

IBM SPSS Complex Samples 24



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、59ページの『特記事項』に記載されている情報をお読みください。

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

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

原典： IBM SPSS Complex Samples 24

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

第 1 章 コンプレックス・サンプルのプロシージャーの概要	1
コンプレックス・サンプルのプロパティ	1
コンプレックス・サンプルのプロシージャーの使用法	2
計画ファイル	2
参考文献	3
第 2 章 コンプレックス・デザインからのサンプリング	5
新規サンプル計画の作成	5
サンプル・ウィザード: デザイン変数	5
サンプル・ウィザードのナビゲートのためのツリー・コントロール	6
サンプル・ウィザード: 抽出方法	6
サンプル・ウィザード: サンプル・サイズ	7
等しくないサイズの定義	8
サンプル・ウィザード: 出力変数	8
サンプル・ウィザード: 計画の要約	9
サンプル・ウィザード: サンプルの抽出: 選択オプション	9
サンプル・ウィザード: サンプルの抽出: 出力ファイル	9
サンプル・ウィザード: 終了	10
既存のサンプル計画の変更	10
サンプル・ウィザード: 計画の要約	10
既存のサンプル計画の実行	11
CSPLAN および CSSELECT コマンドの追加機能	11
第 3 章 分析のためのコンプレックス・サンプルの準備	13
分析計画の新規作成	13
分析準備ウィザード: デザイン変数	13
分析ウィザードのナビゲートのためのツリー・コントロール	14
分析準備ウィザード: 推定方法	14
分析準備ウィザード: サイズ	14
等しくないサイズの定義	15
分析準備ウィザード: 計画の要約	15
分析準備ウィザード: 終了	15
既存の分析計画の変更	16
分析準備ウィザード: 計画の要約	16
第 4 章 コンプレックス・サンプルの計画 17	
第 5 章 コンプレックス・サンプルの度数分布表	19
コンプレックス・サンプルの度数分布表の統計量	19
コンプレックス・サンプルの欠損値	20
コンプレックス・サンプルのオプション	20

第 6 章 コンプレックス・サンプルの記述統計	23
コンプレックス・サンプルの記述統計の統計量	23
コンプレックス・サンプルの記述統計の欠損値	24
コンプレックス・サンプルのオプション	24
第 7 章 コンプレックス・サンプルのクロス集計表	25
コンプレックス・サンプルのクロス集計表の統計量	25
コンプレックス・サンプルの欠損値	26
コンプレックス・サンプルのオプション	27
第 8 章 コンプレックス・サンプルの比率分析	29
コンプレックス・サンプルの比率分析の統計量	29
コンプレックス・サンプルの比率分析の欠損値	30
コンプレックス・サンプルのオプション	30
第 9 章 コンプレックス・サンプルの一般線型モデル	31
コンプレックス・サンプルの一般線型モデル	31
コンプレックス・サンプルの一般線型モデルの統計量	32
コンプレックス・サンプルの仮説の検定	33
コンプレックス・サンプルの一般線型モデルの推定平均	33
コンプレックス・サンプルの一般線型モデルの保存	34
コンプレックス・サンプルの一般線型モデルのオプション	34
CSGLM コマンドの追加機能	34
第 10 章 コンプレックス・サンプルのロジスティック回帰	37
コンプレックス・サンプルのロジスティック回帰の参照カテゴリー	37
コンプレックス・サンプルのロジスティック回帰モデル	38
コンプレックス・サンプルのロジスティック回帰の統計量	38
コンプレックス・サンプルの仮説の検定	39
コンプレックス・サンプルのロジスティック回帰のオッズ比	40
コンプレックス・サンプルのロジスティック回帰の保存	40
コンプレックス・サンプルのロジスティック回帰のオプション	40
CSLOGISTIC コマンドの追加機能	41

第 11 章 コンプレックス・サンプルの順序回帰 43

コンプレックス・サンプルの順序回帰の応答確率 44
コンプレックス・サンプルの順序回帰のモデル 44
コンプレックス・サンプルの順序回帰の統計量 44
コンプレックス・サンプルの仮説の検定 45
コンプレックス・サンプルの順序回帰のオッズ比 46
コンプレックス・サンプルの順序回帰の保存 46
コンプレックス・サンプルの順序回帰のオプション 47
CSORDINAL コマンドの追加機能 47

第 12 章 コンプレックス・サンプルの Cox 回帰 49

イベントの定義 50
予測 50

時間依存の予測変数の定義 51
サブグループ 51
モデル 51
統計量 52
プロット 53
仮説の検定 53
保存 54
エクスポート 55
オプション 56
CSCOXREG コマンドの追加機能 57

特記事項 59
商標 60

索引 63

第 1 章 コンプレックス・サンプルのプロシーチャーの概要

従来のソフトウェア・パッケージの分析プロシーチャーに特有の前提は、データ・ファイルの観測値は対象母集団の単純なランダム・サンプルを表すということでした。より体系的な方法でサンプルを得る方がコスト効果が高く、かつ簡便であると考えられる企業や研究者が増えるにつれ、この前提は成り立たなくなっています。

「コンプレックス・サンプル」オプションを使用すると、コンプレックス・デザインに従ってサンプルを選択し、データ分析にデザイン仕様を取り入れることが可能です。これにより、有効な結果を確実に得ることができます。

コンプレックス・サンプルのプロパティ

コンプレックス・サンプルは、多くの点で単純な無作為サンプルとは異なっています。単純な無作為サンプルでは、個々のサンプリング単位が母集団全体からランダムに等確率で、置換を行わずに (WOR) 直接選択されます。これに対して、コンプレックス・サンプルは、次の特徴のうちの一部またはすべてを備えています。

層化。 層化サンプリングでは、母集団の中の重なり合わない各サブグループ、つまり階層内で、独立してサンプルが選択されます。階層の例として、社会経済グループ、職種カテゴリー、年齢グループ、民族グループなどがあります。層化することによって、対象となるサブグループに適したサンプル・サイズを確保し、全体の推定値の精度を向上させ、階層間で異なるサンプリング方法を使用することができます。

クラスター化。 クラスター・サンプリングでは、サンプリング単位のグループ、つまりクラスターを選択します。例えば、クラスターを学校、病院、地域とすると、サンプリング単位は学生、患者、市民になります。クラスター化は、複数の段階があるデザインや地域のサンプルで一般的に使用されます。

複数の段階。 複数の段階があるサンプリングでは、第 1 段階のサンプルをクラスターを基に選択します。次に、選択したクラスターからサブサンプルを抽出して、第 2 段階のサンプルを作成します。第 2 段階のサンプルがサブクラスターに基づいている場合、サンプルに第 3 段階を追加できます。例えば、調査の第 1 段階で都市のサンプルを抽出するとします。次に、選択した都市から世帯を抽出します。最後に、選択した世帯の個人に対して調査を行います。「サンプリング」ウィザードおよび「分析準備」ウィザードを使用すると、1 つのデザインに 3 つの段階を指定できます。

無作為でないサンプリング。 無作為な選択が難しい場合に、系統的に (固定間隔で) または順序に従って、単位のサンプリングを行うことができます。

不等選択確率。 含まれている単位の数が等しくないクラスターのサンプリングを行うときは、確立比例 (PPS) サンプリング方法を使用して、クラスターの選択確率が、そのクラスターに含まれる単位の比率に等しくなるようにすることができます。PPS サンプリングでは、より一般的な重み付け方式を使用して単位を選択することもできます。

制限されない抽出。 制限されない抽出では、単位の選択時に置換 (WR) が行われます。このため、個々の単位が複数回選択されて抽出されることがあります。

抽出の重み付け。 抽出の重み付けは、コンプレックス・サンプルの抽出時に自動的に計算されます。理論上は、各抽出単位が対象母集団に現れる「頻度」に応じた値になります。したがって、サンプル全体の重み

付けの合計から、母集団のサイズを推定できます。コンプレックス・サンプルの分析プロシージャーでは、コンプレックス・サンプルを適切に分析するために抽出の重み付けが必要になります。この重み付けは、コンプレックス・サンプル・オプションの内部だけで使用する必要があります。重み付けをケースの反復として扱う「ケースの重み付け」プロシージャーを介して、他の分析プロシージャーで使用しないようにしてください。

コンプレックス・サンプルのプロシージャーの使用法

コンプレックス・サンプルのプロシージャーの使用法は、ユーザーのニーズによって異なります。主に次のようなユーザーが使用します。

- コンプレックス・デザインに従って調査を計画して実行し、後でサンプルを分析する可能性があるユーザー。調査者の主要なツールは、サンプル・ウィザードです。
- 以前に取得したサンプルのデータ・ファイルをコンプレックス・デザインに従って分析するユーザー。コンプレックス・サンプルの分析プロシージャーを使用する前に、分析準備ウィザードを使用する必要があります。

どちらのタイプのユーザーも、コンプレックス・サンプルのプロシージャーにデザイン情報を指定する必要があります。この情報は、再利用しやすいように、**計画ファイル**に保存されます。

計画ファイル

計画ファイルには、コンプレックス・サンプルの仕様が格納されます。計画ファイルには次の 2 つのタイプがあります。

抽出計画。 サンプル・ウィザードで指定する内容は、コンプレックス・サンプルを抽出する際に使用されるサンプル計画を定義します。抽出計画ファイルには、この指定が格納されます。また、抽出計画ファイルには、指定されたサンプル計画に適した推定方法を使用するデフォルトの分析計画も含まれています。

分析計画。 この計画ファイルには、コンプレックス・サンプルの分散の推定値を正しく計算するために、コンプレックス・サンプル分析プロシージャーに必要な情報が格納されます。この計画には、サンプルの構造、各段階での推定方法、および必要な変数への参照 (サンプルの重み付けなど) が含まれます。分析準備ウィザードで、分析計画の作成と編集を行うことができます。

計画ファイルに指定内容を保存しておく、次のようないくつかの利点があります。

- 調査者が、複数の段階がある抽出計画の第 1 段階を指定して、すぐに第 1 段階の単位を抽出した後、第 2 段階用にサンプリング単位に関する情報を収集できます。その後、第 2 段階を含めるように、抽出計画を修正できます。
- 抽出計画ファイルにアクセスできないアナリストが、分析計画を指定することによって、各コンプレックス・サンプル分析プロシージャーからその計画を参照できます。
- 大規模な一般利用のためのサンプルを計画する設計者が、抽出計画ファイルを公開できます。これにより、アナリストに対する指示が簡単になり、各アナリストが自分で分析計画を指定する必要がなくなります。

参考文献

サンプリング手法について詳しくは、次の資料を参照してください。

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

第 2 章 コンプレックス・デザインからのサンプリング

サンプル・ウィザードでは、サンプリング計画ファイルの作成、変更、実行に必要な各ステップを順に実行していきます。このウィザードを使用する前に、対象の母集団を明確に定義しておき、サンプリング単位の一覧を用意し、適切なサンプル計画を検討しておく必要があります。

新規サンプル計画の作成

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「サンプルの選択...」

2. 「サンプルを計画する」を選択して、サンプル計画を保存する計画ファイルの名前を選択します。

3. 「次へ」をクリックして、ウィザードを進めます。

4. 必要に応じて、「デザイン変数」ステップで、階層、クラスター、および入力サンプルの重み付けを定義できます。これらの定義を行った後、「次へ」をクリックします。

5. 必要に応じて、「サンプリング方法」ステップで、項目の選択方法を選択できます。

「Brewer (PPS)」または「Murthy (PPS)」を選択した場合は、「終了」をクリックして、サンプルを抽出できます。それ以外の場合は、「次へ」をクリックした後、次の操作を行います。

6. 「サンプル・サイズ」ステップで、サンプリングする単位の数または比率を指定します。

7. これで、「終了」をクリックしてサンプルを抽出できます。

オプションとして、さらに以下のステップを実行できます。

- 保存する出力変数を選択する。
- デザインに第 2 段階または第 3 段階を追加する。
- どの段階でサンプルを抽出するか、乱数シード、およびユーザー欠損値をデザイン変数の有効な値として扱うかどうかなど、さまざまな選択オプションを設定する。
- 出力データの保存先を選択する。
- 選択内容をコマンド・シンタックスとして貼り付ける。

サンプル・ウィザード: デザイン変数

このステップでは、層化とクラスター化の変数を選択でき、サンプルの重み付けの入力を定義できます。また、段階のラベルを指定することもできます。

以下で階層化。 層化変数の交差分類により、別個の部分母集団、つまりストラータが定義されます。階層ごとに別個のサンプルが取得されます。推定値の精度を高めるためには、ストラータ内の単位を対象の特性に対して可能な限り等質にする必要があります。

クラスター。 クラスター変数は、観測単位のグループ、つまりクラスターを定義します。母集団から観測単位を直接サンプリングすることが不可能であったり、費用がかかる場合に、クラスターは有効です。直接サンプリングする代わりに、母集団からクラスターをサンプリングした後に、選択したクラスターから観測単位をサンプリングすることができます。ただし、クラスターを使用すると、サンプリング単位間に相関が発生する場合があります、これによって精度が失われます。この影響を最小限に抑えるためには、クラスター内

の単位を対象の特性に対して可能な限り異質にする必要があります。複数の段階があるデザインを計画するには、1 つ以上のクラスター変数を定義する必要があります。クラスターは、異なる複数のサンプリング方法を使用する場合にも必要です。詳しくは、『サンプル・ウィザード: 抽出方法』のトピックを参照してください。

サンプルの重み付けの入力。 現在のサンプル計画が、より大きなサンプル計画の一部である場合、その大きなサンプル計画の前の段階でサンプルの重み付けを設定していることがあります。現在の計画の第 1 段階で、このような重み付けを含む数値変数を指定することができます。現在の計画のそれ以降の段階では、サンプルの重み付けは自動的に計算されます。

段階ラベル。 必要に応じて、各段階の文字列ラベルを指定できます。このラベルは出力で使用され、段階ごとの情報を識別するのに役立ちます。

注: ソース変数リストには、ウィザードのすべてのステップで同じ内容が表示されます。つまり、特定のステップでソース・リストから変数を削除すると、すべてのステップのリストからその変数が削除されます。変数をソース・リストに戻すと、すべてのステップのリストにその変数が表示されます。

サンプル・ウィザードのナビゲートのためのツリー・コントロール

サンプル・ウィザードの各ステップの左側には、全ステップのアウトラインが表示されます。アウトライン内で有効になっているステップの名前をクリックして、ウィザードをナビゲートできます。ステップは、その前にあるすべてのステップが有効である場合、つまり前にあるステップでそのステップに最低限必要な指定が行われていれば、有効になります。特定のステップが無効である理由については、個々のステップのヘルプを参照してください。

サンプル・ウィザード: 抽出方法

このステップでは、アクティブなデータ・セットからケースを選択する方法を指定できます。

方法。 このグループのコントロールは、選択方法を選択するために使用されます。一部の抽出タイプでは、抽出時に置換を行う (WR) か、置換を行わない (WOR) かを選択できます。詳しくは、各タイプの説明を参照してください。いくつかの確立比例 (PPS) タイプは、クラスターが定義されている場合にのみ使用できることに注意してください。また、どの PPS タイプも計画の第 1 段階でのみ使用できます。さらに、WR 方法は計画の最後の段階でのみ使用できます。

- **単純な無作為抽出。** 等しい確率で単位が選択されます。置換あり、置換なしのどちらでも、単位を選択できます。
- **単純な系統的抽出。** サンプリング枠 (階層が指定されている場合は、階層) 全体にわたって、固定間隔で単位が選択され、置換なしに抽出されます。最初の間隔内で無作為に選択された単位が、開始点として選択されます。
- **単純な連続値。** 単位は等しい確率で順番に選択され、置換は行われません。
- **PPS。** これは、サイズに比例した確率でランダムに単位を選択する第 1 段階の方法です。任意の単位の、置換を行う選択が可能です。クラスターのみで、置換を行わない抽出が可能です。
- **系統的 (PPS)。** これは、サイズに比例した確率で系統的に単位を選択する第 1 段階の方法です。単位は置換なしに選択されます。
- **連続的 (PPS)。** これは、クラスター・サイズに比例した確率で置換なしに、単位を順番に選択する第 1 段階の方法です。

- **Brewer (PPS)**。これは、クラスター・サイズに比例した確率で置換なしに、各階層から 2 つのクラスターを選択する第 1 段階の方法です。この方法を使用するには、クラスター変数を指定する必要があります。
- **Murthy (PPS)**。これは、クラスター・サイズに比例した確率で置換なしに、各階層から 2 つのクラスターを選択する第 1 段階の方法です。この方法を使用するには、クラスター変数を指定する必要があります。
- **Sampford (PPS)**。これは、クラスター・サイズに比例した確率で置換なしに、各階層から 3 つ以上のクラスターを選択する第 1 段階の方法です。これは Brewer 法の拡張です。この方法を使用するには、クラスター変数を指定する必要があります。
- **分析に WR 推定を使用する**。デフォルトでは、推定方法は選択したサンプリング方法と整合する計画ファイルに指定されます。このため、抽出方法に WOR 推定が暗黙指定されている場合でも、置換を行う推定を使用できます。このオプションは段階 1 でのみ使用できます。

サイズの測定 (MOS)。PPS の方法を選択する場合は、各単位のサイズを定義するサイズの測定方法を指定する必要があります。このようなサイズは、変数で明示的に定義することも、データから計算することもできます。オプションとして、MOS に上限と下限を設定して、MOS 変数内で検出された値、またはデータから計算された値をすべて上書きすることができます。これらのオプションは、段階 1 でのみ使用できます。

サンプル・ウィザード: サンプル・サイズ

このステップでは、現在の段階でサンプリングする単位の数または単位の比率を指定できます。サンプル・サイズは固定することも、階層ごとに変更することもできます。サンプル・サイズを指定するために、前の段階で選択されたクラスターを使用して階層を定義できます。

単位。サンプリングする単位の正確なサンプル・サイズまたは比率を指定できます。

- **値**。1 つの値がすべてのストラータに適用されます。単位の測定基準として「**度数**」を選択した場合は、正整数を入力する必要があります。「**比率**」を選択した場合は、負でない値を入力する必要があります。また、置換を行うサンプリングでない場合は、比率の値を 1 以下の値にしてください。
- **ストラータに不等な値**。「等しくないサイズの定義」ダイアログ・ボックスを使用して、ストラータごとにサイズ値を入力することができます。
- **変数から値を読み取る**。ストラータのサイズ値を含む数値変数を選択できます。

「**比率**」を選択した場合は、サンプリングする単位の数に上限と下限を設定することもできます。

等しくないサイズの定義

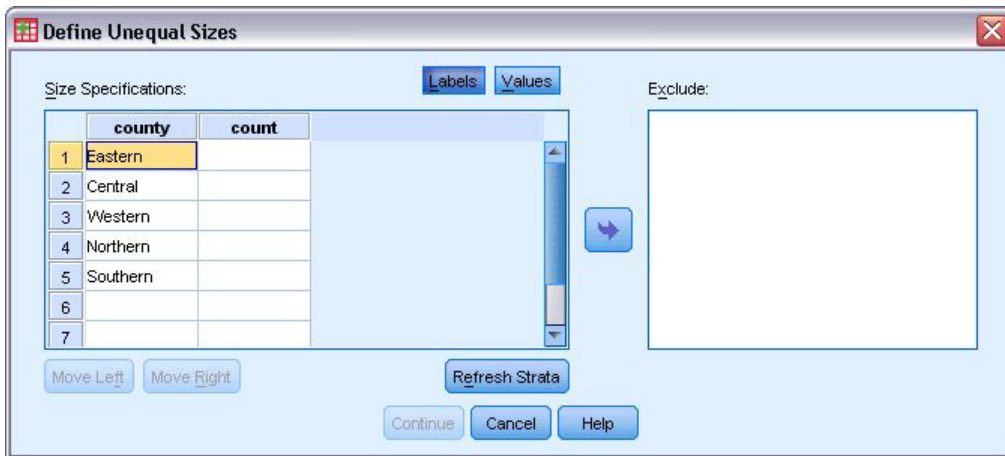


図1. 「等しくないサイズの定義」ダイアログ・ボックス

「等しくないサイズの定義」ダイアログ・ボックスでは、階層ごとにサイズを入力できます。

「サイズ指定」グリッド。 グリッドには、最大 5 個までのストラータ変数またはクラスター変数の交差分類が表示されます。1 行に 1 つの階層とクラスターの組み合わせが表示されます。グリッドの適格な変数には、現在の段階と前の段階にあるすべての層化変数と、前の段階にあるすべてのクラスター変数が含まれます。グリッド内部で変数を並べ替えたり、「除外」リストに移動したりすることができます。右端の列にサイズを入力します。「ラベル」または「値」をクリックすると、グリッド・セル内の層化変数とクラスター変数の表示が値ラベルとデータ値で切り替わります。ラベルのない値を含むセルには、常に値が表示されます。「ストラータを再表示」をクリックすると、グリッド内の変数のラベル付きデータ値の各組み合わせがグリッドに再入力されます。

除外。 階層とクラスターの組み合わせのサブセットにサイズを指定するには、1 つ以上の変数を「除外」リストに移動します。これらの変数はサンプル・サイズの定義には使用されません。

サンプル・ウィザード: 出力変数

このステップでは、サンプルを抽出するときに保存する変数を選択できます。

母集団のサイズ。 特定の段階における母集団の推定単位数。保存される変数のルート名は、「PopulationSize_」になります。

サンプル率。 特定の段階におけるサンプリング比率。保存される変数のルート名は、「SamplingRate_」になります。

サンプル・サイズ。 特定の段階で抽出される単位数。保存される変数のルート名は、「SampleSize_」になります。

サンプルの重み付け。 包含確率の逆数。保存される変数のルート名は、「SampleWeight_」になります。

段階に関する変数の中には、自動的に生成されるものもあります。このような変数には次のものがあります。

包含確率。 特定の段階で抽出される単位の比率。保存される変数のルート名は、「*InclusionProbability_*」になります。

累積重み付け。 それ以前の段階から現在の段階までの累積のサンプルの重み付け。保存される変数のルート名は、「*SampleWeightCumulative_*」になります。

インデックス。 特定の段階において複数回選択される単位を示します。保存される変数のルート名は、「*Index_*」になります。

注: 保存される変数のルート名には、段階の番号を表す整数の接尾辞が含まれます。例えば、段階 1 で保存される母集団のサイズの場合は、*PopulationSize_1_* になります。

サンプル・ウィザード: 計画の要約

これは各段階の最後のステップです。現在の段階におけるサンプル計画の指定の要約が表示されます。ここから、次の段階 (必要な場合は作成します) に進むか、サンプルを抽出するためのオプションを設定することができます。

サンプル・ウィザード: サンプルの抽出: 選択オプション

このステップでは、サンプルを抽出するかどうかを選択できます。ランダム・シードや欠損値の処理などの、他のサンプリング・オプションを制御することもできます。

サンプルの抽出。 サンプルを抽出するかどうかの選択に加えて、サンプリング計画の一部を実行するように選択することもできます。段階の抽出は順番に行う必要があります。つまり、段階 1 の抽出が行われていない場合は、段階 2 でも抽出できません。計画の編集または実行時には、ロックされた段階を再サンプリングすることはできません。

シード。 ここでは、乱数を生成するためのシード値を選択できます。

ユーザー欠損値を含める。 ここでは、ユーザー欠損値が有効かどうかを指定します。有効にした場合、ユーザー欠損値は別のカテゴリーとして処理されます。

データはソート済み。 サンプル・フレームが層化変数の値であらかじめソートされている場合、このオプションを指定すると、選択プロセスを高速化できます。

サンプル・ウィザード: サンプルの抽出: 出力ファイル

このステップでは、サンプリングしたケース、重み付け変数、結合確率、ケース選択規則の出力先を選択できます。

サンプル・データ。 ここにあるオプションによって、サンプルの出力の書き込み先を決定できます。アクティブなデータ・セットに追加するか、新しいデータ・セットに書き込むか、または外部 IBM® SPSS® Statistics データ・ファイルに保存することができます。データ・セットは現在のセッション中は使用可能ですが、データ・セットを明示的にデータ・ファイルとして保存しない限り、以降のセッションでは使用できません。データ・セット名は、変数の命名規則に従っている必要があります。外部ファイルまたは新しいデータ・セットを指定した場合、選択したケースのサンプリングの出力変数とアクティブなデータ・セット内の変数が書き込まれます。

結合確率。 ここにあるオプションによって、結合確率の書き込み先を決定できます。結合確率は、外部 IBM SPSS Statistics データ・ファイルに保存されます。「WOR (PPS)」、「Brewer (PPS)」、「Sampford (PPS)」、「Murthy (PPS)」のいずれかの方法を選択して、WR 推定を指定しない場合に、結合確率が生成されます。

ケース選択規則。 一度に 1 段階ずつサンプルを構成する場合は、ケース選択規則をテキスト・ファイルに保存できます。これは後の段階でサブフレームを構成するときに役立ちます。

サンプル・ウィザード: 終了

これは最後のステップです。計画ファイルを保存してすぐにサンプルを抽出するか、選択内容をシンタックス・ウィンドウに貼り付けることができます。

既存の計画ファイルの段階に変更を加えた場合は、編集した計画を新しいファイルに保存するか、既存のファイルを上書きできます。既存の段階に変更を加えずに段階を追加した場合は、既存の計画ファイルが自動的に上書きされます。計画を新しいファイルに保存する場合は、「ウィザードで生成したシンタックスをシンタックス ウィンドウに貼り付ける」を選択して、シンタックス・コマンドでファイル名を変更します。

既存のサンプル計画の変更

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「サンプルの選択...」

2. 「サンプル デザインを編集する」を選択してから、編集する計画ファイルを選択します。
3. 「次へ」をクリックして、ウィザードを進めます。
4. 「計画の要約」ステップでサンプリング計画を確認した後、「次へ」をクリックします。

以降のステップは、新規計画の場合とほとんど同じです。詳しくは、個々のステップのヘルプを参照してください。

5. 「終了」ステップにナビゲートして、編集した計画ファイルに新しい名前を指定するか、既存の計画ファイルを上書きするように選択します。

オプションとして、以下の操作を実行できます。

- 既にサンプリングが行われた段階を指定する。
- 計画から段階を削除する。

サンプル・ウィザード: 計画の要約

このステップでは、サンプリング計画を確認し、既にサンプリングが行われた段階を表示することができます。計画を編集している場合は、計画から段階を削除することもできます。

以前に抽出された段階。 拡張サンプリング・フレームが使用可能でない場合は、複数の段階があるサンプリング計画を一度に 1 段階ずつ実行する必要があります。ドロップダウン・リストから、既にサンプリングが行われた段階を選択します。実行済みの段階はすべてロックされ、「サンプルの抽出: 選択オプション」ステップでは使用できません。また、計画の編集時にも変更できません。

段階の削除。 複数の段階があるデザインから、段階 2 および段階 3 を削除することができます。

既存のサンプル計画の実行

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「サンプルの選択...」

2. 「サンプルを抽出する」を選択してから、実行する計画ファイルを選択します。

3. 「次へ」をクリックして、ウィザードを進めます。

4. 「計画の要約」ステップでサンプリング計画を確認した後、「次へ」をクリックします。

5. 段階に関する情報を含む個々のステップは、サンプル計画の実行時にはスキップされます。これで、いつでも「終了」ステップに進むことができます。

オプションとして、既にサンプリングを行った段階を指定することもできます。

CSPLAN および CSSELECT 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。

- 出力変数にカスタム名を指定する。
- ビューアーの出力を制御する。例えば、サンプルの計画時または変更時に表示される計画の段階ごとの要約や、サンプル計画の実行時に表示されるサンプリングされたケースの階層ごとの分布の要約を表示しないようにしたり、ケース処理の要約を要求したりすることができます。
- アクティブなデータ・セット内の変数のサブセットを選択して、外部サンプル・ファイルや別のデータ・セットに書き込む。

シンタックスの詳細については、「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 3 章 分析のためのコンプレックス・サンプルの準備

分析準備ウィザードでは、コンプレックス・サンプルのさまざまな分析プロシージャで使用する分析計画の作成と変更に必要な各ステップを順に実行していきます。このウィザードを使用する前に、コンプレックス・デザインに従ってサンプルを抽出しておく必要があります。

サンプルの抽出に使用するサンプリング計画ファイル (サンプリング計画には、デフォルトの分析計画が含まれています) にアクセスできない場合は、新しい計画を作成することが最も有効な方法です。サンプルの抽出に使用するサンプリング計画ファイルにアクセスできる場合は、サンプリング計画ファイルに含まれているデフォルトの分析計画を使用するか、デフォルトの分析の指定を上書きして、変更内容を新しいファイルに保存することができます。

分析計画の新規作成

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「分析の準備...」

2. 「計画ファイルの作成」を選択してから、分析計画を保存する計画ファイルの名前を選択します。
3. 「次へ」をクリックして、ウィザードを進めます。
4. 「計画変数」ステップで、サンプルの重み付けを含む変数を指定します。必要に応じて、階層とクラスターを定義します。
5. これで、「終了」をクリックして計画を保存できます。

オプションとして、さらに以下のステップを実行できます。

- 「推定方法」ステップで、標準誤差の推定方法を選択する。
- 「サイズ」ステップで、抽出される単位の数、または単位ごとの包含確率を指定する。
- デザインに第 2 段階または第 3 段階を追加する。
- 選択内容をコマンド・シンタックスとして貼り付ける。

分析準備ウィザード: デザイン変数

このステップでは、層化とクラスター化の変数を指定し、サンプルの重み付けを定義することができます。また、段階にラベルを指定することもできます。

階層。 層化変数の交差分類により、別個の部分母集団、つまりストラータが定義されます。サンプル全体は、各ストラータからの独立したサンプルの組み合わせを表します。

クラスター。 クラスター変数は、観測単位のグループ、つまりクラスターを定義します。複数の段階で抽出されるサンプルでは、初期の段階でクラスターが選択された後、選択されたクラスターからサブサンプル単位が抽出されます。置換を行うクラスターのサンプリングによって取得されたデータ・ファイルを分析するときは、重複インデックスをクラスター変数として指定する必要があります。

サンプルの重み付け。 サンプルの重み付けは、第 1 段階で指定する必要があります。現在の計画のそれ以降の段階では、サンプルの重み付けは自動的に計算されます。

段階ラベル。 必要に応じて、各段階の文字列ラベルを指定できます。このラベルは出力で使用され、段階ごとの情報を識別するのに役立ちます。

注: ソース変数リストには、このウィザードのすべてのステップで同じ内容が表示されます。つまり、特定のステップでソース・リストから変数を削除すると、すべてのステップのリストからその変数が削除されます。変数をソース・リストに戻すと、すべてのステップでその変数が表示されます。

分析ウィザードのナビゲートのためのツリー・コントロール

分析ウィザードの各ステップの左側には、全ステップのアウトラインが表示されます。アウトライン内で有効になっているステップの名前をクリックして、ウィザードをナビゲートできます。ステップは、その前にあるすべてのステップが有効である場合、つまり前にあるステップでそのステップに最低限必要な指定が行われていれば、有効になります。特定のステップが無効である理由については、個々のステップのヘルプを参照してください。

分析準備ウィザード: 推定方法

このステップでは、段階に対する推定方法を指定できます。

WR (置換を行うサンプリング)。 コМПレックス (複雑な) サンプリング計画で分散を推定するときには、WR 推定には有限母集団からのサンプリングの訂正 (FPC) は含まれません。単純なランダム・サンプリング (SRS) で分散を推定するときには、FPC を含めるか除外するかを選択できます。

分析の重みがスケールされていて、重みを合計しても母集団のサイズに達しない場合は、SRS の分散推定に FPC を含めないようにすることをお勧めします。SRS の分散の推定値は、デザイン効果などの統計量の計算に使用されます。WR 推定は、デザインの最終段階でのみ指定できます。WR 推定を選択した場合は、ウィザードで段階をさらに追加することはできません。

等しい WOR (置換を行わない等確率サンプリング)。 等しい WOR 推定には、有限母集団の訂正が含まれ、単位が等しい確率でサンプリングされると仮定されます。等しい WOR は、デザインの任意の段階に指定できます。

不等な WOR (置換を行わない不等確率サンプリング)。 不等な WOR では、有限母集団の訂正を使用することに加えて、サンプリング単位 (通常はクラスター) が不等確率で選択されます。この推定方法は、第 1 段階でのみ使用できます。

分析準備ウィザード: サイズ

このステップは、現在の段階の包含確率または母集団のサイズを指定するために使用されます。サイズは固定することも、ストラータごとに変更することもできます。サイズを指定するために、前の段階で指定されたクラスターを使用してストラータを定義できます。「推定方法」として「等しい WOR」を選択した場合にのみ、このステップが必要です。

単位。 単位のサンプリングに使用した正確な母集団のサイズ、または確率を指定できます。

- **値。** 1 つの値がすべてのストラータに適用されます。単位の測定基準として「母集団のサイズ」を選択した場合は、負でない整数を入力する必要があります。「包含確率」を選択した場合は、0 から 1 までの値を入力する必要があります。
- **ストラータに不等な値。** 「等しくないサイズの定義」ダイアログ・ボックスを使用して、ストラータごとにサイズ値を入力することができます。
- **変数から値を読み取る。** ストラータのサイズ値を含む数値変数を選択できます。

等しくないサイズの定義

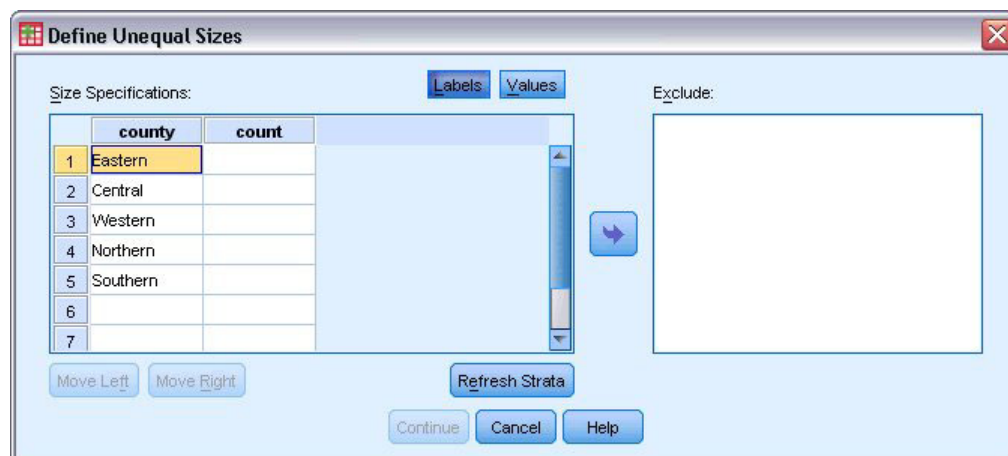


図2. 「等しくないサイズの定義」ダイアログ・ボックス

「等しくないサイズの定義」ダイアログ・ボックスでは、階層ごとにサイズを入力できます。

「サイズ指定」グリッド。 グリッドには、最大 5 個までのストラータ変数またはクラスター変数の交差分類が表示されます。1 行に 1 つの階層とクラスターの組み合わせが表示されます。グリッドの適格な変数には、現在の段階と前の段階にあるすべての層化変数と、前の段階にあるすべてのクラスター変数が含まれます。グリッド内部で変数を並べ替えたり、「除外」リストに移動したりすることができます。右端の列にサイズを入力します。「ラベル」または「値」をクリックすると、グリッド・セル内の層化変数とクラスター変数の表示が値ラベルとデータ値で切り替わります。ラベルのない値を含むセルには、常に値が表示されます。「ストラータを再表示」をクリックすると、グリッド内の変数のラベル付きデータ値の各組み合わせがグリッドに再入力されます。

除外。 階層とクラスターの組み合わせのサブセットにサイズを指定するには、1 つ以上の変数を「除外」リストに移動します。これらの変数はサンプル・サイズの定義には使用されません。

分析準備ウィザード: 計画の要約

これは各段階の最後のステップです。現在の段階における分析計画の指定の要約が表示されます。ここから、次の段階 (必要な場合は作成します) に進むか、分析の指定を保存することができます。

別の段階を追加できない場合は、次の理由が考えられます。

- 「デザイン変数」ステップでクラスター変数が指定されていない。
- 「推定方法」ステップで、WR 推定を選択した。
- これが分析の第 3 段階であり、ウィザードでサポートされる段階は最大で 3 つである。

分析準備ウィザード: 終了

これは最後のステップです。計画ファイルをここで保存するか、選択内容をシンタックス・ウィンドウに貼り付けることができます。

既存の計画ファイルの段階に変更を加えた場合は、編集した計画を新しいファイルに保存するか、既存のファイルを上書きできます。既存の段階に変更を加えずに段階を追加した場合は、既存の計画ファイルが自動

的に上書きされます。計画を新しいファイルに保存する場合は、「ウィザードで生成したシンタックスをシンタックス ウィンドウに貼り付ける」を選択して、シンタックス・コマンドでファイル名を変更します。

既存の分析計画の変更

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「分析の準備...」

2. 「計画ファイルの編集」を選択してから、分析計画を保存する計画ファイルの名前を選択します。

3. 「次へ」をクリックして、ウィザードを進めます。

4. 「計画の要約」ステップで分析計画を確認した後、「次へ」をクリックします。

以降のステップは、新規計画の場合とほとんど同じです。詳しくは、個々のステップのヘルプを参照してください。

5. 「終了」ステップにナビゲートして、編集した計画ファイルに新しい名前を指定するか、既存の計画ファイルを上書きするように選択します。

オプションとして、計画から既存の段階を削除することができます。

分析準備ウィザード：計画の要約

このステップでは、分析計画を確認し、計画から段階を削除することができます。

段階の削除。 複数の段階があるデザインから、段階 2 および段階 3 を削除することができます。計画には 1 つ以上の段階が必要であるため、デザインの段階 1 については編集は可能ですが、削除はできません。

第 4 章 コンプレックス・サンプルの計画

コンプレックス・サンプルの分析プロシージャーで有効な結果を提供するためには、分析計画ファイルまたはサンプル計画ファイルからの分析の指定が必要です。

計画。 分析計画ファイルまたはサンプル計画ファイルのパスを指定します。

結合確率。 PPS WOR 法を使用して抽出されたクラスターに対して「不等な WOR」推定を使用するためには、結合確率を格納している別のファイルまたは開いているデータ・セットを指定する必要があります。このファイルまたはデータ・セットは、サンプリング・ウィザードでサンプリングを行う際に作成されます。

第 5 章 コンプレックス・サンプルの度数分布表

「コンプレックス・サンプルの度数分布表」プロシージャでは、選択した変数の度数分布表が作成され、1 変量の統計量が表示されます。必要に応じて、1 つ以上のカテゴリ変数で定義したサブグループによる統計量を要求することもできます。

例: 「コンプレックス・サンプルの度数分布表」プロシージャを使用すると、National Health Interview Survey (NHIS) の結果に基づき、この公共利用のためのデータに適した分析計画を使用して、米国国民のビタミン使用に関する表形式の 1 変量統計量を取得することができます。

統計: このプロシージャでは、セルの母集団のサイズと表パーセントの推定値、さらに各推定値の標準誤差、信頼区間、変動係数、デザイン効果、デザイン効果の平方根、累積値、および重み付けのない度数が得られます。これらに加えて、セルの比率が等しいかどうかの検定のために、カイ 2 乗および尤度比の統計量が計算されます。

コンプレックス・サンプルの度数分布表データについての考慮事項

データ: 度数分布表を生成する変数は、カテゴリ型である必要があります。部分母集団変数は文字列にも数値にも設定できますが、カテゴリ型である必要があります。

仮定: データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルの度数分布表を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「度数...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。
3. 「続行」をクリックします。
4. 「コンプレックス・サンプルの度数分布表」ダイアログ・ボックスで、1 つ以上の度数変数を選択します。

オプションとして、部分母集団を定義する変数を指定できます。各部分母集団について、統計量が別々に計算されます。

コンプレックス・サンプルの度数分布表の統計量

セル。 このグループでは、セルの母集団のサイズと表パーセントの推定値を要求できます。

統計量。 このグループでは、母集団のサイズまたは表パーセントに関連した統計量を作成します。

- **標準誤差。** 推定値の標準誤差。
- **信頼区間。** 指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。** 推定値に対する推定値の標準誤差の比率。
- **重み付けのない度数。** 推定値の計算に使用される単位数。

- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **累積値。** 変数の各値の累積の推定値。

等しいセル比率の検定。 これを選択すると、変数のカテゴリーの度数が等しいという仮説に対するカイ 2 乗検定および尤度比検定が作成されます。変数ごとに別個の検定が実行されます。

コンプレックス・サンプルの欠損値

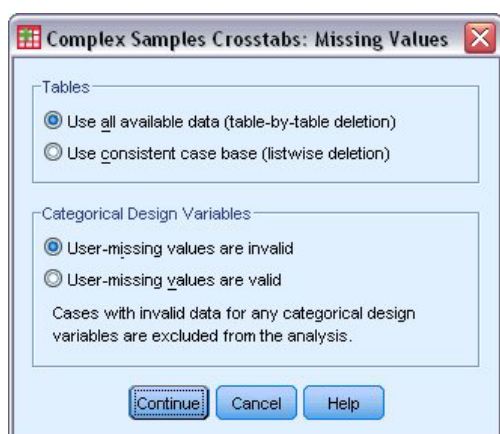


図 3. 「欠損値」ダイアログ・ボックス

テーブル。 このグループでは、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。** 欠損値は、テーブルごとに決定されます。したがって、統計量の計算に使用されるケースは、度数分布表またはクロス集計に応じて変わる場合があります。
- **一貫したケース基準を使用。** 欠損値は、変数全体で決定されます。したがって、統計量の計算に使用されるケースは、すべてのテーブルで一致しています。

カテゴリー計画変数。 このグループでは、ユーザー欠損値が有効か無効かを決定します。

コンプレックス・サンプルのオプション

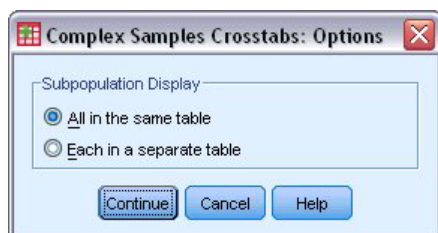


図 4. 「オプション」ダイアログ・ボックス

部分母集団の表示。 部分母集団を同じテーブルに表示するか、別々のテーブルに表示するかを選択できます。

第 6 章 コンプレックス・サンプルの記述統計

「コンプレックス・サンプルの記述統計」プロシージャでは、複数の変数に関する 1 変量の要約統計量が表示されます。必要に応じて、1 つ以上のカテゴリー変数で定義したサブグループによる統計量を要求することもできます。

例: 「コンプレックス・サンプルの記述統計」プロシージャを使用すると、National Health Interview Survey (NHIS) の結果に基づき、この公共利用のためのデータに適した分析計画を使用して、米国国民の活動水準に関する 1 変量の記述統計量を取得できます。

統計: このプロシージャでは、各推定値の平均と合計、および t 検定、標準誤差、信頼区間、変動係数、重み付けのない度数、母集団のサイズ、デザイン効果、デザイン効果の平方根が得られます。

コンプレックス・サンプルの記述統計データについての考慮事項

データ: 測定は、スケール変数である必要があります。部分母集団変数は文字列にも数値にも設定できませんが、カテゴリー型である必要があります。

仮定: データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルの記述統計を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「記述統計...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。

3. 「続行」をクリックします。

4. 「コンプレックス・サンプルの記述統計」ダイアログ・ボックスで、1 つ以上の測定変数を選択します。

オプションとして、部分母集団を定義する変数を指定できます。各部分母集団について、統計量が別々に計算されます。

コンプレックス・サンプルの記述統計の統計量

集計。 このグループでは、測定変数の平均と合計の推定値を要求できます。さらに、指定した値に対する推定値の t 検定を要求できます。

統計量。 このグループでは、平均または合計に関連する統計量が作成されます。

- **標準誤差。** 推定値の標準誤差。
- **信頼区間。** 指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。** 推定値に対する推定値の標準誤差の比率。
- **重み付けのない度数。** 推定値の計算に使用される単位数。
- **母集団のサイズ。** 母集団の推定単位数。

- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

コンプレックス・サンプルの記述統計の欠損値

測定変数の統計量。 このグループでは、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。** 欠損値は変数ごとに決定されます。したがって、統計量の計算に使用されるケースは測定変数によって異なる場合があります。
- **一貫したケース基準を確保。** 欠損値は、変数全体で決定されます。したがって、統計量の計算に使用されるケースはすべて同じになります。

カテゴリ計画変数。 このグループでは、ユーザー欠損値が有効か無効かを決定します。

コンプレックス・サンプルのオプション

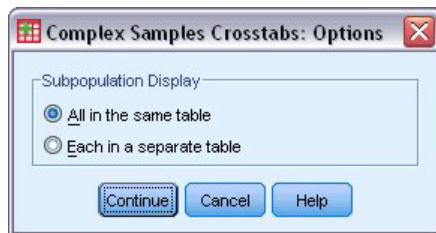


図5. 「オプション」ダイアログ・ボックス

部分母集団の表示。 部分母集団を同じテーブルに表示するか、別々のテーブルに表示するかを選択できます。

第 7 章 コンプレックス・サンプルのクロス集計表

「コンプレックス・サンプルのクロス集計」プロシージャでは、選択した変数のペアについてのクロス集計表が作成され、2次元統計量が表示されます。必要に応じて、1つ以上のカテゴリ変数で定義したサブグループによる統計量を要求することもできます。

例: 「コンプレックス・サンプルのクロス集計表」プロシージャを使用すると、National Health Interview Survey (NHIS) の結果に基づき、この公共利用のためのデータに適した分析計画を使用して、米国国民の喫煙頻度とビタミン使用についての交差分類の統計量を取得することができます。

統計: このプロシージャでは、セルの母集団のサイズと、行パーセント、列パーセント、表パーセントの推定値、さらに各推定値の標準誤差、信頼区間、変動係数、期待値、デザイン効果、デザイン効果の平方根、残差、調整済み残差、重み付けのない度数が得られます。2 x 2 列のテーブルでは、オッズ比、相対リスク、リスクの差が計算されます。さらに、行変数と列変数の独立性の検定のために、Pearson および尤度比の統計量が計算されます。

コンプレックス・サンプルのクロス集計データについての考慮事項

データ: 行変数と列変数はカテゴリ型である必要があります。部分母集団変数は文字列にも数値にも設定できますが、カテゴリ型である必要があります。

仮定: データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルのクロス集計表を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「クロス集計表...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。

3. 「続行」をクリックします。

4. 「コンプレックス・サンプルのクロス集計表」ダイアログ・ボックスで、1つ以上の行変数と1つの列変数を選択します。

オプションとして、部分母集団を定義する変数を指定できます。各部分母集団について、統計量が別々に計算されます。

コンプレックス・サンプルのクロス集計表の統計量

セル。 このグループでは、セルの母集団のサイズと、行、列、および表のパーセント (割合) の推定値を要求できます。

統計量。 このグループでは、母集団のサイズと、行、列、および表のパーセント (割合) に関連する統計量が作成されます。

- **標準誤差。** 推定値の標準誤差。
- **信頼区間。** 指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。** 推定値に対する推定値の標準誤差の比率。

- **期待値。** 行と列の変数が独立しているという仮説に基づいた、推定値の期待値。
- **重み付けのない度数。** 推定値の計算に使用される単位数。
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **残差。** 期待値は、2 つの変数の間に関係がないと想定した場合に期待されるセルのケース数です。正の残差は、行変数と列変数が独立であると想定される場合に、期待されたケース数よりセルの実際のケース数が多いことを示します。
- **調整済み残差。** セルの残差 (観測値から期待値を引いた値) をその標準誤差の推定値で割った値。結果として得られる標準化残差は、標準偏差を単位として平均より上または下で表されます。

2 X 2 列のテーブルの集計。 このグループでは、行と列の変数にそれぞれ 2 つのカテゴリーがある表の統計量が作成されます。それぞれが、ある因子の存在とあるイベントの発生の関連の強さの測定値になります。

- **オッズ比。** 因子の発生がまれなときに、オッズ比を相対リスクの推定値として使用できます。
- **相対リスク。** 因子が存在するイベントのリスクと、因子が存在しないイベントのリスクとの比率。
- **リスクの差分。** 因子が存在するイベントのリスクと、因子が存在しないイベントのリスクとの差。

行と列の独立性の検定。 これを選択すると、行と列の変数が独立しているという仮説に対するカイ 2 乗検定および尤度比検定が作成されます。変数のペアごとに別々の検定が実行されます。

コンプレックス・サンプルの欠損値

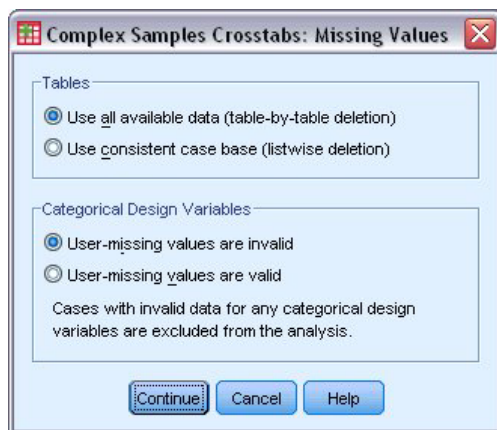


図 6. 「欠損値」ダイアログ・ボックス

テーブル。 このグループでは、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。** 欠損値は、テーブルごとに決定されます。したがって、統計量の計算に使用されるケースは、度数分布表またはクロス集計に応じて変わる場合があります。
- **一貫したケース基準を使用。** 欠損値は、変数全体で決定されます。したがって、統計量の計算に使用されるケースは、すべてのテーブルで一貫しています。

カテゴリー計画変数。 このグループでは、ユーザー欠損値が有効か無効かを決定します。

コンプレックス・サンプルのオプション

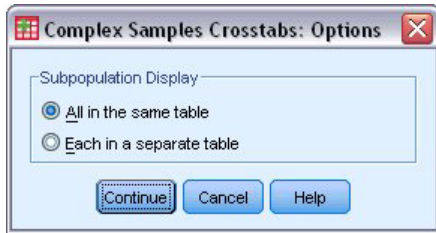


図7. 「オプション」ダイアログ・ボックス

部分母集団の表示。 部分母集団を同じテーブルに表示するか、別々のテーブルに表示するかを選択できます。

第 8 章 コンプレックス・サンプルの比率分析

「コンプレックス・サンプルの比率分析」プロシージャでは、変数の比率に関する 1 変量の要約統計量が表示されます。必要に応じて、1 つ以上のカテゴリー変数で定義したサブグループによる統計量を要求することもできます。

例: 「コンプレックス・サンプルの比率分析」プロシージャを使用すると、コンプレックス・デザインに従って実行された全州での調査の結果に基づき、そのデータに適した分析計画を使用して、現在の資産額と前回査定された資産額との比率に関する記述統計量を取得できます。

統計: このプロシージャでは、比率の推定値、 t 検定、標準誤差、信頼区間、変動係数、重み付けのない度数、母集団のサイズ、デザイン効果、およびデザイン効果の平方根が得られます。

コンプレックス・サンプルの比率分析データについての考慮事項

データ: 分子と分母は正の値を持つスケール変数である必要があります。部分母集団変数は文字列にも数値にも設定できますが、カテゴリー型である必要があります。

仮定: データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルの比率分析を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「比率...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。

3. 「続行」をクリックします。

4. 「コンプレックス・サンプルの比率分析」ダイアログ・ボックスで、1 つ以上の分子変数と分母変数を選択します。

オプションとして、サブグループを定義する変数を指定して、サブグループに対する統計量を作成することができます。

コンプレックス・サンプルの比率分析の統計量

統計量。 このグループでは、比率の推定値に関連する統計量が作成されます。

- **標準誤差。** 推定値の標準誤差。
- **信頼区間。** 指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。** 推定値に対する推定値の標準誤差の比率。
- **重み付けのない度数。** 推定値の計算に使用される単位数。
- **母集団のサイズ。** 母集団の推定単位数。
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

t 検定。 指定した値に対する推定値の t 検定を要求することができます。

コンプレックス・サンプルの比率分析の欠損値

比率。 このグループでは、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。** 欠損値は、比率ごとに決定されます。したがって、統計量の計算に使用されるケースは、分子と分母のペアごとに異なる場合があります。
- **一貫したケース基準を確保。** 欠損値は、変数全体で決定されます。したがって、統計量の計算に使用されるケースはすべて同じになります。

カテゴリ計画変数。 このグループでは、ユーザー欠損値が有効か無効かを決定します。

コンプレックス・サンプルのオプション

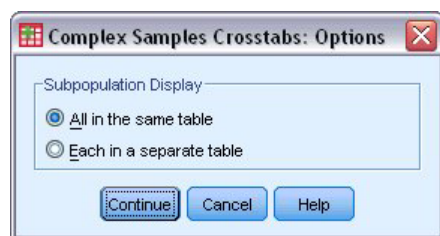


図 8. 「オプション」ダイアログ・ボックス

部分母集団の表示。 部分母集団を同じテーブルに表示するか、別々のテーブルに表示するかを選択できます。

第 9 章 コンプレックス・サンプルの一般線型モデル

「コンプレックス・サンプルの一般線型モデル」(CSGLM) プロシージャーでは、コンプレックス (複雑な) サンプルング方法で抽出したサンプルに対して、線型回帰分析や、分散分析および共分散分析を実行します。必要に応じて、部分母集団の分析も要求できます。

例: あるスーパーチェーンが、コンプレックス・デザインに従って、特定の顧客グループを対象に、購買習慣について調査しました。そのスーパーチェーンは、調査結果と各顧客の前月の購買額を前提に、顧客を性別によって分類し、サンプルング計画を取り入れて、顧客の買い物の頻度が 1 カ月間の購買額に関係しているかどうかを調べます。

統計: このプロシージャーでは、モデル・パラメーターに対する推定値、標準誤差、信頼区間、 t 検定、デザイン効果、およびデザイン効果の平方根、さらにはパラメーター推定値間の相関と共分散が得られます。従属変数と独立変数のモデルの適合度および記述統計も提供されます。このほかに、モデル因子と因子交互作用のレベルに関する推定周辺平均値を要求できます。

コンプレックス・サンプルの一般線型モデル・データについての考慮事項

データ: 従属変数は量的な変数です。因子はカテゴリ型です。共変量は、従属変数に関連する量的な変数です。部分母集団変数は文字列にも数値にも設定できますが、カテゴリ型である必要があります。

仮定. データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルの一般線型モデルを取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「一般線型モデル...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。

3. 「続行」をクリックします。

4. 「コンプレックス・サンプルの一般線型モデル」ダイアログ・ボックスで、従属変数を選択します。

オプションとして、以下の操作を実行できます。

- 因子および共変量の変数をデータの内容に合わせて選択する。
- 部分母集団を定義する変数を指定する。部分母集団変数の選択したカテゴリのみを対象として、分析が実行されます。

コンプレックス・サンプルの一般線型モデル

モデル効果の指定. このプロシージャーでは、デフォルトで、メイン・ダイアログ・ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。あるいは、交互作用効果と入れ子項目を含むカスタム・モデルを作成することもできます。

非入れ子項目

選択した因子と共変量について、次の項目を選択できます。

交互作用。 選択したすべての変数について、最も高いレベルの交互作用項を作成します。

主効果。 選択した各変数に主効果の項目を作成します。

2 次まで。 選択した変数の 2 次までの考えられるすべての交互作用を作成します。

3 次まで。 選択した変数の 3 次までの考えられるすべての交互作用を作成します。

4 次まで。 選択した変数の 4 次までの考えられるすべての交互作用を作成します。

5 次まで。 選択した変数の 5 次までの考えられるすべての交互作用を作成します。

入れ子項目

このプロシージャで、モデルに入れ子項目を作成できます。入れ子項目は、因子または共変量の値と、別の因子のレベルとの間に交互作用がない場合の因子または共変量の効果のモデル化に役立ちます。例えば、あるスーパーチェーンが、いくつかの店舗で顧客の消費性向を追跡調査すると仮定します。1 人の顧客が頻繁に利用するのは、これらの店舗のうちの 1 つだけであるため、「顧客」効果は「店舗の場所」効果の入れ子になっていると言えます。

さらに、入れ子項目に、同一の共変量に関わる多項式の項などの交互作用効果を含めたり、複数レベルの入れ子を追加したりすることができます。

制限。 入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A*A$ の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A(A)$ の指定は無効です。
- 共変量内で効果を入れ子にすることはできません。したがって、 A が因子で X が共変量である場合、 $A(X)$ の指定は無効です。

切片。 通常、モデルには切片が含まれます。データが原点を通ると仮定できる場合は、切片を除外できます。モデル内に切片を含める場合でも、切片に関連する統計を抑制するように選択できます。

コンプレックス・サンプルの一般線型モデルの統計量

モデル・パラメーター。 このグループでは、モデルのパラメーターに関連する統計量の表示を制御できます。

- **推定値。** 係数の推定値を表示します。
- **標準誤差。** 係数の推定値ごとの標準誤差を表示します。
- **信頼区間。** 係数の推定値ごとの信頼区間を表示します。区間の信頼度レベルは、「オプション」ダイアログ・ボックスで設定されます。
- **t 検定。** 係数の推定値ごとの t 検定を表示します。各検定の帰無仮説は、係数の値が 0 であることです。
- **パラメーター推定値の共分散。** モデル係数の共分散行列の推定値を表示します。
- **パラメーター推定値の相関。** モデル係数の相関行列の推定値を表示します。
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

モデルの適合度。 R^2 および平方平均誤差の平方根の統計量を表示します。

従属変数と共変量の母集団の平均値。 従属変数、共変量、および因子に関する要約情報を表示します。

サンプル・デザインの情報。 重み付けのない度数や母集団のサイズを含む、サンプルに関する要約情報を表示します。

コンプレックス・サンプルの仮説の検定

検定統計量。 このグループでは、仮説の検定に使用される統計量の種類を選択できます。「 F 」、「調整済み F 」、「カイ 2 乗」、「調整済みカイ 2 乗」の中から選択できます。

サンプリングの自由度。 このグループでは、すべての検定統計量の p 値の計算に使用されるサンプリング計画の自由度を制御できます。サンプリング計画に基づく場合は、値は基本サンプリング単位数とサンプリングの第 1 段階の階層数の差になります。あるいは、正の整数を指定することで、カスタム自由度を設定することもできます。

多重比較の調整。 多重対比を用いた仮説検定を行う場合、全体の有意水準を、含まれている対比の有意水準を基に調整することができます。このグループでは、調整方法を選択できます。

- **最小有意差。** この方法では、帰無仮説の値と異なる線型対比が存在するという仮説を棄却する全体の確率は制御されません。
- *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。
- *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。
- *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。

コンプレックス・サンプルの一般線型モデルの推定平均

「推定平均」ダイアログ・ボックスでは、「モデル」サブダイアログ・ボックスで指定した因子と因子交互作用のレベルについての、モデルで推定される周辺平均を表示できます。また、母集団全体の平均を表示するように要求することもできます。

項。 選択した因子と因子交互作用に対する推定平均が計算されます。

対比。 対比によって、推定平均を比較するための仮説検定のセットアップ方法が決まります。

- **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。
- **全平均。** 各レベルの平均 (参照カテゴリーを除く) をすべてのレベルの平均 (全平均) と比較します。因子のレベルの次数は任意です。
- **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。** 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。

参照カテゴリ。 単純対比と全平均の対比では、参照カテゴリ、つまり比較対象とする因子レベルを指定する必要があります。

コンプレックス・サンプルの一般線型モデルの保存

変数の保存。 このグループでは、モデルで予測された値と残差を、新しい変数として作業ファイルに保存することができます。

IBM SPSS Statistics データとしてモデルをエクスポート。 パラメーター推定値、標準誤差、有意確率値、および自由度を含むパラメーター相関行列またはパラメーター分散共分散行列を格納するデータ・セットを、IBM SPSS Statistics 形式で書き込みます。行列ファイル内の変数の順序は次のようになります。

- **rowtype_**。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)、および DF (サンプリング計画の自由度) の各値 (および値ラベル) を取ります。モデル・パラメーターごとに、行タイプ COV (または CORR) を持つ別個のケースが存在し、このほかに、その他の各行タイプごとに別々のケースが存在します。
- **varname_**。行タイプ COV または CORR では、すべてのモデル・パラメーターの番号付きリストに対応する P1、P2、... の値を取り、パラメーター推定値テーブルに示されているパラメーターの文字列に対応する値ラベルを持ちます。その他の行タイプでは、セルはブランクになります。
- **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に対応し、変数ラベルはパラメーター推定値テーブルに示されているパラメーターの文字列に対応します。これらの変数は、行タイプに応じた値を取ります。冗長パラメーターの場合、すべての共分散は 0 に、相関はシステム欠損値に設定されます。すべてのパラメーター推定値は 0 に設定され、標準誤差、有意水準、残差自由度はすべてシステム欠損値に設定されます。

注: 行列ファイルを読み取る他のプロシージャで、ここでエクスポートされたすべての行タイプが受け入れられない場合、このファイルを直接その後の分析に使用することはできません。

XML としてモデルをエクスポート。 パラメーター推定値と、パラメーター分散共分散行列 (選択した場合) 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、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。

コンプレックス・サンプルの一般線型モデルのオプション

ユーザー欠損値。 すべてのデザイン変数、および従属変数と共変量が、有効なデータを持っている必要があります。これらのいずれかの変数のデータが無効なケースは、分析から除外されます。この制御によって、ストラータ変数、クラスター変数、部分母集団変数、および因子変数の中で、ユーザー欠損値を有効として扱うかどうかを決定できます。

信頼区間。 これは、係数推定と推定周辺平均の信頼区間レベルです。50 以上 100 未満の値を指定します。

CSGLM 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。

- 線型に組み合わせた効果や値に対し、ユーザー指定による効果の検定を指定する (CUSTOM サブコマンドを使用)。
- 推定周辺平均の計算時に、共変量を平均値以外の値に固定する (EMMEANS サブコマンドを使用)。
- 多項式対比の計量を指定する (EMMEANS サブコマンドを使用)。
- 特異性の検査のための許容値を指定する (CRITERIA サブコマンドを使用)。
- 保存した変数のユーザー指定の名前を作成する (SAVE サブコマンドを使用)。
- 一般推定可能関数テーブルを作成する (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 10 章 コンプレックス・サンプルのロジスティック回帰

「コンプレックス・サンプルのロジスティック回帰」プロシージャーでは、コンプレックス (複雑な) サンプリング方法によって抽出されたサンプルに対して、2 値または多項の従属変数のロジスティック回帰分析を実行します。必要に応じて、部分母集団の分析も要求できます。

例: ある融資担当者が、コンプレックス・デザインに従って、複数の異なる支店で融資を受けた顧客の過去の記録を収集しました。担当者は、顧客の債務不履行の確率が年齢、職歴、クレジットカードの負債額に関係しているかどうかを、サンプル計画を取り入れながら調べます。

統計: このプロシージャーでは、モデル・パラメーターに対する推定値、指数化推定、標準誤差、信頼区間、 t 検定、デザイン効果、およびデザイン効果の平方根、さらにはパラメーター推定値間の相関と共分散が得られます。従属変数および独立変数の擬似 R^2 統計量、分類テーブル、記述統計量も提供されます。

コンプレックス・サンプルのロジスティック回帰データについての考慮事項

データ: 従属変数はカテゴリ変数です。因子はカテゴリ型です。共変量は、従属変数に関連する量的な変数です。部分母集団変数は文字列にも数値にも設定できますが、カテゴリ型である必要があります。

仮定. データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルのロジスティック回帰の取得

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「ロジスティック回帰...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。
3. 「続行」をクリックします。
4. 「コンプレックス・サンプルのロジスティック回帰」ダイアログ・ボックスで、従属変数を選択します。

オプションとして、以下の操作を実行できます。

- 因子および共変量の変数をデータの内容に合わせて選択する。
- 部分母集団を定義する変数を指定する。部分母集団変数の選択したカテゴリのみを対象として、分析が実行されます。

コンプレックス・サンプルのロジスティック回帰の参照カテゴリ

「コンプレックス・サンプルのロジスティック回帰」プロシージャーのデフォルトでは、最高値を持つカテゴリが参照カテゴリになります。このダイアログ・ボックスでは、参照カテゴリとして、最高値、最低値、またはカスタム・カテゴリを指定できます。

コンプレックス・サンプルのロジスティック回帰モデル

モデル効果の指定。 このプロシージャーでは、デフォルトで、メイン・ダイアログ・ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。あるいは、交互作用効果と入れ子項目を含むカスタム・モデルを作成することもできます。

非入れ子項目

選択した因子と共変量について、次の項目を選択できます。

交互作用。 選択したすべての変数について、最も高いレベルの交互作用項を作成します。

主効果。 選択した各変数に主効果の項目を作成します。

2 次まで。 選択した変数の 2 次までの考えられるすべての交互作用を作成します。

3 次まで。 選択した変数の 3 次までの考えられるすべての交互作用を作成します。

4 次まで。 選択した変数の 4 次までの考えられるすべての交互作用を作成します。

5 次まで。 選択した変数の 5 次までの考えられるすべての交互作用を作成します。

入れ子項目

このプロシージャーで、モデルに入れ子項目を作成できます。入れ子項目は、因子または共変量の値と、別の因子のレベルとの間に交互作用がない場合の因子または共変量の効果のモデル化に役立ちます。例えば、あるスーパーチェーンが、いくつかの店舗で顧客の消費性向を追跡調査すると仮定します。1 人の顧客が頻繁に利用するのは、これらの店舗のうちの 1 つだけであるため、「顧客」効果は「店舗の場所」効果の入れ子になっていると言えます。

さらに、入れ子項目に、同一の共変量に関わる多項式の項などの交互作用効果を含めたり、複数レベルの入れ子を追加したりすることができます。

制限。 入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A*A$ の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A(A)$ の指定は無効です。
- 共変量内で効果を入れ子にすることはできません。したがって、 A が因子で X が共変量である場合、 $A(X)$ の指定は無効です。

切片。 通常、モデルには切片が含まれます。データが原点を通ると仮定できる場合は、切片を除外できます。モデル内に切片を含める場合でも、切片に関連する統計を抑制するように選択できます。

コンプレックス・サンプルのロジスティック回帰の統計量

モデルの適合度。 モデル全体のパフォーマンスを測定する統計量の表示を制御します。

- **擬似 R^2 乗。** ロジスティック回帰モデルには、線型回帰の R^2 統計量に完全に対応するものはありません。代わりに、 R^2 統計量に類似したプロパティを模倣する測定値が複数存在します。
- **分類テーブル。** 従属変数についての、観測カテゴリーとモデルで予測されたカテゴリーの表形式の交差分類を表示します。

パラメーター。このグループでは、モデルのパラメーターに関連する統計量の表示を制御できます。

- **推定値。** 係数の推定値を表示します。
- **指数化推定。** 自然対数の底を係数の推定値でべき乗した値を表示します。推定値のプロパティが統計検定に適している場合は、指数化推定、つまり $\exp(B)$ の解釈が容易になります。
- **標準誤差。** 係数の推定値ごとの標準誤差を表示します。
- **信頼区間。** 係数の推定値ごとの信頼区間を表示します。区間の信頼度レベルは、「オプション」ダイアログ・ボックスで設定されます。
- **t 検定。** 係数の推定値ごとの t 検定を表示します。各検定の帰無仮説は、係数の値が 0 であることです。
- **パラメーター推定値の共分散。** モデル係数の共分散行列の推定値を表示します。
- **パラメーター推定値の相関。** モデル係数の相関行列の推定値を表示します。
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

モデル変数の要約統計量。 従属変数、共変量、および因子に関する要約情報を表示します。

サンプル・デザインの情報。 重み付けのない度数や母集団のサイズを含む、サンプルに関する要約情報を表示します。

コンプレックス・サンプルの仮説の検定

検定統計量。 このグループでは、仮説の検定に使用される統計量の種類を選択できます。「 F 」、「調整済み F 」、「カイ 2 乗」、「調整済みカイ 2 乗」の中から選択できます。

サンプリングの自由度。 このグループでは、すべての検定統計量の p 値の計算に使用されるサンプリング計画の自由度を制御できます。サンプリング計画に基づく場合は、値は基本サンプリング単位数とサンプリングの第 1 段階の階層数の差になります。あるいは、正の整数を指定することで、カスタム自由度を設定することもできます。

多重比較の調整。 多重対比を用いた仮説検定を行う場合、全体の有意水準を、含まれている対比の有意水準を基に調整することができます。このグループでは、調整方法を選択できます。

- **最小有意差。** この方法では、帰無仮説の値と異なる線型対比が存在するという仮説を棄却する全体の確率は制御されません。
- *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。
- *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。
- *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。

コンプレックス・サンプルのロジスティック回帰のオッズ比

「オッズ比」ダイアログ・ボックスでは、指定した因子と共変量について、モデルで推定されるオッズ比を表示できます。参照カテゴリーを除く従属変数の各カテゴリーごとに、1組のオッズ比が計算されます。

因子。 ここで選択した因子ごとに、指定の参照カテゴリーのオッズに対する、その因子の各カテゴリーのオッズの比率を表示します。

共変量。 ここで選択した共変量ごとに、(共変量の平均値 + 指定した変化の単位)でのオッズと、平均値でのオッズとの比率を表示します。

このプロシージャでは、因子または共変量のオッズ比の計算時に、その他のすべての因子は最高レベルで固定され、その他のすべての共変量は平均値に固定されます。因子または共変量とモデルの他の予測値との間に交互作用がある場合、オッズ比は指定した変数の変化だけでなく、因子または共変量と交互作用がある変数の値にも依存します。指定した共変量とモデル内の共変量自体との間に交互作用がある場合 (例えば、*age*age*)、オッズ比はその共変量の変化と共変量の値の両方に依存します。

コンプレックス・サンプルのロジスティック回帰の保存

変数の保存。 このグループでは、モデルで予測されたカテゴリーと予測された確率を新しい変数としてアクティブなデータ・セットに保存することができます。

IBM SPSS Statistics データとしてモデルをエクスポート。 パラメーター推定値、標準誤差、有意確率値、および自由度を含むパラメーター相関行列またはパラメーター分散共分散行列を格納するデータ・セットを、IBM SPSS Statistics 形式で書き込みます。行列ファイル内の変数の順序は次のようになります。

- **rowtype_.** 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)、および DF (サンプリング計画の自由度) の各値 (および値ラベル) を取ります。モデル・パラメーターごとに、行タイプ COV (または CORR) を持つ別個のケースが存在し、このほかに、その他の各行タイプごとに別々のケースが存在します。
- **varname_.** 行タイプ COV または CORR では、すべてのモデル・パラメーターの番号付きリストに対応する P1、P2、... の値を取り、パラメーター推定値テーブルに示されているパラメーターの文字列に対応する値ラベルを持ちます。その他の行タイプでは、セルはブランクになります。
- **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に対応し、変数ラベルはパラメーター推定値テーブルに示されているパラメーターの文字列に対応します。これらの変数は、行タイプに応じた値を取ります。冗長パラメーターの場合、すべての共分散は 0 に、相関はシステム欠損値に設定されます。すべてのパラメーター推定値は 0 に設定され、標準誤差、有意水準、残差自由度はすべてシステム欠損値に設定されます。

注: 行列ファイルを読み取る他のプロシージャで、ここでエクスポートされたすべての行タイプが受け入れられない場合、このファイルを直接その後の分析に使用することはできません。

XML としてモデルをエクスポート。 パラメーター推定値と、パラメーター分散共分散行列 (選択した場合) 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、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。

コンプレックス・サンプルのロジスティック回帰のオプション

推定。 このグループでは、モデルの推定に使用されるさまざまな基準を制御できます。

- 「最大反復回数」。アルゴリズムで実行される反復の最大回数。負でない整数を指定します。

- **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 倍に縮小されます。正の整数を指定します。
- **パラメーター推定値の変化に基づいて反復回数を制限。** これを選択した場合、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よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は負でない値である必要があります。
- **対数尤度の変化に基づいて反復回数を制限。** これを選択した場合、対数尤度関数の絶対変化または相対変化が指定した値よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は負でない値である必要があります。
- **データ・ポイントの完全分離を確認。** これを選択した場合、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。このプロシージャーによって各ケースが正しく分類されるモデルを作成できる場合に、分離が発生します。
- **反復履歴を表示。** 0 回目 (初期推定値) から数えて n 回反復するごとの、パラメーター推定値と統計量を表示します。反復履歴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。

ユーザー欠損値。 すべてのデザイン変数、および従属変数と共変量が、有効なデータを持っている必要があります。これらのいずれかの変数のデータが無効なケースは、分析から除外されます。この制御によって、ストラータ変数、クラスター変数、部分母集団変数、および因子変数の中で、ユーザー欠損値を有効として扱うかどうかを決定できます。

信頼区間。 これは、係数の推定値、指数化された係数の推定値、およびオッズ比の信頼区間レベルです。50 以上 100 未満の値を指定します。

CSLOGISTIC 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。

- 線型に組み合わせた効果や値に対し、ユーザー指定による効果の検定を指定する (CUSTOM サブコマンドを使用)。
- 因子と共変量のオッズ比の計算時に、他のモデル変数の値を固定する (ODDSRATIOS サブコマンドを使用)。
- 特異性の検査のための許容値を指定する (CRITERIA サブコマンドを使用)。
- 保存した変数のユーザー指定の名前を作成する (SAVE サブコマンドを使用)。
- 一般推定可能関数テーブルを作成する (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 コンプレックス・サンプルの順序回帰

「コンプレックス・サンプルの順序回帰」プロシージャでは、コンプレックス・サンプリング方法で抽出されたサンプルに対して、2 値または順序の従属変数の回帰分析を実行します。必要に応じて、部分母集団の分析も要求できます。

例: 議会開会前に法案の提出を検討している議員は、その法案に市民の支持があるかどうか、法案への支持が有権者の人口統計にどの程関係しているかに関心を持っています。そこで、世論調査会社がコンプレックス・サンプリング計画に基づいて、対面式の調査を計画し実施することになりました。コンプレックス・サンプルの順序回帰を使用して、有権者の人口統計に基づいて法案の支持レベルのモデルを当てはめることができます。

コンプレックス・サンプルの順序回帰データについての考慮事項

データ: 従属変数は順序変数です。因子はカテゴリ型です。共変量は、従属変数に関連する量的な変数です。部分母集団変数は文字列にも数値にも設定できますが、カテゴリ型である必要があります。

仮定: データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

コンプレックス・サンプルの順序回帰を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「順序回帰...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。
3. 「続行」をクリックします。
4. 「コンプレックス・サンプルの順序回帰」ダイアログ・ボックスで、従属変数を選択します。

オプションとして、以下の操作を実行できます。

- 因子および共変量の変数をデータの内容に合わせて選択する。
- 部分母集団を定義する変数を指定する。部分母集団変数の選択したカテゴリのみを対象として分析が実行されますが、分散はデータ・セット全体に基づいて正しく推定されます。
- リンク関数を選択します。

リンク関数。 リンク関数は、モデルの推定を可能にする累積確率の変換関数です。以下の 5 つのリンク関数を使用できます。

- **ロジット。** $f(x)=\log(x/(1-x))$ 。通常は、分布が均一なカテゴリに使用されます。
- **補ログ・マイナス・ログ。** $f(x)=\log(-\log(1-x))$ 。通常は、順序の高いカテゴリがより確率が高いときに使用されます。
- **負ログ・マイナス・ログ。** $f(x)=-\log(-\log(x))$ 。通常は、順序の低いカテゴリがより確率が高いときに使用されます。
- **プロビット。** $f(x)=\Phi^{-1}(x)$ 。通常は、潜在変数が正規分布しているときに使用されます。
- **コーチット(コーシーの逆関数)。** $f(x)=\tan(\pi(x-0.5))$ 。通常は、潜在変数に多数の極値が存在するときに使用されます。

コンプレックス・サンプルの順序回帰の応答確率

「応答確率」ダイアログ・ボックスでは、応答の累積確率（従属変数の特定のカテゴリまでのすべてのカテゴリに属する確率）が、従属変数の値の増加に応じて増加するか、従属変数の値の減少に応じて増加するかを指定できます。

コンプレックス・サンプルの順序回帰のモデル

モデル効果の指定。 このプロシージャでは、デフォルトで、メイン・ダイアログ・ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。あるいは、交互作用効果と入れ子項目を含むカスタム・モデルを作成することもできます。

非入れ子項目

選択した因子と共変量について、次の項目を選択できます。

交互作用。 選択したすべての変数について、最も高いレベルの交互作用項を作成します。

主効果。 選択した各変数に主効果の項目を作成します。

2 次まで。 選択した変数の 2 次までの考えられるすべての交互作用を作成します。

3 次まで。 選択した変数の 3 次までの考えられるすべての交互作用を作成します。

4 次まで。 選択した変数の 4 次までの考えられるすべての交互作用を作成します。

5 次まで。 選択した変数の 5 次までの考えられるすべての交互作用を作成します。

入れ子項目

このプロシージャで、モデルに入れ子項目を作成できます。入れ子項目は、因子または共変量の値と、別の因子のレベルとの間に交互作用がない場合の因子または共変量の効果のモデル化に役立ちます。例えば、あるスーパーチェーンが、いくつかの店舗で顧客の消費性向を追跡調査すると仮定します。1 人の顧客が頻繁に利用するのは、これらの店舗のうちの 1 つだけであるため、「顧客」効果は「店舗の場所」効果の入れ子になっていると言えます。

さらに、入れ子項目に、同一の共変量に関わる多項式の項などの交互作用効果を含めたり、複数レベルの入れ子を追加したりすることができます。

制限。 入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A*A$ の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A(A)$ の指定は無効です。
- 共変量内で効果を入れ子にすることはできません。したがって、 A が因子で X が共変量である場合、 $A(X)$ の指定は無効です。

コンプレックス・サンプルの順序回帰の統計量

モデルの適合度。 モデル全体のパフォーマンスを測定する統計量の表示を制御します。

- **擬似 R 2 乗。** 順序回帰モデルには、線型回帰の R^2 統計量に完全に対応するものはありません。代わりに、 R^2 統計量に類似したプロパティを模倣する測定値が複数存在します。
- **分類テーブル。** 従属変数についての、観測カテゴリーとモデルで予測されたカテゴリーの表形式の交差分類を表示します。

パラメーター。 このグループでは、モデルのパラメーターに関連する統計量の表示を制御できます。

- **推定値。** 係数の推定値を表示します。
- **指数化推定。** 自然対数の底を係数の推定値でべき乗した値を表示します。推定値のプロパティが統計検定に適している場合は、指数化推定、つまり $\exp(B)$ の解釈が容易になります。
- **標準誤差。** 係数の推定値ごとの標準誤差を表示します。
- **信頼区間。** 係数の推定値ごとの信頼区間を表示します。区間の信頼度レベルは、「オプション」ダイアログ・ボックスで設定されます。
- **t 検定。** 係数の推定値ごとの t 検定を表示します。各検定の帰無仮説は、係数の値が 0 であることです。
- **パラメーター推定値の共分散。** モデル係数の共分散行列の推定値を表示します。
- **パラメーター推定値の相関。** モデル係数の相関行列の推定値を表示します。
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度を、標準誤差と比較可能な単位で表したものです。値が 1 から遠ざかるほど効果が大きいことを示します。

平行線。 このグループでは、各応答カテゴリー (最後の応答カテゴリーを除く) に異なる回帰直線が当てはめられる場合に、平行線以外のモデルに関連する統計量を要求できます。

- **Wald 検定。** すべての累積応答で回帰パラメーターが等しいという帰無仮説に関する検定を行います。平行線以外のモデルが推定され、等しいパラメーターの Wald 検定が適用されます。
- **パラメーター推定値。** 平行線以外のモデルの、係数と標準誤差の推定値を表示します。
- **パラメーター推定値の共分散。** 平行線以外のモデルの、係数の共分散行列の推定値を表示します。

モデル変数の要約統計量。 従属変数、共変量、および因子に関する要約情報を表示します。

サンプル・デザインの情報。 重み付けのない度数や母集団のサイズを含む、サンプルに関する要約情報を表示します。

コンプレックス・サンプルの仮説の検定

検定統計量。 このグループでは、仮説の検定に使用される統計量の種類を選択できます。「 F 」、「調整済み F 」、「カイ 2 乗」、「調整済みカイ 2 乗」の中から選択できます。

サンプリングの自由度。 このグループでは、すべての検定統計量の p 値の計算に使用されるサンプリング計画の自由度を制御できます。サンプリング計画に基づく場合は、値は基本サンプリング単位数とサンプリングの第 1 段階の階層数の差になります。あるいは、正の整数を指定することで、カスタム自由度を設定することもできます。

多重比較の調整。 多重対比を用いた仮説検定を行う場合、全体の有意水準を、含まれている対比の有意水準を基に調整することができます。このグループでは、調整方法を選択できます。

- **最小有意差。** この方法では、帰無仮説の値と異なる線型対比が存在するという仮説を棄却する全体の確率は制御されません。
- *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。
- *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。
- *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。

コンプレックス・サンプルの順序回帰のオッズ比

「オッズ比」ダイアログ・ボックスでは、指定した因子と共変量について、モデルで推定される累積オッズ比を表示できます。この機能は、ロジット・リンク関数を使用したモデルでのみ使用できます。最後のカテゴリを除く、従属変数の全カテゴリに対して 1 つの累積オッズ比が計算されます。比例オッズ・モデルでは、オッズ比がすべて等しいと仮定します。

因子。 ここで選択した因子ごとに、指定の参照カテゴリのオッズに対する、その因子の各カテゴリの累積オッズの比率を表示します。

共変量。 ここで選択した共変量ごとに、(共変量の平均値 + 指定した変化の単位) での累積オッズと、平均値でのオッズとの比率を表示します。

このプロシージャでは、因子または共変量のオッズ比の計算時に、その他のすべての因子は最高レベルで固定され、その他のすべての共変量は平均値に固定されます。因子または共変量とモデルの他の予測値との間に交互作用がある場合、オッズ比は指定した変数の変化だけでなく、因子または共変量と交互作用がある変数の値にも依存します。指定した共変量とモデル内の共変量自体との間に交互作用がある場合 (例えば、*age*age*)、オッズ比はその共変量の変化と共変量の値の両方に依存します。

コンプレックス・サンプルの順序回帰の保存

変数の保存。 このグループでは、モデルで予測されたカテゴリ、予測カテゴリの確率、観測カテゴリの確率、累積確率、および予測された確率を新しい変数としてアクティブなデータ・セットに保存することができます。

IBM SPSS Statistics データとしてモデルをエクスポート。 パラメーター推定値、標準誤差、有意確率値、および自由度を含むパラメーター相関行列またはパラメーター分散共分散行列を格納するデータ・セットを、IBM SPSS Statistics 形式で書き込みます。行列ファイル内の変数の順序は次のようになります。

- **rowtype_.** 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)、および DF (サンプリング計画の自由度) の各値 (および値ラベル) を取ります。モデル・パラメーターごとに、行タイプ COV (または CORR) を持つ別個のケースが存在し、このほかに、その他の各行タイプごとに別々のケースが存在します。
- **varname_.** 行タイプ COV または CORR では、すべてのモデル・パラメーターの番号付きリストに対応する P1、P2、... の値を取り、パラメーター推定値テーブルに示されているパラメーターの文字列に対応する値ラベルを持ちます。その他の行タイプでは、セルはブランクになります。
- **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に対応し、変数ラベルはパラメーター推定値テーブルに示されているパラメーターの文字列に対応します。これらの変数は、行

タイプに応じた値を取ります。冗長パラメーターの場合、すべての共分散は 0 に、相関はシステム欠損値に設定されます。すべてのパラメーター推定値は 0 に設定され、標準誤差、有意水準、残差自由度はすべてシステム欠損値に設定されます。

注: 行列ファイルを読み取る他のプロシージャで、ここでエクスポートされたすべての行タイプが受け入れられない場合、このファイルを直接その後の分析に使用することはできません。

XML としてモデルをエクスポート。 パラメーター推定値と、パラメーター分散共分散行列 (選択した場合) 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、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。

コンプレックス・サンプルの順序回帰のオプション

「推定方法」。パラメーターの推定方法を選択することができます。Newton-Raphson 法、Fisher スコア法、および Fisher スコア法を何回か反復実行した後 Newton-Raphson 法に切り替える混合型の方法の中からいずれかを選択できます。混合型の方法における Fisher スコア法の段階で、その最大反復回数に到達する前に収束が達成された場合も、Newton-Raphson 法のアルゴリズムは続行されます。

推定。 このグループでは、モデルの推定に使用されるさまざまな基準を制御できます。

- 「最大反復回数」。アルゴリズムで実行される反復の最大回数。負でない整数を指定します。
- **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 倍に縮小されます。正の整数を指定します。
- **パラメーター推定値の変化に基づいて反復回数を制限。** これを選択した場合、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よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は負でない値である必要があります。
- **対数尤度の変化に基づいて反復回数を制限。** これを選択した場合、対数尤度関数の絶対変化または相対変化が指定した値よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は負でない値である必要があります。
- **データ・ポイントの完全分離を確認。** これを選択した場合、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。このプロシージャによって各ケースが正しく分類されるモデルを作成できる場合に、分離が発生します。
- **反復履歴を表示。** 0 回目 (初期推定値) から数えて n 回反復するごとの、パラメーター推定値と統計量を表示します。反復履歴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。

ユーザー欠損値。 従属変数や共変量だけでなく、スケール・デザイン変数が有効なデータを持っている必要があります。これらのいずれかの変数のデータが無効なケースは、分析から除外されます。この制御によって、ストラータ変数、クラスター変数、部分母集団変数、および因子変数の中で、ユーザー欠損値を有効として扱うかどうかを決定できます。

信頼区間。 これは、係数の推定値、指数化された係数の推定値、およびオッズ比の信頼区間レベルです。50 以上 100 未満の値を指定します。

CSORDINAL 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。

- 線型に組み合わせた効果や値に対し、ユーザー指定による効果の検定を指定する (CUSTOM サブコマンドを使用)。

- 因子と共変量の累積オッズ比の計算時に、他のモデル変数の値を平均値以外の値に固定する (ODDSRATIOS サブコマンドを使用)。
- オッズ比の要求時に、ラベルなしの値を因子のカスタム参照カテゴリーとして使用する (ODDSRATIOS サブコマンドを使用)。
- 特異性の検査のための許容値を指定する (CRITERIA サブコマンドを使用)。
- 一般推定可能関数テーブルを作成する (PRINT サブコマンドを使用)。
- 25 個を超える確率変数を保存する (SAVE 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 コンプレックス・サンプルの Cox 回帰

「コンプレックス・サンプルの Cox 回帰」プロシージャーでは、コンプレックス・サンプリング方法によって抽出されたサンプルに対して、生存分析を実行します。必要に応じて、部分母集団の分析も要求できます。

例。 政府の法執行機関が、管轄区域での累犯率に注意を払っています。累犯率の尺度の 1 つは、犯罪者の 2 度目の逮捕までの時間です。この法執行機関は、Cox 回帰を使用して再逮捕までの時間をモデル化したと考えていますが、年齢カテゴリー全体に比例ハザードの仮定が適用できないことを心配しています。

医療研究者は、虚血性脳卒中のリハビリ・プログラムを終えた患者の生存時間を調査しています。死亡以外の重大な事象の発生とこれらの事象の時間が記録されるにつれて患者の病歴が変わるため、被験者ごとに複数のケースがある可能性があります。また、虚血性脳卒中の発生時がリスクの起点となる一方で、リハビリ・プログラム後に生存している患者のみがサンプルに含まれるため、観測される生存時間はリハビリの長さによって「膨張する」という意味で、サンプルは左側が切り捨てられます。

コンプレックス・サンプルの Cox 回帰データについての考慮事項

生存時間。 このプロシージャーでは、Cox 回帰を生存時間、すなわち事象の発生前の時間の長さの分析に適用します。生存時間を指定する方法には、区間の開始時間に応じて次の 2 とおりがあります。

- **時間=0。** 通常は、被験者ごとに区間の開始時についての完全な情報があり、終了時間を含む変数があるだけです (あるいは、日時の変数から終了時間を含む単一の変数を作成します。以下の説明を参照してください)。
- **被験者により変化。** これは、左側の切り捨て (遅延した投入とも呼ばれる) を行う場合に適していません。例えば、脳卒中の発作後のリハビリ・プログラムを終えた患者の生存時間を分析する場合、発作の発生時をリスクの起点と見なします。ただし、リハビリ・プログラム後生存している患者のみを含むサンプルの場合、観測される生存時間はリハビリの長さによって「膨張する」という意味で、サンプルは左側が切り捨てられます。調査に投入した時間としてリハビリを終えた時間を指定することで、この理由を説明できます。

日時の変数。 日時の変数を使用して、区間の開始と終了を直接定義することはできません。日時の変数がある場合、それらを使用して生存時間を含む変数を作成する必要があります。左側の切り捨てがない場合は、調査の投入の日付と観測の日付の差に基づいて、終了時間を含む変数を作成するだけです。左側の切り捨てがある場合は、調査の開始日付と投入の日付の差に基づいて開始時間を含む変数を作成し、調査の開始日付と観測の日付の差に基づいて終了時間を含む変数を作成します。

事象の状態。 被験者が区間内に関心の対象となる事象を経験したかどうかを記録する変数が必要です。関心の対象となる事象が発生していない被験者は、右側が打ち切られます。

被験者識別。 単一の被験者の観測値を複数のケースに分割することで、区分的に一定な時間依存の予測変数を容易に組み込むことができます。例えば、脳卒中後の患者の生存時間を分析する場合、病歴を表す変数は予測変数として役立ちます。時間の経過とともに、患者は病歴が変わる重大な医療事象を経験する可能性があります。次の表は、そのようなデータ・セットを構造化する方法を示しています。「患者 ID」は被験者の識別子です。「終了時間」は観測された区間を示します。「状態」は主要な医療事象を記録します。「心臓発作の既往歴」と「出血の既往歴」は区分的に一定な時間依存の予測変数です。

表 1. 区分的に一定な時間依存の予測変数を導入するためのデータ構造：

患者 ID	終了時間	状態	心臓発作の既往歴	出血の既往歴
1	5	心臓発作	いいえ	いいえ
1	7	出血	はい	いいえ
1	8	死亡	はい	はい
2	24	死亡	いいえ	いいえ
3	8	心臓発作	いいえ	いいえ
3	15	死亡	はい	いいえ

仮定: データ・ファイル内のケースは、「コンプレックス・サンプルの計画」ダイアログ・ボックスで選択したファイル内の指定に従って分析されるコンプレックス・デザインからのサンプルを表します。

通常、Cox 回帰モデルでは比例ハザード（つまり、あるケースと他のケースとのハザードの比率は、時間の経過と共に変化してはならない）を仮定します。この仮定が適用できない場合、モデルに時間依存の予測変数を追加する必要がある場合があります。

Kaplan-Meier 分析。 予測変数を選択せずに（または、選択した予測変数をモデルに投入せずに）、「オプション」タブでベースライン生存曲線の計算用に積極限法を選択すると、このプロシージャーでは Kaplan-Meier 型の生存分析が実行されます。

コンプレックス・サンプルの Cox 回帰を取得するには

1. メニューから次の項目を選択します。

「分析」 > 「コンプレックス サンプル」 > 「Cox 回帰...」

2. 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルを選択します。

3. 「続行」をクリックします。

4. 調査から投入時間と終了時間を選択して、生存時間を指定します。

5. 事象の状態変数を選択します。

6. 「事象の定義」をクリックして、1 つ以上の事象の値を定義します。

オプションとして、被験者の識別子を選択することができます。

イベントの定義

最終事象が発生したことを示す値を指定します。

- **個別値。** 1 つ以上の値を指定します。グリッドに値を入力するか、定義済みの値ラベルが付いた値のリストから選択を行います。
- **値の範囲。** 値の範囲を指定します。最小値と最大値を入力するか、定義済みの値ラベルが付いたリストから各値を選択します。

予測

「予測変数」タブでは、モデル効果を作成するための因子と共変量を指定できます。

因子。 因子はカテゴリ予測変数です。因子には数値か文字列を指定できます。

共変量。 共変量はスケール予測変数です。共変量には数値を指定する必要があります。

時間依存予測変数。 比例ハザードの仮定が適用できない状況があります。つまり、ハザード率が時間によって変化し、1 つ (または 1 つ以上) の予測変数の値が時点によって異なるような場合です。このような場合には、時間依存の予測変数を指定する必要があります。詳しくは、『時間依存の予測変数の定義』のトピックを参照してください。時間依存の予測変数は、因子または共変量としても選択できます。

時間依存の予測変数の定義

「時間依存の予測変数の定義」ダイアログ・ボックスでは、組み込み時間変数 $T_$ に依存する予測変数を作成できます。この変数を使用して、次の 2 とおりの一般的な方法で時間依存の共変量を定義できます。

- 非比例ハザードを可能にする拡張 Cox 回帰モデルを推定する場合は、時間依存の予測変数を時間変数 $T_$ と対象の共変量の関数として定義します。時間変数と予測変数の単純な積が一般的な例ですが、より複雑な関数も指定できます。
- 時間によって値が異なる可能性があるが、時間との系統的な関連性のない変数があります。このような場合には、**セグメント化した時間依存の予測変数**を定義する必要があります。これは論理式を使用して定義できます。論理式は、真の場合は値 1、偽の場合は値 0 を取ります。一連の論理式を使用すると、1 組の測定から時間依存の予測変数を作成できます。例えば、4 週間の調査期間に、週に 1 度血圧を測定する場合 ($BP1$ から $BP4$ で識別)、時間依存の予測変数を $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq 2 \ \& \ T_ < 3) * BP3 + (T_ \geq 3 \ \& \ T_ < 4) * BP4$ と定義できます。どのケースでも、括弧内の条件の 1 つだけが 1 となり、残りの条件はすべて 0 になります。つまり、この関数は、時間が 1 週間未満の場合は $BP1$ を使用し、1 週間より長く 2 週間未満の場合は $BP2$ を使用する、ということを示しています。

注: 上記の血圧の例のように、セグメント化した時間依存の予測変数がセグメント内で一定の場合、被験者を複数のケースに分割すると、区分的に一定な時間依存予測変数がより簡単に指定できるようになります。詳しくは、49 ページの『第 12 章 コンプレックス・サンプルの Cox 回帰』の被験者識別についての説明を参照してください。

「時間依存の予測変数の定義」ダイアログ・ボックスでは、関数作成用のコントロールを使用して時間依存の共変量の式を作成したり、「数式」テキスト域に式を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、数値定数は小数点としてドットを使用する米国方式で入力する必要があります。この結果生成される変数には、指定した名前が付けられ、「予測変数」タブに因子または共変量として表示されます。

サブグループ

ベースライン・ストラータ。 この変数の値ごとに、ベースライン・ハザード関数および生存関数が個別に計算されます。これに対して、ストラータ全体で 1 組のモデル係数が推定されます。

部分母集団変数。 部分母集団を定義する変数を指定します。部分母集団変数の選択したカテゴリーのみを対象として、分析が実行されます。

モデル

モデル効果の指定。 このプロシージャーでは、デフォルトで、メイン・ダイアログ・ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。あるいは、交互作用効果と入れ子項目を含むカスタム・モデルを作成することもできます。

非入れ子項目

選択した因子や共変量について、次の項を作成できます。

交互作用。 選択したすべての変数について、最も高いレベルの交互作用項を作成します。

主効果。 選択した各変数に主効果の項目を作成します。

2 次まで。 選択した変数の 2 次までの考えられるすべての交互作用を作成します。

3 次まで。 選択した変数の 3 次までの考えられるすべての交互作用を作成します。

4 次まで。 選択した変数の 4 次までの考えられるすべての交互作用を作成します。

5 次まで。 選択した変数の 5 次までの考えられるすべての交互作用を作成します。

入れ子項目

このプロシージャで、モデルに入れ子項目を作成できます。入れ子項目は、因子または共変量の値と、別の因子のレベルとの間に交互作用がない場合の因子または共変量の効果のモデル化に役立ちます。例えば、あるスーパーチェーンが、いくつかの店舗で顧客の消費性向を追跡調査すると仮定します。1 人の顧客が頻繁に利用するのは、これらの店舗のうちの 1 つだけであるため、「顧客」効果は「店舗の場所」効果の入れ子になっていると言えます。

さらに、入れ子項目に、同一の共変量に関わる多項式の項などの交互作用効果を含めたり、複数レベルの入れ子を追加したりすることができます。

制限。 入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A*A$ の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意である必要があります。したがって、 A が因子である場合、 $A(A)$ の指定は無効です。
- 共変量内で効果を入れ子にすることはできません。したがって、 A が因子で X が共変量である場合、 $A(X)$ の指定は無効です。

統計量

サンプル・デザインの情報。 重み付けのない度数や母集団のサイズを含む、サンプルに関する要約情報を表示します。

イベントおよび調査要約。 打ち切られたケースの数と割合に関する要約情報を表示します。

事象の時点のリスク・セット。 各ベースライン・ストラータ内のイベント時間ごとのイベント数とリスク数を表示します。

パラメーター。 このグループでは、モデルのパラメーターに関連する統計量の表示を制御できます。

- **推定値。** 係数の推定値を表示します。
- **指数化推定。** 自然対数の底を係数の推定値でべき乗した値を表示します。推定値のプロパティが統計検定に適している場合は、指数化推定、つまり $\exp(B)$ の解釈が容易になります。
- **標準誤差。** 係数の推定値ごとの標準誤差を表示します。
- **信頼区間。** 係数の推定値ごとの信頼区間を表示します。区間の信頼度レベルは、「オプション」ダイアログ・ボックスで設定されます。

- **t 検定。** 係数の推定値ごとの t 検定を表示します。各検定の帰無仮説は、係数の値が 0 であることです。
- **パラメーター推定値の共分散。** モデル係数の共分散行列の推定値を表示します。
- **パラメーター推定値の相関。** モデル係数の相関行列の推定値を表示します。
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比率。これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。
- **デザイン効果の平方根。** これはコンプレックス・デザインを指定する効果の尺度です。値が 1 から遠ざかるほど効果が大きいことを示します。

モデル仮定。 このグループでは、比例ハザード仮定の検定を作成できます。検定では、適合モデルと代替モデルとの比較が行われます。代替モデルには、各予測変数 x に対して、時間依存予測変数 $x*_TF$ ($_TF$ は指定された時間関数) が含まれます。

- **時間関数。** 代替モデルの $_TF$ の形式を指定します。同一関数の場合、 $_TF=T_$ になります。対数関数の場合は、 $_TF=\log(T_)$ になります。**Kaplan-Meier** の場合は、 $_TF=1-S_{KM}(T_)$ ($S_{KM}(\cdot)$ は、生存関数の Kaplan-Meier 推定) になります。**ランク** の場合、 $_TF$ は観測された終了時間の中の $T_$ のランク順位です。
- **代替モデルのパラメーター推定値。** 代替モデルのパラメーターごとの、推定値、標準誤差、および信頼区間を表示します。
- **代替モデルの共分散行列。** 代替モデルのパラメーター間の推定共分散の行列を表示します。

ベースライン生存関数とベースライン累積ハザード関数。 ベースライン生存関数およびベースライン累積ハザード関数と、それぞれの標準誤差を表示します。

注: モデルに「予測変数」タブで定義された時間依存の予測変数が含まれる場合、このオプションは使用できません。

プロット

「作図」タブでは、ハザード関数、生存関数、生存関数のログ・マイナス・ログ、および 1 マイナス 生存関数のプロットを要求できます。また、指定した関数に沿って信頼区間を作図するように選択することもできます。信頼度レベルは「オプション」タブで設定されます。

予測変数のパターン。 要求されたプロット、および「エクスポート」タブでエクスポートされる生存ファイルに使用される予測値のパターンを指定できます。「予測変数」タブで定義された時間依存予測変数がモデルに含まれる場合は、以下のオプションは使用できません。

- **因子のプロット。** デフォルトでは、各因子が最も高いレベルで評価されます。必要に応じて、別のレベルを入力または選択します。あるいは、因子のチェック・ボックスを選択することによって、1 つの因子のレベルごとに別個の線を作図することもできます。
- **共変量のプロット。** 各共変量はその平均値で評価されます。必要に応じて、別の値を入力または選択します。

仮説の検定

検定統計量。 このグループでは、仮説の検定に使用される統計量の種類を選択できます。「 F 」、「調整済み F 」、「カイ 2 乗」、「調整済みカイ 2 乗」の中から選択できます。

サンプリングの自由度。 このグループでは、すべての検定統計量の p 値の計算に使用されるサンプリング計画の自由度を制御できます。サンプリング計画に基づく場合は、値は基本サンプリング単位数とサンプリングの第 1 段階の階層数の差になります。あるいは、正の整数を指定することで、カスタム自由度を設定することもできます。

多重比較の調整。 多重対比を用いた仮説検定を行う場合、全体の有意水準を、含まれている対比の有意水準を基に調整することができます。このグループでは、調整方法を選択できます。

- **最小有意差。** この方法では、帰無仮説の値と異なる線型対比が存在するという仮説を棄却する全体の確率は制御されません。
- **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。
- **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。
- **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。

保存

変数の保存。 このグループでは、モデル関連の変数をアクティブなデータ・セットに保存して、さらに結果の診断や報告で利用できるようにします。モデルに時間依存の予測変数が含まれる場合は、このグループは使用できません。

- **生存関数。** ケースごとの、観測された時間および予測値での生存の確率 (生存関数の値) を保存します。
- **生存関数に対する信頼区間の下限。** ケースごとの、観測された時間および予測値での生存関数に対する信頼区間の下限値を保存します。
- **生存関数に対する信頼区間の上限。** ケースごとの、観測された時間および予測値での生存関数に対する信頼区間の上限値を保存します。
- **累積ハザード関数。** ケースごとの、観測された時間および予測値での累積ハザード、つまり $-\ln(\text{survival})$ を保存します。
- **累積ハザード関数に対する信頼区間の下限。** ケースごとの、観測された時間および予測値での累積ハザード関数に対する信頼区間の下限値を保存します。
- **累積ハザード関数に対する信頼区間の上限。** ケースごとの、観測された時間および予測値での累積ハザード関数に対する信頼区間の上限値を保存します。
- **線型予測の予測値。** 参照値で修正された予測値に回帰係数を掛けた線型結合を保存します。線型予測値は、ベースライン・ハザードに対するハザード関数の比です。比例ハザードのモデルでは、この値が時間によらず常に一定です。
- **Schoenfeld 残差。** モデル内の打ち切られていない各ケースおよび各非冗長パラメーターにおいて、Schoenfeld 残差は、モデル・パラメーターに関連する予測変数の観測値と、観測されたイベント時間のリスク・セット内にあるケースに対する予測変数の期待値の差です。Schoenfeld 残差は、比例ハザードの仮定の評価に使用できます。例えば、予測変数 x の場合、時間依存の予測変数 $x \cdot \ln(T_)$ の Schoenfeld 残差対時間のプロットでは、比例ハザードが適用される場合、0 の位置に横線が表示されます。モデル内の非冗長パラメーターごとに別個の変数が保存されます。Schoenfeld 残差は、打ち切られていないケースについてのみ計算されます。
- **Martingale 残差。** 各ケースにおいて、Martingale 残差は、観測された打ち切り (打ち切られた場合 0、打ち切られていない場合 1) と、観測時間中のイベントの期待値の差です。

- **最大対数尤度比残差。** 最大対数尤度比残差は、Martingale 残差を、0 に対してより対称に現れるように「調整」したものです。予測値に対する最大対数尤度比残差のプロットでは、パターンは示されません。
- **Cox-Snell 残差。** 各ケースにおいて、Cox-Snell 残差は観測時間中のイベントの期待値です。つまり、観測された打ち切りから Martingale 残差を引いた値です。
- **スコア残差。** モデル内の各ケースおよび各非冗長パラメーターにおいて、スコア残差は擬尤度の最初の微分係数に対するケースの寄与率です。モデル内の非冗長パラメーターごとに別個の変数が保存されます。
- **DFBeta 残差。** モデル内の各ケースおよび各非冗長パラメーターにおいて、DFBeta 残差は、ケースがモデルから除外された場合のパラメーター推定値の変化の近似です。DFBeta 残差が相対的に大きいケースは、分析に不適切な影響を与えている場合があります。モデル内の非冗長パラメーターごとに別個の変数が保存されます。
- **集計残差。** 複数のケースが単一の被験者を表す場合、被験者の集計残差は、同じ被験者に属する全ケースの対応するケース残差を単に合計したものです。Schoenfeld 残差は、打ち切られていないケースのみを対象として定義されるため、Schoenfeld 残差では、集計された残差は集計されていない残差と同じです。これらの残差は、「期間とイベント」タブで被験者の識別子が指定されている場合にのみ使用できます。

保存された変数の名前。 名前の自動生成によって、すべての作業を保存することができます。カスタム名を指定すると、最初にデータ・エディターで、保存された変数を削除する作業なしに、前回実行した結果の破棄または置き換えができます。

エクスポート

IBM SPSS Statistics データとしてモデルをエクスポート。 パラメーター推定値、標準誤差、有意確率値、および自由度を含むパラメーター相関行列またはパラメーター分散共分散行列を格納するデータ・セットを、IBM SPSS Statistics 形式で書き込みます。行列ファイル内の変数の順序は次のようになります。

- **rowtype_。** 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)、および DF (サンプリング計画の自由度) の各値 (および値ラベル) を取ります。モデル・パラメーターごとに、行タイプ COV (または CORR) を持つ別個のケースが存在し、このほかに、その他の各行タイプごとに別々のケースが存在します。
- **varname_。** 行タイプ COV または CORR では、すべてのモデル・パラメーターの番号付きリストに対応する P1、P2、... の値を取り、パラメーター推定値テーブルに示されているパラメーターの文字列に対応する値ラベルを持ちます。その他の行タイプでは、セルはブランクになります。
- **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に対応し、変数ラベルはパラメーター推定値テーブルに示されているパラメーターの文字列に対応します。これらの変数は、行タイプに応じた値を取ります。冗長パラメーターの場合、すべての共分散は 0 に、相関はシステム欠損値に設定されます。すべてのパラメーター推定値は 0 に設定され、標準誤差、有意水準、残差自由度はすべてシステム欠損値に設定されます。

注: 行列ファイルを読み取る他のプロシージャで、ここでエクスポートされたすべての行タイプが受け入れられない場合、このファイルを直接その後の分析に使用することはできません。

IBM SPSS Statistics データとして生存関数をエクスポート。 ベースライン、および「プロット」タブで指定された予測値のパターンで評価された、失敗の時間または事象の時間ごとの生存関数、生存関数の標準誤差、生存関数に対する信頼区間の上限と下限、および累積ハザード関数を格納するデータ・セットを IBM SPSS Statistics 形式で書き込みます。行列ファイル内の変数の順序は次のようになります。

- **ベースライン・ストラータ変数。** ストラータ変数の値ごとに個別の生命表が作成されます。
- **生存時間変数。** 事象の時間。固有の事象の時間ごとに異なるケースが作成されます。
- **Sur_0, LCL_Sur_0, UCL_Sur_0。** ベースライン生存関数、およびその信頼区間の上限と下限。
- **Sur_R, LCL_Sur_R, UCL_Sur_R。** 「参照」パターン (出力のパターン値の表を参照) で評価された生存関数、およびその信頼区間の上限と下限。
- **Sur_##, LCL_Sur_##, UCL_Sur_##, ...。** 「作図」タブで指定された各予測値のパターンで評価された生存関数、およびその信頼区間の上限と下限。パターンを番号 ## と一致させるには、出力のパターン値の表を参照してください。
- **Haz_0, LCL_Haz_0, UCL_Haz_0。** ベースライン累積ハザード関数、およびその信頼区間の上限と下限。
- **Haz_R, LCL_Haz_R, UCL_Haz_R。** 「参照」パターン (出力のパターン値の表を参照) で評価された累積ハザード関数、およびその信頼区間の上限と下限。
- **Haz_##, LCL_Haz_##, UCL_Haz_##, ...。** 「作図」タブで指定された各予測値のパターンで評価された累積ハザード関数、およびその信頼区間の上限と下限。パターンを番号 ## と一致させるには、出力のパターン値の表を参照してください。

XML としてモデルをエクスポート。 パラメーター推定値およびベースライン生存関数を含む、生存関数の予測に必要なすべての情報を XML (PMML) 形式で保存します。このモデル・ファイルを使用して、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。

オプション

推定。 このコントロールでは、回帰係数の推定の基準を指定します。

- **最大反復回数。** アルゴリズムで実行される反復の最大回数。負でない整数を指定します。
- **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 倍に縮小されます。正の整数を指定します。
- **パラメーター推定値の変化に基づいて反復回数を制限。** これを選択した場合、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よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は正の値である必要があります。
- **対数尤度の変化に基づいて反復回数を制限。** これを選択した場合、対数尤度関数の絶対変化または相対変化が指定した値よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は正の値である必要があります。
- **反復履歴を表示。** パラメーター推定値および擬似対数尤度に関する反復の履歴が表示されます。また、パラメーター推定値および擬似対数尤度の変化の最終評価が出力されます。反復履歴表には、0 回目 (初期推定値) から数えて n 回の反復ごとのデータが出力されます。ここで、 n は増分を示す値です。反復履歴を要求した場合、 n の値にかかわらず、最後の反復の出力は常に表示されます。
- **パラメーター推定値の均衡を破る方法。** 同一の観測された失敗の時間がある場合、いずれかの方法を使用して均衡を破ります。Efron の方法の方が、計算コストが高くなります。

生存関数。 このコントロールでは、生存関数に関する計算の基準を指定します。

- **ベースライン生存関数の推定方法。** **Breslow** (あるいは、Nelson-Aalan または経験的) の方法では、観測された失敗の時間にステップがある非減少のステップ関数によってベースライン累積ハザードを推定し、 $\text{生存} = \exp(-\text{累積ハザード})$ という関係によってベースラインの生存を計算します。**Efron** の方法の方が計算コストが高いため、同順位がない場合は、Breslow の方法に変更します。**積極限法**では、非増加の右連続関数によってベースラインの生存を推定します。モデルに予測変数がない場合は、この方法を Kaplan-Meier の推定に変更します。

- **生存関数の信頼区間。** 信頼区間は、元の単位を使用、対数変換、ログ・マイナス・ログ変換の 3 つの方法で計算できます。信頼区間の範囲が 0 から 1 の間になることが保証されるのはログ・マイナス・ログ変換のみですが、一般的には対数変換で「最適」に実行されると考えられます。

ユーザー欠損値。 すべての変数は、分析の対象となるケースに対して有効な値を取る必要があります。このコントロールでは、カテゴリー・モデル (因子変数、イベント変数、ストラータ変数、および部分母集団変数を含む) およびサンプリングのデザイン変数の中で、ユーザー欠損値を有効な値として扱うかどうかを決定できます。

信頼区間 (%)。 これは、係数の推定値、指数化された係数の推定値、生存関数の推定値、および累積ハザード関数の推定値に使用される信頼区間レベルです。0 以上 100 未満の値を指定します。

CSCOXREG コマンドの追加機能

コマンド言語を使用すると、次の操作も実行できます。

- カスタム仮説検定を実行する (CUSTOM サブコマンドおよび /PRINT LMATRIX を使用)。
- 許容度の指定 (/CRITERIA SINGULAR を使用)。
- 一般推定可能関数テーブル (/PRINT GEF を使用)。
- 複数の予測値のパターン (複数の PATTERN サブコマンドを使用)。
- ルート名の指定時の保存された変数の最大数 (SAVE サブコマンドを使用)。ダイアログでは、CSCOXREG のデフォルトである 25 個の変数を使用できます。

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。

特記事項

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

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

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

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

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

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

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

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

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

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119

Armonk, NY 10504-1785

US

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

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

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

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

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

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

著作権使用許諾:

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

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

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

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

商標

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

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

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

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

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

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

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

索引

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

[ア行]

応答確率
コンプレックス・サンプルの順序回帰 44
オッズ比
コンプレックス・サンプルのクロス集計表 25
コンプレックス・サンプルの順序回帰 46
コンプレックス・サンプルのロジスティック回帰 40
重み付けのない度数
コンプレックス・サンプルの記述統計 23
コンプレックス・サンプルのクロス集計表 25
コンプレックス・サンプルの度数分布表 19
コンプレックス・サンプルの比率分析 29

[カ行]

カイ 2 乗
コンプレックス・サンプル 33, 39, 45
コンプレックス・サンプルの Cox 回帰 53
擬似 R² 乗統計量
コンプレックス・サンプルの順序回帰 44
コンプレックス・サンプルのロジスティック回帰 38
期待値
コンプレックス・サンプルのクロス集計表 25
逆 Hermert 対比
コンプレックス・サンプルの一般線型モデル 33
行パーセント
コンプレックス・サンプルのクロス集計表 25
クラスター
サンプル・ウィザード 5
分析準備ウィザード 13

計画ファイル 2
系統的な抽出
サンプル・ウィザード 6
欠損値
コンプレックス・サンプル 20, 26
コンプレックス・サンプルの一般線型モデル 34
コンプレックス・サンプルの記述統計 24
コンプレックス・サンプルの順序回帰 47
コンプレックス・サンプルの比率分析 30
コンプレックス・サンプルのロジスティック回帰 40
合計
コンプレックス・サンプルの記述統計 23
コンプレックス・サンプリング
サンプル計画 5
分析計画 13
コンプレックス・サンプル
オプション 20, 24, 27, 30
仮説の検定 33, 39, 45
欠損値 20, 26
コンプレックス・サンプルの Cox 回帰
イベントの定義 50
オプション 56
仮説の検定 53
サブグループ 51
時間依存の予測変数 51
統計 52
日付変数と時間変数 49
プロット 53
変数の保存 54
モデル 51
モデルのエクスポート 55
予測 50
Kaplan-Meier 分析 49
コンプレックス・サンプルの一般線型モデル 31
オプション 34
コマンドの追加機能 34
推定平均 33
統計 32
変数の保存 34
モデル 31
コンプレックス・サンプルの記述統計 23
欠損値 24
統計 23

コンプレックス・サンプルのクロス集計表 25
統計 25
コンプレックス・サンプルの順序回帰 43
応答確率 44
オッズ比 46
オプション 47
統計 44
変数の保存 46
モデル 44
コンプレックス・サンプルの度数分布表 19
統計 19
コンプレックス・サンプルの比率分析 29
欠損値 30
統計 29
コンプレックス・サンプルのロジスティック回帰 37
オッズ比 40
オプション 40
コマンドの追加機能 41
参照カテゴリー 37
統計 38
変数の保存 40
モデル 38

[サ行]

最小有意差
コンプレックス・サンプル 33, 39, 45
コンプレックス・サンプルの Cox 回帰 53
サイズの測定
サンプル・ウィザード 6
最大対数尤度比残差
コンプレックス・サンプルの Cox 回帰 54
残差
コンプレックス・サンプルの一般線型モデル 34
コンプレックス・サンプルのクロス集計表 25
参照カテゴリー
コンプレックス・サンプルの一般線型モデル 33
コンプレックス・サンプルのロジスティック回帰 37
サンプリング
コンプレックス・デザイン 5
サンプリングの推定
分析準備ウィザード 14

サンプル重み付けの入力
 サンプル・ウィザード 5
 サンプル計画 5
 サンプルの重み付け
 サンプル・ウィザード 8
 分析準備ウィザード 13
 サンプル率
 サンプル・ウィザード 8
 サンプル・サイズ
 サンプル・ウィザード 7, 8
 サンプル・デザインの情報
 コンプレックス・サンプルの Cox 回
 帰 52
 時間依存の予測変数
 コンプレックス・サンプルの Cox 回
 帰 51
 集計残差
 コンプレックス・サンプルの Cox 回
 帰 54
 自由度
 コンプレックス・サンプル 33, 39, 45
 コンプレックス・サンプルの Cox 回
 帰 53
 信頼区間
 コンプレックス・サンプルの一般線型
 モデル 32, 34
 コンプレックス・サンプルの記述統計
 23
 コンプレックス・サンプルのクロス集
 計表 25
 コンプレックス・サンプルの順序回帰
 44
 コンプレックス・サンプルの度数分布
 表 19
 コンプレックス・サンプルの比率分析
 29
 コンプレックス・サンプルのロジステ
 イック回帰 38
 信頼度レベル
 コンプレックス・サンプルの順序回帰
 47
 コンプレックス・サンプルのロジステ
 イック回帰 40
 推定周辺平均
 コンプレックス・サンプルの一般線型
 モデル 33
 スコア残差
 コンプレックス・サンプルの Cox 回
 帰 54
 全平均の対比
 コンプレックス・サンプルの一般線型
 モデル 33
 層化
 サンプル・ウィザード 5
 分析準備ウィザード 13

相対リスク
 コンプレックス・サンプルのクロス集
 計表 25

[夕行]

対数尤度収束
 コンプレックス・サンプルの順序回帰
 47
 コンプレックス・サンプルのロジステ
 イック回帰 40
 対比
 コンプレックス・サンプルの一般線型
 モデル 33
 多項式の対比
 コンプレックス・サンプルの一般線型
 モデル 33
 段階 2 分
 コンプレックス・サンプルの順序回帰
 47
 コンプレックス・サンプルのロジステ
 イック回帰 40
 単純対比
 コンプレックス・サンプルの一般線型
 モデル 33
 単純な無作為抽出
 サンプル・ウィザード 6
 逐次 Bonferroni 補正
 コンプレックス・サンプル 33, 39, 45
 コンプレックス・サンプルの Cox 回
 帰 53
 逐次 Sidak 補正
 コンプレックス・サンプル 33, 39, 45
 コンプレックス・サンプルの Cox 回
 帰 53
 逐次抽出
 サンプル・ウィザード 6
 抽出方法
 サンプル・ウィザード 6
 調整済み F 統計量
 コンプレックス・サンプル 33, 39, 45
 コンプレックス・サンプルの Cox 回
 帰 53
 調整済みカイ 2 乗
 コンプレックス・サンプル 33, 39, 45
 コンプレックス・サンプルの Cox 回
 帰 53
 調整済み残差
 コンプレックス・サンプルのクロス集
 計表 25
 デザイン効果
 コンプレックス・サンプルの Cox 回
 帰 52
 コンプレックス・サンプルの一般線型
 モデル 32

デザイン効果 (続き)
 コンプレックス・サンプルの記述統計
 23
 コンプレックス・サンプルのクロス集
 計表 25
 コンプレックス・サンプルの順序回帰
 44
 コンプレックス・サンプルの度数分布
 表 19
 コンプレックス・サンプルの比率分析
 29
 コンプレックス・サンプルのロジステ
 イック回帰 38
 デザイン効果の平方根
 コンプレックス・サンプルの Cox 回
 帰 52
 コンプレックス・サンプルの一般線型
 モデル 32
 コンプレックス・サンプルの記述統計
 23
 コンプレックス・サンプルのクロス集
 計表 25
 コンプレックス・サンプルの順序回帰
 44
 コンプレックス・サンプルの度数分布
 表 19
 コンプレックス・サンプルの比率分析
 29
 コンプレックス・サンプルのロジステ
 イック回帰 38

[ハ行]

パラメーター収束
 コンプレックス・サンプルの順序回帰
 47
 コンプレックス・サンプルのロジステ
 イック回帰 40
 パラメーター推定値
 コンプレックス・サンプルの Cox 回
 帰 52
 コンプレックス・サンプルの一般線型
 モデル 32
 コンプレックス・サンプルの順序回帰
 44
 コンプレックス・サンプルのロジステ
 イック回帰 38
 パラメーター推定値の共分散
 コンプレックス・サンプルの一般線型
 モデル 32
 コンプレックス・サンプルの順序回帰
 44
 コンプレックス・サンプルのロジステ
 イック回帰 38

パラメーター推定値の相関
コンプレックス・サンプルの一般線型
モデル 32
コンプレックス・サンプルの順序回帰
44
コンプレックス・サンプルのロジステ
ィック回帰 38
反復
コンプレックス・サンプルの順序回帰
47
コンプレックス・サンプルのロジステ
ィック回帰 40
反復測定
コンプレックス・サンプルの一般線型
モデル 33
反復履歴
コンプレックス・サンプルの順序回帰
47
コンプレックス・サンプルのロジステ
ィック回帰 40
標準誤差
コンプレックス・サンプルの一般線型
モデル 32
コンプレックス・サンプルの記述統計
23
コンプレックス・サンプルのクロス集
計表 25
コンプレックス・サンプルの順序回帰
44
コンプレックス・サンプルの度数分布
表 19
コンプレックス・サンプルの比率分析
29
コンプレックス・サンプルのロジステ
ィック回帰 38
表パーセント
コンプレックス・サンプルのクロス集
計表 25
コンプレックス・サンプルの度数分布
表 19
比例ハザード検定
コンプレックス・サンプルの Cox 回
帰 52
部分母集団
コンプレックス・サンプルの Cox 回
帰 51
分析計画 13
分離
コンプレックス・サンプルの順序回帰
47
コンプレックス・サンプルのロジステ
ィック回帰 40
分類テーブル
コンプレックス・サンプルの順序回帰
44

分類テーブル (続き)
コンプレックス・サンプルのロジステ
ィック回帰 38
ベースライン・ストラータ
コンプレックス・サンプルの Cox 回
帰 51
平均
コンプレックス・サンプルの記述統計
23
平行線の検定
コンプレックス・サンプルの順序回帰
44
変動係数 (COV)
コンプレックス・サンプルの記述統計
23
コンプレックス・サンプルのクロス集
計表 25
コンプレックス・サンプルの度数分布
表 19
コンプレックス・サンプルの比率分析
29
包含確率
サンプル・ウィザード 8
母集団のサイズ
コンプレックス・サンプルの記述統計
23
コンプレックス・サンプルのクロス集
計表 25
コンプレックス・サンプルの度数分布
表 19
コンプレックス・サンプルの比率分析
29
サンプル・ウィザード 8

[ヤ行]

予測値
コンプレックス・サンプルの一般線型
モデル 34
予測確率
コンプレックス・サンプルの順序回帰
46
コンプレックス・サンプルのロジステ
ィック回帰 40
予測カテゴリー
コンプレックス・サンプルの順序回帰
46
コンプレックス・サンプルのロジステ
ィック回帰 40

[ラ行]

リスクの差分
コンプレックス・サンプルのクロス集
計表 25

累積値
コンプレックス・サンプルの度数分布
表 19
累積確率
コンプレックス・サンプルの順序回帰
46
列パーセント
コンプレックス・サンプルのクロス集
計表 25

B

Bonferroni
コンプレックス・サンプル 33, 39, 45
コンプレックス・サンプルの Cox 回
帰 53
Breslow の推定方法
コンプレックス・サンプルの Cox 回
帰 56
Brewer 抽出法
サンプル・ウィザード 6

C

Cox-Snell 残差
コンプレックス・サンプルの Cox 回
帰 54

E

Efron の推定方法
コンプレックス・サンプルの Cox 回
帰 56

F

F 統計量
コンプレックス・サンプル 33, 39, 45
コンプレックス・サンプルの Cox 回
帰 53
Fisher スコア法(F)
コンプレックス・サンプルの順序回帰
47

H

Helmert 対比
コンプレックス・サンプルの一般線型
モデル 33

M

Martingale 残差

コンプレックス・サンプルの Cox 回
帰 54

Murthy 抽出法

サンプル・ウィザード 6

N

Newton-Raphson 法

コンプレックス・サンプルの順序回帰
47

P

PPS 抽出

サンプル・ウィザード 6

R

R2 統計量

コンプレックス・サンプルの一般線型
モデル 32

S

Sampford 抽出法

サンプル・ウィザード 6

Schoenfeld の偏残差

コンプレックス・サンプルの Cox 回
帰 54

Sidak 補正

コンプレックス・サンプル 33, 39, 45
コンプレックス・サンプルの Cox 回
帰 53

T

t 検定

コンプレックス・サンプルの一般線型
モデル 32

コンプレックス・サンプルの順序回帰
44

コンプレックス・サンプルのロジステ
ィック回帰 38



Printed in Japan