



---

## ハイライト

- 層別化抽出によりサンプルの精度向上や代表的なサンプルの確保が可能
  - クラスター・サンプリングによりサンプリング単位のグループを選択可能
  - 多段サンプリングにより、第1段階のサンプルを選択し、次に第2段階のサンプルを作成可能
- 

# IBM SPSS Complex Samples

## コンプレックス・サンプルの統計を正確に計算

サンプル調査を実施するときには、コンプレックス・サンプル・データの正確な推定値を生成する専用の統計パッケージを使用します。IBM SPSS Complex Samples は、コンプレックス・サンプル・デザインの統計情報とその標準誤差を正確かつ簡単に計算できる専用の統計手法を提供します。これは次の作業に適用できます。

- 調査研究 – 調査データについての記述的および推測的な統計を入手
- 市場調査 – 顧客満足度データの分析
- 保健調査 – 健康、栄養摂取、アルコール摂取、交通事故死亡者数などの公衆衛生上の課題に関する大規模な公共利用データ・セットの分析
- 社会科学 – 公共調査のデータ・セットに対する2次的研究の実施
- 世論調査 – 政策問題に対する意見に見られる特徴の分析

SPSS Complex Samples には、コンプレックス・サンプルを扱う上で必要なすべての機能が用意されています。これには、次のようなものがあります。

- スキームのデザインやサンプルの抽出などの手順を段階を踏んで案内する直観的なサンプリング・ウィザード
- 米国疾病対策センター (CDC) による国民健康調査データなどのサンプリング済み公共データ・セットの前処理で効果的な、使いやすい分析準備ウィザード
- コンプレックス・サンプル一般線形モデル (CSGLM) による数値結果予測
- コンプレックス・サンプル順序回帰 (CSORDINAL) による順序結果予測
- コンプレックス・サンプル・カテゴリー回帰 (CSLOGISTIC) によるカテゴリーの結果予測
- コンプレックス・サンプル Cox 回帰 (CSCOXREG) によるイベント発生までの時間予測



SPSS Complex Samples を使用することで、計画段階から分析段階全体にわたるサンプリングまで、正確で信頼性の高い結果を容易に入手できます。複数段階のデザインから得られたデータを分析する場合、SPSS Complex Samples では最大で 3 つの状態が考慮されるので、より正確な分析結果を得ることができます。

SPSS Complex Samples では、デザインの影響を評価できるばかりでなく、部分母集団の評価で他の部分母集団も考慮されるので、より正確なデータを生成できます。

SPSS Complex Samples では、次のタイプのサンプル・デザイン情報を使用できます。

- ・ 層別化サンプリング – 調査対象の母集団のサブグループ内でサンプリングすることで、サンプルの精度が向上し、主要なグループの代表的サンプルとして使用できます。たとえば、特定の人数の男性または女性、特定のカテゴリーの仕事に従事する人々、特定の年齢集団の人々をサブグループとすることができます。
- ・ クラスター・サンプリング – サンプリング単位のグループであるクラスターを調査に合わせて選択します。クラスターには、学生、患者、市民をサンプリング単位として学校、病院、地域などを含めることができます。クラスター化によって、多くの場合は調査の費用効率が向上します。
- ・ 多段サンプリング – 初期段階 (第 1 段階) のサンプルは母集団の構成要素のグループに基づいて選択し、続いて第 1 段階で選択した単位ごとに下位サンプルを採取することで第 2 段階を作成します。この選択を繰り返すことでより高い段階のサンプルを選択できます。

### より確実に得られる結果

調査担当者は、得られた結果に高い信頼度があることを期待しています。これまでの統計ソフトウェアの大半は、データが単純なランダム・サンプリングで得られることを前提としています。しかし、単純なランダム・サンプリングは、ほとんどの大規模な調査には不向きで、費用効率も高くありません。

このようなサンプル・データを従来の統計手法で分析すると、不正確な結果になる危険性があります。たとえば、統計量の推定標準誤差が小さくなりすぎるため、精度について誤った印象を持ちやすくなります。SPSS Complex Samples では、調査分析にサンプル・デザインが組み込まれていることから、コンプレックス・サンプル・データで測定した母集団について、統計的により有効な推定が実現します。

### 効率的で容易な作業

コンプレックス・サンプルの調査結果を容易に把握し、扱うことができるのは SPSS Complex Samples だけです。わかりやすいインターフェースでデータを分析し、結果を解釈できます。作業が完了したところでデータ・セットを公開し、それにサンプリング計画や分析計画を追加できます。

それぞれの計画はテンプレートとして利用でき、作成時の決定をすべて収めておくことができます。これにより、時間の短縮と精度の向上が望めます。また、その計画とデータを使用して、結果の複製や処理対象外のデータの利用を図る他のユーザーにも、同様の効果が期待できます。

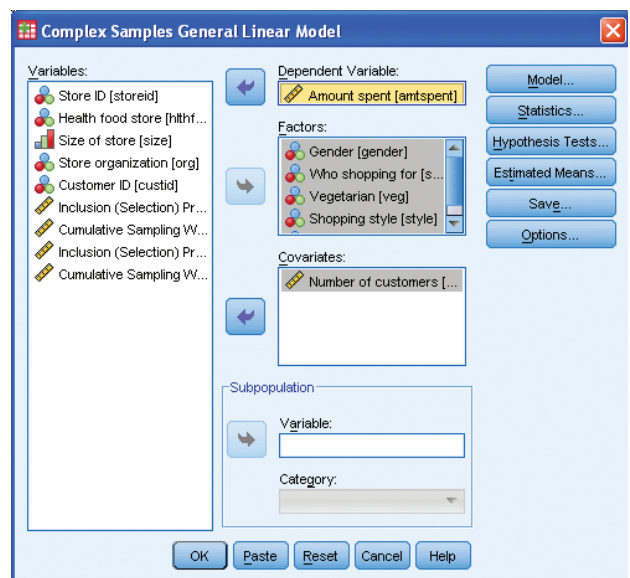
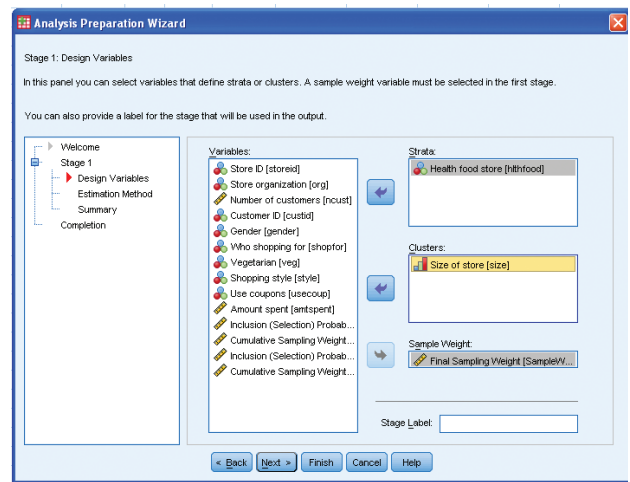


図 1: 食料雑貨品店では、顧客の性別に合わせた管理やサンプル・デザインへの組み込みのために、顧客の買い物の頻度が、買物の総額と関係するかどうかを判断する必要があります。店舗では、最初に分析準備ウィザード (上の図) でサンプル・デザインを指定します。次に、コンプレックス・サンプルの一般線型モデル (下の図) でモデルを設定します。

■ Accurate analysis of survey data

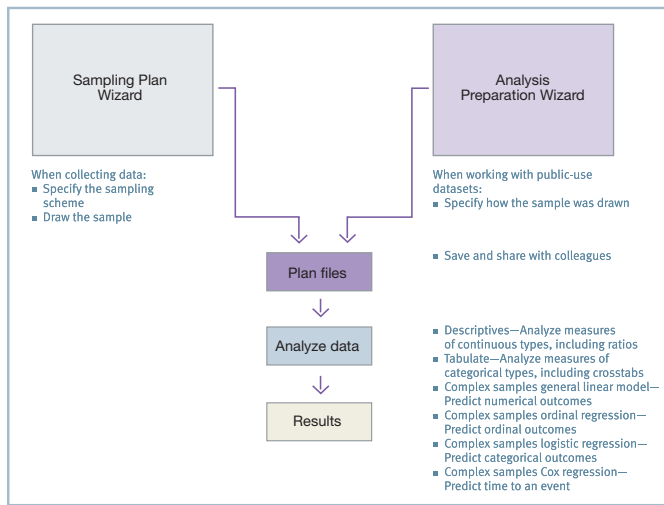


図 2: SPSS Complex Samples では調査データの正確な分析も容易です。最初にいずれかのウィザードから作業を開始します (データ・ソースに応じて適切なウィザードを選択します)。次に対話式インターフェースで、計画を作成し、データを分析して、結果を解釈します。

SPSS Complex Samples での作業はウィザードで開始します。このウィザードでは、考慮が必要なさまざまな要素の入力を求められます。独自のサンプルを作成する場合は、サンプリング・ウィザードを使用して、サンプリング方法を定義します。CDC が提供するデータ・セットなどのサンプリング済みの公共データ・セットを使用する場合は、分析準備ウィザードを使用してサンプルの定義方法や標準誤差の推定方法を指定できます。サンプルを作成するか、標準誤差を指定すると、計画を作成してデータを分析し、結果を生成できます (上記のワークフローの図を参照)。

SPSS Complex Samples の習得は容易で、短時間で活用できるようになります。このソフトウェアでデータを使用する方法については、オンライン・ヘルプ・システム、対話型のケース・スタディー、またはオンライン・チュートリアルを参照してください。SPSS Complex Samples では、次のような成果が得られます。

- ・ 合計、平均、比率などの統計量の正確な推定値の取得
- ・ これらの統計量に存在する標準誤差の算出
- ・ 正確な信頼区間および仮説検証の生成
- ・ 数値結果の予測
- ・ 順序結果の予測
- ・ カテゴリーの結果の予測
- ・ イベント発生までの時間の予測

## 分析資産の共有

分析資産を共有し、効率的に提供するには、社内外のコンプライアンス要件に合致する方法でそれらの資産を保護し、より多くのビジネス・ユーザーが閲覧および利用できるような分析結果を公表します。また、IBM SPSS Collaboration and Deployment Services による IBM SPSS Complex Samples の機能拡張を検討します。[ibm.com/spss/cds](http://ibm.com/spss/cds)

SPSS Complex Samples は、クライアント専用ソフトウェアとしてインストールできますが、より高度なパフォーマンスと拡張性が必要な場合は、サーバー版をご利用いただけます。

## 機能

### コンプレックス・サンプル計画 (CSPLAN)

この手順により、SPSS Complex Samples モジュールにある他の関連する手順で使用するコンプレックス・サンプルのデザインや分析仕様を作成するための、サンプリングの枠組みを指定する共通の機能が得られます。CSPLAN は、実際にはサンプルも分析データも抽出しません。事例をサンプリングするには、CSPLAN で作成したサンプル・デザインを CSSELECT 手順の入力として使用します (次のページで説明)。サンプル・データを分析するには、CSPLAN で作成した分析デザインを CSDDES、CRIPPTIVES、CSTABULATE、CSGLM、CSLOGISTIC、または CSORDINAL のいずれかの手順の入力として使用します (次のページの説明を参照)。

- ・ サンプル・デザインの作成: アクティブなファイルからのサンプリング単位の取得に使用
- ・ 分析デザインの作成: コンプレックス・サンプルの分析に使用
- ・ サンプル・デザインを作成するときに、この手順で適切な分析デザインを計画ファイルに自動的に保存。計画ファイルはサンプルのデザイン向けに作成されているので、サンプルの選択と分析の両方に使用可能です。
- ・ サンプル・デザインまたは分析デザインの表示
- ・ 外部ファイルにある計画の指定
- ・ PLANVARS サブコマンドを使用して、サンプルを抽出する際に計画単位で作成される変数を指定。また、この変数を選択プロセスや推定プロセスの入力として使用します。
  - 推定プロセスで IBM SPSS Complex Samples の分析手法で使用する単位ごとに、サンプルに対する最終的な重み付けを指定
  - CSSELECT 手順でサンプル・デザインを実行したときにサンプル全体に対して生成される重み付けを表示
  - 複数段階のデザインでサンプルに対する最終的な重み付けを計算する際に使用する重み付けを選択

- CSPLAN 手順の出力を PRINT サブコマンドで制御
  - デザインの段階ごとの仕様を反映した内容を持つ計画仕様の要約を表示
  - MATRIX の仕様一覧を表示
- DESIGN サブコマンドを使用した、デザインの通知段階。このサブコマンドを使用して、層別化変数とクラスター変数の定義や、特定の段階を示す説明的なラベルの作成なども可能です。
- METHOD サブコマンドによるサンプル抽出方法の指定。単純なランダム・サンプリングや系統的なランダム・サンプリングなどのさまざまな等確率手法と不等確率手法から選択できます。サイズに比例した確率でサンプリングする手法 (PPS) も利用できます。母集団からの単位の抽出では、置換を伴う方法 (WR) と置換を伴わない方法 (WOR) が可能です。
  - SIMPLE\_WOR: 等確率で単位を選択。置換せずに単位を抽出します。
  - SIMPLE\_WR: 等確率で単位を選択。置換して単位を抽出します。
  - SIMPLE\_SYSTEMATIC: サンプリング枠全体またはサンプリング層全体から固定間隔で単位を選択。ランダムな開始点は最初の間隔の中で選択します。
  - SIMPLE\_CHROMY: 等確率で順番に単位を選択。置換せずに単位を抽出します
  - PPS\_WOR: サイズに比例した確率で単位を選択。置換せずに単位を抽出します。
  - PPS\_WR: サイズに比例した確率で単位を選択。置換して単位を抽出します。
  - PPS\_SYSTEMATIC: サイズに比例した確率で、系統的なランダム・サンプリングによって単位を選択。置換せずに単位を抽出します。
  - PPS\_CHROMY: サイズに比例した確率で順番に単位を選択。置換せずに単位を抽出します。
  - PPS\_BREWER: サイズに比例した確率で層ごとに 2 つの単位を選択。置換せずに単位を抽出します。
  - PPS\_MURTHY: サイズに比例した確率で層ごとに 2 つの単位を選択。置換せずに単位を抽出します。
  - PPS\_SAMPFORD: Brewer の手法を拡張し、サイズに比例した確率で層ごとに 3 つ以上の単位を選択。置換せずに単位を抽出します。
  - 抽出する単位の個数やパーセントの制御: デザインの段階ごとに設定します。段階ごとのサンプリングの重み付けのような、サンプル・デザイン実行の際に作成する出力変数を選択することもできます。
- 推定手法: 置換を伴う方法、等確率で第 1 段階での置換を伴わない方法、および不等確率で置換を伴わない方法があります。
- 単純なランダム・サンプリングで分散を推定する際に、有限母集団修正を扱うかどうかを選択可能
- 置換を伴わない不等確率推定: 第 1 段階でのみ要求可能
- 変数指定: 推定プロセスの入力で使用する変数の指定。これには、サンプルに対する全体的な重み付けや包含確率などがあります。
- SIZE サブコマンドを使用して、現在の段階で抽出するサンプリング単位の数を指定
- 現在の段階で抽出する単位のパーセンテージを指定。たとえば、RATE サブコマンドを使用して抽出率を指定
- RATE を指定する際に、抽出する単位の最小個数を指定。これは四捨五入によって特定の層の抽出率が極端に低くなる場合に便利です。
- RATE を指定する際に、抽出する単位の最大個数を指定。四捨五入によって特定の層の抽出率が目的の値より大きくなる場合に便利です。
- PPS デザイン で母集団単位のサイズを指定。サイズを設定した変数を指定するか。CSSELECT 手順でサンプル枠をスキャンする時にサイズが決まるように要求
- STAGEVARS サブコマンドを使用して、サンプル・デザインを実行する際に段階ごとのサンプル情報変数を取得。取得できる情報は次のとおりです。
  - 段階ごとの包含 (選択) 確率を使用して、特定の段階で母集団から抽出する単位の割合
  - 指定の段階の累積抽出重み付けを使用するそれ以前の段階に関する情報
  - 指定の段階で選択した単位に重複したインデックスを割り当てることで、置換を伴うサンプルを実行するときに 2 回以上選択された単位を一意的に特定
  - 指定の段階の母集団のサイズ
  - 指定の段階で抽出する単位の個数
  - 段階ごとの抽出率
  - 指定の段階でサンプリングに割り当てる重み付け

- ESTIMATOR サブコマンドを使用して、現在の段階で使用する推定手法を選択。指定できる方法は次のとおりです。
  - 等確率選択 (置換なし)
  - 不等確率選択 (置換なし)
  - 選択 (置換あり)
- POPSIZE サブコマンドを使用して、サンプル要素ごとの母集団サイズを指定
- INCLPROB サブコマンドを使用して、指定の段階で母集団から抽出する単位の割合を指定

### コンプレックス・サンプルの選択 (CSSELECT)

CSSELECT は、確率に基づいたコンプレックス・サンプルを母集団から選択します。単位は、CSPLAN 手順で作成したサンプル・デザインに従って選択されます。

- CRITERIA サブコマンドを使用して、実行の範囲を制御し、シード値を指定
- CLASSMISSING サブコマンドを使用して、分類 (層化またはクラスター化) 変数のうち、ユーザーが指定していない値を有効な値として扱うかどうかを指定
- 最新のメルセンヌ・ツイスタ (乱数生成アルゴリズム) を使用して、サンプルを選択
- DATA サブコマンドを使用して、入力ファイルおよび出力ファイルに関する一般的なオプションを指定
  - サンプル重み付け変数や計画ファイルで要求されている段階ごとの出力変数 (包含確率など) を CSSELECT 手順で記述するときに、既存の変数の名前を変更するかどうかを指定
- 指定した変数を保持するオプションまたは除外するオプションを使用して、抽出した単位を外部ファイルに書き込み
- 計画ファイルでサンプリング手法として PPS\_WR を指定している場合に、第 1 段階の結合包含確率を自動的に外部ファイルに保存
- 選択する単位の特性を規定する規則を収めたファイルを作成するかどうかを選択
- PRINT サブコマンドを使用して、出力の表示を制御
  - 選択したケースが層全体にわたって示す分布を要約。計画段階ごとに情報のレポートが得られます。
  - ケース処理の要約を作成

### コンプレックス・サンプルの記述 (CSDESCRIPTIVES)

CSDESCRIPTIVES は、コンプレックス・サンプリング手法で抽出したサンプルの平均、合計、比率を推定し、それらの標準誤差、デザイン効果、信頼区間、仮説検定を算出します。この手順では、サンプルの選択に使用したサンプル・デザインを考慮して変動を推定します。たとえば、等確率、PPS 手法、WR サンプリング手順、WOR サンプリング手順などが考慮されます。また、必要に応じて、部分母集団を分析します。

- CSPLAN 手順で記述された計画ファイル名を指定。この計画ファイルは、PLAN サブコマンドによる分析デザイン仕様を収めています。
- 結合包含確率ファイルの名前を指定
- SUMMARY サブコマンドを使用して、MEAN サブコマンドと SUM サブコマンドで使用する分析変数を指定
- MEAN サブコマンドと SUM サブコマンドを使用して、SUMMARY サブコマンドで指定した変数の平均と合計を推定することを要求
  - TTEST キーワードを使用して、母集団の平均または合計に対して T 検定を実施し、帰無仮説値を計算することを要求。SUBPOP サブコマンドで部分母集団を定義している場合は、母集団全体の検定だけでなく、部分母集団ごとの検定でも帰無仮説値が使用されます。
- RATIO サブコマンドを使用して、SUMMARY サブコマンドで指定した変数の比率を推定することを要求
  - TTEST キーワードを使用し、母集団の比率に対して T 検定を実施して、帰無仮説値を求めることを要求。
- 平均、合計、または比率の推定に次のような構文を関連付け
  - 平均、合計、または比率の推定で使用する、データ・セットに存在する有効な観測値の個数
  - 平均、合計、または比率の推定で使用する母集団のサイズ
  - 平均、合計、または比率の推定で発生する標準誤差
  - 変動係数
  - デザイン効果
  - デザイン効果の平方根
  - 信頼区間
- SUBPOP サブコマンドを使用して、分析の対象となる部分母集団を指定
  - すべての部分母集団の結果を 1 つの同じ表または別々の表に表示

- ・ 欠落データの扱いを指定
  - 統計量の計算に使用する分析変数のすべての有効データに基づいて統計量を計算。指定した両方の変数でデータが有効なすべてのケースを使用して比率を計算します。さまざまなサンプル・サイズに基づいて、さまざまな変数の統計量を計算できます。
  - 統計量の計算で、すべての分析変数でデータが有効なケースのみを使用。さまざまな変数の統計を、常に同じサンプル・サイズに基づいて計算します。
  - 層、クラスター、および部分母集団の変数で、ユーザーが指定していない値を除外
  - 層、クラスター、および部分母集団の変数で、ユーザーが指定していない値を対象とする。これらの変数でユーザーが指定していない値を有効なデータとして扱います。

### コンプレックス・サンプル表 (CSTABULATE)

CSTABULATE は、コンプレックス抽出手法で抽出されたサンプルの一元度数分布表または二元クロス集計表を表示し、さらにそれに関連する標準誤差、デザイン効果、信頼区間、および仮説検定を表示します。この手順では、サンプルの選択に使用したサンプル・デザインを考慮して変動を推定します。たとえば、等確率、PPS 手法、WR サンプルング手順、WOR サンプルング手順などが考慮されます。必要に応じて、部分母集団の表を作成することもできます。

- ・ CSPLAN 手順で記述した XML ファイル を指定。このファイルは、PLAN サブコマンドを使用して分析デザインを指定しているものです。
- ・ 結合包含確率ファイルの名前を指定
- ・ 表の中で以下の統計量を使用
  - 母集団サイズ: 表の各セルおよび周辺の行と列で母集団サイズを推定
    - ・ 標準誤差: 母集団サイズの推定ごとに標準誤差を計算
- ・ 行パーセンテージと列パーセンテージ: ある行または列に存在する各セルの母集団サイズの推定を、その行または列の母集団サイズ推定に対するパーセンテージで表現。この機能は二元クロス集計表で使用できます。
  - 表パーセンテージ: 表に存在する各セルの母集団サイズの推定を、その表の母集団サイズ推定に対するパーセンテージで表現
  - 変動係数
  - デザイン効果
  - デザイン効果の平方根
  - 信頼区間: 信頼区間を 0 から 100 の値で指定

- 重み付けのないカウント: 母集団サイズの推定ごとに、重み付けのないカウントをデータ・セットの有効観測値の個数として使用
- 累積母集団サイズ推定: 累積母集団サイズ推定を使用。この機能は一元度数分布表でのみ使用できます。
- 累積パーセンテージ: 母集団サイズ推定に対応する累積パーセンテージを使用。この機能は一元度数分布表でのみ使用できます。
- 母集団サイズ推定の期待値: クロス集計表で、2 つの変数の各セルの母集団サイズ推定が統計的に独立している場合、母集団サイズ推定の期待値を使用。この機能は二元クロス集計表で使用できます。
- 残差: 集団サイズ推定の観測値と期待値の差をセルごとに表示。この機能は二元クロス集計表で使用できます。
- Pearson 残差: この機能は二元クロス集計表でのみ使用できます。
- 調整済み Pearson 残差: この機能は二元ロス集計表でのみ使用できます。
- ・ 表全体に対して、次の統計量と検定を使用
  - 等しいセル比率の検定
  - 独立性の検定
  - オッズ比
  - 相対リスク
  - リスクの差分
- ・ SUBPOP サブコマンドを使用して、分析の対象となる部分母集団を指定
  - すべての部分母集団の結果を 1 つの同じ表または別々の表に表示 欠落データの扱いを指定
  - 表の作成に使用する作表変数の有効なデータすべてに基づいて表を作成。さまざまなサンプル・サイズに基づいて、さまざまな変数の表を作成できます。
  - 表を作成する際に、すべての作表変数でデータが有効なケースのみを使用。異なる変数の表でも、必ず同じサンプル・サイズに基づきます。
  - 層、クラスター、および部分母集団の変数で、ユーザーが指定していない値を除外
  - 層、クラスター、および部分母集団の変数で、ユーザーが指定していない値を対象とする。これらの変数でユーザーが指定していない値を有効なデータとして扱います。

## コンプレックス・サンプル一般線型モデル (CSGLM)

この手順では、コンプレックス・サンプル手法で抽出したサンプルに使用する線型回帰、分散分析 (ANOVA) モデル、および共分散分析 (ANCOVA) モデルを構築できます。この手順では、サンプルの選択に使用したサンプル・デザインを考慮して変動を推定します。たとえば、等確率、PPS 手法、WR サンプルング手順、WOR サンプルング手順などが考慮されます。また、必要に応じて部分母集団の分析も可能です。

- モデル
  - 主効果
  - すべての n 元相互作用
  - 完全交差
  - ユーザー指定 (入れ子にした項など)
- 統計量
  - モデルのパラメーター: 係数推定、各係数推定の標準誤差、各係数推定の t 検定、各係数推定の信頼区間、各係数推定のデザイン効果、および各係数推定のデザイン効果の平方根
  - 従属変数と共変量の母集団平均
  - モデルの適合
  - サンプル・デザインの情報
- 仮説の検定
  - 検定の統計: Wald の F 検定、調整済み Wald の F 検定、Wald のカイ二乗検定、調整済み Wald のカイ二乗検定
  - 多重比較の調整: 最小有意差、ボンフェローニ、逐次ボンフェローニ、シダック、逐次シダック
  - サンプル抽出の自由度: サンプル・デザインに基づくか、ユーザーによる指定
- 推定平均: モデル内の因子や相互作用の推定マージナル平均を要求
  - 対比: 単純、偏差、ヘルマート、反復、または多項式
- モデル変数は使用中のファイルに保存できるほか、パラメーター行列を収めた外部ファイルにエクスポート可能
  - 変数: 予測された値と残差
  - パラメーターの共分散行列とそれに関する統計量およびパラメーターの相関行列とそれに関する統計量は、IBM SPSS Statistics データ・ファイルとしてエクスポート可能
  - パラメーター推定やパラメーター共分散行列は、XML ファイルとしてエクスポート可能

- 出力
  - サンプル・デザインの情報 (層、PSU など)
  - 回帰係数推定と t 検定
  - 従属変数、共変量、因子の要約情報
  - サンプルの要約情報 (重み付けのないカウントや母集団サイズなど)
  - パラメーター推定の信頼限界とユーザー指定の信頼水準
  - モデル効果の Wald の F 検定
  - デザイン効果
  - 複数の R<sup>2</sup> 値
  - 対比係数 (L) 行列のセット
  - 回帰係数推定の分散共分散行列
  - 二乗平均誤差
  - 回帰係数の共分散行列と相関行列
- 欠落データの扱い
  - 欠落した値をリスト単位で除外
- その他
  - ユーザー指定の共通因子 df (すべての検定統計量で p 値の計算に使用)
  - 共線性の診断
  - モデルを部分母集団に適合可能

## コンプレックス・サンプル順序回帰 (CSORDINAL)

CSORDINAL は、2 値あるいは順序多値の従属変数に対して回帰分析を実行します。ここでは、コンプレックス・サンプル手法で抽出したサンプルに対して、選択された累積リンク関数を使用します。この手順では、サンプルの選択に使用したサンプル・デザインを考慮して変動を推定します。たとえば、等確率、PPS 手法、WR サンプルング手順、WOR サンプルング手順などが考慮されます。また、必要に応じて部分母集団の分析を実行します。

- モデル
  - 主効果
  - すべての n 元相互作用
  - 完全交差
  - ユーザー指定 (入れ子にした項など)
- 統計量:
  - モデルのパラメーター: 係数推定、指数化推定、各係数推定の標準誤差、各係数推定の t 検定、各係数推定の信頼区間、各係数推定のデザイン効果、各係数推定のデザイン効果の平方根、パラメーター推定値の共分散、およびパラメーター推定値の相関行列
  - モデルの適合: 擬 R<sup>2</sup> 乗と分類テーブル

- 平行線の検定: 等しい傾きの Wald 検定、一般化モデル (傾きが等しくないモデル) のパラメーター推定、一般化モデル (傾きが等しくないモデル) のパラメーター推定の共分散
- モデル変数の要約統計量
- サンプル・デザインの情報
- 仮説の検定
  - 検定の統計: Wald の F 検定、調整済み Wald の F 検定、Wald のカイ二乗検定、調整済み Wald のカイ二乗検定
  - 多重比較の調整: 最小有意差、ボンフェローニ、逐次ボンフェローニ、シダック、逐次シダック
  - サンプル抽出の自由度: サンプル・デザインに基づくか、ユーザーによる指定
- モデル変数は使用中のファイルに保存できるほか、パラメーター行列を収めた外部ファイルにエクスポート可能
  - 変数: 予測されたカテゴリー、予測されたカテゴリーの確率、観測されたカテゴリーの確率、累積確率 (カテゴリーごとに変数 1 つ)、予測された確率 (カテゴリーごとに 1 つの変数)
  - IBM SPSS Statistics データ・ファイルとしてエクスポート: パラメーター共分散行列とその他の統計量、パラメーター相関行列とその他の統計量
  - XML としてエクスポート: パラメーター推定やパラメーター共分散行列を XML ファイルとしてエクスポート
- 3つの推定手法: ニュートン・ラフソン、フィッシャー・スコアリング、フィッシャー・スコアリングに続いてニュートン・ラフソンを実行
- モデルを指定する累積リンク関数: コーチット、補ログ・マイナス・ログ、負ログ・マイナス・ログ、プロビット
- 指定した因子や共変量の累積オッズ比。このコマンドは LOGIT リンクでのみ使用可能
- 出力
  - サンプル・デザインの情報 (層、PSUなど)
  - 従属変数、共変量、因子の要約情報
  - サンプルの要約情報 (重み付けのないカウントや母集団サイズなど)
  - パラメーター推定の信頼限界とユーザー指定の信頼水準
  - モデルの要約統計量

- モデル効果に関する Wald の f 検定、調整済み Wald の f 検定、Wald のカイ二乗検定、調整済み Wald のカイ二乗検定
- デザイン効果
- 分類テーブル
- 対比係数 (L) 行列のセット
- 回帰係数推定の分散共分散行列
- 一般推定関数テーブル
- 回帰係数の相関行列
- 欠落データの扱い
  - 欠落した値をリスト単位で除外
- その他
  - ユーザー指定の共通因子 df (すべての検定統計量で p 値の計算に使用)
  - 共線性の診断
  - 部分母集団へのモデルの適合

### コンプレックス・サンプル・ロジスティック回帰 (CSLOGISTIC)

この手順は、コンプレックス・サンプリング手法で抽出したサンプルに対し、二項ロジスティック回帰分析あるいは多項ロジスティック回帰 (MLR) 分析を実行します。CSLOGISTIC では、サンプルの選択に使用されたサンプル・デザインを考慮して変動を推定します。たとえば、等確率、PPS 手法、WR サンプリング手順、WOR サンプリング手順などが考慮されます。また、必要に応じて部分母集団の分析も可能です。

- モデル
  - 主効果
  - すべての n 元相互作用
  - 完全交差
  - ユーザー指定 (入れ子にした項など)
- 統計量
  - モデルのパラメーター: 係数推定、指数化推定、各係数推定の標準誤差、各係数推定の t 検定、各係数推定の信頼区間、各係数推定のデザイン効果、各係数推定のデザイン効果の平方根、パラメーター推定値の共分散、パラメーター推定値の相関行列
  - モデルの適合: 擬 R2 乗と分類テーブル
  - モデル変数の要約統計量
  - サンプル・デザインの情報



- 仮説の検定
  - 検定の統計: Wald の F 検定、調整済み Wald の F 検定、Wald のカイ二乗検定、調整済み Wald のカイ二乗検定
- 多重比較の調整: 最小有意差、ボンフェローニ、逐次ボンフェローニ、シダック、逐次シダック
- サンプル抽出の自由度: サンプル・デザインに基づくか、ユーザーによる指定
- モデル変数は使用中のファイルに保存できるほか、パラメーター行列を収めた外部ファイルにエクスポート可能
  - 変数: 予測されたカテゴリーと予測された確率
  - パラメーターの共分散行列とそれに関する統計量およびパラメーターの相関行列とそれに関する統計量は、IBM SPSS Statistics データ・ファイルとしてエクスポート可能
  - パラメーター推定やパラメーター共分散行列は、XML ファイルとしてエクスポート可能
- 出力
  - サンプル・デザインの情報 (層、PSU など)
  - 従属変数、共変量、因子の要約情報
  - サンプルの要約情報 (重み付けのないカウントや母集団サイズなど)
  - パラメーター推定の信頼限界とユーザー指定の信頼水準
  - モデルの要約統計量
  - モデル効果の Wald の F 検定
  - デザイン効果
  - 分類テーブル
  - 対比係数 (L) 行列のセット
  - 回帰係数推定の分散共分散行列
  - 二乗平均誤差
  - 回帰係数の共分散行列と相関行列
- 欠落データの扱い
  - 欠落した値をリスト単位で除外
- その他
  - ユーザー指定の共通因子 df (すべての検定統計量で p 値の計算に使用)
  - 共線性の診断
  - モデルを部分母集団に適合可能

### コンプレックス・サンプル Cox 回帰 (CSCOXREG)\*

この手順は、生存時間の分析に Cox 比例ハザード回帰を適用します。生存時間とは、コンプレックス・サンプル手法で抽出したサンプルに特定のイベントが発生するまでの時間を指します。CSCOXREG では連続型とカテゴリー型の予測子を扱うことができ、これらは時間依存であることも可能です。CSCOXREG を使用することで、サブグループ間の相違の検討が容易になり、一連の予測子の効果も分析できるようになります。また、この手順では、1 つの対象に複数のケース (来院、遭遇、観察など) が存在するデータを扱うことができます。

- 期間とイベント: 生存時間の変数、および対象となるイベントが発生したことを示す値を指定
  - 生存時間
    - 区間の開始 (リスクの開始)
    - 時間 0
    - 対象により変化
    - 区間の終了
    - 個別値または値の範囲としてのイベント
- 予測子:
  - 因子
  - 共変量
  - 時間依存の予測子
- サブグループ: 分析を層別化するか、部分母集団に限定
- モデル
  - 主効果
  - すべての n 元相互作用
  - ユーザー指定 (入れ子にした項など)
- 統計量:
  - サンプル・デザインの情報
  - イベントおよび調査の要約
  - イベント時のリスク・セット
  - モデルのパラメーター: 係数推定、指数化推定、各係数推定の標準誤差、各係数推定の t 検定、各係数推定の信頼区間、各係数推定のデザイン効果、各係数推定のデザイン効果の平方根、パラメーター推定値の共分散、およびパラメーター推定値の相関行列
  - モデルの前提
    - 比例ハザードの検定
    - 代替モデルのパラメーター推定値
    - 代替モデルの共分散行列
  - ベースラインでの生存と累積
  - ハザード関数

\* マルチスレッド・アルゴリズムによって、マルチプロセッサまたはマルチコア・マシンでのパフォーマンスとスケーラビリティが向上します。

- 作図:
  - 生存関数
  - ハザード関数
  - 生存関数のログ・マイナス・ログ・プロット
  - 1 マイナス生存関数
  - 信頼区間を表示するオプション
  - 指定されたレベルでの因子と共変量のプロット
- 仮説の検定
  - 検定の統計: F 検定、調整済み F 検定、カイ二乗検定、調整済みカイ二乗検定
  - 多重比較の調整: 最小有意差、ボンフェローニ、逐次ボンフェローニ、シダック、逐次シダック
  - サンプル抽出の自由度: サンプル・デザインに基づくか、ユーザーによる指定
- モデル変数は使用中のファイルに保存できるほか、パラメーター行列を収めた外部ファイルにエクスポート可能
  - 変数: 生存関数、生存関数の信頼区間の下限、生存関数の信頼区間の上限、累積ハザード関数、累積ハザード関数の信頼区間の下限、累積ハザード関数の信頼区間の上限、線型予測子の予測値、Schoenfeld 残差 (モデル・パラメーターごとに 1 つの変数)、Martingale 残差、最大対数尤度比残差、Cox Snell 残差、得点残差 (モデル・パラメーターごとに 1 つの変数)、DFBeta 残差 (モデル・パラメーターごとに 1 つの変数)、集計 Martingale 残差、集計最大対数尤度比残差、集計 Cox Snell 残差、集計得点残差 (モデル・パラメーターごとに 1 つの変数)、集計 DFBeta 残差 (モデル・パラメーターごとに 1 つの変数)
- モデルと生存関数のエクスポート
- IBM SPSS Statistics データ・ファイルとしてエクスポート
- 生存関数を IBM SPSS Statistics データ・ファイルとしてエクスポート
- モデルを XML ファイルとしてエクスポート
- 推定基準、生存関数と信頼区間の計算方法、およびユーザーが指定していない値の処理方法を指定するオプション
  - 推定: 最大反復回数、最大段階の二分、パラメーター推定の変化に基づいて反復回数を制限、対数尤度の変化に基づいて反復回数を制限、反復の履歴を表示、パラメーター推定の均衡を破る方法 (Efron または Breslow)
  - 生存関数: 生存関数ベースラインの推定手法 (Efron、Breslow、積極限法)、生存関数の信頼区間 (元の単位または変換された単位)
  - 信頼区間の水準を指定
  - 欠落データの扱い (有効値または無効値のどちらとして扱うか)

#### システム要件

- 要件はプラットフォームによって異なります。

[ibm.com/spss/requirements](https://www.ibm.com/spss/requirements)

## IBM Business Analytics について

IBM Business Analytics ソフトウェアは、業績改善に取り組む意思決定者に対し、実践的な洞察を提供します。IBM は、ビジネス・インテリジェンス、予測分析と高度な分析、財務パフォーマンスと戦略の管理、ガバナンス、リスクおよびコンプライアンス (GRC)、そしてアナリティック・アプリケーションからなる包括的なポートフォリオを用意しています。

IBM ソフトウェアは、ビジネスの傾向やパターンあるいは異常の発見、仮定に基づくシナリオの比較、潜在的な脅威や機会の予測、重要なビジネス・リスクの特定および管理、さらには経営資源に関する計画、予算および予測を実現します。IBM の世界中のお客様は、この充実したアナリティクスを使うことで、業績への理解を深める一方、成果への予測を高め、目標への確かな道筋をつけることができます。



---

日本アイ・ビー・エム株式会社  
〒103-8510  
東京都中央区日本橋箱崎町19-21

IBM のホーム・ページはこちらからご覧になれます。

**ibm.com**

IBM、IBM ロゴ、ibm.com および SPSS 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、[ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。すべての製品が、IBM が営業を行っているすべての国において利用可能なわけではありません。

本書に掲載されている情報は特定物として現存するままの状態を提供され、第三者の権利の不侵害の保証、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なしで提供されています。

IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。

© Copyright IBM Corporation 2012



Please Recycle

**Business Analytics** software