

**IBM SPSS Advanced
Statistics 23**

IBM

お願い

本書および本書で紹介する製品をご使用になる前に、103 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM SPSS Statistics バージョン 23 リリース 0 モディフィケーション 0 および新しい版で明記されない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

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

原典： IBM SPSS Advanced Statistics 23

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

第 1 章 Advanced Statistics の概要 . . . 1

第 2 章 GLM 多変量解析 3

GLM 多変量のモデル	4
項の構築	4
平方和	5
GLM 多変量の対比	5
対比の種類	6
GLM 多変量のプロファイル・プロット	6
GLM 多変量におけるその後の比較	7
GLM の保存	8
GLM 多変量のオプション	9
GLM コマンドの追加機能	10

第 3 章 GLM 反復測定 11

GLM 反復測定の因子の定義	13
GLM 反復測定のモデル	13
項の構築	14
平方和	14
GLM 反復測定の対比	15
対比の種類	15
GLM 反復測定のプロファイル・プロット	16
GLM 反復測定におけるその後の比較	16
GLM 反復測定の保存	17
GLM 反復測定のオプション	18
GLM コマンドの追加機能	19

第 4 章 分散成分分析 21

分散成分のモデル	22
項の構築	22
分散成分のオプション	22
平方和 (分散成分)	23
分散成分の新規ファイルへの保存	23
VARCOMP コマンドの追加機能	24

第 5 章 線型混合モデル 25

線型混合モデルの被験者/反復変数の選択	26
線型混合モデルの固定効果	27
ネストなし項目の構築	27
入れ子項目の構築	27
平方和	27
線型混合モデルのランダム効果	28
線型混合モデルの推定	29
線型混合モデルの統計	29
線型混合モデルの EM 平均	30
線型混合モデルの保存	30
MIXED コマンドの追加機能	31

第 6 章 一般化線型モデル 33

一般化線型モデルにおける応答	35
一般化線型モデルの参照カテゴリー	36

一般化線型モデルにおける予測	36
一般化線型モデルのオプション	36
一般化線型モデルにおけるモデル	37
一般化線型モデルにおける推定	38
一般化線型モデルにおける初期値	38
一般化線型モデルにおける統計	39
一般化線型モデルにおける EM 平均	40
一般化線型モデルにおける保存	41
一般化線型モデルにおけるエクスポート	42
GENLIN コマンドの追加機能	43

第 7 章 一般化推定方程式 45

一般化推定方程式におけるモデルの種類	47
一般化推定方程式における応答	48
一般化推定方程式の参照カテゴリー	49
一般化推定方程式における予測	49
一般化推定方程式のオプション	50
一般化推定方程式におけるモデル	50
一般化推定方程式における推定	51
一般化推定方程式の初期値	52
一般化推定方程式における統計	53
一般化推定方程式における EM 平均	54
一般化推定方程式における保存	54
一般化推定方程式におけるエクスポート	55
GENLIN コマンドの追加機能	56

第 8 章 一般化線型混合モデル 57

一般化線型混合モデルの取得	58
対象	58
固定効果	60
カスタム項目の追加	61
ランダム効果	62
ランダム効果ブロック	62
重み付けとオフセット	63
一般的な作成オプション	63
推定	64
推定平均	65
保存	65
モデル・ビュー	66
モデルの要約	66
データ構造	67
予測と観測	67
分類	67
固定効果	67
固定係数	68
ランダム効果共分散	68
共分散パラメーター	69
推定平均: 有意効果	69
推定平均: カスタム効果	69

第 9 章 モデル選択対数線型分析	71
対数線型分析の範囲の定義	71
対数線型分析のモデル	72
項の構築	72
モデル選択対数線型分析のオプション	72
HILOGLINEAR コマンドの追加機能	72
第 10 章 一般対数線型分析	73
一般対数線型分析のモデル	74
項の構築	74
一般対数線型分析のオプション	74
一般対数線型分析の保存	75
GENLOG コマンドの追加機能	75
第 11 章 ロジット対数線型分析	77
ロジット対数線型分析のモデル	78
項の構築	78
ロジット対数線型分析のオプション	79
ロジット対数線型分析の保存	79
GENLOG コマンドの追加機能	79
第 12 章 生命表	81
生命表の状態変数の事象の定義	82
生命表の範囲の定義	82
生命表のオプション	82
SURVIVAL コマンドの追加機能	83
第 13 章 Kaplan-Meier 生存分析	85
Kaplan-Meier の状態変数のイベントの定義	86
Kaplan-Meier の因子レベルの比較	86
Kaplan-Meier の新規変数の保存	86
Kaplan-Meier のオプション	87

KM コマンドの追加機能	87
第 14 章 Cox 回帰分析	89
Cox 回帰におけるカテゴリー変数の定義	90
Cox 回帰のプロット	90
Cox 回帰における新規変数の保存	91
Cox 回帰のオプション	91
Cox 回帰における状態変数のイベントの定義	92
COXREG コマンドの追加機能	92
第 15 章 時間依存の共変量の計算	93
時間依存の共変量の計算	93
時間依存の共変量を使用する Cox 回帰の追加機能	94
第 16 章 カテゴリー変数のコード化方式 95	
偏差	95
単純	95
Helmert	96
差	96
多項式	96
反復	97
特殊	97
インジケーター	98
第 17 章 共分散構造	99
特記事項	103
商標	105
索引	107

第 1 章 Advanced Statistics の概要

Advanced Statistics オプションでは、Statistics Base オプションより高度なモデリング・オプションを提供するプロシージャーを使用できます。

- GLM 多変量は、GLM 1 変量で提供される一般線型モデルを拡張し、複数の従属変数を使用できるようにしたものです。機能が拡張された GLM 反復測定では、複数の従属変数の反復測定が可能です。
- 分散成分分析とは、従属変数の変動を固定成分とランダム成分に分解する特定のツールです。
- 線型混合モデルは、相関および非定常変動が見られるデータも扱えるように、一般線型モデルを拡張したものです。したがって、線型混合モデルにはデータの平均値だけでなく、分散および共分散をモデリングできる柔軟性が備わっています。
- 一般化線型モデル (GZLM) では、誤差項に関する正規性の仮定が緩和されています。このモデルで必要とされるのは、従属変数と予測値が変換またはリンク関数を介して線型関係にあることのみです。一般化推定方程式 (GEE) を使用すると、GZLM が拡張され、反復測定が可能になります。
- 一般対数線型分析ではクロス分類された度数データにモデルを適合させることができ、モデル選択対数線型分析はモデルを選択する場合に役立ちます。
- ロジット対数線型分析では対数線型モデルを適合させて、1 つのカテゴリ-従属と 1 つ以上のカテゴリ-予測との関係を分析することができます。
- 生存分析は生命表から使用できます。これは、場合によっては因子変数のレベルごとにイベントまでの時間変数の分布を調べるためのものです。Kaplan-Meier 生存分析では、場合によっては因子変数のレベルごとにイベントまでの時間変数の分布を調べたり、層化変数のレベルごとに別の分析を生成します。また、Cox 回帰分析では、指定されたイベントまでの時間を、特定の共変量の値に基づいてモデリングします。

第 2 章 GLM 多変量解析

「GLM 多変量」プロシージャーでは、1 つ以上の因子変数または共変量を使用して、複数の従属変数の回帰分析と分散分析を行います。各因子変数により母集団がいくつかのグループに分けられます。この一般線型モデル・プロシージャーを使用することで、従属変数の結合分布のさまざまなグループ平均に対する因子変数の効果について、帰無仮説を検定できます。因子間の交互作用と個々の因子の効果を調べることができます。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) 変数が共変量として指定されます。

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同じ数のケースが含まれている場合、その計画は釣り合っています。多変量モデルの場合は、モデルの効果による平方和と誤差平方和が、1 変量の分散分析に見られるスカラー形式ではなく、行列形式になります。この行列は SSCP (平方和と交差積) 行列と呼ばれます。複数の従属変数を指定すると、各従属変数についての 1 変量の分散分析だけでなく、Pillai のトレース、Wilks のラムダ、Hotelling のトレース、Roy の最大根基準を近似 F 統計量と併用して、多変量分散分析を行うことができます。「GLM 多変量」では、仮説の検定の他に、パラメーターの推定も行います。

仮説を検定する際は、一般的に使用されている事前 対比を使用できます。さらに、全体的な F 検定で有意性が判明していれば、その後の検定を使用して、特定の平均値間の差を評価できます。推定周辺平均からはモデル内のセルの予測平均値を推定でき、これらの平均値のプロファイル・プロット (交互作用プロット) を使用して一部の関係を簡単に視覚化できます。その後の多重比較検定は従属変数ごとに個別に行われます。

残差、予測値、Cook の距離、てこ比の値は、データ・ファイルに新規変数として保存し、仮定の確認に使用できます。また、残差の平方和と交差積の正行列である残差 SSCP 行列、残差 SSCP 行列を残差の自由度で割った残差の分散共分散行列、および残差の分散共分散行列を標準化した形式である残差の相関行列も使用できます。

「WLS 重み」を使用すると、重み付き最小二乗法 (WLS) 分析の場合、観測値に異なる重みを与えるために使用する変数を指定し、それにより異なる測定精度を補正できる可能性があります。

例: プラスチック製造業者は、プラスチック・フィルムの引き裂き抵抗、光沢、および不透明度という 3 つのプロパティを測定します。2 つの比率の押し出しと、量が異なる 2 つの添加物を試し、押し出し率と添加物の量の各組み合わせについて、3 つのプロパティを測定します。ここで製造業者がわかったことは、押し出し率と添加物の量を使用して得られる結果が個々には有意だが、2 つの因子の交互作用は有意ではないということです。

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、およびタイプ IV 平方和を使用できます。タイプ III がデフォルトです。

統計: その後の範囲検定と多重比較: 最小有意差、Bonferroni、Sidak、Scheff、Ryan-Einot-Gabriel-Welsch の多重 F 、Ryan-Einot-Gabriel-Welsch の多重範囲、Student-Newman-Keuls、Tukey の HSD 検定、Tukey の b 、Duncan、Hochberg の GT2、Gabriel、Waller Duncan の t 検定、Dunnnett (片側と両側)、Tamhane の T2、Dunnnett の T3、Games-Howell、および Dunnnett 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 の等分散性検定、従属変数の共分散行列の等分散性に関する Box の M 検定、および Bartlett の球面性検定。

作図。レベルと広がり、残差、およびプロファイル (交互作用)。

GLM 多変量データの考慮事項

データ: 従属変数は量的である必要があります。因子はカテゴリ型で、数値または文字列値を持つことができます。共変量は、従属変数に関連する量的変数です。

仮定: 従属変数の場合、データは多変量正規母集団からのベクトルのランダム・サンプルです。母集団では、すべてのセルの分散共分散行列は同じです。データは対称でなければなりません。分散分析は正規性からの逸脱には頑健です。仮定の確認には、等分散性の検定 (Box の M を含む) およびレベルと広がり の図を使用できます。また、残差と残差プロットを調べることもできます。

関連プロシージャー: 分散分析を行う前にデータを調べる場合は、「探索」プロシージャーを使用します。単一の従属変数の場合は、「GLM 1 変量」を使用します。各被験者に対して同じ従属変数を何度か測定した場合は、「GLM 反復測定」を使用します。

GLM 多変量の表の取得

1. メニューから次の項目を選択します。

「分析」 > 「一般線型モデル」 > 「多変量...」

2. 少なくとも 2 つの従属変数を選択してください。

オプションとして、「固定因子」、「共変量」、および「WLS 重み」を指定することができます。

GLM 多変量のモデル

「**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。交互作用のサブセットのみを指定する場合や、因子と共変量の交互作用を指定する場合は、「**カスタム**」を選択します。モデルに含める項はすべて指示する必要があります。

「**因子と共変量**」。因子と共変量がリストされます。

「**モデル**」。モデルは、使用するデータの性質によって異なります。「**カスタム**」の選択後、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。欠損セルがない釣り合い型モデルまたは不釣り合い型モデルの場合、最も一般的に使用される方法はタイプ III 平方和です。

「**モデルに切片を含める**」。通常、切片はモデルに含まれます。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

選択した因子や共変量について、次の項を作成できます。

「**交互作用**」。選択したすべての変数の最上位交互作用項を作成します。これはデフォルトです。

「**主効果**」。選択した変数ごとに主効果の項目を作成します。

「**2 次まで**」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「**3 次まで**」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「4 次まで」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「5 次まで」。選択した変数で考えられるすべての 5 次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。タイプ III が最も一般的に使用され、デフォルトです。

タイプ I。この方法は、平方和の階層的分解法としても知られています。モデル内の各項はその前の項に対してのみ調整されます。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 次交互作用効果の前に主効果が指定され、2 次交互作用効果の前に 1 次交互作用効果が指定されているといったような分散分析の釣り合い型モデル。
- 高次項の前に低次項が指定されている多項式回帰モデル。
- 最初に指定された効果が 2 番目に指定された効果内にネストされ、2 番目に指定された効果が 3 番目に指定された効果内にネストされているような純粋なネスト・モデル。(このネスト形式はシンタックスを使用してのみ指定可能です。)

タイプ II。この方法は、他のすべての該当する効果に対して調整されたモデルの効果の平方和を計算します。該当する効果とは、調査されている効果が含まれて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。
- 純粋にネストになっているデザイン (ネストの形式は、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。この方法では、計画内の効果の平方和を、その効果を含まない他の効果に対して調整されており、その効果を含む効果 (存在する場合) に直交している平方和として計算します。タイプ III の平方和には、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多くの場合、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。デザインにおける効果 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ー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。
- 空白セルを含む釣り合い型モデルまたは不釣り合い型モデル。

GLM 多変量の対比

対比は、効果レベルが互いに有意差があるかどうかを検定する場合に使用します。対比はモデルの因子ごとに指定できます。対比は、パラメーターの線型結合を表します。

仮説検定は $LBM = 0$ という帰無仮説に基づきます。ここで、 L は対比係数行列であり、 M は単位行列 (従属変数の数と同じディメンションを持つ)、 B はパラメーターのベクトルです。対比が指定されている場合は、因子に対応する列が対比と一致するように L 行列が作成されます。残りの列は、 L 行列が推定できるように調整されます。

F 統計量および Bonferroni タイプの同時信頼区間 (すべての従属変数に関する対比の差異についてのスチューデント t 分布に基づく) を使用する 1 変量検定に加え、Pillai のトレース、Wilks のラムダ、Hotelling のトレース、Roy の最大根基準を使用する多変量検定を行うことができます。

使用できる対比には、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、参照カテゴリーを最後のカテゴリーにするか最初のカテゴリーにするかを選択できます。

対比の種類

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

「**単純**」。各レベルの平均を指定されたレベルの平均と比較します。このタイプの対比は、制御グループがある場合に役立ちます。最初または最後のカテゴリーを参照として選択できます。

「**差分**」。各レベル (最初を除く) の平均を前のレベルの平均と比較します。(逆 Helmert 対比と呼ばれることもあります。)

「**Helmert**」。因子の各レベル (最終を除く) の平均を後続レベルの平均と比較します。

「**反復**」。各レベル (最後を除く) の平均を後続レベルの平均と比較します。

「**多項式**」。1 次効果、2 次効果、3 次効果などを比較します。第 1 自由度にはすべてのカテゴリーにおける 1 次効果が含まれ、第 2 自由度には 2 次効果が含まれ、以下同様に続きます。このような対比は、多項式トレンドを推定する場合に頻繁に使用されます。

GLM 多変量のプロファイル・プロット

プロファイル・プロット (交互作用プロット) は、モデルの周辺平均を比較する場合に役立ちます。プロファイル・プロットは、因子の 1 レベルにおける従属変数の推定周辺平均 (共変量に対して調整される) を各点で示す折れ線グラフです。第 2 因子のレベルを使用して、個別の線を作成することができます。第 3 因子の各レベルを使用して、個別のプロットを作成できます。すべての因子をプロットで使用することができます。プロファイル・プロットは従属変数ごとに作成されます。

1 つの因子のプロファイル・プロットには、推定周辺平均がレベル全体で増えているか減っているかが示されます。2 つ以上の因子の場合、因子間に交互作用がないことが平行線で示されます。これは、1 つの因子のみのレベルしか調べることができないことを意味します。非平行線は交互作用を示します。

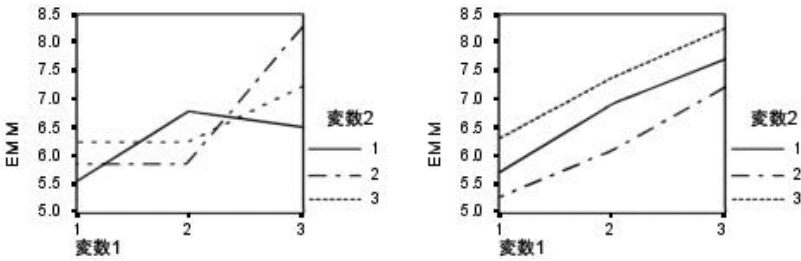


図1. 非平行プロット (左) と平行プロット (右)

横軸の因子を選択し、オプションとして、個別の線と個別のプロットの因子を選択してプロットを指定したら、そのプロットを「作図」リストに追加する必要があります。

GLM 多変量におけるその後の比較

その後の多重比較検定。 平均値に差があることが判明した後で、その後の範囲検定とペアごとの多重比較でどの平均値が異なるのかを判別することができます。比較は未調整値に対して行われます。その後の検定は、従属変数ごとに個別に実行されます。

Bonferroni 検定と Tukey の HSD 検定が一般的に使用される多重比較検定です。**Bonferroni 検定** (スチューデントの t 統計に基づく) では、多重比較が行われるという事実に応じて観測された有意レベルが調整されます。**Sidak の検定**でも有意レベルが調整され、限界が Bonferroni 検定の場合よりもより厳しくなります。**Tukey の HSD 検定**ではスチューデント化された範囲の統計量を使用して、グループ間ですべてのペアごとの比較を行い、実験ごとの誤差率をすべてのペアごとの比較の集合の誤差率に設定します。多数の平均値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。少数のペアの場合は Bonferroni の方が有効です。

Hochberg の GT2 は Tukey の HSD 検定に似ていますが、studentized maximum modulus が使用されます。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**でも studentized maximum modulus を使用しますが、一般的に、セル・サイズが等しくない場合は Hochberg の GT2 より有効です。セル・サイズのばらつきが大きい場合は、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、一連の処理を単一の制御平均値と比較します。最後のカテゴリーがデフォルトの制御カテゴリーです。代わりに、最初のカテゴリーを選択することもできます。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 (制御カテゴリーを除く) の平均値が制御カテゴリーのものと同しくないことを検定するには、両側検定を使用します。因子の任意のレベルの平均値が制御カテゴリーのものより小さいかどうかを検定するには、「< 対照カテゴリー」を選択します。同様に、因子の任意のレベルの平均値が制御カテゴリーのものより大きいかどうかを検定するには、「> 対照カテゴリー」を選択します。

Ryan, Einot, Gabriel, および Welsch (R-E-G-W) は 2 つの多重ステップダウン範囲検定を開発しました。多重ステップダウン・プロシージャでは、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い場合は、平均値のサブセットが等しいかどうかを検定されます。**R-E-G-W F** は F 検定に基づき、**R-E-G-W Q** はスチューデント化された範囲に基づきます。これらの検定は、Duncan の多重範囲検定および Student-Newman-Keuls (これも多重ステップダウン・プロシージャです) より有効ですが、セル・サイズが等しくない場合はお勧めできません。

分散が等しくない場合は、**Tamhane の T2** (t 検定に基づく控えめなペアごとの比較検定)、**Dunnnett の T3** (studentized maximum modulus 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** (これが公平な場合もある)、または **Dunnnett の C** (スチューデント化された範囲に基づくペアごとの比較検定) を使用します。

Duncan の多重範囲検定、Student-Newman-Keuls (**S-N-K**)、および **Tukey の b** は、グループ平均の順位付けおよび範囲値の計算を行う範囲検定です。これらの検定は、前述の検定ほど頻繁には使用されません。

Waller-Duncan の t 検定では Bayesian のアプローチを使用します。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用されます。

Scheffé 検定の有意レベルは、この機能で使用可能なペアごとの比較だけでなく、グループ平均で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。その結果、Scheffé 検定は多くの場合、他の検定より控えめになるため、有意確率では平均値間の差がより大きくなければなりません。

最小有意差 (**LSD**) のペアごとの多重比較検定は、グループのすべてのペア間の多重個別 t 検定と同等です。この検定の欠点は、観測された有意レベルを多重比較用に調整する試みが行われなことです。

表示される検定。ペアごとの比較は、LSD、Sidak、Bonferroni、Games-Howell、Tamhane の T2 と T3、Dunnnett の C、および Dunnnett の T3 用に提供されます。範囲検定の等分散性サブセットは、S-N-K、Tukey の b 、Duncan、R-E-G-W F 、R-E-G-W Q 、および Waller 用に提供されます。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

モデルが予測した値、残差、および関連測定値はデータ・エディターで新規変数として保存できます。これらの変数の多くは、データに関する仮定を調べるために使用できます。値を保存して別の IBM® SPSS® Statistics セッションで使用するには、現在のデータ・ファイルを保存する必要があります。

「**予測値**」。モデルがケースごとに予測する値。

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

「**診断**」。独立変数の一般的でない値の組み合わせのケースと、モデルに大きな影響を与える可能性のあるケースを識別するための測定方法。

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

「**残差**」。非標準化残差は、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。標準化残差、スチューデント化された残差、および削除済み残差も使用できます。WLS 変数を選択すると、重み付けした非標準化残差が使用できます。

- 「**非標準化**」。観測した値と、モデルによって予測された値との差。
- 「**重み付き**」。重み付きの、標準化していない残差。既に WLS 変数を選択している場合に限り使用することができます。

- 「標準化」。残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。平均値は 0 であり、標準偏差は 1 です。
- 「スチューデント化」。独立変数の各ケースの値と独立変数の平均値との距離に応じて、ケースごとに異なる標準偏差の推定量で残差を割ったもの。
- 「削除済み」。ケースを回帰係数の計算から除外した場合のそのケースの残差。従属変数の値と調整済み予測値の差です。

「係数の統計量」。モデルにおけるパラメーター推定値の分散共分散行列を、現在のセッションにある新しいデータ・セット、または IBM SPSS Statistics の外部データ ファイルに書き込みます。また、それぞれの従属変数に対しても、パラメーター推定値の行、パラメーター推定値に対応する t 統計量の有意確率値の行、および残差自由度の行があります。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取る別のプロシージャで、この行列ファイルを使用できます。

GLM 多変量のオプション

このダイアログ・ボックスでオプションの統計を使用できます。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」。セルにおける母集団周辺平均の推定値が必要な因子と交互作用を選択します。これらの平均は共変量 (存在する場合) に対して調整されます。交互作用は、カスタム・モデルを指定している場合にのみ、使用できます。

- 「主効果の比較」。被験者間因子と被験者内因子の両方について、モデル内の主効果に対する推定周辺平均間で無修正のペアワイズ比較を行います。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のみ使用できます。
- 「信頼区間の調整」。最小有意差 (LSD)、Bonferroni、または Sidak の信頼区間と有意性に対する調整を選択します。この項目は、「主効果の比較」が選択されている場合にのみ使用できます。

「表示」。「記述統計」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、および度数が生成されます。「効果サイズの推定値」では、各効果および各パラメーター推定値の偏イータ 2 乗値が示されます。イータの 2 乗統計量は、因子に起因する総変動の比率を記述するものです。観測値に基づいて対立仮説を設定する際に検定力を取得するには、「観測検定力」を選択します。パラメーター推定値、標準誤差、 t 検定、信頼区間、および各検定の観測検定力を生成するには、「パラメータ推定値」を選択します。仮説、誤差「SSCP 行列」、「残差 SSCP 行列」に加え、残差分散共分散行列の Bartlett の球面性検定を表示することができます。

「等分散性の検定」では、被験者間因子の場合にのみ、被験者間因子のレベルの組み合わせ全体の従属変数ごとに、Levene の等分散性検定を行います。この等分散性の検定には、被験者間因子のすべてのレベルの組み合わせにおける従属変数の共分散行列の等分散性に関する Box の M 検定も含まれます。レベルと広がりの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確認に役立ちます。この項目は、因子が存在しない場合は無効になります。各従属変数について、標準化された残差による予測値で観測された値の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を調べる場合に役立ちます。従属変数と独立変数との関係がモデルで適切に記述できるかどうかを確認するには、「不適合度検定」を選択します。「一般の推定可能関数」を使用すれば、一般推定可能関数に基づくカスタム仮説検定を構成できます。任意の対比係数行列における行は、一般推定可能関数の線型結合です。

「有意レベル」。その後の検定で使用される有意レベルと、信頼区間を構成するために使用される信頼度レベルを調整できます。指定された値は、検定の観測検定力の計算にも使用できます。有意レベルの指定時に、信頼区間の関連レベルがダイアログ・ボックスに表示されます。

GLM コマンドの追加機能

これらの機能は、1 変量、多変量、反復測定分析に適用できます。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 計画のネスト効果の指定 (DESIGN サブコマンドを使用)。
- 効果の検定、または効果や値の線型の組み合わせの検定の指定 (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)。
- カスタムの **L** 行列、**M** 行列、または **K** 行列の作成 (LMATRIX、MMATRIX、KMATRIX サブコマンドを使用)。
- 偏差対比または単純対比での中間参照カテゴリーの指定 (CONTRAST サブコマンドを使用)。
- 多項式の対比の行列を指定 (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)。
- 任意の因子に対する推定周辺平均、または因子リスト内の因子間の因子交互作用の計算 (EMMEANS サブコマンドを使用)。
- 一時変数の名前の指定 (SAVE サブコマンドを使用)。
- 相関行列データ・ファイルの作成 (OUTFILE サブコマンドを使用)。
- 被験者間分散分析テーブルから取得した統計を含む行列データ・ファイルの作成 (OUTFILE サブコマンドを使用)。
- 新規データ・ファイルへのデザイン行列の保存 (OUTFILE サブコマンドを使用)。

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

第 3 章 GLM 反復測定

「GLM 反復測定」プロシージャーでは、各被験者またはケースに対して同じ測定を数回行う場合に分散分析を行います。被験者間因子を指定すると、母集団がいくつかのグループに分けられます。この一般線型モデル・プロシージャーを使用して、被験者間因子と被験者内因子の両方の効果について、帰無仮説を検定できます。因子間の交互作用と個々の因子の効果を調べることができます。さらに、定数共変量の効果と被験者間因子と共変量の交互作用を含めることができます。

2 重多変量の反復測定計画の場合、従属変数は、被験者内因子のさまざまなレベルに対する複数の変数測定を表します。例えば、被験者ごとに 3 つの異なる時間に脈と呼吸の両方を測定できます。

「GLM 反復測定」プロシージャーでは、反復測定データに対する 1 変量と多変量の両方の分析を行います。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同じ数のケースが含まれている場合、その計画は釣り合っています。多変量モデルの場合は、モデルの効果による平方和と誤差平方和が、1 変量の分散分析に見られるスカラー形式ではなく、行列形式になります。この行列は SSCP (平方和と交差積) 行列と呼ばれます。「GLM 反復測定」では、仮説の検定の他に、パラメーターの推定も行います。

被験者間因子の仮説検定には、一般的に使用されている事前 対比を使用できます。さらに、全体的な F 検定で有意性が判明していれば、その後の検定を使用して、特定の平均値間の差を評価できます。推定周辺平均からはモデル内のセルの予測平均値を推定でき、これらの平均値のプロファイル・プロット (交互作用プロット) を使用して一部の関係を簡単に視覚化できます。

残差、予測値、Cook の距離、てこ比の値は、データ・ファイルに新規変数として保存し、仮定の確認に使用できます。また、残差の平方和と交差積の正方向行列である残差 SSCP 行列、残差 SSCP 行列を残差の自由度で割った残差の分散共分散行列、および残差の分散共分散行列を標準化した形式である残差の相関行列も使用できます。

「WLS 重み」を使用すると、重み付き最小二乗法 (WLS) 分析の場合、観測値に異なる重みを与えるために使用する変数を指定し、それにより異なる測定精度を補正できる可能性があります。

例: 12 人の生徒が不安度検定でのスコアに基づいて、高い不安度グループまたは低い不安度グループに割り当てられます。不安度は、被験者をグループ分けすることから、「被験者間因子」と呼ばれます。生徒にはそれぞれ学習課題に関する 4 つの試行が与えられ、各試行ごとのエラーの数が記録されます。各試行のエラーは別々の変数に記録され、被験者内因子 (試行) は 4 つの試行に対して 4 つのレベルで定義されません。試行の効果は有意であることがわかりますが、不安度別試行の交互作用は有意ではありません。

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、およびタイプ IV 平方和を使用できます。タイプ III がデフォルトです。

統計: その後の範囲検定と多重比較 (被験者間因子について): 最小有意差、Bonferroni、Sidak、Scheffé、Ryan-Einot-Gabriel-Welsch の多重 F 、Ryan-Einot-Gabriel-Welsch の多重範囲、Student-Newman-Keuls、Tukey の HSD 検定、Tukey の b 、Duncan、Hochberg の GT2、Gabriel、Waller Duncan の t 検定、Dunnett (片側と両側)、Tamhane の T2、Dunnett の T3、Games-Howell、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 の等分散性検定、Box の M 、および Mauchly の球面性検定。

作図: レベルと広がり、残差、およびプロファイル (交互作用)。

GLM 反復測定データの考慮事項

データ: 従属変数は量的である必要があります。被験者間因子では、男性と女性など、サンプルが個別のサブグループに分けられます。このような因子はカテゴリー型であり、数値または文字列値を持つことができます。被験者内因子は、「反復測定」ダイアログ・ボックスで定義されます。共変量は、従属変数に関連する量的変数です。反復測定分析の場合、これらは被験者内変数のそれぞれのレベルで常に一定でなければなりません。

データ・ファイルには、被験者の測定グループごとの変数セットが含まれている必要があります。このセットには、グループ内の測定の反復ごとに 1 つの変数があります。被験者内因子は、そのグループを対象に、反復回数に等しいレベル数で定義されます。例えば、体重測定を複数の異なる日に行うことができます。同じプロパティの測定が 5 日間で行われた場合、被験者内因子は 5 つのレベルで **日数** として指定できます。

被験者内因子が複数ある場合は、被験者ごとの測定回数が各因子のレベル数の積と等しくなります。例えば、1 日 3 回の測定を 4 日行くと、合計測定数は被験者ごとに 12 回となります。被験者内因子は、**日数 (4)** と **回数 (3)** で指定できます。

仮定: 反復測定分析は、1 変量と多変量という 2 とおりの方法で行うことができます。

1 変量のアプローチ (分割プロットによるアプローチ、または混合モデルによるアプローチという名でも知られています) では、従属変数を被験者内因子のレベルに対する応答数と見なします。被験者の測定値は、多変量の正規分布からのサンプルでなければならず、分散共分散行列は、被験者間効果によって形成されるセル全体で同じです。従属変数の分散共分散行列で、特定の仮定が立てられます。分散共分散行列が循環形式である場合、1 変量アプローチに使用された F 統計量の妥当性は確実です (Huynh および Mandeville, 1979)。

この仮定の検定には、Mauchly の球面性の検定を使用できます。この検定では、正規直交型の変換従属変数の分散共分散行列について球面性の検定を行います。反復測定分析の場合は、自動的に Mauchly の検定が表示されます。小さいサンプル・サイズの場合、この検定はそれほど有効ではありません。大きいサンプル・サイズでは、この検定は、結果に対する逸脱の影響が小さくても有意な場合があります。検定の有意性が大きい場合は、球面性の仮説を前提とすることができます。ただし、有意性が小さく、球面性の仮定に反しているような場合には、自由度の分子と分母を調整して、1 変量 F 統計量を確認できます。「GLM 反復測定」プロシージャでは、**イプシロン**という、この調整の 3 つの推定値を使用できます。自由度の分子と分母の両方にイプシロンを掛ける必要があります、 F 比の有意性は新しい自由度を使用して評価する必要があります。

多変量のアプローチは、被験者の測定値を多変量の正規分布からのサンプルと見なし、分散共分散行列は、被験者間効果によって形成されるセル全体で同じです。セル全体で分散共分散行列が同じかどうかを検定する場合は、Box の M 検定を使用できます。

関連プロシージャ: 分散分析を行う前にデータを調べる場合は、「探索」プロシージャを使用します。それぞれの被験者に関する反復測定がない場合は、「GLM 1 変量」または「GLM 多変量」を使用します。各被験者に対する測定が 2 種類 (例えば、事前検定や事後検定) しかなく、被験者間因子が存在しない場合は、「対応のあるサンプルの t 検定」プロシージャを使用できます。

GLM 反復測定の取得

1. メニューから次の項目を選択します。

「分析」 > 「一般線型モデル」 > 「反復測定...」

2. 被験者内因子の名前とそのレベル数を入力します。
3. 「追加」をクリックします。
4. 被験者内因子ごとにこれらの手順を繰り返します。

2 重多変量の反復測定計画の測定因子を定義するには、以下のようになります。

5. 測定名を入力します。
6. 「追加」をクリックします。

因子と測定をすべて定義したら、以下のようになります。

7. 「定義」をクリックします。
8. リストの被験者内因子 (およびオプションとして、測定) の各組み合わせに対応する従属変数を選択します。

変数の位置を変更するには、上矢印と下矢印を使用します。

被験者内因子に変更を加える場合は、メイン・ダイアログ・ボックスを閉じずに、「反復測定の因子の定義」ダイアログ・ボックスを再度開くことができます。オプションとして、被験者間因子と共変量を指定できます。

GLM 反復測定の因子の定義

GLM 反復測定では、同じ属性のさまざまな測定値を表す関連従属変数のグループを分析します。このダイアログ・ボックスでは、GLM 反復測定で使用する被験者内因子を 1 つ以上定義できます。被験者内因子の指定順序は重要であることに注意してください。各因子は、その前の因子内のレベルを構成します。

反復測定を使用する場合は、データを正しく設定する必要があります。このダイアログ・ボックスで被験者内因子を定義する必要があります。これらの因子は、データ内の既存の変数ではなく、ここで定義する因子であることに注意してください。

例: 体重減少に関する研究で、数人の体重を 5 週間にわたって毎週測定すると仮定します。データ・ファイルでは、対象者がそれぞれ被験者またはケースとなります。数週間にわたる体重は変数 *weight1*、*weight2* などのように記録されます。別の変数には各人の性別が記録されます。体重は各被験者ごとに繰り返し測定され、被験者内因子を定義することでグループ分けできます。因子は、*week* という名前にして、5 つのレベルがあると定義することができます。メイン・ダイアログ・ボックスでは、変数 *weight1 ... weight5* を使用して、*week* の 5 つのレベルを割り当てます。男性と女性にグループ分けされたデータ・ファイル内の変数 (*gender*) は、男性と女性の相違点を研究するための被験者間因子として指定できます。

測定. 被験者を毎回複数の測定方法で検定した場合には、その測定方法を定義します。例えば、脈拍と呼吸数を被験者ごとに 1 週間毎日測定するとします。これらの測定方法は、データ・ファイルには変数として存在しませんが、ここで定義されます。複数の測定方法を持つモデルは、2 重多変量反復測定モデルと呼ばれることもあります。

GLM 反復測定のモデル

「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。交互作用のサブセットのみを指定する場合や、因子と共変量の交互作用を指定する場合は、「カスタム」を選択します。モデルに含める項はすべて指示する必要があります。

「**被験者間**」。被験者間因子と共変量がリストされます。

「**モデル**」。モデルは、使用するデータの性質によって異なります。「**カスタム**」の選択後、分析対象の被験者内効果と交互作用、および被験者間効果と交互作用を選択できます。

「**平方和**」。被験者間モデルの場合の平方和の計算方法。欠損セルがない釣り合い型または不釣り合い型被験者間モデルの場合、最も一般的に使用される方法はタイプ III 平方和です。

項の構築

選択した因子や共変量について、次の項を作成できます。

「**交互作用**」。選択したすべての変数の最上位交互作用項を作成します。これはデフォルトです。

「**主効果**」。選択した変数ごとに主効果の項目を作成します。

「**2 次まで**」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「**3 次まで**」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「**4 次まで**」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「**5 次まで**」。選択した変数で考えられるすべての 5 次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。タイプ III が最も一般的に使用され、デフォルトです。

タイプ I。この方法は、平方和の階層的分解法としても知られています。モデル内の各項はその前の項に対してのみ調整されます。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 次交互作用効果の前に主効果が指定され、2 次交互作用効果の前に 1 次交互作用効果が指定されているといったような分散分析の釣り合い型モデル。
- 高次項の前に低次項が指定されている多項式回帰モデル。
- 最初に指定された効果が 2 番目に指定された効果内にネストされ、2 番目に指定された効果が 3 番目に指定された効果内にネストされているような純粋なネスト・モデル。(このネスト形式はシンタックスを使用してのみ指定可能です。)

タイプ II。この方法は、他のすべての該当する効果に対して調整されたモデルの効果の平方和を計算します。該当する効果とは、調査されている効果が含まれて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。
- 純粋にネストになっているデザイン (ネストの形式は、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。この方法では、計画内の効果の平方和を、その効果を含まない他の効果に対して調整されており、その効果を含む効果 (存在する場合) に直交している平方和として計算します。タイプ III の平方和には、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多くの場合、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。デザインにおける効果 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ー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。
- 空白セルを含む釣り合い型モデルまたは不釣り合い型モデル。

GLM 反復測定の対比

対比は、被験者間因子のレベルに差があるかどうかを検定する場合に使用します。対比はモデルの被験者間因子ごとに指定できます。対比は、パラメーターの線型結合を表します。

仮説検定は $LBM=0$ という帰無仮説に基づきます。ここで、**L** は対比係数行列であり、**B** はパラメーターのベクトル、**M** は従属変数の平均変換に対応する平均行列です。この変換行列は、「反復測定: オプション」ダイアログ・ボックスで「**変換行列**」を選択することで表示できます。例えば、4 つの従属変数と 4 つのレベルの被験者内因子があり、多項式対比 (デフォルト) が被験者内因子で使用されている場合、**M** 行列は (0.5 0.5 0.5 0.5) になります。対比が指定されている場合は、被験者間因子に対応する列が対比と一致するように **L** 行列が作成されます。残りの列は、**L** 行列が推定できるように調整されます。

使用できる対比には、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、参照カテゴリーを最後のカテゴリーにするか最初のカテゴリーにするかを選択できます。

被験者内因子には、「なし」以外の対比を選択する必要があります。

対比の種類

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

「**単純**」。各レベルの平均を指定されたレベルの平均と比較します。このタイプの対比は、制御グループがある場合に役立ちます。最初または最後のカテゴリーを参照として選択できます。

「**差分**」。各レベル (最初を除く) の平均を前のレベルの平均と比較します。(逆 Helmert 対比と呼ばれることもあります。)

「**Helmert**」。因子の各レベル (最終を除く) の平均を後続レベルの平均と比較します。

「**反復**」。各レベル (最後を除く) の平均を後続レベルの平均と比較します。

「**多項式**」。1 次効果、2 次効果、3 次効果などを比較します。第 1 自由度にはすべてのカテゴリーにおける 1 次効果が含まれ、第 2 自由度には 2 次効果が含まれ、以下同様に続きます。このような対比は、多項式トレンドを推定する場合に頻繁に使用されます。

GLM 反復測定のプロファイル・プロット

プロファイル・プロット (交互作用プロット) は、モデルの周辺平均を比較する場合に役立ちます。プロファイル・プロットは、因子の 1 レベルにおける従属変数の推定周辺平均 (共変量に対して調整される) を各点で示す折れ線グラフです。第 2 因子のレベルを使用して、個別の線を作成することができます。第 3 因子の各レベルを使用して、個別のプロットを作成できます。すべての因子をプロットで使用することができます。プロファイル・プロットは従属変数ごとに作成されます。プロファイル・プロットには、被験者間因子と被験者内因子の両方を使用できます。

1 つの因子のプロファイル・プロットには、推定周辺平均がレベル全体で増えているか減っているかが示されます。2 つ以上の因子の場合、因子間に交互作用がないことが平行線で示されます。これは、1 つの因子のみのレベルしか調べることができないことを意味します。非平行線は交互作用を示します。

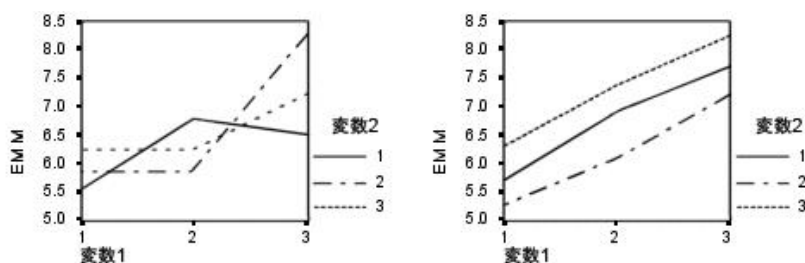


図2. 非平行プロット (左) と平行プロット (右)

横軸の因子を選択し、オプションとして、個別の線と個別のプロットの因子を選択してプロットを指定したら、そのプロットを「作図」リストに追加する必要があります。

GLM 反復測定におけるその後の比較

その後の多重比較検定。 平均値に差があることが判明した後で、その後の範囲検定とペアごとの多重比較でどの平均値が異なるのかを判別することができます。比較は未調整値に対して行われます。被験者間因子がない場合、このような検定は使用できません。その後の多重比較検定は、被験者内因子のレベル全体の平均に対して実行されます。

Bonferroni 検定 と **Tukey の HSD 検定** が一般的に使用される多重比較検定です。**Bonferroni 検定** (スチューデントの t 統計に基づく) では、多重比較が行われるという事実に応じて観測された有意レベルが調整されます。**Sidak の検定** でも有意レベルが調整され、限界が Bonferroni 検定の場合よりもより厳しくなります。**Tukey の HSD 検定** ではスチューデント化された範囲の統計量を使用して、グループ間ですべてのペアごとの比較を行い、実験ごとの誤差率をすべてのペアごとの比較の集合の誤差率に設定します。多数の平均値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。少数のペアの場合は Bonferroni の方が有効です。

Hochberg の GT2 は Tukey の HSD 検定に似ていますが、studentized maximum modulus が使用されます。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** でも studentized maximum modulus を使用しますが、一般的に、セル・サイズが等しくない場合は Hochberg の GT2 より有効です。セル・サイズのばらつきが大きい場合は、Gabriel の検定の方が公平になることがあります。

Dunnnett のペアごとの多重比較 t 検定 では、一連の処理を単一の制御平均値と比較します。最後のカテゴリがデフォルトの制御カテゴリです。代わりに、最初のカテゴリを選択することもできます。両側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 (制御カテゴリを除く) の平均値が制御カテゴリのものと同しくないことを検定するには、両側検定を使用します。因子の任意のレベルの平均値

が制御カテゴリーのものより小さいかどうかを検定するには、「< 対照カテゴリー」を選択します。同様に、因子の任意のレベルの平均値が制御カテゴリーのものより大きいかどうかを検定するには、「> 対照カテゴリー」を選択します。

Ryan, Einot, Gabriel、および Welsch (R-E-G-W) は 2 つの多重ステップダウン範囲検定を開発しました。多重ステップダウン・プロシージャでは、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い場合は、平均値のサブセットが等しいかどうかを検定されます。**R-E-G-W F** は F 検定に基づき、**R-E-G-W Q** はスチューデント化された範囲に基づきます。これらの検定は、Duncan の多重範囲検定および Student-Newman-Keuls (これも多重ステップダウン・プロシージャです) より有効ですが、セル・サイズが等しくない場合はお勧めできません。

分散が等しくない場合は、**Tamhane の T2** (t 検定に基づく控えめなペアごとの比較検定)、**Dunnett の T3** (studentized maximum modulus 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** (これが公平な場合もある)、または **Dunnett の C** (スチューデント化された範囲に基づくペアごとの比較検定) を使用します。

Duncan の多重範囲検定、Student-Newman-Keuls (**S-N-K**)、および **Tukey の b** は、グループ平均の順位付けおよび範囲値の計算を行う範囲検定です。これらの検定は、前述の検定ほど頻繁には使用されません。

Waller-Duncan の t 検定では Bayesian のアプローチを使用します。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用されます。

Scheffé 検定の有意レベルは、この機能で使用可能なペアごとの比較だけでなく、グループ平均で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。その結果、Scheffé 検定は多くの場合、他の検定より控えめになるため、有意確率では平均値間の差がより大きくなければなりません。

最小有意差 (**LSD**) のペアごとの多重比較検定は、グループのすべてのペア間の多重個別 t 検定と同等です。この検定の欠点は、観測された有意レベルを多重比較用に調整する試みが行われなことです。

表示される検定。ペアごとの比較は、LSD、Sidak、Bonferroni、Games-Howell、Tamhane の T2 と T3、Dunnett の C、および Dunnett の T3 用に提供されます。範囲検定の等分散性サブセットは、S-N-K、Tukey の b 、Duncan、R-E-G-W F 、R-E-G-W Q 、および Waller 用に提供されます。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 反復測定の保存

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

「**予測値**」。モデルがケースごとに予測する値。

- 「**非標準化**」。モデルが予測する従属変数の値。
- 「**標準誤差**」。独立変数の値が同じケースを対象とした、従属変数の平均値の標準偏差の推定値。

「**診断**」。独立変数の一般的でない値の組み合わせのケースと、モデルに大きな影響を与える可能性のあるケースを識別するための測定方法。Cook の距離と非中心化てこ比の値が使用できます。

- 「**Cook の距離**」。特定のケースを回帰係数の計算から除外した場合にすべてのケースの残差がどの程度変化するかを示す指標。Cook の D が大きいときは、回帰統計量の計算からケースを除外すると係数が大きく変化することを示します。

- 「てこ比の値」。非中心化てこ比の値。モデルの適合度に対する各観測値の相対的な影響度。

「残差」。非標準化残差は、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。標準化残差、スチューデント化された残差、および削除済み残差も使用できます。

- 「非標準化」。観測した値と、モデルによって予測された値との差。
- 「標準化」。残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。平均値は 0 であり、標準偏差は 1 です。
- 「スチューデント化」。独立変数の各ケースの値と独立変数の平均値との距離に応じて、ケースごとに異なる標準偏差の推定量で残差を割ったもの。
- 「削除済み」。ケースを回帰係数の計算から除外した場合のそのケースの残差。従属変数の値と調整済み予測値の差です。

「係数の統計量」。パラメーター推定値の分散共分散行列をデータ・セットまたはデータ・ファイルに保存します。また、それぞれの従属変数に対しても、パラメーター推定値の行、パラメーター推定値に対応する t 統計量の有意確率値の行、および残差自由度の行があります。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取る別のプロシージャで、この行列データを使用できます。データ・セットは、今後、同じセッションで使用可能ですが、セッション終了前に明示的に保存しない限り、ファイルとして保存されません。データ・セット名は、変数命名規則に従う必要があります。

GLM 反復測定オプション

このダイアログ・ボックスでオプションの統計を使用できます。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」。セルにおける母集団周辺平均の推定値が必要な因子と交互作用を選択します。これらの平均は共変量 (存在する場合) に対して調整されます。被験者内因子と被験者間因子の両方を選択できます。

- 「主効果の比較」。被験者間因子と被験者内因子の両方について、モデル内の主効果に対する推定周辺平均間で無修正のペアワイズ比較を行います。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のみ使用できます。
- 「信頼区間の調整」。最小有意差 (LSD)、Bonferroni、または Sidak の信頼区間と有意性に対する調整を選択します。この項目は、「主効果の比較」が選択されている場合にのみ使用できます。

「表示」。「記述統計」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、および度数が生成されます。「効果サイズの推定値」では、各効果および各パラメーター推定値の偏イータ 2 乗値が示されます。イータの 2 乗統計量は、因子に起因する総変動の比率を記述するものです。観測値に基づいて対立仮説を設定する際に検定力を取得するには、「観測検定力」を選択します。パラメーター推定値、標準誤差、 t 検定、信頼区間、および各検定の観測検定力を生成するには、「パラメータ推定値」を選択します。仮説、誤差「SSCP 行列」、「残差 SSCP 行列」に加え、残差分散共分散行列の Bartlett の球面性検定を表示することができます。

「等分散性の検定」では、被験者間因子の場合にのみ、被験者間因子のレベルの組み合わせ全体の従属変数ごとに、Levene の等分散性検定を行います。この等分散性の検定には、被験者間因子のすべてのレベルの組み合わせにおける従属変数の共分散行列の等分散性に関する Box の M 検定も含まれます。レベルと広がりの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確認に役立ちます。この項目は、因子が存在しない場合は無効になります。各従属変数について、標準化された残差による予測値で観測された値の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を調べる場合に役立ちます。従属変数と独立変数との関係がモデルで適切に記述できるかどうかを確認

するには、「**不適合度検定**」を選択します。「**一般の推定可能関数**」を使用すれば、一般推定可能関数に基づくカスタム仮説検定を構成できます。任意の対比係数行列における行は、一般推定可能関数の線型結合です。

「**有意レベル**」。その後の検定で使用される有意レベルと、信頼区間を構成するために使用される信頼度レベルを調整できます。指定された値は、検定の観測検定力の計算にも使用できます。有意レベルの指定時に、信頼区間の関連レベルがダイアログ・ボックスに表示されます。

GLM コマンドの追加機能

これらの機能は、1 変量、多変量、反復測定分析に適用できます。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 計画のネスト効果の指定 (DESIGN サブコマンドを使用)。
- 効果の検定、または効果や値の線型の組み合わせの検定の指定 (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)。
- カスタムの **L** 行列、**M** 行列、または **K** 行列の作成 (LMATRIX、MMATRIX、KMATRIX サブコマンドを使用)。
- 偏差対比または単純対比での中間参照カテゴリーの指定 (CONTRAST サブコマンドを使用)。
- 多項式の対比の行列を指定 (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)。
- 任意の因子に対する推定周辺平均、または因子リスト内の因子間の因子交互作用の計算 (EMMEANS サブコマンドを使用)。
- 一時変数の名前の指定 (SAVE サブコマンドを使用)。
- 相関行列データ・ファイルの作成 (OUTFILE サブコマンドを使用)。
- 被験者間分散分析テーブルから取得した統計を含む行列データ・ファイルの作成 (OUTFILE サブコマンドを使用)。
- 新規データ・ファイルへのデザイン行列の保存 (OUTFILE サブコマンドを使用)。

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

第 4 章 分散成分分析

混合効果モデルにおける「分散成分」プロシージャーでは、従属変数の分散に対するランダム効果ごとの寄与率を推定します。このプロシージャーは、分割プロット、1 変量の反復測定、およびランダム・ブロック計画などの混合モデルの分析に特に役立ちます。分散成分を計算することで、分散を減らすためにどこに焦点を絞るべきか判断することができます。

分散成分の推定には、最小ノルム 2 次不偏推定量 (MINQUE)、分散分析 (ANOVA)、最尤法 (ML)、および制限付き最尤法 (REML) という 4 つの方法を使用できます。これらの方法でさまざまな指定を使用できます。

すべての方法のデフォルトの出力に、分散成分推定値が含まれます。ML 法または REML 法を使用する場合は、漸近分散共分散行列表も表示されます。その他の使用可能な出力には、ANOVA 表と ANOVA 法の期待平均平方、および ML 法と REML 法の反復の記述などがあります。「分散成分」プロシージャーには、「GLM 1 変量」プロシージャーとの完全な互換性があります。

「WLS 重み」を使用すると、重み付き分析の場合、観測値に異なる重みを付けるために使用する変数を指定し、それにより測定精度における変動を補正できる可能性があります。

例: 農業学校で、子豚の同腹子 6 グループの 1 カ月後の体重増を測定します。同腹子変数は、6 つのレベルを持つランダム因子です。(研究対象の 6 つの同腹子グループは、豚の同腹子のさらに大きな母集団からのランダム・サンプルです)。研究者は、同腹子内の子豚間の差よりも、同腹子グループ間の差の方が、体重増の差異をもたらすはるかに大きな原因であることを発見しました。

分散成分のデータの考慮事項

データ: 従属変数は量的です。因子はカテゴリー型です。これらは最高 8 バイトまでの数値または文字列値を持つことができます。少なくとも 1 つの因子はランダムである必要があります。つまり、因子レベルは可能なレベルのランダム・サンプルである必要があります。共変量は、従属変数に関連する量的変数です。

仮定: どの方法においても、ランダム効果のモデル・パラメーターの平均値は 0 で、有限定数分散があり、相互に相関がないものと仮定します。異なるランダム効果のモデル・パラメーターにも相関はありません。

残差の項も平均値が 0 で、有限定数分散があります。どのランダム効果のモデル・パラメーターとも相関はありません。異なる観測値からの残差項は相関がないものと見なされます。

これらの仮定に基づいて、同じレベルのランダム因子からの観測値に相関を持たせません。この点から、分散成分モデルは一般線型モデルと区別されます。

ANOVA と MINQUE には正規性の仮定は必要ありません。どちらも、正規性の仮定からの適度の逸脱については、許容範囲としています。

ML と REML では、モデル・パラメーターと残差項が正規分布している必要があります。

関連プロシージャー: 分散成分の分析を行う前にデータを調べる場合は、「探索」プロシージャーを使用します。仮説検定には、GLM 1 変量、GLM 多変量、GLM 反復測定を使用します。

分散成分表の取得

1. メニューから次の項目を選択します。

「分析」 > 「一般線型モデル」 > 「分散成分...」

2. 従属変数を選択します。

3. データに応じて、「固定因子」、「ランダム因子」、および「共変量」の変数を選択します。重み付け変数の指定には、「WLS 重み」を使用します。

分散成分のモデル

「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。交互作用のサブセットのみを指定する場合や、因子と共変量の交互作用を指定する場合は、「カスタム」を選択します。モデルに含める項はすべて指示する必要があります。

「因子と共変量」。因子と共変量がリストされます。

「モデル」。モデルは、使用するデータの性質によって異なります。「カスタム」の選択後、分析対象の主効果と交互作用を選択できます。モデルにはランダム因子を含める必要があります。

「モデルに切片を含める」。通常、切片はモデルに含まれます。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

選択した因子や共変量について、次の項を作成できます。

「交互作用」。選択したすべての変数の最上位交互作用項を作成します。これはデフォルトです。

「主効果」。選択した変数ごとに主効果の項目を作成します。

「2 次まで」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「3 次まで」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「4 次まで」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「5 次まで」。選択した変数で考えられるすべての 5 次交互作用を作成します。

分散成分のオプション

「方法」。分散成分の 4 つの推定方法のいずれかを選択できます。

- 「MINQUE」(最小ノルム 2 次不偏推定値) では、固定効果に対して不変の推定値を生成します。データが正規分布しており、推定値が正しければ、この方法により、すべての不偏推定値の中で最も小さい分散が生成されます。ランダム効果先行の重み方式を選択できます。
- 「ANOVA」(分散分析) では、効果ごとにタイプ I またはタイプ III の平方和を使用して、不偏推定値を計算します。ANOVA 法では負の分散推定値が生成される場合があり、間違ったモデル、不適切な推定方法、あるいはさらにデータが必要であることを示している可能性があります。

- 「最尤法」(ML) では、反復を使用して、実際に観測されたデータに最も一致する推定値を生成します。これらの推定値は偏る可能性があります。この方法は漸近的に正規です。ML と REML の推定値は、解釈においては不変です。この方法は、固定効果の推定に使用される自由度を考慮しません。
- 「制限された最尤法」(REML) の推定値により、平行データの (すべてではないが) 多くのケースに対する ANOVA 推定値が減ります。この方法は固定効果に合わせて調整されるため、ML 法の場合より標準誤差が小さくなければなりません。この方法は、固定効果の推定に使用される自由度を考慮します。

「ランダム効果先行」。「一様」とは、すべてのランダム効果と残差項が観測値に同等の影響をもたらすことを意味します。「ゼロ」スキームは、ランダム効果分散をゼロと仮定するのと同じです。MINQUE 法にのみ使用可能です。

「平方和」。「タイプ I」の平方和は、分散成分の文献で頻繁に使用される階層モデルで使用されます。「タイプ III」(GLM におけるデフォルト) を選択すると、分散推定値を GLM 1 変量でタイプ III 平方和を使用した仮説の検定に使用できます。ANOVA 法にのみ使用可能です。

「基準」。収束基準と最大反復数を指定できます。ML 法または REML 法にのみ使用可能です。

「表示」。ANOVA 法の場合、平方和と期待平均平方の表示を選択できます。「最尤法」または「制限された最尤法」を選択した場合は、反復履歴を表示できます。

平方和 (分散成分)

モデルには、平方和のタイプを選択できます。タイプ III が最も一般的に使用され、デフォルトです。

タイプ I。この方法は、平方和の階層的分解法としても知られています。モデル内の各項はその前の項に対してのみ調整されます。タイプ I 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。

- 1 次交互作用効果の前に主効果が指定され、2 次交互作用効果の前に 1 次交互作用効果が指定されているといったような分散分析の釣り合い型モデル。
- 高次項の前に低次項が指定されている多項式回帰モデル。
- 最初に指定された効果が 2 番目に指定された効果内にネストされ、2 番目に指定された効果が 3 番目に指定された効果内にネストされているような純粋なネスト・モデル。(このネスト形式はシンタックスを使用してのみ指定可能です。)

タイプ III。デフォルト。この方法では、計画内の効果の平方和を、その効果を含まない他の効果に対して調整されており、その効果を含む効果 (存在する場合) に直交している平方和として計算します。タイプ III の平方和には、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、このタイプは多くの場合、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 の平方和の方法は、一般に次のような場合に使用します。

- タイプ I にリストされているモデル。
- 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。

分散成分の新規ファイルへの保存

このプロシージャの一部の結果を新しい IBM SPSS Statistics データ・ファイルに保存することができます。

「分散成分推定値」。分散成分の推定値と推定値ラベルをデータ・ファイルまたはデータ・セットに保存します。これらは、より多くの統計量を計算したり、GLM プロシージャで詳しい分析を行う場合に使用できます。例えば、信頼区間の計算や仮説の検定に使用できます。

「成分共変量」。分散共分散行列または相関行列をデータ・ファイルまたはデータ・セットに保存します。「最尤法」または「制限された最尤法」が指定されている場合にのみ使用可能です。

「作成された値の保存先」。分散成分の推定値や行列を含むファイルのデータ・セット名や外部ファイル名を指定できます。データ・セットは、今後、同じセッションで使用可能ですが、セッション終了前に明示的に保存しない限り、ファイルとして保存されません。データ・セット名は、変数命名規則に従う必要があります。

MATRIX コマンドを使用すれば、データ・ファイルから必要なデータを抽出して、信頼区間の計算や検定を行うことができます。

VARCOMP コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 計画のネスト効果の指定 (DESIGN 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)。

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

第 5 章 線型混合モデル

「線型混合モデル」プロシージャーでは、相関および非定常変動が見られるデータも扱えるように、一般線型モデルを拡張します。したがって、線型混合モデルはデータの平均値だけでなく、データの分散および共分散をモデリングできる柔軟性を提供します。

また、「線型混合モデル」プロシージャーは、混合線型モデルとして定式化できる他のモデルを当てはめるための柔軟なツールでもあります。そのようなモデルには、マルチレベル・モデル、階層線型モデル、およびランダム係数モデルがあります。

例: 食料雑貨店チェーンは、さまざまなクーポンが顧客支出に対して与える影響に関心を持っています。常連客のランダム・サンプリングを行い、10 週間の各顧客の支出を調べます。1 週間ごとに異なるクーポンが顧客に送付されます。線型混合モデルを使用して、10 週間にわたる各被験者の反復観測値による相関に合わせて調整しながら、支出に対するさまざまなクーポンの影響を推定します。

方法: 最尤法 (ML) および制限付き最尤法 (REML) の推定。

統計: 記述統計: 従属変数のサンプル・サイズ、平均値、および標準偏差と、因子のそれぞれ異なるレベルの組み合わせの共変量。因子レベル情報: 各因子のレベルをソートした値と、それらの度数。また、固定効果のパラメーター推定値および信頼区間、共分散行列のパラメーターの Wald 検定および信頼区間。異なる仮説を評価する場合は、タイプ I とタイプ III の平方和を使用できます。タイプ III がデフォルトです。

線型混合モデルのデータの考慮事項

データ: 従属変数は量的である必要があります。因子はカテゴリ型である必要があります、数値または文字列値を持つことができます。共変量および重み付け変数は量的である必要があります。被験者および反復変数はどのようなタイプでもかまいません。

仮定. 従属変数は固定因子、ランダム因子、および共変量と線型関係があると仮定されます。固定効果は従属変数の平均値をモデリングします。ランダム効果は従属変数の共分散構造をモデリングします。多重ランダム効果は互いに独立していると思われ、それぞれ別の共分散行列が計算されます。ただし、同じランダム効果で指定したモデル項は相関している場合があります。反復測定は、残差の共分散構造をモデリングします。また、従属変数は正規分布から取得されると仮定されます。

関連プロシージャー: 分析の実行前にデータを調べる場合は、「探索」プロシージャーを使用します。相関変動または非定常変動が存在することが確実な場合は、「GLM 1 変量」または「GLM 反復測定」プロシージャーを使用できます。ランダム効果が分散成分の共分散構造を持ち、反復測定がない場合は、代わりに「分散成分分析」プロシージャーを使用できます。

線型混合モデル分析の取得

1. メニューから次の項目を選択します。

「分析」 > 「混合モデル」 > 「線型...」

2. オプションで、1 つ以上の被験者変数を選択します。
3. オプションで、1 つ以上の反復変数を選択します。
4. オプションで、残差共分散構造を選択します。

5. 「**続行**」をクリックします。
6. 従属変数を選択します。
7. 少なくとも 1 つの因子または共変量を選択します。
8. 「**固定**」または「**変量**」をクリックして、少なくとも固定効果モデルまたはランダム効果モデルを指定します。

オプションで、重み付け変数を選択します。

線型混合モデルの被験者/反復変数の選択

このダイアログ・ボックスでは、被験者および反復観測値を定義する変数を選択し、残差の共分散構造を選択できます。

「**被験者**」。被験者は観測用の単位で、他の被験者から独立していると思なされます。例えば医学研究では、ある患者の血圧測定値は、他の患者の測定値からは独立していると思なすことができます。被験者ごとに反復測定を行い、それらの観測値間の相関をモデル化する場合、被験者の定義は特に重要になります。例えば、病院に連続で通院している間のある患者の血圧測定値は相関していることが予想されます。

被験者は、複数の変数の因子レベルの組み合わせによって定義することもできます。例えば、性別 と年齢 カテゴリー を被験者変数として指定すると、65 歳を超える男性 は互いに似ていますが、65 歳以下の男性 と女性 からは独立しているという確信をモデル化できます。

「被験者」リストで指定された変数はすべて、残差共分散構造の被験者を定義するために使用されます。一部またはすべての変数を、ランダム効果の共分散構造の被験者を定義するために使用できます。

「**反復**」。このリストで指定された変数は、反復観測値を識別するために使用されます。例えば、週 という 1 つの変数を使用して、医学研究において 10 週間の観測値を識別できます。また、月 と日 を共に使用すると、1 年間にわたって毎日の観測値を識別できます。

「**反復共分散タイプ**」。残差に対する共分散構造を指定します。使用可能な構造は次のとおりです。

- 前従属: 1 次
- AR(1)
- AR(1): 不均質
- ARMA(1,1)
- 複合対称
- 複合対称: 相関行列
- 複合対称: 不均質
- 対角
- 因子分析的: 1 次
- 因子分析的: 1 次、不均質
- Huynh-Feldt
- スケーリングされた単位
- Toeplitz
- Toeplitz: 不均質
- 無構造
- 無構造: 相関

詳しくは、99 ページの『第 17 章 共分散構造』のトピックを参照してください。

線型混合モデルの固定効果

固定効果。 デフォルトのモデルはありません。したがって、明示的に固定効果を指定する必要があります。ネスト項目またはネストなし項目を作成することもできます。

切片を含める。 通常、切片はモデルに含まれます。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。

平方和。 平方和の計算方法。欠損セルがないモデルの場合は、タイプ III 法が最も一般的に使用されます。

ネストなし項目の構築

選択した因子や共変量について、次の項を作成できます。

「因子」。選択した変数で考えられるすべての交互作用および主効果を作成します。これはデフォルトです。

「交互作用」。選択したすべての変数の最上位交互作用項を作成します。

「主効果」。選択した変数ごとに主効果の項目を作成します。

「2 次まで」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「3 次まで」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「4 次まで」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「5 次まで」。選択した変数で考えられるすべての 5 次交互作用を作成します。

入れ子項目の構築

このプロシージャでは、モデルにネスト項目を作成できます。ネスト項目は、値が他の因子のレベルと交互作用しない因子または共変量の効果をモデル化するのに有用です。例えば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査するとします。各顧客は、それらの店舗のうち 1 つの店舗にのみ頻繁に通うため、顧客 効果は店舗の場所 効果内にネストされているとすることができます。

さらに、交互作用効果を含めたり、ネスト項目に対して複数レベルのネストを追加することができます。

制限。 ネスト項目には、次のような制限があります。

- 交互作用内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A*A$ の指定は無効です。
- ネスト効果内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A(A)$ の指定は無効です。
- 共変量内では効果をネストすることはできません。したがって、 A が因子で、 X が共変量の場合、 $A(X)$ の指定は無効です。

平方和

モデルには、平方和のタイプを選択できます。タイプ III が最も一般的に使用され、デフォルトです。

タイプ I。この方法は、平方和の階層的分解法としても知られています。モデル内の各項はその前の項に対してのみ調整されます。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 次交互作用効果の前に主効果が指定され、2 次交互作用効果の前に 1 次交互作用効果が指定されているといったような分散分析の釣り合い型モデル。
- 高次項の前に低次項が指定されている多項式回帰モデル。
- 最初に指定された効果が 2 番目に指定された効果内にネストされ、2 番目に指定された効果が 3 番目に指定された効果内にネストされているような純粋なネスト・モデル。(このネスト形式はシンタックスを使用してのみ指定可能です。)

「**タイプ III**」。デフォルト。この方法では、計画内の効果の平方和を、その効果を含まない他の効果に対して調整されており、その効果を含む効果 (存在する場合) に直交している平方和として計算します。タイプ III の平方和には、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多くの場合、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 の平方和の方法は、一般に次のような場合に使用します。

- タイプ I にリストされているモデル。
- 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。

線型混合モデルのランダム効果

「**共分散タイプ**」。これを使用して、ランダム効果モデルの共分散構造を指定できます。各ランダム効果に個別の共分散行列が推定されます。使用可能な構造は次のとおりです。

- 前従属: 1 次
- AR(1)
- AR(1): 不均質
- ARMA(1,1)
- 複合対称
- 複合対称: 相関行列
- 複合対称: 不均質
- 対角
- 因子分析的: 1 次
- 因子分析的: 1 次、不均質
- Huynh-Feldt
- スケーリングされた単位
- Toeplitz
- Toeplitz: 不均質
- 無構造
- 無構造: 相関行列
- 分散成分

詳しくは、99 ページの『第 17 章 共分散構造』のトピックを参照してください。

「**ランダム効果**」。デフォルトのモデルはありません。したがって、明示的にランダム効果を指定する必要があります。ネスト項目またはネストなし項目を作成することもできます。ランダム効果モデルに切片項を含めることを選択することもできます。

複数のランダム効果モデルを指定できます。最初のモデルを構築した後で、次のモデルを構築する場合は「**次へ**」をクリックします。スクロールして既存のモデルに戻る場合は、「**前へ**」をクリックします。各ランダム効果モデルは他のすべてのランダム効果モデルから独立している（つまり、それぞれ別の共分散行列が計算される）ものと仮定されます。同じランダム効果モデルで指定した項は関連している場合があります。

「**被験者のグループ化**」。リストされる変数は、「被験者および反復測定の定義」ダイアログ・ボックスで被験者変数として選択した変数です。ランダム効果モデルの被験者を定義するには、これらの変数の一部またはすべてを選択します。

線型混合モデルの推定

「**方法**」。 「**最尤推定値**」または「**制限最尤推定値**」を選択します。

「**反復**」：使用可能なオプションは次のとおりです。

- 「**最大反復回数**」。負でない整数を指定してください。
- 「**段階 2 分の最大数**」。対数尤度が増加するか、最大段階 2 分に達するまで、反復ごとにステップ・サイズが 0.5 の因子ずつ減少します。正の整数を指定してください。
- 「**すべての反復の記述**」。対数尤度関数値、および 0 回目の反復（初期推定値）から始まる反復の n 回ごとのパラメーター推定値を含む表を表示します。反復の記述の出力を選択した場合、 n の値にかかわらず、最後の反復は常に出力されます。

「**対数尤度収束**」。対数尤度関数の絶対変化または相対変化が、指定した値（負以外でなければなりません）より小さい場合に収束とみなされます。指定された値が 0 の場合、この基準は使用されません。

「**パラメーター収束**」。パラメーター推定値の最大絶対変化または最大相対変化が、指定した値（負以外でなければなりません）より小さい場合に収束とみなされます。指定された値が 0 の場合、この基準は使用されません。

「**Hessian 収束**」。 「**絶対**」を指定した場合は、Hessian に基づく統計量が、指定した値よりも小さい場合に収束とみなされます。「**相対値**」を指定した場合は、指定した値と対数尤度の絶対値の積よりも統計が小さい場合に収束とみなされます。指定された値が 0 の場合、この基準は使用されません。

「**最大スコアリング・ステップ数**」。反復回数が n に達するまで、Fisher スコア法のアルゴリズムを使用するよう要求します。負でない整数を指定してください。

「**特異性許容度**」。この値は、特異性の検査時に許容度として使用されます。正の値を指定してください。

線型混合モデルの統計

「**統計の要約**」。次の表が作成されます。

- 「**記述統計**」。従属変数のサンプル・サイズ、平均値、および標準偏差と、共変量（指定した場合）を表示します。これらの統計は、因子の異なるレベルの組み合わせごとに表示されます。
- 「**ケース処理要約**」。因子のソート値、反復測定変数、反復測定の被験者、およびランダム効果の被験者とその度数を表示します。

「モデル統計量」。次の表が作成されます。

- 「パラメーター推定値」。固定効果およびランダム効果のパラメーター推定値と、それらの近似標準誤差を表示します。
- 「共分散パラメーターの検定」。共分散パラメーターの漸近標準誤差と Wald 検定を表示します。
- 「パラメーター推定値の相関」。固定効果パラメーター推定値の漸近相関行列を表示します。
- 「パラメーター推定値の共分散」。固定効果パラメーター推定値の漸近分散共分散行列を表示します。
- 「ランダム効果の共分散」。ランダム効果の推定共分散行列を表示します。このオプションは、1 つ以上のランダム効果が指定されている場合にのみ使用できます。ランダム効果に被験者変数が指定されている場合は、共通ブロックが表示されます。
- 「残差の共分散」。推定残差分散共分散行列を表示します。このオプションは、反復変数が指定されている場合にのみ使用できます。被験者変数が指定されている場合は、共通ブロックが表示されます。
- 「対比係数行列」。このオプションは、固定効果とカスタム仮説の検定に使用される推定可能関数を表示します。

「信頼区間」。信頼区間を構築する場合は、常にこの値を使用します。0 以上 100 未満の値を指定します。デフォルト値は 95 です。

線型混合モデルの EM 平均

「適合モデルの推定周辺平均」。このグループでは、セル内の従属変数のモデル予測推定周辺平均と、指定した因子に対するそれらの標準誤差を要求することができます。さらに、主効果の因子レベルの比較を要求することができます。

- 「因子と因子交互作用」。このリストには、「固定」ダイアログ・ボックスで指定された因子と因子交互作用、および OVERALL 項が含まれています。共変量から作成したモデル項は、このリストから除外されます。
- 「平均値の表示」。このプロシージャでは、このリストに選択した因子と因子交互作用の推定周辺平均を計算します。OVERALL を選択した場合は、すべての因子を非表示にして、従属変数の推定周辺平均が表示されます。選択した因子または因子交互作用は、関連する変数がメイン・ダイアログ・ボックスの「因子」リストから除外されない限り、選択されたままの状態になることに注意してください。
- 「主効果の比較」。このオプションを使用して、選択した主効果のレベルのペアワイズ比較を要求できます。「信頼区間の調整」では、多重比較に応じて信頼区間および有意確率値に調整を適用できます。使用できる方法は LSD (調整なし)、Bonferroni、および Sidak です。最後に、各因子に対して、比較の対象となる参照カテゴリーを選択できます。参照カテゴリーを選択しない場合は、すべてのペアワイズ比較が構成されます。参照カテゴリーのオプションは、「最初」、「最後」、または「カスタム」(この場合は、参照カテゴリーの値を入力します) です。

線型混合モデルの保存

このダイアログ・ボックスを使用して、さまざまなモデル結果を作業ファイルに保存できます。

「固定予測値」。効果のない回帰平均に関連する変数を保存します。

- 「予測値」。ランダム効果のない回帰平均。
- 「標準誤差」。推定値の標準誤差。
- 「自由度」。推定値に関連付けられている自由度。

「予測値と残差」。モデルの当てはめ値に関連する変数を保存します。

- 「予測値」。モデルの当てはめ値。
- 「標準誤差」。推定値の標準誤差。
- 「自由度」。推定値に関連付けられている自由度。
- 「残差」。データ値から予測値を引いた値。

MIXED コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 効果の検定、または効果や値の線型の組み合わせの検定の指定 (TEST 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)。
- 指定した共変量値の推定周辺平均の計算 (EMMEANS サブコマンドの WITH キーワードを使用)。
- 交互作用の単純な主効果の比較 (EMMEANS サブコマンドを使用)。

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

第 6 章 一般化線型モデル

一般化線型モデルは、従属変数が、指定したリンク関数によって因子および共変量と線型の関係になるように一般線型モデルを拡張したものです。さらにこのモデルを使用すると、従属変数に非正規分布を持たせることができるようになります。正規分布応答の線型回帰、バイナリー・データに対するロジスティック・モデル、度数データに対する対数線型モデル、打ち切りデータに対する補ログ・マイナス・ログ・モデル、さらに一般的なモデル式を使用したその他多くの統計モデルなど、広く使用されている統計モデルを網羅しています。

例。 運送会社では一般化線型モデルを使用して、異なる期間に建造された複数のタイプの船の損害数にポワソン回帰を当てはめることができます。生成されるモデルは、損害を最も受けやすい船のタイプの判別に役立ちます。

自動車保険会社では一般化線型モデルを使用して、自動車の損害請求にガンマ回帰を当てはめることができます。生成されたモデルは、請求サイズに最も寄与する因子の判別に役立ちます。

医療研究者は一般化線型モデルを使用して、打ち切りデータに補ログ・マイナス・ログ回帰を当てはめ、病状が再発する時間を予測できます。

一般化線型モデルのデータの考慮事項

データ: 応答には、スケール、度数、2 値、または試行におけるイベントがあります。因子はカテゴリ型と見なされます。共変量、スケール重み付け、およびオフセットはスケールと見なされます。

仮定: ケースは独立した観測値と見なされます。

一般化線型モデルを取得するには

メニューから次の項目を選択します。

「分析」 > 「一般化線型モデル」 > 「一般化線型モデル...」

1. 分布およびリンク関数を指定します (各種オプションの詳細については下記参照)。
2. 「応答」タブで、従属変数を選択します。
3. 「予測」タブで、従属変数の予測に使用する因子と共変量を選択します。
4. 「モデル」タブで、選択した因子および共変量を使用してモデル効果を指定します。

「モデルの種類」タブでは、モデルに対して分布およびリンク関数を指定でき、応答の種類によって分類されているいくつかの一般的なモデルに対してショートカットが提供されます。

モデル・タイプ

スケール応答。使用可能なオプションは次のとおりです。

- **線型。** 分布に「正規」、リンク関数に「同一」を指定します。
- **対数リンクを使用するガンマ。** 分布に「ガンマ」、リンク関数に「対数」を指定します。

順序応答。使用可能なオプションは次のとおりです。

- **順序ロジスティック。** 分布に「多項 (順序)」、リンク関数に「累積ロジット」を指定します。

- **順序プロビット**。分布に「多項 (順序)」、リンク関数に「累積プロビット」を指定します。

度数。使用可能なオプションは次のとおりです。

- **ポワソン対数線型**。分布に「ポワソン」、リンク関数に「対数」を指定します。
- **対数リンクを使用した負の 2 項**。分布に「負の 2 項 (補助パラメーターに 1 の値)」、リンク関数に「対数」を指定します。補助パラメーターの値を推定するプロシージャーにするには、「負の 2 項分布」のカスタム・モデルを指定し、パラメーター・グループで「推定値」を選択します。

2 値応答またはイベント/試行データ。使用可能なオプションは次のとおりです。

- **2 値ロジスティック**。分布に「2 項」、リンク関数に「ロジット」を指定します。
- **2 値プロビット**。分布に「2 項」、リンク関数に「プロビット」を指定します。
- **打ち切り**。分布に「2 項」、リンク関数に「補ログ・マイナス・ログ」を指定します。

混合。使用可能なオプションは次のとおりです。

- **対数リンクを使用した Tweedie**。分布に「Tweedie」、リンク関数に「対数」を指定します。
- **同一リンクを使用した Tweedie**。分布に「Tweedie」、リンク関数に「同一」を指定します。

カスタム。独自の分布とリンク関数の組み合わせを指定します。

分布

この選択により、従属変数の分布を指定します。非正規分布と非同一リンク関数を指定する機能は、一般線型モデルから一般化線型モデルに移行する上での重要な改善点です。分布とリンク関数は多くの組み合わせが可能で、その中のいくつかは特定のデータ・セットに適しているため、経験的理論の考察、または最も適合すると考えられる組み合わせから選択を導き出すことができます。

- **2 項**。この分布はその応答が 2 値応答またはイベント数を表す変数にのみ適しています。
- **ガンマ**。この分布は、正の値が大きくなるほどゆがむ正のスケール値を持つ変数に適しています。データ値が 0 以下または欠損値である場合、対応するケースは分析には使用されません。
- **逆ガウス**。この分布は、正の値が大きくなるほどゆがむ正のスケール値を持つ変数に適しています。データ値が 0 以下または欠損値である場合、対応するケースは分析には使用されません。
- **負の 2 項**。この分布は k の成功を観測するために必要な試行回数として考えることができ、負ではない整数値の変数に適しています。データ値が非整数である、0 未満である、または欠損値の場合、対応するケースは分析には使用されません。負の 2 項分布の補助パラメーターの値は、0 以上の数値ですが、パラメーターに固定値を設定し、プロシージャーで推定することも可能です。補助パラメーターが 0 に設定されると、この分布を使用することはポワソン分布を使用することに相当します。
- **正規**。これは、値が対称で、中心 (平均) 値に関してベル型の分布であるスケール変数に適しています。従属変数は数値型である必要があります。
- **ポワソン**。この分布は一定期間の対象のイベントの発生回数として考えることができ、負ではない整数値の変数に適しています。データ値が非整数である、0 未満である、または欠損値の場合、対応するケースは分析には使用されません。
- **Tweedie**。この分布は、ガンマ分布のポワソン混合により表現できる変数に適しています。分布は連続分布 (負ではない実数をとる) および離散分布 (単一の値での正の確率質量、0) のプロパティを組み合わせるという意味から「混合」です。従属変数は数値型で 0 以上のデータ値である必要があります。データ値が 0 未満である、または欠損値の場合、対応するケースは分析には使用されません。Tweedie 分布のパラメーターの固定値は、1 より大きい 2 未満の数値です。
- **多項**。この分布は順序型応答を表す変数に適しています。従属変数は数値または文字列のどちらかで、少なくとも 2 つの異なる有効データ値を持つ必要があります。

リンク関数

リンク関数は従属変数の変換を行い、それによりモデルの推定ができます。次の関数を使用することができます。

- 同一。 $f(x)=x$ 。従属変数は変換されません。このリンクはどの分布でも使用できます。
- 補ログ・マイナス・ログ。 $f(x)=\log(-\log(1-x))$ 。これは 2 項分布にのみ適しています。
- 累積コーチット。 $f(x) = \tan(\pi (x - 0.5))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- 累積補ログ・マイナス・ログ。 $f(x)=\ln(-\ln(1-x))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- 累積ロジット。 $f(x)=\ln(x / (1-x))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- 累積の負ログ・マイナス・ログ。 $f(x)=-\ln(-\ln(x))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- 累積プロビット。 $f(x)=\Phi^{-1}(x)$ 。応答の各カテゴリーの累積確率に適用されます。 Φ^{-1} は逆標準正規累積分布関数です。これは多項分布にのみ適しています。
- 対数。 $f(x)=\log(x)$ 。このリンクはどの分布でも使用できます。
- 対数補完。 $f(x)=\log(1-x)$ 。これは 2 項分布にのみ適しています。
- ロジット。 $f(x)=\log(x / (1-x))$ 。これは 2 項分布にのみ適しています。
- 負の 2 項。 $f(x)=\log(x / (x+k^{-1}))$ 。 k は負の 2 項分布の補助パラメーターです。これは負の 2 項分布にのみ適しています。
- 負ログ・マイナス・ログ。 $f(x)=-\log(-\log(x))$ 。これは 2 項分布にのみ適しています。
- オズベき乗。 $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$ ($\alpha \neq 0$ の場合)。 $f(x)=\log(x)$ ($\alpha=0$ の場合)。 α には数値を指定する必要があり、その数値は実数である必要があります。これは 2 項分布にのみ適しています。
- プロビット。 $f(x)=\Phi^{-1}(x)$ 。 Φ^{-1} は逆標準正規累積分布関数です。これは 2 項分布にのみ適しています。
- べき乗。 $f(x)=x^\alpha$ 、 $\alpha \neq 0$ の場合。 $f(x)=\log(x)$ 、 $\alpha=0$ の場合。 α には数値を指定する必要があり、その数値は実数である必要があります。このリンクはどの分布でも使用できます。

一般化線型モデルにおける応答

多くの場合、従属変数を簡単に指定できます。ただし、2 値のみをとる変数や試行で発生するイベントを記録する応答については、特別な注意が必要です。

- 2 値応答。従属変数が 2 値のみをとる場合、パラメーター推定値に参照カテゴリーを指定できます。2 値応答変数は数値型または文字列型です。
- 一連の試行内で発生するイベント数。応答が一連の試行内で発生するイベント数の場合、従属変数にはイベント数が含まれ、試行回数を含む追加の変数を選択できます。ただし、試行回数がすべての被験者間で同じ場合、試行回数は固定値を使用して指定できます。試行回数は各ケースのイベント数以上である必要があります。イベント数は負でない整数、試行回数は正の整数である必要があります。

順序多項モデルの場合、応答のカテゴリー順を、昇順、降順、またはデータ順 (データ順は、データで最初に検出された値が最初のカテゴリーを定義し、最後に検出された値が最後のカテゴリーを定義することを意味します) に指定できます。

「スケール重み」。スケール・パラメーターは、応答の分散に関連する推定モデル・パラメーターです。スケール重みは観測ごとに異なる「既知」の値です。スケール重み変数が指定されている場合、応答の分散に

関するスケール・パラメーターは、観測ごとにその変数で除算されます。0 以下または欠損しているスケール重み値があるケースは、分析では使用されません。

一般化線型モデルの参照カテゴリー

2 値応答については、従属変数に参照カテゴリーを選択できます。これはパラメーター推定値や保存値などの特定の出力に影響を与えますが、モデル適合度は変更しないでください。例えば、2 値応答の値が 0 と 1 の場合は次のとおりです。

- デフォルトでは、プロシーチャーは最後の (最大値の) カテゴリー、または参照カテゴリー (1) を作成します。この場合、モデル保存確率は特定のケースが 0 の値をとる確率を推定し、パラメーター推定値はカテゴリー 0 の尤度に関連するものとして解釈される必要があります。
- 最初の (最小値の) カテゴリー、または参照カテゴリーとして 0 を指定する場合、モデル保存確率は特定のケースが 1 の値をとる確率を推定します。
- カスタムのカテゴリーを指定し、変数に定義済みラベルがある場合、リストから値を選択して参照カテゴリーを設定できます。これは、モデルを指定している途中で、特定の変数がどのようにコード化されたか正確にわからないときに便利です。

一般化線型モデルにおける予測

「予測変数」タブでは、モデル効果の作成やオプションのオフセットの指定に使用する因子や共変量を指定できます。

因子。 因子はカテゴリー予測値です。数値または文字列です。

共変量。 共変量はスケール予測値です。数値型である必要があります。

注: バイナリー・フォーマットの 2 項応答の場合、プロシーチャーは逸脱度、および部分母集団によるカイ 2 乗適合度統計量を計算します。これは選択した因子と共変量の観測値の交差分類に基づいています。部分母集団の数に一貫性を持たせるため、プロシーチャーを複数回実行する間には同じ予測値を維持する必要があります。

オフセット。 オフセット項は「構造的な」予測値です。係数はモデルの推定対象外ですが、値 1 を持つものとみなされます。したがって、オフセットの値は単純に対象の線型予測値に追加されます。これは特にポワソン回帰モデルで有用です。このモデルでは各ケースには対象のイベントに対して様々なレベルのエクスポージャーがあります。

例えば、個々のドライバーの事故率をモデリングする場合、3 年間で過失責任事故が 1 回のドライバーと、25 年間で過失責任事故が 1 回のドライバーの間には、重要な差があります。ドライバーの経験の自然対数がオフセット項として含まれる場合、事故の回数は、対数リンクのあるポワソンまたは負の 2 項応答としてモデル化できます。

分布およびリンク・タイプのその他の組み合わせには、オフセット変数の他の変換が必要です。

一般化線型モデルのオプション

これらのオプションは、「予測変数」タブで指定されたすべての因子に適用されます。

ユーザーの欠損値。 因子は分析に含められる 1 つのケースに対して有効な値を持つ必要があります。これらを制御することで、ユーザーの欠損値を因子変数で有効な値として扱うかどうかを決定できます。

カテゴリ順。 因子の最後のレベルの決定と関係します。これは推定アルゴリズムの冗長パラメーターに関連付けられます。カテゴリ順を変更すると、因子レベル効果の値が変更されます。これらのパラメーター推定値は「最後の」レベルと比較して計算されるためです。因子は最小値から最大値に昇順で、最大値から最小値に降順で、または「データ順」でソートすることができます。つまり、データ内で検出された最初の値が最初のカテゴリを定義し、検出された最後の固有値が最後のカテゴリを定義します。

一般化線型モデルにおけるモデル

「**モデル効果の指定**」。デフォルトのモデルは切片のみであるため、他のモデル効果は明示的に指定する必要があります。ネスト項目またはネストなし項目を作成することもできます。

ネストなし項目

選択した因子や共変量について、次の項を作成できます。

「**主効果**」。選択した変数ごとに主効果の項目を作成します。

「**交互作用**」。選択したすべての変数の最上位交互作用項を作成します。

「**因子**」。選択した変数で考えられるすべての交互作用および主効果を作成します。

「**2 次まで**」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「**3 次まで**」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「**4 次まで**」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「**5 次まで**」。選択した変数で考えられるすべての 5 次交互作用を作成します。

ネスト項目

このプロシージャでは、モデルにネスト項目を作成できます。ネスト項目は、値が他の因子のレベルと交互作用しない因子または共変量の効果をモデル化するのに有用です。例えば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査するとします。各顧客は、それらの店舗のうち 1 つの店舗にのみ頻繁に通うため、顧客 効果は店舗の場所 効果内にネストされているということが可能です。

また、同じ共変量が関係する多項式の項のような交互作用効果を含めたり、ネスト項目に対して複数レベルのネストを追加したりすることができます。

制限。 ネスト項目には、次のような制限があります。

- 交互作用内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A*A$ の指定は無効です。
- ネスト効果内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A(A)$ の指定は無効です。
- 共変量内では効果をネストすることはできません。したがって、 A が因子で、 X が共変量の場合、 $A(X)$ の指定は無効です。

切片。 通常、切片はモデルに含まれます。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。

多項順序分布のモデルには単一の切片項はありませんが、その代わりに隣接するカテゴリ間の転移点を定義するしきい値パラメーターがあります。しきい値は常にモデルに含まれます。

一般化線型モデルにおける推定

「パラメーター推定」。このグループのコントロールを使用して、推定方法を指定したり、パラメーター推定値に初期値を指定することができます。

- 「**方法**」。パラメーター推定方法を選択できます。Newton-Raphson 法、Fisher スコア法、または混合型の方法の中から選択します。混合型では、Newton-Raphson 法に切り替えられる前に、Fisher スコア法の反復が実行されます。Fisher スコア法の最大反復回数に達する前に、混合型の Fisher スコア法のフェーズ中に収束に達した場合、アルゴリズムは Newton-Raphson 法で続行されます。
- 「**スケール・パラメーター法**」。スケール・パラメーター推定法を選択できます。最尤法では、スケール・パラメーターとモデル効果を組み合わせて推定します。このオプションは、応答に負の 2 項、ポワソン、または多項分布が含まれている場合は有効ではないことに注意してください。逸脱度と Pearson のカイ 2 乗オプションは、これらの統計の値からスケール・パラメーターを推定します。または、スケール・パラメーターに固定値を指定することもできます。
- 「**初期値**」。このプロシージャーでは、パラメーターに対する初期値が自動的に計算されます。または、パラメーター推定値に初期値を指定することもできます。
- 「**共分散行列**」。モデルに基づく推定量は、ヘッセ行列の一般化逆行列の負の値です。頑健推定量 (Huber/White/サンドウィッチ推定量とも呼ばれます) は「修正された」モデルに基づく推定量で、分散やリンク関数の指定が不適切な場合でも、精度の高い共分散の推定を行うことができます。

反復。使用可能なオプションは次のとおりです。

- **最大反復回数**。アルゴリズムが実行する最大反復回数です。負でない整数を指定してください。
- **最大段階 2 分**。対数尤度が増加するか、最大段階 2 分に達するまで、反復ごとにステップ・サイズが 0.5 の因子ずつ減少します。正の整数を指定してください。
- **分離データを確認**。選択すると、パラメーター推定値が固有値になるように、アルゴリズムによって検定が行われます。プロシージャーが各ケースを正しく分類するモデルを作成できたときに分離が発生します。このオプションは、バイナリー・フォーマットの多項応答および 2 項応答で使用できます。

収束基準。使用可能なオプションは次のとおりです。

- **パラメーター収束**。選択すると、パラメーター推定値の絶対変化または相対変化が、指定した値 (正の値) より少ない反復の後にアルゴリズムは停止します。
- **対数尤度収束**。選択すると、対数尤度関数の絶対変化または相対変化が、指定した値 (正の値) より少ない反復の後にアルゴリズムは停止します。
- **Hessian 収束**。「絶対」を指定した場合、Hessian 収束に基づいた統計が指定された正の値より小さい場合に収束とみなされます。「相対」を指定した場合は、指定した正の値と対数尤度の絶対値の積よりも統計が小さい場合に収束とみなされます。

「**特異性許容度**」。特異な (または反転できない) 行列には線型従属列があります。これは推定アルゴリズムに深刻な問題をもたらす可能性があります。特異性に近似する行列でさえ不適切な結果を導く可能性があるため、プロシージャーは決定要因が許容値より小さい行列を特異性があるものとして処理します。正の値を指定してください。

一般化線型モデルにおける初期値

初期値を指定する場合は、モデル内のすべてのパラメーター (冗長パラメーターを含む) に対して初期値を指定する必要があります。データ・セットでは、変数の順序は左から右に、`RowType_`、`VarName_`、`P1`、`P2`、... とする必要があります。`RowType_` と `VarName_` は文字列変数で、`P1`、`P2`、... はパラメーターの番号付きリストに対応する数値型変数です。

- 初期値はレコードで変数 *RowType_* に対して値 *EST* が指定されます。実際の初期値は、変数 *P1*、*P2*、... に指定された値になります。プロシーチャーは、*RowType_* に *EST* 以外の値があるすべてのレコードを無視します。また、*RowType_* が *EST* であるレコードが最初に現れた後は、それ以降のレコードも無視されます。
- モデルに含まれている場合は切片を、または応答に多項分布がある場合はしきい値パラメーターを、最初の初期値にする必要があります。
- スケール・パラメーターおよび負の 2 項パラメーター (応答に負の 2 項分布がある場合) を、最後の初期値に指定する必要があります。
- 分割ファイルが有効な場合、変数はファイル分割変数、または分割ファイルを作成したときに指定した順序の変数で始める必要があります。上記で説明したとおり順番に、*RowType_*、*VarName_*、*P1*、*P2*、... のようになります。分割は指定したデータ・セット内で、元のデータ・セットと同じ順序で行う必要があります。

注: 変数名 *P1*、*P2*、... は必須ではありません。プロシーチャーはパラメーターにあらゆる有効な値を受け入れます。パラメーターへの変数のマッピングは、変数名ではなく変数の位置に基づいているためです。最後のパラメーターよりも後にある変数はすべて無視されます。

初期値のファイル構造は、モデルをデータとしてエクスポートする際に使用するものと同じです。したがって、一回プロシーチャーを実行して取得した最後の値を、それ以降に実行するプロシーチャーの入力として使用することができます。

一般化線型モデルにおける統計

モデル効果。使用可能なオプションは次のとおりです。

- **分析タイプ**。作成する分析のタイプを指定します。タイプ I 分析は、一般的に、何らかの理由に基づいてモデル内の予測変数を順序付ける場合に適しています。これに対し、タイプ III はより一般的に使用できます。Wald または 尤度比統計は、カイ 2 乗統計グループでの選択に基づいて計算されます。
- **信頼区間**。50 より大きい 100 未満の信頼度レベルを指定します。Wald 区間は、パラメーターに漸近正規分布があるという仮定に基づいています。プロファイルの尤度区間はより正確になりますが、計算が効率的ではないことがあります。プロファイル尤度区間に対する許容度は、区間の計算に使用される反復アルゴリズムの停止に使用される基準です。
- **対数尤度関数**。対数尤度関数の表示形式を制御します。完全な関数には、パラメーター推定値に対して一定の追加の項が含まれます。この項はパラメーター推定に影響を与えることはなく、ソフトウェア製品によっては表示対象から除外されたままになります。

印刷。出力できる内容は以下のとおりです。

- 「**ケース処理要約**」。分析対象となるケースおよび分析対象から除外されるケースの数と割合、および「**相関データの集計**」表が表示されます。
- 「**記述統計**」。記述統計量に加え、従属変数、共変量、および因子に関する要約情報が表示されます。
- 「**モデル情報**」。データ・セット名、従属変数またはイベント変数と試行変数、オフセット変数、スケール重み変数、確率分布、およびリンク関数が表示されます。
- 「**適合度統計量**」。逸脱度とスケール逸脱度、Pearson のカイ 2 乗と尺度付き Pearson カイ 2 乗、対数尤度、赤池情報量基準 (AIC)、有限サンプル修正 AIC (AICC)、ベイズ情報量基準 (BIC)、一致 AIC (CAIC) が表示されます。
- 「**モデル要約統計量**」。モデル適合度のオムニバス検定に関する尤度比統計量や、効果ごとのタイプ I またはタイプ III の対比に関する統計量を含むモデル適合度検定が表示されます。

- 「**パラメーター推定値**」。パラメーター推定値およびそれに対応する検定統計量と信頼区間が表示されます。オプションで、未調整のパラメーター推定値に加えて、指数化されたパラメーターの推定値を表示できます。
- 「**パラメーター推定値の共分散行列**」。推定パラメーター分散共分散行列が表示されます。
- 「**パラメーター推定値の相関行列**」。推定パラメーター相関行列が表示されます。
- 「**対比係数 (L) 行列**」。デフォルトの効果の対比係数が表示されます。また、「EM 平均」タブで要求されている場合は、推定周辺平均の対比係数も表示されます。
- 「**一般の推定可能関数**」。対比係数 (L) 行列を生成するための行列が表示されます。
- 「**反復履歴**」。パラメーター推定値および対数尤度に関する反復の履歴が表示され、勾配ベクトルおよびヘッセ行列の最終評価が印刷されます。反復履歴表には、0 回目の反復 (初期推定値) から始まる反復の n 回ごとのパラメーター推定値が表示されます。ここで、 n は印刷間隔の値を表します。反復履歴が要求されている場合は、 n の値に関係なく、最後の反復が常に表示されます。
- **LaGrange 乗数検定**。スケール・パラメーターの妥当性を評価するための LaGrange 乗数検定統計量が表示されます。正規分布、ガンマ分布、逆ガウス分布、Tweedie 分布の場合、このスケール・パラメーターは、最大対数尤度比または Pearson カイ 2 乗を使用して計算されるか、固定値に設定されます。負の 2 項分布の場合は、固定補助パラメーターが検定対象となります。

一般化線型モデルにおける EM 平均

このタブでは、因子のレベルおよび因子の交互作用の推定周辺平均を表示できます。また、全体の推定平均の表示を要求することもできます。推定周辺平均は順序多項モデルでは利用できません。

「**因子と交互作用**」。このリストには、「予測変数」タブで指定された因子と「モデル」タブで指定された因子交互作用が表示されます。共変量はこのリストから除外されます。項はこのリストから直接選択することができます。または「**乗算***」ボタンを使用して交互作用の項に結合させることもできます。

「**平均値の表示**」。選択した因子と因子交互作用に対して、推定平均が計算されます。対比により、どのように仮説の検定が設定されて推定平均を比較するのかが決定されます。単純対比では、比較される他のものに対する参照カテゴリーまたは因子レベルが必要です。

- 「**ペアごと**」。ペアワイズ比較では、指定された因子または暗黙の因子のすべてのレベルでの組み合わせに対して計算が行われます。これは因子交互作用に対して使用できる唯一の対比です。
- **単純**。各レベルの平均値を、指定した水準の平均値と比較します。この種類の対比は、対照群がある場合に有用です。
- 「**偏差**」。各レベルの因子は全平均と比較されます。全平均の対比は直交ではありません。
- **差分**。各レベル (最初を除く) の平均を前のレベルの平均と比較します。逆 Helmert 対比と呼ぶこともあります。
- **Helmert**。因子の各レベル (最後のレベルを除く) の平均をその後のレベルの平均と比較します。
- **反復測定**。各レベル (最後を除く) の平均値を後続レベルの平均値と比較します。
- **多項式**。1 次効果、2 次効果、3 次効果などを比較します。最初の自由度はすべてのカテゴリーにわたる 1 次効果を含み、2 番目の自由度は 2 次効果を含みます (以下同様)。多くの場合、これらの対比は多項式トレンドの推定に使用します。

「**スケール**」。推定周辺平均は、応答に対して計算される場合、従属変数の元のスケールに基づいて行われます。または線型予測値に対して計算される場合は、リンク関数により変換された従属変数に基づいて行われます。

「多重比較の調整」。複数の対比で仮説検定を実行すると、全体の有意水準は、含められた対比の有意水準から調整されます。このグループから、調整方法を選択できます。

- 「最小有意差」。この方法では、線型対比のいくつかが帰無仮説の値とは異なるという仮説を拒否する全体的な確率は制御されません。
- *Bonferroni*. この方法では、多重比較を検定するときに、有意確率を調整します。
- *Sequential Bonferroni* (逐次 *Bonferroni*). 個々の仮説を棄却する点であり保守的ではないが、同じ全体の有意水準を維持する逐次ステップダウン棄却 *Bonferroni* 手続き。
- *Sidak*. この方法では、*Bonferroni* の方法より限界が厳しくなります。
- *Sequential Sidak* (逐次 *Sidak*). 個々の仮説を棄却する点であり保守的ではないが、同じ全体の有意水準を維持する逐次ステップダウン棄却 *Sidak* 手続き。

一般化線型モデルにおける保存

チェック・マークの付いた項目は指定した名前でも保存されます。その際、新しい変数と同じ名前を持つ既存の変数を上書きするか、名前の競合を避けるために接尾辞を付加して新しい変数名を固有にするかを選択できます。

- 「**応答の平均値の予測値**」。ケースごとのモデル予測値を、元の応答の計量値として保存します。応答分布が 2 項分布で従属変数が 2 値変数の場合、このプロシージャーでは予測確率が保存されます。応答分布が多項分布である場合、項目ラベルは「**累積予測確率**」となり、このプロシージャーでは、保存対象として指定されたカテゴリーの数まで、応答の各カテゴリー (最後のカテゴリーを除く) の累積予測確率を保存します。
- 「**応答の平均値に対する信頼区間の下限**」。応答の平均に対する信頼区間の下限を保存します。応答分布が多項分布の場合、項目ラベルは「**累積予測確率に対する信頼区間の下限**」となり、このプロシージャーでは、保存対象として指定されたカテゴリーの数まで、応答の各カテゴリー (最後のカテゴリーを除く) の下限を保存します。
- 「**応答の平均値に対する信頼区間の上限**」。応答の平均に対する信頼区間の上限を保存します。応答分布が多項分布の場合、項目ラベルは「**累積予測確率に対する信頼区間の上限**」となり、このプロシージャーでは、保存対象として指定されたカテゴリーの数まで、応答の各カテゴリー (最後のカテゴリーを除く) の上限を保存します。
- 「**予測カテゴリー**」。2 項分布および 2 値従属変数、または多項分布を含むモデルの場合は、各ケースの予測応答カテゴリーが保存されます。このオプションは他の応答分布には使用できません。
- **線型予測の予測値**。ケースごとのモデル予測値を、線型予測 (指定したリンク関数で変換された応答) の計量値として保存します。応答分布が多項分布の場合、このプロシージャーでは、保存対象として指定されたカテゴリーの数まで、応答の各カテゴリー (最後のカテゴリーを除く) の予測値を保存します。
- 「**線型予測の予測値に関する推定標準誤差**」。応答分布が多項分布の場合、このプロシージャーでは、保存対象として指定されたカテゴリーの数まで、応答の各カテゴリー (最後のカテゴリーを除く) の推定標準誤差を保存します。

次の項目は、応答分布が多項分布の場合は使用できません。

- 「**Cook の距離**」。特定のケースを回帰係数の計算から除外した場合にすべてのケースの残差がどの程度変化するかを示す指標。Cook の D が大きいときは、回帰統計量の計算からケースを除外すると係数が大きく変化することを示します。
- 「**てこ比の値**」。回帰の適合性に対する 1 つのポイントの影響度を測定します。中心化てこ比の範囲は、0 (適合性への影響なし) から $(N-1)/N$ までです。
- 「**未調整の残差**」。観測した値と、モデルによって予測された値との差。

- 「**Pearson 残差**」。生の残差の符号が付いた、Pearson カイ 2 乗統計量に対するケースの寄与率の平方根。
- 「**標準化 Pearson 残差**」。Pearson 残差に、スケール・パラメーターと「1 - ケースのてこ比」の積の逆数の平方根を乗算した値。
- 「**最大対数尤度比残差**」。生の残差の符号が付いた、最大対数尤度比統計量に対するケースの寄与率の平方根。
- 「**標準化最大対数尤度比残差**」。最大対数尤度比残差に、スケール・パラメーターと「1 - ケースのてこ比」の積の逆数の平方根を乗算した値。
- 「**尤度残差**」。生の残差の符号が付いた、標準化 Pearson 残差と標準化最大対数尤度比残差の 2 乗の (ケースのてこ比に基づく) 重み付け平均の平方根。

一般化線型モデルにおけるエクスポート

モデルをデータとしてエクスポートする。パラメーター相関、またはパラメーター推定値、標準誤差、有意確率値、自由度を持つ共分散行列を含む、IBM SPSS Statistics 形式でデータ・セットを記述します。行列ファイル内の変数の順序は次のとおりです。

- **分割変数**。この変数が使用されると、あらゆる変数が分割を定義します。
- **RowType**。値 (値ラベル) は、*COV* (共分散)、*CORR* (相関)、*EST* (パラメーター推定値)、*SE* (標準誤差)、*SIG* (有意水準)、*DF* (サンプリング計画自由度) のようになります。各モデル・パラメーターに対する行タイプ *COV* (または *CORR*) を持つ別のケースがあり、さらにその他の行タイプごとに別のケースがあります。
- **VarName**。 *P1*、*P2* などの値をとります。すべての推定モデル・パラメーター (スケールまたは負の 2 項パラメーターを除く) の番号付きリストに対応し、行タイプ *COV* または *CORR* については、「パラメーター推定値」テーブルに表示されたパラメーター文字列に対応する値ラベルが表示されます。その他の行タイプについてはセルは空白です。
- **P1、P2、...**。これらの変数は、すべてのモデル・パラメーターの番号付きリスト (スケールおよび負の 2 項パラメーターなど、適宜) に対応しています。「パラメーター推定値」テーブルに表示されたパラメーター文字列に対応する変数ラベルを持ち、行タイプに基づいた値をとります。

冗長パラメーターについては、すべての共分散は 0 に設定され、相関はシステム欠損値に設定されません。つまり、すべてのパラメーター推定値は 0 に設定され、すべての標準誤差、有意水準、残差自由度はシステム欠損値に設定されます。

スケール・パラメーターについては、共分散、相関、有意水準、自由度はシステム欠損値に設定されません。スケール・パラメーターが最尤法で推定される場合は、標準誤差が設定されます。最尤法で推定されない場合は、システム欠損値が設定されます。

負の 2 項パラメーターについては、共分散、相関、有意水準、自由度はシステム欠損値に設定されません。負の 2 項パラメーターが最尤法で推定される場合は、標準誤差が設定されます。最尤法で推定されない場合は、システム欠損値が設定されます。

分割がある場合、パラメーターのリストはすべての分割間で集計される必要があります。分割では、無関係なパラメーターが存在する場合がありますが、これは冗長と同じではありません。無関係なパラメーターについては、すべての共分散または相関、パラメーター推定値、標準誤差、有意水準、自由度はシステム欠損値に設定されます。

この行列ファイルを初期値として使用し、より詳細なモデル推定を行うことができます。このファイルは、行列ファイルを読み込むその他のプロシージャでさらにモデル分析を行うために、すぐには使用できない

ことに注意してください。ここにエクスポートされたすべての行タイプをそれらのプロシージャで承認する必要があります。そのような場合でも、この行列ファイル内のすべてのパラメーターが、ファイルを読み込むプロシージャに対して同じ意味を持つことに注意する必要があります。

XML としてモデルをエクスポートする。パラメーター推定値と、(選択されていれば) パラメーター分散共分散行列を XML (PMML) 形式で保存します。このモデル・ファイルを使用して、モデル情報を他のデータ・ファイルに適用して、スコアリングに使用できます。

GENLIN コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 数値のリストでパラメーター推定値の初期値を指定する (CRITERIA サブコマンドを使用)。
- 推定周辺平均を計算する際、平均以外の値で共変量を固定する (EMMEANS サブコマンドを使用)。
- 推定周辺平均に対するカスタムの多項式の対比を指定する (EMMEANS サブコマンドを使用)。
- 指定した対比タイプを使用して、比較のために表示された推定周辺平均に対する因子のサブセットを指定する (EMMEANS サブコマンドの TABLES および COMPARE キーワードを使用)。

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

第 7 章 一般化推定方程式

「一般化推定方程式」プロシージャでは、反復測定や、クラスター・データなど他の相関のある観測値を分析できるように、一般化線型モデルを拡張します。

例: 公衆衛生当局では、一般化推定方程式を使用して反復測定ロジスティック回帰を当てはめ、子供に与える大気汚染の影響を調査することができます。

一般化推定方程式のデータの考慮事項

データ: 応答には、スケール、度数、2 値、または試行におけるイベントがあります。因子はカテゴリー型と見なされます。共変量、スケール重み付け、およびオフセットはスケールと見なされます。被験者反復測定または被験者内反復測定の定義に使用する変数は、応答の定義に使用することはできませんが、モデル内のその他の役割を果たすことはできます。

仮定: ケースは、被験者内では従属関係にあり、被験者間では相互に独立していると見なされます。被験者内の従属関係を表す相関行列は、モデルの一部として推定されます。

一般化推定方程式の取得

メニューから次の項目を選択します。

「分析」 > 「一般化線型モデル」 > 「一般化推定方程式...」

1. 1 つ以上の被験者変数を選択します (詳細なオプションについては下記参照)。

指定した変数の値の組み合わせにより、データ・セット内の**被験者**を一意的に定義する必要があります。例えば、1 つの病院内の被験者を定義するには 1 つの患者 ID 変数のみで十分ですが、複数の病院間で患者の ID 番号が固有でない場合は、病院 ID と患者 ID の組み合わせが必要になることもあります。反復測定の設定では、被験者ごとに複数の観測値が記録されるため、各被験者がデータ・セット内の複数のケースを占めることがあります。

2. 「モデルの種類」タブで、分布およびリンク関数を指定します。
3. 「応答」タブで、従属変数を選択します。
4. 「予測」タブで、従属変数の予測に使用する因子と共変量を選択します。
5. 「モデル」タブで、選択した因子および共変量を使用してモデル効果を指定します。

オプションとして、「反復」タブで以下を指定できます。

「**被験者内変数**」。被験者内変数の値の組み合わせにより、被験者内の測定順序が定義されます。そのため、被験者内変数および被験者変数の組み合わせにより、各測定が一意的に定義されます。例えば、期間、病院 ID、および患者 ID の組み合わせにより、ケースごとに、特定の病院内の特定の患者に対する特定の外来診療が定義されます。

データ・セットのソートが既に行われているため、各被験者の反復測定がケースの連続するブロックに正しい順序で示される場合は、必ずしも被験者内変数を指定する必要はなく、「**被験者変数および被験者内変数に基づいてケースを並べ替え**」の選択を解除し、(一時的な) ソートの実行に必要な処理時間を節約することができます。通常は、測定の順序が正しくなるように被験者内変数を利用することをお勧めします。

被験者変数および被験者内変数を応答の定義に使用することはできませんが、これらの変数はモデル内の他の関数を実行することができます。例えば、病院 ID をモデル内の因子として使用することができます。

「共分散行列」。モデルに基づく推定量は、ヘッセ行列の一般化逆行列の負の値です。頑健推定量 (Huber/White/サンドウィッチ推定量とも呼ばれます) は「修正された」モデルに基づく推定量で、作業相関行列の指定が不適切な場合でも、精度の高い共分散の推定を行うことができます。ここでの指定内容は、一般化推定方程式の線型モデル部分に含まれるパラメーターに適用されますが、「推定」タブでの指定内容は、最初の一般化線型モデルにのみ適用されます。

「作業中の相関行列」。この相関行列は被験者内の従属関係を表します。この行列のサイズは測定数と、被験者内変数の値の組み合わせによって決まります。構造は次のいずれかを指定できます。

- 「独立」。反復測定は相関関係を持ちません。
- 「AR(1)」。反復測定は 1 次自己回帰関係を持ちます。任意の 2 つの要素間の相関は、要素が隣接する場合はローに等しく、要素が 3 番目の要素で分離されている場合はロー² となり、以下同様に続きます。 $-1 << 1$ になるように制限されます。
- 「交換可能」。この構造では、要素間に均質な相関があります。これは複合対称構造とも呼ばれます。
- 「M 従属」。連続する測定間の相関係数は同じであり、1 つ隔たりのある測定ペアの相関係数も同じであり、他の測定と $m-1$ 隔たりのある測定ペアにいたるまで関係は同様です。例えば、3 年生から 7 年生までの生徒に、毎年標準テストを実施するとします。この構造では、3 年生と 4 年生、4 年生と 5 年生、5 年生と 6 年生、6 年生と 7 年生の点数は同じ相関を持つと仮定され、3 年生と 5 年生、4 年生と 6 年生、5 年生と 7 年生は同じ相関を持ち、3 年生と 6 年生、4 年生と 7 年生は同じ相関を持つと仮定されます。 m を超えて離れている測定は、相関がないものと見なされます。この構造を選択する場合は、作業相関行列の次数より小さい m 値を指定してください。
- 「無構造」。これは完全に一般的な相関行列です。

このプロシージャーでは、非冗長パラメーターの数による相関推定値の調整がデフォルトで行われます。データにおける被験者レベルでの複製変更に対して推定値を不変にするのであれば、この調整を行わない方が適切な場合があります。

- 「最大反復回数」。一般化推定方程式アルゴリズムが実行する最大反復回数です。負でない整数を指定してください。ここでの指定内容は、一般化推定方程式の線型モデル部分に含まれるパラメーターに適用されますが、「推定」タブでの指定内容は、最初の一般化線型モデルにのみ適用されます。
- 「行列の更新」。作業中の相関行列の要素は、アルゴリズムが反復されるたびに更新されるパラメーター推定値に基づいて推定されます。作業中の相関行列がまったく更新されない場合は、推定プロセス全体を通して最初の作業中の相関行列が使用されます。この行列が更新される場合は、作業中の相関行列要素の更新の反復回数を指定することができます。1 より大きい値を指定すると、処理時間が短縮されることがあります。

「収束基準」。ここでの指定内容は、一般化推定方程式の線型モデル部分に含まれるパラメーターに適用されますが、「推定」タブでの指定内容は、最初の一般化線型モデルにのみ適用されます。

- 「パラメーター収束」。選択すると、パラメーター推定値の絶対変化または相対変化が、指定した値 (正の値) より少ない反復の後にアルゴリズムは停止します。
- 「Hessian 収束」。Hessian に基づく統計値が指定した値 (正の値) よりも小さい場合に、収束が想定されます。

一般化推定方程式におけるモデルの種類

「モデルの種類」タブでは、モデルに対して分布およびリンク関数を指定でき、応答の種類によって分類されているいくつかの一般的なモデルに対してショートカットが提供されます。

モデル・タイプ

スケール応答。使用可能なオプションは次のとおりです。

- **線型**。分布に「正規」、リンク関数に「同一」を指定します。
- **対数リンクを使用するガンマ**。分布に「ガンマ」、リンク関数に「対数」を指定します。

順序応答。使用可能なオプションは次のとおりです。

- **順序ロジスティック**。分布に「多項 (順序)」、リンク関数に「累積ロジット」を指定します。
- **順序プロビット**。分布に「多項 (順序)」、リンク関数に「累積プロビット」を指定します。

度数。使用可能なオプションは次のとおりです。

- **ポワソン対数線型**。分布に「ポワソン」、リンク関数に「対数」を指定します。
- **対数リンクを使用した負の 2 項**。分布に「負の 2 項 (補助パラメーターに 1 の値)」、リンク関数に「対数」を指定します。補助パラメーターの値を推定するプロシージャーにするには、「負の 2 項分布」のカスタム・モデルを指定し、パラメーター・グループで「推定値」を選択します。

2 値応答またはイベント/試行データ。使用可能なオプションは次のとおりです。

- **2 値ロジスティック**。分布に「2 項」、リンク関数に「ロジット」を指定します。
- **2 値プロビット**。分布に「2 項」、リンク関数に「プロビット」を指定します。
- **打ち切り**。分布に「2 項」、リンク関数に「補ログ・マイナス・ログ」を指定します。

混合。使用可能なオプションは次のとおりです。

- **対数リンクを使用した Tweedie**。分布に「Tweedie」、リンク関数に「対数」を指定します。
- **同一リンクを使用した Tweedie**。分布に「Tweedie」、リンク関数に「同一」を指定します。

カスタム。独自の分布とリンク関数の組み合わせを指定します。

分布

この選択により、従属変数の分布を指定します。非正規分布と非同一リンク関数を指定する機能は、一般線型モデルから一般化線型モデルに移行する上での重要な改善点です。分布とリンク関数は多くの組み合わせが可能で、その中のいくつかは特定のデータ・セットに適しているため、経験的理論の考察、または最も適合すると考えられる組み合わせから選択を導き出すことができます。

- **2 項**。この分布はその応答が 2 値応答またはイベント数を表す変数にのみ適しています。
- **ガンマ**。この分布は、正の値が大きくなるほどゆがむ正のスケール値を持つ変数に適しています。データ値が 0 以下または欠損値である場合、対応するケースは分析には使用されません。
- **逆ガウス**。この分布は、正の値が大きくなるほどゆがむ正のスケール値を持つ変数に適しています。データ値が 0 以下または欠損値である場合、対応するケースは分析には使用されません。
- **負の 2 項**。この分布は k の成功を観測するために必要な試行回数として考えることができ、負ではない整数値の変数に適しています。データ値が非整数である、0 未満である、または欠損値の場合、対応するケースは分析には使用されません。負の 2 項分布の補助パラメーターの値は、0 以上の数値ですが、パラメーターに固定値を設定し、プロシージャーで推定することも可能です。補助パラメーターが 0 に設定されると、この分布を使用することはポワソン分布を使用することに相当します。

- **正規**。これは、値が対称で、中心 (平均) 値に関してベル型の分布であるスケール変数に適しています。従属変数は数値型である必要があります。
- **ポワソン**。この分布は一定期間の対象のイベントの発生回数として考えることができ、負ではない整数値の変数に適しています。データ値が非整数である、0 未満である、または欠損値の場合、対応するケースは分析には使用されません。
- **Tweedie**。この分布は、ガンマ分布のポワソン混合により表現できる変数に適しています。分布は連続分布 (負ではない実数をとる) および離散分布 (単一の値での正の確率質量、0) のプロパティを組み合わせるという意味から「混合」です。従属変数は数値型で 0 以上のデータ値である必要があります。データ値が 0 未満である、または欠損値の場合、対応するケースは分析には使用されません。Tweedie 分布のパラメーターの固定値は、1 より大きい 2 未満の数値です。
- **多項**。この分布は順序型応答を表す変数に適しています。従属変数は数値または文字列のどちらかで、少なくとも 2 つの異なる有効データ値を持つ必要があります。

リンク関数

リンク関数は従属変数の変換を行い、それによりモデルの推定ができます。次の関数を使用することができます。

- **同一**。 $f(x)=x$ 。従属変数は変換されません。このリンクはどの分布でも使用できます。
- **補ログ・マイナス・ログ**。 $f(x)=\log(-\log(1-x))$ 。これは 2 項分布にのみ適しています。
- **累積コーチット**。 $f(x) = \tan(\pi (x - 0.5))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- **累積補ログ・マイナス・ログ**。 $f(x)=\ln(-\ln(1-x))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- **累積ロジット**。 $f(x)=\ln(x / (1-x))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- **累積の負ログ・マイナス・ログ**。 $f(x)=-\ln(-\ln(x))$ 。応答の各カテゴリーの累積確率に適用されます。これは多項分布にのみ適しています。
- **累積プロビット**。 $f(x)=\Phi^{-1}(x)$ 。応答の各カテゴリーの累積確率に適用されます。 Φ^{-1} は逆標準正規累積分布関数です。これは多項分布にのみ適しています。
- **対数**。 $f(x)=\log(x)$ 。このリンクはどの分布でも使用できます。
- **対数補完**。 $f(x)=\log(1-x)$ 。これは 2 項分布にのみ適しています。
- **ロジット**。 $f(x)=\log(x / (1-x))$ 。これは 2 項分布にのみ適しています。
- **負の 2 項**。 $f(x)=\log(x / (x+k^{-1}))$ 。 k は負の 2 項分布の補助パラメーターです。これは負の 2 項分布にのみ適しています。
- **負ログ・マイナス・ログ**。 $f(x)=-\log(-\log(x))$ 。これは 2 項分布にのみ適しています。
- **オズベキ乗**。 $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$ ($\alpha \neq 0$ の場合)。 $f(x)=\log(x)$ ($\alpha=0$ の場合)。 α には数値を指定する必要があります、その数値は実数である必要があります。これは 2 項分布にのみ適しています。
- **プロビット**。 $f(x)=\Phi^{-1}(x)$ 。 Φ^{-1} は逆標準正規累積分布関数です。これは 2 項分布にのみ適しています。
- **べき乗**。 $f(x)=x^\alpha$ 、 $\alpha \neq 0$ の場合。 $f(x)=\log(x)$ 、 $\alpha=0$ の場合。 α には数値を指定する必要があります、その数値は実数である必要があります。このリンクはどの分布でも使用できます。

一般化推定方程式における応答

多くの場合、従属変数を簡単に指定できます。ただし、2 値のみをとる変数や試行で発生するイベントを記録する応答については、特別な注意が必要です。

- **2 値応答**。従属変数が 2 値のみをとる場合、パラメーター推定値に参照カテゴリーを指定できます。2 値応答変数は数値型または文字列型です。
- **一連の試行内で発生するイベント数**。応答が一連の試行内で発生するイベント数の場合、従属変数にはイベント数が含まれ、試行回数を含む追加の変数を選択できます。ただし、試行回数がすべての被験者間で同じ場合、試行回数は固定値を使用して指定できます。試行回数は各ケースのイベント数以上である必要があります。イベント数は負でない整数、試行回数は正の整数である必要があります。

順序多項モデルの場合、応答のカテゴリー順を、昇順、降順、またはデータ順 (データ順は、データで最初に検出された値が最初のカテゴリーを定義し、最後に検出された値が最後のカテゴリーを定義することを意味します) に指定できます。

「**スケール重み**」。スケール・パラメーターは、応答の分散に関連する推定モデル・パラメーターです。スケール重みは観測ごとに異なる「既知」の値です。スケール重み変数が指定されている場合、応答の分散に関するスケール・パラメーターは、観測ごとにその変数で除算されます。0 以下または欠損しているスケール重み値があるケースは、分析では使用されません。

一般化推定方程式の参照カテゴリー

2 値応答については、従属変数に参照カテゴリーを選択できます。これはパラメーター推定値や保存値などの特定の出力に影響を与えますが、モデル適合度は変更しないでください。例えば、2 値応答の値が 0 と 1 の場合は次のとおりです。

- デフォルトでは、プロシーチャーは最後の (最大値の) カテゴリー、または参照カテゴリー (1) を作成します。この場合、モデル保存確率は特定のケースが 0 の値をとる確率を推定し、パラメーター推定値はカテゴリー 0 の尤度に関連するものとして解釈される必要があります。
- 最初の (最小値の) カテゴリー、または参照カテゴリーとして 0 を指定する場合、モデル保存確率は特定のケースが 1 の値をとる確率を推定します。
- カスタムのカテゴリーを指定し、変数に定義済みラベルがある場合、リストから値を選択して参照カテゴリーを設定できます。これは、モデルを指定している途中で、特定の変数がどのようにコード化されたか正確にわからないときに便利です。

一般化推定方程式における予測

「**予測変数**」タブでは、モデル効果の作成やオプションのオフセットの指定に使用する因子や共変量を指定できます。

因子。因子はカテゴリー予測値です。数値または文字列です。

共変量。共変量はスケール予測値です。数値型である必要があります。

注: バイナリー・フォーマットの 2 項応答の場合、プロシーチャーは逸脱度、および部分母集団によるカイ 2 乗適合度統計量を計算します。これは選択した因子と共変量の観測値の交差分類に基づいています。部分母集団の数に一貫性を持たせるため、プロシーチャーを複数回実行する間は同じ予測値を維持することが必要です。

オフセット。オフセット項は「構造的な」予測値です。係数はモデルの推定対象外ですが、値 1 を持つものとみなされます。したがって、オフセットの値は単純に対象の線型予測値に追加されます。これは特にポワソン回帰モデルで有用です。このモデルでは各ケースには対象のイベントに対して様々なレベルのエクスポージャーがあります。

例えば、個々のドライバーの事故率をモデリングする場合、3年間で過失責任事故が1回のドライバーと、25年間で過失責任事故が1回のドライバーの間には、重要な差があります。ドライバーの経験の自然対数がオフセット項として含まれる場合、事故の回数は、対数リンクのあるポワソンまたは負の2項応答としてモデル化できます。

分布およびリンク・タイプのその他の組み合わせには、オフセット変数の他の変換が必要です。

一般化推定方程式のオプション

これらのオプションは、「予測変数」タブで指定されたすべての因子に適用されます。

ユーザーの欠損値。 因子は分析に含まれる1つのケースに対して有効な値を持つ必要があります。これらを制御することで、ユーザーの欠損値を因子変数で有効な値として扱うかどうかを決定できます。

カテゴリー順。 因子の最後のレベルの決定と関係します。これは推定アルゴリズムの冗長パラメーターに関連付けられます。カテゴリー順を変更すると、因子レベル効果の値が変更されます。これらのパラメーター推定値は「最後の」レベルと比較して計算されるためです。因子は最小値から最大値に昇順で、最大値から最小値に降順で、または「データ順」でソートすることができます。つまり、データ内で検出された最初の値が最初のカテゴリーを定義し、検出された最後の固有値が最後のカテゴリーを定義します。

一般化推定方程式におけるモデル

「**モデル効果の指定**」。デフォルトのモデルは切片のみであるため、他のモデル効果は明示的に指定する必要があります。ネスト項目またはネストなし項目を作成することもできます。

ネストなし項目

選択した因子や共変量について、次の項を作成できます。

「**主効果**」。選択した変数ごとに主効果の項目を作成します。

「**交互作用**」。選択したすべての変数の最上位交互作用項を作成します。

「**因子**」。選択した変数で考えられるすべての交互作用および主効果を作成します。

「**2次まで**」。選択した変数で考えられるすべての2次交互作用を作成します。

「**3次まで**」。選択した変数で考えられるすべての3次交互作用を作成します。

「**4次まで**」。選択した変数で考えられるすべての4次交互作用を作成します。

「**5次まで**」。選択した変数で考えられるすべての5次交互作用を作成します。

ネスト項目

このプロシージャでは、モデルにネスト項目を作成できます。ネスト項目は、値が他の因子のレベルと交互作用しない因子または共変量の効果をモデル化するのに有用です。例えば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査するとします。各顧客は、それらの店舗のうち1つの店舗にのみ頻繁に通うため、顧客効果は店舗の場所効果内にネストされているとすることができます。

さらに、交互作用効果を含めたり、ネスト項目に対して複数レベルのネストを追加することができます。

制限。 ネスト項目には、次のような制限があります。

- 交互作用内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A*A$ の指定は無効です。
- ネスト効果内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A(A)$ の指定は無効です。
- 共変量内では効果をネストすることはできません。したがって、 A が因子で、 X が共変量の場合、 $A(X)$ の指定は無効です。

切片。通常、切片はモデルに含まれます。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。

多項順序分布のモデルには単一の切片項はありませんが、その代わりに隣接するカテゴリ間の転移点を定義するしきい値パラメーターがあります。しきい値は常にモデルに含まれます。

一般化推定方程式における推定

「パラメーター推定」。このグループのコントロールを使用して、推定方法を指定したり、パラメーター推定値に初期値を指定することができます。

- 「**方法**」。パラメーターの推定方法を選択できます。Newton-Raphson 法、Fisher スコア法、または混合型の方法の中から選択します。混合型では、Newton-Raphson 法に切り替えられる前に、Fisher スコア法の反復が実行されます。Fisher スコア法の最大反復回数に達する前に、混合型の Fisher スコア法のフェーズ中に収束に達した場合、アルゴリズムは Newton-Raphson 法で続行されます。
- 「**スケール・パラメーター法**」。スケール・パラメーター推定法を選択できます。

最尤法では、スケール・パラメーターとモデル効果を同時に推定します。このオプションは、応答に負の 2 項分布、ポワソン分布、または 2 項分布が含まれている場合は有効ではないことに注意してください。一般化推定方程式には尤度という概念は取り入れられないため、ここでの指定内容は最初の一般化線型モデルにのみ適用されます。その後、このスケール・パラメーター推定値は一般化推定方程式に渡されます。一般化推定方程式では、Pearson カイ 2 乗をその自由度で割った値で、スケール・パラメーターが更新されます。

逸脱オプションおよび Pearson カイ 2 乗オプションでは、最初の一般化線型モデルにおけるこれらの統計値に基づいて、スケール・パラメーターの推定が行われます。このスケール・パラメーター推定値は一般化推定方程式に渡されます。一般化推定方程式では、このスケール・パラメーター推定値が固定値として扱われます。

または、スケール・パラメーターに固定値を指定します。これは、最初の一般化線型モデルおよび一般化推定方程式の推定時に固定値として扱われます。

- 「**初期値**」。このプロシージャでは、パラメーターに対する初期値が自動的に計算されます。または、パラメーター推定値に初期値を指定することもできます。

このタブで指定された反復および収束基準は、最初の一般化線型モデルにのみ適用できます。一般化推定方程式の当てはめに使用される推定基準については、「反復」タブの記述を参照してください。

反復。使用可能なオプションは次のとおりです。

- **最大反復回数**。アルゴリズムが実行する最大反復回数です。負でない整数を指定してください。
- **最大段階 2 分**。対数尤度が増加するか、最大段階 2 分に達するまで、反復ごとにステップ・サイズが 0.5 の因子ずつ減少します。正の整数を指定してください。

- **分離データを確認。** 選択すると、パラメーター推定値が固有値になるように、アルゴリズムによって検定が行われます。プロシージャーが各ケースを正しく分類するモデルを作成できたときに分離が発生します。このオプションは、バイナリー・フォーマットの多項応答および 2 項応答で使用できます。

収束基準。 使用可能なオプションは次のとおりです。

- **パラメーター収束。** 選択すると、パラメーター推定値の絶対変化または相対変化が、指定した値 (正の値) より少ない反復の後にアルゴリズムは停止します。
- **対数尤度収束。** 選択すると、対数尤度関数の絶対変化または相対変化が、指定した値 (正の値) より少ない反復の後にアルゴリズムは停止します。
- **Hessian 収束。** 「絶対」を指定した場合、Hessian 収束に基づいた統計が指定された正の値より小さい場合に収束とみなされます。「相対」を指定した場合は、指定した正の値と対数尤度の絶対値の積よりも統計が小さい場合に収束とみなされます。

「特異性許容度」。特異な (または反転できない) 行列には線型従属列があります。これは推定アルゴリズムに深刻な問題をもたらす可能性があります。特異性に近似する行列でさえ不適切な結果を導く可能性があるため、プロシージャーは決定要因が許容値より小さい行列を特異性があるものとして処理します。正の値を指定してください。

一般化推定方程式の初期値

このプロシージャーは、初期の一般化線型モデルを推定します。このモデルの推定値は、一般化推定方程式の線型モデル部分におけるパラメーター推定値の初期値として使用されます。行列の要素はパラメーター推定値に基づいているため、作業中の相関行列に対する初期値は必要ありません。このダイアログ・ボックスで指定される初期値は、「推定」タブの最大反復数を 0 に設定しない限りは、一般化推定方程式ではなく初期の一般化線型モデルの出発点として使用されます。

初期値を指定する場合は、モデル内のすべてのパラメーター (冗長パラメーターを含む) に対して初期値を指定する必要があります。データ・セットでは、変数の順序は左から右に、*RowType_*、*VarName_*、*P1*、*P2*、... とする必要があります。*RowType_* と *VarName_* は文字列変数で、*P1*、*P2*、... はパラメーターの番号付きリストに対応する数値型変数です。

- 初期値はレコードで変数 *RowType_* に対して値 *EST* が指定されます。実際の初期値は、変数 *P1*、*P2*、... に指定された値になります。プロシージャーは、*RowType_* に *EST* 以外の値があるすべてのレコードを無視します。また、*RowType_* が *EST* であるレコードが最初に現れた後は、それ以降のレコードも無視されます。
- モデルに含まれている場合は切片を、または応答に多項分布がある場合はしきい値パラメーターを、最初の初期値にする必要があります。
- スケール・パラメーターおよび負の 2 項パラメーター (応答に負の 2 項分布がある場合) を、最後の初期値に指定する必要があります。
- 分割ファイルが有効な場合、変数はファイル分割変数、または分割ファイルを作成したときに指定した順序の変数で始める必要があります。上記で説明したとおり順番に、*RowType_*、*VarName_*、*P1*、*P2*、... のようになります。分割は指定したデータ・セット内で、元のデータ・セットと同じ順序で行う必要があります。

注: 変数名 *P1*、*P2*、... は必須ではありません。プロシージャーはパラメーターにあらゆる有効な値を受け入れます。パラメーターへの変数のマッピングは、変数名ではなく変数の位置に基づいているためです。最後のパラメーターよりも後にある変数はすべて無視されます。

初期値のファイル構造は、モデルをデータとしてエクスポートする際に使用するものと同じです。したがって、一回プロシージャーを実行して取得した最後の値を、それ以降に実行するプロシージャーの入力として使用することができます。

一般化推定方程式における統計

モデル効果。使用可能なオプションは次のとおりです。

- 「分析タイプ」。モデル効果を検定するために生成する分析タイプを指定します。タイプ I 分析は、一般的に、何らかの理由に基づいてモデル内の予測変数を順序付ける場合に適しています。これに対し、タイプ III はより一般的に使用できます。Wald または一般化スコア統計量は、カイ 2 乗統計量グループでの選択に基づいて計算されます。
- 「信頼区間」。50 より大きい 100 未満の信頼度レベルを指定します。Wald 区間は常に選択されたカイ 2 乗統計量のタイプにかかわらず生成され、パラメーターに漸近正規分布が適用されるという仮定に基づきます。
- 「対数疑似尤度関数」。対数疑似尤度関数の表示形式を制御します。完全な関数には、パラメーター推定値に対して一定の追加の項が含まれます。この項はパラメーター推定に影響を与えることはなく、ソフトウェア製品によっては表示対象から除外されたままになります。

印刷。出力できる内容は以下のとおりです。

- 「ケース処理要約」。分析対象となるケースおよび分析対象から除外されるケースの数と割合、および「相関データの集計」表が表示されます。
- 「記述統計」。記述統計量に加え、従属変数、共変量、および因子に関する要約情報が表示されます。
- 「モデル情報」。データ・セット名、従属変数またはイベント変数と試行変数、オフセット変数、スケール重み変数、確率分布、およびリンク関数が表示されます。
- 「適合度統計量」。モデル選択に対する赤池情報量基準の 2 つの拡張を示します。1 つの独立モデル基準の準尤度 (QIC) で最適な相関構造を選択し、もう 1 つの QIC 基準で予測の最適サブセットを選択します。
- 「モデル要約統計量」。モデル適合度のオムニバス検定に関する尤度比統計量や、効果ごとのタイプ I またはタイプ III の対比に関する統計量を含むモデル適合度検定が表示されます。
- 「パラメーター推定値」。パラメーター推定値およびそれに対応する検定統計量と信頼区間が表示されます。オプションで、未調整のパラメーター推定値に加えて、指数化されたパラメーターの推定値を表示できます。
- 「パラメーター推定値の共分散行列」。推定パラメーター分散共分散行列が表示されます。
- 「パラメーター推定値の相関行列」。推定パラメーター相関行列が表示されます。
- 「対比係数 (L) 行列」。デフォルトの効果の対比係数が表示されます。また、「EM 平均」タブで要求されている場合は、推定周辺平均の対比係数も表示されます。
- 「一般の推定可能関数」。対比係数 (L) 行列を生成するための行列が表示されます。
- 「反復履歴」。パラメーター推定値および対数尤度に関する反復の履歴が表示され、勾配ベクトルおよびヘッセ行列の最終評価が印刷されます。反復履歴表には、0 回目の反復 (初期推定値) から始まる反復の n 回ごとのパラメーター推定値が表示されます。ここで、 n は印刷間隔の値を表します。反復履歴が要求されている場合は、 n の値に関係なく、最後の反復が常に表示されます。
- 「作業中の相関行列」。被験者内の従属関係を表す行列の値が表示されます。その構造は「反復」タブでの指定内容によって異なります。

一般化推定方程式における EM 平均

このタブでは、因子のレベルおよび因子の交互作用の推定周辺平均を表示できます。また、全体の推定平均の表示を要求することもできます。推定周辺平均は順序多項モデルでは利用できません。

「因子と交互作用」。このリストには、「予測変数」タブで指定された因子と「モデル」タブで指定された因子交互作用が表示されます。共変量はこのリストから除外されます。項はこのリストから直接選択することができます。または「乗算*」ボタンを使用して交互作用の項に結合させることもできます。

「平均値の表示」。選択した因子と因子交互作用に対して、推定平均が計算されます。対比により、どのように仮説の検定が設定されて推定平均を比較するのかが決定されます。単純対比では、比較される他のものに対する参照カテゴリーまたは因子レベルが必要です。

- 「ペアごと」。ペアワイズ比較では、指定された因子または暗黙の因子のすべてのレベルでの組み合わせに対して計算が行われます。これは因子交互作用に対して使用できる唯一の対比です。
- 単純。各レベルの平均値を、指定した水準の平均値と比較します。この種類の対比は、対照群がある場合に有用です。
- 「偏差」。各レベルの因子は全平均と比較されます。全平均の対比は直交ではありません。
- 差分。各レベル (最初を除く) の平均を前のレベルの平均と比較します。逆 Helmert 対比と呼ぶこともあります。
- Helmert。因子の各レベル (最後のレベルを除く) の平均をその後のレベルの平均と比較します。
- 反復測定。各レベル (最後を除く) の平均値を後続レベルの平均値と比較します。
- 多項式。1 次効果、2 次効果、3 次効果などを比較します。最初の自由度はすべてのカテゴリーにわたる 1 次効果を含み、2 番目の自由度は 2 次効果を含みます (以下同様)。多くの場合、これらの対比は多項式トレンドの推定に使用します。

「スケール」。推定周辺平均は、応答に対して計算される場合、従属変数の元のスケールに基づいて行われます。または線型予測値に対して計算される場合は、リンク関数により変換された従属変数に基づいて行われます。

「多重比較の調整」。複数の対比で仮説検定を実行すると、全体の有意水準は、含められた対比の有意水準から調整されます。このグループから、調整方法を選択できます。

- 「最小有意差」。この方法では、線型対比のいくつかが帰無仮説の値とは異なるという仮説を拒否する全体的な確率は制御されません。
- Bonferroni。この方法では、多重比較を検定するときに、有意確率を調整します。
- Sequential Bonferroni (逐次 Bonferroni)。個々の仮説を棄却する点であまり保守的ではないが、同じ全体の有意水準を維持する逐次ステップダウン棄却 Bonferroni 手続き。
- Sidak。この方法では、Bonferroni の方法より限界が厳しくなります。
- Sequential Sidak (逐次 Sidak)。個々の仮説を棄却する点であまり保守的ではないが、同じ全体の有意水準を維持する逐次ステップダウン棄却 Sidak 手続き。

一般化推定方程式における保存

チェック・マークの付いた項目は指定した名前でも保存されます。その際、新しい変数と同じ名前を持つ既存の変数を上書きするか、名前の競合を避けるために接尾辞を付加して新しい変数名を固有にするかを選択できます。

- 「応答の平均値の予測値」。ケースごとのモデル予測値を、元の応答の計量値として保存します。応答分布が 2 項分布で従属変数が 2 値変数の場合、このプロシージャーでは予測確率が保存されます。応

答分布が多項分布である場合、項目ラベルは「累積予測確率」となり、このプロシージャーでは、保存対象として指定されたカテゴリの数まで、応答の各カテゴリ（最後のカテゴリを除く）の累積予測確率を保存します。

- 「**応答の平均値に対する信頼区間の下限**」。応答の平均に対する信頼区間の下限を保存します。応答分布が多項分布の場合、項目ラベルは「**累積予測確率に対する信頼区間の下限**」となり、このプロシージャーでは、保存対象として指定されたカテゴリの数まで、応答の各カテゴリ（最後のカテゴリを除く）の下限を保存します。
- 「**応答の平均値に対する信頼区間の上限**」。応答の平均に対する信頼区間の上限を保存します。応答分布が多項分布の場合、項目ラベルは「**累積予測確率に対する信頼区間の上限**」となり、このプロシージャーでは、保存対象として指定されたカテゴリの数まで、応答の各カテゴリ（最後のカテゴリを除く）の上限を保存します。
- 「**予測カテゴリ**」。2 項分布および 2 値従属変数、または多項分布を含むモデルの場合は、各ケースの予測応答カテゴリが保存されます。このオプションは他の応答分布には使用できません。
- **線型予測の予測値**。ケースごとのモデル予測値を、線型予測（指定したリンク関数で変換された応答）の計量値として保存します。応答分布が多項分布の場合、このプロシージャーでは、保存対象として指定されたカテゴリの数まで、応答の各カテゴリ（最後のカテゴリを除く）の予測値を保存します。
- 「**線型予測の予測値に関する推定標準誤差**」。応答分布が多項分布の場合、このプロシージャーでは、保存対象として指定されたカテゴリの数まで、応答の各カテゴリ（最後のカテゴリを除く）の推定標準誤差を保存します。

次の項目は、応答分布が多項分布の場合は使用できません。

- 「**未調整の残差**」。観測した値と、モデルによって予測された値との差。
- 「**Pearson 残差**」。生の残差の符号が付いた、Pearson カイ 2 乗統計量に対するケースの寄与率の平方根。

一般化推定方程式におけるエクスポート

モデルをデータとしてエクスポートする。パラメーター相関、またはパラメーター推定値、標準誤差、有意確率値、自由度を持つ共分散行列を含む、IBM SPSS Statistics 形式でデータ・セットを記述します。行列ファイル内の変数の順序は次のとおりです。

- **分割変数**。この変数が使用されると、あらゆる変数が分割を定義します。
- **RowType_**。値（値ラベル）は、*COV*（共分散）、*CORR*（相関）、*EST*（パラメーター推定値）、*SE*（標準誤差）、*SIG*（有意水準）、*DF*（サンプリング計画自由度）のようになります。各モデル・パラメーターに対する行タイプ *COV*（または*CORR*）を持つ別のケースがあり、さらにその他の行タイプごとに別のケースがあります。
- **VarName_**。*P1*、*P2* などの値をとります。すべての推定モデル・パラメーター（スケールまたは負の 2 項パラメーターを除く）の番号付きリストに対応し、行タイプ *COV* または *CORR* については、「パラメーター推定値」テーブルに表示されたパラメーター文字列に対応する値ラベルが表示されます。その他の行タイプについてはセルは空白です。
- **P1、P2、...**。これらの変数は、すべてのモデル・パラメーターの番号付きリスト（スケールおよび負の 2 項パラメーターなど、適宜）に対応しています。「パラメーター推定値」テーブルに表示されたパラメーター文字列に対応する変数ラベルを持ち、行タイプに基づいた値をとります。

冗長パラメーターについては、すべての共分散は 0 に設定され、相関はシステム欠損値に設定されません。つまり、すべてのパラメーター推定値は 0 に設定され、すべての標準誤差、有意水準、残差自由度はシステム欠損値に設定されます。

スケール・パラメーターについては、共分散、相関、有意水準、自由度はシステム欠損値に設定されず。スケール・パラメーターが最尤法で推定される場合は、標準誤差が設定されます。最尤法で推定されない場合は、システム欠損値が設定されます。

負の 2 項パラメーターについては、共分散、相関、有意水準、自由度はシステム欠損値に設定されず。負の 2 項パラメーターが最尤法で推定される場合は、標準誤差が設定されます。最尤法で推定されない場合は、システム欠損値が設定されます。

分割がある場合、パラメーターのリストはすべての分割間で集計される必要があります。分割では、無関係なパラメーターが存在する場合がありますが、これは冗長と同じではありません。無関係なパラメーターについては、すべての共分散または相関、パラメーター推定値、標準誤差、有意水準、自由度はシステム欠損値に設定されます。

この行列ファイルを初期値として使用し、より詳細なモデル推定を行うことができます。このファイルは、行列ファイルを読み込むその他のプロシージャーでさらにモデル分析を行うために、すぐには使用できないことに注意してください。ここにエクスポートされたすべての行タイプをそれらのプロシージャーで承認する必要があります。そのような場合でも、この行列ファイル内のすべてのパラメーターが、ファイルを読み込むプロシージャーに対して同じ意味を持つことに注意する必要があります。

XML としてモデルをエクスポートする。パラメーター推定値と、(選択されていれば) パラメーター分散共分散行列を XML (PMML) 形式で保存します。このモデル・ファイルを使用して、モデル情報を他のデータ・ファイルに適用して、スコアリングに使用できます。

GENLIN コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 数値のリストでパラメーター推定値の初期値を指定する (CRITERIA サブコマンドを使用)。
- 固定された作業中の相関行列を指定する (REPEATED サブコマンドを使用)。
- 推定周辺平均を計算する際、平均以外の値で共変量を固定する (EMMEANS サブコマンドを使用)。
- 推定周辺平均に対するカスタムの多項式の対比を指定する (EMMEANS サブコマンドを使用)。
- 指定した対比タイプを使用して、比較のために表示された推定周辺平均に対する因子のサブセットを指定する (EMMEANS サブコマンドの TABLES および COMPARE キーワードを使用)。

完全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。

第 8 章 一般化線型混合モデル

一般化線型混合モデルは、線型モデルを次のように拡張します。

- 対象は、指定したリンク関数を介して因子および共変量に線型に相関します。
- 対象は非正規分布を持つことができます。
- 観測は相関している場合があります。

一般化線型混合モデルには、単純な線型回帰から、非正規分布の縦断的データを取り扱う複雑なマルチレベル・モデルまで、さまざまなモデルがあります。

例。 地区の教育委員会は一般化線型混合モデルを使用して、実験的な教育方法が数学のスコアを向上させるのに効果的であるかどうかを判断できます。同じクラスの生徒同士は、同じ教師から教わっているため相関しているはずであり、同じ学校内のクラスも同様に相関する場合があります。そのため、異なる変動要因を構成するように学校やクラスのレベルでランダム効果を含めることができます。詳しくは、のトピックを参照してください。

医学研究者は一般化線型混合モデルを使用して、新しい抗けいれん薬により患者がてんかん性発作を起こす割合をどのくらい減らせるかを割り出すことができます。同じ患者からの反復測定は通常正の相関性があるため、いくつかのランダム効果を含む混合モデルが適しています。対象フィールド（発作の回数）は正の整数値を取るため、ポワソン分布および対数リンクを使用した一般化線型混合モデルが適切な場合があります。詳しくは、のトピックを参照してください。

テレビ、電話、およびインターネット・サービスのケーブル・プロバイダーの経営陣は、一般化線型混合モデルを使用して潜在的な顧客についての詳細を把握することができます。ありうる回答は名義型測定レベルを持つため、その会社のアナリストはランダム切片のある一般化ロジット混合モデルを使用して、特定の調査応答者の回答内の、サービス・タイプ全体（テレビ、電話、インターネット）のサービス利用に関する質問への回答間の相関を捕捉できます。詳しくは、のトピックを参照してください。

「データ構造」タブでは、観測値が相関するとき、データ・セット内のレコード間の構造的な関係を指定できます。データ・セット内のレコードが独立した観測値を示す場合、このタブでは何も指定する必要はありません。

被験者。 指定したカテゴリ・フィールドの値を組み合わせることで、データ・セット内の被験者を一意的に定義する必要があります。例えば、1 つの病院内の被験者を定義するには 1 つの患者 ID フィールドのみで十分ですが、複数の病院間で患者の ID 番号が固有でない場合は、病院 ID と患者 ID の組み合わせが必要になることもあります。反復測定の設定では、各被験者に対して複数の観測値が記録されるため、データ・セット内には被験者ごとに複数のレコードが存在する可能性があります。

被験者は観測用の単位で、他の被験者から独立していると見なされます。例えば医学研究では、ある患者の血圧測定値は、他の患者の測定値からは独立していると見なすことができます。被験者ごとに反復測定を行い、それらの観測値間の相関をモデル化する場合、被験者の定義は特に重要になります。例えば、病院に連続で通院している間のある患者の血圧測定値は相関していることが予想されます。

「データ構造」タブで「被験者」として指定されているフィールドはすべて、残差共分散構造の被験者を定義するのに使用されます。また、ランダム効果ブロック上のランダム効果共分散構造の被験者を定義するために可能なフィールドのリストを表示します。

反復測定。ここで指定するフィールドは、反復観測値を特定するために使用されます。例えば、週 という 1 つの変数を使用して、医学研究において 10 週間の観測値を識別できます。また、月 と日 を共に使用すると、1 年間にわたって毎日の観測値を識別できます。

共分散グループの定義。ここで指定するカテゴリー・フィールドは、反復効果共分散パラメーターの独立したセットを定義します。グループ化フィールドの交差分類により定義される各カテゴリーに対して 1 つです。すべての被験者は同じ共分散のタイプを持ちます。つまり、同じ共分散グループ内の被験者は、パラメーターに同じ値を持ちます。

反復共分散タイプ。残差に対する共分散構造を指定します。使用可能な構造は次のとおりです。

- 1 次自己回帰 (AR1)
- 自己回帰移動平均 (1,1) (ARMA11)
- 複合対称
- 対角
- スケーリングされた単位
- Toeplitz
- 無構造
- 分散成分

一般化線型混合モデルの取得

この機能を使用するには Advanced Statistics 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」 > 「混合モデル」 > 「一般化線型...」

1. 「データ構造」タブでデータ・セットの被験者の構造を定義します。
2. 「フィールドと効果」タブで、1 つの対象を指定する必要があります。測定レベル、イベント/試行は任意に指定できますが、イベントおよび試行の指定は連続型である必要があります。オプションとして、分布とリンク関数、固定効果、および任意のランダム効果ブロック、オフセット、または分析の重みを指定します。
3. 「作成オプション」をクリックして、オプションの作成設定を指定します。
4. 「モデル オプション」をクリックして、アクティブ・データ・セットにスコアを保存し、モデルを外部ファイルにエクスポートします。
5. 「実行」をクリックして、プロシージャーを実行し、モデル・オブジェクトを作成します。

対象

これらの設定では、リンク関数を使用して、対象、対象の分布、対象と予測値の関係を定義します。

対象。対象は必須です。対象は任意の測定レベルを持つことができ、対象の測定レベルはどの分布とリンク関数が適切であるかを制限します。

- **分母に試行数を使用する。**対象の応答が一連の試行内で発生するイベント数である場合、対象フィールドにはイベント数が含まれます。また試行回数を含んでいる追加フィールドを選択できます。例えば、新しい農薬を検定する際は、さまざまな濃度の農薬をアリのサンプルに噴霧し、死んだアリの数と各サンプル内のアリの数を記録します。この場合、死んだアリの数を記録するフィールドは対象 (イベント)

フィールドとして指定する必要がある、各サンプル内のアリの数を記録するフィールドは、試行フィールドとして指定する必要があります。アリの数が各サンプルに対して同じである場合、試行回数は固定値を使用して指定することができます。

試行回数は、各レコードのイベント数以上である必要があります。イベント数は負でない整数、試行回数は正の整数である必要があります。

- **参照カテゴリーのカスタマイズ。**カテゴリー対象に参照カテゴリーを選択できます。これは、パラメーター推定値などの特定の出力に影響を与えますが、モデル適合度は変更しないでください。例えば、対象がデフォルトで 0、1、2 の値をとる場合、プロシージャは最後の (最大値の) カテゴリー、すなわち 2 を参照カテゴリーとします。この場合、パラメーター推定値はカテゴリー 2 の尤度に相対するカテゴリー 0 または カテゴリー 1 の尤度に関連するものとして解釈される必要があります。カスタムのカテゴリーを指定し、対象に定義済みラベルがある場合、リストから値を選択して参照カテゴリーを設定できます。これは、モデルを指定している途中で、特定のフィールドがどのようにコード化されたか正確にわからないときに便利です。

対象の分布および線型モデルとの関連 (リンク)。予測値の値を指定することで、モデルは指定した形状に従う対象の値の分布、および指定したリンク関数を使用して予測値と線型に関連する対象の値を予測します。いくつかの一般的なモデルへのショートカットが用意されています。または短いリストでは表示されていない、当てはめたい特定の分布やリンク関数の組み合わせがある場合は、「**カスタム**」設定を選択します。

- **線型モデル。**同一リンクを持つ正規分布を指定します。これは、線型回帰または分散分析モデルを使用して対象が予測される際に有用です。
- **ガンマ回帰。**対数リンクを持つガンマ分布を指定します。これは、対象にすべての正の値が含まれ、値が大きくなるほどゆがむ場合に使用されます。
- **対数線型。**対数リンクを持つポワソン分布を指定します。これは、対象が一定期間内の出現回数を表すときに使用されます。
- **負の 2 項回帰。**対数リンクを持つ負の 2 項分布を指定します。これは、対象と分母が k の成功を観測するために必要な試行回数を表すときに使用されます。
- **多項ロジスティック回帰。**多項分布を指定します。これは対象が複数カテゴリーの応答である場合に使用されます。累積ロジット・リンク (順序結果) または一般化ロジット・リンク (マルチカテゴリー名義型レスポンス) のいずれかを使用します。
- **2 値ロジスティック回帰。**ロジット・リンクを持つ 2 項分布を指定します。これは対象がロジスティック回帰モデルで予測される 2 値応答である場合に使用されます。
- **2 値プロビット。**プロビット・リンクを持つ 2 項分布を指定します。これは対象が基礎の正規分布を使用した 2 値応答である場合に使用されます。
- **打ち切り。**補ログ・マイナス・ログ・リンクを持つ 2 項分布を指定します。これは終了イベントのない観測がある場合の生存分析で有用です。

分布

この選択により、対象の分布を指定します。非正規分布と非同リンク関数を指定する機能は、線型混合モデルから一般化線型混合モデルに移行する上での重要な改善点です。分布とリンク関数は多くの組み合わせが可能で、その中のいくつかは特定のデータ・セットに適しているため、経験的理論の考察、または最も適合すると考えられる組み合わせから選択を導き出すことができます。

- **2 項。**この分布はその応答が 2 値応答またはイベント数である対象にのみ適しています。
- **ガンマ。**この分布は、正の値が大きくなるほどゆがむ正のスケール値を持つ対象に適しています。データ値が 0 以下または欠損値である場合、対応するケースは分析には使用されません。

- **逆ガウス**。この分布は、正の値が大きくなるほどゆがむ正のスケール値を持つ対象に適しています。データ値が 0 以下または欠損値である場合、対応するケースは分析には使用されません。
- **多項**。この分布は、複数カテゴリの応答を表す対象に適しています。モデルの形式は、対象の測定レベルによって異なります。

名義型対象は、結果的に名義型多項モデルになります。その中では、モデル・パラメーターの別のセットが、対象の各カテゴリ（参照カテゴリ以外）に対して推定されます。特定の予測値に対するパラメーター推定値は、その予測値と参照カテゴリに相対する対象の各カテゴリの尤度の関係を示します。

順序型対象は、結果的に順序型多項モデルになります。従来型の切片項は、対象のカテゴリの累積確率に関連するしきい値パラメーターのセットで置き換えられます。

- **負の 2 項**。負の 2 項回帰は対数リンクを持つ負の 2 項分布を使用します。これは対象が顕著な差異がある出現回数を表す際に使用されます。
- **正規**。これは、値が対称で、中心（平均）値に関してベル型の分布である連続型対象に適しています。
- **ポワソン**。この分布は一定期間の対象のイベントの発生回数として考えることができ、負ではない整数値の変数に適しています。データ値が非整数である、0 未満である、または欠損値の場合、対応するケースは分析には使用されません。

リンク関数

リンク関数は対象を変換するもので、モデルの推定ができるようになります。次の関数を使用することができます。

- **同一**。 $f(x)=x$ 。対象は変換されません。このリンクは多項以外のどの分布でも使用できます。
- **補ログ・マイナス・ログ**。 $f(x)=\log(-\log(1-x))$ 。これは 2 項分布または多項分布にのみ適しています。
- **コーチット**。 $f(x) = \tan(\pi (x - 0.5))$ 。これは 2 項分布または多項分布にのみ適しています。
- **対数**。 $f(x)=\log(x)$ 。このリンクは多項以外のどの分布でも使用できます。
- **対数補完**。 $f(x)=\log(1-x)$ 。これは 2 項分布にのみ適しています。
- **ロジット**。 $f(x)=\log(x / (1-x))$ 。これは 2 項分布または多項分布にのみ適しています。
- **負ログ・マイナス・ログ**。 $f(x)=-\log(-\log(x))$ 。これは 2 項分布または多項分布にのみ適しています。
- **プロビット**。 $f(x)=\Phi^{-1}(x)$ 。 Φ^{-1} は逆標準正規累積分布関数です。これは 2 項分布または多項分布にのみ適しています。
- **べき乗**。 $f(x)=x^\alpha$ ($\alpha \neq 0$ の場合)。 $f(x)=\log(x)$ ($\alpha=0$ の場合)。 α には数値を指定する必要があるため、その数値は実数である必要があります。このリンクは多項以外のどの分布でも使用できます。

固定効果

固定効果因子は一般的に、そのフィールドの対象の値がすべてデータ・セットに表示され、スコアリングに使用できるフィールドとして考えられます。デフォルトでは、ダイアログの他の場所で指定されていない定義済みの入力ロールがあるフィールドは、モデルの固定効果部分に入力されます。カテゴリ（名義、順序）フィールドはモデル内で因子として使用され、連続型フィールドは共変量として使用されます。





1 つ以上のフィールドをソース・リストで選択し、効果リストにドラッグしてモデルに効果を入力します。作成される効果のタイプは、選択をドロップするホット・スポットによって異なります。

- **主**。ドロップされたフィールドが、別の主効果として効果リストの下部に表示されます。

- **2 次**。ドロップされたフィールドのすべての可能なペアが、2 次交互作用として効果リストの下部に表示されます。
- **3 次**。ドロップされたフィールドのすべての可能なトリプレットが、3 次の交互作用として効果リストの下部に表示されます。
- *。すべてのドロップされたフィールドの組み合わせが、1 つの交互作用として効果リストの下部に表示されます。

「効果ビルダー」の右側のボタンを押すと、さまざまなアクションを実行できます。

表 1. 効果ビルダー・ボタンの説明:

アイコン	説明
	削除する項目を選択して削除ボタンをクリックすることで、固定効果モデルから項目を削除できます。
	並べ替える項目を選択して上下矢印をクリックすることで、固定効果モデル内の項目を並べ替えることができます。
	
	「カスタム項目の追加」ボタンをクリックして『カスタム項目の追加』ダイアログを表示し、モデルにネスト項目を追加します。

切片を含める。通常、切片はモデルに含まれます。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。

カスタム項目の追加

このプロシージャでは、モデルにネスト項目を作成できます。ネスト項目は、値が他の因子のレベルと交互作用しない因子または共変量の効果をモデル化するのに有用です。例えば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査するとします。各顧客は、それらの店舗のうち 1 つの店舗にのみ頻繁に通うため、顧客 効果は店舗の場所 効果内にネストされているということが可能です。

また、同じ共変量が関係する多項式の項のような交互作用効果を含めたり、ネスト項目に対して複数レベルのネストを追加したりすることができます。

制限。ネスト項目には、次のような制限があります。

- 交互作用内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A*A$ の指定は無効です。
- ネスト効果内のすべての因子は固有である必要があります。したがって、 A が因子の場合、 $A(A)$ の指定は無効です。
- 共変量内では効果をネストすることはできません。したがって、 A が因子で、 X が共変量の場合、 $A(X)$ の指定は無効です。

ネスト項目の作成

1. 他の因子内でネストされた因子または共変量を選択し、矢印ボタンをクリックします。
2. 「(内部)」をクリックします。
3. 前の因子または共変量がネストされた中で因子を選択し、矢印ボタンをクリックします。
4. 「項目を追加」をクリックします。

オプションで、交互作用効果を含めたり、ネスト項目に対して複数レベルのネストを追加することができます。

ランダム効果

ランダム効果因子は、データ・ファイル内のフィールド値が、値のより大きな母集団からの無作為サンプルを考慮できるフィールドです。これらは対象の過剰なばらつきを説明するのに有用です。デフォルトでは、「データ構造」タブで複数の被験者を選択した場合、最も内側の被験者以降の被験者ごとにランダム効果ブロックが作成されます。例えば、「データ構造」タブで被験者として「学校」、「クラス」、「生徒」を選択した場合、次のランダム効果ブロックが自動的に作成されます。

- ランダム効果 1: 被験者は学校 (効果なし、切片のみ)
- ランダム効果 2: 被験者は学校 * クラス (効果なし、切片のみ)

次のような方法でランダム効果ブロックを操作できます。

1. 新しいブロックを追加するには、「**ブロックの追加...**」をクリックします。『ランダム効果ブロック』ダイアログが開きます。
2. 既存のブロックを編集するには、編集するブロックを選択して「**ブロックの編集...**」をクリックします。『ランダム効果ブロック』ダイアログが開きます。
3. 1 つ以上のブロックを削除するには、削除するブロックを選択して「削除」ボタンをクリックします。





ランダム効果ブロック

1 つ以上のフィールドをソース・リストで選択し、効果リストにドラッグしてモデルに効果を入力します。作成される効果のタイプは、選択をドロップするホット・スポットによって異なります。カテゴリー (名義、順序) フィールドはモデル内で因子として使用され、連続型フィールドは共変量として使用されます。

- **主。**ドロップされたフィールドが、別の主効果として効果リストの下部に表示されます。
- **2 次。**ドロップされたフィールドのすべての可能なペアが、2 次交互作用として効果リストの下部に表示されます。
- **3 次。**ドロップされたフィールドのすべての可能なトリプレットが、3 次の交互作用として効果リストの下部に表示されます。
- ***** すべてのドロップされたフィールドの組み合わせが、1 つの交互作用として効果リストの下部に表示されます。

「効果ビルダー」の右側のボタンを押すと、さまざまなアクションを実行できます。

表 2. 効果ビルダー・ボタンの説明:

アイコン	説明
	削除する項目を選択して削除ボタンをクリックすることで、モデルから項目を削除できます。
	並べ替える項目を選択して上下矢印をクリックすることで、モデル内の項目を並べ替えることができます。
	
	「カスタム項目の追加」ボタンをクリックして 61 ページの『カスタム項目の追加』ダイアログを表示し、モデルにネスト項目を追加します。

切片を含める。デフォルトでは、切片はランダム効果モデルに含まれません。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。

共分散グループの定義ここで指定するカテゴリ・フィールドは、ランダム効果共分散パラメーターの独立したセットを定義します。グループ化フィールドの交差分類により定義される各カテゴリに対して 1 つです。各ランダム効果ブロックに対してグループ化フィールドの異なるセットを指定することができます。すべての被験者は同じ共分散のタイプを持ちます。つまり、同じ共分散グループ内の被験者は、パラメーターに同じ値を持ちます。

被験者の組み合わせ。「データ構造」タブで事前設定された被験者の組み合わせから、ランダム効果の被験者を指定できます。例えば、学校、クラス、生徒がこの順番で「データ構造」タブで被験者として定義されている場合、「被験者の組み合わせ」ドロップダウン・リストには「なし」、「学校」、「学校 * クラス」、「学校 * クラス * 生徒」のオプションが表示されます。

ランダム効果共分散タイプ。残差に対する共分散構造を指定します。使用可能な構造は次のとおりです。

- 1 次自己回帰 (AR1)
- 自己回帰移動平均 (1,1) (ARMA11)
- 複合対称
- 対角
- スケーリングされた単位
- Toeplitz
- 無構造
- 分散成分

重み付けとオフセット

分析の重み付け。スケール・パラメーターは、応答の分散に関連する推定モデル・パラメーターです。分析の重み付けは、観測ごとに異なる「既知」の値です。分析の重み付けフィールドが指定された場合、応答の分散に関するスケール・パラメーターは、各観測に対する分析の重み付け値によって除算されます。値が 0 以下または欠損値である分析重み付け値のあるレコードは、分析では使用されません。

オフセット。オフセット項は「構造的な」予測値です。係数はモデルの推定対象外ですが、値 1 を持つものとみなされます。したがって、オフセットの値は単純に対象の線型予測値に追加されます。これは特にポワソン回帰モデルで有用です。このモデルでは各ケースには対象のイベントに対して様々なレベルのエクスポージャーがあります。

例えば、個々のドライバーの事故率をモデリングする場合、3 年間で過失責任事故が 1 回のドライバーと、25 年間で過失責任事故が 1 回のドライバーの間には、重要な差があります。ドライバーの経験の自然対数がオフセット項として含まれる場合、事故の回数は、対数リンクのあるポワソンまたは負の 2 項応答としてモデル化できます。

分布およびリンク・タイプのその他の組み合わせには、オフセット変数の他の変換が必要です。

一般的な作成オプション

これらの選択により、モデルの作成に使用する、より高度な基準が指定されます。

ソート順。これらのコントロールは、「最後の」カテゴリを決定するために、対象と因子 (カテゴリ入力) に対するカテゴリの順序を決定します。対象のソート順設定は、対象がカテゴリではない場合、またはカスタムの参照カテゴリが 58 ページの『対象』設定で指定されていない場合は無視されます。

停止規則。アルゴリズムが実行する反復の最大回数を指定できます。アルゴリズムは、内部ループと外部ループで構成される 2 重の反復プロセスを使用します。最大反復回数に指定される値は、両方のループに適用されます。負でない整数を指定してください。デフォルトは 100 です。

推定後設定。これらの設定により、表示のためにモデルの出力がどのように計算されるかが決定されます。

- **信頼度レベル。**これはモデル係数の区間推定値の計算に使用される信頼度のレベルです。0 より大きい 100 未満の値を指定してください。デフォルトは 95 です。
- **自由度。**有意差検定に対する自由度の計算方法を指定します。標本サイズが十分に大きい場合、またはデータが均衡している場合は、「**すべての検定に固定 (残差方法)**」を選択してください。選択しない場合は、モデルはスケリングされた単位または対角など、より簡単な共分散タイプを使用します。これはデフォルトです。標本サイズが小さい場合、またはデータが不均衡の場合は、「**検定全体で異なる (Satterthwaite 近似値)**」を選択してください。選択しない場合、モデルは無構造などの複雑な共分散タイプを使用します。
- **固定効果と係数の検定。**これはパラメーター推定値共分散行列を計算する方法です。モデルの仮定に反していると考えられる場合は、頑強な推定を選択してください。

推定

このモデル構築アルゴリズムでは、内部ループと外部ループで構成される 2 重の反復プロセスを使用します。以下の設定が内部ループに適用されます。

パラメーター収束。

パラメーター推定値の最大絶対変化または最大相対変化が、指定した値 (負以外でなければなりません) より小さい場合に収束とみなされます。指定された値が 0 の場合、この基準は使用されません。

対数尤度収束。

対数尤度関数の絶対変化または相対変化が、指定した値 (負以外でなければなりません) より小さい場合に収束とみなされます。指定された値が 0 の場合、この基準は使用されません。

Hessian 収束。

「絶対」を指定した場合は、Hessian に基づく統計量が、指定した値よりも小さい場合に収束とみなされます。「相対値」を指定した場合は、指定した値と対数尤度の絶対値の積よりも統計が小さい場合に収束とみなされます。指定された値が 0 の場合、この基準は使用されません。

Fisher スコア法の最大ステップ数。

負でない整数を指定してください。値 0 を指定すると Newton-Raphson 法を使用します。0 より大きい値を指定すると、反復回数が n に達するまで Fisher スコア法のアルゴリズムを使用します。ここで、 n は指定した整数で、それ以降は Newton-Raphson 法が使用されます。

特異性許容度。

この値は、特異性の検査時に許容度として使用されます。正の値を指定してください。

注: デフォルトでは、許容度が 1E-6 の最大「絶対」変化が検査される「パラメータ収束」が使用されます。この設定は、バージョン 22 より前のバージョンで取得される結果とは異なる結果を生成する場合があります。バージョン 22 より前のバージョンの結果を再現するには、「パラメータ収束」基準に「相対値」を使用し、デフォルトの許容値を 1E-6 のままにしてください。

推定平均

このタブでは、因子のレベルおよび因子の交互作用の推定周辺平均を表示できます。推定周辺平均は多項モデルでは利用できません。

項。全体がカテゴリ・フィールドから構成されている固定効果のモデル項が表示されます。モデルで推定周辺平均を作成する場合は各項を確認してください。

- **対比タイプ。**対比フィールドのレベルに使用する対比のタイプを指定します。「なし」を選択すると、対比は作成されません。「ペアごと」を選択すると、指定した因子のすべてのレベルの組み合わせに対してペアワイズ比較が作成されます。これは因子交互作用に対して使用できる唯一の対比です。「偏差」の対比は因子の各レベルを全平均と比較します。「単純」対比は、最後以外の因子の各レベルを最後のレベルと比較します。「最後の」レベルは「作成オプション」で指定された因子のソート順によって決まります。これらの対比タイプのすべてが直交ではないことに注意してください。
- **対比フィールド。**選択した対比タイプを使用して、1つの因子(比較されるレベル)を指定します。対比タイプに「なし」を選択すると、対比フィールドを選択することはできず、選択する必要もありません。

連続型フィールド。連続型フィールドのリストは、連続型フィールドを使用する固定効果の項から抽出されます。推定周辺平均を計算する際、共変量は指定した値に固定されます。平均値を選択するか、カスタムの値を指定してください。

推定平均を表示する単位。対象の元の尺度、あるいはリンク関数変換のどちらに基づいて推定周辺平均を計算するかを指定します。「元の目標スケール」は、対象の推定周辺平均を計算します。対象がイベント/試行オプションを使用して指定された場合、イベント数ではなく、イベント/試行比率に対して推定周辺平均が計算されることに注意してください。「リンク関数の変換」は、線型予測値の推定周辺平均を計算します。

多重比較の調整方法。複数の対比で仮説検定を実行すると、全体の有意水準は、含められた対比の有意水準から調整されます。これにより、調整方法を選択できます。

- **最小有意差。**この方法では、線型対比のいくつかが無効仮説の値とは異なるという仮説を拒否する全体的な確率は制御されません。
- *Sequential Bonferroni (逐次 Bonferroni)*. 個々の仮説を棄却する点であまり保守的ではないが、同じ全体の有意水準を維持する逐次ステップダウン棄却 Bonferroni 手続き。
- *Sequential Sidak (逐次 Sidak)*. 個々の仮説を棄却する点であまり保守的ではないが、同じ全体の有意水準を維持する逐次ステップダウン棄却 Sidak 手続き。

最小有意差法は、逐次 Sidak 法よりも保守的ではなく、さらに逐次 Bonferroni 法よりも保守的ではありません。つまり、最小有意差法は、逐次 Sidak と少なくとも同数の独立仮説を拒否し、さらに逐次 Bonferroni 法と少なくとも同数の独立仮説を拒否します。

保存

チェックが付けられた項目が、指定された名前でも保存されます。既存のフィールド名と競合する名前にはできません。

予測値。対象の予測値を保存します。デフォルトのフィールド名は、*PredictedValue* です。

カテゴリ対象の予測確率。対象がカテゴリの場合、このキーワードは最初の n カテゴリから、「保存する最大カテゴリ数」に指定された値までの予測確率を保存します。計算された値は順序型対象の累積確率です。デフォルトのルート名は、*PredictedProbability* です。予測カテゴリの予測確率を保存するには、信頼度を保存します (下記参照)。

信頼区間。予測値または予測確率に対する信頼区間の上限と下限を保存します。多項以外のすべての分布については、2 つの変数および *_Lower* および *_Upper* がサフィックスとして付けられた *CI* という名前のデフォルト・ルートが作成されます。

多項分布および名義型対象については、従属変数のカテゴリごとにフィールドが 1 つ作成されます。最初の n カテゴリから、「保存する最大カテゴリ数」に指定された値までの予測確率の上限と下限を保存します。デフォルトのルート名は *CI* で、デフォルトのフィールド名は、*CI_Lower_1*、*CI_Upper_1*、*CI_Lower_2*、*CI_Upper_2* などです。これは対象カテゴリの順序に対応しています。

多項分布および順序型対象については、最後を除く従属変数カテゴリごとにフィールドが 1 つ作成されます (詳しくは、63 ページの『一般的な作成オプション』のトピックを参照してください)。最初の n カテゴリから、最後を除く「保存する最大カテゴリ数」に指定された値までの累積予測確率の上限と下限を保存します。デフォルトのルート名は *CI* で、デフォルトのフィールド名は、*CI_Lower_1*、*CI_Upper_1*、*CI_Lower_2*、*CI_Upper_2* などです。これは対象カテゴリの順序に対応しています。

Pearson 残差。各レコードの Pearson の残差を保存します。これはモデル適合度の推定後の診断に使用されます。デフォルトのフィールド名は、*PearsonResidual* です。

信頼度。カテゴリ対象の予測値に信頼度を保存します。計算された信頼度は、予測値の確率 (最も高い予測確率)、または最も高い予測確率と 2 番目に高い予測確率との差に基づいています。デフォルトのフィールド名は、*Confidence* です。

モデルのエクスポート。モデルを外部 *.zip* ファイルに書き出します。このモデル・ファイルを使用して、モデル情報を他のデータ・ファイルに適用して、スコアリングに使用できます。固有で有効なファイル名を指定します。ファイルの指定が既存ファイルを指す場合、ファイルが上書きされます。

モデル・ビュー

プロシージャでは、ビューアでモデル・オブジェクトが作成されます。このオブジェクトをアクティブ化 (ダブルクリック) すると、モデルの双方向ビューが表示されます。

デフォルトでは、「モデルの要約」ビューが表示されます。別のモデル・ビューを表示するには、ビューのサムネイルから選択します。

モデル・オブジェクトの代わりに、「オプション」ダイアログ・ボックス (「編集」 > 「オプション」) の「出力」タブの「出力表示」グループの「ピボット テーブルとグラフ」を選択して、ピボット・テーブルとグラフを生成できます。以下のトピックでは、モデル・オブジェクトについて説明します。

モデルの要約

このビューはスナップショットで、モデルとその適合についてまとめた概要です。

テーブル。テーブルは『対象の設定』で指定された対象、確率分布、リンク関数を識別します。対象がイベントや試行によって定義されている場合、セルはイベント・フィールドや試行フィールド、または固定の試行回数を表示するために分割されます。また、有限サンプル補正済み赤池情報量基準 (AICC) およびベイズ情報基準 (BIC) が表示されます。

- **補正済み赤池。**-2 (制限) 対数尤度に基づいて混合モデルを選択し、比較するための指標。値が小さいほどモデルが良好であることを示します。AICC は、小さな標本サイズに適応するように AIC を「修正」します。標本サイズが大きくなるに従い、AICC は AIC に収束します。
- **ベイズ。**-2 対数尤度に基づいてモデルを選択し、比較するための指標。値が小さいほどモデルが良好であることを示します。BIC も、パラメータが多すぎるモデル (入力が大量にあるにコンプレックス モデルなど) に「ペナルティ」を科しますが、AIC より厳密に科します。

グラフ。対象がカテゴリーの場合、グラフには最終のモデルの精度が表示されます。これは正確な分類の割合です。

データ構造

このビューでは、指定したデータ構造の要約が表示され、被験者と反復測定が正しく指定されていることを確認できます。最初の被験者の観測情報が各被験者フィールドと反復測定フィールド、および対象に表示されます。さらに、各被験者フィールドと反復測定フィールドのレベルの数が表示されます。

予測と観測

イベント/試行として指定された対象をはじめとする連続型対象について、縦軸が予測値、横軸が観測値を表す分割散布図を表示します。点は 45 度の線上にあるのが理想です。つまり、このビューにはモデルにより特に不正であると予測されるレコードがあるかどうかを示されます。

分類

カテゴリー対象については、ヒート・マップの予測値に対する観測値の交差分類、および全体の正解のパーセントが表示されます。

テーブルのスタイル。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・リストからアクセスできます。

- **行パーセント。**セルの行パーセント (行合計のパーセントで表現されるセルの度数) が表示されます。これはデフォルトです。
- **セルの度数。**セルのセル度数が表示されます。ヒート・マップの陰影付けは、そのまま行パーセントをベースにします。
- **ヒート・マップ。**セルの値は表示しません。陰影付けのみ表示します。
- **圧縮。**セルの行または列のヘッダー、セルの値を表示しません。この方法は、対象にカテゴリーが多い場合に有用です。

欠損値。対象に欠損値があるレコードがある場合、レコードはすべての有効な行の下の「(欠損値)」行に表示されます。欠損値のあるレコードは、全体の正解のパーセントに寄与しません。

複数の対象。複数のカテゴリー対象がある場合、各対象は別々のテーブルに表示され、「対象」ドロップダウン・リストを使用して表示する対象を制御します。

大型テーブル。表示する対象に 100 を超えるカテゴリーがある場合、テーブルは表示されません。

固定効果

このビューには、モデルの各固定効果のサイズが表示されます。

スタイル。さまざまな表示のスタイルがあり、「スタイル」ドロップダウン・リストからアクセスできます。

- **ダイアグラム。**「固定効果」設定で指定された順番で先頭から最後まで効果がソートされたグラフです。ダイアグラムの接続線は効果有意性に基づいて重み付けが行われ、より有意のある効果 (p - 値よりも小さい) は、より太い線幅で表示されます。これはデフォルトです。
- **テーブル。**これは全体のモデルおよび個々のモデル効果に対する分散分析テーブルです。個々の効果は「固定効果」設定で指定された順番で先頭から最後までソートされます。

有意。「有意」スライダーを使用して、ビューで表示する効果を制御します。スライダーの値よりも大きい有意性を持つ効果は表示されません。これでモデルが変更されるわけではありませんが、最も重要な効果に集中することができます。デフォルトの値は 1.00 であるため、有意性に基づいてフィルターされる効果はありません。

固定係数

このビューでは、モデルの各固定係数の値が表示されます。因子 (カテゴリー予測値) はモデル内で指標コード化されているため、因子を含む効果には、一般的に複数の関連する係数があることに注意してください。冗長係数に対応するカテゴリーを除く、各カテゴリーごとに 1 つです。

スタイル。さまざまな表示のスタイルがあり、「スタイル」ドロップダウン・リストからアクセスできます。

- **ダイアグラム。**このグラフでは、最初に切片を表示し、「固定効果」設定で指定された順番で先頭から最後まで効果がソートされます。因子を含んでいる効果内では、係数はデータ値の昇順でソートされます。ダイアグラムの接続線は、係数の有意性を基に色付けと重み付けが行われ、より有意の係数 (p - 値よりも小さい) は、より太い線幅で表示されます。これはデフォルトのスタイルです。
- **テーブル。**個々のモデルの係数に対する値、有意差検定、信頼区間を表示します。切片の表示後、「固定効果」設定で指定された順番で先頭から最後まで効果がソートされます。因子を含んでいる効果内では、係数はデータ値の昇順でソートされます。

多項。多項分布が効果にある場合、「多項」ドロップダウン・リストを使用して表示する対象カテゴリーを制御します。リストの値のソート順は、「作成オプション」設定の仕様により決まります。

指数。2 値ロジスティック回帰 (2 項分布およびロジット・リンク)、名義ロジスティック回帰 (多項分布およびロジット・リンク)、負の 2 項回帰 (負の 2 項分布および対数リンク)、対数線型モデル (ポワソン分布および対数リンク) など、特定のモデル・タイプに対する指数係数推定値と信頼区間を表示します。

有意。「有意」スライダーを使用して、ビューに表示する係数を制御します。スライダーの値よりも大きい有意確率値を持つ係数は表示されません。これでモデルが変更されるわけではありませんが、最も重要な係数に集中することができます。デフォルトの値は 1.00 であるため、有意性に基づいてフィルターされる係数はありません。

ランダム効果共分散

このビューには、ランダム効果共分散行列 (**G**) が表示されます。

スタイル。さまざまな表示のスタイルがあり、「スタイル」ドロップダウン・リストからアクセスできます。

- **共分散値。**「固定効果」設定で指定された順番で先頭から最後まで効果がソートされた、共分散行列のヒート・マップです。corrgram 内の色は、キーに示されているセルの値に対応します。これはデフォルトです。
- **Corrgram。**共分散行列のヒート・マップです。
- **圧縮。**これは行と列のヘッダーを除いた共分散行列のヒート・マップです。

ブロック。複数のランダム効果ブロックがある場合は、「ブロック」ドロップダウン・リストから表示するブロックを選択します。

グループ。ランダム効果ブロックにグループ仕様がある場合は、「グループ」ドロップダウン・リストで表示するグループ・レベルを選択します。

多項。多項分布が効果にある場合、「多項」ドロップダウン・リストを使用して表示する対象カテゴリーを制御します。リストの値のソート順は、「作成オプション」設定の仕様により決まります。

共分散パラメーター

このビューには、共分散パラメーター推定値、および残差とランダム効果に関連する統計が表示されます。これらは高度ですが基本的な結果で、共分散構造が適しているかどうかについての情報を示します。

要約表。これは、残差 (**R**) およびランダム効果 (**G**) 共分散行列のパラメーター数、固定効果 (**X**) およびランダム効果 (**Z**) デザイン行列のランク (列数)、データ構造を定義する被験者フィールドにより定義される被験者数に関するクイック・リファレンスです。

共分散パラメーター・テーブル。選択した効果について、推定値、標準誤差、信頼区間が各共分散パラメーターに表示されます。表示されるパラメーター数は、効果に対する共分散構造により異なり、ランダム効果ブロックについては、ブロック内の効果数です。非対角パラメーターが有意でない場合、より簡単な共分散構造を使用することができます。

効果。ランダム効果ブロックがある場合は、「効果」ドロップダウン・リストで表示する残差またはランダム効果ブロックを選択します。残差効果は常に使用可能です。

グループ。残差またはランダム効果ブロックにグループ仕様がある場合は、「グループ」ドロップダウン・リストで表示するグループ・レベルを選択します。

多項。多項分布が効果にある場合、「多項」ドロップダウン・リストを使用して表示する対象カテゴリーを制御します。リストの値のソート順は、「作成オプション」設定の仕様により決まります。

推定平均: 有意効果

これらのグラフは、10 の「最も有意な」固定の全因子の効果を表示します。3 次の交互作用から始まり、次に 2 次の交互作用、最後に主効果を表示します。横軸が主効果 (または交互作用で最初にリストされる効果) の各値、縦軸が対象のモデル推定値を表すグラフが表示されます。交互作用の 2 番目にリストされる効果の各値に対して別の線が作成され、3 次の交互作用の 3 番目にリストされる効果の各値に対してはグラフが別途作成されます。その他すべての予測値は定数が保持されます。これは、対象の各予測値の係数の影響を視覚化するのに有用です。予測値が有意でない場合は推定平均は作成されないことに注意してください。

信頼度。これは、「作成オプション」の一部として指定された信頼度レベルを使用して、周辺平均に対する信頼限界の上限と下限を表示します。

推定平均: カスタム効果

これらはユーザーからリクエストのあった固定の全因子の効果のテーブルとグラフです。

スタイル。さまざまな表示のスタイルがあり、「スタイル」ドロップダウン・リストからアクセスできます。

- **ダイアグラム。**このスタイルは、横軸が主効果（または交互作用で最初にリストされる効果）の各値、縦軸が対象のモデル推定値を表す折れ線グラフを表示します。交互作用の 2 番目にリストされる効果の値に対して別の線が作成され、3 次の交互作用の 3 番目にリストされる効果の各値に対しては別のグラフが作成されます。その他すべての予測値は定数が保持されます。

対比が要求された場合、対比フィールドのレベルを比較するための別のグラフが表示されます。交互作用については、対比フィールド以外の各レベルの効果を組み合わせたグラフが表示されます。**ペアワイズ**対比では、これは間隔ネットワーク図です。つまり、ネットワークのノード間の距離がサンプル間の差に相当する比較テーブルを図で表現したものです。黄色の線は統計の有意差に相当し、黒の線は非有意差に相当します。ネットワークの線上にマウス・ポインタを乗せると、線で接続されたノード間の差の調整済み有意差を示すツールチップが表示されます。

全平均の対比については、縦軸が対象のモデル推定値、横軸が対比フィールドの値を表す棒グラフが表示されます。交互作用については、対比フィールド以外の各レベルの効果を組み合わせたグラフが表示されます。棒は対比フィールドの各レベルと全体の平均との差を示します。これは黒の横線で表示されます。

単純な対比については、縦軸が対象のモデル推定値、横軸が対比フィールドの値を表す棒グラフが表示されます。交互作用については、対比フィールド以外の効果の各レベルを組み合わせたグラフが表示されます。棒は対比フィールドの各レベル（最後のレベル以外）と最後のレベル間の差を示します。これは黒の横線で表示されます。

- **テーブル。**このスタイルは対象のモデル推定値、標準誤差、効果内のフィールドの各レベルの組み合わせに対する信頼区間のテーブルを表示します。その他すべての予測値は定数が保持されます。

対比が要求された場合、推定、標準誤差、有意差検定、各対比に対する信頼区間が別のテーブルに表示されます。交互作用については、対比フィールド以外の各レベルの効果の組み合わせに対する行セットが別途表示されます。さらに、全体の検定結果のテーブルが表示されます。交互作用については、対比フィールド以外の各レベルの効果の組み合わせに対する全体の検定が別途行われます。

信頼度。これは、「作成オプション」の一部として指定された信頼度レベルを使用して、周辺平均に対する信頼限界の上限と下限の表示を切り替えます。

レイアウト。これは、ペアワイズ対比ダイアグラムのレイアウトを切り替えます。円レイアウトはネットワーク・レイアウトよりも対比の表示は少ないですが、線の重なり合いを回避できます。

第 9 章 モデル選択対数線型分析

「モデル選択対数線型分析」プロシージャでは、多重クロス集計表 (分割表) を分析します。反復比例当てはめアルゴリズムを使用して、階層対数線型モデルを多次元クロス集計表に当てはめます。このプロシージャは、関連するカテゴリー変数を見つける際に役立ちます。モデルを作成する場合は、強制投入法と変数減少法を使用できます。飽和モデルの場合は、パラメーター推定値と偏連関の検定を要求できます。飽和モデルは、すべてのセルに 0.5 を加えます。

例: 2 つの洗濯用洗剤のうちの 1 つに対するユーザーの好みについての調査で、研究者は各グループ内の人数を数え、軟硬度 (軟、中、硬)、いずれかのブランドの使用経験の有無、洗濯温度 (低温または高温) などのさまざまなカテゴリーを組み合わせました。これにより、温度が軟硬度だけでなく、ブランドの好みとどのような関係があるかがわかりました。

統計: 度数、残差、パラメーター推定値、標準誤差、信頼区間、および偏連関の検定。カスタム・モデルの場合は、残差プロットと正規確率プロット。

モデル選択対数線型分析のデータの考慮事項

データ: 因子変数はカテゴリー型です。分析対象となる変数はすべて数値型である必要があります。カテゴリー型の文字列変数は、モデル選択分析を開始する前に、数値変数に再度割り当てることができます。

レベル数の多い変数を多数指定することは避けてください。そのような指定を行うと、多数のセルで観測数が少数になり、カイ 2 乗値が役に立たなくなることがあります。

関連プロシージャ: 「モデル選択」プロシージャは、モデルに必要な項を識別する際に役立ちます。その後、一般対数線型分析またはロジット対数線型分析を使用して、モデルの評価を続行できます。自動再割り当てを使用すれば、文字列変数を再度割り当てることができます。数値変数のカテゴリーが空白の場合は、再割り当てを使用して、連続した整数値を作成します。

モデル選択対数線型分析の取得

メニューから次の項目を選択します。

「分析」 > 「対数線型」 > 「モデルの選択...」

1. 2 つ以上の数値カテゴリー因子を選択します。
2. 「因子」リストで 1 つ以上の因子変数を選択して、「**範囲の定義**」をクリックします。
3. 因子変数ごとに値の範囲を定義します。
4. 「モデルの作成」グループでオプションを選択します。

オプションとして、セルの重み付け変数を選択して、構造 0 を指定できます。

対数線型分析の範囲の定義

各因子変数のカテゴリー範囲を示す必要があります。最小値と最大値は、因子変数の最低カテゴリーと最高カテゴリーに対応します。両方の値は整数でなければならず、最小値は最大値より小さくなければなりません。範囲外の値を持つケースは除外されます。例えば、最小値に 1 を指定し、最大値に 3 を指定した場合は、1、2、および 3 の値のみが使用されます。この処理を因子変数ごとに繰り返します。

対数線型分析のモデル

「モデルの指定」。飽和モデルには因子の主効果と因子間の交互作用のすべてが含まれます。不飽和モデルの生成クラスを指定するには、「カスタム」を選択します。

「生成クラス」。生成クラスは因子が表示される最高次の項のリストです。階層モデルには、生成クラスを定義する項とそれより低次の関係のすべてが含まれます。「因子」リストで変数 A 、 B 、および C を選択し、「項の構築」ドロップダウン・リストから「交互作用」を選択します。生成されるモデルには、指定された 3 次交互作用 $A*B*C$ 、2 次交互作用 $A*B$ 、 $A*C$ 、 $B*C$ 、および A 、 B 、 C の主効果が含まれます。生成クラスには低次の関係を指定しないでください。

項の構築

選択した因子や共変量について、次の項を作成できます。

「交互作用」。選択したすべての変数の最上位交互作用項を作成します。これはデフォルトです。

「主効果」。選択した変数ごとに主効果の項目を作成します。

「2 次まで」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「3 次まで」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「4 次まで」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「5 次まで」。選択した変数で考えられるすべての 5 次交互作用を作成します。

モデル選択対数線型分析のオプション

「表示」。「度数」または「残差」、あるいはその両方を選択できます。飽和モデルでは、観測度数と期待度数は同じであり、残差は 0 になります。

「作図」。カスタム・モデルの場合は、「残差」または「正規確率」、あるいはその両方のタイプのプロットを選択できます。これは、モデルとデータとの適合度の判別に役立ちます。

「飽和モデルの表示」。飽和モデルの場合は、「パラメータ推定値」を選択できます。パラメータ推定値は、モデルから除外できる項の判別に役立つ場合があります。偏連関の検定をリストする関連表も使用できます。多数の因子を含む表では、このオプションを選択すると計算により多くの時間がかかります。

「モデルの基準」。反復比例当てはめアルゴリズムを使用して、パラメータ推定値を取得します。「最大反復回数」、「収束基準」、または「デルタ」(飽和モデルのセル度数のすべてに加える値) を指定すれば、1 つ以上の推定基準をオーバーライドできます。

HILOGLINEAR コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 行列形式でセルの重みを指定する (CWEIGHT サブコマンドを使用)。
- 単一コマンドでの数モデルの分析の生成 (DESIGN サブコマンドを使用)。

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

第 10 章 一般対数線型分析

「一般対数線型分析」プロシージャーでは、クロス集計表または分割表の各交差分類カテゴリーに分類される観測値の度数を分析します。表の各交差分類はセルを構成し、各カテゴリー変数は因子と呼ばれます。従属変数は、クロス集計表のセルに含まれるケースの数 (度数) であり、説明変数は因子と共変量です。このプロシージャーでは、Newton-Raphson 法を使用して、階層および非階層の対数線型モデルの最尤パラメーターを推定します。ポワソン分布または多項分布を分析できます。

表のセルを定義する場合は、最大 10 個の因子を選択できます。セル構造変数を使用することで、不完全な表に構造 0 を定義したり、モデルにオフセット項を含めたり、対数比モデルを当てはめたり、あるいは周辺表の調整方法を実装できます。対比変数を使用すれば、一般化対数オッズ比 (GLOR) を計算できます。

モデル情報と適合度統計量は自動的に表示されます。また、さまざまな統計や作図を表示したり、アクティブなデータ・セットに残差や予測値を保存したりすることもできます。

例: フロリダの自動車事故報告からのデータを使用して、シートベルトの着用と負傷の程度 (致命的か致命的でないか) との関係を判別します。オッズ比により、関係の有効な証拠が示されます。

統計: 観測度数と期待度数; 未調整、調整済み、および最大対数尤度比残差; 計画行列; パラメーター推定値; オッズ比; 対数オッズ比; GLOR; Wald 統計量; および信頼区間。作図: 調整済み残差、最大対数尤度比残差、および正規確率。

一般対数線型分析のデータの考慮事項

データ: 因子はカテゴリー型で、セル共変量は連続型です。モデル内に共変量がある場合、セル内のケースの共変量の平均値がそのセルに適用されます。対比変数は連続型です。これらは一般化対数オッズ比を計算する際に使用されます。対比変数の値は、期待セル度数の対数の線型結合に対する係数です。

セル構造変数は重みを割り当てます。例えば、一部のセルが構造 0 である場合、セル構造変数の値は 0 か 1 になります。集計データに重みを付ける際に、セル構造変数を使用しないでください。代わりに、「データ」メニューから「ケースの重み付け」を選択してください。

仮定: 「一般対数線型分析」では、ポワソンと多項という 2 つの分布を使用できます。

ポワソン分布では、次のように仮定します。

- 研究前に合計サンプル・サイズが固定されません。すなわち分析は合計サンプル・サイズを条件としません。
- セルに存在する観測イベントは、他のセルのセル度数から統計的に独立しています。

多項分布では、次のように仮定します。

- 合計サンプル・サイズが固定されます。すなわち分析は合計サンプル・サイズを条件とします。
- セル度数は統計的に独立していません。

関連プロシージャー: クロス集計を調べる場合は「クロス集計表」プロシージャーを使用します。1 つ以上のカテゴリー変数を応答変数と見なし、その他の変数を説明変数と見なすのが自然な場合は、「ロジット対数線型」プロシージャーを使用します。

一般対数線型分析の取得

1. メニューから次の項目を選択します。

「分析」 > 「対数線型」 > 「一般...」

2. 「一般対数線型分析」ダイアログ・ボックスで、最大 10 個の因子変数を選択します。

オプションとして、以下のようにすることができます。

- セル共変量を選択する。
- セル構造変数を選択して、構造 0 を定義するか、オフセット項を含める。
- 対比変数を選択する。

一般対数線型分析のモデル

「モデルの指定」。飽和モデルには、因子変数に関連する主効果と交互作用がすべて含まれます。共変量の項は含まれていません。交互作用のサブセットのみを指定する場合や、因子と共変量の交互作用を指定する場合は、「カスタム」を選択します。

「因子と共変量」。因子と共変量がリストされます。

「モデルに含まれる項」。モデルは、使用するデータの性質によって異なります。「カスタム」の選択後、分析対象の主効果と交互作用を選択できます。モデルに含める項はすべて指示する必要があります。

項の構築

選択した因子や共変量について、次の項を作成できます。

「交互作用」。選択したすべての変数の最上位交互作用項を作成します。これはデフォルトです。

「主効果」。選択した変数ごとに主効果の項目を作成します。

「2 次まで」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「3 次まで」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「4 次まで」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「5 次まで」。選択した変数で考えられるすべての 5 次交互作用を作成します。

一般対数線型分析のオプション

「一般対数線型分析」プロシージャでは、モデル情報と適合度統計量が表示されます。さらに、次の中から 1 つ以上を選択できます。

「表示」。観測セル度数と期待セル度数 (未調整、調整済み、および最大対数尤度比残差)、モデルの計画行列、モデルのパラメーター推定値など、いくつかの統計量を表示できます。

「作図」。カスタム・モデルにのみ使用可能な作図には、2 つの散布図行列 (観測セル度数および期待セル度数に対する調整済み残差または最大対数尤度比残差) が含まれます。調整済み残差または最大対数尤度比残差の正規確率および傾向化除去正規プロットを表示することもできます。

「信頼区間」。パラメーター推定値の信頼区間を調整することができます。

「基準」。Newton-Raphson 法を使用して、最尤パラメーター推定値を取得します。最大反復回数、収束基準、デルタ (初期近似値のすべてのセルに追加される定数) の新しい値を入力できます。デルタは飽和モデルのセルに残ります。

一般対数線型分析の保存

アクティブなデータ・セットに新規変数として保存する値を選択します。新しい変数名の接尾辞 n を 1 ずつ増やして、保存された変数ごとに固有名を作成します。

集計データ (分割表のセル) がデータ・エディターで個々の観測値に記録されている場合でも、保存値はそのデータを参照します。非集計データの残差や予測値を保存すると、データ・エディターで分割表のセルの保存値がそのセルのケースごとに入力されます。保存値を理解するには、データを集計してセル度数を取得する必要があります。

保存できる残差のタイプは、未調整、標準化、調整済み、最大対数尤度比の 4 つです。予測値を保存することもできます。

- 残差。単純残差や未調整残差とも呼びます。観測したセル度数と期待度数の差です。
- 標準化残差。残差を標準誤差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。
- 調整済み残差。標準化残差を、推定された標準誤差で割った値。選択したモデルが正しいときには調整済み残差が漸近的に標準正規分布になるので、正規性をチェックする目的では標準化残差よりも効果的です。
- 最大対数尤度比残差。尤度比カイ 2 乗統計量 (G^2) に対する個々の寄与の符号付き平方根。符号としては、残差 (観測度数から期待度数を引いた値) の符号を採用します。最大対数尤度比残差は漸近標準正規分布になります。

GENLOG 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。

- 観測対象のセル度数と期待されるセル度数の線型結合の計算、および残差、標準化残差、その組み合わせの調整済み残差の出力 (GERESID サブコマンドを使用)。
- 冗長検査のためのデフォルトしきい値の変更 (CRITERIA サブコマンドを使用)。
- 標準化残差の表示 (PRINT サブコマンドを使用)。

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

第 11 章 ロジット対数線型分析

「ロジット対数線型分析」プロシージャーでは、従属 (または応答) 変数と独立 (または説明) 変数との間の関係を分析します。従属変数は常にカテゴリ型ですが、独立変数はカテゴリ型 (因子) にすることができます。その他の独立変数 (セル共変量) は連続型にすることができますが、これらはケースごとには適用されません。セルに対する共変量の加重平均はそのセルに適用されます。従属変数のオッズ対数は、パラメーターの線型結合として表されます。多項分布は自動的に仮定されるため、これらのモデルは多項ロジット・モデルと呼ばれることもあります。このプロシージャーでは、Newton-Raphson アルゴリズムを使用してロジット対数線型モデルのパラメーターを推定します。

従属変数と因子変数は 1 から 10 までの組み合わせから選択できます。セル構造変数を使用することで、不完全な表に構造 0 を定義したり、モデルにオフセット項を含めたり、対数比モデルを当てはめたり、あるいは周辺表の調整方法を実装できます。対比変数を使用すれば、一般化対数オッズ比 (GLOR) を計算できます。対比変数の値は、期待セル度数の対数の線型結合に対する係数です。

モデル情報と適合度統計量は自動的に表示されます。また、さまざまな統計や作図を表示したり、アクティブなデータ・セットに残差や予測値を保存したりすることもできます。

例: フロリダで 219 匹のワニを対象とした研究が行われたとします。そのサイズや生息地である 4 つの湖によって、ワニの食料の種類がどう変化するでしょうか。この研究から、小型のワニは魚類よりは虫類を好むというオッズが大型のワニの場合より 0.7 倍低いことと、魚類ではなく主には虫類を選択するというオッズが湖 3 で最高であることがわかりました。

統計: 観測度数と期待度数; 未調整、調整済み、および最大対数尤度比残差; 計画行列; パラメーター推定値; 一般化対数オッズ比; Wald 統計量; および信頼区間。作図: 調整済み残差、最大対数尤度比残差、および正規確率プロット。

ロジット対数線型分析のデータの考慮事項

データ: 従属変数はカテゴリ型です。因子はカテゴリ型です。セルの共変量を連続型にすることはできませんが、共変量がモデルに含まれている場合は、セル内のケースの共変量の平均値がそのセルに適用されます。対比変数は連続型です。これらは一般化対数オッズ比 (GLOR) の計算に使用されます。対比変数の値は、期待セル度数の対数の線型結合に対する係数です。

セル構造変数は重みを割り当てます。例えば、一部のセルが構造 0 である場合、セル構造変数の値は 0 か 1 になります。集計データに重みを付ける際に、セル構造変数を使用しないでください。代わりに、「データ」メニューの「ケースの重み付け」を使用してください。

仮定: 説明変数のカテゴリの各組み合わせ内の度数は、多項分布すると仮定します。多項分布では、次のように仮定します。

- 合計サンプル・サイズが固定されます。すなわち分析は合計サンプル・サイズを条件とします。
- セル度数は統計的に独立していません。

関連プロシージャー: 分割表を表示する場合は「クロス集計表」プロシージャーを使用します。観測度数と一連の説明変数との間の関係を分析する場合は、「一般対数線型分析」プロシージャーを使用します。

ロジット対数線型分析の取得

1. メニューから次の項目を選択します。

「分析」 > 「対数線型」 > 「ロジット...」

2. 「ロジット対数線型分析」ダイアログ・ボックスで、1 つ以上の従属変数を選択します。

3. 1 つ以上の因子変数を選択します。

従属変数と因子変数の総数は 10 以下にする必要があります。

オプションとして、以下のようにすることができます。

- セル共変数を選択する。
- セル構造変数を選択して、構造 0 を定義するか、オフセット項を含める。
- 1 つ以上の対比変数を選択する。

ロジット対数線型分析のモデル

「モデルの指定」。飽和モデルには、因子変数に関連する主効果と交互作用がすべて含まれます。共変量の項は含まれていません。交互作用のサブセットのみを指定する場合や、因子と共変量の交互作用を指定する場合は、「カスタム」を選択します。

「因子と共変量」。因子と共変量がリストされます。

「モデルに含まれる項」。モデルは、使用するデータの性質によって異なります。「カスタム」の選択後、分析対象の主効果と交互作用を選択できます。モデルに含める項はすべて指示する必要があります。

項は、従属項に可能なすべての組み合わせを使用し、モデル・リストの各項にそれぞれの組み合わせを一致させて計画に追加します。「従属変数に対して定数項を含む」を選択した場合は、モデル・リストに単位項 (1) も追加されます。

例えば、変数 $D1$ と $D2$ が従属変数であるとして、従属項リストは「ロジット対数線型分析」プロシージャで作成されます ($D1$ 、 $D2$ 、 $D1*D2$)。「モデルに含まれる項」リストに $M1$ と $M2$ が含まれ、さらに定数が含まれている場合は、モデル・リストに 1、 $M1$ 、および $M2$ が含まれます。結果として得られる計画には、次のように、各モデル項と各従属項の組み合わせが含まれます。

$D1$ 、 $D2$ 、 $D1*D2$

$M1*D1$ 、 $M1*D2$ 、 $M1*D1*D2$

$M2*D1$ 、 $M2*D2$ 、 $M2*D1*D2$

「従属変数に対して定数項を含む」。カスタム・モデルに従属変数の定数を含めます。

項の構築

選択した因子や共変量について、次の項を作成できます。

「交互作用」。選択したすべての変数の最上位交互作用項を作成します。これはデフォルトです。

「主効果」。選択した変数ごとに主効果の項目を作成します。

「2 次まで」。選択した変数で考えられるすべての 2 次交互作用を作成します。

「3 次まで」。選択した変数で考えられるすべての 3 次交互作用を作成します。

「4 次まで」。選択した変数で考えられるすべての 4 次交互作用を作成します。

「5 次まで」。選択した変数で考えられるすべての 5 次交互作用を作成します。

ロジット対数線型分析のオプション

「ロジット対数線型分析」プロシージャでは、モデル情報と適合度統計量が表示されます。さらに、次のオプションから 1 つ以上を選択できます。

「表示」。観測セル度数と期待セル度数、生の残差、調整済み残差、最大対数尤度比残差、モデルの計画行列、モデルのパラメーター推定値など、統計量をいくつか表示できます。

「作図」。カスタム・モデルに使用可能な作図には、2 つの散布図行列（観測セル度数および期待セル度数に対する調整済み残差または最大対数尤度比残差）が含まれます。調整済み残差または最大対数尤度比残差の正規確率および傾向化除去正規プロットを表示することもできます。

「信頼区間」。パラメーター推定値の信頼区間を調整することができます。

「基準」。Newton-Raphson 法を使用して、最尤パラメーター推定値を取得します。最大反復回数、収束基準、デルタ（初期近似値のすべてのセルに追加される定数）の新しい値を入力できます。デルタは飽和モデルのセルに残ります。

ロジット対数線型分析の保存

アクティブなデータ・セットに新規変数として保存する値を選択します。新しい変数名の接尾辞 n を 1 つ増やして、保存された変数ごとに固有名を作成します。

集計データ（分割表のセル）がデータ・エディターで個々の観測値ごとに記録されている場合でも、保存値はそのデータを参照します。非集計データの残差や予測値を保存すると、データ・エディターで分割表のセルの保存値がそのセルのケースごとに入力されます。保存値を理解するには、データを集計してセル度数を取得する必要があります。

保存できる残差のタイプは、未調整、標準化、調整済み、最大対数尤度比の 4 つです。予測値を保存することもできます。

- 残差。単純残差や未調整残差とも呼びます。観測したセル度数と期待度数の差です。
- 標準化残差。残差を標準誤差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。
- 調整済み残差。標準化残差を、推定された標準誤差で割った値。選択したモデルが正しいときには調整済み残差が漸近的に標準正規分布になるので、正規性をチェックする目的では標準化残差よりも効果的です。
- 最大対数尤度比残差。尤度比カイ 2 乗統計量 (G^2) に対する個々の寄与の符号付き平方根。符号としては、残差（観測度数から期待度数を引いた値）の符号を採用します。最大対数尤度比残差は漸近標準正規分布になります。

GENLOG コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 観測セル度数と期待セル度数の線型結合の計算、およびその結合の残差、標準化残差、調整済み残差の印刷 (GERESID サブコマンドを使用)。
- 冗長検査のためのデフォルトしきい値の変更 (CRITERIA サブコマンドを使用)。

- 標準化残差の表示 (PRINT サブコマンドを使用)。

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

第 12 章 生命表

雇用期間（就職から退職までの期間）など、2 つの事象の間の時間分布を調べる必要があるさまざまな状況があります。ただし、通常、この種のデータには、2 つ目の事象がそれ以上記録されないケース（調査の終了時点でまだ対象者が会社に在籍している場合など）がいくつか含まれます。これはいくつかの理由により発生する可能性があります。場合によっては、調査が終了する前に事象がまったく発生しなかったり、調査終了前のある時点で状況を追跡できなくなったり、また、調査とは関係のない理由（従業員が病気で休職しているなど）で続行できなくなることがあります。このようなケースをまとめて**調査済みケース**と呼びます。この種の調査に、 t 検定や線型回帰など従来の手法を使用することは適切ではありません。

この種のデータに有効な統計手法をフォローアップ**生命表**といいます。この生命表の基本的な考えは、観測期間をより短い時間間隔に分割するということです。各間隔において、少なくともその間隔の間観測された人すべてを対象に、その時間間隔で発生する最終事象の確率を計算します。次に、各間隔から推定された確率を使用して、さまざまな時点で発生する事象の確率全体を推定します。

例: 新しいニコチン・パッチ療法が従来のパッチ療法よりも禁煙に役立つかを調べるとします。この場合、喫煙者を 2 つのグループ（従来の療法を受けたグループと試験的な療法を受けたグループ）に分けて研究を行うことができます。データから生命表を構築することで、この 2 つのグループの全体的な禁煙率を比較し、試験的な療法が従来の療法より改善されているかどうかを判別できます。また、生存関数またはハザード関数を作図して視覚的に比較し、さらに詳細な情報を得ることもできます。

統計: グループごとの各時間間隔に対する、参加数、離脱数、リスクにさらされた数、終結事象の数、終結する比率、生存率、累積生存率（および標準誤差）、確率密度（および標準誤差）、ハザード比（および標準誤差）、グループごとの生存時間の中央値、およびグループ間の生存分布を比較するための Wilcoxon (Gehan) 検定。作図: 累積生存関数、対数累積生存関数、密度関数、ハザード比関数、および 1 マイナス累積生存関数に対する作図。

生命表のデータの考慮事項

データ: 時間変数は量的である必要があります。状態変数は、整数としてコード化された二分変数またはカテゴリ変数である必要があります。この場合、事象は単一値または連続値の範囲としてコード化されません。因子変数は、整数としてコード化されたカテゴリ変数である必要があります。

仮定: 対象事象の確率は、初期事象以降の時間にのみ依存する必要がある、絶対時間に対して一定であると仮定されます。つまり、異なる時間で調査に入るケース（異なる時間に治療を始める患者など）は同様でなければなりません。また、調査済みケースと未調査ケースとの間にシステム上の違いがあってはなりません。例えば、調査済みケースの多くがより深刻な状態の患者である場合、結果が偏ることがあります。

関連プロシージャー: 「生命表」プロシージャーでは、この種の分析（一般的には生存分析と呼ばれます）に対して保険数理法を使用します。「Kaplan-Meier 生存分析」プロシージャーでは、観測期間のより短い時間間隔への分割に依存しない、少し異なる生命表の計算方法を使用します。この方法が推奨されるのは、観測数が少なく、各生存時間間隔における観測数がわずかしかないような場合です。生存時間との関連に疑いのある変数、または制御対象の変数（共変量）がある場合は、「Cox 回帰」プロシージャーを使用します。同じケースに対して各時点で共変量の値が異なる可能性がある場合は、「時間依存の Cox 回帰」を使用します。

生命表の作成

1. メニューから次の項目を選択します。

「分析」 > 「生存」 > 「生命表...」

2. 数値 生存変数を 1 つ選択します。
3. 調べる時間間隔を指定します。
4. 状態変数を選択して、最終事象が発生したケースを定義します。
5. 「イベント定義」をクリックし、イベントが発生したことを示す状態変数の値を指定します。

オプションとして、1 次因子変数を選択することができます。生存変数の保険数理表は、因子変数のカテゴリごとに生成されます。

2 次の因子ごとの 変数を選択することもできます。生存変数の保険数理表は、1 次および 2 次の因子変数のあらゆる組み合わせに対して生成されます。

生命表の状態変数の事象の定義

状態変数に対して選択された値の出現は、それらのケースに対して終結事象が発生したことを示します。その他のすべてのケースは調査済みと見なされます。対象事象を識別する単一値または値の範囲を入力します。

生命表の範囲の定義

指定範囲内の因子変数に対する値があるケースは分析に含まれ、各表 (要求されている場合は作図も) は範囲内の固有値ごとに生成されます。

生命表のオプション

生命表分析のさまざまな側面を制御することができます。

「生命表」。出力での生命表の表示を抑制するには、「生命表」の選択を解除します。

「作図」。生存関数の作図を要求できます。因子変数を定義している場合は、その因子変数によって定義されたサブグループごとに作図が生成されます。使用できる作図は、累積生存関数、対数累積生存関数、ハザード関数、密度関数、および 1 マイナス累積生存関数です。

- 「累積生存関数」。累積生存関数を線型スケールで表示します。
- 「対数累積生存関数」。累積生存関数を対数スケールで表示します。
- 「ハザード」。累積ハザード関数を線型スケールで表示します。
- 「密度関数」。密度関数を表示します。
- 「累積死亡関数」。線型スケールで 1 マイナス累積生存関数を作図します。

「1 次因子のレベルの比較」。1 次の制御変数がある場合、このグループ内の代替項目のいずれかを選択して、Wilcoxon (Gehan) 検定を実行し、サブグループの生存を比較できます。検定は 1 次因子に対して実行されます。2 次因子を定義している場合、検定は 2 次変数のレベルごとに実行されます。

SURVIVAL コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 複数の従属変数を指定する。
- 不等なスペースの間隔を指定する。
- 複数の状態変数を指定する。
- 必ずしもすべての因子および制御変数を含まない比較を指定する。
- 厳密ではない概算比較を計算する。

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

第 13 章 Kaplan-Meier 生存分析

雇用期間 (就職から退職までの期間) など、2 つの事象の間の時間分布を調べる必要があるさまざまな状況があります。ただし、この種のデータには通常、調査済みケースがいくつか含まれています。調査済みケースとは、2 つ目の事象がそれ以上記録されないケースのことです (例えば、調査の終了時点でまだ対象者が会社に在籍している場合など)。Kaplan-Meier プロシージャは、調査済みケースが存在する場合に、事象までの時間のモデルを推定するための手法です。Kaplan-Meier モデルは、事象の各発生時点の条件付き確率の推定に基づくものであり、これらの確率の積極限を使用して、各時点における生存率を推定します。

例: AIDS の新しい治療法には延命に対する効果があるでしょうか。この場合、AIDS 患者を 2 つのグループ (従来の治療を受けるグループと試験的な治療を受けるグループ) に分けて研究を行うことができます。データから Kaplan-Meier モデルを構築することで、この 2 つのグループの全体的な生存率を比較し、試験的な治療が従来の治療より改善されているかどうかを判別できます。また、生存関数またはハザード関数を作図して視覚的に比較し、さらに詳細な情報を得ることもできます。

統計: 時間、状況、累積生存と標準誤差を含む生命表、累積イベント数、残存数、および標準誤差と 95% の信頼区間を含む生存時間の平均値と中央値。作図: 累積生存関数、ハザード関数、対数累積生存関数、および 1 マイナス累積生存関数。

Kaplan-Meier のデータの考慮事項

データ: 時間変数は連続型である必要がありますが、状態変数はカテゴリー型または連続型にすることができます。因子およびストラータ変数はカテゴリー型である必要があります。

仮定: 対象事象の確率は、初期事象以降の時間にのみ依存する必要がありますが、絶対時間に対して一定であると仮定されます。つまり、異なる時間で調査に入るケース (異なる時間に治療を始める患者など) は同様でなければなりません。また、調査済みケースと未調査ケースとの間にシステム上の違いがあってはなりません。例えば、調査済みケースの多くがより深刻な状態の患者である場合、結果が偏ることがあります。

関連プロシージャ: Kaplan-Meier プロシージャでは、各事象の発生時間における生存関数またはハザード関数を推定する生命表計算の手法を使用します。「生命表」プロシージャでは、観測期間のより短い時間間隔への分割に依存する生存分析に対して、保険数理法を使用します。これは、多数のサンプルを扱う際に役立つ場合があります。生存時間との関連に疑いのある変数、または制御対象の変数 (共変量) がある場合は、「Cox 回帰」プロシージャを使用します。同じケースに対して各時点で共変量の値が異なる可能性がある場合は、「時間依存の Cox 回帰」を使用します。

Kaplan-Meier 生存分析の取得

1. メニューから次の項目を選択します。

「分析」 > 「生存」 > 「Kaplan-Meier...」

2. 時間変数を選択します。

3. 状態変数を選択して、最終事象が発生したケースを識別します。この変数は数値と短い文字列 のどちらでもかまいません。次に、「イベント定義」をクリックします。

オプションとして、因子変数を選択して、グループ間の違いを調べることができます。また、ストラータ変数を選択して、変数のレベル (階層) ごとに別の分析を生成することもできます。

Kaplan-Meier の状態変数のイベントの定義

最終事象が発生したことを示す値を 1 つ以上入力します。単一値、値の範囲、または値のリストを入力することができます。「値の範囲」オプションを使用できるのは、状態変数が数値である場合のみです。

Kaplan-Meier の因子レベルの比較

さまざまなレベルの因子について、生存分布の等質性を検定する統計を要求できます。使用可能な統計は、「対数順位」、「Breslow」、および「Tarone-Ware」です。これらのいずれかを選択して、「階層全体でプール」、「階層ごと」、「階層全体でペアごと」、「階層ごとにペアごと」から実行する比較を指定します。

- 「対数順位」。生存分布の等質性を比較するための検定。この検定では、すべての時点に等しく重みを割り当てます。
- 「Breslow」。生存分布の等質性を比較するための検定。各時点での危険なケースの個数によって時点に重みを付けます。
- 「Tarone-Ware」。生存分布の等質性を比較するための検定。各時点での危険なケースの個数の平方根によって時点に重みを付けます。
- 「階層全体でプール」。単一の検定ですべての因子レベルを比較し、生存曲線の等質性を検定します。
- 「階層全体でペアごと」。因子レベルの各個別ペアを比較します。ペアごとのトレンド検定は使用できません。
- 「階層ごと」。階層ごとにすべての因子レベルの等質性を個別に検定します。ストラータ変数がない場合は検定を実行しません。
- 「階層ごとにペアごと」。階層ごとに因子レベルの各個別ペアを比較します。ペアごとのトレンド検定は使用できません。ストラータ変数がない場合は検定を実行しません。

因子レベルの線型傾向因子のレベル全体で線型トレンドに対する検定を行うことができます。このオプションは、因子レベルを (ペアごとではなく) 全体的に比較する場合にのみ使用できます。

Kaplan-Meier の新規変数の保存

Kaplan-Meier 表の情報を新規変数として保存できます。これらの変数は、その後の分析に使用して、仮説の検定や仮定の確認を行うことができます。生存、生存の標準誤差、ハザード、および累積イベントを新規変数として保存できます。

- 「累積生存関数」。累積生存確率の推定値。デフォルトの変数名は、接頭辞 `sur_` の後に連続番号を付加したものです。例えば、`sur_1` が既に存在する場合、Kaplan-Meier では `sur_2` という変数名を割り当てます。
- 「生存の標準誤差」。累積生存確率推定値の標準誤差。デフォルトの変数名は、接頭辞 `se_` の後に連続番号を付加したものです。例えば、`se_1` が既に存在する場合、Kaplan-Meier では `se_2` という変数名を割り当てます。
- 「ハザード」。累積ハザード関数の推定値。デフォルトの変数名は、接頭辞 `haz_` の後に連続番号を付加したものです。例えば、`haz_1` が既に存在する場合、Kaplan-Meier では `haz_2` という変数名を割り当てます。
- 「累積イベント」。ケースを生存時間およびステータス・コードでソートしたときのイベントの累積度数。デフォルトの変数名は、接頭辞 `cum_` の後に連続番号を付加したものです。例えば、`cum_1` が既に存在する場合、Kaplan-Meier では `cum_2` という変数名を割り当てます。

Kaplan-Meier のオプション

Kaplan-Meier 分析ではさまざまな出力タイプを要求できます。

統計: 生命表、生存の平均値と中央値、四分位など、計算した生存関数について表示される統計を選択できます。因子変数を含めた場合は、グループごとに別の統計が生成されます。

作図: 作図することにより、生存関数、死亡関数、ハザード関数、および対数生存関数を視覚的に調べることができます。因子変数を含めた場合は、グループごとに関数が作図されます。

- 「累積生存関数」。累積生存関数を線型スケールで表示します。
- 「累積死亡関数」。線型スケールで 1 マイナス累積生存関数を作図します。
- 「ハザード」。累積ハザード関数を線型スケールで表示します。
- 「対数累積生存関数」。累積生存関数を対数スケールで表示します。

KM コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 追跡不能なケースを調査済みケースとは別のカテゴリーとして扱う度数分布表を取得する。
- 線型傾向の検定に対して不等な間隔を指定する。
- 生存時間変数の四分位以外のパーセンタイルを取得する。

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

第 14 章 Cox 回帰分析

Cox 回帰では、時間事象データに対して予測モデルを作成します。このモデルは、予測値変数に与えられた特定の値に対して、対象のイベントが所定の時間 t に発生する確率を予測する生存関数を生成します。生存関数の形状および予測値の回帰係数は、観測された被験者から推定されます。このモデルは予測値変数の測定値がある新しいケースに適用できます。観測時間中に対象のイベントを経験していない調査済み被験者からの情報が、モデルの推定に対して有効に寄与することに注意してください。

例: 喫煙による肺ガン発生の危険性は、男性と女性とで異なるでしょうか。Cox 回帰モデルを構築する際に、喫煙量 (1 日あたりに吸うタバコの本数) と性別を共変量として入力すれば、肺ガンが発病するまでの期間における、性別および喫煙量が与える影響についての仮説を検定できます。

統計: 各モデル: $-2LL$ 、尤度比統計量、および全体のカイ 2 乗。モデルの変数: パラメーター推定値、標準誤差、Wald 統計量。モデルにない変数: スコア統計および残差カイ 2 乗。

Cox 回帰のデータの考慮事項

データ: 時間変数は量的である必要がありますが、状態変数はカテゴリ型や連続型にすることができます。独立変数 (共変量) は連続型またはカテゴリ型にすることができます。カテゴリ型の場合、ダミー・コーディングまたはインジケータ・コーディングする必要があります (プロシージャーには自動的にカテゴリ変数を再コーディングするオプションがあります)。ストラータ変数は、整数または短い文字列としてコーディングされる、カテゴリ型である必要があります。

仮定: 観測は独立しており、ハザード比は経時的に一定である必要があります。つまり、ケース間のハザードの比例関係は時間の経過と共に変化してはならないということです。後者の仮定は**比例ハザードの仮定**と呼ばれます。

関連プロシージャー: 比例ハザードの仮定が適用されない場合 (上記参照)、「時間依存の Cox 回帰」プロシージャーを使用する必要がある場合があります。共変量がない場合、またはカテゴリ共変量が 1 つしかない場合は、「生命表」または「Kaplan-Meier」プロシージャーを使用して、サンプルに対する生存関数またはハザード関数を調べることができます。サンプル内に調査済みデータがない場合 (つまり、すべてのケースで終結事象が発生した場合)、「線型回帰」プロシージャーを使用して、予測と事象までの時間との間の関係をモデリングできます。

Cox 回帰分析の取得

1. メニューから次の項目を選択します。

「分析」 > 「生存」 > 「Cox 回帰...」

2. 時間変数を選択します。時刻値が負になっているケースは分析されません。

3. 状態変数を選択し、「イベント定義」をクリックします。

4. 共変量を 1 つ以上選択します。交互作用の項を含めるには、交互作用に含まれる変数をすべて選択し、「>a*b>」をクリックします。

オプションで、ストラータ変数を定義することで、異なるグループに対して別のモデルを計算できます。

Cox 回帰におけるカテゴリー変数の定義

「Cox 回帰」プロシージャでカテゴリー変数を処理する方法の詳細を指定することができます。

「共変量」。メイン・ダイアログ・ボックスで、単独で、あるいは交互作用の一部として指定された、層内のすべての共変量をリストします。これらの一部が文字列変数またはカテゴリー型である場合は、カテゴリー共変量としてのみ使用可能です。

「カテゴリー共変量」。カテゴリーとして識別される変数をリストします。各変数には、使用される対比コードを示す、括弧で囲まれた表記が含まれます。文字列変数 (名前に続く記号 < で示される) は「カテゴリー共変量」リストに既に存在します。「共変量」リストからその他のカテゴリー共変量を選択して、「カテゴリー共変量」リストに移動します。

「対比の変更」。対比方法を変更できます。次の対比方法を使用できます。

- 「指標」。対比はカテゴリー・メンバーシップの有無を示します。参照カテゴリーは、ゼロ行として対比行列に示されます。
- 「単純」。予測変数の各カテゴリー (参照カテゴリーを除く) は参照カテゴリーと比較されます。
- 「差分」。予測変数の各カテゴリー (最初のカテゴリーを除く) は前のカテゴリーの平均効果と比較されます。逆 Helmert 対比とも呼ばれます。
- 「Helmert」。予測変数の各カテゴリー (最後のカテゴリーを除く) は後続のカテゴリーの平均効果と比較されます。
- 「反復」。予測変数の各カテゴリー (最初のカテゴリーを除く) は前にあるカテゴリーと比較されません。
- 「多項式」。直交多項式の対比。カテゴリーは等間隔であると仮定されます。多項式対比は数値変数にのみ使用可能です。
- 「偏差」。予測変数の各カテゴリー (参照カテゴリーを除く) は全体的な効果と比較されます。

「偏差」、「単純」、または「指標」を選択する場合は、参照カテゴリーとして「最初」または「最後」を選択します。方法は、「変更」をクリックするまでは実際に変更されないことに注意してください。

文字列共変量はカテゴリー共変量でなければなりません。「カテゴリー共変量」リストから文字列変数を削除するには、メイン・ダイアログ・ボックスの「共変量」リストから、その変数を含むすべての項を削除する必要があります。

Cox 回帰のプロット

プロットは、推定モデルを評価し、結果を解釈する場合に役立ちます。生存関数、ハザード関数、ログ・マイナス・ログ関数、および死亡関数を作図できます。

- 「累積生存関数」。累積生存関数を線型スケールで表示します。
- 「ハザード」。累積ハザード関数を線型スケールで表示します。
- 「ログ・マイナス・ログ」。ln(-ln) 変換が推定値に適用された後の累積生存推定値。
- 「累積死亡関数」。線型スケールで 1 マイナス累積生存関数を作図します。

これらの関数は共変量の値によって異なるため、時間に対する関数を作図するには、共変量に対して定数値を使用する必要があります。デフォルトでは各共変量の平均を定数値として使用しますが、「値の変更」制御グループを使用して、プロットに独自の値を入力できます。

「個別の線」テキスト・ボックスに共変量を移動すれば、カテゴリ共変量の値ごとに個別の線を作図できます。このオプションは、「作図する共変量の値の配置」リストで名前の後に「(カテゴリ)」と表示される、カテゴリ共変量にのみ使用可能です。

Cox 回帰における新規変数の保存

各種分析結果は新規変数として保存できます。これらの変数は、その後の分析に使用して、仮説の検定または仮定の確認を行うことができます。

「モデル変数を保存」。回帰に対する生存関数とその標準誤差、ログ・マイナス・ログ推定値、ハザード関数、偏残差、DfBeta、および線型予測 $X*Beta$ を新規変数として保存できます。

- 「生存関数」。任意の時刻での累積生存関数の値。その期間まで生存する確率に等しくなります。
- 「ログ・マイナス・ログ生存関数」。 $\ln(-\ln)$ 変換を推定値に適用した後の累積生存推定値。
- 「ハザード関数」。累積ハザード関数の推定値 (Cox-Snell 残差ともいう) を保存します。
- 「偏残差」。偏残差を生存時間に対してプロットすると、比例ハザードの仮定を検定することができます。最終モデルでは、各共変量に対して 1 つの変数が保存されます。偏残差は、少なくとも 1 つの共変量を持つモデルに限って使用することができます。
- 「DfBeta」。ケースを除外した場合に推定される係数の変化。最終モデルでは、各共変量に対して 1 つの変数が保存されます。DfBeta は、少なくとも 1 つの共変量を持つモデルに限って使用することができます。
- $X*Beta$ ($X * \text{ベータ}$)。線型予測得点。ケースごとの、平均値中心化共変量の値とそれらに対応するパラメーター推定値の積の合計。

時間依存の共変量を使用して Cox を実行する場合、保存できる変数は DfBeta および線型予測変数 $X*Beta$ だけです。

「モデル情報を XML ファイルにエクスポート」。パラメーター推定値は XML 形式で指定されたファイルにエクスポートされます。このモデル・ファイルを使用して、モデル情報を他のデータ・ファイルに適用して、スコアリングに使用できます。

Cox 回帰のオプション

分析および出力をさまざまな側面から制御することができます。

「モデル統計量」。 $\exp(B)$ の信頼区間および推定値の相関を含む、モデル・パラメーターの統計量を取得することができます。これらの統計量を要求できるのは、各ステップまたは最後のステップのみです。

「ステップワイズにおける確率」。ステップワイズ法を選択した場合は、モデルに対する投入または除去の確率を指定することができます。投入する F の有意レベルが「投入」値より小さい場合は変数が投入され、有意レベルが「除去」値より大きい場合は変数が除去されます。「投入」値は「除去」値より小さくしなければなりません。

「最大反復回数」。プロシーチャーで解を検索する時間を制御する、モデルの最大反復回数を指定できません。

「ベースライン関数の表示」。ベースラインのハザード関数と累積生存を共変量の平均値で表示できます。この表示は、時間依存の共変量を指定した場合は使用できません。

Cox 回帰における状態変数のイベントの定義

最終事象が発生したことを示す値を 1 つ以上入力します。単一値、値の範囲、または値のリストを入力することができます。「値の範囲」オプションを使用できるのは、状態変数が数値である場合のみです。

COXREG コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 追跡不能なケースを調査済みケースとは別のカテゴリとして扱う度数分布表を取得する。
- 偏差、単純、指標の対比方法で、最初と最後以外の参照カテゴリを選択する。
- 多項式の対比方法に対するカテゴリの不等な間隔を指定する。
- 追加の反復基準を指定する。
- 欠損値の処理を制御する。
- 保存された変数の名前を指定する。
- 外部 IBM SPSS Statistics データ・ファイルに出力を書き込む。
- 処理の間、外部のスクラッチ・ファイルに各分割ファイル・グループのデータを保持する。これにより、大型のデータ・セットで分析を実行する際、メモリー・リソースを節約できます。これは時間依存の共変量には利用できません。

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

第 15 章 時間依存の共変量の計算

Cox 回帰モデルを計算する際に、比例ハザードの仮定が適用されない特定の状況があります。つまり、ハザード比が時間の経過と共に変化し、共変量のいずれか (または複数) の値が時点によって異なる場合があります。このような場合は、拡張 Cox 回帰モデルを使用する必要があります。これにより、**時間依存共変量**を指定できるようになります。

このようなモデルを分析するには、まず、時間依存共変量を定義する必要があります。(コマンド・シンタックスを使用すれば、複数の時間依存共変量を指定できます。)この作業を容易にするために、時間を表すシステム変数を使用できます。この変数は $T_.$ と呼ばれます。この変数を使用すると、時間依存共変量を次の一般的な 2 とおりの方法で定義できます。

- 特定の共変量に関する比例ハザードの仮定を検定する場合、あるいは非比例ハザードを許可する拡張 Cox 回帰モデルを推定する場合は、時間変数 $T_.$ と対象の共変量の関数として時間依存の共変量を定義することで実行できます。時間変数と共変量の単純な積などが一般的な例と言えますが、より複雑な関数も指定できます。時間依存共変量の係数の有意性を検定することで、比例ハザードの仮定が妥当であるかがわかります。
- 変数の中には、時間によって値が異なるが、時間との系統的な関連性がないものもあります。このような場合は、**セグメント化した時間依存の共変量**を定義する必要があります。この作業は、**論理式**を使用して行うことができます。この論理式では、真の場合は 1、偽の場合は 0 という値を使用します。一連の論理式を使用することで、一連の測定から時間依存共変量を作成できます。例えば、4 週にわたる研究で血圧 ($BP1$ から $BP4$ として識別) を週に 1 度測定する場合は、時間依存共変量を $(T_ < 1) * BP1 + (T_ >= 1 \ \& \ T_ < 2) * BP2 + (T_ >= 2 \ \& \ T_ < 3) * BP3 + (T_ >= 3 \ \& \ T_ < 4) * BP4$ と定義できます。どのケースでも括弧内の 1 つの項目のみが 1 になり、残りはすべて 0 になることに注意してください。つまり、時間が 1 週間より短い場合は、 $BP1$ を使用し、1 週間より長く 2 週間より短い場合は、 $BP2$ を使用するなどとなります。

「時間依存の共変量の計算」ダイアログ・ボックスで、関数作成用のコントロールを使用して、時間依存共変量の式を作成することも、「 $T_COV_.$ の式」テキスト領域に直接入力することもできます。文字列定数は引用符またはアポストロフィで囲み、数値定数は米国形式 (小数点としてドットを使用) で入力する必要があります。結果として得られる変数は $T_COV_.$ と呼ばれ、Cox 回帰モデルに共変量として含める必要があります。

時間依存の共変量の計算

1. メニューから次の項目を選択します。

「分析」 > 「生存」 > 「時間依存の Cox 回帰...」

2. 時間依存共変量の式を入力します。
3. 「モデル」をクリックして、Cox 回帰分析を続行します。

注: 新しい変数の $T_COV_.$ は、必ず Cox 回帰モデルに共変量として含めるようにしてください。

詳しくは、89 ページの『第 14 章 Cox 回帰分析』のトピックを参照してください。

時間依存の共変量を使用する Cox 回帰の追加機能

コマンド・シンタックス言語を使用して、複数の時間依存の共変量を指定することもできます。その他のコマンド・シンタックス機能は、時間依存共変量の有無に関わらず、Cox 回帰で使用できます。

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

第 16 章 カテゴリー変数のコード化方式

多くのプロシージャーで、1 つのカテゴリー独立変数を一連の対比変数で自動置換するように要求できます。その後この対比変数セットは、ブロックとして式に投入されるか、または式から除去されます。対比変数セットをどのようにコード化するかを指定できます。通常これは CONTRAST サブコマンドで行います。この付録では、CONTRAST で要求された各種の対比タイプが、実際にどのように機能するかを説明し、例を示します。

偏差

全平均からの偏差。行列の項では、これらの対比が次の形式を取ります。

```
mean ( 1/k  1/k  ...  1/k  1/k)
df(1) (1-1/k -1/k  ... -1/k -1/k)
df(2) (-1/k  1-1/k  ... -1/k -1/k)
      .
      .
df(k-1) (-1/k  -1/k  ...  1-1/k -1/k)
```

k は独立変数のカテゴリー数です。デフォルトでは最後のカテゴリーが除外されます。例えば、カテゴリーが 3 つある独立変数の全平均との対比は次のようになります。

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3  2/3 -1/3)
```

最後のカテゴリー以外のカテゴリーを除外するには、DEVIATION キーワードの後に、除外するカテゴリーの番号を括弧に入れて指定します。例えば、次のサブコマンドを実行すると、1 番目と 3 番目のカテゴリーの偏差が求められ、2 番目のカテゴリーが除外されます。

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

FACTOR に 3 つのカテゴリーがあるとします。得られる対比行列は次のようになります。

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3  2/3)
```

単純

単純対比。因子の各レベルを最後のレベルと比較します。一般的な行列形式は次のとおりです。

```
mean (1/k  1/k  ...  1/k  1/k)
df(1) ( 1  0  ...  0  -1)
df(2) ( 0  1  ...  0  -1)
      .
      .
df(k-1) ( 0  0  ...  1  -1)
```

k は独立変数のカテゴリー数です。例えば、カテゴリーが 4 つある独立変数の単純対比は次のようになります。

```
(1/4  1/4  1/4  1/4)
( 1  0  0  -1)
( 0  1  0  -1)
( 0  0  1  -1)
```

最後のカテゴリではなく別のカテゴリを参照カテゴリとして使用するには、SIMPLE キーワードの後に、参照カテゴリのシーケンス番号を括弧に入れて指定します。これは必ずしもそのカテゴリに関連付けられた値ではありません。例えば、次の CONTRAST サブコマンドを実行すると、2 番目のカテゴリを除去する対比行列が得られます。

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

FACTOR に 4 つのカテゴリがあるとします。得られる対比行列は次のようになります。

```
(1/4  1/4  1/4  1/4)
(  1  -1   0   0)
(  0  -1   1   0)
(  0  -1   0   1)
```

Helmert

Helmert 対比。 独立変数のカテゴリを、後続のカテゴリの平均値と比較します。一般的な行列形式は次のとおりです。

```
mean (1/k  1/k  ...  1/k  1/k  1/k)
df(1) (  1 -1/(k-1) ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) (  0  1 ... -1/(k-2) -1/(k-2) -1/(k-2))
      .
      .
df(k-2) (  0  0 ...  1 -1/2 -1/2)
df(k-1) (  0  0 ...  0  1 -1)
```

k は独立変数のカテゴリ数です。例えば、カテゴリが 4 つある独立変数の Helmert 対比行列の形式は次のとおりです。

```
(1/4  1/4  1/4  1/4)
(  1 -1/3 -1/3 -1/3)
(  0  1 -1/2 -1/2)
(  0  0  1 -1)
```

差

逆 Helmert 対比。 独立変数のカテゴリを、その変数の以前のカテゴリの平均値と比較します。一般的な行列形式は次のとおりです。

```
mean (  1/k  1/k  1/k ... 1/k)
df(1) (  -1   1   0 ... 0)
df(2) ( -1/2 -1/2  1 ... 0)
      .
      .
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

k は独立変数のカテゴリ数です。例えば、カテゴリが 4 つある独立変数の逆 Helmert 対比は次のようになります。

```
( 1/4  1/4  1/4  1/4)
( -1   1   0   0)
(-1/2 -1/2  1   0)
(-1/3 -1/3 -1/3  1)
```

多項式

直交多項式の対比。 第 1 自由度にはすべてのカテゴリにわたる 1 次効果が含まれ、第 2 自由度には 2 次効果、第 3 自由度には 3 次効果というように、効果の次数が上がっていきます。

与えられたカテゴリ変数で測定される治療のレベルの間に間隔を指定することができます。距離を省略した場合のデフォルトである等間隔を、1 から k までの連続整数として指定できます (k はカテゴリの数)。変数 DRUG に 3 つのカテゴリがある場合、次のサブコマンド

```
/CONTRAST(DRUG)=POLYNOMIAL
```


は、次のサブコマンドと同じです。

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

ただし、等間隔が常に必要となるわけではありません。例えば、*DRUG* が 3 つのグループに投与される異なる薬品量を表すとします。2 番目のグループへの投薬量が 1 番目のグループの 2 倍であり、3 番目のグループへの投薬量が 1 番目のグループの 3 倍である場合、治療カテゴリーは等間隔であり、この状況に適した距離は次のように連続整数で表されます。

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

一方、2 番目のグループへの投薬量が 1 番目のグループの 4 倍であり、3 番目のグループへの投薬量が 1 番目のグループの 7 倍である場合、適切な距離は次のようになります。

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

どちらの場合も、対比を指定した結果、*DRUG* の第 1 自由度には投薬量レベルの 1 次効果が含まれ、第 2 自由度には 2 次効果が含まれます。

多項式対比は、トレンドを検定する場合や、応答曲面の性質を調べる場合に特に役立ちます。また、多項式対比を使用して、曲線回帰などの非線型曲線の当てはめを実行することもできます。

反復

独立変数の隣接するレベルを比較します。一般的な行列形式は次のとおりです。

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      .
      .
df(k-1) ( 0 0 0 ... 1 -1)
```

k は独立変数のカテゴリー数です。例えば、カテゴリーが 4 つある独立変数の反復対比は次のようになります。

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

この対比は、プロファイル分析や、差分スコアが必要な場合に役立ちます。

特殊

ユーザー定義の対比。与えられた独立変数のカテゴリーと同じ数の行と列がある正方行列の形式で、特殊対比を入力することができます。MANOVA および LOGLINEAR の場合、最初に入力された行は、常に平均効果 (等効果) であり、与えられた変数について他の独立変数 (存在する場合) を平均する方法を示す重みのセットを表します。一般に、この対比は 1 を並べたベクトルです。

行列の残りの行には、変数のカテゴリー間の比較を示す特殊対比が入ります。通常は、直交対比が最も有用です。直交対比は統計的に独立しており、冗長性がありません。対比は、次の場合に直交となります。

- 各行で、対比係数の合計が 0 である。
- 非連結行の全ペアの対応する係数の積和も 0 である。

例えば、治療に 4 つのレベルがあり、それぞれのレベルの治療を相互に比較するとします。適切な特殊対比は次のとおりです。

```
(1  1  1  1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0  2 -1 -1) compare 2nd with 3rd and 4th
(0  0  1 -1) compare 3rd with 4th
```

これを指定するには、MANOVA、LOGISTIC REGRESSION、および COXREG に対して次の CONTRAST サブコマンドを使用します。

```
/CONTRAST(TREATMNT)=SPECIAL( 1  1  1  1
                             3 -1 -1 -1
                             0  2 -1 -1
                             0  0  1 -1)
```

LOGLINEAR に対しては、次を指定する必要があります。

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1  1  1  1
                                   3 -1 -1 -1
                                   0  2 -1 -1
                                   0  0  1 -1)
```

平均値の行を除く各行の合計は 0 です。非連結行の各ペアの積和も 0 です。

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

特殊対比は、直交でなくてもかまいません。ただし、相互の線型結合であってはなりません。線形結合であると、プロシージャーから線型従属が報告され、処理が停止されます。Helmert 対比、逆 Helmert 対比、および多項式対比は、いずれも直交対比です。

インジケータ

指示変数のコード化。これはダミー・コード化とも呼ばれ、LOGLINEAR または MANOVA では使用できません。新しくコード化される変数の数は、 $k-1$ です。参照カテゴリーのケースは、 $k-1$ 個の変数すべてで 0 にコード化されます。 i 番目のカテゴリーのケースは、1 にコード化される i 番目以外のすべての指示変数で 0 にコード化されます。

第 17 章 共分散構造

このセクションでは、共分散構造の追加情報について説明します。

前従属: 1 次。この共分散構造には不均質の分散、および隣接する要素間で不均質の相関があります。2 つの隣接しない要素間の相関は、対象の要素間に存在する要素の間の相関の積となります。

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3) \\
 (\sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3) \\
 (\sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3) \\
 (\sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2)
 \end{array}$$

AR(1)。これは均質の分散の 1 次自己回帰構造です。任意の 2 つの要素間の相関は、要素が隣接する場合は ρ に等しく、要素が 3 番目の要素で分離されている場合は ρ^2 となり、以下同様に続きます。 $-1 < \rho < 1$ になるように制限されます。

$$\begin{array}{cccc}
 (\sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \sigma^2\rho^3) \\
 (\sigma^2\rho & \sigma^2 & \sigma^2\rho & \sigma^2\rho^2) \\
 (\sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \sigma^2\rho) \\
 (\sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho & \sigma^2)
 \end{array}$$

AR(1): 不均質。これは不均質の分散の 1 次自己回帰構造です。任意の 2 つの要素間の相関は要素が隣接する場合は r に等しく、2 つの要素が 3 番目の要素で分離されている場合は r^2 となり、以下同様に続きます。 -1 から 1 になるように制限されます。

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3) \\
 (\sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2) \\
 (\sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho) \\
 (\sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2)
 \end{array}$$

ARMA(1,1)。これは 1 次自己回帰の移動平均構造です。均質の分散があります。2 つ要素間の相関は要素が隣接する場合は $\phi\rho$ に等しく、要素が 3 番目の要素で分離されている場合は $\phi\rho^2$ となり、以下同様に続きます。それぞれ自己回帰および移動平均パラメーターであり、値は -1 から 1 の間に制限されます。

$$\begin{array}{cccc}
 (\sigma^2 & \sigma^2\phi\rho & \sigma^2\phi\rho^2 & \sigma^2\phi\rho^3) \\
 (\sigma^2\phi\rho & \sigma^2 & \sigma^2\phi\rho & \sigma^2\phi\rho^2) \\
 (\sigma^2\phi\rho^2 & \sigma^2\phi\rho & \sigma^2 & \sigma^2\phi\rho) \\
 (\sigma^2\phi\rho^3 & \sigma^2\phi\rho^2 & \sigma^2\phi\rho & \sigma^2)
 \end{array}$$

複合対称。この構造には定数分散および定数共分散があります。

$$\begin{array}{cccc}
 (\sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1) \\
 (\sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1) \\
 (\sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1) \\
 (\sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2)
 \end{array}$$

複合対称: 相関行列。この共分散構造には均質の分散および要素間の均質の相関があります。

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho)$
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho)$
$(\sigma^2\rho$	$\sigma^2\rho$	σ^2	$\sigma^2\rho)$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2)$

複合対称: 不均質。 この共分散構造には不均質の分散および要素間の定数相関があります。

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho)$
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho)$
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho)$
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	$\sigma_4^2)$

対角。 この共分散構造には不均質の分散があり、要素間の相関はゼロです。

$(\sigma_1^2$	0	0	0)
(0	σ_2^2	0	0)
(0	0	σ_3^2	0)
(0	0	0	$\sigma_4^2)$

因子分析的: 1 次。 この共分散構造には不均質の分散があります。これは要素間で不均質な項、および要素間で均質な項から構成されます。任意の 2 要素間の共分散は不均質の分散項の積の平方根です。

$(\lambda_1^2 + d$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1)$
$(\lambda_2\lambda_1$	$\lambda_2^2 + d$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2)$
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d$	$\lambda_4\lambda_3)$
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d)$

因子分析的: 1 次、不均質。 この共分散構造には不均質の分散があります。これは要素間で不均質な 2 項から構成されます。任意の 2 要素間の共分散は、不均質の分散項の最初の積の平方根です。

$(\lambda_1^2 + d_1$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1)$
$(\lambda_2\lambda_1$	$\lambda_2^2 + d_2$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2)$
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d_3$	$\lambda_4\lambda_3)$
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d_4)$

Huynh-Feldt。 これは循環行列で、任意の 2 要素間の共分散は、分散の平均から定数を引いたものと等しくなります。分散も共分散も定数ではありません。

$(\sigma_1^2$	$[\sigma_1^2 + \sigma_2^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_2^2]/2 - \lambda$	σ_2^2	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	σ_3^2	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$	$\sigma_4^2)$

スケーリングされた単位。 この構造には定数分散があります。どの要素間にも相関がないものと仮定されます。

$(\sigma^2$	0	0	0)
(0	σ^2	0	0)
(0	0	σ^2	0)
(0	0	0	$\sigma^2)$

Toeplitz. この共分散構造には均質の分散および要素間の不均質の相関があります。隣接する要素間の相関は、隣接する要素のペア間で均質です。3 番目の要素で分離される要素間の相関も均質です。

$$\begin{array}{cccc}
 (\sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 & \sigma^2\rho_3) \\
 (\sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2) \\
 (\sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1) \\
 (\sigma^2\rho_3 & \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2)
 \end{array}$$

Toeplitz: 不均質. この共分散構造には、不均質の分散および要素間の不均質の相関があります。隣接する要素間の相関は、隣接する要素のペア間で均質です。3 番目の要素で分離される要素間の相関も均質です。

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3) \\
 (\sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_1 & \sigma_4\sigma_2\rho_2) \\
 (\sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_4\sigma_3\rho_1) \\
 (\sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2)
 \end{array}$$

無構造. これは完全に一般的な共分散行列です。

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_{2\ 1} & \sigma_{31} & \sigma_{41}) \\
 (\sigma_{2\ 1} & \sigma_2^2 & \sigma_{32} & \sigma_{4\ 2}) \\
 (\sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{4\ 3}) \\
 (\sigma_{41} & \sigma_{4\ 2} & \sigma_{4\ 3} & \sigma_4^2)
 \end{array}$$

無構造: 相関行列. この共分散構造には不均質の分散および不均質の相関があります。

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41}) \\
 (\sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42}) \\
 (\sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43}) \\
 (\sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2)
 \end{array}$$

分散成分. この構造は、指定した各ランダム効果にスケーリングされた単位 (ID) 構造を割り当てます。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。

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

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

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

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

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

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

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

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

IBM Software Group

ATTN: Licensing

200 W. Madison St.

Chicago, IL; 60606

U.S.A.

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

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

この文書に含まれるいかなるパフォーマンス・データも、管理環境下で決定されたものです。そのため、他の操作環境で得られた結果は、異なる可能性があります。一部の測定が、開発レベルのシステムで行われた可能性があります。その測定値が、一般に利用可能なシステムのものと同じである保証はありません。さらに、一部の測定値が、推定値である可能性があります。実際の結果は、異なる可能性があります。お客様は、お客様の特定の環境に適したデータを確かめる必要があります。

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

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

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

著作権使用許諾:

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

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

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

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

商標

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

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

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

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

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

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

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

索引

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

[ア行]

イータの 2 乗
GLM 多変量 9
GLM 反復測定 18
一般化推定方程式 45
アクティブなデータ・セットへの変数の保存 54
応答 48
カテゴリー因子のオプション 50
初期値 52
推定基準 51
推定周辺平均 54
統計 53
モデル指定 50
モデルのエクスポート 55
モデルの種類 47
予測 49
2 値応答の参照カテゴリー 49
一般化線型混合モデル 57
オフセット 63
カスタム項目 61
共分散パラメーター 69
固定係数 68
固定効果 60, 67
推定周辺平均 65
推定平均 69
対象の分布 58
データ構造 67
フィールドの保存 65
分析の重み付け 63
分類テーブル 67
モデルのエクスポート 65
モデルの要約 66
モデル・ビュー 66
予測と観測 67
ランダム効果 62
ランダム効果共分散 68
ランダム効果ブロック 62
リンク関数 58
一般化線型モデル 33
アクティブなデータ・セットへの変数の保存 41
一般化線型混合モデル 57
応答 35

一般化線型モデル (続き)
カテゴリー因子のオプション 36
初期値 38
推定基準 38
推定周辺平均 40
統計 39
分布 33
モデル指定 37
モデルのエクスポート 42
モデル・タイプ 33
予測 36
リンク関数 33
2 値応答の参照カテゴリー 36
一般化対数オッズ比
一般対数線型分析 73
一般推定可能関数
一般化推定方程式 53
一般化線型モデル 39
一般対数線型分析
因子 73
基準 74
コマンドの追加機能 75
作図 74
残差 75
信頼区間 74
セル共変量 73
セル度数の分布 73
セルの構造 73
対比 73
表示オプション 74
変数の保存 75
モデル指定 74
予測値の保存 75
因子
GLM 反復測定 13
因子レベル情報
線型混合モデル 29
オッズ比
一般対数線型分析 73
オッズべき乗リンク関数
一般化推定方程式 47
一般化線型モデル 33
重み付き予測値
GLM 8
GLM 反復測定 17

[カ行]

階層対数線型モデル 71
階層的分解法 5, 14
分散成分 23

階層モデル
一般化線型混合モデル 57
カスタム・モデル
分散成分 22
モデル選択対数線型分析 72
GLM 反復測定 13
観測平均値
GLM 多変量 9
GLM 反復測定 18
ガンマ分布
一般化推定方程式 47
一般化線型モデル 33
記述統計
一般化推定方程式 53
一般化線型モデル 39
線型混合モデル 29
GLM 多変量 9
GLM 反復測定 18
逆ガウス分布
一般化推定方程式 47
一般化線型モデル 33
共分散行列
一般化推定方程式 51, 53
一般化線型モデル 38, 39
線型混合モデル 29
GLM 8
共分散構造 99
線型混合モデル 99
共分散の分析
GLM 多変量 3
共分散パラメーター検定
線型混合モデル 29
共変量
Cox 回帰 90
クロス集計表
モデル選択対数線型分析 71
ケース処理要約
一般化推定方程式 53
一般化線型モデル 39
効果サイズの推定
GLM 多変量 9
GLM 反復測定 18
交互作用項 4, 14, 22, 72, 74, 78
線型混合モデル 27
交差積
仮説と誤差行列 9
恒等リンク関数
一般化推定方程式 47
一般化線型モデル 33
項目の構築 4, 14, 22, 72, 74, 78

固定効果	
線型混合モデル	27
固定予測値	
線型混合モデル	30
混合モデル	
一般化線型混合モデル	57
線型	25

[サ行]

最小有意差	
GLM 多変量	7
GLM 反復測定	16
最大対数尤度比残差	
一般化線型モデル	41
最尤推定値	
分散成分	22
削除済み残差	
GLM	8
GLM 反復測定	17
作図	
一般対数線型分析	74
ロジット対数線型分析	79
残差	
一般化推定方程式	54
一般化線型モデル	41
一般対数線型分析	75
線型混合モデル	30
モデル選択対数線型分析	72
ロジット対数線型分析	79
残差 SSCP	
GLM 多変量	9
GLM 反復測定	18
残差プロット	
GLM 多変量	9
GLM 反復測定	18
残差分散共分散行列	
線型混合モデル	29
参照カテゴリー	
一般化推定方程式	49, 50
一般化線型モデル	36
信頼区間	
一般対数線型分析	74
線型混合モデル	29
ロジット対数線型分析	79
GLM 多変量	9
GLM 反復測定	18
推定周辺平均	
一般化推定方程式	54
一般化線型モデル	40
線型混合モデル	30
GLM 多変量	9
GLM 反復測定	18
スケール・パラメーター	
一般化推定方程式	51
一般化線型モデル	38

スコアリング	
線型混合モデル	29
すべての因子によるモデル	
分散成分	22
GLM 反復測定	13
正規確率プロット	
モデル選択対数線型分析	72
正規分布	
一般化推定方程式	47
一般化線型モデル	33
制限最尤推定値	
分散成分	22
生成クラス	
モデル選択対数線型分析	72
生存関数	
生命表	81
生存分析	
時間依存 Cox 回帰	93
生命表	81
Cox 回帰	89
Kaplan-Meier	85
生命表	81
因子変数	82
因子レベルの比較	82
コマンドの追加機能	83
作図	82
生存関数	81
生存状態変数	82
統計	81
ハザード比	81
表の表示の抑制	82
例	81
Wilcoxon (Gehan) 検定	82
セグメント化した時間依存の共変量	
Cox 回帰	93
線型混合モデル	25, 99
共分散構造	99
交互作用項	27
項目の構築	27
固定効果	27
コマンドの追加機能	31
推定基準	29
推定周辺平均	30
変数の保存	30
モデル	29
ランダム効果	28
相関行列	
一般化推定方程式	53
一般化線型モデル	39
線型混合モデル	29

[タ行]

対数順位検定	
Kaplan-Meier	86
対数線型分析	71

対数線型分析 (続き)	
一般化線型混合モデル	57
一般対数線型分析	73
ロジット対数線型分析	77
対数補数リンク関数	
一般化推定方程式	47
一般化線型モデル	33
対数尤度収束	
一般化推定方程式	51
一般化線型モデル	38
線型混合モデル	29
対数リンク関数	
一般化推定方程式	47
一般化線型モデル	33
対比	
一般対数線型分析	73
ロジット対数線型分析	77
Cox 回帰	90
対比係数行列	
一般化推定方程式	53
一般化線型モデル	39
多項分布	
一般化推定方程式	47
一般化線型モデル	33
多項ロジスティック回帰	
一般化線型混合モデル	57
多項ロジット・モデル	77
縦方向モデル	
一般化線型混合モデル	57
多変量 GLM	3
多変量回帰	3
多変量分散分析	3
段階 2 分	
一般化推定方程式	51
一般化線型モデル	38
線型混合モデル	29
調査済みケース	
生命表	81
Cox 回帰	89
Kaplan-Meier	85
適合度	
一般化推定方程式	53
一般化線型モデル	39
てこ比の値	
一般化線型モデル	41
GLM	8
GLM 反復測定	17
等分散性の検定	
GLM 多変量	9
GLM 反復測定	18
特異性許容度	
線型混合モデル	29
度数	
モデル選択対数線型分析	72

[ナ行]

ネスト項目

- 一般化推定方程式 50
- 一般化線型モデル 37
- 線型混合モデル 27

[ハ行]

ハザード比

- 生命表 81

パラメーター収束

- 一般化推定方程式 51
- 一般化線型モデル 38
- 線型混合モデル 29

パラメーター推定値

- 一般化推定方程式 53
- 一般化線型モデル 39
- 一般対数線型分析 73
- 線型混合モデル 29
- モデル選択対数線型分析 72
- ロジット対数線型分析 77

- GLM 多変量 9
- GLM 反復測定 18

パラメーター分散共分散行列

- 線型混合モデル 29

反復

- 一般化推定方程式 51
- 一般化線型モデル 38
- モデル選択対数線型分析 72

反復測定変数

- 線型混合モデル 26

反復履歴

- 一般化推定方程式 53
- 一般化線型モデル 39
- 線型混合モデル 29

被験者数

- 線型混合モデル 26

非標準化残差

- GLM 8
- GLM 反復測定 17

標準化残差

- GLM 8
- GLM 反復測定 17

標準誤差

- GLM 8
- GLM 多変量 9
- GLM 反復測定 17, 18

標準偏差

- GLM 多変量 9
- GLM 反復測定 18

比例ハザード・モデル

- Cox 回帰 89

負の 2 項分布

- 一般化推定方程式 47
- 一般化線型モデル 33

負の 2 項リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

負ログ・マイナス・ログ関数

- 一般化推定方程式 47
- 一般化線型モデル 33

プロビット分析

- 一般化線型混合モデル 57

プロビット・リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

プロファイル・プロット

- GLM 多変量 6
- GLM 反復測定 16

分割表

- 一般対数線型分析 73

分散成分 21

- オプション 22
- 結果の保存 23
- コマンドの追加機能 24
- モデル 22

分散分析

- 一般化線型混合モデル 57
- 分散成分 22
- GLM 多変量 3
- GLM 反復測定 11

分離

- 一般化推定方程式 51
- 一般化線型モデル 38

平方和 5, 14

- 仮説と誤差行列 9
- 線型混合モデル 27
- 分散成分 23

べき乗推定値

- GLM 多変量 9
- GLM 反復測定 18

べき乗リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

変数減少法

- モデル選択対数線型分析 71

飽和モデル

- モデル選択対数線型分析 72

補ログ・マイナス・ログ・リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

ポワソン回帰

- 一般化線型混合モデル 57
- 一般対数線型分析 73

ポワソン分布

- 一般化推定方程式 47
- 一般化線型モデル 33

[マ行]

マルチレベル・モデル

- 一般化線型混合モデル 57

文字列共変量

- Cox 回帰 90

モデル情報

- 一般化推定方程式 53
- 一般化線型モデル 39

モデル選択対数線型分析 71

- 因子範囲の定義 71
- オプション 72
- コマンドの追加機能 72
- モデル 72

モデル・ビュー

- 一般化線型混合モデル 66

[ヤ行]

尤度残差

- 一般化線型モデル 41

予測値

- 一般対数線型分析 75
- 線型混合モデル 30
- ロジット対数線型分析 79

[ラ行]

ランダム効果

- 線型混合モデル 28

ランダム効果共分散行列

- 線型混合モデル 29

ランダム効果先行

- 分散成分 22

リンク関数

- 一般化線型混合モデル 58

累積コーチット・リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

累積負ログ・マイナス・ログ関数

- 一般化推定方程式 47
- 一般化線型モデル 33

累積プロビット・リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

累積補ログ・マイナス・ログ・リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

累積ロジット・リンク関数

- 一般化推定方程式 47
- 一般化線型モデル 33

レベルと広がり

- GLM 多変量 9
- GLM 反復測定 18

ロジスティック回帰

- 一般化線型混合モデル 57

ロジット対数線型分析 77
 因子 77
 基準 79
 作図 79
 残差 79
 信頼区間 79
 セル共変量 77
 セル度数の分布 77
 セルの構造 77
 対比 77
 表示オプション 79
 変数の保存 79
 モデル指定 78
 予測値 79
 ロジット・リンク関数
 一般化推定方程式 47
 一般化線型モデル 33

[数字]

2 項分布
 一般化推定方程式 47
 一般化線型モデル 33

B

Bartlett の球面性の検定
 GLM 多変量 9
 Bonferroni
 GLM 多変量 7
 GLM 反復測定 16
 Box の M 検定
 GLM 多変量 9
 Breslow 検定
 Kaplan-Meier 86

C

Cook の距離
 一般化線型モデル 41
 GLM 8
 GLM 反復測定 17
 Cox 回帰 89
 イベントの定義 92
 カテゴリー共変量 90
 共変量 89
 コマンドの追加機能 92
 作図 90
 時間依存の共変量 93, 94
 新変数の保存 91
 ステップワイズ法による投入と除去
 91
 生存関数 91
 生存状態変数 92
 対比 90

Cox 回帰 (続き)
 統計 89, 91
 ハザード関数 91
 反復 91
 ベースライン関数 91
 偏残差 91
 文字列共変量 90
 例 89
 DfBeta 91

D

Duncan の多重範囲検定
 GLM 多変量 7
 GLM 反復測定 16
 Dunnett の C(U)
 GLM 多変量 7
 GLM 反復測定 16
 Dunnett の t 検定
 GLM 多変量 7
 GLM 反復測定 16
 Dunnett の T3(3)
 GLM 多変量 7
 GLM 反復測定 16

F

Fisher スコア法
 線型混合モデル 29
 Fisher の LSD
 GLM 多変量 7
 GLM 反復測定 16

G

Gabriel のペアごとの比較検定
 GLM 多変量 7
 GLM 反復測定 16
 Games-Howell のペアごとの比較検定
 GLM 多変量 7
 GLM 反復測定 16
 Gehan 検定
 生命表 82
 GLM
 行列の保存 8
 変数の保存 8
 GLM 多変量 3, 10
 因子 3
 オプション 9
 共変量 3
 従属変数 3
 診断 9
 推定周辺平均 9
 その後の検定 7
 表示 9

GLM 多変量 (続き)
 プロファイル・プロット 6
 GLM 反復測定 11
 因子の定義 13
 オプション 18
 コマンドの追加機能 19
 診断 18
 推定周辺平均 18
 その後の検定 16
 表示 18
 プロファイル・プロット 16
 変数の保存 17
 モデル 13
 GLOR
 一般対数線型分析 73

H

Hessian 収束
 一般化推定方程式 51
 一般化線型モデル 38
 Hochberg の GT2(H)
 GLM 多変量 7
 GLM 反復測定 16

K

Kaplan-Meier 85
 イベントの定義 86
 因子レベルの線型傾向 86
 因子レベルの比較 86
 コマンドの追加機能 87
 作図 87
 四分位 87
 新変数の保存 86
 生存時間の平均値と中央値 87
 生存状態変数 86
 生命表 87
 統計 85, 87
 例 85

L

L 行列
 一般化推定方程式 53
 一般化線型モデル 39
 LaGrange 乗数検定
 一般化線型モデル 39
 Levene の検定
 GLM 多変量 9
 GLM 反復測定 18

M

- Mauchly の球面性検定
 - GLM 反復測定 18
- MINQUE(M)
 - 分散成分 22

N

- Newman-Keuls
 - GLM 多変量 7
 - GLM 反復測定 16
- Newton-Raphson 法
 - 一般対数線型分析 73
 - ロジット対数線型分析 77

P

- Pearson 残差
 - 一般化推定方程式 54
 - 一般化線型モデル 41

R

- Ryan-Einot-Gabriel-Welsch の多重 F
 - GLM 多変量 7
 - GLM 反復測定 16
- Ryan-Einot-Gabriel-Welsch の多重範囲
 - GLM 多変量 7
 - GLM 反復測定 16
- R-E-G-W F(R)
 - GLM 多変量 7
 - GLM 反復測定 16
- R-E-G-W Q(Q)
 - GLM 多変量 7
 - GLM 反復測定 16

S

- Scheffé 検定
 - GLM 多変量 7
 - GLM 反復測定 16
- Sidak の t 検定
 - GLM 多変量 7
 - GLM 反復測定 16
- SSCP
 - GLM 多変量 9
 - GLM 反復測定 18
- Student-Newman-Keuls の検定
 - GLM 多変量 7
 - GLM 反復測定 16

T

- t 検定
 - GLM 多変量 9
 - GLM 反復測定 18
- Tamhane の T2(M)
 - GLM 多変量 7
 - GLM 反復測定 16
- Tarone-Ware 検定
 - Kaplan-Meier 86
- Tukey の b 検定
 - GLM 多変量 7
 - GLM 反復測定 16
- Tukey の HSD 検定
 - GLM 多変量 7
 - GLM 反復測定 16
- Tweedie 分布
 - 一般化推定方程式 47
 - 一般化線型モデル 33

W

- Wald 統計
 - 一般対数線型分析 73
 - ロジット対数線型分析 77
- Waller-Duncan の t 検定
 - GLM 多変量 7
 - GLM 反復測定 16
- Wilcoxon 検定
 - 生命表 82



Printed in Japan