

IBM SPSS Regression V27

IBM

注記

本書および本書で紹介する製品をご使用になる前に、[27 ページの『特記事項』](#)に記載されている情報をお読みください。

本書は、IBM® SPSS® Statistics バージョン V27 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slashと表示されたり、バックslashが円記号と表示されたりする場合があります。

原典：

IBM SPSS Regression V27

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

© Copyright International Business Machines Corporation .

目次

第 1 章 回帰	1
2 項ロジスティック回帰のプロシーチャーの選択.....	1
ロジスティック回帰.....	2
ロジスティック回帰の規則の設定.....	3
ロジスティック回帰の変数選択方法.....	3
ロジスティック回帰のカテゴリー変数の定義.....	3
ロジスティック回帰の新規変数の保存.....	4
ロジスティック回帰のオプション.....	4
LOGISTIC REGRESSION コマンドの追加機能.....	5
多項ロジスティック回帰.....	5
多項ロジスティック回帰.....	6
多項ロジスティック回帰の参照カテゴリー.....	7
多項ロジスティック回帰の統計.....	7
多項ロジスティック回帰の基準.....	7
多項ロジスティック回帰のオプション.....	8
多項ロジスティック回帰の保存.....	8
NOMREG コマンドの追加機能.....	9
プロビットの回帰.....	9
プロビット分析の範囲の定義.....	10
プロビット分析のオプション.....	10
PROBIT コマンドの追加機能.....	10
分位の回帰.....	11
分位の回帰: 基準.....	12
分位の回帰: モデル.....	13
分位の回帰: 表示.....	14
分位の回帰: 保存.....	15
分位の回帰: エクスポート.....	15
非線型回帰.....	16
条件論理 (非線型回帰).....	17
非線型回帰のパラメーター.....	17
非線型回帰の一般的なモデル.....	17
非線型回帰の損失関数.....	18
非線型回帰のパラメーター制約.....	18
非線型回帰の新規変数の保存.....	19
非線型回帰のオプション.....	19
非線型回帰の結果の解釈.....	19
NLR コマンドの追加機能.....	20
重み付け推定.....	20
重み付け推定のオプション.....	21
WLS コマンドの追加機能.....	21
2 段階最小二乗回帰.....	21
2 段階最小二乗回帰のオプション.....	22
2SLS コマンドの追加機能.....	22
カテゴリー変数のコード化方式.....	22
偏差.....	23
単純.....	23
Helmert.....	23
差.....	24
多項式.....	24
反復.....	25
特殊.....	25

インジケータ	26
特記事項	27
商標	28
索引	29

第1章 回帰

以下の回帰分析機能が、SPSS Statistics Standard Edition または Regression オプションに含まれています。

2 項ロジスティック回帰のプロシーチャーの選択

ロジスティック回帰プロシージャおよび多項ロジスティック回帰プロシージャを使用して、2 項ロジスティック回帰モデルの当てはめを行うことができます。それぞれのプロシージャには、もう一方のプロシージャにはないオプションがあります。2つのプロシージャには重要な理論上の違いがあります。ロジスティック回帰プロシージャでは、データの入力方法や共変量パターンの数がケースの総数より少ないかどうかに関係なく、個々のケース・レベルのデータを使用してすべての予測、残差、影響力の統計、および適合度検定が出力されます。一方、多項ロジスティック回帰プロシージャでは、ケースが内部で集計されて、予測値の共変量パターンが同一である複数の部分母集団が構成され、これらの部分母集団に基づいて予測、残差、および適合度検定が出力されます。すべての予測値がカテゴリーであるか、または連続型予測値の値の数が限られている場合(そのために個々の共変量パターンに複数のケースが存在する場合)は、部分母集団アプローチで有効な適合度検定と有益な残差が得られますが、個別のケース・レベル・アプローチでは得られません。

ロジスティック回帰

以下に示す独自機能を備えています。

- モデルの適合度を調べる Hosmer-Lemeshow 検定
- ステップワイズ分析
- モデルのパラメーター化を定義する対比
- 分類のための代替分割点
- 分類プロット
- 1つのケース・セット上で一連のホールドアウト・ケースに当てはめられたモデル
- 予測、残差、および影響力の統計の保存

多項ロジスティック回帰

以下に示す独自機能を備えています。

- モデルの適合度を調べる Pearson と逸脱カイ 2 乗検定
- 適合度検定用のデータをグループ化するための部分母集団の指定
- 部分母集団別の度数、予測度数、および残差のリスト表示
- 過分散に対する分散推定値の修正
- パラメーター推定値の共分散行列
- パラメーターの線型結合の検定
- ネストされたモデルの明示的指定
- 差分変数を使用した 1 対 1 一致条件付きロジスティック回帰モデルの当てはめ

注:

- これらのプロシージャはいずれも、2 項分布とロジットリンク関数を使用する一般化線型モデルである 2 値データのモデルの当てはめを行います。使用するデータに対して別のリンク関数がより適切である場合、一般化線型モデル プロシージャを使用する必要があります。
- 2 値データ(あるいはその他の点で相関があるレコード)の反復測定の場合、一般化線型混合モデル プロシージャまたは一般化推定方程式プロシージャを検討してください。

ロジスティック回帰

ロジスティック回帰は、一連の予測変数の値に基づいて特性または結果の有無を予測する必要がある場合に役立ちます。線型回帰モデルに似ていますが、従属変数が二分変数であるモデルに適しています。ロジスティック回帰係数を使用して、モデル内の各独立変数のオッズ比を推定することができます。ロジスティック回帰は、判別分析よりも広い範囲の調査に適用されます。

例: 生活様式のどの特性が、冠状動脈性心臓病 (CHD) の危険因子になるかを調べます。喫煙状況、食生活、運動、飲酒、および CHD の状態を測定した患者のサンプルがある場合に、4 つの生活様式を変数として使用するモデルを作成して、患者のサンプルにおける CHD の有無を予測できます。このモデルを使用して、各因子のオッズ比の推定値を導き出すことで、例えば、喫煙者は非喫煙者と比べて CHD になる確率がどの程度高いかなどがわかります。

統計: 各分析については、ケースの総数、選択したケース、有効なケース。各カテゴリ変数については、パラメーターのコード化。各ステップについては、投入または削除された変数、反復履歴、-2 対数尤度、適合度、Hosmer-Lemeshow 適合度統計、モデルのカイ 2 乗、改良カイ 2 乗、分類テーブル、変数間の相関、観測グループと予測確率グラフ、残差のカイ 2 乗。式内の各変数については、係数 (B)、B の標準誤差、Wald 統計、推定オッズ比 ($\exp(B)$)、 $\exp(B)$ の信頼区間、モデルから項が除外された場合は対数尤度。式にない各変数については、スコア統計。各ケースについては、観測グループ、予測確率、予測グループ、残差、標準化残差。

方法: 変数のブロック投入を使用するか、またはステップワイズ法のいずれか (条件付き変数増加法、尤度比変数増加法、Wald 変数増加法、条件付き変数減少法、尤度比変数減少法、または Wald 変数減少法) を使用して、モデルを推定できます。

ロジスティック回帰データの考慮事項

データ: 従属変数は二分変数でなければなりません。独立変数には、区間レベルまたはカテゴリを使用できます。カテゴリの場合は、変数をダミー変数または指示変数としてコード化する必要があります (プロシージャーで、カテゴリ変数を自動的に再コードすることもできます)。

仮定: ロジスティック回帰は、判別分析と同じようには分布仮定に依存しません。ただし、予測変数が多変量正規分布の場合は、より安定した解が得られることがあります。また、他の形式の回帰と同様に、予測変数間の多重共線性により、推定値の偏りや標準誤差の誇張が生じることがあります。このプロシージャーは、所属グループが実際にカテゴリ変数である場合に最も効果的です。所属グループが連続型変数の値 (例えば「高い IQ」と「低い IQ」) に基づく場合は、連続型変数自体の豊富な情報量を利用できるように、線型回帰の使用を検討してください。

関連プロシージャー: データの多重共線性を検査するには、散布図プロシージャーを使用します。多変量正規性と等分散共分散行列の仮定が満たされる場合は、判別分析プロシージャーを使用することで、より迅速に解が得られます。すべての予測変数がカテゴリである場合は、対数線型プロシージャーも使用できます。従属変数が連続型の場合は、線型回帰プロシージャーを使用します。ROC 曲線プロシージャーを使用して、ロジスティック回帰プロシージャーで保存された確率を作図することができます。

ロジスティック回帰分析の実施

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「二項ロジスティック ...」

2. 二値従属変数を 1 つ選択します。この変数は、数値または文字列です。

3. 共変量を 1 つ以上選択します。交互作用項を包含するには、交互作用に関係するすべての変数を選択し、「>a*b>」を選択します。

変数をグループ (ブロック) で投入するには、ブロックに対して共変量を選択し、「次へ」をクリックして新規のブロックを指定します。すべてのブロックを指定するまで、この作業を繰り返します。

オプションとして、分析するケースを選択できます。選択変数を選択し、ルール基準を入力します。

ロジスティック回帰の規則の設定

選択規則によって定義されたケースが、モデルの推定に使用されます。例えば、変数と「等しい」を選択し、値5を指定した場合は、選択した変数の値が5であるケースのみがモデルの推定で使用されます。

統計および分類の結果は、選択されたケースと選択されなかったケースの両方に対して生成されます。これにより、生成されたモデルを検証するために、既存のデータに基づいて新規ケースを分類したり、データを学習用および検証用サブセットに分割したりすることが可能になります。

ロジスティック回帰の変数選択方法

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

- 「入力」。変数選択手続きの1つ。ブロック内のすべての変数を1つのステップで投入します。
- *Forward Selection (Conditional)* (変数増加法 (条件付き))。ステップワイズ選択法の1つ。スコア統計量の有意性に基づいて投入をテストし、条件付きパラメーター推定値に基づく尤度比統計量の確率に基づいて除去をテストします。
- *Forward Selection (Likelihood Ratio)* (変数増加法 (尤度比))。ステップワイズ選択法の1つ。スコア統計量の有意性に基づいて投入をテストし、最大偏尤度推定値に基づく尤度比統計量の確率に基づいて除去をテストします。
- *Forward Selection (Wald)* (変数増加法 (Wald))。ステップワイズ選択法の1つ。スコア統計量の有意性に基づいて投入をテストし、Wald 統計量の確率に基づいて除去をテストします。
- *Backward Elimination (Conditional)* (変数減少法 (条件付き))。変数減少ステップワイズ選択。条件付きパラメーター推定値に基づく尤度比統計の確率に基づいて除去の判定を行います。
- *Backward Elimination (Likelihood Ratio)* (変数減少法 (尤度比))。変数減少ステップワイズ選択。最大偏尤度推定値に基づく尤度比統計の確率に基づいて除去の判定を行います。
- *Backward Elimination (Wald)* (変数減少法 (Wald))。変数減少ステップワイズ選択。Wald 統計量の確率に基づいて除去の判定を行います。

出力の有意確率値は、1つのモデルの当てはめに基づいています。したがって、ステップワイズ法を使用した場合は、一般に有意確率値が無効になります。

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

ロジスティック回帰のカテゴリ変数の定義

ロジスティック回帰プロシージャーでのカテゴリ変数の処理方法を、詳しく指定することができます。

「**共変量**」。メインのダイアログ・ボックスで指定された単独または交互作用の一部としての共変量が、層を問わずすべてリストされます。そのうちの一部が文字列変数またはカテゴリ変数である場合は、カテゴリ共変量としてのみ使用できます。

「**カテゴリ共変量**」。カテゴリ変数として識別された変数がリストされます。変数ごとに、使用する対比コード化が括弧に入れて表記されます。文字列変数 (名前の後の記号 < で示されます) は、「カテゴリ共変量」リストに既に含まれています。「共変量」リストからその他のカテゴリ共変量を選択して、「カテゴリ共変量」リストに移動します。

「**対比の変更**」。対比方法を変更することができます。選択可能な対比方法は次のとおりです。

- 「**指標**」。対比は、所属カテゴリの有無を示します。参照カテゴリは、対比行列でゼロの行として表されます。
- 「**単純**」。参照カテゴリを除く予測変数の各カテゴリが、参照カテゴリと比較されます。
- 「**差分**」。最初のカテゴリを除く予測変数の各カテゴリが、先行するカテゴリの平均効果と比較されます。逆 Helmert 対比とも呼ばれます。

- 「**Helmert**」。最後のカテゴリーを除く予測変数の各カテゴリーが、後続のカテゴリーの平均効果と比較されます。
- **反復**: 最後のカテゴリーを除く予測変数の各カテゴリーが、次のカテゴリーと比較されます。
- 「**多項式**」。直交多項式の対比。各カテゴリーが等間隔で配置されていると仮定されます。多項式対比は、数値型変数にのみ使用できます。
- 「**偏差**」。参照カテゴリーを除く予測変数の各カテゴリーが、全体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。「変更」をクリックするまで方法は実際に変更されないの注意してください。

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、メインのダイアログ・ボックスで、その変数を包含するすべての項を「共変量」リストから削除する必要があります。

ロジスティック回帰の新規変数の保存

ロジスティック回帰の結果を新規の変数としてアクティブ・データ・セットに保存することができます。

「**予測値**」。モデルで予測された値を保存します。選択可能なオプションは「確率」と「所属グループ」です。

- *Probabilities* (確率). ケースごとにイベント発生の予測確率を保存します。出力のテーブルには、新しい変数の名前と内容が表示されます。「イベント」は、値が高い方の従属変数のカテゴリーです。例えば、従属変数の値が 0 と 1 である場合は、カテゴリー 1 の予測確率が保存されます。
- *Predicted Group Membership* (予測された所属グループ). 判別得点に基づいた、事後確率が最大になるグループ。モデルがケースの所属先として推定したグループ。

「**影響力**」。ケースが予測値に与える影響を測定する統計の値を保存します。選択可能なオプションは「Cook」、「てこ比の値」、および「DfBeta」です。

- *Cook's* (Cook の統計量). Cook の影響統計量に類似したロジスティック回帰での値。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。
- *Leverage Value* (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。
- *DfBeta(s)*. ベータ値の差は、特定のケースを除外した結果として生じる回帰係数の変化です。値は、モデル内の項ごとに (定数項を含む) 計算されます。

「**残差**」。残差を保存します。選択可能なオプションは、「標準化されていない」、「ロジット」、「スチューデント化」、「標準化」、および「逸脱」です。

- *Unstandardized Residuals* (標準化されていない残差). 観測した値と、モデルによって予測された値との差。
- *Logit Residual* (ロジット残差). ロジット・スケールで予測した場合のケースの残差。ロジット残差は、残差を予測確率で割り、 $(1 - \text{予測確率})$ を掛けた値です。
- *Studentized Residual* (スチューデント化された残差). ケースをモデルから除外した場合の、モデルの逸脱度の変化。
- *Standardized Residuals* (標準化残差). 残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼びます。平均値は 0 であり、標準偏差は 1 です。
- *Deviance* (逸脱). モデルの逸脱に基づく残差。

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

ロジスティック回帰のオプション

ロジスティック回帰分析では、オプションを指定することができます。

「**統計と作図**」。統計と作図を要求することができます。選択可能なオプションは、「分類プロット」、「Hosmer-Lemeshow の適合度」、「残差のケースごとのリスト」、「推定値の相関」、「反復履歴」、および

「Exp(B)の信頼区間」です。「表示」グループでいずれかの選択肢を選択して、「ステップごと」または「最後のステップ」(最終モデルの場合のみ)で統計と作図を表示します。

- *Hosmer-Lemeshow goodness-of-fit statistic (Hosmer-Lemeshowの適合度統計量)*. この適合度統計量は、特に連続共変量を持つモデルおよび標本サイズが小さい調査の場合に、ロジスティック回帰で使用される従来の適合度統計量よりも頑健です。ケースをリスクの十分位数にグループ化し、それぞれの十分位数内で観測確率と期待確率を比較した結果に基づいています。

「ステップワイズにおける確率」。式に対して変数を投入または削除する基準を制御できます。「投入」または「削除」で、変数の投入または削除を行う基準を指定します。

- *Probability for Stepwise (ステップワイズにおける確率)*. スコア統計量の確率が「投入」の値より小さい場合は変数をモデルに投入し、確率が「削除」の値より大きい場合は変数を除去します。デフォルト設定をオーバーライドするには、「投入」と「削除」に正の整数を入力してください。「投入」は「除去」より小さくしなければなりません。

「分類の打ち切り」。ケースを分類するための分割点を決定できます。予測値が分類の打ち切りを超えるケースは正に分類され、予測値が打ち切りを下回るケースは負に分類されます。デフォルトを変更するには、0.01から0.99までの値を入力します。

「最大反復回数」。モデルの反復を終了する基準となる最大反復回数を変更できます。

「モデル内に定数項を含める」。モデル内に定数項を含めるかどうかを指定できます。無効にした場合は、定数項が0になります。

LOGISTIC REGRESSION コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変数の値または変数ラベルを使用して、ケースごとに出力を識別する。
- 反復レポートの間隔を制御する。反復のたびにパラメーター推定値を印刷するのではなく、 n 回目の反復の度にパラメーター推定値を要求できます。
- 反復を終了して冗長性をチェックするための基準を変更する。
- ケースごとのリスト作成で変数リストを指定する。
- 処理中に各分割ファイル・グループのデータを外部のスクラッチ・ファイルに保持することで、メモリーを節約する。

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。

多項ロジスティック回帰

多項ロジスティック回帰は、一連の予測変数の値に基づいて被験者を分類する場合に役立ちます。このタイプの回帰はロジスティック回帰に似ていますが、従属変数が2つのカテゴリーに制限されていないため、より一般的です。

例: 映画のマーケティングをより効果的に行うために、映画会社では、映画ファンが見に行く可能性が高い映画の種類を予測したいと考えています。多項ロジスティック回帰を実行することにより、観客の年齢、性別、および同伴者の状況が、観客が好む映画の種類に及ぼす影響の大きさを特定できます。こうして、特定の映画の宣伝キャンペーンを、その映画を見に行く可能性が高い人々に合わせて実施することができます。

統計: 反復履歴、パラメーター係数、漸近分散共分散行列と相関行列、モデル効果と部分効果の尤度比検定、 -2 対数尤度。Pearson および逸脱のカイ 2 乗適合度。Cox と Snell、Nagelkerke、および McFadden R^2 。分類: 応答カテゴリー別の観測度数と予測度数。クロス集計: 共変量パターンおよび応答カテゴリー別の観測度数と予測度数(残差あり)、および比率。

方法: 多項ロジット・モデルは、すべての因子によるモデル、またはユーザー指定モデルに適しています。パラメーター推定は、反復最尤法アルゴリズムを使用して実行されます。

多項ロジスティック回帰データの考慮事項

データ: 従属変数は、カテゴリ変数でなければなりません。独立変数には、因子または共変量を使用できません。一般に、因子はカテゴリ変数でなければならず、共変量は連続型変数でなければなりません。

仮定: 任意の2つのカテゴリのオッズ比が、他のすべての応答カテゴリから独立していると仮定されます。例えば、新製品を市場に投入する場合、この仮定によると、他のすべての製品の市場占有率が同じ比率で影響を受けることになります。また、共変量パターンが与えられた場合、応答は独立多項変数であると仮定されます。

多項ロジスティック回帰分析の実施

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「多項ロジスティック回帰...」

2. 従属変数を1つ選択します。

3. 因子はオプションであり、数値またはカテゴリのいずれかを指定できます。

4. 共変量はオプションですが、指定する場合は数値でなければなりません。

多項ロジスティック回帰

多項ロジスティック回帰プロシージャでは、デフォルトで因子主効果と共変量主効果を含むモデルが作成されますが、このダイアログ・ボックスで、カスタム・モデルを指定したり、ステップワイズ・モデル選択を要求したりすることができます。

「**モデルの指定**」。主効果モデルには、共変量主効果と因子主効果が含まれますが、交互作用効果は含まれません。すべての因子によるモデルには、すべての主効果とすべての因子間交互作用が含まれます。共変量の交互作用は含まれません。カスタム・モデルを作成して因子交互作用または共変量交互作用のサブセットを指定するか、またはモデル項のステップワイズ選択を要求することができます。

「**因子と共変量**」。因子と共変量がリストされます。

「**強制投入の項**」。強制投入リストに追加された項は、常にモデルに包含されます。

「**ステップワイズの項**」。ステップワイズ・リストに追加された項は、ユーザーが選択した次のいずれかのステップワイズ法に従ってモデルに包含されます。

- 「**変数増加法投入**」。この方法は、モデルにステップワイズの項がない状態で開始されます。各ステップでは、モデルから除外されたすべてのステップワイズの項が、モデルに追加されても統計的に有意な寄与率を持たなくなるまで、最も有意な項がモデルに追加されます。
- 「**変数減少法**」。この方法は、ステップワイズ・リストで指定されたすべての項をモデルに投入することで開始されます。各ステップでは、残りのすべてのステップワイズの項がモデルに対して統計的に有意な寄与率を持つまで、最も有意でないステップワイズの項がモデルから除外されます。
- 「**変数増加ステップワイズ法**」。この方法は、変数増加法投入で選択されるモデルで開始されます。その後、モデル内のステップワイズの項に対する変数減少法と、モデルから除外された項に対する変数増加法投入をアルゴリズムが交互に繰り返します。これは、投入または削除の基準を満たす項がなくなるまで続行されます。
- 「**変数減少ステップワイズ法**」。この方法は、変数減少法で選択されるモデルから開始されます。その後、アルゴリズムが、モデルから除外された項に対する変数増加法投入と、モデル内のステップワイズの項に対する変数減少法を交互に繰り返します。これは、投入または削除の基準を満たす項がなくなるまで続行されます。

「**モデルに切片を含む**」。モデルに対して、切片項を包含したり除外したりすることができます。

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に対して、最上位の交互作用項を作成します。

「**主効果**」。選択した変数ごとに、主効果項を作成します。

「**2要因まで**」。選択した変数に対し、考えられる2要因の交互作用をすべて作成します。

「**3 要因まで**」。選択した変数に対し、考えられる 3 要因の交互作用をすべて作成します。

「**4 要因まで**」。選択した変数に対し、考えられる 4 要因の交互作用をすべて作成します。

「**5 要因まで**」。選択した変数に対し、考えられる 5 要因の交互作用をすべて作成します。

多項ロジスティック回帰の参照カテゴリー

多項ロジスティック回帰プロシージャーでは、デフォルトで最後のカテゴリーが参照カテゴリーになります。このダイアログ・ボックスでは、参照カテゴリーおよびカテゴリーの順序を制御できます。

「**参照カテゴリー**」。最初のカテゴリー、最後のカテゴリー、またはカスタム・カテゴリーを指定します。

「**カテゴリーの順序**」。昇順の場合は、最低値によって最初のカテゴリーが定義され、最高値によって最後のカテゴリーが定義されます。降順の場合は、最高値によって最初のカテゴリーが定義され、最低値によって最後のカテゴリーが定義されます。

多項ロジスティック回帰の統計

多項ロジスティック回帰に対して、次の統計を指定することができます。

「**ケース処理要約**」。このテーブルには、指定したカテゴリー変数に関する情報が含まれています。

「**モデル**」。モデル全体の統計です。

- 「**疑似 R² 乗**」。Cox と Snell、Nagelkerke、および McFadden の各 R² 乗統計を出力します。
- 「**ステップの要約**」。このテーブルでは、ステップワイズ法の各ステップで投入または削除された効果が要約されます。「**モデル**」ダイアログ・ボックスでステップワイズ・モデルが指定されていない場合、このテーブルは作成されません。
- 「**モデル適合情報**」。このテーブルでは、切片のみのモデルまたは帰無仮説モデルが、適合モデルと比較されます。
- 「**情報基準**」。このテーブルでは、赤池情報量基準 (AIC) および Schwarz のベイズ情報量基準 (BIC) が出力されます。
- 「**セルの確率**」。共変量パターンおよび応答カテゴリー別の観測度数、期待度数 (残差あり)、および比率のテーブルが出力されます。
- 「**分類テーブル**」。観測応答と予測応答を対比するテーブルが出力されます。
- 「**適合度カイ 2 乗統計量**」。Pearson カイ 2 乗統計量と尤度比カイ 2 乗統計量が出力されます。すべての因子と共変量、または因子と共変量のユーザー定義サブセットで決定される共変量パターンについて、統計量が計算されます。
- 「**単調性測定方法**」。一致ペア、不一致ペア、および同順位ペアの数に関する情報を含んだテーブルが表示されます。このテーブルには、Somers の D、Goodman と Kruskal のガンマ、Kendall のタウ a、および一致指数 C も表示されます。

「**パラメーター**」。モデル・パラメーターに関連した統計です。

- 「**推定値**」。ユーザー指定の信頼度レベルで、モデル・パラメーターの推定値が出力されます。
- 「**尤度比検定**」。モデルの部分効果の尤度比検定が出力されます。モデル全体の検定は、自動的に出力されます。
- 「**漸近相関**」。パラメーター推定相関の行列が出力されます。
- 「**漸近共分散**」。パラメーター推定共分散の行列が出力されます。

「**部分母集団の定義**」。セル確率および適合度検定で使用される共変量パターンを定義するために、因子と共変量のサブセットを選択することができます。

多項ロジスティック回帰の基準

多項ロジスティック回帰に対して、次の基準を指定することができます。

「**反復回数**」。アルゴリズムを繰り返し実行する最大回数、段階 2 分の最大段階数、対数尤度とパラメーターの変更に対する収束許容度、反復アルゴリズムの進行状況を出力する頻度、およびプロシージャーでデータの完全分離または準完全分離のチェックを開始する必要がある反復を指定できます。

- 「**対数尤度収束**」。対数尤度関数での絶対変化が指定した値を下回ると、収束したと見なされます。値が0であると、この基準は使用されません。負でない値を指定してください。
- 「**パラメーター収束**」。パラメーター推定値での絶対変化がこの値を下回ると、収束したと見なされます。値が0であると、この基準は使用されません。

「**デルタ**」。1より小さく負でない値を指定できます。この値は、共変量パターン別の応答カテゴリーのクロス集計表にある各空セルに追加されます。これによって、アルゴリズムが安定し、推定値に偏りが生じなくなります。

「**特異性許容度**」。特異性の検査に使用する許容範囲を指定できます。

多項ロジスティック回帰のオプション

多項ロジスティック回帰に対して、次のオプションを指定することができます。

「**散らばりの尺度**」。パラメーター分散共分散行列の推定値を修正するために使用される散らばりの尺度値を指定できます。「**逸脱**」では、逸脱関数(尤度比カイ2乗)統計量を使用して尺度値が推定されます。「**Pearson**」では、Pearsonカイ2乗統計量を使用して尺度値が推定されます。さらに、独自の尺度値を指定することもできます。これは正の数値でなければなりません。

「**ステップワイズのオプション**」。このオプションにより、ステップワイズ法を使用してモデルを作成する場合の統計基準を制御できます。このオプションは、「**モデル**」ダイアログ・ボックスでステップワイズ・モデルが指定されない限り無視されます。

- 「**投入の確率**」。これは、変数投入に対する尤度比統計の確率です。指定した確率が大きいほど、変数がモデルに投入されやすくなります。この基準は、変数増加法投入、変数増加ステップワイズ法、または変数減少ステップワイズ法が選択されない限り無視されます。
- 「**投入検定**」。これは、ステップワイズ法で項目を投入するための方法です。尤度比検定とスコア検定のいずれかを選択します。この基準は、変数増加法投入、変数増加ステップワイズ法、または変数減少ステップワイズ法が選択されない限り無視されます。
- 「**除去の確率**」。これは、変数の削除に対する尤度比統計量の確率です。指定した確率が大きいほど、変数がモデルに保持されやすくなります。この基準は、変数減少法、変数増加ステップワイズ法、または変数減少ステップワイズ法が選択されていない場合は無視されます。
- 「**除去検定**」。これは、ステップワイズ法で項目を削除するための方法です。尤度比検定とWald検定のいずれかを選択します。この基準は、変数減少法、変数増加ステップワイズ法、または変数減少ステップワイズ法が選択されていない場合は無視されます。
- 「**モデルの最小ステップ効果**」。変数減少法または変数減少ステップワイズ法を使用する場合に、モデルに包含する項の最小数を指定します。切片は、モデル項としてカウントされません。
- 「**モデルの最大ステップ効果**」。変数増加法投入または変数増加ステップワイズ法を使用する場合に、モデルに包含する項の最大数を指定します。切片は、モデル項としてカウントされません。
- 「**階層制約投入および項目削除**」。このオプションでは、モデル項の包含に関する制約を設定するかどうかを選択できます。階層の場合は、包含するすべての項について、その項の一部である低次の項を最初にモデルに包含する必要があります。例えば、階層要件が有効である場合は、*Marital status*Gender* 交互作用を追加する前に、*Marital status* と *Gender* の両方の因子をモデルに包含する必要があります。3つのラジオ・ボタン・オプションにより、階層を決定する際の共変量の役割が決まります。

多項ロジスティック回帰の保存

「**保存**」ダイアログ・ボックスでは、変数を作業ファイルに保存したり、モデル情報を外部ファイルにエクスポートしたりすることができます。

「**保存変数**」。次の変数を保存することができます。

- 「**推定応答確率**」。因子/共変量パターンが応答カテゴリーに分類される確率の推定値です。推定確率は応答変数のカテゴリーと同じ数だけ存在し、最大25個まで保存されます。
- 「**予測カテゴリー**」。因子/共変量パターンの期待確率が最大である応答カテゴリーです。
- 「**予測カテゴリー確率**」。最大の推定応答確率です。
- 「**実際のカテゴリー確率**」。因子/共変量パターンが観測カテゴリーに分類される確率の推定値です。

「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とその共分散 (オプション) が、指定したファイルに XML (PMML) 形式でエクスポートされます。このモデル・ファイルを使用して、スコアリングのために他のデータ・ファイルにモデル情報を適用できます。

NOMREG 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。

- 従属変数の参照カテゴリーを指定する。
- ユーザー欠損値があるケースを包含する。
- 帰無仮説をパラメーターの線型結合として指定することにより、仮説検定をカスタマイズする。

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。

プロビットの回帰

このプロシージャーでは、刺激の強さと、刺激に対して一定の応答を示すケースの割合との関係を測定します。このプロシージャーは、ある独立変数のレベルが影響しているか、または原因となっていると思われる二分出力がある場合に役立ちます。また、特に実験データに適しています。このプロシージャーでは、メディアン有効量のように、一定の割合の応答を導くために必要な刺激の強さを推定することができます。

例: 新しい殺虫剤がアリの駆除にどの程度有効か、またどの濃度で使用するのが適切かを調べます。アリのサンプルにさまざまな濃度の殺虫剤を散布する実験を行い、駆除されたアリの数と殺虫剤を散布したアリの数を記録します。このデータにプロビットの回帰を適用すると、濃度と殺虫力の関係の強さを判別できます。また、例えば、殺虫剤を散布したアリの 95% を確実に駆除したい場合の適切な殺虫剤の濃度を特定できます。

統計: 回帰係数と標準誤差、切片と標準誤差、Pearson のカイ 2 乗の適合度、観測度数と期待度数、独立変数の有効レベルの信頼区間。プロット: 変換された応答プロット。

このプロシージャーでは、Gill、Murray、Saunders、および Wright が開発した NPSOL[®] で提案され、実装されているアルゴリズムを使用してモデル・パラメーターが推定されます。

プロビットの回帰データの考慮事項

データ: 独立変数の各値 (独立変数が複数ある場合は値の各組み合わせ) について、対象の応答を示す値を含んだケースの数を応答変数とし、独立変数の値を含んだケースの総数を総観測変数とする必要があります。因子変数には、整数でコード化したカテゴリー変数を使用します。

仮定: 観測値は独立している必要があります。観測調査で見られるように、観測値の数に比べて独立変数の値が多い場合は、カイ 2 乗と適合度の統計が有効とならないことがあります。

関連プロシージャー: プロビット分析は、ロジスティック回帰と密接に関連しています。実際に、ロジット変換を選択すると、このプロシージャーでは実質的にロジスティック回帰が計算されます。一般に、プロビット分析は計画実験に適しているのに対し、ロジスティック回帰は観測調査に適しています。その違いは出力に顕著に現れます。プロビット分析プロシージャーは、さまざまな応答比率の有効値 (メディアン有効量を含む) の推定値をレポートします。一方、ロジスティック回帰プロシージャーは、独立変数のオッズ比の推定値をレポートします。

プロビットの回帰分析の実施

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「プロビット...」

2. 応答度数変数を選択します。この変数は、検定刺激への応答を示すケースの数を表します。この変数の値を負にすることはできません。
3. 総観測変数を選択します。この変数は、刺激が適用されたケースの数を示します。この変数の値は負にすることができず、各ケースの応答度数変数の値以上である必要があります。

オプションで、因子変数を選択することができます。その場合は、「**範囲の定義**」を使用してグループの範囲を定義します。

4. 共変量を1つ以上選択します。この変数には、各観測値に適用される刺激のレベルが格納されます。共変量を変換する場合は、「**変換**」ドロップダウン・リストから適用する変換を選択します。変換を適用せず、かつ制御グループがある場合は、その制御グループが分析に取り込まれます。
5. 「**プロビット**」または「**ロジット**」モデルを選択します。

Probit Model (プロビット・モデル)

プロビット変換 (累積標準正規分布関数の逆関数) を応答率に適用します。

Logit Model (ロジット・モデル)

ロジット (オッズの対数) 変換を応答率に適用します。

プロビット分析の範囲の定義

分析する因子変数のレベルを指定することができます。因子レベルは連続する整数としてコード化する必要があります。指定した範囲内のすべてのレベルが分析されます。

プロビット分析のオプション

プロビット分析では、オプションを指定することができます。

「**統計**」。オプションの統計として、「**度数**」、「**相対メディアン潜在性**」、「**平行性の検定**」、および「**フィデューシアル信頼区間**」を要求できます。

- **Relative Median Potency (相対メディアン潜在性)**. 因子レベルの各ペアに対するメディアン潜在性の比率を表示します。各相対メディアン潜在性の 95% 信頼限界も表示します。因子変数がない場合または複数の共変量がある場合は、相対メディアン潜在性を使用できません。
- **Parallelism Test (平行性の検定)**. すべての因子レベルが共通の傾きを持つという仮説を検定します。
- **Fiducial Confidence Intervals (フィデューシアル信頼区間)**. 特定の応答確率を得るために必要な作用因の適用量の信頼区間。

複数の共変量を選択した場合は、「**フィデューシアル信頼区間**」と「**相対メディアン潜在性**」を使用できません。「**相対メディアン潜在性**」と「**平行性の検定**」は、因子変数を選択した場合にのみ使用できます。

「**自然応答比**」。刺激が存在しない場合でも自然応答比を指定することができます。指定可能な選択肢は、「**なし**」、「**データから計算**」、および「**値**」です。

- **Calculate from Data (データから計算)**. サンプル・データから自然応答率を推定します。データには、共変量の値が 0 の制御レベルを表すケースがなければなりません。プロビットは、制御レベルの応答比を初期値として使用して自然応答率を推定します。
- **Value (値)**. モデルの自然応答率を設定します (この項目は、自然応答率が事前に判明している場合に選択します)。自然応答率を入力してください (比率は 1 より小さくしなければなりません)。例えば、刺激が 0 のときに時間の 10% で応答が発生する場合は、0.10 と入力します。

「**基準**」。反復パラメーター推定アルゴリズムのパラメーターを制御できます。「**最大反復回数**」、「**ステップの限界**」、および「**最適性許容度**」に対するデフォルトを変更できます。

PROBIT 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。

- プロビット・モデルとロジット・モデルの両方の分析を要求する。
- 欠損値の処理を制御する。
- 底 10 対数または自然対数以外の底で共変量を変換する。

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。

分位の回帰

回帰分析は、定量的モデリングで広く使用されている統計手法です。多重線型回帰は、研究者が複数の変数の値を使用してスケール結果の平均値を説明または予測するための基本的かつ標準的なアプローチです。しかし、多くの場合、知りたいのはスケール結果の中央値などの分位値です。

分位の回帰は、一連の予測 (独立) 変数と、目標 (従属) 変数の特定のパーセンタイル (または分位。多くの場合は中央値) の間の関係をモデル化します。これは、最小 2 乗回帰と比べて、以下の 2 つの主要な利点があります。

- 分位の回帰では目標変数の分布について仮定を設けない。
- 分位の回帰は異常な観測値の影響を受けにくい傾向がある。

分位の回帰は、生態学、医療、金融経済などの業界の調査に広く使用されています。

例

総世帯収入と収入に占める食費の比率との間の関係を調べます。エンゲルの法則は、収入が上昇すると、食品に対する絶対的な支出が増えても、収入に占める食費の比率は低下するという経済学の説です。分位の回帰をこれらのデータに適用すると、(ある収入を得ている 100 世帯における) 90% の世帯をカバーする食費が分かります。この場合、食費の平均に関心はありません。

統計量

分位の回帰、シンプレックス手法、Frisch-Newton 内点非線形最適法アルゴリズム、Barrodale と Roberts、Bofinger、Hall Sheather、帯域幅、有意水準、行列操作、収束基準、回帰重み、定数項、予測目標、予測残差、作表、予測プロット、パラメータ推定値、共分散行列、相関行列、観測値、信頼区間。

このプロシージャは、Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles, *Econometrica*, 46, 33-50 で提案されたアルゴリズムを使用します。

分位の回帰データの考慮事項

データ

単一の数値型の従属変数が必要です。目標変数は連続型の変数でなければなりません。カテゴリ型の予測変数の場合は、予測変数を連続型変数またはダミー変数にすることができます。分析を実行するには、切片項または少なくとも 1 つの予測変数が必要です。

仮定

分位の回帰は、目標変数の分布に対して仮定を設けず、異常な観測値の影響も受けにくい回帰です。

関連プロシージャ

分位分析は、最小 2 乗回帰に関連しています。

分位の回帰分析の実施

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「分位...」

ダイアログでは、分位の回帰分析に使用する目標、因子、共変量、および重みの変数を指定できます。このダイアログには、複雑な分析や大規模なデータセットの場合にメモリーを節約するオプションも用意されています。

2. 数値型の目標変数を選択します。分析を実行するために必要な目標変数は 1 つだけです。数値型変数だけを使用できます。
3. 必要に応じて、1 つ以上の因子変数を選択します。スケール変数を使用することはできません。
4. 必要に応じて、1 つ以上の共変量変数を選択します。文字型変数を使用することはできません。

注: 「因子」リストと「共変量」リストが両方とも空であり、かつ「モデル」ダイアログで「モデルに切片を含む」を選択した場合は、以下のメッセージが表示されます。

```
No effects have been specified. Therefore, an intercept only model will be fit.
Do you want to fit an intercept-only model?
```

5. 必要に応じて、回帰重み変数を選択します。文字型変数を使用することはできません。
6. 必要に応じて、「複雑な分析または大きなデータセット向けのメモリを保持する」を選択します。この設定は、処理中に外部ファイルにデータを保持するかどうかを制御します。この設定を有効にすると、複雑な分析や大規模なデータセットでの分析を実行するときにメモリリソースを節約できます。

分位の回帰: 基準

「基準」ダイアログは以下のオプションを提供します。

分位

分位を指定するためのオプションを提供します。

単一分位を指定 (Specify single quantiles)

これを選択した場合、分析を実行するには少なくとも1つの値が必要です。複数の値を指定できます。それぞれの値は [0, 1] の範囲になければなりません。それぞれの値をブランク・スペースで区切って、複数の値を指定できます。分位値リストの値を操作するには、「追加」、「変更」、「削除」の各ボタンを使用します。

値はすべて固有でなければなりません (値が重複してはなりません)。デフォルト値は 0.5 です。

グリッド分位を指定 (Specify grid quantiles)

これを選択した場合、「開始」値 (value1) から「終了」値 (value2) まで、「基準」 (value3) の増分で、分位のグリッドを指定できます。これを指定した場合、有効な [value1 TO value2 BY value3] のセットは1つだけ許可されます。これは $0 \leq \text{value1} \leq \text{value2} \leq 1$ を満たしている必要があります。value1 = value2 の場合は、value3 に関係なく、単一の value1 を指定することと同じになります。

推定方法

モデル推定方法を指定するためのオプションを提供します。

プログラムにより自動的に選択

適切な推定方法をプロシージャに自動的に選択させます。これはデフォルト設定です。

シンプレックス アルゴリズム

Barrodale および Roberts によって開発されたシンプレックス アルゴリズムを呼び出します。

Frisch-Newton 内点非線形最適法

Frisch-Newton 内点非線形最適法のアルゴリズムを呼び出します。

事後推定

パラメーター推定値の分散共分散および予測目標値に対する信頼区間の事後推定のためのオプションを提供します。

ケースが IID であると仮定する

この設定を選択すると、誤差項が独立同分布であると仮定されます。この設定を選択しない場合は、大規模なモデルの場合に計算時間が大幅に長くなることがあります。この設定はデフォルトで選択されています。

帯域幅タイプ

パラメータ推定値の分散共分散行列を推定するために使用する帯域幅の方式 (「Bofinger」または「Hall-Sheather」) を決定します。「Bofinger」がデフォルト設定です。

数値的な方法

以下のオプションを提供します。

特異性許容度

内点法における行列操作の許容値を指定します。指定する値は単一の倍精度値 (0, 10^{-3}) でなければなりません。デフォルト設定は 10^{-12} です。

収束基準

数値解法の収束基準を指定します。指定する値は単一の倍精度値 (0, 10^{-3}) でなければなりません。デフォルト設定は 10^{-6} です。

最大反復回数

反復の最大回数を指定します。指定する値は単一の正整数でなければなりません。デフォルト値は 2000 です。

欠損値

欠損値の処理方法を指定できます。

ユーザー欠損値とシステム欠損値の両方を除外

選択すると、ユーザー欠損値とシステム欠損値の両方が除外されます。

ユーザー欠損値を有効として取り扱う

選択すると、ユーザー欠損値が有効として扱われます。

信頼区間 (%)

有意水準を指定します。指定する場合の値は、0 から 100 までの単一の倍精度値でなければなりません。デフォルト値は 95 です。

分位の回帰: モデル

「モデル」ダイアログには、モデルで使用する効果と重みを指定するためのオプションが用意されています。省略した場合、または単独で指定した場合、モデルは、定数項、ならびに共変量リストの共変量および因子リストの因子によるすべての主効果を含みます。

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。「**項目の構築**」を選択した場合は、ネストなし項目に対して以下の効果および交互作用のオプションを指定できます。

主効果

選択した変数ごとに、主効果の項目が作成されます。

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 2 要因の交互作用がすべて作成されます。

3次まで

選択した変数で考えられる 3 要因の交互作用がすべて作成されます。

4次まで

選択した変数で考えられる 4 要因の交互作用がすべて作成されます。

5次まで

選択した変数で考えられる 5 要因の交互作用がすべて作成されます。

「**入れ子項目の構築**」を選択した場合は、ネスト項目を構築できます。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、これらの場所のうち 1 つの場所にのみ頻繁に訪れるため、「顧客」効果は「店舗の場所」効果内でネストされていると言えます。

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。

ネストに関する注意:

- 2 つの因子の間の交互作用の効果を含めるには、キーワード BY またはアスタリスク (*) を使用して、交互作用に関わる因子を結合します。
- 交互作用効果内の因子はそれぞれ異ならなければなりません。
- 項を別の項の中にネストするための効果を含めるには、括弧の組を使用します。
- 複数の括弧の対が存在する場合、それぞれの対が、別の括弧の対の内部に囲まれている (ネストされている) 必要があります。
- 多重ネストも可能です。
- ネストされた効果の間の交互作用はサポートされていません。

制限: 入れ子項目には、次の制限があります。

- 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。

モデルに切片を含む

選択すると、モデルに切片項が含まれます。選択しなかった場合は、分析を実行するために少なくとも1つの予測変数が必要です。この設定はデフォルトで有効になっています。

分位の回帰: 表示

「表示」ダイアログは、出力およびプロットを調節する設定を提供します。

印刷

以下の出力オプションが選択可能です。

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分散共分散行列

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。

作図と作表

以下のプロット・オプションを使用できます。

次のパラメータ推定値をプロット

特定の数の上位の効果、またはモデル内のすべての効果に対するパラメータ推定値のプロットを選択できます。「上位 xx 件の効果 (Top xx effects)」の設定は、1つの共変量と1つまたは2つの因子との交互作用内でプロットするカテゴリの数 (混合効果の場合はカテゴリの組み合わせの数) を制御します。値は単一の正整数でなければなりません (デフォルト設定は50です)。

注:

- 指定した整数値がカテゴリまたは組み合わせの数より大きい場合は、すべての効果について予測プロットが作成されます。
- この設定が有効なのは、「基準」ダイアログの「分位値」設定に複数の値が指定されている場合だけです。分位値が1つしか指定されていない場合、プロットは作成されません。

予測対観測のプロットを表示

予測値と観測値のプロットの作成を制御します。有効にすると、複数の点 (分位を表す異なる色が付いた点) が含まれたプロットが1つ作成されます。この設定は、デフォルトでは無効になっています。

モデルの効果を予測

有効な場合は、以下のオプションが使用可能です。

上位 x 件の効果を作図または作表 (Plot or tabulate the top x effects)

予測プロットまたは予測表を作成する上位の効果の数を指定します。3がデフォルト値です。

注: 指定した値がモデル内の有効な効果の数より大きい場合は、すべての効果について予測プロットまたは予測表が作成されます。

ユーザー指定の効果を作図または作表

有効な効果のガイドラインは以下のとおりです。

- 1つの共変量 (共変量自体の高い有意水準を含む) を持つ効果: プロットを1つ作成し、それぞれの分位によって予測される線を引きます。

- 1つの因子を持つ効果: 分位ごとに、因子のカテゴリの予測を表にします。
- 2つの因子の交互作用を持つ効果: 分位ごとに、2つの因子のカテゴリの予測を表にします。
- 1つの共変量と1つまたは2つの因子との交互作用を持つ効果: 分位ごとにプロットを1つ作成し、交互作用効果内の各カテゴリまたはカテゴリの組み合わせごとに線を引きます。
- プロットする組み合わせの最大数は、「混合効果における最大 xx 個のカテゴリまたはカテゴリの組み合わせをプロット (Plot maximum xx categories of combinations of categories in a mixed effect)」に指定された値によって決まります。

「モデル効果」から「予測の線」リストに移動した効果がプロットに使用されます。指定した効果が定数である場合、プロットは作成されません(モデルの構築から除外されます)。

「モデル効果」から「予測表」リストに移動した効果が作表に使用されます。指定した効果が定数である場合、表は作成されません(モデルの構築から除外されます)。

混合効果における最大 xx 個のカテゴリまたはカテゴリの組み合わせをプロット (Plot maximum xx categories or combinations of categories in a mixed effect)

プロットするカテゴリの組み合わせの最大数を制御します。デフォルト値は 10 です。

分位の回帰: 保存

「保存」ダイアログは、モデルをスコアリングするためのオプションを提供します。

応答の予測値

選択すると、予測された目標値が記録されます。

残差

選択すると、予測された残差が記録されます。

予測区間の下限

選択すると、予測区間の下限が記録されます。

予測区間の上限

選択すると、予測区間の上限が記録されます。

注: 変数名は保存オプションごとに指定できます。ルート名を指定する場合は、有効な変数名でなければなりません。「基準」ダイアログの「分位値」設定に複数の値を指定した場合は、ルート名の後にアンダースコア (_) 文字と分かりやすい分位の接尾辞を付けたものが使用されます。

分位の回帰: エクスポート

「エクスポート」ダイアログは、エクスポートする統計、および統計のエクスポート方法(外部データ ファイルかデータセットか)を指定するためのオプション、および処理中のデータの処理方法(通常どおりに処理するか、処理中に外部スクラッチ ファイルに保持するか)を制御するためのオプションを提供します。

パラメータ推定値の分散共分散行列

選択すると、パラメータ推定値の共分散行列を外部データ ファイルまたは事前に宣言したデータセットに書き込むオプションが使用可能になります。

パラメータ推定値の相関行列

選択すると、パラメータ推定値の相関行列を外部データ ファイルまたは事前に宣言したデータセットに書き込むオプションが使用可能になります。

共分散/相関行列は、単一データセットまたは外部データ ファイル(複数の回帰の分位が存在する場合)に保存されます

複数の分位が存在する場合は、このオプションを使用して、単一もしくは複数のデータセットまたは外部データ ファイルへの共分散行列や相関行列の保存を切り替えます。無効にすると、行列が単一の外部データ ファイルまたはデータセットに保存されます。有効にすると、行列が複数の外部データ ファイルまたはデータセットに保存されます。この設定が効力を持つのは、「基準」ダイアログの「分位値」設定に複数の値が指定されている場合だけです。

注: このオプションを使用できるのは、「パラメータ推定値の分散共分散行列」または「パラメータ推定値の相関行列」を選択した場合だけです。

モデル情報を XML ファイルにエクスポート

選択すると、モデル情報を特定の XML ファイル名および場所にエクスポートするためのオプションが提供されます。

XML としてエクスポート

「モデル情報を XML ファイルにエクスポート」を選択すると、パラメータ推定値と分散共分散行列のエクスポート、またはパラメータ推定値のみのエクスポートを選択できます。「パラメータ推定値と分散共分散行列」がデフォルト設定です。

ファイル名の規則

- 「基準」ダイアログの「分位値」設定に値を 1 つ指定した場合は、外部データ ファイルまたはデータセットの指定に savefile および dataset が使用されます。
- 「基準」ダイアログの「分位値」設定に複数の値を指定した場合は、各分位が外部データ ファイルまたはデータセットに保存されます。
- 自動的に、データ ファイルまたはデータセットの名前にはアンダースコア文字「_」と、その後に分かりやすい分位の接尾辞とが付加されます。例えば、「分位値」として 0.25、0.50、および 0.75 を指定すると、データ ファイル名の .sav 拡張子の前に接尾辞 _25、_50、および _75 が付加されます。
- 各分位の接尾辞には、必要に応じて追加の桁を指定できます。
- 「分位値」の先行ゼロおよび小数点は接尾辞に使用されません。
- 「分位値」に指数表記を指定した場合は、接尾辞として出力するときには小数値に変換されます。

非線型回帰

非線型回帰は、従属変数と一連の独立変数の間の関係の非線型モデルを検出する方法です。線型モデルの推定に限定される従来の線型回帰とは異なり、非線型回帰では従属変数と独立変数の間に任意の関係があるモデルを推定することができます。これを行うために、反復推定アルゴリズムが使用されます。このプロシージャは、 $Y = A + BX^{**2}$ 形式の単純な多項式モデルでは必要ありません。 $W = X^{**2}$ を定義することで、単純な線型モデル $Y = A + BW$ が得られます。これは、線型回帰プロシージャなどの従来の方法を使用して推定できます。

例: 人口を時間に基づいて予測することが可能かを調べます。ある散布図が、人口と時間の間に強い関係が見られることを示しています。ただし、その関係は非線型のため、非線型回帰プロシージャの特殊な推定方法が必要です。ロジスティック人口増加モデルなどの適切な式を設定することで、モデルの適切な推定値を得て、実際には測定されなかった時点の人口を予測できます。

統計: 各反復については、パラメーター推定値および残差平方和。各モデルについては、回帰平方和、残差、無修正総和と修正総和、パラメーター推定値、漸近標準誤差、パラメーター推定値の漸近相関行列。

注: 制約条件付き非線型回帰では、Gill、Murray、Saunders、および Wright が開発した NPSOL[®] で提案され、実装されているアルゴリズムを使用してモデル・パラメーターが推定されます。

非線型回帰データの考慮事項

データ: 従属変数および独立変数は、量的でなければなりません。宗教、専攻、居住地区などのカテゴリ変数は、2 値 (ダミー) 変数またはその他のタイプの対比変数へと再コード化する必要があります。

仮定: 結果が有効となるのは、従属変数と独立変数の関係を正確に表す関数を指定した場合に限られます。また、適切な開始値を選択することが非常に重要になります。モデルの機能的形式を正しく指定した場合でも、使用する開始値が不適切であると、モデルが収束しなかったり、グローバルな最適解ではなくローカルな最適解が得られたりすることがあります。

関連プロシージャ: 最初は非線型に見えるモデルの多くは、線型モデルに変換でき、線型回帰プロシージャを使用して分析することができます。適切なモデルが不明である場合は、曲線推定プロシージャを使用して、データにおける有用な機能的関係を特定することができます。

非線型回帰分析の実施

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「非線形...」

2. アクティブ・データ・セット内の変数のリストから、数値型従属変数を1つ選択します。
3. モデル式を作成するには、「モデル式」フィールドに式を入力するか、成分(変数、パラメーター、関数)をこのフィールドに貼り付けます。
4. 「パラメーター」をクリックして、モデルのパラメーターを特定します。

分割モデル(ドメインのどの部分にあるかによって形式が変わるモデル)は、単一モデル・ステートメント内で条件論理を使用して指定する必要があります。

条件論理(非線型回帰)

条件論理を使用して、分割モデルを指定することができます。モデル式または損失関数で条件論理を使用するには、一連の項の和を各条件に1つ形成します。各項は、論理式(括弧内)と、その論理式が真である場合に生成される式の積で構成されます。

例えば、 $X \leq 0$ の場合に0、 $0 < X < 1$ の場合にX、 $X >= 1$ の場合に1と等しくなる分割モデルについて考えると、式は次のようになります。

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X >= 1) * 1$$

括弧内の論理式は、すべて1(真)または0(偽)に評価されます。したがって次のようになります。

$X \leq 0$ の場合、上記は $1 * 0 + 0 * X + 0 * 1 = 0$ になります。

$0 < X < 1$ の場合、上記は $0 * 0 + 1 * X + 0 * 1 = X$ になります。

$X > 1$ の場合、上記は $0 * 0 + 0 * X + 1 * 1 = 1$ になります。

別の論理式と結果式に置き換えることで、さらに複雑な例も簡単に作成できます。 $0 < X < 1$ などの二重不等式は、 $(X > 0 \ \& \ X < 1)$ などの複合式で記述する必要がありますので注意してください。

論理式では文字列変数を使用することができます。

$$(city='New York') * costliv + (city='Des Moines') * 0.59 * costliv$$

この式により、ニューヨークの住人用に1つの式が得られ(変数 *costliv* の値)、デモインの住人用に別の式(この値の59%)が得られます。文字列定数は、上記のように引用符またはアポストロフィで囲む必要があります。

非線型回帰のパラメーター

パラメーターは、非線型回帰プロシージャで推定されるモデルの一部です。パラメーターには、付加定数、乗法係数、指数、または関数評価で使用される値があります。定義したすべてのパラメーターが、メインのダイアログ・ボックスのパラメーター・リストに(初期値とともに)表示されます。

「名前」。各パラメーターに名前を指定する必要があります。この名前は、有効な変数名であること、およびメインのダイアログ・ボックスで指定したモデル式で使用した名前であることが必要です。

「初期値」。パラメーターの初期値を指定できます。可能であれば、期待する最終解にできるだけ近い値を指定します。不適切な初期値を指定すると、収束が失敗したり、解の収束がローカルになったり(グローバルではなく)、物理的に不可能となったりする可能性があります。

「前回の分析結果を初期値として使用」。このダイアログ・ボックスから非線型回帰を既に実行している場合は、このオプションを選択することで、前回の実行の値からパラメーターの初期値を取得できます。これにより、アルゴリズムが徐々に収束している場合に検索を継続できます(初回の初期値はメインのダイアログ・ボックスのパラメーター・リストに表示されたままとなります)。

注: この選択は、セッションが終了するまでダイアログ・ボックスで有効のままとなります。モデルを変更する場合は、必ず選択を解除してください。

非線型回帰の一般的なモデル

次の表に、数多くある公開非線型回帰モデルのモデル・シンタックスの例を示します。無作為に選択したモデルが、使用するデータにうまく適合することはまずありません。パラメーターには適切な開始値が必要です。また、モデルによっては、収束させるために制約が必要となります。

表 1. モデル・シンタックスの例

名前	モデル式
漸近回帰	$b1 + b2 * \exp(b3 * x)$
漸近回帰	$b1 - (b2 * (b3 ** x))$
密度	$(b1 + b2 * x) ** (-1 / b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
対数修正	$(b1 + b3 * x) ** b2$
対数ロジスティック	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Metcherlich の収穫逓減の法則	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$
3 次の有理式	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
2 次の有理式	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x ** b4)$
収穫密度	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

非線型回帰の損失関数

非線型回帰の**損失関数**は、アルゴリズムによって最小化される関数です。残差平方和を最小化する場合は「**残差平方和**」を選択し、別の関数を最小化する場合は「**ユーザーの定義した損失関数**」を選択します。

「**ユーザーの定義した損失関数**」を選択した場合は、選択したパラメーター値で合計 (全ケースが対象) を最小化する必要がある損失関数を定義する必要があります。

- ほとんどの損失関数には、残差を表す特殊変数 *RESID_* が関係します (デフォルトの残差平方和損失関数は、*RESID_**2* として明示的に入力できます)。損失関数で予測値を使用する必要がある場合、その値は従属変数から残差を引いた値と等しくなります。
- 条件論理を使用して、条件付き損失関数を指定することができます。

「ユーザー定義の損失関数」フィールドに式を入力するか、このフィールドに式の成分を貼り付けることができます。文字列定数は引用符 (アポストロフィ) で囲む必要があります。また数値定数は、小数点にドットを使用する米国形式で入力する必要があります。

非線型回帰のパラメーター制約

制約は、解を求める反復でパラメーターに指定できる値を制限します。線型式はステップの実行前に評価されるため、線型制約を使用してオーバーフローするおそれがあるステップを防止できます。非線型式は、ステップの実行後に評価されます。

各式または不等式には、次の要素が必要です。

- モデル内の少なくとも 1 つのパラメーターを包含する式。式を入力するか、またはキーパッドを使用して数値、演算子、または括弧を式に貼り付けることができます。必要なパラメーターを式の残りの部分と

ともに入力するか、画面左側のパラメーター・リストから貼り付けることができます。制約では通常の変数を使用できません。

- 3つの論理演算子 (<=、=、>=) のいずれか。
- 数値定数。この数値定数と式が論理演算子を使用して比較されます。定数を入力します。数値定数は、小数点にドットを使用する米国形式で入力する必要があります。

非線型回帰の新規変数の保存

アクティブなデータ・ファイルに、新規の変数を複数保存することができます。選択可能なオプションは、「残差」、「予測値」、「微分係数」、および「損失関数値」です。これらの変数を後続の分析で使用して、モデルの適合を検定したり、問題のあるケースを特定したりできます。

- *Residuals* (残差). 残差を変数名 `resid` で保存します。
- *Predicted Values* (予測値). 変数名 `pred_` を使用して予測値を保存します。
- *Derivatives* (微分係数). モデル・パラメーターごとに1つの微分係数が保存されます。微分係数名は、パラメーター名の先頭6文字の前に「d.」を付けて作成されます。
- *Loss Function Values* (損失関数値). このオプションは、ユーザー独自の損失関数を指定した場合に使用することができます。変数名 `loss_` が損失関数の値に割り当てられます。

非線型回帰のオプション

オプションによって、非線型回帰分析のさまざまな側面を制御することができます。

Bootstrap Estimates (ブートストラップ推定値). 元のデータ・セットから反復したサンプルを使用することによって統計量の標準誤差を推定する方法。これは、元のデータ・セットと同じサイズのサンプルを (置き換えを行って) 抽出することで行います。これらの各サンプルについて、非線型式が推定されます。次に、各パラメーター推定値の標準誤差を、ブートストラップ推定値の標準偏差として計算します。元のデータ・セットから得たパラメーター値を、各ブートストラップ・サンプルの初期値として使用します。これには、逐次2次計画法のアルゴリズムが必要です。

「推定方法」。可能であれば、推定方法を選択することができます (このダイアログ・ボックスまたは他のダイアログ・ボックスでの選択内容によっては、逐次2次計画法のアルゴリズムが必要となります)。選択可能なオプションは、「逐次2次計画法」と「Levenberg-Marquardt法」です。

- *Sequential Quadratic Programming* (逐次2次計画法). この手法は、制約条件のあるモデルにも制約条件のないモデルにも使用できます。制約条件のあるモデル、ユーザー定義の損失関数、またはブートストラップを指定した場合は、自動的に逐次2次計画法を使用します。「最大反復回数」および「ステップの限界」には新しい値を入力できます。また、「最適性許容度」、「関数精度」、および「無限ステップ・サイズ」の各ドロップダウン・リストの選択内容は変更できます。
- *Levenberg-Marquardt* (Levenberg-Marquardt法). 制約条件のないモデルの場合のデフォルトのアルゴリズムです。制約条件のあるモデル、ユーザー定義の損失関数、またはブートストラップを指定した場合は Levenberg-Marquardt 法を使用できません。「最大反復回数」には新しい値を入力できます。また、「平方和の収束」および「パラメーター収束」の各ドロップダウン・リストの選択内容は変更できます。

非線型回帰の結果の解釈

非線型回帰では、計算上の困難が生じる問題が頻繁に発生します。

- パラメーターの初期値の選択は、収束に影響します。初期値には、合理的、かつ可能であれば期待する最終解に近いものを選択してください。
- 問題によっては、あるアルゴリズムよりも別のアルゴリズムを使用した方が、よい結果を得られる場合があります。別のアルゴリズム (選択可能な場合) は、「オプション」ダイアログで選択します (損失関数または一定のタイプの制約を指定する場合は、Levenberg-Marquardt アルゴリズムを使用できません)。
- 反復が最大回数に達したために反復が停止した場合、「最終」モデルはおそらく適切な解ではありません。「パラメーター」ダイアログで「前回の分析結果を初期値として使用」を選択して反復を続行するか、さらに適切な方法として別の初期値を選択します。
- 大きなデータ値のべき乗または大きなデータ値によるべき乗を必要とするモデルでは、オーバーフローやアンダーフロー (コンピューターで表現できないほど数値が大きくなる、または小さくなる現象) が発生

することがあります。適切な初期値を選択するか、パラメーターに制約を設定することで、この現象を回避できる場合があります。

NLR 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。

- パラメーター推定値の初期値の読み込み元となるファイルを指定する。
- 複数のモデル・ステートメントと損失関数を指定する。これにより、分割モデルの指定が簡単になります。
- プログラムで計算される微分係数ではなく、独自の微分係数を指定する。
- 生成するブートストラップ・サンプルの数を指定する。
- 追加の反復基準を指定する(微分係数チェック用の限界値を設定したり、残差と微分係数間の相関の収束基準の定義したりするなど)。

CNLR (制約条件付き非線型回帰) コマンドに基準を追加することで、次のことが可能になります。

- 主要な反復の中で実行できる副次的な反復の最大回数を指定する。
- 微分係数のチェックに使用する限界値を設定する。
- ステップの限界を設定する。
- 初期値が指定された範囲内にあるかを判別するためのクラッシュ許容度を指定する。

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

重み付け推定

標準の線型回帰モデルでは、調査中の母集団内で分散が一定していると仮定します。この仮定が当てはまらない場合(例えば、ある属性が高いケースが、その属性が低いケースよりも変動が大きい場合など)、最小二乗法(OLS)を使用する線型回帰では、最適なモデルを推定できなくなります。変動の差を別の変数から予測できる場合は、重み付け推定プロシージャで、重み付き最小二乗法(WLS)を使用して線型回帰モデルの係数を計算できます。この場合、より正確な観測値(変動が少ない観測値)に、回帰係数を決定する際により大きな重みを与えるようにします。重み付け推定プロシージャは、一連の重み変換を検定し、データに最も適合するものを示します。

例: インフレーションと失業が株価の変動に与える影響を調べます。株価が高い株式は、株価が安い株式よりも変動が大きい場合が多いため、最小二乗法では最適な推定値を得られません。重み付け推定を使用すると、線型モデルの計算で価格変化の変動性に株価が与える影響を説明できます。

統計: 検定された重み付けソース変数のべき乗ごとの対数尤度値、多重 R 、 R^2 乗、調整済み R^2 乗、WLS モデルの分散分析テーブル、非標準化パラメーター推定値と標準化パラメーター推定値、および WLS モデルの対数尤度。

重み付け推定データの考慮事項

データ: 従属変数および独立変数は、量的でなければなりません。宗教、専攻、居住地区などのカテゴリ変数は、2 値(ダミー)変数またはその他のタイプの対比変数へと再コード化する必要があります。重み付け変数は量的であること、および従属変数の変動に関連していることが必要です。

仮定: 独立変数の各値に対して、従属変数の分布が正規分布でなければなりません。従属変数と各独立変数の関係が線型であること、およびすべての観測値が独立していることが必要です。従属変数の分散は独立変数のレベルによって異なる場合がありますが、その差を重み付け変数に基づいて予測できる必要があります。

関連プロシージャ: 探索プロシージャを使用して、データを検査することができます。探索では、分散の正規性と等分散性を検定し、グラフィカルに表示することができます。従属変数が独立変数のすべてのレベルで等分散していると思われる場合は、線型回帰プロシージャを使用できます。データがある仮定(正規性など)に違反していると思われる場合は、データの変換を試みます。データが線型に関連しておらず、変換も効果がない場合は、曲線推定プロシージャで代替モデルを使用します。従属変数が二分変数である場合(例えば、特定の販売が終了したか否か、品目に不具合があるか否かなど)は、ロジスティック

回帰プロシーチャーを使用します。従属変数が打ち切られている場合 (例えば、手術後の生存期間など) は、Custom Tables and Advanced Statistics で使用可能な、「生命表」、「Kaplan-Meier」、または「Cox 回帰」を使用します。データが独立していない場合 (例えば同じ人を複数の条件下で観測する場合) は、Custom Tables and Advanced Statistics で使用可能な「反復測定」プロシーチャーを使用します。

重み付け推定分析の実施

1. メニューから次の項目を選択します。
「分析」 > 「回帰」 > 「重み付け推定...」
2. 従属変数を 1 つ選択します。
3. 1 つ以上の独立変数を選択します。
4. 分散不均一性のソースとなる変数を重み付け変数として選択します。

Weight Variable (重み付け変数)

この変数の逆数のべき乗でデータに重みを付けます。指定された範囲の指数値について回帰式を計算し、対数尤度関数を最大化する指数を示します。

Power Range (べき乗の範囲)

これは、重みを計算するために重み付け変数とともに使用します。べき乗の範囲内の値ごとに 1 つの回帰式が適合します。「べき乗の範囲」検定ボックスおよび「から」テキスト・ボックスに入力する値は -6.5 から 7.5 まで (境界を含む) でなければなりません。べき乗の値の範囲はこの最低値から最高値までであり、指定した値によって増分が決定されます。べき乗の範囲内の値の総数は 150 までに制限されています。

重み付け推定のオプション

重み付け推定分析では、次のオプションを指定できます。

「新しい変数として最適な重み付けを保存」。重み付け変数をアクティブ・ファイルに追加します。この変数は `WGT_n` と呼ばれます。この `n` は、変数に固有の名前を付けるために選択された数字です。

「分散分析と推定値の表示」。統計の出力の表示方法を制御できます。選択可能なオプションは、「最適べき乗のみ表示」と「各べき乗を表示」です。

WLS 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。

- べき乗に単一値を指定する。
- べき乗値のリスト、または値の範囲とべき乗の値のリストの混合を指定する。

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

2 段階最小二乗回帰

標準の線型回帰モデルでは、従属変数の誤差と独立変数の間に相関がないと仮定します。この仮定が当てはまらない場合 (変数間の関係が双方向の場合など)、最小二乗法 (OLS) を使用する線型回帰では、最適なモデルを推定できなくなります。2 段階最小二乗回帰では、誤差項と相関していない操作変数を使用して、問題のある予測の推定値を計算し (第 1 段階)、これらの計算値を使用して従属変数の線型回帰モデルを推定します (第 2 段階)。計算値は、誤差と相関がない変数に基づくため、2 段階モデルの結果は最適となります。

例: ある商品の需要が、その価格および消費者の収入に関連しているかを調べます。このモデルの難しさは、価格と需要が互いに互恵的な効果をもたらす点です。つまり、価格が需要に影響し、需要も価格に影響する可能性があります。2 段階最小二乗回帰モデルでは、消費者の収入およびラグのある価格を使用して、需要の測定誤差との相関がない代理の価格を計算します。この代理価格は、最初に指定されたモデルでの価格と置き換えられた後に推定されます。

統計: 各モデルについては、標準化回帰係数と非標準化回帰係数、多重 R 、 R^2 乗、調整済み R^2 乗、推定値の標準誤差、分散分析テーブル、予測値、および残差。また、各回帰係数の 95% 信頼区間、パラメーター推定値の相関および共分散行列。

2 段階最小二乗回帰データの考慮事項

データ: 従属変数および独立変数は、量的でなければなりません。宗教、専攻、居住地区などのカテゴリー変数は、2 値 (ダミー) 変数またはその他のタイプの対比変数へと再コード化する必要があります。内生説明変数は、カテゴリーではなく量的にする必要があります。

仮定: 独立変数の各値に対して、従属変数の分布が正規分布でなければなりません。従属変数の分布の分散は、独立変数のすべての値に対して一定でなければなりません。従属変数と各独立変数の関係は、線型でなければなりません。

関連プロシージャー: 従属変数の誤差と相関している予測変数がないと思われる場合は、線型回帰プロシージャーを使用できます。データがいずれかの仮定 (正規性または定数分散など) に違反していると思われる場合は、そのデータを変換します。データが線型に関連しておらず、変換も効果がない場合は、曲線推定プロシージャーで代替モデルを使用します。従属変数が二値である場合 (特定の販売が終了したか否かなど) は、ロジスティック回帰プロシージャーを使用します。データが独立していない場合 (例えば同じ人をさまざまな条件下で観測する場合) は、反復測定プロシージャーを使用します。

2 段階最小二乗回帰分析の実施

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「2 段階最小二乗...」

2. 従属変数を 1 つ選択します。

3. 説明 (予測) 変数を 1 つ以上選択します。

4. 操作変数を 1 つ以上選択します。

- *Instrumental* (操作変数). 2 段階最小二乗分析の最初の段階で内生変数の予測値を計算するために使用する変数です。同じ変数が「説明変数」と「操作変数」の両方のリスト・ボックスに表示される場合があります。操作変数の数は、説明変数の数以上でなければなりません。リストされている説明変数と操作変数がすべて同じ場合の結果は、線型回帰プロシージャーでの結果と同じです。

操作変数として指定されていない説明変数は、内生変数と見なされます。通常、「説明変数」リスト内のすべての外生変数は、操作変数としても指定されます。

2 段階最小二乗回帰のオプション

分析には、次のオプションを選択できます。

「新規変数の保存」。新規の変数をアクティブ・ファイルに保存することができます。選択可能なオプションは、「予測値」と「残差」です。

「パラメータの分散共分散行列を表示」。パラメーター推定値の共分散行列を出力することができます。

2SLS 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、同時に複数の式を推定することもできます。シンタックスについて詳しくは、「コマンドシンタックスのリファレンス」を参照してください。

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。


```
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て、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数です。例えば、カテゴリが 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、プロフィール分析や、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、相互の線形結合であってはなりません。線形結合であると、プロシージャから線形従属が報告され、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。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。この資料の他の言語版を IBM から入手できる場合があります。ただし、これを入手するには、本製品または当該言語版製品を所有している必要がある場合があります。

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

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

〒 103-8510

東京都中央区日本橋箱崎町 19 番 21 号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス 渉外

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

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

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

IBM は、お客様が提供するいかなる情報も、お客様に対してな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 Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119

Armonk, NY 10504-1785

US

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

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

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

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

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。これらの名前はすべて架空のものであり、名前や住所が類似する個人や企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾:

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

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

© Copyright IBM Corp. 2020. 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. 1989 - 2020. 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]

逸脱関数

散らばりの尺度値の推定用 [8](#)

重み付け推定

新しい変数として最適な重み付けを保存 [21](#)

コマンドの追加機能 [21](#)

対数尤度 [20](#)

統計 [20](#)

反復履歴 [21](#)

分散分析と推定値の表示 [21](#)

例 [20](#)

[カ行]

カスタム・モデル

多項ロジスティック回帰 [6](#)

カテゴリ共変量 [3](#)

空セル

多項ロジスティック回帰 [7](#)

共分散行列

多項ロジスティック回帰 [7](#)

共変量

ロジスティック回帰 [3](#)

[サ行]

参照カテゴリ

多項ロジスティック回帰 [7](#)

収獲密度モデル

非線型回帰 [17](#)

収束基準

多項ロジスティック回帰 [7](#)

主効果モデル

多項ロジスティック回帰 [6](#)

信頼区間

多項ロジスティック回帰 [7](#)

ステップワイズ選択

多項ロジスティック回帰 [6](#)

ロジスティック回帰 [3](#)

すべての因子によるモデル

多項ロジスティック回帰 [6](#)

制約条件付き回帰

非線型回帰 [18](#)

切片

包含または除外 [6](#)

セル確率テーブル

多項ロジスティック回帰 [7](#)

線型回帰

重み付け推定 [20](#)

2段階最小二乗回帰 [21](#)

漸近回帰

非線型回帰 [17](#)

相関行列

相関行列 (続き)

多項ロジスティック回帰 [7](#)

相対メデアン潜在性

プロビット分析 [10](#)

[タ行]

対数修正モデル

非線型回帰 [17](#)

対数尤度

重み付け推定 [20](#)

多項ロジスティック回帰 [7](#)

対比

ロジスティック回帰 [3](#)

多項ロジスティック回帰

基準 [7](#)

コマンドの追加機能 [9](#)

参照カテゴリ [7](#)

統計 [7](#)

保存 [8](#)

モデル [6](#)

モデル情報のエクスポート [8](#)

段階 2 分

多項ロジスティック回帰 [7](#)

散らばりの尺度値

多項ロジスティック回帰 [8](#)

定数項

線型回帰 [4](#)

適合度

多項ロジスティック回帰 [7](#)

てこ比の値

ロジスティック回帰 [4](#)

デルタ

空セルの修正 [7](#)

特異性

多項ロジスティック回帰 [7](#)

[ハ行]

パラメーター推定値

多項ロジスティック回帰 [7](#)

パラメーター制約

非線型回帰 [18](#)

反復

多項ロジスティック回帰 [7](#)

プロビット分析 [10](#)

ロジスティック回帰 [4](#)

反復履歴

多項ロジスティック回帰 [7](#)

非線型回帰

一般的な非線形モデル [17](#)

結果の解釈 [19](#)

コマンドの追加機能 [20](#)

残差 [19](#)

条件論理 [17](#)

初期値 [17](#)

新規変数の保存 [19](#)

非線型回帰 (続き)

- 推定方法 [19](#)
- 損失関数 [18](#)
- 逐次2次計画法 [19](#)
- 統計 [16](#)
- パラメーター [17](#)
- パラメーター制約 [18](#)
- 微分係数 [19](#)
- ブートストラップ推定値 [19](#)
- 分割モデル [17](#)
- 予測値 [19](#)
- 例 [16](#)
- Levenberg-Marquardt アルゴリズム [19](#)

非線形モデル

- 非線型回帰 [17](#)
- フィデューシアル信頼区間
 - プロビット分析 [10](#)
- プロビットの回帰
 - 統計 [9](#)
 - 例 [9](#)
- プロビット分析
 - 基準 [10](#)
 - コマンドの追加機能 [10](#)
 - 自然応答比 [10](#)
 - 相対メディアン 潜在性 [10](#)
 - 統計 [10](#)
 - 範囲の定義 [10](#)
 - 反復 [10](#)
 - フィデューシアル信頼区間 [10](#)
 - 平行性の検定 [10](#)

分位の回帰

- 基準 [12](#)
- 統計 [11](#)
- 表示 [14](#)
- 保存 [15](#)
- モデル [13](#)
- 例 [11](#)
- export [15](#)

分離

- 多項ロジスティック回帰 [7](#)

分類

- 多項ロジスティック回帰 [5](#)

分類テーブル

- 多項ロジスティック回帰 [7](#)

平行性の検定

- プロビット分析 [10](#)

変数減少法

- ロジスティック回帰 [3](#)

変数増加法

- ロジスティック回帰 [3](#)

[マ行]

密度モデル

- 非線型回帰 [17](#)

文字列共変量

- ロジスティック回帰 [3](#)

[ヤ行]

尤度比

- 散らばりの尺度値の推定用 [8](#)
- 適合度 [7](#)

[ラ行]

ロジスティック回帰

- 影響の測定 [4](#)
- カテゴリー共変量 [3](#)
- 規則の設定 [3](#)
- 係数 [2](#)
- コマンドの追加機能 [5](#)
- 残差 [4](#)
- 新変数の保存 [4](#)
- ステップワイズにおける確率 [4](#)
- 選択規則の定義 [3](#)
- 対比 [3](#)
- 定数項 [4](#)
- 統計 [2](#)
- 統計と作図 [4](#)
- 反復 [4](#)
- 表示オプション [4](#)
- 分類の打ち切り [4](#)
- 変数選択方法 [3](#)
- 文字列共変量 [3](#)
- 予測値 [4](#)
- 例 [2](#)
- 2項 [1](#)
- Hosmer-Lemeshow の適合度統計 [4](#)
- ロジスティック回帰分析 [2](#)

[数字]

- 2項ロジスティック回帰 [1, 2](#)

2段階最小二乗回帰

- コマンドの追加機能 [22](#)
- 新変数の保存 [22](#)
- 操作変数 [21](#)
- 統計 [21](#)
- パラメーターの共分散 [22](#)
- 例 [21](#)

2次の有理式モデル

- 非線型回帰 [17](#)

3次の有理式モデル

- 非線型回帰 [17](#)

C

Cook の距離

- ロジスティック回帰 [4](#)

Cox と Snell の R² 乗

- 多項ロジスティック回帰 [7](#)

D

DfBeta

- ロジスティック回帰 [4](#)

G

Gauss モデル

- 非線型回帰 [17](#)

Gompertz モデル

- 非線型回帰 [17](#)

H

Hosmer-Lemeshow の適合度統計
ロジスティック回帰 [4](#)

J

Johnson-Schumacher モデル
非線型回帰 [17](#)

M

McFadden の R² 乗
多項ロジスティック回帰 [7](#)

Metcherlich の収穫逓減の法則
非線型回帰 [17](#)

Michaelis Menten モデル
非線型回帰 [17](#)

Morgan-Mercer-Florin モデル
非線型回帰 [17](#)

N

Nagelkerke の R² 乗
多項ロジスティック回帰 [7](#)

P

Peal-Reed モデル
非線型回帰 [17](#)

Pearson カイ 2 乗
散らばりの尺度値の推定用 [8](#)
適合度 [7](#)

R

Richards モデル
非線型回帰 [17](#)

S

SPSS ロジスティック回帰 [2](#)

V

Verhulst モデル
非線型回帰 [17](#)

Von Bertalanffy モデル
非線型回帰 [17](#)

W

Weibull モデル
非線型回帰 [17](#)

