

IBM SPSS Missing Values

23



お願い

本書および本書で紹介する製品をご使用になる前に、23 ページの『特記事項』に記載されている情報をお読みください。

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

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

原典： IBM SPSS Missing Values 23

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

第 1 章 欠損値の概要	1	欠損データ値の代入	13
		方法	14
		制約	15
		出力	16
第 2 章 欠損値分析	3	MULTIPLE IMPUTATION コマンドの追加機能	17
欠損値のパターンの表示	4	多重代入データでの作業	17
欠損値の記述統計の表示	5	多重代入データの分析	18
統計量の推定と欠損値への代入	6	多重代入オプション	21
EM 推定のオプション	7		
回帰推定のオプション	7	特記事項	23
予測された変数と予測値変数	8	商標	25
MVA コマンドの追加機能	8		
第 3 章 多重代入	11	索引	27
パターン分析	12		

第 1 章 欠損値の概要

欠損値のあるケースは、通常モデリング・プロシージャーでは単に分析から破棄されるため、重要な課題となります。欠損値が少数 (おおまかに見て、全ケース数の 5% 未満) であり、かつ当該の値がランダムに欠損していると考えられる (つまり、値が欠損しているかどうかは他の値に依存しない) 場合には、通常の方法であるリストワイズ除去が比較的「安全」です。欠損値オプションは、リストワイズ除去が十分かどうかを判断する際の助けとなり、十分でない場合には欠損値を処理するための方法を提供します。

欠損値分析プロシージャー対多重代入プロシージャー

欠損値オプションには、欠損値を処理するためのプロシージャーが 2 セットあります。

- 多重代入プロシージャーでは、最終的には欠損値の多重代入を目的として欠損データのパターンの分析を行います。つまり、複数のバージョンのデータ・セットが生成され、それぞれに独自の代入値のセットがある、ということです。統計分析が実行される際、すべての代入データ・セットに対するパラメーター推定値がプールされ、一般的には代入が 1 つのみの場合よりも正確な推定値が提供されます。
- 欠損値分析での、欠損データを分析するための記述ツールのセットはやや異なり (中でも特に、Little の MCAR 検定)、さまざまな単一代入方法が含まれています。一般に、単一代入より多重代入の方が優れていると見なされている、ということに注意してください。

欠損値タスク

欠損値の分析は、以下の基本的なステップにより開始します。

1. **欠損の調査。**「欠損値分析」と「パターン分析」を使用してデータ中の欠損値のパターンを探索し、多重代入が必要かどうか判断します。
2. **欠損値の代入。**「欠損データ値の代入」を使用して、欠損値を多重代入します。
3. **「完全な」データの分析。**多重代入データをサポートする、いずれかのプロシージャーを使用します。多重代入データ・セット分析、およびこうしたデータをサポートするプロシージャーのリストについては、18 ページの『多重代入データの分析』を参照してください。

第 2 章 欠損値分析

「欠損値分析」プロシージャーでは、以下の 3 つの主要な機能を実行します。

- 欠損データのパターンの記述。欠損値はどこにあるか。範囲はどのくらいか。変数ペアに、複数のケースにわたって値が欠損する傾向があるか。データ値は極値か。値の欠損の仕方はランダムか。
- 欠損値にまつわるさまざまな方法 (リストワイズ法、ペアワイズ法、回帰法、EM (期待値最大化) 法) による、平均値、標準偏差、共分散、および相関の推定。ペアワイズ法では、ペアワイズにおいて完全なケース数も表示します。
- 回帰法または EM 法の使用による、欠損値への推定値の入力 (代入)。ただし一般には、多重代入の方がより正確な結果を取得できるとされています。

欠損値分析により、不完全なデータに起因する懸案事項のいくつかに対処することができます。欠損値のあるケースが欠損値のないケースと系統的に異なる場合、その結果は誤ったものになる可能性があります。また、当初の計画よりも情報が少ないため、欠損データは、算出される統計量の精度を低下させる場合があります。別の懸案事項として、多くの統計手法の背後にある仮定は完全なケースに基づいており、欠損値があると必要な理論が複雑化する場合がある、ということがあります。

例: 白血病の治療法を評価する中で、いくつかの変数を測定するとします。ただし、各患者に対してすべての測定値が使用可能であるとは限りません。欠損データのパターンを表示し、集計したら、ランダムであることがわかりました。EM 分析を使用して、平均値、相関、および共分散を推定します。これは、データが完全にランダムに欠損しているかどうかを判断するためにも使用されます。この後、欠損値は代入値に置き換えられ、さらなる分析に向けて新規データ・ファイルに保存されます。

統計: 非欠損値の数、平均値、標準偏差、欠損値の数、極値の数などの 1 変量統計量。リストワイズ法、ペアワイズ法、EM 法、または回帰法を使用しての推定平均値、共分散行列、および相関行列。EM 結果による Little の MCAR 検定。さまざまな方法による、平均値の集計。 t 検定 (欠損値対非欠損値で定義されるグループの場合)。すべての変数について、変数ごとのケースで表示される欠損値パターン。

データに関する考慮事項

データ: データは、カテゴリと量的 (スケールと連続) のいずれも可能です。ただし、統計量を推定して欠損データを代入できるのは、量的変数の場合のみです。いずれの変数でも、システム欠損値としてコード化されていない欠損値は、ユーザー欠損値として定義する必要があります。例えば、調査票のある項目での回答「わからない」が 5 とコード化されており、これを欠損値として扱う場合、この項目では 5 をユーザー欠損値としてコード化する必要があります。

度数による重み付け。 度数による重み付け (反復重み付け) は、このプロシージャーで処理します。反復重み付けの値が負または 0 であるケースは無視されます。非整数である重み付けは切り捨てられます。

仮定: リストワイズ推定、ペアワイズ推定、および回帰推定は、欠損値パターンはデータ値に依存しない、という仮定に基づいています。(この条件を、完全にランダムに欠損 (MCAR) といいます。) したがって、データが MCAR である場合、すべての推定方法 (EM 法も含む) では、相関および共分散の一貫した不偏推定値が得られます。MCAR の仮定に反すると、リストワイズ法、ペアワイズ法、および回帰法によって生成される推定値に偏りが生じることがあります。データが MCAR ではない場合は、EM 推定を使用する必要があります。

EM 推定は、欠損データのパターンは観測データにのみ関連する、という仮定に基づいています。(この条件を、**ランダムに欠損 (MAR)** といいます。) この仮定では、入手可能な情報を使用して推定値を調整できます。例えば、教育と収入の調査において、教育レベルの低い被験者の方が収入の欠損値が多くなる可能性があります。この場合、データは **MCAR** ではなく **MAR** です。つまり、**MAR** では、収入が記録される確率は被験者の教育レベルに依存します。確率は、その教育レベル内の収入ではなく、教育によって異なってくる、ということです。収入が記録される確率が各教育レベル内の収入の値によっても異なる (高収入の人は収入を報告しないなどの) 場合、そのデータは **MCAR** でも **MAR** でもありません。この状況はまれではありません。該当する場合は、いずれの方法も適切ではありません。

関連プロシージャー: 多くのプロシージャーでは、リストワイズ推定またはペアワイズ推定を使用することができます。線型回帰および因子分析では、欠損値を平均値で置き換えることができます。予測アドオン・モジュールには、欠損値を時系列で置き換えるのに使用可能な方法が複数あります。

欠損値分析の取得方法

1. メニューから次の項目を選択します。

「分析」 > 「欠損値分析...」

2. 統計量を推定し、オプションで欠損値を代入するための量的 (スケール) 変数を 1 つ以上選択します。

オプションとして、以下の選択が可能です。

- カテゴリー変数 (数値または文字列) を選択し、カテゴリー数の制限を入力します (**最大カテゴリ数**)。
- 「**パターン**」をクリックして、欠損データのパターンを集計します。詳しくは、トピック『欠損値のパターンの表示』を参照してください。
- 「**記述統計**」をクリックして、欠損値の記述統計を表示します。詳しくは、トピック 5 ページの『欠損値の記述統計の表示』を参照してください。
- 統計量 (平均値、共分散、相関) の推定と (場合によっては) 欠損値の代入をするための方法を選択します。詳しくは、トピック 6 ページの『統計量の推定と欠損値への代入』を参照してください。
- EM または回帰を選択した場合は、「**変数**」をクリックして、推定をするために使用されるサブセットを指定します。詳しくは、トピック 8 ページの『予測された変数と予測値変数』を参照してください。
- ケースのラベル変数を選択します。この変数は、個々のケースを表示するパターン・テーブル内のケースにラベルを付けるために使用します。

欠損値のパターンの表示

欠損データのパターンと程度を示すテーブルにはさまざまなものがあり、その表示を選択できます。これらのテーブルにより、以下のことを確認できます。

- 欠損値のある場所
- 変数ペアに、個々のケースにおいて値が欠損する傾向があるか
- データ値は極値か

表示

欠損データのパターンを表示するのに使用可能なテーブルが 3 種類あります。

集計されたケース。 分析変数の欠損値パターンが集計され、各パターンについて度数が表示されます。「**欠損値のパターンで変数を並べ替え**」を使用して、パターンの類似性によって個数と変数をソートするかどうかを指定します。「**n % 未満のケースのパターンを省略する**」を使用して、頻度の高くないパターンを除去します。

欠損値のあるケース。欠損値または極値のある各ケースが、各分析変数について集計されます。「欠損値のパターンで変数を並べ替え」を使用して、パターンの類似性によって個数と変数をソートするかどうかを指定します。

すべてのケース。各ケースが集計され、各変数の欠損値および極値が示されます。ケースは、「並べ替え」に変数が指定されていない限り、データ・ファイルでの出現順にリストされます。

個々のケースを表示するテーブルでは、以下の記号が使用されます。

+ 極大値

- 極小値

S システム欠損値

A ユーザー欠損値の 1 番目のタイプ

B ユーザー欠損値の 2 番目のタイプ

C ユーザー欠損値の 3 番目のタイプ

変数

分析に含まれている変数について、追加情報を表示できます。「追加情報」に追加した変数は、欠損パターンのテーブルで個別に表示されます。量的 (スケール) 変数の場合は平均値が表示され、カテゴリ変数の場合は各カテゴリについてパターンのあるケースの数が表示されます。

- **並べ替え。** ケースは、指定した変数の値の昇順または降順でリストされます。「すべてのケース」の場合にのみ使用可能です。

欠損値パターンの表示方法

1. 「欠損値分析」のメインダイアログ・ボックスで、欠損値パターンの表示対象の変数を選択します。
2. 「パターン」をクリックします。
3. 表示するパターン・テーブルを選択します。

欠損値の記述統計の表示

1 変量統計量

1 変量統計量は、欠損データのおおまかな程度を特定する助けとなります。各変数について、以下のものが表示されます。

- 非欠損値の数
- 欠損値の数とパーセント

量的 (スケール) 変数の場合は、以下のものも表示されます。

- 平均値
- 標準偏差
- 極大値と極小値の数

指示変数統計

各変数に対して、指示変数が作成されます。このカテゴリ変数は、個々のケースに変数があるか欠損しているかを示します。指示変数は、不一致、*t* 検定、および度数分布表を作成するのに使用されます。

不一致のパーセント。 変数のペアごとに、一方の変数は欠損値でもう一方の変数は非欠損値、というケースのパーセントを表示します。表の各対角要素には、1 つの変数についての欠損値のパーセントが含まれません。

指示変数で作成したグループの *t* 検定。 各量的変数について、スチューデント *t* 統計量を使用して、2 つのグループの平均値を比較します。グループに変数があるか欠損しているかが明示されます。2 つのグループの *t* 統計量、自由度、欠損値と非欠損値の個数、および平均値が表示されます。*t* 統計量に関連付けられた両側確率を表示することもできます。分析の結果、複数の検定を行うことになった場合、これらの確率を有意差検定に使用しないでください。こうした確率は、1 つの検定を計算する場合にのみ適しています。

カテゴリ変数と指示変数のクロス集計表。 カテゴリ変数ごとに、表が表示されます。それぞれのカテゴリについて、他の変数の非欠損値の度数とパーセントがこの表に表示されます。欠損値のタイプごとのパーセントも表示されます。

n % 未満のケースの欠損変数を省略する。 計算対象となるケースがごく少数である統計量を除外することで、表を小さくすることができます。

記述統計の表示方法

1. 「欠損値分析」のメインダイアログ・ボックスで、欠損値の記述統計の表示対象とする変数を選択します。
2. 「記述統計」をクリックします。
3. 表示する記述統計を選択します。

統計量の推定と欠損値への代入

リストワイズ法 (完全なケースのみ)、ペアワイズ法、EM (期待値最大化) 法、および/または回帰法を使用した平均値、標準偏差、共分散、および相関の推定を選択することができます。また、欠損値の代入 (置換値の推定) を選択することもできます。一般に、欠損値の問題を解決するには、単一代入より多重代入の方が優れていると見なされている、ということに注意してください。Little の MCAR 検定もなお、代入が必要かどうかを決定するのに有用です。

リストワイズ法

この方法は、完全なケースのみを使用します。いずれかの分析変数に欠損値が含まれていれば、そのケースは計算から除外します。

ペアワイズ法

この方法は、分析変数をペアで調べて、いずれの変数にも非欠損値が含まれている場合にのみケースを使用します。度数、平均値、および標準偏差は、ペアごとに個別に計算されます。ケース中の他の欠損値は無視されるため、2 つの変数の相関と共分散が他の何らかの変数で欠損している値に左右されることはありません。

EM 法

この方法は、一部欠損データについて分布を仮定し、当該分布の下での尤度に基づく推論を行います。各反復は、E ステップと M ステップから成っています。E ステップでは、パラメーターの観測値と現在の推定値を前提に、「欠損」データの条件付き期待値を求めます。次に、「欠損」データがこれらの期待値で置

換されます。M ステップでは、欠損データが入力されているものとして、パラメーターの最大尤度推定値が計算されます。欠損値が直接入力されていないため、「欠損」のように括弧で囲みます。代わりに、これらの関数は、対数尤度で使用されます。

値が完全にランダムに欠損しているかどうか (MCAR) を検定するための Roderick J. A. Little のカイ 2 乗統計量が、EM 行列の脚注として印刷されます。この検定の帰無仮説では、データが完全にランダムに欠損しており、 p 値は 0.05 レベルで有意となります。値が 0.05 未満の場合、データは完全にランダムには欠損していません。データはランダムに欠損している (MAR) か、ランダムに欠損していない (NMAR) のいずれかです。いずれか一方を仮定することはできないので、データを分析し、どのように欠損しているかを判断する必要があります。

回帰法

この方法では、多重線型回帰の推定値を計算します。また、ランダム成分で推定値を増加させるというオプションがあります。このプロシージャでは、各予測値に対し、ランダムに抽出した完全なケース、ランダムな正規偏差、または t 分布の (残差平均平方の平方根で計られた) ランダム偏差からの残差を加えることができます。

EM 推定のオプション

EM 法では反復プロセスを使用して、欠損値のある量的 (スケール) 変数の平均値、共分散行列、および相関を推定します。

分布。 EM 法では、指定されている分布の尤度に基づいて推論します。デフォルトでは、正規分布と仮定します。分布の裾が正規分布の裾より長いことがわかっている場合、プロシージャで学生 t 分布から自由度 n で尤度関数を構成するよう要求することができます。混合正規分布では、裾の長い分布にも対応しています。混合正規分布の標準偏差の比、および 2 つの分布の混合比率を指定します。混合正規分布では、分布の標準偏差のみが異なる、と仮定します。平均値は同じである必要があります。

最大反復回数。 反復の最大数を設定して、真の共分散を推定します。プロシージャは、反復がこの数に達したときに、推定値が収束されていなくても、停止します。

完成したデータの保存。 欠損値の代わりに代入値を持つデータ・セットを保存することができます。ただし、代入値を使用し共分散に基づく統計量では、そうしたパラメーター値を過小評価するので、注意が必要です。過小評価度は、同時には観測されていないケースの数に比例します。

EM オプションの指定方法

1. 「欠損値分析」のメインダイアログ・ボックスで、EM 法を使用して欠損値を推定する対象の変数を選択します。
2. 「推定」グループで「EM」を選択します。
3. 予測された変数と予測値変数を指定するには、「変数」をクリックします。詳しくは、トピック 8 ページの『予測された変数と予測値変数』を参照してください。
4. 「EM」をクリックします。
5. 目的の EM オプションを選択します。

回帰推定のオプション

回帰法では、多重線型回帰を使用して欠損値を推定します。予測された変数の平均値、共分散行列、および相関行列が表示されます。

推定調整。 回帰法により、回帰推定にランダム成分を追加することができます。残差、正規変量、スチューデントの t 変量、または調整なしを選択できます。

- *Residuals* (残差). 完全なケースの観測残差から無作為に誤差項を選択して回帰推定に加えます。
- *Normal Variates* (正規変量). 期待値が 0 で標準偏差が回帰の平方平均誤差項の平方根に等しい分布から誤差項を無作為に抽出します。
- *Student's t Variates* (スチューデントの T 変量). 指定された自由度を持つ t 分布から無作為に誤差項を抽出し、平方平均誤差の平方根 (RMSE) でスケールします。

最大予測数。 推定処理で使用する予測 (独立) 変数の最大数を設定します。

完成したデータの保存。 回帰法により推定された値で欠損値を置き換え、現在のセッションまたは外部の IBM® SPSS® Statistics データ・ファイルにデータ・セットを書き込みます。

回帰オプションの指定方法

1. 「欠損値分析」のメインダイアログ・ボックスで、回帰法を使用して欠損値を推定する対象の変数を選択します。
2. 「推定」グループで「回帰」を選択します。
3. 予測された変数と予測値変数を指定するには、「変数」をクリックします。詳しくは、トピック『予測された変数と予測値変数』を参照してください。
4. 「回帰」をクリックします。
5. 目的の回帰オプションを選択します。

予測された変数と予測値変数

デフォルトで、EM 推定および回帰推定には、すべての量的変数が使用されます。必要に応じて、推定では、予測された変数と予測値変数として、特定の変数を選択できます。両方のリストに任意の変数を指定できますが、変数の使用を制限することが推奨されるような状況もあります。例えば、結果変数の値を推定することに不安を感じるアナリストもいます。また、さまざまな推定にそれぞれ異なる変数を使用して、プロシーチャーを複数回実行することが推奨される場合もあります。例えば、看護師の評価に関する項目セットと医師の評価に関する項目セットとが別になっている場合、看護師項目を使用してその欠損項目の推定を実行する一方で、別の推定を実行して医師項目の推定値を求めることが推奨されます。

回帰法を使用する場合には、また別の考慮事項が発生します。多重回帰では、使用する独立変数のサブセットが大きいと、生成される予測値の質が、より小さいサブセットの場合よりも低下する可能性があります。したがって、変数が使用されるには、投入する F 値制限である 4.0 に到達する必要があります。この制限は、シンタックスで変更することができます。

予測された変数と予測値変数の指定方法

1. 「欠損値分析」のメインダイアログ・ボックスで、回帰法を使用して欠損値を推定する対象の変数を選択します。
2. 「推定」グループで「EM」または「回帰」を選択します。
3. 「変数」をクリックします。
4. 予測された変数と予測値変数として、すべての変数ではなく特定の変数を使用する場合は、「変数を選択」を選択して、変数をそれに適したリストに移動します。

MVA 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。

- MPATTERN、DPATTERN、または TPATTERN サブコマンドで DESCRIBE キーワードを使用して、欠損値パターン、データ・パターン、および集計されたパターンに別個の記述変数を指定します。
- DPATTERN サブコマンドを使用して、データ・パターン・テーブルに複数のソート変数を指定します。
- DPATTERN サブコマンドを使用して、データ・パターンに複数のソート変数を指定します。
- EM サブコマンドを使用して、許容度と収束基準を指定します。
- REGRESSION サブコマンドを使用して、許容度と投入する F 値を指定します。
- EM および REGRESSION サブコマンドを使用して、EM と 回帰に別々の変数リストを指定します。
- TTESTS、TABULATE、および MISMATCH のそれぞれに対し、表示されるケースを非表示にするパーセントとして別々の値を指定します。

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

第 3 章 多重代入

多重代入の目的は、欠損値に対し適用可能な値を生成することで、データの「完全な」セットを複数作成することです。多重代入データ・セットを処理する分析プロシージャは、「完全な」データ・セットごとに出力を作成します。加えて、元のデータ・セットに欠損値がなかったらどのような結果になっていたかを推定するプール出力も作成します。このようにしてプールされた結果は、一般に、複数回にわたる単一代入法で生成された結果よりも正確です。

多重代入データでの考慮事項












分析変数。分析変数として可能なものは以下のとおりです。

- *Nominal* (名義データ)。本質的な順位を持たないカテゴリーを表す値である場合 (従業員の勤務先企業での部署など)、変数を名義変数として取り扱うことができます。名義変数の例としては、地域、郵便番号、宗教上の所属などが挙げられます。
- *Ordinal* (順序データ)。値が本質的な順位を持つカテゴリーを表す場合 (例えば、サービス満足度のレベルを「非常に不満」から「非常に満足」までの順位で評価する場合) は、変数を順序変数として扱うことができます。順序変数の例としては、満足度や信頼度を表す得点や嗜好得点などが挙げられます。
- *Scale* (スケール データ)。意味のある測定基準を持つ順序カテゴリーを値が表しており、値の間の距離の比較が可能である場合は、変数をスケール (連続型) として扱うことができます。スケール変数の例としては、年齢や、千ドル単位で表した所得が挙げられます。

このプロシージャでは、変数に割り当てられている測定レベルはすべて適切なものである、と仮定しますが、変数に対する測定レベルは一時的に変更することができます。そのためには、ソース変数リストで変数を右クリックし、ポップアップ・メニューから測定レベルを選択します。変数に対する測定レベルを完全に変更する場合は、

変数リストで各変数の横にあるアイコンが示す測定レベルとデータ型は、以下のとおりです。

表 1. 測定レベル・アイコン

	数値	文字列	日付	時間
スケール (連続)		該当なし		
順序				
名義				

度数による重み付け。度数による重み付け (反復重み付け) は、このプロシージャで処理します。反復重み付けの値が負または 0 であるケースは無視されます。非整数である重み付けは、最も近い整数に丸められます。

分析重み付け。分析 (回帰または抽出) 重み付けは、欠損値の集計および代入モデルの当てはめに組み込まれています。分析重み付けが負または 0 であるケースは除外されます。

コンプレックス・サンプル。「多重代入」プロシージャーでは、階層、クラスターその他のコンプレックス・サンプル構造を明示的に処理するわけではありませんが、最終的な標本の重み付けは、分析重み付け変数の形式で受け入れることができます。また、「コンプレックス・サンプル」プロシージャーは現時点で、多重代入されたデータ・セットを自動的に分析しない、ということに注意してください。プールをサポートするプロシージャーの完全なリストについては、18ページの『多重代入データの分析』を参照してください。

欠損値。ユーザー欠損値とシステム欠損値のいずれも、無効値として扱われます。つまり、いずれのタイプの欠損値も、値が代入されると置き換えられ、代入モデルでの予測値として使用される変数の無効値として扱われます。欠損値の分析では、ユーザー欠損値もシステム欠損値もまた、欠損として取り扱われます。

結果の複製 (欠損データ値の代入)。代入の結果を正確に複製するには、使用するプロシージャーの設定を同じにするのに加え、使用する乱数ジェネレーターの初期化の値、データの順序、変数の順序も同じにします。

- **乱数の生成。**このプロシージャーでは、代入値の計算時に乱数の生成を行います。ランダム化された結果として今後も同じものを再作成するには、「欠損データ値の代入」プロシージャーを実行するごとに事前に、乱数ジェネレーターに同じ初期化の値を使用します。
- **ケースの順序。**値はケースの順序に従って代入されます。
- **変数の順序。**完全条件指定 (FCS) 代入方法では、「分析変数」リストで指定された順序に従って値を代入します。

多重代入専用のダイアログは 2 つあります。

- **パターン分析は、データにある欠損値のパターンの記述方法を提供します。**これは、代入前の検討段階で有用です。
- **欠損データ値の代入は、多重代入を生成するのに使用します。**多重代入データ・セットをサポートするプロシージャーで、完全なデータ・セットを分析することができます。多重代入データ・セットの分析、およびこうしたデータをサポートするプロシージャーのリストについては、18ページの『多重代入データの分析』を参照してください。

パターン分析

パターン分析は、データにある欠損値のパターンの記述方法を提供します。これは、代入前の検討段階で有用です。

例:ある通信プロバイダーが、顧客データベースでのサービス利用パターンをよりよく把握しようとしています。この企業には、顧客が利用するサービスについては完全なデータがありますが、自社で収集した人口統計情報には多くの欠損値があります。欠損値のパターンを分析することは、代入での次の手順を決定するための助けとなります。詳しくは、トピックを参照してください。

メニューから次の項目を選択します。

「分析」 > 「多重代入」 > 「パターン分析...」

1. 2 つ以上の分析変数を選択します。このプロシージャーにより、これらの変数の欠損データのパターンが分析されます。

オプション設定

分析重み付け。この変数には、分析 (回帰または抽出) 重み付けが含まれます。このプロシージャーでは、欠損値の集計に分析重み付けが含まれます。分析重み付けが負または 0 であるケースは除外されます。

出力。オプション出力として選択可能なものは以下のとおりです。

- **欠損値の集計。**これは、色分けされている円グラフを表示します。このグラフには、分析変数、ケース、または 1 つ以上の欠損値がある個別のデータ値の数およびパーセントが示されます。
- **欠損値のパターン。**これは、欠損値の集計されたパターンを表示します。それぞれのパターンは、分析変数に不完全および完全なデータの同じパターンがあるケースのグループに対応します。この出力を使用すると、データに対して単調な代入方法を使用することができるかどうか、また使用できない場合はデータが単調なパターンにどの程度近似しているかを判断できます。このプロシージャーでは、分析変数を順序付けして、単調なパターンを明らかにしたり、単調なパターンに近似させたりします。再度順序付けしても単調でないパターンが存在しない場合、データに単調なパターンがあるのは分析変数がそのように順序付けられている場合である、と結論付けることができます。
- **欠損値が最大度数である変数。**これは、分析変数を欠損値のパーセントの降順で並べた表を表示します。この表には、スケール変数の記述統計量 (平均値および標準偏差) が含まれます。

表示する変数の最大数と、表示対象とする変数の欠損の最小パーセントは、制御することができます。両方の基準を満たす変数のセットを表示します。例えば、変数の最大数を 50、欠損の最小パーセントを 25 に設定すると、欠損値が 25% 以上の変数最大 50 個を表に表示する、という要求が出されます。分析変数が 60 個あるが欠損値が 25% 以上なのは 15 個のみである場合、出力される変数は 15 個のみとなります。

欠損データ値の代入

「欠損データ値の代入」は、多重代入を生成するのに使用されます。多重代入データ・セットをサポートするプロシージャーで、完全なデータ・セットを分析することができます。多重代入データ・セットの分析、およびこうしたデータをサポートするプロシージャーのリストについては、18 ページの『多重代入データの分析』を参照してください。

例: ある通信プロバイダーが、顧客データベースでのサービス利用パターンをよりよく把握しようとしています。この企業には、顧客が利用するサービスについては完全なデータがありますが、自社で収集した人口統計情報には多くの欠損値があります。さらに、これらの値は完全にランダムには欠損していないため、データ・セットを完全なものにするためには多重代入が使用されます。詳しくは、トピックを参照してください。

メニューから次の項目を選択します。

「分析」 > 「多重代入」 > 「欠損データ値の代入...」

1. 代入モデルには、2 つ以上の変数を選択します。このプロシージャーにより、これらの変数の欠損データについて、複数の値が代入されます。
2. 計算する代入の数を指定します。デフォルトでは、この値は 5 です。
3. 代入データの書き込み先となるデータ・セットまたは IBM SPSS Statistics 形式のデータ・ファイルを指定します。

出力データ・セットは、欠損データを含む元のケース・データと、各代入についての代入値を含むケースのセットから成っています。例えば、元のデータ・セットに 100 ケースがあり、代入が 5 つである場合、出力データ・セットには 600 ケースあることとなります。入力データ・セットのすべての変数が出力データ・セットに含まれます。既存の変数のディクショナリー・プロパティー (名前、ラベルなど) は、新しいデータ・セットにコピーされます。ファイルには新規の変数 *Imputation_* も含まれます。これは、代入を示す数値変数 (0 は元のデータ、1..n は代入値を持つケースの場合) です。

出力データ・セットが作成される場合、プロシージャーでは自動的に *Imputation_* 変数を分割変数として定義します。このプロシージャーの実行時に分割が有効になっている場合、出力データ・セットには、分割変数の値の組み合わせごとに、代入が 1 セット含まれます。

オプション設定

分析重み付け。この変数には、分析 (回帰または抽出) 重み付けが含まれます。プロシージャーには、回帰モデルおよび分類モデルで欠損値の代入に使用される分析重み付けが含まれます。分析重み付けは、平均値、標準偏差、標準誤差などの代入値の集計でも使用されます。分析重み付けが負または 0 であるケースは除外されます。

測定レベルが不明なフィールド

測定レベルの警告は、データ・セットにある 1 つ以上の変数 (フィールド) について、測定レベルが不明である場合に表示されます。測定レベルはこのプロシージャーでの結果の計算に影響するため、すべての変数に対して測定レベルが定義されている必要があります。

「データをスキャン」。アクティブ・データ・セットのデータを読み取り、その時点で測定レベルが不明なフィールドがあれば、デフォルトの測定レベルを割り当てます。データ・セットが大きいと、時間がかかる場合があります。

手動で割り当て。測定レベルが不明なフィールドをすべてリストするダイアログを開きます。このダイアログを使用することで、こうしたフィールドに測定レベルを割り当てることができます。測定レベルの割り当ては、データ・エディターの変数ビューでも可能です。

測定レベルはこのプロシージャーで重要であるため、このプロシージャーを実行するためのダイアログにアクセスできるのは、すべてのフィールドに測定レベルが定義されてからです。

方法

「方法」タブでは、欠損値がどのように代入されるか (使用されるモデルの種類を含む) を指定します。カテゴリー予測値は、(ダミー) コード化された指標です。

代入方法。「自動」方法では、データをスキャンして、データが単調な欠損値パターンを示す場合には単調方法を使用します。それ以外の場合は、完全条件指定が使用されます。使用する方法が確定している場合、その方法を「カスタム」方法として指定することができます。

- **完全条件指定。**これは、反復によるマルコフ連鎖モンテカルロ (MCMC) 法であり、欠損データのパターンが任意 (単調または非単調) の場合に使用できます。

変数リストで指定された順序での反復ごとおよび変数ごとに、完全条件指定 (FCS) 方法は、モデル内の他のすべての使用可能な変数を予測値として使用して 1 変数 (1 従属変数) モデルを当てはめてから、当てはまった変数について欠損値を代入します。この方法は最大反復回数に達するまで続き、最大反復回数での代入値が、代入データ・セットに保存されます。

最大反復回数。これは、FCS 方法で使用されるマルコフ連鎖がとる反復の回数 (ステップの数) を指定します。FCS 方法が自動的に選択された場合は、デフォルトである 10 を反復回数とします。FCS を明示的に選択した場合は、反復回数をカスタマイズできます。マルコフ連鎖が収束していない場合は、反復回数を増やす必要がある場合があります。「出力」タブでは、FCS 反復履歴データを保存してプロットすることで、収束を評価することができます。

- **単調。**これは非反復方法であり、データに単調な欠損値パターンがある場合のみ使用できます。単調なパターンが存在するのは、変数を順序付けできる場合です。例えば、変数に非欠損値があり、先行する

すべての変数にも非欠損値がある、という場合です。この方法を「カスタム」方法として指定する場合は、変数を、単調なパターンを示す順序でリストに指定してください。

単調な順序の変数ごとに、単調方法は、モデル内のすべての先行する変数を予測値として使用して 1 変量 (1 従属変数) モデルを当てはめてから、当てはまった変数について欠損値を代入します。こうした代入値は、代入データ・セットに保存されます。

2 次の交互作用を含める。 代入方法が自動的に選択される場合、各変数の代入モデルには、予測値変数についての定数項や主効果が含まれます。特定の方法を選択する場合は、オプションで、カテゴリ予測値変数の中から 2 次の交互作用として可能なものをすべて含めることができます。

スケール変数のモデル・タイプ。 代入方法が自動的に選択される場合、スケール変数の 1 変量モデルとして、線型回帰が使用されます。特定の方法を選択する場合は、代わりに、スケール変数のモデルとして、予測平均マッチング (PMM) を選択することができます。PMM は、回帰モデルで計算される代入値を、最も近い観測値に一致させる、線型回帰のバリエーションです。

ロジスティック回帰は常に、カテゴリ変数についての 1 変量モデルとして使用されます。モデルの種類にかかわらず、カテゴリ予測値は、指標の (ダミー) コード化により処理されます。

特異性許容度。 特異な (または非可逆的な) 行列に線型従属列があると、このことが推定アルゴリズムで重大な問題を引き起こす可能性があります。特異に近い行列の場合でも十分な結果が得られなくなる可能性があるため、プロシージャーでは、行列式が許容度を下回る行列については、特異として処理することになります。正の値を指定してください。

制約

「制約条件」タブにより、代入時の変数の役割を制限することができます。また、スケール変数の代入値を妥当な範囲に制限することもできます。加えて、分析対象を欠損値の最大パーセントより少ない変数に制限することもできます。

変数の集計データをスキャン。 「データをスキャン」をクリックすると、リストには、分析変数と、それぞれについて、観測された欠損のパーセント、最小値、および最大値が表示されます。集計は、「ケース」テキスト・ボックスでの指定に応じて、すべてのケースに基づく場合もあれば、最初の n ケースのスキャンに制限される場合もあります。「データを再スキャン」をクリックすると、分布の集計が更新されます。

制約の定義

- **役割。** これによりカスタマイズした変数のセットは、代入や、予測値としての扱いが可能になります。通常、各分析変数は、代入モデルの予測値および従属変数の両方であると考えられています。「役割」を使用して、「予測変数としてのみ使用」する変数の代入をオフにするか、変数が予測値として使用されないようにする (「代入のみ」) ことにより、予測モデルをよりコンパクトにすることができます。これは、カテゴリ変数や予測値としてのみ使用される変数に対して指定することのできる、唯一の制約です。
- **最小と最大。** これらの列では、スケール変数に対し、最大および最小の代入値として許容される値を指定することができます。代入値がこの範囲の外になる場合、プロシージャーでは、別の値を引き出します。これは、範囲内にある値を見つけるか、引き出しの最大数に達する (後述の「最大引き出し数」を参照) まで実行されます。これらの列は、「線型回帰」がスケール変数モデルの種類として「方法」タブで選択されている場合のみ使用可能です。
- **丸め。** 変数の中にはスケールとして使用されることのあるものがありますが、値はその特性上、さらなる制限を受けます。例えば、世帯の人数は整数である必要がありますし、スーパーなどで支払う金額に

は貨幣の最小単位の小数点以下の数値は出現しません。この列には、受容可能な最小単位を指定することができます。例えば、整数値を得るために丸め単位として 1 を指定する、近似値の銭に丸めた値を得るために 0.01 を指定するなどです。一般的には、丸め単位の整数倍で最も近い値に丸められます。作用する丸め値によって代入値 6.64823 (丸め前) がどのように異なってくるかを以下の表に示します。

表 2. 丸めの結果：

丸め単位	6.64832 が丸められたあとの値
10	10
1	7
0.25	6.75
0.1	6.6
0.01	6.65

大量の欠損データを持つ変数を除外。 通常、分析変数は、代入モデルを推定するのに十分なデータがある場合、欠損値の数に関わらず、代入されたり、予測値として使用されたりします。欠損値のパーセントが高い変数を除外する、という選択をすることができます。例えば、「欠損の最大パーセント」に 50 を指定すると、欠損値が 50% を超える分析変数は、代入されることも、代入モデルで予測値として使用されることもなくなります。

最大引き出し数。 スケール変数の代入値に最小値や最大値が指定されている場合 (前述の「最小と最大」を参照)、プロシージャーでは、指定された範囲内にある値のセットを見つけるまで、ケースについての値の引き出しを試行します。1 ケースにつき指定されている引き出し数以内で値のセットを得られない場合、プロシージャーでは別のモデル・パラメーターのセットを引き出し、ケース引き出し処理を繰り返します。範囲内にある値のセットが、ケースおよびパラメーター引き出し数として指定されている回数以内で得られない場合は、エラーが発生します。

こうした値を増やすと処理時間も長くなることに注意してください。プロシージャーに時間がかかる場合、または引き出し値としてふさわしいものを見つけることができない場合、指定されている最小値と最大値を調べて、これらが妥当かどうかを確認してください。

出力

表示。 出力の表示を制御します。代入についての全体的な要約が常に表示されます。これには、代入指定、反復 (完全条件指定方法の場合)、代入された従属変数、代入から除外された従属変数、および代入順序に関連する表が含まれます。分析変数の制約が指定されている場合は、これも表示されます。

- **代入モデル。** ここでは、従属変数と予測値の代入モデルに加え、1 変数モデルの種類、モデル効果、および代入される値の数が表示されます。
- **記述統計。** ここでは、値が代入される従属変数の記述統計が表示されます。スケール変数の場合、記述統計には、(代入前の) 元の入力データの平均値、個数、標準偏差、最小値、および最大値に加え、(代入による) 代入値と完全なデータ (元の値と代入値 (代入による)) が含まれます。カテゴリ変数の場合、記述統計には、(代入前の) 元の入力データの個数とカテゴリ別のパーセントに加え、(代入による) 代入値と完全なデータ (元の値と代入値 (代入による)) が含まれます。

反復履歴。 完全条件指定による代入方法が使用される場合は、FCS による代入の反復履歴データを含むデータ・セットを要求することができます。データ・セットには、反復による平均値と標準偏差、および値が代入される各スケール従属変数についての代入が含まれます。データをプロットすると、モデル収束の評価の助けとなります。詳しくは、トピックを参照してください。

MULTIPLE IMPUTATION コマンドの追加機能

このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変数のサブセットを指定して、その記述統計量を表示する (IMPUTATIONSUMMARIES サブコマンド)。
- 欠損パターンの分析、および代入の両方を、プロシーチャーの実行 1 回で指定する。
- モデル・パラメーターについて、何らかの変数を代入する場合に許容される最大数を指定する (MAXMODELPARAM キーワード)。

シンタックスの完全な情報については、シンタックス参照コマンド を参照してください。

多重代入データでの作業

多重代入 (MI) データ・セットが作成されると、変数 *Imputation_* (変数ラベルは 代入番号) が追加され、これによりデータ・セットが昇順にソートされます。元のデータ・セットからのケースは、値 0 です。代入値のケースには、1 から *M* の番号が付与されます。ここで、*M* は代入の数です。

データ・セットを開いた際に *Imputation_* があるのは、そのデータ・セットが MI データ・セットである可能性がある、ということです。

分析用の多重代入データ・セットのアクティブ化

データ・セットは、「グループの比較」 オプションを使用して (*Imputation_* をグループ化変数として) 分割する必要があります。分析において MI データ・セットとして扱われるようにするためです。他の変数に分割を定義することもできます。

メニューから次の項目を選択します。

「データ」 > 「分割ファイル...」

1. 「グループの比較」を選択します。
2. ケースのグループ化の基準とする変数として、代入番号 (*Imputation_*) を選択します。

また、マーキングをオンにする (後述参照) と、代入番号 (*Imputation_*) を基準としてファイルが分割されます。

観測値と代入値との区別

セルの背景色、フォント、太字などを代入値に適用することで、代入値と観測値とを区別することができます。「欠損値の代入」の現行セッションで新規のデータ・セットを作成した場合、マーキングはデフォルトでオンになります。保存したデータ・ファイルを開き、それが代入を含んでいる場合、マーキングはオフになります。

マーキングをオンにするには、データ・エディターのメニューから以下の項目を選択します。

「表示」 > 「代入データのマーク...」

または、データ・エディターのデータ・ビューで編集バーの右端にある代入マーキング・ボタンをクリックすることで、マーキングをオンにすることもできます。

代入間での移動

1. メニューから次の項目を選択します。

「編集」 > 「代入に移動...」

2. ドロップダウン・リストから代入 (または元のデータ) を選択します。

または、データ・エディターのデータ・ビューにある編集バーのドロップダウン・リストから代入を選択することもできます。

ケースの相対的な位置は、代入を選択する際にも保持されます。例えば、元のデータ・セットに 1000 ケースある場合、ケース 1034 (1 番目の代入での 34 番目のケース) がグリッドの先頭に表示されます。ドロップダウンで代入 2 を選択した場合、ケース 2034 (2 番目の代入での 34 番目のケース) がグリッドの先頭に表示されることとなります。ドロップダウンで「元のデータ」を選択した場合、ケース 34 がグリッドの先頭に表示されることとなります。異なる代入同士で値を容易に比較できるように、代入間をナビゲートする際にも列の位置は保持されます。

代入値の変換と編集

代入データに対して変換を行うことが必要な場合があります。例えば、給与変数のすべての値のログを取得して、結果を新しい変数に保存する場合などです。代入データを使用して計算された値は、元のデータを使用して計算した値と異なる場合、代入済みとして扱われます。

データ・エディターのセルで代入値を編集した場合でも、そのセルは代入済みとして扱われます。このような方法で代入値を編集することは推奨されていません。

多重代入データの分析

多重代入データ・セットの分析から得られた結果のプールは、多くのプロシージャでサポートされています。代入マーキングが有効な場合、プールをサポートするプロシージャの横に、特殊なアイコンが表示されます。例えば、「分析」メニューの「記述統計」サブメニューにある「度数」、「記述統計」、「探索」、および「クロス集計表」はすべてプールをサポートしますが、「比率」、「P-P プロット」、および「Q-Q プロット」はサポートしません。

表形式出力とモデル PMML の両方をプールすることができます。プールされた出力を要求するための新しいプロシージャはありません。代わりに、「オプション」ダイアログの新しいタブで、多重代入の出力を包括的に制御できます。

- **表形式出力のプール。** デフォルトでは、多重代入 (MI) データ・セットに対してサポートされているプロシージャを実行する場合、結果は自動的に、代入ごとに、元の (代入されていない) データについて、および代入間での変動を考慮に入れてプールされた (最終) 結果について生成されます。プールされた統計は、プロシージャにより異なってきます。
- **PMML のプール。** プール PMML は、PMML をエクスポートするサポート対象プロシージャで取得することもできます。プール PMML は、非プール PMML と同じ方法で要求され、非プール PMML の代わりに保存されます。

サポート対象でないプロシージャは、プール出力とプール PMML ファイルのいずれも生成しません。

プールのレベル

出力は、以下の 2 つのレベルのいずれかを使用してプールされます。

- **Naive 組み合わせ。** プールされたパラメーターのみ使用可能です。
- **1 変量の組み合わせ。** プールされたパラメーターとその標準誤差、検定統計および有効な自由度、 p 値、信頼区間、およびプールの診断 (欠損情報の割合、相対効率、分散の相対増加) が (使用可能な場合に) 表示されます。

係数 (回帰と相関)、平均値 (および平均値の差)、および個数は、通常はプールされます。統計の標準誤差が使用可能な場合は、1 変量プールが使用されます。そうでない場合は、Naïve プールが使用されます。

プールをサポートするプロシージャ

以下のプロシージャでは、MI データ・セットが、それぞれの出力に指定されたプールのレベルでサポートされています。

度数。以下のフィーチャーがサポートされています。

- 「統計量」テーブルでは、1 変量プールにおける平均値 (標準誤差の平均値も要求されている場合) と、Naïve プールにおける有効 N と欠損値 N がサポートされています。
- 度数分布表では、Naïve プールにおける度数がサポートされています。

記述統計。以下のフィーチャーがサポートされています。

- 「記述統計量」テーブルでは、1 変量プールにおける平均値 (標準誤差の平均値も要求されている場合) と、Naïve プールにおける N がサポートされています。

クロス集計表。以下のフィーチャーがサポートされています。

- クロス集計表では、Naïve プールにおける個数がサポートされています。

平均値。以下のフィーチャーがサポートされています。

- 「レポート」テーブルでは、1 変量プールにおける平均値 (標準誤差の平均値も要求されている場合) と、Naïve プールにおける N がサポートされています。

1 サンプルの T 検定。以下のフィーチャーがサポートされています。

- 「統計量」テーブルでは、1 変量プールにおける平均値と、Naïve プールにおける N がサポートされています。
- 「検定」テーブルでは、1 変量プールにおける平均値の差がサポートされています。

独立したサンプルの T 検定。以下のフィーチャーがサポートされています。

- 「グループ統計量」テーブルでは、1 変量プールにおける平均値と、Naïve プールにおける N がサポートされています。
- 「検定」テーブルでは、1 変量プールにおける平均値の差がサポートされています。

対応のあるサンプルの T 検定。以下のフィーチャーがサポートされています。

- 「統計量」テーブルでは、1 変量プールにおける平均値と、Naïve プールにおける N がサポートされています。
- 「相関」テーブルでは、Naïve プールにおける相関と N がサポートされています。
- 「検定」テーブルでは、1 変量プールにおける平均値がサポートされています。

一元配置分散分析。以下のフィーチャーがサポートされています。

- 「記述統計量」テーブルでは、1 変量プールにおける平均値と、Naïve プールにおける N がサポートされています。
- 「対比の検定」テーブルでは、1 変量プールにおける対比の値がサポートされています。

線型混合モデル。以下のフィーチャーがサポートされています。

- 「記述統計量」テーブルでは、Naïve プールにおける平均値と N がサポートされています。
- 「固定効果の推定値」テーブルでは、1 変量プールにおける推定値がサポートされます。

- 「共分散パラメーターの推定値」テーブルでは、1 変量プールにおける推定値がサポートされます。
- 「推定周辺平均: 推定」テーブルでは、1 変量プールにおける平均値がサポートされています。
- 「推定周辺平均: ペアごとの比較」テーブルでは、1 変量プールにおける平均値の差がサポートされています。

一般化線型モデルと一般化推定方程式。 これらのプロシージャーでは、プール PMML がサポートされています。

- 「カテゴリ変数情報」テーブルでは、Naïve プールにおける N とパーセントがサポートされています。
- 「連続型変数情報」テーブルでは、Naïve プールにおける N と平均値がサポートされています。
- 「パラメーター推定値」テーブルでは、1 変量プールにおける係数と B がサポートされています。
- 「推定周辺平均: 推定係数」テーブルでは、Naïve プールにおける平均値がサポートされています。
- 「推定周辺平均: 推定」テーブルでは、1 変量プールにおける平均値がサポートされています。
- 「推定周辺平均: ペアごとの比較」テーブルでは、1 変量プールにおける平均値の差がサポートされています。

2 変量の相関分析。 以下のフィーチャーがサポートされています。

- 「記述統計量」テーブルでは、Naïve プールにおける平均値と N がサポートされています。
- 「相関」テーブルでは、1 変量プールにおける相関と N がサポートされています。相関はプール前に Fisher の z 変換を使用して変換され、プール後に後方変換される、ということに注意してください。

偏相関。 以下のフィーチャーがサポートされています。

- 「記述統計量」テーブルでは、Naïve プールにおける平均値と N がサポートされています。
- 「相関」テーブルでは、Naïve プールにおける相関がサポートされています。

線型回帰。 このプロシージャーでは、プール PMML がサポートされています。

- 「記述統計量」テーブルでは、Naïve プールにおける平均値と N がサポートされています。
- 「相関」テーブルでは、Naïve プールにおける相関と N がサポートされています。
- 「係数」テーブルでは、1 変量プールにおける B と、Naïve プールにおける相関がサポートされています。
- 「相関係数」テーブルでは、Naïve プールにおける相関がサポートされています。
- 「残差の統計量」テーブルでは、Naïve プールにおける平均値と N がサポートされています。

2 項ロジスティック回帰。 このプロシージャーでは、プール PMML がサポートされています。

- 「式に含まれる変数」テーブルでは、1 変量プールにおける B がサポートされています。

多項ロジスティック回帰。 このプロシージャーでは、プール PMML がサポートされています。

- 「パラメーター推定値」テーブルでは、1 変量プールにおける係数と B がサポートされています。

順序回帰。 以下のフィーチャーがサポートされています。

- 「パラメーター推定値」テーブルでは、1 変量プールにおける係数と B がサポートされています。

判別分析。 このプロシージャーでは、プールされたモデル XML がサポートされています。

- 「グループ統計量」テーブルでは、Naïve プールにおける平均値と有効 N がサポートされています。
- 「プールされたグループ内行列」テーブルでは、Naïve プールにおける相関がサポートされています。

- 「標準判別関数係数」テーブルでは、Naïve プールにおける非標準化係数がサポートされています。
- 「グループ重心の関数」テーブルでは、Naïve プールにおける非標準化係数がサポートされています。
- 「分類関数係数」テーブルでは、Naïve プールにおける係数がサポートされています。

カイ 2 乗検定。 以下のフィーチャーがサポートされています。

- 「記述」テーブルでは、Naïve プールにおける平均値と N がサポートされています。
- 度数分布表では、Naïve プールにおいて観測された N がサポートされています。

2 項検定。 以下のフィーチャーがサポートされています。

- 「記述」テーブルでは、Naïve プールにおける平均値と N がサポートされています。
- 「検定」テーブルでは、Naïve プールにおける N、観測比率、および検定比率がサポートされています。

ラン検定。 以下のフィーチャーがサポートされています。

- 「記述」テーブルでは、Naïve プールにおける平均値と N がサポートされています。

1 サンプル Kolmogorov-Smirnov 検定。 以下のフィーチャーがサポートされています。

- 「記述」テーブルでは、Naïve プールにおける平均値と N がサポートされています。

2 個の独立サンプルの検定。 以下のフィーチャーがサポートされています。

- 「ランク」テーブルでは、Naïve プールにおける平均ランクと N がサポートされています。
- 度数分布表では、Naïve プールにおける N がサポートされています。

複数の独立サンプルの検定。 以下のフィーチャーがサポートされています。

- 「ランク」テーブルでは、Naïve プールにおける平均ランクと N がサポートされています。
- 度数分布表では、Naïve プールにおける個数がサポートされています。

2 個の対応サンプルの検定。 以下のフィーチャーがサポートされています。

- 「ランク」テーブルでは、Naïve プールにおける平均ランクと N がサポートされています。
- 度数分布表では、Naïve プールにおける N がサポートされています。

複数の関連サンプルの検定。 以下のフィーチャーがサポートされています。

- 「ランク」テーブルでは、Naïve プールにおける平均ランクがサポートされています。

Cox 回帰。 このプロシージャーでは、プール PMML がサポートされています。

- 「式に含まれる変数」テーブルでは、1 変量プールにおける B がサポートされています。
- 「共変量の平均値」テーブルでは、Naïve プールにおける平均値がサポートされています。

多重代入オプション

「多重代入」タブでは、多重代入に関連する 2 種類の設定を制御します。

代入されたデータの外観。 デフォルトでは、代入されたデータを含むセルと、代入されていないデータを含むセルとは、背景色が異なります。代入されたデータの外観は独特のものになっているので、そのセルは、データ・セットをスクロールすれば容易に探し出すことができます。デフォルトのセル背景色とフォントは、変更することができます。また、代入されたデータを太字で表示することもできます。

分析の出力。 このグループは、多重代入されたデータ・セットが分析されるたびに生成されるビューアー出力のタイプを制御します。デフォルトでは、元の (代入を行う前の) データ・セットと、代入された各データ・セットについて、出力が作成されます。また、代入されたデータのプールをサポートするプロシージャで生成される、プールされた結果は、最終的なものとなります。1 変量プールが実行されると、プールの診断も表示されます。ただし、参照しない出力は非表示にすることができます。

多重代入オプションの設定方法

メニューから以下の項目を選択します。

「編集」 > 「オプション」

「多重代入」タブをクリックします。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。

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

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

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

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

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

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

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

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

IBM Software Group

ATTN: Licensing

200 W. Madison St.

Chicago, IL; 60606

U.S.A.

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

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

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

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

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

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

著作権使用許諾:

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

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

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

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

商標

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

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

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

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

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

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

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

索引

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

[カ行]

- 回帰
 - 欠損値分析での 7
- カテゴリの集計
 - 欠損値分析での 5
- 完全条件指定
 - 多重代入での 14
- 共分散
 - 欠損値分析での 7
- 極値の個数
 - 欠損値分析での 5
- ケースの集計
 - 欠損値分析での 4
- ケースのソート
 - 欠損値分析での 4
- 欠損値
 - 1 変量統計量 5
- 欠損値分析 3
 - 回帰 7
 - 記述統計 5
 - 期待値最大化 8
 - 欠損値への代入 6
 - コマンドの追加機能 8
 - 統計量の推定 6
 - パターン 4
 - 方法 6
 - EM 7
 - MCAR 検定 6
- 欠損指示変数
 - 欠損値分析での 5
- 欠損データ値の代入 13
 - 出力 16
 - 制約 15
 - 代入方法 14

[サ行]

- 残差
 - 欠損値分析での 7
- 指示変数
 - 欠損値分析での 5
- スチューデント t 検定
 - 欠損値分析での 7

- 正規変量
 - 欠損値分析での 7
- 相関
 - 欠損値分析での 7

[タ行]

- 多重代入 11, 17, 18
 - 欠損データ値の代入 13
 - パターン分析 12
- 単調代入
 - 多重代入での 14
- 度数分布表
 - 欠損値分析での 5

[ハ行]

- パターン分析 12
- 反復履歴
 - 多重代入での 16
- 標準偏差
 - 欠損値分析での 5
- 不一致
 - 欠損値分析での 5
- 不完全なデータ
 - 「欠損値分析」を参照 3
- ペアワイズ除去
 - 欠損値分析での 3
- 平均値
 - 欠損値分析での 5, 7

[ラ行]

- リストワイズ除去
 - 欠損値分析での 3

E

- EM
 - 欠損値分析での 7

L

- Little の MCAR 検定 6
 - 欠損値分析での 3

M

- MCAR 検定
 - 欠損値分析での 3

T

- t 検定
 - 欠損値分析での 5



Printed in Japan