バージョン12

基本的な統計分析

「真の発見の旅とは、新しい風景を探すことではなく、新たな視点を持つことである。」
マルセル・プルースト

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

このマニュアルを引用する場合は、次の正式表記を使用してください: SAS Institute Inc. 2015. 『JMP® 12 基本的な統計分析』Cary, NC: SAS Institute Inc.

JMP® 12 基本的な統計分析

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

2015年3月

技術ライセンスに関する通知

- Scintilla - Copyright © 1998-2014 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

- Telerik RadControls: Copyright © 2002-2012, Telerik. Usage of the included Telerik RadControls outside of JMP is not permitted.
- ZLIB Compression Library - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS

OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- bzip2, the associated library "libbzip2", and all documentation, are Copyright © 1996-2010, Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- R software is Copyright © 1999-2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.

  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED.IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE

LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar licence but with different Copyright notices) all the files are:

  Copyright © 1998 - 2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT.IN NO EVENT SHALL THE DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

目次
基本的な統計分析

10 ブートストラップ

11 表の作成

12 モデル化ユーティリティ

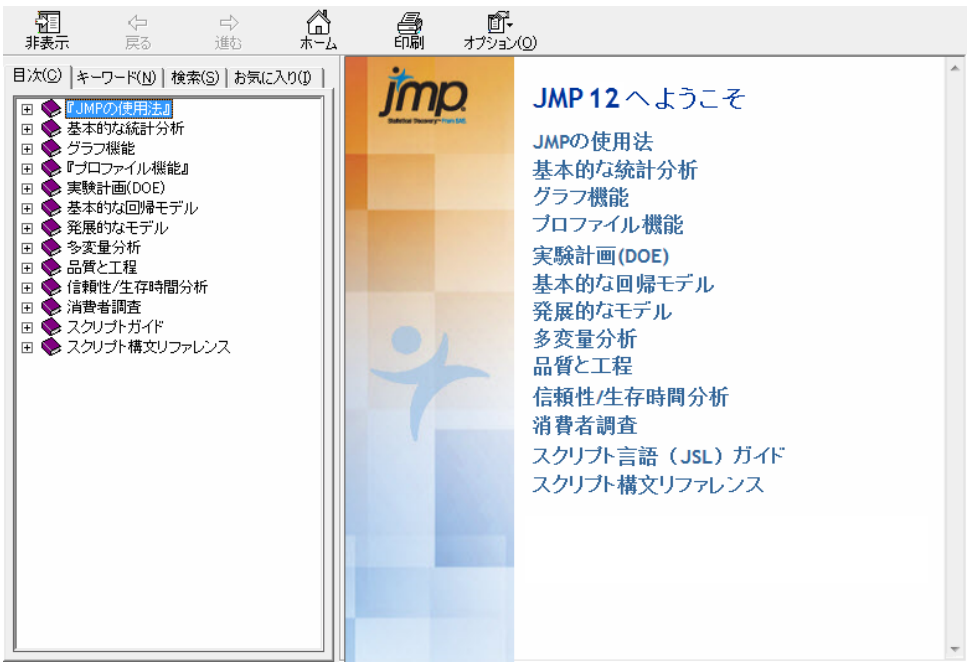
第 1 章

JMP の概要
マニュアルとその他のリソース

この章には以下の情報が記載されています。

- 本書の表記法
- JMP のマニュアル
- JMP ヘルプ
- その他のリソース
  - その他の JMP のドキュメンテーション
  - チュートリアル
  - 索引
  - Web リソース

図 1.1 JMP ヘルプのホームウィンドウ（Windows）



目次

表記規則

マニュアルの内容と画面に表示される情報を対応付けるために、次の表記規則を使っています。

- サンプルデータ名、列名、パス名、ファイル名、ファイル拡張子、およびフォルダ名は「」で囲んで表記しています。
- スクリプトのコードは`Lucida Sans Typewriter`フォントで表記しています。
- スクリプトコードの結果（ログに表示されるもの）は`Lucida Sans Typewriter`（斜体）フォントで表記し、先に示すコードよりインデントされています。
- クリックまたは選択する項目は [] で囲んで太字で表記しています。これには以下の項目があります。
  - ボタン
  - チェックボックス
  - コマンド
  - 選択可能なリスト項目
  - メニュー
  - オプション
  - タブ名
  - テキストボックス
- 次の項目は太字で表記しています。
  - 重要な単語や句、JMP に固有の定義を持つ単語や句
  - マニュアルのタイトル
  - 変数名
  - スクリプトの出力
- JMP Pro のみの機能には JMP Pro アイコン  がついています。JMP Pro の機能の概要については http://www.jmp.com/software/pro/ をご覧ください。

注: 特別な情報および制限事項には、この文のように「注:」という見出しがついています。

ヒント: 役に立つ情報には「ヒント」という見出しがついています。

JMPのマニュアル

JMPには、印刷版、PDF版、電子本など、さまざまな形式のマニュアルが用意されています。

- PDF版は［**ヘルプ**］>［**ドキュメンテーション**］メニューまたはJMPオンラインヘルプのフッタから開くことができます。
- 検索しやすいようにすべてのドキュメンテーションが1つのPDFファイルにまとめられた『JMPドキュメンテーションライブラリ』と呼ばれるファイルがあります。『JMPドキュメンテーションライブラリ』のPDFファイルは［**ヘルプ**］>［**ドキュメンテーション**］メニューから開くことができます。

JMPドキュメンテーションライブラリ

以下の表は、JMPライブラリに含まれている各ドキュメンテーションの目的および内容をまとめたものです。

マニュアル	目的	内容
『はじめてのJMP』	JMPをあまりご存知ない方を対象とした入門ガイド	JMPの紹介と、データを作成および分析し始めるための情報
『JMPの使用法』	JMPのデータテーブルと、基本操作を理解する	一般的なJMPの概念と、データの読み込み、列プロパティの変更、データの並べ替え、SASへの接続など、JMP全体にわたる機能の説明
『基本的な統計分析』	このマニュアルを見ながら、基本的な分析を行う	［分析］メニューからアクセスできる以下のプラットフォームの説明： • 一変量の分布 • 二変量の関係 • 対応のあるペア • 表の作成 ブートストラップを使用した標本分布の近似方法やモデル化ユーティリティの説明も含まれています。

マニュアル	目的	内容
『グラフ機能』	データに合った理想的なグラフを見つける	［グラフ］メニューからアクセスできる以下のプラットフォームの説明： • グラフビルダー • 重ね合わせプロット • 三次元散布図 • 等高線図 • バブルプロット • パラレルプロット • セルプロット • ツリーマップ • 散布図行列 • 三角図 • チャート このマニュアルには背景マップやカスタムマップの作成方法も記載されています。
『プロファイル機能』	対話式のプロファイルツールの使い方を学ぶ。任意の応答曲面の断面を表示できるようになります。	［グラフ］メニューに表示されるすべてのプロファイルについて。誤差因子の分析が、ランダム入力を使用したシミュレーションの実行とともに含まれています。
『実験計画 (DOE)』	実験の計画方法と適切な標本サイズの決定方法を学ぶ	**［実験計画（DOE）］**メニューと［分析］>［モデル化］メニューの「スクリーニング」に関するすべてのトピックについて。

マニュアル	目的	内容
『基本的な回帰モデル』	「モデルのあてはめ」プラットフォームとその多くの手法について学ぶ	［分析］メニューの「モデルのあてはめ」プラットフォームで使用できる、以下の手法の説明： • 標準最小2乗 • ステップワイズ • 正則化回帰 • 混合モデル • MANOVA • 対数線形-分散 • 名義ロジスティック • 順序ロジスティック • 一般化線形モデル
『発展的なモデル』	付加的なモデリング手法について学ぶ	［分析］>［モデル化］メニューで使用できる以下のプラットフォームの説明： • パーティション • ニューラル • モデルの比較 • 非線形回帰 • Gauss過程 • 時系列分析 • 応答スクリーニング ［分析］>［モデル化］メニューの「スクリーニング」プラットフォームについては『実験計画(DOE)』で説明しています。
『多変量分析』	複数の変数を同時に分析するための手法について理解を深める	［分析］>［多変量］メニューで使用できる以下のプラットフォームの説明： • 多変量の相関 • クラスター分析 • 主成分分析 • 判別分析 • PLS

マニュアル	目的	内容
『品質と工程』	工程を評価し、向上させるためのツールについて理解を深める	［分析］>［品質と工程］メニューで使用できる以下のプラットフォームの説明： • 管理図ビルダーと個々の管理図 • 測定システム分析 • 変動性図/計数値用ゲージチャート • 工程能力分析 • パレート図 • 特性要因図
『信頼性/生存時間分析』	製品やシステムにおける信頼性を評価し、向上させる方法、および人や製品の生存時間データを分析する方法について学ぶ	［分析］>［信頼性/生存時間分析］メニューで使用できる以下のプラットフォームの説明： • 寿命の一変量 • 寿命の二変量 • 再生モデルによる分析 • 劣化分析と破壊劣化分析 • 信頼性予測 • 信頼性成長 • 信頼性ブロック図 • 生存時間分析 • 生存時間(パラメトリック)のあてはめ • 比例ハザードのあてはめ
『消費者調査』	消費者選好を調査し、その洞察を使用してより良い製品やサービスを作成するための方法を学ぶ	［分析］>［消費者調査］メニューで使用できる以下のプラットフォームの説明： • カテゴリカル • 多重対応分析 • 因子分析 • 選択モデル • アップリフト • 項目分析
『スクリプトガイド』	パワフルなJMPスクリプト言語（JSL）の活用方法について学ぶ	スクリプトの作成やデバッグ、データテーブルの操作、ディスプレイボックスの構築、JMPアプリケーションの作成など。

マニュアル	目的	内容
『スクリプト構文リファレンス』	JSL関数、その引数、およびオブジェクトやディスプレイボックスに送信するメッセージについて理解を深める	JSLコマンドの構文、例、および注意書き。

注: [ドキュメンテーション]メニューでは、印刷可能な2つのリファレンスカードも用意されています。『メニューカード』はJMPのメニューをまとめた表で、『クイックリファレンス』はJMPのショートカットキーをまとめた表です。

JMPヘルプ

JMPヘルプは、一連のマニュアルの簡易版です。JMPのヘルプは、次のいくつかの方法で開くことができます。

- Windowsでは、F1キーを押すとヘルプシステムウィンドウが開きます。
- データテーブルまたはレポートウィンドウの特定の部分のヘルプを表示します。[**ツール**]メニューからヘルプツール を選択した後、データテーブルやレポートウィンドウの任意の位置でクリックすると、その部分に関するヘルプが表示されます。
- JMPウィンドウ内で[**ヘルプ**]ボタンをクリックします。
- Windowsの場合、[**ヘルプ**]メニューの[**ヘルプの目次**]、[**ヘルプの検索**]、[**ヘルプの索引**]の各オプションを使用して、JMPヘルプ内を検索し、目的の内容を表示します。Macの場合、[**ヘルプ**] > [**JMPヘルプ**]を選択します。

JMPを習得するためのその他のリソース

JMPのマニュアルとJMPヘルプの他、次のリソースもJMPの学習に役立ちます。

- チュートリアル（「チュートリアル」（25ページ）を参照）
- サンプルデータ（「サンプルデータテーブル」（25ページ）を参照）
- 索引（「統計用語とJSL用語の習得」（25ページ）を参照）
- 使い方ヒント（「JMPを使用するためのヒント」（25ページ）を参照）
- Webリソース（「JMP User Community」（26ページ）を参照）
- 専門誌『JMPer Cable』（「JMPer Cable」（26ページ）を参照）
- JMPに関する書籍（「JMP関連書籍」（26ページ）を参照）
- JMPスターター（「「JMPスターター」ウィンドウ」（27ページ）を参照）

チュートリアル

［**ヘルプ**］>［**チュートリアル**］を選択して、JMP のチュートリアルを表示できます。［**チュートリアル**］メニューの最初の項目は［**チュートリアルディレクトリ**］です。この項目を選択すると、すべてのチュートリアルをカテゴリ別に整理した新しいウィンドウが開きます。

JMP に慣れていない方は、まず［**初心者用チュートリアル**］を試してみてください。JMP のインターフェースおよび基本的な使用方法を学ぶことができます。

他のチュートリアルでは、円グラフの作成、グラフビルダーの使用など、JMP の具体的な活用法を学習できます。

サンプルデータテーブル

JMP のマニュアルで取り上げる例は、すべてサンプルデータを使用しています。［**ヘルプ**］>［**サンプルデータライブラリ**］を選択してサンプルデータディレクトリを開きます。

サンプルデータテーブルを文字コード順に並べた一覧を表示する、またはカテゴリごとにサンプルデータを表示するには、［**ヘルプ**］>［**サンプルデータ**］を選択します。

サンプルデータテーブルは次のディレクトリにインストールされています。

Windows の場合: C:¥Program Files¥SAS¥JMP¥<バージョン番号>¥Samples¥Data

Macintosh の場合: \Library\Application Support\JMP\<バージョン番号>\Samples\Data

JMP Pro では、サンプルデータが（JMP ではなく）JMPPRO ディレクトリにインストールされています。シングルユーザーライセンス版の JMP（JMP シュリンクラップ）では、サンプルデータが JMPSW ディレクトリにインストールされています。

統計用語と JSL 用語の習得

［**ヘルプ**］メニューには、次の索引が用意されています。

統計の索引　統計用語が説明されています。

スクリプトの索引　JSL 関数、オブジェクト、ディスプレイボックスに関する情報を検索できます。スクリプトの索引からサンプルスクリプトを編集して実行することもできます。

JMP を使用するためのヒント

JMP を最初に起動すると、「使い方ヒント」ウィンドウが表示されます。このウィンドウには、JMP を使う上でのヒントが表示されます。

「使い方ヒント」ウィンドウを表示しないようにするには、［**起動時にヒントを表示する**］のチェックを外します。再表示するには、［**ヘルプ**］>［**使い方ヒント**］を選択します。または、「環境設定」ウィンドウで非表示に設定することもできます。詳細については、『JMP の使用法』を参照してください。

ツールヒント

次のような項目の上にカーソルを置くと、その項目を説明するツールヒントが表示されます。

- メニューまたはツールバーのオプション
- グラフ内のラベル
- レポートウィンドウ内の結果（テキスト）（カーソルで円を描くと表示される）
- 「ホームウィンドウ」内のファイル名またはウィンドウ名
- スクリプトエディタ内のコード

ヒント： JMP 環境設定で、ツールヒントを表示しないよう設定できます。［**ファイル**］>［**環境設定**］>［**一般**］（Macintosh の場合は［**JMP**］>［**環境設定**］>［**一般**］）を選択し、［**メニューのヒントを表示**］のチェックを外します。

JMP User Community

JMP User Community では、さまざまな方法でJMPをさらに習得したり、他のSASユーザとのコミュニケーションを図ったりできます。ラーニングライブラリには1ページ構成のガイド、チュートリアル、デモなどが用意されており、JMPを使い始める上でとても便利です。また、JMPのさまざまなトレーニングコースに登録して、自己教育を進めることも可能です。

その他のリソースとして、ディスカッションフォーラム、サンプルデータやスクリプトファイルの交換、Webcastセミナー、ソーシャルネットワークグループなども利用できます。

WebサイトのJMPリソースにアクセスするには［**ヘルプ**］>［**JMP User Community**］を選択します。

JMPer Cable

JMPer Cableは、JMPユーザを対象とした年刊の専門誌です。JMPer Cableは次のJMP Webサイトで閲覧可能です。

http://www.jmp.com/about/newsletters/jmpercable/（英語）

JMP関連書籍

JMP関連書籍は、次のJMP Webページで紹介されています。

http://www.jmp.com/ja_jp/academic/books-for-jmp-users.html

「JMP スターター」ウィンドウ

JMP またはデータ分析にあまり慣れていないユーザは、「JMP スターター」ウィンドウから開始するとよいでしょう。カテゴリ分けされた項目には説明がついており、ボタンをクリックするだけで該当の機能を起動できます。「JMP スターター」ウィンドウには、[**分析**]、[**グラフ**]、[**テーブル**]、および［**ファイル**］メニュー内の多くのオプションがあります。

- 「JMP スターター」ウィンドウを開くには、[**表示**]（Macintosh では［**ウィンドウ**]）> [**JMP スターター**] を選択します。
- Windows で JMP の起動時に自動的に「JMP スターター」を表示するには、［**ファイル**］>［**環境設定**］>［**一般**］を選び、「開始時の JMP ウィンドウ」リストから［**JMP スターター**］を選択します。Macintosh では、［**JMP**］>［**環境設定**］>［**起動時に JMP スターターウィンドウを表示する**］を選択します。

第 2 章

基本的な統計分析について

基本的な分析手法の概要

本書では、JMP でよく使用される基本的な分析機能について解説します。

- 「一変量の分布」プラットフォームでは、1 変数の分布を分析できます。ヒストグラムを表示し、平均や標準偏差などをレポートします。データがどのように分布しているかわかれば、さらに具体的な分析を行い、検討を進めることができます。詳細については、第 3 章「一変量の分布」を参照してください。
- 「二変量の関係」プラットフォームでは、X 変数と Y 変数の尺度によって、自動的に分析が選択されます。詳細については、第 4 章「「二変量の関係」プラットフォームの概要」を参照してください。次の 4 種類の分析を実行できます。
  - 「二変量」プラットフォームでは、2 つの連続変数における関係を分析します。詳細については、第 5 章「二変量分析」を参照してください。
  - 「一元配置」プラットフォームでは、X 変数のグループごとに、連続尺度の Y 変数の分布がどのように異なるかを分析します。詳細については、第 6 章「一元配置分析」を参照してください。
  - 「分割表」プラットフォームでは、X 変数のグループごとに、カテゴリカルな Y 変数の分布がどのように異なるかを分析します。詳細については、第 7 章「分割表分析」を参照してください。
  - 「ロジスティック」プラットフォームでは、カテゴリカルな応答変数 (Y) の確率を、連続尺度の説明変数（X）にあてはめます。詳細については、第 8 章「ロジスティック分析」を参照してください。
- 「対応のあるペア」プラットフォームでは、相関がある 2 変数間の平均を比較し、その差を評価します。詳細については、第 9 章「対応のあるペア分析」を参照してください。
- ブートストラップ分析は、データから何度もサンプリングすることによって統計量の標本分布を近似する手法です。反復ごとに、データは復元抽出（重複抽出）され、抽出されたデータから統計量が計算されます。この計算を反復して、統計量の分布を求めます。詳細については、第 10 章「ブートストラップ」を参照してください。
- 「表の作成」プラットフォームでは、記述統計量のテーブルを対話形式で作成できます。詳細については、第 11 章「表の作成」を参照してください。
- モデル化ユーティリティは、データ分析におけるデータクリーニングと前処理に役立ちます。各ユーティリティにはデータ探索のための強力なツールがあり、これらのツールを使ってデータの理解をさらに深めることができます。詳細については、第 12 章「モデル化ユーティリティ」を参照してください。

第 3 章

一変量の分布

「一変量の分布」プラットフォームを使用する

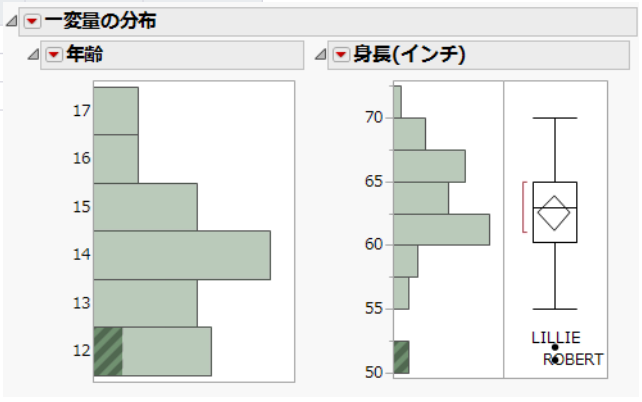
「一変量の分布」プラットフォームでは、1変数の分布を分析できます。ヒストグラムを表示し、平均や標準偏差などをレポートします。「一変量」という用語は単に、分析対象の変数が2つ（二変量）や複数（多変量）ではなく、1つであることを示しています。ただし、複数の変数の分布を1つのレポート内に表示して調べることもできます。レポートの内容は、カテゴリカル変数（名義または順序変数）か、連続変数かによって異なります。

データがどのように分布しているかわかれば、さらに具体的な分析を行い、検討を進めることができます。

「一変量の分布」レポートウィンドウは対話的に操作できます。ヒストグラムの棒をクリックすると、他のヒストグラムやデータテーブル内で、対応するデータが強調表示されます。詳細については、図3.1を参照してください。

図3.1「一変量の分布」プラットフォームの例

	名前	年齢	性別	身長(インチ)	体重(ポンド)
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M		
9	BARBARA	13	F		
10	ALICE	13	F		
11	SUSAN	13	F		



目次

「一変量の分布」プラットフォームの概要

「一変量の分布」プラットフォームでは、変数がカテゴリカル（名義尺度または順序尺度）か連続尺度かによって、変数の処理方法が異なります。

カテゴリカル変数

カテゴリカル変数の場合、最初に表示されるグラフはヒストグラムです。ヒストグラムとは、順序尺度または名義尺度の変数の各水準をそれぞれ1本の棒で表したものです。モザイク図（分割された棒グラフ）を追加することもできます。

レポートには度数と割合が表示されます。信頼区間や割合の検定を追加することもできます。

連続尺度の変数

分析対象の変数が数値で連続尺度の場合は、ヒストグラムと外れ値の箱ひげ図が最初に表示されます。ヒストグラムは、連続変数の値をいくつかのグループに分け、各グループをそれぞれ1本の棒で表したものです。次のプロットも必要に応じて表示できます。

- 分位点の箱ひげ図
- 正規分位点プロット
- 幹葉図
- 累積確率プロット

レポートには、分位点と要約統計量が表示されます。次のようなレポートオプションが用意されています。

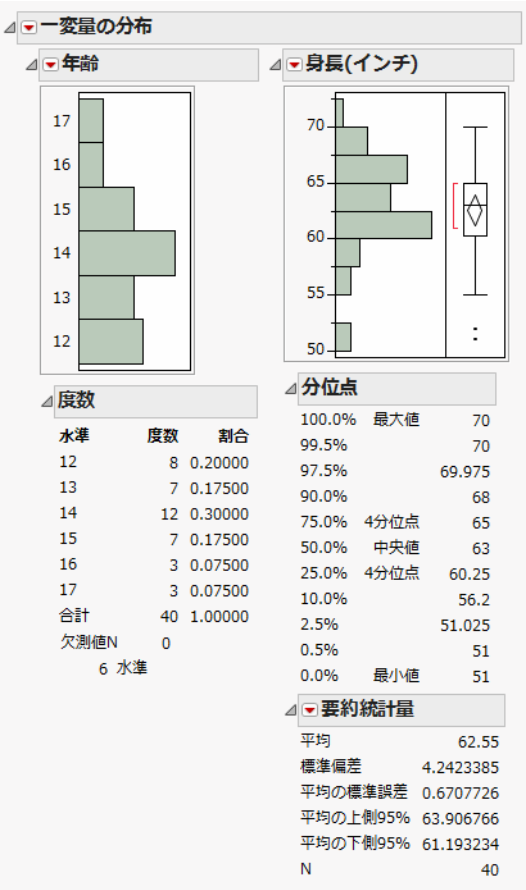
- 順位、確率スコア、正規分位点などをデータテーブルの新しい列に保存する。
- ある定数に対して、列の平均や標準偏差を検定する。
- さまざまな確率分布をあてはめたり、ノンパラメトリックな密度曲線を求めたりする。
- 品質管理を目的とした工程能力分析を実施する。
- 信頼区間、予測区間、許容区間。

「一変量の分布」プラットフォームの例

40名の生徒のデータが手元にあるとします。これらの生徒の年齢と身長の分布を確認してみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**一変量の分布**］を選びます。
3. **「年齢」**と**「身長 (インチ)」**を選択し、［**Y, 列**］をクリックします。
4. ［**OK**］をクリックします。

図3.2「一変量の分布」プラットフォームの例



ヒストグラムを見ると、次のことがわかります。

- 年齢の分布は一様ではありません。
- 「身長(インチ)」のヒストグラムには、離れている点が2つあります（**外れ値**である可能性があります）。

「身長(インチ)」のヒストグラムで「50」の棒をクリックし、外れ値の可能性が高い点について検討します。

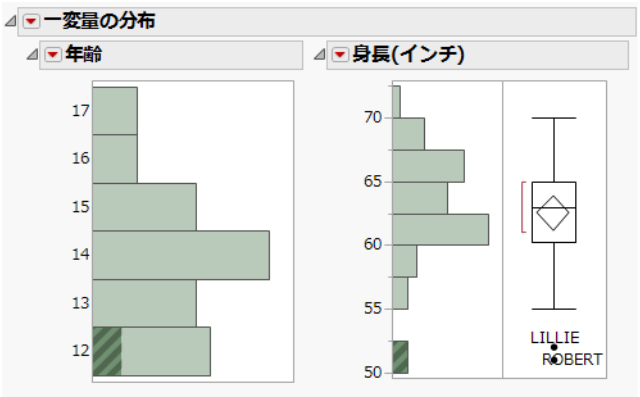
- 「年齢」のヒストグラムで、対応する年齢が強調表示されます。外れ値の年齢は12であることがわかります。
- データテーブル内では、選択した棒に対応する行が強調表示されます。外れ値の名前はLillieとRobertであることがわかります。

「身長(インチ)」のヒストグラムの外れ値にラベルを追加します。

1. 両方の外れ値を選択します。

2. いずれかの外れ値を右クリックして［**行ラベル**］を選択します。
   データテーブルでは、該当行にラベルアイコンが追加されます。
3. 必要に応じて、ラベル全体が表示されるように箱ひげ図の幅を広げてください。

図3.3 ラベルを付けた外れ値



「一変量の分布」プラットフォームの起動

「一変量の分布」プラットフォームを起動するには、［**分析**］>［**一変量の分布**］を選択します。

図3.4「一変量の分布」の起動ウィンドウ

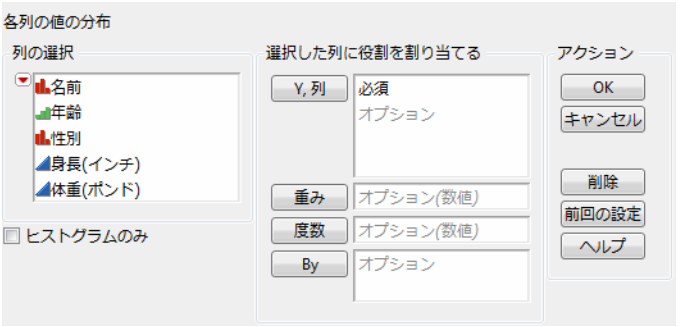


表 3.1「一変量の分布」起動ウィンドウの説明

Y, 列	分析の対象となる変数を割り当てます。選択した変数ごとに、ヒストグラムと、いくつかのレポートが表示されます。
重み	オブザベーションに異なる重みを付けるための変数を指定します。平均や標準偏差などのモーメントの統計量は、重みを考慮して計算されます。
度数	度数の変数を指定します。データがすでに要約されている場合に便利です。度数の変数には、Y 値の発生頻度を含む変数を指定してください。度数の合計は、「要約統計量」レポートに全体の度数（[**N**]）として表示されます。「要約統計量」レポートに表示されるすべての統計量（平均や標準偏差など）が、**度数**を考慮して計算されます。
By	**By** 変数の水準ごとに個別のレポートが作成されます。複数の **By** 変数を割り当てた場合、それらの **By** 変数の水準の組み合わせごとに個別のレポートが作成されます。
ヒストグラムのみ	レポートウィンドウに、ヒストグラムだけを表示し、それ以外の情報は表示しません。

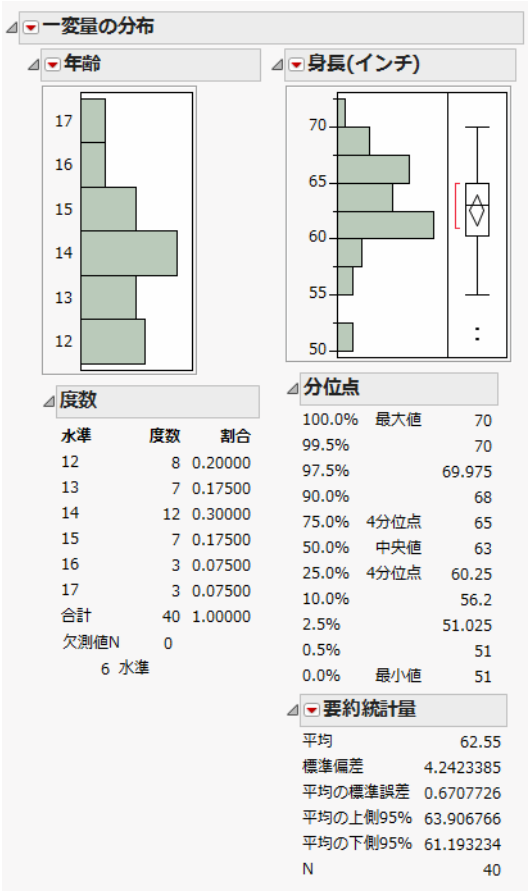
起動ウィンドウ全般については、『JMP の使用法』を参照してください。

［**OK**］をクリックすると、「一変量の分布」レポートウィンドウが開きます。詳細については、「「一変量の分布」レポート」（36 ページ）を参照してください。

「一変量の分布」レポート

「「一変量の分布」プラットフォームの例」（33 ページ）の手順に従って作業すると、図 3.5 のレポートが作成されます。

図3.5 最初に表示される「一変量の分布」レポートウィンドウ



注: データテーブルで除外した行は、ヒストグラムで非表示になります。

「一変量の分布」レポートウィンドウには、変数ごとにヒストグラムとレポートが表示されます。なお、次のような操作もできます。

- 現在の変数を、別の変数に置き換えることができます。データテーブルの「列」パネルから変数をドラッグし、ヒストグラムの軸の上にドロップすると、変数が置き換えられます。
- 新しい変数を追加できます。データテーブルの「列」パネルから変数をドラッグし、既存のヒストグラムの外にドロップすると、その変数が新たに追加されます。変数は、既存のヒストグラムの、一番先頭、間、または、一番後ろのいずれにも追加できます。

注: 変数を削除するには、赤い三角ボタンのメニューから［**削除**］を選択します。

- 「一変量の分布」の横にある赤い三角ボタンをクリックするとメニューが開き、レポートウィンドウ内のすべての変数に適用されるオプションが表示されます。詳細については、「「一変量の分布」プラットフォームのオプション」（44 ページ）を参照してください。
- 各変数名の横にある赤い三角ボタンをクリックするとメニューが開き、該当する変数にのみ適用されるオプションが表示されます。「カテゴリカル変数のオプション」（45 ページ）または「連続変数のオプション」（46 ページ）を参照してください。Ctrl キーを押したまま変数のオプションを選択すると、尺度が同じすべての変数にそのオプションが適用されます。
- ヒストグラムはデータを視覚的に表示します。詳細については、「ヒストグラム」（38 ページ）を参照してください。
- カテゴリカル変数のレポートには、「度数」レポートが含まれます。詳細については、「「度数」レポート」（41 ページ）を参照してください。
- 連続変数のレポートには、「分位点」レポートと「要約統計量」レポートが含まれます。「「分位点」レポート」（41 ページ）および「「要約統計量」レポート」（42 ページ）を参照してください。

ヒストグラム

ヒストグラムはデータを視覚的に表示します。変数がカテゴリカル（名義尺度もしくは順序尺度）の場合には、棒グラフが表示されます。棒グラフは、変数の水準ごとに 1 本の棒で度数を表します。変数が連続尺度である場合は、ヒストグラムが表示されます。ヒストグラムは、変数の数値をいくつかのグループに分け、各グループをそれぞれ 1 つの棒で度数を表します。

表 3.2 ヒストグラムの操作

データの強調表示	ヒストグラムの棒をクリックすると、対応する行がデータテーブル内で強調表示されます。他にもヒストグラムがある場合は、そのヒストグラムでも対応する部分が強調表示されます。詳細については、「棒の強調表示と行の選択」（40 ページ）を参照してください。
サブセットの作成	ヒストグラムの棒をダブルクリックするか、ヒストグラムの棒を右クリックして［**サブセット**］を選択すると、該当するデータだけを抽出した新しいデータテーブルが作成されます。
ヒストグラム全体のサイズの変更	ヒストグラムの枠線の上にマウスを置くと、カーソルの形が双方向の矢印に変わります。カーソルの形が変化したら、枠線をクリックしてドラッグすると、グラフのサイズが変更されます。詳細については、『JMP の使用法』を参照してください。
軸の再スケール	（連続尺度の変数のみ）軸をクリックしてドラッグすると、スケールが変更されます。 または、軸の上にマウスを置くと、手のひらの形にカーソルが変わります。カーソルの形が変化したら、軸をダブルクリックし、「軸の指定」ウィンドウでスケールを設定しても、スケールが変更されます。

表 3.2 ヒストグラムの操作（**続き**）

ヒストグラムにおける棒の幅の変更	（連続尺度の変数のみ）ヒストグラムの棒の幅は、複数の方法で変更できます。詳細については、「連続変数のヒストグラムにおける棒の幅の変更」（39 ページ）を参照してください。
棒の選択方法	複数のヒストグラムにおける棒を選択するには、いくつかの方法があります。詳細については、「複数のヒストグラムにまたがる選択」（41 ページ）を参照してください。

ヒストグラムに関するオプションを表示するには、次の操作を実行します。

- ヒストグラム上で右クリックします。『JMP の使用法』を参照してください。
- 変数名の横にある赤い三角ボタンをクリックし、［**ヒストグラムオプション**］を選択します。表示されるオプションは、変数の尺度によって若干異なります。「カテゴリカル変数のオプション」（45 ページ）または「連続変数のオプション」（46 ページ）を参照してください。

連続変数のヒストグラムにおける棒の幅の変更

連続変数に対するヒストグラムでは、次を使用して棒の幅を変更できます。

- 手のひらツール
- ［**棒の幅の設定**］オプション
- ［**目盛り間隔**］オプション

手のひらツールの使用

手のひらツールを使用すると、データをすばやく調べることができます。

1. ［**ツール**］>［**手のひら**］を選択します。

注：（Windows のみ）デフォルトの設定では、ウィンドウタイトルの下にあるバーの上にマウスを置かないと、メニューバーは表示されません。この設定は、［**ファイル**］>［**環境設定**］>［**Windows のみ**］で変更することもできます。

2. ヒストグラム上の任意の位置に手のひらツールを配置します。
3. ヒストグラムの棒をクリックしてドラッグします。

ヒストグムの各棒を、オブザベーションを含む入れ物と考えてください。手のひらツールによって、その入れ物の大きさが変更されます。

- 手のひらツールを左へ動かすと、区間が結合されて棒の幅が広くなります。1 本 1 本の棒の幅が広くなると、棒の数は減ります。
- 手のひらツールを右へ動かすと、棒の幅が狭くなり、棒の数が増えます。
- 手のひらツールを上下へ動かすと、軸上での棒の位置がシフトし、それに応じて各棒の位置や高さも変化します。

［棒の幅の設定］オプションの使用

［**棒の幅の設定**］オプションによって、ヒストグラムにおける棒の幅を正確に設定できます。［**棒の幅の設定**］オプションを使用するには、変数名の横にある赤い三角ボタンをクリックし、メニューから［**ヒストグラムオプション**］>［**棒の幅の設定**］を選択します。そして、［棒の幅］に数値を指定すると、ヒストグラムの棒がその幅になります。

［目盛り間隔］オプションの使用

［**目盛り間隔**］オプションでも、棒の幅を正確に設定できます。［**目盛り間隔**］オプションを使用するには、軸をダブルクリックし、［目盛り間隔］の値を変更します。

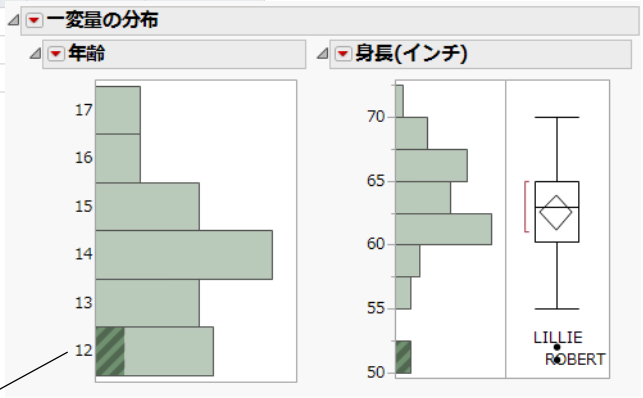
棒の強調表示と行の選択

ヒストグラムの棒をクリックすると、その棒と、データテーブル内の対応する行が強調表示されます。また、その他のグラフでも対応するデータの部分が強調表示されます。図3.6は、「**身長(インチ)**」のヒストグラムにおいて1本の棒を強調表示した結果です。データテーブル内では、その棒に対応する行が選択されています。

ヒント： 特定のヒストグラムの棒の選択を解除するには、Ctrl キーを押しながら、強調表示されている棒をクリックします。

図3.6 棒と行の強調表示

	名前	年齢	性別	身長(インチ)	体重(ポンド)
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F		
10	ALICE	13	F		
11	SUSAN	13	F		



棒を選択すると、データテーブルの行と他の出力の対応部分が強調表示される。

複数のヒストグラムにまたがる選択

複数のヒストグラムにまたがる選択では、論理OR、もしくは、論理ANDのいずれかで、棒を選択していくことができます。

- すでに選択されているものに新しいものを追加していくときには、Shift キーを押しながら別の棒を選択していきます。これは、**OR**演算と同じです。
- 選択を絞り込んでいくときには、Ctrl キーと Shift キーを同時に押しながら棒を選択していきます。これは、**AND**演算と同じです。

関連情報

- 「複数のヒストグラムにおけるデータ選択の例」（64ページ）

「度数」レポート

名義変数と順序変数の場合は、「度数」レポートに変数の水準と、各水準の発生頻度および全体に占める割合が表示されます。

「度数」レポートには、カテゴリカル変数（名義または順序変数）の水準ごとに、表3.3に示す情報が表示されます。欠測値は分析から除外されます。

表3.3「度数」レポートの説明

水準	応答変数の値をリストしたもの。
度数	応答変数の水準ごとの行数をリストしたもの。ただし、度数変数を指定した場合、度数変数の値の合計が表示されます。
割合	応答変数の水準ごとに、その割合（確率）をリストしたもの。割合は、その水準の度数を全体の度数（表の下端に表示されている）で割ったものです。
割合の標準誤差	割合の標準誤差をリストしたもの。デフォルトでは、この列は非表示になっています。表示するには、度数表内を右クリックし、[**列**] > [**割合の標準誤差**] を選択します。
累積割合	割合の累積合計が表示されます。デフォルトでは、この列は非表示になっています。表示するには、度数表内を右クリックし、[**列**] > [**累積割合**] を選択します。

「分位点」レポート

連続変数の場合には、「分位点」レポートに分位点（パーセント点）が表示されます。

関連情報

- 「分位点の統計的詳細」（70ページ）

「要約統計量」レポート

連続変数の場合には、「要約統計量」レポートに平均や標準偏差などが表示されます。表示する要約統計量は、自由に選べます。「要約統計量」の赤い三角ボタンのメニューから［**要約統計量のカスタマイズ**］を選択して、表示したい統計量を選んでください。

- 表3.4に、デフォルトで表示される統計量を示します。
- 表3.5に、**「要約統計量のカスタマイズ」**ウィンドウで追加できる統計量を示します。

表3.4「要約統計量」レポートの説明

平均	算術平均。標本平均。平均は、応答変数の母集団分布における期待値の推定値です。非欠測値に関して、それらの合計を個数で割って求めます。
標準偏差	データのばらつきを示す指標。なお、正規分布は、平均と標準偏差だけによって定義されます。標本サイズが大きく、データが正規分布に従っている場合、これら2つのパラメータによって、次のようにデータを簡単に要約することができます。 • データの68%は、平均から±1標準偏差の区間内にあります。 • データの95%は、平均から±2標準偏差の区間内にあります。 • データの99.7%は、平均から±3標準偏差の区間内にあります。
平均の標準誤差	平均（標本平均）の標準偏差の推定値。
平均の上側95%信頼限界と平均の下側95%信頼限界	平均に関する両側95%信頼区間。「この区間に、母平均が含まれている」という判断の信頼性が高い区間を表します。
N	非欠測値の個数。

表3.5 その他の要約統計量

重みの合計	（起動ウィンドウで）［重み］の役割を割り当てられた列の合計。平均を計算する際、*N*の代わりに分母として使用されます。
合計	応答の値の合計。
分散	不偏分散。標準偏差の2乗。
歪度	歪みや対称性に関する指標。
尖度	尖りや裾の重さに関する指標。
変動係数	単位がパーセントの変動係数。標準偏差を平均で割り、100を掛けた値です。変動係数は相対的な変動の評価に効果的です。たとえば、異なる単位で測定されたデータの変動を比較するときに使用できます。
欠測値N	欠測値の個数。

表 3.5 その他の要約統計量（**続き**）

ゼロの個数	0の値の数。
一意な値の個数	一意（ユニーク）な値の個数。
無修正平方和	修正を加えていない平方和。データそのものの2乗和。
修正平方和	修正を加えた平方和。平均からの偏差の2乗和。
自己相関	（［度数］変数を指定していない場合のみ表示されます。）1次の自己相関。行と、その1つ前の行との間に、相関が見られるかどうかを示します。データにおける系列相関を検出するのに役立ちます。
最小値	0パーセント点。
最大値	100パーセント点。
中央値（メディアン）	メディアン。50パーセント点。
最頻値	モード。データ内で最も出現頻度が高い値。最頻値が複数ある場合、その中の最小値が表示されます。
トリム平均	データの大きい方からpパーセントと小さい方からpパーセントを除外して計算した平均値。pの値は、ウィンドウの最下部にある［**トリム平均に用いるパーセントを入力**］に入力される値です。「重み」変数を指定した場合は、［トリム平均］オプションは表示されません。
幾何平均	データの積のn乗根。
範囲	データの最大値と最小値の差。
四分位範囲	第3四分位点と第1四分位点の差。
中央絶対偏差	（［重み］変数を指定した場合は表示されません。）中央値からの偏差の絶対値の中央値。
ロバスト 平均	ロバストな平均値。外れ値の影響を受けにくいHuberのM推定で求められた平均値です。Huber and Ronchetti（2009）を参照してください。
ロバスト 標準偏差	ロバストな標準偏差。外れ値の影響を受けにくいHuberのM推定で求められた標準偏差です。Huber and Ronchetti（2009）を参照してください。
平均の信頼区間に対する(1-α)を入力	平均の信頼区間に対するα水準を指定します。
トリム平均に用いるパーセントを入力	トリム平均に使用するパーセントを指定します。指定したパーセントのデータが、データの上下から除外されます。

要約統計量のオプション

「要約統計量」の赤い三角ボタンのメニューには、次のオプションが含まれています。

要約統計量のカスタマイズ　表示したい統計量を、一覧から選択してください。すべての統計量を選択したり、逆に、すべての選択を解除したりもできます。用意されている統計量については、詳細については、表3.5を参照してください。

すべての最頻値を表示　複数の最頻値があるときに、そのすべてが表示されます。

関連情報

- 「要約統計量の統計的詳細」（71ページ）

「一変量の分布」プラットフォームのオプション

「一変量の分布」の横にある赤い三角ボタンをクリックするとメニューが開きます。このメニューで用意されているオプションは、「一変量の分布」プラットフォームのすべてのレポートとグラフに適用されます。

表3.6「一変量の分布」プラットフォームのオプションの説明

スケールの統一	すべての軸で最小値、最大値、間隔が同じになり、分布が簡単に比較できるようになります。
積み重ねて表示	ヒストグラムとレポートを横に並べて表示します。複数の変数の分布レポートがある場合は、縦に積み重ねて表示します。このオプションの選択を解除すると、レポートウィンドウのレイアウトは元に戻ります。
複数行に配置	横に並べて表示するプロットの数を入力します。複数のプロットを横一列に配列するのではなく、縦列に配置したい場合に便利です。
Flash(.SWF)形式で保存	ヒストグラムをAdobe Flash Playerに対応した.SWFファイル形式で保存します。保存したファイルはプレゼンテーションやWebページで使用できます。SWFファイルの使用に必要なコードを含むHTMLページも同時に保存されます。 このオプションの詳細については、http://www.jmp.com/support/swfhelp/jaを参照してください。
スクリプト	このメニューには、すべてのプラットフォームに共通するスクリプト関連のコマンドがリストされています。分析をやり直すコマンド、分析のJSLコマンドをウィンドウやファイルに保存するコマンドなどがあります。詳細は、『JMPの使用法』を参照してください。

カテゴリカル変数のオプション

各変数名の横にある赤い三角ボタンをクリックするとメニューが開きます。このメニューで用意されているオプションは、該当する変数にのみ適用されます。ここでは、カテゴリカル変数（名義尺度または順序尺度の変数）に適用されるオプションについて説明します。

連続尺度の変数に適用されるオプションについては、「連続変数のオプション」（46 ページ）を参照してください。

表 3.7 カテゴリカル変数のオプションの説明

［表示オプション］サブメニューには、次のようなオプションが含まれます	
度数	「度数」レポートの表示／非表示を切り替えます。詳細については、「「度数」レポート」（41 ページ）を参照してください。
横に並べる	ヒストグラムとレポートの表示方向を縦／横に切り替えます。
軸を左に	**度数**、**割合**、**密度**の各軸を右側から左側へ移動させます。 このオプションは、**［横に並べる］**が選択されている場合のみ使用可能です。
［ヒストグラムオプション］サブメニューには、次のようなオプションが含まれます。	
ヒストグラム	ヒストグラムの表示／非表示を切り替えます。詳細については、「ヒストグラム」（38 ページ）を参照してください。
縦に表示	ヒストグラムの表示方向を縦／横に切り替えます。
標準誤差バー	ヒストグラムの水準ごとに標準誤差バーが描かれます。
棒の間を離す	ヒストグラムの棒を離して表示します。
ヒストグラムの色	ヒストグラムの棒の色を変更します。
度数軸	それぞれの棒に含まれている値の度数を表す軸が表示されます。
割合軸	それぞれの棒に含まれている値の割合を表す軸が表示されます。
密度軸	ヒストグラムの棒の長さを示す軸が追加されます。 度数と割合は、それぞれ次のように計算されます。 割合 =（棒の幅）× 密度 度数 =（棒の幅）× 密度 ×（合計度数）
パーセントの表示	ヒストグラムの各棒に、その棒の度数が全体に占める割合が表示されます。
度数の表示	ヒストグラムの各棒に、その棒の度数が表示されます。

表 3.7 カテゴリカル変数のオプションの説明（続き）

モザイク図	1 変数に対するモザイク図が表示されます。モザイク図とは、長方形を縦に積み重ねた一種の棒グラフであり、長方形の面積がグループの度数に比例しています。
棒の表示順序	ヒストグラム、モザイク図、「度数」レポートを度数の大きい順または小さい順に並べ替えます。新しい順序を列プロパティとして保存するには、[**保存**] > [**値の順序**] オプションを使用します。
割合の検定	仮説の割合を検定したレポートが表示されます。詳細は、「[割合の検定] オプションの例」（64 ページ）を参照してください。
信頼区間	割合に対する信頼区間が計算されます。信頼区間の計算方法は、スコア法です。このメニューでは、信頼区間の α 水準を選択できます。メニューにある値を選択するか、[**その他**] を選択して独自の値を入力します。
[**保存**] サブメニューには、次のようなオプションが含まれます。	
水準番号	データテーブルの中に、「**水準 < 列名 >**」という新しい列が作成されます。各オブザベーションの水準番号は、ヒストグラム上でそのオブザベーションが属している棒の番号です。
値の順序	（このオプションは、[**棒の表示順序**] オプションと併用してお使いください。）データテーブルの列に、「値の順序」列プロパティを作成します。保存された「値の順序」列プロパティでは、現在、指定されている順序でデータが並びかえられています。
スクリプトをログに保存	現在のレポートを生成するためのスクリプトコマンドがログウィンドウに表示されます。[**表示**] > [**ログ**] を選択して、ログウィンドウを表示します。
削除	「一変量の分布」レポートから、変数とそのすべてのレポートが完全に削除されます。

連続変数のオプション

各変数名の横にある赤い三角ボタンをクリックするとメニューが開きます。このメニューで用意されているオプションは、該当する変数にのみ適用されます。ここでは、連続変数に適用されるオプションについて説明します。

名義尺度または順序尺度の変数に適用されるオプションについては、「カテゴリカル変数のオプション」（45 ページ）を参照してください。

表 3.8 連続尺度の変数のオプションの説明

［表示オプション］サブメニューには、次のようなオプションが含まれます	
分位点	「分位点」レポートの表示／非表示を切り替えます。詳細については、「「分位点」レポート」（41 ページ）を参照してください。
分位点の間隔の設定	分位点の間隔を変更するか、デフォルトの値に戻します。
カスタム分位点	計算する分位点を、値または増分によって指定できます。信頼区間の信頼水準を変更することもできます。このオプションでは、2 種類のカスタム分位点とそれらの信頼限界が算出されます。分位点を求める方法として、重み付き平均で求める方法と、平滑化した経験尤度法で求める方法の 2 種類が使用されます。 • 重み付き平均で求められる分位点の計算方法については、「分位点の統計的詳細」（70 ページ）を参照してください。 • 重み付き平均の分位点に対する信頼限界には、分布によらないノンパラメトリックな方法が使われていますが、その詳細については、Hahn and Meeker（1991）のセクション 5.2 を参照してください。 • 平滑化した経験尤度法による分位点は、カーネル密度推定に基づいて求められます。これらの分位点および信頼限界の推定方法の詳細については、Chen and Hall（1993）を参照してください。
要約統計量	「要約統計量」レポートの表示／非表示を切り替えます。詳細については、「「要約統計量」レポート」（42 ページ）を参照してください。
要約統計量のカスタマイズ	「要約統計量」レポートに表示される統計量を追加または削除します。詳細については、「「要約統計量」レポート」（42 ページ）を参照してください。
横に並べる	ヒストグラムとレポートの表示方向を縦／横に切り替えます。
軸を左に	**度数**、**割合**、**密度**、**正規分位点プロット**の各軸を右側から左側へ移動させます。 このオプションは、**［横に並べる］**が選択されている場合のみ使用可能です。
［ヒストグラムオプション］サブメニューには、次のようなオプションが含まれます。	
ヒストグラム	ヒストグラムの表示／非表示を切り替えます。詳細については、「ヒストグラム」（38 ページ）を参照してください。

表 3.8 連続尺度の変数のオプションの説明（**続き**）

シャドウグラム	ヒストグラムの代わりにシャドウグラムを表示します。ヒストグラムのバンド幅を変更すると、ヒストグラムの形状が変化します。シャドウグラムは、いくつかのバンド幅に対するカーネル密度推定の結果を合わせたものです。シャドウグラムにすると、分布の主要な特長は把握しにくくなります。 シャドウグラムでは次のオプションは使用できません。 • 標準誤差バー • 度数の表示 • パーセントの表示
縦に表示	ヒストグラムの表示方向を縦／横に切り替えます。
標準誤差バー	標準誤差を使って、ヒストグラムの水準ごとに標準誤差バーが作成されます。手のひらツールで棒の数を変更すると、標準誤差バーも自動的に調整されます。「連続変数のヒストグラムにおける棒の幅の変更」（39 ページ）および「標準誤差バーの統計的詳細」（70 ページ）を参照してください。
棒の幅の設定	ヒストグラムの棒の幅を変更します。詳細については、「連続変数のヒストグラムにおける棒の幅の変更」（39 ページ）を参照してください。
ヒストグラムの色	ヒストグラムの棒の色を変更します。
度数軸	それぞれの棒に含まれている値の度数を表す軸が表示されます。 **注：**ヒストグラムの棒のサイズを変更すると、度数軸も変化します。
割合軸	それぞれの棒に含まれている値の割合を表す軸が表示されます。 **注：**ヒストグラムの棒のサイズを変更すると、割合軸も変化します。
密度軸	密度軸が追加されます。密度とは、ヒストグラムの棒の長さを指します。度数と割合は、それぞれ次のように計算されます。 割合 =（棒の幅）× 密度 度数 =（棒の幅）× 密度 ×（合計度数） ［分布のあてはめ］オプションで作成した密度曲線を検討する場合は、密度軸によって、密度曲線上の値を知ることができます。 **注：**ヒストグラムの棒のサイズを変更すると、密度軸も変化します。
パーセントの表示	ヒストグラムの各棒に、その棒の度数が全体に占める割合が表示されます。
度数の表示	ヒストグラムの各棒に、その棒の度数が表示されます。

表 3.8 連続尺度の変数のオプションの説明（続き）

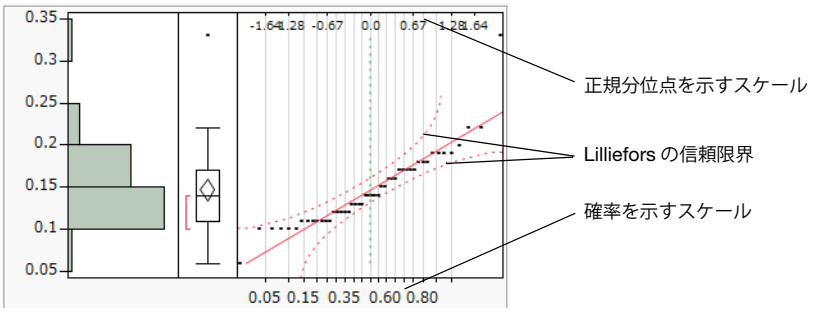
正規分位点プロット	変数がどの程度正規分布に従っているかを示す正規分位点プロットが作成されます。詳細については、「正規分位点プロット」（50 ページ）を参照してください。
外れ値の箱ひげ図	外れ値の箱ひげ図が表示されます。「外れ値の箱ひげ図」は、外れ値を探すのに役立ちます。詳細については、「外れ値の箱ひげ図」（50 ページ）を参照してください。
分位点の箱ひげ図	分位点の箱ひげ図が表示されます。「分位点」レポートに示されている分位点が表示されます。詳細については、「分位点の箱ひげ図」（51 ページ）を参照してください。
幹葉図	ヒストグラムの一種である幹葉図が作成されます。詳細については、「幹葉図」（52 ページ）を参照してください。
累積確率プロット	経験累積分布関数のプロットが作成されます。詳細については、「累積確率プロット」（52 ページ）を参照してください。
平均の検定	平均の一標本検定が実行されます。詳細については、「平均の検定」（53 ページ）を参照してください。
標準偏差の検定	標準偏差の一標本検定が実行されます。詳細については、「標準偏差の検定」（54 ページ）を参照してください。
信頼区間	平均と標準偏差に対する信頼区間が計算されます。詳細については、「連続変数における信頼区間」（55 ページ）を参照してください。
予測区間	次に無作為抽出される 1 つの観測値、もしくは、次に無作為抽出される標本の平均や標準偏差が含まれる区間が計算されます。詳細については、「予測区間」（56 ページ）を参照してください。
許容区間	少なくとも指定した割合だけのオブザベーションを含むと主張できる区間が計算されます。詳細については、「許容区間」（57 ページ）を参照してください。
工程能力分析	工程が特定の仕様限界に適合しているかどうかを測定します。詳細については、「工程能力分析」（57 ページ）を参照してください。
連続分布のあてはめ	連続分布をデータにあてはめます。詳細については、「分布のあてはめ」（59 ページ）を参照してください。
離散分布のあてはめ	離散分布をデータにあてはめます。詳細については、「分布のあてはめ」（59 ページ）を参照してください。
保存	連続変数またはカテゴリカル変数についての情報が保存されます。詳細については、「連続変数における保存コマンド」（55 ページ）を参照してください。
削除	「一変量の分布」レポートから、変数とそのすべてのレポートが完全に削除されます。

正規分位点プロット

［正規分位点プロット］オプションを選択すると、変数がどの程度正規分布に従っているかを確認できます。ある変数が正規分布に従っている場合、正規分位点プロットはほぼまっすぐな対角線になります。このプロットは、分位点－分位点プロットまたはQ-Qプロットとも言います。

正規分位点プロットには、Lilliefors の信頼限界（Conover 1980）、確率スケール、正規分位点スケールも表示されます。

図 3.7 正規分位点プロット



次の点に注目してください。

- *Y* 軸は、列の値を表します。
- *X* 軸は、各値の経験累積確率を示します。

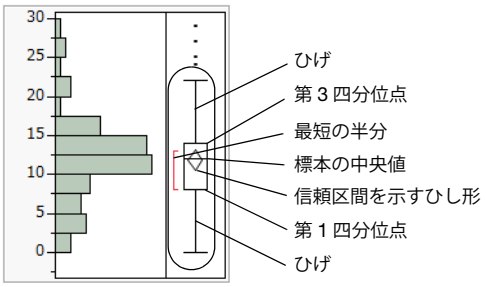
関連情報

- 「正規分位点プロットの統計的詳細」（72 ページ）

外れ値の箱ひげ図

外れ値の箱ひげ図では、分布を観察したり、外れ値を見つけたりすることができます。一般に、箱ひげ図には、いくつかの分位点が表されています。

図 3.8 外れ値の箱ひげ図



外れ値の箱ひげ図には、次のような情報が描かれます。

- 箱を横切る線は標本の中央値（メディアン）を示します。
- ひし形は、平均と、平均の95% 信頼区間を表します。ひし形の中央にある水平線が平均を示し、ひし形の上下の点が、それぞれ平均の上側と下側の95% 信頼限界を表します。
- 箱の両端は25% と75% の分位点で、それぞれ第1および第3**四分位点**とも呼ばれます。
- 第1四分位点と第3四分位点の差が**四分位範囲**です。
- 箱の両端には**ひげ**と呼ばれる線が付いています。ひげは、箱の端から、次の式で計算された範囲内で最も遠くにある点まで伸びています。

  第1四分位点 - 1.5×（四分位範囲）

  第3四分位点 + 1.5×（四分位範囲）

  上の式で計算された範囲内にあるデータ点の一番大きい値と小さい値（外れ値は除く）が、ひげの両端になります。
- 箱の外側に表示されている括弧は、最短の半分（**shortest half**）を示します。これは、オブザベーションの50% が密集する最も短い範囲です（Rousseuw and Leroy 1987）。

箱ひげ図からのひし形や最短半分の削除

外れ値の箱ひげ図から、信頼区間を示すひし形や、「最短の半分」を示す括弧を削除するには、次の手順に従います。

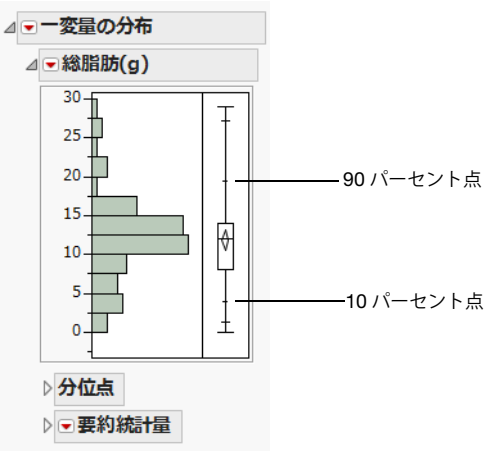
1. 外れ値の箱ひげ図を右クリックし、［**カスタマイズ**］を選択します。
2. **「箱ひげ図」**を選択します。
3. ［**信頼のひし形**］または［**最短の半分**］のチェックを外します。

「グラフをカスタマイズ」ウィンドウの詳細については、『JMP の使用法』を参照してください。

分位点の箱ひげ図

［分位点の箱ひげ図］を選択すると、「分位点」レポートに示されている分位点が表示されます。分布が対称であれば、箱ひげ図に表示される分位点も対称に並んでいるはずです。これを目安にすれば、分布が対称かは一目で判断できます。たとえば、分位点を示すマークの間隔が一方では狭く、もう一方では広い場合、その分布は間隔の広い側に歪んでいることがわかります。詳細については、図3.9を参照してください。

図 3.9 分位点の箱ひげ図



p パーセント点は、値を小さい方から数えて *p*% のところにある値です。たとえば、小さい方から数えて 10% のところにある点が 10 パーセント点、90% のところにある点が 90 パーセント点となります。

幹葉図

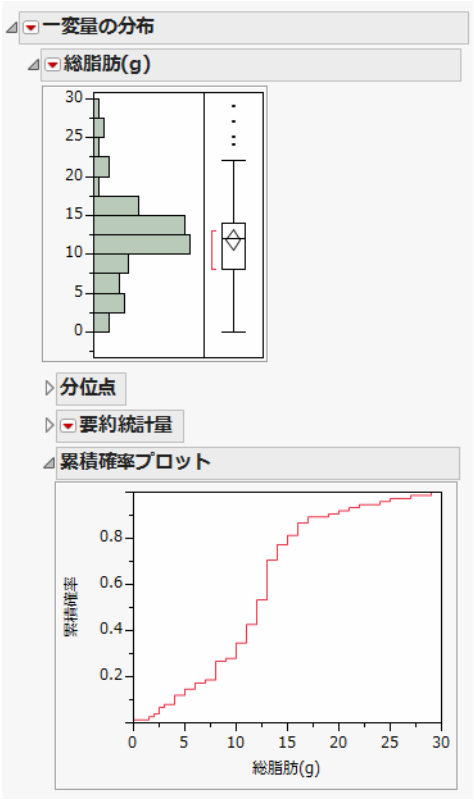
プロットの各行に表示された**幹**の値は、データ値の最初のいくつかの桁を表します。**葉**の値は、各データ値の残りの桁を示します。幹と葉を結合すると、データ値になります。幹葉図には、テーブル上の実測値を丸めた値が表示される場合もあります。幹葉図の値はマウスやブラシツールで選択できます。

注: 幹葉図は、度数が小数点であるデータをサポートしていません。

累積確率プロット

経験累積分布関数のプロットが作成されます。累積確率プロットでは、*X* 軸上の所定値以下に収まるデータの割合を確認できます。

図 3.10 累積確率プロット



たとえば、この累積確率プロットでは、データのおよそ 30% で総脂肪の値が 10 グラムを下回っています。

平均の検定

［**平均の検定**］オプションを選択すると、検定を設定するためのウィンドウが開きます。そこで指定された検定方法によって、一標本の平均に対する検定が実行されます。標準偏差の値を指定すると、*z* 検定が実行されます。標準偏差を入力しなかった場合は、標本から推定された標準偏差を使って、*t* 検定が実行されます。ノンパラメトリックな Wilcoxon の符号付順位検定を実行することもできます。

［**平均の検定**］オプションは、帰無仮説における平均を変えて、何回でも実行できます。実行するたびに、新しい「平均の検定」レポートが追加されます。

表 3.9 「平均の検定」レポートの説明

平均の検定で計算される統計量	
t 検定（または z 検定）	検定統計量の値と、片側検定と両側検定の *p* 値がリストされます。
Wilcoxon の符号付順位検定	（Wilcoxon の符号付順位検定を選択した場合のみ表示）Wilcoxon の符号付順位検定の統計量と、片側検定と両側検定の *p* 値がリストされます。この検定では、0 の値は Pratt 法を使用して処理されます。「中央値が仮説値と等しい」という帰無仮説を検定するノンパラメトリックな検定の 1 つです。なお、「中央値が仮説値と等しい」を帰無仮説とする場合は、分布の対称性を仮定する必要があります。詳細については、「Wilcoxon の符号付順位検定の統計的詳細」（72 ページ）を参照してください。
p 値	
p 値 (Prob>|t|)	母平均が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本の *t* 値の絶対値よりも、**t** 値の絶対値が大きくなる確率です。これは両側 *t* 検定の *p* 値です。
p 値 (Prob>t)	母平均が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本の *t* 値より大きい *t* 値を得る確率です。これは上側の片側検定の *p* 値です。
p 値 (Prob<t)	母平均が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本の *t* 値より小さい値 *t* を得る確率です。これは下側の片側検定の *p* 値です。

表 3.10［平均の検定］オプションの説明

p 値アニメーション	対話的なグラフに、p 値を描きます。平均の仮説値などを変更した場合に、p 値にどのような影響が及ぶかを確認できます。
検出力アニメーション	対話的なグラフに、検出力とベータを描きます。真の平均や、平均の仮説値によって、検出力とベータがどのように変化するを確認できます。
検定結果の削除	平均の検定結果を削除します。

標準偏差の検定

［**標準偏差の検定**］オプションを選択すると、一標本の標準偏差に対する検定（Neter, Wasserman and Kutner 1990）が実行されます。［**標準偏差の検定**］オプションは、帰無仮説における標準偏差を変えて、何回でも実行できます。実行するたびに、新しい「標準偏差の検定」レポートが追加されます。

表 3.11「標準偏差の検定」レポートの説明

検定統計量	カイ 2 乗検定統計量の値。詳細については、「標準偏差の検定の統計的詳細」（74 ページ）を参照してください。

表 3.11「標準偏差の検定」レポートの説明（**続き**）

最小 p 値	母標準偏差が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本のカイ 2 乗値より大きいカイ 2 乗値を得る確率です。詳細については、「標準偏差の検定の統計的詳細」（74 ページ）を参照してください。
p 値（Prob>ChiSq）	母標準偏差が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本のカイ 2 乗値より大きいカイ 2 乗値を得る確率です。これは片側 *t* 検定の *p* 値です。
p 値 (Prob<ChiSq)	母標準偏差が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本のカイ 2 乗値より小さいカイ 2 乗値を得る確率です。これは片側 *t* 検定の *p* 値です。

連続変数における信頼区間

［**信頼区間**］サブメニューの各オプションを選択すると、平均と標準偏差の信頼区間が計算されます。［**0.90**］、［**0.95**］、［**0.99**］の各オプションを選択した場合、両側信頼区間が計算されます。［**信頼区間**］**>**［**その他**］オプションを選択した場合、信頼水準を入力し、片側か両側のどちらかを選択できます。また、既知の σ を入力することもできます。既知の σ を入力した場合は、*t* 値ではなく *z* 値によって、平均の信頼区間が計算されます。

「信頼区間」レポートには、平均と標準偏差のパラメータ推定値、および 1 - α の上側信頼限界と下側信頼限界が表示されます。

連続変数における保存コマンド

［保存］メニューのコマンドを使用して、連続尺度の変数に関する情報を保存できます。各保存コマンドを選択すると、現在のデータテーブルに新しい列が作成されます。新しい列の名前は、選択した保存コマンド名の後ろに変数名（下の表で**<列名>**と表記）が付いた形になります。詳細については、表 3.12 を参照してください。

［保存］メニューのコマンドを繰り返し選択し、ヒストグラムの棒を組み合わせる前と後など、異なる状況で同じ情報を何度でも保存できます。［保存］メニューのコマンドを 2 回以上実行した場合、列の名前には列名 1、列名 2、列名 3 というように連番が振られます。

表 3.12［保存］メニューのコマンドの説明

コマンド	データテーブルに追加される列	説明
水準番号	**水準 <列名>**	各オブザベーションの水準番号は、ヒストグラム上でそのオブザベーションが属している棒の番号です。棒に対して、小さいものから順に 1 から始まる通し番号が付けられます。
水準中間点	**中間点 <列名>**	各オブザベーションの中間点の値は、水準の下側の境界に、水準の幅の半分を足したものです。

表 3.12［保存］メニューのコマンドの説明（続き）

コマンド	データテーブルに追加される列	説明
順位	**順位 <列名>**	対応する列の値の 1 から始まる順位が保存されます。重複している値には、データテーブル内での順序に従って、連続した順位が割り当てられます。
平均順位	**平均順位 <列名>**	データに重複がない場合は、平均順位は、通常の順位と同じです。同じ値が k 回起こっている場合、それらの値の順位を合計し、k で割ったものが平均順位です。
確率スコア	**確率 <列名>**	非欠測値の数を N としたとき、ある値の確率スコアは、その値の平均順位を $N + 1$ で割って算出されます。この列の値は、経験累積分布関数（CDF）に似ています。
正規分位点	**正規分位点 <列名>**	正規分位点がデータテーブルに保存されます。詳細については、「正規分位点プロットの統計的詳細」（72 ページ）を参照してください。
標準化	**標準化 <列名>**	標準化したデータがデータテーブルに保存されます。詳細については、「標準化データの統計的詳細」（74 ページ）を参照してください。
中心化	**中心化 <列名>**	0 で中心化した値が保存されます。
仕様限界	（なし）	工程能力分析での仕様限界が、現在のデータテーブルの対応列の列プロパティとして保存されます。次に工程分析を再実行したとき、自動的にこの仕様限界が読み込まれ、表示されます。
スクリプトをログに保存	（なし）	スクリプトをログウィンドウに出力します。スクリプトを実行すると、分析を再現できます。

予測区間

予測区間とは、次に無作為抽出される 1 つの観測値、もしくは、次に無作為抽出される標本の平均や標準偏差が含まれる区間を計算したものです。これらの予測区間は、「正規分布に従う母集団から、標本が無作為抽出されている」という仮定の下で計算されます。片側予測区間と両側予測区間のいずれかを選択できます。

変数の［**予測区間**］オプションを選択すると、「予測区間」ウィンドウが表示されます。このウィンドウで、信頼水準、予測する標本サイズ、片側または両側限界を指定します。

関連情報

- 「予測区間の統計的詳細」（75 ページ）
- 「予測区間の例」（66 ページ）

許容区間

少なくとも指定した割合だけのオブザベーションを含むと主張できる区間を指します。つまり、平均や標準偏差ではなく、母集団の割合に対する信頼区間ということができます。許容区間は、Hahn and Meeker（1991）と Tamhane and Dunlop（2000）で詳細に論述されています。

変数の［**許容区間**］オプションを選択すると、「許容区間」ウィンドウが表示されます。このウィンドウで、信頼水準、オブザベーションの割合、片側または両側限界を指定します。これらの許容区間は、「正規分布に従う母集団から、標本が無作為抽出されている」という仮定の下で計算されます。

関連情報

- 「許容区間の統計的詳細」（75 ページ）
- 「許容区間の例」（68 ページ）

工程能力分析

［**工程能力分析**］オプションでは、工程が特定の仕様限界に適合しているかどうかを測定します。変数の［**工程能力分析**］オプションを選択すると、「工程能力分析」ウィンドウが表示されます。このウィンドウで、仕様限界、分布の種類、シグマについての情報を入力します。

注 : 仕様限界を列のプロパティとしてデータテーブルに保存するには、［**保存**］**>**［**仕様限界**］を選択します。次に工程能力分析を再実行したとき、保存済みの仕様限界が自動的に読み込まれます。

「工程能力分析」レポートは、「工程能力分析」と分布の種類（「長期シグマ」、「指定されたシグマ」など）の 2 つのセクションに分かれています。

工程能力分析の説明

「工程能力分析」のウィンドウ、レポート、オプションを以下の表で説明します。

表 3.13「工程能力分析」ウィンドウの説明

<分布の種類>	デフォルトでは、工程能力の統計量や仕様限界から外れているものの割合は、正規分布の仮定に基づいて算出されます。非正規分布について工程能力分析を実行するには、「あてはめた分布のオプション」（61 ページ）の「仕様限界」の説明を参照してください。
<シグマの種類>	シグマ（σ）の推定値が選択した手法で計算されます。詳細については、「工程能力分析の統計的詳細」（76 ページ）を参照してください。

表 3.14「工程能力分析」レポートの説明

仕様	仕様限界のリストです。
値	仕様限界と目標値のフィールドにユーザが入力した値のリストです。

表 3.14「工程能力分析」レポートの説明（**続き**）

比率、実測値 %	「比率」列には、「実測値 %」列の数値について説明する次のラベルが表示されます。 • 「LSL 未満」は、下限仕様限界未満のデータの割合を示します。 • 「USL 超え」は、上限仕様限界を超えたデータの割合を示します。 • 「限界外の合計数」は、下限仕様限界未満と上限仕様限界超の割合の合計を示します。
工程能力	工程能力指数の種類。詳細については、表 3.19 を参照してください。 **注：**環境設定で**「Ppk 工程能力ラベル設定」**をオンにした場合、全データから推定した工程能力統計量には、「P_{pk}」というラベルが付きます。この環境設定を確認するには、「環境設定」ウィンドウを開き（**［ファイル］> ［環境設定］**）、**［プラットフォーム］> ［工程能力］**を選択します。
指数	工程能力指数の値。
上側信頼限界	上側信頼限界。
下側信頼限界	下側信頼限界。
比率、パーセント	「比率」列には、「パーセント」列の数値について説明するラベルが表示されます。 • 「LSL 未満」は、あてはめた分布から計算された、下限仕様限界未満の割合を示します。 • 「USL 以上」は、あてはめた分布から計算された、上限仕様限界を超える割合を示します。 • 「限界外の合計数」は、あてはめた分布から計算された、下限仕様限界未満と上限仕様限界超の割合を合計したものです。
PPM（100 万分の 1）	「PPM」列の値は、**「パーセント」**列の値に 1 万を掛けたものです。
シグマクオリティ	「シグマクオリティ」は、シックスシグマ分析の基準として使用するもので、**Process Sigma（工程シグマ）**とも言います。詳細については、「工程能力分析の統計的詳細」（76 ページ）を参照してください。

表 3.15「工程能力分析」のオプションの説明

Z ベンチ	ベンチマーク Z の値を、**「指数」**という列に表示します。この指標については、『AIAG Statistical Process Control』マニュアルで説明されています。平均と仕様限界との差を、標準偏差を単位として表した指標です。Z USL および Z LSL は、それぞれ上側仕様限界と下側仕様限界までの平均からの距離を表します。詳細については、「工程能力分析の統計的詳細」（76 ページ）を参照してください。

表 3.15「工程能力分析」のオプションの説明（続き）

工程能力アニメーション	仕様限界や平均が変化したときの工程能力指数の変化を、対話的に確認することができます。このオプションは、正規分析に基づく工程能力分析でのみ使用できます。

関連情報

- 「工程能力分析の統計的詳細」（76 ページ）
- 「工程能力分析の例」（69 ページ）

分布のあてはめ

［**連続分布のあてはめ**］または［**離散分布のあてはめ**］の各オプションを使用して、連続変数または離散変数に分布をあてはめます。

ヒストグラム上に曲線が表示され、レポートウィンドウに新たに「パラメータ推定値」レポートが表示されます。赤い三角ボタンのメニューには、追加のオプションが表示されます。詳細については、「あてはめた分布のオプション」（61 ページ）を参照してください。

注：「寿命の一変量」プラットフォームでも、様々な分布をあてはめることができます。ただし、「一変量の分布」プラットフォームとは、パラメータ化が異なるものもあります。また、「寿命の一変量」プラットフォームでは、打ち切りデータを扱えます。『品質と工程』を参照してください。

連続分布のあてはめ

［**連続分布のあてはめ**］オプションでは、次のような連続分布が用意されています。

- 正規分布。左右対称で、ほとんどのデータが中央にある分布です。JMP は、不偏推定量を使用して正規分布のパラメータを求めます。
- 対数正規分布。0 を下限値とします。右に大きく歪んでおり、非常に大きな値がいくつかある場合に使います。正規分布に従うデータの指数をとると、対数正規分布になります。JMP は、最尤推定を使用して対数正規分布のパラメータを求めます。
- Weibull 分布、閾値つき Weibull 分布、極値分布。これらの分布は、特に、工業製品の故障時間や、生物学の生存時間を分析するときに使われています。
- 指数分布。特に、工業製品の故障時間や、生物学の生存時間を分析するときに使われています。また、あるイベントが生じる間隔をモデル化するのにも使われます。たとえば、コンピュータのサーバーにクエリーが送られてくる時間、お客様がサービスカウンターに来る間隔、交換機が着信する間隔などです。
- ガンマ分布。この分布も、下限値は 0 です。
- ベータ分布。（0, 1）の区間を範囲としています。標準ベータ分布は、割合などの 0 から 1 の範囲に収まるデータに対してよく使われています。

- 正規混合分布。正規分布を混合した分布です。多峰性の分布にも対応した柔軟な分布です。[**二重正規混合**]、[**三重正規混合**]、または［**その他**］オプションを選択すると、指定した数だけの分布を混合した正規混合分布があてはめられます。
- ［平滑曲線］オプションを選択すると、カーネル密度推定が行われます。カーネル密度推定は、ノンパラメトリックな密度推定の一種で、密度曲線は滑らかなです。ヒストグラム上に平滑曲線が表示され、プロットの下にスライダが表示されます。スライダを使ってカーネル標準偏差を変更することで、平滑化の度合いを調節できます。［**カーネル標準偏差**］の初期値は、データの標準偏差に基づいて算出されます。
- Johnson Su 分布、Johnson Sb 分布、Johnson Sl 分布。これら 3 つで構成される分布システムは、歪度と尖度のあらゆる組み合わせに対応しており、柔軟にデータにあてはまります。
- 一般化対数分布（Glog 分布）。生物検定法（バイオアッセイ）など、分散が一定でない非正規なデータに役立ちます。

すべての分布の比較

［**すべて**］オプションは、現在のデータに適用可能なすべての分布をあてはめます。「分布の比較」レポートに、あてはめられた各分布の統計量が表示されます。チェックボックスで分布を選択することによって、その分布のレポート、およびヒストグラム上に描かれる密度曲線の表示／非表示を切り替えることができます。デフォルトでは、データに最も適合している分布のチェックボックスがオンになっています。

「表示分布」のリストは、AICc の値で昇順に並んでいます。

すべての分布をあてはめる際に、データに負の値がある場合は、正の値のみを対象とする分布は除外されます。データに整数以外の値がある場合は、離散分布は除外されます。すべての場合において、閾値パラメータがある分布（ベータ分布や Johnson Sb 分布など）は除外されます。

関連情報

- 「［連続分布のあてはめ］の統計的詳細」（79 ページ）
- 「あてはめた分布の分位点」（88 ページ）
- 「あてはめた分布に対するオプション」（88 ページ）

離散分布のあてはめ

［離散分布のあてはめ］の各オプションは、離散値のデータに対して、Poisson 分布や二項分布などをあてはめます。分布の種類は次のとおりです。

- Poisson
- 二項
- ガンマ Poisson
- ベータ二項

関連情報

- 「［離散分布のあてはめ］の統計的詳細」（85 ページ）

- 「あてはめた分布の分位点」（88 ページ）
- 「あてはめた分布に対するオプション」（88 ページ）

あてはめた分布のオプション

あてはめた分布の各レポートには、赤い三角ボタンがあり、クリックすると追加のオプションが表示されます。

表 3.16 分布のあてはめのオプションの説明

診断プロット	分位点プロットまたは確率プロットが作成されます。詳細については、「診断プロット」（62 ページ）を参照してください。
密度曲線	分布のパラメータ推定値を基に密度曲線が計算され、ヒストグラムに重ねて表示されます。
適合度	あてはめた分布に対して、適合度検定が計算されます。詳細については、「適合度」（63 ページ）を参照してください。
分布パラメータの指定	一部のパラメータを任意の値に設定し、他のパラメータの推定値を計算することができます。また、入力したパラメータ値がデータにあてはまるかどうかが検定され、その結果が「妥当性の尤度比検定」レポートとして表示されます。
分位点	このオプションを選択すると、指定した下側累積確率に対して、尺度化も中心化もしていない分位点が戻されます。
K シグマから仕様限界を設定	設定すべき仕様限界（規格限界）が決まっておらず、データから求めたい場合にこのオプションを使ってください。 仕様限界は、工程を工学的に検討して決めるのが普通です。工学的な検討が行えず、かつ、現在のデータが信頼に足るベンチマークである（工程能力が高い）場合には、現在のデータに分布をあてはめて、その分位点を仕様限界としてもよいでしょう。詳細については、「あてはめた分布に対するオプション」（88 ページ）を参照してください。
仕様限界	指定した仕様限界と目標値を基に、一般化された工程能力指数が計算されます。詳細については、「仕様限界」（63 ページ）を参照してください。
あてはめたモデルの分位点を保存	あてはめた分布の分位点が、データテーブルの列に保存されます。詳細については、「あてはめた分布の分位点」（88 ページ）を参照してください。
密度関数の保存	あてはめた分布の密度関数が、データテーブルの列に保存されます。密度関数では、パラメータ推定値が使用されます。
仕様限界の保存	仕様限界を列プロパティとして保存します。詳細については、「あてはめた分布に対するオプション」（88 ページ）を参照してください。

表3.16 分布のあてはめのオプションの説明（続き）

変換の計算式を保存	正規分布に変換するための計算式が保存されます。この計算式は、あてはめた分布に基づいて、データを正規分布に近づくように変換するものです。このオプションは、Johnson 分布か一般化対数分布があてはめられた場合にのみ使用できます。
あてはめの削除	分布のあてはめに関する情報がレポートウィンドウから削除されます。

診断プロット

［**診断プロット**］オプションを選択すると、分位点プロットまたは確率プロットが作成されます。あてはめた分布に応じて、作成されるプロットは4つの形式のいずれかになります。

表3.17 プロット形式の説明

プロット形式	対象の分布
X軸：データ、Y軸：あてはめた分布の分位点	• 閾値つき Weibull • ガンマ • ベータ • Poisson • ガンマ Poisson • 二項 • ベータ二項
X軸：データ、Y軸：あてはめた分布の確率	• 正規 • 正規混合分布法 • 指数
X軸：対数スケールのデータ、Y軸：あてはめた分布の確率	• Weibull 損失関数 • 対数正規 • 極値損失関数
X軸：標準正規分布の分位点、Y軸：あてはめた分布の確率	• Johnson Sl • Johnson Sb • Johnson Su • 一般化対数

表3.18では、「診断プロット」の横の赤い三角ボタンをクリックすると表示されるオプションについて説明しています。

表 3.18「診断プロット」のオプションの説明

回転	X 軸と Y 軸を入れ換えます。
信頼限界	正規分位点プロットでは Lilliefors の 95% 信頼限界、その他の分位点プロットでは $a = 0.01$、$b = 0.99$ のときの 95% 等精度信頼帯（95% equal precision band）を表示します（Meeker and Escobar 1998）。
あてはめ線	参照線としてまっすぐな対角線を表示します。変数の値がこの参照線の付近にあるなら、選択した分布が変数にあてはまっていることがわかります。
中央値の参照線	応答変数の中央値を示す水平線を表示します。

適合度

［**適合度**］オプションを選択すると、あてはめた分布に対して、適合度検定が計算されます。ここでの適合度検定は、カイ 2 乗検定ではなく、経験分布関数（EDF; Empirical Distribution Function）に基づく検定です。経験分布関数に基づく検定は、カイ 2 乗検定よりも検出力が高く、ヒストグラムの中間点によって検定結果が変わらないという点でカイ 2 乗検定より優れています。

- 正規分布については、標本サイズが 2000 以下のときは正規性に関する Shapiro-Wilk の検定が、標本サイズが 2000 より大きいときは KSL（Kolmogorov-Smirnov-Lilliefors）の検定が計算されます。
- Poisson 分布などの離散分布において標本サイズが 30 以下の場合は、2 つの正確な Kolmogorov 片側検定を組み合わせて両側検定とした方法（Conover 1972）が使われます。この両側検定は、ほぼ正確です。また、離散分布において標本サイズが 30 を超える場合は、Pearson のカイ 2 乗検定が使われます。

関連情報

- 「あてはめた分布に対するオプション」（88 ページ）

仕様限界

［**仕様限界**］オプションを選択すると、仕様限界と目標値の入力用ウィンドウが開きます。指定した値を基に、一般化された工程能力指数が計算されます。この計算は、正規分布での 3σ に対応する分位点を用いて計算されます。正規分布では、中央値（平均値）から下側に 3σ 離れた位置は、0.135 パーセント点です。また、上側に 3σ 離れた位置は、99.865 パーセント点です。一般化された工程能力指数は、あてはめた分布のこれらのパーセント点から計算されます。正規分布を仮定した標準の計算式では、該当するパーセント点から中央値までの距離は 3σ です。

関連情報

- 「あてはめた分布に対するオプション」（88 ページ）

「一変量の分布」プラットフォームのその他の例

ここでは、「一変量の分布」プラットフォームを使った例をさらに紹介します。

複数のヒストグラムにおけるデータ選択の例

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Companies.jmp」を開きます。
2. ［分析］>［一変量の分布］を選びます。
3. 「タイプ」と「会社規模」を選択し、［Y, 列］をクリックします。
4. ［OK］をクリックします。

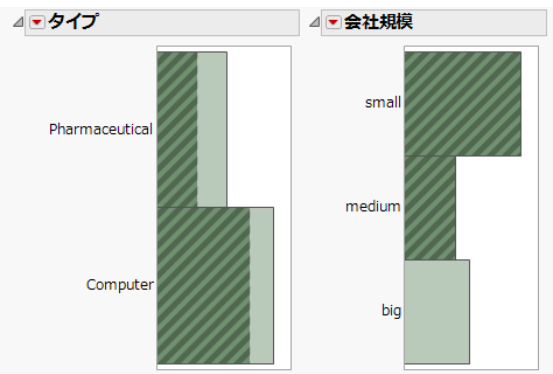
   小規模企業について、「タイプ」の分布を確認しましょう。
5. 「small」の横にある棒をクリックします。

   製薬（Pharmaceutical）よりもコンピュータ関連（Computer）の会社において、小規模企業（small）が多いことがわかります。次に、選択範囲を拡げて、小規模企業だけでなく、中規模企業も選択しましょう。
6. 「会社規模」のヒストグラム上で、Shift キーを押しながら、「medium」の棒をクリックします。

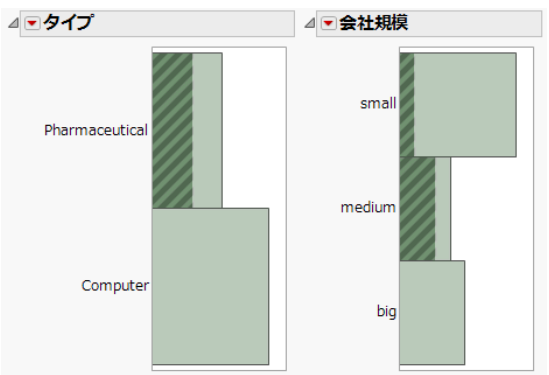
   小規模（small）および中規模（medium）における「タイプ」の分布を確認できます。図 3.11 の左側を参照してください。次は、選択範囲を絞って、小規模および中規模の製薬会社だけを表示することにします。
7. 「Computer」に対する選択を解除するために、「タイプ」のヒストグラム上で、Ctrl キーと Shift キーを押しながら、「Computer」の棒をクリックします。

   小規模（small）、もしくは中規模（medium）の製薬会社（Pharmaceutical）が、どれぐらいデータにあるかが分かります。図 3.11 の右側を参照してください。

図 3.11 複数のヒストグラムにおけるデータの選択



Shift キーを使用して選択範囲を広げる。



Ctrl キーと Shift キーを使って選択範囲を絞る。

［割合の検定］オプションの例

3 水準以上のカテゴリカルな変数に対して、割合の検定を実行する例を述べます。

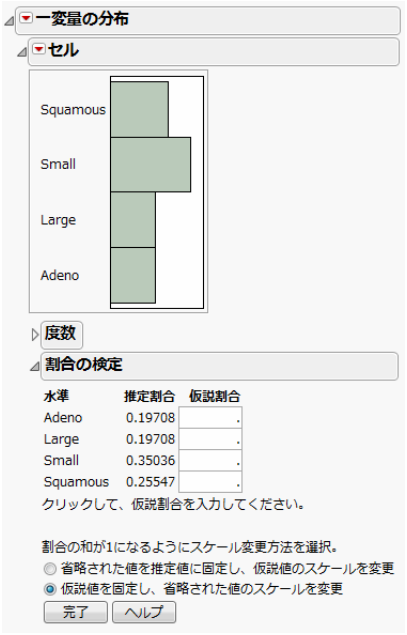
1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「VA Lung Cancer.jmp」を開きます。
2. ［分析］>［一変量の分布］を選びます。

3. 「**セル**」を選択し、[**Y, 列**] をクリックします。
4. [**OK**] をクリックします。
5. 「セル」の横にある赤い三角ボタンをクリックし、メニューから [**割合の検定**] を選択します。
   図 3.12 の左側を参照してください。

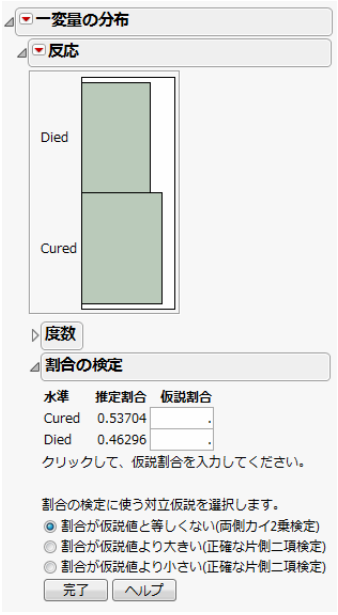
2 水準のカテゴリカルな変数に対しても、同じ操作で割合の検定を実行できます。以下に、例を述べます。

1. [**ヘルプ**] > [**サンプルデータライブラリ**] を選択し、「Penicillin.jmp」を開きます。
2. [**分析**] > [**一変量の分布**] を選びます。
3. 「**反応**」を選択し、[**Y, 列**] をクリックします。
4. [**OK**] をクリックします。
5. 「反応」の横にある赤い三角ボタンをクリックし、メニューから [**割合の検定**] を選択します。
   図 3.12 の右側を参照してください。

図 3.12「割合の検定」のオプションの例



水準が 3 つ以上の変数の
レポートオプション



水準が 2 つの変数の
レポートオプション

「割合の検定」の実行例

3 水準以上の場合には、先ほど述べた操作例のあとに、次のような操作を行ってください。

1. 図 3.12 の左側における 「仮説割合」のすべてのフィールドに「0.25」と入力します。

2. ［仮説値を固定し、省略された値のスケールを変更］ボタンをクリックします。
3. ［完了］をクリックします。

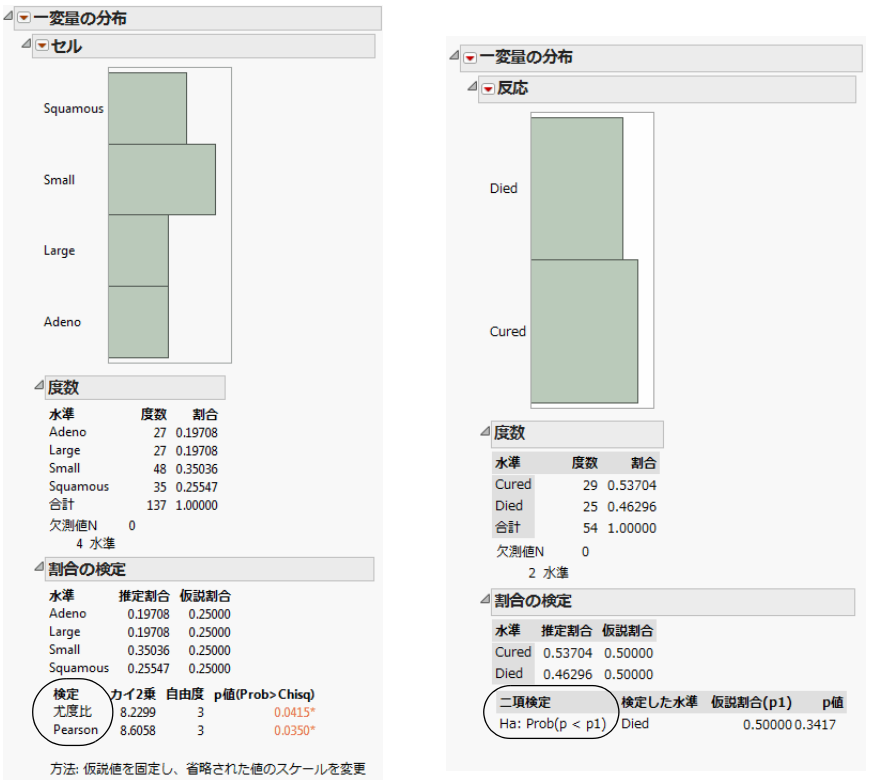
尤度比検定と Pearson のカイ 2 乗検定が計算されます。図 3.13 の左側を参照してください。

2 水準の場合には、先ほど述べた操作例のあとに、次のような操作を行ってください。

1. 図 3.12 の右側における 「仮説割合」の両方のフィールドに「0.5」と入力します。
2. ［割合が仮説値より小さい］ボタンをクリックします。
3. ［完了］をクリックします。

二項検定の正確な p 値が計算されます。図 3.13 の右側を参照してください。

図 3.13「割合の検定」レポートの例



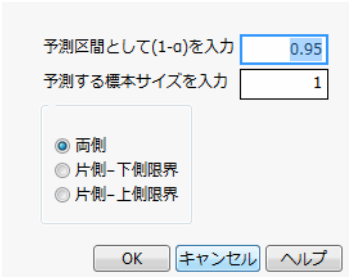
予測区間の例

将来、観測される 10 個のオゾン量の観測値に関して、予測区間を計算したいとしましょう。

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Cities.jmp」を開きます。
2. ［分析］>［一変量の分布］を選びます。

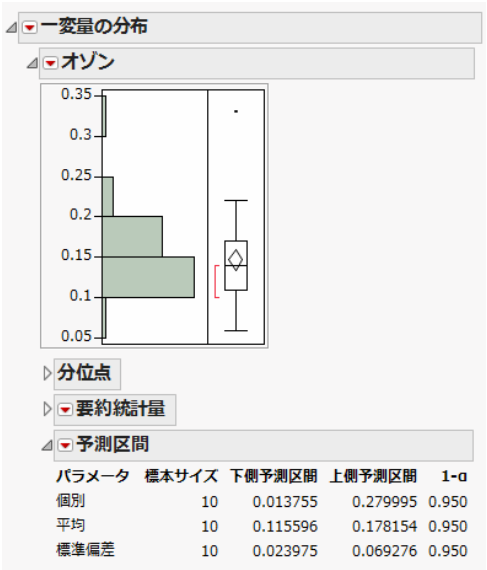
3. 「**オゾン**」を選択し、［**Y, 列**］をクリックします。
4. ［**OK**］をクリックします。
5. 「オゾン」の横にある赤い三角ボタンをクリックし、メニューから［**予測区間**］を選択します。

図 3.14「予測区間」ウィンドウ



6. 「予測区間」ウィンドウで、「**予測する標本サイズを入力**」の横に「10」と入力します。
7. ［**OK**］をクリックします。

図 3.15「予測区間」レポートの例



この例では、95% の信頼水準で、次のように結論することができます。

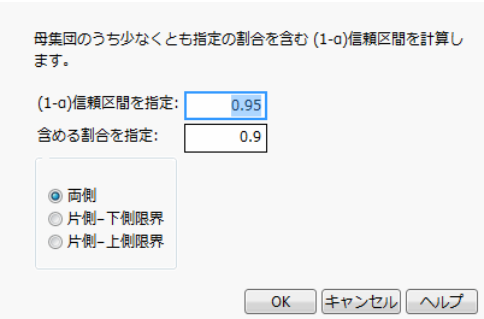
- 将来、観測される 10 個の観測値は、すべて、0.013755 ～ 0.279995 の区間にある。
- 将来、観測される 10 個の観測値の平均値は、0.115596 ～ 0.178154 の区間にある。
- 将来、観測される 10 個の観測値の標準偏差は 0.023975 ～ 0.069276 の区間にある。

許容区間の例

オゾン量の測定値の 90% が含まれる区間を推定したいとしましょう。

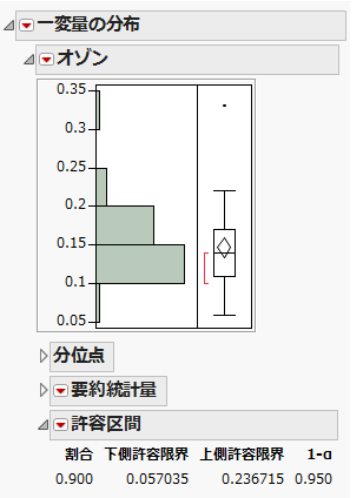
1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Cities.jmp」を開きます。
2. ［分析］>［一変量の分布］を選びます。
3. 「オゾン」を選択し、［Y, 列］をクリックします。
4. ［OK］をクリックします。
5. 「オゾン」の横にある赤い三角ボタンをクリックし、メニューから［許容区間］を選択します。

図 3.16「許容区間」ウィンドウ



6. デフォルトの値をそのまま使用し、［OK］をクリックします。

図 3.17「許容区間」レポートの例



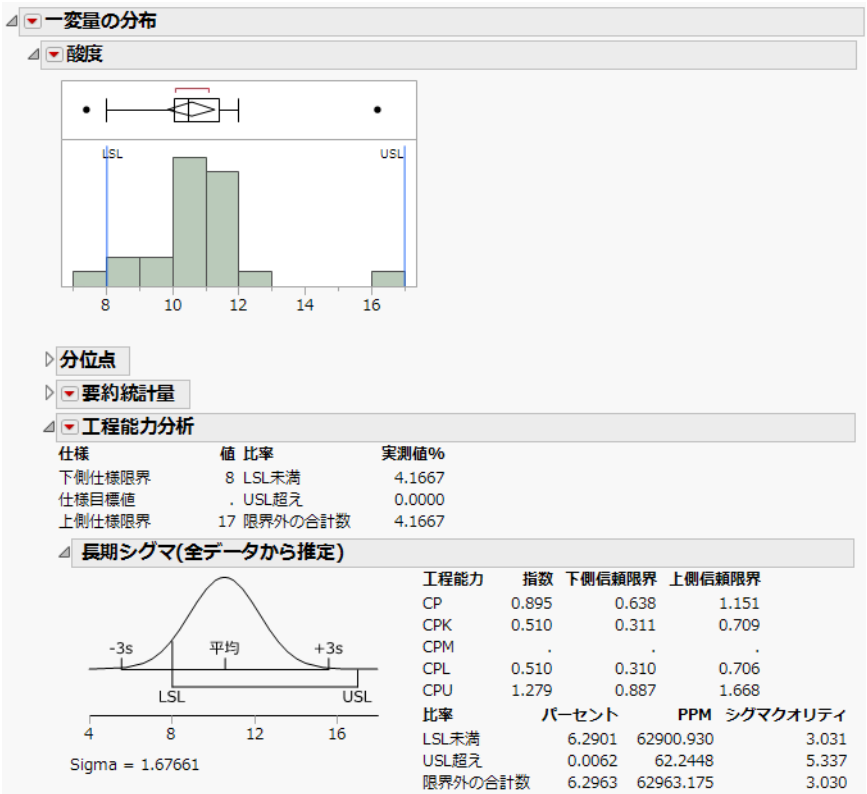
この例では、「下限許容限界」と「上限許容限界」の結果から、95% の信頼水準で、「母集団の少なくとも 90% が 0.057035 ～ 0.236715 の区間にある」と結論できます。

工程能力分析の例

会社で製造しているタイヤの摩擦度を調べたいとしましょう。この例では、下限仕様限界は100、上限仕様限界は200とします。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Pickles.jmp」を開きます。
2. ［**分析**］>［**一変量の分布**］を選びます。
3. 「**酸度**」を選択し、［**Y, 列**］をクリックします。
4. ［**OK**］をクリックします。
5. 「**酸度**」の横にある赤い三角ボタンをクリックし、メニューから［**工程能力分析**］を選択します。
6. 「**下限仕様限界**」に「8」と入力します。
7. 「**上限仕様限界**」に「17」と入力します。
8. 残りの設定項目についてはデフォルトの値をそのまま使用し、［**OK**］をクリックします。
9. 「**酸度**」の横にある赤い三角ボタンをクリックし、メニューから［**ヒストグラムオプション**］>［**縦に表示**］を選択します。

図3.18「工程能力分析」レポートの例



工程能力分析を実行すると、ヒストグラムに仕様限界線も表示されます。ヒストグラムを見ると、データの中には下限仕様限界を下回るものや、上限仕様限界に近いものがあるのが分かります。「工程能力分析」レポートに、計算結果が表示されます。CPK などの数値結果（CPK=0.510）からも、現在の工程能力は、与えられた仕様限界に対して十分でないことがわかります。

「一変量の分布」プラットフォームの統計的詳細

ここでは、「一変量の分布」で使われている統計分析の詳細について説明します。

標準誤差バーの統計的詳細

ヒストグラムで表示される標準誤差のバーは、$\sqrt{np_i(1-p_i)}$ で求めた標準誤差を表しています。この式で、$p_i=n_i/n$ です。

分位点の統計的詳細

ここでは、分位点の計算方法について説明します。

データに非欠測値が n 個あり、そのデータから p パーセント点を計算するとします。まず、n 個のデータ値を昇順に並べ、それぞれのデータ値を $y_1, y_2, \ldots, y_n$ とします。そして、p パーセント点に対する、データの順位を、$p/100(n+1)$ とします。

- 計算された順位が整数のときは、その順位に相当するデータ値が p パーセント点です。
- 計算された順位が整数でないときは、補間を行って p パーセント点を求めます。p パーセント点（q_p）は次の式で計算されます。

  $$q_p = (1-f)y_i + (f)y_{i+1}$$

  ここで、

  - n は、非欠測値の個数
  - $y_1, y_2, \ldots, y_n$ は、データを小さい順に並べた値
  - y_{n+1} は y_n として扱う
  - i、f はそれぞれ、$(n+1)p$ の整数部と小数部
  - $(n+1)p = i+f$

たとえば、データテーブルに 15 の行があり、連続尺度の列の 75 パーセント点と 90 パーセント点を計算するとしましょう。まず、データ値を昇順に並べてから、各分位点に該当する順位を次のように計算します。

$$\frac{75}{100}(15+1) = 12 \text{ および } \frac{90}{100}(15+1) = 14.4$$

75 パーセント点は y_{12} の値です。90 パーセント点は、補間によって求められます。それは、14 番目と 15 番目のデータの重み付き平均で、$y_{90} = 0.6y_{14} + 0.4y_{15}$ で計算されます。

要約統計量の統計的詳細

ここでは、「要約統計量」レポートに表示される個々の統計量について説明します。

平均

非欠測値の合計を、非欠測値の個数で割った値です。なお、［**重み**］や［**度数**］の列を起動ウィンドウで指定した場合は、次のように計算されます。

1. 各データ値に、対応する重みや度数を掛け合わせます。
2. 乗算後の値を合計し、それを、重みや度数の合計で割ります。

標準偏差

平均周りの分布の散らばりを示す指標です。通常は s で表されます。標準偏差は標本から計算された分散 s^2 の平方根です。

$$s = \sqrt{s^2} \quad \text{ここで、}$$

$$s^2 = \sum_{i=1}^{N} \frac{w_i(y_i - \bar{y}_w)^2}{N-1}$$

$\bar{y}_w$ =（重みつき）平均

平均の標準誤差

標準偏差 s を、N の平方根で割ったものです。ただし、［重み］や［度数］の列を起動ウィンドウで指定した場合は、重みや度数の合計の平方根で割ります。

歪度

平均周りの3次のモーメントを表す指標です。次式によって計算されます。

$$\sum w_i^{\frac{3}{2}} z_i^3 \frac{N}{(N-1)(N-2)} \quad \text{この式で、} \quad z_i = \frac{x_i - \bar{x}}{s}$$

上の式で、w_i は重みを表す係数（重み付けをしない場合は1）です。

尖度

平均周りの4次のモーメントを表す指標です。次式によって計算されます。

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} w_i^2 \left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

ここで、w_i は重みです（重み付けをしない場合、すべての重みは1です）。この式によって求められた尖度は、正規分布の場合には0となります。

正規分位点プロットの統計的詳細

各値の経験累積確率は、次のような式で計算されます。

$$\frac{r_i}{N+1}$$

ここで、r_iはデータ値の順位、Nは非欠測値の個数を表します。

正規分位点の値は次のような式で計算されます。

$$\Phi^{-1}\left(\frac{r_i}{N+1}\right)$$

ここで、Φは正規分布の累積確率分布関数です。

これらの正規分位点の値は、正規分布を仮定したときの順序統計量の期待値に対する Van Der Waerden の近似値です。

Wilcoxon の符号付順位検定の統計的詳細

Wilcoxon の符号付順位検定は、1 母集団、または対応のあるデータに対して用いることができ、分布が対称な場合には中央値に対する検定となります。対応のあるデータに対する検定は、対応のあるペアの差に対する 1 母集団の検定と等価です。

Wilcoxon 符号付順位検定では、同順位も考慮されます。また、データ値にゼロがある場合は、Pratt の方法で計算されます。Lehman（2006）および Pratt（1959）を参照してください。

1 標本の中央値の検定

- N 個の観測値を、以下のように表すとします。

  X_1、X_2、...、X_N
- 分布が対称である場合、Wilcoxon の符号付き順位検定の帰無仮説は、以下の通りです。

  H_0: 中央値が m に等しい
- 内部的な計算では、次式により観測値と仮説値との差が使われます。

  $D_j = X_j - m$

2 つの母集団の中央値が等しいことの検定

対応のあるデータに対しても、次のような枠組みで、Wilcoxon の符号付順位検定を適用できます。

- 2 つの母集団のそれぞれに、N 個の観測値があるとします。

  X_1、X_2、..., X_N および Y_1、Y_2、...、Y_N

- 差の分布が対称である場合、Wilcoxonの符号付き順位検定の帰無仮説は、以下の通りです。

  H_0: $_{X-Y}$の中央値が0である
- 1標本の問題とするには、次のような式で観測値間の差を計算します。

  $D_j = X_j - Y_j$

Wilcoxonの符号付順位検定の統計量

Wilcoxonの符号付順位検定の統計量は、正の符号付き順位の合計に基づきます。符号付き順位は次のように定義されます。

- 差の絶対値$|D_j|$を、小さい値から順番に順位付けします。
- この時、差が0のものがあったとしても、順位付けにそれらの値も含め、順位は1から始めます。
- また、差の絶対値に同じ値があった場合には、それらには平均順位（**中間順位**）を与えます。

以下において、差D_jの順位（中間順位）をRと記します。D_jの**符号付順位**は次のように定義されます。

- 差D_jが正の場合、符号付順位はRとなります。
- 差D_jが0の場合、符号付順位は0となります。
- 差D_jが負の場合、符号付順位は$-R$となります。

次のように定義します。

R^+は、正の符号付順位の合計。

d_0は、0と等しい符号付順位の数。

符号付順位の統計量は次のように計算されます。

$$W = R^+ - \frac{1}{4}[N(N+1) - d_0(d_0+1)]$$

Wilcoxonの符号付順位検定のp値

$N \le 20$の場合は、正確なp値が計算されます。$N > 20$の場合は、同順位のデータがある場合にはそれらに対する修正をしながら、Studentのt分布に近似して計算されます。Iman（1974）およびLehmann（1998）を参照してください。

次式で求められた統計量tが、自由度$N-1$のt分布に近似されます。

$$t = W\sqrt{\frac{N-1}{NV - W^2}}$$

ここで、分散Vは次式により計算されます。

$$V = \frac{1}{24}\left[N(N+1)(2N+1) - d_0(d_0+1)(2d_0+1) - \frac{1}{2}\sum_{i>0} d_i(d_i+1)(d_i-1)\right]$$

Vの計算式の最後の和は、同順位のペアに対する修正です。i > 0に対するd_iは、符号付順位が0以外で、かつ、符号付順位が同じグループに属するデータの個数を示します（符号付順位に同順位がまったくなければ、すべてが$d_i = 1$となるので、この項は0となります）。

標準偏差の検定の統計的詳細

検定統計量は次の式で計算されます。

$$\frac{(n-1)s^2}{\sigma^2}$$

検定統計量は、母集団が正規分布に従うとき、自由度n - 1のカイ2乗分布に従います。

「最小p値」は、両側検定のp値で、次の式で計算されます。

2×min(p1,p2)

ここで、p1は下側の片側検定のp値で、p2は上側の片側検定のp値です。

正規分位点の統計的詳細

正規分位点の値は次のような式で計算されます。

$\Phi^{-1}\left(\frac{r_i}{N+1}\right)$ ここで、

- Φ は正規分布の累積確率分布関数
- r_iはi番目のオブザベーションの順位
- Nは欠測値でないオブザベーションの個数

標準化データの統計的詳細

標準化された値は、次の式で計算されます。

$\frac{X-\overline{X}}{S_X}$ ここで、

- Xは、元のデータ値
- $\overline{X}$ は、平均
- S_X は標準偏差

予測区間の統計的詳細

JMP では、次式により予測区間を計算しています。

- 将来の **m** 個の観測値に対する予測区間

$$[\underset{\sim}{y}_m, \tilde{y}_m] = \bar{X} \pm t_{(1-\alpha/2m;n-1)} \times \sqrt{1+\frac{1}{n}} \times s \quad \text{ただし} \quad m \geq 1$$

- 将来の **m** 個の観測値から計算される平均に対する予測区間

$$[Y_l, Y_u] = \bar{X} \pm t_{(1-\alpha/2,\, n-1)} \times \sqrt{\frac{1}{m}+\frac{1}{n}} \times s \quad (m \geq 1 \text{ に対して})$$

- 将来の **m** 個の観測値から計算される標準偏差に対する予測区間

$$[s_l, s_u] = \left[s \times \sqrt{\frac{1}{F_{(1-\alpha/2;(n-1,\, m-1))}}},\; s \times \sqrt{F_{(1-\alpha/2;(m-1,\, n-1))}} \right] \quad \text{ただし} \quad m \geq 2$$

上式において、m は将来の観測値の個数、**n** は現在、分析に用いているデータの標本サイズです。

- なお、予測区間が片側の場合には、1-α の分位点を用いて計算します。

これらの式は、Hahn and Meeker（1991）の 61 ～ 64 ページに記載されています。

許容区間の統計的詳細

ここでは、片側許容区間と両側許容区間の統計的詳細について説明します。

片側許容区間

許容区間は、片側区間の場合、次式によって計算されます。

上側の場合$=\bar{x} + g's$

下側の場合$=\bar{x} - g's$

この式で、

$g' = t(1-\alpha, n-1, \Phi^{-1}(p) \cdot \sqrt{n}) / \sqrt{n}$ は、Odeh and Owen（1980）の表 1 に記載されています。

t は非心 t 分布の分位点、Φ^{-1} は標準正規分布の分位点です。

両側区間

両側区間は次の式で計算されます。

$$[\tilde{T}_{p_L}, \tilde{T}_{p_U}] = [\bar{x} - g_{(1-\alpha/2;p_L,n)}s \;,\; \bar{x} + g_{(1-\alpha/2;p_U,n)}s]$$

この式で、

sは標準偏差、$g_{(1-\alpha/2;p,n)}$はOdeh and Owen（1980）の表4に記載されている定数です。

gは、許容区間内に母集団が収まる割合に基づき計算されます。この割合は、次のように定義されます（Tamhane and Dunlop（2000））。

$$\Phi\left(\frac{\bar{x}+gs-\mu}{\sigma}\right)-\Phi\left(\frac{\bar{x}-gs-\mu}{\sigma}\right)$$

Φは標準正規分布の 累積分布関数を表します。この割合に対して、gは次式を解くことにより計算されます。

$$P\left\{\Phi\left(\frac{\overline{X}+gs-\mu}{\sigma}\right)-\Phi\left(\frac{\overline{X}-gs-\mu}{\sigma}\right)\geq 1-\gamma\right\}=1-\alpha$$

1-γが、母集団のうち、許容区間に含まれるものの割合となります。

詳細については、Hahn and Meeker（1991）の表A.1a、A.1b、A.11a、A.11bを参照してください。

工程能力分析の統計的詳細

いくつかのオプションがありますが、正規分布に基づく工程能力分析では、基本的に同じ計算式が使用されています。オプションは、シグマ（σ）の計算方法が異なるだけです。

- ［**σ (全データから推定)**］では、全体のシグマを使用します。この統計量はP_{pk}とも呼ばれており、次式によりシグマが算出されます。

$$\sigma=\sqrt{\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}}$$

注 : 環境設定で「**Ppk 工程能力ラベル設定**」をオンにした場合、全データから推定した工程能力統計量には、「P_{pk}」というラベルが付きます。このオプションを確認するには、［**ファイル**］**>**［**環境設定**］を選択し、［**プラットフォーム**］**>**［**一変量の分布**］を選択します。

- ［**σ の指定**］では、既知のシグマで工程能力分析を行います。シグマの値は計算されるのではなく、ユーザが入力します。
- ［**移動範囲**］では、移動範囲を計算する際の範囲を入力することができます。次式により、シグマが算出されます。

$\sigma=\frac{\bar{R}}{d_2(n)}$この式で、

$\bar{R}$ は移動範囲の平均

$d_2(n)$は、母標準偏差が1である**n**個の独立した正規分布に従う確率変数の範囲の期待値

- ［**σ（グループに分けて推定）、サブグループサイズの指定**］では、データの並び順によって、データをサブグループに分割して、次式によりシグマが計算されます。ここで、サイズ n_j のサブグループの数を r としています。

$$\sigma = \sqrt{\frac{1}{r}\sum_{i=1}^{r}\sum_{j=1}^{n_j}\frac{(X_{ij}-\bar{X}_i)^2}{n_j-1}} \text{ この式で、 } \bar{X}_i = \frac{1}{n_j}\sum_{j=1}^{nj} X_{ij}$$

- この式は、一般には「誤差の標準偏差（RMSE、誤差の平均平方の平方根）」と呼ばれます。

注： 次の表の信頼区間は、有意水準 0.05 を使って計算されています。

表 3.19 工程能力指数と計算式の説明

指数	指数名	計算式
CP	工程能力指数 C_p	(USL - LSL)/6*s* ここで • USL は上限仕様限界 • LSL は下限仕様限界
CP の信頼区間	CP の下側信頼限界	$CP\sqrt{\frac{\chi^2_{\alpha/2,\,n-1}}{n-1}}$
	CP の上側信頼限界	$CP\sqrt{\frac{\chi^2_{1-\alpha/2,\,n-1}}{n-1}}$
CPK（AIAG の PPK）	工程能力指数 C_{pk}	min(CPL, CPU)
CPK の信頼区間 （Bissell（1990）を参照）	下側信頼限界	$\hat{C}_{pk}\left[1-\Phi^{-1}(1-\alpha/2)\sqrt{\frac{1}{9n\hat{C}_{pk}^2}+\frac{1}{2(n-1)}}\right]$
	上側信頼限界	$\hat{C}_{pk}\left[1+\Phi^{-1}(1-\alpha/2)\sqrt{\frac{1}{9n\hat{C}_{pk}^2}+\frac{1}{2(n-1)}}\right]$

表 3.19 工程能力指数と計算式の説明（続き）

指数	指数名	計算式
CPM	工程能力指数 C_{pm}	$\frac{\min(\text{Target} - \text{LSL}, \text{USL} - \text{Target})}{3\sqrt{s^2 + (\bar{x} - \text{Target})^2}}$ **注：**CPM の信頼区間は、目標値が「下側信頼下限界」から「上側信頼限界」の範囲にない場合は表示されません。目標値がこの範囲に含まれる場合に限り表示されます。CPM の信頼区間が算出されない場合は、ログにメッセージが書き込まれます。
CPM の信頼区間	CPM の下側信頼限界	$\text{CPM}\sqrt{\frac{\chi^2_{\alpha/2,\gamma}}{\gamma}}$、ここで $\gamma = \frac{n\left(1 + \left(\frac{\bar{x} - \text{Target}}{s}\right)^2\right)^2}{1 + 2\left(\frac{\bar{x} - \text{Target}}{s}\right)^2}$
	CPM の上側信頼限界	$\text{CPM}\sqrt{\frac{\chi^2_{1-\alpha/2,\gamma}}{\gamma}}$ ここで γ = 同上。
CPL	下側仕様の工程能力比	(平均 - LSL)/3*s*
CPU	上側仕様の工程能力比	（USL - 平均）/3*s*

- 工程能力指数が 1.33 以上であるべきと考えられています。工程能力指数が 1.33 である場合、正規分布に従っているならば、10 万個の製品のうち不適合品となる個数の期待値は、約 6 個です。
- CPL の正確な 100(1 - α)% 下側および上側信頼限界が、Chou et al.（1990）による手法を一般化した式によって計算されます。Chou et al. (1990) によると、CPU の 100(1 - α)% の下側信頼限界（CPULCL）は、次式を満たします。

  $\Pr\{T_{n-1}(\delta = 3\sqrt{n})\text{CPLLCL} \le 3\text{CPL}\sqrt{n}\} = 1 - \alpha$

  $T_{n\text{-}1}(\delta)$ は、自由度が n - 1、非心度パラメータが δ の非心 *t* 分布に従います。
- CPU の正確な 100(1 - α)% 下側および上側信頼限界も、Chou et al.（1990）による手法を一般化した式によって計算されます。Chou et al. (1990) によると、CPU の 100(1 - α)% の下側信頼限界（CPULCL）は、次式を満たします。

  $\Pr\{T_{n-1}(\delta = 3\sqrt{n})\text{CPULCL} \ge 3\text{CPU}\sqrt{n}\} = 1 - \alpha$

  $T_{n\text{-}1}(\delta)$ は、自由度が n - 1、非心度パラメータが δ の非心 *t* 分布に従います。

注: 本稿執筆時点では裏付けとなる研究がないため、工程能力が標準偏差から算出された場合だけしか、工程能力指数の信頼区間は計算されません。

- シグマクオリティは次のように定義されます。

  $$\text{シグマクオリティ} = \text{Normal Quantile}\left(1 - \frac{\text{不適合の\%}}{100}\right) + 1.5$$

  $$\text{シグマクオリティ（USL 以上）} = \text{Normal Quantile}\left(1 - \frac{\text{USL 超えの\%}}{100}\right) + 1.5$$

  $$\text{シグマクオリティ（LSL 以下）} = \text{Normal Quantile}\left(1 - \frac{\text{LSL 未満の\%}}{100}\right) + 1.5$$

  たとえば、100 万個のオブザベーションの中に 3 つの不適合があった場合、シグマクオリティは 6.03 という値になります。「シグマクオリティ」においては、「限界外の合計数」の行における値が、「USL 超え」と「LSL 未満」の合計にはなりません。これは、シグマクオリティは正規分布の分位点から求められており、それらは加法的でないためです。
- ベンチマーク Z の値は次の式で計算されます。

  Z USL = (USL-Xbar)/σ = 3 * CPU
  Z LSL = (Xbar-LSL)/σ = 3 * CPL
  Z Bench = 「1 - P(LSL) - P(USL)」に対する分位点
  ここで、
  P(LSL) = Prob(X < LSL) = Z LSL における下側累積確率
  P(USL) = Prob(X > USL) = 1 - 「Z LSL における下側累積確率」

［連続分布のあてはめ］の統計的詳細

ここでは、［連続分布のあてはめ］メニューで用意されている統計分布について説明します。

正規

［**正規**］オプションを選択すると、正規分布のパラメータが推定されます。正規分布は、分布が左右対称で、ほとんどのデータが中央にある分布です。正規分布は、負の値や小数点などが、データにあってもかまいません。［**正規**］を選択し、正規分布がデータにどれほど良くあてはまるかを調べてみることができます。

正規分布には、次のパラメータがあります。

- μ（平均）は、*X* 軸上での分布の位置を定義します。
- σ（標準偏差）は、分布のばらつきまたは広がりを定義します。

標準正規分布では μ = 0、σ = 1 です。「パラメータ推定値」レポートには、μ と σ の推定値、および上側と下側の 95% 信頼限界が表示されます。

確率密度関数: $\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ ただし $-\infty < x < \infty;\ -\infty < \mu < \infty;\ 0 < \sigma$

$E(x) = \mu$

$Var(x) = \sigma^2$

対数正規

［**対数正規**］オプションを選択すると、2パラメータの対数正規分布があてはめられ、パラメータμ（尺度）とσ（形状）が推定されます。対数正規分布に従う変数Yを、$X = \ln(Y)$と変換すると、Xは正規分布に従います。データはゼロより大きくなければなりません。

確率密度関数: $\frac{1}{\sigma\sqrt{2\pi}}\frac{\exp\left[\frac{-(\log(x)-\mu)^2}{2\sigma^2}\right]}{x}$ ただし $0 \le x;\ -\infty < \mu < \infty;\ 0 < \sigma$

$E(x) = \exp(\mu + \sigma^2/2)$

$Var(x) = \exp(2(\mu + \sigma^2)) - \exp(2\mu + \sigma^2)$

Weibull、閾値つきWeibull、極値

Weibull分布は、α（尺度パラメータ）とβ（形状パラメータ）によって決められる分布です。特に機械装置や生物学上の寿命を推定するときのモデルとしてよく使用されます。［**Weibull**］オプションの結果は、［閾値つきWeibull］オプションで閾値パラメータ（θ）を0としたときと同じです。［閾値つきWeibull］オプションのデフォルトでは、閾値パラメータは、データの最小値に設定されます。閾値パラメータが既知の場合は、［**分布パラメータの指定**］オプションで、その値を指定してください。詳細については、「あてはめた分布のオプション」（61ページ）を参照してください。

閾値つきWeibull分布の確率密度関数は、次のような式で計算されます。

確率密度関数: $\frac{\beta}{\alpha^\beta}(x-\theta)^{\beta-1}\exp\left[-\left(\frac{x-\theta}{\alpha}\right)^\beta\right]$ ただし $\alpha,\beta > 0;\ \theta < x$

$E(x) = \theta + \alpha\Gamma\left(1 + \frac{1}{\beta}\right)$

$Var(x) = \alpha^2\left\{\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right\}$

上の式で、$\Gamma(\cdot)$はガンマ関数です。

［極値］分布は、2パラメータのWeibull分布における(α,β)を、$\delta = 1/\beta$および$\lambda = \ln(\alpha)$に変換した分布です。

指数

指数分布は、工業製品の故障時間や、生物学分野の生存時間を分析するときに使われています。また、あるイベントが生じる間隔をモデル化するのにも使われます。たとえば、コンピュータのサーバーにクエリーが送られてくる時間、お客様がサービスカウンターに来る間隔、交換機が着信する間隔などです。

βが1であるWeibull 分布は、指数分布です（このとき、Weibull分布のαは、指数分布のσ対応しています）。また、αが1であるガンマ分布も、指数分布です。

確率密度関数: $\frac{1}{\sigma}\exp(-x/\sigma)$ ただし $0 < \sigma;\ 0 \le x$

$E(x) = \sigma$

$Var(x) = \sigma^2$

Devore（1995）などによると、指数分布は**無記憶性**の分布です。「無記憶性」とは、*t*時間後にまだ正常に機能している部品の、t時点以降での生存時間分布（*t*時点以降の生存時間の条件付き分布）が、元の生存時間分布と同じになることです。

ガンマ

［**ガンマ**］オプションを選択すると、ガンマ分布があてはめられ、$\alpha > 0$、$\sigma > 0$ というパラメータが推定されます。αは分布の形状を、σは分布の尺度を表します。**「閾値」**（θ）と呼ばれる3番目のパラメータは、分布の下限値を示します。この閾値パラメータは、デフォルトでは0に設定されていますが（負のデータ値がない場合）、［**分布パラメータの指定**］オプションを選択すれば、任意の値に変更できます。詳細については、「あてはめた分布のオプション」（61ページ）を参照してください。

確率密度関数: $\frac{1}{\Gamma(\alpha)\sigma^{\alpha}}(x-\theta)^{\alpha-1}\exp(-(x-\theta)/\sigma)$ ただし $\theta \le x;\ 0 < \alpha,\sigma$

$E(x) = \alpha\sigma + \theta$

$Var(x) = \alpha\sigma^2$

- $\sigma = 1$の時のガンマ分布は、**標準**ガンマ分布と呼ばれています。σを変更すると、分布が*X*軸に沿って伸縮します。このようなパラメータは、一般に、尺度パラメータと呼ばれています。
- $\sigma = 2$、$\alpha = \nu/2$、$\theta = 0$のときのガンマ分布は、カイ2乗分布$\chi^2_{(\nu)}$になります。
- $\alpha = 1$、$\theta = 0$のときのガンマ分布は、指数分布になります。

標準のガンマ分布の密度関数は、$\alpha \le 1$の場合、常に減少します。$\alpha > 1$の場合、0から最大値に達するまで増加した後、減少します。

ベータ

ベータ分布の中で、標準ベータ分布は、(0, 1) の区間しか取りません。標準ベータ分布は、割合などの 0 から 1 の範囲に収まるデータに対してよく使われています。[**ベータ**] オプションを選択すると、ベータ分布の 2 つの形状パラメータ $\alpha > 0$、$\beta > 0$ が推定されます。また、下限と上限を決める θ と σ というパラメータも、同時に推定されます。ベータ分布では、$\theta \le x \le (\theta + \sigma)$ の区間にのみ値があります。θ は最小値、σ は範囲を表しています。標準ベータ分布は、$\theta = 0$、$\sigma = 1$ です。

[**分布パラメータの指定**] オプションを選択すれば、パラメータを任意の値に設定できます。その際、上側の閾値はデータの最大値以上、下側の閾値はデータの最小値以下にしなければなりません。[分布パラメータの指定] オプションの詳細については、「あてはめた分布のオプション」(61 ページ) を参照してください。

確率密度関数: $\dfrac{1}{B(\alpha, \beta)\sigma^{\alpha+\beta-1}}(x-\theta)^{\alpha-1}(\theta+\sigma-x)^{\beta-1}$ ただし $\theta \le x \le \theta+\sigma$; $0 < \sigma, \alpha, \beta$

$$E(x) = \theta + \sigma\frac{\alpha}{\alpha+\beta}$$

$$Var(x) = \frac{\sigma^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

上の式で、$B(\cdot)$ はベータ関数です。

正規混合分布法

[**正規混合**] オプションでは、正規分布の混合分布をあてはめます。多峰性の分布にも対応した柔軟な分布です。

2 つもしくは 3 つの正規分布の混合分布をあてはめるには、[**二重正規混合**] もしくは [**三重正規混合**] オプションをそれぞれ選択します。または、[**その他**] オプションを選択して、k 個の正規分布の混合分布をあてはめることもできます。グループごとに 1 つずつ個別の平均、標準偏差、全体に占める割合が推定されます。

確率密度関数: $\displaystyle\sum_{i=1}^{k}\frac{\pi_i}{\sigma_i}\phi\left(\frac{x-\mu_i}{\sigma_i}\right)$

$$E(x) = \sum_{i=1}^{k}\pi_i\mu_i$$

$$Var(x) = \sum_{i=1}^{k}\pi_i(\mu_i^2+\sigma_i^2) - \left(\sum_{i=1}^{k}\pi_i\mu_i\right)^2$$

上の式で、μ_i、σ_i、π_i はそれぞれ、第 i 群の平均、標準偏差、割合です。$\phi(\cdot)$ は標準正規分布の確率密度関数です。

平滑曲線

［**平滑曲線**］オプションは、ノンパラメトリックな密度の推定値（カーネル密度推定）を計算し、平滑曲線をあてはめます。ヒストグラム上に平滑曲線が表示され、プロットの下にスライダが表示されます。スライダを使ってカーネル標準偏差を変更することで、平滑化の度合いを調節できます。［**カーネル標準偏差**］の初期値は、データの標準偏差に基づいて算出されます。

Johnson Su、Johnson Sb、Johnson Sl

Johnson の分布システムは、変換すると正規分布になる3つの分布で構成されています。これら3つの分布は、Yに関して有界でないJohnson Su分布、上下に有界なJohnson Sb分布（0 < Y < 1）、そして対数正規分布のJohnson Sl分布です。

注: Sは「システム」、添え字は範囲を示します。JMPで採用されている手法とは異なりますが、各Johnsonシステムの選択基準については、Slifker and Shapiro（1980）が参考になります。

Johnson分布システムは、柔軟性が高いことで人気があります。これら3つで構成される分布システムは、歪度と尖度のあらゆる組み合わせに対応しており、柔軟にデータにあてはまります。

Zを標準正規変量とすると、分布システムは次のように定義されます。

$$Z = \gamma + \delta f(Y)$$

Johnson Suの場合は、

$$f(Y) = \ln\left(Y + \sqrt{1 + Y^2}\right) = \sinh^{-1} Y$$

$$Y = \frac{X - \theta}{\sigma} \qquad -\infty < X < \infty$$

Johnson Sbの場合は、

$$f(Y) = \ln\left(\frac{Y}{1 - Y}\right)$$

$$Y = \frac{X - \theta}{\sigma} \qquad \theta < X < \theta + \sigma$$

Johnson Slの場合は、$\sigma = \pm 1$ で

$$f(Y) = \ln(Y)$$

$$Y = \frac{X - \theta}{\sigma} \qquad \theta < X < \infty \quad \sigma = 1 \text{ の場合}$$
$$-\infty < X < \theta \quad \sigma = -1 \text{ の場合}$$

Johnson Su

確率密度関数: $\frac{\delta}{\sigma}\left[1+\left(\frac{x-\theta}{\sigma}\right)^2\right]^{-1/2}\phi\left[\gamma+\delta\sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)\right]$ ただし $-\infty < x, \theta, \gamma < \infty$; $0 < \theta, \delta$

Johnson Sb

確率密度関数: $\phi\left[\gamma+\delta\ln\left(\frac{x-\theta}{\sigma-(x-\theta)}\right)\right]\left(\frac{\delta\sigma}{(x-\theta)(\sigma-(x-\theta))}\right)$ ただし $\theta < x < \theta+\sigma$; $0 < \sigma$

Johnson Sl

確率密度関数: $\frac{\delta}{|x-\theta|}\phi\left[\gamma+\delta\ln\left(\frac{x-\theta}{\sigma}\right)\right]$ ただし $\sigma = 1$ の場合 $\theta < x$、 $\sigma = -1$ の場合 $\theta > x$

上の式で、$\phi(\cdot)$ は標準正規分布の確率密度関数です。

注: デフォルトのレポートには、パラメータの信頼区間は表示されません。Johnson の分布は正規分布へと変換するためのものであり、パラメータの信頼区間にはあまり意味がないためです。パラメータの信頼区間を表示するには、レポートを右クリックし、[**列**] > [**下側 95% 信頼限界**] と [**上側 95% 信頼限界**] を選択します。

一般化対数（Glog）

この分布は、生物検定法など、正規分布に従うことがほとんどなく、分散が一定でないデータに役立ちます。[一般化対数] 分布は、μ（位置）、σ（尺度）、λ（形状）の 3 つのパラメータで記述されます。

確率密度関数: $\phi\left\{\frac{1}{\sigma}\left[\log\left(\frac{x+\sqrt{x^2+\lambda^2}}{2}\right)-\mu\right]\right\}\frac{x+\sqrt{x^2+\lambda^2}}{\sigma(x^2+\lambda^2+x\sqrt{x^2+\lambda^2})}$

ただし $0 \le \lambda$; $0 < \sigma$; $-\infty < \mu < \infty$

一般化対数分布は、正規分布へと変換するためのものであり、正規分布とは次のような関係にあります。

$z = \frac{1}{\sigma}\left[\log\left(\frac{x+\sqrt{x^2+\lambda^2}}{2}\right)-\mu\right] \sim N(0,1)$ のとき、$x \sim Glog(\mu,\sigma,\lambda)$

$\lambda = 0$ の場合、一般化対数分布は LogNormal (μ,σ) に等しくなります。

注: デフォルトのレポートには、パラメータの信頼区間は表示されません。一般化対数分布は正規分布へと変換するためのものであり、パラメータの信頼区間にはあまり意味がないためです。パラメータの信頼区間を表示するには、レポートを右クリックし、[**列**] > [**下側 95% 信頼限界**] と [**上側 95% 信頼限界**] を選択します。

すべて

［すべて］オプションを選択すると、用意されているほとんどすべての分布があてはめられ、「分布の比較」レポートにおいて、AICcが小さい順に、それらの分布の名前が表示されます。

AICcは、次のような式で計算されます。

$$\text{AICc} = -2\log L + 2\nu + \frac{2\nu(\nu+1)}{n-(\nu+1)}$$

ここで、

- logLは対数尤度
- nは標本サイズ
- νはパラメータの個数

すべての分布をあてはめる際に、データに負の値がある場合は、正の値のみを対象とする分布は除外されます。データに整数以外の値がある場合は、離散分布は除外されます。すべての場合において、閾値パラメータがある分布（ベータ分布やJohnson Sb 分布など）は除外されます。

［離散分布のあてはめ］の統計的詳細

ここでは、［離散分布のあてはめ］メニューで用意されている統計分布について説明します。

Poisson

Poisson分布では、尺度パラメータλ > 0が推定されます。

確率関数: $\frac{e^{-\lambda}\lambda^{x}}{x!}$ ただし $0 \le \lambda < \infty$; $x = 0,1,2,\ldots$

$E(x) = \lambda$

$\text{Var}(x) = \lambda$

Poisson分布は離散分布なので、分布関数は整数ごとに変化する階段関数になります。

ガンマ Poisson

ガンマPoisson分布は、平均μが異なる複数のPoisson分布から、データが生成されているときに役立ちます。例として、複数の交差点で発生した事故の発生数などが挙げられます。この場合、各交差点における事故発生数の平均（μ）は、交差点によって異なると考えられます。

ガンマPoisson分布は、x|μが平均μのPoisson分布に従い、μがパラメータ(α,τ)のガンマ分布に従うという仮定から導出できます。ガンマPoisson分布には、λ=ατとσ=τ+1の2つのパラメータがあります。パラメータσは過分散パラメータです。σ > 1の場合は、過大分散、つまり通常のPoisson分布よりも分散が大きくなります。σ = 1の場合、xの分布は平均λのPoisson分布になります。

確率関数: $\frac{\Gamma\left(x+\frac{\lambda}{\sigma-1}\right)}{\Gamma(x+1)\Gamma\left(\frac{\lambda}{\sigma-1}\right)}\left(\frac{\sigma-1}{\sigma}\right)^{x}\sigma^{-\frac{\lambda}{\sigma-1}}$ ただし $0<\lambda$; $1\le\sigma$; $x=0,1,2,...$

$E(x)=\lambda$

$Var(x)=\lambda\sigma$

上の式で、$\Gamma(\cdot)$ はガンマ関数です。

前述のとおり、$x|\mu$ は平均 μ の Poisson 分布に従い、μ はパラメータ (α,τ) のガンマ分布に従っていると仮定されています。このプラットフォームでは、$\lambda=\alpha\tau$ および $\sigma=\tau+1$ がパラメータとして推定されます。α と τ の推定値は、次式により計算できます。

$$\hat{\tau} = \hat{\sigma} - 1$$

$$\hat{\alpha} = \frac{\hat{\lambda}}{\hat{\tau}}$$

ただし、σ の推定値が 1 の場合は、上記の式を使用できません。その場合、ガンマ Poisson 分布は平均 λ の Poisson 分布に等しくなり、$\hat{\lambda}$ がその λ の推定値となります。

α の推定値が整数の場合、ガンマ Poisson 分布は、次の確率関数を持つ負の二項分布に等しくなります。

$$p(y) = \binom{y+r-1}{y}p^{r}(1-p)^{y} \quad \text{ただし} \quad 0\le y$$

ここで $r=\alpha$ および $(1\text{-}p)/p=\tau$

「Samples/Scripts」フォルダの中の「demoGammaPoisson.jsl」を実行すると、パラメータ λ と σ のガンマ Poisson 分布とパラメータ λ の Poisson 分布を比較できます。

二項

［**二項**］オプションでは、一定の標本サイズか、標本サイズを含む列のいずれかを指定することができます。

確率関数: $\binom{n}{x}p^{x}(1-p)^{n-x}$ ただし $0\le p\le 1$; $x=0,1,2,...,n$

$E(x)=np$

$Var(x)=np(1\text{-}p)$

上の式で、n は独立した試行の回数を示します。

注: 二項分布のパラメータに対する信頼区間の計算には、スコア信頼区間（Agresti 1998）が使用されています。

ベータ二項

ベータ二項分布は、確率πが異なる複数の二項分布から、データが生成されているときに役立ちます。たとえば、複数の製造ラインにおける不適合品の個数などが挙げられます。この場合、不適合となる確率（π）は、製造ラインによって異なると考えられます。

ベータ二項分布は、x|πが二項分布(n,π)に従い、πがベータ分布(α,β)に従うという仮定から導出できます。ベータ二項分布には、p = α/(α+β) と δ = 1/(α+β+1) の2つのパラメータがあります。パラメータδは、過分散パラメータです。δ > 0の場合は、過大分散、つまり通常の二項分布よりも分散が大きくなります。δ < 0の場合は、過小分散です。δ = 0の場合、xは二項分布(n,p)に従います。ベータ二項分布は、$n \geq 2$の場合にのみ存在します。

確率関数: $$\binom{n}{x}\frac{\Gamma\left(\frac{1}{\delta}-1\right)\Gamma\left[x+p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-x+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left(n+\frac{1}{\delta}-1\right)}$$

ただし $0 \leq p \leq 1$; $\max\left(-\frac{p}{n-p-1}, -\frac{1-p}{n-2+p}\right) \leq \delta \leq 1$; $x = 0,1,2,\ldots,n$

$E(x) = np$

$Var(x) = np(1-p)[1+(n-1)\delta]$

上の式で、$\Gamma(\cdot)$ はガンマ関数です。

前述のとおり、$x|\pi$は二項分布(n,π)に従い、πはベータ分布(α,β)に従うと仮定されています。このプラットフォームでは、p = α/(α+β) と δ = 1/(α+β+1) のパラメータ推定値が計算されます。αとβの推定値は、次式で計算できます。

$$\hat{\alpha} = \hat{p}\left(\frac{1-\hat{\delta}}{\hat{\delta}}\right)$$

$$\hat{\beta} = (1-\hat{p})\left(\frac{1-\hat{\delta}}{\hat{\delta}}\right)$$

ただし、δの推定値が0の場合は、上記の式が使用できません。その場合、ベータ二項分布は、確率pの二項分布となり、$\hat{p}$がそのpの推定値となります。

なお、ベータ二項分布におけるパラメータの信頼区間は、プロファイル尤度法によって計算されます。

「Samples/Scripts」フォルダにある「demoBetaBinomial.jsl」を実行すると、過分散パラメータδのベータ二項分布とパラメータpとのn = 20の二項分布を比較できます。

すべての分布の比較

［すべて］オプションを選択すると、用意されているほとんどすべての分布があてはめられ、「分布の比較」レポートにおいて、AICcが小さい順に、それらの分布の名前が表示されます。

AICcは、次のような式で計算されます。

$$AICc = -2\log L + 2\nu + \frac{2\nu(\nu+1)}{n-(\nu+1)}$$

ここで、

- logLは対数尤度
- nは標本サイズ
- νはパラメータの個数

あてはめた分布の分位点

「診断プロット」に表示される分位点と、[**あてはめたモデルの分位点を保存**]コマンドで保存される分位点は、次の方法で計算されます。

1. データを昇順に並べ、順位を付けます。この時、同順位（タイ）のデータに対しても、別々の順位を割り当てます。
2. 各データに対して、$p_{[i]}$ = 順位/(n+1) を計算します。
3. $quantile_{[i]} = Quantile_d(p_{[i]})$を計算します。$Quantile_d$は、あてはめた分布の分位点関数です。また、i = 1,2,...,nです。

あてはめた分布に対するオプション

ここでは、あてはめた分布に対する、適合度検定と仕様限界について説明します。

適合度

表3.20 JMPの適合度検定の説明

分布	パラメータ	適合度検定
正規[a]	μとσが未知	Shapiro-Wilk（n ≤ 2000) KSL（Kolmogorov-Smirnov-Lilliefors）（n > 2000）
	μとσが既知	KSL（Kolmogorov-Smirnov-Lilliefors）
	μとσのどちらかが既知	（なし）
対数正規	μとσが未知または既知	KolmogorovのD検定
Weibull損失関数	αとβが未知または既知	Cramér-von MisesのW^2検定
閾値つきWeibull	α、β、θが未知または既知	Cramér-von MisesのW^2検定
極値損失関数	αとβが未知または既知	Cramér-von MisesのW^2検定

表 3.20 JMP の適合度検定の説明（続き）

分布	パラメータ	適合度検定
指数	σ が既知または未知	Kolmogorov の *D* 検定
ガンマ	α と σ が既知	Cramér-von Mises の W^2 検定
	α と σ のどちらかが未知	（なし）
ベータ	α と β が既知	Kolmogorov の *D* 検定
	α と β のどちらかが未知	（なし）
二項	ρ が既知または未知、*n* が既知	Kolmogorov の *D* 検定（n ≤ 30） Pearson のカイ 2 乗（χ^2）（n > 30）
ベータ二項	ρ と δ が既知または未知	Kolmogorov の *D* 検定（n ≤ 30） Pearson のカイ 2 乗（χ^2）（n > 30）
Poisson	λ が既知または未知	Kolmogorov の *D* 検定（n ≤ 30） Pearson のカイ 2 乗（χ^2）（n > 30）
ガンマ Poisson	λ または σ が既知または未知	Kolmogorov の *D* 検定（n ≤ 30） Pearson のカイ 2 乗（χ^2）（n > 30）

a. Johnson の 3 つの分布と一般化対数分布では、データを正規分布に変換してから、正規性の検定が行われます。

仕様限界

非正規分布に一般化された工程能力指数は、次のように計算されます。次式において、*T* は目標値、*LSL* および *USL* は、それぞれ下側仕様限界および上側仕様限界、P_α はあてはめた分布の分位点です。

$$C_{pl} = \frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}$$

$$C_{pu} = \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}$$

$$C_p = \frac{USL - LSL}{P_{0.99865} - P_{0.00135}}$$

$$C_{pk} = \min\left(\frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}\right)$$

$$K = 2 \times \frac{\left|\frac{1}{2}(USL + LSL) - P_{0.5}\right|}{USL - LSL}$$

$$C_{pm} = \frac{\min\left(\frac{T - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - T}{P_{0.99865} - P_{0.5}}\right)}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

分布が正規分布の場合には、これらは、通常の工程能力指数と等しくなります。通常の工程能力指数については、詳細については、表3.19を参照してください。

Kシグマから仕様限界を設定

このオプションを選択した後、工程能力分析におけるKの値を入力し、仕様限界が片側なのか、両側なのかを選択してください。まず、正規分布において「平均 ± 標準偏差のK倍」の範囲外になる確率（上側または下側）が計算されます。そして、あてはめた分布において、その確率に相当する分位点が求められます。その分位点が、工程能力分析の仕様限界として設定されます。このオプションは［**分位点**］オプションで計算される工程能力分析と似ていますが、このオプションを指定した場合には、確率の代わりにKを指定します。Kは、正規分布において、仕様限界が平均値から標準偏差の何倍離れているかを示す値です。

たとえば、K = 3と指定した場合、正規分布において、平均値から標準偏差の3倍だけ移動した分位点は、それぞれ下側確率0.00135と0.99865に対応した分位点です。よって、あてはめた分布の0.00135に対する分位点が下側仕様限界に、0.99865に対する分位点が上側仕様限界に設定されます。そして、工程能力分析がこれらの仕様限界に基づいて実行されます。

第 4 章

「二変量の関係」プラットフォームの概要

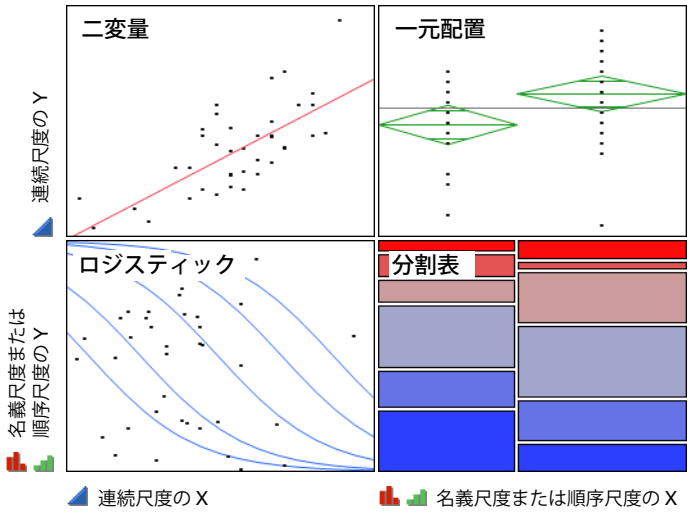
２変数の関係を調べる

「二変量の関係」プラットフォームでは、Ｘ変数とＹ変数の尺度によって、自動的に分析が選択されます。

次の４種類の分析を実行できます。

- 二変量分析
- 一元配置の分散分析
- ロジスティック回帰
- 分割表分析

図4.1 ４種類の分析



目次

「二変量の関係」プラットフォームの概要

「二変量の関係」プラットフォームは、次の4つの分析プラットフォームで構成されています。

プラットフォーム	尺度	説明	参照
二変量	X：連続 Y：連続	2つの連続変数における関係を分析します。	詳細については、「二変量分析」を参照してください。
一元配置	X：名義／順序 Y：連続	X変数のグループごとに、連続尺度のY変数の分布がどのように異なるかを分析します。	詳細については、「一元配置分析」を参照してください。
ロジスティック	X：連続 Y：名義／順序	カテゴリカルな応答変数の確率を、連続な説明変数（X）にあてはめます。	詳細については、「ロジスティック分析」を参照してください。
分割表	X：名義／順序 Y：名義／順序	X変数のグループごとに、カテゴリカルなY変数の分布がどのように異なるかを分析します。	詳細については、「分割表分析」を参照してください。

「二変量の関係」プラットフォームの起動

「二変量の関係」プラットフォームを起動するには、［**分析**］ > ［**二変量の関係**］を選択します。

図4.2「二変量の関係」起動ウィンドウ

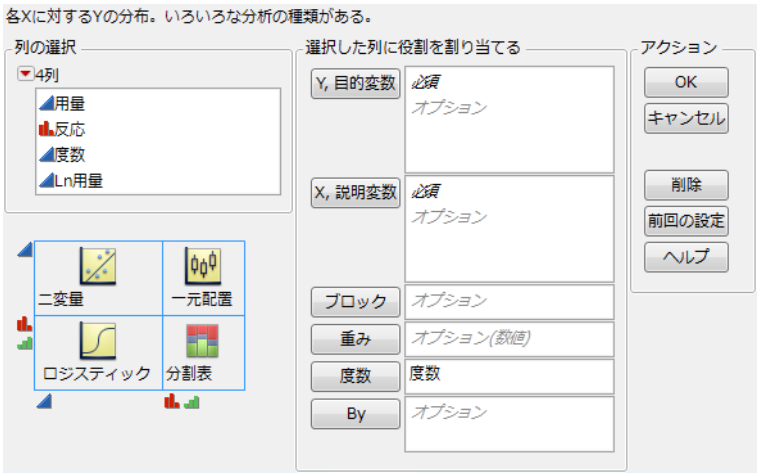


表 4.1 「二変量の関係」起動ウィンドウの説明

二変量、一元配置、ロジスティック、分割表	このグリッドには、変数の尺度に応じて、どの分析が実行されるかが示されています。列に役割を割り当てると、実行される分析名がグリッドの上部に表示されます。
ブロック	（オプション。一元配置分析または分割表分析の場合のみ） • 分散分析の場合には、指定された因子を第2 の因子として、交互作用のない二元配置分散分析が行われます。分散分析でブロック変数を指定する場合には、釣り合いがとれていなければいけません。つまり、ブロックごとの各グループのセル度数がすべて等しくないといけません。プロットでは、Y変数の値がブロック変数ごとに中心化されます。 • 分割表の場合には、指定されたブロックの列を、層別因子として、Cochran-Mantel-Haenszel 検定が行われます。

起動ウィンドウの詳細については、『JMP の使用法』を参照してください。

「JMP スターター」ウィンドウからの各分析の実行

「JMP スターター」ウィンドウから、特定の分析（**「二変量」**、**「一元配置」**、**「ロジスティック」**、または **「分割表」**）を実行できます。この方法では、列の尺度（Y変数とX変数）に応じた分析を自分自身で指定する必要があります。詳細については、表 4.2 を参照してください。

「JMP スターター」ウィンドウから特定の分析を実行するには、次の手順に従います。

1. カテゴリリストの［**基本統計**］をクリックします。
2. 実行する分析のボタンをクリックします。

プラットフォームの起動ウィンドウはほとんど同じですが、［Y］ボタンと［X］ボタンの名前が、選択された分析によって次のように異なっています。

表 4.2 プラットフォームとボタン

プラットフォームまたは分析	［Y］ボタン	［X］ボタン
二変量の関係	Y, 目的変数	X, 説明変数
二変量	Y, 目的変数	X, 説明変数
一元配置	Y, 目的変数	X, グループ変数
ロジスティック	Y, カテゴリカル応答変数	X, 連続量の説明変数
分割表	Y, 応答カテゴリ	X, グループ化カテゴリ

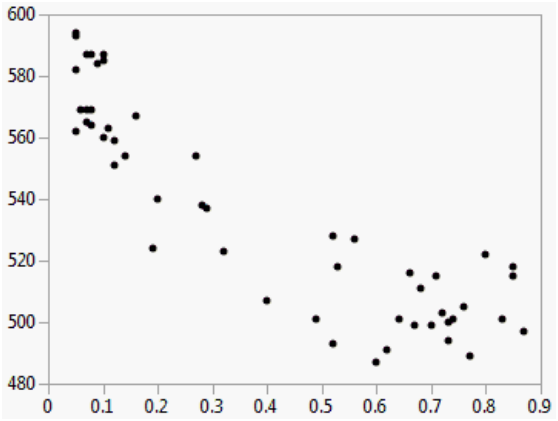
第 5 章

二変量分析
2 つの連続変数の関係を調べる

「二変量」プラットフォームでは、2つの連続変数における関係を分析します。**2つの連続変数**に対して「二変量の関係」を実行すると、「二変量」プラットフォームが呼び出されます。「二変量」という用語は単に、分析対象の変数が1つ（一変量）や複数（多変量）ではなく、2つであることを表しています。

結果は、散布図によって図示されます。散布図の各点は、単一の被験者または個体のXスコアとYスコアを表します。つまり、各点が2つの変数の値を示しています。散布図では、2つの変数間の関係の程度や傾向を一目で把握できます。単回帰や多項式回帰などの様々なモデルのあてはめを追加していくことができます。

図 5.1 二変量分析の例



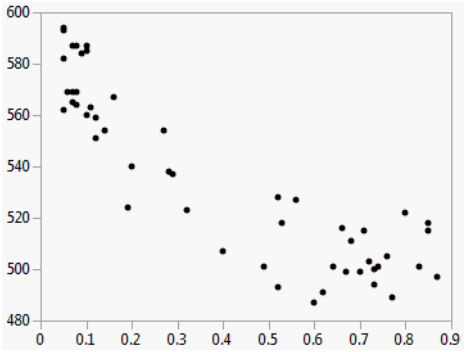
目次

二変量分析の例

この例では、「SAT.jmp」サンプルデータテーブルを使用します。米国の 50 州およびコロンビア特別区に対して、大学進学適性全国テスト（SAT）における言語と数学の点数が記録されています。2004 年度について、受験率と言語テストの点数との関係を調べましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「SAT.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**2004 言語**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**2004 受験率 (%)**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。

図 5.2 受験率に対するテストスコアの例



受験率が低い州ほど、言語テストの点数が高いことがわかります。

「二変量」プラットフォームの起動

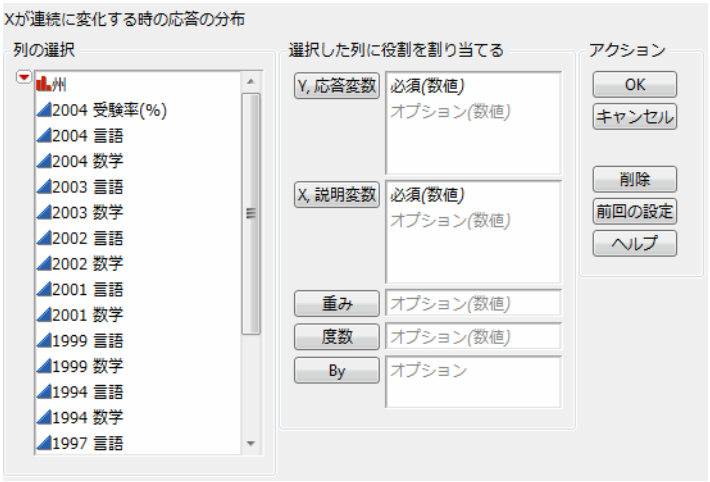
二変量分析は、「二変量の関係」プラットフォームまたは「二変量」プラットフォームを使用して実行できます。どちらのプラットフォームを使用しても結果は同じです。

- 「二変量の関係」プラットフォームを起動するには、［**分析**］>［**二変量の関係**］を選択します。

  または

- 「二変量」プラットフォームを起動するには、「JMP スターター」ウィンドウのカテゴリリストで［**基本統計**］をクリックし、［**二変量**］をクリックします。

図 5.3「二変量」起動ウィンドウ



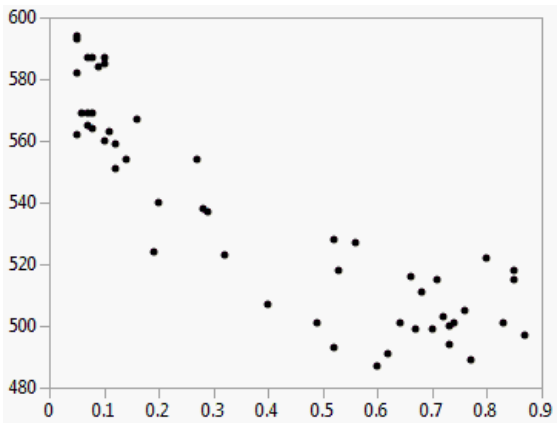
この起動ウィンドウの詳細については、「「二変量の関係」プラットフォームの概要」の章（91 ページ）を参照してください。

[OK] をクリックすると、二変量に対する散布図が表示されます。詳細については、「二変量に対する散布図」（98 ページ）を参照してください。

二変量に対する散布図

図 5.4 は、「二変量分析の例」（97 ページ）の手順に従って作成したものです。

図 5.4 二変量に対する散布図



注: データテーブルで非表示にした行は、二変量の散布図でも非表示になります。

「二変量の関係」レポートには、まず、*X* と *Y* の組み合わせごとに、散布図が描かれます。プロット上の変数は、ドラッグ & ドロップ操作で置き換えることができます。使用中の変数を入れ替える場合は、一方の軸の変数をドラッグし、他方の軸の上にドロップします。データテーブル内の他の変数を使いたい場合は、「列」パネルで該当する変数をクリックし、軸の上までドラッグします。

この散布図は、JMP の他のグラフと同じような操作を行えます（散布図のサイズの変更、矢印ツールやブラシツールによる点の強調表示、点のラベルの指定など）。これらの機能の詳細については、『JMP の使用法』を参照してください。

赤い三角ボタンをクリックすると、メニューが表示されます。このメニューには、さまざまなあてはめのコマンドが用意されています。コマンドを選択すると、プロットに曲線をあてはめたり、統計量のレポートを表示したり、追加のメニューを表示したりできます。詳細については、「あてはめ用コマンドとオプション」（99 ページ）を参照してください。

あてはめ用コマンドとオプション

注: Y 変数または X 変数を複数指定した場合は、[あてはめのグループ] メニューが表示されます。このメニューのオプションを使って、レポートの配置を変えたり、適合度の良い順に並べ替えたりできます。詳細については、『基本的な回帰モデル』を参照してください。

「二変量の関係」の赤い三角ボタンのメニューには、表示オプション、あてはめオプション、およびコントロールオプションがあります。

表 5.1 表示オプションとコントロールオプションの説明

点の表示	散布図上の点の表示／非表示を切り替えます。チェックマークが付いている場合は、点が表示されています。
ヒストグラム軸	散布図の *X* 軸と *Y* 軸にヒストグラムを追加表示します。チェックマークが付いている場合は、ヒストグラム軸が表示されています。詳細については、「ヒストグラム軸」（102 ページ）を参照してください。
グループ別	分類変数（グループ変数）を選択できます。グループ変数の水準ごとに個別の分析が行われ、散布図に回帰曲線または楕円が重ねて表示されます。詳細については、「グループ別」（119 ページ）を参照してください。
スクリプト	すべてのプラットフォームに共通なオプションが表示されます。分析を再実行するオプションや、JSL コマンドをウィンドウやファイルに保存するオプションがあります。詳細は、『JMP の使用法』を参照してください。

あてはめのコマンドを選択すると、次の情報が追加で表示されます。

- 直線、曲線、密度などが、散布図に重ねて表示されます。
- 赤い三角ボタンのメニューが、レポートウィンドウに追加されます。
- 分析結果が、レポートウィンドウに追加されます。

図 5.5［平均のあてはめ］コマンドの例

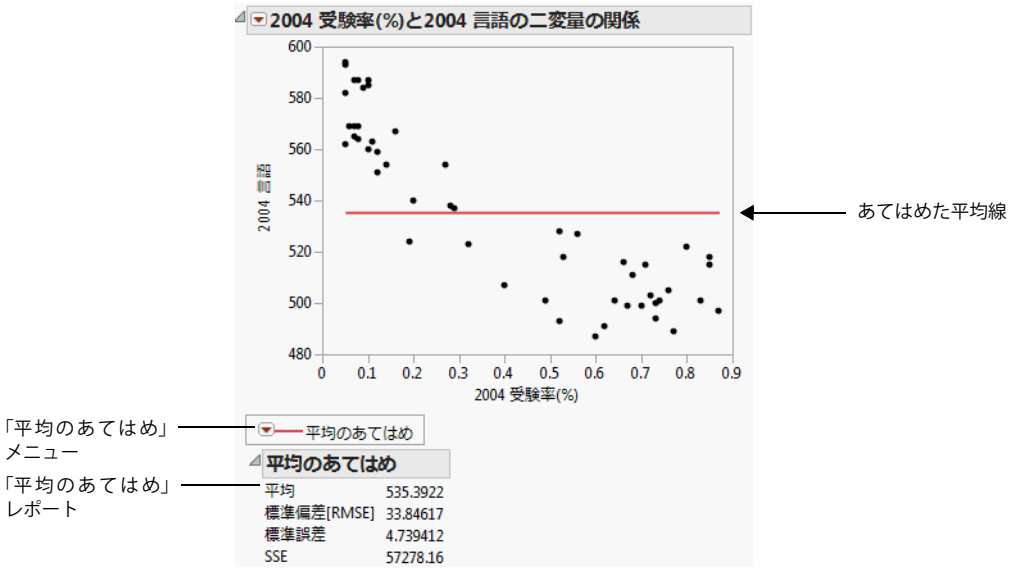


表 5.2 あてはめコマンドの説明

平均のあてはめ	応答変数 Y の平均を示す水平線を、散布図に表示します。	詳細については、「平均のあてはめ」（102 ページ）を参照してください。
直線のあてはめ	最小 2 乗法によって、散布図に回帰直線をあてはめます。	詳細については、「直線のあてはめと多項式のあてはめ」（103 ページ）を参照してください。
多項式のあてはめ	最小 2 乗法によって、指定された次数の多項式曲線をあてはめます。	詳細については、「直線のあてはめと多項式のあてはめ」（103 ページ）を参照してください。
その他のあてはめ	Y や X を変換した後に、直線や曲線をあてはめます。変換には、自然対数、平方根、2 乗、逆数、指数があります。多項式の中心化をオフにしたり、切片と傾きの制約を指定したり、多項式モデルをあてはめることも可能です。	詳細については、「その他のあてはめ」（110 ページ）を参照してください。

表 5.2 あてはめコマンドの説明（**続き**）

ノンパラメトリックな曲線	滑らかな曲線をあてはめるためのオプションが用意されています。	詳細については、「ノンパラメトリックな曲線」（112 ページ）を参照してください。
直交のあてはめ	Y だけではなく、X の値もばらつくような状況で役立つ直交回帰をあてはめるオプションがあります。X と Y の分散比に対して、いくつかのサブオプションがあります。	詳細については、「直交のあてはめ」（115 ページ）を参照してください。
ロバスト	ここで用意されているオプションでは、データにおける外れ値の影響を少なくしながら、回帰モデルをあてはめます。	詳細については、「ロバスト」（116 ページ）を参照してください。
確率楕円	X と Y から推定される二変量正規分布から計算された確率楕円をプロットします。	詳細については、「確率楕円」（116 ページ）を参照してください。
ノンパラメトリック密度	滑らかな密度等高線をプロットします。密度等高線はデータ点の密度を表します。	詳細については、「ノンパラメトリック密度」（118 ページ）を参照してください。

注: あてはめた線は、［**あてはめの削除**］コマンドで削除できます。詳細は、「あてはめ用メニューのオプション」（120 ページ）を参照してください。

あてはめ用コマンドの種類

あてはめを行うコマンドは、回帰のあてはめと、密度推定の 2 つに大別されます。

カテゴリ	説明	あてはめコマンド
回帰のあてはめ	回帰分析は、観測されたデータ点に線をあてはめます。このあてはめ手法には、最小 2 乗のあてはめ、スプライン曲線のあてはめ、カーネル平滑化、直交のあてはめ、ロバストなあてはめがあります。	**平均のあてはめ** **直線のあてはめ** **多項式のあてはめ** **その他のあてはめ** **ノンパラメトリックな曲線** **直交のあてはめ** **ロバスト**
密度推定	密度推定では、二変量の分布が点にあてはめられます。楕円形の等高線で表される二変量正規密度と、一般的なノンパラメトリック密度が用意されています。	**確率楕円** **ノンパラメトリック密度**

同じあてはめ用コマンドの繰り返し

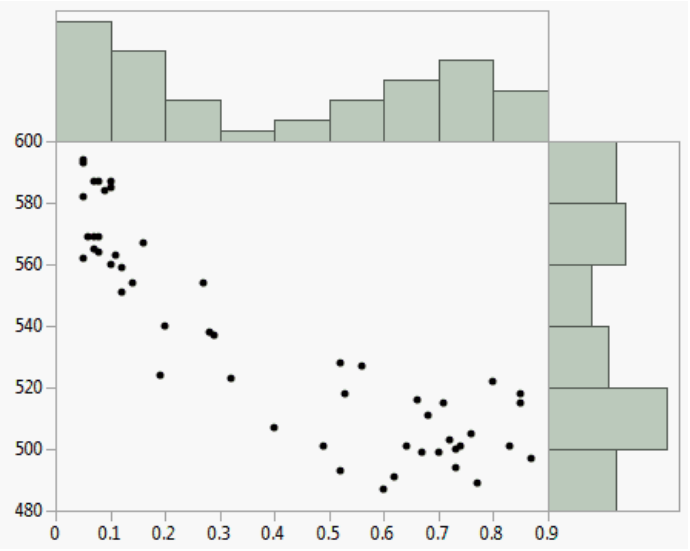
同じあてはめ用コマンドを、繰り返し選択できます。あてはめを行うたびに散布図に新しい線が追加されます。ある分析を試した後、点を除外してから同じ分析を実行し、それらの分析結果を散布図上で比較できます。

複数の分析結果に同じあてはめ用コマンドを一括で適用したい場合は、Ctrl キーを押しながらコマンドを選択します。

ヒストグラム軸

［**ヒストグラム軸**］は、散布図の *X* 軸と *Y* 軸にヒストグラムを表示するオプションです。X 変数と Y 変数の周辺分布が、ヒストグラムによって描かれます。

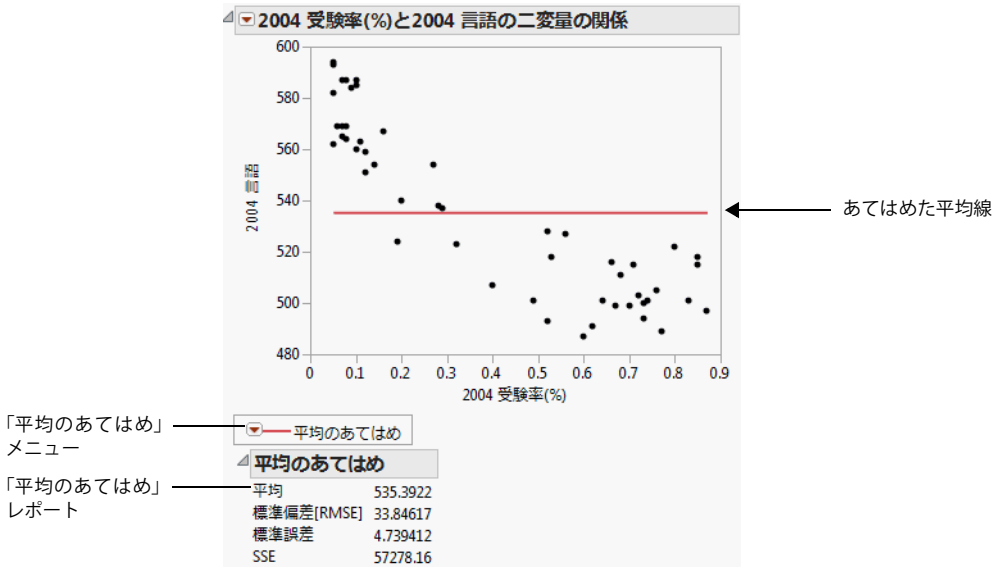
図 5.6 ヒストグラム軸の例



平均のあてはめ

［**平均のあてはめ**］コマンドを選ぶと、応答変数 **Y** の平均を示す水平線が、散布図に追加されます。平均の水平線は、他のあてはめ（直線、信頼曲線、多項式曲線などのあてはめ）と比較するための参照線として使えます。

図 5.7 平均のあてはめの例



「平均のあてはめ」レポート

［平均のあてはめ］レポートは、平均のあてはめに関連する要約統計量をまとめたものです。

表 5.3「平均のあてはめ」レポートの説明

平均	応答変数の平均。モデル内で効果が指定されていないときの応答の予測値です。
標準偏差 [RMSE]	応答変数の標準偏差。誤差の平均平方の平方根で、「誤差の標準偏差（RMSE）」と呼ばれます。
標準誤差	応答変数の平均の標準偏差。RMSE を標本サイズの平方根で割った値です。
SSE	単純な平均モデルの誤差平方和。モデルのあてはめでは、「分散分析」レポートに**「誤差」**の平方和として表示されます。

関連情報

- 「あてはめ用メニュー」（119 ページ）

直線のあてはめと多項式のあてはめ

［**直線のあてはめ**］コマンドは、最小 2 乗法により、散布図に回帰直線をあてはめます。［**多項式のあてはめ**］コマンドは、最小 2 乗法により、指定された次数の多項式曲線をあてはめます。

図 5.8 直線と多項式のあてはめ

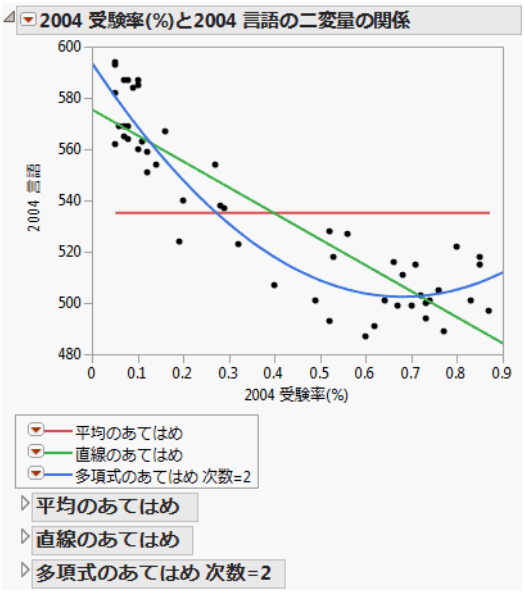


図 5.8 は、平均線、回帰直線、2 次多項式を比較したものです。

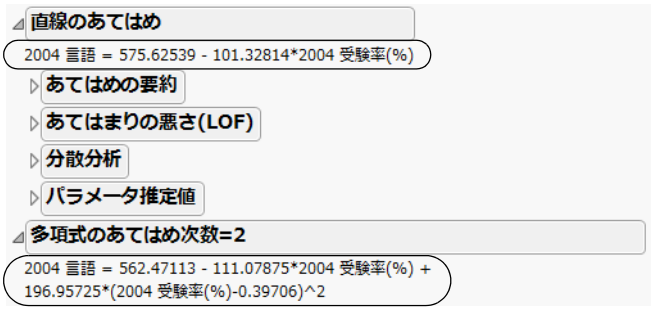
次の点に注目してください。

- **［直線のあてはめ］**は、1 次の多項式をあてはめるのと同じことです。
- **［平均のあてはめ］**は、次数が 0 の多項式をあてはめるのと同じことです。

「直線のあてはめ」と「多項式のあてはめ」のレポート

「直線のあてはめ」と「多項式のあてはめ」のレポート冒頭には、あてはめられた式が表示されます。

図 5.9 あてはめの式の例



注: 式はクリックして編集できます。

「直線のあてはめ」と「多項式のあてはめ 次数＝X」の各レポートには、少なくとも3つのレポートが表示されます。Xが同じ値を持つオブザベーションが複数ある場合は、4つ目の「あてはまりの悪さ (LOF)」レポートも表示されます。

「あてはめの要約」レポート

次図の「あてはめの要約」レポートは、同じデータに直線と2次多項式をあてはめたときの要約です。このレポートを比較すると、直線から2次多項式にすることで、どの程度、モデルのあてはまりが良くなるのかが分かります。**「R2 乗」**が大きくなるほど、**「誤差の標準偏差 (RMSE)」**の値が小さくなるほど、モデルのあてはまりは良くなっています。

図 5.10 直線と多項式の「あてはめの要約」レポート

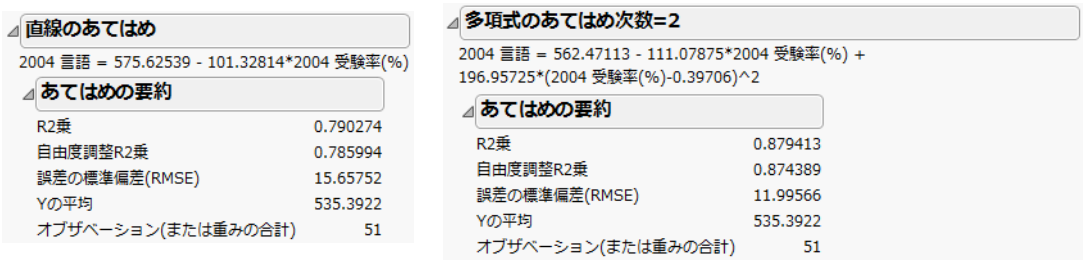


表 5.4「あてはめの要約」レポートの説明

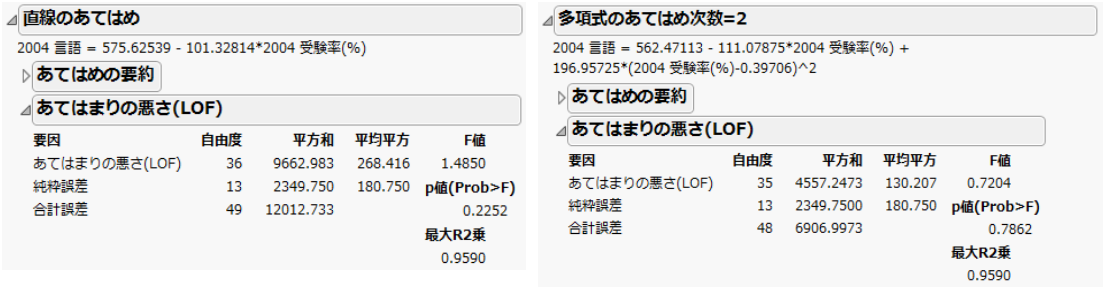
R2 乗	モデルによって説明できる変動の割合を表します。残りの変動は、モデルで説明できないため、ランダムな誤差とみなされます。モデルが完全にあてはまっていると、**「R2 乗」**は1になります。 図5.10の**「R2 乗」**の値から、2次多項式は、ほんの少しだけ直線よりもあてはまりが良いことがわかります。 「「あてはめの要約」レポートの統計的詳細」(133ページ)を参照してください。
自由度調整R2 乗	パラメータ数の異なるモデルでも比較できるように**「R2 乗」**に調整を加えたもので、計算の際に自由度が考慮されます。 詳細については、「「あてはめの要約」レポートの統計的詳細」(133ページ)を参照してください。
誤差の標準偏差 (RMSE)	ランダムな誤差の標準偏差の推定値です。「分散分析」レポートにある**「誤差」**の平均平方の平方根です。詳細については、図5.12を参照してください。
応答の平均	応答変数の標本平均（算術平均）。これは、モデルで効果が指定されていないときのYの予測値です。
オブザベーション	あてはめを推定するために使われたオブザベーションの数。ただし、重み変数が使われた場合は、重みの合計です。

「あてはまりの悪さ (LOF)」レポート

注:「あてはまりの悪さ (LOF)」レポートは、Xが同じ値を持つ行が複数ある場合にのみ表示されます。

Xが同じ値をもつオブザベーションが複数ある場合には、あてはめられたモデルが正しいかどうかに関係なく、誤差の大きさを推定できます。このような反復（Xが同じ値をもつオブザベーション）から計算された誤差を **純粋誤差**（pure error）と言います。純粋誤差は、データの誤差のうち、どのようなモデルを構築しても説明や予測ができない変動を表します。なお、自由度が非常に少ない（反復されたXの値が少ない）場合は、「あてはまりの悪さ（LOF）」検定はあまり役に立たない可能性があります。

図 5.11 直線および多項式のあてはめの「あてはまりの悪さ (LOF)」レポートの例



モデルの残差誤差から、純粋誤差を引いたものを、**あてはまりの悪さ**（LOF; Lack Of Fit）の誤差といいます。指定したモデルが不適切だと、LOF 誤差が純粋誤差よりも有意に大きくなります。その場合は、別のモデルをあてはめるようにしてください。「あてはまりの悪さ (LOF)」レポートには、LOF 誤差が 0 かどうかの検定結果も表示されます。

表 5.5「あてはめの要約」レポートの説明

要因	変動の 3 つの要因（**「あてはまりの悪さ (LOF)」**、**「純粋誤差」**、**「合計誤差」**）。
自由度	誤差の各要因の**自由度**（DF）。 • **「合計誤差」**の自由度は、「分散分析」レポートの**「誤差」**の行に表示されている自由度です（「「分散分析」レポート」（107 ページ）を参照）。この自由度は、**データ全体**の自由度から、**モデル**の自由度を引いたものです。**「誤差」**の自由度は、あてはまりの悪さ（LOF）と純粋誤差の 2 つの自由度に分けることができます。 • **「純粋誤差」**の自由度は、X値が同じである行をグループにまとめ、それぞれのグループの自由度を合算したものです。詳細については、「「あてはまりの悪さ (LOF)」レポートの統計的詳細」（133 ページ）を参照してください。 • **「あてはまりの悪さ (LOF)」**の自由度は、**「合計誤差」**と**「純粋誤差」**の自由度の差です。

表 5.5「あてはめの要約」レポートの説明（**続き**）

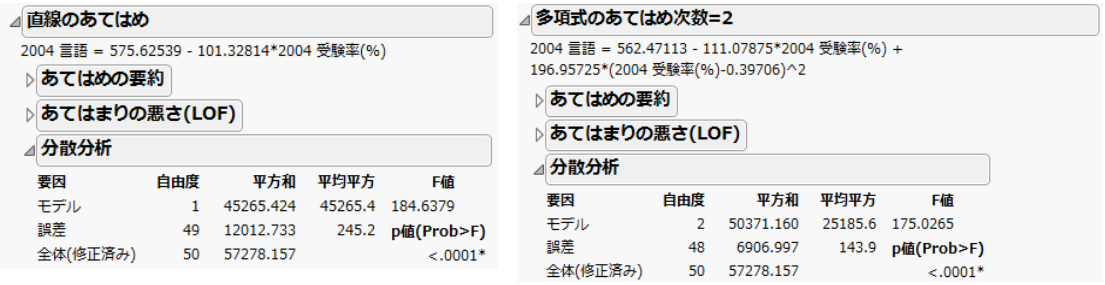
平方和	誤差の各要因の平方和（SS）。 • **「合計誤差」**の平方和は、「分散分析」レポートの**「誤差」**の行に表示されている平方和です（「「分散分析」レポート」（107 ページ）を参照）。 • **「純粋誤差」**の平方和は、*X* 値が同じである行をグループにまとめ、それぞれのグループの平方和を合算したものです。この平方和は、モデルの *X* 効果ではまったく説明できない、純粋にランダムな誤差の大きさを推定したものです。詳細については、「「あてはまりの悪さ (LOF)」レポートの統計的詳細」（133 ページ）を参照してください。 • **「あてはまりの悪さ (LOF)」**の平方和は、**「合計誤差」**と**「純粋誤差」**の平方和の差です。LOF の平方和が大きいときは、モデルがデータに良くあてはまっていない可能性があります。この後で説明する「*F* 値」によって、あてはまりの悪さによって生じる変動が十分に小さいかどうか、つまり、純粋誤差に対して無視できる大きさであるかが検定されます。
平均平方	平方和を関連する自由度で割った値。この計算により、平方和が平均（平均平方）に換算されます。統計的検定で使用する *F* 値は平均平方の比です。
F 値	「あてはまりの悪さ (LOF)」の平均平方の、「純粋誤差」の平均平方に対する比。「LOF 誤差は 0 である」という帰無仮説を検定します。
p 値（Prob > F）	LOF に起因する変動がない場合に、まったくの偶然だけで、現在の **F** 値よりも大きな **F** 値を得る確率。**p** 値が大きい場合は、あてはまりの悪さの誤差は有意ではありません。
最大 R2 乗	モデルに含まれている変数で達成できる最大の R^2 乗。 「「あてはまりの悪さ (LOF)」レポートの統計的詳細」（133 ページ）を参照してください。

「分散分析」レポート

回帰モデルに対する分散分析表では、標本全体の変動が、いくつかの成分に分割されます。これらの成分は、モデルの有効性を評価する「*F* 値」の計算に使用されます。「*F* 値」に関連する確率（p 値）が小さいとき、「そのモデルは、Y の平均だけのモデルよりも、良くあてはまっている」と見なすことができます。

図 5.12 は、線形のあてはめ（[**直線のあてはめ**]）と 2 次のあてはめ（[**多項式のあてはめ**]）の「分散分析」レポートです。どちらも平均だけのモデルよりも統計的にあてはまりが良いことがわかります。

図 5.12 直線および多項式のあてはめの「分散分析」レポートの例



直線のあてはめ

2004 言語 = 575.62539 - 101.32814*2004 受験率(%)

あてはめの要約

あてはまりの悪さ(LOF)

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	45265.424	45265.4	184.6379
誤差	49	12012.733	245.2	p値(Prob>F)
全体(修正済み)	50	57278.157		<.0001*

多項式のあてはめ次数=2

2004 言語 = 562.47113 - 111.07875*2004 受験率(%) +
196.95725*(2004 受験率(%)-0.39706)^2

あてはめの要約

あてはまりの悪さ(LOF)

分散分析

要因	自由度	平方和	平均平方	F値
モデル	2	50371.160	25185.6	175.0265
誤差	48	6906.997	143.9	p値(Prob>F)
全体(修正済み)	50	57278.157		<.0001*

表 5.6「分散分析」レポートの説明

要因	変動の 3 つの要因（**「モデル」**、**「誤差」**、 **「全体 (修正済み)」**）。
自由度	変動の各要因の自由度（DF）。 • 自由度は、非欠測値の標本サイズ（*N*）から、使用したパラメータ数を引いて求めます。標本全体の変動においては、全体平均を表す 1 つのパラメータだけが使用されるので、標本サイズから自由度が 1 つ引かれます。例では、49 となっています。「全体 (修正済み)」の自由度は、**「モデル」**項と**「誤差」**項に分かれます。 • 「直線のあてはめ」では、切片のパラメータに、傾きのパラメータを 1 つ追加したモデルが推定されます。分散分析表における **「モデル」**の自由度は、1 です。また、2 次の「多項式のあてはめ」では、切片のパラメータに、2 つのパラメータ（β_1 と β_2）を追加したモデルが推定されます。分散分析表における「モデル」の自由度は 2 となります。 • **「誤差」**の自由度は、**「全体 (修正済み)」**と**「モデル」**の自由度の差です。
平方和	変動の各要因の平方和（SS）。 • 図 5.12 の例では、各応答から標本平均までの距離の平方和（**「全体 (修正済み)」**）は 57,2758.157 です。これは、基本モデル（単純な平均モデル）の平方和で、その他のモデルとの比較に使用されます。 • 線形回帰においては、各点から直線までの距離の平方和は 12,012.733 に減少します。これが、線形モデルでは説明できない**「誤差」**（残差）の平方和です。2 次多項式をあてはめると、残差の平方和は 6,906.997 になり、直線のときよりわずかに多く変動が説明されていることになります。言い換えると、2 次多項式の方が、直線よりもモデル平方和が大きくなっており、より多くの変動を説明しています。**「全体 (修正済み)」**の平方和から**「誤差」**の平方和を引くと、モデルで説明される平方和が求められます。

表 5.6「分散分析」レポートの説明（**続き**）

平均平方	平方和を関連する自由度で割った値。統計的検定で使用される「*F* 値」は、次に示す 2 つの平均平方の比です。 • 「直線のあてはめ」の**「モデル」**平均平方は 45,265.4 です。この値は、モデルのパラメータが 0 であるという仮説のもとでの誤差分散の推定値です。 • **「誤差」**平均平方である 245.2 は、誤差分散の推定値です。
F 値	モデルの平均平方を誤差の平均平方（MSE）で割ったもの。これは、（切片を除く）すべての回帰パラメータが 0 であるという仮説を検定します。この仮説が真のとき、「誤差」と「モデル」の両方の平均平方が誤差分散の推定値となり、その比は *F* 分布に従います。0 でないパラメータがある場合、「*F* 値」は通常、偶然だけから期待される値よりも大きくなります。
p 値（Prob > F）	真のモデルが、平均だけのモデルであった場合に、まったくの偶然だけで、現在より大きな *F* 値が得られる確率（*p* 値）を表します。有意確率が 0.05 以下だと、効果がある証拠だと考えられています。

「パラメータ推定値」レポート

「直線のあてはめ」の「パラメータ推定値」レポートには、項として切片と *X* 変数が 1 つ表示されています。

k 次多項式のあてはめでは、切片の推定値と、*X* 変数の *k* 個のべき乗それぞれに 1 つの推定値があります。

図 5.13 直線および多項式のあてはめの「パラメータ推定値」レポートの例

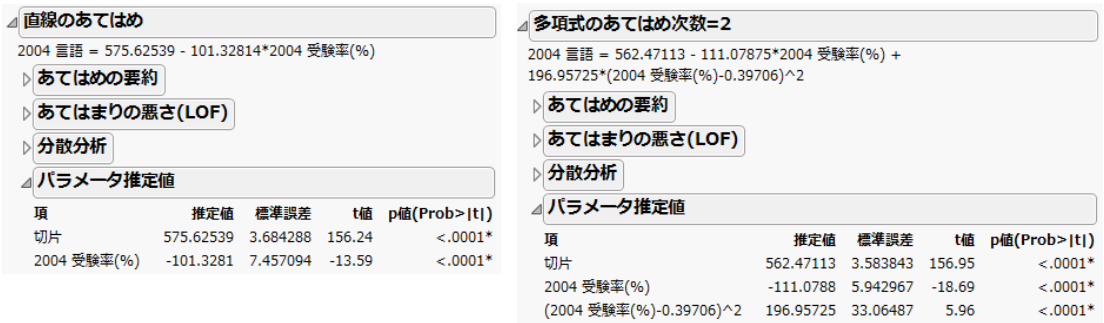


表 5.7「パラメータ推定値」レポートの説明

項	モデルの各パラメータの名前。定数項である切片は、どのモデルにも含まれます。
推定値	線形モデルのパラメータ推定値。回帰モデルの予測式は、これらの推定値と、対応する変数との線形結合で表されます。
標準誤差	パラメータ推定値の標準誤差を推定した値。この標準誤差を使って、検定と信頼区間が計算されます。

表 5.7「パラメータ推定値」レポートの説明（続き）

t 値	各パラメータが 0 であるという仮説を検定する統計量。t 値は、パラメータ推定値とその標準誤差の比です。仮説が真のとき、*t* 値は Student の *t* 分布に従います。
p 値 (Prob>|t|)	それぞれの *t* 値から計算した有意確率。これは、帰無仮説が真のときに、まったくの偶然だけで、計算された t 値より（絶対値が）大きい *t* 値が得られる確率を示します。値が 0.05（場合によっては 0.01）より小さいときは、パラメータがゼロとは有意に異なる証拠と解釈されます。

追加の統計量を表示するには、レポートを右クリックし、[列] の各オプションを選択します。次の統計量はデフォルトでは表示されません。

下側 95%　パラメータ推定値の下側 95% 信頼限界。

上側 95%　パラメータ推定値の上側 95% 信頼限界。

標準 β　標準化されたパラメータ推定値。異なるスケールで測定した X 変数の効果を比較する場合に便利です。詳細については、「「パラメータ推定値」レポートの統計的詳細」（134 ページ）を参照してください。

VIF　分散拡大係数（Variance Inflation Factor）。X 変数の相関行列の逆行列における対角要素。

計画の標準誤差　計画行列に基づく、パラメータ推定値の標準誤差。「「パラメータ推定値」レポートの統計的詳細」（134 ページ）を参照してください。

関連情報

- 「[直線のあてはめ] の統計的詳細」（132 ページ）
- 「あてはめ用メニュー」（119 ページ）

その他のあてはめ

[**その他のあてはめ**] コマンドは、Y や X を変換した後に、直線や曲線をあてはめます。変換には、自然対数、平方根、2 乗、逆数、指数があります。切片と傾きの制約を指定したり、指定した次数の多項式をあてはめたり、多項式の中心化をオフすることも可能です。

表 5.8「変換または制約を指定」ウィンドウの説明

Y の変換	ここで選択したオプションに従って Y 変数が変換されます。
X の変換	ここで選択したオプションに従って X 変数が変換されます。
次数	指定した次数の多項式をあてはめます。

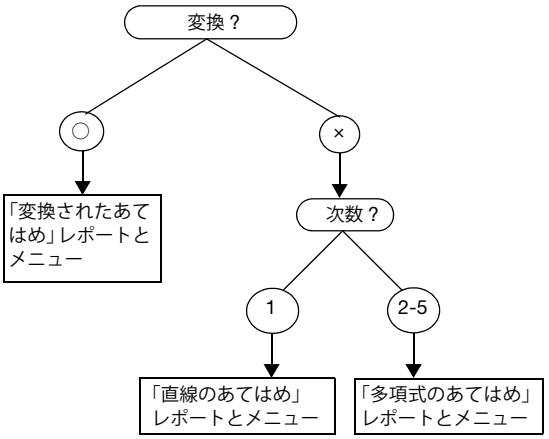
表 5.8「変換または制約を指定」ウィンドウの説明（続き）

多項式の中心化	多項式の中心化をオフにするには、[**多項式の中心化**] のチェックを外します。詳細については、図 5.19 を参照してください。なお、X 変数を変換した場合、多項式の中心化は行われません。多項式の中心化オプションを用いると、多重共線性の影響が軽減され、回帰係数の数値計算が安定します。
切片の制約	このチェックボックスをオンにすると、指定された数値にモデルの切片が固定されます。
傾きの制約	このチェックボックスをオンにすると、指定された数値にモデルの傾きが固定されます。

その他のあてはめのレポートとメニュー

「変換または制約を指定」ウィンドウで選択した値によって、表示されるレポートやメニューは異なります。図 5.14 では、選択内容によってどのレポートとメニューが表示されるかをフローチャートで示しています。

図 5.14 その他のあてはめのフローチャート



「変換されたあてはめ」レポート

「変換されたあてはめ」レポートには、「「直線のあてはめ」と「多項式のあてはめ」のレポート」（104 ページ）で説明されているレポートが表示されます。

また、Y を変換した場合は、このほかにも、「元のスケールで測定したあてはめ」レポートも表示されます。このレポートに表示される統計量は、Y 変数に関して、変換前のスケールで計算されています。

関連情報

- 「「その他のあてはめ」コマンドの例」（124 ページ）
- 「「直線のあてはめ」と「多項式のあてはめ」のレポート」（104 ページ）
- 「あてはめ用メニュー」（119 ページ）

ノンパラメトリックな曲線

［ノンパラメトリックな曲線］メニューでは、滑らかな曲線をあてはめるためのオプションが用意されています。

- ［スプライン曲線のあてはめ］は、罰則を課した最小2乗法によって、曲線をあてはめます。パラメータλによって、滑らかさの度合を調整してください。
- ［カーネル平滑化］は、局所的な重み付けにより、曲線をあてはめます。パラメータ α によって、局所的な重み付けの度合いを調整してください。
- ［値ごとのあてはめ］は、各X値での応答平均を計算します。

スプライン曲線のあてはめ

［**スプライン曲線のあてはめ**］コマンドは、平滑化スプライン曲線をあてはめます。曲線の滑らかさ（柔軟性）は、ラムダ（λ）値によって変わります。λは、スプライン曲線を調整するパラメータです。λの値を小さくすると、目的関数における誤差平方和の比重が大きくなり、データに曲線をあてはめようとする傾向が強くなり、曲線のカーブが多くなります。逆に、λの値を大きくすると、あてはめが硬くなり、スプライン曲線は直線に近付きます。

次の点に注目してください。

- 平滑化スプライン曲線は、Xに対するYの期待値を調べるのに役立ちます。
- 曲線の各部分に近い点ほど、その各部分の形状に大きな影響を及ぼしています。λ の値を小さくすると、局所的に点が曲線に与える影響が大きくなり、曲線が柔軟になります。
- メニューに表示されていないラムダ値を使用する場合は、［**スプライン曲線のあてはめ**］>［**その他**］を選択します。X変数のスケールが変わると、あてはめたモデルも変わります。これを避けるためには、「**Xの標準化**」オプションを選択します。そうすれば、X変数のスケールや単位をどのように変更しても、あてはめたモデルの結果はどれも同じになります。
- 異なるλ値を試してみてください。「平滑化スプライン曲線によるあてはめ」レポートの下にある「**λ の変更**」スライダを動かしてλの値を変更できます。ただし、λはデータのスケールによって変化するので注意が必要です。たとえば、X軸がインチ単位の場合とセンチメートル単位の場合では、λの値が異なります。

「平滑化スプライン曲線によるあてはめ」レポート

「平滑化スプライン曲線によるあてはめ」レポートには、「**R2 乗**」と「**誤差平方和 (SSE)**」が表示されます。これらの値は、スプライン曲線をその他のモデルや別のスプライン曲線と比較するのに役立ちます。

表 5.9「平滑化スプライン曲線によるあてはめ」レポートの説明

R2 乗	データの全変動のうち、平滑化スプライン曲線モデルによって説明されている割合。詳細は、「「平滑化スプライン曲線によるあてはめ」レポートの統計的詳細」（134 ページ）を参照してください。
誤差平方和（SSE）	各点からスプライン曲線までの距離の平方和。これは、スプラインモデルをあてはめた後もまだ説明できていない「誤差」（**残差**）を表します。
λの変更	λ値を変更できます。値を入力するか、スライダを動かして変更します。

関連情報

- 「あてはめ用メニュー」（119 ページ）
- 「［スプライン曲線のあてはめ］の統計的詳細」（132 ページ）

カーネル平滑化

［**カーネル平滑化**］コマンドは、幾つかの抽出された点に関して、局所的に重み付けて直線や 2 次曲線といった単純なモデルを次々にあてはめていくことにより、全体の曲線を求めます。最大で合計 128 個の局所的な回帰モデルが組み合わされ、データ全体にわたって滑らかな曲線が作成されます。この局所的に重み付けられた回帰は、**loess**、もしくは、提案された当初の名前である **lowess** とも呼ばれています。Lowess という呼び名は、locally weighted scatterplot smoother (局所重み付き散布図平滑化）が由来です。Cleveland（1979）を参照してください。

この手法は、変数間の関係を特定して、その後に必要な分析やモデル式を考察するのに役立ちます。

「局所的な平滑化」レポート

「局所的な平滑化」レポートには、「**R2 乗**」と「**誤差平方和 (SSE)**」が表示されます。これらの値は、平滑化曲線をその他のモデルや別の平滑化曲線と比較するのに役立ちます。

表 5.10「局所的な平滑化」レポートの説明

R2 乗	データの全変動のうち、平滑化モデルによって説明されている割合。詳細は、「「平滑化スプライン曲線によるあてはめ」レポートの統計的詳細」（134 ページ）を参照してください。
誤差平方和（SSE）	各点からあてはめた平滑化曲線までの距離の平方和。これは、平滑化モデルをあてはめた後もまだ説明できていない「誤差」（**残差**）を表します。
局所的なあてはめ（λ）	個々の局所的なあてはめに対し、多項式の次数を選択します。2 次多項式を使用すると、局所的なでこぼこが滑らかになります。λは、この手法であてはめる多項式の次数を表します。λには、0 次（平均）、1 次（線形）、または 2 次を指定できます。

表 5.10「局所的な平滑化」レポートの説明（**続き**）

重み関数	それぞれの局所的な回帰において、近傍のデータをどのように重み付けするかを指定します。デフォルトでは、重み付けの関数として、トリキューブ（3次式の3乗）が使われます。このような重み付け関数は、局所的な回帰をあてはめていくときに、各点（*xi, yi*）が局所的な回帰に与える影響を決めます。重みは、*x* から *xi* までの距離が大きくなるにつれて小さくなり、やがてゼロになります。
滑らかさ（α）	個々の局所的なあてはめを構成する点の数を指定します。スライダーで値を調整するか、値を直接入力します。α は、平滑化パラメータで、正の数値なら何でもかまいませんが、通常は 0.25 ～ 1 の値を使用します。α を大きくすると、曲線の滑らかさが増します。
ロバスト性	モデルから離れた点に小さな重みを与えることにより、外れ値による影響を小さくします。ここには、重みを再計算する計算過程の反復回数（パスの回数）を指定します。この反復計算の目的は、外れ値に対する重みを小さくすることで、自動的に外れ値をなくすことです。

関連情報

- 「あてはめ用メニュー」（119 ページ）

値ごとのあてはめ

［**値ごとのあてはめ**］コマンドは、X 変数の一意の値ごとに値をあてはめます。X 変数の一意の値ごとに応答の平均があてはめられます。

「値ごとのあてはめ」レポート

「値ごとのあてはめ」レポートには、モデルのあてはめに関する要約統計量が表示されます。

表 5.11「値ごとのあてはめ」レポートの説明

オブザベーションの数	オブザベーションの合計数。
異なる X の値の個数	X の異なる値の個数。
自由度	純粋誤差の自由度。
平方和	純粋誤差の平方和。
平均平方	純粋誤差の平均平方。

関連情報

- 「あてはめ用メニュー」（119 ページ）

直交のあてはめ

［**直交のあてはめ**］コマンドでは、*X* と *Y* 両方の変動を考慮して直線があてはめられます。

「直交のあてはめ］のオプション

次の表では、分散比を指定するオプションについて説明します。

一変量分散 , 主成分	*X* と *Y* の一変数ずつから求めた分散の推定値を使います。この結果は、標準化したデータに対する主成分分析の第 1 主成分です。このオプションは、測定システムの分析には向いていません。得られたデータにおける分散の比は、母集団における分散の比と異なる場合が多いからです。
等分散性	誤差分散が等しいと仮定し、分散比を 1 とします。この結果は、標準化していないデータに対する主成分分析の第 1 主成分です。散布図において、*X* 軸と *Y* 軸のスケールを同じにし、正規密度楕円を表示すると、得られた線が楕円の長軸になります。
Y→X のあてはめ	*Y* の分散が 0 であるとし、分散比を 0 とします。
指定された分散比	このオプションでは、任意の分散比を入力できます。*X* と *Y* の測定誤差の大きさが事前にわかっている場合に活用できます。

「直交回帰」レポート

「直交回帰」レポートは、直交回帰モデルの要約統計量をまとめたものです。

次の表では、「直交回帰」レポートについて説明します。

表 5.12「直交回帰」レポートの説明

変数	直線のあてはめに使用された変数の名前。
平均	各変数の平均。
標準偏差	各変数の標準偏差。
分散比	直線のあてはめに使用された分散比。
相関	2 つの変数間の相関。
切片	あてはめた直線の切片。
傾き	あてはめた直線の傾き。
下側信頼限界	傾きの下側信頼限界。
上側信頼限界	傾きの上側信頼限界。
Alpha	信頼区間の計算に使用する α 水準を入力します。

関連情報

- 「あてはめ用メニュー」（119ページ）
- 「[直交のあてはめ] の統計的詳細」（132ページ）
- 「[直交のあてはめ] コマンドの使用例」（127ページ）

ロバスト

[**ロバスト**] オプションには、データセットの外れ値の影響を受けにくい2通りの手法があります。データに外れ値がある場合、ロバストではない方法を用いると、推定や決定が間違ったものになってしまう可能性があります。

ロバストなあてはめ

[ロバストなあてはめ] オプションで用意されている手法は、応答変数の外れ値が推定結果に与える影響が少ないものです。HuberのM推定が使用されます。HuberのM推定では、Huberの損失関数を最小にするパラメータ推定値を見つけます。この推定方法では、外れ値に対して、ある種の罰則（ペナルティ）が課せられます。Huberの損失関数は、誤差が小さい場合は2次曲線として増加し、誤差が大きい場合は線形的に増加します。ロバストなあてはめの詳細については、Huber（1973）およびHuber and Ronchetti（2009）を参照してください。

関連情報

- 「あてはめ用メニュー」（119ページ）
- 「[ロバストなあてはめ] コマンドの使用例」（128ページ）

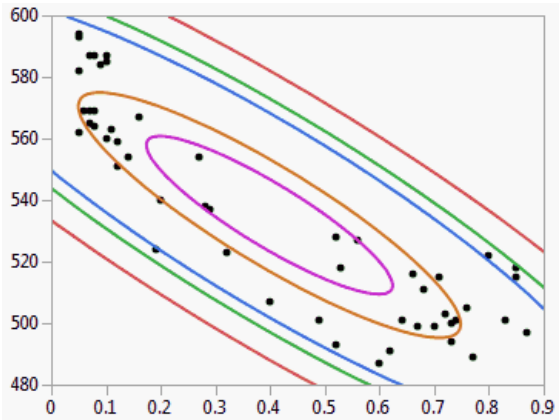
Cauchyのあてはめ

このオプションでは、誤差がCauchy分布に従うと仮定されます。Cauchy分布は正規分布よりも裾が広く、その結果、外れ値が推定に与える影響が小さくなります。このオプションは、データにある外れ値の割合が大きい場合に有用です。しかし、データが正規分布に近く、外れ値が少ない場合は、このオプションの推定結果は間違ったものになる可能性があります。[Cauchyのあてはめ] オプションは、最尤推定によってパラメータ推定値を算出します。

確率楕円

[**確率楕円**] オプションは、サブメニューで選択された確率の確率楕円を描きます。楕円内に位置する確率が、選択された確率になっている楕円が描かれます。

図 5.15 確率楕円の例



確率楕円は、X と Y から推定された二変量正規分布から計算されます。二変量正規分布は、X 変数と Y 変数の平均と標準偏差、および、その相関によって決められます。なお、［**その他**］オプションを選択すると、0 より大きく 1 未満の、任意の確率を指定することができます。

確率楕円は、確率密度が等しいところを示す等高線です。また、信頼領域を示す等高線でもあります。二変量正規分布を仮定した上で、指定された割合だけ点が含まれる範囲を表しています。

確率楕円は、2 変数間の相関を調べるのに役立ちます。2 変数間の相関が 1 または -1 へ近付くにつれて、楕円は対角線方向に長くなります。2 変数に相関がない場合、楕円は（対角線方向に伸びず）水平もしくは垂直に伸びます。

「相関」レポート

確率楕円ごとに作成される「相関」レポートには、X 変数と Y 変数の相関係数が表示されます。

注 : 複数の変数に対して、確率楕円と相関を行列形式で表示させたい場合には、メニューバーから［**分析**］> ［**多変量**］> ［**多変量の相関**］を選択してください。

表 5.13「相関」レポートの説明

変数	楕円の作成に使用された変数の名前。
平均	X 変数と Y 変数のそれぞれの平均。
標準偏差	X 変数と Y 変数のぞれぞれの標準偏差。 平均と標準偏差の詳細については、「一変量の分布」の章の「「要約統計量」レポート」（42 ページ）を参照してください。

表 5.13「相関」レポートの説明（**続き**）

相関	Pearson の相関係数。二変量間に完全な線形関係がある場合、相関係数は 1（正の相関）または -1（負の相関）になり、関係が存在しない場合、相関は 0 に近くなります。 詳細は、「「相関」レポートの統計的詳細」（135 ページ）を参照してください。
p 値	X 変数と Y 変数の間に相関関係がないときに、まったくの偶然だけで、計算された相関よりも絶対値の大きい相関が得られる確率。
数	計算に使用されたオブザベーションの数。

関連情報

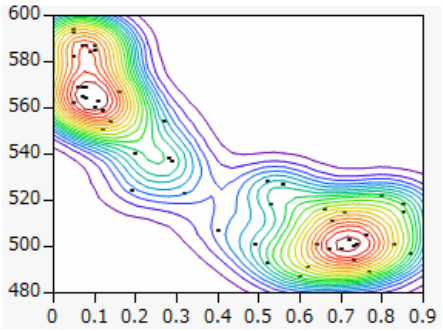
- 「あてはめ用メニュー」（119 ページ）
- 「グループ別確率楕円の例」（130 ページ）

ノンパラメトリック密度

プロットに数千にのぼる点が密集している場合、点が塊になって表示され、密度のパターンが見えなくなってしまいます。［ノンパラメトリック密度］コマンドを使用すると、密度のパターンがわかりやすくなります。

このオプションでは、各座標におけるノンパラメトリックな密度推定値が、滑らかな曲面で表されます。レポートの散布図上には、密度が等しくなっている等高線が表示されます（図 5.16）。等高線は 5% 間隔で表示されます。最も低い等高線より下に位置する点は全体の 5%、その次の等高線より下に位置するのは 10%、というように続いていきます。最も高い等高線より下には約 95% の点が位置します。

図 5.16 ノンパラメトリック密度の例



ノンパラメトリック密度等高線を描くグリッド点の個数を変更して、等高線をより滑らかに描画させることができます。デフォルト値は 51 個ですが、場合によっては、高密度な部分の等高線がぎざぎざになる場合があります。

個数を変更するには、Shift キーを押しながら「二変量の関係」の赤い三角ボタンをクリックし、メニューから［**ノンパラメトリック密度**］を選択してください。デフォルトの 51 よりも大きい値を入力できます。

「ノンパラメトリック二変量の密度」レポート

「ノンパラメトリック二変量の密度」レポートには、ノンパラメトリック密度の計算に使用された標準偏差が表示されます。

関連情報

- 「あてはめ用メニュー」（119ページ）

グループ別

［**グループ別**］オプションは、分類変数 (グループ変数) を指定するためのものです。「二変量」プラットフォームでは、グループ変数を指定すると、グループ変数の水準ごとに分析が行われ、回帰直線や楕円が散布図に重ねて描かれます。グループ変数のどの水準の結果なのかは、散布図の下の凡例を見ると分かります。この凡例の1つ1つにポップアップメニューがあり、情報を保存したり、削除したりできます。

グループ変数が指定されている場合には、［**グループ別**］オプションにチェックマークが付きます。グループ変数を変更するには、まず、［**グループ別**］オプションを選択してチェックマークを外します。これで、既存のグループ変数が削除されたことになります。次に、［**グループ別**］オプションを再び選択し、前回同様にグループ変数を選択します。

［グループ別］オプションは、次のような用途で使用できます。

- グラフに複数の回帰直線を重ね合わせると、傾きが簡単に比較できます。
- 複数の確率楕円を重ね合わせると、グループ変数の水準ごとに点がクラスター（塊）に分かれて表示されます。

関連情報

- 「グループ別確率楕円の例」（130ページ）
- 「グループ別回帰直線の例」（131ページ）

あてはめ用メニュー

各種のあてはめを実行するコマンドを選択すると、レポートが表示されるとともに、あてはめ用のメニューも新たに追加されます。次の表は、あてはめ用コマンドとそれに対応するあてはめ用メニューをまとめたものです。

あてはめ用コマンド	あてはめ用メニュー
平均のあてはめ	平均のあてはめ
直線のあてはめ	直線のあてはめ

あてはめ用コマンド	あてはめ用メニュー
多項式のあてはめ	多項式のあてはめ 次数=X*
その他のあてはめ	直線のあてはめ 多項式のあてはめ 次数=X* 変換されたあてはめ X* 制約のあるあてはめ
スプライン曲線のあてはめ	平滑化スプライン曲線によるあてはめ、λ=X*
平滑線	局所的な平滑化
値ごとのあてはめ	値ごとのあてはめ
直交のあてはめ	直交回帰 分散比=X*
ロバストなあてはめ	ロバストなあてはめ
Cauchy のあてはめ	Cauchy のあてはめ
確率楕円	二変量正規楕円 P=X*
ノンパラメトリック密度	分位点密度等高線

*X は、何らかの文字または数字です。

あてはめ用メニューのオプション

選択したオプションのあてはめ用メニューには、該当するあてはめに適用できるオプションがあります。

- 表5.14は、ほとんどのあてはめに表示されるオプションを示しています。
- 表5.15は、複数のあてはめに表示されるオプションを示しています。
- 表5.16は、「二変量正規楕円」のみに表示されるオプションを示しています。
- 表5.17は、「分位点密度等高線」メニューのみに表示されるオプションを示しています。

表5.14 ほとんどのあてはめに適用できるオプション

オプション	説明
あてはめ線	モデルのあてはめを説明する直線または曲線の表示／非表示を切り替えます。「二変量正規楕円」レポートの場合、このオプションは等高線の境界を示す楕円の表示／非表示を切り替えます。このオプションは、「分位点密度等高線」にはありません。
線の色	パレットが開き、各あてはめ線に割り当てる色を選択できます。このオプションは、「分位点密度等高線」にはありません。

表5.14 ほとんどのあてはめに適用できるオプション（**続き**）

オプション	説明
線種	パレットが開き、各あてはめ線に割り当てる線種を選択できます。このオプションは、「分位点密度等高線」にはありません。
線の幅	あてはめ線の幅を3つの中から選択します。デフォルトでは最も細い線が使われています。このオプションは、「分位点密度等高線」にはありません。
レポート	あてはめのレポートの表示／非表示を切り替えます。二変量のプロットは変更されません。
あてはめの削除	グラフとレポートからあてはめ線を削除します。

表5.15 複数のあてはめに適用できるオプション

オプション	説明
回帰の信頼区間	期待値（平均）に対する信頼限界の表示／非表示を切り替えます。このオプションは、スプライン曲線のあてはめ、確率楕円、値ごとのあてはめ、および直交のあてはめではグレー表示になっていて使用できません。
個別の値に対する信頼区間	個別の予測値に対する信頼限界の表示／非表示を切り替えます。信頼区間は、誤差における変動とパラメータ推定値における変動の両方を考慮しています。このオプションは、平均のあてはめ、スプライン曲線のあてはめ、確率楕円、値ごとのあてはめ、直交のあてはめではグレー表示になっていて使用できません。
予測値の保存	データテーブル内に**「予測値 列名」**という新しい列を作成します。**「列名」**の部分にはY変数の名前が挿入されます。この列には、計算式の形式で予測式が保存されます。データテーブルに行を追加すると、保存された予測式によって、その行の予測値が自動的に計算されます。このオプションは、値ごとのあてはめ、確率楕円のあてはめではグレー表示になっていて使用できません。 **注：**［**予測値の保存**］と［**残差の保存**］のコマンドは、あてはめごとに適用できます。これらのコマンドを何度も実行した場合や、グループ変数と併用した場合は、データテーブル内に作成された列がどのあてはめのものかわかるように列名を変更してください。

表 5.15 複数のあてはめに適用できるオプション（**続き**）

オプション	説明
残差の保存	**注：**データテーブル内に**「残差 列名」**という新しい列を作成します。**「列名」**の部分には *Y* 変数の名前が挿入されます。この列には、各行の実測値（観測値）と予測値の差が保存されます。**［予測値の保存］**コマンドとは異なり、このコマンドでは新しい列に計算式は保存されません。このオプションは、値ごとのあてはめ、確率楕円のあてはめではグレー表示になっていて使用できません。 **注：［予測値の保存］**と**［残差の保存］**のコマンドは、あてはめごとに適用できます。これらのコマンドを何度も実行した場合や、グループ変数と併用した場合は、データテーブル内に作成された列がどのあてはめのものかわかるように列名を変更してください。
平均の信頼限界の計算式	平均の信頼区間の計算式が保存されます。
個別の信頼限界の計算式	個々の応答値に対する信頼区間の計算式が保存されます。
残差プロット（「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」のみ）	「予測値と残差のプロット」、「予測値と実測値のプロット」、「行番号と残差のプロット」、「残差の正規分位点プロット」の 4 種類の診断プロットを表示します。詳細については、「診断プロット」（124 ページ）を参照してください。
α 水準の設定	さまざまなあてはめの信頼区間を計算するための α 水準を指定します。
回帰の信頼区間内を塗る	**［回帰の信頼区間］**コマンドと同じ曲線を描き、曲線間に陰影を付けます。
個別の値に対する信頼区間内を塗る	**［個別の値に対する信頼区間］**コマンドと同じ曲線を描き、曲線間に陰影を付けます。
係数の保存	新しいデータテーブルが作成され、スプライン曲線の係数が、**X**、**A**、**B**、**C**、**D** という列に保存されます。**X** 列の値は節点を指します。**A**、**B**、**C**、**D** はそれぞれ、現在の **X** 値から次に大きな **X** 値までの区間における 3 次スプライン曲線の切片、1 次係数、2 次係数、3 次係数です。

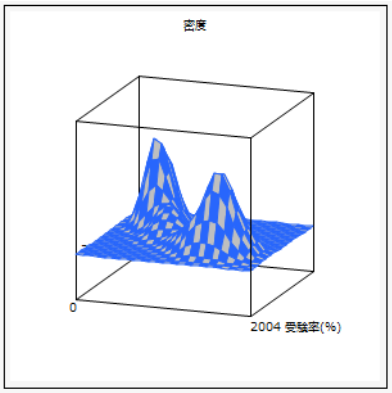
表 5.16「二変量正規楕円」のみに適用できるオプション

オプション	説明
楕円内を塗る	確率楕円の内側の領域を塗ります。
内側の点を選択	確率楕円の内側の点を選択します。
外側の点を選択	確率楕円の外側の点を選択します。

表 5.17「分位点密度等高線」のみに適用できるオプション

オプション	説明
カーネルの制御	標準偏差を変更するためのスライダを各変数に表示します。ノンパラメトリック密度を計算するための X と Y の標準偏差を変更できます。
5% の等高線	5% の等高線の表示／非表示を切り替えます。
等高線	10% の等高線の表示／非表示を切り替えます。
等高線間を塗りつぶす	等高線と等高線の間の領域を塗りつぶします。
密度による点の選択	指定された分位点範囲内の点を選択します。
密度分位点による色分け	密度分位点に応じて点を色分けします。
密度分位点の保存	新しい列を作成し、各点の密度分位点を保存します。
メッシュプロット	分析対象の2変数をグリッドにして、密度推定値を3次元でプロットします。詳細については、図 5.17 を参照してください。
峰型クラスター	データテーブル内に新しい列を作成し、クラスター値を保存します。 **注：**峰型クラスター値を先に保存してから密度グリッドを保存すると、グリッドテーブルにもクラスター値が含まれます。クラスター値は、プロットの点に色やマークを付けるときに役立ちます。
密度グリッドの保存	密度推定値とその分位点を新しいデータテーブルに保存します。「三次元散布図」や「等高線図」プラットフォームなどで密度をグラフ化したいときは、このグリッドデータを用いてください。

図 5.17 メッシュプロットの例



診断プロット

［**残差プロット**］オプションを選択すると、残差プロットを始め、モデルの適合度を診断するプロットが表示されます。プロットの種類は次のとおりです。

予測値と残差のプロット　予測値と残差を プロットしたものです。残差のヒストグラムも作成されます。

予測値と実測値のプロット　予測値と実測値を プロットしたものです。

行番号と残差のプロット　行番号と残差を プロットしたものです。

Xと残差のプロット　X変数と残差を プロットしたものです。

残差の正規分位点プロット　残差の正規分位点プロットです。

「二変量」プラットフォームのその他の例

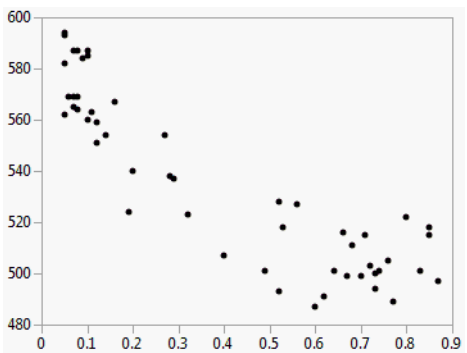
ここでは、「二変量」プラットフォームを使った例をさらに紹介します。

「その他のあてはめ」コマンドの例

Yを対数変換、Xを平方根変換するには、次の手順に従います。

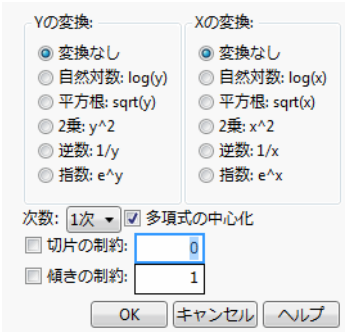
1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「SAT.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**2004 言語**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**2004 受験率 (%)**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。

図5.18 受験率に対するテストスコアの例



6. 「二変量の関係」の赤い三角ボタンをクリックし、メニューから［**その他のあてはめ**］を選択します。「変換または制約を指定」ウィンドウが表示されます。このウィンドウの説明については、表5.8を参照してください。

図5.19「変換または制約を指定」ウィンドウ



7. 「Yの変換」で、［**自然対数 : log(y)**］を選択します。
8. 「Xの変換」で、［**平方根 : sqrt(x)**］を選択します。
9. ［**OK**］をクリックします。

図 5.20 その他のあてはめのレポートの例

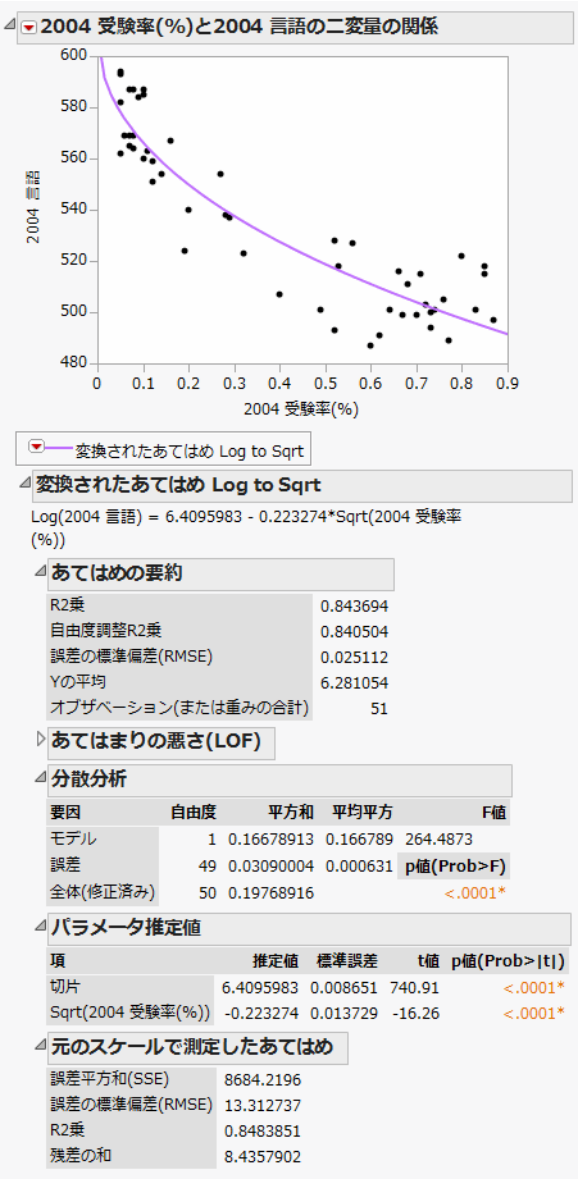


図 5.20 に示すように、あてはめられた線は元のスケールでプロットされます。プロットされた線の近くにデータが分布しているので、このモデルはデータに良くあてはまっているようです。

［直交のあてはめ］コマンドの使用例

この例は、2段階に分かれています。まず、「一変量の分布」プラットフォームを使用して、変数を標準化します。次に、標準化された変数を使用して、直交回帰モデルをあてはめます。

変数の標準化

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**一変量の分布**］を選びます。
3. 「**身長 (インチ)**」と「**体重 (ポンド)**」を選択し、［**Y, 列**］をクリックします。
4. ［**OK**］をクリックします。
5. Ctrlキーを押しながら、「**身長 (インチ)**」の赤い三角ボタンをクリックし、［**保存**］>［**標準化**］を選択します。

   Ctrlキーを押しながらコマンドを選択すると、レポートウィンドウに表示されているすべての変数に、選択されたコマンドが一括して適用されます。上記の操作により、「Big Class.jmp」サンプルデータテーブルに、新しい列が2つ追加されます。
6. 「一変量の分布」レポートウィンドウを閉じます。

標準化された変数を使用した直交回帰モデルのあてはめ

1. 「Big Class.jmp」サンプルデータテーブルから、［**分析**］>［**二変量の関係**］を選択します。
2. 「**標準化体重（ポンド）**」を選択し、［**Y, 目的変数**］をクリックします。
3. 「**標準化身長（インチ）**」を選択し、［**X, 説明変数**］をクリックします。
4. ［**OK**］をクリックします。
5. 赤い三角ボタンのメニューから、［**直線のあてはめ**］を選択します。
6. 赤い三角ボタンをクリックし、メニューから［**直交のあてはめ**］を選択します。次に、以下のサブメニューをそれぞれ選択します。
   - **等分散性**
   - **Y→Xのあてはめ**
   - ［**指定された分散比**］>「0.2」を入力
   - ［**指定された分散比**］>「5」を入力

図 5.21「直交のあてはめ」のオプション例

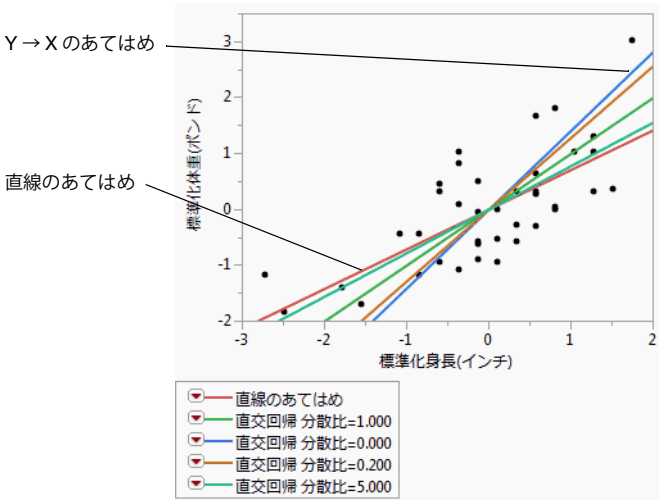


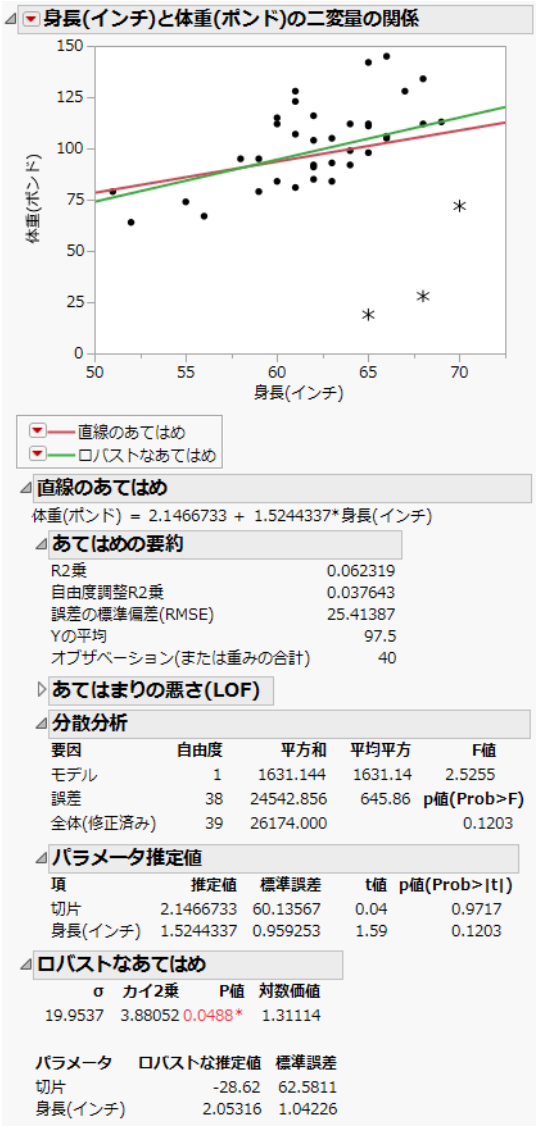
図5.21の散布図は、「身長(インチ)」と「体重(ポンド)」を標準化し、いろいろな直線をあてはめたものです。分散比を変化させたときの「直交のあてはめ」の様子がわかります。X変数の分散をゼロとした場合は、普通の回帰直線です。その対極にあるのが**XのYへの回帰**で、そこではY変数の分散をゼロとしています。他の直線は、これら2つの間に位置し、分散比に応じて少しずつ傾きが変わっています。分散比が大きくなるにつれて、Yの変動だけが考慮されるようになり、直線の傾きがYのXへの回帰直線（普通の回帰直線）に近付きます。逆に、分散比を小さくすると、直線の傾きはXのYへの回帰直線に近付きます。

［ロバストなあてはめ］コマンドの使用例

「Weight Measurements.jmp」サンプルデータには、40名の学生の身長と体重が記録されています。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Weight Measurements.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**体重(ポンド)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**身長(インチ)**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンのメニューから、［**直線のあてはめ**］を選択します。
7. 赤い三角ボタンのメニューから、［**ロバスト**］>［**ロバストなあてはめ**］を選択します。

図 5.22 ロバストなあてはめの例



標準の「分散分析」レポートを見ると、p 値が 0.1203 であるため、身長と体重に線形関係はないという誤った結論に導かれてしまう可能性がありますが、「ロバストなあてはめ」レポートを確認すれば、p 値が 0.0489 になっているため、線形関係は成立すると結論付けられるでしょう。一部の測定値が不自然に低いようなので、ユーザーによる誤入力が疑われます。これらの測定値が分析に好ましくない影響を与えたものと考えられます。

グループ別確率楕円の例

この例では、「Hot Dogs.jmp」サンプルデータを使用します。「**タイプ**」列には、ホットドッグの種類（牛肉、合い挽き、鳥肉）が記録されています。3種類のホットドッグでグループ分けして、コストを分析します。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Hot Dogs.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**1 オンスあたりのコスト ($)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**蛋白質 1 ポンドあたりのコスト ($)**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**グループ別**］を選択します。
7. リストから「**タイプ**」を選択します。
8. ［**OK**］をクリックします。もう一度［**グループ別**］オプションを確認すると、横にチェックマークが付いていることがわかります。
9. 赤い三角ボタンをクリックし、メニューから［**確率楕円**］>［**0.90**］を選択します。

点を「**タイプ**」別に色分けするには、次の手順に従います。

10. 散布図上を右クリックし、［**行の凡例**］を選択します。
11. 列のリストで「**タイプ**」を選択し、［**OK**］をクリックします。

図 5.23 グループ別の例

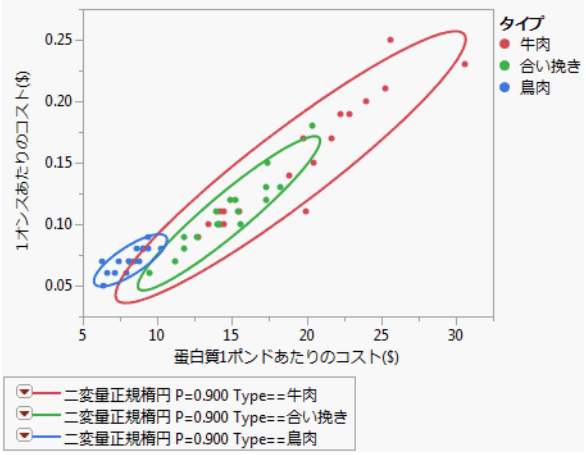


図 5.23 は、異なる種類のホットドッグの確率楕円を、コストを変数としてプロットしたものです。種類ごとに点がクラスターを形成しているのがわかります。

グループ別回帰直線の例

次の例では、グループ別に回帰直線を求めて、その傾きを比較します。

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Big Class.jmp」を開きます。
2. ［分析］>［二変量の関係］を選びます。
3. 「体重 (ポンド)」を選択し、［Y, 目的変数］をクリックします。
4. 「身長 (インチ)」を選択し、［X, 説明変数］をクリックします。
5. ［OK］をクリックします。

図 5.24 左側の回帰直線を描画するには、次の手順に従います。

6. 赤い三角ボタンをクリックし、メニューから［直線のあてはめ］を選択します。

図 5.24 右側の回帰直線を描画するには、次の手順に従います。

7. ［直線のあてはめ］メニューから［あてはめの削除］を選択します。
8. 赤い三角ボタンをクリックし、メニューから［グループ別］を選択します。
9. リストから「性別」を選択します。
10. ［OK］をクリックします。
11. 赤い三角ボタンをクリックし、メニューから［直線のあてはめ］を選択します。

図 5.24 データ全体とグループ別の回帰分析の例

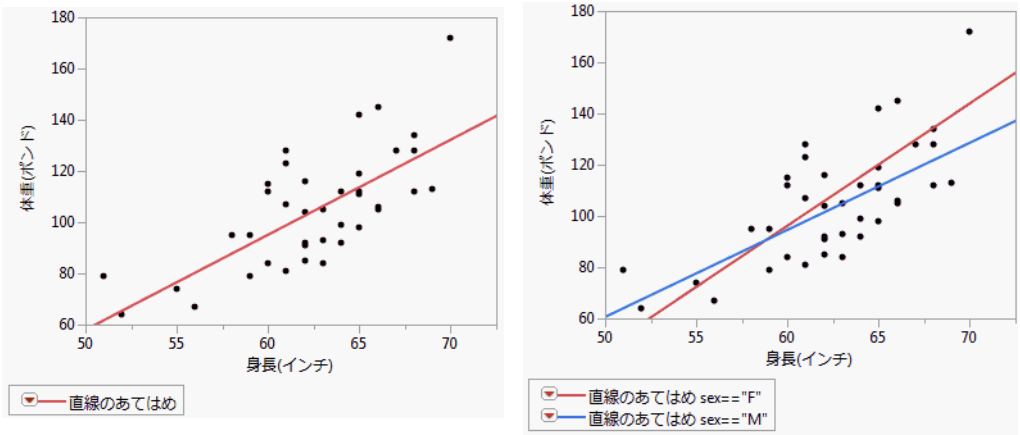


図 5.24 右側の散布図に描かれている回帰直線は、それぞれ男子および女子のデータから推定されています。左側の散布図に描かれている回帰直線は、データ全体から推定されています。

「二変量」プラットフォームの統計的詳細

ここでは、「二変量」プラットフォームの統計的詳細について説明します。

［直線のあてはめ］の統計的詳細

［**直線のあてはめ**］コマンドでは、残差の平方和を最小にするような直線のパラメータβ_0とβ_1が求められます。なお、i番目の行に対して、直線のモデルは$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$と表されます。

2次多項式は放物線、3次多項式は3次曲線です。次数をkとしたとき、i行目のモデルは、次式で表されます。

$$y_i = \sum_{j=0}^{k} \beta_j x_i^j + \varepsilon_i$$

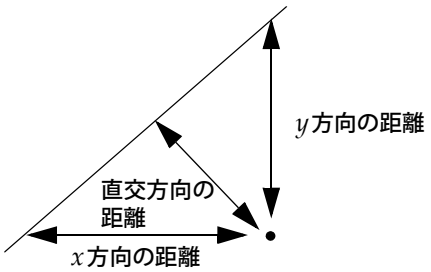
［スプライン曲線のあてはめ］の統計的詳細

3次の平滑化スプライン法では、複数の3次多項式を、つなぎ目（節点）においても滑らかで連続な曲線になるように、つなぎ合わせます。また、ペナルティ（曲線の曲がり具合を、曲線全体の範囲で積分したもの）と誤差平方和の合計が最小になるような推定値を求めます。この手法の詳細については、Reinsch（1967）の論文かEubank（1988）の本を参照してください。

［直交のあてはめ］の統計的詳細

標準的な最小2乗法では、X変数が所与で、Xの関数に誤差を加えたものがY変数であると仮定されます。しかし、Xの測定値にもランダムなばらつきがあるときは、y方向の距離ではなく、直交方向の距離の平方和を最小にする直線を求めた方がよいでしょう。詳細については、図5.25を参照してください。ただし、直交方向に引かれた垂線の長さは、XとYのスケールに左右されます。XとYのスケールをどのように決めるべきかということは、単なるグラフ上の問題ではなく、統計的な問題です。

図5.25 回帰直線への垂線



直交のあてはめを行うには、Y の誤差の、Xの誤差に対する分散比を指定する必要があります。ここで、「分散」はデータから計算される分散ではなく、誤差の母分散であることに注意してください $(\sigma_y^2)/(\sigma_x^2)$ の比は、標準的な最小2乗法では σ_x^2 が0なので、無限大になります。大きな分散比を指定して「直交のあてはめ」を行うと、求めれた直線は、標準的な最小2乗法の回帰直線に近くなります。逆に、分散比を0にすると、YのXへの回帰ではなく、XのYへの回帰に等しい結果が得られます。

この直交のあてはめが最も使われるのは、2つの測定システム（測定器）で、同じ対象を測定したときです。この場合、XとYの測定値には、同じような種類の測定誤差があります。この分析において、測定誤差の分散はどうやって求めたらよいでしょうか。二変量のデータからは、誤差のうちどれだけがどちらの測定システムで生じているのかがわからないので、測定誤差の分散を求めることはできません。そのため、何らかの適当な分散比（たとえば、1.0）を設定するか、2つの測定システムそれぞれで同じ対象を繰り返し測定して、誤差分散を求める必要があります。

この手法の長所は、XとY両方の予測値が計算できることです。この予測値は、直線上でデータ点に最も近い座標です。ただし、この「近さ」は、分散比に対応したスケールで計算されます。

信頼区間は、Tan and Iglewicz（1999）で説明されている方法で計算されます。

「あてはめの要約」レポートの統計的詳細

R2 乗

応答変数が連続尺度の場合、「分散分析」レポートの統計量から、次式によって**「R2 乗」**は計算されます。

$$\frac{\text{モデルの平方和}}{\text{修正済みの全平方和}}$$

自由度調整 R2 乗

R2 乗値は平方和の比でしたが、自由度調整済み R2 乗値は、次のように、平均平方の比です。

$$1 - \frac{\text{誤差の平均平方}}{\text{全体の修正済みの平均平方}}$$

「誤差」の平均平方は、「分散分析」レポートに表示されています。詳細については、図5.12を参照してください。分母の平均平方は、**「全体　(修正済み)」**の**平方和**を、対応する自由度で割ったものです。

「あてはまりの悪さ (LOF)」レポートの統計的詳細

「純粋誤差」の自由度

たとえば、「Big Class.jmp」データテーブルには、身長が同じ生徒が複数名います。一般に、X 値が同じであるグループが g 個あるとき、それらの自由度を合算した、純粋誤差の自由度（DF_p）は、次式で表されます。

$$DF_p = \sum_{i=1}^{g} (n_i - 1)$$

ここで、n_iはi番目のグループの被験者数を表します。

「純粋誤差」の平方和

一般に、X値が同じであるグループがg個あるとき、それらの平方和を合算した、純粋誤差の平方和（SS_p）は、次式で表されます。

$$SS_p = \sum_{i=1}^{g} SS_i$$

ここで、SS_iはi番目のグループにおける平方和（各グループで、グループの平均を引いて計算された平方和）です。

最大 R2 乗

「純粋誤差」は、指定されたモデルに関係なく不変です。また、あらゆるモデルの中での誤差分散の最小値を示します。**「最大 R2 乗」**は、次式で計算されます。

$$1 - \frac{\text{純粋誤差の平方和}}{\text{修正済みの全平方和}}$$

「パラメータ推定値」レポートの統計的詳細

標準 β

標準 β は、次の式で計算されます。

$$\hat{\beta}(s_x / s_y)$$

ここで、$\hat{\beta}$はパラメータ推定値、sxとsyはX変数とY変数の標準偏差です。

計画の標準誤差

この値は、パラメータ推定値の標準誤差をRMSEで割って求めます。

「平滑化スプライン曲線によるあてはめ」レポートの統計的詳細

JMPの平滑化スプラインにおけるR2乗は、1 - SSE/SSTによって計算されています。ここで、SSEは平滑化スプラインの誤差平方和です。また、SSTは修正済みの全平方和で、「直線のあてはめ」の「分散分析」レポートに表示されます。

「相関」レポートの統計的詳細

「相関」レポートに表示される相関は、Pearson の相関係数です。Pearson の相関係数は、"*r*" という記号で表されることが多いです。次式で計算されます。

$$r_{xy} = \frac{s^2_{xy}}{\sqrt{s^2_x s^2_y}}$$ この式で、$$s^2_{xy} = \frac{\sum w_i (x_i - \overline{x}_i)(y_i - \overline{y}_i)}{df}$$

この式の w_i は、「重み」列を指定してあるときは i 番目のオブザベーションの重み、「重み」列がないときは 1 です。

第 6 章

一元配置分析

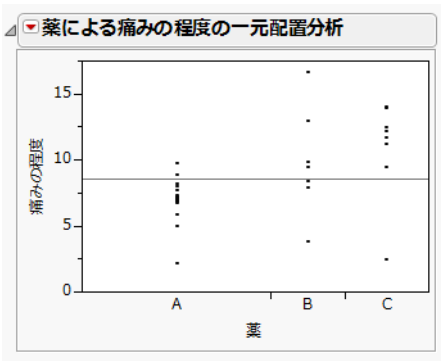
連続尺度のＹ変数とカテゴリカルなＸ変数の関係を調べる

「一元配置」または「二変量の関係」プラットフォームでは、カテゴリカルなＸ変数のグループによって、連続尺度のＹ変数における分布がどのように異なるかを検討できます。たとえば、薬剤の種類（X）によって、連続尺度で測定した痛みの程度（Y）がどのように異なるかを調べることができます。

「一元配置」は、「二変量の関係」のプラットフォームにおいて、**Ｙが連続変数で、Ｘが名義／順序変数**の場合に使われるものです。追加で行った統計分析の情報は、プロットにも重ねて描かれていきます。結果を対話的に操作して、次のような分析を追加で行うことができます。

- 一元配置の分散分析を実行する。グループごとに平均を求めて、それらが等しいかどうかを検定する
- ノンパラメトリックな検定を行う
- 等分散性を検定する
- 検定結果を比較円で示しながら、平均の多重比較を行う
- 外れ値の箱ひげ図を、グループごとに表示する
- 一元配置分散分析に関して、検出力などを求める

図6.1 一元配置分析



目次

一元配置分析の概要

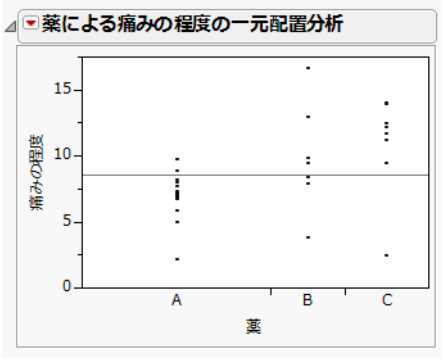
一元配置の分散分析では、群（グループ）ごとの平均に差があるかどうかを検定します。応答変数の総変動は、群内変動と群間変動に分割できます。群内変動に比べて群間変動が大きいときは、群ごとの平均に有意差があると判定します。

一元配置分析の例

この例では、「Analgesics.jmp」サンプルデータを使用します。33 人の被験者が 3 種類の鎮痛剤（A、B、C）を服用し、痛みの程度を数値で評価しています。A、B、および C の平均値に有意差があるかどうかを調べてみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Analgesics.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**痛みの程度**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**薬**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。

図 6.2 一元配置分析の例



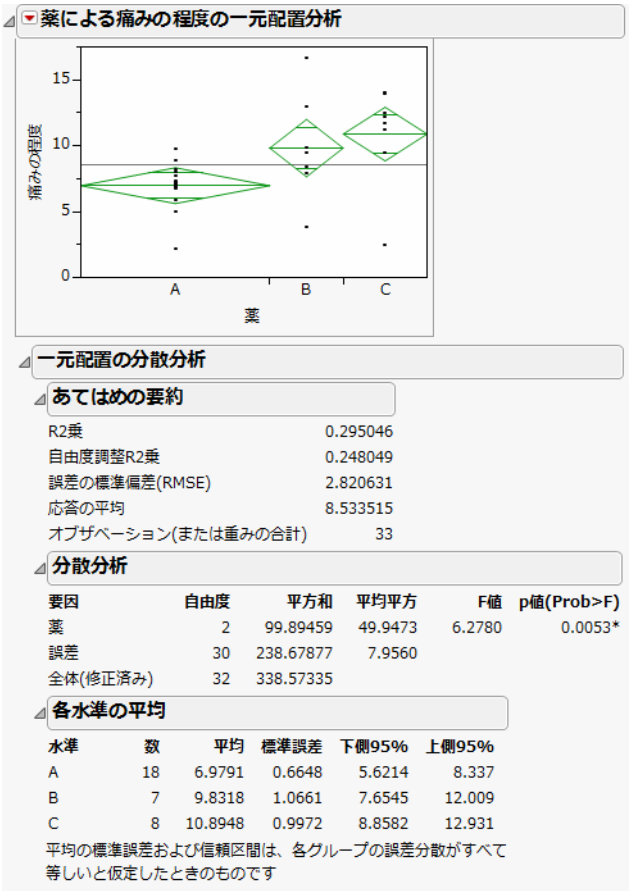
A 薬のスコアが、他の薬と比べて全体的に低いことがわかります。なお、X 軸の目盛りが等間隔ではありませんが、これは、各薬ごとのデータの個数に目盛りの幅が比例しているためです。

データの分散分析を行います。

6. 「薬による痛みの程度の一元配置分析」の赤い三角ボタンをクリックし、メニューから［**平均/ANOVA**］を選択します。

注 : X 変数の水準が 2 つしかない場合は、［**平均/ANOVA**］オプションの代わりに［**平均/ANOVA/ プーリングした t 検定**］が表示され、レポートウィンドウにはプーリングした *t* 検定のレポートも表示されます。

図 6.3［平均/ANOVA］オプションの例



次の点を確認してください。

- 信頼区間を示す平均のひし形が表示されます。
  - 各ひし形の中央に引かれている水平線はグループ平均を示します。一見したところでは、各薬の平均に有意差があるように見えます。
  - ひし形の縦の長さがグループ平均の 95% 信頼区間を表します。

  詳細については、「平均のひし形と標本サイズに比例した X 軸」（153 ページ）を参照してください。
- 「あてはめの要約」レポートは、分析に関する全体的な要約情報をまとめたものです。
- 「分散分析」レポートには、分散分析（ANOVA; Analysis of Variance）の基本的な情報が表示されます。この例では**「Prob>F」**（*p* 値）が 0.0053 です。これは、薬の平均には有意差がありそうだという、グラフからの直観的な判断が正しいことを裏付けています。
- 「各水準の平均」レポートには、カテゴリカルな因子の水準ごとに、平均、標本サイズ、標準誤差が表示されます。

「一元配置」プラットフォームの起動

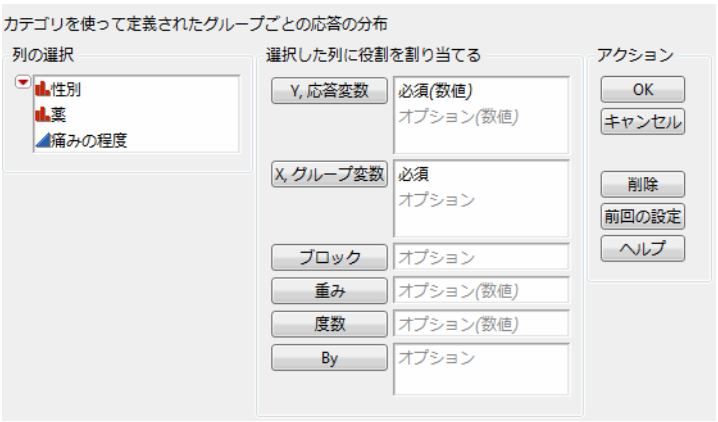
一元配置の分散分析は、「二変量の関係」または「一元配置」プラットフォームを使用して実行できます。どちらを使用しても結果は同じです。

- 「二変量の関係」プラットフォームを起動するには、[**分析**] > [**二変量の関係**] を選択します。

  または

- 「一元配置」プラットフォームを起動するには、「JMP スターター」ウィンドウのカテゴリリストで [**基本統計**] をクリックし、[**一元配置**] をクリックします。

図 6.4「一元配置」起動ウィンドウ



この起動ウィンドウの詳細については、「「二変量の関係」プラットフォームの概要」の章（91 ページ）を参照してください。

[**OK**] をクリックすると、「一元配置分析」レポートウィンドウが開きます。詳細については、「一元配置プロット」（141 ページ）を参照してください。

一元配置プロット

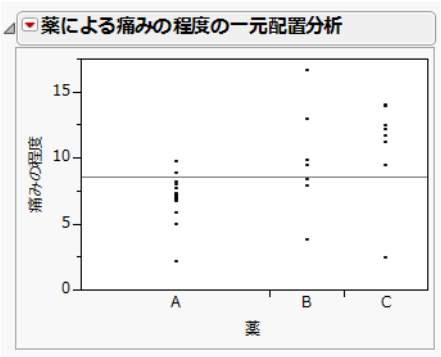
一元配置プロットは、説明変数（X）の値ごとに、応答変数 (Y) の点の分布を示します。Y の分布を、X の水準間で比較できます。X における一意な各値は、「水準」とも言います。

プロットで使われている変数は、ドラッグ & ドロップ操作によって、別の変数に置き換えることができます。X 軸と Y 軸の変数を入れ替えるには、一方の軸の変数をドラッグし、他方の軸の上にドロップします。また、データテーブルの他の変数に置き換えるには、「列」パネルで該当する変数をクリックし、プロットの軸上までドラッグします。

「一元配置分析」の赤い三角ボタンをクリックし、メニューから各種オプションを選択して、レポート、プロット、検定結果をレポートウィンドウに追加できます。詳細については、「「一元配置」プラットフォームのオプション」（142 ページ）を参照してください。

図 6.5 は、「一元配置分析の例」（139 ページ）の手順に従って作成したものです。

図 6.5 一元配置プロット



「一元配置」プラットフォームのオプション

注： Y 変数または X 変数を複数指定した場合は、[あてはめのグループ] メニューが表示されます。このメニューのオプションを使って、レポートの配置を変えたり、適合度の良い順に並べ替えたりできます。詳細については、『基本的な回帰モデル』を参照してください。

オプションを選択すると、追加のグラフ要素がプロットに描画され、また、レポートウィンドウに数値結果が追加されます。

表 6.1 オプションと要素の例

プラットフォームオプション	プロットに追加されるグラフ要素	レポートウィンドウに追加される結果
分位点	箱ひげ図	「分位点」レポート
平均/ANOVA	平均のひし形	「一元配置の分散分析」（ANOVA）レポート
平均と標準偏差	平均線、平均誤差バー、標準偏差線	「平均と標準偏差」レポート
平均の比較	比較円 （［ノンパラメトリックな多重比較］オプションを除く）	「平均の比較」レポート

次の表では、「一元配置分析」の赤い三角ボタンのメニューにあるプラットフォームオプションについて説明します。特定の条件下でなければ表示されないオプションもあります。

表 6.2「一元配置分析」プラットフォームオプションの説明

分位点	グループごとに次の分位点が表示されます。 • 0%（最小値） • 10% • 25% • 50%（中央値） • 75% • 90% • 100%（最大値） ［**表示オプション**］サブメニューの［**箱ひげ図**］オプションが選択状態になります。 詳細については、「分位点」（147 ページ）を参照してください。
平均/ANOVA または **平均/ANOVA/ プーリングした t 検定**	各グループの平均が求められ、一元配置分散分析によって平均間に差があるかどうかが検定されます。詳細については、「平均/ANOVA および平均/ANOVA/ プーリングした t 検定」（149 ページ）を参照してください。 ［**平均/ANOVA/ プーリングした t 検定**］オプションは、X 変数の水準数が 2 つの場合にのみ表示されます。
平均と標準偏差	グループごとに要約統計量が表示されます。ここで計算される平均の標準誤差は、各グループの誤差分散がすべて等しいと仮定した推定値ではなく、グループごとに計算された標準偏差が使われます。 プロットには、平均線、平均誤差バー、標準偏差線が表示されます。これらの要素の簡単な説明については、「表示オプション」（146 ページ）を参照してください。さらに詳しい解説については、「平均線、平均誤差バー、標準偏差線」（153 ページ）を参照してください。
個々の分散を用いた t 検定	分散が等しくないと仮定したときの *t* 検定が計算されます。詳細については、「「t 検定」レポート」（150 ページ）を参照してください。 このオプションは、X 変数の水準数が 2 つの場合にのみ表示されます。

表 6.2「一元配置分析」プラットフォームオプションの説明（続き）

平均分析法	サブメニューが開き、平均分析（Analysis of Means）を実行するコマンドが 5 つ表示されます。これらは、平均、分散、および範囲を比較するためのコマンドです。詳細については、「平均分析法」（154 ページ）を参照してください。
平均の比較	サブメニューが開き、複数のグループ平均を比較するための多重比較法が表示されます。詳細については、「平均の比較」（158 ページ）を参照してください。
ノンパラメトリック	サブメニューが開き、グループの位置をノンパラメトリックに比較するための方法が表示されます。詳細については、「ノンパラメトリック」（163 ページ）を参照してください。
等分散性の検定	4 種類の等分散性の検定を実行します。さらに Welch 検定も行います。Welch 検定は、グループ内の分散が等しくないときに、平均を比較する分散分析です。詳細については、「等分散性の検定」（167 ページ）を参照してください。
同等性の検定	平均の差が、特定の閾値より小さいことを検定します。詳細については、「同等性の検定」（169 ページ）を参照してください。
ロバスト	外れ値の影響を受けにくい 2 つの手法が用意されています。詳細については、「ロバスト」（170 ページ）を参照してください。
検出力	特定の仮説検定に関して、検出力などを計算します。詳細については、「検出力」（170 ページ）を参照してください。 「検出力の詳細」ウィンドウとレポートは、「モデルのあてはめ」プラットフォームでも表示されます。検出力の計算の詳細および例については、『基本的な回帰モデル』を参照してください。
α 水準の設定	サブメニューから、よく使われている α 水準を選択するか、[**その他**] を選択して任意の α 水準を入力します。α 水準を変更すると、次の処理が行われます。 • 信頼限界が再計算されます。 • プロット上に平均のひし形が表示されている場合は調整されます。 • レポート内の上側信頼限界と下側信頼限界の値が変更されます。 • すべての「平均の比較」レポートの棄却値と比較円が変更されます。 • すべての「ノンパラメトリックな多重比較」レポートの棄却値が変更されます。

表 6.2「一元配置分析」プラットフォームオプションの説明（続き）

正規分位点プロット	サブメニューが開き、各グループの分位点プロットに関する次のオプションが表示されます。 • ［**分位点 - 実測値プロット**］を選択すると、X 軸が正規分布の分位点、Y 軸が実測値を表す分位点プロットが作成されます。カテゴリカルな X 因子の水準ごとの分位点プロットが表示されます。 • ［**実測値 - 分位点プロット**］は、［**分位点 - 実測値プロット**］の X 軸と Y 軸を逆にしたものです。 • ［**あてはめ線**］を選択すると、X 変数の水準ごとに、プロット上に参照線として直線が引かれます。このオプションは、プロットを作成（［分位点 - 実測値プロット］または［実測値 - 分位点プロット］を選択）してからでないと使用できません。
累積確率プロット	グループごとの累積分布関数がプロットされます。詳細については、「累積確率プロット」（172 ページ）を参照してください。
密度	グループ間の密度を比較します。詳細については、「密度」（173 ページ）を参照してください。
対応のある列を設定	対応のあるモデルの分析を行うときに、対応のある変数を指定します。このオプションは、異なるグループのオブザベーションが同じ個体から取られた場合など、一元配置分析で使うデータが対応のあるデータから取り出されたものである場合に使用します。 プロット上では、対応のある点が線でつながれます。 詳細については、「対応のある列を設定」（173 ページ）を参照してください。
保存	現在のデータテーブルの新しい列に、次の数値を保存します。 • ［**残差の保存**］では、X の各水準ごとに、Y から Y の平均を引いた値が保存されます。 • ［**標準化データの保存**］では、X の各水準ごとに標準化された Y の値が保存されます。これは、各水準ごとに、中心化した Y を標準偏差で割ったものです。 • ［**正規分位点の保存**］では、カテゴリカルな X 変数の各水準内で計算された正規分位点が保存されます。 • ［**予測値の保存**］では、X の水準ごとに応答変数の平均が保存されます。
表示オプション	プロットに対して要素を追加または削除します。詳細については、「表示オプション」（146 ページ）を参照してください。

表 6.2「一元配置分析」プラットフォームオプションの説明（続き）

スクリプト	このメニューには、すべてのプラットフォームに共通するスクリプト関連のコマンドがリストされています。分析をやり直すコマンド、分析の JSL コマンドをウィンドウやファイルに保存するコマンドなどがあります。詳細は、『JMP の使用法』を参照してください。

表示オプション

［表示オプション］サブメニューを使用して、プロットに要素を追加したり、プロットから要素を削除したりします。条件が合わないと表示されないオプションもあります。

表 6.3［表示オプション］の説明

すべてのグラフ	プロット全体の表示／非表示を切り替えます。
点	プロット上のデータ点の表示／非表示を切り替えます。
箱ひげ図	各グループの箱ひげ図の表示／非表示を切り替えます。
平均のひし形	ひし形の中央にある水平線は、各グループの平均に位置しています。また、その長さは、X 軸の長さに比例しています。ひし形の上下の点は、グループの上側と下側の 95% 信頼限界を表します。詳細については、「平均のひし形と標本サイズに比例した X 軸」（153 ページ）を参照してください。
平均線	各グループの平均を表す線が引かれます。詳細については、「平均線、平均誤差バー、標準偏差線」（153 ページ）を参照してください。
平均の信頼区間	各グループ平均の上側と下側の 95% 信頼限界に線が引かれます。
平均誤差バー	各グループの平均と、平均から上下へ 1 標準誤差の区間が表示されます。詳細については、「平均線、平均誤差バー、標準偏差線」（153 ページ）を参照してください。
全体平均	プロット上に、Y 変数の全体平均を示す線が引かれます。
標準偏差線	各グループの平均から上下へ 1 標準偏差だけ離れたところに、線が表示されます。詳細については、「平均線、平均誤差バー、標準偏差線」（153 ページ）を参照してください。
比較円	比較円の表示／非表示を切り替えます。このオプションは、**［平均の比較］**のオプションが選択されている場合にだけ使用できます。詳細については、「比較円の統計的詳細」（192 ページ）を参照してください。
平均をつなぐ	グループの平均をつなぐ直線が表示されます。

表 6.3［表示オプション］の説明（続き）

平均の平均	グループ平均の平均を表す線が引かれます。
標本サイズに比例したX軸	X軸が、各水準の標本サイズに比例した長さに区切られます。詳細については、「平均のひし形と標本サイズに比例したX軸」（153ページ）を参照してください。
点の拡散	点がX軸方向に拡散して表示されます。
点をずらす	Yの値が等しいために重なってしまっている点を少しずつずらして表示します。4*(一様分布乱数 – 0.5)^{5}の乱数により、-0.125～0.125の範囲で点が水平方向に移動します。
対応を表す線	（［**対応のある列を設定**］オプションが選択されている場合にのみ表示されます）。対応のある点を線でつなぎます。
対応を表す点線	（［**対応のある列を設定**］オプションが選択されている場合にのみ表示されます）。テーブル内の欠損セルのセル平均を点線でつなぎます。線の終点は、二元配置の分散分析モデルで計算されます。
ヒストグラム	元のプロットの隣に、横に並べたヒストグラムが表示されます。
ロバスト 平均線	（［ロバスト］のオプションが選択されている場合にのみ表示されます）。各グループのロバスト平均を表す線が引かれます。

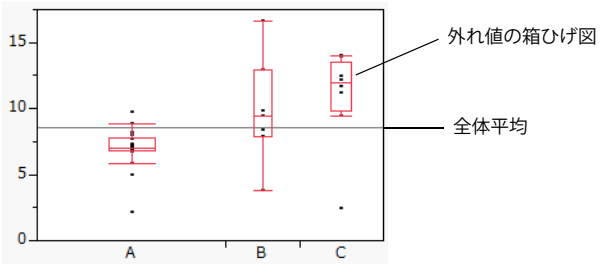
分位点

「分位点」レポートには、X変数の水準ごとに所定のパーセント点が表示されます。中央値は50パーセント点で、25パーセント点と75パーセント点は**四分位点**とも呼ばれます。

［**分位点**］オプションを選択すると、プロットに次のような要素が追加されます。

- Y変数の全体平均を示す線
- Xの水準別に点の分布を要約した箱ひげ図

図 6.6 外れ値の箱ひげ図と全体平均



注: これらの要素を非表示にするには、「一元配置分析」の赤い三角ボタンをクリックし、[**表示オプション**] > [**箱ひげ図**] または [**全体平均**] をクリックします。

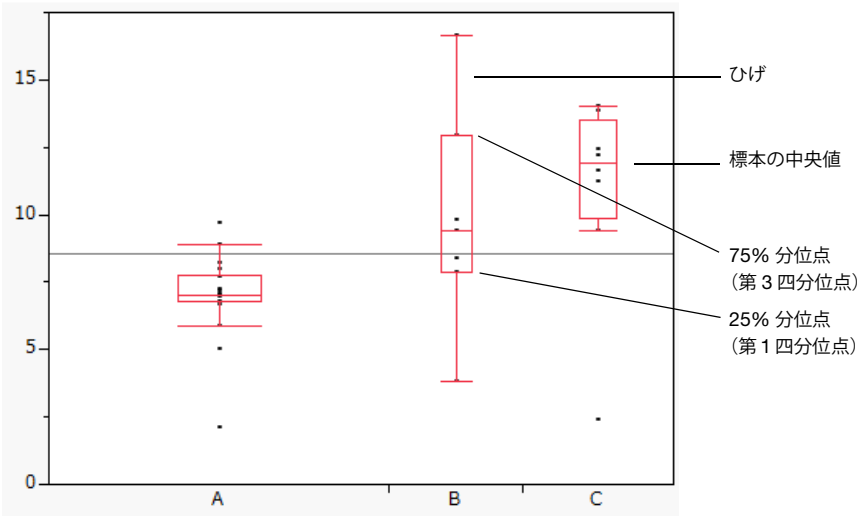
外れ値の箱ひげ図

外れ値の箱ひげ図は、データの分布をグラフの形で要約したものです。外れ値の箱ひげ図について、次の点を確認してください（図6.7を参照）。

- 箱を横切る線は標本の中央値（メディアン）を示します。
- 箱の両端は25%と75%の分位点で、それぞれ第1および第3四分位点とも呼ばれます。
- 第1四分位点と第3四分位点の差が**四分位範囲**です。
- どの箱にも両端に**ひげ**と呼ばれる線が付いています。ひげは、箱の端から、次の式で計算された範囲内で最も遠くにある点まで伸びています。

  第3四分位点 + 1.5*（四分位範囲）
  第1四分位点 - 1.5*（四分位範囲）
  上の式で計算された範囲内にあるデータ点の一番大きい値と小さい値（外れ値は除く）が、ひげの両端になります。

図6.7 外れ値の箱ひげ図の例



平均/ANOVA および平均/ANOVA/ プーリングした t 検定

［**平均/ANOVA**］オプションは、分散分析を実行します。X 変数の水準が 2 つしかない場合は、このオプションは［**平均/ANOVA/ プーリングした t 検定**］という名前で表示されます。その場合は、他のレポートに加えて、「*t* 検定」レポートも作成され、誤差分散をプールして求めた検定（誤差分散が等しいと仮定した検定）が表示されます。

表 6.4 レポートウィンドウの要素の説明

エレメント	参照
一元配置プロットに平均のひし形が表示されます。	「表示オプション」（146 ページ）および「平均のひし形と標本サイズに比例した X 軸」（153 ページ）を参照してください。
レポート	「「あてはめの要約」レポート」（149 ページ）を参照してください。
	「「分散分析」レポート」（151 ページ）を参照してください。
	「「各水準の平均」レポート」（152 ページ）を参照してください。
	「「t 検定」レポート」（150 ページ）を参照してください。 **注：**このレポートは、［**平均 /ANOVA/ プーリングした t 検定**］オプションを選択した場合にのみ表示されます。
	「「ブロック平均」レポート」（152 ページ）を参照してください。 **注：**このレポートは、起動ウィンドウでブロック変数を指定した場合にのみ表示されます。

「あてはめの要約」レポート

「あてはめの要約」レポートは、一元配置の分散分析の結果をまとめたものです。

レポートに表示される統計量は次のとおりです。

表 6.5「あてはめの要約」レポートの説明

R2 乗	Y の変動のうち、X の各水準に平均をあてはめることにより説明される変動の割合を表します。残りの変動は、モデルで説明できないため、ランダムな誤差とみなされます。R^2 が 1 ということは、グループ平均をあてはめることで変動のすべてが説明され、誤差がないことを意味します。R^2 が 0 だと、このあてはめが予測モデルとして全体平均より優れていないことを示します。詳細は、「「あてはめの要約」レポートの統計的詳細」（194 ページ）を参照してください。 R^2 は**決定係数**とも言います。

表 6.5「あてはめの要約」レポートの説明（続き）

自由度調整 R2 乗	パラメータ数が異なるモデルでも比較できるように、自由度によって R^2 を調整したもの。詳細は、「「あてはめの要約」レポートの統計的詳細」（194 ページ）を参照してください。
誤差の標準偏差（RMSE）	ランダムな誤差の標準偏差の推定値です。「分散分析」レポートにある**「誤差」**の平均平方の平方根です。
応答の平均	Y 変数の全体平均（算術平均）。
オブザベーション（または重みの合計）	あてはめを推定するために使われたオブザベーションの数。ただし、重み変数が使われた場合は、重みの合計です。 詳細については、「「あてはめの要約」レポートの統計的詳細」（194 ページ）を参照してください。

「t 検定」レポート

注 : このレポートは、[**平均 /ANOVA/ プーリングした t 検定**] オプションを選択した場合にのみ表示されます。

t 検定には 2 種類あります。

- 等分散性を仮定して、平均を比較する検定。[**平均 /ANOVA/ プーリングした t 検定**] オプションを選択した場合は、「*t* 検定」レポートが表示されます。これは、分散が等しいと仮定した場合の *t* 検定です。
- 等分散性を仮定しないで、平均を比較する検定。赤い三角ボタンのメニューから [**個々の分散を用いた t 検定**] オプションを選択した場合も、「*t* 検定」レポートが表示されます。これは、分散が等しくないと仮定した場合の *t* 検定です。

レポートには次の情報が表示されます。

表 6.6「t 検定」レポートの説明

t 検定プロット	帰無仮説が真であると仮定したときの、平均の差の標本分布を示します。赤色の縦線は平均の実際の差を表しています。陰影が付いた領域は p 値に相当します。
差	2 つの X 水準に見られる差の推定値を表示します。プロットでは、2 水準を比較した「差」の値が赤い線で示されます。
差の標準誤差	差の標準誤差を示します。
差の上側信頼限界	差の上側信頼限界を示します。
差の下側信頼限界	差の下側信頼限界を示します。

表 6.6「t 検定」レポートの説明（**続き**）

信頼率	信頼水準（1 - α）を示します。信頼水準を変更するには、プラットフォームの赤い三角ボタンをクリックし、［**α 水準の設定**］サブメニューから新しいα水準を選択します。
t 値	*t* 検定統計量の値。
自由度	*t* 検定で使用された自由度。
p 値 (Prob>|t|)	両側検定の *p* 値。
p 値 (Prob>t)	下側の片側検定の *p* 値。
p 値 (Prob<t)	上側の片側検定の *p* 値。

「分散分析」レポート

「分散分析」レポートでは、標本の合計変動が 2 つの要素に分かれています。そのうち、平均平方の比は *F* 値と呼ばれます。*F* 値の確率が小さいときは、そのモデルが Y の全体平均よりも統計的に良いあてはめであることを意味します。

注: ブロック列を指定した場合は、「分散分析」レポートに**ブロック**変数も表示されます。

「分散分析」レポートには、次のような情報が表示されます。

表 6.7「分散分析」レポートの説明

要因	変動の 3 つの要因（モデル要因、**「誤差」**、 **「全体 (修正済み)」**）。
自由度	各変動因の自由度（DF）。 • **「全体 (修正済み)」**の自由度は、 N - 1 で求められます。N は分析に使用されたオブザベーションの数です。 • X 変数の水準数を k とすると、モデルの自由度は k - 1 になります。 **「誤差」**の自由度は、**「全体 (修正済み)」**と**モデル**の自由度の差（N - k）です。
平方和	変動の各要因の平方和（SS）。 • 全体（**「全体 (修正済み)」**の平方和は、各応答から全体平均までの距離の平方和です。 この**全体（修正済み）**の平方和は、基本モデルの平方和として、その他のモデルとの比較に使用されます。 • 「誤差」（残差）の平方和は、各点からそれぞれのグループ平均までの距離の平方和です。この平方和は、分散分析モデルをあてはめた後に残った、説明できない部分の平方和です。 全体平方和から誤差平方和を引いたものがモデルの平方和です。全変動のうちモデルによって説明されている部分を表します。

表 6.7「分散分析」レポートの説明（続き）

平均平方	平方和を関連する自由度で割った値。 • **モデル**の平均平方は、グループ平均が等しいという仮説のもとで推定した誤差の分散です。 • **「誤差」**の平均平方は、モデルの平均平方とは無関係に推定した誤差の分散で、モデルの仮説による制約を受けません。
F 値	モデルの平均平方を「誤差」の平均平方（MSE）で割ったもの。グループ平均が等しい（実際に差がない）という仮説が真のとき、誤差とモデルの平均平方が両方とも誤差分散の推定値となり、その比は *F* 分布に従います。分散分析モデルをあてはめた結果、誤差の変動が全体の変動よりも有意に減少した場合、「*F* 値」は期待値よりも大きくなります。
p 値 (Prob>F)	グループ平均の間に実際には差がないときに、まったくの偶然だけで現在の *F* 値よりも大きな *F* 値を得る確率。観測有意確率が 0.05 以下だと、グループ平均に有意差がある証拠だとみなされます。

「各水準の平均」レポート

「各水準の平均」レポートは、名義／順序尺度の X の水準ごとに、Y 変数の情報をまとめたものです。

「各水準の平均」レポートには、次のような情報が表示されます。

表 6.8「各水準の平均」レポートの説明

水準	X 変数の水準。
数値	各グループ内のオブザベーションの個数。
平均	各グループの平均。
標準誤差	グループ平均の標準偏差の推定値。標準誤差は、Y の分散がどの水準でも等しいという仮定のもとに計算されます。この標準誤差は、「あてはめの要約」レポートにある誤差の標準偏差（RMSE）を、グループごとの標本サイズの平方根で割ったものです。
下側 95% と上側 95%	グループ平均の下側 95% 信頼限界と上側 95% 信頼限界。

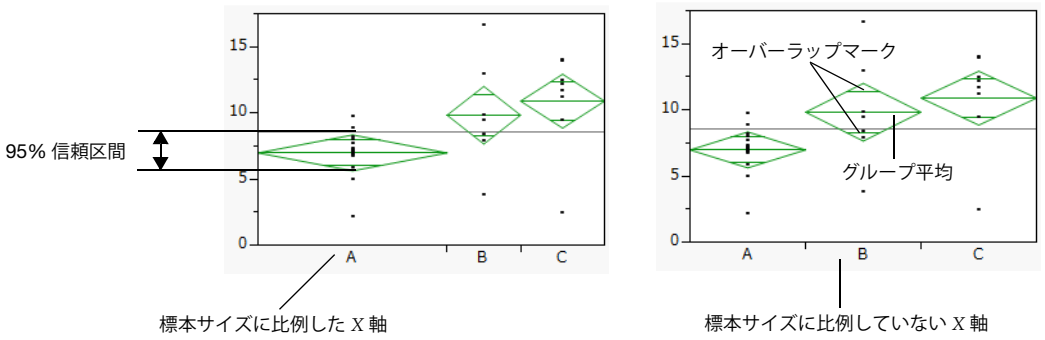
「ブロック平均」レポート

起動ウィンドウでブロック変数を指定した場合は、［**平均/ANOVA**］コマンドと［**平均/ANOVA/ プーリングした t 検定**］コマンドを選択すると、「ブロック平均」レポートが作成されます。このレポートには、各ブロックの平均と、各ブロックのオブザベーションの個数が表示されます。

平均のひし形と標本サイズに比例した X 軸

平均のひし形は、標本平均とその信頼区間を表します。

図 6.8 平均のひし形と標本サイズに比例した X 軸の例



次の点を確認してください。

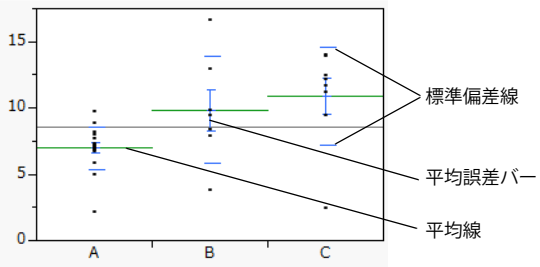
- 各ひし形の上下の点が、各グループの（1 - α）x 100% の信頼区間を表します。信頼区間は、分散がすべてのオブザベーションにわたって等しいという前提のもとに計算されるので、ひし形の高さは、グループ内にあるオブザベーション数の平方根の逆数に比例します。
- **［標本サイズに比例した X 軸］**オプションが選択されているときは、X 軸に沿った各グループの幅（ひし形の横幅）が、そのグループ（X 変数の各水準）の標本サイズに比例します。そのため、幅の狭いひし形は縦長になる傾向があります。それは、データ点の数が少ないと信頼区間が広くなるためです。
- 各ひし形の中央を横切る線はグループ平均を表します。
- グループ平均から上下に離れた位置に引かれている線はオーバーラップマークです。グループの標本サイズが等しい場合、オーバーラップマークを見ると、2 つのグループ平均が所定の信頼水準において有意に異なるかどうかがわかります。オーバーラップマークは、グループ平均 $\pm (\sqrt{2})/2 \times \mathrm{CI}/2$ で計算されます。標本サイズが等しいとき、あるひし形の両側のオーバーラップマークが 2 つとも、別のひし形のオーバラップマークを超えていない場合は、所定の信頼水準において 2 つのグループが異ならないことを意味します。
- 平均のひし形は、プラットフォームのメニューから **［平均/ANOVA/ プーリングした t 検定］**または **［平均/ANOVA］**オプションを選択すると自動的に表示されますが、赤い三角ボタンをクリックし、メニューから **［表示オプション］** > **［平均のひし形］**を選択すれば、いつでも表示／非表示を切り替えることができます。

平均線、平均誤差バー、標準偏差線

平均線を表示するには、**［表示オプション］** > **［平均線］**を選択します。平均線は、X 変数の水準ごとに応答変数の平均を示します。

平均誤差バーと標準偏差線は、赤い三角ボタンをクリックし、メニューから［**平均と標準偏差**］オプションを選択すると表示されます。詳細については、図6.9を参照してください。平均誤差バーと標準偏差線の表示／非表示を個別に切り替える場合は、［**表示オプション**］>［**平均誤差バー**］または［**標準偏差線**］をそれぞれ選択します。

図6.9 平均線、平均誤差バー、標準偏差線



平均分析法

平均分析（ANOM; analysis of means）法は、平均、分散、もしくは、位置や尺度に関するその他の指標を、複数のグループ間で比較します。次のような場合に、この分析は用いることができます。

- いずれかのグループ平均が、全体平均と有意に異なるかどうかを確認する場合
- いずれかのグループの標準偏差が、データ全体から計算される誤差の標準偏差（RMSE）と有意に異なるかどうかを確認する場合
- いずれかのグループの範囲が、範囲の全体平均と有意に異なるかどうかを確認する場合

注：「分割表」プラットフォームでは、応答のカテゴリが2つの場合は［**割合の平均分析**］を使用できます。詳細については、「分割表分析」の章（197ページ）を参照してください。

平均分析法全般に関する詳細と、JMPで実装されている平均分析法については、Nelson et al.（2005）を参照してください。

位置の指標に対する平均分析

次のオプションによって、グループ間で平均や位置が差があるかどうかを検定できます。

- 平均分析
- 順位変換平均分析

平均分析

グループの平均を全体平均と比較するには、［平均分析］を使用します。この分析法では、データがほぼ正規分布に従っているものと仮定します。詳細については、「平均分析法の例」（174ページ）を参照してください。

順位変換平均分析

ノンパラメトリック版の**平均分析**法です。この方法は、データが明らかに正規分布に従っておらず、正規分布に変換できない場合に使用します。各グループで計算した順位平均を、全体の順位平均と比較します。［順位変換平均分析］では、順位を正規スコアに変換したデータに対して、通常の平均分析の手順と棄却値が適用されます。

順位変換

n 個の観測値があるとします。それらを変換したものは次のように計算されます。

- すべての観測値に小さい値から順番に順位を与えます。この時、同順位も考慮します。同順位のものには、平均順位を割り当てます。
- 順位を R_1、$_R2$、...、R_n とします。
- i 番目の観測値に対しての変換は、次のように求められます。

$$\text{変換された } R_i = Normal\ Quantile\left[\left(\frac{R_i}{2n+1}\right) + 0.5\right]$$

変換された R_i の値に平均分析の手順が適用されます。順位は一様分布に従うので、上式で変換された順位の分布は正規分布を半分に折ったものになります。詳細については、Nelson et al. (2005) を参照してください。

尺度の指標に対する平均分析

次のオプションによって、グループごとのばらつきが等しいかどうかを検定できます。

- 分散の平均分析
- 分散の平均分析 -Levene の ADM
- 範囲の平均分析

分散の平均分析

グループの標準偏差（または分散）を誤差の標準偏差（RMSE）（または MSE）と比較するには、［分散の平均分析］を使用します。この分析法では、データがほぼ正規分布に従っているものと仮定します。この方法を使用するには、各グループのオブザベーション数が 4 個以上でなければなりません。分散の平均分析の詳細については、Wludyka and Nelson（1997）および Nelson et al.（2005）を参照してください。例として、「分散に対する平均分析法の例」（175 ページ）を参照してください。

分散の平均分析 -Levene の ADM

［分散の平均分析 -Levene の ADM］では、**中央値からの偏差の絶対値**（ADM; absolute deviations from the median）のグループ平均と、ADM の全体平均を比較します。この手法は分布に対してロバストですので、データが正規分布に従わず、また、正規分布に変換できないと思われる場合は、この手法を使用してください。［分散の平均分析 -Levene の ADM］は、分散に対する平均分析のノンパラメトリック版です。詳細については、Levene（1960）または Brown and Forsythe（1974）を参照してください。

範囲の平均分析

この手法では、各グループの範囲を、範囲の全体平均と比較します。この検定は、尺度の指標として範囲を用いて、各グループのばらつきに差があるかどうかを検定します。詳細については、Wheeler（2003）を参照してください。

注: ［範囲の平均分析］は、バランス（釣り合い）のとれたデータ、および、特定の標本サイズにのみ使用できます。詳細については、「［範囲の平均分析］の制限」（156 ページ）を参照してください。

［範囲の平均分析］の制限

他の［平均分析］とは異なり、「範囲の平均分析」の決定限界は統計表に記載されている棄却値のみを使用します。このため、［範囲の平均分析］は次のすべての条件を満たす場合しか使用できません。

- 各グループの標本サイズが同じ
- 各グループの標本サイズが次の場合: 2–10、12、15、20
- グループ数が 2 ～ 30 個
- 有意水準が 0.10、0.05、0.01

平均分析法のグラフ

平均分析法を実行すると、レポートウィンドウに次の情報を示すグラフが表示されます。

- 上側決定限界（UDL; Upper Decision Limits）
- 下側決定限界（LDL; Lower Decision Limits）
- 決定限界の間で、次の位置を示す水平（中心）線
  - 平均分析: 全体平均
  - 順位変換平均分析: 順位変換データの全体平均
  - 分散の平均分析: 誤差の標準偏差（RMSE）（Y 軸のスケールが分散の場合は MSE）
  - 分散の平均分析 -Levene の ADM: 平均からの偏差の絶対値の全体平均
  - 範囲の平均分析: グループ範囲の全体平均

グループの統計量が決定限界外にプロットされている場合、そのグループの統計量と、すべてのグループの統計量の全体平均の間には統計的な有意差があることを示します。

平均分析のオプション

平均分析法を実行すると、それぞれの平均分析のレポートに対して、赤い三角ボタンのメニューが追加されます。

表 6.9「平均分析法」のオプションの説明

有意水準の設定	サブメニューから、よく使われているα水準を選択するか、[**その他**] を選択して任意のα水準を入力します。有意水準を変更すると、上側決定限界と下側決定限界が変更されます。 **注：**[範囲の平均分析] の場合、選択できるのは 0.10、0.05、および 0.01 のみです。
要約レポートの表示	• [平均分析] の場合は、グループ平均と決定限界を示すレポートが作成されます。 • [順位変換平均分析] の場合は、順位のグループ平均と決定限界を示すレポートが作成されます。 • [分散の平均分析] の場合は、グループごとの標準偏差（もしくは、分散）と決定限界を示すレポートが作成されます。 • [分散の平均分析 -Levene の ADM] の場合は、ADM のグループ平均と決定限界を示すレポートが作成されます。 • [範囲の平均分析] の場合は、グループ範囲と決定限界を示すレポートが作成されます。
Y 軸スケールを分散に	（[**分散の平均分析**] のみ）Y 軸のスケールを標準偏差から分散に変更します。
表示オプション	次の表示オプションがあります。 • [**決定限界の表示**] は、決定限界線の表示／非表示を切り替えます。 • [**決定限界の陰影の表示**] は、決定限界の陰影の表示／非表示を切り替えます。 • [**中心線の表示**] は、中心線の表示／非表示を切り替えます。 • **点のオプション** – [**垂線の表示**] は、垂線を表示します。これがデフォルトのオプションです。 – [**接続線の表示**] は、各グループの平均をつなぐ線を表示します。 – [**点のみを表示**] は、各グループの平均を示す点のみを表示します。

平均の比較

注: 平均の比較方法としては、平均分析（ANOM）法もあります。詳細については、「平均分析法」（154 ページ）を参照してください。

［平均の比較］の各オプションを使用して、グループ平均の多重比較を実行できます。これらのオプションの結果は、誤差分散が等しいと仮定して、プーリングした分散推定値を使っています。各オプションを選択すると、プロットの横に比較円が追加され、レポートウィンドウに特定のレポートが追加されます。比較円の詳細については、「比較円の使用」（159 ページ）を参照してください。

オプション	説明	参照	［ノンパラメトリック］メニューのオプション
各ペア ,Student の t 検定	Student の *t* 検定を使用してペアごとの比較が計算されます。ペアごとの検定をたくさん行うと、仮説検定における実際の α 水準（第 1 種の誤り）が、検定に使用する名目的な α 水準よりも大きくなります。	詳細については、「各ペア ,Student の t 検定」（160 ページ）を参照してください。	［ノンパラメトリック］> ［ノンパラメトリックな多重比較］> ［ペアごと Wilcoxon 検定］
すべてのペア , Tukey の HSD 検定	平均間のすべての差に対して検定が行われます。これは **Tukey** の HSD（honestly significant difference）検定、または **Tukey-Kramer** の HSD 検定と呼ばれます（Tukey 1953, Kramer 1956）。この検定は、グループごとの標本サイズが同じときには、指定された有意水準の正確な検定です。グループごとの標本サイズが異なるときは、保守的な結果になります（Hayter 1984）。	詳細については、「すべてのペア ,Tukey の HSD 検定」（161 ページ）を参照してください。	［ノンパラメトリック］> ［ノンパラメトリックな多重比較］> ［すべてのペア Steel-Dwass 検定］
最適値との比較 (Hsu の MCB)	平均が未知の最大平均より小さいかどうか、未知の最小平均より大きいかどうかが検定されます。これは Hsu の MCB 検定と呼ばれます（Hsu 1981）。	詳細については、「最適値との比較 (Hsu の MCB)」（161 ページ）を参照してください。	なし

オプション	説明	参照	［ノンパラメトリック］メニューのオプション
コントロール群との比較 (Dunnett)	平均がコントロール群の平均と異なるかどうかが検定されます。これは Dunnett の検定と呼ばれます（Dunnett 1955）。	詳細については、「コントロール群との比較 (Dunnett)」（162 ページ）を参照してください。	［ノンパラメトリック］> ［ノンパラメトリックな多重比較］> ［コントロール群との比較 Steel 検定］

注: ブロック列を指定した場合は、多重比較の前に、データがブロック平均に合わせて調整されます。

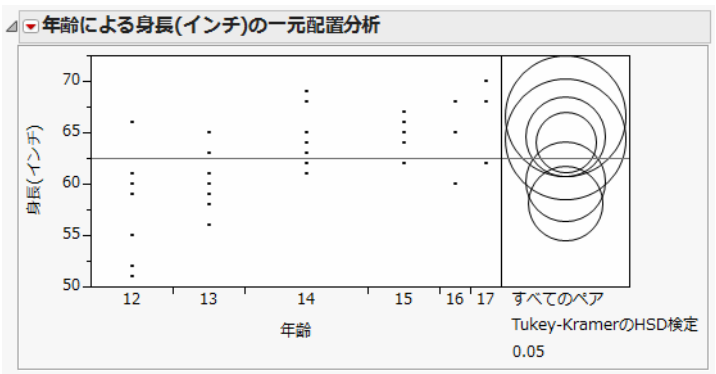
関連情報

- 「「平均の比較」で用意されているすべての検定」（181 ページ）

比較円の使用

多重比較検定では、まず、**比較円**プロットが作成されます。比較円プロットはグループ平均を比較するためのグラフです。図 6.10 は、［すべてのペア ,Tukey の HSD 検定］の比較円を示しています。他の多重比較検定でも同じような比較円が描かれますが、円の直径が異なります。

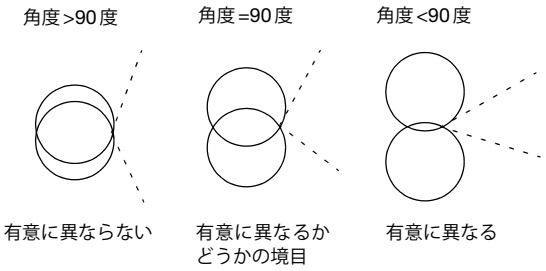
図 6.10 グラフによるグループ平均の比較



グループ平均の各ペアを、比較円の交わりを見て比較することができます。外側の交角によってグループ平均が有意に異なるかどうかがわかります。詳細については、図 6.11 を参照してください。

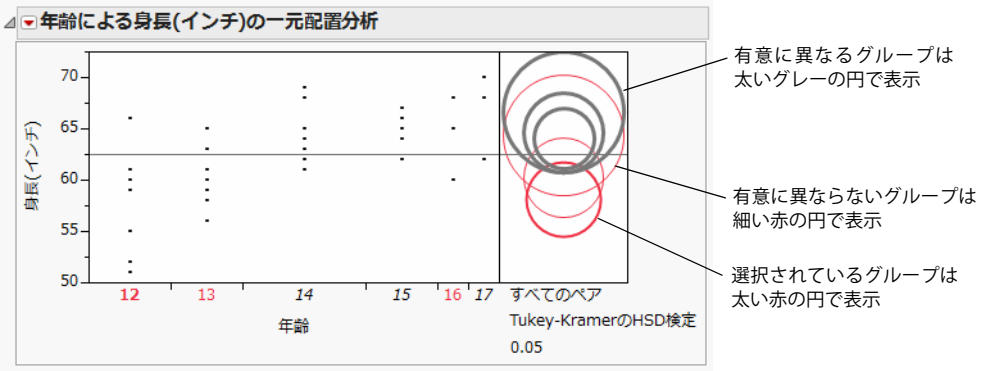
- 平均が有意に異なる場合、比較円は交わらないか交わりが浅く、外側の交角が 90 度以下になります。
- 比較円が 90 度以上の交角で交わっているか入れ子になっている場合は、平均が有意には異ならないことを意味します。

図 6.11 交角と有意性



交角がほぼ 90 度になっていて、見た目では判断がつきにくい場合には、該当する比較円をクリックすると、平均が有意に異なるかどうかを確認できます。詳細については、図 6.12 を参照してください。選択を解除するには、円の外側の白い領域をクリックします。

図 6.12 比較円の選択



関連情報

- 「比較円の統計的詳細」(192 ページ)

各ペア ,Student の t 検定

[各ペア ,Student の t 検定] では、グループ水準の各ペアに対して Student の *t* 検定が行われます。この検定は、個々の比較だけを考慮したものです。

関連情報

- 「[各ペア ,Student の t 検定] の例」(176 ページ)

すべてのペア ,Tukey の HSD 検定

［**すべてのペア ,Tukey の HSD 検定**］（または Tukey-Kramer の HSD 検定）は、すべてのペアを比較するときの有意水準を保護します。そのため、Tukey 検定の最小有意差 (HSD) は、各ペアの比較を行う Student の *t* 検定の最小有意差 (LSD) より大きくなります。また、グラフでは比較円が大きくなり、p 値は大きくなります。

Tukey 検定の分位点「q*」は、q* = (1/sqrt(2)) * q という式で計算されます。ここで、q はスチューデント化された範囲の分位点です。詳細については、Neter, Wasserman, and Kutner（1990）の T 統計量に関する説明を参照してください。

関連情報

- 「［すべてのペア ,Tukey の HSD 検定］の例」（177 ページ）

最適値との比較 (Hsu の MCB)

［**最適値との比較 (Hsu の MCB)**］では、各平均が平均の最大値または最小値と異なるかどうかを検定します。Hsu の MCB 検定は、最大値と最小値に対して水準ごとに実行されます。

Hsu の MCB 検定では、最小有意差 (LSD) の計算に使われる分位点が水準によって異なるので、比較円のグラフはあまりうまく機能しません。比較円のグラフはあまりうまく機能しません。標本サイズが等しくない限り、検定対象の水準に応じて比較円をスケールし直さなければならないからです。比較円は最大分位点を使って作成されます。正確に有意差を評価したい場合には、グラフではなく、検定の p 値を見てください。

次の点を念頭に置いてください。

- 平均より**上**に有意に離れた平均がある場合、その平均は最大値ではないとみなされます。
- 平均より**下**に有意に離れた平均がある場合、その平均は最小値ではないとみなされます。
- 他のすべての平均より有意に離れて**上**にある平均は、最大値とみなされます。
- 他のすべての平均より有意に離れて**下**にある平均は、最小値とみなされます。

注 : MCB で最大値または最小値とみなされない平均は、Gupta（1965）の部分集合 (最大平均または最小平均をもつ可能性があるとして選択された群の集合) にも含まれません。

最大値の表には、行の平均から列の平均を引き、さらに最小有意差 (LSD) を引いたものが表示されています。正の値がある列の平均は、その正の値である行の平均は列の平均より有意に大きいため、最大値とはみなされません。

最小値の表には、行の平均から列の平均を引き、最小有意差 (LSD) を加算したものが表示されています。負の値がある列の平均は、その負の値である行の平均は列の平均より有意に小さいため、最小値とはみなされません。

関連情報

- 「［最適値との比較 (Hsu の MCB)］の例」（179 ページ）

コントロール群との比較 (Dunnett)

［コントロール群との比較 (Dunnett)］検定では、各群の平均が、コントロール群の平均と比較されます。この検定で比較を行う回数は、Student の *t* 検定と Tukey-Kramer の検定の中間となるため、LSD も Student の *t* 検定と Tukey-Kramer 検定の間の値になります。

Dunnett の検定のレポートに表示される |*d*| 分位点は、Student の *t* 検定と同じように使用できます。LSD 閾値行列には、差の絶対値から LSD を引いた値が表示されます。値が正のときは、その平均がコントロール群の平均より LSD 以上大きく、したがってコントロール群の平均と有意に異なることを意味します。

関連情報

- 「［コントロール群との比較 (Dunnett)］の例」（180 ページ）

「平均の比較」のオプション

「平均の比較」の各検定のレポートには、赤い三角ボタンのメニューがあり、カスタマイズ用のオプションが用意されています。

表 6.10「平均の比較」の各オプションの説明

差の行列	平均の差を、表にまとめて表示します。
棄却限界値	信頼区間に使用される *t* 値、または該当する他の分位点を表示します。
LSD 閾値行列	それぞれの比較における差が最小有意差（LSD; Least Significant Difference）を上回るかどうかを示す行列を表示します。 **注：**Hsu の MCB 検定と Dunnett 検定では、［差の行列］、［棄却限界値］、［LSD 閾値行列］だけが表示されます。
文字の接続レポート	平均の有意差をアルファベットで示した、伝統的なレポートが作成されます。同じ文字でつながっていないグループの平均は、有意に異なります。
差の順位レポート	正の値を取る差とその信頼区間が順番に並べて表示されます。Student の t 検定と Tukey-Kramer の HSD 検定で選択することができ、選択するとレポートの最後に「差の順位」レポートが追加されます。 このレポートは差を大きいものから順に並べたリストです。プロット上には信頼区間を示す帯が重ねて表示されます。差の信頼区間にゼロが含まれていない場合、平均が有意に異なることを示しています。
詳細な比較レポート	各比較の詳細なレポートを表示します。比較ごとに、水準間の差、標準誤差と信頼区間、*t* 値、*p* 値、および自由度が表示されます。検定結果を図示したグラフが、各レポートの右側に表示されます。 このオプションは、［すべてのペア ,Tukey の HSD 検定］と［ノンパラメトリックな多重比較］では使用できません。

ノンパラメトリック

ノンパラメトリックな検定は、グループの平均または中央値がすべてのグループで同じ位置にあるかどうかを検定するのに役立ちます。ただし、通常の分散分析で仮定される正規性が、ここでは仮定されません。ノンパラメトリックな検定では、「順位スコア」と呼ばれる、Y 変数の順位を特定の関数で変換した値が使用されます（Hajek 1969）。

注: ブロック列を指定してある場合は、ノンパラメトリックな検定を行う前に、データがブロック平均に合わせて調整されます。

表 6.11 ノンパラメトリックな検定の説明

Wilcoxon 検定[a]	Wilcoxon の順位スコアに基づく検定を実行します。Wilcoxon の順位スコアは、データの順位そのものです。Wilcoxon の検定は、誤差がロジスティック分布に従っている場合に、最強力の順位検定となります。因子の水準が 3 つ以上の場合は、Kruskal-Wallis 検定が実行されます。 Wilcoxon 検定は Mann-Whitney 検定とも呼ばれます。
メディアン検定	メディアン順位スコアに基づく検定を実行します。メディアン順位スコアは、順位がメディアン（中央値）の順位の上か下かによって 1 または 0 の値を取ります。メディアン検定は、誤差が二重指数分布に従っている場合に、最強力の順位検定となります。
Van der Waerden の検定	Van der Waerden の順位スコアに基づく検定を実行します。Van der Waerden の順位スコアは、データの順位をオブザベーション数+1 で割り、正規分布関数の逆関数を使って正規スコアに変換したものです。Van der Waerden の検定は、誤差が正規分布に従っている場合に、最強力の順位検定となります。
Kolmogorov-Smirnov 検定	経験分布関数（EDF）に基づいて、応答の分布がグループ間で同じかどうかを検定します。近似検定と正確検定の両方が行われます。この検定は、X 変数の水準が 2 つの場合にだけ実行できます。
正確検定	サブメニューが開き、Wilcoxon、メディアン、Van der Waerden、Kolmogorov Smirnov の正確検定を実行するためのオプションが表示されます。これらのオプションは、JMP Pro において、X 変数が 2 水準である時だけに使えます。 JMP PRO 正確検定は JMP Pro でのみ実行できます。

表 6.11 ノンパラメトリックな検定の説明（続き）

ノンパラメトリックな多重比較	サブメニューが開き、ノンパラメトリックな多重比較検定を実行するためのオプションが表示されます。これらは順位を用いた検定で、［ペアごと Wilcoxon 検定］以外では多重比較の調整が行われます。これらの検定の詳細については、Dunn（1964）および Hsu（1996）を参照してください。 **ペアごと Wilcoxon 検定** 多重比較の調整を行わず、各ペアに対して Wilcoxon 検定を実行します。これは、［平均の比較］メニューの［**各ペア ,Student の t 検定**］オプションのノンパラメトリック版です。 **すべてのペア Steel-Dwass 検定** すべてのペアに対して Steel-Dwass 検定を実行します。これは、［平均の比較］メニューの［**すべてのペア ,Tukey の HSD 検定**］オプションのノンパラメトリック版です。 **コントロール群との比較 Steel 検定** 各水準をコントロール水準と比較します。これは、［平均の比較］メニューの［**コントロール群との比較 (Dunnett)**］オプションのノンパラメトリック版です。 **コントロール群との比較 併合順位の Dunn 検定** ［コントロール群との比較 Steel 検定］オプションと同様、各水準をコントロール水準と比較しますが、この Dunn 検定では、比較対象のペアにおける順位でなく、すべてのデータを通じた順位が計算に使われます。表内の p 値には、Bonferroni の調整が施されています。つまり、未調整の p 値に比較の数を掛けたものです。調整済み p 値が 1 を超えた場合は、1 と表示されます。 **すべてのペア 併合順位の Dunn 検定** ［すべてのペア Steel-Dwass 検定］オプションと同様に各ペアを比較しますが、この Dunn 検定では、比較対象のペアにおける順位でなく、すべてのデータを通じた順位が計算に使われます。表内の p 値には、Bonferroni の調整が施されています。つまり、未調整の p 値に比較の数を掛けたものです。調整済み p 値が 1 を超えた場合は、1 と表示されます。

a. Wilcoxon 検定、メディアン検定、van der Waerden の検定では、X 変数に水準が 3 つ以上ある場合、一元配置検定のカイ 2 乗近似が計算されます。X 変数の水準が 2 つの場合は、「カイ 2 乗近似」というレポートのほかに、「正規近似」というレポートも出力されます。

ヒント： 反復測定データなどに対するノンパラメトリックな Friedman 検定は、JMP では直接サポートされていませんが、次のような方法で実行できます。まず、各ブロック内での順位を計算します。そして、「二変量の関係」プラットフォームを使用して、順位に対して分割表の分析を実行します。このとき、ブロック変数を指定してください。

ノンパラメトリック検定のレポート

すべてのノンパラメトリック検定のレポートを、次の表で説明します。

- 表 6.12（165 ページ）
- 表 6.13（166 ページ）
- 表 6.14（166 ページ）

表 6.12 Wilcoxon 検定、メディアン検定、Van der Waerden 検定の説明

水準	因子の水準。
度数	各水準の度数。
スコア和	各水準の順位スコアの和。
スコアの期待値	水準間に差はないという帰無仮説のもとでのスコアの期待値。
スコア平均	各水準の平均順位スコア。
(平均 - 平均 0)/標準偏差 0	標準化したスコア。「平均 0」は帰無仮説のもとでの平均スコアの期待値で、「標準偏差 0」は帰無仮説のもとでの平均スコアの標準偏差です。帰無仮説は、グループの平均または中央値がすべてのグループで同じ位置にあるというものです。
カイ 2 乗	カイ 2 乗検定統計量の値。
自由度	検定の自由度。
p 値（Prob>ChiSq）	検定の p 値。
S	順位スコアの和。X 変数の水準が 2 つの場合に表示されます。
Z	正規近似検定の検定統計量。X 変数の水準が 2 つの場合に表示されます。
p 値 (Prob>|Z|)	正規近似検定の p 値。X 変数の水準が 2 つの場合に表示されます。
p 値 (Prob≥S)	片側検定の p 値。X 変数の水準が 2 つで、正確検定が実行された場合に表示されます。 JMP PRO 正確検定は JMP Pro でのみ実行できます。
p 値 (Prob≥|S-Mean|)	両側検定の p 値。X 変数の水準が 2 つで、正確検定が実行された場合に表示されます。 JMP PRO 正確検定は JMP Pro でのみ実行できます。

表 6.13 Kolmogorov Smirnov 検定の説明

水準	因子の水準。
度数	各水準の度数。
最大時の経験分布関数	標本全体の経験分布関数（EDF）の平均からの乖離が最大となっているときの乖離。
最大時の平均からの乖離	各水準の経験分布関数（EDF）の平均からの乖離が最大であるときの EDF の値。
KS	Kolmogorov-Smirnov 検定の検定統計量。
KSa	漸近的な Kolmogorov-Smirnov 検定の検定統計量。
D=max|F1-F2|	2 群の EDF の差で、その絶対値が最大のもの。
p 値 (Prob > D)	検定に対する p 値。水準間または標本間に差はないという帰無仮説のもとで、D が観測値 d よりも大きくなる確率です。
D+=max(F1-F2)	第1群の EDF から、第2群の EDF を引いた差の最大値。片側検定の検定統計量。
p 値 (Prob > D+)	水準間に差はないという帰無仮説のもとで、D+ が観測値 d+ よりも大きくなる確率。
D-=max(F2-F1)	第2群の EDF から、第1群の EDF を引いた差の最大値。片側検定の検定統計量。
p 値 (Prob > D-)	D- が観測値 d- より大きくなる確率。

表 6.14 ノンパラメトリックな多重比較検定の説明

q*	信頼区間で使用されている分位点の値。
Alpha	信頼区間で使用されている α 水準。
水準	現在の比較で使用されているペア。
スコア平均の差	スコア平均間の差。
差の標準誤差	スコア平均間の差に対する標準誤差。
Z	標準化された検定統計量。帰無仮説のもとで漸近的に標準正規分布に従います。
p 値	検定統計量 Z から計算された両側検定の漸近的な有意確率（p 値）。
Hodges-Lehmann	位置の差に対する Hodges-Lehmann 推定値。2 標本のオブザベーションのすべての組み合わせの差の中央値です。
下側信頼限界	Hodges-Lehmann 推定値の下側信頼限界。
上側信頼限界	Hodges-Lehmann 推定値の上側信頼限界。

等分散性の検定

グループごとの分散が等しくない場合、通常の分散分析が必要とする仮定が満たされないので、通常の分散分析の F 検定は適切ではありません。JMP には、4 種類の等分散性の検定の他に、グループ標本の分散が等しくないときにも妥当な分散分析が用意されています。等分散性の検定のうちの 3 つは、各グループ内での広がりを測定するための新しい Y 変数を作成し、それに対して分散分析を行います。もう 1 つは Bartlett の検定です。Bartlett の検定は、正規分布のもとでの尤度比検定に類似しています。

注: 等分散性の検定には、分散の平均分析（ANOMV）法も使用できます。詳細については、「平均分析法」（154 ページ）を参照してください。

表 6.15 等分散性の検定の説明

O'Brien	グループ平均が元の Y 変数のグループごとの不偏分散と等しくなるような従属変数が新しく作成されます。O'Brien 検定は、グループごとの分散に対する ANOVA であると言えます（O'Brien 1979、Olejnik and Algina 1987）。
Brown-Forsythe	グループの中央値と各観測値との差の絶対値を応答として ANOVA を実行し、F 検定を行います（Brown and Forsythe 1974）。
Levene	グループの平均と各観測値との差の絶対値を応答として ANOVA を実行し、F 検定を行います（Levene 1960）。ばらつきが（SAS のデフォルトが $z_{ij}^2 = (y_{ij} - \bar{y}_i)^2$ であるのとは対照的に）$z_{ij} = \left\lvert y_{ij} - \bar{y}_i \right\rvert$ として計算されます。
Bartlett	標本分散の重み付きの算術平均を、標本分散の重み付きの幾何平均と比べます。幾何平均は必ず算術平均以下になり、算術平均と幾何平均が等しくなるのは、すべての値が等しい場合です。すべての標本分散が等しいと、それらの算術平均と幾何平均は等しくなります。グループ分散の間に見られる変動が大きいほど、2 つの平均の差が大きくなります。2 つの平均から計算された値が、χ^2 分布によって近似されます（χ^2 分布は、ある定式化における F 分布の一種です）。値が大きいということは、算術平均／幾何平均の比が大きい、つまりグループ分散の変動が大きいことを意味します。Bartlett のカイ 2 乗検定統計量を自由度で割ったものが、表の中の F 値です。Bartlett の検定は、正規性の仮定が満たされないケースに対して頑健ではありません（Bartlett and Kendall 1946）。

検定の対象が 2 つのグループだけである場合、等分散性の検定として標準的な F 検定も行われます。F 検定は、小さい方の分散値に対する大きい方の分散推定値の比です。両側検定にする目的で、F 分布の p 値に 2 が掛けられます。

注: ブロック列を指定した場合は、分散検定を行う前に、データがブロック平均に合わせて調整されます。

「分散が等しいことを調べる検定」レポート

「分散が等しいことを調べる検定」レポートには、標準偏差、グループの平均と各観測値との差の絶対値を平均した値、グループの中央値と各観測値との差の絶対値を平均した値、および検定の要約が表示されます

等分散性の検定からグループの分散が有意に異なることがわかったときは、通常の ANOVA 検定ではなく、Welch の検定を使用した方がよいでしょう。Welch の統計量は、通常の ANOVA の *F* 検定を基にしていますが、グループごとの不偏分散の逆数で加重された平均が計算に使われています（Welch 1951; Brown and Forsythe 1974b; Asiribo, Osebekwin and Gurland 1990）。水準が 2 つしかないとき、Welch の一元配置分散分析は分散が等しくない場合の *t* 検定と等価です。

表 6.16「分散が等しいことを調べる検定」レポートの説明

水準	因子の水準。
度数	各水準の度数。
標準偏差	X 変数の水準ごとに応答の標準偏差を表示したもの。O'Brien 検定における変換された変数のグループ平均は、標準偏差の 2 乗（分散）です。データが 1 つしかない水準については、標準偏差は計算されません。
平均からの平均絶対偏差	グループ平均からの Y の偏差の絶対値を平均した値。Levene 検定における変換された変数のグループ平均は、この平均絶対偏差になります。
中央値からの平均絶対偏差	グループ中央値からの Y の偏差の絶対値を平均した値。Brown-Forsythe 検定における変換された変数のグループ平均は、この平均絶対偏差になります。
検定	実行された検定の種類。
F 値	検定ごとの方法で計算された *F* 統計量。 詳細については、「「分散が等しいことを調べる検定」レポートの統計的詳細」（194 ページ）を参照してください。
分子自由度	各検定の分子で使用される自由度。因子の水準数を *k* とすると、分子には *k*-1 個の自由度があります。O'Brien、Brown-Forsythe、および Levene の統計量の計算では、データが 1 つしかない水準は無視されます。その場合の分子自由度は、計算に使用された水準数から 1 を引いた数だけあります。
分母自由度	各検定の分母で使用される自由度。O'Brien、Brown-Forsythe、および Levene では、検定統計量の計算に使用された因子水準の数だけ自由度が引かれます。因子の水準数を k とすると、分母自由度は *n* - *k* となります。
p 値	分散がすべての水準にわたって等しい場合に、まったくの偶然だけで現在の *F* 値より大きな *F* 値を得る確率。

表 6.17「Welch の検定」レポートの説明

F 値	等分散性の検定の F 統計量。 詳細については、「「分散が等しいことを調べる検定」レポートの統計的詳細」（194 ページ）を参照してください。
分子自由度	検定の分子で使用される自由度。因子の水準数を *k* とすると、分子には *k*-1 個の自由度があります。データ内で一度しか起こらなかった水準は、Welch の一元配置の計算に使用されません。その場合の分子自由度は、計算に使用された水準数から 1 を引いた数だけあります。
分母自由度	検定の分母で使用される自由度。 詳細については、「「分散が等しいことを調べる検定」レポートの統計的詳細」（194 ページ）を参照してください。
p 値 (Prob>F)	平均がすべての水準にわたって等しい場合に、まったくの偶然だけで現在の *F* 値より大きな *F* 値を得る確率。有意確率が 0.05 以下の場合には、いくつかの平均が他の平均と異なっているという証拠があるとみなされます。
t 検定	F 値と t 値の関係を示すために出力されます。t 値は、F 値の平方根です。この結果は、**X** 変数の水準が 2 つしかない時にだけ表示されます。

関連情報

- 「等分散性検定の例」（182 ページ）
- 「「分散が等しいことを調べる検定」レポートの統計的詳細」（194 ページ）

同等性の検定

同等性の検定では、実質的な差が平均にあるかどうかを検定します。差に対して閾値を設定し、差がそれより小さい場合は、平均は実質的に等しいと結論付けます。同等性検定の最も単純なものは、実質的に同等とみなす区間の上限と下限のそれぞれで、1 回ずつ片側 *t* 検定を行う方法です。2 つの片側検定で帰無仮説が棄却されれば（平均間の差が、下限以上で、かつ、上限以下であると検定により結論できれば）、2 つのグループの平均は実質的に等しいと主張できます。[**同等性の検定**] オプションは、このような Two One-Sided Tests（TOST; 2 つの片側検定）アプローチを採用しています。

関連情報

- 「同等性の検定の例」（183 ページ）

ロバスト

データに外れ値があると、ロバストではない方法を用いると、推定や決定が間違ったものになってしまう可能性があります。[**ロバスト**] オプションには、外れ値の影響を受けにくい2つの手法、[ロバストなあてはめ] と [Cauchy のあてはめ] があります。

ロバストなあてはめ

[ロバストなあてはめ] オプションで用意されている手法は、応答変数の外れ値が推定結果に与える影響が少ないものです。Huber の M 推定が使用されます。Huber の M 推定では、Huber の損失関数を最小にするパラメータ推定値を見つけます。この推定方法では、外れ値に対して、ある種の罰則（ペナルティ）が課せられます。Huber の損失関数は、誤差が小さい場合は 2 次曲線として増加し、誤差が大きい場合は線形に増加します。ロバストなあてはめの詳細については、Huber（1973）および Huber and Ronchetti（2009）を参照してください。

関連情報

- 「[ロバストなあてはめ] オプションの例」（184 ページ）

Cauchy のあてはめ

このオプションでは、誤差が Cauchy 分布に従うと仮定されます。Cauchy 分布は正規分布よりも裾が広く、その結果、外れ値が推定に与える影響が小さくなります。このオプションは、データにある外れ値の割合が大きい場合に有用です。しかし、データが正規分布に近く、外れ値が少ない場合は、このオプションの推定結果は間違ったものになる可能性があります。[Cauchy のあてはめ] オプションは、最尤推定によってパラメータ推定値を算出します。

検出力

[**検出力**] オプションを選択すると、特定の仮説検定に対して、検出力に関連した結果が表示されます

- **最小有意値**（LSV: Least Significant Value）は、*p* 値が特定の値になるようなパラメータ (もしくはパラメータ関数）の値です。つまり、ある値以下に *p* 値がなって有意と判断される最小の効果の大きさを表します。最小有意値（LSV）は、確率のスケールではなく、パラメータのスケール上で有意度を測定していると言うことができます。その値によって、計画とデータがどのぐらい敏感かがわかります。
- **最小有意数**（LSN: Least Significant Number）は、*p* 値が特定の値になるような標本サイズです。つまり、Alpha（有意水準）、Sigma（誤差の標準偏差）、Delta（効果の大きさ）が与えられているときに、結果が有意となるほど推定値の分散が減少するために必要な標本サイズが、最小有意数（LSN）です。有意性を得るためにさらに多くのデータが必要な場合は、最小有意数（LSN）を参考にすれば、どれだけのデータが必要かがわかります。最小有意数（LSN）の値だけオブザベーションがあるとき、検出力は約 50% になります。

- **検出力**とは、グループ間に実際に差がある場合に、検定が有意となる確率（*p* 値 < α となる確率）を指します。これは、標本サイズ、効果の大きさ、誤差の標準偏差、および有意水準の関数です。検出力からは、その実験で、与えられた有意水準において差（効果の大きさ）をどれだけ的確に検出できるかがわかります。

注: 一元配置でグループが2つしかないときは、「検出力の詳細」で計算された最小有意値（LSV）の値が「多重比較」表の最小有意差（LSD）と等しくなります。

「検出力の詳細」ウィンドウとレポート

「検出力の詳細」ウィンドウとレポートは、「モデルのあてはめ」プラットフォームで実行される一般的な検出力計算と同じです。検出力の計算の詳細については、『基本的な回帰モデル』を参照してください。

「α」、「σ」、「δ」、「数」の各列には値を1～2つ入力するか、連続した値の範囲として開始値、終了値、および間隔を入力します。詳細については、図6.26を参照してください。入力した値のすべての組み合わせに対して、検出力が計算されます。

表6.18「検出力の詳細」レポートの説明

α	0～1の値を取る有意水準（通常は0.05、0.01、0.10のどれかを使用します）。デフォルトでは0.05に設定されています。
σ	モデル内の残差の標準誤差（誤差の標準偏差に対する推定値）。デフォルトとして、誤差の平均平方の平方根から推定されたRMSEが表示されています。
δ	標準化していない効果の大きさ（effect size; 効果量）。効果の大きさを計算する方法については、『基本的な回帰モデル』を参照してください。最初のフィールドには仮説の平方和を *n* で割ったものの平方根、つまり $\delta = \sqrt{SS/n}$ が表示されます。
数（*n*）	すべてのグループの標本サイズの合計。デフォルトでは、最初のフィールドに現在の標本サイズが表示されます。
検出力を求める	「α」、「σ」、「δ」、および「数」の関数として、検出力（有意な結果が出る確率）が計算されます。
最小有意数を求める	指定された「α」、「σ」、「δ」で、おおよそ50%の検出力を得るために必要なオブザベーションの個数。
最小有意値を求める	*p* 値がα以下となるパラメータまたは線形検定の値。これは「α」、「σ」、「数」、および推定値の標準誤差の関数です。この機能は、X変数の水準数が2つの場合にだけ使用でき、通常は個々のパラメータに適用します。

表 6.18「検出力の詳細」レポートの説明（**続き**）

調整済み検出力と信頼区間	事後的に検出力を調べるときは、誤差の標準偏差と検定するパラメータに対して推定値を使用します • 推定値を単純に代入してF分布の非心度パラメータを計算するとバイアスが生じるため、バイアスの少ない推定値を使って計算したものが、調整済み検出力です。 • また、調整済み検出力の信頼区間は、非心度推定値の信頼区間に基づいています。 調整済み検出力と信頼限界が計算されるのは、元の「δ」の推定値に対してだけで、δにその他の値を指定して計算することはできません（調整済み検出力は、δの推定値におけるランダムな変動を調整したものです）。

関連情報

- 「[検出力] オプションの例」（186 ページ）
- 「検出力の統計的詳細」（193 ページ）

正規分位点プロット

2 種類の正規分位点プロットを作成できます。

- [**分位点 - 実測値プロット**] を選択すると、X 軸が正規分位点で、Y 軸が応答変数の実測値であるプロットが作成されます。正規分位点は、X 変数の水準ごとに計算されます。
- [**実測値 - 分位点プロット**] を選択すると、X 軸が応答変数の実測値で、X 軸が正規分位点であるプロットが作成されます。正規分位点は、X 変数の水準ごとに計算されます。

[**あてはめ線**] オプションを選択すると、分位点プロット上のあてはめ線の表示／非表示が切り替わります。

関連情報

- 「正規分位点プロットの例」（187 ページ）

累積確率プロット

累積確率プロットには、「一元配置」レポートのすべてのグループの累積分布関数が表示されます。X 変数の水準に対する応答変数の分布を比較する場合に役立ちます。

関連情報

- 「累積確率プロットの例」（188 ページ）

密度

［**密度**］オプションで描かれるグラフは、応答変数の分布や構成を、X変数の各水準で比較します。3つのオプションがあります。

- ［**密度の比較**］を選択すると、各グループの密度推定値を表す滑らかな曲線が表示されます。
- ［**密度の構成**］を選択すると、グループの度数で重み付けした密度の合計がグラフで表示されます。「密度の構成」プロットでは、応答変数の各値において、全体の密度が、どのように各グループによって構成されているかが分かります。
- ［**密度の割合**］を選択すると、応答変数の各値において、全体の密度に占める各グループの割合がグラフで表示されます。

関連情報

- 「［密度］の各オプションの例」（189 ページ）

対応のある列を設定

［**対応のある列を設定**］オプションでは、対応がある場合の分析を行うために、対応付けのための変数（ID変数）を指定します。［**対応のある列を設定**］オプションは、異なるグループのデータが同じ個体から取られた場合など、データに対応がある場合に使用します。

注： この分析の特殊なケースが、対応のある *t* 検定です。**「対応のあるペア」**プラットフォームでも対応のあるデータを分析できます。ただし、ペアとなっているデータが異なる行ではなく、異なる列に入っている必要があります。

［**対応のある列を設定**］オプションでは、主に2つの処理が行われます。

- グループ変数（「二変量の関係」におけるX変数）と対応付けのための変数の両方を含んだ加法モデルが（反復比例あてはめアルゴリズムを使用して）あてはめられます。反復比例あてはめアルゴリズムは個体が何百もあるときに非常に有効です。個体数が多い場合に、線形モデルを普通にあてはめると、計算に時間がかかり、膨大なメモリが必要になります。
- グループにわたって対応する点が線でつながれます。同じIDを持つオブザベーションが複数ある場合は、それらのオブザベーションの平均から線が引かれます。

［**対応のある列を設定**］オプションを選択すると、自動的に［**対応を表す線**］オプションも選択され、対応する点が線でつながれます。この線の表示／非表示を切り替えるには、［**表示オプション**］**>**［**対応を表す線**］を選択します。

「対応のあるあてはめ」レポートには、効果とその *F* 検定が表示されています。これは、「モデルのあてはめ」プラットフォームで交互作用項のあるモデルとないモデルの2つを実行したときに計算される検定と同じです。水準が2つしかないときは、*F* 検定は対応のある *t* 検定と等しくなります。

注:「モデルのあてはめ」プラットフォームの詳細については、『基本的な回帰モデル』を参照してください。

関連情報

- 「[対応のある列を設定] オプションの例」(190 ページ)

「一元配置」プラットフォームのその他の例

ここでは、「一元配置」プラットフォームの例をさらに紹介します。

平均分析法の例

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Analgesics.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「痛みの程度」を選択し、[Y, 目的変数] をクリックします。
4. 「薬」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、メニューから [平均分析法] > [平均分析] を選択します。

図 6.13「平均分析」チャートの例

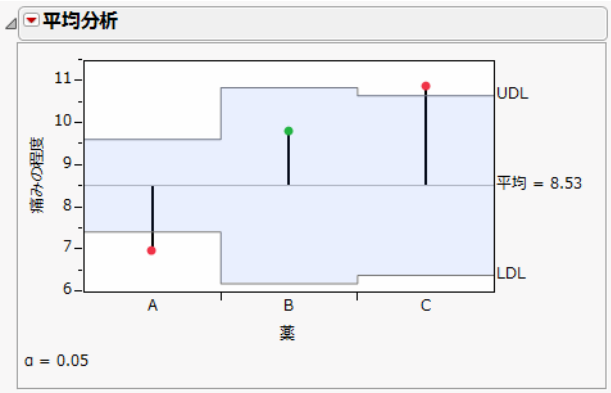


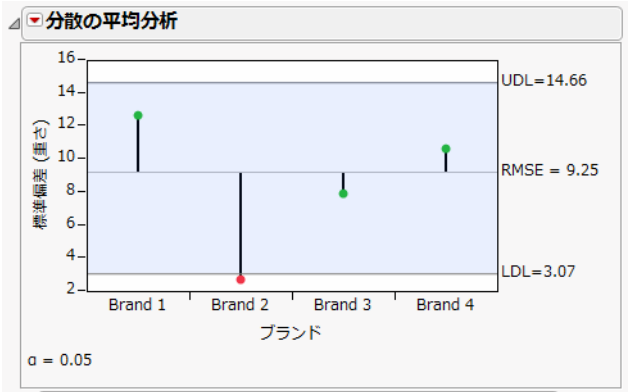
図 6.13 の例では、薬 A と C の平均は、全体平均と有意に異なります。薬 A の平均は低く、薬 C の平均は高くなっています。各薬で標本サイズが異なっているため、薬ごとに決定限界が違っています。

分散に対する平均分析法の例

この例では、「Spring Data.jmp」サンプルデータを使用します。4つのブランドのバネ製品について、0.10インチ伸ばすのに必要な重量を調べました。ブランドごとに6製品ずつ調べました。ANOMV（ANalysis Of Means for Variance; 分散の平均分析）の統計的検定は、非正規分布に対して頑健ではありません。そこで、データが正規分布に従っていることが確認されました。ANOMVによって、ブランド間で分散が有意に異なるかどうかを調べます。

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Spring Data.jmp」を開きます。
2. ［分析］>［二変量の関係］を選びます。
3. 「重さ」を選択し、［Y, 目的変数］をクリックします。
4. 「ブランド」を選択し、［X, 説明変数］をクリックします。
5. ［OK］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［平均分析法］>［分散の平均分析］を選択します。
7. 「分散の平均分析」の横にある赤い三角ボタンをクリックし、メニューから［要約レポートの表示］を選択します。

図6.14「分散の平均分析」チャートの例



分散の平均分析の要約

水準	N	下側限界	グループ標準偏差	上側限界	限界外
Brand 1	6	3.07047	12.6248	14.66183	
Brand 2	6	3.07047	2.666583	14.66183	下側
Brand 3	6	3.07047	7.91547	14.66183	
Brand 4	6	3.07047	10.63982	14.66183	

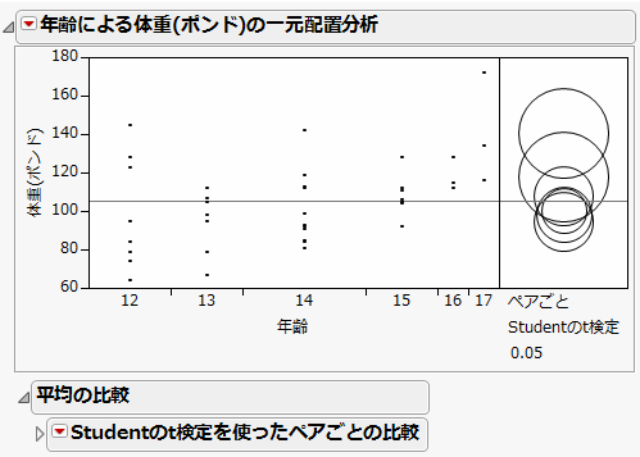
図6.14から、「Brand 2」の標準偏差が下側決定限界を下回っていることがわかります。つまり、「Brand 2」は、他のブランドと比較して、分散が有意に小さいことを示しています。

［各ペア ,Student の t 検定］の例

この例では、「Big Class.jmp」サンプルデータテーブルを使用して、「年齢」を因子、「体重 (ポンド)」を応答変数とした一元配置分析を行います。そして、各ペアの平均を、t 検定で比較します。

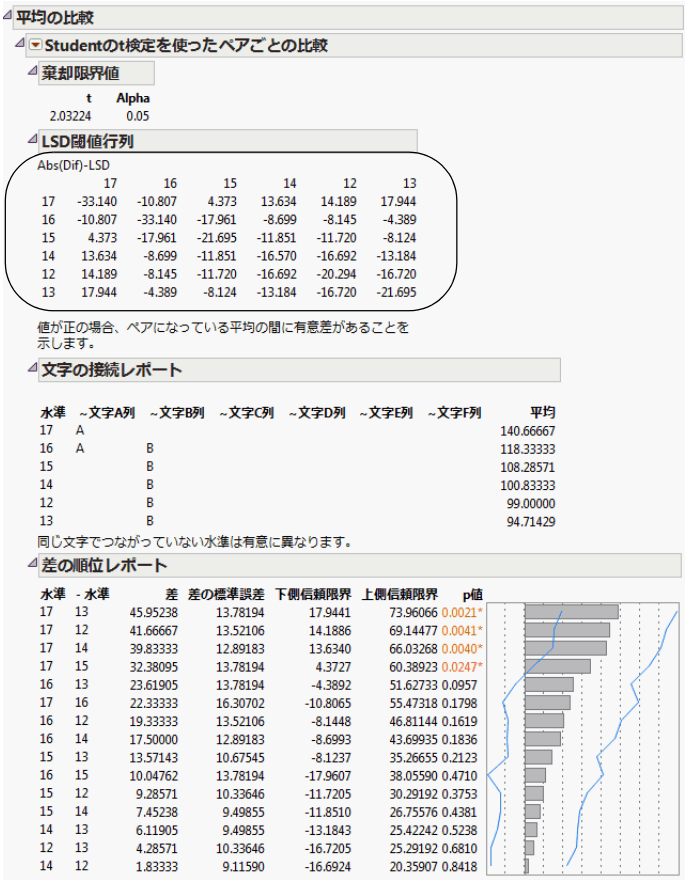
1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**体重 (ポンド)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**年齢**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**平均の比較**］>［**各ペア ,Student の t 検定**］を選択します。

図 6.15 ［各ペア ,Student の t 検定］の比較円の例



平均の差に対する検定が統計的に有意になるのは、観測された平均の差が、検定が有意となる差よりも大きい場合です。この「検定が有意となる差」を、LSD（Least Significant Diference; 最小有意差）と呼びます。LSD は Student の *t* 検定だけでなく、他の検定でも同じような働きをします。比較円のグラフでは、2 つの円の中心間の距離が、観測された平均の差を表します。一方、LSD は、2 つの円の交角が直角となった場合の中心間の距離です。

図 6.16「Student の *t* 検定を使ったペアごとの比較」レポートの例



平均の比較

Studentのt検定を使ったペアごとの比較

棄却限界値

t	Alpha
2.03224	0.05

LSD閾値行列

Abs(Dif)-LSD

	17	16	15	14	12	13
17	-33.140	-10.807	4.373	13.634	14.189	17.944
16	-10.807	-33.140	-17.961	-8.699	-8.145	-4.389
15	4.373	-17.961	-21.695	-11.851	-11.720	-8.124
14	13.634	-8.699	-11.851	-16.570	-16.692	-13.184
12	14.189	-8.145	-11.720	-16.692	-20.294	-16.720
13	17.944	-4.389	-8.124	-13.184	-16.720	-21.695

値が正の場合、ペアになっている平均の間に有意差があることを示します。

文字の接続レポート

水準	~文字A列	~文字B列	~文字C列	~文字D列	~文字E列	~文字F列	平均
17	A						140.66667
16	A	B					118.33333
15		B					108.28571
14		B					100.83333
12		B					99.00000
13		B					94.71429

同じ文字でつながっていない水準は有意に異なります。

差の順位レポート

水準	- 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
17	13	45.95238	13.78194	17.9441	73.96066	0.0021*
17	12	41.66667	13.52106	14.1886	69.14477	0.0041*
17	14	39.83333	12.89183	13.6340	66.03268	0.0040*
17	15	32.38095	13.78194	4.3727	60.38923	0.0247*
16	13	23.61905	13.78194	-4.3892	51.62733	0.0957
17	16	22.33333	16.30702	-10.8065	55.47318	0.1798
16	12	19.33333	13.52106	-8.1448	46.81144	0.1619
16	14	17.50000	12.89183	-8.6993	43.69935	0.1836
15	13	13.57143	10.67545	-8.1237	35.26655	0.2123
16	15	10.04762	13.78194	-17.9607	38.05590	0.4710
15	12	9.28571	10.33646	-11.7205	30.29192	0.3753
15	14	7.45238	9.49855	-11.8510	26.75576	0.4381
14	13	6.11905	9.49855	-13.1843	25.42242	0.5238
12	13	4.28571	10.33646	-16.7205	25.29192	0.6810
14	12	1.83333	9.11590	-16.6924	20.35907	0.8418

図 6.16 の先頭に表示されている行列は、観測された平均の差の絶対値から、LSD（最小有意差；検定が有意となる差）を引いたものです。値が正の場合、観測された平均の差が LSD よりも大きいことを意味していますので、2 つの平均には有意差があると結論できます。

［すべてのペア ,Tukey の HSD 検定］の例

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**体重 (ポンド)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**年齢**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**平均の比較**］>［**すべてのペア ,Tukey の HSD 検定**］を選択します。

図 6.17［すべてのペア ,Tukey の HSD 検定］の比較円の例

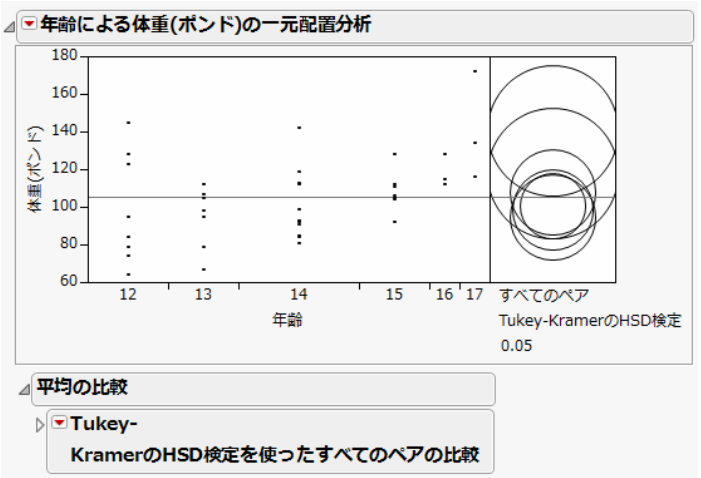


図 6.18「Tukey-Kramer の HSD 検定を使ったすべてのペアの比較」レポートの例

平均の比較

Tukey-KramerのHSD検定を使ったすべてのペアの比較

棄却限界値

q*	Alpha
3.01825	0.05

LSD閾値行列

Abs(Dif)-HSD

	17	16	15	14	12	13
17	-49.219	-26.885	-9.216	0.923	0.857	4.355
16	-26.885	-49.219	-31.550	-21.411	-21.477	-17.978
15	-9.216	-31.550	-32.221	-21.217	-21.912	-18.650
14	0.923	-21.411	-21.217	-24.609	-25.681	-22.550
12	0.857	-21.477	-21.912	-25.681	-30.140	-26.912
13	4.355	-17.978	-18.650	-22.550	-26.912	-32.221

値が正の場合、ペアになっている平均の間に有意差があることを示します。

文字の接続レポート

水準	~文字A列	~文字B列	~文字C列	~文字D列	~文字E列	~文字F列	平均
17	A						140.66667
16	A	B					118.33333
15	A	B					108.28571
14		B					100.83333
12		B					99.00000
13		B					94.71429

同じ文字でつながっていない水準は有意に異なります。

差の順位レポート

水準	- 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
17	13	45.95238	13.78194	4.3551	87.54968	0.0233*
17	12	41.66667	13.52106	0.8568	82.47657	0.0431*
17	14	39.83333	12.89183	0.9226	78.74404	0.0423*
17	15	32.38095	13.78194	-9.2163	73.97825	0.2029
16	13	23.61905	13.78194	-17.9783	65.21635	0.5325
17	16	22.33333	16.30702	-26.8853	71.55192	0.7443
16	12	19.33333	13.52106	-21.4766	60.14323	0.7091
16	14	17.50000	12.89183	-21.4107	56.41071	0.7512
15	13	13.57143	10.67545	-18.6497	45.79256	0.7982
16	15	10.04762	13.78194	-31.5497	51.64492	0.9769
15	12	9.28571	10.33646	-21.9123	40.48369	0.9442
15	14	7.45238	9.49855	-21.2166	36.12134	0.9683
14	13	6.11905	9.49855	-22.5499	34.78801	0.9867
12	13	4.28571	10.33646	-26.9123	35.48369	0.9983
14	12	1.83333	9.11590	-25.6807	29.34736	1.0000

図 6.18 の先頭に表示されている行列は、観測された平均の差の絶対値から、HSD（Tukey 検定における最小有意差）を引いたものです。値が正の場合、観測された平均の差が HSD よりも大きいことを意味していますので、2 つの平均には有意差があると結論できます。q*（レポートの先頭に表示されている値）は、Tukey 検定の HSD を計算するのに使用される分位点です。計算上では、ペアごとの比較における、Student の *t* 統計量と同じような役割を果たします。

［最適値との比較 (Hsu の MCB)］の例

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**体重 (ポンド)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**年齢**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**平均の比較**］>［**最適値との比較 (Hsu の MCB)**］を選択します。

図 6.19 ［最適値との比較 (Hsu の MCB)］の比較円の例

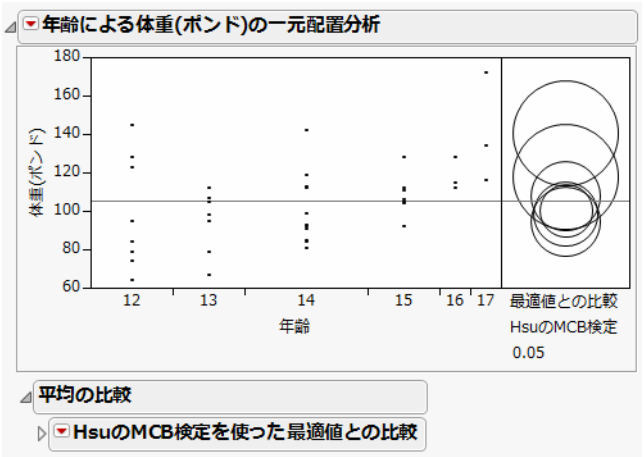


図 6.20「Hsu の MCB 検定を使った最適値との比較」レポートの例

HsuのMCB検定を使った最適値との比較

棄却限界値

d	Alpha
2.23157	0.05
2.23157	
2.34056	
2.38159	
2.35278	
2.34056	

最小/最大との比較

水準	p値(vs. 最大)	p値(vs. 最小)
17	.	0.0039*
16	0.2162	0.1268
15	0.0441*	0.2978
14	0.0075*	0.6416
12	0.0080*	0.7071
13	0.0036*	0.9365

LSD閾値行列

平均[i]-平均[j]-LSD

	17	16	15	14	12	13
17	-36.390	-14.057	0.124	9.130	9.855	13.695
16	-58.724	-36.390	-22.210	-13.203	-12.479	-8.638
15	-63.136	-40.803	-24.987	-15.169	-15.034	-11.415
14	-68.602	-46.269	-29.684	-19.418	-19.614	-16.113
12	-71.840	-49.507	-33.479	-23.544	-23.495	-19.907
13	-76.708	-54.374	-38.558	-28.741	-28.605	-24.987

列内に正の値がある場合、平均が最大値より有意に小さいことを示します。

平均[i]-平均[j]+LSD

	17	16	15	14	12	13
17	36.390	58.724	64.638	70.536	73.479	78.210
16	14.057	36.390	42.305	48.203	51.145	55.876
15	-1.626	20.708	24.987	30.074	33.605	38.558
14	-11.064	11.269	14.780	19.418	23.281	28.351
12	-11.493	10.840	14.907	19.877	23.495	28.479
13	-15.197	7.136	11.415	16.503	20.034	24.987

列内に負の値がある場合、平均が最小値より有意に大きいことを示します。

「最小/最大との比較」レポートは、各水準の平均を、その水準以外での最小平均、および、最大平均と比較したものです。たとえば、最小平均をもつ 13 才と有意に異なる水準は、17 才だけであることがわかります。

［コントロール群との比較 (Dunnett)］の例

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**体重 (ポンド)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**年齢**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**平均の比較**］>［**コントロール群との比較 (Dunnett)**］を選択します。
7. コントロール群として使用するグループを選択します。この例では、年齢「12」を選択します。

   または、行をクリックして散布図上で該当する点を強調表示してから、［**平均の比較**］>［［**コントロール群との比較 (Dunnett)**］オプションを選択します。選択した行がコントロール群として使用されます。

8. ［**OK**］をクリックします。

図 6.21［コントロール群との比較(Dunnett)］の比較円の例

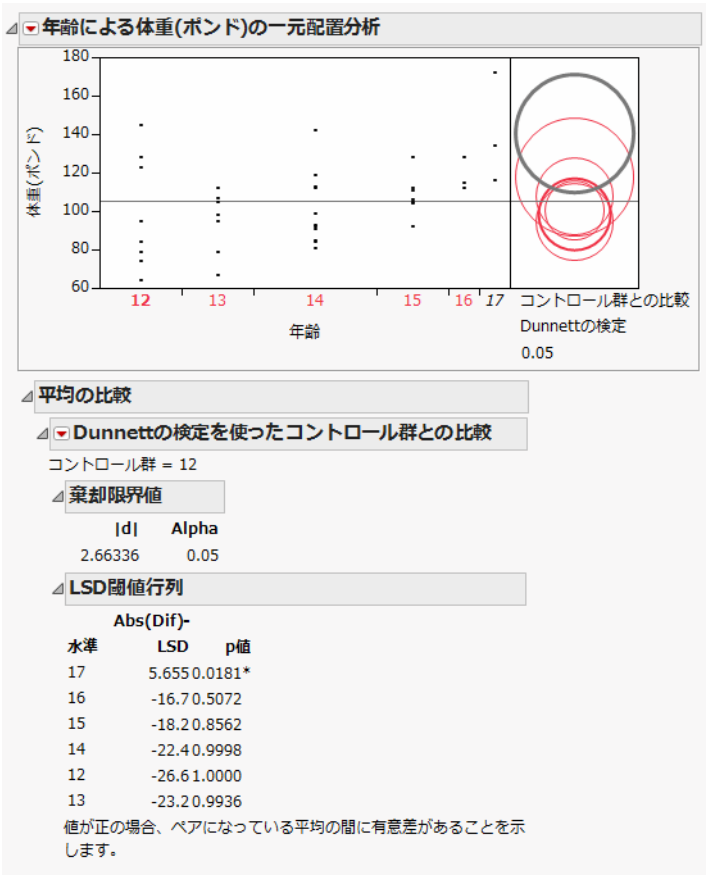


図 6.21 の比較円を見ると、水準 17 だけが、コントロール群である水準 12 と有意に異なることがわかります。

「平均の比較」で用意されているすべての検定

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**体重 (ポンド)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**年齢**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、［**平均の比較**］サブメニューから各オプションを選択します。

4種類の検定はどれもグループ平均の差を検定するものですが、結果は異なります。図6.22は、4つの検定の比較円を示しています。この例のDunnett検定では、「年齢」が17のグループをコントロール群としています。

図6.22 4つの多重比較検定の比較円

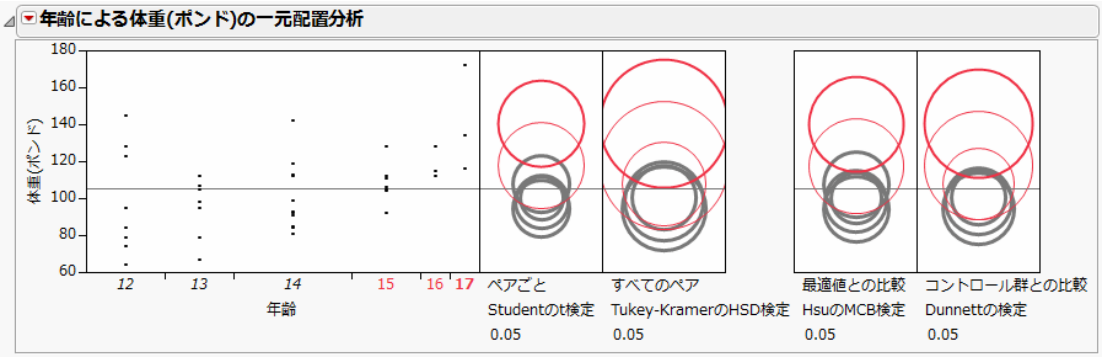


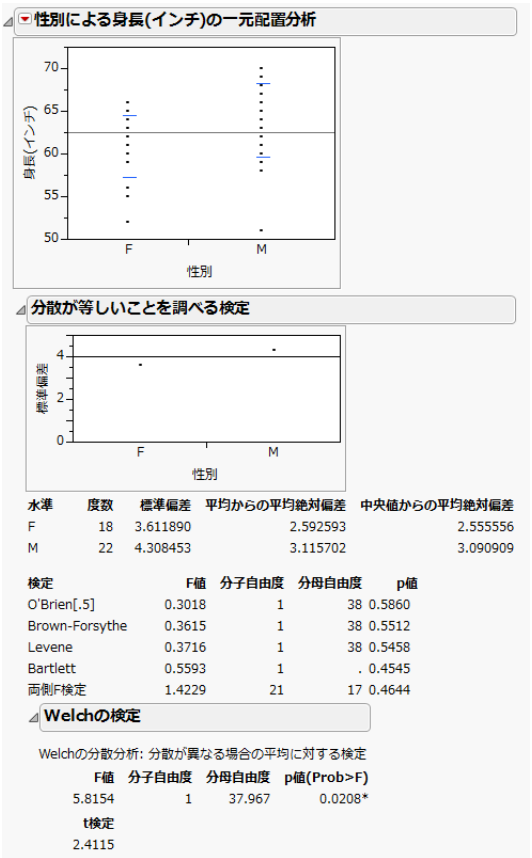
図6.22のStudentのt検定とHsu検定では、「年齢」が15才のグループ（上から3番目の円）が、17才のグループと有意に異なり、グレーで表示されています。しかし、Tukey検定とDunnett検定では、15才と17才には有意差が認められず、赤で表示されています。

等分散性検定の例

この例では、分析者は、男性と女性の身長に関して、平均ではなく、分散の比較に興味があるとします。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**身長 (インチ)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**性別**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**等分散性の検定**］を選択します。

図 6.23「分散が等しいことを調べる検定」レポートの例



両側 F 検定の p 値が大きいことから、「2 つの分散は等しくない」とは結論できません。

同等性の検定の例

この例では、「Big Class.jmp」サンプルデータテーブルを使用して、男性と女性の身長差が 6 インチ未満かどうかを調べます。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**身長 (インチ)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**性別**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**同等性の検定**］を選択します。
7. 「実質的に 0 とみなす差」に「6」と入力します。

8. ［**OK**］をクリックします。

図 6.24 同等性の検定の例

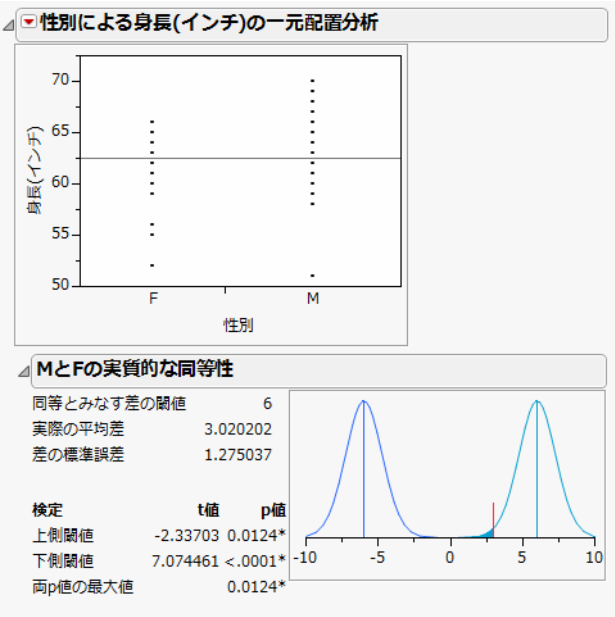


図 6.24 のような正規分位点プロットが描かれます。

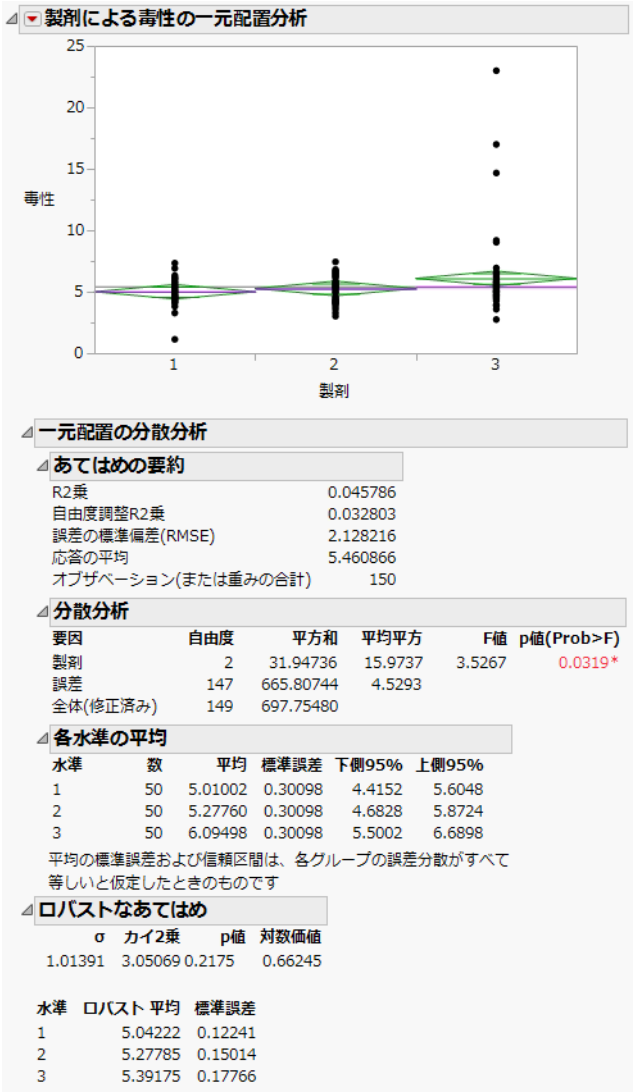
- 「上側閾値」の検定は、帰無仮説における平均の差を 6 としたときの下側の片側検定です。
- 「下側閾値」の検定は、帰無仮説における平均の差を -6 としたとの上側の片側検定です。
- 2 つの片側検定の両方で p 値が小さいため、母平均の差は、-6 から 6 までの間にあると結論付けることができます。したがって、平均は実質的に等しいということができます。

［ロバストなあてはめ］オプションの例

「Drug Toxicity.jmp」サンプルデータテーブルには、3 種類の製剤における毒性のデータが記録されています。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Drug Toxicity.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**毒性**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**製剤**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンのメニューから、［**平均/ANOVA**］を選択します。
7. 赤い三角ボタンのメニューから、［**ロバスト**］>［**ロバストなあてはめ**］を選択します。

図6.25 ロバストなあてはめの例

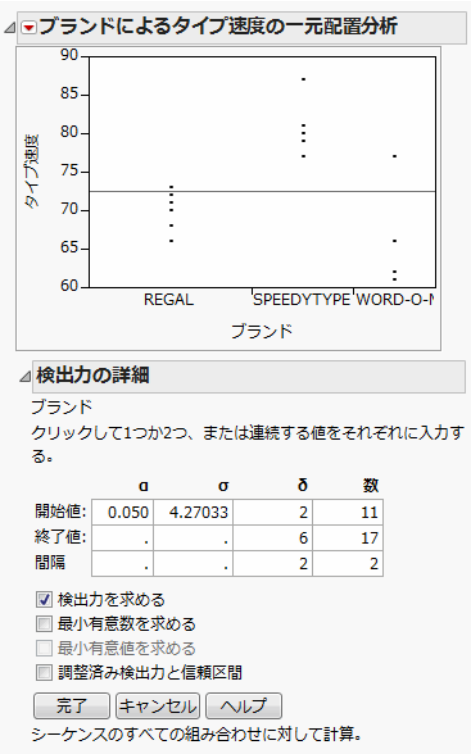


標準の「分散分析」レポートを見ると、p 値が 0.0319 であるため、3 つの製剤は有意に異なるという誤った結論に導かれてしまう可能性がありますが、「ロバストなあてはめ」レポートを確認すれば、p 値が 0.21755 になっているため、3 つの製剤に有意差があるという誤った結論には行きつかないでしょう。いくつかのオブザベーションで毒性が不自然に高く記録され、データに思わしくない影響を生んでいることが伺えます。

［検出力］オプションの例

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Typing Data.jmp」を開きます。
2. ［分析］>［二変量の関係］を選びます。
3. 「タイプ速度」を選択し、［Y, 目的変数］をクリックします。
4. 「ブランド」を選択し、［X, 説明変数］をクリックします。
5. ［OK］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［検出力］を選択します。
7. 「開始値」行の「δ」（3 番目のボックス）に「2」、「数」に「11」と入力します。
8. 「終了値」行の「δ」に「6」、「数」に「17」と入力します。
9. 「間隔」行の「δ」と「数」に「2」と入力します。
10. ［検出力を求める］チェックボックスをオンにします。

図 6.26「検出力の詳細」ウィンドウの例



11. ［完了］をクリックします。

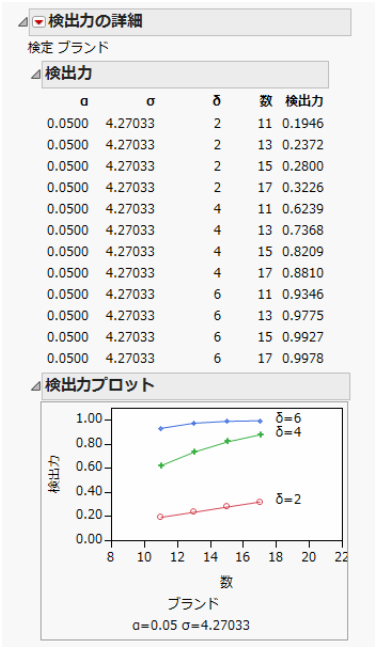
注：［完了］ボタンは、必要なオプションがすべて適用されるまでグレー表示のままです。

「δ」と「数」の組み合わせごとに検出力が計算され、「検出力」レポートに表示されます。

検出力の値をプロットにするには、次の手順に従います。

12. レポートの下部にある赤い三角ボタンをクリックし、メニューから［**検出力プロット**］を選択します。

図 6.27「検出力」レポートの例



13. プロット上にすべてのデータが表示されていない場合は、「検出力」軸をクリックし、縦にドラッグする必要があります。

検出力は「δ」と「数」の組み合わせごとにプロットされています。一般的な性質として、「数」（標本サイズ）と「δ」（平均の差）が大きくなるほど検出力は上がりますが、この例でもその性質を確認することができます。

正規分位点プロットの例

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**身長 (インチ)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**性別**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**正規分位点プロット**］>［**分位点 - 実測値プロット**］を選択します。

図 6.28 正規分位点プロットの例

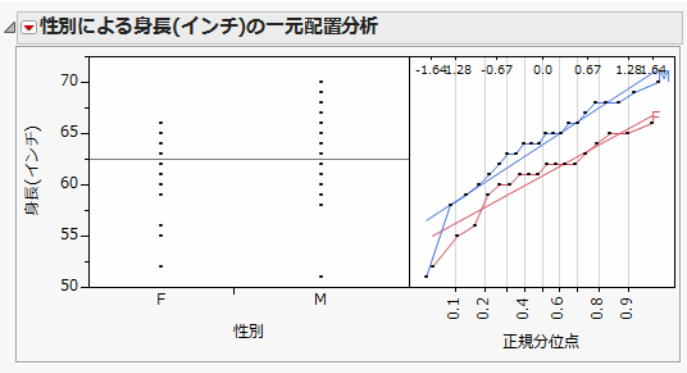


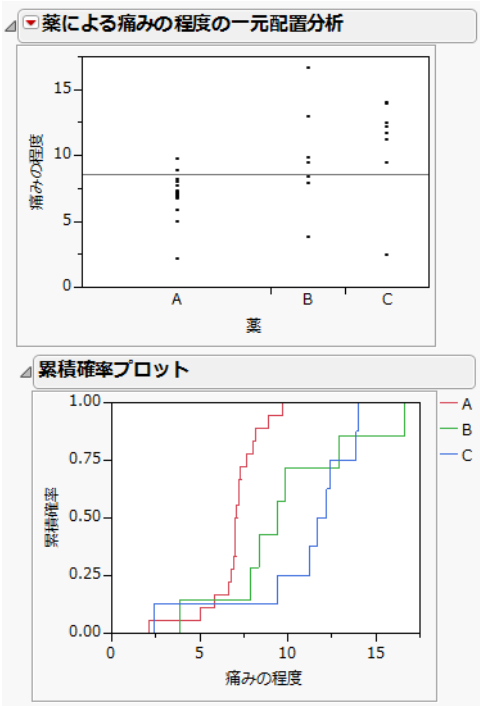
図 6.28 のような正規分位点プロットが描かれます。

- あてはめ線はデフォルトで表示されます。
- データ点はあてはめ線のごく近くに分布しています。つまり、データは正規分布に従っていることを示唆しています。

累積確率プロットの例

1. ［**ヘルプ**］ > ［**サンプルデータライブラリ**］を選択し、「Analgesics.jmp」を開きます。
2. ［**分析**］ > ［**二変量の関係**］を選びます。
3. 「**痛みの程度**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**薬**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**累積確率プロット**］を選択します。

図 6.29 累積確率プロットの例

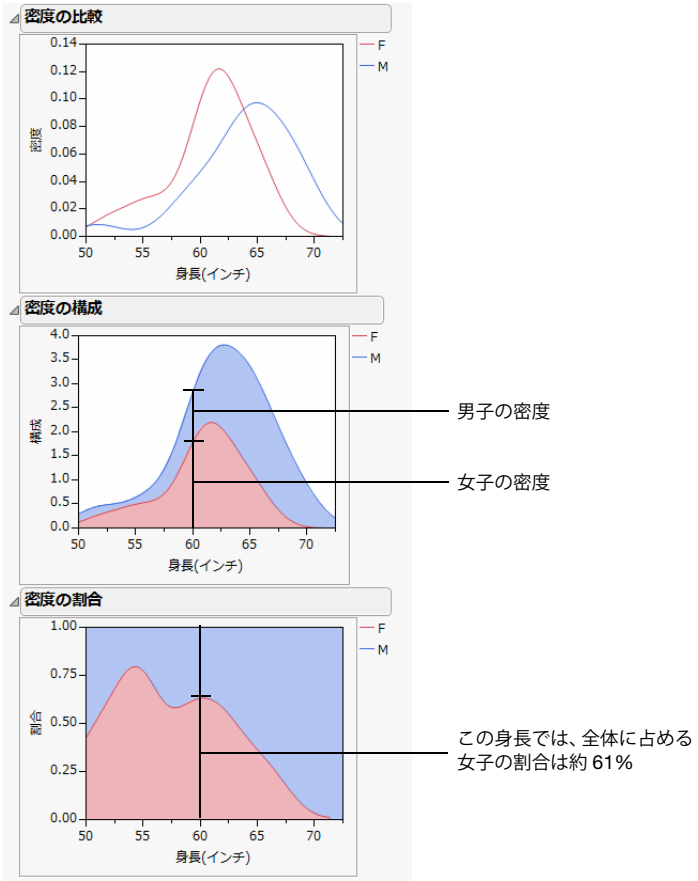


累積確率プロッでは、X 変数の各水準が 1 本の線で表されています。また、横軸は Y の値になっています。

［密度］の各オプションの例

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**身長 (インチ)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**性別**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**密度**］>［**密度の比較**］、［**密度**］>［**密度の構成**］、および［**密度**］>［**密度の割合**］を選択します。

図 6.30［密度］の各オプションの例



［対応のある列を設定］オプションの例

この例では、6頭の動物が移動した距離を、季節ごとに記録した「Matching.jmp」サンプルデータを使用します。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Matching.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選択します。
3. 「**距離 (マイル)**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**季節**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**対応のある列を設定**］を選択します。
7. 対応のある列として「**個体**」を選択します。
8. ［**OK**］をクリックします。

図 6.31「対応のあるあてはめ」レポートの例

「個体」を対応のある変数として、季節ごとの移動距離がプロットに表示されます。**「種別」**と**「個体」**がラベルの列として設定されており、また、各個体の最初の測定値において、ラベルが表示されるように設定されています。

「対応のあるあてはめ」レポートには、**「季節」**効果と**「個体」**効果、およびその *F* 検定が表示されています。これは、「モデルのあてはめ」プラットフォームで交互作用項のあるモデルとないモデルの 2 つを実行したときに計算される検定と同じです。水準が 2 つしかないときは、*F* 検定は対応のある *t* 検定と等しくなります。

注:「モデルのあてはめ」プラットフォームの詳細については、『基本的な回帰モデル』を参照してください。

「一元配置」プラットフォームの統計的詳細

この節では、いくつかのオプションとレポートの統計的詳細を解説します。

比較円の統計的詳細

2つの平均の差が統計的に有意かどうかを確認する方法の1つは、観測された平均の差が**最小有意差**（**LSD; Least Significant Difference**）より大きいかどうかを見ることです。最小有意差は、2つの平均の差の標準誤差をStudentのt統計量に掛けたもので、次式で表されます。

$$\text{LSD} = t_{\alpha/2}\text{std}(\hat{\mu}_1 - \hat{\mu}_2)$$

2つの独立した平均の差の標準誤差は、次の関係から計算されます。

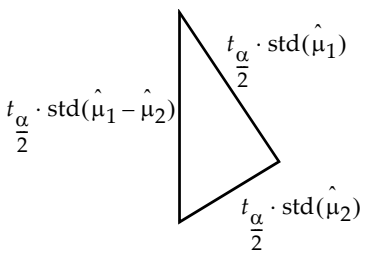
$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{std}(\hat{\mu}_1)]^2 + [\text{std}(\hat{\mu}_2)]^2$$

平均の間に相関がない場合、統計量の関係は次のようになります。

$$\text{LSD}^2 = [t_{\alpha/2}\text{std}((\hat{\mu}_1 - \hat{\mu}_2))]^2 = [t_{\alpha/2}\text{std}(\hat{\mu}_1)]^2 + [t_{\alpha/2}\text{std}\hat{\mu}_2]^2$$

これらの平方値を図にすると、ピタゴラスの定理に基づく直角三角形になります（図6.32）。

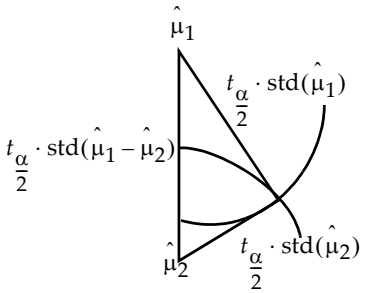
図6.32 2平均の差の関係



三角形の斜辺が、平均を比較する際の目安になります。観測された平均の差が、斜辺の長さ（最小有意差）より大きい場合に限り、平均は有意に異なります。

では、2つの平均がちょうど境界線上にあり、観測された平均の差が最小有意差に等しい場合にはどうなるでしょうか。縦軸に2つの平均を置き、それを頂点として三角形を描きます。また、それぞれの平均を中心として、それぞれの信頼区間の長さを直径とした円を描きます。

図 6.33 *t* 検定統計量の幾何学的関係



各円の半径は対応する三角形の辺の長さと等しく、$t_{\alpha/2}\text{std}(\hat{\mu}_i)$ で表されます。

円は三角形の 2 辺と同じ角度で交わるため、次のような関係が成り立ちます。

- 平均の差が最小有意差と等しい場合、平均を中心とした信頼区間の円は直角に交わります。つまり、接線の角度が直角になります。

では、平均の差が最小有意差より大きい場合や、逆に小さい場合に、円がどのように交わるかを考えてみましょう。

- 円の外側の交角が直角より大きいとき、平均は有意には**異なりません**。円の外側の交角が直角より小さいとき、平均は有意に**異なります**。外側の交角が 90 度より小さいということは、平均の差が最小有意差より大きいことを意味します。
- 円が交わらない場合、平均は有意に異なり、円が入れ子になる場合、平均は有意には異なりません。詳細については、図 6.11 を参照してください。

Student の *t* 分布の分位点ではなく、他の分位点を用いれば、いろいろな多重比較検定に対しても同様のグラフを描くことができます。

検出力の統計的詳細

検出力を計算するには、非心度の *F* 分布を使用します。計算式は、次のとおりです（O'Brien and Lohr 1984）。

Power = Prob($F > F_{crit}$, ν_1, ν_2, *nc*)

ここで、

- 上の式は、非心度パラメータ nc、分子自由度 ν_1、分母自由度 ν_2 の *F* 分布に従う確率変数が、Fcrit よりも大きくなる確率を指します。$F(nc, v_1, v_2)$ は ν_1 および ν_2 の自由度を持つ *F* 分布の累積確率 1 - α における分位点です。
- $\nu_1 = r\text{-}1$ は分子自由度。
- $\nu_2 = r(n - 1)$ は分母自由度。
- *n* は、各グループの標本サイズ。
- *r* はグループの数。

- $nc = n(CSS)/\sigma^2$は非心度パラメータ。

  $CSS = \sum_{g=1}^{r} (\mu_g - \mu)^2$は仮説の平方和。

- μ_gはg番目のグループの平均。
- μは全体平均。
- σ^2は平均2乗誤差（MSE）による推定値。

「あてはめの要約」レポートの統計的詳細

R2乗

連続尺度の応答のあてはめでは、R^2はモデルの「分散分析」レポートの統計量を使用して、次のように計算されます。

$$\frac{\text{モデルの平方和}}{\text{修正済みの全平方和}}$$

自由度調整R2乗

R2乗値は平方和の比でしたが、自由度調整済みR2乗値は、次のように、平均平方の比です。

$$1 - \frac{\text{誤差の平均平方}}{\text{全体の修正済みの平均平方}}$$

ここで、「誤差」の平均平方は「分散分析」レポートにある値です。 分母の平均平方は、「**全体 (修正済み)**」の平方和を、対応する自由度で割ったものです。詳細については、「「分散分析」レポート」（151ページ）を参照してください。

「分散が等しいことを調べる検定」レポートの統計的詳細

F値

O'Brienの検定では、元の応答変数を変換して、新しい応答変数を作成し、それに対して一元配置分散分析を行います。変換後の応答変数は、そのグループ平均が、元の応答変数のグループごとの不偏分散と等しくなるように作成されます。この変換後の応答変数は、次式で計算されます。

$$r_{ijk} = \frac{(n_{ij} - 1.5)n_{ij}(y_{ijk} - \overline{y}_{ij\cdot})^2 - 0.5s_{ij}^2(n_{ij} - 1)}{(n_{ij} - 1)(n_{ij} - 2)}$$

nは、観測値y_{ijk}の個数を表します。

Brown-Forsythe 検定は、$z_{ij} = \left|y_{ij} - \tilde{y}_i\right|$ に対する一元配置分散分析の F 検定です。ここで、$\tilde{y}_i$ は i 番目の水準における Y の中央値です。

Leven 検定は、$z_{ij} = \left|y_{ij} - \overline{y_{i.}}\right|$ に対する一元配置分散分析の F 検定です。ここで、$\overline{y_{i.}}$ は i 番目の水準の応答平均です。

Bartlett 検定は、次式により計算されます。

$$T = \frac{v\log\left(\sum_i \frac{v_i}{v} s_i^2\right) - \sum_i v_i \log(s_i^2)}{1 + \left(\frac{\sum_i \frac{1}{v_i} - \frac{1}{v}}{3(k-1)}\right)}$$ この式で、$v_i = n_i - 1$ および $v = \sum_i v_i$

n_i は i 番目の水準における度数、s_i^2 は i 番目の水準における Y の不偏分散です。この Bartlett 検定の検定統計量は、カイ 2 乗分布に従います。表中の Bartlett 検定の F 値に出力されている値は、この Bartlett のカイ 2 乗検定統計量を自由度で割ったものです。

Welch の検定の F 値

Welch の検定の F 値は、次の式で計算されます。

$$F = \frac{\left[\frac{\sum_i w_i(\bar{y}_i - \tilde{y}_{..})^2}{k-1}\right]}{\left\{1 + \frac{2(k-2)}{k^2-1}\left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i - 1}\right]\right\}}$$ この式で $w_i = \frac{n_i}{s_i^2}$、$u = \sum_i w_i$、$\tilde{y}_{..} = \sum_i \frac{w_i \bar{y}_{i.}}{u}$

n_i は i 番目の水準における度数、$\bar{y}_{i.}$ は i 番目の水準における Y の平均、s_i^2 は i 番目の水準における Y の不偏分散です。

Welch 検定の分母自由度

Welch 検定の分母自由度は、次の近似式によって計算されます。

$$df = \frac{1}{\left(\frac{3}{k^2-1}\right)\left[\sum_i \frac{1 - \frac{w_i}{u}^2}{n_i - 1}\right]}$$

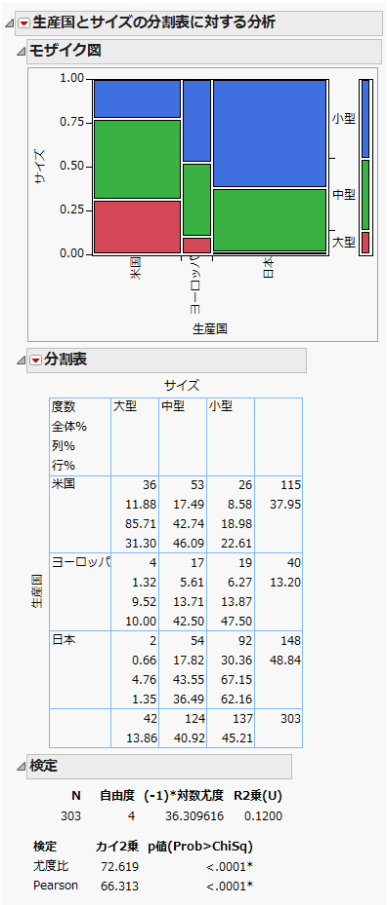
ここで、w_i、n_i、u は F 値の式と同じ定義です。

第 7 章

分割表分析
2つのカテゴリカルな変数の関係を調べる

「分割表」または「二変量の関係」プラットフォームは、カテゴリカルなY変数の分布を、カテゴリカルなX変数の水準ごとに比較します。カテゴリカル変数とは、名義尺度や順序尺度の変数を指します。分割表分析は、「二変量の関係」プラットフォームにおいて、XおよびYが、**両方ともカテゴリカル変数**のときに行われる分析です。分析結果として、モザイク図、度数、割合などが表示されます。また、割合の平均分析や対応分析なども行えます。

図7.1 分割表分析の例



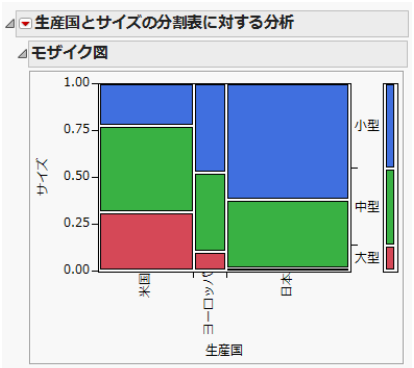
目次

分割表分析の例

この例では、車に関する調査の「Car Poll.jmp」サンプルデータを使用します。データには、調査対象者の性別、配偶者の有無、年齢などが記録されています。また、所有する車について、生産国やサイズ、タイプなどの情報も記録されています。ここでは、車のサイズ（小型、中型、大型）と生産国の関係について調べます。

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Car Poll.jmp」を開きます。
2. ［分析］>［二変量の関係］を選びます。
3. 「サイズ」を選択し、［Y, 目的変数］をクリックします。
4. 「生産国」を選択し、［X, 説明変数］をクリックします。
5. ［OK］をクリックします。

図7.2 分割表分析の例



生産国とサイズの分割表に対する分析

モザイク図

分割表

サイズ

生産国	度数 全体% 列% 行%	大型	中型	小型	
米国		36	53	26	115
		11.88	17.49	8.58	37.95
		85.71	42.74	18.98	
		31.30	46.09	22.61	
ヨーロッパ		4	17	19	40
		1.32	5.61	6.27	13.20
		9.52	13.71	13.87	
		10.00	42.50	47.50	
日本		2	54	92	148
		0.66	17.82	30.36	48.84
		4.76	43.55	67.15	
		1.35	36.49	62.16	
		42	124	137	303
		13.86	40.92	45.21	

検定

N	自由度	(-1)*対数尤度	R2乗(U)
303	4	36.309616	0.1200

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	72.619	<.0001*
Pearson	66.313	<.0001*

図 7.2 のモザイク図と凡例から、次のことがわかります。

- 日本車において、大型車に分類されるものはごくわずかです。
- ヨーロッパ車の大半は、小型車か中型車に分類されます。
- 米国車の大半は、大型車か中型車に分類されます。

「分割表」プラットフォームの起動

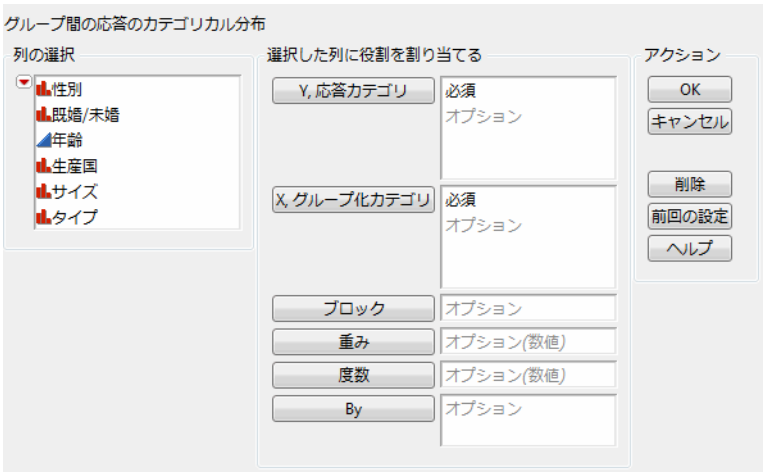
分割表分析は、「二変量の関係」プラットフォームまたは「分割表」プラットフォームを使用して実行できます。どちらを使用しても結果は同じです。

- 「二変量の関係」プラットフォームを起動するには、[**分析**] > [**二変量の関係**] を選択します。

  または

- 「分割表」プラットフォームを起動するには、「JMP スターター」ウィンドウのカテゴリリストで [**基本統計**] をクリックし、[**分割表**] をクリックします。

図 7.3 「分割表」起動ウィンドウ



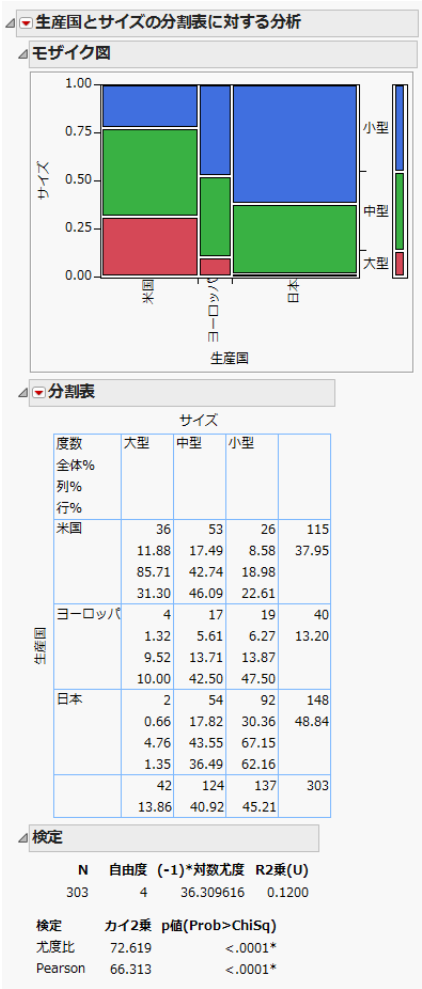
この起動ウィンドウの詳細については、「「二変量の関係」プラットフォームの概要」の章（91 ページ）を参照してください。

[**OK**] をクリックすると、「分割表」レポートウィンドウが開きます。詳細については、「「分割表」レポート」（201 ページ）を参照してください。

「分割表」レポート

図7.4は、「分割表分析の例」（199ページ）の手順に従って作成したものです。

図7.4「分割表」レポートの例



度数 全体% 列% 行%	大型	中型	小型	
米国	36 11.88 85.71 31.30	53 17.49 42.74 46.09	26 8.58 18.98 22.61	115 37.95
ヨーロッパ	4 1.32 9.52 10.00	17 5.61 13.71 42.50	19 6.27 13.87 47.50	40 13.20
日本	2 0.66 4.76 1.35	54 17.82 43.55 36.49	92 30.36 67.15 62.16	148 48.84
	42 13.86	124 40.92	137 45.21	303

注: データテーブルで除外した行は、モザイク図で非表示になります。

「分割表に対する分析」レポートには最初、「モザイク図」、「分割表」、「検定」レポートが表示されます。赤い三角ボタンをクリックし、表示されるメニューのオプションを使用して、他の分析や検定を追加できます。これらのレポートやオプションの詳細については、「「分割表」プラットフォームのオプション」（202ページ）を参照してください。

「分割表」プラットフォームのオプション

注 : Y 変数を複数指定した場合は、[あてはめのグループ] メニューが表示されます。このメニューのオプションを使って、レポートの配置を変えたり、適合度の良い順に並べ替えたりできます。詳細については、『基本的な回帰モデル』を参照してください。

「分割表に対する分析」の横にある赤い三角ボタンをクリックし、表示されるメニューのオプションを選択することにより、追加の分析や検定を実行できます。

モザイク図	「分割表」に表示されているデータをグラフで示します。詳細については、「モザイク図」(203 ページ)を参照してください。
分割表	2 元度数表。列は Y の水準と同数、行は X の水準と同数だけあります。詳細については、「分割表」(206 ページ)を参照してください。
検定	連続量のデータに対して作成される「分散分析」レポートに相当します。すべての X 水準において Y 水準の割合が等しいかどうかを示す検定結果を表示します。詳細については、「検定」(208 ページ)を参照してください。
α 水準の設定	信頼区間の α 水準を変更します。サブメニューから、よく使われている有意水準 ([0.10]、[0.05]、[0.01]) を選択するか、[**その他**] を選択して任意の有意水準を入力します。
割合の平均分析	応答変数の水準数が 2 つの場合にのみ表示されます。X 変数の水準ごとの応答の割合を、割合の全体平均と比較します。詳細については、「割合の平均分析」(209 ページ)を参照してください。
対応分析	度数表において度数のパターンが似ている行や列をグラフに示します。対応分析のグラフには、分割表の各行と各列につき 1 つの点がプロットされます。詳細については、「対応分析」(210 ページ)を参照してください。
Cochran-Mantel-Haenszel 検定	3 つ目の分類変数で層別したときに、2 つのカテゴリカル変数の間に関係が見られるかどうかを検定します。詳細については、「Cochran-Mantel-Haenszel 検定」(211 ページ)を参照してください。
一致性の統計量	X と Y における水準がまったく同じ場合にのみ表示されます。カッパ統計量 (Agresti 1990) とその標準誤差、信頼区間、仮説検定、Bowker の対称性検定 (McNemar の検定) が表示されます。詳細については、「一致性の統計量」(211 ページ)を参照してください。
相対リスク	リスク比を計算します。X と Y が両方とも 2 水準である場合にのみ表示されます。詳細については、「相対リスク」(212 ページ)を参照してください。

オッズ比	オッズ比を計算します。XとYが両方とも2水準である場合にのみ表示されます。詳細は、「［オッズ比］オプションの統計的詳細」（227ページ）を参照してください。 また、レポートには、オッズ比に対する信頼区間も表示されます。［**αの水準の設定**］オプションを使用してα水準を変更できます。
割合の2標本検定	2標本の割合に対する検定を実行します。X変数の2水準間で、Y変数の割合が異なるかどうかを比較します。XとYが両方とも2水準である場合にのみ表示されます。詳細については、「割合の2標本検定」（212ページ）を参照してください。
関連の指標	分割表の変数間における関連度を表す指標を計算します。詳細については、「関連の指標」（213ページ）を参照してください。
Cochran-Armitageの傾向検定	ある変数の水準間に見られる、二項割合の傾向を調べます。この検定は、片方の変数が2水準の変数で、もう一方が順序変数の場合にだけ適切な検定です。詳細については、「Cochran-Armitageの傾向検定」（215ページ）を参照してください。
JMP PRO **正確検定**	次の正確検定を実行するためのオプションが表示されます。 • Fisherの正確検定 • Cochran-Armitageの傾向検定 • 一致性に対する検定 「正確検定」（215ページ）を参照してください。

モザイク図

モザイク図は、2元度数表（分割表）をグラフにしたものです。モザイク図は長方形に分割したプロットで、それぞれの長方形における縦の長さが、X変数の各水準内における、Y変数の各水準の割合を表します。モザイク図は1981年にHartiganとKleinerによって考案され、Friendly（1994）が改良を加えました。

図7.5は、「分割表分析の例」（199ページ）の手順に従って作成したものです。

図 7.5 モザイク図の例

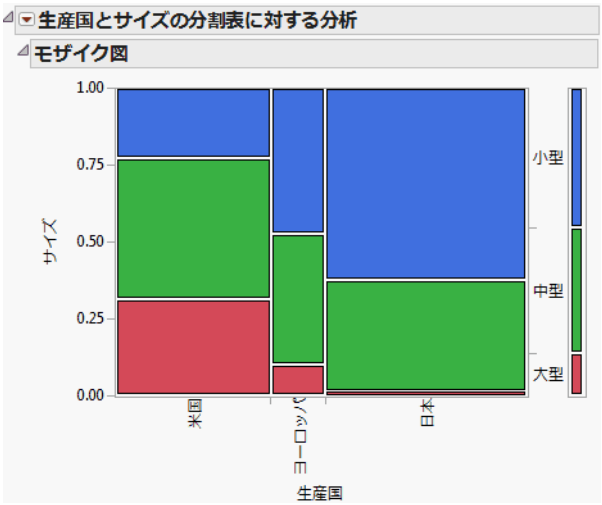


図 7.5 のモザイク図で、次の点に注目してください。

- X 軸上の長さは、X 変数（「生産国」）の各水準のオブザベーション数を示します。
- 右側の Y 軸は、全水準（「米国」、「ヨーロッパ」、「日本」）を組み合わせたときに、「小型」車、「中型」車、「大型」車が全体に占める割合を示します。
- 左側の Y 軸は、軸全体を 1（標本合計）としたときの応答確率を表します。

モザイク図の長方形をクリックすると、その領域が強調表示され、関連するデータテーブルで対応するデータが選択されます。

モザイク図で使われている変数は、ドラッグ＆ドロップ操作によって、別の変数に置き換えることができます。X 軸と Y 軸の変数を入れ替えるには、一方の軸の変数をドラッグし、他方の軸の上にドロップします。また、データテーブルの他の変数に置き換えるには、「列」パネルで該当する変数をクリックし、モザイク図の軸上までドラッグします。

コンテキストメニュー

モザイク図上を右クリックして、色を変更したり、セルのラベルを指定したりすることができます。

表 7.1 コンテキストメニューのオプション

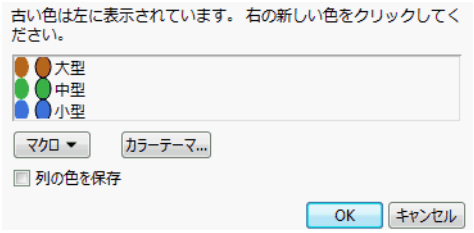
色の設定	水準に割り当てられている色を表示します。詳細については、「色の設定」（205 ページ）を参照してください。
セルのラベル	モザイク図に表示するラベルを指定します。次のいずれかのオプションを選択します。 **ラベルなし**　ラベルを表示しません。ラベルが表示されている場合は削除します。 **度数の表示**　各セルにオブザベーション数を表示します。 **パーセントの表示**　各セルにオブザベーションの割合を表示します。 **ラベルを表示**　各セルに対応する Y 変数の水準を表示します。 **行ラベルの表示**　セルが表しているすべての行の行番号を表示します。

注： コンテキストメニューの他のオプションの詳細については、『JMP の使用法』を参照してください。

色の設定

［**色の設定**］オプションを選択すると、「値の色を選択」ウィンドウが開きます。

図 7.6「値の色を選択」ウィンドウ



デフォルトのモザイク図の色は、応答列が順序尺度か名義尺度か、「値の色」列プロパティが設定されているかどうかによって異なります。水準の色を変更するには、2 列目の色の楕円形をクリックし、新しい色を選択します。

表 7.2「値の色を選択」ウィンドウの説明

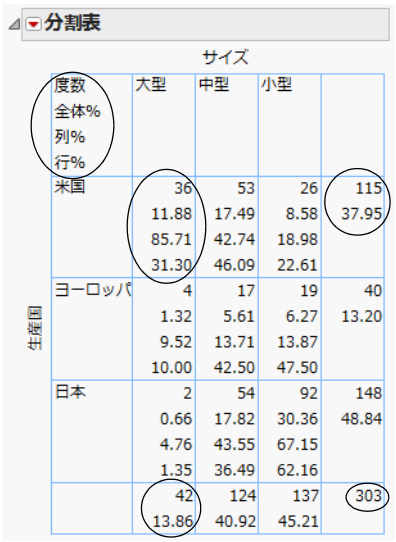
マクロ	2つの水準間における色のグラデーションを、次のような方法で設定できます。 • 2つの水準を選択し（それら2つの水準の範囲をドラッグするか、Shift キーを押しながら最初と最後の水準をクリックする）、[**選択した点の間でグラデーション**] を選択すると、それらの範囲にだけグラデーションが適用されます。 • [**両端の間でグラデーション**] オプションを選択すると、変数のすべての水準にグラデーションが適用されます。 • [**元の色に戻す**] を選択すると、変更内容を取り消すことができます。
カラーテーマ	カラーテーマに従って各値の色を変更します。
列の色を保存	このチェックボックスをオンにすると、関連するデータテーブルの列に新しい列プロパティ（[**値の色**]）が追加されます。データテーブルからこのプロパティを編集する場合は、[**列**] > [**列情報**] を選択します。

分割表

「分割表」は2元度数表です。列はYの水準と同数、行はXの水準と同数だけあります。

図7.7は、「分割表分析の例」（199ページ）の手順に従って作成したものです。

図 7.7 分割表の例



この分割表について、次の点を確認してください。

- この分割表において、「度数」、「全体 %」、「列 %」、「行 %」は、行と列が交わる各セル（たとえば、「米国」と「大型」が交差するセルなど）内の数値を指しています。
- 最後の列には、各行の合計度数と割合が表示されています。
- 最下行には、各列の合計度数と割合が表示されています。

たとえば、図7.7で、「米国」製の「大型」車に着目してください。次の表で、車の分割表から導き出せる結論について説明します。

表7.3 例の分割表による結果

数値	説明	表のラベル
36	米国製大型車の台数	度数
11.88%	全車両のうち米国製大型車が占める割合（36/303）[a]	全体 %
85.71%	大型車のうち米国車が占める割合（36/42）[b]	列 %
31.30%	米国車のうち大型車が占める割合（36/115）[c]	行 %
37.95%	全車両のうち米国車が占める割合（115/303）	（なし）
13.86%	全車両のうち大型車が占める割合（42/303）	（なし）

a. 303は、調査対象の全車両の台数です。
b. 42は、調査対象の大型車の台数です。
c. 115は、調査対象の米国車の台数です。

ヒント： 分割表における「度数」、「全体 %」、「列 %」、「行 %」の数値は、「分割表」の赤い三角ボタンをクリックし、その項目を選択すると、表示／非表示を切り替えられます。

表7.4「分割表」レポートの説明

度数	セルの度数、行／列の合計度数、全体の合計度数（全体の標本サイズ）。
全体 %	セルの度数と行／列の合計度数が全体の合計度数に占める割合。
行 %	セルの度数が行の合計度数に占める割合。
列 %	セルの度数が列の合計度数に占める割合。
期待値	行と列が独立していると仮定して計算された各セルの期待度数（*E*）。対応する行合計と列合計の積を全体合計で割った値です。
偏差	観測されたセル度数（*O*）から期待値（*E*）を引いたもの。
セルのカイ2乗	$(O - E)^2 / E$ で求めた各セルのカイ2乗。
累積列度数	列の累積合計。

表 7.4「分割表」レポートの説明（続き）

累積列 %	列の累積割合。
累積行度数	行の累積合計。
累積行 %	行の累積割合。

検定

「検定」レポートには、すべてのX水準においてY水準の割合が等しいかどうかを調べる2つの検定の結果が表示されます。

図7.8のレポートを作成するには、「分割表分析の例」（199ページ）の手順に従います。

図 7.8「検定」レポートの例

検定

N	自由度	(-1)*対数尤度	R2乗(U)
303	4	36.309616	0.1200

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	72.619	<.0001*
Pearson	66.313	<.0001*

カイ 2 乗統計量について、次の点を確認してください。

- これらのカイ2乗検定は、2つのカテゴリカル変数が両方とも応答（Y変数）のとき、その2つの変数が独立であることを検定します。
- 2つの変数のうち、Xが所与である場合、カイ2乗統計量は、Y変数の分布がXの全水準で同じであることを検定します。

表 7.5「検定」レポートの説明

N	オブザベーションの合計数。
自由度	検定に関係する自由度。 自由度は $(c - 1)(r - 1)$ に等しくなります。c は列数、r は行数です。
(-1)* 対数尤度	対数尤度の符号を逆にしたもの。適合度や不確定性を表す指標です（連続尺度の応答変数における平方和と同じような指標です）。

表 7.5「検定」レポートの説明（続き）

R2 乗 (U)	全体の不確定性のうち、モデルによって説明される割合。 • R^2 が 1 のときは、カテゴリカルな応答が因子によって完全に予測できることを意味します。 • R^2 が 0 ということは、標本全体で応答確率が一定であるとすることに比べて、モデルを使っても予測が向上しないことを意味しています。 詳細は、「「検定」レポートの統計的詳細」（227 ページ）を参照してください。
検定	「すべての標本グループにおいて、応答確率は等しい」という仮説を検定する 2 種類のカイ 2 乗検定。詳細は、「「検定」レポートの統計的詳細」（227 ページ）を参照してください。
p 値（Prob>ChiSq）	X 変数と Y 変数の間に関係が存在しないときに、まったくの偶然だけで、現在のカイ 2 乗値より大きなカイ 2 乗値を得る確率。両方の変数が 2 水準しかない場合は、Fisher の正確検定も実行されます（両側検定と片側検定の p 値が表示されます）。

Fisher の正確検定

このレポートには、2×2 分割表に対する Fisher の正確検定の結果が表示されます。2×2 分割表の場合は、Fisher の正確検定が自動的に実行されます。2×2 および *r*×*c* 分割表に対する Fisher の正確検定の詳細については、「正確検定」（215 ページ）を参照してください。

割合の平均分析

応答変数が 2 水準である場合、このオプションによって、X 変数の水準ごとの応答の割合を、割合の全体平均と比較できます。この方法では、二項分布に対する正規近似を計算に用いているため、標本サイズが小さすぎる場合は、警告が表示されます。

注： 平均分析法の解説については、Nelson, Wludyka, and Copeland（2005）を参照してください。分散の平均分析（ANOMV）法の解説については、Wludyka and Nelson（1997）を参照してください。

表 7.6「割合の平均分析」のオプションの説明

有意水準の設定	分析に使用する α 水準を選択します。
要約レポートの表示	X 変数の水準ごとに、応答変数の割合と決定限界を表示します。このレポートには、決定限界を超過しているかどうかも示されます。
応答水準の切り替え	分析に使用する応答変数のカテゴリを変更します。
表示オプション	［決定限界の表示］、［決定限界の陰影の表示］、［中心線の表示］、［点のオプション］ の各オプションがあります。

関連情報

- 「割合の平均分析の例」（216 ページ）

対応分析

対応分析は、度数表の中から度数のパターンが似ている行や列を見つけるための手法です。対応分析のグラフでは、各行と各列につき 1 つの点がプロットされます。水準数が多く、モザイク図から有用な情報を引き出すのが難しい場合は、対応分析が有効です。

対応分析プロットの理解

行ごとの割合、つまり、行内にある度数をその行の合計度数で割ったものを、**行プロファイル**と言います。2 つの行の行プロファイルが似ているとき、対応分析のグラフにおいて、それらの行を表す点は互いに近い位置にあります。プロットにおける 2 点間の平方距離は、「カイ 2 乗距離」という 2 行の等質性を示す距離を近似したものです。

列プロファイルも、行プロファイルと同じように定義され、同じように解釈できます。行の点と列の点の距離には意味がありませんが、原点からの方向には意味があり、その関係を見ることでプロットを解釈できます。

「対応分析」のオプション

「対応分析」の横にある赤い三角ボタンをクリックし、表示されるメニューのオプションを使用して、三次元散布図を作成したり、データテーブルに列プロパティを追加したりすることができます。

表 7.7 「対応分析」のオプションの説明

三次元対応分析	三次元散布図を作成します。
値の順序を各列に保存	対応分析における第 1 次元のスコアで値を並べた順序が、X 列と Y 列の「値の順序」列プロパティに保存されます。

「詳細」レポート

「詳細」レポートは、対応分析の統計情報をまとめたもので、プロットに使われている値をリストしています。

表 7.8「詳細」レポートの説明

特異値	分割表の特異値分解。式については、「対応分析の「詳細」レポートの統計的詳細」（228 ページ）を参照してください。
慣性	特異値の平方。各次元で説明できる相対的な変動を反映しています。
割合	合計慣性に対する各慣性の割合。

表 7.8「詳細」レポートの説明（続き）

累積	慣性の割合の累積。最初の 2 つの特異値が慣性のほとんどを捉えているときは、2 次元の対応分析プロットで、表内に存在する関係を十分に示すことができます。
X 変数（チーズ）c1、c2、c3	対応分析プロット上に表示される値（図 7.11）。
Y 変数（評価）、c1、c2、c3	対応分析プロット上に表示される値（図 7.11）。

関連情報

- 「対応分析の例」（217 ページ）

Cochran-Mantel-Haenszel 検定

Cochran-Mantel-Haenszel 検定では、3 つ目の分類変数で層別したときに、2 つのカテゴリカル変数の間に関係が見られるかどうかを調べます。

表 7.9 Cochran-Mantel-Haenszel 検定の説明

スコアの相関	Y と X の両方が順序尺度か間隔尺度であるときに使います。対立仮説は、層別変数の少なくとも 1 つの層において、X と Y の間に線形関係があるというものです。
X 間でのスコア比較	Y が順序尺度か間隔尺度であるときに使います。対立仮説は、層別変数の少なくとも 1 つの層において、ある列の行スコアの平均が、他の列の行スコアの平均と異なるというものです。
Y 間でのスコア比較	X が順序尺度か間隔尺度であるときに使います。対立仮説は、層別変数の少なくとも 1 つの層において、ある行の列スコアの平均が、他の行の列スコアの平均と異なるというものです。
カテゴリの 一般連関	層別変数の少なくとも 1 つの水準で、X と Y の間に何らかの関連があることを検定します。

関連情報

- 「Cochran-Mantel-Haenszel 検定の例」（219 ページ）

一致性の統計量

2 変数の水準がまったく同じ場合には、[**一致性の統計量**] オプションが使用できます。このオプションは、カッパ統計量（Agresti 1990）の算出と、Bowker の対称性検定を行います。カッパ統計量に対しては、標準誤差、信頼区間、仮説検定が計算されます。

この節で紹介しているカッパ統計量とp値は近似値ですが、正確な値も計算できます。詳細については、「正確検定」（215ページ）を参照してください。

表7.10「一致性の統計量」レポートの説明

カッパ	カッパ統計量。
標準誤差	カッパ統計量の標準誤差。
下側95%	カッパ統計量の下側信頼区間。
上側95%	カッパ統計量の上側信頼区間。
p値(Prob>Z)	カッパ統計量の片側検定のp値。「カッパ統計量が0である」という帰無仮説を検定します。
p値(Prob>|Z|)	カッパ統計量の両側検定のp値。
カイ2乗	分割表の対称性に対するBowker検定の検定統計量です。Bowker検定の帰無仮説は、「分割表における確率が対称性である（すべてのセルにおいて$p_{ij}=p_{ji}$である）」というものです。なお、XとYが2水準の場合、Bowker検定は、McNemar検定と等しくなります。
p値（Prob>ChiSq）	Bowker検定のp値。

関連情報

- 「[一致性の統計量] オプションの例」（220ページ）
- 「[一致性の統計量] オプションの統計的詳細」（226ページ）

相対リスク

分割表が2×2表である場合、**[相対リスク]** オプションを選択すると、リスク比が計算されます。レポートには信頼区間も表示されます。この手法の詳細は、Agresti（1990）の3.4.2節で説明されています。

[相対リスク] オプションを選択すると、「相対リスクカテゴリの選択」ウィンドウが開きます。1対の応答と因子について、または、応答と因子のすべての組み合わせについて相対リスクを計算することができます。

関連情報

- 「[相対リスク] オプションの例」（222ページ）

割合の2標本検定

分割表が2×2表である場合は、割合の差に対する信頼区間も計算できます。また、信頼区間に相当する検定も計算されます。

表 7.11「割合の 2 標本検定」レポートの説明

説明	行われている検定の種類。
割合の差	X 変数の 2 水準間における割合の差。
下側 95%	割合の差に対する両側信頼区間の下限。この信頼区間は、調整を伴った Wald 法によって計算されています。
上側 95%	割合の差に対する両側信頼区間の上限。この信頼区間は、調整を伴った Wald 法によって計算されています。
調整済み Wald 検定	両側検定と片側検定を示します。
p 値	検定の p 値。
対象となる応答<変数>のカテゴリ	検定で使用する応答の水準を選択します。

関連情報

- 「割合の 2 標本検定の例」（223 ページ）

関連の指標

［**関連の指標**］オプションを選択すると、分割表の変数間における関連を示す、複数の指標が計算されます。

関連の指標の詳細については、次の文献を参考にしてください。

- Brown and Benedetti（1977）
- Goodman and Kruskal（1979）
- Kendall and Stuart（1979）
- Snedecor and Cochran（1980）
- Somers（1962）

表 7.12「関連の指標」レポートの説明

ガンマ	順序関係が一致するペアと一致しないペアの個数から計算される相関係数です。同順位のペアは無視します。-1 から 1 の値を取ります。
Kendall のタウ-b	ガンマに似ていますが、同順位のペアに対して、修正が行われます。-1 から 1 の値を取ります。
Stuart のタウ-c	ガンマに似ていますが、表のサイズによる調整、および、同順位のペアに対する修正が行われます。-1 から 1 の値を取ります。

表 7.12「関連の指標」レポートの説明（**続き**）

SomersのD	Kendallのタウ-bに、非対称性の修正を加えた指標です。 • 「C|R」は、行変数Xを独立変数、列変数Yを従属変数とすることを示します。 • 同様に、「R|C」は、列変数Yを独立変数、行変数Xを従属変数とすることを示します。 SomersのDは、同順位のペアに対する修正を、独立変数の同順位に対してだけ行うという点で、タウ-bとは異なります。-1から1の値を取ります。
非対称ラムダ	• 「非対称ラムダ（C|R)」は、行変数Xで列変数Yを予測したときに、予測的中率がどれぐらい向上するかを示します。 • 「非対称ラムダ（R|C)」は、列変数Yから行変数Xを予測したときに、予測的中率がどれぐらい向上するかを示します。 0から1の値を取ります。
対称ラムダ	2つの非対称ラムダ係数の平均のような指標です。0から1の値を取ります。
不確実性係数	• 「不確実性係数（C|R)」は、行変数Xによって説明される列変数Yの不確実性の割合を示します。 • 「不確実性係数（C|R)」は、列変数Yによって説明される列変数Xの不確実性の割合を示します。 0から1の値を取ります。
不確実性係数(対称)	上記2つの不確実性係数を、対称性をもつように変えた指標です。0から1の値を取ります。

各推定値、および、その標準誤差と信頼区間が表示されます。次の点を念頭に置いてください。

- ガンマ、Kendallのタウ-b、Stuartのタウ-c、SomersのDは、X変数の増加に伴い、Y変数も増加する傾向にあるかどうかを示す順序的な関係を示す指標です。これらの統計量は、順序関係の一致するペアと一致しないペアの個数から計算されます。データのペアにおいて、X値がより大きい方が、Y値もより大きくなっているペアを、「順序関係が一致するペア」と言います。逆に、X値が大きい方が、Y値が小さくなっているペアを「順序関係が一致しないペア」と言います。これらの指標は、両変数が順序尺度の場合だけに適しています。
- ラムダおよび不確実性係数は、順序尺度と名義尺度の変数に適しています。

関連情報

- 「[関連の指標] オプションの例」（224 ページ）

Cochran-Armitageの傾向検定

Cochran-Armitageの傾向検定は、順序変数の水準間における二項割合の傾向を調べます。この検定は、片方の変数が2水準の変数で、もう一方が順序変数の場合にだけ適切な検定です。2水準の変数が応答変数、順序尺度の変数が説明変数として扱われます。帰無仮説は「割合の傾向がない」、つまり、「説明変数の全水準において二項割合が同じ」というものです。

この節で紹介している検定統計量とp値は近似値です。正確な傾向検定も実行できます。「正確検定」（215ページ）を参照してください。

関連情報

- 「Cochran-Armitageの傾向検定の例」（225ページ）

正確検定

JMP PRO 次の表では、「分割表」プラットフォームで実行できる3つの正確検定について説明します。

表7.13 正確検定

Fisherの正確検定	$r \times c$分割表に対するFisherの正確検定を実行します。この検定は、2変数の関連を調べます。Fisherの正確検定は、行和および列和を固定して、超幾何分布を用いてp値を求めます。 この検定は大標本近似ではありません。標本サイズが小さい場合や、度数が0のセルがある場合など、尤度比検定やPearson検定では妥当な結果が得られないときに適しています。 レポートには次の情報も表示されます。 **表の確率(P)** 現在、観測されている表が生じる確率。検定のp値ではありません。 **両側p値(Prob≤P)** 両側検定のp値。 2x2分割表の場合、Fisherの正確検定が自動的に実行されます。詳細については、「検定」（208ページ）を参照してください。
正確Cochran-Armitage傾向検定	正確なCochran-Armitage傾向検定を実行します。この検定は、いずれかの変数が2水準である場合にのみ実行できます。傾向検定の詳細については、「Cochran-Armitageの傾向検定」（215ページ）を参照してください。

表 7.13 正確検定（続き）

正確一致性検定	2 変数間の一致性に対する正確検定を実行します。これはカッパ統計量に対する正確検定です。2 変数の水準がまったく同じである場合にのみ実行できます。一致性検定の詳細については、「一致性の統計量」（211 ページ）を参照してください。

「分割表」プラットフォームのその他の例

ここでは、「分割表」プラットフォームのオプションを使った例をさらに紹介します。

割合の平均分析の例

この例では、「Office Visits.jmp」サンプルデータを使用します。このデータには、ある地域の 6 つの診療所において、予約どおりの時間に患者が診察できたかどうかが記録されています。6 つの診療所ごとに、1 週間分のデータから、60 の診察予約データを無作為抽出しました。予約時間から 5 分以内に診察室に通された場合には、定刻どおりの診察とみなします。予約時間どおりに診察を受けた患者の割合を調べてみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Office Visits.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. **「時間どおり」**を選択し、［**Y, 目的変数**］をクリックします。
4. **「診療所」**を選択し、［**X, 説明変数**］をクリックします。
5. **「度数」**を選択し、［**度数**］をクリックします。
6. ［**OK**］をクリックします。
7. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［**割合の平均分析**］を選択します。
8. 「割合の平均分析」の横にある赤い三角ボタンをクリックし、メニューから［**要約レポートの表示**］および［**応答水準の切り替え**］を選択します。

図 7.9 割合の平均分析の例

割合の平均分析の要約

水準	N	下側限界	グループ割合	上側限界	限界外
A	60	0.367911	0.166667	0.676534	下側
B	60	0.367911	0.8	0.676534	上側
C	60	0.367911	0.616667	0.676534	
D	60	0.367911	0.133333	0.676534	下側
E	60	0.367911	0.583333	0.676534	
F	60	0.367911	0.833333	0.676534	上側

図 7.9 は予約時間どおりに診察を受けた患者の割合を、診療所ごとに示しています。図 7.9 のようなグラフが描かれます。

- 予約時間どおりに受診できた割合が一番高いのは、診療所 F です。診療所 B がそれに次いでいます。
- 診療所 D は、予約時間どおりに受診できた割合が一番低く、診療所 A がそれに次いでいます。
- 診療所 C と E は平均に近く、決定限界を超えていません。

対応分析の例

この例では、「Cheese.jmp」サンプルデータを使用します。このデータは、Newell によるチーズの試食試験の結果で、McCullagh and Nelder（1989）で取り上げられています。試験では、4 種類のチーズ添加物ごとに、9 段階評価の度数が記録されました。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Cheese.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**評価**」を選択し、［**Y, 目的変数**］をクリックします。

   「**評価**」は、1（最下位）から 9（第 1 位）までを範囲としています。
4. 「**チーズ**」を選択し、［**X, 説明変数**］をクリックします。

   「A」、「B」、「C」、「D」はチーズの 4 つの添加物を表します。
5. 「**度数**」を選択し、［**度数**］をクリックします。
6. ［**OK**］をクリックします。

図7.10 チーズデータのモザイク図

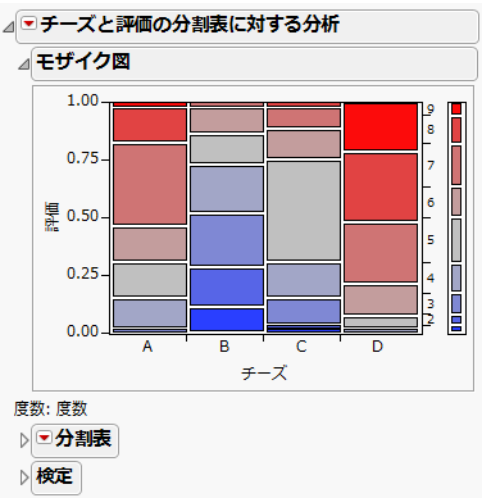


図7.10のモザイク図を見ると、分布はそれぞれ異なるようですが、9つも水準があるのでモザイク図だけでは傾向が読めません。このようなときは、対応分析が役立ちます。

7. 対応分析プロットを表示するには、「分割表に対する分析」の横にある赤い三角ボタンをクリックし、[**対応分析**]を選択します。

図7.11 対応分析プロットの例

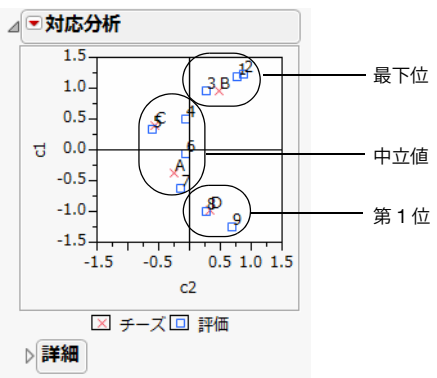


図 7.11 は、対応分析をグラフ化したもので、軸には c1 と c2 という名前が付いています。上の例で、次の点に注意してください。

- c1 は、全体的な満足度に対応しているようです。c1 軸の上部に 1（最下位）、下部に 9（第 1 位）という評価が分布しています。
- チーズ D は、評価が 8 と 9 である割合が、他よりも高くなっています。
- チーズ B は、評価が 1、2、3 である割合が、他よりも高くなっています。

- チーズCとAは、4、5、6、7という中間的な評価を得ています。

8. 「対応分析」の横にある赤い三角ボタンをクリックし、メニューから［**三次元対応分析**］を選択します。

図 7.12 三次元散布図の例

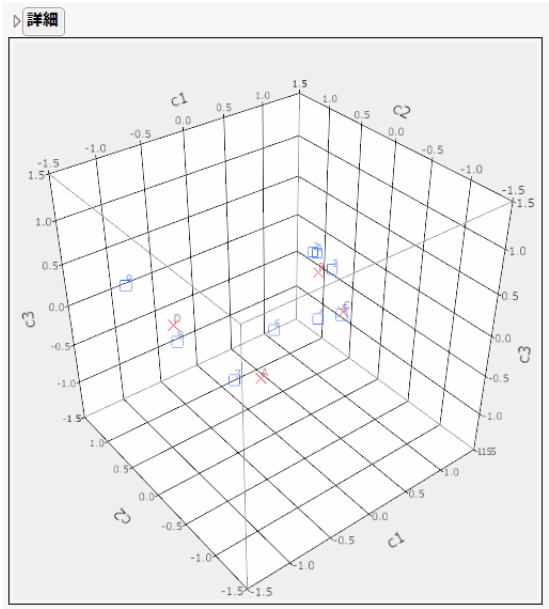


図 7.12 のような正規分位点プロットが描かれます。

- c1 軸を見ると、評価 1 から評価 9 までが、正から負に並んでいます。
- c2 軸を見ると、AとCが 0 の右側（正）、BとDが 0 の左側（負）に位置しています。
- c1 は全体的な満足度を表しています。

Cochran-Mantel-Haenszel 検定の例

この例では、「Hot Dogs.jmp」サンプルデータを使用します。ホットドッグに関して、タイプと味の関係を調べてみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Hot Dogs.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**タイプ**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**味**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［**Cochran-Mantel-Haenszel 検定**］を選択します。

7. 層別変数として**「脂肪に対する蛋白質の比率」**を選択し、**［OK］**をクリックします。

図 7.13 Cochran-Mantel-Haenszel 検定の例

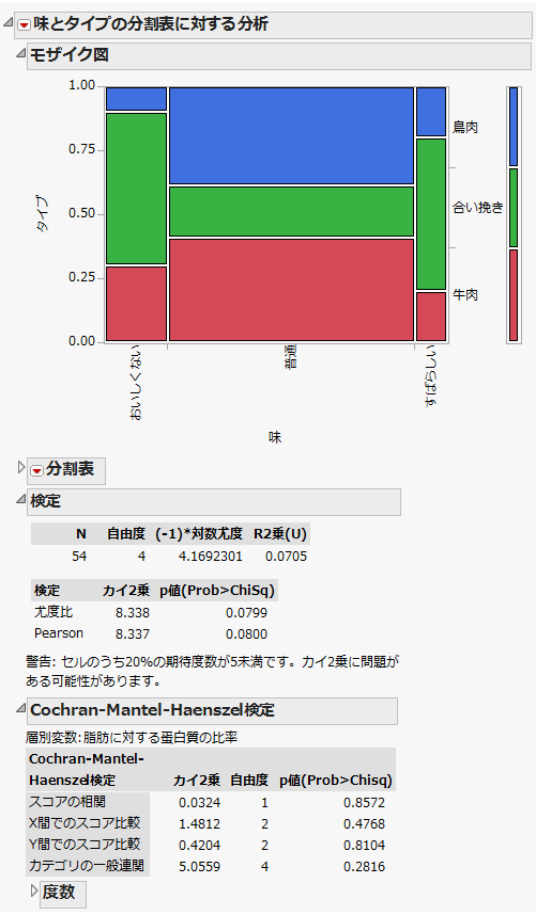


図 7.13 からは次のようなことがわかります。

- 「検定」レポートを見ると、Pearson のカイ 2 乗検定に対する p 値は 0.0800 で、5% よりは少し大きいですが、ホットドッグの「味」と「タイプ」の間に何らかの関係があるかもしれないことを示唆しています。
- 一方、「Cochran-Mantel-Haenszel 検定」レポートを見ると、一般的な連関性に対する p 値は 0.2816 であり、5% よりもかなり大きくなっています。

［一致性の統計量］オプションの例

この例では、「Attribute Gauge.jmp」サンプルデータを使用します。3 人の判定者が 50 個の部品をそれぞれ 3 回評価した結果が記録されています。判定者 A と B の関係を調べてみましょう。

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Attribute Gauge.jmp」を開きます。
2. ［分析］>［二変量の関係］を選びます。
3. 「A」を選択し、［Y, 目的変数］をクリックします。
4. 「B」を選択し、［X, 説明変数］をクリックします。
5. ［OK］をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［一致性の統計量］を選択します。

図7.14「一致性の統計量」レポートの例

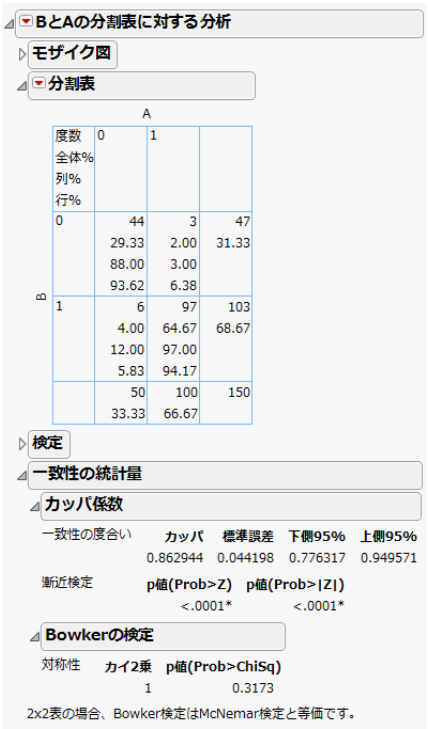


図7.14を見ると、一致性の統計量は0.86と、1に近い大きな値となっています。また、そのp値は<.0001と小さくなっています。分割表を見ても、対角線上における度数が大きく、一致性が高いことを裏付けています。2名の評価者が同じ評価をしている場合（つまり、ともに0と評価するか、または、ともに1と評価している場合）、一致性の統計量は大きくなります。

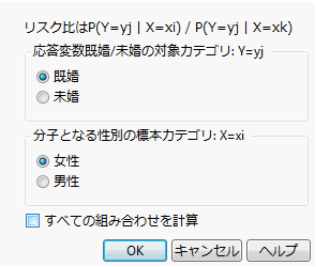
［相対リスク］オプションの例

この例では、「Car Poll.jmp」サンプルデータを使用して、調査対象者における性別ごとの既婚の割合を調べてみます。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Car Poll.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**既婚 / 未婚**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**性別**」を選択し、［**X, 説明変数**］をクリックします。
5. ［**OK**］をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［**相対リスク**］を選択します。

   「相対リスクカテゴリの選択」ウィンドウが表示されます。

図 7.15「相対リスクカテゴリの選択」ウィンドウ



「相対リスクカテゴリの選択」ウィンドウについて、次の点を確認してください。

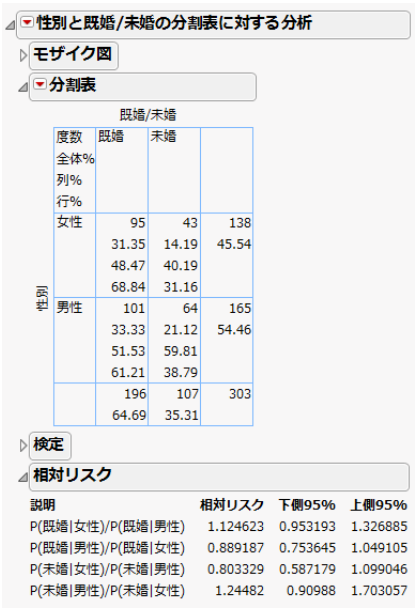
- 1 対の応答と因子についてのみ相対リスクを計算するには、このダイアログボックスで対象の変数を指定します。たとえば、図 7.15 のウィンドウで［**OK**］をクリックすると、次のように計算されます。

$$\frac{P(Y = \text{既婚} \mid X = \text{女性})}{P(Y = \text{既婚} \mid X = \text{男性})}$$

- 応答と因子のすべての組み合わせ（2 × 2 =4）について相対リスクを計算するには、［**すべての組み合わせを計算**］チェックボックスをオンにします。詳細については、図 7.16 を参照してください。

7. ［**すべての組み合わせを計算**］チェックボックスをオンにして、すべての組み合わせについて計算を行います。他のすべての設定はデフォルト値のままにしておきます。

図7.16「相対リスク」レポートの例



次の手順に従って、相対リスクがどのように計算されているのか確認してみましょう。

1. 「相対リスク」レポートの最初の項目（P(既婚 | 女性)/P(既婚 | 男性)）を調べます。
2. P(既婚 女性) および P(既婚 | 男性) といった確率は、分割表に表示されています。これらの確率は、分割表では、女性および男性の各行における、既婚の「行 %」に相当します。

   P(既婚 | 女性) = 0.6884

   P(既婚 | 男性) = 0.6121

したがって、計算式は次のようになります。

$$P(\text{既婚} \mid \text{女性})/P(\text{既婚} \mid \text{男性}) = \frac{0.6884}{0.6121} = 1.1247$$

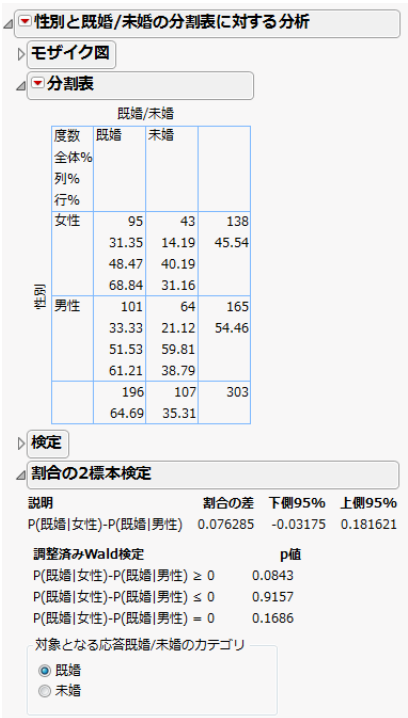
割合の 2 標本検定の例

この例では、「Car Poll.jmp」サンプルデータを使用して、女性と男性について既婚者の割合を調べてみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Car Poll.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**既婚 / 未婚**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**性別**」を選択し、［**X, 説明変数**］をクリックします。

5. ［OK］をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［割合の２標本検定］を選択します。

図7.17「割合の２標本検定」レポートの例



この例では、女性と男性で既婚者の割合を比較します。「分割表」の「行 %」で次の値を確認します。

P(既婚 | 女性) = 0.6884

P(既婚 | 男性) = 0.6121

2つの割合の差は0.0763で、これがレポートの「割合の差」に表示されています。割合の差に対する両側信頼区間は、[-0.03175, 0.181621] となっています。この信頼区間に対応した修正Wald法による両側検定のp値は、0.1686 となっています。これは、Pearsonのカイ2乗検定によるp値（0.1665）に近い値です。一般に、割合の差の検定には、修正Wald検定よりも、Pearsonのカイ2乗検定の方がよく使用されます。

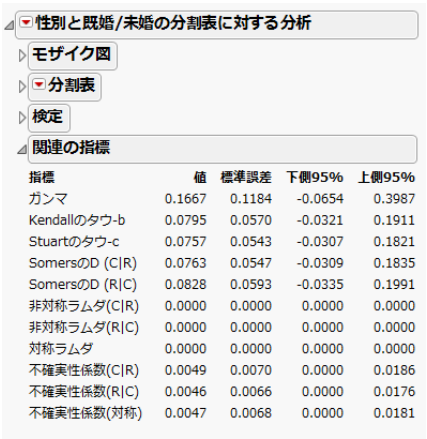
［関連の指標］オプションの例

この例では、「Car Poll.jmp」サンプルデータを使用して、女性と男性について既婚者の割合を調べてみましょう。

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Car Poll.jmp」を開きます。

2. ［分析］>［二変量の関係］を選びます。
3. 「既婚 / 未婚」を選択し、［Y, 目的変数］をクリックします。
4. 「性別」を選択し、［X, 説明変数］をクリックします。
5. ［OK］をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［関連の指標］を選択します。

図 7.18「関連の指標」レポートの例


性別と既婚/未婚の分割表に対する分析
モザイク図
分割表
検定
関連の指標

指標	値	標準誤差	下側95%	上側95%
ガンマ	0.1667	0.1184	-0.0654	0.3987
Kendallのタウ-b	0.0795	0.0570	-0.0321	0.1911
Stuartのタウ-c	0.0757	0.0543	-0.0307	0.1821
SomersのD (C|R)	0.0763	0.0547	-0.0309	0.1835
SomersのD (R|C)	0.0828	0.0593	-0.0335	0.1991
非対称ラムダ(C|R)	0.0000	0.0000	0.0000	0.0000
非対称ラムダ(R|C)	0.0000	0.0000	0.0000	0.0000
対称ラムダ	0.0000	0.0000	0.0000	0.0000
不確実性係数(C|R)	0.0049	0.0070	0.0000	0.0186
不確実性係数(R|C)	0.0046	0.0066	0.0000	0.0176
不確実性係数(対称)	0.0047	0.0068	0.0000	0.0181

分析の対象である変数（性別と既婚 / 未婚）が名義尺度なので、ラムダと不確実性係数を使用します。いずれの値も小さくなっているので、関連性は弱いようです。

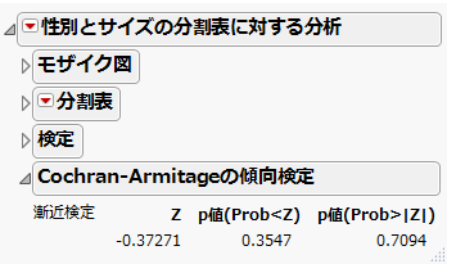
Cochran-Armitage の傾向検定の例

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Car Poll.jmp」を開きます。

この検定の目的に合わせて、「サイズ」を順序尺度に変更します。

2. 「列」パネルで、「サイズ」の横にあるアイコンを右クリックし、［順序尺度］を選択します。
3. ［分析］>［二変量の関係］を選びます。
4. 「性別」を選択し、［Y, 目的変数］をクリックします。
5. 「サイズ」を選択し、［X, 説明変数］をクリックします。
6. ［OK］をクリックします。
7. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから［Cochran-Armitage の傾向検定］を選択します。

図7.19「Cochran-Armitageの傾向検定」レポートの例



両側検定のp値（0.7094）は大きい値になっています。このことから、各サイズの車両を購入する男性と女性の割合に関係があるとは結論できません。

「分割表」プラットフォームの統計的詳細

ここでは、「分割表」プラットフォームのいくつかの統計手法の詳細について説明します。

［一致性の統計量］オプションの統計的詳細

2名の評価者が、n個の対象を判定したとします。2人の判定者の判定が完全に一致している場合は、カッパ係数が1になります。観測された一致が、偶然によって一致する期待値を上回ると、カッパ係数は正の値になります。カッパ係数が大きいほど、一致性の高いことを示します。実際にはほとんどあり得ませんが、観測された一致が、偶然によって一致する期待値を下回ると、カッパは負になります。カッパ係数の最小値は、周辺度数に応じて-1から0までの間の値を取ります。

カッパ統計量は、次式により計算されます。

$$\hat{\kappa} = \frac{P_0 - P_e}{1 - P_e} \text{ この式で、} P_0 = \sum_i p_{ii} \text{ および} P_e = \sum_i p_{i.} p_{.i}$$

単純なカッパ係数の漸近分散は、次のような推定式で計算されます。

$$\text{var} = \frac{A + B - C}{(1 - P_e)^2 n} \text{ この式で} A = \sum_i p_{ii}[1 - (p_{i.} + p_{.i})(1 - \hat{\kappa})]^2 \text{、} B = (1 - \hat{\kappa})^2 \sum_{i \neq j} \sum p_{ij} (p_{.i} + p_{j.})^2 \text{ および}$$

$$C = [\hat{\kappa} - P_e(1 - \hat{\kappa})]^2$$

Fleiss, Cohen, and Everitt（1969）を参照してください。

対称性に対するBowker検定は、「分割表における確率が対称である（p_{ij}=p_{ji}である）」という帰無仮説を検定します。

［オッズ比］オプションの統計的詳細

オッズ比は次の式で計算されます。

$$\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}$$

上の式で、p_{ij} は 2×2 分割表の第 i 行第 j 列の度数です。

「検定」レポートの統計的詳細

R2 乗 (U)

R2 乗 (U) は次の式で計算されます。

$$\frac{\text{モデルの「(-1)* 対数尤度」}}{\text{全体の修正済み「(-1)* 対数尤度」}}$$

分母の「(-1)* 対数尤度」は、標本全体において応答確率が一定であるモデルのものです。

検定

次の 2 種類の検定が行われます。

「尤度比」カイ 2 乗検定の検定等計量は、「検定」レポートに表示されている「(-1)* 対数尤度」を 2 倍した値です。この「(-1)* 対数尤度」は、グループごとに応答確率をあてはめたモデルの対数尤度から、標本全体において応答確率が一定であるモデルの対数尤度を引いたものです。尤度比カイ 2 乗値は、G^2 という記号で示されることがあり、次式で求めます。

$$G^2 = 2\left[\sum_{ij}(-n_{ij})\ln(p_j) - \sum_{ij} -n_{ij}\ln(p_{ij})\right] \text{この式で、} p_{ij} = \frac{n_{ij}}{N} \text{ および } p_j = \frac{N_j}{N}$$

この式は簡略に次式のようにも表せます。

$$G^2 = 2\sum_i\sum_j n_{ij}\ln\left(\frac{n_{ij}}{e_{ij}}\right)$$

「Pearson」カイ 2 乗検定の検定統計量は、各セルにおける観測度数と期待度数の差の平方和から計算されます。Pearson のカイ 2 乗検定は、大標本では度数が正規分布に従う傾向にあるという性質を用いています。Pearson のカイ 2 乗統計量を算出するには、次式がよく知られています。

$$\chi^2 = \sum\frac{(O - E)^2}{E}$$

上式において、Oはセルの観測度数、Eはセルの期待度数です。和記号は、すべてのセルの合計を表しています。2×2分割表に対してときどき採用されている連続修正は、JMPでは行っていません。

対応分析の「詳細」レポートの統計的詳細

対応分析の「詳細」レポートに表示される特異値は、次の行列の特異値です。

$$D_r^{-0.5}(\mathrm{P} - rc')\mathrm{D}_c^{-0.5}$$

ここで、

- Pは、各セルの度数を合計度数で割ったものを要素としてもつ行列。
- rはPの行和。cはPの列和。
- Dは、rとcを対角要素とする対角行列。

第 8 章

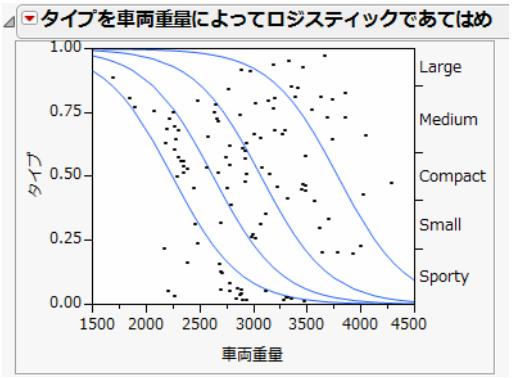
ロジスティック分析

カテゴリカルなＹ変数と連続尺度のＸ変数の関係を調べる

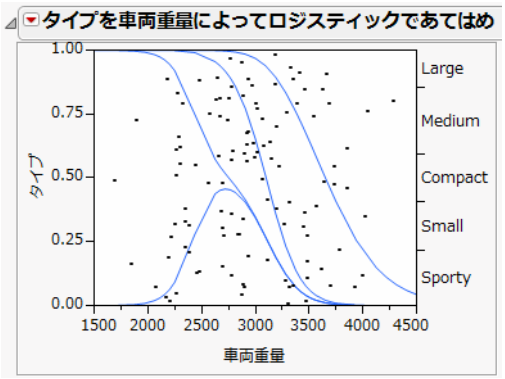
「ロジスティック」プラットフォームは、カテゴリカルな応答変数（Y）の確率を、連続尺度の説明変数（X）にあてはめます。モデルをあてはめた結果、各 X 値における応答変数の確率を予測する式が作成されます。「ロジスティック」プラットフォームは、「二変量の関係」プラットフォームのうち、応答変数が**名義尺度**もしくは**順序尺度**で、説明変数が**連続尺度**の場合に実行される分析です。応答変数が名義尺度のときには名義ロジスティック回帰、順序尺度のときには順序ロジスティック回帰が行われます。

- 名義ロジスティック回帰では、応答変数の値ごとに、各値が生じる確率のロジスティック曲線が計算されます。
- 順序ロジスティック回帰では、応答変数の値が与えられた値以下になる確率がモデル化されます。そのため、ロジスティック曲線は1つだけ計算され、それを水平に移動させることですべての順序カテゴリの確率が推定されます。このモデルは、名義ロジスティックよりも複雑でなく、順序尺度の応答変数に適しています。

図 8.1 ロジスティック回帰の例



順序ロジスティック回帰



名義ロジスティック回帰

目次

ロジスティック回帰の概要

ロジスティック回帰は古くから普及している手法で、医薬品開発における用量反応データや、マーケティングの購買選択など、いろいろな分野で応用されています。非常に単純な手法であるにもかかわらず、初歩的な統計の授業では残念ながらあまり取り上げられません。ロジスティック回帰の文献（Hosmer and Lemeshow 1989）のほかに、多くのカテゴリカル分析のテキスト（Agresti 1998）で取り上げられています。正規分布の分布関数を使用する手法もあり、その場合の分析は**プロビット分析**と言います。また、同じ状況でも、カテゴリカル変数を *X*、連続変数を *Y* として考えて、ロジスティック回帰ではなく判別分析を行う人もいます。判別分析では、連続量のデータが固定された説明変数ではなく正規分布に従ったランダムな応答であると仮定されます。

単純ロジスティック回帰は、「モデルのあてはめ」プラットフォームで行われるカテゴリカル応答の分析を、単純かつグラフィカルにしたものです。より複雑なロジスティック回帰モデルについては、『基本的な回帰モデル』を参照してください。

名義ロジスティック回帰

名義ロジスティック回帰では、ある水準に応答変数がなる確率を、説明変数の滑らかな関数によって表します。あてはめた確率は 0 ～ 1 の間をとります。また、説明変数のある値に対する確率を、すべての応答変数の水準で合計すると、1 になります。

ロジスティック確率プロットでは *Y* 軸が確率を表します。応答変数の水準数を *k* とすると、**k** - 1 本の滑らかな曲線が作成され、それによって合計確率（1）が応答水準ごとに分割されます。ロジスティック回帰のあてはめでは、「観測された応答データが生じる確率の自然対数」を合計した値の符号を逆にしたものが最小化されます。この推定方法を「最尤法」と呼びます。

順序ロジスティック回帰

Y が順序尺度の場合、順序性を考慮したロジスティック回帰が行われます。応答変数が特定の水準以下になる累積確率がロジスティック曲線で表されます。どの水準に対しても曲線の形状は同じで、ただ水平方向に位置がずれているだけです。

順序ロジスティックモデルでは、応答変数の水準数を *r* としたとき、累積ロジスティック曲線の **r** - 1 個の異なる切片を推定しますが、傾きは 1 つです。実際に必要になることはまずありませんが、各パラメータ推定値を個別に検定することも可能です。

順序モデルは計算されるパラメータが名義モデルより少ないため、適切であれば順序モデルを使用した方が効率的です。特に、順序尺度の応答変数に水準が何百もあるときに適しています。

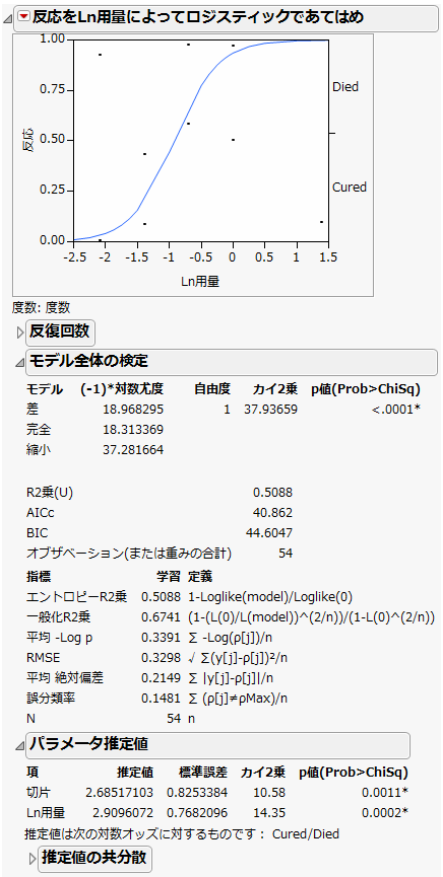
名義ロジスティック回帰の例

この例では、「Penicillin.jmp」サンプルデータを使用します。このデータは、ウサギを12匹ずつ5群に分け、連鎖球菌を注入した実験から得たものです。体内に連鎖球菌が認められたウサギに対して、用量を変えてペニシリンを投与します。ウサギの治癒率に対して、用量の自然対数（**「Ln用量」**）が効果を持つかどうかを調べます。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Penicillin.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. **「反応」**を選択し、［**Y, 目的変数**］をクリックします。
4. **「Ln用量」**を選択し、［**X, 説明変数**］を選択します。

   「度数」が自動的に［**度数**］に設定されます。これは、［**度数**］の役割が、あらかじめ「度数」列に割り当てられているためです。
5. ［**OK**］をクリックします。

図8.2 名義ロジスティックのレポートの例



あてはめられたモデルを示すグラフが作成されます。治癒の予測確率が「Ln用量」の関数として表示されます。p値は統計的に有意で、「薬の用量は、ウサギが治癒する確率に影響している」と結論できます。

「ロジスティック」プラットフォームの起動

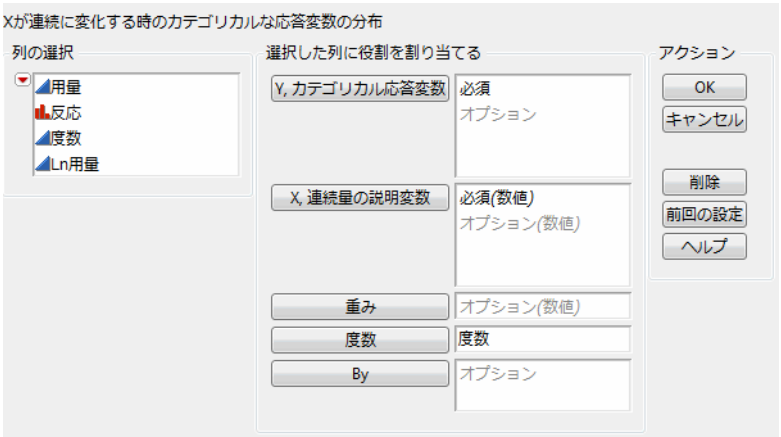
ロジスティック分析は、「二変量の関係」または「ロジスティック」プラットフォームで実行できます。どちらを使用しても結果は同じです。

- 「二変量の関係」プラットフォームを起動するには、[**分析**] > [**二変量の関係**] を選択します。

  または

- 「ロジスティック」プラットフォームを起動するには、「JMP スターター」ウィンドウのカテゴリリストで [**基本統計**] をクリックし、[**ロジスティック**] をクリックします。

図8.3「ロジスティック」起動ウィンドウ



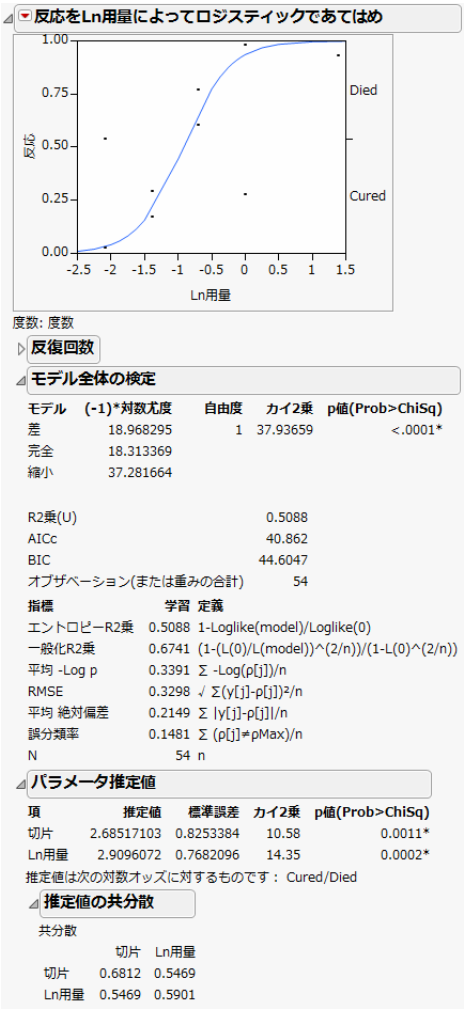
この起動ウィンドウの詳細については、「「二変量の関係」プラットフォームの概要」の章（91 ページ）を参照してください。

[**OK**] をクリックすると、「ロジスティック」レポートウィンドウが開きます。詳細については、「「ロジスティック」レポート」（234 ページ）を参照してください。

「ロジスティック」レポート

図 8.4 は、「名義ロジスティック回帰の例」（232 ページ）の手順に従って作成したものです。

図 8.4 「ロジスティック」レポートの例



「ロジスティック」レポートウィンドウには、ロジスティックプロット、「反復回数」レポート、「モデル全体の検定」レポート、「パラメータ推定値」レポートが表示されます。

注： 赤い三角ボタンをクリックするとメニューが開き、最初に表示されるレポートウィンドウに情報を追加するための各種オプションが表示されます。詳細については、「「ロジスティック」プラットフォームのオプション」（238 ページ）を参照してください。

ロジスティックプロット

ロジスティックプロットを見ると、ロジスティック回帰の全体像を把握できます。Xのそれぞれの値において、応答変数のカテゴリごとの確率にY軸が区切られています。曲線と曲線の間の縦方向の間隔が確率を表し、すべてのYカテゴリにわたって確率を合計すると1になります。

プロットで使われている変数は、ドラッグ&ドロップ操作によって、別の変数に置き換えることができます。X軸とY軸の変数を入れ替えるには、一方の軸の変数をドラッグし、他方の軸の上にドロップします。また、データテーブルの他の変数に置き換えるには、「列」パネルで該当する変数をクリックし、プロットの軸上までドラッグします。

関連情報

- 「ロジスティックプロットの例2」（242ページ）

反復履歴

「反復履歴」レポートには、各反復計算とモデルの収束判定基準が表示されます。「反復履歴」は名義ロジスティック回帰の場合にのみ表示されます。

モデル全体の検定

「モデル全体の検定」レポートには、「データ全体で応答確率が一定である」と仮定したときよりも、現在のモデルがどれぐらい良くあてはまっているかが示されます。このレポートは、連続尺度の応答に対する分散分析表に相当します。ここでは、モデルがデータを説明しているかどうかが、カイ2乗近似の尤度比検定で検定されます。観測された確率の自然対数を取り、その合計の符号を逆にしたものが**「(-1)*対数尤度」**です。カテゴリカルな応答における「(-1) *対数尤度」は、連続尺度の応答における平方和に相当します。現在のモデルの「(-1) *対数尤度」から、全体で確率が一定であると仮定したモデルの「(-1) *対数尤度」を引いた差を2倍したものが、ここでの尤度比検定のカイ2乗統計量です。この尤度比検定は、「X変数が応答変数に対して、まったく効果を持たない」という帰無仮説を検定します。

「R2乗(U)」（R^2）は0～1の値を取ります。モデルのあてはまりがよい場合はR^2値が高い値になりますが、カテゴリカルなモデルでは高い値になることはほとんどありません。

表8.1「モデル全体の検定」の説明

モデル （「要因」ともいう）	• **「縮小」**モデルは、切片だけを含むモデルです。 • **「完全」**モデルは、すべての効果と切片を含むモデルです。 • **「差」**は、完全モデルと縮小モデルの対数尤度の差です。
自由度	モデルに関係する自由度。

表 8.1「モデル全体の検定」の説明（**続き**）

(-1)* 対数尤度	データの変動やばらつきを表す指標。**不確定性**ともいう。 **「完全」**（完全モデル）は、あてはめたモデルの「(-1) * 対数尤度」（不確定性）です。ロジスティック回帰では、線形式をロジスティック変換して、確率が予測されます。この予測された確率から計算される不確定性が最小になるように、モデルがあてはめられます。 **「縮小」**（縮小モデル）は、応答確率が全体で一定であると仮定し、応答全体の割合で応答確率を推定したときの「(-1) * 対数尤度」（不確定性）です。これは、モデルに効果がないときの不確定性を表します。 上の 2 つの「(-1)* 対数尤度」の差は、モデルをあてはめた結果生じた対数尤度の減少で、それに 2 を掛けた値が尤度比カイ 2 乗です。 詳細は、『基本的な回帰モデル』を参照してください。
カイ 2 乗	尤度比カイ 2 乗。「応答確率が標本全体で一定であるとしたときよりも、モデルの適合度は良いわけではない」という帰無仮説を検定する値です。これは「差」の**「(-1)* 対数尤度」**を、2 倍したものです。つまり、あてはめたモデルの対数尤度から、標本全体で応答確率が一定であるとしたモデルの対数尤度を引いた差を、2 倍したものです。 詳細は、「「モデル全体の検定」レポートの統計的詳細」（248 ページ）を参照してください。
p 値（Prob>ChiSq）	カイ 2 乗検定の観測有意確率（*p* 値）。これは、まったくの偶然だけで、現在のカイ 2 乗値よりも大きいカイ 2 乗値を得る確率です。通常、p 値が 0.05 より小さいとモデルが有意だと考えられます。
R2 乗 (U)	不確定性のうちモデルのあてはめに起因する部分。 説明変数が応答に対して効果を持たないという仮説を検定するには、次の 2 つの数値の差を調べます。 • あてはめたモデルの対数尤度 • 水平線のモデル（切片だけのモデル）の対数尤度 詳細は、「「モデル全体の検定」レポートの統計的詳細」（248 ページ）を参照してください。
AICc	修正済みの赤池の情報量規準（Corrected AIC）。『基本的な回帰モデル』を参照してください。
BIC	ベイズ情報量規準。『基本的な回帰モデル』を参照してください。
オブザベーション （または**重みの合計**）	計算に使用したすべてのデータの標本サイズ。ただし、**重み**変数を指定した場合は、重みの合計です。

表 8.1「モデル全体の検定」の説明（続き）

指標	次の適合度指標が表示されます。 **エントロピー R2 乗**　あてはめたモデルの対数尤度と、切片だけのモデルの対数尤度を比較します。 **一般化 R2 乗**　R2 乗を、一般化した指標です。正規分布に従う連続尺度の応答変数に対する通常の R2 乗に相当します。エントロピー R2 乗に似ていますが、対数尤度ではなく尤度の 2/n 乗根が計算に使われています。 **平均 -Log p**　-log(p) の平均。p は、実際に発生したイベントの確率の予測値です。 **RMSE**　誤差の標準偏差。応答と p（実際に発生したイベントの確率の予測値）の差を示します。 **平均 絶対偏差**　応答と p（実際に発生したイベントの確率の予測値）の差の絶対値の平均。 **誤分類率**　確率の予測値が最も大きい応答カテゴリが、観測されたカテゴリではない確率。 ［エントロピー R2 乗］と［一般化 R2 乗］は、1 に近いほど適合度が良いことを示します。［平均 -Log p］、［RMSE］、［平均 絶対偏差］、［誤分類率］は、値が小さいほど、適合度が良いことを示します。
学習	適合度指標の値。
定義	適合度指標の定義式。

パラメータ推定値

名義ロジスティックモデルでは、応答水準の数を k としたとき、$k - 1$ 個の比較ごとに切片と傾きのパラメータがあてはめられます。「パラメータ推定値」レポートには、次のような推定値が表示されます。実際に必要になることはまずありませんが、各パラメータ推定値を個別に検定することも可能です。

表 8.2「パラメータ推定値」レポートの説明

項	ロジスティックモデル内のパラメータ。応答変数の各水準（最後の水準を除く）における因子ごとに、切片と傾きが表示されます。
推定値	ロジスティックモデルで計算されたパラメータ推定値。
標準誤差	各パラメータ推定値の標準誤差。各項を 0 と比較する統計量の計算に使用されます。
カイ 2 乗	各パラメータが 0 であるという仮説を検定する Wald 検定。Wald のカイ 2 乗は (推定値 / 標準誤差)^2 という式で計算されます。
p 値（Prob>ChiSq）	カイ 2 乗検定の観測有意確率。

推定値の共分散

パラメータ推定値の分散および共分散の推定値が表示されます。分散推定値の平方根は、**「標準誤差」**で表示される数値に一致します。

「ロジスティック」プラットフォームのオプション

注: Y 変数を複数指定した場合は、[あてはめのグループ] メニューが表示されます。このメニューのオプションを使って、レポートの配置を変えたり、適合度の良い順に並べ替えたりできます。詳細については、『基本的な回帰モデル』を参照してください。

次の表では、「ロジスティックであてはめ」の赤い三角ボタンをクリックすると表示されるオプションについて説明します。

オッズ比	「パラメータ推定値」レポートにオッズ比を追加します。詳細については、『基本的な回帰モデル』を参照してください。 このオプションは、応答の水準数が 2 つの場合のみ使用できます。
逆推定	与えられた Y の値から X の値を推定します。詳細は、「逆推定」（240 ページ）を参照してください。 このオプションは、応答の水準数が 2 つの場合のみ使用できます。
ロジスティックプロット	ロジスティックプロットの表示／非表示を切り替えます。
プロットのオプション	「プロットのオプション」サブメニューには次のオプションが表示されます。 **点の表示**　点の表示／非表示を切り替えます。 **応答率折れ線の表示**　各 X 値に複数の点がある場合に役立つコマンドです。各 X 値のグループごとに応答の割合が計算されます。このように計算された応答の割合を、ロジスティック曲線と比較することができます。度数が少なくて応答の割合が 0 か 1 になってしまうグループが多数生じるのを防ぐため、点が 3 つ以上ある X 値のグループに対してだけ、応答の割合は表示されます。 **線の色**　プロットの曲線の色を指定できます。
ROC 曲線	受診者動作特性（ROC）曲線は、X の値ごとに、「1- 特異度」に対して感度をプロットしたものです。詳細については、「ROC 曲線」（239 ページ）を参照してください。

リフトチャート	このコマンドを選択すると、リフトチャートが描画されます。リフトチャートは、ROC 曲線と同じ情報を使って計算されますが、順序付けした情報の上位の部分に特に重点を置きます。*Y* 軸には、予測確率が上位のグループにおいて、該当する応答水準をもつ割合がどれぐらい高いかが、集団全体における割合に対する比として表示されます。リフトチャートの詳細については、『基本的な回帰モデル』を参照してください。
確率の計算式の保存	データテーブル内の新しい列に計算式が保存されます。詳細については、「確率の計算式の保存」（240 ページ）を参照してください。
スクリプト	このメニューには、すべてのプラットフォームに共通するスクリプト関連のコマンドがリストされています。分析をやり直すコマンド、分析の JSL コマンドをウィンドウやファイルに保存するコマンドなどがあります。詳細は、『JMP の使用法』を参照してください。

ROC 曲線

ある疾患にかかっているかどうかを、次のように *X* の値によって診断するとします。

- *X* の値が特定の閾値より大きいときは、その疾患にかかているとします。
- *X* の値が閾値より小さいときは、その疾患にかかっていないと診断するとします。

たとえば、特定の種類のガンを診断する検査では、ある血液成分の濃度を測定します。診断における閾値をいろいろな値に変更し、それに伴って変化した偽陽性率と偽陰性率をプロットしてみれば、診断の良し悪しが分かります。分析の目的は、真陰性と真陽性を最も高める *X* の基準値を求めることです。受診者動作特性（ROC）曲線を見ると、これらの値がどれだけ急激に変化するかがわかります。理想的な診断では、ROC 曲線における曲線下の面積（AUC: Area Under Curve）が大きくなります。

次の 2 つは、医学の分野で使用されている概念です。

- **感度**　ある状態が存在しているときに、その状態が存在していることを、与えられた *X* の値によって（診断や基準値によって）正確に予測する確率。*X* の値が与えられているときに、状態が本当は存在しているのに存在していないと予測される確率は、「1-感度」です。
- **特異度**　ある状態が存在していないときに、その状態が存在しないことを、診断によって正確に予測する確率。

受診者動作特性（ROC）曲線は、*X* の値ごとに、「1-特異度」に対して感度をプロットしたものです。ROC 曲線より下の面積 (AUC: Area Under Curve) は、曲線によって示唆された情報を要約するための指標として使用されます。

単純ロジスティック回帰の応答が 2 値の場合は、ROC 曲線コマンドがプラットフォームのポップアップメニューに表示されます。[**ROC 曲線**] オプションを選択すると、どちらの水準を**「陽性」**にするかを指定するウィンドウが開きます。

検査による予測が完全であれば、1つの値を境にして、その上にすべての異常な標本値が位置し、下にすべての正常な値が位置するようになります。その場合、曲線は完璧な感度を持ち、グリッド上の（0,1）の点を通ります。ROC曲線がこの理想の点に近ければ近いほど、その判別能力は高くなります。逆に、予測することがまったくできない場合は、ROC曲線はグリッドの対角線上を通ります（DeLong, et al. 1988）。

ROC曲線は、偽陽性率と偽陰性率の関係をグラフ化したものです。関係は通常、レポートのプロットの下に表示される曲線下の面積（AUC）に基づいて評価されます。プロット内には、ROC曲線に接する45度の線が黄色で表示されます。黄色い線と曲線との交点は、偽陰性と偽陽性のコストが等しいと仮定したときに、コストが最小になる点を示しています。

関連情報

- 「ROC曲線の例」（244ページ）

確率の計算式の保存

［確率の計算式の保存］オプションを選択するとデータテーブルに新しい列が作成されます。これらのデータテーブル列には、次の情報が保存されます。

- *X*因子の線形結合式（ロジット）
- 各水準が生じる確率の予測式
- 最も生じる確率が高い応答変数の値を計算する予測式

逆推定

逆推定は予測の反対で、与えられた*Y*の値から*X*の値を推定するものです。ただし、ロジスティック回帰では、*Y*の値ではなく*Y*の1水準の確率を使います。逆推定は、応答のカテゴリが2つ（2値応答）の場合にしか機能しません。

「モデルのあてはめ」プラットフォームにも、逆推定の信頼限界を計算するオプションがあります。逆推定の詳細については、『基本的な回帰モデル』を参照してください。

関連情報

- 「十字ツールを使用した逆推定の例」（245ページ）
- 「［逆推定］オプションを使用した逆推定の例」（246ページ）

ロジスティック回帰のその他の例

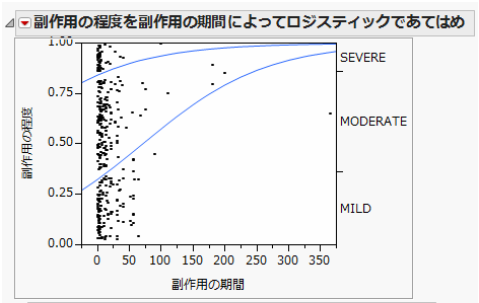
ここでは、ロジスティック回帰を使った例をさらに紹介します。

順序ロジスティック回帰の例

この例では、「AdverseR.jmp」サンプルデータを使用します。副作用の程度を、治療期間の関数として、順序ロジスティック回帰によりモデル化します。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「AdverseR.jmp」を開きます。
2. 「**副作用の程度**」の左側のアイコンを右クリックし、順序尺度に変更します。
3. ［**分析**］>［**二変量の関係**］を選びます。
4. 「**副作用の程度**」を選択し、［**Y, 目的変数**］をクリックします。
5. 「**副作用の期間**」を選択し、［**X, 説明変数**］をクリックします。
6. ［**OK**］をクリックします。

図 8.5 順序ロジスティックのレポートの例



副作用の程度を副作用の期間によってロジスティックであてはめ

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	3.59292	1	7.185838	0.0073*
完全	350.19962			
縮小	353.79254			

R2乗(U)	0.0102
AICc	706.467
BIC	718.032
オブザベーション(または重みの合計)	357

指標	学習	定義
エントロピーR2乗	0.0102	1-Loglike(model)/Loglike(0)
一般化R2乗	0.0231	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
平均 -Log p	0.9810	Σ -Log(ρ[j])/n
RMSE	0.6104	√ Σ(y[j]-ρ[j])²/n
平均 絶対偏差	0.5966	Σ |y[j]-ρ[j]|/n
誤分類率	0.4762	Σ (ρ[j]≠ρMax)/n
N	357	n

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片[MILD]	-0.7315899	0.1309775	31.20	<.0001*
切片[MODERATE]	1.6713299	0.163161	104.93	<.0001*
副作用の期間	0.0103692	0.0044661	5.39	0.0202*

このレポートの内容は、名義尺度のレポートと同様です。詳細については、「「ロジスティック」レポート」（234 ページ）を参照してください。

プロットにおける点の X 座標は、データ値をそのまま反映しています。一方、点の Y 軸は、同じ Y 値をもつデータをランダムに散らばせて表示しています。ランダムに散らばせることにより、点の間に間隔をあけ、各点を見やすくしています。

点の数が多いところでは曲線と曲線の間隔が広くなり、点の数が少ないところでは曲線どうしが接近しています。データ点が曲線の間隔を押し広げる傾向にあるのは、確率の積が最大になるようにモデルが推定されるからです。つまり、混み合った点の確率が小さくならないように、曲線は求められています。計算の詳細については、『基本的な回帰モデル』を参照してください。

「モデル全体の検定」レポートと「パラメータ推定値」レポートの詳細については、「「ロジスティック」レポート」（234 ページ）を参照してください。「パラメータ推定値」レポートを見ると、切片パラメータは最後の応答水準を除いたすべての水準について計算されていますが、傾きパラメータは 1 つしかありません。切片パラメータは応答水準間の間隔を示し、常に単調増加しています。

ロジスティックプロットの例 2

この例では、「Car Physical Data.jmp」サンプルデータを使用して、ロジスティックプロットの例をもう 1 つ紹介します。「車両重量」によって、116 台の車両のサイズ（**「タイプ」**）を予測するとしましょう。車両サイズは小さいものから順に、「Sporty」、「Small」、「Compact」、「Medium」、「Large」とします。

1. ［**ヘルプ**］ > ［**サンプルデータライブラリ**］を選択し、「Car Physical Data.jmp」を開きます。
2. **「列」**パネルで、**「タイプ」**の左側のアイコンを右クリックし、［**順序尺度**］を選択します。
3. **「タイプ」**を右クリックし、［**列情報**］を選択します。
4. 「列プロパティ」メニューから［**値の順序**］を選択します。
5. 上から順に「Sporty」、「Small」、「Compact」、「Medium」、「Large」となるようにデータを移動します。
6. ［**OK**］をクリックします。
7. ［**分析**］ > ［**二変量の関係**］を選びます。
8. **「タイプ」**を選択し、［**Y, 目的変数**］をクリックします。
9. **「車両重量」**を選択し、［**X, 説明変数**］をクリックします。
10. ［**OK**］をクリックします。

    レポートウィンドウが表示されます。

図8.6「タイプ」を「車両重量」によってロジスティックであてはめたプロットの例

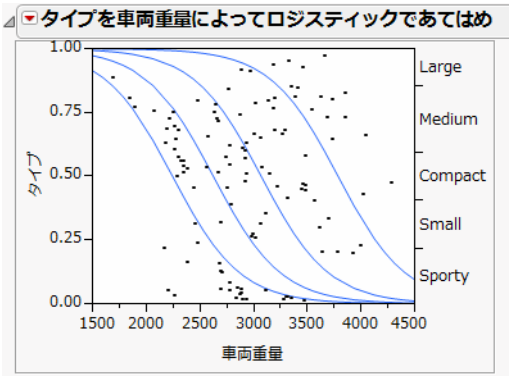


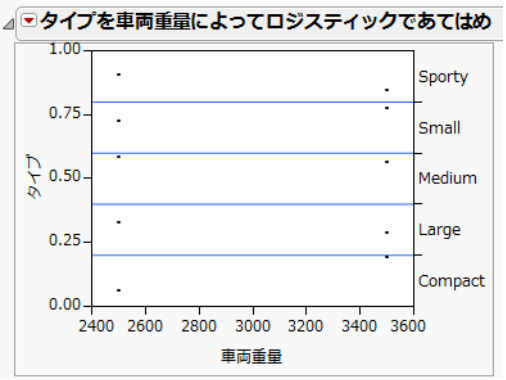
図8.6を見ると、次のことがわかります。

- 最初（一番下）の曲線は、所定の「車両重量」の車両が「Sporty」である確率を表しています。
- 2番目の曲線は、車両が「Small」または「Sporty」である確率を表しています。1番目と2番目の曲線の垂直距離が、車両が「Small」である確率です。
- プロットを見ると、予想どおり、重量が重い車両ほど「Large」である確率が高くなります。
- プロット上に描かれている点のX座標での位置はデータの値を反映していますが、Y軸方向では応答カテゴリの内側でランダムに散らばっています。

X変数が応答変数に対して効果を持たない場合はロジスティック曲線は水平になります。つまり、応答変数の各水準の確率が、説明変数の範囲全体にわたって一定になります。図8.7のロジスティックプロットが、ロジスティック曲線が水平となる例です。このような場合では、**「タイプ」**の確率を予測するのに、**「車両重量」**は役立ちません。

図8.7 予測が無意味なデータとロジスティックプロットの例

	車両重量	タイプ
1	2500	Large
2	2500	Medium
3	2500	Compact
4	2500	Small
5	2500	Sporty
6	3500	Large
7	3500	Medium
8	3500	Compact
9	3500	Small
10	3500	Sporty



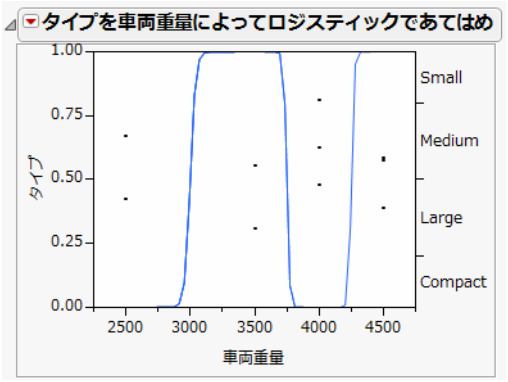
注: 図8.7および図8.8のプロットを再現するには、図中のデータテーブルを作成してから、この節の冒頭で紹介した手順7 ～ 10を実行します。

説明変数の値から応答変数の値を完全に予測できる場合、ロジスティック曲線は事実上垂直になります。応答の予測値はどの説明変数の値においてもほぼ確か（確率がほとんど1）です。図8.8のロジスティックプロットでは、「**車両重量**」が「**タイプ**」をほぼ完全に予測しています。

注: この場合、パラメータ推定値が非常に大きくなり、回帰レポート上に「**不安定**」と表示されます。

図8.8 予測が完全となるデータとロジスティックプロット

	車両重量	タイプ
1	2500	Small
2	2500	Small
3	3500	Compact
4	3500	Compact
5	4000	Medium
6	4000	Medium
7	4000	Medium
8	4500	Large
9	4500	Large
10	4500	Large



ROC 曲線の例

ROC曲線を表示するには、次の手順に従います。

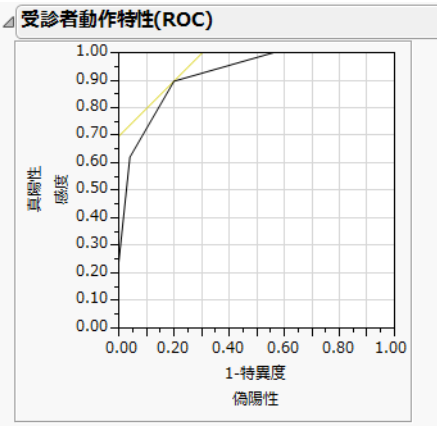
1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Penicillin.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**反応**」を選択し、［**Y, 目的変数**］をクリックします。
4. 「**Ln 用量**」を選択し、［**X, 説明変数**］を選択します。

   「**度数**」が自動的に［**度数**］に設定されます。これは、［**度数**］の役割が、あらかじめ「度数」列に割り当てられているためです。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**ROC 曲線**］を選択します。
7. 「**Cured**」を陽性として選択します。
8. ［**OK**］をクリックします。

注: この例では、名義尺度の応答のROC曲線を示します。順序尺度のROC曲線の詳細については、『発展的なモデル』の対話的パーティショニングの章を参照してください。

この例では、「**反応**」を応答として、「**Ln 用量**」を説明変数としています。応答を予測する確率が、前述した説明によって計算され、それらの結果に基づきROC曲線が描かれます。なお、「ROCテーブル」では、「感度－(1 － 特異度)」が最も大きい行にアスタリスクが付けられます。

図8.9 ROC曲線と「ROCテーブル」の例



ROCテーブル

X	確率	1-特異度	感度	感度-(1-特異度)		真陽性	真陰性	偽陽性	偽陰性
.	.	0.0000	0.0000	0.0000		0	25	0	29
1.38629	0.9988	0.0000	0.2414	0.2414		7	25	0	22
0.00000	0.9361	0.0400	0.6207	0.5807		18	24	1	11
-0.69315	0.6611	0.2000	0.8966	0.6966	*	26	20	5	3
-1.38629	0.2061	0.5600	1.0000	0.4400		29	11	14	0
-2.07944	0.0334	1.0000	1.0000	0.0000		29	0	25	0
-2.07944	0.0334	1.0000	1.0000	0.0000		29	0	25	0

ROC曲線が対角線より明らかに上にあるため、モデルの予測能力は高いと言えます。

十字ツールを使用した逆推定の例

ウサギを用いた試験をもとに、50%の確率でウサギが治癒する投与量（の自然対数）を求めるとします。治癒率が50%になるときの投与量は、**ED50**（Effective Dose 50）と呼ばれています。この例では、十字ツールを使用して、グラフ上で逆推定の大まかな値を求めてみます。

治癒率が50%となる「**Ln 用量**」を求めるには、次の手順に従います。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Penicillin.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**反応**」を選択し、［**Y, 目的変数**］をクリックします。

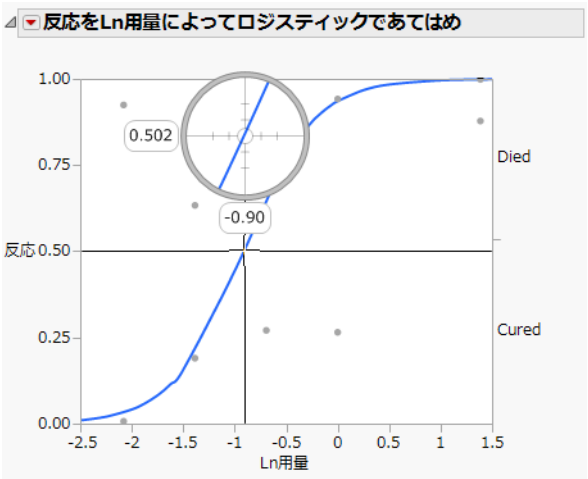
4. 「**Ln 用量**」を選択し、[**X, 説明変数**] を選択します。

   「**度数**」が自動的に [**度数**] に設定されます。これは、[**度数**] の役割が、あらかじめ「度数」列に割り当てられているためです。

5. [**OK**] をクリックします。
6. 十字ツールをクリックします。
7. まず、縦軸（「**反応**」）が 0.5 となる位置に十字ツールを合わせ、次に、ロジスティック曲線との交点まで、カーソルを動かします。
8. 十字ツールのカーソルと、ロジスティック曲線が交わった点で、横軸における「**Ln 用量**」の値を読みます。

この例では、「**Ln 用量**」の値が約 -0.9 のときに、治癒率と致死率が等しくなります。

図 8.10 ロジスティックプロット上の十字ツールの例



[逆推定] オプションを使用した逆推定の例

応答の水準が 2 つのときは、[**逆推定**] オプションを選択すれば正確な逆推定が計算されます。低い方の応答カテゴリが生じる確率に対する X 値と、その X 値の信頼区間が求められます。

[**逆推定**] オプションを使用するには、次の手順に従います。

1. [**ヘルプ**] > [**サンプルデータライブラリ**] を選択し、「Penicillin.jmp」を開きます。
2. [**分析**] > [**二変量の関係**] を選びます。
3. 「**反応**」を選択し、[**Y, 目的変数**] をクリックします。
4. 「**Ln 用量**」を選択し、[**X, 説明変数**] を選択します。

   「**度数**」が自動的に [**度数**] に設定されます。これは、[**度数**] の役割が、あらかじめ「度数」列に割り当てられているためです。

5. [**OK**] をクリックします。

6. 赤い三角ボタンをクリックし、[**逆推定**] を選択します。詳細については、図 8.11 を参照してください。
7. 「**信頼水準**」の入力ボックスに「0.95」と入力します。
8. 「信頼水準」のドロップダウンリストから「**両側**」を選択します。
9. 対象となる応答確率を指定します。この例では、「0.5」と「0.9」を入力します。つまり、治癒率が 0.5 と 0.9 であるときの「**Ln 用量**」の値を求めます。
10. [**OK**] をクリックします。

    逆推定プロットが表示されます。

図 8.11「逆推定」ウィンドウ

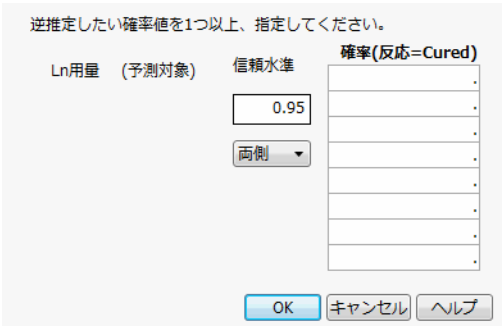
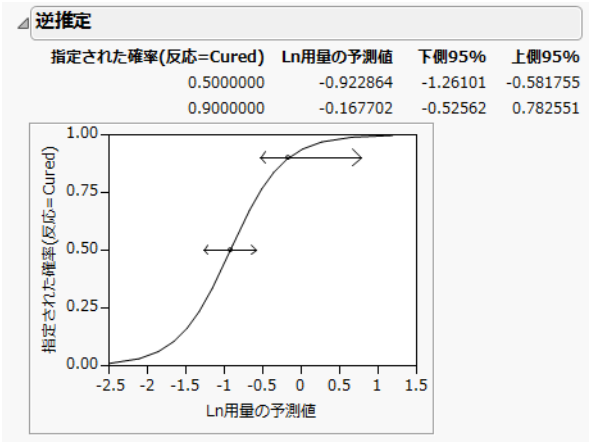


図 8.12 逆推定プロットの例



X の推定値と信頼区間がレポートと確率プロットに示されます。たとえば、治癒率が 90% になるときの「**Ln 用量**」の両側信頼区間は、-0.526 から 0.783 までとなっています。

「ロジスティック」プラットフォームの統計的詳細

「モデル全体の検定」レポートの統計的詳細

カイ 2 乗

モデル全体に対する尤度比検定のカイ 2 乗統計量は、次式により計算されます。この統計量は、教科書では、ときどき、G^2 という記号で表されています。

$$G^2 = 2(\sum -\ln p(\text{ 縮小モデル }) - \sum -\ln p(\text{ モデル }))$$

ここでの総和は、すべてのオブザベーションに対する和を表しています。

R2 乗 (U)

縮小モデルの対数尤度に対するこの統計量の比が R^2 です。式では、次のように表されます。

$$\frac{\text{差の「(-1)* 対数尤度」}}{\text{縮小モデルの「(-1)* 対数尤度」}}$$

これらの統計量の値は、「モデル全体の検定」レポートに表示されています。

第 9 章

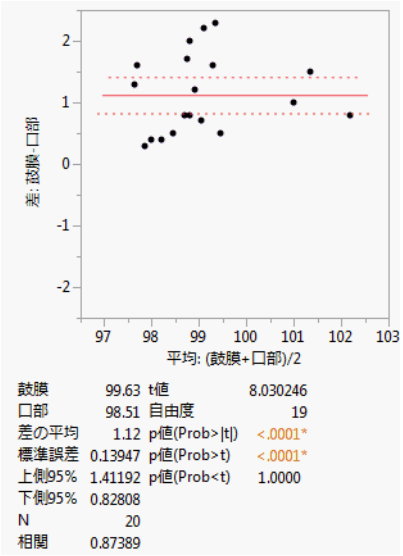
対応のあるペア分析

同一対象に関する測定値を比較する

「対応のあるペア」プラットフォームでは、相関がある 2 変数間の平均を比較し、その差を評価します。たとえば、同じ被験者から測定した、治療前後の血圧を比較します。応答に相関があることを考慮した統計手法として、**対応のある t 検定**が使用されます。

このプラットフォームでは、ペアの平均と差をプロットした散布図が描かれます。また、両側検定、下側の片側検定、上側の片側検定という 3 つの対立仮説に対する検定が実行されます。なお、3 列以上の応答を指定すると、それらの組み合わせに対して、処理が行われます。また、グループ列を指定した場合、単純な反復測定分散分析モデルに基づき、検定が行われます。

図 9.1 対応のあるペア分析の例



目次

「対応のあるペア」プラットフォームの概要

「対応のあるペア」プラットフォームは、行ごとの対応がある2つの応答変数を、対応のある *t* 検定によって比較します。対応のあるデータの例としては、同じ被験者で反復測定した処置前後データが挙げられます。また、同一の対象を、2つの異なる計器で測定したデータも該当します。

「対応のあるペア」プラットフォームを使用するには、応答データが2列に保存されていなければいけません。すべての測定値が1列に入っているときは、次のいずれかの操作を実行してください。

- ［**テーブル**］メニューの［**列の分割**］オプションで列を2つに分けます。その後、「対応のあるペア」プラットフォームを実行してください。
- 応答データが2列に含まれている場合には、まず、その差を計算する3つ目の列を作成し、次に、差の列の平均が0かどうかを「一変量の分布」プラットフォームで検定することでも、同じ分析を実行できます。
- 応答データが1列で保存されている場合は、2元配置分散分析でも、同じ分析を実行できます。1つの因子によって2つの応答のいずれであるかを、もう1つの因子によって各被験者を区別します。「二変量の関係」の「一元配置」プラットフォームでブロック変数（被験者の列）を指定するか、「モデルのあてはめ」プラットフォームで二元配置分散分析を実行します。応答を区別する因子に対する検定が、対応のある *t* 検定に一致します。

注: データに対応があるときは、対応のない *t* 検定を行わないようにしてください。データを1列に積み重ね、ブロック変数を指定せずに、「二変量の関係」で一元配置分散分析を行うのは、適切ではありません。それは、応答の間にある相関を無視して検定を行うと、応答に負の相関がある場合には効果を過大評価し、応答に正の相関がある場合には効果を過小評価してしまうからです。

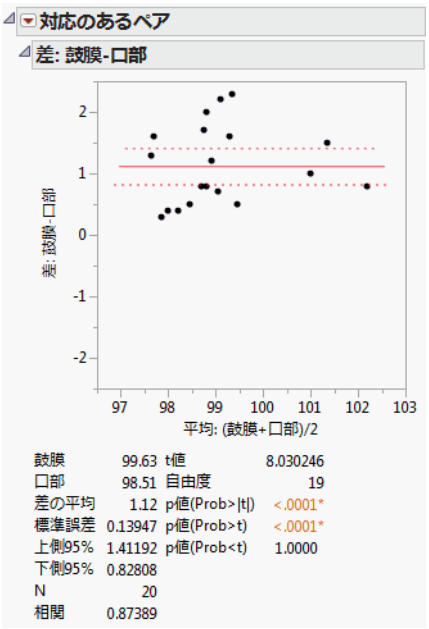
対応のあるペアの比較例

この例では、「Therm.jmp」サンプルデータを使用します。このデータには、20人の被験者に関して、口腔体温計と鼓膜体温計（耳式体温計）で計測した体温が記録されています。2種類の体温計で計測された体温が等しいかどうかを調べます。ここでは、被験者の個体差には興味がなく、2つの体温計における測定値の差に興味があります。

1. ［**ヘルプ**］**>**［**サンプルデータライブラリ**］を選択し、「Therm.jmp」を開きます。
2. ［**分析**］**>**［**対応のあるペア**］を選択します。
3. 「**口部**」と「**鼓膜**」を選択し、［**Y, 対応のある応答**］をクリックします。
4. ［**OK**］をクリックします。

   レポートウィンドウが表示されます。

図9.2「対応のあるペア」レポートウィンドウ



この結果から、鼓膜体温計の方が口腔体温計に比べ、平均して1.12度高い測定値が得られることがわかります。「p値(Prob>|t|)」が小さい値であるため、この差は偶然の結果ではなく、統計的に有意と判断できます。

なお、この分析は、どちらの体温計が正しいかを判定するものではなく、体温計間に差があることを示しているにすぎません。

「対応のあるペア」プラットフォームの起動

「対応のあるペア」プラットフォームを起動するには、［**分析**］**>**［**対応のあるペア**］を選択します。

図9.3「対応のあるペア」起動ウィンドウ

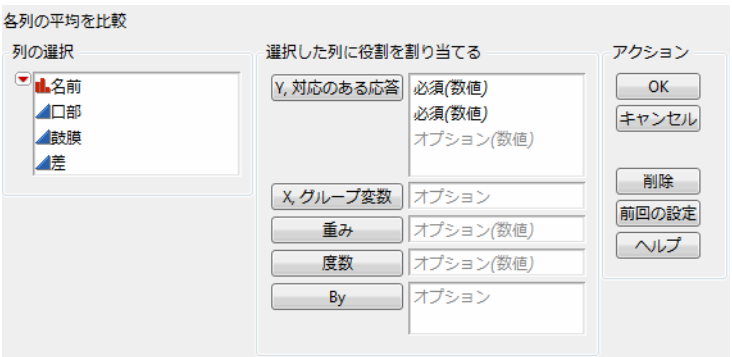


表 9.1「対応のあるペア」起動ウィンドウの説明

Y, 対応のある応答	2つの応答列を指定します。3つ以上の応答変数を分析する方法については、「複数のY列」（253ページ）を参照してください。
X, グループ変数	（オプション）グループ間の差を比較するためのグループ変数を指定します。詳細は、「グループ効果を含めた分析」（255ページ）を参照してください。
重み	（オプション）この役割を割り当てた列の数値は、分析において各行の重みとして使用されます。
度数	（オプション）この役割を割り当てた列の数値は、分析において各行の度数として使用されます。
By	（オプション）By変数の水準ごとに、個別に分析が行われます。

起動ウィンドウの詳細については、『JMPの使用法』を参照してください。

［**OK**］をクリックすると、「対応のあるペア」レポートウィンドウが開きます。詳細については、「「対応のあるペア」レポート」（254ページ）を参照してください。

複数のY列

対応するペア分析では複数の応答変数を処理できます。応答変数の数が奇数の場合、可能な組み合わせがすべて分析対象となります。次の表は、応答変数の数が3つの場合の例を示しています。

Y1-Y2	Y1-Y3
	Y2-Y3

応答変数の数が偶数の場合は、可能な組み合わせをすべて分析するかどうかを確認するメッセージが表示されます。すべての組み合わせを分析しない場合は、隣接する応答変数をペアとして分析します。次の表は、応答変数の数が4つの場合の例を示しています。

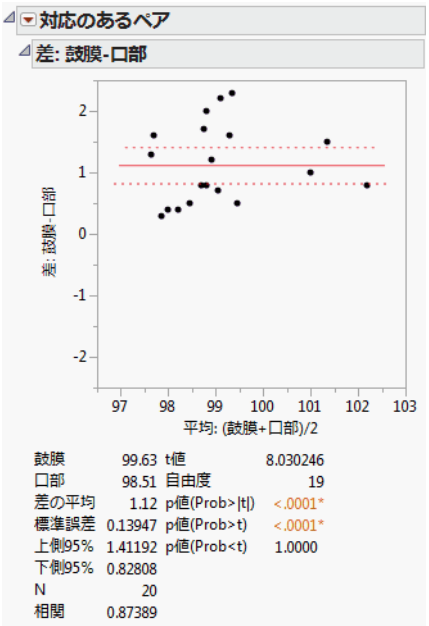
Y1-Y2	Y3-Y4

「対応のあるペア」レポート

「対応のあるペアの比較例」（251 ページ）の手順に従って作業すると、図 9.4 のレポートウィンドウが作成されます。

「対応のあるペア」レポートには、Tukey の差 - 平均プロット（Tukey diffrence-mean plot）、要約統計量、対応のある t 検定の結果が表示されます。詳細については、「「差」のプロットとレポート」（254 ページ）を参照してください。[**X, グループ変数**] を指定した場合は、「グループごと」レポートも表示されます。詳細については、「グループ効果を含めた分析」（255 ページ）を参照してください。

図 9.4「対応のあるペア」レポートの例



注: 赤い三角ボタンをクリックするとメニューが開き、最初に表示されるレポートウィンドウにレポートを追加するための各種オプションが表示されます。詳細については、「「対応のあるペア」プラットフォームのオプション」（256 ページ）を参照してください。

「差」のプロットとレポート

「差」プロットには差と平均が表示されます。「差」プロットについて、次の点を確認してください。

- 水平線が平均の差を示し、その上下に 95% 信頼区間を表す点線が表示されています。0 が信頼区間の中に挟まれている場合は、平均が 0.05 の水準において有意には異ならないことを意味します。この例では、信頼区間が 0 の水平線より上に位置しているため、差が有意であることがわかります。

- 参照枠を追加した場合は、ペアの平均は垂直線で表されます。参照枠の詳細については、「「対応のあるペア」プラットフォームのオプション」（256 ページ）を参照してください。

「差」レポートには、各応答変数の平均、差の平均、差の信頼区間が表示されます。また、対応のある t 検定の結果も表示されます。

グループ効果を含めた分析

注：「グループごと」レポートは、［**X, グループ変数**］を指定した場合にのみ表示されます。

グループ効果を含めた分析は、単純な反復測定分析です。（「モデルのあてはめ」プラットフォームで**「MANOVA」**手法を使用して反復測定分析を行うと同じ結果が出ます。）

表 9.2「グループごと」レポートの説明

差の平均	対応のある 2 列の差を計算し、それらのグループ別平均を比較したものです。言い換えると、これは個体内因子と個体間因子の交互作用（一次因子と二次因子の交互作用）を表しています。
平均の平均	対応のある 2 列の平均を計算し、それらのグループ別平均を比較したものです。言い換えると、個体間因子（一次因子）の主効果を表しています。
グループ別検定	次の 2 つに関して、グループごとの平均が異なるかを、*F* 検定によって検定します。 • **「差の平均」**は、応答の差がグループによって異なるかどうかを検定します。 • **「平均の平均」**は、応答の平均がグループによって異なるかどうかを検定します。

関連情報

- 「グループ効果を含めたペア比較の例」（257 ページ）

「対応のあるペア」プラットフォームのオプション

次の表では、「対応のあるペア」の赤い三角ボタンをクリックすると表示されるオプションについて説明します。

表 9.3

平均値と差のプロット	「平均値と差」のプロットの表示／非表示を切り替えます。このプロットの詳細については、「「差」のプロットとレポート」（254 ページ）を参照してください。
行ごとに差をプロット	「行番号ごとの差」のプロットの表示／非表示を切り替えます。
参照枠	「平均値と差」のプロットにおける参照枠の表示／非表示を切り替えます。この参照枠は、図中に傾いて表示される矩形です。参照枠は、かなり押しつぶされて表示される場合もあります。赤い縦線は、「平均の平均」を表します。参照枠は、差の範囲がデータ範囲の半分より大きい場合にはデフォルトで表示されます。
Wilcoxon の符号付順位検定	「Wilcoxon の符号付順位検定」の表示／非表示を切り替えます。Wilcoxon の符号付順位検定は、対応のあるデータの差に適用されます。これは、対応のある t 検定のノンパラメトリック版で、正の値を取る差と負の値を取る差の大きさを比較するものです。この検定では、0 の差は Pratt 法を用いて処理します。また、差の分布が対称であると仮定します。詳細は、「一変量の分布」の章の「Wilcoxon の符号付順位検定の統計的詳細」（72 ページ）を参照してください。また、Lehman（2006）および Conover（1999、350 ページ）も参照してください。
符号検定	符号検定の表示／非表示を切り替えます。これは、対応のある t 検定のノンパラメトリック版です。差の符号（正負）だけを検定に使用します。
α 水準の設定	分析で使用する α 水準を変更します。レポートとプロット内の信頼区間に影響します。
スクリプト	このメニューには、すべてのプラットフォームに共通するスクリプト関連のコマンドがリストされています。分析をやり直すコマンド、分析の JSL コマンドをウィンドウやファイルに保存するコマンドなどがあります。詳細は、『JMP の使用法』を参照してください。

グループ効果を含めたペア比較の例

この例では、「Dogs.jmp」サンプルデータを使用して、「対応のあるペア」でグループを含めた分析と、「MANOVA のあてはめ」(「モデルのあてはめ」を使用）を用いる方法を紹介します。両者の検定結果は、同じです。

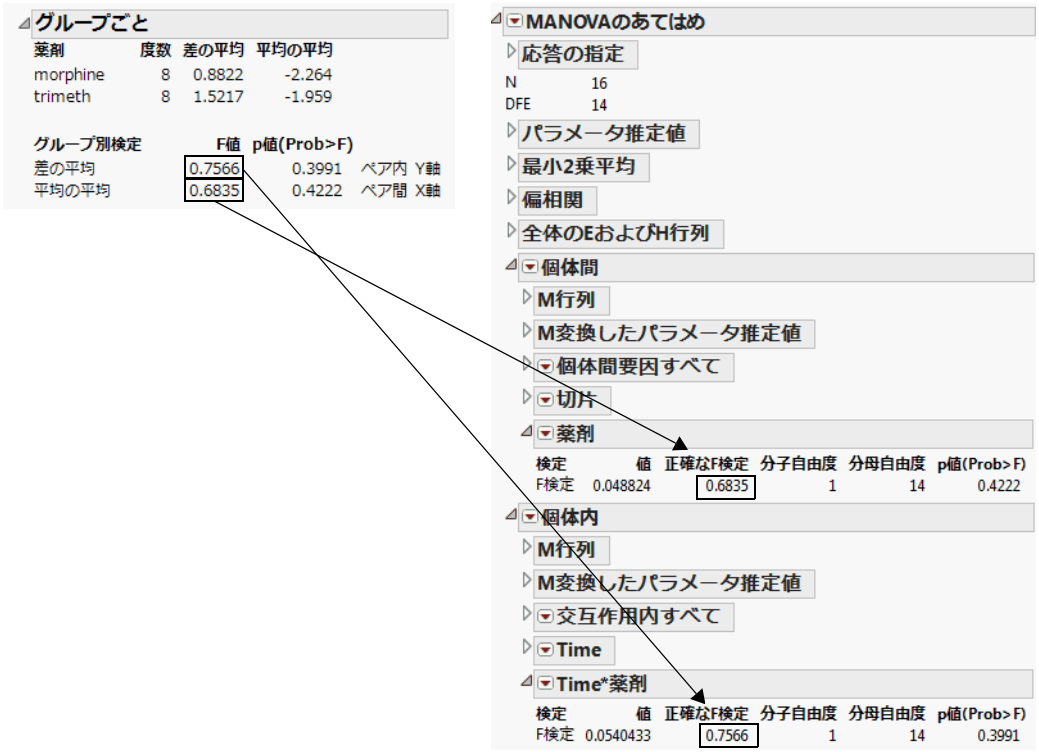
1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Dogs.jmp」を開きます。
2. ［**分析**］>［**対応のあるペア**］を選択します。
3. 「**Log(ヒスタミン0)**」と「**Log(ヒスタミン1)**」を選択し、［**Y, 対応のある応答**］をクリックします。
4. 「**薬剤**」を選択し、［**X, グループ変数**］をクリックします。
5. ［**OK**］をクリックします。

図9.5の左側のレポートが表示されます。

次に、「モデルのあてはめ」で同じデータテーブルを使用してレポートを作成します。

1. ［**分析**］>［**モデルのあてはめ**］を選択します。
2. 「**Log(ヒスタミン0)**」と「**Log(ヒスタミン1)**」を選択し、［**Y**］をクリックします。
3. 「**薬剤**」を選択し、［**追加**］をクリックします。
4. 「手法」から「**MANOVA**」を選択します。
5. ［**実行**］をクリックします。
6. 「応答の指定」レポートで、［**応答の選択**］メニューから［**反復測定**］を選択します。
7. ［**OK**］をクリックします。

図 9.5「対応のあるペア」の「グループごと」レポートと「モデルのあてはめ」の反復測定データに対する MANOVA レポートの例



「グループごと」レポートの「差の平均」の「F 値」が、「個体内」レポートの「時間*薬剤」の「正確な F 検定」の値に対応しています。「グループごと」レポートの「平均の平均」の「F 値」が、「個体間」レポートの「薬剤」の「正確な F 検定」の値に対応しています。

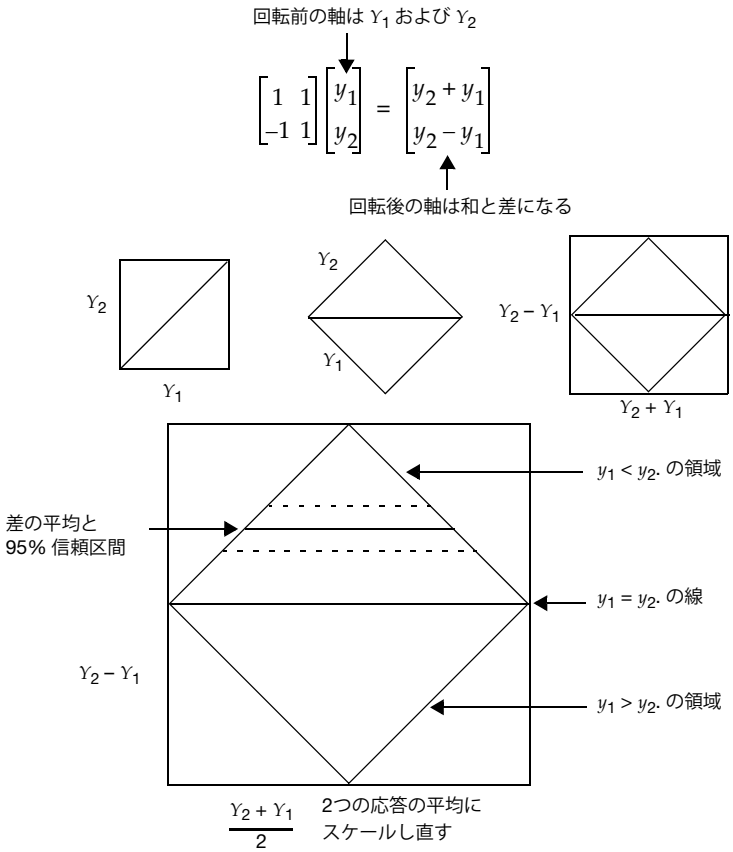
「対応のあるペア」プラットフォームの統計的詳細

ここでは、対応のあるペア分析の統計的詳細について説明します。

対応のあるペアのグラフ

このプラットフォームのグラフは、X 軸に 2 変数の平均、Y 軸に 2 変数の差をプロットしています。「Tukey の平均-差プロット」（Tukey mean-difference plot）と呼ばれています（Cleveland 1994, p. 130）。これは、2 変数の散布図を 45 度回転させたものと同じです。45 度の回転と再スケールによって元の座標軸が差と平均に変換されます。

図9.6 45度の回転による差と平均への変換例

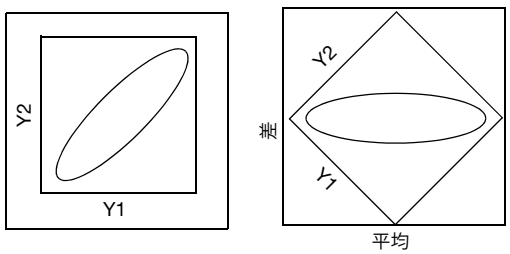


応答の相関

対応のあるデータが同じ個体から複数回にわたって測定されたものである場合は、多くの場合、正の相関があります。ただし、一方が他方を阻害する応答の場合などには、負の相関があるかもしれません。

図9.7は、2つの応答変数に正の相関があると差の分散（Y軸上）が小さくなることを示しています。負の相関があるときは楕円が反対の方向に伸び、回転させたグラフではY軸上での分散が大きくなります。

図9.7 回転前後の正の相関の例



第 10 章

ブートストラップ
標本再抽出によって統計量の分布を近似する

JMP PRO

ブートストラップは、データから標本を何度も無作為抽出することにより、統計量の標本分布を近似する手法です。反復ごとに、データは復元抽出（重複抽出）され、抽出されたデータから統計量が計算されます。この計算を反復して、統計量の分布を求めます。

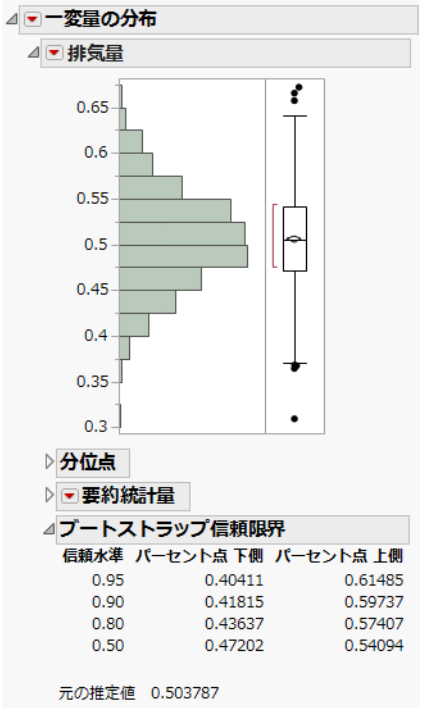
ブートストラップは、以下のような状況で、統計量（平均、標準誤差など）を推測するのに役立ちます。

- 統計量の標本分布を理論的に導出するのが複雑であるか、不明である
- 推測を行うための仮定が満たされないため、パラメトリックな推測が行えない

JMP のブートストラップは、各行が互いに独立なデータを分析するプラットフォームで実行できます。

［ブートストラップ］オプションは、通常の赤い三角ボタンのコマンドではなく、レポートの表を右クリック（コンテキストクリック）すると呼び出されるメニューに用意されています。

図10.1 ブートストラップで作成される分布の例



目次

ブートストラップの例

あるタイヤメーカーが、エンジンの排気量（in^3）からエンジンの馬力を予測しようとしています。2つの変数には直線関係があるとし、傾きについて推測するとします。直線の傾きがわかれば、排気量の増加に伴う、馬力の増加が分かります。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Car Physical Data.jmp」を開きます。
2. ［**分析**］>［**二変量の関係**］を選びます。
3. 「**馬力**」に［**Y, 目的変数**］の役割を割り当てます。
4. 「**排気量**」に［**X, 説明変数**］の役割を割り当てます。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、メニューから［**直線のあてはめ**］を選択します。
7. 「**パラメータ推定値**」レポートの「**推定値**」列を右クリックし、開いたメニューから［**ブートストラップ**］を選択します（図10.2）。

図10.2［ブートストラップ］オプションの選択

［**選択された列の分割**］オプションを選択する場合は、どの列で右クリックしたかによって、分析対象となる列が変わる点に注意してください。詳細は、「「ブートストラップ」ウィンドウのオプション」（264ページ）を参照してください。

8. 「**ブートストラップ標本数**」に「1000」と入力します。
9. ［**小数の重み**］にチェックマークをつけます。他の2つのオプションは選択したままにしておきます。
10. ［**OK**］をクリックします。

    ブートストラップ処理が行われ、傾きと切片の結果を積み重ねていないデータテーブルが作成されます。

    次に、ブートストラップによって生成された傾きの分布を分析します。
11. 積み重ねていないデータテーブルを選択して、［**分析**］>［**一変量の分布**］を選択します。

ヒント： ［一変量の分布］を実行する前に、必ず適切なデータテーブルを選択してください。そうしないと、「ブートストラップ信頼限界」レポートが表示されません。

12.**「排気量」**に［**Y, 列**］を割り当てます。

13.［**OK**］をクリックします。

「一変量の分布」レポートには、「ブートストラップ信頼限界」レポートが含まれます（図10.3）。

図10.3 ブートストラップのレポート

ブートストラップ信頼限界

信頼水準	パーセント点 下側	パーセント点 上側
0.95	0.40255	0.61002
0.90	0.41757	0.59258
0.80	0.43552	0.56803
0.50	0.46642	0.53833

元の推定値 0.503787

この例では、元のデータの傾きの推定値が0.504です。信頼水準が95%であれば、傾きは0.4026～0.6100であると推定できます。つまり、排気量が1単位増加すると、馬力は0.4026～0.6100だけ増加すると言えます。

ブートストラップ分析の実行

ブートストラップを実行するには、特定のプラットフォームを実行した後、そのレポート内の表において右クリックし、メニューから［**ブートストラップ**］を選択します。そして、ブートストラップ標本の標本数を入力し、小数の重みを使用するかどうか、また、選択した列を分割するかどうかを指定します。詳細は、「「ブートストラップ」ウィンドウのオプション」（264ページ）を参照してください。レポートに含まれるすべての統計量がブートストラップ法で処理されます。

ブートストラップでは、元データから、標本が何回も無作為抽出されます。無作為抽出される標本のデータは非表示の状態で、処理は行われます。

「ブートストラップ」ウィンドウのオプション

レポート内の列を右クリックし、開いたメニューから［**ブートストラップ**］を選択すると、次のようなオプションを含む「ブートストラップ」ウィンドウが表示されます。

ブートストラップ標本数　データからの無作為抽出と、そこからの統計量の計算といった一連の処理を反復する回数。回数を増やせば、それだけ統計量の状態をより正確に推定することができます。

小数の重み　このオプションをチェックすると、標本抽出の際に、小数の重みを用います。

［**小数の重み**］を選択しなかった場合、抽出結果は次のようになります。

- 標本抽出の際に使われる、各オブザベーションに対する重みは、整数です。
- 各オブザベーションが推定に与える影響は、相対的な重みに左右されます。たとえば、重みが2のオブザベーションは、その標本に2回出現します。
- 重みが0のこともあります。重みが0のオブザベーションは、その標本で行われる推定から除外されます。

［**小数の重み**］を選択した場合は、次のようになります。

- 重みが、整数ではなく、小数点以下の値をもつ数値になります。各オブザベーションが推定に与える影響は、相対的な重みに左右されます。
- 重みが0になることはありません。すべてのオブザベーションが推定に含まれます。

どちらの場合も、重みの合計は n（元のデータのオブザベーション数）です。

選択された列の分割　結果のデータテーブルにおいて、統計量は縦に積み重なっています。このオプションをチェックすると、統計量ごとの個別の列に分割したデータテーブルも作成します。たとえば、［**ブートストラップ**］オプションを選択するために右クリックした列が、3つの統計量を含んでいたとします。ブートストラップを行うと、3つの統計量を縦に積み重ねたテーブルのほかに、それらを3つの列に分割したテーブルが作成されます。例として、「結果を積み重ねたテーブル」（265ページ）を参照してください。

分割が成功したら積み重ねたテーブルを破棄　（［**選択された列の分割**］オプションが選択されている場合にのみ適用）［**選択された列の分割**］が成功した場合、結果を積み重ねたテーブルは表示されません。結果を積み重ねていないテーブルだけが表示されます。

結果を積み重ねたテーブル

ブートストラップを行うと、結果が縦に積み重なったデータテーブルが作成されます（図10.4）。［**分割が成功したら積み重ねたテーブルを破棄**］オプションを選択した場合、このテーブルは表示されないことがあります。

図10.4 結果を積み重ねたテーブル

	X	Y	項	~バイアス	推定値	標準誤差	t値	p値 Prob>|t|	BootID•
1	排気量	馬力	切片		50.443509471	6.7436209999	7.48	<.0001	0
2	排気量	馬力	排気量		0.5037874592	0.0398202611	12.65	<.0001	0
3	排気量	馬力	切片		49.066021402	6.7923828018	7.22	<.0001	1
4	排気量	馬力	排気量		0.5229410644	0.0403481627	12.96	<.0001	1
5	排気量	馬力	切片		44.945440879	7.2524201403	6.20	<.0001	2
6	排気量	馬力	排気量		0.5583090878	0.0402667802	13.87	<.0001	2
7	排気量	馬力	切片		46.581500604	6.339193916	7.35	<.0001	3
8	排気量	馬力	排気量		0.5409317075	0.0362753845	14.91	<.0001	3
9	排気量	馬力	切片		47.20954584	6.5502975195	7.21	<.0001	4

縦に積み重なったデータテーブルには、次のような特徴があります。

- 分析で使用したデータテーブル列が含まれます。
- ブートストラップを行った表のすべての列が含まれます。この例では、「二変量の関係」の「パラメータ推定値」表が元になっています。この表には、**「項」**、**「バイアス」**、**「推定値」**、**「標準誤差」**、**「t値」**、**「p値(Prob>|t|)」**といった列があります。
- **「BootID•」**列の値は、ブートストラップ処理の反復を識別する番号です。**「BootID•」**が 0 の行には、元の推定値が含まれています。この行には、x印のマーカーと、「除外」の行属性が与えられます。

［選択された列の分割］オプションを選択した場合は、結果を積み重ねていないデータテーブルも作成されます。詳細については、「結果を積み重ねていないテーブル」（266ページ）を参照してください。

結果を積み重ねていないテーブル

レポート上の列を右クリックして［**ブートストラップ**］を選択し、［**選択された列の分割**］を選択すると、結果を積み重ねていないデータテーブルが作成されます。その際、レポート上で右クリックした列が分割の対象となります。たとえば、図10.5では、図10.4の**「推定値」**列が、2つの列（**「排気量」**と**「切片」**）に分割されています。この2列は、**「項」**に含まれている2つの統計量です。

図10.5 結果を積み重ねていないテーブル

	X	Y	BootID•	切片	排気量
1	排気量	馬力	0	50.443509471	0.5037874592
2	排気量	馬力	1	25.672934626	0.6585617977
3	排気量	馬力	2	46.401614885	0.5020573438
4	排気量	馬力	3	50.517072009	0.4978244894
5	排気量	馬力	4	57.240010782	0.4434724344
6	排気量	馬力	5	63.399681795	0.4014883027
7	排気量	馬力	6	57.666919893	0.4119716824
8	排気量	馬力	7	51.943743427	0.4663459323
9	排気量	馬力	8	46.291323632	0.5444316084

「BootID•」列の値は、ブートストラップ処理の反復を識別する番号です。**「BootID•」**が0の行には、元の推定値が含まれています。この行には、x印のマーカーと、「除外」の行属性が与えられます。

作成した統計量データの分析

ブートストラップ抽出で作成された統計量のデータテーブルは、「一変量の分布」プラットフォームで分析してください。［**分析**］>［**一変量の分布**］をクリックし、分析したい列に、「Y」の役割を割り当てます。「一変量の分布」プラットフォームは、**「BootID•」**列がデータテーブルにある場合、「ブートストラップ信頼限界」レポートを自動的に追加します（図10.6）。

図10.6「ブートストラップ信頼限界」レポート

ブートストラップ信頼限界

信頼水準	パーセント点 下側	パーセント点 上側
0.95	0.40255	0.61002
0.90	0.41757	0.59258
0.80	0.43552	0.56803
0.50	0.46642	0.53833

元の推定値 0.503787

「ブートストラップ信頼限界」レポートには、次のような分位点が表示されます。

信頼水準 分位点に対応する累積確率。ブートストラップ・パーセント点信頼区間の信頼水準。

パーセント点 下側 信頼区間の下限。ブートストラップ抽出で作成された統計量のデータにおいて、信頼水準を p としたとき、$100((1-p)/2)$ 番目の分位点。

パーセント点 上側 信頼区間の上限。ブートストラップ抽出で作成された統計量のデータにおいて、信頼水準を p としたとき、$100(p + (1-p)/2)$ 番目の分位点。

元の推定値 元のデータから計算された統計量。

「ブートストラップ信頼限界」レポートの解釈の詳細については、「ブートストラップの例」（263ページ）を参照してください。

第 11 章

表の作成

対話的に要約テーブルを作成する

「表の作成」プラットフォームでは、記述統計量の表を対話的に作成できます。簡単かつ柔軟な方法で、要約データを表形式にできます。表は、グループ列、分析列、および各種統計量により構成されます。

図11.1 表の作成例

	性別											
	F						M					
	年齢						年齢					
	12	13	14	15	16	17	12	13	14	15	16	17
	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)
最小値	64	67	81	92	112	116	79	79	92	104	128	134
平均	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
最大値	145	112	142	112	115	116	128	105	119	128	128	172

		性別							
		女性				男性			
		既婚/未婚				既婚/未婚			
		既婚		未婚		既婚		未婚	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	サイズ	年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
米国	大型	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	中型	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	小型	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
ヨーロッパ	大型	34.0	7.071	28.0	.	.	.	26.0	.
	中型	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	小型	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
日本	大型	25.0	.	.	.	32.0	.	.	.
	中型	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	小型	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
生産国									
米国		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
ヨーロッパ		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
日本		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

		平均		
タイプ	会社規模	利益($M)	売上($M)	従業員一人あたりの利益
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
	すべて	240.87	5652.02	6159.015
Pharmaceutical	big	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
	small	156.95	1083.75	38337.19
	すべて	690.08	5070.25	23546.12
すべて	すべて	409.32	5433.86	12679.18
タイプ				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
すべて		409.32	5433.86	12679.18

目次

「表の作成」プラットフォームの使用例

男子生徒と女子生徒の身長を測定したデータがあります。このデータを基に、男子生徒と女子生徒それぞれの平均身長と全体の平均を示す 図11.2のような表を作成します。

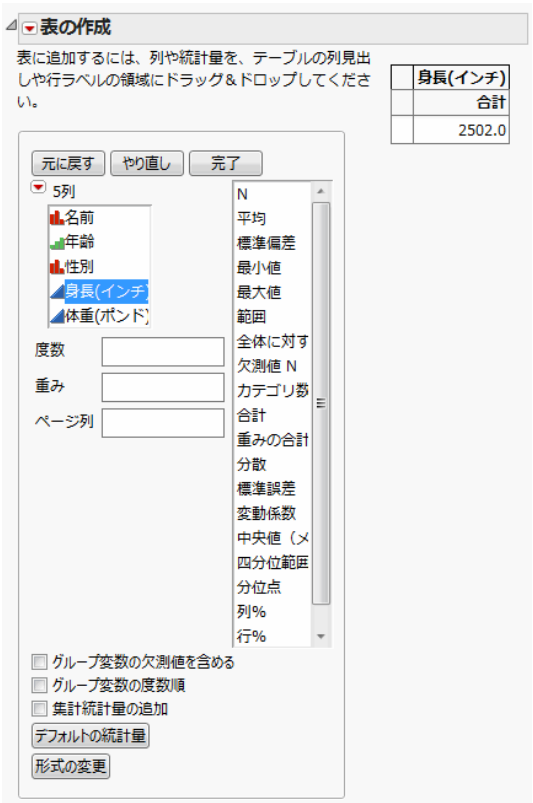
図11.2 平均身長を示す表

	身長(インチ)
性別	平均
F	60.9
M	63.9
すべて	62.6

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Big Class.jmp」を開きます。
2. ［分析］>［表の作成］を選択します。

   身長が分析対象の変数であるため、これを表の一番上に表示します。
3. **「身長(インチ)」**をクリックし、「列のドロップゾーン」にドラッグします。

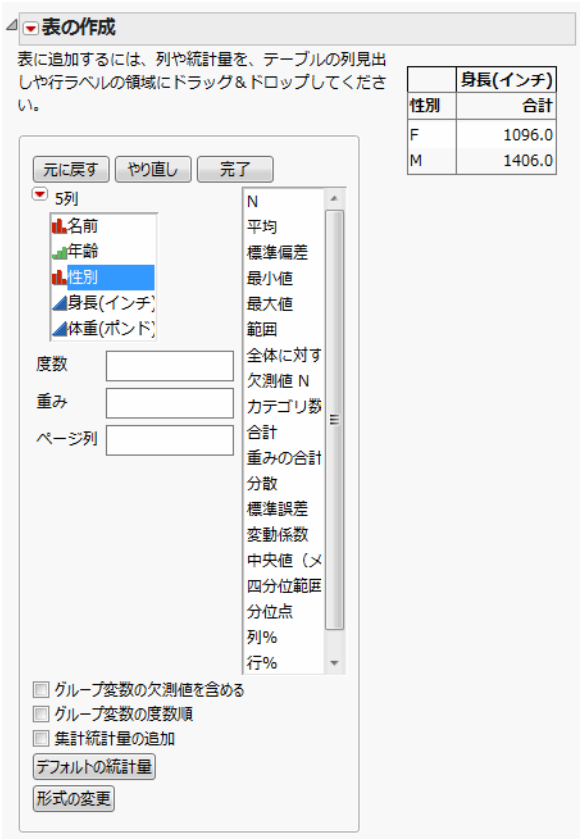
図11.3「身長」変数の追加後



性別の統計量が必要なため、性別を左側に表示します。

4. 「**性別**」をクリックしてドラッグし、数値「2502.0」の横にある空のセルにドロップします。

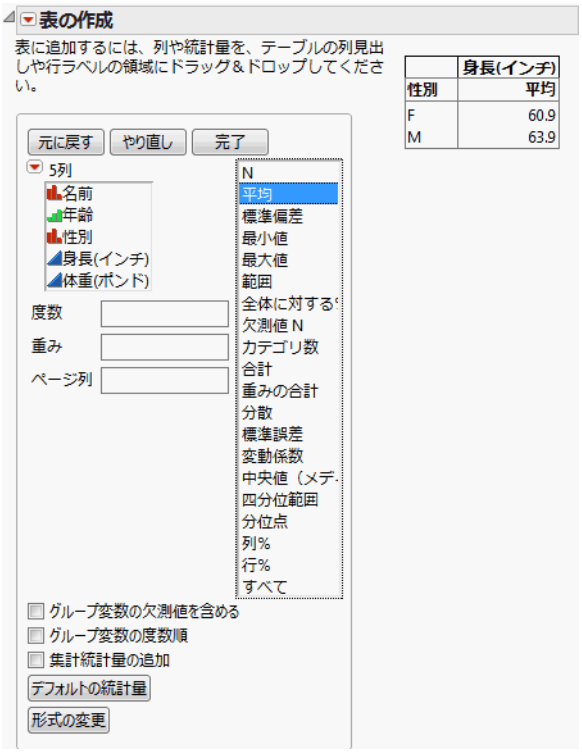
図 11.4「性別」変数の追加後



合計の代わりに平均を表示します。

5. ［**平均**］をクリックしてドラッグし、「**合計**」の上にドロップします。

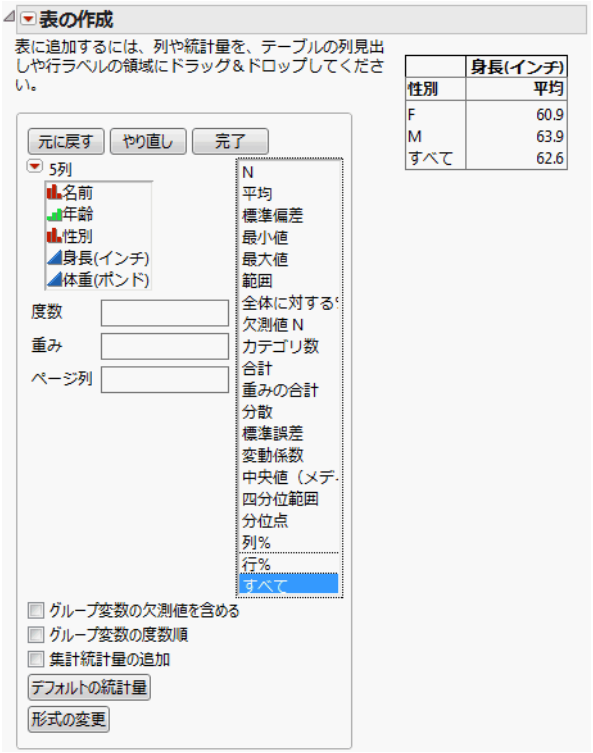
図11.5［平均］の追加後



さらに、男子と女子を合わせた全体の平均も表示します。

6. ［**すべて**］をクリックしてドラッグし、「**性別**」の上にドロップします。または、［**集計統計量の追加**］チェックボックスをオンにします。

図11.6［すべて］の追加後



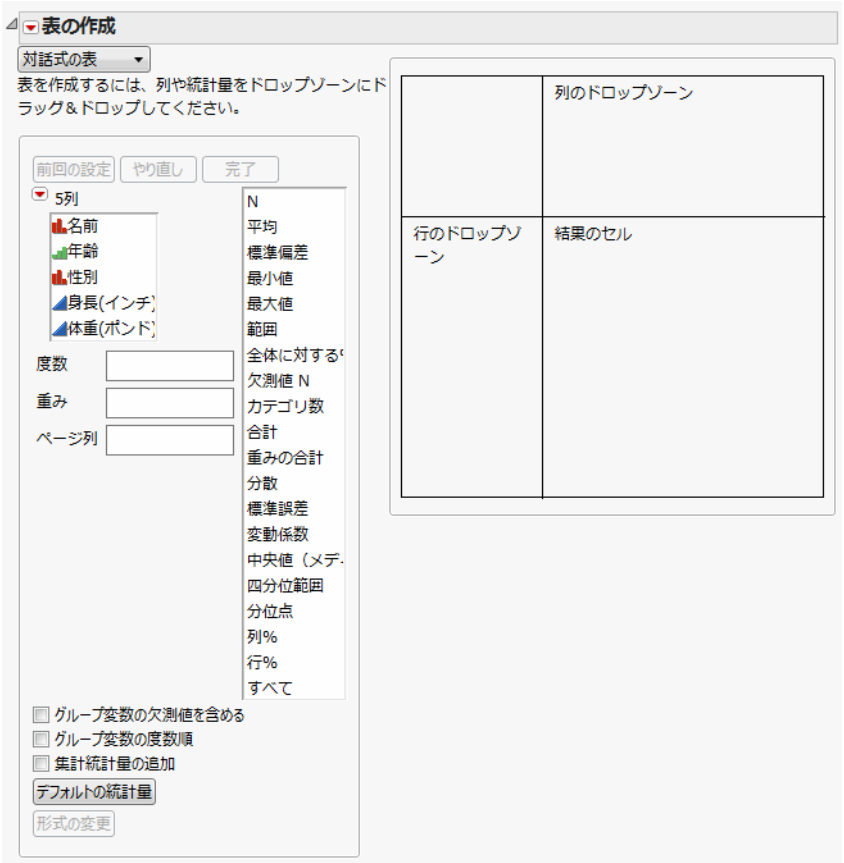
7. （オプション）［**完了**］をクリックします。

完成した表に、女子と男子の平均身長と全体の平均身長が表示されます。

「表の作成」プラットフォームの起動

「表の作成」プラットフォームを起動するには、[分析] > [表の作成] を選択します。

図11.7「表の作成」の対話的な表



注: 赤い三角ボタンのメニューのオプションについては、「「表の作成」プラットフォームのオプション」（284 ページ）を参照してください。

「表の作成」ウィンドウには、次のオプションがあります。

対話式の表／ダイアログボックス 2つのモード間で切り替えます。項目をドラッグ&ドロップしながら独自の表を作成していく場合は、[対話式の表] を、決まった形式の簡単な表を作成する場合は [ダイアログボックス] を選択します。詳細については、「ダイアログボックスの使用」(276 ページ)を参照してください。

[統計量] オプション 標準的な統計量のリストが表示されます。統計量をリストから表にドラッグして表に組み込むことができます。詳細については、「統計量の追加」（277 ページ）を参照してください。

列のドロップゾーン　列または統計量をドラッグしてこのゾーンにドロップし、列を作成します。

注:［統計量］オプションに表示されているものと同じ名前の列がデータテーブルにある場合は、必ず列リストから列名をドラッグ＆ドロップしてください。そうしないと、表で同名の統計量が代用される場合があります。

行のドロップゾーン　列または統計量をドラッグしてこのゾーンにドロップし、行を作成します。

結果のセル　ドラッグ＆ドロップした列や統計量に基づいて結果のセルが表示されます。

度数　この役割を割り当てた列の値は、各行の度数（繰り返し数）を表します。このオプションは、分析対象のデータが要約されたものである場合に、度数を割り当てるのに便利です。

重み　データに重み（重要度、影響度など）をつけるのに使う変数の列を指定します。

ページ列　名義尺度または順序尺度の列のカテゴリごとに個別の表を作成します。詳細については、「ページ列の使用例」（292ページ）を参照してください。

グループ変数の欠測値を含める　グループ変数にある欠測値を1つのグループとして集計した結果を、表に追加します。これをチェックしない場合、欠測値はテーブルに含まれません。列プロパティとして定義した欠測値のコードが考慮されます。

グループ変数の度数順　度数の大きい順に並べ直してグループを表示します。

集計統計量の追加　すべての行と列の集計統計量を追加します。

デフォルトの統計量　分析列または非分析列（グループ変数など）をドラッグ＆ドロップしたときにデフォルトで表示される統計量を変更できます。

形式の変更　特定の統計量の表示形式を変更できます。詳細については、「数値の表示形式の変更」（279ページ）を参照してください。

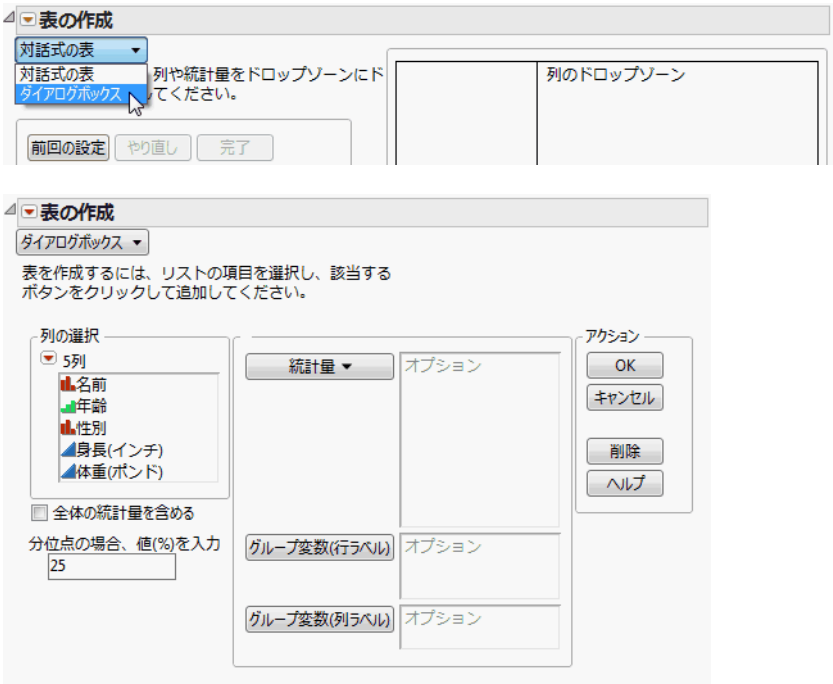
プロットスケールの変更　（赤い三角ボタンのメニューで［**チャートの表示**］が選択されている場合にのみ表示）統一したカスタムスケールを指定できます。

プロットスケールの統一　（赤い三角ボタンのメニューで［**チャートの表示**］が選択されている場合にのみ表示）このチェックボックスをオフにすると、各列のデータに基づき、それぞれの棒グラフのスケールが決められます。

ダイアログボックスの使用

ドラッグ＆ドロップを使って対話式の表を作成する代わりに、ダイアログボックスを使って簡単な表を作成することもできます。この場合、［**分析**］>［**表の作成**］の選択後に、メニューから［**ダイアログボックス**］を選択してください（図11.8を参照）。表の作成後、必要であれば赤い三角ボタンのメニューから［**設定パネルの表示**］を選択し、新しい項目を表にドラッグ＆ドロップして、より複雑な表にすることもできます。

図11.8 ダイアログボックスの使用



ダイアログボックスには、次のオプションがあります。

全体の統計量を含める　グループ列のカテゴリごとに集計した要約統計量を表示します。

分位点の場合、値(%)を入力　指定した分位点を求めます。たとえば、75% の分位点を求めた場合、データの 75% がその分位点より小さくなっています。すべてのグループ列に適用されます。

統計量　列を選択したら、その列に関して求めたい統計量をこのリストから選択します。「統計量の追加」(277 ページ）を参照

グループ変数(行ラベル)　行ラベルとして使用する列を選択します。

グループ変数(列ラベル)　列ラベルとして使用する列を選択します。

統計量の追加

ヒント： 列と統計量の両方を同時に選択して、表にドラッグできます。

表には標準的な統計量を含めることができます。使用できる統計量は、設定パネルに表示されます。列の操作と同じく、任意の統計量をリストから表にドラッグできます。次の点を念頭に置いてください。

- グループ列によって定義されるカテゴリに属する全オブザベーションから、分析列の値に基づいて統計量が計算されます。
- 統計量も、行テーブルと列テーブルのどちらにでも指定できますが、どちらかに統一しておく必要があります。
- 連続尺度の列をデータ領域にドラッグすると、分析列として扱われます。

「表の作成」では、次の統計量を使用できます。

N その列にある欠測値でないデータの数です。分析列がない場合のデフォルトの統計量です。

平均 列の値の算術平均です。つまり、欠測値以外の値の合計を［**重みの合計**］で割った値です（定義されている場合は、［**重み**］変数を掛けます）。

標準偏差 欠測値以外の値について計算される標本の標準偏差です。つまり、分散の平方根です。

最小値 欠測値を除く、列内の最小値です。

最大値 欠測値を除く、列内の最大値です。

範囲 ［**最大値**］と［**最小値**］の差です。

全体に対する % データ全体の合計に対する割合を計算します。分母はすべてのオブザベーションの合計で、分子はカテゴリごとの合計です。分析列がない場合、［全体に対する %］は度数の合計に対する割合になります。分析列がある場合、［全体に対する %］は分析列の全合計に対する割合になります。つまり、分母はオブザベーションすべての分析列の合計で、分子はカテゴリに属するオブザベーションの分析列の合計になります。テーブルに統計量のキーワードをドラッグして、異なる割合を求めることができます。

- ［全体に対する %］の見出しに、表の中から 1 つまたは複数のグループ列をドロップすると、分母の定義が変わります。このとき、各グループごとの合計が分母に使われます。
- 行テーブルと列テーブルにグループ列を指定して 2 元表を作成したとき、列合計に対する割合を取得したい場合は、行テーブルのすべてのグループ列をドラッグして、［**全体に対する %**］の見出しにドロップします（［列 %］と同様）。同様に、行合計に対する割合を取得したい場合は、列テーブルのすべてのグループ列をドラッグして、［**全体に対する %**］の見出しにドロップします（［行 %］と同様）。

欠測値 N 欠測値の数です。

カテゴリ数 カテゴリの数です。

合計 列内のすべての値の合計です。表に他の統計量が表示されていない場合の、分析列のデフォルトの統計量です。

重みの合計 列内のすべての［**重み**］の合計です。［**重み**］に割り当てられた列がない場合、［**重みの合計**］は欠測値以外の値の個数になります。

分散 欠測値以外の値について計算される分散（不偏分散）です。つまり、平均からの偏差の 2 乗の合計を、欠測値以外の値の個数から 1 を引いた数で割った値です。

標準誤差　平均の標準誤差です。つまり、標準偏差を**N**の平方根で割った値です。列に「重み」役割が割り当てられている場合、分母は重みの和の平方根です。

変動係数　ばらつきの度合いを示す指標で、標準偏差を平均で割り、100を掛けて求められます。

中央値（メディアン）　50%のパーセント点です。下半分のデータが中央値より小さく、上半分のデータが中央値より大きくなっています。

四分位範囲　第3四分位点と第1四分位点の差です。

分位点　指定した分位点を求めます。たとえば、75%の分位点を求めた場合、データの75%がその分位点より小さくなっています。異なる分位点を指定するには、［**分位点**］をクリックして表にドラッグし、表示されるボックスに分位点を入力します。

列 %　分析列がない場合は、セルの度数が列の合計度数に占める割合を示します。分析列がある場合、［列 %］は分析列の値の列合計に対する割合になります。統計量を上部に配置した表の場合、複数の行テーブルがある表（縦に積み重ねて表示されているもの）に［列 %］を追加できます。

行 %　分析列がない場合は、セルの度数が行の合計度数に占める割合を示します。分析列がある場合、［行 %］は分析列の値の行合計に対する割合になります。統計量を左側に配置した表の場合、複数の列テーブルがある表（横に並べて表示されているもの）に［行 %］を追加できます。

すべて　グループ列のカテゴリごとに集計した要約統計量を表示します。

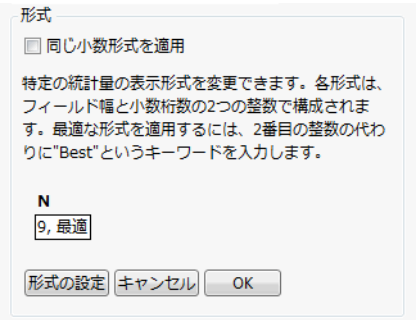
数値の表示形式の変更

各セルの形式は、分析列と統計量によって決まります。度数の場合、デフォルトでは小数点以下は表示されません。統計量を含むセルの形式は、分析列および指定した統計量の形式に基づき、適切な形式に設定されます。デフォルトの形式を変更するには、次の手順を行います。

1. 「表の作成」ウィンドウの下部にある［**形式の変更**］ボタンをクリックします。
2. 表示されるパネルで、フィールドの幅、カンマ、小数点以下の桁数を入力します。詳細については、図11.9を参照してください。
3. （オプション）最適な形式を自動的に選択するには、テキストボックスに**「最適」**と入力します。

   すると、各セル値の精度を考慮した上で、その値を表示するのに最適な表示形式が選ばれます。
4. 変更内容を確定し、「形式」セクションを閉じるには［**OK**］をクリックします。または、［形式の設定］をクリックすると、**「形式」**セクションが開いたまま、変更内容が反映されます。

図11.9 数値の表示形式の変更



「表の作成」の出力

「表の作成」の出力ウィンドウには、行テーブルや列テーブルが上下左右に連結された表を表示できます。列テーブルや行テーブルが1つのみの場合も考えられます。

図11.10「表の作成」の出力

	性別											
	F						M					
	年齢						年齢					
	12	13	14	15	16	17	12	13	14	15	16	17
	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)
最小値	64	67	81	92	112	116	79	79	92	104	128	134
平均	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
最大値	145	112	142	112	115	116	128	105	119	128	128	172

		性別							
		女性				男性			
		既婚/未婚				既婚/未婚			
		既婚		未婚		既婚		未婚	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	サイズ	年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
米国	大型	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	中型	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	小型	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
ヨーロッパ	大型	34.0	7.071	28.0	.	.	.	26.0	.
	中型	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	小型	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
日本	大型	25.0	.	.	.	32.0	.	.	.
	中型	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	小型	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
生産国									
米国		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
ヨーロッパ		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
日本		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

		平均		
タイプ	会社規模	利益($M)	売上($M)	従業員一人あたりの利益
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
	すべて	240.87	5652.02	6159.015
Pharmaceutical	big	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
	small	156.95	1083.75	38337.19
	すべて	690.08	5070.25	23546.12
すべて	すべて	409.32	5433.86	12679.18
タイプ				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
すべて		409.32	5433.86	12679.18

表を対話的に作成する際は、次のような作業を繰り返します。

- 項目（列または統計量）を適切なリスト内でクリックし、（行または列の）ドロップゾーンまでドラッグします。「表の編集」（283 ページ）および「列テーブルと行テーブル」（282 ページ）を参照してください。
- ドラッグ & ドロップ操作を繰り返してテーブルに適宜項目を追加します。表は、項目を追加するたびに更新されます。すでに列見出しや行ラベルがある場合は、既存の項目を基準にして別の項目の追加場所を決めます。

クリック操作やドラッグ操作の際には、次の点に注意してください。

- JMP では、尺度を基に列の役割を判断します。連続尺度の列は分析列と想定されます。詳細については、「分析列」（281 ページ）を参照してください。順序尺度または名義尺度の列はグループ列と想定されます。詳細については、「グループ列」（281 ページ）を参照してください。
- 複数の列を表にドラッグ & ドロップした場合は、次のように処理されます。
  - 複数の列が共通の値のセットを持っている場合は、1 つのテーブルに結合され、列名とそれらの列が持つカテゴリから成る集計表が生成されます。この時、各セルは、指定された各列と、それらの列の各値から構成されます。
  - 複数の列が共通の値のセットをもっていない場合は、個別のテーブルに分かれます。
  - このデフォルトの結果は、列を右クリックして［**テーブルの結合**］または［**テーブルを分離**］を選択すれば変更できます。詳細は、「列の右クリックメニュー」（285 ページ）を参照してください。
- 列を入れ子状にする場合は、表に第 1 列を作成してから、追加の列をその第 1 列にドラッグします。
- 適切に作成された表では、グループ列、分析列、統計量がそれぞれまとまって配置されます。分析列のリストの中に統計量が配置されたり、グループ列のリストの中に分析列が挿入されたりすることはありません。
- 「表の作成」上の設定パネルだけではなく、データテーブルのテーブルパネルからも、「表の作成」テーブルに列をドラッグすることができます。

分析列

分析列は、統計量の計算に使用する任意の数値列であり、連続尺度です。グループ列によって区別されるカテゴリごとに、分析列の統計量が計算されます。

分析列は、行テーブルと列テーブルのどちらにでも指定できますが、どちらかに統一しておく必要があります。

グループ列

データをカテゴリ別に分類するために使う列を、グループ列と呼びます。文字、整数値、小数値などを使用できますが、カテゴリ数は制限する必要があります。グループ列は名義尺度または順序尺度です。

次の点を念頭に置いてください。

- グループ列が入れ子状になっている場合は、列の値の入れ子階層に基づいてカテゴリが作成されます。たとえば、「性別」と「既婚／未婚」の2つのグループ列がある場合、「女性で既婚」、「女性で未婚」、「男性で既婚」、「男性で未婚」の4つのカテゴリを持つ表が作成されます。
- グループ列は、列テーブルと行テーブルのどちらでも使用できます。これらのグループ列を組み合わせて、表のセルのカテゴリが決まります。
- グループ列の値に欠測値がある場合、そのオブザベーションはデフォルトでは表に含まれません。[**グループ変数の欠測値を含める**]オプションをクリックすると、それらのオブザベーションを含めることができます。
- 欠測値として扱うコードまたは値を指定するには、「欠測値のコード」列プロパティを使用します。[**グループ変数の欠測値を含める**]オプションをクリックすると、それらのオブザベーションを含めることができます。欠測値のコードの詳細については、『JMPの使用法』を参照してください。

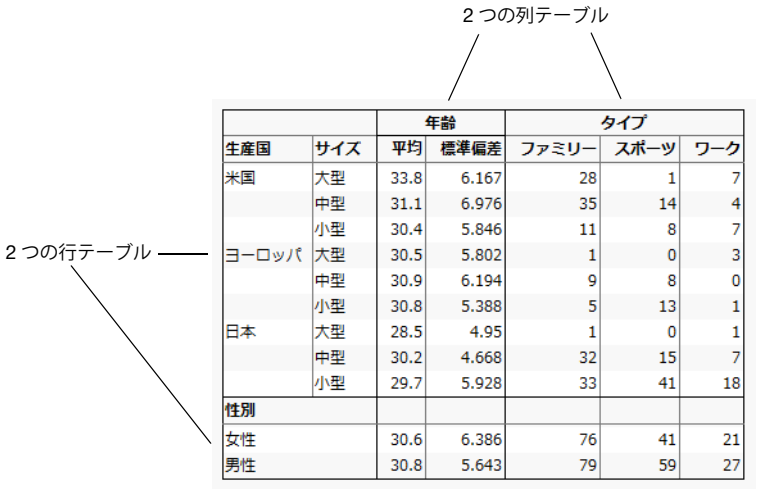
列テーブルと行テーブル

表は、列見出しと行ラベルを基準として作成されます。これらのサブテーブルは、**行テーブル**および**列テーブル**と呼ばれます。詳細については、図11.11を参照してください。

行テーブルと列テーブルの例

1. [**ヘルプ**] > [**サンプルデータライブラリ**]を選択し、「Car Poll.jmp」を開きます。
2. [**分析**] > [**表の作成**]を選択します。
3. **「サイズ」**をドラッグし、「行のドロップゾーン」にドロップします。
4. **「生産国」**をドラッグし、**「サイズ」**見出しの左側にドロップします。
5. [**平均**]をドラッグし、**「N」**見出しの上にドロップします。
6. [**標準偏差**]をドラッグし、**「平均」**見出しの下にドロップします。
7. **「年齢」**をドラッグし、**「平均」**見出しの上にドロップします。
8. **「タイプ」**をドラッグし、表の右端にドロップします。
9. **「性別」**をドラッグし、表の下にドロップします。

図11.11 行テーブルと列テーブル



		年齢		タイプ		
生産国	サイズ	平均	標準偏差	ファミリー	スポーツ	ワーク
米国	大型	33.8	6.167	28	1	7
	中型	31.1	6.976	35	14	4
	小型	30.4	5.846	11	8	7
ヨーロッパ	大型	30.5	5.802	1	0	3
	中型	30.9	6.194	9	8	0
	小型	30.8	5.388	5	13	1
日本	大型	28.5	4.95	1	0	1
	中型	30.2	4.668	32	15	7
	小型	29.7	5.928	33	41	18
性別						
女性		30.6	6.386	76	41	21
男性		30.8	5.643	79	59	27

複数の列テーブルがある場合は、左側のラベルを複数の列テーブルが共有します。この例では、「生産国」と「性別」が共有されています。同様に、複数の行テーブルがある場合は、上部の見出しを複数の行テーブルが共有します。この例では、「年齢」と「タイプ」の両方がテーブル間で共有されています。

表の編集

表に追加した項目を編集するにはいくつかの方法があります。

項目の削除

表に項目を追加した後で、次のいずれかの方法で項目を削除できます。

- 項目をドラッグし、表の外側にドロップします。
- 最後の項目を削除するには、[**元に戻す**] をクリックします。
- 項目を右クリックし、[**削除**] を選びます。

列ラベルの削除

グループ列の各カテゴリの上部には列名が表示されますが、列名が必要ない場合もあります。列テーブルから列名を削除するには、列名を右クリックし、[**列ラベルの削除**] を選びます。列ラベルを再度挿入するには、カテゴリのいずれかを右クリックし、[**列ラベルの復元**] を選びます。

統計量とラベルの編集

統計量の名前やラベルを編集できます。たとえば、「平均」の代わりに「平均値」とすることができます。編集する項目を右クリックし、[**項目ラベルの変更**] を選びます。表示されるボックスに、新しいラベルを入力します。または、名前やラベルをダブルクリックして編集ボックスへ直接入力することもできます。

統計量の名前を、別の統計量の名前に変更した場合、実際に計算される統計量も別のものになるので注意してください。別の統計量の名前にすることは、統計量をいったん表から削除し、その別の統計量を追加するのとまったく同じです。

「表の作成」プラットフォームのオプション

「表の作成」の横にある赤い三角ボタンをクリックすると、メニューに次のオプションが表示されます。

表の表示　要約データを表形式で表示します。

チャートの表示　要約統計量の表のデータを棒グラフで表示します。簡単な棒グラフで、要約統計量の相対的な大きさを視覚的に比較できます。デフォルトでは、すべての列の棒グラフに同じスケールが適用されます。[**プロットスケールの統一**] チェックボックスをオフにすると、各列のデータに基づき、それぞれの棒グラフのスケールが決まります。[**プロットスケールの変更**] ボタンを使用して、統一したカスタムスケールを指定できます。チャートの 0 の線は、左端か中央に位置します。データがすべて負か正のどちらかに統一されている場合は、0 を基準線としてチャートが作成され、それ以外の場合は、チャートの中央が 0 になります。

設定パネルの表示　インタラクティブに操作するための設定パネルを表示します。

濃淡の表示　表内に複数の行がある場合に、グレーの濃淡を適用します。

ツールヒントの表示　表上でマウスを動かすと、関連項目のヒントが表示されます。

テスト集計パネルの表示　元のテーブルから無作為に抽出した値を使ってテスト集計を行うための、テスト集計パネルを表示します。膨大な量のデータがある場合に特に便利です。詳細については、「テスト集計パネルの表示」（284 ページ）を参照してください。

データテーブルに出力　表からデータテーブルを作成します。行テーブルが複数ある場合は、それぞれのラベルの構成が異なるため、行テーブルごとにデータテーブルが作成されます。

フルパスの列名　作成されるデータテーブルの列名に、グループ列のフルパス（列名=水準名 の形式）を使用します。

スクリプト　スクリプトの保存、分析のやり直し、データテーブルの表示などのオプションがあります。詳細については、『JMP の使用法』を参照してください。

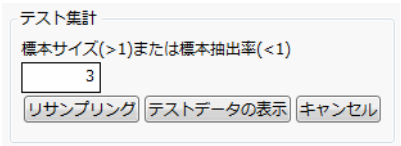
注：「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMP の使用法』を参照してください。

テスト集計パネルの表示

大量のデータを含むデータテーブルを使用する場合、そのデータテーブルのサブセットを使ってさまざまな表のレイアウトを試し、最適なレイアウトを見つけることができます。そのような場合、JMP では、標本サイズを指定して無作為抽出を行った後、表を作成します。テスト集計機能を使うには、次の手順を行います。

1. 「表の作成」の横にある赤い三角ボタンをクリックし、[**テスト集計パネルの表示**] を選びます。
2. 図 11.12 に示すように、「**標本サイズ (>1) または標本抽出率 (<1)**」の下のボックスに、標本のサイズを入力します。標本のサイズは、アクティブなテーブルに対する割合、または新しいテーブルの行数のどちらかとして認識されます。

図 11.12 テスト集計パネル



3. [**リサンプリング**] をクリックします。
4. JMP データテーブルでサンプリングされたデータを確認するには、[**テストデータの表示**] ボタンをクリックします。テスト集計パネルを非表示にすると、データテーブルのすべてのデータを使って表が再生成されます。

列の右クリックメニュー

「表の作成」で列を右クリックすると、次のオプションが表示されます。

削除　選択した列を削除します。

グループ変数として使用　分析列をグループ列に変更します。

分析列として使用　グループ列を分析列に変更します。

項目ラベルの変更　（個別の列または入れ子状の列の場合にのみ表示）新しいラベルを入力できます。

テーブルの結合 (カテゴリごとの列)　（個別の列または入れ子状の列の場合にのみ表示）個々の列または入れ子状の列を結合します。詳細については、「複数の列を持つテーブルの例」（290 ページ）を参照してください。

テーブルを分離　（結合したテーブルの場合にのみ表示）列ごとに個別のテーブルを作成します。

グループ変数を入れ子にする　グループ変数を縦または横に入れ子にします。

「表の作成」プラットフォームのその他の例

この例は、次の手順で進めます。

1. 「度数表の作成」（286 ページ）
2. 「統計量の表の作成」（288 ページ）
3. 「表の配置の変更」（289 ページ）

度数表の作成

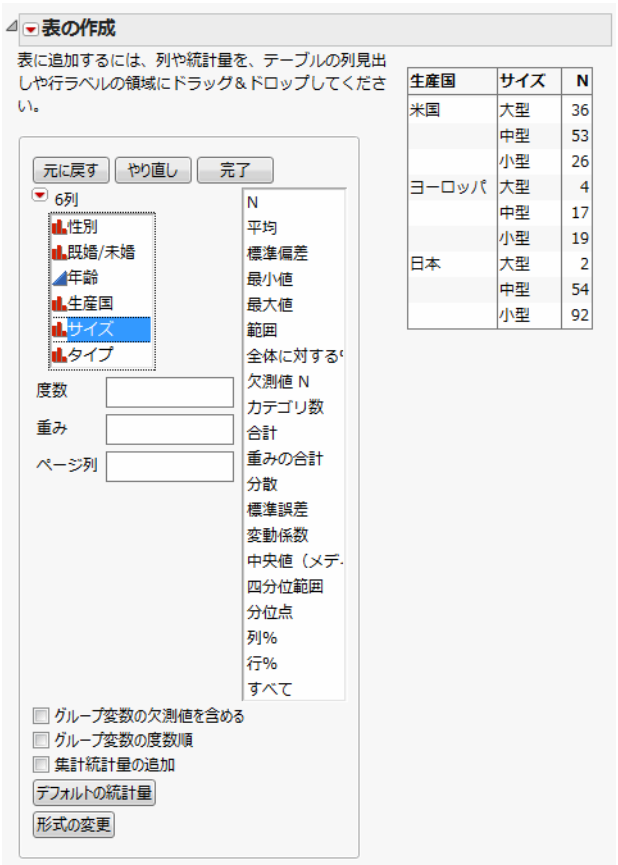
日本車、ヨーロッパ車、米国車の所有者数を車のサイズ別に表示する表を作成してみましょう。図 11.3 のような表を作成します。

図 11.13 車の所有者数を示す表

生産国	サイズ	N
米国	大型	36
	中型	53
	小型	26
ヨーロッパ	大型	4
	中型	17
	小型	19
日本	大型	2
	中型	54
	小型	92

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Car Poll.jmp」を開きます。
2. ［**分析**］>［**表の作成**］を選択します。
3. **「生産国」**をクリックし、「行のドロップゾーン」にドラッグします。
4. **「サイズ」**をクリックしてドラッグし、**「生産国」**見出しの右側にドロップします。

図11.14「生産国」と「サイズ」を表に追加



統計量の表の作成

次に、各サイズの車の所有者の平均年齢と標準偏差を表示してみましょう。図11.15のような表を作成します。

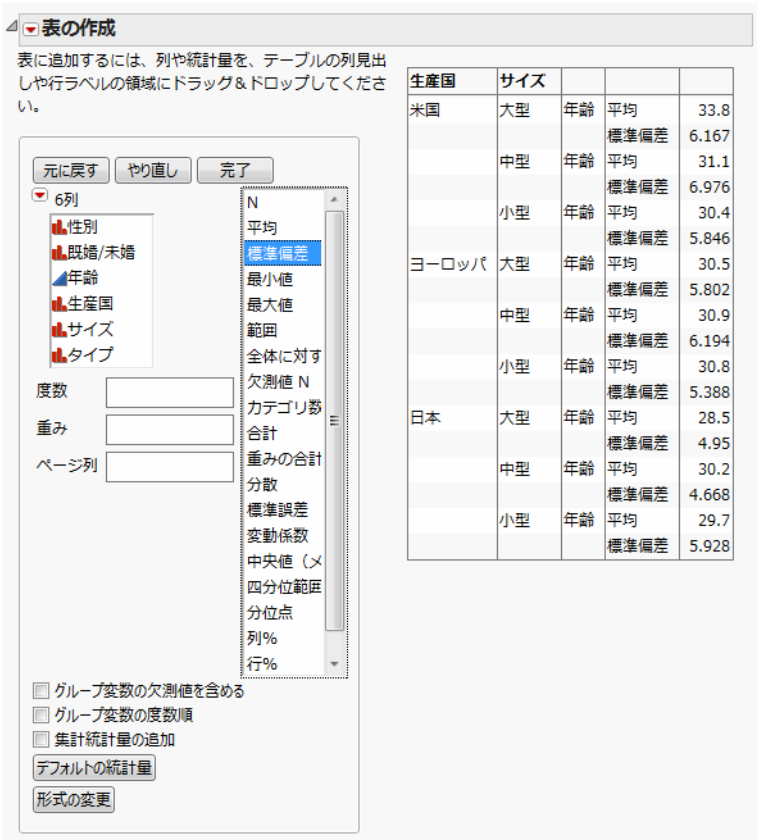
図11.15 年齢と標準偏差を年齢別に示した表

生産国	サイズ			
米国	大型	年齢	平均	33.8
			標準偏差	6.2
	中型	年齢	平均	31.1
			標準偏差	7.0
	小型	年齢	平均	30.4
			標準偏差	5.8
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.8
	中型	年齢	平均	30.9
			標準偏差	6.2
	小型	年齢	平均	30.8
			標準偏差	5.4
日本	大型	年齢	平均	28.5
			標準偏差	4.9
	中型	年齢	平均	30.2
			標準偏差	4.7
	小型	年齢	平均	29.7
			標準偏差	5.9

1. 図 11.14 の状態から開始します。**「年齢」**をクリックしてドラッグし、**「サイズ」**見出しの右側にドロップします。
2. **［平均］**をクリックしてドラッグし、「合計」の上にドロップします。
3. **［標準偏差］**をクリックしてドラッグし、「平均」の下側にドロップします。

   表の「平均」の下に「標準偏差」が表示されます。［標準偏差］を「平均」の上にドロップすると、「平均」の上に「標準偏差」が表示されます。

図11.16「年齢」、［平均］、［標準偏差］を表に追加



生産国	サイズ			
米国	大型	年齢	平均	33.8
			標準偏差	6.167
	中型	年齢	平均	31.1
			標準偏差	6.976
	小型	年齢	平均	30.4
			標準偏差	5.846
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
			標準偏差	6.194
	小型	年齢	平均	30.8
			標準偏差	5.388
日本	大型	年齢	平均	28.5
			標準偏差	4.95
	中型	年齢	平均	30.2
			標準偏差	4.668
	小型	年齢	平均	29.7
			標準偏差	5.928

表の配置の変更

「サイズ」を表の上部に移動し、クロス表形式にします。図11.17のような表を作成します。

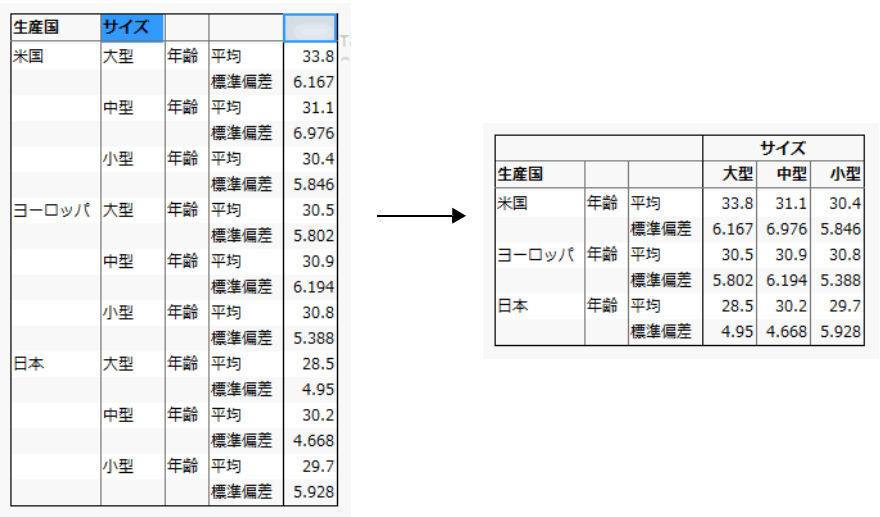
図11.17「サイズ」を上部に移動

	サイズ					
	大型		中型		小型	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	年齢	年齢	年齢	年齢	年齢	年齢
米国	33.8	6.167	31.1	6.976	30.4	5.846
ヨーロッパ	30.5	5.802	30.9	6.194	30.8	5.388
日本	28.5	4.95	30.2	4.668	29.7	5.928

表の配置を変更するには、次の手順を行います。

1. 図 11.16 の状態から開始します。**「サイズ」**見出しをクリックしてドラッグし、表見出しの一番右側にドロップします。詳細については、図11.18を参照してください。

図11.18「サイズ」の移動



生産国	サイズ			
米国	大型	年齢	平均	33.8
			標準偏差	6.167
	中型	年齢	平均	31.1
			標準偏差	6.976
	小型	年齢	平均	30.4
			標準偏差	5.846
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
			標準偏差	6.194
	小型	年齢	平均	30.8
			標準偏差	5.388
日本	大型	年齢	平均	28.5
			標準偏差	4.95
	中型	年齢	平均	30.2
			標準偏差	4.668
	小型	年齢	平均	29.7
			標準偏差	5.928

			サイズ		
生産国			大型	中型	小型
米国	年齢	平均	33.8	31.1	30.4
		標準偏差	6.167	6.976	5.846
ヨーロッパ	年齢	平均	30.5	30.9	30.8
		標準偏差	5.802	6.194	5.388
日本	年齢	平均	28.5	30.2	29.7
		標準偏差	4.95	4.668	5.928

2. 「年齢」をクリックしてドラッグし、**「大型」**、**「中型」**、**「小型」**見出しの下にドロップします。
3. ［平均］と［標準偏差］の両方を選択してドラッグし、**「大型」**見出しの下にドロップします。

これで、見やすい表になりました。車のサイズと生産国別に、所有者の平均年齢と標準偏差がすぐにわかります。

複数の列を持つテーブルの例

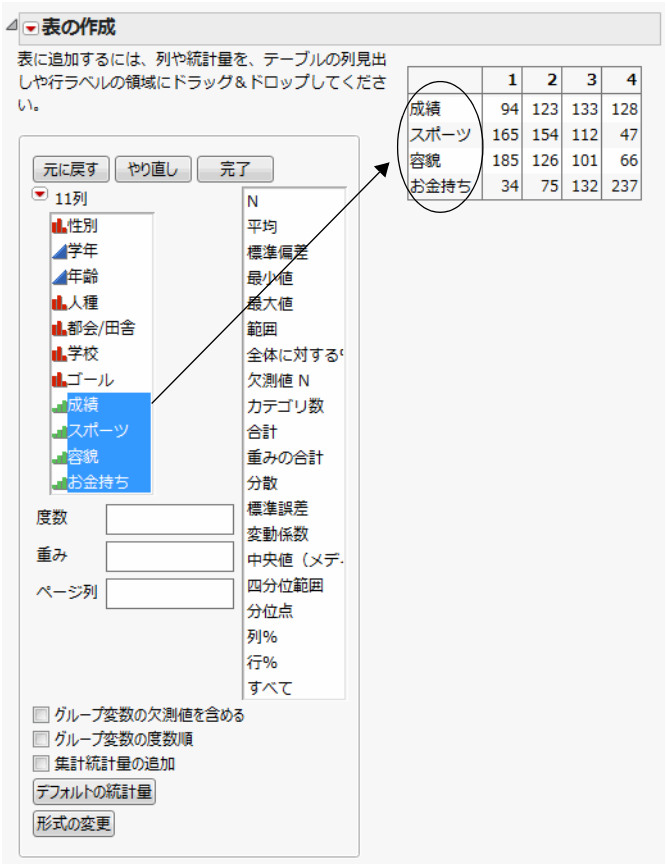
人気者になるのに重要だと思う要素（成績、スポーツ、要望、お金持ち）を、生徒が評価した結果のデータがあります。これらの要素をすべて1つのテーブルにまとめ、統計量や他の因子もあわせて表示してみましょう。図11.19のような表を作成します。

図11.19 人口統計データの追加

		全体に対する%														
		都会/田舎														
		Rural					Suburban					Urban				
性別		1	2	3	4	すべて	1	2	3	4	すべて	1	2	3	4	すべて
boy	成績	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	スポーツ	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	容貌	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
	お金持ち	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
girl	成績	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	スポーツ	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	容貌	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	お金持ち	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

1. ［ヘルプ］>［サンプルデータライブラリ］を選択し、「Children's Popularity.jmp」を開きます。
2. ［分析］>［表の作成］を選択します。
3. 「成績」、「スポーツ」、「容貌」、「お金持ち」を選択してドラッグし、「行のドロップゾーン」にドロップします。

図11.20 カテゴリの列



4つの列は同じ水準値を持つため、これらをまとめたテーブルが表示されます。

各要素に対する4段階評価の割合を表にします。

4. 「性別」をドラッグし、左側の空の見出しにドロップします。
5. ［全体に対する %］をドラッグし、数字の見出しの上にドロップします。
6. ［すべて］をドラッグし、数字の見出し「4」の右側にドロップします。

図11.21「性別」、「全体に対する %」、「すべて」を表に追加

		全体に対する%				
性別		1	2	3	4	すべて
boy	成績	8.16%	12.76%	14.02%	12.55%	47.49%
	スポーツ	26.57%	12.55%	6.49%	1.88%	47.49%
	容貌	9.21%	15.48%	12.34%	10.46%	47.49%
	お金持ち	3.56%	6.69%	14.64%	22.59%	47.49%
girl	成績	11.51%	12.97%	13.81%	14.23%	52.51%
	スポーツ	7.95%	19.67%	16.95%	7.95%	52.51%
	容貌	29.50%	10.88%	8.79%	3.35%	52.51%
	お金持ち	3.56%	9.00%	12.97%	26.99%	52.51%

人の属性データを追加して、さらに細かく集計します。

7. **「都会/田舎」**をドラッグし、**「全体に対する %」**見出しの下にドロップします。

図11.22「都会/田舎」を表に追加

		全体に対する%														
		都会/田舎														
		Rural					Suburban					Urban				
性別		1	2	3	4	すべて	1	2	3	4	すべて	1	2	3	4	すべて
boy	成績	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	スポーツ	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	容貌	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
	お金持ち	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
girl	成績	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	スポーツ	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	容貌	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	お金持ち	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

これで、田舎（「Rural」）、郊外（「Suburban」）、都会（「Urban」）の男子（「boys」）にとって、人気者になるために一番重要な要素は「スポーツ」だということがわかります。田舎、郊外、都会の女子（「girls」）は、「容貌」を一番大事な要素だと考えているようです。

ページ列の使用例

男子生徒と女子生徒の身長を測定したデータがあります。平均身長を生徒の年齢別にまとめた表を作成してみましょう。さらにデータを性別で層別し、性別ごとの表を作成します。このとき、層別の列をページ列として追加します。ページ列を指定すると、グループごとにページが作成されます。図11.23のような表を作成します。

図11.23 生徒の性別ごとの平均身長

	身長(インチ)
年齢	合計
12	293
13	177
14	313
15	126
16	125
17	62

女子

	身長(インチ)
年齢	合計
12	172
13	245
14	457
15	326
16	68
17	138

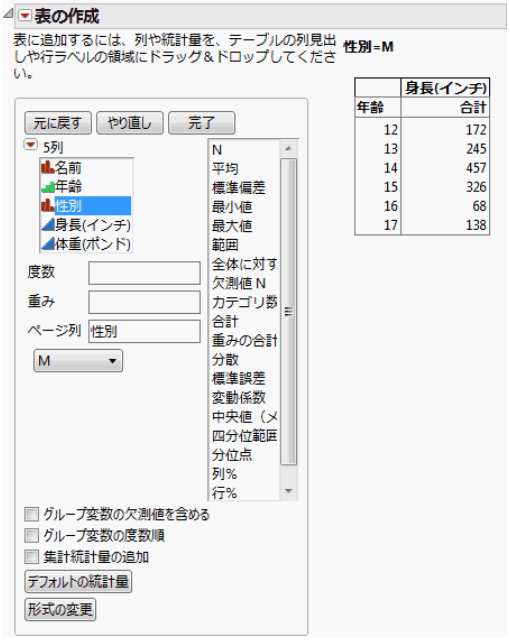
男子

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Big Class.jmp」を開きます。
2. ［**分析**］>［**表の作成**］を選択します。

   身長が分析対象の変数であるため、これを表の一番上に表示します。
3. **「身長 (インチ)」**をクリックし、「列のドロップゾーン」にドラッグします。

   年齢別の統計量が必要なため、年齢を左側に表示します。
4. **「年齢」**をクリックしてドラッグし、数値「2502」の横にある空のセルにドロップします。
5. **「性別」**をクリックしてドラッグし、**「ページ列」**のフィールドにドロップします。
6. 「ページ列」リストから「F」を選択すると、女子の平均身長だけが表示されます。
7. 「ページ列」リストから「M」を選択すると、男子の平均身長だけが表示されます。［**選択行なし**］をクリックして、性別を指定せず、全体の値を表示させることもできます。

図11.24 ページ列の使用例



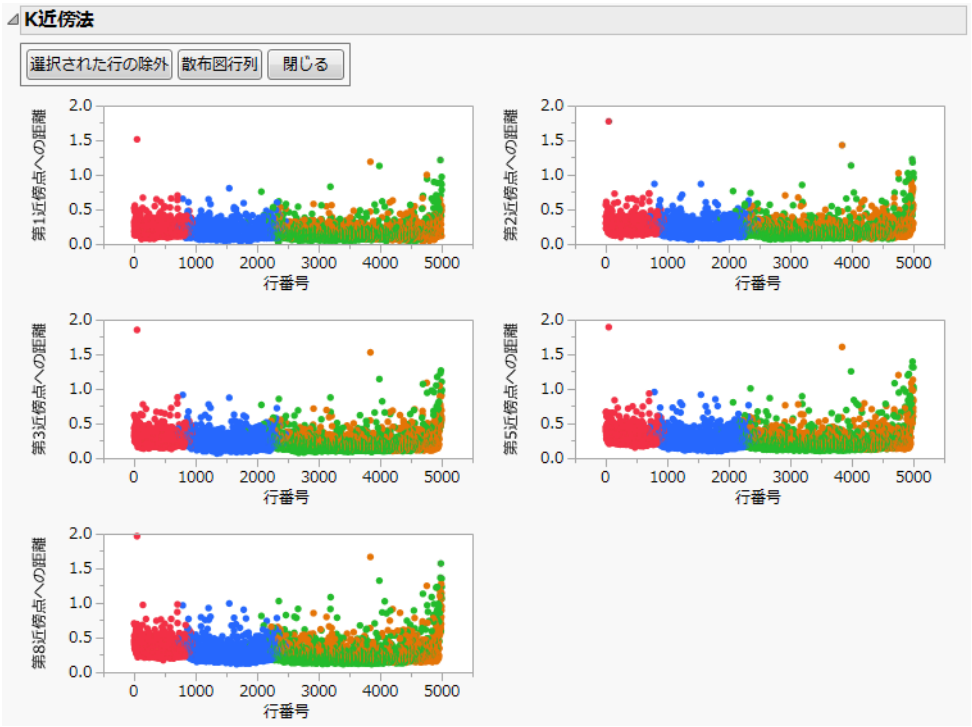
第 12 章

モデル化ユーティリティ
外れ値の除去、欠測値の補間、変数の選択

「モデル化ユーティリティ」には、データのクリーニングや予備処理に関する一連の分析が用意されています。「モデル化ユーティリティ」はデータ探索の強力なツールであり、データをさらに深く理解するのに役立ちます。「モデル化ユーティリティ」では、以下の処理を行えます。

- 一変量や多変量における外れ値を調べる。
- データの欠測値を探索し、補完する。
- JMP PRO 学習セット・検証セット・テストセットにデータ全体を分割する検証列を作成する。
- 説明変数が多いデータにおいて、強い効果をもつ説明変数を選び出す。

図12.1 多変量のk近傍法外れ値の例



目次

「外れ値を調べる」ユーティリティ

データの外れ値を調べることは、重要です。外れ値が生じる理由としては、データの収集や記録における入力ミス、測定システムの不備、あるいは、「999」などの欠測値コードやエラーコードを生のデータ値として扱ってしまっている、などが考えられます。外れ値は推定に影響します。どのような統計分析でも、外れ値の方向にバイアス（偏り）がかかります。たとえば、外れ値があると、標本分散が過大に推定されてしまいます。しかし、場合によっては、外れ値を除去せずにそのまま残しておくべきかもしれません。なぜなら、外れ値を除去することにより、逆に、標本分散が過小に推定され、逆の方向にバイアスがかかるかもしれないからです。

外れ値を削除するにしても保持するにしても、まずそれらを見つけなければなりません。外れ値を視覚的に見つける方法はいくつかあります。たとえば、箱ひげ図、ヒストグラム、および散布図を描くことによって、極端な値を見つけられることはよくあります。詳細については、『はじめてのJMP』の「データの視覚化」の章を参照してください。

「外れ値を調べる」ユーティリティには、一変量や多変量における外れ値を調べるための4つのオプションがあります。

分位点範囲の外れ値 一変量の分位点に基づいて、極端な値としての外れ値を識別します。このツールは、データ中の欠測値コードやエラーコードを見つけるのに便利です。まずこのツールから、外れ値の探索を始めると良いでしょう。

ロバスト推定による外れ値 各列の中心と散らばりをロバストに推定し、遠く離れている値をそれらの値に基づき外れ値として識別します。

多変量ロバスト推定による外れ値 「多変量の相関」プラットフォームで「ロバスト」オプションを指定したときに計算される平均と共分散行列から、Mahalanobisの距離を求め、外れ値を識別します。

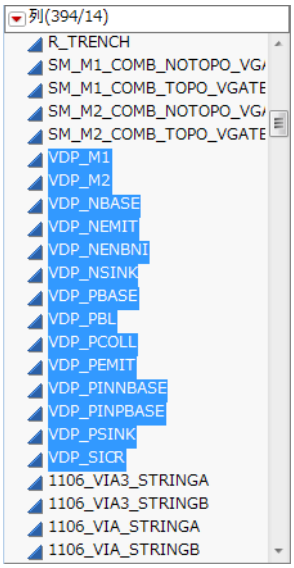
多変量のk近傍法外れ値 k番目の近傍点からの距離が遠いものを、外れ値として識別します。

「外れ値を調べる」ユーティリティの例

「Probe.jmp」サンプルデータテーブルには、5800個の半導体ウエハについて測定された、387個の特徴（「**Responses**」列グループ）が含まれています。「**ロットID**」列と「**ウエハ番号**」列は、一意にウエハを識別します。ここでの目的は、データのいくつかの列を調べて、外れ値を識別することです。「外れ値を調べる」ユーティリティを使って外れ値を識別し、その後「一変量の分布」プラットフォームを使って分析してみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Probe.jmp」を開きます。
2. 「**Responses**」列グループを展開し、「**VDP_M1**」列から「**VDP_SICR**」列までを選択します。これで、14個の列が選択されています（図12.2を参照）。

図12.2 選択された列



3. ［**列**］>［**モデル化ユーティリティ**］>［**外れ値を調べる**］を選択します。
4. ［**分位点範囲の外れ値**］をクリックします。

   外れ値のレポートには、各列と、それぞれの外れ値の個数と値がリストされます。
5. ［**外れ値のある列のみ表示**］を選択すると、外れ値のある列のみが表示されます。

   いくつかの列の外れ値の値が9999であることに注目してください。9は多くの業界で欠測値のコードとして使用されています。
6. 「「9」を含むデータ」レポートで、各列を選択します。
7. ［**欠測値のコードに最大「9」を追加**］をクリックします。

   ［**名前を付けて保存**］コマンドを使って元のデータを保存するように求める警告ダイアログが表示されます。
8. ［**OK**］をクリックします。
9. 「分位点範囲の外れ値」レポートで、［**再スキャン**］をクリックします。
10. ［**検索を整数に限定**］を選択します。

    連続尺度のデータでは、エラーコードや欠測値コードなどを整数の値で表すことがよくあります。今、分析している列には、「9999」以外のエラーコードは含まれていないようです。
11. ［**検索を整数に限定**］の選択を解除します。

データを調べる

1. 「分位点範囲の外れ値」レポートで、残りのすべての列を選択します。
2. ［**行の選択**］をクリックします。

3. ［分析］>［一変量の分布］を選びます。
4. 先ほど選択した列を［Y, 列］の役割に割り当てます。「分位点範囲の外れ値」レポートでこれらの列を選択していたので、「一変量の分布」起動ウィンドウでもこれらの列がすでに選択されています。
5. ［OK］をクリックします。

   図 12.3 に、作成されるレポートの一部を示します。

図 12.3 外れ値が選択された列の分布

「VDP_M1」列と「VDP_PEMIT」列では、選択された外れ値のいくつかは大半のデータに幾分近いものとなっています。その他の列では、選択された外れ値は十分に離れており、分析から除外しても良さそうです。

除外した外れ値の精査

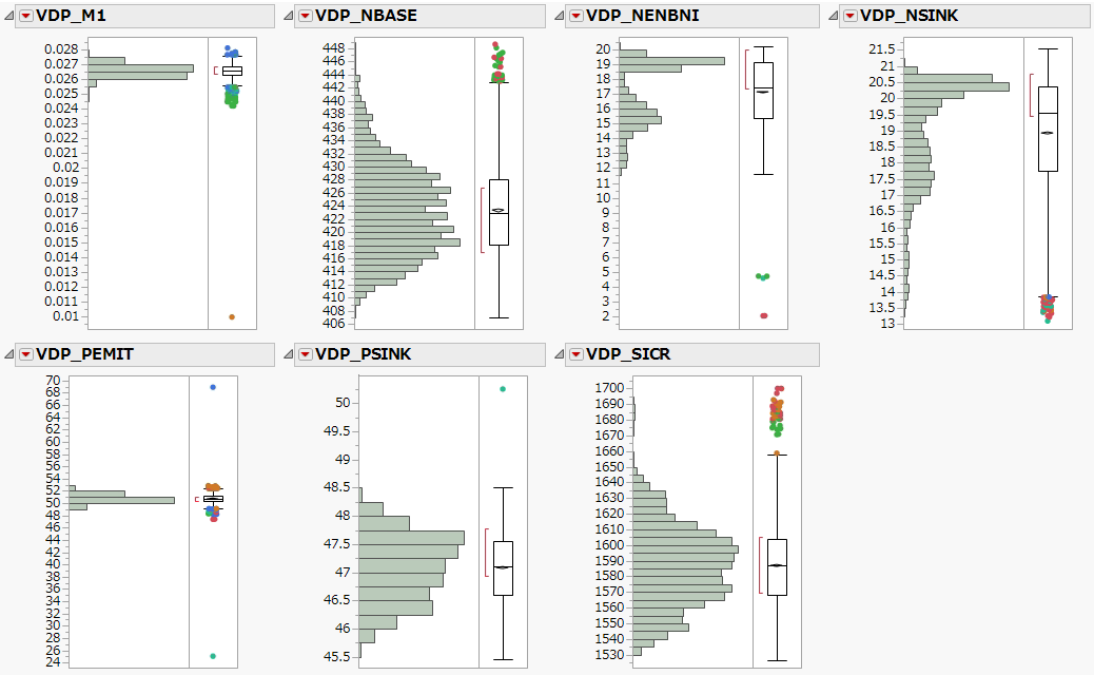
1. 「分位点範囲の外れ値」レポートで、Ctrl キーを押しながら「VDP_M1」と「VDP_PEMIT」の選択を解除します。
2. 残りの列は選択されたままにして、［行の除外］をクリックします。
3. 「Q」を 20 に変更します。
4. ［再スキャン］をクリックします。
5. レポートで「VDP_M1」と「VDP_PEMIT」を選択します。［行の選択］をクリックします。

データを再度調べる

1. 「一変量の分布」レポートを再度調べます。これで、選択された外れ値は、大半のデータからかけ離れていて分析から除外しても良さそうです。
2. 「分位点範囲の外れ値」レポートで、［**行の除外**］をクリックします。
3. 「一変量の分布」レポートの赤い三角ボタンのメニューをクリックします。
4. ［**スクリプト**］>［**分析のやり直し**］を選択します。

   図12.4に、作成されるレポートの一部を示します。

図12.4 外れ値が除外された列の分布



外れ値を除外したデータの分布は、より情報を把握しやすいものとなります。

「外れ値を調べる」ユーティリティの起動

「外れ値を調べる」ユーティリティを起動するには、まず関心のある列を選択して、その後［**列**］>［**モデル化ユーティリティ**］>［**外れ値を調べる**］を選択します。

分位点範囲の外れ値

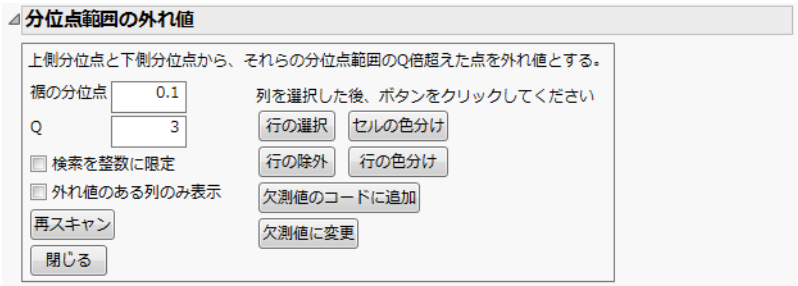
「分位点範囲の外れ値」ユーティリティでは、各列の分位点によって極端な値を識別します。分位点の計算には特定の確率分布を仮定する必要がないので、外れ値を検出するのに分位点を用いるのは便利です。データは単純に小さい値から大きい値へと並べられます。たとえば、20% の分位点は、データの 20% がそれより小さい値です。外れ値は、分位点の範囲（2 つの分位点の差）の定数倍によって定義されます。分位点の計算方法についての詳細は、「一変量の分布」の章の「分位点の統計的詳細」（70 ページ）を参照してください。

「分位点範囲の外れ値」ユーティリティは、データにある欠測値コードを識別するのにも便利です。前述したように、一部の業界では、欠測値を「9」として入力しています（「999」や「9999」など）。このユーティリティでは、上側分位点以上のデータ値のなかで「9」が連続している整数をすべて、欠測値コードの候補として取り上げます。このユーティリティを使って、それらの候補をデータテーブルの「欠測値のコード」列プロパティに追加することもできます。

［分位点範囲の外れ値］のオプション

「分位点範囲の外れ値」パネルでは、外れ値の計算方法と管理方法を指定できます。図 12.5 は、デフォルトの「分位点範囲の外れ値」ウィンドウです。

図 12.5「分位点範囲の外れ値」ウィンドウ



上側分位点もしくは下側分位点から、分位点範囲（上側分位点と下側分位点の差）を Q 倍したもの以上離れている点がすべて外れ値とみなされます。Q の値と何パーセントの分位点にするかは、変更できます。

裾の分位点　分位点範囲を計算するために使われる下側分位点の累積確率。なお、上側分位点の累積確率は、1 からこの値を引いたものとみなされます。たとえば、この「裾の分位点」に 0.1 が指定された場合、分位点範囲は、90% 分位点から 10% 分位点を引いたものです。デフォルトの値は 0.1 です。

Q　外れ値を定義するのに使われる乗数。「裾の分位点」で定義された下側分位点もしくは上側分位点から、分位点範囲を Q 倍したもの以上離れている値が外れ値とみなされます。Q の値を大きくするほど、外れ値の検出がより控えめになります。デフォルトの値は 3 です。

検索を整数に限定　外れ値の候補として取り上げるデータ値を整数だけに限定します。この機能は、欠測値コードやエラーコードを見つけるのに役立ちます。

外れ値のある列のみ表示　レポートに表示する列を、外れ値のある列だけに限定します。

特定の方法によって外れ値を探し出した後、レポートに表示されているこれらの外れ値に対して、さまざまな処理を行えます。ある列における外れ値に対して処理をしたい場合には、まず、「分位点範囲の外れ値」レポートでその列を選択してください。

行の選択 選択した列で外れ値を含む行が、データテーブルで選択されます。

行の除外 外れ値を含む行に対して、「行の除外」属性を有効にします。この処理が終わった後、「分位点範囲の外れ値」レポートを更新するには、[**再スキャン**] をクリックしてください。

セルの色分け データテーブルにおいて、外れ値を含むセルが色分けされます。

行の色分け データテーブルにおいて、外れ値を含む行が色分けされます。

欠測値のコードに追加 選択した列の外れ値が、その列の「欠測値のコード」列プロパティに追加されます。このオプションを使うと、欠測値コードやエラーコードを、欠測値として定義できます。欠測値コードやエラーコードは、整数であることが多く、また、9がいくつか並んだ正または負の整数で表されることが多いです。この処理が終わった後、「分位点範囲の外れ値」レポートを更新するには、[**再スキャン**] をクリックします。

欠測値に変更 データテーブルにおいて、外れ値が欠測値に置換されます。データ値を欠測値に置換するには注意が必要です。データが無効か、正確でないとわかっている場合にのみ、データ値を欠測値に置換してください。この処理が終わった後、「分位点範囲の外れ値」レポートを更新するには、[**再スキャン**] をクリックしてください。

再スキャン 何らかの外れ値の処理をした後に、レポートを更新したい場合には、この [**再スキャン**] を行ってください。

閉じる 「分位点範囲の外れ値」パネルを閉じます。

「分位点範囲の外れ値」レポート

「分位点範囲の外れ値」レポートには、すべての列が、指定のオプションを使って検出された外れ値とともに表示されます。また、上側分位点と下側分位点、そして上側閾値と下側閾値が表示されます。これらの閾値の外側にある値は、外れ値とみなされます。各列の外れ値の個数も表示されます。各外れ値の値は、レポートの最後の列に表示されます。1つの列で複数回現れる外れ値については、その回数が括弧内に示されます。レポートから外れ値のない列を削除するには、[**外れ値のある列のみ表示**] を選択します。

このレポートを読み取る際には、以下の点に注目してください。

- エラーコード。連続尺度のデータの場合、整数の値が大きいほど疑わしく、エラーコードと考えられます。たとえば、上側分位点が0.5程度で、下側分位点が-0.5程度である場合、1049や-777といった極端に大きな整数値はエラーコードである可能性が高いでしょう。
- ゼロ。欠測値を「0」とコーディングしている場合もあるでしょう。たとえば、データの大半が比較的大きな値で、0が外れ値として存在している場合、その0は欠測値である可能性が高いでしょう。

「9」を含むデータ

「分位点範囲の外れ値」ウィンドウのおける「「9」を含むデータ」レポートには、欠測値コードかもしれないデータ値を含む列が表示されます。欠測値コードの候補として取り上げられるデータ値は、上側分位点よりも大きな値で、かつ、9が連続しているもの（たとえば9999）のなかで、最大となっているものです。それらが頻出している場合、それらの外れ値は実際には欠測値コードであると考えられます。それらの度数が少ない場合は、単なる外れ値であるのか、それとも、欠測値コードであるのかを、さらに調べる必要があります。「「9」を含むデータ」レポートには、上側分位点も表示されます。

このレポートは、欠測値コードと類推されるデータ値が存在している場合にのみ表示されます。

欠測値のコードに最大「9」を追加　選択した外れ値が「欠測値のコード」列プロパティに追加されます。この処理の後、「分位点範囲の外れ値」レポートを更新するには、［**再スキャン**］をクリックしてください。

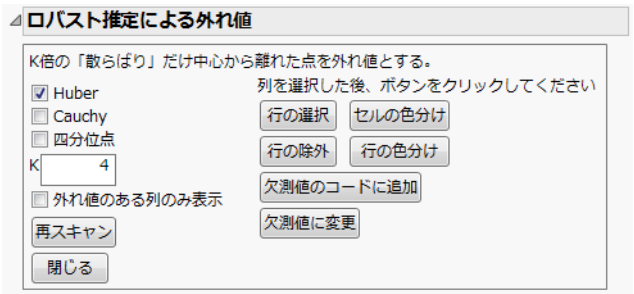
「9」を含む最大値を欠測値に変更　データテーブルから、選択した外れ値を削除します。

注：データを変更する処理（［**欠測値に変更**］や［**行の除外**］など）を最初に選択した際、元のデータを保持するために［**名前を付けて保存**］コマンドを使ってデータテーブルを新しいファイルとして保存するよう求める警告ダイアログが表示されます。このウィンドウが表示されたら、［**OK**］をクリックしてください。また、新しいデータを保存する場合は、［**ファイル**］>［**名前を付けて保存**］を選択し、新しい名前でファイルを保存してください。

ロバスト推定による外れ値

パラメータをロバストに推定する方法は、ロバストでない方法よりも、外れ値の影響を受けにくくなっています。［ロバスト推定による外れ値］には、中心とちらばりを推定するいくつかの推定法があり、中心から大きく離れた値が外れ値とみなされます。図12.6は、デフォルトの「ロバスト推定による外れ値」ウィンドウです。

図12.6「ロバスト推定による外れ値」ウィンドウ



「ロバスト推定による外れ値」オプション

指定されたロバストな方法によって中心とちらばりが推定され、中心から*k*倍のちらばりだけ離れた値が外れ値とみなされます。「ロバスト推定による外れ値」ウィンドウには、ロバストな推定法を選択したり、*k*を指定したりするオプションがあり、また、検出された外れ値を処理するためのツールがあります。

Huber HuberのM推定を使用して、中心とちらばりを求めます。これがデフォルトのオプションです。Huber and Ronchetti（2009）を参照してください。

Cauchy Cauchy分布に従うと仮定して、中心とちらばりを推定します。Cauchy分布を仮定した推定は、破綻点（breakpoints）が高く、通常、Huber推定よりもロバストです。ただし、複数のクラスターにデータが分かれている場合、ある1つのクラスターに近い半分のデータだけしか考慮せず、残りのデータを無視する傾向があります。

四分位点 四分位範囲（IQR; interquartile range）に基づいて、中心とちらばりを推定します。中央値が中心の推定値として使われます。また、IQRを1.34898で割った値がちらばりの推定値として使われます。正規分布においては、IQRを1.34898で割った値は、標準偏差です。

K 中心からちらばりのk倍以上離れているデータ値を外れ値とみなします。Kの値を大きくするほど、外れ値の検出がより控えめになります。デフォルトの値は4です。

外れ値のある列のみ表示 レポートに表示する列を、外れ値のある列だけに限定します。

特定の方法によって外れ値を探し出した後、レポートに表示されているこれらの外れ値に対して、さまざまな処理を行えます。ある列における外れ値に対して処理をしたい場合には、まず、「分位点範囲の外れ値」レポートでその列を選択してください。

行の選択 選択した列で外れ値を含む行が、データテーブルで選択されます。

行の除外 外れ値を含む行に対して、「行の除外」属性を有効にします。この処理が終わった後、「分位点範囲の外れ値」レポートを更新するには、**［再スキャン］**をクリックしてください。

セルの色分け データテーブルにおいて、外れ値を含むセルが色分けされます。

行の色分け データテーブルにおいて、外れ値を含む行が色分けされます。

欠測値のコードに追加 選択した列の外れ値が、その列の「欠測値のコード」列プロパティに追加されます。このオプションを使うと、欠測値コードやエラーコードを、欠測値として定義できます。欠測値コードやエラーコードは、整数であることが多く、また、9がいくつか並んだ正または負の整数で表されることが多いです。この処理が終わった後、「分位点範囲の外れ値」レポートを更新するには、**［再スキャン］**をクリックします。

欠測値に変更 データテーブルにおいて、外れ値が欠測値に置換されます。この処理が終わった後、「分位点範囲の外れ値」レポートを更新するには、**［再スキャン］**をクリックしてください。

再スキャン 何らかの外れ値の処理をした後に、レポートを更新したい場合には、この［再スキャン］を行ってください。

閉じる 「ロバスト推定と外れ値」パネルを閉じます。

多変量ロバスト推定による外れ値

「多変量ロバスト推定による外れ値」ツールは、「多変量の相関」プラットフォームの［ロバスト］オプションを使用して、多変量間の関係を調べます。「多変量の相関」プラットフォームの詳細については、『多変量分析』の「相関と多変量の手法」に関する章を参照してください。

外れ値分析

［外れ値分析］では、各点から多変量正規分布の中心までのMahalanobisの距離を計算します。この距離は、特定の相関行列をもつ多変量正規分布の確率密度を表す等高線に対応しています。中心から離れるほど、外れ値である確率が高くなります。Mahalanobisの距離の詳細については、『多変量分析』の「相関と多変量の手法」に関する章で「「多変量の相関」プラットフォームのオプション」の節を参照してください。

行を除外した後、分析を再実行するか、またはユーティリティを閉じます。分析を再実行すると、除外されていない行だけで多変量分布の中心が再計算されます。なお、除外された行をデータテーブルにて非表示にしない限り、それらの行はグラフに表示されたままです。

「Mahalanobisの距離」の赤い三角ボタンのメニューから、［**Mahalanobisの距離を保存**］を選択すると、Mahalanobisの距離をデータテーブルに保存できます。

図12.7「多変量ロバスト推定による外れ値」の「Mahalanobisの距離」プロット

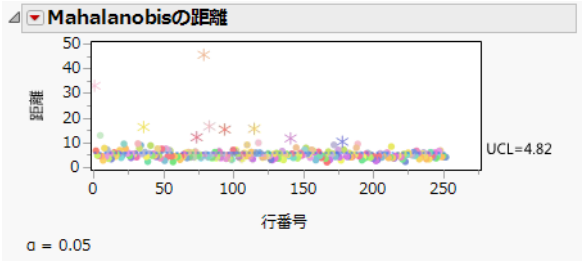


図12.7は、16列から計算されたMahalanobisの距離です。プロットには、4.82の位置に上側管理限界（UCL; upper control limit）も描かれています。このUCLは、どの程度のデータを外れ値として扱うかを考えるのに役立ちます。ただし、どれを外れ値とするかは、各状況に応じて判断してください。UCLの詳細については、Mason and Young（2002）を参照してください。

［多変量のロバスト推定による外れ値分析］のオプション

［多変量のロバスト推定による外れ値分析］の赤い三角ボタンのメニューには、多変量のデータを分析するための数々のオプションがあります。これらのオプションの説明については、『多変量分析』の「相関と多変量の手法」に関する章で「「多変量の相関」プラットフォームのオプション」の節を参照してください。

多変量のk近傍法外れ値

外れ値を検出する方法の1つとして、他のほとんどの点から離れている点を選ぶことが考えれます。ある点の他の点からの距離を特徴づける方法の1つとしては、k番目の近傍点からの距離を用いることが考えられます。「多変量のk近傍法外れ値」ユーティリティは、各点に対して、k番目の近傍点からのユークリット距離を表示します。グラフを少なくするためにフィボナッチ数列によっていくつかの値をスキップしながら、$K = 1, 2, 3, \ldots, k$の近傍点に対するグラフを描きます。

このアプローチは、*k*の値に影響を受けやすいです。*k*が小さすぎると、少数の外れ値どうしが距離を縮めて、外れ値を見落としてしまいがちです。*k*が大きすぎると、外れ値とみなされた点が、*k*個未満のデータ点からなるクラスターに属しているだけかもしれません。つまり、*k*が小さいと、外れ値は検出しにくくなり、*k*が大きいと外れ値は検出されにくくなります。ただし、*k*を大きめにすると、有意でない変数を分析に含めることによるマイナスの影響を減らせます。

このユーティリティを起動するには、「外れ値を調べる」ウィンドウの「コマンド」セクションで［**多変量のk近傍法外れ値**］を選択します。*k*の値を指定し（デフォルトは8）、［**OK**］をクリックします。

k近傍法外れ値で外れ値を識別する

「K近傍法」レポートには、1番目から*k*番目までの近傍点に関して、近傍点への距離が表示されます。どのグラフにおいても、近傍点から離れているデータ点が、外れ値と考えられます。

プロット内でマウスを使って、外れ値の疑いがある点を選択できます。選択したデータ点は、［**散布図行列**］をクリックすることでさらに調べることができます。

選択したデータ点を外れ値と判断した場合は、［**選択された行の除外**］をクリックして、それらのデータ点をその後の分析から除外できます。

行を除外した後、分析を再実行するか、またはユーティリティを閉じます。分析を再実行すると、除外されていない行だけですべての近傍点が再計算されます。なお、除外された行をデータテーブルにて非表示にしない限り、それらの行はグラフに表示されたままです。

「外れ値を調べる」ユーティリティのその他の例

「多変量のk近傍法外れ値」の例

「Water Treatment.jmp」サンプルデータには、都市の排水処理工場内の38個のセンサーで毎日計測された値が含まれています。これらのデータから外れ値の可能性がある値を検出しましょう。外れ値の原因としては、センサーの故障、嵐、その他の状態が考えられます。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Water Treatment.jmp」を開きます。
2. 「**Sensor Measurements**」列グループ内の列を選択します。
3. ［**列**］>［**モデル化ユーティリティ**］>［**外れ値を調べる**］>［**多変量のk近傍法外れ値**］を選択します。
4. k近傍法のkを13とします。［**OK**］をクリックします。

図12.8 多変量のk近傍法外れ値の例での外れ値

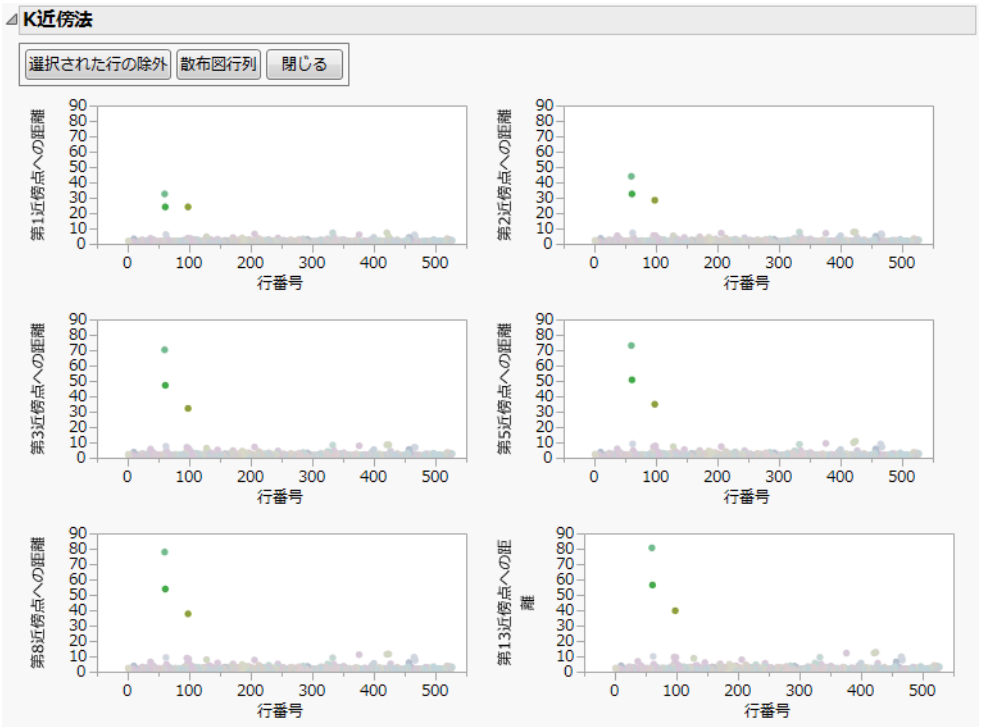


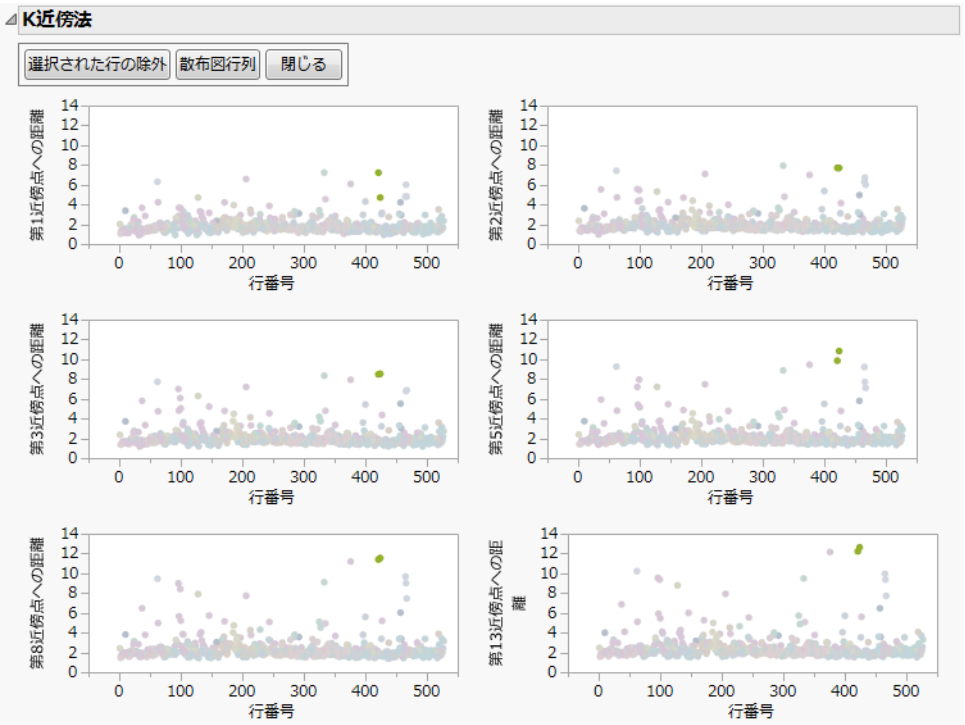
図12.8の「k近傍法」プロット内では、3つの外れ値が選択されています。これら3つのデータ点それぞれが、排水処理工場内の二次沈殿槽が正常に動作していないとされるデータに対応しています。これら3つのデータ点は不良な機器によるものであるため、これらを今後の分析から除外します。

5. 3つの外れ値を選択して、[**選択された行の除外**]をクリックします。

   ユーティリティを再実行するかウィンドウを閉じるかを尋ねるダイアログが表示されます。

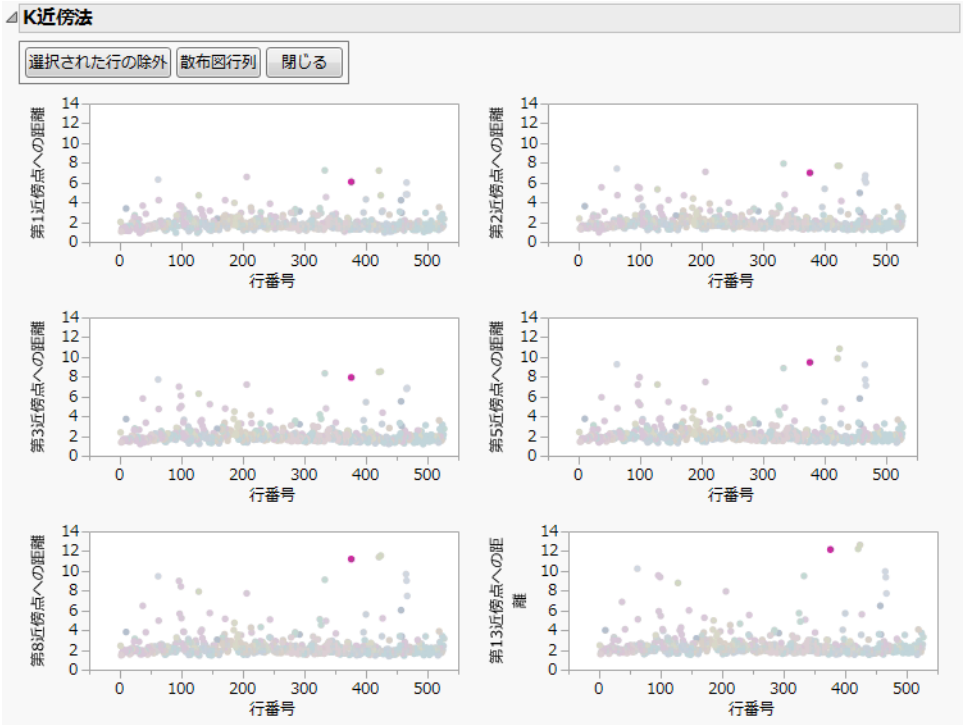
6. [**再実行**]をクリックします。
7. k近傍法のkを13とします。[**OK**]をクリックします。

図12.9 多変量のk近傍法の例での外れ値



次に、行400に近い薄緑色の2つの外れ値に注目してください。この2つの点は、*k*が大きくなるにつれ、距離はどのように変化しているでしょうか。これらのデータ点は13番目の近傍点からは比較的遠い距離にありますが、調べてみると、この時は排水処理での固形物負荷が高かったことが分かりました。これは実際の現場で生じる状況なので、これらの2点は分析から除外しないこととします。ただし、今後の分析ではこの2点に注意する必要があるでしょう。

図12.10 多変量のk近傍法の例での外れ値



では次に、行375あたりにある濃いピンク色の外れ値に注目してください。他の変数との関係において、この点がどのあたりに位置しているのかを調べてみましょう。

8. 行375あたりにある外れ値を選択します。
9. ［**散布図行列**］をクリックします。

   これだけの量の散布図があると、視認で外れ値を確認することが得策とは限りません。しかし、ここでは、問題となっているデータ点が、1変数や2変数で外れているかどうかを見てみましょう。
10. 行列の一番下までスクロールします。この点の「**RD-SED-G**」の値が非常に低いことに注目してください。この点は、「**RD-SED-G**」の値が小さいだけでなく、そのような低い値を持つ数少ないデータ点の1つです。
11. 「**RD-SED-G**」と「**SED-S**」の間の関係を見るため、右にスクロールします。問題のデータ点は、「**RD-SED-G**」が低く、「**SED-S**」が高くなっています。これら2つの変数間には何らかの関係があるようです。「**RD-SED-G**」が低いときに「**SED-S**」が高くなることはあまりありませんが、まったくないとも言えません。この散布図だけからは、この点を分析から必ず除外すべきとは言えません。

「欠測値を調べる」ユーティリティ

データの欠測はさまざまな理由で発生します。たとえば、機器が故障していた場合や、観測値が検出範囲を超えていて記録されなかった場合などです。欠測値は、統計分析の結果に影響する場合があります。たとえば、もし寿命の調査において、多くの健康な人々のデータが欠測していると、それらを考慮せずに分析した結果は、寿命を短く見積もってしまう方向にバイアスがかかります。欠測値は、その存在を識別するだけでなく、統計分析を行う前に理解しておかなければなりません。

いくつかのプラットフォームでは、説明変数に欠測値がある場合でも、予測値を算出できます。詳細については、『基本的な回帰モデル』の「モデルのあてはめ」の章の「欠測値をカテゴリとして扱う」の節を参照してください。ただし、多変量分析の多くでは、データに1つでも欠測値がある行は、分析から除外されます。このツールでは、データにある欠測値を識別し、それらの欠測値を補完できます。

「欠測値を調べる」ユーティリティの例

「Arrhythmia.jmp」サンプルデータには、452人の患者の心電図（ECG）データが含まれています。このデータは、元々、心電図のさまざまなパターンから不整脈を分類するために収集されたものです。ただし、このデータテーブルには欠測値があります。ここでは、主にこれらの欠測値を調べて、必要に応じて値を補完してみましょう。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Arrhythmia.jmp」を開きます。
2. テーブル全体の欠測値を調べるため、データテーブル内で280列すべてを選択します。
3. ［**列**］>［**モデル化ユーティリティ**］>［**欠測値を調べる**］を選択します。
4. ［**欠測値のある列のみ表示**］を選択します。

図12.11 欠測値に関する情報

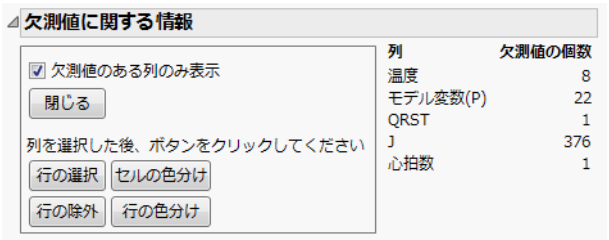


図12.11に示す「欠測値に関する情報」から、欠測値のある列は5つだけであることがわかります。「**J**」列には最も多くの欠測値が含まれており、［モデルのあてはめ］などの［欠測値をカテゴリとして扱う］プラットフォームで使用する場合を除き、分析に有益ではありません。

5. 「欠測値に関する情報」で、「**温度**」、「**モデル変数(P)**」、「**QRST**」、「**心拍数**」の列を選択します。
6. ［**多変量正規分布による補完**］をクリックします。

共分散の推定値として縮小させた推定値を用いるかどうかを尋ねるウィンドウが表示されます。

7. ［**はい**］をクリックします。

   ［**名前を付けて保存**］コマンドを使って元のデータを保存するよう求める警告ダイアログが表示されます。

8. ［**OK**］をクリックします。

図12.12 欠測値補完に関する情報

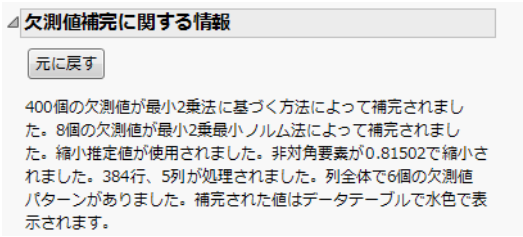


図12.12に示す「欠測値補完に関する情報」には、補完された欠測値の数と具体的な情報が表示されます。［多変量正規分布による補完］が使用された4つの列には、欠測値がなくなりました。

「欠測値を調べる」ユーティリティの起動

「欠測値を調べる」ユーティリティを起動するには、まず関心のある列を選択して、その後［**列**］>［**モデル化ユーティリティ**］>［**欠測値を調べる**］を選択します。

欠測値に関する情報

「欠測値に関する情報」には、各列と欠測値の個数が表示されます。

欠測値のある列のみ表示　欠測値がない列をリストから除外します。

閉じる　欠測値の列セクションを閉じます。

行の選択　選択した列内の欠測値の行を選択します。

行の除外　選択した行の「行の除外」属性を有効にします。

セルの色分け　選択した列内の欠測値を色分けします。

行の色分け　欠測値の行を色分けします。

欠測値のクラスター分析

［欠測値のクラスター分析］は、行内および列内の欠測値に対して階層型クラスター分析を行います。この機能によって、意図せずに欠測値になってしまっている変数の集合を見つけることができます。欠測値は赤で、欠測値以外は青で示されます。マウスを使ってデータを選択できます。たとえば、多くのデータが欠測値ではない列を選択したい場合には、ほとんどが青色となっている領域に対応する列を樹形図で選択します。

欠測値のスナップショット

［欠測値のスナップショット］には、欠測値が視覚的に表示されます。これは、特に、全期間においてはデータが収集されなかった経時測定データを視覚化するのに便利です。

多変量正規分布による補完

統計分析の多くにおいて、欠測値をどのように扱うかは問題となります。ほとんどの統計分析では、欠測値が1つでもある行を分析から除外します。これらの行は不完全なケース（incomplete cases）と呼ばれます。完全なケース（complete cases）だけを使って統計分析を行うのは処理は簡単なのですが、そうなると不完全なケースに含まれる情報を失ってしまいます。また、1つでも欠測値を含む行を除外することは、完全なケースと不完全なケースとの間にある系統的な違いを無視してしまうことにもなり、その結果、特に完全なケースが少ない場合には、統計的推測が母集団を表すものではなくなります。

［多変量正規分布による補完］ユーティリティは、連続尺度のデータに対してだけ使え、多変量正規分布に基づいて欠測値を補完します。

［多変量正規分布による補完］では、共分散の推定値として縮小させた推定値（shrinkage estimation）を使用できます。縮小させた推定値を使用すると、共分散行列の推定値を改善できます。縮小させた推定値の詳細については、Schafer and Strimmer（2005）を参照してください。

欠測値補完に関する情報

「欠測値補完に関する情報」には、欠測値補完に関する次のような情報が表示されます。

- 補完の手法（最小2乗法または最小2乗最小ノルム法）
- 補完された個数
- 縮小推定値の使用の有無
- 非対角要素を収縮した係数
- 処理された行数および列数
- 欠測値パターンの数

補完が完了すると、「欠測値に関する情報」にはどの列にも欠測値が存在しないことが自動的に示されます。補完された値はデータテーブル内で明るい青色で表示されます。

補完を取り消すには［**元に戻す**］をクリックします。すると、補完された値が元の欠測値に戻ります。

多変量の特異値分解補完

「多変量の特異値分解補完」ユーティリティは、特異値分解（SVD; Singular Value Decomposition）を使って欠測値を補完します。このユーティリティは、変数の数が数百を超えるようなデータに便利です。変数の数が多いデータテーブルでペアごとの共分散行列を計算するのは実践的ではありません。そのため、そのような場合には特異値分解の手法をお勧めします。

データが大きい場合は、進捗バーが表示されて、特異値分解が完了した次元が示されます。いつでも補完を停止し、その次元の数を使用できます。

欠測値補完に関する情報

「欠測値補完に関する情報」には、欠測値補完に関する次のような情報が表示されます。

- 補完の手法
- 補完された個数
- 処理された行数および列数

補完が完了すると、「欠測値に関する情報」にはどの列にも欠測値が存在しないことが自動的に示されます。補完された値はデータテーブル内で明るい青色で表示されます。

補完を取り消すには［**元に戻す**］をクリックします。すると、補完された値が元の欠測値に戻ります。

「検証列の作成」ユーティリティ

JMP PRO

検証（validation）とは、データの一部をモデルパラメータの推定に使用し、残りのデータでモデルの予測能力を評価することを指します。データが複雑な場合は、これにより、モデルがオーバーフィットするリスクを軽減できます。

検証列は、データを2つまたは3つの部分に分けます。

- モデルパラメータの推定に使うデータを、学習セットといいます。
- 予測能力が高いモデルを選ぶのに用いるデータを、検証セットといいます。
- モデルが選択された後、モデルの予測能力をチェックするデータを、テストセットといいます。

検証列は、［モデルのあてはめ］プラットフォームで検証の手法の1つとして使用できます。

「検証列の作成」ユーティリティの例

「Lipid Data.jmp」データテーブルには、カリフォルニア州のある病院で収集された、95人の患者の血液測定値、身体測定値、およびアンケート調査のデータが含まれています。ここで、今後の分析における検証方法の1つとして、検証列を使用したいとします。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Lipid Data.jmp」を開きます。
2. ［**分析**］>［**一変量の分布**］を選びます。
3. 「**性別**」に［Y, 列］を割り当てます。［**OK**］をクリックします。

図12.13「Lipid Data.jmp」での「**性別**」の分布

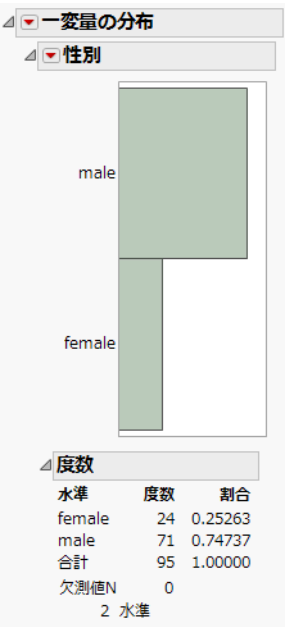


図12.13は、データ内の「**性別**」の分布を示しています。被験者の男性と女性の割合は同じではないことに注目してください。女性のデータのほうが少ないので、検証セットと学習セット全体では性別のバランスをとる必要があります。

4. ［**列**］>［**モデル化ユーティリティ**］>［**検証列の作成**］を選択します。
5. ［**層化無作為抽出**］をクリックします。
6. 検証データを層化抽出する際に層とする列として、「**性別**」を選択します。
7. ［**OK**］をクリックします。

   データテーブルに「検証」列が追加されます。モザイク図を作成すると、検証セットと学習セットの分布を確認できます。

8. ［**分析**］>［**二変量の関係**］を選びます。
9. 「**検証**」を［Y, 目的変数］に、「**性別**」を［X, 説明変数］に割り当てます。
10. ［**OK**］をクリックします。

図12.14 検証セットと学習セットにおける性別の分布

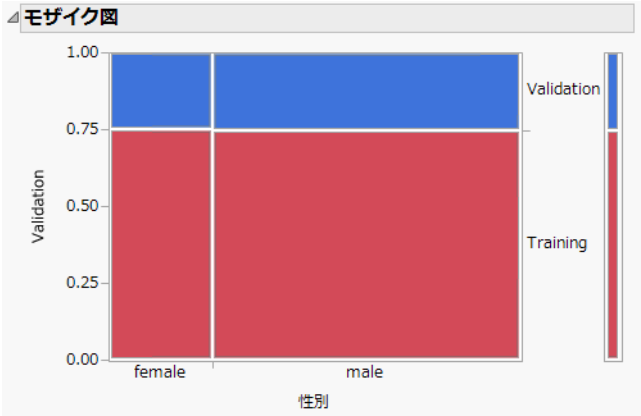


図12.14は、検証セットと学習セットにおける**「性別」**の分布を示しています。男性と女性それぞれの約75%が学習セットに、男性と女性それぞれの約25%が検証セットに含まれていることがわかります。

「検証列の作成」ユーティリティの起動

「検証列の作成」ユーティリティを起動するには、2通りの方法があります。

- ［**列**］>［**モデル化ユーティリティ**］>［**検証列の作成**］を選択します。
- プラットフォームの起動ウィンドウで［**検証**］をクリックします。次の点を念頭に置いてください。
  - プラットフォームが「検証」列に対応している必要があります。
  - 「列の選択」リストで、列が選択されていてはいけません。
  - この操作により、データテーブルに「検証」列が追加されます。また、起動ウィンドウでは、追加された新しい列が「検証」の役割に割り当てられます。

注:「検証列の作成」ユーティリティをプラットフォームの起動ウィンドウから起動した場合、［列］メニューから起動した場合とは異なり、［単純無作為抽出］と［層化無作為抽出］の手法は使用できません。デフォルトの設定では、データの70%が学習データとして、30%が検証セットとして使用されます。

「検証列の作成」レポート

学習、検証、テストの各セットに含めたい行の数または割合を入力します。デフォルトでは、データの75%が学習データとして、25%が検証セットとして使用されます。

［検証列の作成］では、新しい検証列の名前を指定することもできます。デフォルトの名前は**「検証」**です。

異なるセットを作成するために、2つの異なる手法が用意されています。

単純無作為抽出 入力された配分に基づいてデータをセットに分けます。たとえば、デフォルトの値が入力されている場合、各行は学習セットに含まれる確率が0.75、検証セットに含まれる確率が0.25となります。このオプションは、特定の集団に偏った分布をしておらず、また、データ量が多い場合に便利です。

層化無作為抽出 指定した列のグループごとに均等な配分でデータをセットに分けます。［層化無作為抽出］オプションは、データテーブル内の新しい**「検証」**列に列「ノート」プロパティを追加します。「ノート」には、データがどのように層化無作為抽出され、**「検証」**列が作成されたかが示されます。学習、検証、およびテストの各セットにおいて、指定した列のグループ構成と同じ割合でデータを抽出したい場合に、このオプションを使用します。

「説明変数のスクリーニング」ユーティリティ

［説明変数のスクリーニング］は、有意な説明変数をデータから選び出します。数百個の説明変数があり、そこから有意なものだけを選択するとします。それには、ステップワイズ回帰も使えますが、常に役立つとは限らず、機能しない場合もあります。［説明変数のスクリーニング］では、100個のディシジョンツリーから構成されるランダム森を作成し、各ツリーにおける寄与を平均化し、各説明変数の最終的な寄与を算出します。ディシジョンツリーの詳細については、『発展的なモデル』の「パーティションモデル」の章を参照してください。

「説明変数のスクリーニング」ユーティリティの例

「Bands Data.jmp」データテーブルには、印刷会社の輪転グラビア印刷機から得られた測定データが含まれています。データセットには539個のレコードと38個の変数が含まれています。応答Yは**「印刷縞の有無」**列で、その値は「BAND」（縞が生じてしまった）と「NOBAND」（縞なし）です。ここでの関心は、どの印刷設定が印刷結果に最も寄与しているかを理解することです。

1. ［**ヘルプ**］>［**サンプルデータライブラリ**］を選択し、「Bands Data.jmp」を開きます。
2. ［**列**］>［**モデル化ユーティリティ**］>［**説明変数のスクリーニング**］を選択します。
3. **「粒子のスクリーニング」**列から**「クロームの有無」**列までを説明変数Xの列として選択します。
4. ［**OK**］をクリックします。
5. **「印刷縞の有無」**を 予測したい応答Yの列として選択します。
6. ［**OK**］をクリックします。

図12.15 説明変数のスクリーニングでの列の順位付け

説明変数のスクリーニング

印刷編の有無

予測子	寄与	割合	順位
溶剤(%)	13.0981	0.2311	1
インク(%)	9.7249	0.1716	2
プレス	6.6378	0.1171	3
プレス速度	5.3692	0.0947	4
粒子のスクリーニング	4.7262	0.0834	5
ローラー硬度	4.4068	0.0778	6
ワニス(%)	2.4510	0.0432	7
インクの種類	2.2002	0.0388	8
プレスの種類	2.0053	0.0354	9
製紙工場の所在地	1.4952	0.0264	10
ブレード圧	0.8668	0.0153	11
用紙の種類	0.7375	0.0130	12
ユニット番号	0.6275	0.0111	13
シリンダのサイズ	0.4568	0.0081	14
シリンダ上でのタイプ	0.4034	0.0071	15
校正カット	0.3017	0.0053	16
湿度	0.2290	0.0040	17
ワックス	0.2248	0.0040	18
硬化剤	0.2170	0.0038	19
粗度	0.1418	0.0025	20
メッキ槽	0.1138	0.0020	21
粘度	0.1136	0.0020	22
電極間距離	0.0804	0.0014	23
キャリパー	0.0427	0.0008	24
インク温度	0.0044	0.0001	25
光沢インクでの校正	0.0000	0.0000	26
ブレードの製造元	0.0000	0.0000	27
直接蒸気	0.0000	0.0000	28
溶剤の種類	0.0000	0.0000	29
ESA電圧	0.0000	0.0000	30
ESA アンペア	0.0000	0.0000	31
電流密度	0.0000	0.0000	32
クロームの含有	0.0000	0.0000	33

列は、応答変数に対する寄与度の順に並び替えられて順位付けされます。図12.15の順位21～33の列は、わずかしか応答に寄与していません。これらの列は応答に何ら影響していないと考えられるので、今後の分析から除外できます。最も大切な列は、「**溶剤(%)**」、「**インク(%)**」、「**プレス**」と判断できます。

「説明変数のスクリーニング」ユーティリティの起動

［**列**］>［**モデル化ユーティリティ**］>［**説明変数のスクリーニング**］を選択して、「説明変数のスクリーニング」モデル化ユーティリティを起動します。

図12.16 説明変数Xを選択するための「説明変数のスクリーニング」起動ウィンドウ

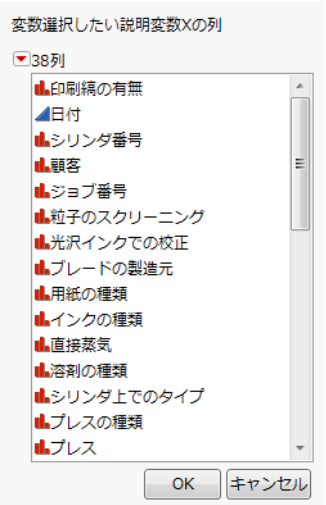


図12.16は、説明変数Xを選択するための起動ウィンドウです。ここでは、スクリーニングしたい列を選択します。

図12.17 応答変数Yを選択するための「説明変数のスクリーニング」起動ウィンドウ

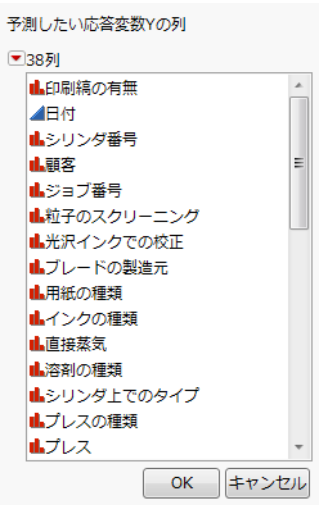


図12.17は、応答変数Yを選択するための起動ウィンドウです。

「説明変数のスクリーニング」レポート

このレポートには、説明変数と、その寄与度および順位が表示されます。寄与度が高い説明変数は、Yを予測するのに重要であると考えられます。

「寄与」列には、モデルに対する各説明変数の寄与が表示されます。説明変数の寄与が0.0000の場合、その説明変数は重要ではないと考えられます。レポートの「割合」列には、各変数の寄与率が表示されます。

「説明変数のスクリーニング」レポートで重要な説明変数を選択した後、［**分析**］**>**［**モデルのあてはめ**］を選択できます。そして、重要な説明変数だけを使用し、それ以外の変数は除外して、モデルをあてはめることができます。この「説明変数のスクリーニング」ユーティリティでは、モデルのあてはめのいくつかのステップを保存できます。

付録 A

参考文献

Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, New York: John Wiley and Sons, Inc.

Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley and Sons, Inc.

Agresti, A., and Coull, B. (1998), "Approximate is Better Than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119–126

Akaike, H. (1974), "Factor Analysis and AIC," *Pschychometrika*, 52, 317–332.

Akaike, H. (1987), "A new Look at the Statistical Identification Model," *IEEE Transactions on Automatic Control*, 19, 716–723.

American Society for Quality Statistics Division (2004), *Glossary and Tables for Statistical Quality Control*, Fourth Edition, Milwaukee: Quality Press.sis of Variances–Heterogeneity and the Logarithmic Transformation," *JRSS* Suppl 8, 128–138.

Bartlett, M.S. (1966), *An Introduction to Stochastic Processes*, Second Edition, Cambridge: Cambridge University Press.

Bissell, A. F. (1990), "How Reliable is Your Capability Index?", *Applied Statistics*, 30, 331-340.

Brown, M.B. and Benedetti, J.K. (1977), "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables," *Journal of the American Statistical Association* 72, 305-315.

Brown, M.B. and Forsythe, A.B. (1974a), "The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means," *Technometrics* 16:1, 129–132.

Brown, M.B. and Forsythe, A.B. (1974), "Robust tests for the equality of variances," *Journal of the American Statistical Association*, 69, 364–367.

Chen, Song Xi and Hall, Peter. (1993), "Empirical Likelihood Confidence Intervals for Quantiles," *The Annals of Statistics* 21(3): 1166-1181.

Chou, Y.M., Owen, D.B., and Borrego, S.A. (1990), "Lower confidence limits on process capability indices," *Journal of Quality Technology*, 22(3): 223-229.

Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368): 829–836.

Cleveland, W.S. (1994). *Visualizing Data*, Summit, N.J: Hobart Press.

Cohen, J. (1960), "A coefficient of agreement for nominal scales," *Education Psychological Measurement*, 20: 37–46.

Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*, Second Edition, New York: John Wiley and Sons.

Conover, W. J. (1972). "A Kolmogorov Goodness-of-fit Test for Discontinuous Distributions". *Journal of the American Statistical Association* 67: 591–596.

Conover, W.J. (1980), *Practical Nonparametric Statistics*, New York: John Wiley and Sons, Inc.

DeLong, E., Delong, D, and Clarke-Pearson, D.L. (1988), "Comparing the Areas Under Two or more Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics* 44, 837–845.

Devore, J. L. (1995), *Probability and Statistics for Engineering and the Sciences*, Duxbury Press, CA.

Dunn, O.J. (1964), "Multiple Comparisons Using Rank Sums," *Technometrics* 6, 241-252.

Dunnett, C.W. (1955), "A multiple comparison procedure for comparing several treatments with a control" *Journal of the American Statistical Association*, 50, 1096–1121.

Dwass, M. (1955), "A Note on Simultaneous Confidence Intervals," *Annals of Mathematical Statistics* 26: 146–147.

Eubank, R.L. (1999), *Nonparametric Regression and Spline Smoothing*, Second Edition, Boca Raton, Florida: CRC.

Fisher, L. and Van Ness, J.W. (1971), "Admissible Clustering Procedures," *Biometrika*, 58, 91–104.

Fleiss, J.L., Cohen J., and Everitt, B.S. (1969), "Large-Sample Standard Errors of Kappa and Weighted Kappa," *Psychological Bulletin*, 72: 323–327.

Friendly, M. (1991), "Mosaic Displays for Multiway Contingency Tables," *New York University Department of Psychology Reports*: 195.

Goodman, L.A. and Kruskal, W.H. (1979), *Measures of Association for Cross Classification*, New York: Springer-Verlag (reprint of JASA articles).

Gupta, S.S. (1965), On Some Multiple Decision (selection and ranking), Rules., *Technometrics* 7, 225–245.

Hahn, G. J. and Meeker, W. Q. (1991) *Statistical Intervals: A Guide for Practitioners*. New York: Wiley.

Hajek, J. (1969), *A Course in Nonparametric Statistics*, San Francisco: Holden–Day.

Hartigan, J.A. and Kleiner, B. (1981), "Mosaics for Contingency Tables," *Proceedings of the 13th Symposium on the Interface between Computer Science and Statistics*, Ed. Eddy, W. F., New York: Springer–Verlag, 268–273.

Hayter, A.J. (1984), "A proof of the conjecture that the Tukey–Kramer multiple comparisons procedure is conservative," *Annals of Mathematical Statistics*, 12 61–75.

Hosmer, D.W. and Lemeshow, S. (1989), *Applied Logistic Regression*, New York: John Wiley and Sons.

"Hot Dogs," (1986), *Consumer Reports* (June), 364–367.

Hsu, J. (1981), "Simultaneous confidence intervals for all distances from the 'best'," *Annals of Statistics*, 9, 1026–1034.

Hsu, J. (1989), *Tutorial Notes on Multiple Comparisons*, American Statistical Association, Washington, DC.

Hsu, J.C. (1996), *Multiple Comparisons: Theory and Methods*, Chapman and Hall.

Huber, Peter J. (1973), "Robust Regression: Asymptotics, Conjecture, and Monte Carlo," *Annals of Statistics*, Volume 1, Number 5, 799-821.

Huber, P.J. and Ronchetti, E.M. (2009), *Robust Statistics*, Second Edition, Wiley.

Iman, R.L. (1974), "Use of a t-statistic as an Approximation to the Exact Distribution of Wilcoxon Signed Ranks Test Statistic," *Communications in Statistics—Simulation and Computation*, 795–806.

Kendall, M. and Stuart, A. (1979), *The Advanced Theory of Statistics, Volume 2*, New York: Macmillan Publishing Company, Inc.

Kramer, C.Y. (1956), "Extension of multiple range tests to group means with unequal numbers of replications," *Biometrics*, 12, 309–310.

Lehmann, E.L. (2006), Nonparametrics: Statistical Methods Based on Ranks, 2nd edition, Springer.

Levene, H. (1960), "Robust tests for the equality of variances" In I. Olkin (ed), *Contributions to probability and statistics*, Stanford Univ. Press.

Mason, R.L. and Young, J.C. (2002), *Multivariate Statistical Process Control with Industrial Applications*, Philadelphia: ASA-SIAM.

McCullagh, P. and Nelder, J.A. (1983), *Generalized Linear Models*, London: Chapman and Hall Ltd.

Meeker, W.Q. and Escobar, L.A. (1998), *Statistical Methods for Reliability Data*, pp. 60–62, New York: John Wiley and Sons.

Nelson, P.R., Wludyka, P.S., and Copeland, K.A.F. (2005), *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*, Philadelphia: Society for Industrial and Applied Mathematics.

Neter, J., Wasserman, W. and Kutner, M.H. (1990), *Applied Linear Statistical Models*, Third Edition, Boston: Irwin, Inc.

O'Brien, R.G. (1979), "A general ANOVA method for robust tests of additive models for variances," *Journal of the American Statistical Association*, 74, 877–880.

O'Brien, R., and Lohr, V. (1984), "Power Analysis For Linear Models: The Time Has Come," *Proceedings of the Ninth Annual SAS User's Group International Conference*, 840–846.

Odeh, R. E. and Owen, D. B. (1980) *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*. New York: Marcel Dekker, Inc.

Olejnik, S.F. and Algina, J. (1987), "Type I Error Rates and Power Estimates of Selected Parametric and Nonparametric Tests of Scale," *Journal of Educational Statistics* 12, 45–61.

Pratt, J.W. (1959), "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures", Journal of the American Statistical Association, 54:287, 655-667.

Reinsch, C.H. (1967), *Smoothing by Spline Functions*, Numerische Mathematik, 10, 177–183.

Rousseuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: John Wiley and Sons.

Schafer, J. and Strimmer, K. (2005), "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics," *Statistical Applications in Genetics and Molecular Biology*, 4:1, 1-30.

Slifker, J. F. and Shapiro, S. S. (1980). *Technometrics*, 22, 239-246.

Snedecor, G.W. and Cochran, W.G. (1980), *Statistical Methods*, 7th edition, Ames, Iowa: Iowa State University Press.

Somers, R.H. (1962), "A New Asymmetric Measure of Association for Ordinal Variables," *American Sociological Review*, 27, 799-811.

Tan, Charles Y., and Iglewicz, Boris (1999), "Measurement-methods Comparisons and Linear Statistical Relationship," *Technometrics*, 41:3, 192–201.

Tamhane, A. C. and Dunlop, D. D. (2000) *Statistics and Data Analysis*, Prentice Hall.

Tukey, J. (1953), "A problem of multiple comparisons," Dittoed manuscript of 396 pages, Princeton University.

Welch, B.L. (1951), "On the comparison of several mean values: an alternative approach," *Biometrika* 38, 330–336.

Wheeler, D.J. (2003), *Range Based Analysis of Means*, SPC Press, Knoxville, TN.

Wludyka, P.S. and Nelson, P.R. (1997), "An Analysis-of-Means-Type Test for Variances From Normal Populations", *Technometrics*, 39:3, 274-285.

Wludyka, P.S. and Nelson, P.R. (1999), "Two Non-Parametric, Analysis-of-Means-Type Tests for Homogeneity of Variances," *Journal of Applied Statistics*, 26:2, 243-256.

索引
基本的な統計分析

T

V

W-Z

ア

イ

ウ

オ

カ

キ

ク

ケ

コ

サ

シ

ス

セ

ソ

タ

チ

ツ

テ

ト

ニ

ノ

ハ

ヒ

フ

ヘ

ホ

マ

ミ