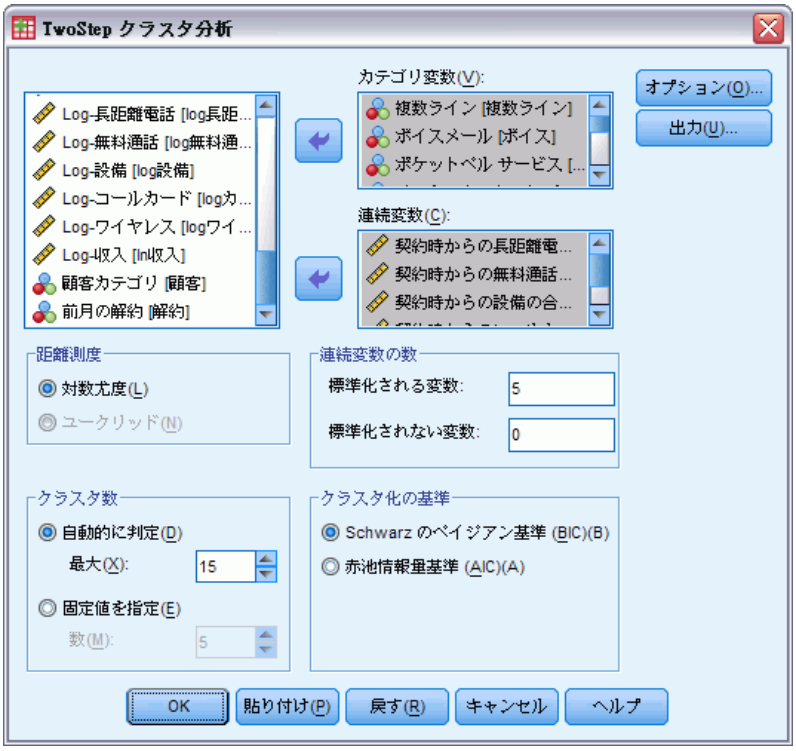
[TwoStep クラスタ分析] 手続きは、データセット内の自然なグループ（またはクラスタ）を明らかにすることを目的として設計された探索ツールです。この手続きで使用されるアルゴリズムには、従来のクラスタ分析技術とは異なるいくつかの優れた特徴があります。

- **カテゴリ変数と連続変数の処理。**変数が独立していると仮定することにより、カテゴリ変数および連続変数について、多項分布と正規分布を結合できます。
- **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、手続きで、最適なクラスタ数を自動的に判断できます。
- **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 アルゴリズムで大きなデータ ファイルを解析できます。

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

© Copyright SPSS Inc. 1989, 2010

図 24-1
[TwoStep クラスタ分析] ダイアログ ボックス



距離測度。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。連続変数は正規分布しているものと仮定し、カテゴリ変数は多項分布しているものと仮定します。すべての変数は独立しているものと仮定します。
- **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。この測定方法は、すべての変数が連続している場合にだけ使用できます。

クラスタ数。この項目を選択すると、クラスタの個数の判定方法を指定できます。

- **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」クラスタの個数を自動的に判定します。オプションとして、手続きで考慮するクラスタの個数の最大値を正の整数で入力できます。
- **固定値を指定。**解に含まれるクラスタの個数を固定できます。正の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変数の標準化設定の集計を示します。 詳細は、 p. 186 TwoStep クラスタ分析のオプション を参照してください。

クラスタ化の基準。自動クラスタ化アルゴリズムで、クラスタの個数を判定する方法を指定します。ベイズの情報量基準 (BIC) または赤池情報量基準 (AIC) のどちらかを指定できます。

データ。この手続きは、連続変数とカテゴリ変数の両方に使用できます。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。異なる無作為な順序で並べ替えら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できます。ファイル サイズが非常に大きいことが原因でこの操作の実行が難しい場合は、異なる無作為な順序で並べ替えられたケースのサンプルを使用し、複数回に分けて実行することで代用します。

仮定。尤度距離測度は、クラスタ モデル内の変数は独立しているものと仮定します。さらに、各連続変数は正規分布であると仮定し、カテゴリ変数は多項分布であると仮定します。経験的内部検定は、この手続きが独立および分布仮定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、把握するようにしてください。

2 つの連続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。連続変数とカテゴリ変数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。カテゴリ変数が、指定された多項分布になってい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。

階層クラスタ分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > 分類 > TwoStep クラスタ...

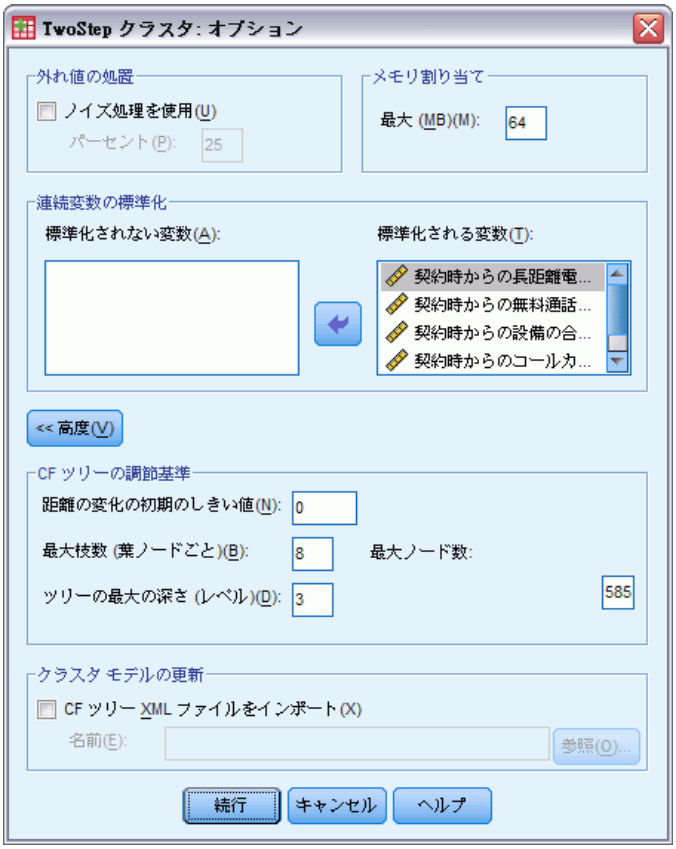
▶ 1 つ以上のカテゴリ変数または連続変数を選択します。

オプションとして、次の選択が可能です。

- クラスタ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、およびクラスタ モデル投入の設定を選択する。
- モデル ビューア出力を要求する。
- 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 クラスタ分析のオプション

図 24-2
[TwoStep クラスタ オプション] ダイアログ ボックス



外れ値の処置。クラスタ機能 (CF) ツリーが満杯の場合は、このグループで、クラスタ化の間に外れ値を特別に処理できます。CF ツリーが葉ノードでこれ以上ケースを受け入れられない場合、および葉ノードが分割できない場合、その CF ツリーは満杯です。

- ノイズ処理を選択したが、CF ツリーが埋まっている場合は、空きがある葉の中にあるケースを「ノイズ」葉に配置した後に、CF ツリーが再生されます。ツリー内のケースが最大葉サイズに対する指定パーセントよりも少ない場合、そのリーフには、空きがあると見なされます。ツリーが再生された後、可能であれば外れ値が CF に配置されます。不可能であれば、外れ値は廃棄されます。
- ノイズ処理を選択せず、CF ツリーが満杯の場合は、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。最終クラスタ化の後、クラスタに割り当てられなかった値は、外れ値としてラベル付

けされます。外れ値クラスタには、-1 という ID 番号が与えられ、クラスタの個数に含まれません。

メモリ割り当て。このグループで、クラスタ 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) で指定できます。手続きがこの最大値を超えると、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 以上の値を指定します。

- 自分の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の個数、または望ましいクラスタの個数を検出できない場合があります。

変数の標準化。クラスタ化アルゴリズムは、標準化された連続変数で使用できます。標準化されていない連続変数は、[標準化される変数] リストの変数として残しておきます。コンピュータ操作に費やす時間と手間を省くには、すでに標準化してある連続変数を [標準化されない変数] リストの変数として選択します。

[高度] オプション

CF ツリーの調節基準。次のクラスタ化アルゴリズム設定は、クラスタ化の特徴 (CF) ツリーだけに適用されるもので、慎重に変更する必要があります。

- **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。指定したケースを CF ツリーの葉に挿入したときにしきい値未満である場合、その葉は分割されません。このしきい値を超える場合は分割されます。
- **最大枝数 (葉ノードごと)。**葉ノードが持つことのできる子ノードの最大数。
- **ツリーの最大の深さ。**CF ツリーが持つことの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、手続きによって生成されることができる CF ツリー ノードの最大数を示します。b は最大枝数、d は最大ツリー深さを示します。CF ツリーが大きすぎると、システム リソースが浪費され、手続きの動作にとって逆効果となる場合があることに注意してください。最低でも、各ノードに 16 バイトは必要です。

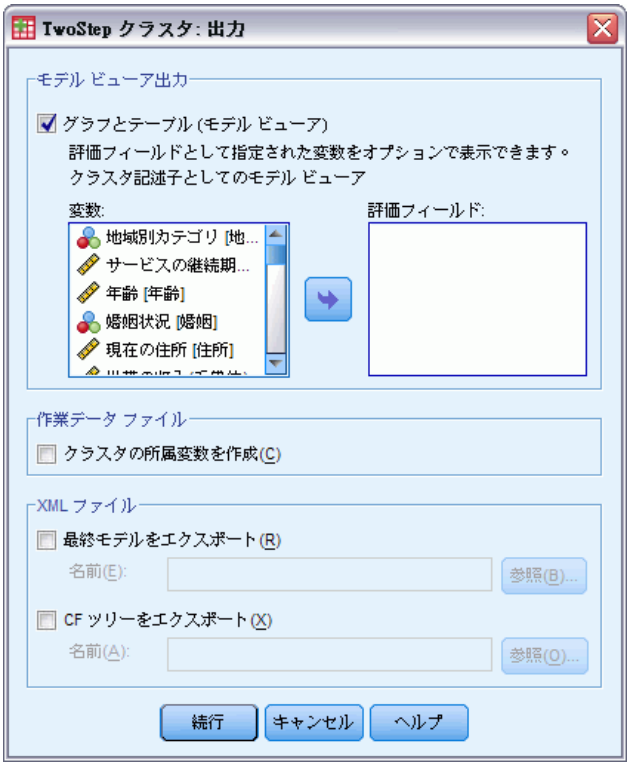
クラスタ モデルの更新。先に実行した分析で生成されたクラスタ モデルをインポートおよび更新できます。入力ファイルには、CF ツリーが XML 形式で格納されています。モデルは、アクティブ ファイルのデータで更新されます。メイン ダイアログ ボックスで、先の分析で指定したときと同じ順序で変数名を選択する必要があります。同じファイル名で新しいモデル情報を記述しない限り、XML ファイルは変更されません。 詳細は、p. 188 TwoStep クラスタ分析の出力 を参照してください。

クラスタ 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 ツリーの世代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 ツリーの基準調節設定が使用され、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 モデルの更新を行うとき、手続きでは、元のクラスタ モデルの作成に、アクティブなデータセット内の選択ケースは使用されなかったものと仮定します。また、モデルの更新に使用されるケースは、元のモデルの作成に使用されたケースと同じ分布のもの、つまり、連続変数の平均値と分散、およびカテゴリ変数のレベルは、どちらのケース グループについても同じであると仮定します。「新しい」ケースと「古い」ケースが不均質の分布のものである場合は、結合されたケース グループで [TwoStep クラスタ分析] 手続きを実行して、最良の結果を求めてください。

TwoStep クラスタ分析の出力

図 24-3
[2 ステップ クラスタ: 出力] ダイアログ ボックス



モデル ビューア出力。このグループで、クラスタ化の結果を表示するためのオプションを指定します。

- **図表と表** 表および図表など、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、モデル要約と、クラスタ機能グリッドが表示されます。モデル ビューのグラフィカル出力には、クラスタ品質表、クラスタ サイズ、変数の重要度、クラスタ比較グリッド、およびセル情報があります。
- **評価フィールド。**クラスタ作成に使用されなかった変数のクラスタ データを計算します。[表示] サブダイアログで選択して、モデル ビューアの入力フィールドとともに評価フィールドを表示します。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。このグループで、変数をアクティブなデータセットに保存します。

- **クラスタの所属変数を作成。**この変数には、各ケースのクラスタ ID 番号が含まれています。この変数の名前は tsc_n で、n は、所定のセッションでこの手順により完了するアクティブなデータセット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は、最終クラスタ モデルおよび CF ツリーです。

- **最終モデルをエクスポート。**最終クラスタ モデルが、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、得点付け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。**このオプションで、クラスタ 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更新できます。

クラスタ ビューア

通常、クラスタ モデルを使用して、検証された変数に基づいて類似したレコードのグループ（またはクラスタ）を検索します。同じグループのメンバー間の類似性は高く、異なるグループのメンバー間の類似性は低くなります。結果を使用して、明らかでなかった関連性を特定できます。たとえば、顧客の嗜好、収入レベル、購買習慣のクラスタ分析を使用して、特定のマーケティング キャンペーンに応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。

クラスタ表示の結果を解釈するには、次の 2 つの方法があります。

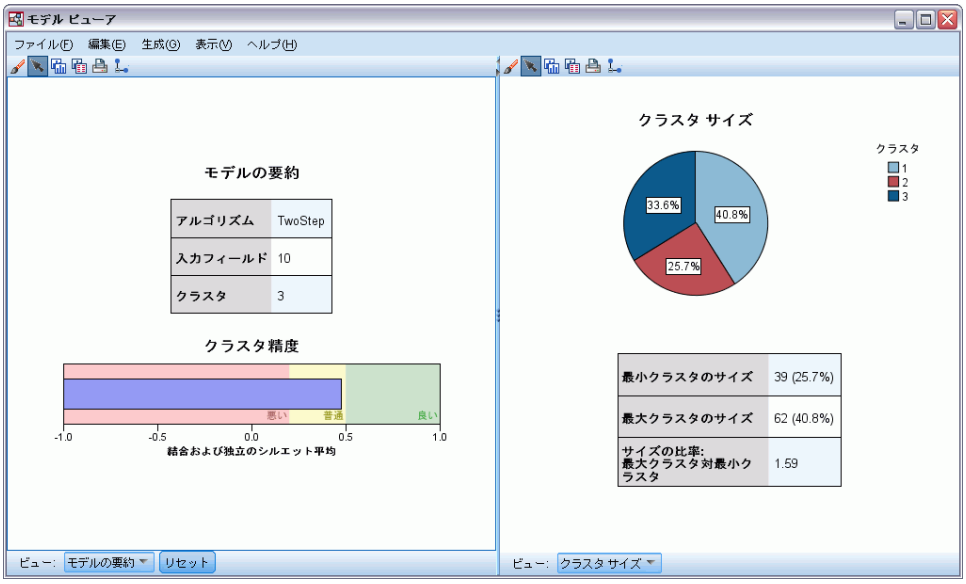
- クラスタを検証して、そのクラスタに特有の特性を確認します。1 つのクラスタに、高収入な借り手がすべて含まれていますか？また、このクラスタには、他のクラスタよりも多くのレコードが含まれていますか？
- クラスタ間でフィールドを検証して、値がクラスタ間でどのように分布しているかを確認します。クラスタ内のメンバーシップは、メンバーの学歴により決定されますか？最高得点の格付けにより、クラスタごとのメンバーシップが区別されますか？

クラスタ ビューアのメイン ビューおよびリンク ビューを使用して、これらの疑問に答えるための手がかりを得ることができます。

クラスタ モデルの詳細を表示するには、ビューアのモデル ビューア オブジェクトを有効化（ダブルクリック）します。

クラスタ ビューア

図 24-4
デフォルト表示のクラスタ ビューア



クラスタ ビューアは 2 つのパネルで構成されています。左側はメイン ビュー、右側はリンク ビューまたは補助ビューです。メイン ビューには、2 種類があります。

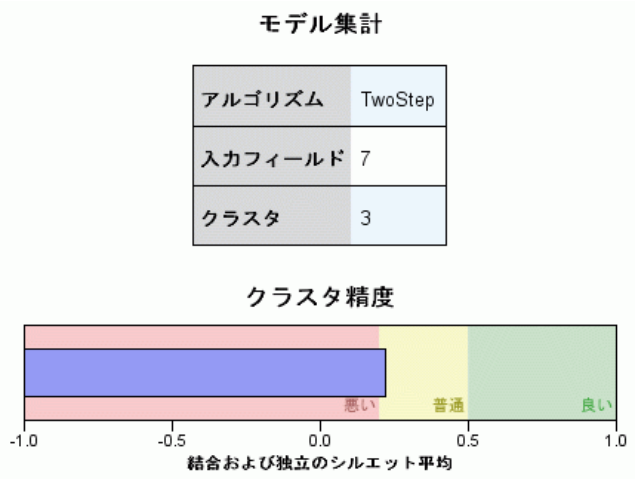
- モデルの要約（デフォルト）。 詳細は、 p. 191［モデルの要約］ビュー を参照してください。
- クラスタ： 詳細は、 p. 192 クラスタ ビュー を参照してください。

リンク/補助ビューには、次の 4 種類あります。

- 予測値の重要度。 詳細は、 p. 196 クラスタ予測値の重要度 ビュー を参照してください。
- クラスタ サイズ（デフォルト）。 詳細は、 p. 197 クラスタ サイズ ビュー を参照してください。
- セルの分布。 詳細は、 p. 198 セルの分布ビュー を参照してください。
- クラスタの比較。 詳細は、 p. 199 クラスタの比較ビュー を参照してください。

[モデルの要約] ビュー

図 24-5
メイン パネルの [モデルの要約] ビュー



[モデルの要約] ビューには、陰影を付けて悪い結果、普通の結果、よい結果を示すクラスタ結合および独立のシルエット平均など、クラスタ モデルについてのスナップショットまたは要約が表示されます。このスナップショットを使用して、品質が悪いかどうかをすばやく確認できます。この場合、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し、よりよい結果を生成することができます。

悪い結果、普通の結果、よい結果は、クラスタ構造の解釈に関する Kaufman と Rousseeuw (1990) の研究に基づきます。[モデルの要約] ビューで、よい結果は 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、普通の結果は弱い証拠の評価を、悪い結果は、重要な証拠のない評価を反映するデータとなります。

すべてのレコードに対するシルエット平均は (B-A) / max(A,B) となります。A は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隣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、すべてのケースはクラスタ中心に直接配置されているということを意味します。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、その他の最近隣クラスタとの間で等距離にあることを意味します。

要約では、次の情報について示す表も表示されます。

- **アルゴリズム:** 「TwoStep」など、使用されるクラスタリング アルゴリズム。

- **入力フィールド**:入力フィールドまたは**予測フィールド**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

図 24-6
メイン パネルの [クラスタ中心] 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、プロファイルを示すクラスタ化機能グリッドがあります。

グリッドの列には次の情報が表示されます。

- **クラスタ**: アルゴリズムで作成されたクラスタ数。
- **ラベル**: 各クラスタに適用されるラベル (デフォルトでは空欄)。セルをダブルクリックして、クラスタの内容を説明するラベル (例: 「高級車購入者」) を入力します。
- **説明**: クラスタの内容についての説明 (デフォルトでは空欄)。セルをダブルクリックして、クラスタについての説明するラベル (例: 「55 歳以上、専門、収入 $100,000 以上」) を入力します。

- **サイズ**: 各クラスタのサイズ (クラスタ サンプル全体の割合)。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割合を示す垂直バー、数値形式のサイズの割合、クラスタ ケースの度数が表示されます。
- **フィールド。**それぞれの入力または予測。デフォルトでは全体の重要度で並べ替えられています。列に表示されるサイズに等しい値がある場合、クラスタ番号の昇順で表示されます。

  フィールド全体の重要度は、セル背景色の濃さで示されます。最も重要なフィールドが濃く、重要でないフィールドが薄くなります。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。

セルの上にマウス ポインタを置くと、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種類によっては、より詳細な情報が表示されます。[クラスタ中心] ビューで、「平均: 4.32」など、セルの統計量やセル値を示します。カテゴリ フィールドの場合、セルは最も頻度の高い (モーダル) カテゴリとそのパーセントを示します。

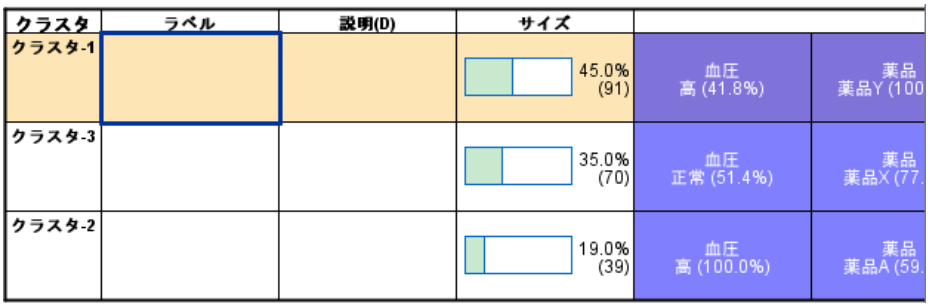
[クラスタ] ビューで、クラスタ情報を表示するさまざまな方法を選択できます。

- クラスタとフィールドを入れ替える。 詳細は、 p. 193 クラスタとフィールドを入れ替え を参照してください。
- フィールドを並べ替える。 詳細は、 p. 194 フィールドの並べ替え を参照してください。
- クラスタを並べ替える。 詳細は、 p. 194 クラスタの並べ替え を参照してください。
- セルの内容を選択する。 詳細は、 p. 194 セルの内容 を参照してください。

クラスタとフィールドを入れ替え

デフォルトでは、クラスタは列に、フィールドは行に表示されます。この表示を逆にするには、[**フィールドの並べ替え基準**] の左側の [**クラスタと特徴を入れ替え**] ボタンをクリックします。たとえば、表示するクラスタが多い場合、データの表示に必要な水平方向のスクロールを少なくするために、この設定が必要になります。

図 24-7
メイン パネルのクラスタの入れ替え



フィールドの並べ替え

[フィールドの並べ替え基準] ボタンを使用して、特徴セルの表示方法を選択できます。

- **全体の重要度。**これはデフォルトのソート順です。フィールドは全体の重要度の降順に並べ替えられ、ソート順はクラスタ間で同じになります。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並べられます。
- **クラスタ内重要度。**フィールドは、各クラスタの重要度に応じて並べ替えられます。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並べられます。このオプションを選択すると、ソート順は通常クラスタ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並べられます。
- **データ順。**フィールドは、データセットのデータ順に並べられます。

クラスタの並べ替え

デフォルトでは、クラスタはサイズの小さい順に並んでいます。[クラスタのソート基準] ボタンを使用して、名前のアルファベット順に並べ替えることができます。または一意のラベルを作成した場合は、ラベルのアルファベット順に並べ替えることができます。

同じラベルを持つフィールドは、クラスタ名で並べられます。クラスタがラベル順に並べられ、クラスタのラベルを編集すると、ソート順は自動的に更新されます。

セルの内容

[セル] ボタンを使用して、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。

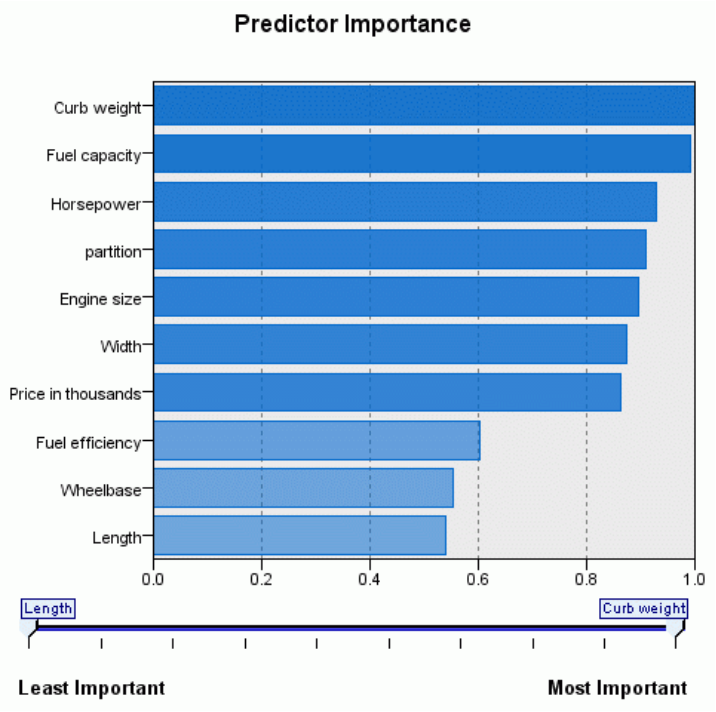
- **クラスタ中心。**デフォルトでは、セルにフィールド名/ラベルと各クラスタ/フィールドの組み合わせの中心傾向が表示されます。連続型フィールドの平均値が、カテゴリ フィールドの最頻値（最も頻繁に発生するカテゴリ）がカテゴリ パーセントとともに表示されます。
- **絶対分布。**各クラスタのフィールド名/ラベルとフィールドの絶対分布が表示されています。カテゴリ フィールドの場合、データ値が低い順に並んでいるカテゴリが重ねられた棒グラフが表示されます。連続型フィールドの場合、各クラスタの同じエンドポイントと区間を使用する平滑密度プロットが表示されます。

  濃い赤はクラスタ分布を示し、淡い赤は全体のデータを示します。
- **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ています。一般的に、相対分布が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。

  濃い赤はクラスタ分布を示し、淡い赤は全体のデータを示します。
- **基本ビュー。**多くのクラスタがある場合、スクロールせずにすべての詳細を確認するのは難しくなります。スクロールを少なくするために、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。

クラスタ予測値の重要度 ビュー

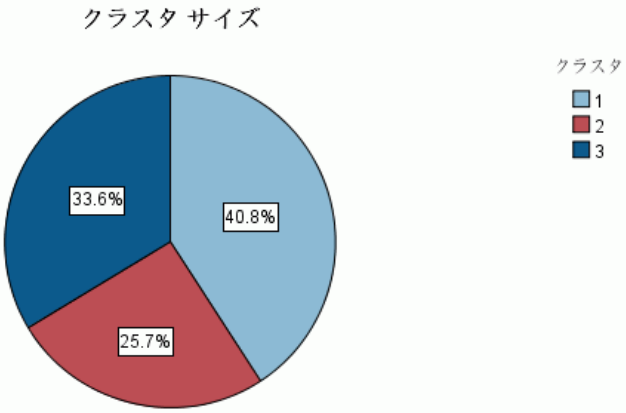
図 24-8
リンク パネルの [クラスタ予測値の重要度] ビュー



[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を表示します。

クラスタ サイズ ビュー

図 24-9
リンク パネルの [クラスタ サイズ] ビュー



最小クラスタのサイズ	39 (25.7%)
最大クラスタのサイズ	62 (40.8%)
サイズの比率: 最大クラスタ対最小クラスタ	1.59

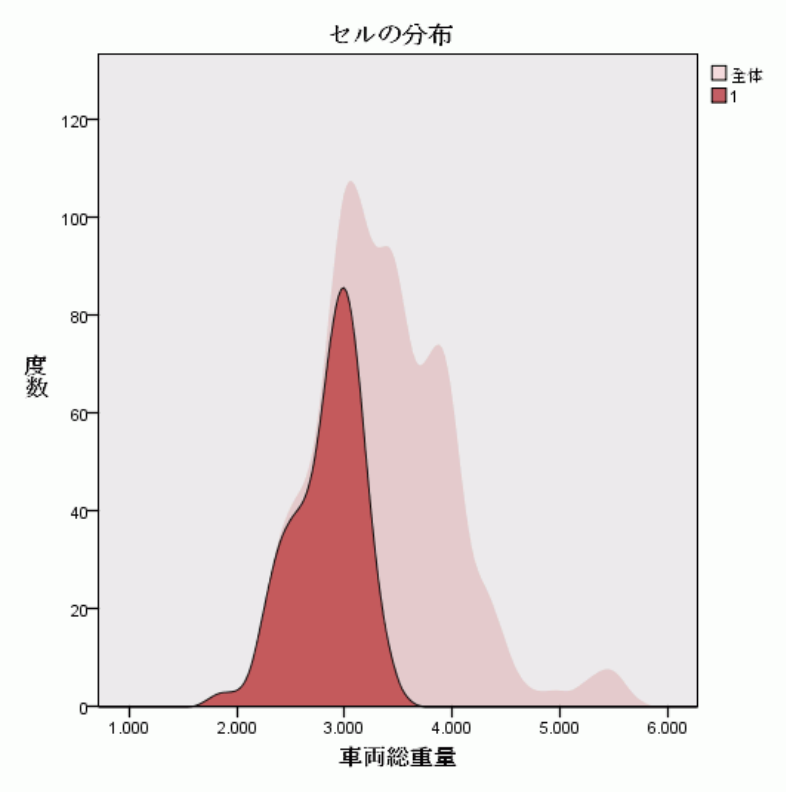
[クラスタ サイズ] ビューには、各クラスタについて示す円グラフが表示されます。各クラスタのサイズのパーセントが各スライスに表示されます。各スライスにマウス ポインタを置くと、そのスライスに度数が表示されます。

円グラフの下の表に、次のサイズ情報について表示されます。

- 最小クラスタのサイズ (度数と全体の割合)
- 最大クラスタのサイズ (度数と全体の割合)
- 最大クラスタの最小クラスタに対するサイズの比率

セルの分布ビュー

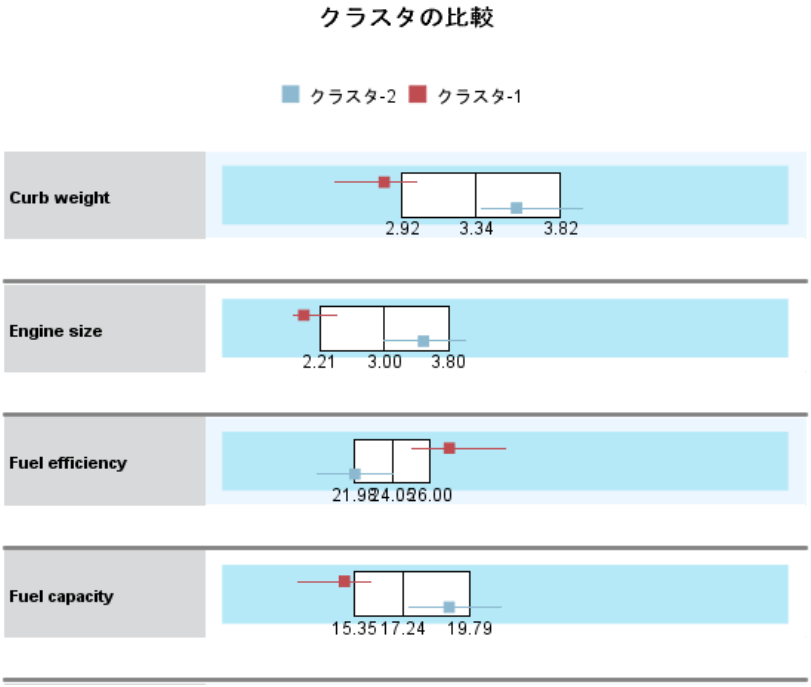
図 24-10
リンク パネルの [セルの分布] ビュー



[セルの分布] ビューには、クラスタのメイン パネルで選択したフィールド セルのデータの分布について、拡張し、より詳細なプロット図が表示されます。

クラスタの比較ビュー

図 24-11
リンク パネルの [クラスタの比較] ビュー



[クラスタの比較] 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、選択したクラスタは列に表示されます。このビューを使用して、クラスタを構成する要素をより良く理解できます。また、全体のデータだけでなく、クラスタ間で比較して、それぞれの差分を確認することもできます。

表示するクラスタを選択するには、[クラスタ] メイン パネルのクラスタ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して、比較する複数のクラスタを選択または選択解除できます。

注 :表示するクラスタは、5 個まで定義することができます。

クラスタは選択された順に表示されます。フィールドの順序は、[**フィールドの並べ替え基準**] オプションで決定します。[**クラスタ内重要度**] を選択した場合、フィールドは常に全体の重要度の順に並べられます。

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図で表示されます。ドットのサイズは、フィールドごとの各クラスタの最も頻度の高い/モーダルなカテゴリを示します。
- 連続型フィールドは箱ひげ図で表示され、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の箱ひげ図が重なります。

- 連続型フィールドの場合、四角形のポイント マーカーと水平線は、それぞれ各クラスタの中央値と 4 分位範囲を示します。
- 各クラスタは、それぞれの色で示され、ビューの最上位に表示されます。

クラスタ ビューアの操作方法

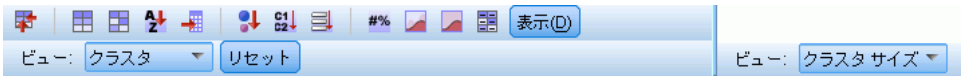
クラスタ ビューアはインタラクティブ表示です。以下を行うことができます。

- フィールドまたはクラスタを選択して詳細を表示する。
- クラスタ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、ビューアをデフォルト設定にリセットし、ダイアログ ボックスを指定して、メイン パネルのクラスタ ビューの内容を指定することもできます。

図 24-12
クラスタ ビューアに表示されるデータを制御するツールバー



[フィールドのソート基準]、[クラスタのソート基準]、[セル]、[表示] のオプションは、メイン パネルで [クラスタ] ビューを選択した場合にのみ使用できます。 詳細は、 p. 192 クラスタ ビュー を参照してください。

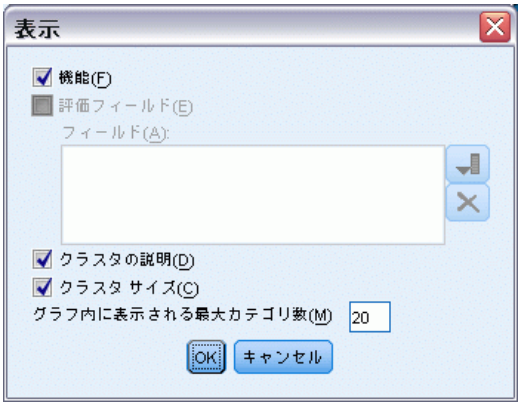
	「クラスタとフィールドを入れ替え」（ p. 193 ）を参照してください。
	「フィールドの並べ替え基準」（ p. 194 ）を参照してください。

	「クラスタの並べ替え基準」（ p. 194 ）を参照してください。
	「セル」（ p. 194 ）を参照してください。

クラスタ ビュー表示の制御

メイン パネルのクラスタ 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。[表示] ダイアログが表示されます。

図 24-13
クラスタ ビューア - 表示オプション



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隠すには、チェック ボックスをオフにします。

評価フィールド。表示する評価フィールド（クラスタ モデルの作成には使用されず、モデル ビューアに送信してクラスタを評価するフィールド）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：使用できる評価フィールドがない場合、このチェック ボックスは使用できません。

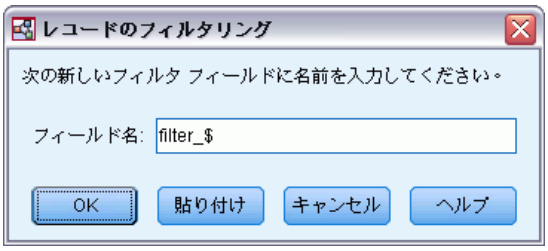
クラスタの説明。デフォルトで選択されています。すべてのクラスタ説明のセルを隠すには、チェック ボックスをオフにします。

クラスタ 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 サイズのセルを隠すには、チェック ボックスをオフにします。

カテゴリの最大数。カテゴリ フィールドのグラフに表示するカテゴリ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

図 24-14
クラスタ ビューア - ケースのフィルタリング



特定のクラスタまたはクラスタのグループのケースについてより詳しく理解するには、さらに分析を行うために、選択したクラスタに基づいてレコードのサブグループを選択できます。

▶ クラスタ ビューアのクラスタ ビューでクラスタを選択します。複数のクラスタを選択するには、Ctrl キーを押しながらクリックします。

▶ 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...

▶ フィルタ変数の名前を入力します。選択したクラスタのレコードには、このフィールドに対し値 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、フィルタの状態を変更するまで、以降の分析から除外されます。

▶ [OK] をクリックします。

章
25

階層クラスタ分析

この手続きでは、個々のクラスタの各ケース（または変数）で始まるアルゴリズムを使用して、相対的に等質なケース（または変数）をグループの選択された特性に基づいて識別し、さらに、最後の 1 つになるまで他のすべてのクラスタを結合させます。元の変数を分析したり、さまざまな標準化変換を選択することもできます。距離と類似度は、[近接] 手続き（近接性分析）で生成します。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。階層クラスタ分析では、テレビ番組（ケース）を視聴者の特性に基づいて等質グループにクラスタ化できます。この分析結果は、コマーシャルの時間帯の特定に使用できます。また、都市（ケース）を等質グループにクラスタし、比較する都市を選択していろいろなマーケティング戦略を検定できます。

統計量 クラスタ凝集経過工程、距離（または類似）行列、および単一の解または解の範囲の所属クラスタ。プロット: デンドログラムおよびつららプロット。

データ。変数は数量、2 値、または度数データです。尺度の違いがクラスタの解に影響することがあるため、変数の尺度は重要です。変数の尺度がまったく異なる場合（たとえば、ある変数はドル単位で測定され、別の変数は年単位で測定される場合）、変数の標準化（[階層クラスタ分析] 手続きで自動的に実行可能）を行う必要があります。

ケースの並び順。同一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内で発生した場合、導き出されるクラスタの解は、ファイル内のケースの並び順によって異なる可能性があります。異なる無作為な順序で並べ替えら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できます。

仮定。使用する距離、または類似度の測定方法は、分析データに適していなければなりません（距離と類似度の測定方法の選択については、[近接] 手続きを参照してください）。また、必要な変数をすべて分析に含める必要があります。重要な変数を省略すると、誤った解が生成される可能性があります。階層クラスタ分析は探索的手法なので、結果は独立サンプルで確認するまで仮説扱いとなります。

© Copyright SPSS Inc. 1989, 2010

階層クラスタ分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > 分類 > 階層クラスタ...

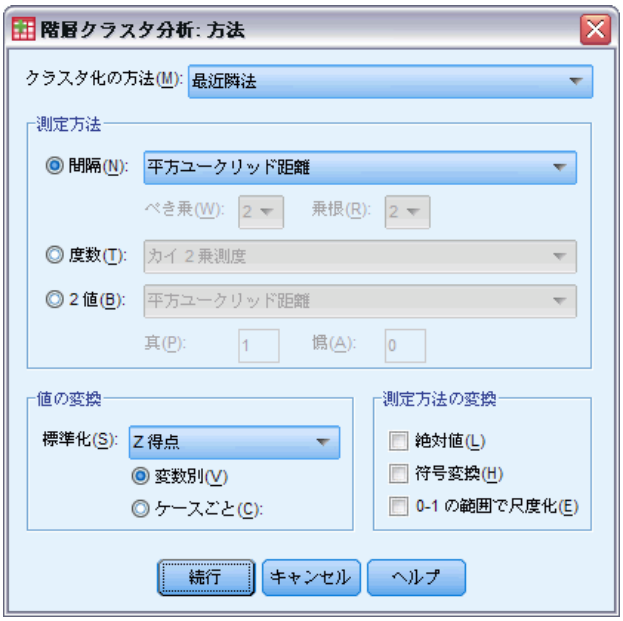
図 25-1
[階層クラスタ分析] ダイアログ ボックス

▶ ケースをクラスタ化するには、最低 1 つの数値型変数を選択します。変数をクラスタ化するには、最低 3 つの数値型変数を選択します。

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することもできます。

階層クラスタ分析の方法

図 25-2
[階層クラスタ分析: 方法] ダイアログ ボックス



クラスタ化の方法。使用できるオプションは [グループ間平均連結法]、[グループ内平均連結法]、[最近隣法]、[最遠隣法]、[重心法]、[メディアン法]、または [Ward 法] です。

測定方法。距離や類似度の測定方法を指定して、クラスタ化に使用できます。データの種類とそれぞれの距離や類似度の測定方法を指定します。

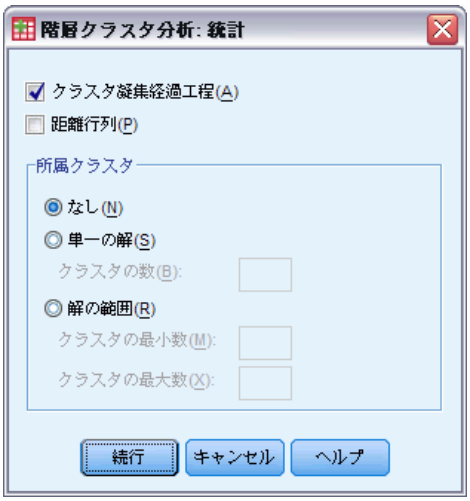
- **間隔。**使用できる測定方法は [ユークリッド距離]、[平方ユークリッド距離]、[コサイン]、[Pearson の相関]、[Chebychev]、[都市ブロック]、[Minkowski]、または [カスタマイズ] です。
- **度数。**使用できる測定方法は [カイ 2 乗測度] または [ファイ 2 乗測度] です。
- **2 値。**使用できる測定方法は [ユークリッド距離]、[平方ユークリッド距離]、[サイズの差異]、[パターンの差異]、[分散]、[散らばり]、[形]、[単純マッチング]、[ファイ 4 分点相関係数]、[ラムダ]、[Anderberg の D]、[dice]、[Hamann]、[Jaccard]、[Kulczynski 1]、[Kulczynski 2]、[Lance と Williams]、[落合]、[Rogers と Tanimoto]、[Russel と Rao]、[Sokal と Sneath 1]、[Sokal と Sneath 2]、[Sokal と Sneath 3]、[Sokal と Sneath 4]、[Sokal と Sneath 5]、[Yule Y]、または [Yule Q] 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を標準化します (2 値データには使用できません)。使用できる標準化方法は、[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]、または [標準偏差を 1] 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。変換した値は、距離を計算した後で適用されます。使用できるオプションは、[絶対値]、[符号変換]、および [0 ～ 1 の範囲で尺度化] です。

階層クラスタ分析の統計

図 25-3
[階層クラスタ分析: 統計] ダイアログ ボックス



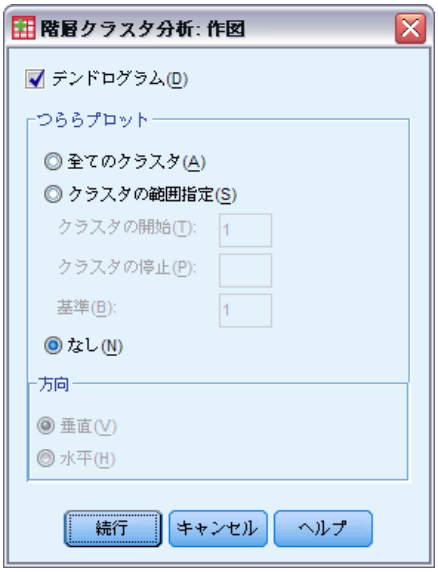
クラスタ凝集経過工程。各段階で結合されたケースまたはクラスタ、結合中のケースまたはクラスタ間の距離、およびケース (または変数) がクラスタに結合した最後のクラスタ レベルを表示します。

距離行列。項目間の距離または類似度を計算します。

所属クラスタ。クラスタを結合する 1 つ以上の段階で、各ケースが割り当てられているクラスタを表示します。使用できるオプションは、[単一の解] および [解の範囲] です。

階層クラスタ分析の作図

図 25-4
[階層クラスタ分析: 作図] ダイアログ ボックス

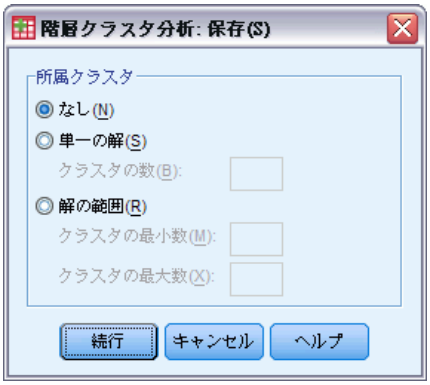


デンドログラム。デンドログラムを表示します。デンドログラムは、形成されたクラスタの結合性の評価に使用し、適当なクラスタ数を保持するための情報を提供します。

つららプロット。すべてのクラスタまたは指定された範囲のクラスタなどの **つららプロット**を表示します。つららプロットは、分析の反復ごとに、ケースのクラスタへの結合のしかたに関する情報を表示します。[方向] では、[垂直] または [水平] をクリックして、作図の方向を指定します。

階層クラスタ分析の新変数の保存

図 25-5
[階層クラスタ分析: 新変数の保存] ダイアログ ボックス



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。保存変数は、その後の分析で、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 シンタックスを使用すると、次の作業も実行できます。

- 1つの分析にいくつかのクラスタの方法を使用。
- 近接行列の読み込み、および分析。
- 後で分析するために、近接行列をディスクに書き込み。
- ユーザー指定の([べき乗])距離速度に、べき乗と平方根の値を任意に指定。
- 保存変数名を指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章

26

大規模ファイルのクラスタ分析

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、選んだ特性に基づくケースの中で相対的に等質のグループの識別を試みます。ただし、このアルゴリズムを使用するには、クラスタの個数を指定する必要があります。あらかじめわかっていれば、初期クラスタ中心を指定できます。ケースを分類するには、クラスタ中心を反復更新するか、分類するだけにするかの 2 つの方法のうち、どちらかを選択できます。保存できるのは、所属クラスタ、距離情報、クラスタ中心の最終値です。オプションとして、値をケースごとの出力のラベル付けに使用する変数を指定できます。また、分散分析の F 統計量も要求できます。この統計は便宜的なものですが（この手続きでは性質の違うグループを形成しようとするため）、統計量のサイズを比較すると、グループの分離に対する変数ごとの寄与率の情報を得ることができます。

例:€ 各グループ内で同じような視聴者層を持つテレビ番組のグループで識別可能なの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化できます。この処理は、コマーシャルの時間帯の特定に使用できます。また、都市（ケース）を等質グループにクラスタし、比較する都市を選択していろいろなマーケティング戦略を検定できます。

統計量 完全な解として初期クラスタ中心、分散分析表。ケースごとのクラスタ情報、クラスタ中心からの距離。

データ。 変数は区間尺度または比例尺度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、階層クラスタ分析手続きを使用します。

ケースと初期クラスタ中心の順序。 初期クラスタ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変化します。［反復］ダイアログの［**移動平均を使用**］オプションを使用すると、初期クラスタ中心の選択方法に関係なく、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している場合、異なる無作為な順序で並べ替えら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できます。初期クラスタ中心を指定し、［**移動平均を使用**］オプションを使用しない場合は、ケースの順序に関連する問題を回避できます。ただし、ケースからクラスタ中心に同一の距離がある場合、初期クラスタ中心の順序は解に影響する場合があります。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。

© Copyright SPSS Inc. 1989, 2010

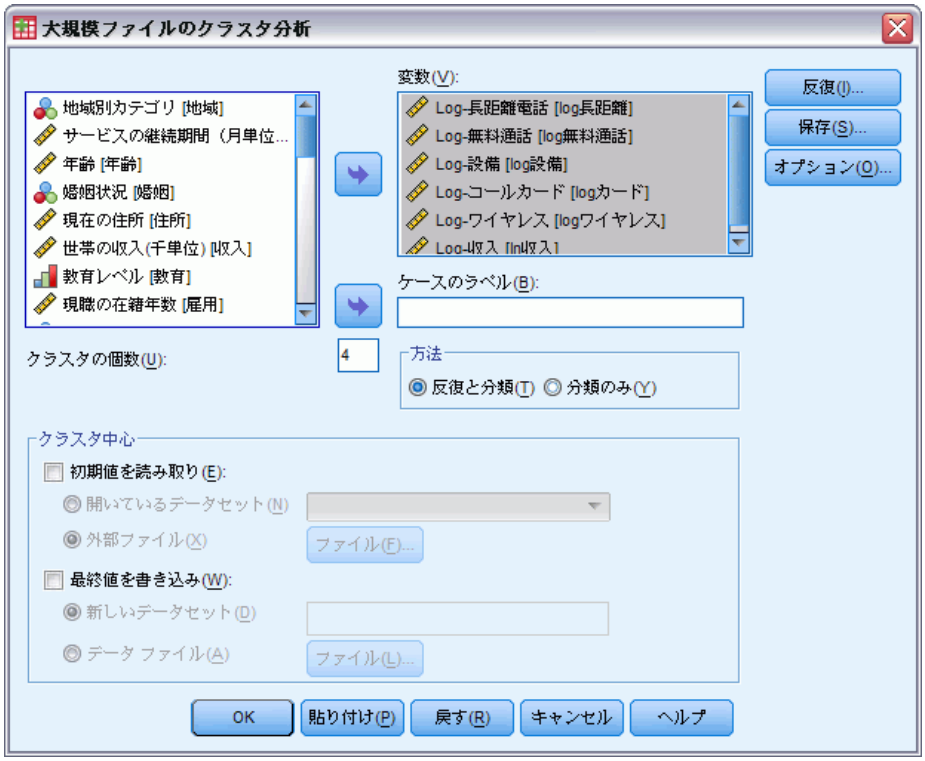
仮定。距離は、単純なユークリッド距離を使用して計算します。距離または類似度の別の測定方法を使用したい場合は、階層クラスタ分析手続きにします。変数の尺度は、重要な考慮事項です。変数が異なる尺度で測定されていると（ある変数はドル単位、別の変数は年単位など）、結果が疑わしくなります。このような場合には、大規模ファイルのクラスタ分析を実行する前に、まず変数を標準化する必要があります（このタスクは、［記述統計］で行うことができます）。この手続きでは、適切な個数のクラスタを選択し、必要な変数がすべて含まれていると仮定します。選んだクラスタの個数が不適切だったり、重要な変数が抜けてたりしていると、適切な結果を得られない場合があります。

大規模ファイルのクラスタ分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > 分類 > 大規模ファイルのクラスタ...

図 26-1
［大規模ファイルのクラスタ分析］ダイアログ ボックス



▶ クラスタ分析で使用する変数を選びます。

- ［クラスタの個数］ボックスでクラスタの個数を指定します。（クラスタの個数は、2 つ以上で、そのデータ ファイルのケース数以下でなければなりません。）
- ［**反復と分類**］または［**分類のみ**］を選択します。
- オプションとして、ケースにラベルを付けるための識別変数を選択できます。

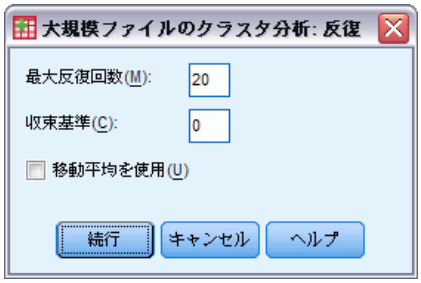
大規模ファイルのクラスタ分析の効率

大規模ファイルのクラスタ分析のコマンドは、［階層クラスタ分析］コマンドで採用するアルゴリズムを始めとする多くのクラスタ化の方法とは異なり、距離の計算でケースのすべてのペア間が対象とならないため、基本的には効率的です。

効率を最高にするには、ケースのサンプルを使用し、［**反復と分類**］をクリックして、クラスタ中心を決めます。［**最終値の書き出し**］を選択します。全体のデータ ファイルを読み込んでから、［**分類のみ**］を方法として選択、［**初期値の読み込み**］を選択し、サンプルから推定した中心を使ってファイル全体を分類します。ファイルやデータセットからの読み込み、およびそれらに対する書き出しを行なうことができます。データセットは、同じセッションの今後で利用可能ですが、セッション終了前に明示的に保存しない限り、保存されません。データセット名は、変数命名規則に従う必要があります。

大規模ファイルのクラスタ分析の反復

図 26-2
［大規模ファイルのクラスタ分析: 反復］ダイアログ ボックス



注: これらのオプションが利用可能となるのは、［**反復と分類**］方法を［大規模ファイルのクラスタ分析］ダイアログ ボックスで選択した場合のみです。

最大反復回数。大規模ファイルのクラスタ分析のアルゴリズムでの最大反復回数に上限を設けます。収束基準が満たされない場合でも、この回数の反復が終われば、それ以上の反復は行われません。入力する値は、1 から 999 の範囲内でなければなりません。

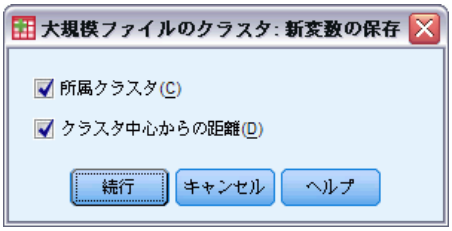
バージョン 5.0 以前の Quick Cluster コマンドで使用するアルゴリズムを再現するには、[**最大反復回数**] ボックスの値を 1 に設定します。

収束基準。反復停止のタイミングを決定します。最初のクラスタ中心間の最小距離の比率を表すため、1 ではなく、0 より大きくなければいけません。たとえば、基準が 0.02 の場合、完全な反復が、最初のクラスタ中心間の最小距離の 2% を超える距離で、いずれかのクラスタ中心が移動しない場合、反復が停止します。

移動平均を使用。ケースごとの割り当てが終わった後にクラスタ中心を更新するよう要求できます。このオプションを選ばないと、新しいクラスタ中心の計算は、すべてのケースの割り当てが終わった後に行われます。

大規模ファイルのクラスタ分析の保存

図 26-3
[大規模ファイルのクラスタ分析: 新変数の保存] ダイアログ ボックス



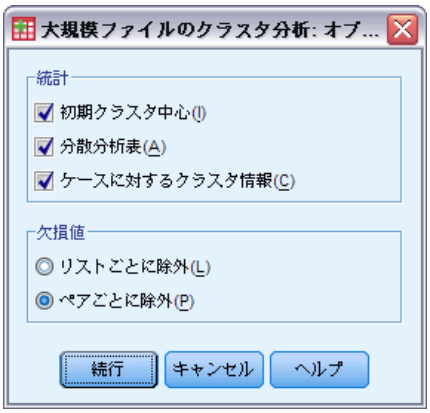
解に関する情報は、新変数として保存し、その後の分析に利用できます。

所属クラスタ。ケースごとの最終所属クラスタを示す新変数を作成します。新変数の値は、1 からクラスタの個数の範囲内です。

クラスタ中心からの距離。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変数を作成します。

大規模ファイルのクラスタ分析のオプション

図 26-4
[大規模ファイルのクラスタ分析: オプション] ダイアログ ボックス



統計量 初期クラスタ中心、分散分析表、各ケースのクラスタ情報の統計から選択できます。

- **初期クラスタ中心(I).** 各クラスタに対する、変数の最初の平均値。デフォルトでは、クラスタの個数と同じ数の、十分に間隔のあるケースがデータから選択されます。初期クラスタ中心は分類の最初の一巡のために使われ、その後に更新されます。
- **分散分析表 (カテゴリ/カテゴリの回帰).** 各クラスタ変数に対して 1 変量 F 検定を含む分散分析表を表示します。F 検定は記述のみなので、結果として作成される確率を解釈することはできません。分散分析表は、すべてのケースが 1 つのクラスタに割り当てられない場合、表示されません。
- **ケースに対するクラスタ情報(C).** それぞれのケースについて、最終クラスタ割り当て、およびケースを分類するためのケースとクラスタ中心の間のユークリッド距離を表示します。さらに、最終クラスタ中心間のユークリッド距離も表示します。

欠損値。 使用できるオプションは、[**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。** クラスタ変数に対して欠損値のある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分析手続きでは、QUICK CLUSTER コマンド シンタックスを使用します。コマンド シンタックスを使用すると、次の作業も実行できます。

- 最初の K-ケース を初期クラスタ中心とします。これにより、推定にデータ パスの使用を避けることができます。
- 初期クラスタ中心をコマンド シンタックスの一部として指定します。
- 保存変数名を指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
27

ノンパラメトリック検定

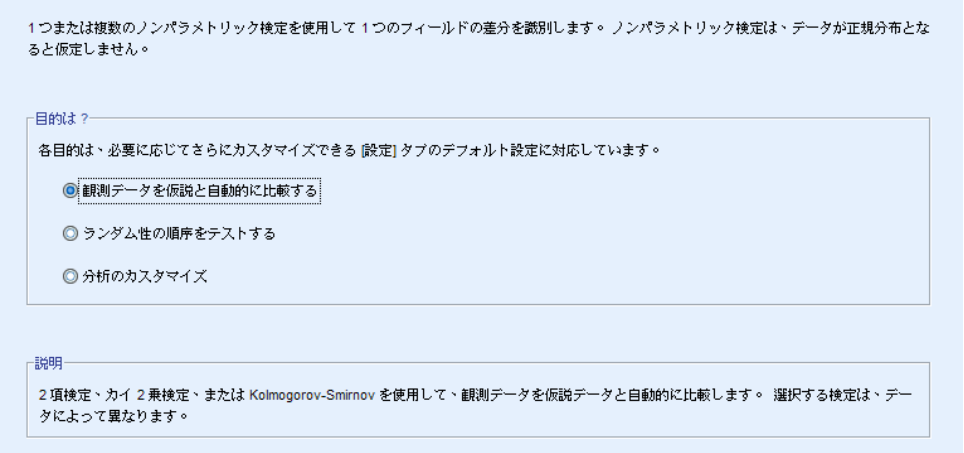
ノンパラメトリック検定は、データの基本的な分布についての仮定は最小限に行います。これらのダイアログで使用できる検定は、データの構成に基づいて、次の 3 つの大きなカテゴリにグループ化されます。

- 1 サンプル検定は 1 つのフィールドを分析します。
- 対応サンプルの検定では、同じセットのケースの 2 つ以上のフィールドを比較します。
- 独立サンプル検定は、別のフィールドのカテゴリでグループ化された 1 つのフィールドを分析します。

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 つまたは複数のノンパラメトリック検定を使用して 1 つの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は、データが正規分布となると仮定しません。

図 27-1
1 サンプルのノンパラメトリック検定の [目的] タブ



目的は ? 目的により、異なるが一般的に使用され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。**カテゴリ数が 2 つだけのカテゴリ フィールドには 2 項検定を、その他すべてのカテゴリ フィールドにはカイ 2 乗検定を、連続型フィールドには Kolmogorov-Smirnov 検定を適用します。
- **ランダム性の順序をテストする。**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。継続して［設定］タブのオプションに変更を行うも、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ます。

1 サンプルのノンパラメトリック検定を行うには

メニューから次の項目を選択します。

分析(A) > ノンパラメトリック検定 > 1 サンプル...

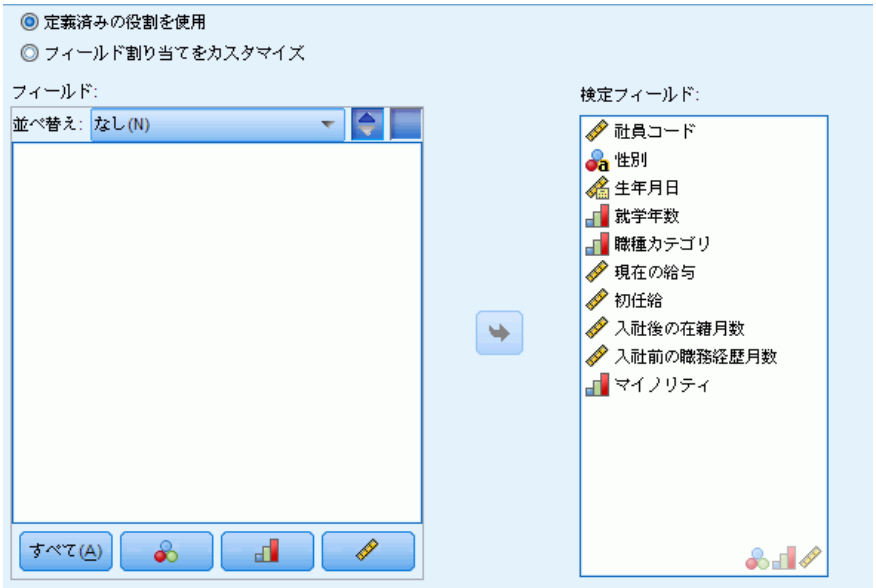
▶ ［実行］をクリックします。

オプションとして、次の選択が可能です。

- ［目的］タブで目的を指定します。
- ［フィールド］タブでフィールドの割り当てを指定します。
- ［設定］タブでエキスパート設定を指定します。

[フィールド] タブ

図 27-2
1 サンプルのノンパラメトリック検定の [フィールド] タブ



[フィールド] タブは、検定する必要のあるフィールドを指定します。

事前定義された役割を使用: このオプションを選択すると、既存のフィールド情報を使用します。事前定義された役割が [入力]、[目標] または [両方] に指定されているすべてのフィールドは、検定フィールドとして使用されます。検定フィールドは、少なくとも 1 つ必要です。

カスタム フィールド割り当ての使用: フィールドの役割を上書きできます。このオプションを選択した後、次のフィールドを指定します。

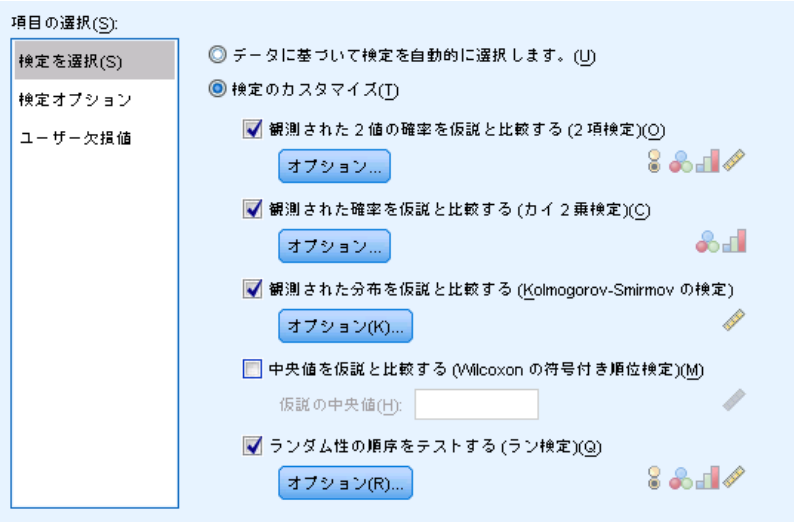
- **検定フィールド。** 1 つまたは複数のフィールドを選択します。

[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デフォルト設定に変更を行うと、[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

図 27-3
1 サンプルのノンパラメトリック検定の [検定の選択] 設定



[フィールド] タブで指定したフィールドで実行する検定を指定します。

データに基づいて自動的に検定を選択します。カテゴリ数が 2 つだけの有効な (欠損値のない) カテゴリ フィールドには 2 項検定を、その他すべてのカテゴリ フィールドにはカイ 2 乗検定を、連続型フィールドには Kolmogorov-Smirnov 検定を適用します。

検定のカスタマイズ。実行する特定の検定を選択できます。

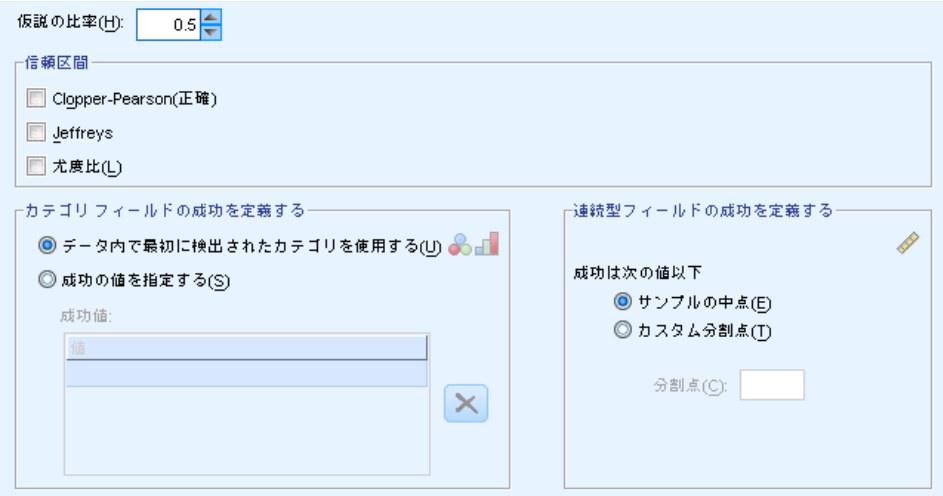
- **観測された 2 値の確率を仮説と比較する (2 項検定)** すべてのフィールドに 2 項検定を適用できます。フラグ型フィールド (カテゴリ数が 2 つだけのカテゴリ フィールド) 観測された分布が、指定された 2 項分布から期待されるものと同じかどうかを検定する 1 サンプル検定を作成します。 また、信頼区間を要求できます。検定の設定に関する詳細は、「2 項検定オプション」を参照してください。
- **観測された確率を仮説と比較する (カイ 2 乗検定)。**名義型フィールドおよび順序型フィールドにカイ 2 乗検定を適用できます。フィールドの観測度数と期待度数の差に基づいて、カイ 2 乗統計量を比較する 1 サンプル検定が作成されます。検定の設定に関する詳細は、「カイ 2 乗検定オプション」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-Smimov の検定)。**連続型フィールドに Kolmogorov-Smirnov 検定が適用されます。1 つのフィールドのサンプル累積分布関数を、正規、一様、ポアソン、または指数分布と等質であるかどうかを検定する 1 サンプル検定を作成します。

検定の設定に関する詳細は、「Kolmogorov-Smirnov 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 signed-rank 検定は連続型フィールドに適用されます。フィールドの中央値に対する 1 サンプル検定を作成します。数値を仮説の中央値として指定します。
- **ランダム性の順序をテストする (ラン検定)。** すべてのフィールドにラン検定を適用できます。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 サンプル検定を作成します。検定の設定に関する詳細は、「ラン検定オプション」を参照してください。

2 項検定オプション

図 27-4
1 サンプルのノンパラメトリック検定の 2 項検定オプション



2 項検定はフラグ型フィールド (カテゴリが 2 つだけのカテゴリ型フィールド) のみを対象としていますが、「成功」を定義するルールに基づき、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される比率、または p を指定します。0 より大きく、1 より小さいの値を指定します。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、次のような方法を使用できます。

- **Clopper-Pearson(正確)。** 累積 2 項分布に基づいた正確な区間。
- **Jeffreys。** Jeffreys を以前に使用し、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。仮説の比率に対して検定されたデータ値である、カテゴリ 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。

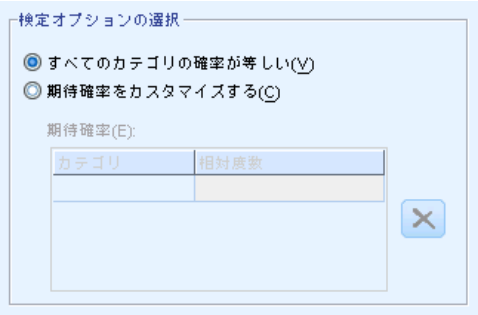
- ［**データ内で最初に検出されたカテゴリを使用する**］を指定すると、サンプル内で最初に検出された値を使用して 2 項検定を実行し、「成功」を定義します。値が 2 つだけの名義型フィールドまたは順序型フィールドにのみ適用できます。このオプションが使用されている［フィールド］タブで指定されたその他すべてのカテゴリ フィールドは検定されません。これはデフォルトです。
- ［**成功の値を指定する**］を指定すると、指定された値のリストを使用して 2 項検定を実行し、「成功」を定義します。文字列値または数値のリストを指定します。リストの値は、サンプルのとおりである必要はありません。

連続型フィールドの成功を定義する。検定値に対して検定されたデータ値である、連続型フィールドの「成功」をどのように定義するのかを指定します。成功は、分割点以下の値に定義されます。

- ［**サンプルの中点**］で、最小値と最大値の平均に分割点を設定します。
- ［**カスタム分割点**］で、分割点の値を指定できます。

カイ 2 乗検定オプション

図 27-5
1 サンプルのノンパラメトリック検定のカイ 2 乗検定オプション



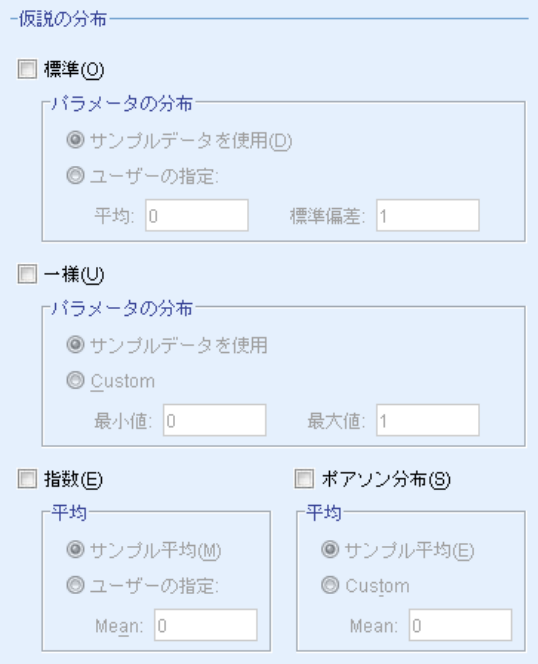
すべてのカテゴリの確率が等しい。サンプルのすべてのカテゴリで同等の度数を作成します。これはデフォルトです。

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ます。リストの値は、サンプルのとおりである必要はありません。［**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。1、2、3 を指定することは度数 10、20、および 30 を指定することと同じであるように、カスタム度数は比率として処理されます。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 番目のカテゴリになる

よう指定します。ユーザー指定の期待確率が指定されると、ユーザー指定のカテゴリ値には、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。指定していない場合、そのフィールドに検定は実行されません。

Kolmogorov-Smirnov オプション

図 27-6
1 サンプルのノンパラメトリック検定の Kolmogorov-Smirnov オプション



検定する必要のある分布、および仮説の分布のパラメータを指定します。

正規分布。[サンプル データの使用] では観測された平均値および標準偏差を使用し、[カスタム] で値を指定できます。

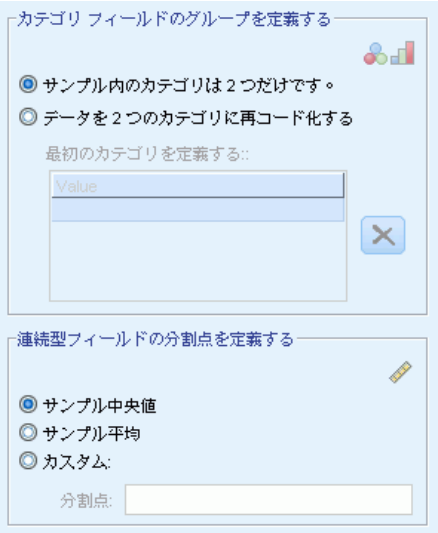
一様分布。[サンプル データの使用] では観測された最小値および最大値を使用し、[カスタム] で値を指定できます。

指数。[サンプルの平均] では観測された平均値を使用し、[カスタム] で値を指定できます。

ポアソン分布。[サンプルの平均] では観測された平均値を使用し、[カスタム] で値を指定できます。

ラン検定オプション

図 27-7
1 サンプルのノンパラメトリック検定のラン検定オプション



ラン検定はフラグ型フィールド（カテゴリが 2 つだけのカテゴリ型フィールド）のみを対象としていますが、グループを定義するルールに基づき、すべてのフィールドに適用されます。

カテゴリ フィールドのグループを定義する

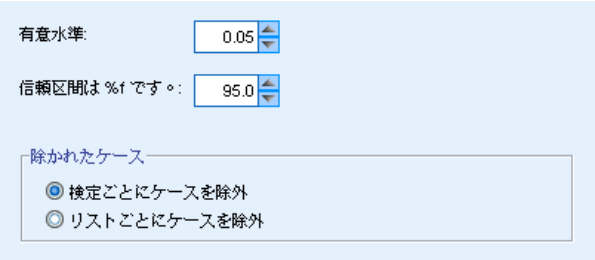
- [**サンプルには 2 つのカテゴリのみ**] を指定すると、サンプルで検出された値を使用してラン検定を実行し。グループを定義します。値が 2 つだけの名義型フィールドまたは順序型フィールドにのみ適用できます。このオプションが使用されている [フィールド] タブで指定されたその他すべてのカテゴリ フィールドは検定されません。
- [**データを 2 つのカテゴリに再コード化する**] を指定すると、値の指定されたリストを使用してラン検定を実行し、グループのいずれかを定義します。サンプルのその他すべての値は他のグループを定義します。リストの値は、サンプルのとおりである必要はありませんが、少なくとも 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をどのように定義するのかを指定します。最初のグループは分割点以下の値に定義されます。

- [**サンプル中央値**] を指定して、分割点をサンプル中央値に設定します。
- [**サンプル平均値**] を指定して、分割点をサンプル平均値に設定します。
- [**カスタム**] で、分割点の値を指定できます。

検定オプション

図 27-8
1 サンプルのノンパラメトリック検定の検定オプションの設定



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。

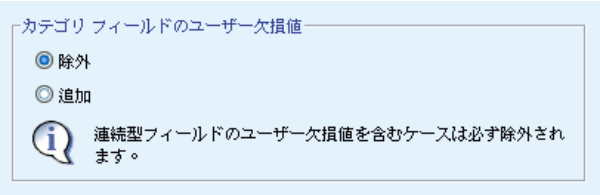
信頼区間 (%)。作成されたすべての信頼区間の確信度を指定します。0 から 100 までの値を指定してください。デフォルトは 95 です。

除かれたケース。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] タブで指定されたフィールドの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の欠損値を持つレコードは、その検定から除外されます。分析で複数の検定を指定する場合、各検定は個別に評価されます。

ユーザー欠損値

図 27-9
1 サンプルのノンパラメトリック検定の [ユーザー欠損値] 設定

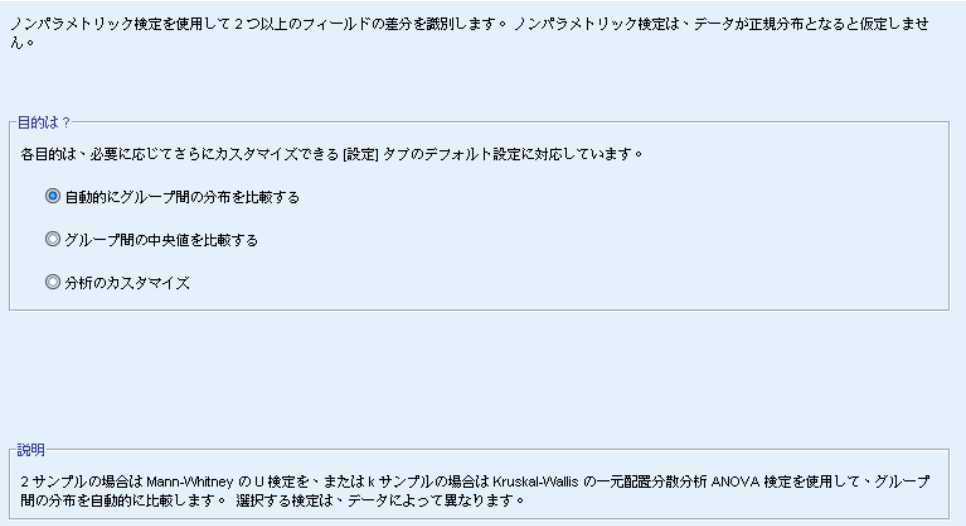


カテゴリ フィールドのユーザー欠損値。カテゴリ フィールドは、分析の対象となるレコードに対して有効な値を取る必要があります。このオプションを使用すると、ユーザー欠損値をカテゴリ フィールドで有効な値として扱うかどうかを決定できます。システム欠損値および連続型フィールドの欠損値は常に無効なものとして処理されます。

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では、1 つまたは複数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識別します。ノンパラメトリック検定は、データが正規分布となると仮定しません。

図 27-10
独立サンプルのノンパラメトリック検定の [目的] タブ



目的は？ 目的により、異なるが一般的に使用され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。**2 サンプルの場合は Mann-Whitney の U 検定を、または k グループの場合は Kruskal-Wallis の一元配置分散分析 ANOVA 検定をデータに適用します。
- **グループ間の中央値を比較する。**中央値検定を使用して、グループ間で観測された中央値を比較します。
- **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。継続して [設定] タブのオプションに変更を行うも、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ます。

独立サンプルのノンパラメトリック検定を行うには

メニューから次の項目を選択します。

分析(A) > ノンパラメトリック検定 > 独立サンプル...

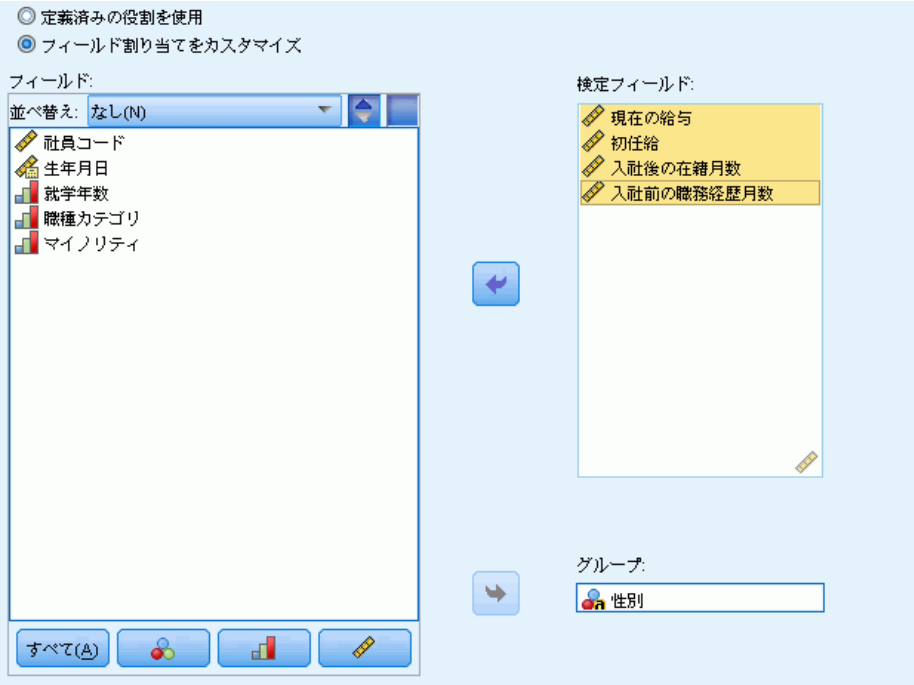
▶ [実行] をクリックします。

オプションとして、次の選択が可能です。

- ［目的］タブで目的を指定します。
- ［フィールド］タブでフィールドの割り当てを指定します。
- ［設定］タブでエキスパート設定を指定します。

［フィールド］タブ

図 27-11
独立サンプルのノンパラメトリック検定の［フィールド］タブ



［フィールド］タブは、検定する必要のあるフィールドおよびグループを定義するの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: このオプションを選択すると、既存のフィールド情報を使用します。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 1 つのカテゴリフィールドがある場合、グループ フィールドとして使用されます。そうでない場合、グループ フィールドはデフォルトで使用されず、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とグループ 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: フィールドの役割を上書きできます。このオプションを選択した後、次のフィールドを指定します。

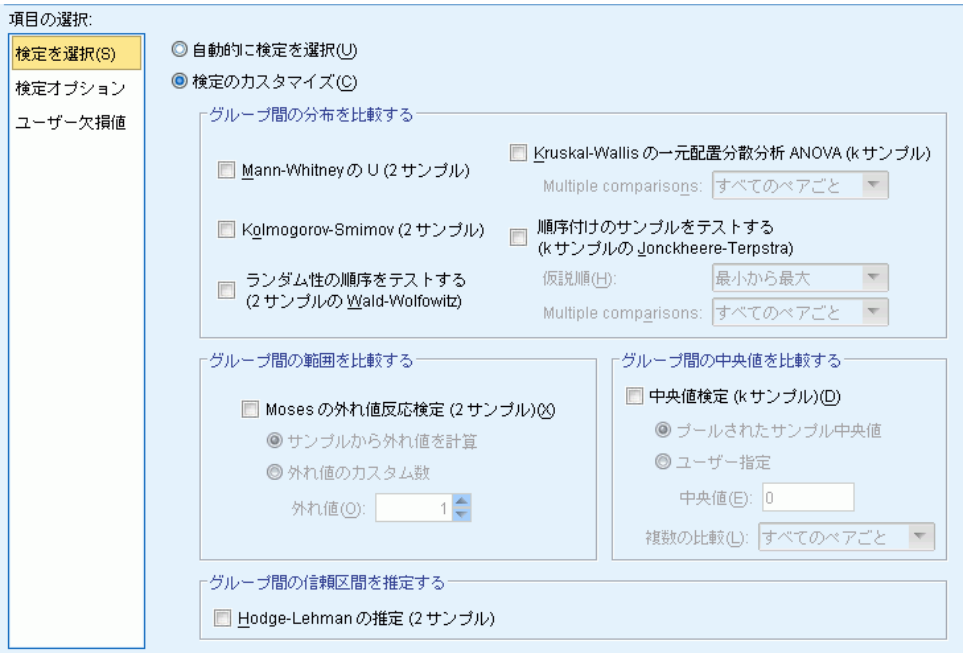
- **検定フィールド。**1 つ以上の連続型フィールドを選択します。
- **グループ。**カテゴリ フィールドを選択します。

[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デフォルト設定に変更を行うと、[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

図 27-12
独立サンプルのノンパラメトリック検定の [検定の選択] 設定



[フィールド] タブで指定したフィールドで実行する検定を指定します。

データに基づいて自動的に検定を選択します。2 サンプルの場合は Mann-Whitney の U 検定を、または k グループの場合は Kruskal-Wallis の一元配置分散分析 ANOVA 検定をデータに適用します。

検定のカスタマイズ。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。**サンプルがおな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、グループがおなじ母集団から抽出されているかどうかを検定します。グループフィールド内で昇順に並んだ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 フィールドに 3 つ以上の値が指定されている場合、検定は行われません。

[Kolmogorov-Smirnov (2 サンプル)] は、2 つの分布間の中央値、ばらつき度、歪度などの差分に対して敏感です。グループ フィールドに 3 つ以上の値が指定されている場合、検定は行われません。

[ランダム性の順序をテストする (2 サンプルの Wald-Wolfowitz)] を指定すると、基準として所属グループを指定して検定を実行します。グループ フィールドに 3 つ以上の値が指定されている場合、検定は行われません。

[Kruskal-Wallis の一元配置分散分析 ANOVA (k サンプル)] は、Mann-Whitney 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、k サンプルの多重比較、[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 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。たとえば、k 個の母集団が k 段階の上昇温度を表す場合があるとします。異なる温度でも同じ応答分布を示すという仮説は、温度が上昇するにつれて応答の大きさが上昇するという対立仮説に対して検定されます。ここで、対立仮説が順序付けされるので、Jonckheere-Terpstra が最適の検定になります。対立仮説の順序を指定します。[最小から最大] を指定すると最初のグループの位置パラメータが 2 番目と等しくなく、2 番目が 3 番目と等しくないという対立仮説を規定します。[最大から最小] を指定すると、最後のグループの位置パラメータが最後から 2 番目と等しくなく、最後から 2 番目が最後から 3 番目と等しくないという対立仮説を規定します。オプションで、k サンプルの多重比較、[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。

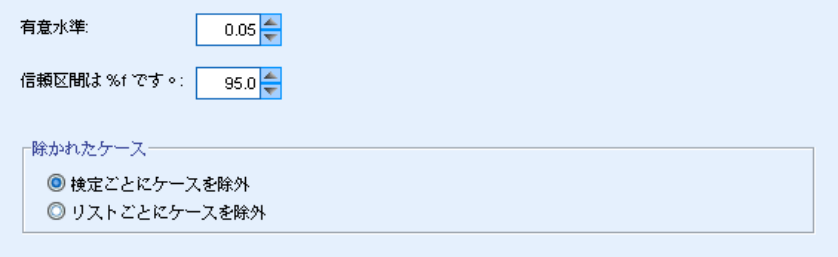
- **グループ間の範囲を比較する。**サンプルの範囲が同じかどうかに関する独立サンプル検定を作成します。[Moses の外れ値反応検定 (2 サンプル)] を指定すると、比較グループに対して対照グループを検定します。グループフィールド内で昇順に並んだ最初の値が対照グループを定義し、2 番目の値が比較グループを定義します。グループ フィールドに 3 つ以上の値が指定されている場合、検定は行われません。
- **グループ間の中央値を比較する。**サンプルの中央値が同じかどうかに関する独立サンプル検定を作成します。[中央値検定 (k サンプル)] では、プールされたサンプル中央値 (データセットの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します。オプションで、k

サンプルの多重比較、[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及び信頼区間を作成します。グループ フィールドに 3 つ以上の値が指定されている場合、検定は行われません。

検定オプション

図 27-13
独立サンプルのノンパラメトリック検定の [検定オプション] 設定



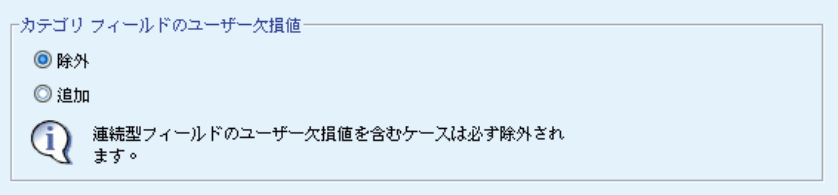
有意水準。 すべての検定の有意水準 (アルファ) を指定します。0 から 1 までの値を指定してください。デフォルトは 0.05 です。

信頼区間 (%)。 作成されたすべての信頼区間の確信度を指定します。0 から 100 までの値を指定してください。デフォルトは 95 です。

除かれたケース。 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されたフィールドの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の欠損値を持つレコードは、その検定から除外されます。分析で複数の検定を指定する場合、各検定は個別に評価されます。

ユーザー欠損値

図 27-14
独立サンプルのノンパラメトリック検定の [ユーザー欠損値] 設定



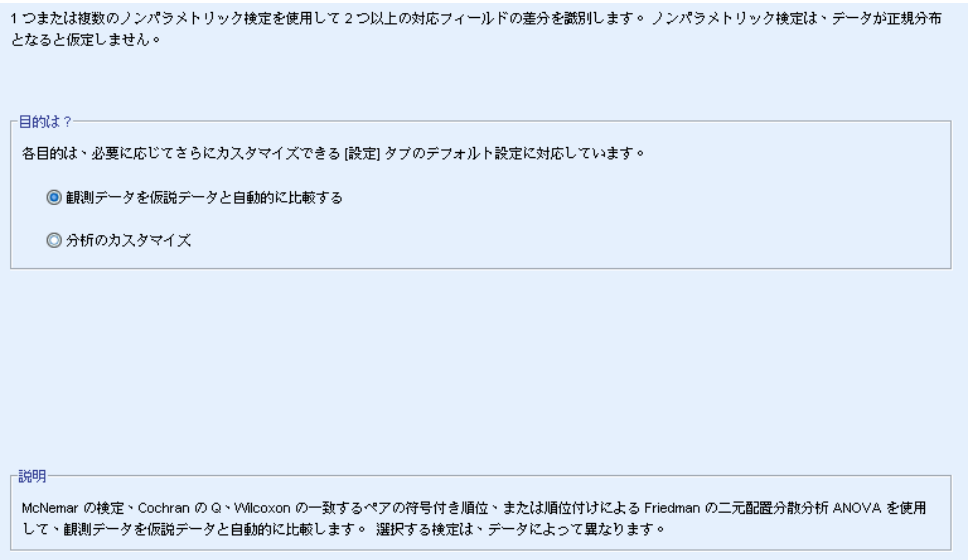
カテゴリ フィールドのユーザー欠損値。カテゴリ フィールドは、分析の対象となるレコードに対して有効な値を取る必要があります。このオプションを使用すると、ユーザー欠損値をカテゴリ フィールドで有効な値として扱うかどうかを決定できます。システム欠損値および連続型フィールドの欠損値は常に無効なものとして処理されます。

対応サンプルのノンパラメトリック検定

1 つまたは複数のノンパラメトリック検定を使用して 2 つ以上の対応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は、データが正規分布となると仮定しません。

データの考慮事項各レコードは、2 つ以上の関連する測定がデータセットの各フィールドに入力されている指定の被験者に対応しています。たとえば、各被験者の体重が定期的に測定され、「ダイエット前の体重」、「中間の体重」、「ダイエット後の体重」などのフィールドに入力されている場合、ダイエット計画の効果に関する研究を、対応サンプル 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。

図 27-15
対応サンプルのノンパラメトリック検定の [目的] タブ



目的は？ 目的により、異なるが一般的に使用される検定の設定を迅速に指定できます。

- **観測データを仮説データと自動的に比較する。**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 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 データに Cochran の Q を、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を、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を適用します。

- **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。継続して［設定］タブのオプションに変更を行うも、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ます。

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、適切な検定が各グループに適用されます。たとえば、**［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行うには

メニューから次の項目を選択します。
分析(A) > ノンパラメトリック検定 > 対応サンプル...

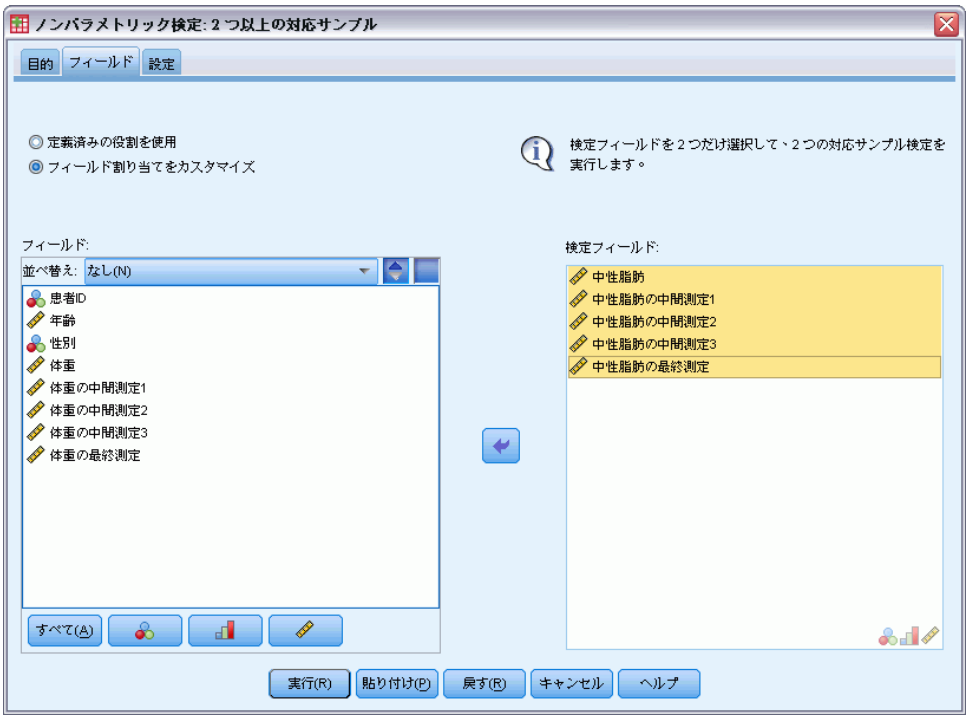
▶ **［実行］**をクリックします。

オプションとして、次の選択が可能です。

- ［目的］タブで目的を指定します。
- ［フィールド］タブでフィールドの割り当てを指定します。
- ［設定］タブでエキスパート設定を指定します。

[フィールド] タブ

図 27-16
対応サンプルのノンパラメトリック検定の [フィールド] タブ



[フィールド] タブは、検定する必要のあるフィールドを指定します。

事前定義された役割を使用: このオプションを選択すると、既存のフィールド情報を使用します。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フィールドは、検定フィールドとして使用されます。検定フィールドは、少なくとも 2 つ必要です。

カスタム フィールド割り当ての使用: フィールドの役割を上書きできます。このオプションを選択した後、次のフィールドを指定します。

- **検定フィールド。** 2 つ以上のフィールドを選択します。各フィールドは各対応サンプルに対応しています。

[設定] タブ

[設定] タブは、手続きで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。その他の目的と互換性のないデフォルト設定に変更を行うと、[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

図 27-17
対応サンプルのノンパラメトリック検定の [検定の選択] 設定

[フィールド] タブで指定したフィールドで実行する検定を指定します。

データに基づいて自動的に検定を選択します。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 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 データに Cochran の Q を、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を、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を適用します。

検定のカスタマイズ。実行する特定の検定を選択できます。

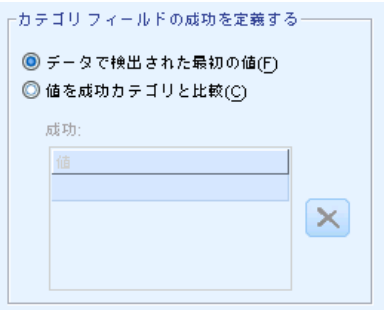
- **2 値のデータの変更の検定。**[McNemar の検定 (2 サンプル)] をカテゴリ フィールドに適用できます。2 つのフラグ型フィールド (値が 2 つのみのカテゴリ型フィールド) 間の値の組み合わせが 等しいかどうかに関する対応サンプルの検定を作成します。[フィールド] タブで 3 つ以上のフィールドが指定されている場合、この検定は実行されていません。検定の設定に関する詳細は、「McNemar の検定:成功の定義」を参照してください。[Cochran の Q (k サンプル)] をカテゴリ　フィールドに適用できます。k 件のフラグ型フィールド (値が 2 つのみのカテゴリ型フィールド) 間の値の組み合わせが 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、k サンプルの多重比較、[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

較を要求できます。検定の設定に関する詳細は、「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)] で、2 つのペアになっている順序型フィールド間の値の組み合わせが等しいかどうかに関するタイプ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 値反応から多値反応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ていません。
- **中央値の差分を仮説と比較。**これらの検定は、それぞれ 2 つの連続型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。[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、検定は実行されません。
- **信頼区間の推定。**2 つのペアになっている連続型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。[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ていません。
- **関連性の数量化。**[Kendall の一致係数 (k サンプル)] は、審判または評価者間の一致どの尺度を作成します。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、k サンプルの多重比較、[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] の比較を要求できます。
- **分布の比較。**[順位付けによる Friedman の二元配置分散分析 ANOVA (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タイプサンプルの検定を作成します。オプションで、k サンプルの多重比較、[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] の比較を要求できます。

McNemar の検定:成功の定義

図 27-18
対応サンプルのノンパラメトリック検定の [McNemar の検定: 成功の定義] 設定



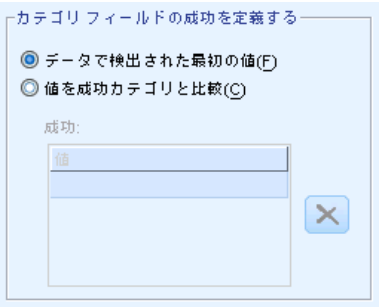
McNemar の検定はフラグ型フィールド（カテゴリが 2 つだけのカテゴリ型フィールド）のみを対象としていますが、「成功」を定義するルールに基づき、すべてのカテゴリ フィールドに適用されます。

カテゴリ フィールドの成功を定義する。カテゴリ 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。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］タブで指定され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ます。リストの値は、サンプルのとおりである必要はありません。

Cochran の Q:成功の定義

図 27-19
対応サンプルノンパラメトリック検定の Cochran の Q:成功を定義



Cochran の Q 検定はフラグ型フィールド（カテゴリが 2 つだけのカテゴリ型フィールド）のみを対象としていますが、「成功」を定義するルールに基づき、すべてのカテゴリ フィールドに適用されます。

カテゴリ フィールドの成功を定義する。カテゴリ 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。

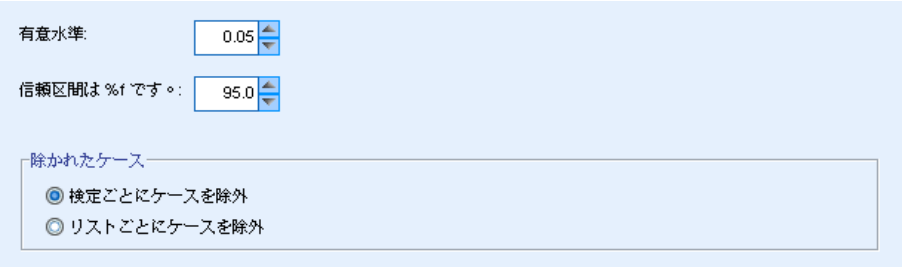
- ［データ内で最初に検出されたカテゴリを使用する］を指定すると、サンプル内で最初に検出された値を使用して検定を実行し、「成功」を定義します。値が 2 つだけの名義型フィールドまたは順序型フィールドにのみ適用できます。このオプションが使用されている［フィールド］タ

ブで指定され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ます。リストの値は、サンプルのとおりである必要はありません。

検定オプション

図 27-20
対応サンプルのノンパラメトリック検定の [検定オプション] 設定



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。

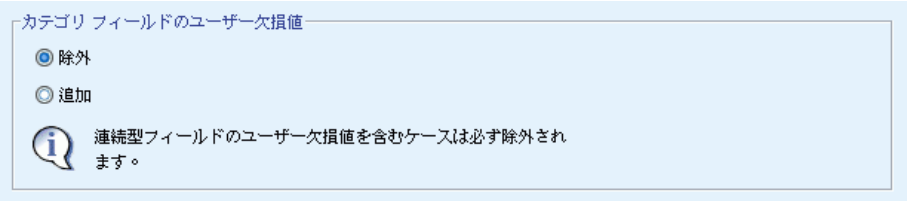
信頼区間 (%)。作成されたすべての信頼区間の確信度を指定します。0 から 100 までの値を指定してください。デフォルトは 95 です。

除かれたケース。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されたフィールドの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の欠損値を持つレコードは、その検定から除外されます。分析で複数の検定を指定する場合、各検定は個別に評価されます。

ユーザー欠損値

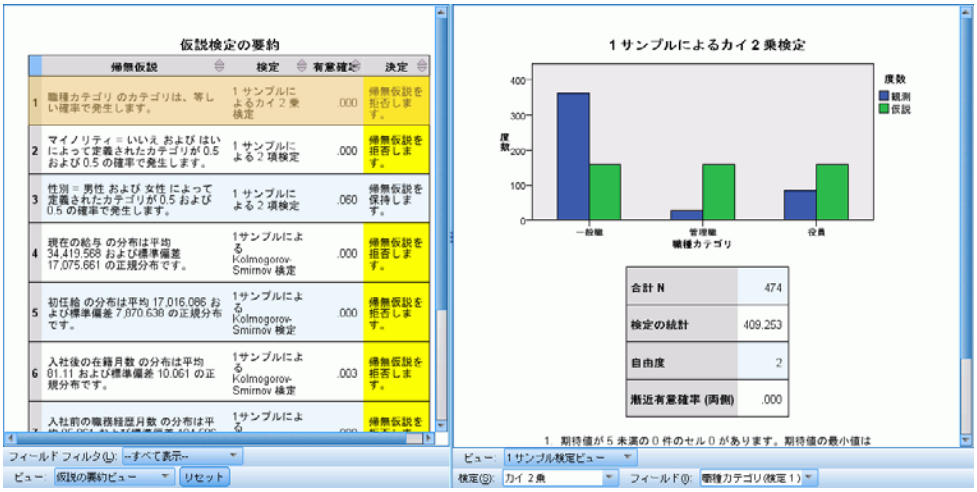
図 27-21
対応サンプルのノンパラメトリック検定の [ユーザー欠損値] 設定



カテゴリ フィールドのユーザー欠損値。カテゴリ フィールドは、分析の対象となるレコードに対して有効な値を取る必要があります。このオプションを使用すると、ユーザー欠損値をカテゴリ フィールドで有効な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なものとして処理されます。

モデル ビュー

図 27-22
ノンパラメトリック検定のモデル ビュー



この手順で、ビューアーにモデル ビューアー オブジェクトを作成します。このオブジェクトを有効化（ダブルクリック）すると、モデルの双方向ビューを取得します。モデル ビューは 2 つのパネルで構成されています。左側はメイン ビュー、右側はリンク ビューまたは補助ビューです。

メイン ビューには、2 種類があります。

- 仮説の要約。これがデフォルト ビューです。詳細は、 p. 237 仮説の要約 を参照してください。
- 信頼区間の要約。 詳細は、 p. 239 信頼区間の要約 を参照してください。

リンク/補助ビューは、次のとおりです。

- 1 サンプル検定。1 サンプル検定が要求されている場合は、これがデフォルト ビューになります。 詳細は、 p. 239 1 サンプル検定 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 サンプル検定が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。 詳細は、 p. 245 対応サンプル検定 を参照してください。

- 独立サンプル検定。対応サンプル検定も 1 サンプル検定が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。 詳細は、 p. 251 独立サンプル検定 を参照してください。
- カテゴリ フィールド情報。 詳細は、 p. 259 カテゴリ フィールド情報 を参照してください。
- 連続型フィールド情報。 詳細は、 p. 260 連続型フィールド情報 を参照してください。
- ペアごとの比較。 詳細は、 p. 261 ペアごとの比較 を参照してください。
- 等質サブセット。 詳細は、 p. 262 等質サブセット を参照してください。

仮説の要約

図 27-23
仮説の要約

仮説検定の要約

	帰無仮説	検定	有意確率	決定
1	職種カテゴリ のカテゴリは、等しい確率で発生します。	1 サンプルによるカイ 2 乗検定	.000	帰無仮説を拒否します。
2	マイノリティ = いいえ および はい によって定義されたカテゴリが 0.5 および 0.5 の確率で発生します。	1 サンプルによる 2 項検定	.000	帰無仮説を拒否します。
3	性別 = 男性 および 女性 によって定義されたカテゴリが 0.5 および 0.5 の確率で発生します。	1 サンプルによる 2 項検定	.060	帰無仮説を保持します。
4	現在の給与 の分布は平均 34,419.568 および標準偏差 17,075.661 の正規分布です。	1サンプルによる Kolmogorov-Smirnov 検定	.000	帰無仮説を拒否します。
5	初任給 の分布は平均 17,016.086 および標準偏差 7,870.638 の正規分布です。	1サンプルによる Kolmogorov-Smirnov 検定	.000	帰無仮説を拒否します。
6	入社後の在籍月数 の分布は平均 81.11 および標準偏差 10.061 の正規分布です。	1サンプルによる Kolmogorov-Smirnov 検定	.003	帰無仮説を拒否します。
7	入社前の職務経歴月数 の分布は平均 95.861 および標準偏差 104.586 の正規分布です。	1サンプルによる Kolmogorov-Smirnov 検定	.000	帰無仮説を拒否します。
8	社員コード の分布は平均 237.5 および標準偏差 136.976 の正規分布です。	1サンプルによる Kolmogorov-Smirnov 検定	.082	帰無仮説を保持します。

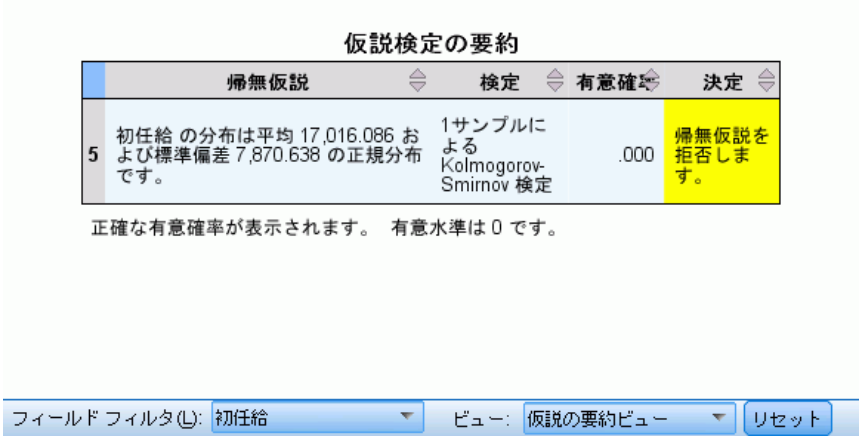
正確な有意確率が表示されます。 有意水準は 0 です。

フィールド フィルタ(L): --すべて表示--
ビュー: 仮説の要約ビュー リセット

［モデル要約］ビューはスナップショットで、ノンパラメトリック検定について一目でわかる要約です。帰無仮説と決定を強調し、重要な p 値に注意を導きます。

- 各行は、各検定に対応します。行をクリックすると、検定に関する詳細情報がリンク 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並べ替えられます。
- [**リセット**] ボタンを使用して、モデル ビューアを元の状態に戻すことができます。
- [**フィールド フィルタ**] ドロップダウン リストを使用して、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。たとえば、[**フィールド フィルタ**] で［初任給］が選択されると、仮説の要約には 2 つの検定のみが表示されます。

図 27-24
仮説の要約:「初任給」でフィルタリング。



信頼区間の要約

図 27-25
信頼区間の要約

信頼区間の要約

信頼区間型	パラメータ	推定値	漸近 95% 信頼区間 下限	漸近 95% 信頼区間 上限
1 サンプルによる 2 項の成功率 (Clopper-Pearson)	確率(性別=男性)。	.544	.498	.590
1 サンプルによる 2 項の成功率 (Jeffreys)	確率(性別=男性)。	.544	.499	.589
1 サンプルによる 2 項の成功率 (Profile Likelihood)	確率(性別=男性)。	.544	.499	.589
1 サンプルによる 2 項の成功率 (Clopper-Pearson)	確率(マイノリティ=いいえ)。	.781	.741	.817

ビュー: 信頼区間の要約ビュー リセット

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並べ替えられます。

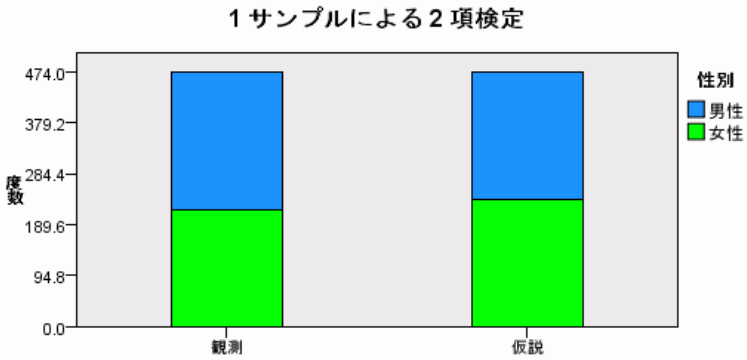
1 サンプル検定

1 サンプル検定ビューには、要求された 1 サンプル 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] ドロップダウンを使用して、指定した種類の 1 サンプル検定を選択できます。
- [フィールド] ドロップダウンを使用して、[検定] ドロップダウン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

図 27-26
1 サンプル検定ビュー、2 項検定



合計 N	474
検定の統計	258.000
標準誤差	10.886
標準化された検定の統計	1.883
漸近有意確率 (両側)	.060

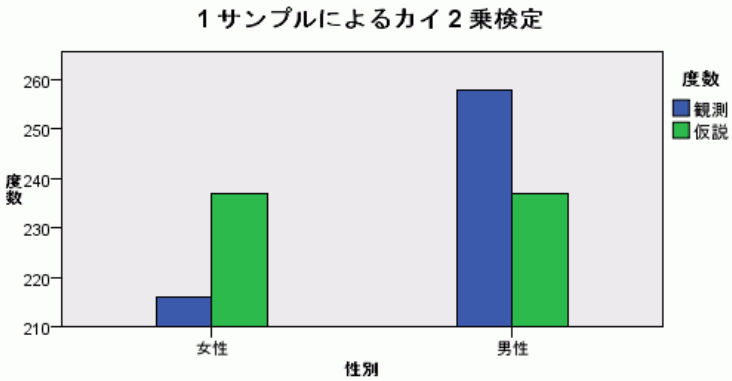
検定(S): 2 項分布　フィールド(I): 性別(検定 1)　ビュー: 1 サンプル検定ビュー

[2 項検定] には、積み上げ棒グラフと検定表が表示されます。

- 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。バーの上でマウスポインタを停止すると、カテゴリの割合が ToolTip に表示されます。棒グラフに表示される差分は、検定フィールドに仮説の 2 項分布がないことを示します。
- 表には検定の詳細が表示されます。

カイ 2 乗検定

図 27-27
1 サンプル検定ビュー、カイ 2 乗検定



合計 N	474
検定の統計	3.722
自由度	1
漸近有意確率 (両側)	.054

1. 期待値が 5 未満の 0 件のセル 0 があります。期待値の最小値は 237 です。

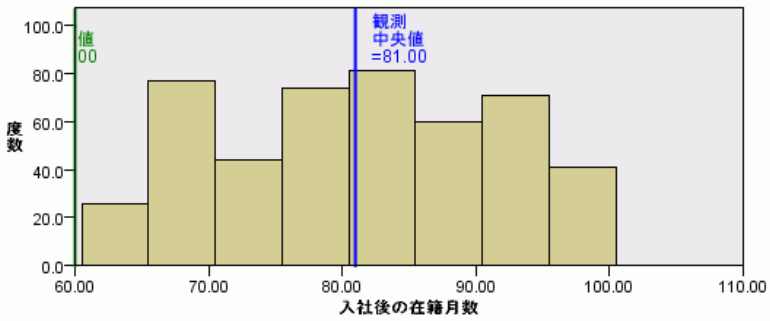
検定(S): カイ 2 乗 フィールド(I): 性別(検定 2) ビュー: 1 サンプル検定ビュー

[カイ 2 乗検定] には、クラスタ棒グラフと検定表が表示されます。

- クラスタ棒グラフには、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。バーの上でマウスポインタを停止すると、ToolTip に観測度数と仮説度数およびそれらの差 (残差) 表示されます。観測と仮説の棒グラフに表示される差分は、検定フィールドに仮説の分布がないことを示します。
- 表には検定の詳細が表示されます。

Wilcoxon の符号付き順位

図 27-28
1 サンプル検定ビュー、Wilcoxon の符号付き順位検定



合計 N	474
検定の統計	112,575.000
標準誤差	2,983.353
標準化された検定の統計	18.867
漸近有意確率 (両側)	.000

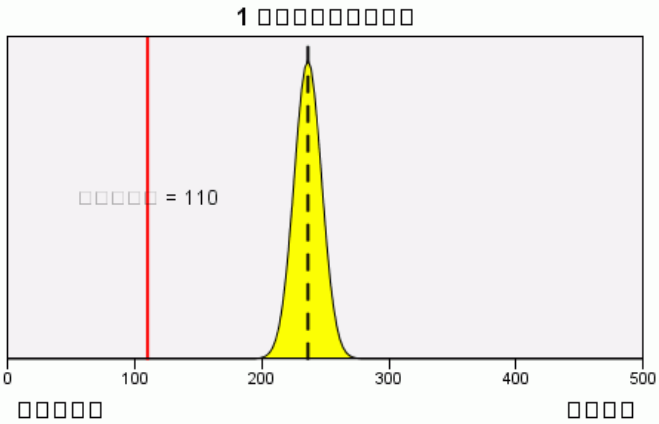
ビュー: 1 サンプル検定ビュー
検定(S): Wilcoxon の符号付き順位　フィールド(I): 入社後の在籍月数(検定 1)

[Wilcoxon の符号付き検定] 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垂直線が表示されます。
- 表には検定の詳細が表示されます。

ラン検定

図 27-29
1 サンプル検定ビュー、ラン検定



合計 N	474
検定の統計	110.000
標準誤差	10.789
標準化された検定の統計	-11.692
漸近有意確率 (両側)	.000

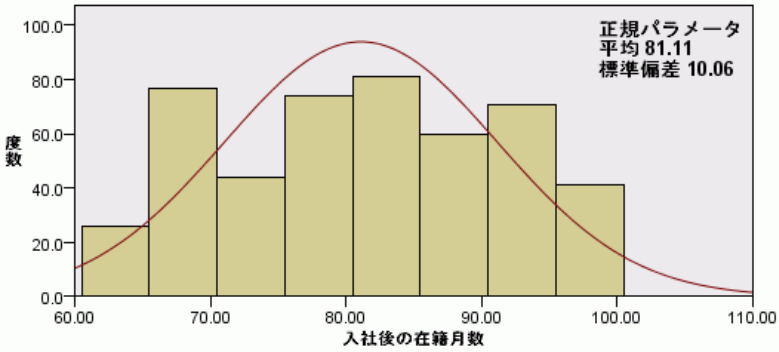
検定(S): 連　フィールド(I): 性別(検定 3)　ビュー: 1 サンプル検定ビュー

［ラン検定］ビューには、グラフと検定表が表示されます。

- グラフには、観測された数のランが垂直線でマークされた正規分布を示します。正確検定が実行された場合、検定は正規分布に基づきません。
- 表には検定の詳細が表示されます。

Kolmogorov-Smirnov 検定

図 27-30
1サンプル検定ビュー、Kolmogorov-Smirnov 検定



合計 N		474
最遠距離差	絶対値	.083
	ポジティブ (正)	.083
	ネガティブ (負)	-.082
検定の統計		1.797
漸近有意確率 (両側)		.003

ビュー: 1 サンプル検定ビュー
検定(S): Kolmogorov-Smirnov　フィールド(I): 入社後の在籍月数(検定 2)

[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。

- ヒストグラムでは、仮説の一様分布、正規分布、ポアソン分布、指数分布の確率密度関数が重ねて表示されます。検定は累積分布に基づき、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。
- 表には検定の詳細が表示されます。

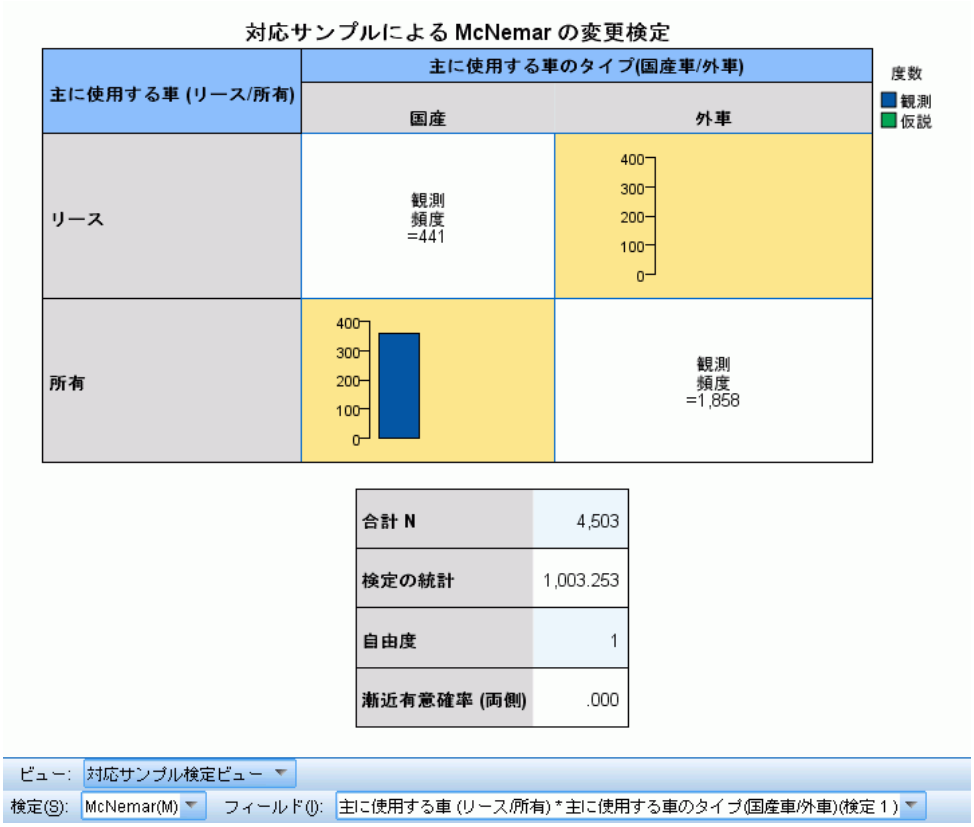
対応サンプル検定

1 サンプル検定ビューには、要求された 1 サンプル 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] ドロップダウンを使用して、指定した種類の 1 サンプル検定を選択できます。
- [フィールド] ドロップダウンを使用して、[検定] ドロップダウン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 の検定

図 27-31
対応サンプル検定ビュー、McNemar 検定

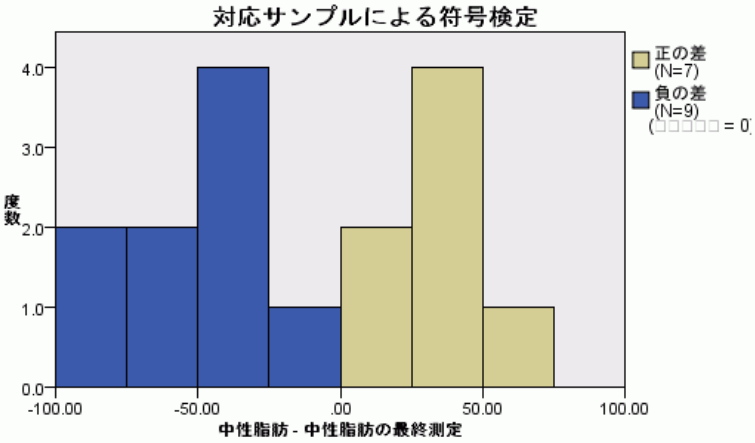


[McNemar 検定] ビューには、クラスタ棒グラフと検定表が表示されます。

- クラスタ棒グラフには、検定フィールドで定義された 2×2 の対角線外セルの観測度数および仮説度数が表示されます。
- 表には検定の詳細が表示されます。

符号検定

図 27-32
対応サンプル検定ビュー、符号検定



合計 N	16
検定の統計	7.000
標準誤差	2.000
標準化された検定の統計	-.250
漸近有意確率 (両側)	.803
正確な有意確率 (両側)	.804

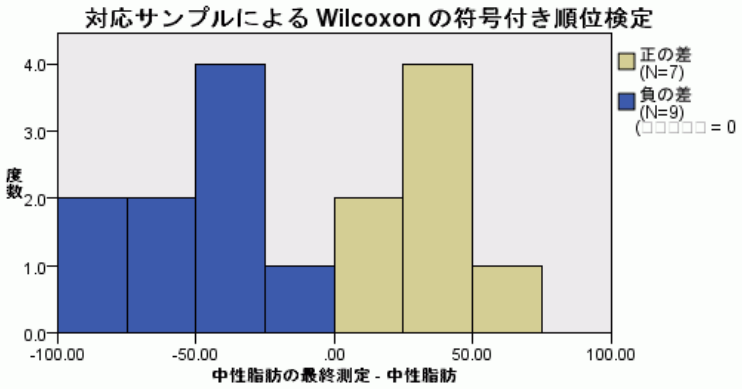
ビュー: 対応サンプル検定ビュー
検定(S): 符号　フィールド(I): 中性脂肪 - 中性脂肪の最終測定(検定 1)

［符号検定］ビューには、積み上げヒストグラムと検定表が表示されます。

- 積み上げヒストグラムには、フィールド間の差分を、差分の符号を積み上げフィールドとして使用して表示します。
- 表には検定の詳細が表示されます。

Wilcoxon の符号付き順位検定

図 27-33
対応サンプル検定ビュー、Wilcoxon の符号付き順位検定



合計 N	16
検定の統計	45.000
標準誤差	19.339
標準化された検定の統計	-1.189
漸近有意確率 (両側)	.234

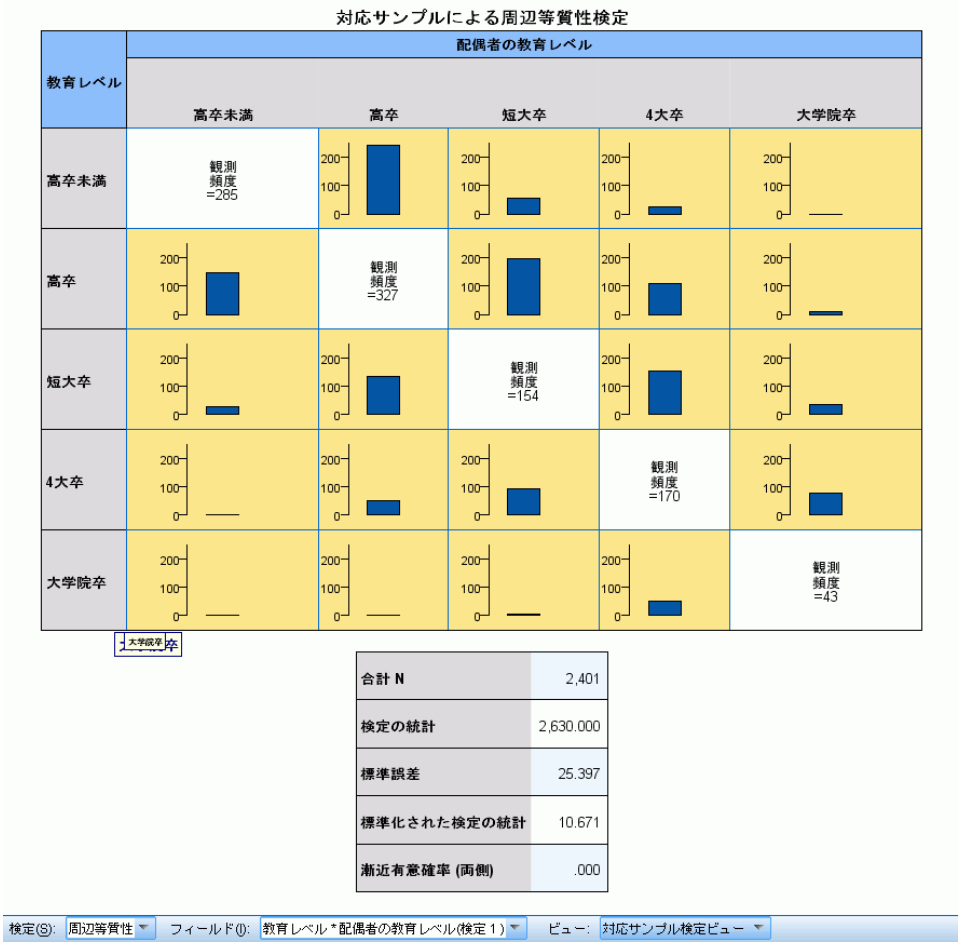
ビュー: 対応サンプル検定ビュー
検定(S): Wilcoxon の符号付き順位 フィールド(I): 中性脂肪の最終測定 - 中性脂肪(検定 2)

[Wilcoxon の符号付き検定] 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、フィールド間の差分を、差分の符号を積み上げフィールドとして使用して表示します。
- 表には検定の詳細が表示されます。

周辺等質性検定

図 27-34
対応サンプル検定ビュー、周辺等質性検定

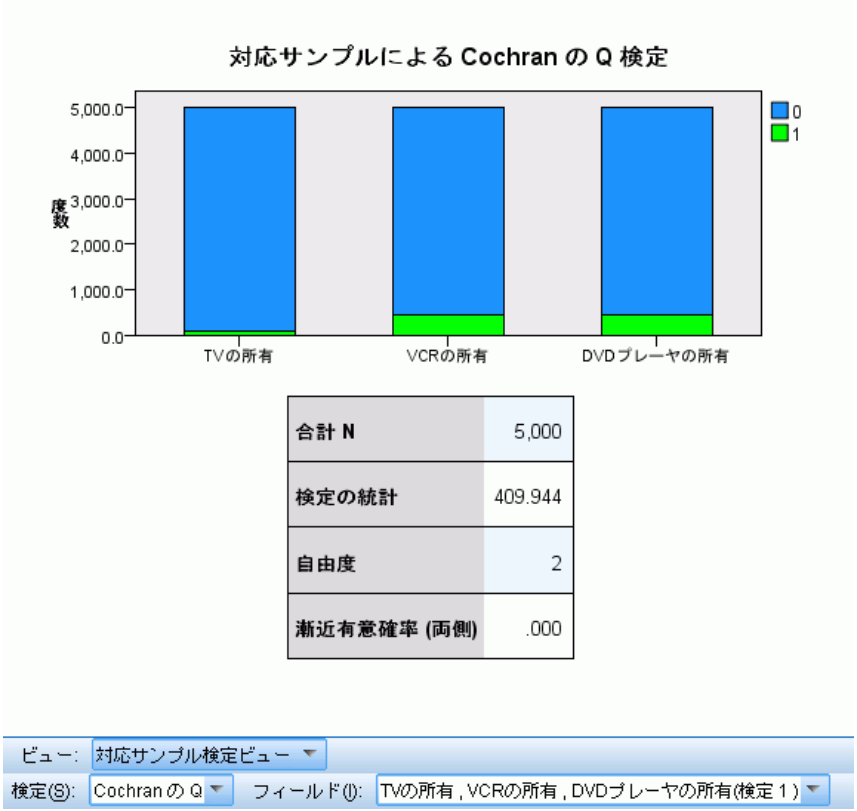


[周辺等質性検定] ビューには、クラスタ棒グラフと検定表が表示されます。

- クラスタ棒グラフには、検定フィールドで定義された 22 の対角線外セルの観測度数が表示されます。
- 表には検定の詳細が表示されます。

Cochran の Q 検定

図 27-35
対応サンプル検定ビュー、Cochran の Q 検定



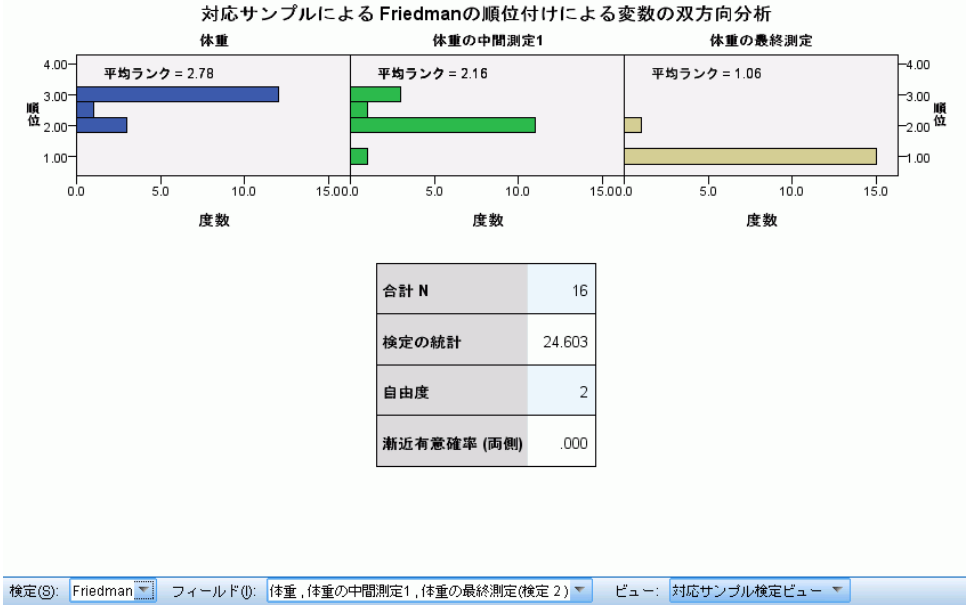
合計 N	5,000
検定の統計	409.944
自由度	2
漸近有意確率 (両側)	.000

[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。バーの上でマウスポインタを停止すると、カテゴリの割合が ToolTip に表示されます。
- 表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

図 27-36
対応サンプル検定ビュー、Friedmanの順位付けによる変数の双方向分析

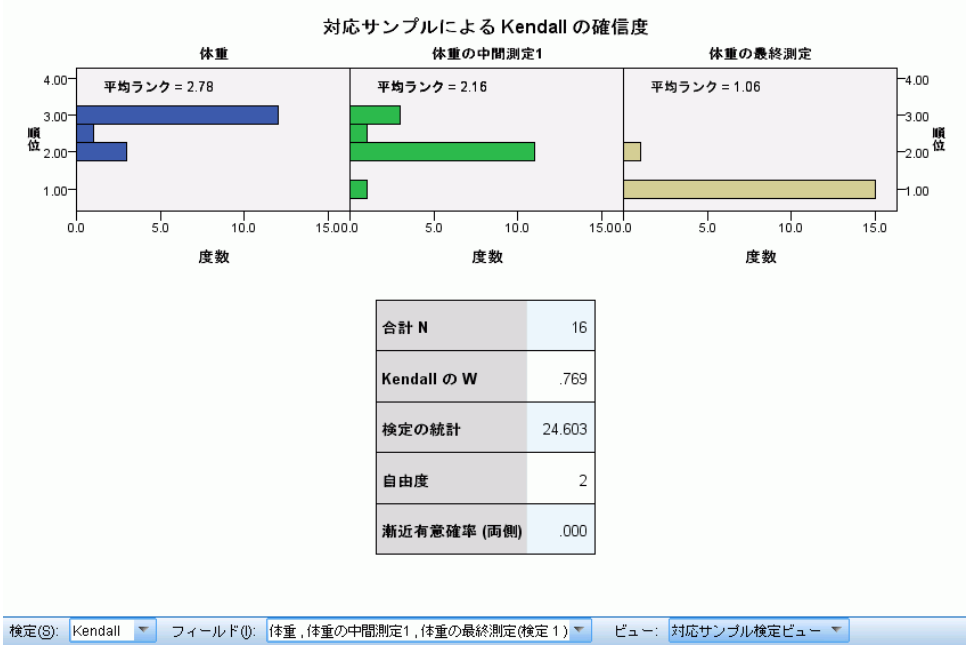


[Friedmanの順位付けによる変数の双方向分析] ビューには、パネル化されたヒストグラムと検定テーブルが表示されます。

- 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。
- 表には検定の詳細が表示されます。

Kendall の一致係数

図 27-37
対応サンプル検定ビュー、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。
- 表には検定の詳細が表示されます。

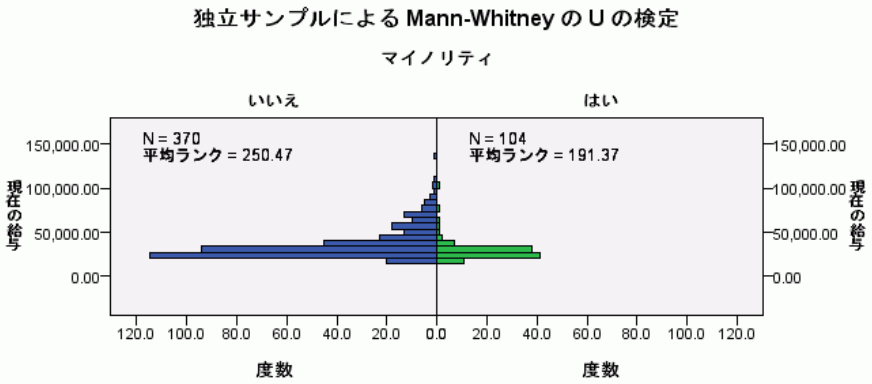
独立サンプル検定

[独立サンプル検定] ビューには、要求された独立サンプル 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**] ドロップダウンを使用して、指定した種類の独立サンプル検定を選択できます。
- [**フィールド**] ドロップダウンを使用して、[**検定**] ドロップダウンで選択した検定を使用して検定された検定とクループ フィールドの組み合わせを選択できます。

Mann-Whitney 検定

図 27-38
独立サンプル検定ビュー、Mann-Whitney 検定



合計 N	474
Mann-Whitney の U	14,442.500
Wilcoxon の W	92,672.500
検定の統計	14,442.500
標準誤差	1,234.117
標準化された検定の統計	-3.887
漸近有意確率 (両側)	.000

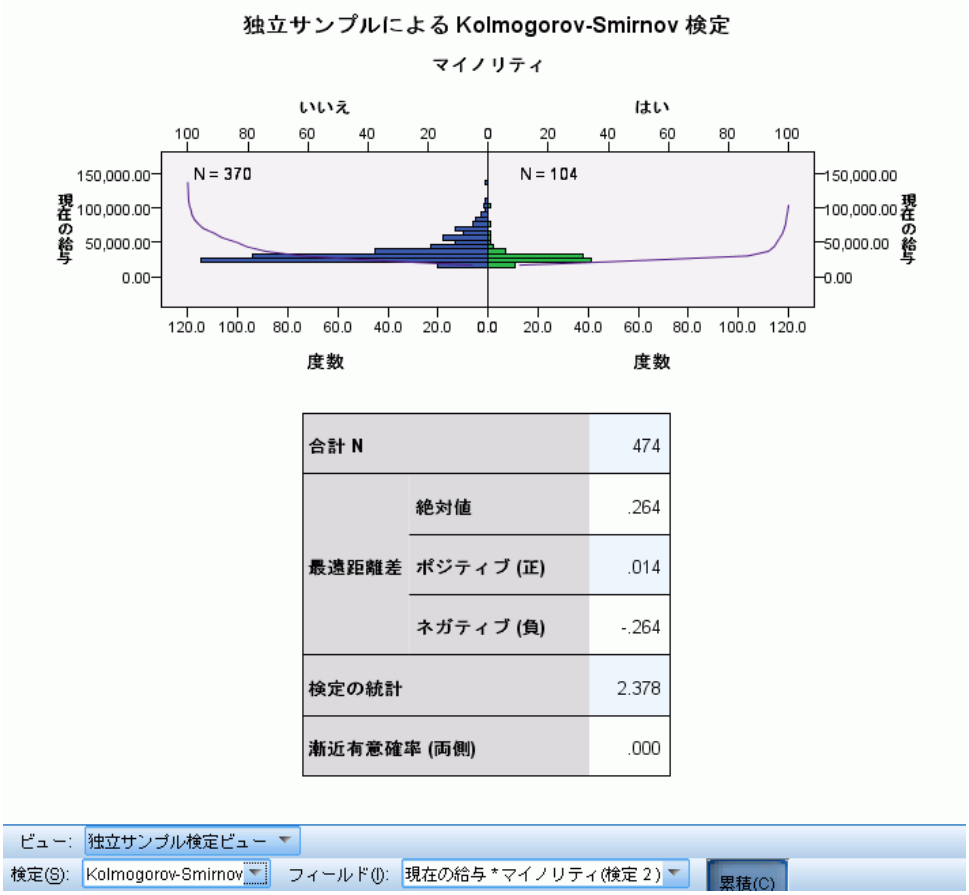
ビュー: 独立サンプル検定ビュー
検定(S): Mann-Whitney　フィールド(I): 現在の給与 * マイノリティ(検定 1)

[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バックツーバック ヒストグラムがグループフィールドのカテゴリごとに表示され、各グループの数とグループの平均順位が示されます。
- 表には検定の詳細が表示されます。

Kolmogorov-Smirnov 検定

図 27-39
独立サンプル検定ビュー、Kolmogorov-Smirnov 検定



合計 N		474
最遠距離差	絶対値	.264
	ポジティブ (正)	.014
	ネガティブ (負)	-.264
検定の統計		2.378
漸近有意確率 (両側)		.000

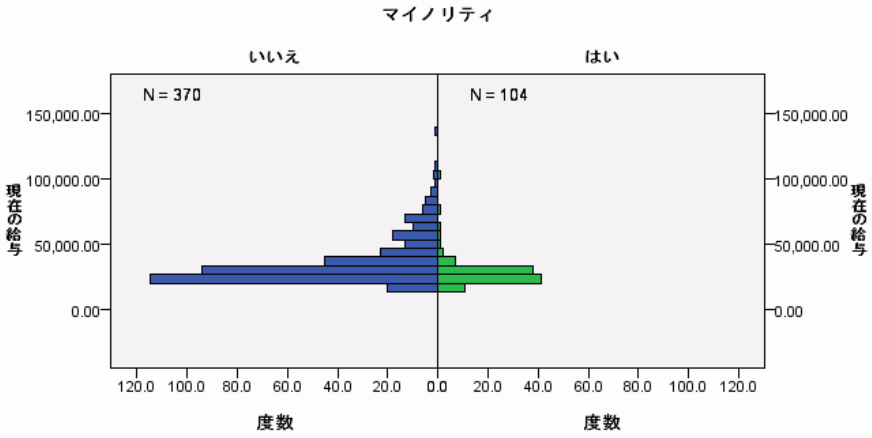
[Kolmogorov-Smirnov 検定] ビューには、人口ピラミッドグラフと検定テーブルが表示されます。

- 人口ピラミッドグラフにはバックツーバック ヒストグラムがグループ フィールドのカテゴリごとに表示され、各グループの数が示されます。観測累積分布のラインは、[累積] ボタンをクリックして表示または非表示にできます。
- 表には検定の詳細が表示されます。

Wald-Wolfowitz のラン検定

図 27-40
独立サンプル検定ビュー、Wald-Wolfowitz のラン検定

独立サンプルによる Wald-Wolfowitz のラン検定。



合計 N		474
可能最小数	検定の統計[1]	97.000
	標準誤差	7.442
	標準化された検定の統計	-8.917
	漸近有意確率 (両側)	.000
可能最大数	検定の統計[1]	199.000
	標準誤差	7.442
	標準化された検定の統計	4.788
	漸近有意確率 (両側)	1.000

[1]The test statistic is the number of runs.
1. There are 55 inter-group ties involving 228 records.

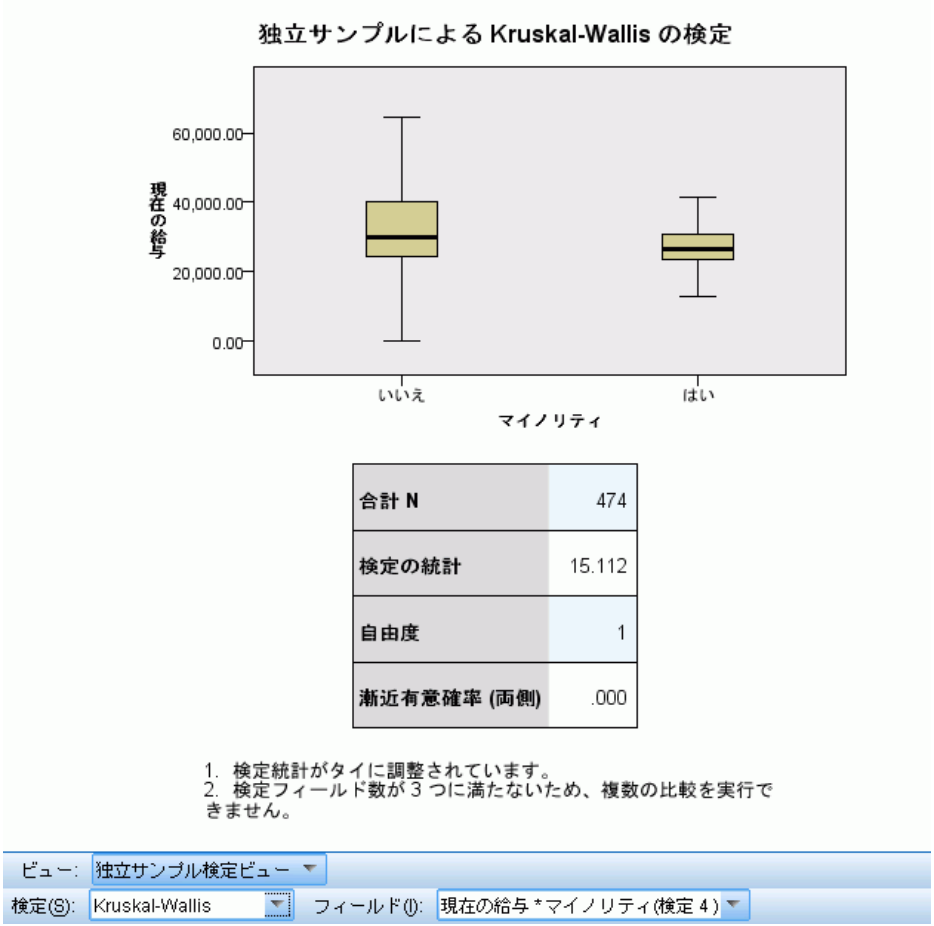
ビュー: 独立サンプル検定ビュー
検定(S): Wald-Wolfowitz　フィールド(I): 現在の給与 * マイノリティ(検定 3)

[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バックツーバック ヒストグラムがグループフィールドのカテゴリごとに表示され、各グループの数が示されます。
- 表には検定の詳細が表示されます。

Kruskal-Wallis 検定

図 27-41
独立サンプル検定ビュー、Kruskal-Wallis の検定

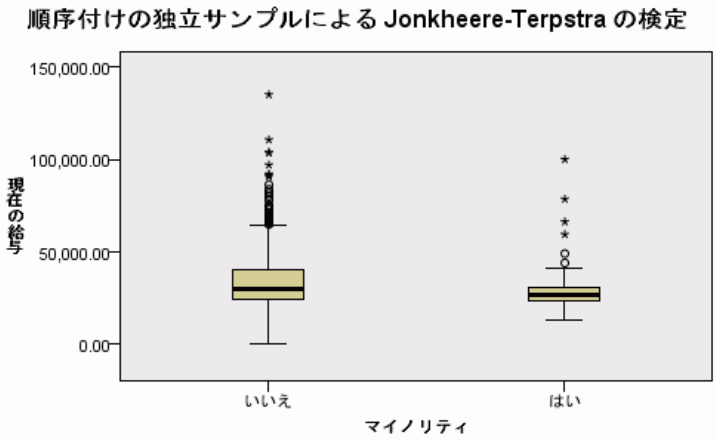


[Kruskal-Wallis の検定] ビューには、箱ひげ図と検定テーブルが表示されます。

- 各箱ひげ図は、グループ フィールドのカテゴリごとに表示されます。ボックスの上でマウスポインタを停止すると、平均順位が ToolTip に表示されます。
- 表には検定の詳細が表示されます。

Jonckheere-Terpstra の検定

図 27-42
独立サンプル検定ビュー、Jonckheere-Terpstra の検定



合計 N	474
検定の統計	14,442.500
標準誤差	1,234.117
標準化された検定の統計	-3.887
漸近有意確率 (両側)	.000

1. 検定フィールド数が 3 つに満たないため、複数の比較を実行できません。

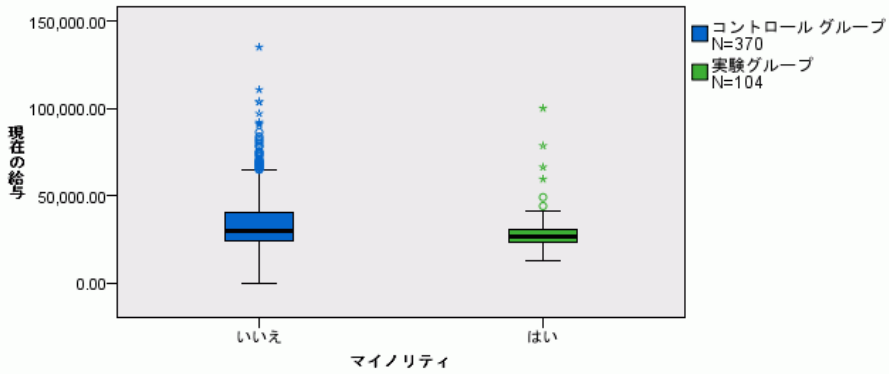
ビュー: 独立サンプル検定ビュー
検定(S): Jonckheere-Terpstra　フィールド(I): 現在の給与 * マイノリティ (検定 5)

[Jonckheere-Terpstra の検定] ビューには、グラフと検定表が表示されます。

- 各箱ひげ図は、グループ フィールドのカテゴリごとに表示されます。
- 表には検定の詳細が表示されます。

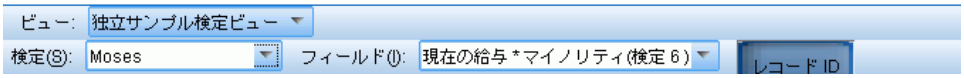
外れ値対応の Moses の検定

図 27-43
独立サンプル検定ビュー、外れ値対応の Moses の検定



合計 N		474
観測された対象グループ	検定の統計[1]	474.000
	正確な有意確率 (片側)	1.000
トリム化対象コントロール グループ	検定の統計[1]	423.000
	正確な有意確率 (片側)	.341
最後からトリム化された外れ値		19.000

[1]The test statistic is the span.

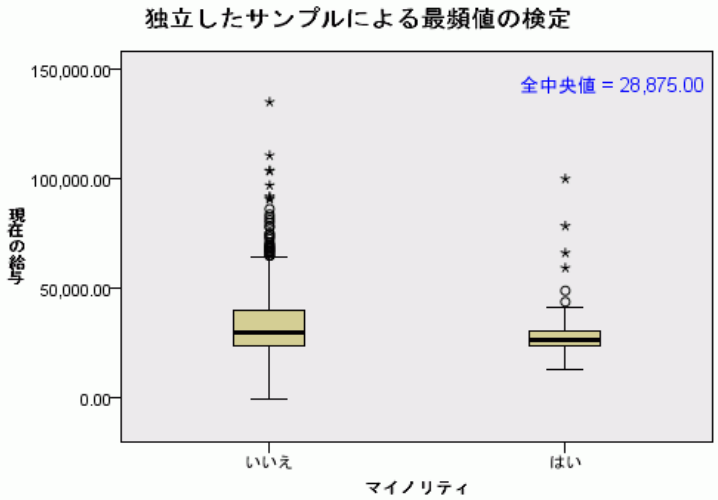


[外れ値対応の Moses の検定] ビューには、箱ひげ図と検定テーブルが表示されます。

- 各箱ひげ図は、グループ フィールドのカテゴリごとに表示されます。ポイント ラベルは、[**レコード ID**] ボタンをクリックして表示または非表示にできます。
- 表には検定の詳細が表示されます。

メディアン検定

図 27-44
独立サンプル検定ビュー、メディアン検定



合計 N		474
中央値		28,875.000
検定の統計		14.240
自由度		1
漸近有意確率 (両側)		.000
Yates の連続修正	カイ 2 乗	13.414
	自由度	1
	漸近有意確率 (両側)	.000

1. 検定フィールド数が 3 つに満たないため、複数の比較を実行できません。

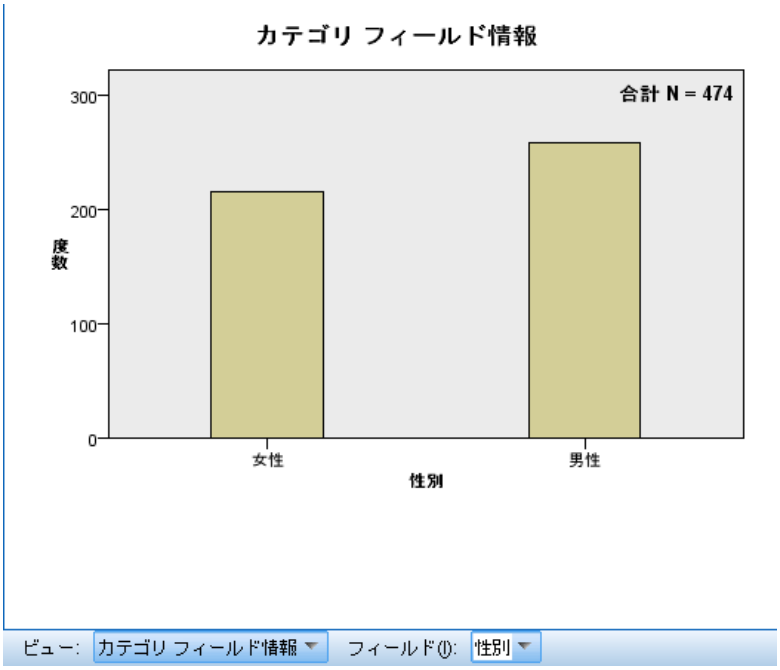
ビュー: 独立サンプル検定ビュー
検定(S): 中央値　フィールド(I): 現在の給与 * マイノリティ (検定 7)

[メディアン検定] ビューには、グラフと検定表が表示されます。

- 各箱ひげ図は、グループ フィールドのカテゴリごとに表示されます。
- 表には検定の詳細が表示されます。

カテゴリ フィールド情報

図 27-45
カテゴリ フィールド情報

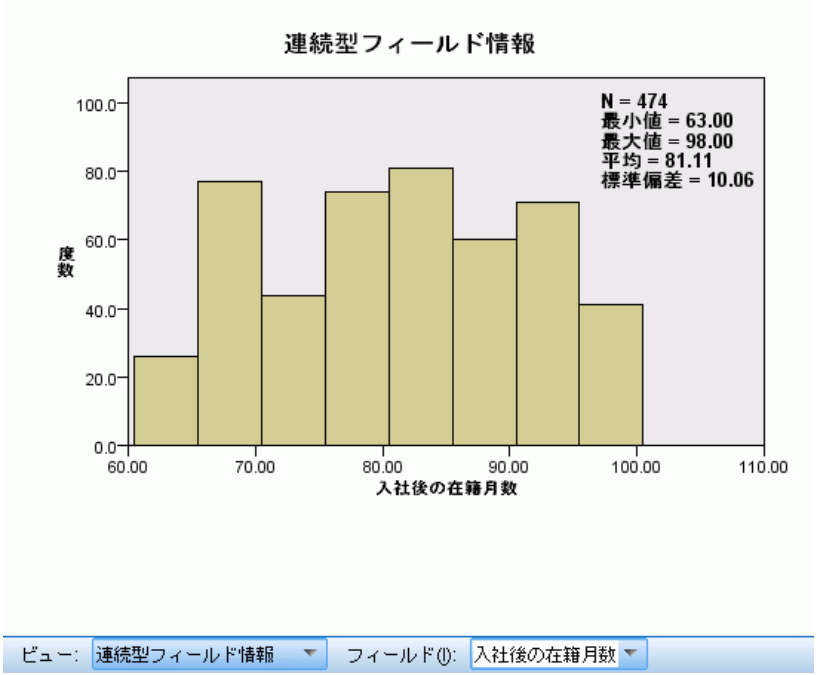


[カテゴリ フィールド情報] ビューには、[**フィールド**] ドロップダウンで選択したカテゴリ フィールドの棒グラフが表示されます。使用できるフィールドのリストは、[仮説の要約] ビューで現在選択されている検定で使用されるカテゴリ フィールドに制限されます。

- バーの上でマウスポインタを停止すると、カテゴリの割合が ToolTip に表示されます。

連続型フィールド情報

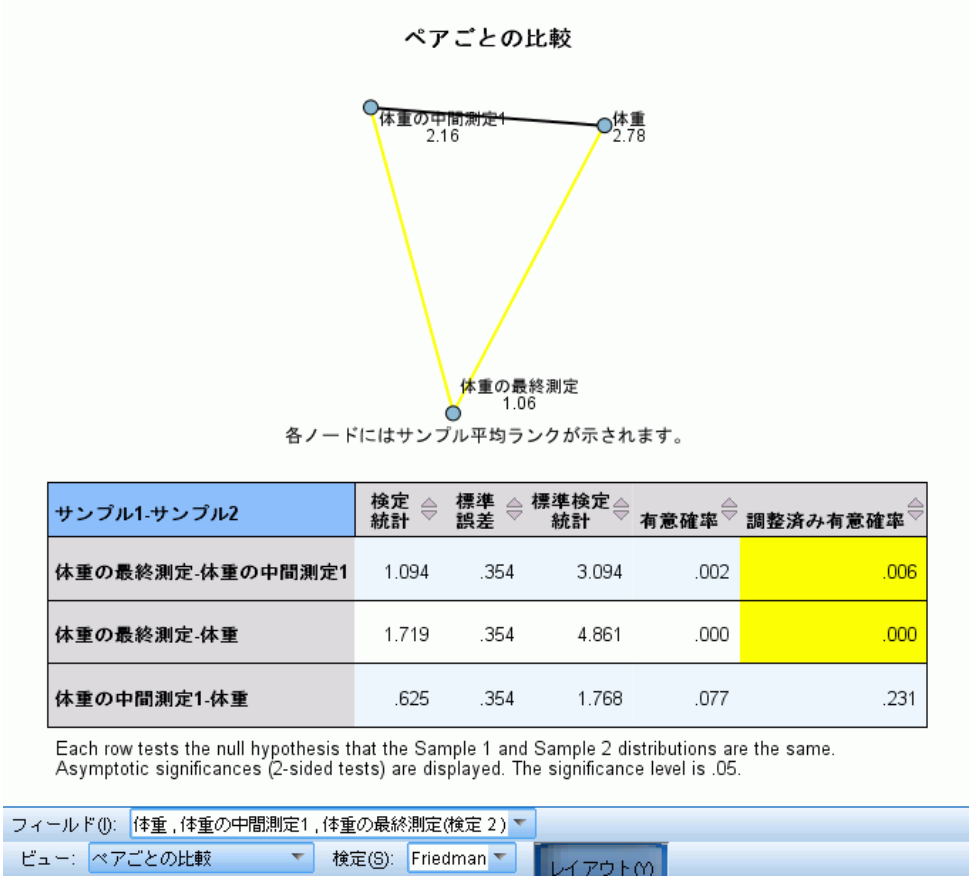
図 27-46
連続型フィールド情報



［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る連続型フィールドに制限されます。

ペアごとの比較

図 27-47
ペアごとの比較



サンプル1-サンプル2	検定統計	標準誤差	標準検定統計	有意確率	調整済み有意確率
体重の最終測定-体重の中間測定1	1.094	.354	3.094	.002	.006
体重の最終測定-体重	1.719	.354	4.861	.000	.000
体重の中間測定1-体重	.625	.354	1.768	.077	.231

[ペアごとの比較] ビューには、ペアごとの複数の比較が要求された場合、k 個のサンプルによる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 グラフと比較表が表示されます。

- 距離のネットワーク グラフは、ネットワークのノード間の距離がサンプル間の差分にタイプする比較表のグラフィカル表示です。黄色い線は統計的に重要な差分に対応し、黒い線は重要でない差分に対応します。線上にマウスポインタを置くと、Tooltip にノード間の差分についての調整済み有意度を線でつないで表示されます。
- 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、各ペアごとの比較に対応します。列の見出しをクリックすると、その列の値によって行が並べ替えられます。

等質サブセット

図 27-48
等質サブセット

等質サブセット

		サブセット		
		1	2	3
サンプル[1]	体重の最終測定	1.063		
	体重の中間測定1		2.156	
	体重			2.781
検定の統計		.[2]	.[2]	.[2]
有意確率 (2 方向)		.	.	.
調整済み有意確率 (2 方向)		.	.	.

等質サブセットは、漸近有意確率に基づきます。 有意水準は .05 です。

[1]各セルにはサンプル平均ランクが示されます。

[2]Unable to compute because the subset contains only one sample.

フィールド(I): 体重 , 体重の中間測定1 , 体重の最終測定(検定 1)
ビュー: 等質サブセット 検定(S): Kendall

[等質サブセット] ビューには、ペアごとの複数の比較が要求された場合、k 個のサンプルによる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 (フィールドごとにデータに表示) に対応します。統計的に重要で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。指定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大きく異なる場合、サンプルごとに各サブセットが存在します。統計的に大きく異なるサンプルがない場合、サブセットは 1 つになります。
- 検定統計量、有意値、調整済み有意値は、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

コマンド シンタックスを使用すると、次の作業も実行できます。

- 手順の 1 回の実行で、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します。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

レガシー ダイアログ

ノンパラメトリック検定も実行する「レガシー」ダイアログが数多くあります。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ています。

カイ 2 乗検定。変数がカテゴリに分類され、観測度数と期待度数の差に基づいて、カイ 2 乗統計量が計算されます。

2 項検定。2 分変数のカテゴリ内の観測度数と、2 項分布からの期待度数を比較します。

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 サンプルによる Kolmogorov-Smirnov 検定。変数の観測累積分布関数と、正規分布、一様分布、指数分布、またはポアソン分布のうち指定した理論分布を比較します。

2 個の独立サンプルの検定。1 つの変数で 2 つのグループに分けられたケースを比較します。使用できる検定は、Mann-Whitney の U 検定、2 サンプルの Kolmogorov-Smirnov 検定、Moses の外れ値反応の検定、および Wald-Wolfowitz のラン検定です。

2 個の対応サンプルの検定。2 個の変数の分布を比較します。使用できる検定は、Wilcoxon の符号付き順位検定、符号検定、および McNemar 検定です。

複数の独立サンプルの検定。1 つの変数で 2 つ以上のグループに分けられたケースを比較します。使用できる検定は、Kruskal-Wallis の検定、メディアン検定、および Jonckheere-Terpstra 検定です。

複数の対応サンプルの検定。2 個以上の変数の分布を比較します。使用できる検定は、Friedman の検定、Kendall の W、および Cochran の Q です。

上記すべての検定で、4 分位ならびに平均値、標準偏差、最小値、最大値、および欠損値を除いたケース数を計算できます。

カイ 2 乗検定

［カイ 2 乗検定］手続きでは、変数をカテゴリに分類してカイ 2 乗統計量を計算します。この適合度検定では、それぞれのカテゴリの中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の値が含まれているか、または各カテゴリに指定した比率の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、橙、赤、黄色のそれぞれの飴玉が同じ比率で入っているかどうかを判別できます。また、それぞれの色の飴玉が、青 5%、茶 30%、緑 10%、橙 20%、赤 15%、黄色 15% の比率で入っているかどうかを調べることができます。

統計量 平均値、標準偏差、最小値、最大値、および 4 分位。欠損値を含まないケース数と欠損ケース数、およびパーセント；各カテゴリの観測ケース数と期待ケース数；残差；およびカイ 2 乗統計量。

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（順序尺度または名義尺度）。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［値の再割り当て］を使用します。

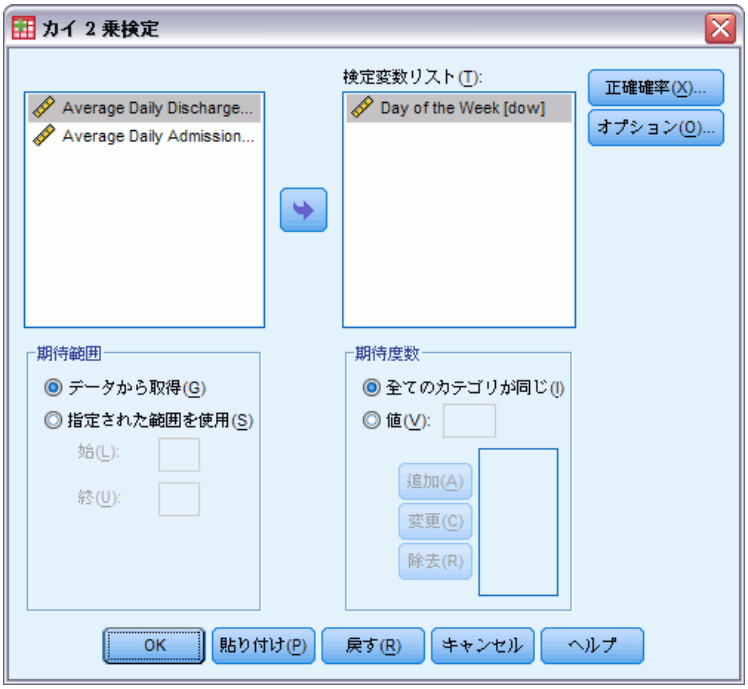
仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。データは、無作為のサンプルと仮定されます。各カテゴリの期待度数は、1 以上です。期待度数が 5 未満となるカテゴリは 20% までです。

カイ 2 乗検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > ノンパラメトリック検定 > レガシー ダイアログ > カイ 2 乗...

図 27-49
[カイ 2 乗検定] ダイアログ ボックス



- 1 つ以上の検定変数を選択します。各変数は、個別の検定を作成します。
- 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

カイ 2 乗検定の期待範囲および期待度数

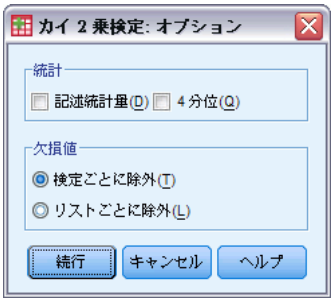
期待範囲。デフォルトでは、変数の各数値がカテゴリとして定義されます。指定された範囲内でカテゴリを定義するには、[**指定された範囲を使用**] をクリックして、[始] および [終] の各ボックスに整数を入力します。カテゴリは範囲内の各整数で定義され、範囲外の値のケースは除外されます。たとえば、[始] ボックスで 1、[終] ボックスで 4 を指定した場合、カイ 2 乗検定では、1 から 4 までの整数が使用されます。

期待値。デフォルトでは、すべてのカテゴリが同じ期待度数を持ちます。カテゴリは、ユーザー指定の期待比率を取ることができます。[**値**] をクリックし、検定変数の各カテゴリに 0 より大きい値を入力して [**追加**] をクリックします。値を追加するたびに、値リストの下端に表示されます。このボックスの値の順序は、検定変数のカテゴリ値の昇順に対応するので重要です。一覧の最初の値が検定変数のグループ値の最低値と対応し、最後の値が最高値と対応します。値の一覧の各値が合計され、値ごとにその合計値で割られ、対応するカテゴリの期待ケースの比率が計算されます。

たとえば、値の一覧が 3、4、5、4 の場合、期待比率は、3/16、4/16、5/16、および 4/16 となります。

カイ 2 乗検定のオプション

図 27-50
[カイ 2 乗検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数ごとに個別の最小値および最大値、または期待度数を指定 (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 (EXPECTED サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

2 項検定

[2 項検定] 手続きでは、2 分変数の 2 つのカテゴリの観測度数を、指定した確率パラメータの 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は 0.5 です。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、1 から最初のグループに指定した比率を引いたものです。

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、結果を記録します (表または裏)。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 でない、つまり、銀貨が歪んでいる可能性があることを示しています。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

データ。検定変数は、数値型変数および 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 [値の再割り当て] 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、使用できる値が 2 つしかない変数です。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、分割点以下の値のケースが最初のグループに、残りのケースが 2 番目のグループに割り当てられます。

仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。データは、無作為のサンプルと仮定されます。

2 項検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > ノンパラメトリック検定 > レガシー ダイアログ > 2 項...

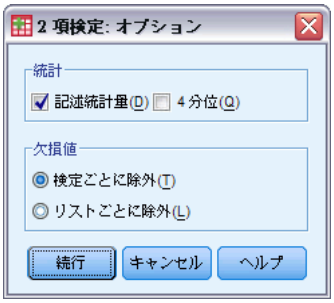
図 27-51
[2 項検定] ダイアログ ボックス

▶ 1 つ以上の数値検定変数を選択します。

▶ 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

2 項検定のオプション

図 27-52
[2 項検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**検定される変数の欠損値のあるケースは、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 (2 項検定)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数が 3 以上のカテゴリを持つ場合に、特定のグループを選択 (他の値を持つグループを除外) (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 (BINOMIAL サブコマンドを使用)。
- 1 つの変数を異なる分割点または確率で検定 (EXPECTED サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』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 つの値が現れる順序がランダムかどうかを検定します。ランは、同じ観測値の連続する長さです。ランが非常に多いかまたは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。1 つの製品を購入するかどうかを確認するために、20 人を調査したとします。20 人のすべてが同じ性別であった場合は、仮定されたサンプルの無作為性が疑問になります。ラン検定を使用して、サンプルが無作為に抽出されたかどうかを判断できます。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

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 [値の再割り当て] を使用します。

仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。連続確率分布のサンプルを使用します。

ラン検定を行うには

▶ メニューから次の項目を選択します。
分析(A) > ノンパラメトリック検定 > レガシー ダイアログ > ラン...

図 27-53
ユーザー指定の分割点を追加

▶ 1 つ以上の数値検定変数を選択します。

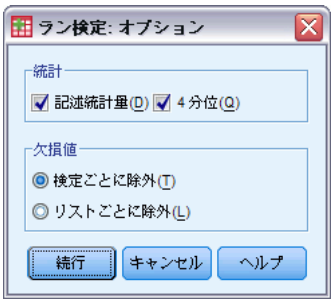
▶ 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

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。観測した平均値、中央値、最頻値、またはユーザーが指定した値のどれかを使用できます。分割点未満の値のケースが 1 つのグループに割り当てられ、分割点以上の値のケースがもう 1 つのグループに割り当てられます。指定した分割点ごとに検定が実行されます。

ラン検定のオプション

図 27-54
[ラン検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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）

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数ごとに分割点を指定（RUNS サブコマンドを使用）。
- 1 つの変数を異なる分割点で検定（RUNS サブコマンドを使用）。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

1サンプルによる Kolmogorov-Smirnov 検定

1 サンプルによる Kolmogorov-Smirnov 検定手続きは、1 つの変数で観測された累積分布関数を、正規、一様、ポアソン、または指数のいずれかの、指定した理論分布と比較します。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取られていると言えるかどうか検定します。

例:€ パラメトリック検定では、多くの場合、正規分布の変数が必要です。1 サンプルによる Kolmogorov-Smirnov 検定を使用すると、変数（収入）が正規分布しているかを検定することができます。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

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 検定では、検定分布のパラメータは事前に指定されているものと仮定しています。この手続きは、サンプルからパラメータを推定します。サンプルの平均値と標準偏差は正規分布のパラメータで、サンプルの最小値と最大値は一様な分布の範囲を定義し、サンプルの平均値はポアソン分布のパラメータです。検定の仮設分布からの逸脱の検出力は、大きく逓減する場合があります。推定のパラメータを持つ

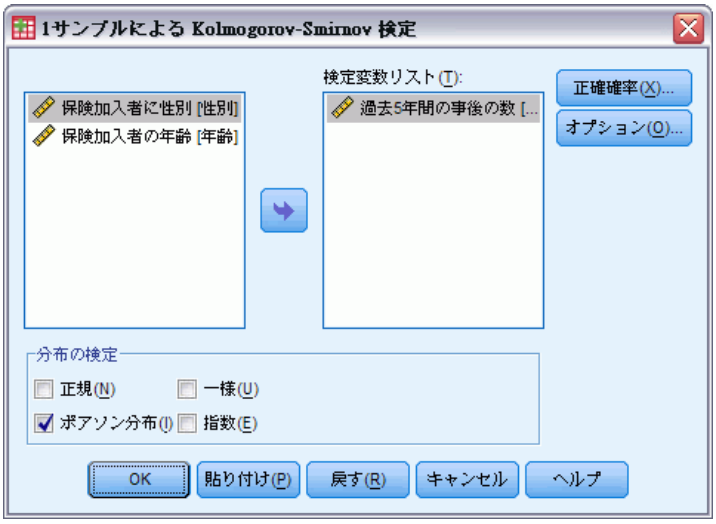
正規分布の検定については、調整済みの K-S Lilliefors の検定を検討してください([探索的分析] 手続きで使用可)。

1 サンプルによる Kolmogorov-Smirnov 検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-55
[1 サンプルによる Kolmogorov-Smirnov 検定] ダイアログ ボックス

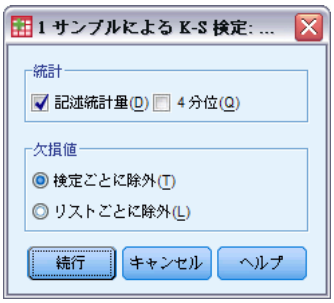


▶ 1 つ以上の数値検定変数を選択します。各変数は、個別の検定を作成します。

▶ 必要に応じて、[オプション] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

1 サンプルによる Kolmogorov-Smirnov 検定のオプション

図 27-56
[1 サンプルによる K-S 検定のオプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR 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 シンタックスを使用すると、検定分布のパラメータを指定することもできます (K-S サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』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] 手続きにより、1 つの変数に関して 2 つのグループのケースを比較できます。

例:€より快適な装着感、外観の改善、および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、古い矯正器および新しい矯正器のそれぞれを、無作為に選択した 10 名の子供たちに装着してもらいました。Mann-Whitney の U 検定により、古い矯正器を装着した子供たちに比べ、新しい矯正器を装着した子供たちの方が装着期間が短かったことがわかります。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 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。

データ。順位付けできる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 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

▶ メニューから次の項目を選択します。
分析(A) > 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...

図 27-57
[2 個の独立サンプルの検定] ダイアログ ボックス

▶ 1 つ以上の変数を選択します。

▶ 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、任意のグループ化変数を選択し、[**グループの定義**] をクリックします。

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）が同じ分布から取り出されたものであるかを検定するには、4 個の検定が使用できます。

[Mann-Whitney の U] は、[2 個の独立サンプルの検定] ダイアログボックスの中で最もよく使用されます。また、2 個のグループについては、Wilcoxon 検定および Kruskal-Wallis の H 検定と同等のものです。[Mann-Whitney の U] は、2 つのグループ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合わせて順位を付けます。観測値が同点の場合は、順位の平均値を割り当てます。同順位の数は、観測の合計数に対して相対的に小さくします。分布が位置的に同じである場合、ランクは 2 個のサンプル間で無作為に混合されます。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、またグループ 2 の得点がグループ 1 の得点より多い場合の回数が計算されます。Mann-Whitney の U 統計量は、これら 2 つの数の小さいほうです。Wilcoxon の順位和合計の 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場合、W は小さいほうの平均順位のグループの順位和で、[2 個の独立サンプル

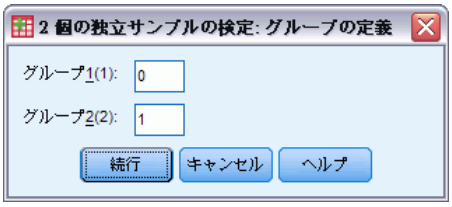
の検定: グループの定義] ダイアログ ボックスに後に名前が載っているグループの順位和になります。

分布の位置と形の両方における差を検出するには、[Kolmogorov-Smirnov の Z] および [Wald-Wolfowitz のラン] が一般的に使われます。Kolmogorov-Smirnov の正規性の検定は、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。最大絶対差が有意である場合、2 個の分布は別々のものとして認識されます。Wald-Wolfowitz のラン検定は、両方のグループの観測値を合わせてランク付けします。2 つのサンプルが同じグループから取り出されたものである場合、2 つのグループは、ランク全体を通して無作為に分散されます。

[Moses の外れ値反応] は、実験変数が何人かの被験者をある方向に、また他の被験者を反対方向に影響を及ぼすことを前提としています。そして、対照群と比較した外れ値応答数を検定します。この検定は、対照グループのスパンに注目し、また、実験群の極値が対照グループと結合したときのスパンに及ぼす影響の尺度となります。対照群は、[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックスの中のグループ 1 の値によって定義されます。両方のグループの観測が組み合わされ、ランク付けされます。対照グループのスパンは、対照グループの最大値と最小値の順位間の差 + 1 で計算します。外れ値はスパンの範囲を歪ませるので、対照ケースの 5% は、対照グループの各最後 (端部) から取り除かれます。

2 個の独立サンプルの検定: グループの定義

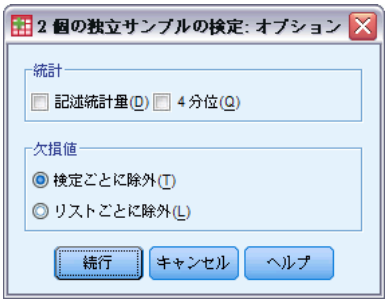
図 27-58
[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックス



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、最初の整数値をグループ 1 に、もう 1 つ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: オプション

図 27-59
[2 個の独立サンプルの検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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 (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 の外れ値反応の検定に対してトリム化するケースの数を指定する (MOSES サブコマンドを使用) こともできます。

複雑なシンタックス情報については、「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] 手続きを使用して、2 つの変数の分布を比較できます。

例。一般に、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 検定を 10 戸の家のデータに当てはめてみると、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。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 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] オプションが (Windows のオペレーティング システム上でのみ可能) インストールされているなら、周辺等質性検定も可能です。

データ。順位付けできる数値変数を使用します。

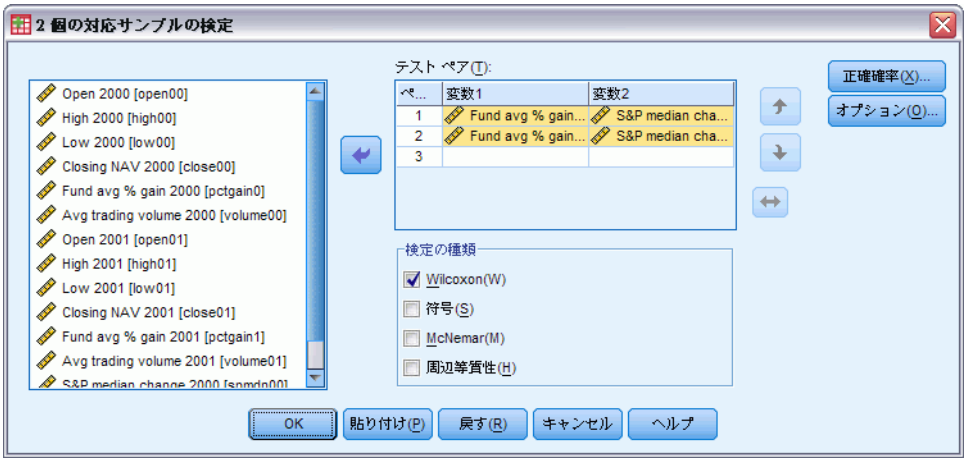
仮定。2 つの変数に対して特定の分布が仮定されていない場合でも、対応のある差の母集団の分布は対称であると仮定されます。

2 個の対応サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-60
[2 個の対応サンプルの検定] ダイアログ ボックス



▶ 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

この項での検定では、2 つの対応した変数の分布が比較されます。データの種類によって、使用する検定が異なります。

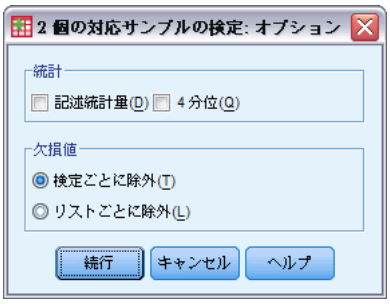
連続的なデータには、符号検定または Wilcoxon の符号付き順位検定を使用します。**符号検定**は、すべてのケースの 2 つの変数間の差を計算し、計算された差が正か負か、または同一かを分類します。2 つの変数の分布が類似している場合は、正と負の数に大きな差は見られません。**Wilcoxon の符号付き順位検定**は、ペア間の差の符号および量の両方に関する情報を考慮します。Wilcoxon の符号付き順位検定は、データに関する情報をより多く含んでいるため、符号検定より強力です。

2 進データには、**McNemar 検定**を使用します。この検定は、各被験者からあるイベントが発生する前と後の 2 回にわたって回答を得る反復測定に使用します。McNemar 検定を使用して、事前の応答比（イベントが発生する前）と事後の応答比（イベントが発生した後）が同等かどうかを決定します。この検定は、計画の前後での実験的介入によって発生する応答の変化を検出する場合に役立ちます。

カテゴリのデータには、**周辺等質性検定**を使用します。この検定は、McNemar の検定を 2 値反応から多値反応に拡張したものです。応答の変化を（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応答の変化を検出する場合に役立ちます。周辺等質性検定は、Exact Tests がインストールされている場合に限り使用できます。

2 個の対応サンプルの検定: オプション

図 27-61
[2 個の対応サンプルの検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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 (2 個の対応サンプル)

コマンド シンタックス言語を使用すると、指定した変数リストのすべての変数の組み合せについて検定することもできます。

シンタックスの詳細は、『Command Syntax Reference』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］手続きでは、1 つの変数で複数のケースのグループを比較します。

例:€ 3 種類の 100 ワット電球が燃え尽きるまでの平均時間に違いはあるのでしょうか。Kruskal-Wallis 一元配置分散分析から、3 種類の電球は平均寿命が違うことを理解できます。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Kruskal-Wallis の H、メディアン検定。

データ。順位付けできる数値変数を使用します。

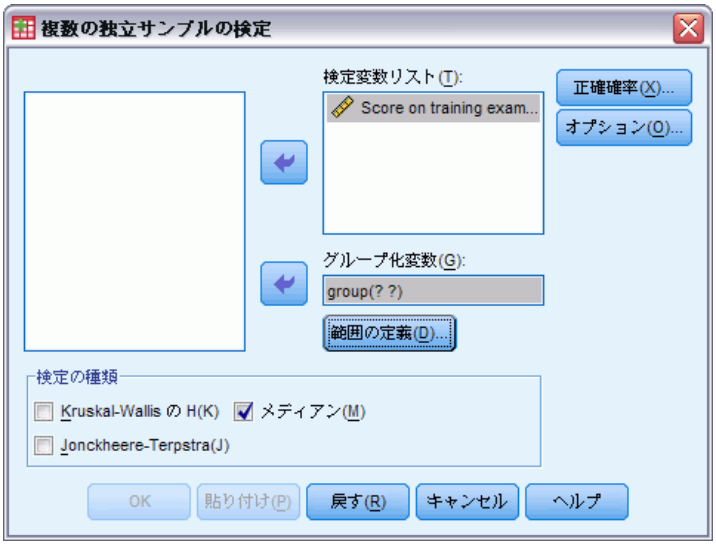
仮定。無作為に抽出された独立サンプルを使用します。Kruskal-Wallis の H 検定では、サンプルの分布の形が近似していることが検定されている必要があります。

複数の独立サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-62
メディアン検定の定義



▶ 1 つ以上の変数を選択します。

▶ 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と最大の整数値を指定します。

複数の独立サンプルの検定の検定の種類

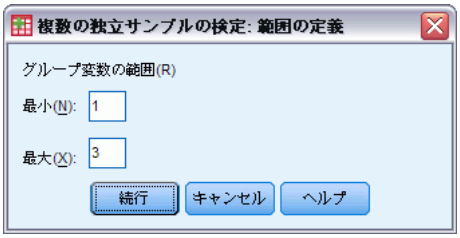
複数の独立サンプルが同じ母集団から発生しているかどうかを判断するには、3 つの検定を利用できます。Kruskal-Wallis の H 検定、メディアン検定、および Jonckheere-Terpstra の検定は、すべて複数の独立標本が同じ母集団から発生しているかどうかを検定します。

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あり、分布の位置の差を検出します。**メディアン検定**はこれよりも一般的な検定ですが、それほど強力ではなく、分布の位置と形の違いを検出します。Kruskal-Wallis の H 検定およびメディアン検定は、標本が抽出される k 個の母集団の 事前の順位付けがないものと仮定しています。

k 個の母集団の自然な事前の順位付け（昇順または降順）があるときには、**Jonckheere-Terpstra 検定**の方が効果的です。たとえば、k 個の母集団が k 段階の上昇温度を表す場合があるとします。異なる温度でも同じ応答分布を示すという仮説は、温度が上昇するにつれて応答の大きさが上昇するという対立仮説に対して検定されます。ここで、対立仮説が順序付けされるので、Jonckheere-Terpstra が最適の検定になります。Jonckheere-Terpstra 検定は、Exact Tests アドオン モジュールをインストールしている場合にだけ利用できます。

複数の独立サンプルの検定の範囲定義

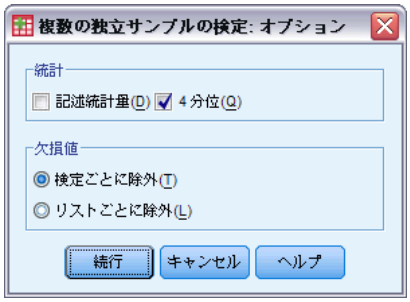
図 27-63
複数の独立サンプルの検定: 範囲の定義ダイアログ ボックス



範囲を定義するには、グループ化変数の最小と最大カテゴリに対応する **最小値**と **最大値**として整数値を入力します。上下限を超える値を含むケースは除外されます。たとえば、1 の最小値と 3 の最大値を指定すると、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なければなりません。また、必ず両方の値を指定しなければなりません。

複数の独立サンプルの検定のオプション

図 27-64
[複数の独立サンプルの検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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 (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、メディアン検定に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)。

シンタックスの詳細は、『Command Syntax Reference』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] 手続きでは、複数の変数の分布を比較します。

例:€ 世間の人が、医者、法律家、警察官、教師という職業に対して抱く権威の大きさに差はあるでしょうか?10 人に依頼して、権威の大きさについてこの 4 種類の職業を順位付けしてもらいました。Friedman の検定では、世間の人がこの 4 つの職業に対して抱く権威の大きさに差があるかどうかを示します。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Friedman、Kendall の W、および Cochran の Q。

データ。順位付けできる数値変数を使用します。

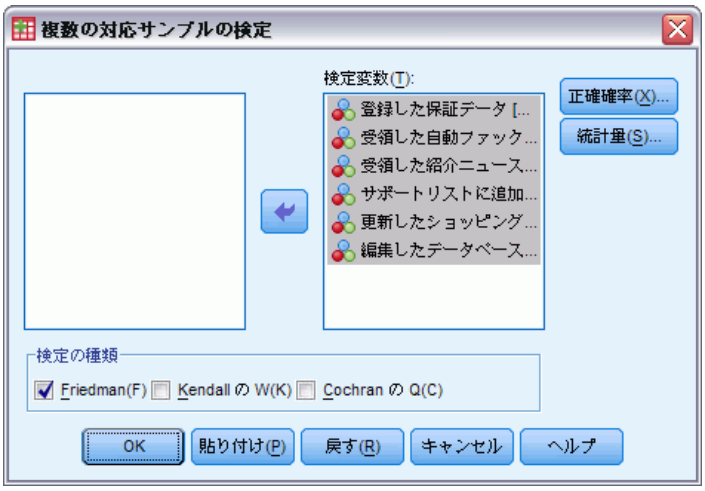
仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。無作為に抽出された対応のあるサンプルを使用します。

複数の対応サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-65
Cochran を検定の種類として選択



▶ 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

複数の対応する変数の分布を比較するには、3 つの検定が利用できます。

Friedman の検定は、1 サンプルの反復測定の分散分析やセルごとに 1 つの観測を持つ 2 元配置分散分析に対応するノンパラメトリック検定です。Friedman の検定は、k 個の対応変数は同じ母集団から発生するという帰無仮説です。各ケースに対して、k 個の変数は 1 から k に順位付けされます。検定の統計量は、それらの順位に基づきます。

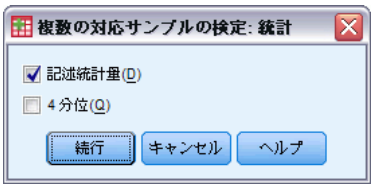
Kendall の W は Friedman の統計の規準です。Kendall の W は、評価の一致度の尺度となっている一致係数として解釈することができます。各ケースは評価者で、各変数は評価されている品目または人です。各変数に対して、順位の合計が計算されます。Kendall の W は、0（まったく一致していない）と 1（完全な一致）の間の値を取ります。

Cochran の Q は、Friedman の検定とまったく同じですが、すべての応答が 2 値データのときに適用します。この検定は、McNemar 検定を k 個のサンプルの場合に拡張したものです。Cochran の Q は、複数の対応のある

2 分変数には同じ平均値があるという仮説を検定します。変数は同一の個体または対応のあ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

図 27-66
[複数の対応サンプルの検定: 統計] ダイアログ ボックス



統計を選択します。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。

NPAR 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は、『Command Syntax Reference』を参照してください。

2 項検定

[2 項検定] 手続きでは、2 分変数の 2 つのカテゴリの観測度数を、指定した確率パラメータの 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は 0.5 です。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、1 から最初のグループに指定した比率を引いたものです。

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、結果を記録します (表または裏)。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 でない、つまり、銀貨が歪んでいる可能性があることを示しています。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

データ。検定変数は、数値型変数および 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 [値の再割り当て] 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、使用できる値が 2 つしかない変数です。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、分割点以下の値のケースが最初のグループに、残りのケースが 2 番目のグループに割り当てられます。

仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。データは、無作為のサンプルと仮定されます。

2 項検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > ノンパラメトリック検定 > レガシー ダイアログ > 2 項...

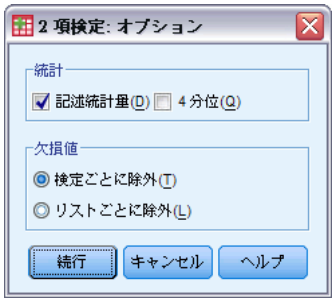
図 27-67
[2 項検定] ダイアログ ボックス

▶ 1 つ以上の数値検定変数を選択します。

▶ 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

2 項検定のオプション

図 27-68
[2 項検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**検定される変数の欠損値のあるケースは、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 (2 項検定)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数が 3 以上のカテゴリを持つ場合に、特定のグループを選択 (他の値を持つグループを除外) (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 (BINOMIAL サブコマンドを使用)。
- 1 つの変数を異なる分割点または確率で検定 (EXPECTED サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』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 つの値が現れる順序がランダムかどうかを検定します。ランは、同じ観測値の連続する長さです。ランが非常に多いかまたは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。 1 つの製品を購入するかどうかを確認するために、20 人を調査したとします。20 人のすべてが同じ性別であった場合は、仮定されたサンプルの無作為性が疑問になります。ラン検定を使用して、サンプルが無作為に抽出されたかどうかを判断できます。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

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［値の再割り当て］を使用します。

仮定。 ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。連続確率分布のサンプルを使用します。

ラン検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > ノンパラメトリック検定 > レガシー ダイアログ > ラン...

図 27-69
ユーザー指定の分割点を追加

▶ 1 つ以上の数値検定変数を選択します。

▶ 必要に応じて、**[オプション]** をクリックし、［統計量］で［記述統計量］および［4 分位］、［欠損値］で欠損値データの扱い方を指定できます。

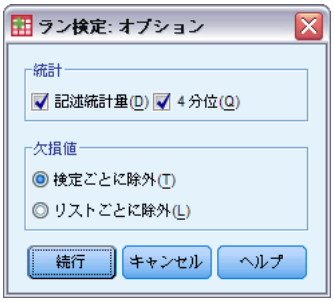
ラン検定の分割点

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。観測した平均値、中央値、最頻値、またはユーザーが指定した値のどれかを使用できます。分割点未満の値のケースが 1 つのグループに割り当てられ、分割点

以上の値のケースがもう 1 つのグループに割り当てられます。指定した分割点ごとに検定が実行されます。

ラン検定のオプション

図 27-70
[ラン検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数ごとに分割点を指定 (RUNS サブコマンドを使用)。
- 1 つの変数を異なる分割点で検定 (RUNS サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

1サンプルによる Kolmogorov-Smirnov 検定

1 サンプルによる Kolmogorov-Smirnov 検定手続きは、1 つの変数で観測された累積分布関数を、正規、一様、ポアソン、または指数のいずれかの、指定した理論分布と比較します。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取られていると言えるかどうか検定します。

例:€ パラメトリック検定では、多くの場合、正規分布の変数が必要です。1 サンプルによる Kolmogorov-Smirnov 検定を使用すると、変数（収入）が正規分布しているかを検定することができます。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

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）。

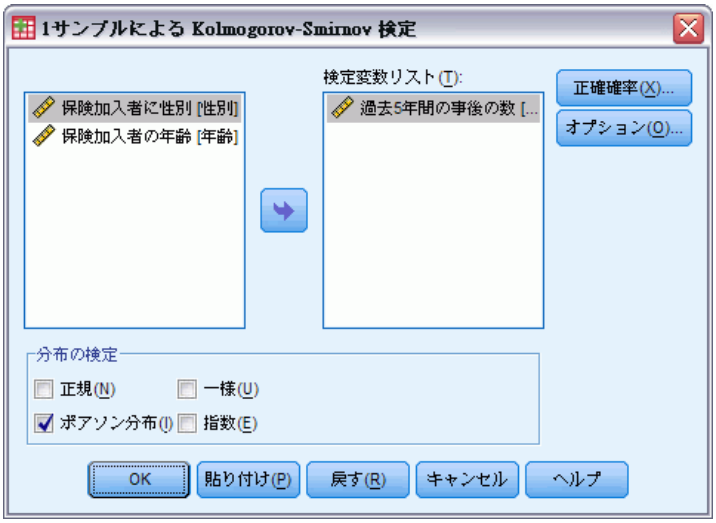
仮定。Kolmogorov-Smirnov 検定では、検定分布のパラメータは事前に指定されているものと仮定しています。この手続きは、サンプルからパラメータを推定します。サンプルの平均値と標準偏差は正規分布のパラメータで、サンプルの最小値と最大値は一様な分布の範囲を定義し、サンプルの平均値はポアソン分布のパラメータです。検定の仮設分布からの逸脱の検出力は、大きく逓減する場合があります。推定のパラメータを持つ正規分布の検定については、調整済みの K-S Lilliefors の検定を検討してください（[探索的分析] 手続きで使用可）。

1 サンプルによる Kolmogorov-Smirnov 検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-71
[1 サンプルによる Kolmogorov-Smirnov 検定] ダイアログ ボックス

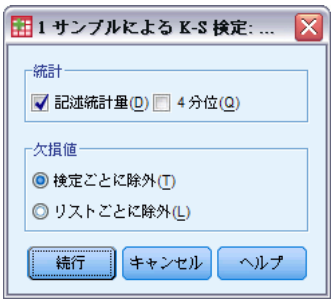


▶ 1 つ以上の数値検定変数を選択します。各変数は、個別の検定を作成します。

▶ 必要に応じて、[オプション] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

1 サンプルによる Kolmogorov-Smirnov 検定のオプション

図 27-72
[1 サンプルによる K-S 検定のオプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR 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 シンタックスを使用すると、検定分布のパラメータを指定することもできます (K-S サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』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] 手続きにより、1 つの変数に関して 2 つのグループのケースを比較できます。

例:€より快適な装着感、外観の改善、および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、古い矯正器および新しい矯正器のそれぞれを、無作為に選択した 10 名の子供たちに装

着してもらいました。Mann-Whitney の U 検定により、古い矯正器を装着した子供たちに比べ、新しい矯正器を装着した子供たちの方が装着期間が短かったことがわかります。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 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。

データ。順位付けできる数値変数を使用します。

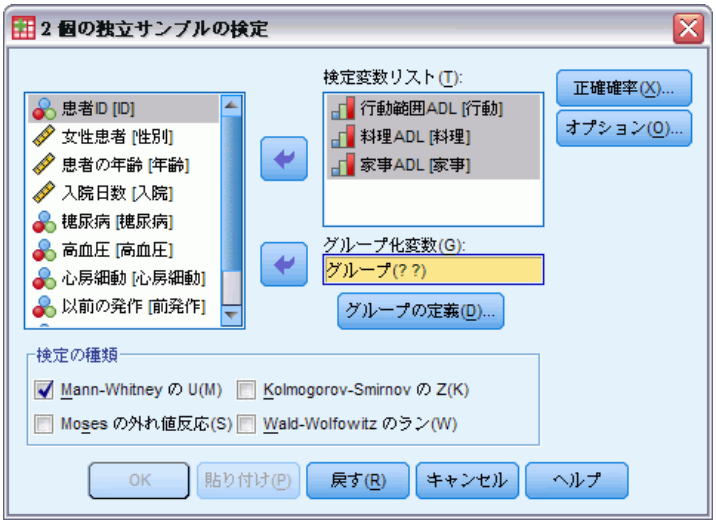
仮定。無作為に抽出された独立サンプルを使用します。Mann-Whitney の U 検定は、2 つの分布の同等性を検定します。この検定を使用して、2 つの分布間の位置の相違を検定するには、それらの分布が同じ形状であると仮定する必要があります。

2 個の独立サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-73
[2 個の独立サンプルの検定] ダイアログ ボックス



▶ 1 つ以上の変数を選択します。

▶ 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、任意のグループ化変数を選択し、[**グループの定義**] をクリックします。

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）が同じ分布から取り出されたものであるかを検定するには、4 個の検定が使用できます。

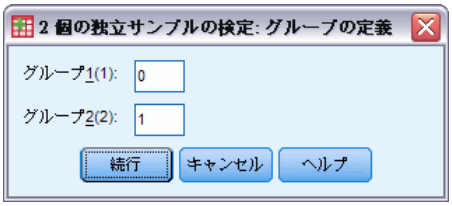
[Mann-Whitney の U] は、[2 個の独立サンプルの検定] ダイアログ ボックスの中で最もよく使用されます。また、2 個のグループについては、Wilcoxon 検定および Kruskal-Wallis の H 検定と同等のものです。[Mann-Whitney の U] は、2 つのグループのサンプルが位置的に同じかどうかを検定します。両方のグループからの観測値を合わせて順位を付けます。観測値が同点の場合は、順位の平均値を割り当てます。同順位の数は、観測の合計数に対して相対的に小さくします。分布が位置的に同じである場合、ランクは 2 個のサンプル間で無作為に混合されます。グループ 1 の得点がグループ 2 の得点より多い場合の回数、またグループ 2 の得点がグループ 1 の得点より多い場合の回数が計算されます。Mann-Whitney の U 統計量は、これら 2 つの数の小さいほうです。Wilcoxon の順位和合計の 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場合、W は小さいほうの平均順位のグループの順位和で、[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックスに後に名前が載っているグループの順位和になります。

分布の位置と形の両方における差を検出するには、[Kolmogorov-Smirnov の Z] および [Wald-Wolfowitz のラン] が一般的に使われます。Kolmogorov-Smirnov の正規性の検定は、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。最大絶対差が有意である場合、2 個の分布は別々のものとして認識されます。Wald-Wolfowitz のラン検定は、両方のグループの観測値を合わせてランク付けします。2 つのサンプルが同じグループから取り出されたものである場合、2 つのグループは、ランク全体を通して無作為に分散されます。

[Moses の外れ値反応] は、実験変数が何人かの被験者をある方向に、また他の被験者を反対方向に影響を及ぼすことを前提としています。そして、対照群と比較した外れ値応答数を検定します。この検定は、対照グループのスパンに注目し、また、実験群の極値が対照グループと結合したときのスパンに及ぼす影響の尺度となります。対照群は、[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックスの中のグループ 1 の値によって定義されます。両方のグループの観測が組み合わされ、ランク付けされます。対照グループのスパンは、対照グループの最大値と最小値の順位間の差 + 1 で計算します。外れ値はスパンの範囲を歪ませるので、対照ケースの 5% は、対照グループの各最後 (端部) から取り除かれます。

2 個の独立サンプルの検定: グループの定義

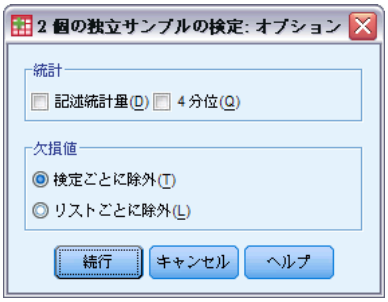
図 27-74
[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックス



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、最初の整数値をグループ 1 に、もう 1 つ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: オプション

図 27-75
[2 個の独立サンプルの検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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 (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 の外れ値反応の検定に対してトリム化するケースの数を指定する (MOSES サブコマンドを使用) こともできます。

複雑なシンタックス情報については、「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] 手続きを使用して、2 つの変数の分布を比較できます。

例。一般に、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 検定を 10 戸の家のデータに当てはめてみると、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。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 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] オプションが (Windows のオペレーティング システム上でのみ可能) インストールされているなら、周辺等質性検定も可能です。

データ。順位付けできる数値変数を使用します。

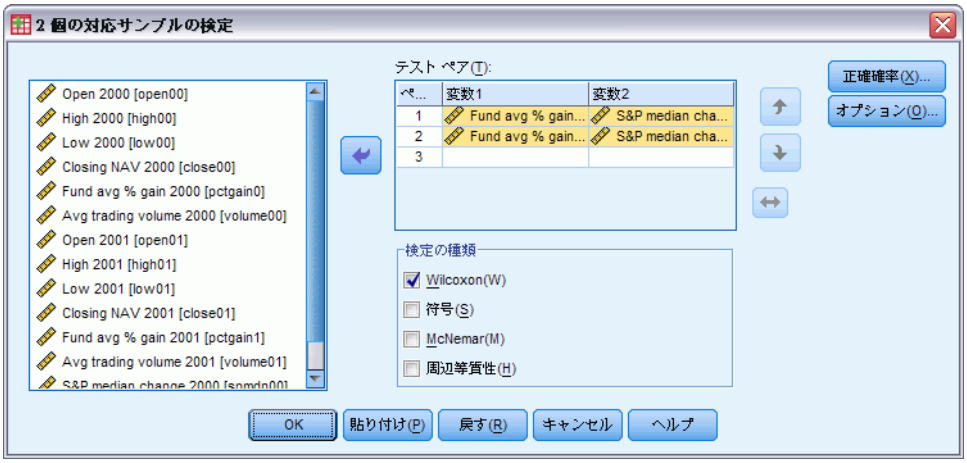
仮定。2 つの変数に対して特定の分布が仮定されていない場合でも、対応のある差の母集団の分布は対称であると仮定されます。

2 個の対応サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-76
[2 個の対応サンプルの検定] ダイアログ ボックス



▶ 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

この項での検定では、2 つの対応した変数の分布が比較されます。データの種類によって、使用する検定が異なります。

連続的なデータには、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、すべてのケースの 2 つの変数間の差を計算し、計算された差が正か負か、または同一かを分類します。2 つの変数の分布が類似している場合は、正と負の数に大きな差は見られません。Wilcoxon の

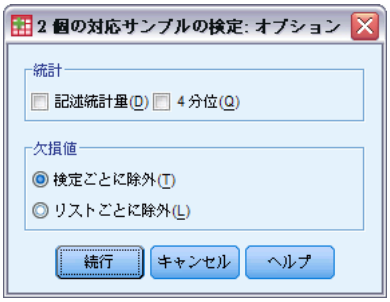
符号付き順位検定は、ペア間の差の符号および量の両方に関する情報を考慮します。Wilcoxon の符号付き順位検定は、データに関する情報をより多く含んでいるため、符号検定より強力です。

2 進データには、**McNemar 検定**を使用します。この検定は、各被験者からあるイベントが発生する前と後の 2 回にわたって回答を得る反復測定に使用します。McNemar 検定を使用して、事前の応答比（イベントが発生する前）と事後の応答比（イベントが発生した後）が同等かどうかを決定します。この検定は、計画の前後での実験的介入によって発生する応答の変化を検出する場合に役立ちます。

カテゴリのデータには、**周辺等質性検定**を使用します。この検定は、McNemar の検定を 2 値反応から多値反応に拡張したものです。応答の変化を（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応答の変化を検出する場合に役立ちます。周辺等質性検定は、Exact Tests がインストールされている場合に限り使用できます。

2 個の対応サンプルの検定: オプション

図 27-77
[2 個の対応サンプルの検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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 (2 個の対応サンプル)

コマンド シンタックス言語を使用すると、指定した変数リストのすべての変数の組み合せについて検定することもできます。

シンタックスの詳細は、『Command Syntax Reference』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] 手続きでは、1 つの変数で複数のケースのグループを比較します。

例:€ 3 種類の 100 ワット電球が燃え尽きるまでの平均時間に違いはあるのでしょうか。Kruskal-Wallis 一元配置分散分析から、3 種類の電球は平均寿命が違うことを理解できます。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Kruskal-Wallis の H、メディアン検定。

データ。順位付けできる数値変数を使用します。

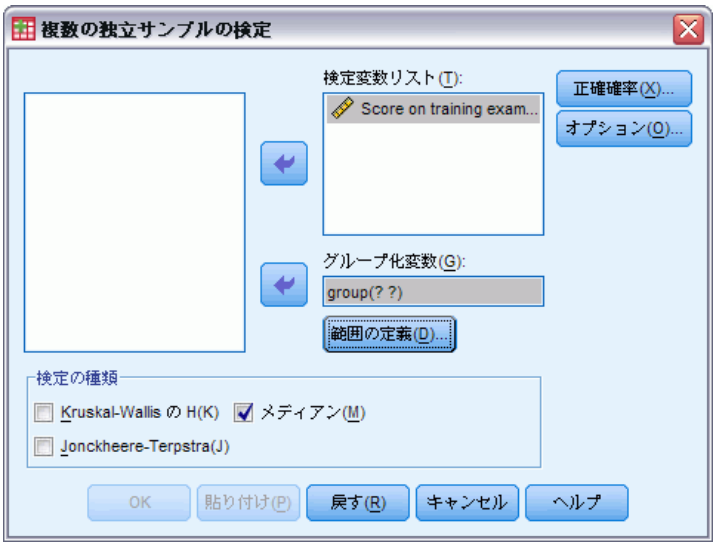
仮定。無作為に抽出された独立サンプルを使用します。Kruskal-Wallis の H 検定では、サンプルの分布の形が近似していることが検定されている必要があります。

複数の独立サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-78
メディアン検定の定義



▶ 1 つ以上の変数を選択します。

▶ 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と最大の整数値を指定します。

複数の独立サンプルの検定の検定の種類

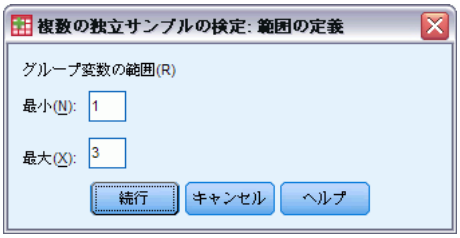
複数の独立サンプルが同じ母集団から発生しているかどうかを判断するには、3 つの検定を利用できます。Kruskal-Wallis の H 検定、メディアン検定、および Jonckheere-Terpstra の検定は、すべて複数の独立標本が同じ母集団から発生しているかどうかを検定します。

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あり、分布の位置の差を検出します。**メディアン検定**はこれよりも一般的な検定ですが、それほど強力ではなく、分布の位置と形の違いを検出します。Kruskal-Wallis の H 検定およびメディアン検定は、標本が抽出される k 個の母集団の 事前の順位付けがないものと仮定しています。

k 個の母集団の自然な事前の順位付け（昇順または降順）があるときには、**Jonckheere-Terpstra 検定**の方が効果的です。たとえば、k 個の母集団が k 段階の上昇温度を表す場合があるとします。異なる温度でも同じ応答分布を示すという仮説は、温度が上昇するにつれて応答の大きさが上昇するという対立仮説に対して検定されます。ここで、対立仮説が順序付けされるので、Jonckheere-Terpstra が最適の検定になります。Jonckheere-Terpstra 検定は、Exact Tests アドオン モジュールをインストールしている場合にだけ利用できます。

複数の独立サンプルの検定の範囲定義

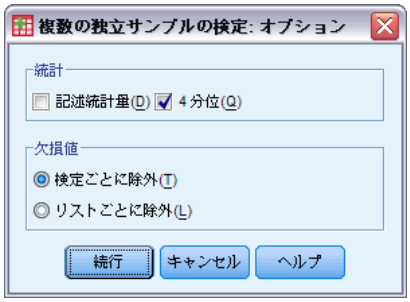
図 27-79
複数の独立サンプルの検定: 範囲の定義ダイアログ ボックス



範囲を定義するには、グループ化変数の最小と最大カテゴリに対応する **最小値**と **最大値**として整数値を入力します。上下限を超える値を含むケースは除外されます。たとえば、1 の最小値と 3 の最大値を指定すると、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なければなりません。また、必ず両方の値を指定しなければなりません。

複数の独立サンプルの検定のオプション

図 27-80
[複数の独立サンプルの検定: オプション] ダイアログ ボックス



統計。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、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 (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、メディアン検定に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)。

シンタックスの詳細は、『Command Syntax Reference』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] 手続きでは、複数の変数の分布を比較します。

例:€ 世間の人が、医者、法律家、警察官、教師という職業に対して抱く権威の大きさに差はあるでしょうか?10 人に依頼して、権威の大きさについてこの 4 種類の職業を順位付けしてもらいました。Friedman の検定では、世間の人がこの 4 つの職業に対して抱く権威の大きさに差があるかどうかを示します。

統計量平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Friedman、Kendall の W、および Cochran の Q。

データ。順位付けできる数値変数を使用します。

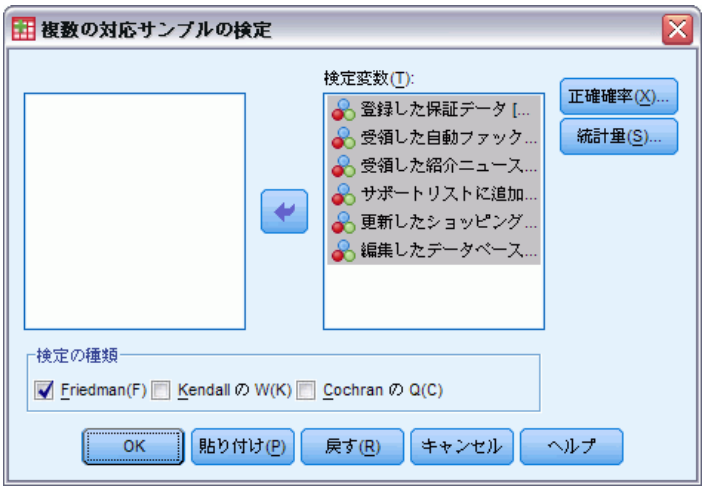
仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。無作為に抽出された対応のあるサンプルを使用します。

複数の対応サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析(A) > 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...

図 27-81
Cochran を検定の種類として選択



▶ 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

複数の対応する変数の分布を比較するには、3 つの検定が利用できます。

Friedman の検定は、1 サンプルの反復測定の分散分析やセルごとに 1 つの観測を持つ 2 元配置分散分析に対応するノンパラメトリック検定です。Friedman の検定は、k 個の対応変数は同じ母集団から発生するという帰無仮説です。各ケースに対して、k 個の変数は 1 から k に順位付けされます。検定の統計量は、それらの順位に基づきます。

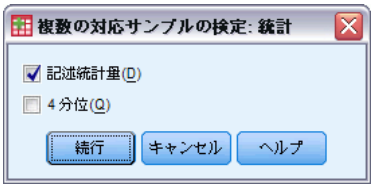
Kendall の W は Friedman の統計の規準です。Kendall の W は、評価の一致度の尺度となっている一致係数として解釈することができます。各ケースは評価者で、各変数は評価されている品目または人です。各変数に対して、順位の合計が計算されます。Kendall の W は、0（まったく一致していない）と 1（完全な一致）の間の値を取ります。

Cochran の Q は、Friedman の検定とまったく同じですが、すべての応答が 2 値データのときに適用します。この検定は、McNemar 検定を k 個のサンプルの場合に拡張したものです。Cochran の Q は、複数の対応のある

2 分変数には同じ平均値があるという仮説を検定します。変数は同一の個体または対応のあ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

図 27-82
[複数の対応サンプルの検定: 統計] ダイアログ ボックス



統計を選択します。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **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。

NPAR 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は、『Command Syntax Reference』を参照してください。

章
28

多重回答の分析

多重 2 分および多重カテゴリ グループの分析には、2 つの手続きが利用できます。[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 次元および 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:€ この例は、市場調査での多重回答項目の使用を示しています。データは架空のものであり、実際のデータとして解釈してはなりません。ある航空会社が、競合する航空会社を評価するために、特定ルートを飛行する旅客を調査するとします。この例では、アメリカン エアラインが、シカゴとニューヨーク間での他の航空会社の利用状況と、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要視されるかを調査します。客室乗務員が、各旅客の搭乗時に簡単なアンケートを手渡します。最初の質問は、「アメリカン、ユナイテッド、トランスワールド、US エア、その他のうち、最近 6 か月間にこの路線で少なくとも 1 回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ので、多重回答質問です。しかし、変数は各ケースに 1 つの値しか設定できないので、この質問は直接コード化できません。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。それには、2 つの方法があります。その 1 つは、それぞれの選択肢（たとえば、アメリカン、ユナイテッド、トランスワールド、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に丸をつけた場合、変数 united にコード 1が、そうでない場合は 0 が割り当てられます。これは、変数のマップの**複合二分法**です。回答をマップするもう 1 つの方法は、**多重カテゴリ法**というもので、利用した航空会社を指定するコードを使用して、質問に対して考えられる最大数の回答を予測し、同じ数の変数を設定する方法です。アンケートのサンプルを詳細に調べると、どの旅客も最近 6 か月間にこの路線で利用した航空会社は、最高 3 社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、その他のカテゴリには 10 社の航空会社名が記されています。多重回答方法を使用し、3 つの変数を定義して、1 = アメリカン、2 = ユナイテッド、3 = トランスワールド、4 = US エア、5 = デルタのようにそれぞれをコード化します。特定の旅客がアメリカンとトランスワールドに丸を付けると、最初の変数はコード 1 になり、2 番目の変数はコード 3 となり、3 番目の変数は欠損値コードになります。別の旅客がアメリカンとデルタに丸を付けた場

合、。最初の変数はコードが 1 になり、2 番目の変数はコードが 5 になり、3 番目の変数は欠損値コードになります。一方、多重 2 分方法では、14 の別個の変数が使用されます。この調査ではどちらのマッピング方法も利用できますが、回答の分類によって選択する方法が決まります。

多重回答グループを定義

[多重回答グループを定義] 手続きでは、基本変数を多重 2 分グループおよび多重カテゴリ グループに分けて、度数分布表およびクロス集計表を求めることができます。多重回答グループは、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] ボックスの一覧でそのグループを選択し、**[除去]** をクリックします。グループを変更するには、同様に選択してグループの定義特性を修正し、**[変更]** をクリックします。

基本変数は、2 分またはカテゴリとしてコード化できます。2 分変数を使用するには、**[2 分]** を選択して多重 2 分グループを作成します。集計値に対して整数値を入力します。集計値が少なくとも 1 回発生する各変数は、多重 2 分割グループのカテゴリになります。**[カテゴリ]** を選択して、成分変数と同じ範囲の値を持つ多重カテゴリ グループを作成します。多重カテゴリ グループのカテゴリの範囲として、最小値と最大値の整数値を入力します。この手続きでは、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。空のカテゴリは表にされません。

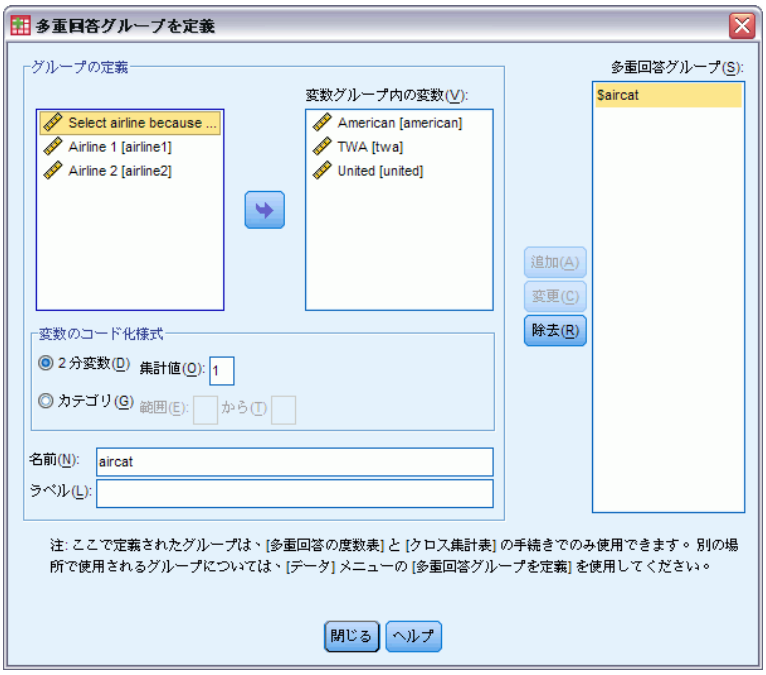
各多重回答グループには、7 文字までの一意の名前を割り当てる必要があります。この手続きでは、ユーザーが割り当てた名前の前にドル記号 ($) を付けます。ただし、予約名 casenum、sysmis、jdate、date、time、length、width は使用できません。多重回答グループの名前は、[多重回答] 手続き専用です。多重回答グループ名を他の手続きで参照することはできません。オプションとして、多重回答グループに記述的な変数ラベルを入力できます。ラベルの最大文字数は、40 文字です。

多重回答グループを定義するには

▶ メニューから次の項目を選択します。

分析(A) > 多重回答 > 変数グループの定義(E)...

図 28-1
[多重回答グループを定義] ダイアログ ボックス



- ▶ 2 つ以上の変数を選択します。
- ▶ 変数が 2 分変数としてコード化されている場合には、数えておき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には、カテゴリの範囲を定義します。
- ▶ 各多重回答グループに対して、一意の名前を入力します。
- ▶ 定義したグループのリストに多重回答グループを追加するには、[追加] をクリックします。

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 (「多重回答グループを定義」を参照)。

多重 2 分グループでは、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定義した変数ラベルによって決められます。変数名を定義しない場合、変数名はラベルとして使用されます。多重カテゴリ グループでは、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付けられます。最初の変数に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は、欠損カテゴリに値ラベルを定義します。

欠損値。欠損値のあるケースは、テーブルごとに除外されます。代わりに、次の中から 1 つまたは両方を選択できます。

- **2 分グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重 2 分グループの表から除外します。これは、2 分グループとして定義された多重回答グループだけに適用します。デフォルトでは、ケースの成分変数のどれにも集計値がない場合、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。すべての変数にではなく、いくつかの変数に欠損値があるケースは、1 つ以上の変数に集計値があれば、そのグループの集計表に含まれます。
- **カテゴリ グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重カテゴリ グループの表から除外します。これは、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のみ適用します。デフォルトでは、ケースの成分変数のどれにも定義された範囲内に有効な値がない場合にだけ、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したそれぞれの変数は、基本変数です。多重回答項目を分析するには、変数を 2 種類の多重回答グループ、つまり、多重 2 分グループまたは多重カテゴリ グループのどちらか 1 つに結合する必要があります。たとえば、航空会社の調査で、3 つの航空会社（アメリカン、ユナイテッド、トランスワールド）のうち最近 6 か月間に利用したのはどの会社かという質問をした場合、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 つ以上の航空会社を言及しないとわかった場合は、各航空会社に 1 つずつ、計 3 つのコードを持った 2 つの変数を作成できます。**多重カテゴリ グループ**を定義する場合、値は基本変数で同じコードを一緒に追加して集計されます。結果としての値のグループは、基本変数のものと同じになります。たとえば、ユナイテッドとして答えた 30 の回答は、航空会社 1 として ユナイテッドを答えた 5 の回答と航空会社 2 として ユナイテッド を答えた 25 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。

統計量。度数を表示する度数分布表、回答のパーセント、ケースのパーセント、有効なケース数、および欠損ケースの数。

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、分布から取り込んだデータについて有効な記述をします。

関連手続き。［多重回答グループを定義］手続きでは、多重回答グループを定義できます。

多重回答の度数分布表を作成するには

▶ 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...

図 28-2
[多重回答の度数表] ダイアログ ボックス

▶ 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 パーセントの表示、欠損値の処理の修正、または対応するクロス集計表の作成など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 分グループでは、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定義した変数ラベルによって決められます。変数名を定義しない場合、変数名はラベルとして使用されます。多重カテゴリ グループでは、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付けられます。最初の変数に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は、欠損カテゴリに値ラベルを定義します。この手続きでは、1 行あたり最大 8 文字で、3 つの行にある列にカテゴリ ラベルを表示します。単語が分割しないようにするには、行と列の項目を反転するか、またはラベルを再定義します。

例。多重 2 分変数グループと多重カテゴリ グループの両方は、この手続きで他の変数とクロス集計できます。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 か月間に少なくとも 1 回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、ユナイテッド、トランスワールド)。航空会社の選択時に、スケジュールとサービ

スのどちらがより重要ですか。1 つだけ選択してください。「。2 分変数または多重カテゴリとしてデータを入力してから、そのデータを 1 セットに結合すると、サービスまたはスケジュールに関連する質問の航空会社の選択肢をクロス集計できます。

統計量。セル、行、列および総数付き集計表、およびセル、行、列、および合計パーセント。セル パーセントは、ケースまたは回答に基づくことができます。

データ。多重回答グループまたは数値カテゴリ変数を使用します。

仮定。度数とパーセントは、分布から取り込んだデータについて有効な記述をします。

関連手続き。［多重回答グループを定義］手続きでは、多重回答グループを定義できます。

多重回答のクロス集計表を作成するには

▶ 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...

図 28-3
［多重回答のクロス集計表］ダイアログ ボックス

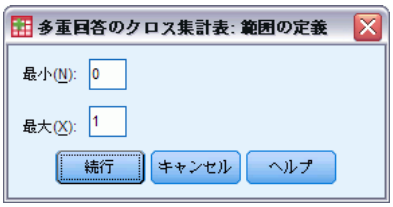
▶ 1 つ以上の数値変数を選択するか、クロス集計表の各次元に多重回答グループを選択します。

▶ 各基本変数の範囲を定義します。

オプションとして、管理変数の各カテゴリに対し、2 元配置のクロス集計表または多重回答グループを作成できます。［層］リストに対して 1 つ以上の項目を選択します。

多重回答のクロス集計表の範囲の定義

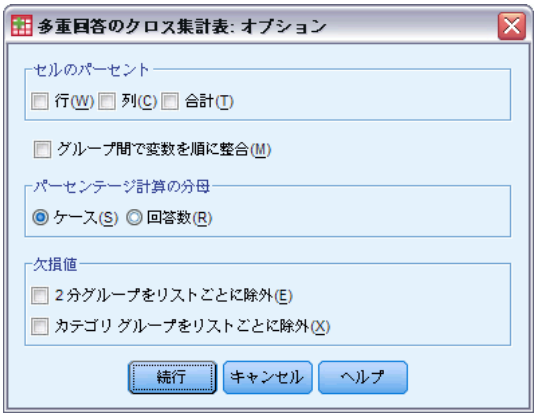
図 28-4
[多重回答のクロス集計表: 範囲の定義] ダイアログ ボックス



値の範囲は、クロス集計表のどの基本変数にも定義する必要があります。集計したい整数の最小カテゴリ値と最大カテゴリ値を入力します。範囲外のカテゴリは、分析から除外されます。範囲内の値は整数であるとみなされます（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

図 28-5
[多重回答のクロス集計表: オプション] ダイアログ ボックス



セルのパーセント。セル度数は常に表示されます。選択すると、行パーセント、列パーセント、および二元配置（全体）パーセントを表示できます。

パーセンテージ計算の分母。セルのパーセントをケース（または回答）に基づいて設定できます。これは、多重カテゴリ グループ全体で変数の結合を選択すると利用できません。セル パーセントは、回答に基づくこともできます。多重 2 分割グループでは、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 グループでは、回答数は定義された範囲内の値数です。

欠損値。次の中から 1 つまたは両方を選択することができます。

- **2 分グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重 2 分グループの表から除外します。これは、2 分グループとして定義された多重回答グループだけに適用します。デフォルトでは、ケースの成分変数のどれにも集計値がない場合、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。すべての変数にではなく、いくつかの変数に欠損値があるケースは、1 つ以上の変数に集計値があれば、そのグループの集計表に含まれます。
- **カテゴリ グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重カテゴリ グループの表から除外します。これは、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のみ適用します。デフォルトでは、ケースの成分変数のどれにも定義された範囲内に有効な値がない場合にだけ、多重カテゴリ 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 グループをクロス集計すると、最初のグループ内にある各変数を 2 番目のグループ内にある各変数と集計して各セルの度数を合計するので、いくつかの回答が 1 つのテーブルで複数回表示されることもあります。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。1 番目のグループにある最初の変数を 2 番目のグループにある最初の変数とペアにし、以後同じようにします。このオプションを選択すると、セル パーセントは、回答者ではなく回答数に基づきます。対応は、多重 2 分グループまたは基本変数には利用できません。

MULT RESPONSE コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 5 次元までのクロス集計表の作成（BY サブコマンドを使用）。
- 値ラベルの抑制を含む出力書式オプションの変更（FORMAT サブコマンドを使用）。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
29

結果の報告

ケースのリストおよび記述統計量は、データを調査、提示するための基本ツールです。データ エディタまたは［ケースの要約］手続きを使ってケースのリストを、［度数分布表］手続きを使って度数と記述統計量を、および［グループの平均］手続きを使ってグループ別の母集団の統計量を得ることができます。それぞれの手続きでは、情報を明確にするためにレイアウトされた書式を使用します。情報を別の書式で表示したい場合は、［報告書: 行の集計］および［報告書: 列の集計］ダイアログ ボックスで、データ表示を必要に応じて指定できます。

報告書の行の集計

［報告書: 行の集計］手続きでは、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。要約統計量を伴うまたは伴わないで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持っている会社が、給与、勤務年数、および各従業員が働く店舗と部門を含む従業員情報を記録しているとします。各店舗、部門、および各店舗内の部門の要約統計量（たとえば、平均給与）を、店舗と部門別（ブレーク変数）に分類した個別の従業員情報（リスト）を示した報告書を作成できます。

データ列。必要とするケースのリストまたは要約統計量のための変数を表示して、データ列の表示書式を指定します。

ブレーク列。報告書をグループに分割する任意のブレーク変数を表示して、要約統計量およびブレーク列の表示書式を指定します。ブレーク変数が複数ある場合は、リスト内の先行するブレーク変数のカテゴリ内にある各ブレーク変数の各カテゴリに対して個別のグループができます。ブレーク変数は、ケースを一定数の有意なカテゴリに分割するカテゴリ変数でなければなりません。各ブレーク変数の個々の値は、データ列の左に、個別の列で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、および表題を含む報告書全体の特徴を指定します。

© Copyright SPSS Inc. 1989, 2010

ケースの表示。それぞれのケースに対して、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、集計報告書よりもかなり長くなるリスト報告書が作成されます。

プレビュー。報告書の最初のページのみを表示します。このオプションは、報告書全体を処理しないで報告書の書式を下見するときに有効です。

データは並べ替え済み。ブレーク変数のある報告書では、報告書を作成する前に、データ ファイルをブレーク変数値でソートする必要があります。データ ファイルがブレーク変数で既に並び替えられている場合、このオプションを選択すると処理時間を節約できます。このオプションは、プレビュー報告書を実行した後には特に有効です。

集計報告書: 行の集計を取得するには

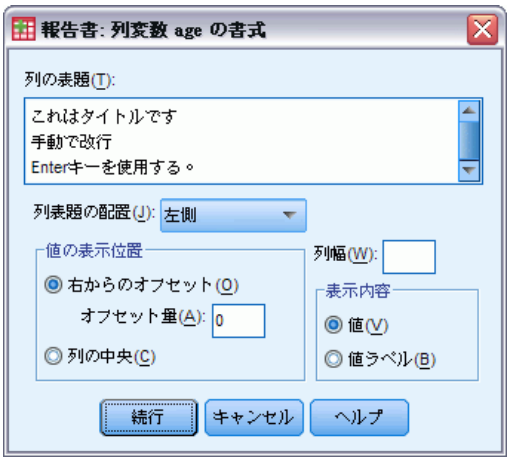
- メニューから次の項目を選択します。
  分析(A) > 報告書 > 報告書: 行の集計...
- データ列に 1 つ以上の変数を選択します。選択された各変数に対して報告書内に 1 つの列が生成されます。
- サブグループで並べ替えられ、表示された報告書に対しては、ブレーク列に 1 つ以上の変数を選択します。
- ブレーク変数で定義されたサブグループの要約統計量を表示する報告書に対しては、ブレーク列変数リスト内のブレーク変数を選択して、ブレーク列グループ内の [**集計**] をクリックして集計項目を指定します。
- 全体の要約統計量を表示する報告書では、[**集計**] をクリックして集計項目を指定します。

図 29-1
[報告書: 行の集計] ダイアログ ボックス

報告書のデータ列/ブレーク列の書式

[書式] ダイアログ ボックスでは、列の表題、列幅、テキストの位置合わせ、およびデータ値または値ラベルの表示を指定します。データ列の [書式] では、報告書ページの右側のデータ列の書式を指定します。ブレーク列の [書式] では、左側のブレーク列の書式を指定します。

図 29-2
[報告書: 列の集計] ダイアログ ボックス



列の表題。選択した変数に対して、列の表題を指定します。長い表題は、列内で自動的に折り返されます。表題を折り返したい場所に手動で改行を挿入するには、Enter キーを押します。

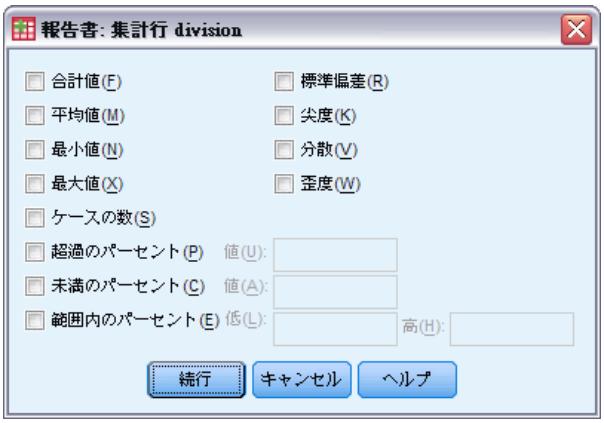
値の表示位置。選択した変数に対して、列内のデータ値または値ラベルの位置合わせを指定します。値またはラベルの位置合わせは、列見出しの位置合わせに影響しません。指定した文字数だけ列の内容をインデントすることもできますが、内容を中央に設定することもできます。

表示内容。選択した変数に対して、データ値または定義済みの値ラベルのどちらを表示するかを指定します。データ値は、常に定義済みの値ラベルを持っていない値に対して表示されます。(列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 ボックスでは、ブレーク グループおよび全体の報告書に対して要約統計量の表示を指定します。[集計行] ダイアログ ボックスは、ブレーク変数で定義された各カテゴリ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 ボックスでは、報告書の終わりに表示される全体の統計量を指定します。

図 29-3
[報告書: 最終集計行] ダイアログ ボックス

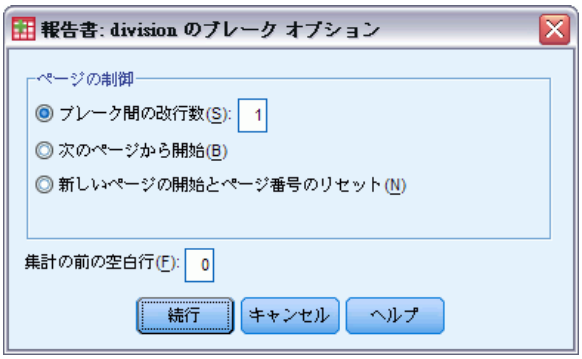


使用できる要約統計量としては、合計、平均値、最小値、最大値、ケースの数、指定した値の上下でのケースのパーセント、指定した範囲値内のケースのパーセント、標準偏差、尖度、分散、および歪度があります。

報告書のブレーク オプション

ブレーク列のオプションでは、ブレーク カテゴリ情報の改行とページ番号を指定します。

図 29-4
[報告書: ブレークのオプション] ダイアログ ボックス



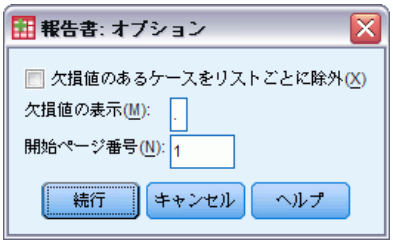
ページの制御。選択したブレーク変数のカテゴリに対して改行およびページ番号を指定します。ブレーク カテゴリ間の空白行数を指定することもできますが、新しいページから各ブレーク カテゴリを開始することもできます。

集計の前の空白行。ブレーク カテゴリ ラベルまたはデータと要約統計量間の空白行の数を指定します。これは、ブレーク カテゴリに個々のケース リストと要約統計量の両方を含む報告書には特に有効です。この報告書では、ケース リストと要約統計量間にスペースを挿入できます。

報告書のオプション

[報告書: オプション] ダイアログ ボックスでは、欠損値の扱いと表示および報告書のページ番号を指定します。

図 29-5
[報告書: オプション] ダイアログ ボックス



欠損値のあるケースをリストごとに除外。報告書の変数のどれかに対して欠損値のあるケースを（報告書から）除去します。

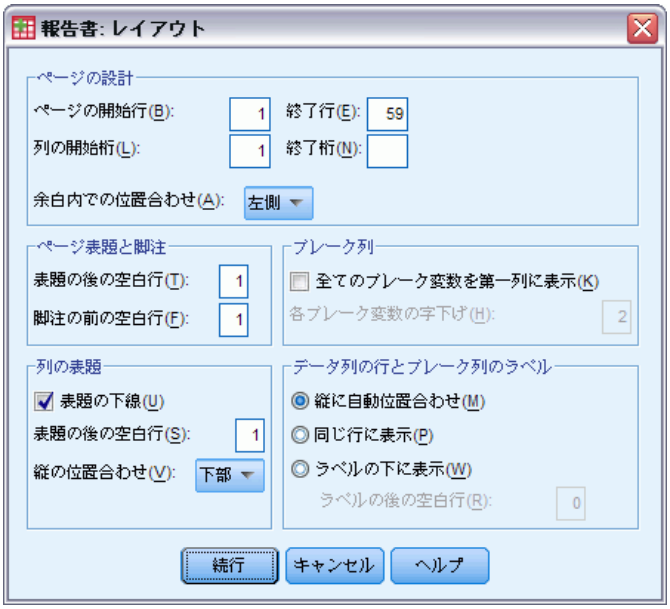
欠損値の表示。データ ファイル内の欠損値を表す記号を指定できます。記号は 1 文字だけ使用でき、この記号を使用して、**システム欠損値**と **ユーザー欠損値**の両方を表すことができます。

開始ページ番号。報告書の最初のページにページ番号を指定できます。

報告書のレイアウト

［報告書: レイアウト］ダイアログ ボックスでは、各報告書ページの幅と長さ、ページ上での報告書の配置、および空白行とラベルの挿入を指定します。

図 29-6
［報告書: レイアウト］ダイアログ ボックス



ページの設計。行単位（上端と下端）と文字単位（左と右）で表されるページ余白および余白内での報告書の位置合わせを指定します。

ページ表題と脚注。ページ表題と脚注を報告書の本文から分離する行数を指定します。

ブレーク列。ブレーク列の表示を指定します。複数のブレーク変数を指定すると、変数を個別の列または最初の列に入れることができます。最初の列にすべてのブレーク変数を置くと、幅の狭い報告書が作成されます。

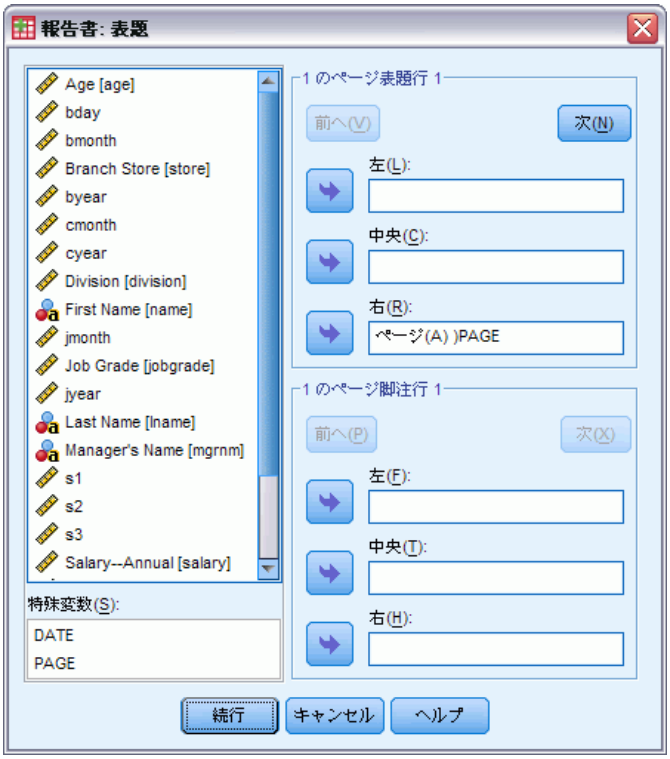
列の表題。表題に付ける下線、表題と報告書の本文との間の空白行、および列表題の縦の位置合わせを指定します。

データ列の行とブレーク列のラベル。各ブレーク カテゴリの最初のブレーク ラベルに関してデータ列情報（データ値または要約統計量）の配置を指定します。データ列情報の最初の行は、ブレーク カテゴリ ラベルと同じ行、またはブレーク カテゴリ ラベル行から指定した行数後のどちらかから開始できます。（列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］ダイアログ ボックスでは、報告書の表題と脚注の内容および配置を指定します。各行の内容を左詰め、中央、または右詰めにして、10 行までのページ表題、および 10 行までのページ脚注を指定できます。

図 29-7
［報告書: 表題］ダイアログ ボックス



変数を表題または脚注に挿入した場合は、現在の値ラベルまたは変数の値が表題または脚注に表示されます。表題では、ページの始めに変数の値に対応する値ラベルが表示されます。脚注では、ページの終わりに変数の値に対応する値ラベルが表示されます。値ラベルがない場合は、実際の値が表示されます。

特殊変数。特殊変数 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表題または脚注の任意の行に挿入できます。データ ファイルに DATE または PAGE という変数がある場合は、この変数を報告書の表題または脚注で使用できません。

報告書の列の集計

［報告書: 列の集計］では、個々の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持っている会社が、給与、勤務年数、および各従業員が働く部門を含む従業員情報を記録しているとします。各部門に給与の要約統計量（たとえば、平均値、最小値、最大値）を示す報告書を作成できます。

データ列。要約統計量を必要とする変数を表示し、各変数に表示される表示書式と要約統計量を指定します。

ブレーク列。報告書をグループに分割する任意のブレーク変数を表示し、ブレーク列の表示書式を指定します。ブレーク変数が複数ある場合は、リスト内の先行するブレーク変数のカテゴリ内にある各ブレーク変数の各カテゴリに対して個別のグループができます。ブレーク変数は、ケースを一定数の有意なカテゴリに分割するカテゴリ変数でなければなりません。

報告書。 欠損値、ページ番号、および表題を含む報告書全体の特徴を指定します。

プレビュー。報告書の最初のページのみを表示します。このオプションは、報告書全体を処理しないで報告書の書式を下見するときに有効です。

データは並べ替え済み。ブレーク変数のある報告書では、報告書を作成する前に、データ ファイルをブレーク変数値でソートする必要があります。データ ファイルがブレーク変数で既に並び替えられている場合、このオプションを選択すると処理時間を節約できます。このオプションは、プレビュー報告書を実行した後には特に有効です。

集計報告書: 列の集計を取得するには

- ▶ メニューから次の項目を選択します。
  分析(A) > 報告書 > 報告書: 列の集計...
- ▶ データ列に 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。
- ▶ 変数の集計項目を変更するには、データ列変数の変数を選択して **[集計]** をクリックします。
- ▶ 1 つの変数に複数の集計項目を使用するには、左側の変数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、それに希望する集計項目の 1 つを割り当てます。この操作を希望する集計項目の数だけ繰り返します。
- ▶ 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、**[列合計の挿入]** をクリックします。これで、合計という変数がデータ列リストに挿入されます。

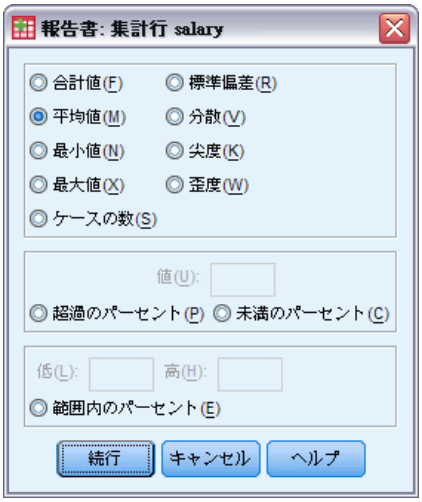
▶ サブグループで並べ替えられ、表示された報告書に対しては、ブレーク列に 1 つ以上の変数を選択します。

図 29-8
[報告書: 列の集計] ダイアログ ボックス

データ列の集計関数

集計行ダイアログ ボックスでは、選択されたデータ列変数に表示される要約統計量を指定します。

図 29-9
[報告書: 最終集計行] ダイアログ ボックス



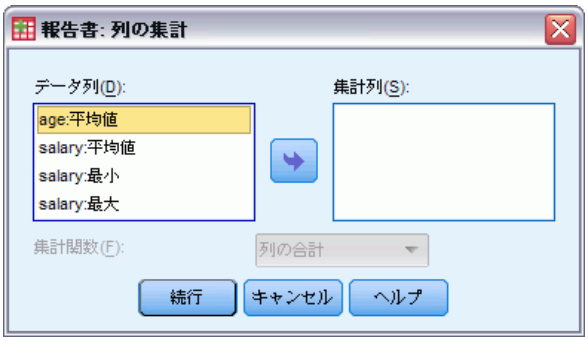
使用できる要約統計量としては、合計、平均値、最小値、最大値、ケースの数、指定した値の上下でのケースのパーセント、指定した範囲値内のケースのパーセント、標準偏差、尖度、分散、および歪度があります。

合計列のデータ列集計

列の集計では、2 つ以上のデータ列を集計する合計の要約統計量を指定します。

使用できる合計の要約統計量としては、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、および列の値を掛けた積があります。

図 29-10
[報告書: 列の集計] ダイアログ ボックス



列の合計。合計の列は、[集計列] リストにある列の合計です。

列の平均。合計の列は、[集計列] リストにある列の平均です。

列の最小。合計の列は、[集計列] リストにある列の最小値です。

列の最大。合計の列は、[集計列] リストにある列の最大値です。

1 列目 - 2 列目。合計の列は、[集計列] リストにある列の差です。[集計列] リストには 2 つの列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] リストにある列の商です。[集計列] リストには 2 つの列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] リストにある第 2 列に対する最初の列のパーセントです。[集計列] リストには 2 つの列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] リストにある列の積です。

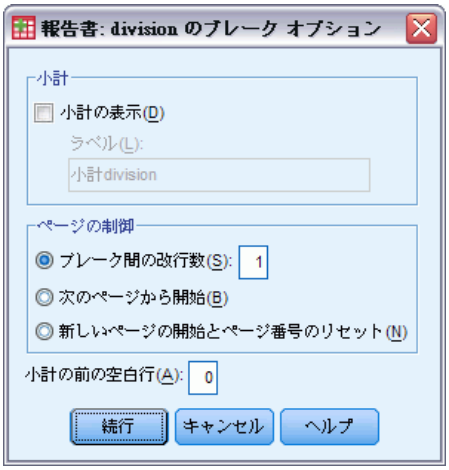
報告書の列の書式

[報告書: 列の集計] ダイアログ ボックスのデータ列とブレーク列の書式オプションは、[報告書: 行の集計] ダイアログ ボックスで説明したものと同じです。

報告書の列の集計でのブレーク列のオプション

ブレーク列のオプションでは、ブレーク カテゴリの小計表示、改行、およびページ番号を指定します。

図 29-11
[報告書: ブレークのオプション] ダイアログ ボックス



小計。ブレーク カテゴリの小計表示を指定します。

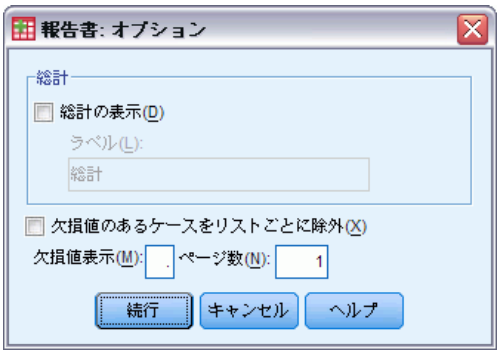
ページの制御。選択したブレーク変数のカテゴリに対して改行およびページ番号を指定します。ブレーク カテゴリ間の空白行数を指定することもできますが、新しいページから各ブレーク カテゴリを開始することもできます。

小計の前の空白行。ブレーク カテゴリ データと小計間の空白行の数を指定します。

報告書の列の集計のオプション

オプションで、列集計報告書内の総計の表示、欠損値の表示、およびページ番号を指定します。

図 29-12
[報告書: オプション] ダイアログ ボックス



総計。列の下端に表示される各列の総計の表示とラベル付けを行います。

欠損値。報告書から欠損値を除外できますが、報告書内の欠損値を示すために 1 つの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] ダイアログ ボックスの [報告書: レイアウト] オプションは、[報告書: 行の集計] ダイアログ ボックスで説明したものと同じです。

REPORT コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 1 つの集計行の列に別の集計関数を表示。
- 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。
- 中央値、最頻値、度数、およびパーセントを集計関数として使用。
- 要約統計量の表示書式をさらに正確に指定。

- 報告書内のさまざまな点に空白行を挿入。
- 報告書のリストで n 番目のケースごとに空白行を挿入。

REPORT シンタックスは複雑なため、ダイアログ ボックスから作成した報告書に近づけるためにシンタックスで新しい報告書を構築するときには、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希望する正確な報告書を作成するようにしてください。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
30

信頼性分析

信頼性分析では、測定尺度とスケールを構成する項目の特性を調査できます。[信頼性分析] 手続きは、一般に使用される多数の尺度の信頼性の測度を計算し、尺度の各項目間の関係についての情報を提供します。級内相関係数は、評価者間の信頼性推定値の計算に役立ちます。

例:€ 質問表は消費者の満足度の測定に役立つでしょうか。信頼性分析を使用して、質問表の項目間の関連の度合いを決定し、反復性の全体的な指標や尺度全体の内部的一貫性を求めることができます。また、問題のある項目を識別して尺度から除去できます。

統計量 各変数と尺度の記述統計、項目での要約統計、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 乗、Tukey の加法性の検定。

モデル。次の信頼性のモデルを使用できます。

- **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。**このモデルでは、繰り返しを通じ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に加えて、すべての項目で平均が等しいと仮定します。

データ。データは 2 分データ、順序データ、または間隔データでなければなりません。また、数値でコード化されている必要があります。

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 変量正規分布でなければなりません。スケールは加法的でなければならず、その結果、各項目は合計得点に対して線型になります。

関連手続き。尺度項目の各次元を詳細に調べる (項目得点のパターンの説明に複数の構成が必要かどうかを調べるため) 場合には、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識別するには、変数を寄せ集める階層的クラスタ分析を使用できます。

© Copyright SPSS Inc. 1989, 2010

信頼性分析を行うには

▶ メニューから次の項目を選択します。
分析(A) > スケール > 信頼性分析...

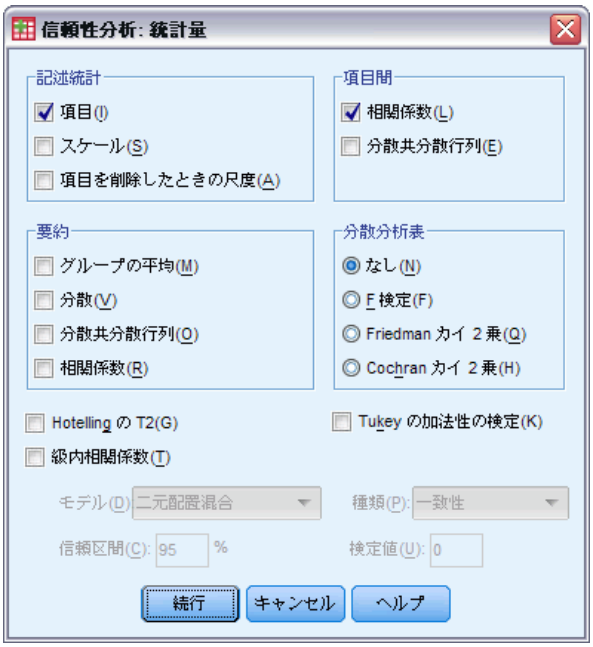
図 30-1
[信頼性分析] ダイアログ ボックス

▶ 加法的尺度の成分の候補として、複数の変数を選択します。

▶ [モデル] ボックスの一覧からモデルを選択します。

信頼性分析の統計

図 30-2
[信頼性分析: 統計量] ダイアログ ボックス



スケールと項目を表すさまざまな統計を選択できます。デフォルトで求める統計は、ケース数、項目数、および次の信頼性推定値です。

- **アルファ モデル。**アルファ係数。2 分変数の場合には、Kuder-Richradson の公式 20 (KR20) 係数と同じです。
- **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）および 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、および真の分散の推定値、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。すべてのケースで、尺度または項目の記述統計量を求めます。

- **項目。**すべてのケースで、項目の記述統計量を求めます。
- **スケール。**尺度の記述統計量を求めます。
- **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に比較した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間の相関、および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で項目分布の記述統計量を求めます。

- **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、項目平均値の範囲と分散、および最大項目平均値と最小項目平均値の比率が表示されます。
- **分散(V).** 項目分散の要約統計量。項目分散の最小、最大、平均、項目分散の範囲、分散、および項目分散の最大と最小の比率が表示されます。
- **分散共分散行列(O).** 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率が表示されます。
- **相関係数(L).** 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率が表示されます。

項目間。項目間の相関行列または分散共分散行列を作成できます。

分散分析表。平均が等しい検定を行います。

- **F 検定(F).** 反復測定（経時的変化）による分散分析表を表示します。
- **Friedman カイ 2 乗(Q).** 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、ランク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、分散分析表の通常の F 検定に代わるものです。
- **Cochran カイ 2 乗(H).** Cochran の Q を表示します。このオプションは、2 分変数データに適しています。Q 統計量は、分散分析表の通常の F 統計量に代わるものです。

Hotelling の T2。スケール上の全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うことができます。

Tukey の加法性の検定。項目間に（乗法的）交互作用がないという仮説の検定を行うことができます。

級内相関係数。ケース内の一致度または一致値を測定します。

- **モデル。**級内相関係数を計算するモデルを選択します。使用できるモデルは二元配置混合、二元配置変量、および一元配置変量です。人的効果が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**] を選択するか、人的効果が変量である場合は [**一元配置変量**] を選択します。
- **型。**指標の種類を選択します。使用できる種類は一致性と絶対一致です。
- **信頼区間。**信頼区間の係数を指定します。デフォルト値は 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る値です。デフォルト値は 0 です。

RELIABILITY コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 相関行列の読み込み、および分析。
- その後の分析で使用する相関行列の書き込み。
- 折半法で等分割以外の指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
31

多次元尺度法

多次元尺度法は、対象またはケース間の一群の距離測度の中に構造を見つける手法です。この手法では、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ます。この場合、空間内の点の間の距離が、指定した非類似度にできるだけ近接するようにします。多くの場合、データを深く理解するために、この概念的な空間の次元を解釈して使用します。

変数が客観的に測定されている場合には、多次元尺度法をデータ削減のための手法として使用できます（必要であれば、[多次元尺度法] 手続きは多変量データの距離を計算します）。また、多次元尺度法は、対象または概念の非類似度を主観的に評価する場合にも適用できます。また、[多次元尺度法] 手続きでは、複数の評価者や複数のアンケートの回答者からのデータがある場合に、それらの複数のデータから非類似度を処理することもできます。

例:€ 人はどのようにして異なる車の関連に気づくでしょうか。車の異なった作りとモデルの類似度を評価した回答者からのデータがある場合は、多次元尺度法を使用して、消費者の認識を表す次元を識別できます。たとえば、価格と車の大きさが 2 次元空間を定義する場合があります。この空間は、回答者によって示される類似度を説明しています。

統計量。各モデル: データ行列、最適尺度化データ行列、S ストレス (Young)、ストレス (Kruskal)、RSQ、刺激座標、平均ストレス、各刺激の RSQ (RMDS モデル)。各差分 (INDSCAL) モデル: 被験者の重みおよび各被験者の奇異性のインデックス。重複した多次元尺度法モデルの各行列: 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)、差異と距離の散布図。

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数量でなければならず、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、変数は数量データ、2 値データ、または度数データとなります。尺度の違いは解に影響することもあるので、変数の尺度は重要な問題です。変数の尺度に大きな違いがある（たとえば、ある変数はドルで測定され、別の変数は年で測定されている）場合には、変数の標準化を検討する必要があります（[多次元尺度法] 手続きでは、標準化は自動的に実行できます）。

© Copyright SPSS Inc. 1989, 2010

仮定。［多次元尺度法］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、［多次元尺度法：オプション］ダイアログボックスで適切な尺度（［順序データ］、［区間データ］、または［比データ］）を選択する必要があります。

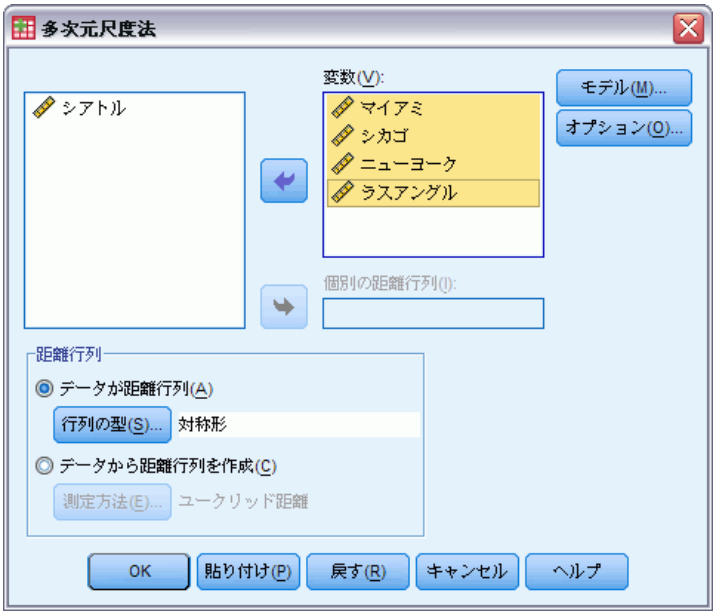
関連手続き。データの削減を目的としている場合で、特に変数が数量のときには、代わりに因子分析を使用できます。類似ケースのグループを識別する場合には、多次元尺度分析を階層クラスタ分析または大規模ファイルのクラスタ分析で補強することを検討します。

多次元尺度法分析を行うには

▶ メニューから次の項目を選択します。

分析(A) > スケール > 多次元尺度法...

図 31-1
［多次元尺度法］ダイアログ ボックス



▶ 分析には最低 4 つの数値型変数を選択します。

▶ ［距離行列］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**］をクリックします。

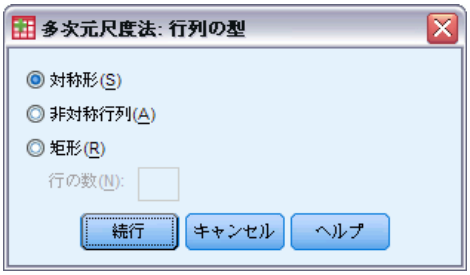
▶ ［**データから距離行列を作成**］を選択すると、個別行列にグループ変数を選択することもできます。グループ変数は、数値型と文字型のどちらでもかまいません。

オプションとして、次の選択も可能です。

- データが距離行列の場合、距離行列の形状を指定します。
- データから距離行列を作成する場合、使用する距離速度を指定します。

多次元尺度法のデータの形式

図 31-2
[多次元尺度法: 型] ダイアログ ボックス



作業データセットが対象グループ内の距離、または 2 組の対象グループ間の距離を表している場合に、正しい結果を得るには、データ行列の形式を指定する必要があります。

注 :[モデル] ダイアログ ボックスの [条件付け] で [行] を指定している場合、[**対称行列**] は選択できません。

多次元尺度法の尺度の作成

図 31-3
[多次元尺度法: データからの尺度の作成] ダイアログ ボックス

多次元尺度法では、非類似度データを使用して尺度解を作成します。データが多変量データ（測定された複数の変数の値）の場合は、非類似度データを作成し、多次元尺度法の解を計算します。データから非類似度の測度を作成する方法を詳細に指定できます。

測定方法。分析で使用する非類似度の測定方法を指定できます。データの種類に応じて［測定方法］のオプションを 1 つクリックします。選択したオプションのドロップダウン リストから測定方法を 1 つ選択します。使用可能な選択肢は次のとおりです。

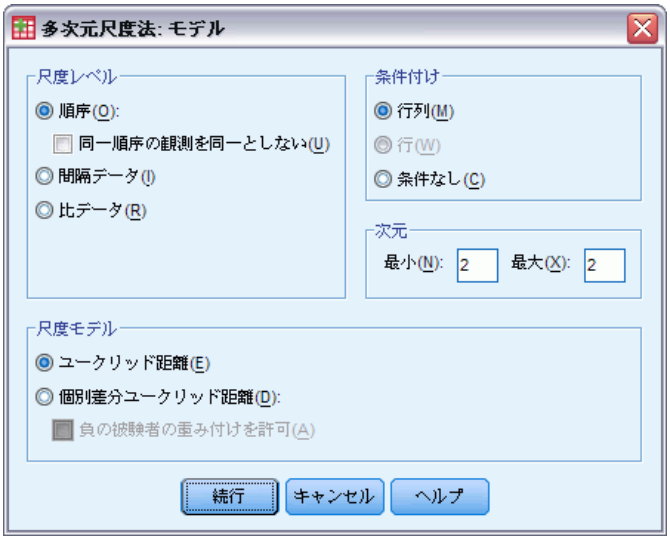
- **間隔。**［ユークリッド距離］、［平方ユークリッド距離］、［Chebychev］、［都市ブロック］、［Minkowski］、または［カスタマイズ］。
- **度数。**［カイ 2 乗測度］または［ファイ 2 乗測度］。
- **2 値。**［ユークリッド距離］、［平方ユークリッド距離］、［サイズの差異］、［パターンの差異］、［分散］、または［Lance と Williams］。

距離行列の作成。分析の単位を選択します。［変数間］または［ケース間］をクリックします。

値の変換。たとえば、変数がまったく異なるスケールで測定されている場合などで、値を標準化してから近接を計算します（2 値データには適用できません）。［標準化］ドロップダウン リストの一覧から標準化の方法を選択します。標準化を行う必要がない場合は、［なし］を選択します。

多次元尺度法のモデル

図 31-4
［多次元尺度法: モデル］ダイアログ ボックス



データとモデル自体の状態を指定することで、多次元尺度法のモデルを正確に推定します。

尺度レベル。データのレベルを指定できます。[順序データ]、[区間データ]、または [比データ] のいずれか 1 つをクリックします。変数が順序データの場合には、**[同一順序の観測を同一としない]** チェック ボックスをオンにすると、変数が連続変数として扱われるため、同順位のデータ（ケースが違っても同じ値を持つデータ）が最適に分析されます。

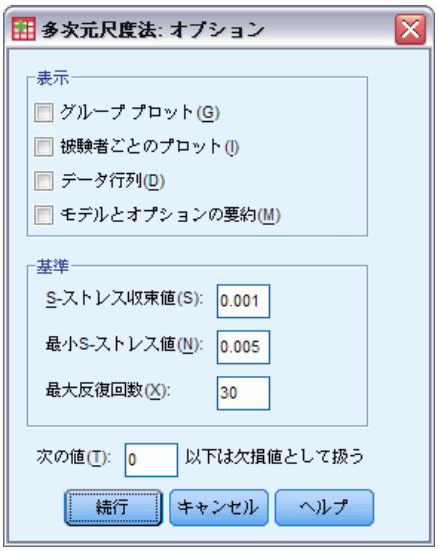
条件付け。意味のある比較を指定します。[行列]、[行]、[条件なし] のいずれか 1 つをクリックします。

次元。尺度法で求める解の次元数を指定します。範囲内の各数値に対して、解が 1 つ計算されます。ここでは、1 から 6 までの整数を指定できます。最小値の 1 を指定できるのは、尺度モデルに **[ユークリッド距離]** を指定した場合だけです。解を 1 つだけ求めるには、最小値と最大値に同じ数を指定します。

尺度モデル。尺度法を実行するための仮定を指定します。[ユークリッド距離] または [個人差ユークリッド距離] (INDSCAL とも言います) のどちらかをクリックします。個人差ユークリッド距離モデルでは、**[負の被験者の重み付けを許可]** を選択できます（ただし、そのデータに適当な場合）。

多次元尺度法のオプション

図 31-5
[多次元尺度法: オプション] ダイアログ ボックス



多次元尺度法でオプションを指定できます。

表示。各種の出力形式を指定できます。[グループ プロット]、[被験者ごとのプロット]、[データ行列]、[モデルとオプションの要約] が使用できます。

基準。反復を停止させる基準を指定します。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。指定した値以下の距離が分析から除外されます。

ALSCAL コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 多次元尺度法の文献で、ASCAL、AINDS、および GEMSCAL とよばれる追加モデルの使用。
- 区間データと比データの多項式変換の実行。
- （距離ではなく）類似度を順序データで分析。
- 名義データの分析。
- さまざまな座標や重み付き行列をファイルに保存し、分析で再使用。
- 多次元展開の制約。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
32

比率統計量

比率統計量手続きでは、2 つのスケール変数間の比率を示す要約統計量の総合的なリストを作成できます。

出力は、グループ化変数値により昇順または降順に並べ替えることができます。比率統計量報告書は出力内で抑制することができ、結果は外部ファイルとして保存することができます。

例:€ 家屋の査定額と販売額間の比率は、5 か国間でほぼ同じなのでしょうか。出力結果から、比率の分布は国によって大きく異なることがわかります。

統計量 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、および濃度インデックス。

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 (名義または順序尺度)。

仮定。 分子と分母の比率を定義する変数は、正の値を取るスケール変数でなければなりません。

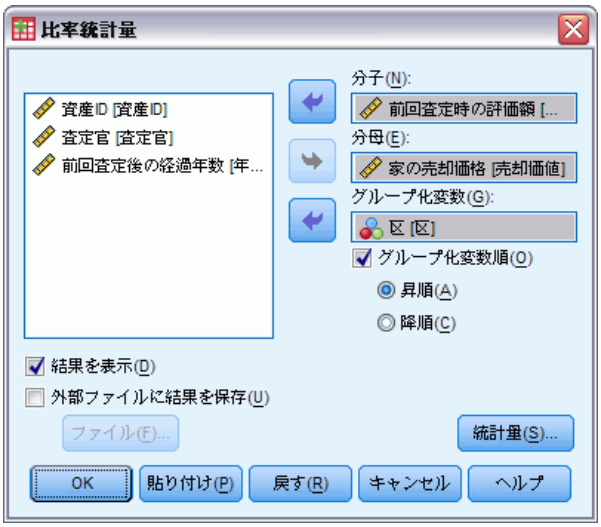
比率統計を行うには

▶ メニューから次の項目を選択します。

分析(A) > 記述統計 > 比率...

© Copyright SPSS Inc. 1989, 2010

図 32-1
[比率統計量] ダイアログ ボックス



▶ 分子変数を選択します。

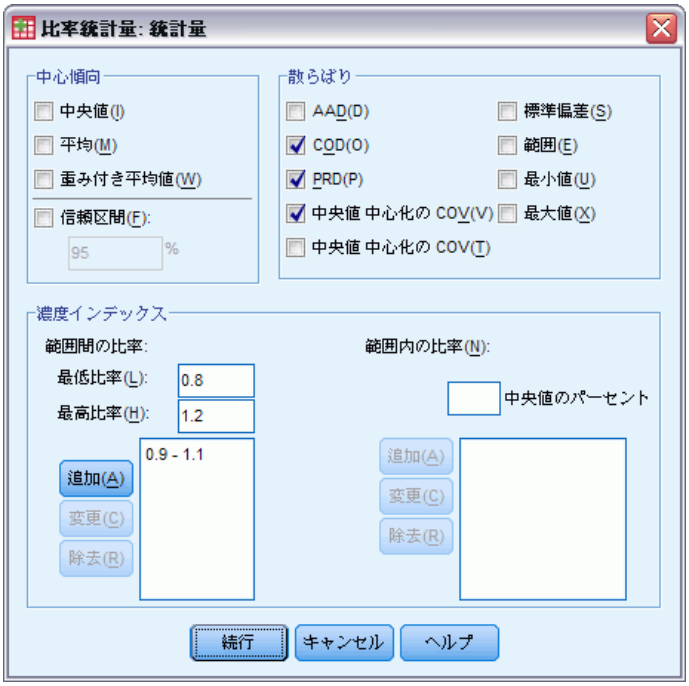
▶ 分母変数を選択します。

次のオプションが選択できます。

- グループ化変数を選択し、結果のグループ順序を指定します。
- ビューアに結果を表示するかどうかを選択します。
- 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名を指定します。

比率統計量

図 32-2
[比率統計量] ダイアログ ボックス



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。**この中央値より小さい比率の数と、大きい比率の数は同じになります。
- **平均値。**比率の合計の結果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、分母によって重みを付けられた比率の平均値にもなります。
- **信頼区間。**平均値、中央値、および重み付き平均値の信頼区間を表示します（要求がある場合）。信頼水準として、0 以上 100 未満の値を指定します。

散らばり。観測値での変動量、または広がり量を測定する統計量です。

- **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（回帰指数として知られている）は、平均値を加重平均で割ったものです。

- **中央値 中心化の COV。**中央値を基準にした変動係数は、中央値からの偏差のルート平均平方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。**標準偏差とは、平均値の比率の平方偏差合計を比率の総数から 1 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。**最少値とは、最少比のことです。
- **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、区間に入る比のパーセントを表します。濃度係数は、次の 2 種類の方法で計算できます。

- **範囲間の比率。**この区間は、区間の最低値および最高値を指定することにより明示的に定義されます。最低値と最高値の比率を入力し、[追加] をクリックして区間を指定します。
- **範囲内の比率。**この区間は、中央値のパーセントを指定することにより黙示的に定義されます。0 ～ 100 までの値を入力し、[追加] をクリックします。区間の下限は（1 − 0.01 × 値）× 中央値に等しく、上限は（1 + 0.01 × 値）× 中央値に等しくなります。

章
33

ROC 曲線

被験者を分類する 2 つのカテゴリに 1 つの変数がある場合に、この手続きを使用して分類方法のパフォーマンスを評価できます。

例:€ 銀行は、ローン返済の履行、不履行について顧客を正しく分類することに関心があります。したがって、そのための意思決定を行う特別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できます。

統計量 信頼区間を持つ ROC 曲線の下の領域と ROC 曲線の座標点。作図(T):ROC 曲線。

方法。ROC 曲線の下の面積の推定値は、ノンパラメトリックまたは 2 負指数モデルを使用するパラメトリックのいずれかで計算できます。

データ。検定変数は量的変数です。ほとんどの検定変数は、判別分析またはロジスティック回帰または任意のスケール上の得点（これは一方のカテゴリあるいは他方のカテゴリに被験者を分類する評価者の「確信の強さ」を示しています）からの確率です。状態変数はどのような種類のものでもよく、被験者が属する真のカテゴリを示しています。状態変数の値は、どのカテゴリを 正とみなすべきかを示します。

仮定。評価者のスケール上の数値が大きければ大きいほど、被験者が一方のカテゴリに属すという確信が強まり、このスケール上の数値が小さければ小さいほど、被験者が他方のカテゴリに属すという確信が強まると仮定します。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と仮定します。

ROC 曲線を作成するには

▶ メニューから次の項目を選択します。

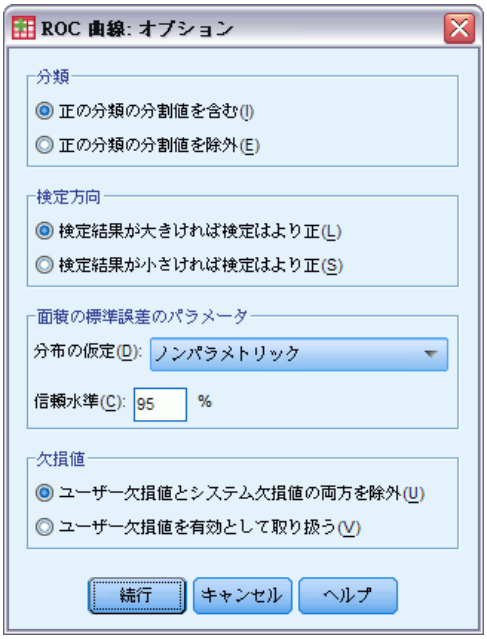
分析(A) > ROC 曲線(V)...

図 33-1
[ROC 曲線] ダイアログ ボックス

- ▶ 1 つ以上の検定変数を選択します。
- ▶ 1 つの状態変数を選択します。
- ▶ 状態変数の 正の値を識別します。

ROC 曲線のオプション

図 33-2
[ROC 曲線: オプション] ダイアログ ボックス



ROC 分析に次のオプションを指定できます。

分類。正の分類を実行するときに、分割点の値を含めるか除外するかを指定できます。この設定は、現在、出力に影響しません。

検定方向。正のカテゴリに対するスケールの方向を指定できます。

面積の標準誤差のパラメータ。曲線の下の面の標準誤差を推定する方法を指定できます。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 50.1～99.9% です。

欠損値。欠損値の扱い方を指定できます。

付録

A

Notices

Licensed Materials - Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shmtl.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

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

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, http://www.winwrap.com.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.

索引

層
　クロス集計表, 26
度数, 8, 33
　書式, 12, 36
　統計量, 10, 12
　度数分布表, 10
　最近隣分析, 159
　探索的, 20, 36
　因子分析, 176, 178
　信頼性分析, 321, 334
　一元配置分散分析, 58, 62
　作図, 114
　記述統計, 16
　報告書: 行の集計, 311
　GLM, 68–69
　OLAP キューブ, 45, 48
　ROC 曲線, 336
　オプション, 35
　階層クラスタ分析, 203
　グループの平均, 40
　クロス集計表, 30
　コードブック, 1
　サブグループ, 38, 43
　線型回帰での保存, 116
　表の抑制, 12
　列の集計報告書, 319
　大規模ファイルのクラスタ分析, 211
平均値, 38
　作図, 21
　統計量, 40
　度数分布表, 10
　因子分析, 180
　線型回帰, 119–120
　因子分析, 175–176
　最近隣分析, 152
　比率統計量, 334
　一元配置分散分析, 62
　記述統計, 16
　報告書: 行の集計, 312
　1 サンプルによる Kolmogorov-Smirnov 検定, 289
　1 サンプルの t 検定, 55
　OLAP キューブ, 45
　ROC 曲線, 338
　オプション, 40
　グループの平均, 40
　独立サンプルの t 検定, 52
　線型回帰での保存, 116
　行の集計報告書, 308
　報告書の列, 310
　対応のあるサンプルの t 検定, 53
　モデルのクラスタリング, 190
　ラン検定, 287
度数分布表
　度数分布表, 8, 10
因子分析, 173, 179
　例, 173
　基準, 169–170, 173, 176, 178
　欠損値, 170, 180
　事前確率, 170, 179
　最近隣分析, 144
　分散共分散行列, 170
　統計量, 119, 173, 175, 178
　線型回帰, 119
　順序回帰, 125, 168, 173, 175–176, 178
　判別分析法, 169
　因子抽出法, 176
　一元配置分散分析, 57, 59, 62
　残差, 116
　統計量, 16, 89
　記述統計, 16
　変換測定方法, 89
　線型回帰, 116
　1 サンプルの t 検定, 55
　2 個の対応サンプルの検定, 292
　GLM, 68–69, 75
　GLM 1 変量, 75
　Mahalanobis の距離(M), 169
　OLAP キューブ, 45
　Rao の V, 169
　ROC 曲線, 338
　Wilks のラムダ(W), 169
　表示オプション, 169–170
　クロス集計表, 27
　クロス集計表, 30
　ケースの選択, 174
　コマンドの追加機能, 172
　コマンドの追加機能, 180
　コマンドの追加機能, 129
　独立サンプルの t 検定, 52
　対応サンプルのノンパラメトリック検定, 232
　線型回帰での保存, 116
　係数の表示書式, 180
　分類変数の保存, 172
　類似度の測定方法, 89

新変数の保存, 116
対応のあるサンプルの t 検定, 53
大規模ファイルのクラスタ分析, 211
プロットのロード, 178
位置モデル, 126, 128
モデル情報のエクスポート, 116, 172
最近隣分析, 139, 173, 176
近隣, 144, 147, 150
特徴選択, 146
最近隣分析, 146
級内相関係数, 323
最近隣分析, 154
探索的, 20–21, 323
線型回帰, 119
記述統計, 323
比率統計量, 334
線型回帰, 119
分散分析表(A), 323
GLM 1 変量, 75
Hotelling の T^2, 323
Kuder-Richardson 20, 323
Tukey の加法性の検定(K), 323
オプション, 152
変数の保存, 149
項目間の相関係数および共分散, 323
線型モデル, 95
モデル ビュー, 153
標準回帰係数
線型回帰, 119
GLM 1 変量, 75
クロス集計表, 27
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, 232
線型モデル, 95
一元配置分散分析
欠損値, 62
多重比較, 59
オプション, 62
一元配置分散分析, 57
対比, 58
因子変数, 57
コマンドの追加機能, 63
多項式の対比, 58
分散共分散行列
判別分析, 170
線型回帰, 119
周辺等質性検定
2 個の対応サンプルの検定, 292
記述統計量
記述統計, 14
1 サンプルのノンパラメトリック検定, 219
順位相関係数
2 変量の相関分析, 78
分散共分散行列
判別分析, 168
記述統計, 14, 122
探索的, 20, 120, 122
線型回帰, 119
変数選択方法, 120
比率統計量, 334
報告書: 行の集計, 311
GLM, 68
TwoStep クラスタ分析, 186
z 得点の保存, 14
オプション, 123
階層クラスタ分析, 205
グループの平均, 40
クロス集計表, 30
ケース選択変数, 114
コマンドの追加機能, 121
コマンドの追加機能, 90
リンク, 123
変数増加法
線型回帰, 113
線型回帰, 116
TwoStep クラスタ分析, 188
比率統計量, 332
度数分布表, 10
一元配置分散分析
その後の検定, 59
多次元尺度法, 326
次元, 329–330
変換値, 328
距離測度, 328
表示オプション, 330
条件付け, 329
コマンドの追加機能, 331
データ行列の定義, 328
距離行列の作成, 328
尺度モデル, 329
尺度レベル, 329
探索的, 18
要約, 36
度数分布表, 10
統計量, 20–22
順序回帰, 125
2 項検定, 285
2 個の独立サンプルの検定, 292, 294
GLM, 66
オプション, 22
カイ 2 乗検定, 266
コマンドの追加機能, 23
列の集計報告書, 319
複数の独立サンプルの検定, 297
べき乗の変換, 22
判別分析, 165
例, 165

要約, 36, 168
統計量, 165
独立変数, 165
最近隣分析, 161
探索的, 20, 168
順序回帰, 125, 168
記述統計量, 168
GLM, 75
OLAP キューブ, 45
グループ化変数, 165
ケースの選択, 167
コマンドの追加機能, 17
ステップワイズ法, 165
範囲の定義, 167
信頼性分析, 321
例, 321
統計量, 321, 334
記述統計量
要約, 36
探索的, 20
一元配置分散分析, 59
四文位分布図
最近隣分析, 160
分散共分散行列
順序回帰, 125
回帰
線型回帰, 111
比率統計量, 334
多次元尺度法, 326
報告書: 列の集計, 316
列の集計報告書, 318
複数の回帰, 111
列の報告書の比較, 317
列の報告書全体の乗算, 317
複数の列の報告書の, 317
最尤法
因子分析, 176
TwoStep クラスタ分析, 186
主因子法, 176
線型回帰, 111
変数選択方法, 113
ブロック, 111
重み付きと成分負荷, 111
分割表, 24
偏相関, 84
集計列, 317
線型回帰, 114
2 変量の相関分析, 80
列の比較, 317
列の集計報告書, 315
合計の列, 317
列の値の乗算, 317
列の報告書の比較, 317
距離行列, 86
例, 86
偏相関, 82
距離, 88, 114
統計量, 86, 88
変換測定方法, 88
報告書: 列の集計, 316
0 次, 84
2 変量の相関分析, 78
クロス集計表, 26
ケース間の距離の計算, 86
非類似度の測定方法, 88
変数間の距離の計算, 86
偏相関分析, 82
統計量, 84
0 次相関, 84
オプション, 84
コマンドの追加機能, 325
コマンドの追加機能, 84
多次元尺度法
例, 326
統計量, 326
平均絶対偏差 (AAD)
比率統計量, 334
変動係数 (COV)
比率統計量, 334
級内相関係数 (ICC)
信頼性分析, 323
価格関連格差 (PRD)
比率統計量, 334
偏相関最小 2 乗法回帰, 134
モデル, 136
変数をエクスポート, 137
逆 Helmert 対比
GLM, 68–69
従属 t 検定
対応のあるサンプルの t 検定, 52
分散分析(N)
一元配置分散分析, 57
GLM 1 変量, 64
モデル, 66
分散分析(N)
グループの平均, 40
分散分析(N)
線型モデル, 105
0 次相関
偏相関, 84
1 サンプルの t 検定, 54
欠損値, 55
信頼区間, 55
オプション, 55
コマンドの追加機能, 56
1 サンプルのノンパラメトリック検定
ラン検定, 222

1 サンプルのノンパラメトリック検定, 215
2 項検定, 219
Kolmogorov-Smirnov 検定, 221
カイ 2 乗検定, 220
フィールド, 217
1サンプルによる Kolmogorov-Smirnov 検定, 287
統計量, 289
検定分布, 287
オプション, 289
コマンドの追加機能, 289
2 項検定
1 サンプルのノンパラメトリック検定, 218–219
2 項検定, 283
統計量, 285
2 分変数, 283
オプション, 285
コマンドの追加機能, 285
2 項検定, 283
統計量, 285
2 分変数, 283
オプション, 285
コマンドの追加機能, 285
2 サンプルの t 検定
独立サンプルの t 検定, 49
2 変量の相関分析
有意確率, 78
2 変量の相関分析
欠損値, 80
オプション, 80
コマンドの追加機能, 80
2 個の独立サンプルの検定, 289, 292
欠損値, 292
オプション, 292
グループ化変数, 291
グループの定義, 291
コマンドの追加機能, 292
検定の種類, 290
2 個の対応サンプルの検定
統計量, 294
オプション, 294
コマンドの追加機能, 294
検定の種類, 293
2 次モデル
曲線推定, 132
3 次モデル
曲線推定, 132
4 分位
度数分布表, 10
5% トリム平均
探索的, 20
Anderson-Rubin 因子得点, 179
Andrews のウェイブ推定量
探索的, 20
Bartlett 因子得点, 179
Bartlett の球面性の検定
因子分析, 175
Bonferroni(B)
一元配置分散分析, 59
Bonferroni の方法
GLM, 71
Box の M 検定
判別分析, 168
Brown-Forsythe 統計量
一元配置分散分析, 62
Chebychev の距離
距離, 88
Clopper-Pearson 区間
1 サンプルのノンパラメトリック検定, 219
Cochran 統計量
クロス集計表, 27
Cochran の Q
複数の対応サンプルの検定, 298
Cochran の Q 検定
対応サンプルのノンパラメトリック検定, 232, 234
Cohen のカッパ
クロス集計表, 27
Cook の距離
線型回帰, 116
Cook の距離
GLM, 73
Cox と Snell の R^2 乗
順序回帰, 125
Cramer の V
クロス集計表, 27
Cronbach のアルファ
信頼性分析, 323
信頼性分析, 321
d
クロス集計表, 27
DfBeta
線型回帰, 116
DfFit(F)
線型回帰, 116
Duncan の多重範囲検定
一元配置分散分析, 59
Duncan の多重範囲検定
GLM, 71
Dunnett の C
一元配置分散分析, 59
Dunnett の t 検定
一元配置分散分析, 59
Dunnett の T3(3)
一元配置分散分析, 59

Dunnett の t 検定
GLM, 71
Dunnett の T3
GLM, 71
Durbin-Watson の統計量
線型回帰, 119
F 統計量
線型モデル, 95
Fisher の直接法
クロス集計表, 27
Fisher の LSD
GLM, 71
Friedman の検定
複数の対応サンプルの検定, 298
Friedman の検定
対応サンプルのノンパラメトリック検定, 232
Gabriel と Howell のペアごとの比較検定
一元配置分散分析, 59
Gabriel のペアごとの比較検定
一元配置分散分析, 59
Gabriel と Howell のペアごとの比較検定
GLM, 71
Gabriel のペアごとの比較検定
GLM, 71
GLM
平方和, 66
行列の保存, 73
その後の検定, 71
プロファイル プロット, 70
変数の保存, 73
モデル, 66
GLM - 1 変量, 64
対比, 68
オプション, 75
GLM - 1 変量, 76
GLM - 1 変量
対比, 69, 75
推定周辺平均, 75
Goodman と Kruskal のガンマ
クロス集計表, 27
Goodman と Kruskal のタウ
クロス集計表, 27
Goodman と Kruskal のラムダ
クロス集計表, 27
Guttman のモデル
信頼性分析, 323
信頼性分析, 321
Hampel の M-推定量
探索的, 20
Helmert 対比
GLM, 68–69
Hochberg の GT2(H)
一元配置分散分析, 59
Hochberg の GT2
GLM, 71
Hodge-Lehman の推定
対応サンプルのノンパラメトリック検定, 232
Hotelling の T^2
信頼性分析, 323
信頼性分析, 321
Huber の M-推定量
探索的, 20
ICC。「級内相関係数」を参照, 323
Jeffreys 区間
1 サンプルのノンパラメトリック検定, 219
k および特徴選択
最近隣分析, 163
k の選択
最近隣分析, 162
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, 232
Kendall の W
複数の対応サンプルの検定, 298
Kendall のタウ b
2 変量の相関分析, 78
クロス集計表, 27
Kendall のタウ c, 27
クロス集計表, 27
Kolmogorov-Smirnov 検定
1 サンプルのノンパラメトリック検定, 221
1 サンプルのノンパラメトリック検定, 218
Kolmogorov-Smirnov の Z
1 サンプルによる Kolmogorov-Smirnov 検定, 287
2 個の独立サンプルの検定, 290
KR20
信頼性分析, 323
Kruskal のタウ
クロス集計表, 27
Kruskal-Wallis の H
2 個の独立サンプルの検定, 295
Kuder-Richardson 20 (KR20)
信頼性分析, 323
Lance と Williams の非類似度測定方法, 88
距離, 88
legal notices, 339
Levene の検定
探索的, 21
Levene の検定
一元配置分散分析, 62
GLM 1 変量, 75
Lilliefors の検定
探索的, 21
M-推定量(M)
探索的, 20

Mahalanobis の距離(M)
線型回帰, 116
Mahalanobis の距離(M)
判別分析, 169
Manhattan 距離
最近隣分析, 144
Mann-Whitney の U
2 個の独立サンプルの検定, 290
Mantel - Haenszel 統計量
クロス集計表, 27
McFadden の R^2 乗
順序回帰, 125
McNemar の検定
2 個の対応サンプルの検定, 292
対応サンプルのノンパラメトリック検定, 232–233
McNemar の検定
クロス集計表, 27
Minkowski の距離
距離, 88
Moses の外れ値反応検定
2 個の独立サンプルの検定, 290
Nagelkerke の R^2 乗
順序回帰, 125
Newman-Keuls
GLM, 71
OLAP キューブ, 43
表題, 48
統計量, 45
outliers
線型回帰, 114
TwoStep クラスタ分析, 186
Pearson 残差
順序回帰, 125
Pearson の相関
2 変量の相関分析, 78
クロス集計表, 27
Pearson のカイ 2 乗
順序回帰, 125
クロス集計表, 27
PLUM
順序回帰, 122
R 統計量
線型回帰, 119
R 統計量
グループの平均, 40
r の相関係数
クロス集計表, 27
r の相関係数
2 変量の相関分析, 78
R-E-G-W の F
一元配置分散分析, 59
R-E-G-W の Q
一元配置分散分析, 59
R^2
線型回帰, 119
R^2 の変化量, 119
グループの平均, 40
R2 乗
線型モデル, 99
Rao の V
判別分析, 169
R-E-G-W の Q
GLM, 71
ROC 曲線, 336
統計量と作図, 338
Ryan-Einot-Gabriel-Welsch 多重範囲
一元配置分散分析, 59
Ryan-Einot-Gabriel-Welsch 多重 F
一元配置分散分析, 59
Ryan-Einot-Gabriel-Welsch 多重 F
GLM, 71
Ryan-Einot-Gabriel-Welsch 多重範囲
GLM, 71
S ストレス
多次元尺度法, 326
Scheffé の検定
一元配置分散分析, 59
Scheffé の検定
GLM, 71
Shapiro-Wilk の検定
探索的, 21
Sidak の t 検定
一元配置分散分析, 59
Sidak の t 検定
GLM, 71
Somers の d
クロス集計表, 27
Spearman の相関係数
2 変量の相関分析, 78
Spearman の相関係数
クロス集計表, 27
Spearman-Brown の信頼性
信頼性分析, 323
Student-Newman-Keuls の検定
一元配置分散分析, 59
Student-Newman-Keuls の検定
GLM, 71
S モデル
曲線推定, 132
t 検定
1 サンプルの t 検定, 54
GLM 1 変量, 75
独立サンプルの t 検定, 49
対応のあるサンプルの t 検定, 52
Tamhane の T2(M)
一元配置分散分析, 59

amhane の T2
GLM, 71
trademarks, 340
Tukey の b 検定
一元配置分散分析, 59
Tukey の HSD 検定
一元配置分散分析, 59
Tukey の加法性の検定(K)
信頼性分析, 323
信頼性分析, 321
Tukey のバイウェイト推定量
探索的, 20
Tukey の HSD 検定
GLM, 71
Tukey の b 検定
GLM, 71
TwoStep クラスタ分析, 183
統計量, 188
オプション, 186
作業ファイルに保存, 188
V
クロス集計表, 27
Wald-Wolfowitz のラン(W)
2 個の独立サンプルの検定, 290
Waller-Duncan の t 検定
一元配置分散分析, 59
Waller-Duncan の t 検定
GLM, 71
Welch 統計量
一元配置分散分析, 62
Wilcoxon の符号付き順位検定
1 サンプルのノンパラメトリック検定, 218
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, 232
Wilcoxon の符号付き順位検定
2 個の対応サンプルの検定, 292
Wilks のラムダ(W)
判別分析, 169
Yates の連続修正
クロス集計表, 27
Z 得点
記述統計, 14
変数として保存, 14

アルファ因子法, 176
アルファ係数
信頼性分析, 323
アルファ係数
信頼性分析, 321
アンサンブル
線型モデル, 97
イメージ因子法, 176
濃度インデックス
比率統計量, 334
イータ
グループの平均, 40
クロス集計表, 27
イータの 2 乗
GLM 1 変量, 75
グループの平均, 40

エカマックス回転
因子分析, 178

直接オブリミン回転
因子分析, 178
重み付き予測値
GLM, 73
オーバーフィット防止基準
線型モデル, 95

カイ 2 乗, 264
期待値, 265–266
期待範囲, 265
1 サンプル検定, 264
オプション, 266
カイ 2 乗距離
距離, 88
1 サンプルのノンパラメトリック検定, 220
カイ 2 乗
独立性, 27
Fisher の直接法, 27
Pearson, 27
Yates の連続修正, 27
クロス集計表, 27
線型と線型による連関, 27
カイ 2 乗検定
1 サンプルのノンパラメトリック検定, 218
尤度比カイ 2 乗
順序回帰, 125
尤度比カイ 2 乗
クロス集計表, 27
階層的分解法, 67
カッパ
クロス集計表, 27
参照カテゴリ
GLM, 68–69
カテゴリ フィールド情報
ノンパラメトリック検定, 259
ガンマ
クロス集計表, 27

逆数モデル
曲線推定, 132
曲線推定, 130
残差の保存, 133
定数を含む, 130
分散分析, 130
モデル, 132
予測, 133
予測区間の保存, 133
予測値の保存, 133

クォーティマックス回転
因子分析, 178
クラスタ分析
効率, 211
階層クラスタ分析, 203
大規模ファイルのクラスタ分析, 209
階層クラスタ分析, 203
例, 203
統計量, 203, 205
距離測度, 205
変換測定方法, 205
クラスタ化の方法, 205
ケースのクラスタ化, 203
類似度の測定方法, 205
変数のクラスタ化, 203
階層クラスタ分析
統計量, 206
距離行列, 206
クラスタ凝集経過工程, 206
所属クラスタ, 206–207
コマンドの追加機能, 208
つららプロット, 207
デンドログラム, 207
新変数の保存, 207
プロット方向, 207
クラスタ ビューア
概要, 190, 200
クラスタ サイズ ビュー, 197
クラスタ ビューア, 192
クラスタ モデルについて, 189
クラスタとフィールドを入れ替え, 193
クラスタとフィールドを入れ替える, 193
クラスタの比較, 199
クラスタのサイズ, 197
クラスタの比較ビュー, 199
クラスタ予測値の重要度ビュー, 196
クラスタの並べ替え, 194
クラスタ表示の並べ替え, 194
セルの分布, 198
セル内容の表示, 194
セルの内容の並べ替え, 194
セルの分布ビュー, 198
予測値の重要度, 196
要約ビュー, 191
基本ビュー, 195
フィールドの並べ替え, 194
フィールド表示の並べ替え, 194
モデルの要約, 191
レコードのフィルタリング, 202
クラスタの度数
TwoStep クラスタ分析, 188
クラスタリング, 189
全体表示, 190
クラスタの表示, 190
手続きの選択, 181
棒グラフ
度数分布表, 12
グループ平均, 43
グループ中央値
OLAP キューブ, 45
グループ中央値
グループの平均, 40
グループ平均, 38
グループ中央値
要約, 36
多重回答グループ
コードブック, 1
グループの比較
OLAP キューブ, 47
グループ間の差分
OLAP キューブ, 47
多重回答グループを定義, 301
2 分変数, 301
カテゴリ, 301
変数グループの名前, 301
変数グループのラベル, 301
クロス集計表
セル表示, 30
クロス表
クロス集計表, 24
クロス集計表, 24
層, 26
書式, 32
制御変数, 26
表の抑制, 24
クロス集計表
統計量, 27
クラスタ棒グラフ, 26
クロス表
多重回答, 304

欠損値
多重回答のクロス集計表, 306
多重回答の度数表, 302

ケース選択変数
　線型回帰, 114
削除ケース残差
　線型回帰, 116
ケースごとの図
　線型回帰, 119
ケースコントロール研究
　対応のあるサンプルの t 検定, 52
ケースの数
　OLAP キューブ, 45
　グループの平均, 40
ケースの数
　要約, 36
ケースの一覧を表示, 33

交互作用項, 67, 128
項目の構築, 67, 128
コードブック, 1
　出力, 3
　統計量, 6

最小有意差
　GLM, 71
サイズの差異測定
　距離, 88
効果サイズの推定
　GLM 1 変量, 75
削除ケース残差
　GLM, 73
サブグループ平均, 43
サブグループ平均, 38
等質サブセット
　ノンパラメトリック検定, 262
　線型モデル, 95
残差
　曲線推定で保存, 133
対応サンプル, 292, 297
　最近隣分析, 147
独立サンプル検定
　ノンパラメトリック検定, 251
対応サンプルのノンパラメトリック検定
　Cochran の Q 検定, 234
　McNemar の検定, 233
独立サンプルのノンパラメトリック検定, 224, 229
　フィールド, 231
　［フィールド］タブ, 225

初期しきい値
　TwoStep クラスタ分析, 186
時系列分析
　ケースの予測, 133
　予測, 133

指数成長モデル
　曲線推定, 132
標準化した値
　記述統計, 14
一般化した最小 2 乗法
　因子分析, 176
独立したサンプルの t 検定, 49
　信頼区間, 52
　欠損値, 52
　文字型変数, 51
　オプション, 52
　グループ化変数, 51
　グループの定義, 51

スチューデント t 検定, 49
スチューデント化された残差
　線型回帰, 116
ステップワイズ法
　線型回帰, 113
変数増加ステップワイズ法
　線型モデル, 95
ストレス
　多次元尺度法, 326
図表
　ケースのラベル, 130
すべての因子によるモデル
　GLM, 66
一致するペアの研究
　対応のあるサンプルの t 検定, 52

成長モデル
　曲線推定, 132
線型モデル
　曲線推定, 132

その後の多重比較, 59

対数モデル
　曲線推定, 132
タウ b
　クロス集計表, 27
タウ c
　クロス集計表, 27
多重回答
　コマンドの追加機能, 307
多重回答のクロス集計表, 304
　値の範囲の定義, 306
　回答に基づいたパーセント, 306
　グループ間で変数を順に整合, 306
　ケースに基づいたパーセント, 306
　欠損値, 306
　セルのパーセント, 306

多重回答の度数表, 302
　欠損値, 302
多重回答分析
　クロス表, 304
　多重回答のクロス集計表, 304
　多重回答の度数表, 302
　度数分布表, 302

つららプロット
　階層クラスタ分析, 207
ツリーの深さ
　TwoStep クラスタ分析, 186

てこ比の値
　線型回帰, 116
てこ比の値
　GLM, 73
デンドログラム
　階層クラスタ分析, 207
自動データ準備
　線型モデル, 100
分類テーブル
　最近隣分析, 164

水準と広がりの図
　探索的, 21
水準と広がりの図
　GLM 1 変量, 75
線型と線型による連関
　クロス集計表, 27
全平均との対比
　GLM, 68
全平均との対比
　GLM, 69

列の集計報告書, 315
仮説の要約
　ノンパラメトリック検定, 237
誤差の集計
　最近隣分析, 164
　OLAP キューブ, 47
歪度の標準誤差
　OLAP キューブ, 45
分布の測定
　度数分布表, 10
　順序回帰, 125
分布の測定
　記述統計, 16
変数の重要度
　最近隣分析, 158
歪度の標準誤差
　要約, 36
　グループの平均, 40
共線性の診断
　線型回帰, 119
　グループの平均, 40
折半法の信頼性
　信頼性分析, 323
　クロス集計表, 30
類似度の測定方法
　距離, 89
平均値の標準誤差
　OLAP キューブ, 45
　グループの平均, 40
多項式の対比
　一元配置分散分析, 58
　GLM, 68–69
類似度の測定方法
　階層クラスタ分析, 205
中心傾向の測定
　度数分布表, 10
等分散性の検定
　一元配置分散分析, 62
中心傾向の測定
　探索的, 20
中心傾向の測定
　比率統計量, 334
独立性の検定
　カイ 2 乗, 27
最近隣の距離
　最近隣分析, 159
　OLAP キューブ, 47
平均値の標準誤差
　要約, 36
平行線の検定
　順序回帰, 125
折半法の信頼性
　信頼性分析, 321
信頼区間の要約
　ノンパラメトリック検定, 239, 245
等分散性の検定
　GLM 1 変量, 75
複数の回帰
　線型回帰, 111
予測値の重要度
　線型モデル, 101
対応のあるサンプルの t 検定, 52
　欠損値, 53
　オプション, 53
　対応のある変数の選択, 52
ノイズ処理
　TwoStep クラスタ分析, 186
複数の対応サンプルの検定, 297
　統計量, 297, 299
　オプション, 297
　グループ化変数, 296

コマンドの追加機能, 297, 299
検定の種類, 296, 298
複数の独立サンプルの検定, 295
報告書の行の集計, 308
欠損値, 312
ソート順, 308
データ列, 308
列の書式, 310
ブレーク列, 308
ブレーク間隔, 311
ページの指定, 311
ページの番号付け, 312
報告書の列の集計, 315
表題, 314
表題の変数, 314
フッター, 314
ページの設計, 313
報告書の列の集計
総計, 319
欠損値, 319
コマンドの追加機能, 319
ページの番号付け, 319
報告書の列の集計
小計, 318
集計列, 317
ページの指定, 318
ノンパラメトリック検定
1サンプルによる Kolmogorov-Smirnov 検定, 287
2 個の独立サンプルの検定, 289
2 個の対応サンプルの検定, 292
複数の独立サンプルの検定, 295, 297
ラン検定, 285
ノンパラメトリック検定
1サンプルによる Kolmogorov-Smirnov 検定, 287
カイ 2 乗, 264
モデル ビュー, 236
ラン検定, 285

バギング
線型モデル, 93
パターン行列
因子分析, 173
パターンの差異測定
距離, 88
パラメータ推定値
順序回帰, 125
パラメータ推定値
GLM 1 変量, 75
バリマックス回転
因子分析, 178
パーセンタイル
度数分布表, 10
探索的, 20
パーセント
クロス集計表, 30
行パーセント
クロス集計表, 30
合計パーセント
クロス集計表, 30

箱ひげ図
探索的, 21
変数の比較, 21
因子の水準の比較, 21
ヒストグラム
度数分布表, 12
探索的, 21
線型回帰, 114
標準化されていない残差
GLM, 73
標準化残差
GLM, 73
標準誤差
GLM, 73

ファイ
クロス集計表, 27
ファイ 2 乗距離
距離, 88
大規模ファイルのクラスタ分析
例, 209
方法, 209, 211
統計量, 213
統計量, 209
収束基準, 211
クラスタ距離, 212
所属クラスタ, 212
クラスタ情報の保存, 212
コマンドの追加機能, 213
連続型フィールド情報
ノンパラメトリック検定, 260
複合成長モデル
曲線推定, 132
ブロック距離
距離, 88
都市ブロック距離
最近隣分析, 144
残差プロット
GLM 1 変量, 75
偏相関プロット
線型回帰, 114
正規確率プロット
探索的, 21

傾向化除去正規プロット
　探索的, 21
正規確率プロット
　線型回帰, 114
プロットのロード
　因子分析, 178
プロファイル プロット
　GLM, 70
分散共分散行列
　GLM, 73
分散分析
　曲線推定, 130
ブースティング
　線型モデル, 93

ペアごとの比較
　ノンパラメトリック検定, 261
平方和, 67
べき乗推定値
　GLM 1 変量, 75
べき乗モデル
　曲線推定, 132
ページの指定
　行の集計報告書, 312
ページの指定
　列の集計報告書, 318
ページの番号付け
　行の集計報告書, 312
ページの番号付け
　列の集計報告書, 319

ホールドアウト サンプル
　最近隣分析, 147

調整済み R^2
　線型回帰, 119
調整済み R2 乗
　線型モデル, 95
重み付き平均値
　比率統計量, 334
重み付き最小 2 乗法
　線型回帰, 111
重み付けのない最小 2 乗法
　因子分析, 176

メディアン検定
　2 個の独立サンプルの検定, 295
メモリ割り当て
　TwoStep クラスタ分析, 186

線型モデル, 91
　目的, 93
　順序回帰, 126, 128
　信頼性分析, 323
　信頼性分析, 321
　情報量基準, 99
　R2 乗統計量, 99
　アンサンブル, 97
　自動データ準備, 94
　モデル選択, 95
　モデル オプション, 98
　モデルの要約, 99
　結合ルール, 97
　確信度レベル, 94
　結果を複製, 98
厳密平行モデル
　信頼性分析, 323
　信頼性分析, 321
線型モデル
　係数, 107
　推定平均値, 109
　残差, 103
　予測対観測, 102
　分散分析表(A), 105
　outliers, 104
　自動データ準備, 100
　予測値の重要度, 101
　モデル構築の要約, 110
モデル ビュー
　最近隣分析, 153
　ノンパラメトリック検定, 236
モード
　度数分布表, 10

ユークリッド距離
　最近隣分析, 144
ユークリッド距離
　距離, 88
平方ユークリッド距離
　距離, 88
ユーザーの指定によるモデル
　GLM, 66

予測
　曲線推定, 133
予測区間
　曲線推定で保存, 133
予測値
　曲線推定で保存, 133

散らばり係数 (COD)
　比率統計量, 334
散らばりの測定
　度数分布表, 10

散らばりの測定
　記述統計, 16
散らばりの測定
　比率統計量, 334
散らばりの測定
　探索的, 20
ラムダ
　クロス集計表, 27
ラン検定
　分割点, 285–287
　1 サンプルのノンパラメトリック検定, 222
　1 サンプルのノンパラメトリック検定, 218
　オプション, 287
　コマンドの追加機能, 287

リスク
　クロス集計表, 27
相対リスク
　クロス集計表, 27
リンク
　順序回帰, 123

結合ルール
　線型モデル, 97

外れ値
　探索的, 20

ロジスティック モデル
　曲線推定, 132
ロー
　2 変量の相関分析, 78
　クロス集計表, 27