



---

## ハイライト

- データをさまざまな角度から容易に検証する
  - 欠損データの問題を素早く診断する
  - 欠損値を推定値に置き換える
  - 欠損データ・タイプおよび極値を表示する
  - 隠れたバイアスを除去する
- 

# IBM SPSS Missing Values

## 空白を埋める際の適切なモデルを構築

アンケート調査や市場調査、社会科学、データ・マイニングなどの多くの専門家が、調査データの検証に IBM SPSS Missing Values を使用しています。

欠損データを無視したり、除外したりすると、偏った無意味な結果につながる危険性があります。SPSS Missing Values を使用して、欠損データを補完し、より有効な結論を引き出すことができます。SPSS Missing Values は、データの妥当性を考慮する際の重要なツールです。データを容易に検証して、欠損データのパターンを明らかにすることができ、その後、要約統計量を推定し、統計アルゴリズムを使用して欠損値を補完できます。

例えば、観察された欠損データのパターンに基づいて、分かりにくい質問を特定することで、アンケートの質問を改善できます。それだけでなく、パターンの不一致のパーセントを示すテーブルを使用することで、ある変数の欠損値が、別の変数の欠損値と関連しているかどうかを判断できます。

例えば、収入に関する質問に答えなかった回答者は、学歴に関する質問にも答えない、などです。この情報を使用して、将来の調査内容の質を高めることができます。



### 迅速かつ容易に欠損データを診断する

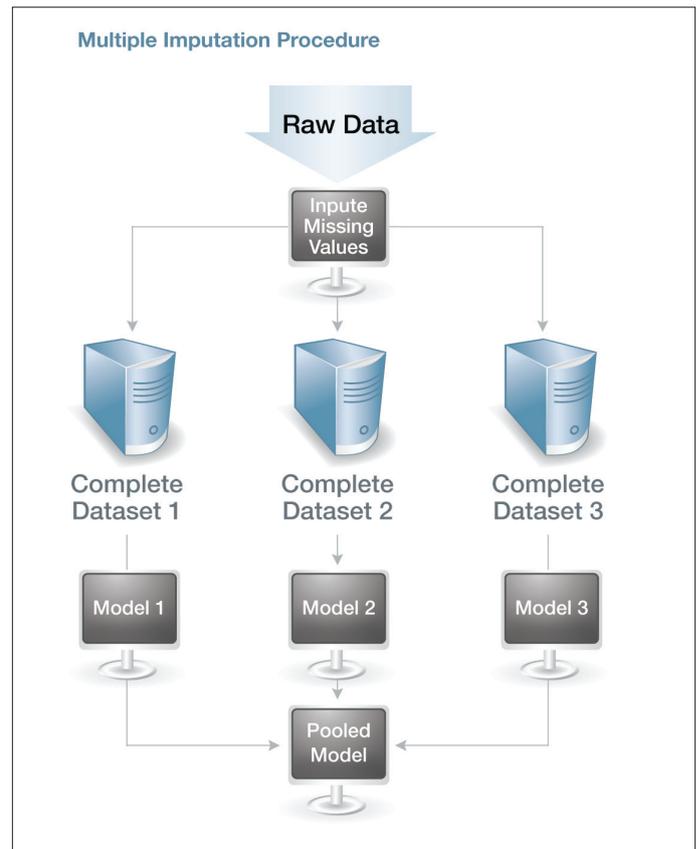
欠損値の全体的な要約レポートを使用して、欠損データが重大な問題になるかどうか素早く診断できます。欠損値パターンのレポートは、データの概要をケースごとに示します。欠損値の各タイプのスナップショットや、各ケースの極値が表示されます。

柔軟性の高い個別分散 t 検定およびカテゴリ変数テーブルのクロス集計を使用して、回答者と非回答者の間に有意差が存在するかどうかを検出します。これらのレポートは、分析において欠損データが問題となりうるかどうかを判断する際に役立ちます。

表形式のパターン・レポートでは、欠損データのパターンの要約を示し、パターンを構成する変数セットを強調表示します。

### 多重代入法を使用して、欠損値を補完

SPSS Missing Values では、データ・セット内の「欠損」のパターンを理解するのに役立ち、欠損値を妥当な推定値に置き換えることができる、多重代入法が使用されています。完全自動代入モードでは、データに基づいて最適な代入方法を選択します。また、代入モデルをカスタマイズすることもできます。



まず、複数の完全なデータ・セットが生成され (通常、3 から 5 個)、それぞれは異なる値に置き換えられます。次に、線形回帰などの通常の手法を使用して個々のデータ・セットをモデル化し、データ・セットごとにパラメーター推定値を生成できます。そして、最終的なパラメーター推定値を得ます。これは、ステップ 2 で得られたパラメーター推定値の個々のパラメーター・セットをプールし、代入内および各代入間の変動量を考慮する推測統計を算出することによって行われます。

個々のデータ・セットの分析および結果のプールは、回帰など既存の SPSS Statistics の手法を選択することによりサポートされます。代入値を含むデータ・セットで作業する場合、既存の手法により、プールされたパラメーター推定値が自動的に生成されます。

### より有意な結果を得る

欠損値を推定値に置き換えて、統計的に有意な結果を得るようにします。欠損値を推定値で置き換え、回答が少ないものも含めて、すべてのグループを分析に含めることにより、隠れたバイアスをデータから除去します。

SPSS Missing Values は、IBM SPSS Statistics ソフトウェア・スイートの一部です。この製品は、単独でご利用いただけますが、次の 3 つのエディションでもご利用いただけます。それは、IBM SPSS Statistics Standard、IBM SPSS Statistics Professional、および IBM SPSS Statistics Premium です。これらのエディションは、重要な機能を結集することにより、チームまたは部署全体が、組織の成功に寄与する分析の実行に必要なフィーチャーおよび機能を確保するための効果的な方法を提供します。

SPSS Missing Values 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。

### より高い価値を創出するコラボレーション

IBM SPSS Collaboration and Deployment Services と統合して使用することで、分析資産の共有および分散を効率的に行い、社内外のコンプライアンス要件を満たす方法で保護し、分析結果を公開して、より多くのビジネス・ユーザーが閲覧および利用することが可能です。それらの有益な機能の詳細については、次の Web サイトでご覧になれます。 [ibm.com/spss/cds](https://www.ibm.com/spss/cds)

## 機能

### パターンの分析

- データ・パターン・テーブルを使用して、すべてのケースおよびすべての変数について、欠損データおよび極端なケースを表示
  - システム欠損値および 3 つのタイプのユーザー定義の欠損値を表示
  - 昇順または降順でソート
  - 指定した変数の実際の値を表示
- 欠損パターン・テーブルを使用して、少なくとも 1 つの欠損値があるすべてのケースについて、欠損値のパターンを表示
  - 類似した欠損値パターンをグループにまとめる
  - 欠損パターンおよび変数によりソート
  - 指定した変数の実際の値を表示
- 個別分散の t 検定テーブルを使用して、関連する変数について、欠損グループと非欠損グループの間の差を判別
  - t 検定、自由度、平均値、p 値、および度数
- カテゴリー変数テーブルの分布を使用して、カテゴリー変数の存在データと欠損データの差を表示
  - ある変数と他の変数の組み合わせによるカテゴリーごとに、結果と欠損データを示すクロス集計を作成

- パターンの不一致のパーセントを示すテーブルを使用して、ある変数の欠損データが別の変数の欠損データにどの程度関連しているかを評価
  - 行列を欠損値パターンまたは変数でソート
- パターンを表形式で表示して、すべての固有パターンを特定。この表で欠損データのパターンを要約し、各パターンの度数を表示。また各変数の平均値と度数を表示
  - 欠損値パターンの要約テーブルを使用して、各欠損値パターンの度数および平均値を表示

### 統計量

- 1 変量: 欠損値を含むケースを除いたすべてのケースの度数、平均値、標準偏差、平均値の標準誤差、すべての変数の欠損値の度数とパーセント、および極値を計算
- リストごと: 欠損値を除外したケースについて、すべての量的変数の平均値、共分散行列、および相関行列を計算
- ペアごと: 度数、平均値、分散、共分散行列、および相関行列を計算

## 多重代入法

- どの変数に代入するかを指定し、代入値の制約 (最小値や最大値など) を指定。また、他の変数の欠損値に代入する場合に予測子として使用される変数を指定することも可能
  - カテゴリー変数および連続変数に値を代入。カテゴリー変数にはロジスティック回帰を使用し、連続変数には線形回帰を使用。予測平均値マッチングは、連続的な結果のオプションであり、代入値の妥当性 (オリジナル・データの範囲内) を確保
  - 欠損データのパターンの検出により、使用する代入方式の決定を支援
  - 次の 3 つの代入法を提供
    - 単調: 欠損パターンが単調な場合のデータに効果的な方法
    - 完全条件指定 (FCS): 不定な欠損パターン (単調または非単調) の場合に適切な反復のマルコフ連鎖モンテカルロ (MCMC) 法
    - 自動: データをスキャンして最適な代入法を決定 (単調またはFCS)
  - 以下を指定:
    - 代入数
    - 代入値の範囲
    - 代入時に交互作用効果が使われるかどうか
- 高い確率で欠損値がある変数の代入を任意でやめる
  - 特異性を調べるための許容レベル
  - 分析 (回帰) の重みが含まれた変数も指定可能。この手法には、欠損値の代入に使用された回帰や分類モデルの分析の重みが組み込まれる。また、分析の重みは代入値の要約にも使用される (例: 平均、標準偏差、標準誤差)
  - データ内の欠損値の全体的な要約と代入の要約、および値が代入された各変数の代入モデルを表示。変数による欠損値の分析と表形式の欠損値パターンを取得。また、代入値の記述統計量も取得
  - ケース、変数、個々のデータ (セル) 値の欠損をグラフ形式で要約
  - 代入値と FCS 反復履歴が含まれる IBM SPSS Statistics データ・ファイルを要求
  - 多重代入データ・セットは対応している分析手法を用いて分析でき、いくつかの代入値に内在する不確実性を考慮した最終的な (結合した) パラメーター推定値を取得

## 分析

- 多重代入に対応する分析手法 (注: 該当する手法が搭載されている適切なモジュールを購入する必要があります)
- 記述的手法: 度数、記述統計、クロス集計、相関\*、ノンパラメトリック相関、部分相関\*
- 平均の比較: 平均、t 検定、ノンパラメトリック検定、一元配置分散分析、1 変量分散分析
- モデル: 一般線形モデル、一般化線形モデル、線形回帰\*、多項ロジスティック回帰\*、二項ロジスティック回帰、判別分析\*、順序回帰、線形混合モデル
- 生存分析手法: Cox 回帰

## プーリング

- 出力のプーリング: プーリングが生成するプールされたパラメーターの 2 つのレベルのうちの 1 つを使用して、出力はプールされる
- プーリング診断:
  - 分散の相対的増加: 代入全体のパラメーター推定値の相対的変動性の測度
  - 欠損情報の割合: 分散の比率の相対的増加。無回答による不確実性の測度
  - 相対的効率: 無限の代入数に相対する多重代入の推定の効率
- プールされたパラメーター推定値のために PMML モデルを取得: 線形回帰、一般化線形モデル、多項ロジスティック回帰、二項ロジスティック回帰、判別分析、Cox 回帰

## 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 ビジネス・アナリティクスについて

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、次の Web サイトをご覧ください。

[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ものではありません。

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