



---

## 特長

- データセットの調査とクレンジングのために使用される 2 つのプロシージャ（「データの検証」と「重複ケースの特定」）について学びます。
  - SPSS Statistics のオンラインのヘルプ・メニューから役に立つケース スタディを参照します。
  - さらに高いスキルを習得するために、一般に公開されているトレーニングについて確認します。
- 

# クイック・スタート・ガイド

## IBM SPSS Statistics の使用を開始する

### はじめに

IBM® SPSS® Statistics とは、汎用的なデータ分析（レポートイング、可視化、統計分析を含む）のための機能を統合したソフトウェア製品群です。

スプレッドシートのユーザーの多くは、データの容量が大きすぎてスプレッドシートでは対応できないため SPSS Statistics に移行しています。複雑なスプレッドシートの作成とメンテナンスに要する時間を懸念するユーザーもいれば、品質保証上の問題を抱えているユーザーもいます。IBM SPSS Statistics は統計の専門家が開発したレポートイングと統計分析のためのさまざまな標準のプロシージャを備えた、分析結果の品質を保証するために専門家がテスト済みの製品です。

本ガイドでは、データセットの調査とクレンジングを行うための 2 つのプロシージャ（「データの検証」と「重複ケースの特定」）について紹介します。

\*本ガイド内では英語版の画面が掲載されていますが、日本語版をダウンロードいただいた場合は日本語版の画面が表示されます。



## データの検証

### ステップ 1

「データの検証」プロシージャは、疑わしいケースか無効なケース、変数、およびデータ値を特定します。

- データ・ファイルの telco\_missing.sav を開きます。本ファイルは、C ドライブの ProgramFiles/IBM/SPSS/23/Samples/English またはプログラムのインストール・ドライブにあります。  
\*ProgramFiles/IBM/SPSS/Statistics/24/Samples/Japanese を選択いただくと日本語のデータをご利用いただけます。
- Statistics データ エディタのウィンドウで「データ」をクリックします。
- 「検証」に移動します。
- 「事前定義の規則を読み込み」をクリックします。[注意: この処理は、本プロシージャを初めて使用する際に行います。Statistics が、これまで定義した規則を置き換えることを示す警告メッセージを表示します。カスタム ルールを定義していないため、このメッセージは無視します。IBM SPSS Statistics が通常使用される標準の規則をプリロードします。]
- 「OK」をクリックします。
- 再度 Statistics のデータ エディタのメニューから「データ」をクリックします。
- 「検証」を選択します。
- 「データの検証」をクリックします (図 1)。

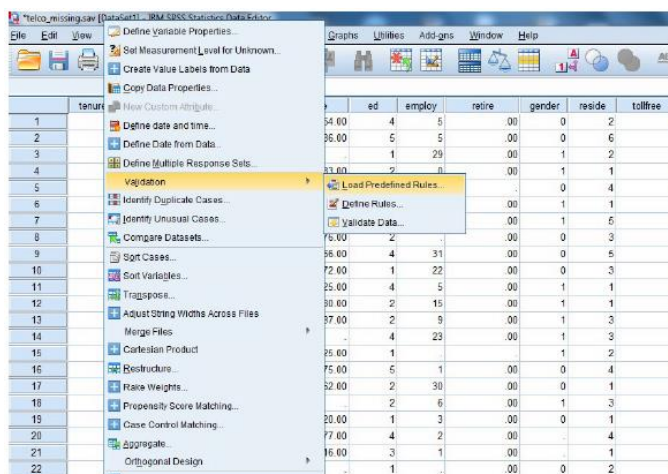


図 1: Statistics のメニューから「データの検証」を選択

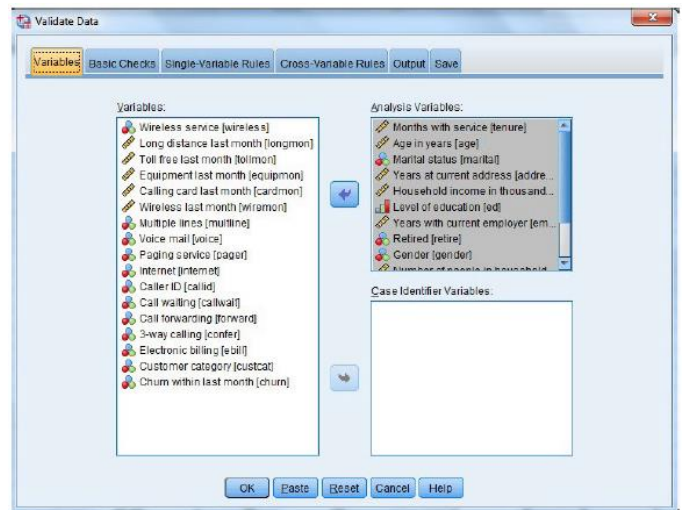


図 2: 「データの検証」プロシージャの設定: telco\_missing.sav ファイルから「勤続月数 (Months with service)」から「世帯人数 (Number of people)」までの変数を選択し、「データの検証」ダイアログ ウィンドウ内の「分析変数」ボックスに配置します。

### ステップ 2

一連の事前に定義された規則が提供され、ユーザーは必要に応じて規則 (単純な規則と複雑な規則を含む) を追加できます。

- 左側のリストから検証したいすべての変数を選択します。
- この変数を「分析変数」リストに移動します (図 2)。
- 「単一変数規則」タブをクリックします。
- このタブは分析変数の統計情報のサマリーを示すため、個々の変数に基づいて検証規則を設定できます。
- 「規則の定義」ボタンをクリックして、新しい規則の修正や追加を行います。このタブの情報を確認するものの、規則は設定しません。
- 「続行」をクリックします。

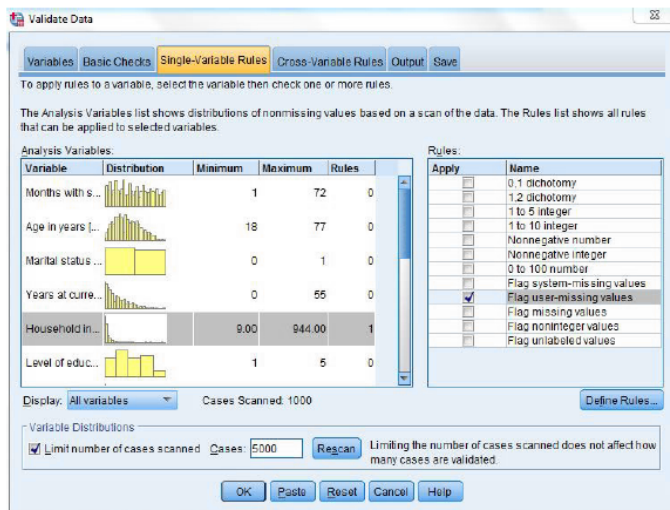


図 3: 選択した変数に規則を適用: 「データの検証」ダイアログ ボックスの「単一変数規則」において、変数の「世帯収入 (Household income、単位: 千ドル)」が分析変数として選択されています。規則リストでは、「ユーザー欠損値にフラグを付ける」が選択されています。

### ステップ 2 (続き)

「分析変数」リストから変数を選択します。

- 規則リストで対象となるボックスにチェックを入れ、選択した変数に規則を適用します (図 3)。
- 「クロス変数規則」のタブをクリックします。[注意: このタブは 2 つ以上の変数に関する規則を指定するために使用します。この論理式は、単一では有効に見えても、組み合わせると無効になる変数値を定義します。この典型的な例としては、性別が男性であり、かつ、妊娠検査の結果が陽性である場合が挙げられます。ここではこのタブの情報を確認しますが、規則は設定しません。]
- 「規則の定義」ボタンをクリックします。
- 「続行」をクリックします。
- 「OK」をクリックします。

### ステップ 3

「データの検証」プロシージャで作成する複数のテーブルは、さまざまな方法で規則違反のデータも使用中の規則を違反していない場合は、規則の違反がない旨を示すメッセージが生成されます。

- 分析変数に基づいて「変数の要約」が提供されます。
- 変数ごとに違反があった規則のリストが示されます。
- 最後の列は特定の変数ごとの規則の違反回数を示します (図 4)。

Variable Summary

	Rule	Number of Violations
Household income in thousands	Flag missing values	179
	Total	179

Case Report<sup>b</sup>

Case	Validation Rule Violations
	Single-Variable <sup>a</sup>
3	Flag missing values (1)
14	Flag missing values (1)
18	Flag missing values (1)
22	Flag missing values (1)
26	Flag missing values (1)
32	Flag missing values (1)
40	Flag missing values (1)
51	Flag missing values (1)
58	Flag missing values (1)
59	Flag missing values (1)

図 4: 変数の要約とケース報告書: 出力ウィンドウは、「データの検証」ダイアログ ウィンドウの規則で特定した要約と個々のケースを表示します。

## 重複ケースの特定

### ステップ 1

「重複ケースの特定」プロシージャを使用すると、選択した任意の変数に基づいて重複値を持つケースを特定できます。

- stroke\_invalid.sav ファイルを開きます。本ファイルは、SPSS Statistics の標準のインストール・ディレクトリーにあります。
- Statistics のデータ エディタ ウィンドウで「データ」をクリックします。
- 「重複ケースの特定」をクリックします (図 5)。

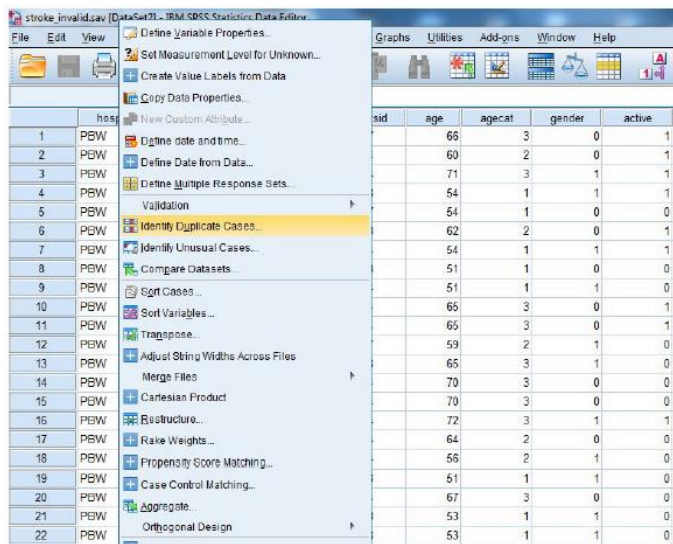


図 5: Statistics のメニューで「重複ケースの特定」を選択

### ステップ 2

プライマリ ケースを自動的に特定するための規則を設定できます。例えば、作成日が最近のケースをプライマリ ケースとして特定したい場合があります。同じグループ内で作成日がそれより前のケースは重複ケースとなります。

- 左側のリストから重複ケースを定義する変数を選択します。
- 全く同じケース (すべてのフィールドが同じケース) を検索するには、すべての変数を選択します。ほとんどの場合、データセット内の一部の変数のみを使用して重複ケースを定義します。
- この変数を「一致するケースを定義」のリストに移動します (図 6)。
- 「OK」をクリックします。

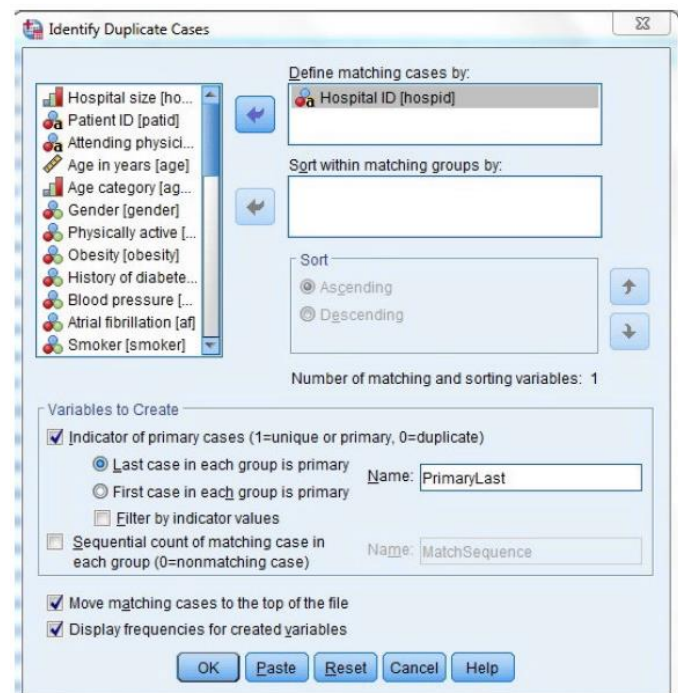


図 6: 「重複ケースの特定」プロシージャの設定: stroke\_invalid.sav ファイルから一連の変数を選択し、「重複ケースの特定」ダイアログ ウィンドウ内の「一致するケースを定義」ボックスに配置します。

### ステップ 3

データセットのうちどれくらいが重複ケースなのか確認できます。

- 出力ウィンドウで、「プライマリとしての最後に一致するケースの識別子」テーブルが結果を示します (図 7)。
- これは度数に基づくテーブルであり、何件のケースがプライマリ ケースで、何件のケースが重複ケースか示します。
- プライマリ ケースの標準の定義では、値が一致する一連のケースの中で最後のケースを指します。この定義を変更して最初のケースを使用することができ、他の変数を使用してグループをソートすることで、より柔軟にプライマリ ケースを定義することができます。

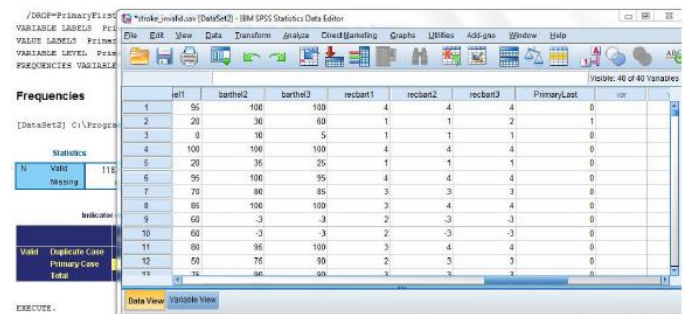


図 8: データ エディタが示す重複ケース

Statistics		
N	Valid	1183
	Missing	0

Indicator of each last matching case as Primary					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Duplicate Case	1167	98.6	98.6	98.6
	Primary Case	16	1.4	1.4	100.0
	Total	1183	100.0	100.0	

図 7: プライマリ ケースとして最後に一致する各ケースの指標: 出力ウィンドウは、「重複ケースの特定」ダイアログ ウィンドウ内で「一致するケースを定義」が特定したプライマリ ケースと重複ケースのサマリーを提供します。

### ステップ 4

さらにデータを操作するには、指示変数が役に立ちます。

- データセット内で新規の変数を作成し、プライマリ ケースか重複ケースか示すことができます (図 8)。[データセットとともにこの変数を保存できます。通常、分析のためにデータをフィルタリングし重複ケースを削除するために、この指示変数を使用することができます。]

## IBM SPSS Statistics のケース スタディ

### データの検証

IBM SPSS Statistics のメニューから、「ヘルプ」->「トピック」->「Case Studies (ケース スタディ)」とクリックします。ケース スタディの Introduction (概要) が開きます。「Data preparation Option (データ準備のオプション)」を展開し、「Validate Data (データの検証)」を選択します。

### 重複ケースの特定

IBM SPSS Statistics のメニューから、「ヘルプ」->「トピック」->「Case Studies (ケース スタディ)」とクリックします。ケース スタディの概要が開きます。「Data preparation Option (データ準備のオプション)」を展開し、「Validate Data (データの検証)」を展開し、「Validating a Medical Database (医療データベースの検証)」を選択します。

## IBM SPSS について一般に公開されている トレーニング

### IBM SPSS Statistics 入門

2 日間の IBM SPSS Statistics の基本的な操作を習得するためのコースです。データの入力と定義、いろいろなファイルからのデータの読み込み、データの集計、グラフ表現などができるようになります。タイトルは『入門』ですが、コースに含まれる内容は非常に充実しています。「Statistics をすぐに使いたい!」「Statistics の基本操作をしっかりマスターしたい!」「効率よく作業を進めたい」といった方にお勧めです。

### IBM SPSS Statistics による基礎統計

2 日間のコースです。データ分析を行うにあたり、必要となる統計的な考え方を Statistics を使用しながら習得して頂きます。はじめに、データのエラーチェックをオプションの IBM SPSS Data Preparation を使用して行います。つぎに、クリーニングされたデータを使用し分析を行う際に把握すべきデータの特徴について、統計用語の説明を交えながら紹介します。Statistics を使った分析の目的は、分析の対象となる集団(母集団)の傾向を知ることです。その考え方を理解し、データの性質に合わせて統計的なデータ(2変量)間の関係性を調べる手順と結果の解釈の仕方を習得します。使用する手法は、カイ 2 乗検定、相関、t 検定です。

### 詳細情報

詳細については、IBM SPSS のトレーニングに関する Web サイトを参照ください。



© Copyright IBM Corporation 2015

日本アイ・ビー・エム株式会社  
IBM Analytics

〒103-8510  
東京都中央区日本橋箱崎町 19-21

Produced in Japan  
2015 年 5 月

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、[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) にある「著作権および商標の情報」をご覧ください。

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とは限りません。

本書に含まれる情報は、特定物として現存するままの状態を提供され、商品性の保証、特定目的適合性の保証、および第三者の権利の不侵害の保証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。



Please Recycle