バージョン 11

消費者調査

「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

11.2

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. 2014.
『JMP® 11 消費者調査』Cary, NC: SAS Institute Inc.

JMP® 11 消費者調査

Copyright © 2014, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-61290-766-6

All rights reserved. Produced in the United States of America.

印刷物の場合 : この出版物のいかなる部分も、出版元である SAS Institute Inc. の書面による許可なく、電子的、機械的、複写など、形式や方法を問わず、複製すること、検索システムへ格納すること、および転送することを禁止します。

Web からのダウンロードや電子本の場合 : この出版物の使用については、入手した時点で、ベンダーが規定した条件が適用されます。

この出版物を、インターネットまたはその他のいかなる方法でも、出版元の許可なくスキャン、アップロード、および配布することは違法であり、法律によって罰せられます。正規の電子版のみを入手し、著作権を侵害する不正コピーに関与または加担しないでください。著作権の保護に関するご理解をお願いいたします。

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

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

2014 年 7 月、第 1 版

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at support.sas.com/publishing or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

技術ライセンスに関する通知

- Scintilla - Copyright © 1998-2012 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.

目次
消費者調査

1 JMP の概要

マニュアルとその他のリソース 11
表記規則 13
JMP のマニュアル 13
JMP ドキュメンテーションライブラリ 14
JMP ヘルプ 18
JMP を習得するためのその他のリソース 18
チュートリアル 18
サンプルデータテーブル 19
統計用語と JSL 用語の習得 19
JMP を使用するためのヒント 19
ツールヒント 19
JMP User Community 20
JMPer Cable 20
JMP 関連書籍 20
「JMP スターター」ウィンドウ 20

2 消費者調査について

消費者調査データに対する分析の概要 21

3 カテゴリカルな応答の分析

アンケート調査データなどの度数表 23
「カテゴリカル」プラットフォームの概要 25
「カテゴリカル」プラットフォームの使用例 25
「カテゴリカル」プラットフォームの起動 27
応答の役割 27
選択した列に役割を割り当てる 31
起動ウィンドウのその他のオプション 32
「カテゴリカル」レポート 33

シェアチャート 35
度数チャート 35
「カテゴリカル」プラットフォームのオプション 36
レポートオプション 36
統計量のオプション 37
自由回答のレポートオプション 50
表の構成レポートのオプション 52
「カテゴリカル」プラットフォームのその他の例 53
多重応答 54
応答の度数 55
指示変数 56
多重応答 区切り文字 57
多重応答 ID別 58
スコアの平均の例 59

4 因子分析

潜在的な因子によるモデル化 61
「因子分析」プラットフォームの概要 63
「因子分析」プラットフォームの例 63
「因子分析」プラットフォームの起動 66
「因子分析」レポート 66
モデルの設定 67
回転方法 68
「因子分析」プラットフォームのオプション 70
因子分析の「モデルの設定」のオプション 71

5 選択モデル

選択モデルのあてはめ 77
「選択モデル」プラットフォームの概要 79
「選択モデル」プラットフォームの例 80
「選択モデル」プラットフォームの起動 81
選択モデルの出力 85
被験者効果 86
グリッドごとの効用最適化 88
「選択モデル」プラットフォームのオプション 90

例：トレードオフの評価 92
データテーブルが1つしかないデータ形式 97
例：ピザのデータを1つのテーブルにまとめた場合 98
セグメント化 100
データに関する特殊規則 105
デフォルトの選択肢集合 105
応答データを含んだ被験者データ 105
ロジスティック回帰 105
データの変換 109
データを2つの分析テーブルに変換する 109
データを1つの分析テーブルに変換する 113
対応のあるデータに対する条件付きロジスティック回帰 115
統計的詳細 116

6 アップリフトモデル

アクションが消費者行動に及ぼす効果をモデル化 119
「アップリフト」プラットフォームの概要 121
「アップリフト」プラットフォームの使用例 122
「アップリフト」プラットフォームの起動 123
「アップリフトモデル」レポート 124
アップリフトモデルのグラフ 124
アップリフトレポートのオプション 127

7 項目分析

テスト結果に対する項目反応分析 129
「項目分析」プラットフォームの概要 131
「項目分析」プラットフォームの起動 134
「項目分析」レポート 136
特性曲線 136
情報曲線 137
デュアルプロット 137
「項目分析」プラットフォームのオプション 139
技術的詳細 139

A 参考文献

索引

消費者調査 143

第 1 章

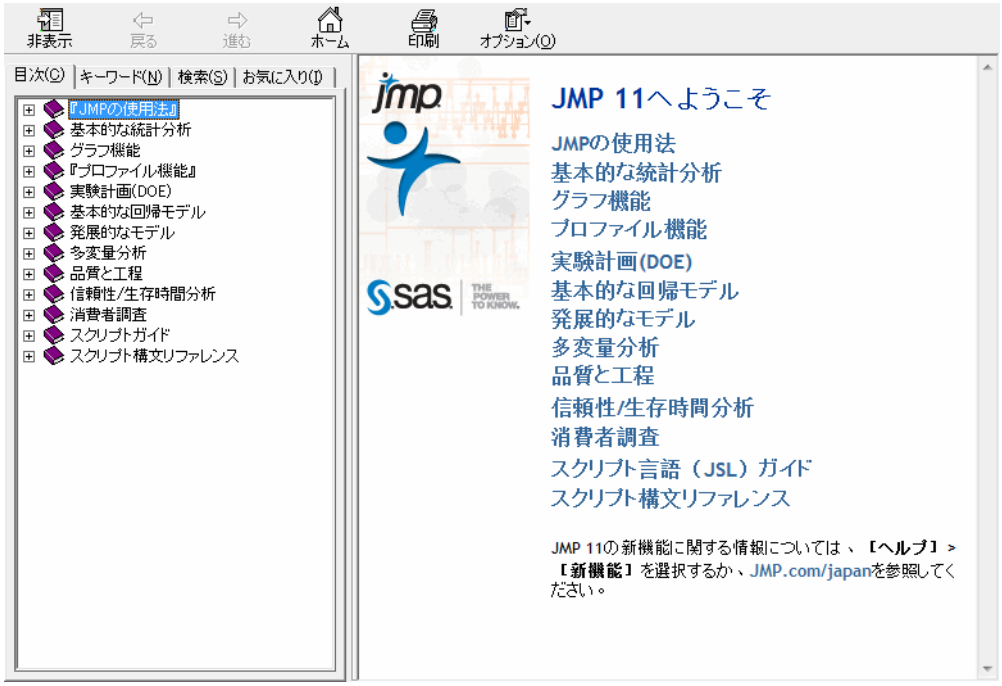
JMP の概要

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 13
JMP のマニュアル 13
JMP ドキュメンテーションライブラリ 14
JMP ヘルプ 18
JMP を習得するためのその他のリソース 18
チュートリアル 18
サンプルデータテーブル 19
統計用語と JSL 用語の習得 19
JMP を使用するためのヒント 19
ツールヒント 19
JMP User Community 20
JMPer Cable 20
JMP 関連書籍 20
「JMP スターター」ウィンドウ 20

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
- JMP Pro のみの機能には JMP Pro アイコン JMP PRO 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。
- 電子本は Amazon、Safari Books Online、および Apple iBookstore でお求めになれます。

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 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）］**メニューのすべてのトピックについて。
『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

マニュアル	目的	内容
『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］>［モデリング］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］>［モデリング］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
『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 • パレート図 • 特性要因図

マニュアル	目的	内容
『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 • 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 • 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。
『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。
- http://jmp.com/support/help/ でヘルプを検索します（英語のみ）。

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」（18ページ）を参照）
- サンプルデータ（「サンプルデータテーブル」（19ページ）を参照）
- 索引（「統計用語とJSL用語の習得」（19ページ）を参照）
- 使い方ヒント（「JMPを使用するためのヒント」（19ページ）を参照）
- Webリソース（「JMP User Community」（20ページ）を参照）
- 専門誌『JMPer Cable』（「JMPer Cable」（20ページ）を参照）
- JMPに関する書籍（「JMP関連書籍」（20ページ）を参照）
- JMPスターター（「「JMPスターター」ウィンドウ」（20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。次の操作はすべて［**ヘルプ**］>［**サンプルデータ**］で表示されるウィンドウで行えます。

- サンプルデータディレクトリを開く。
- すべての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。
- カテゴリ別に整理されたリストからサンプルデータテーブルを見つける。

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。

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/japan/academic/books.shtml

「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

消費者調査について

消費者調査データに対する分析の概要

「どのように製品やサービスを消費者が利用しているか？」、また、「顧客の満足度はどのくらいか？」などを、各企業が調べるときがあります。そして、その結果にもとづき、製品やサービスの品質を良くし、顧客の満足度を向上させています。

JMP のバージョン 11 以降では、消費者調査のデータを分析する機能が備わっています。バージョン 10 以前でも、JMP では、製造業での実験を中心として、実験計画や分析の機能が充実していました。バージョン 11 以降では、消費者調査データの分析機能も、より充実しました。それらの一連の統計分析は、分かりやすいように、すべて［消費者調査］メニューにまとめられています。以下のようなプラットフォームが用意されています。

- 「カテゴリカル」プラットフォームは、アンケート調査におけるカテゴリカルな回答を集計するのに適しています。選択式の質問に対する、さまざまな形式の回答データを集計できます。多重応答（複数回答）も集計できます。度数表やグラフがレポートされます。また、応答の平均スコアを計算したり、それらの差を統計的検定によって比較したりもできます。また、度数表をデータテーブルとして出力できます。詳細については、第 3 章「カテゴリカルな応答の分析」を参照してください。
- 「因子分析」プラットフォームは、因子分析を行います。因子分析は、多変量データを少数の因子で表そうとする分析です。因子分析のモデルは、「多数の観測変数が、少数の潜在変数（因子）によって説明されている」と仮定したモデルです。このプラットフォームでは、因子の個数を決めるときに参考になるグラフや、各因子と各変数の関係の強さを表す因子パターンや因子構造が出力されます。詳細については、第 4 章「因子分析」を参照してください。
- 「選択モデル」は、選択実験のデータを分析するためのプラットフォームです。選択実験では、どちらの商品がより好ましいかを、被験者に選択してもらいます。消費者の選好構造を明らかにし、消費者に好まれる商品を探り出すことができます。詳細については、第 5 章「選択モデル」を参照してください。
- 「アップリフト」プラットフォームは、ダイレクトメールを送付したときのレスポンスが、送付しなかったときのレスポンスに比べて大きい集団を、探し出すのに役立ちます。この分析は、性別や年齢などの考慮すべき背景変数が数多く、データ量が比較的大きい場合でも処理できます。マーケティング分野だけではなく、テイラーメイド治療や個別化医療に応用することも考えられます。詳細については、第 6 章「アップリフトモデル」を参照してください。
- 「項目分析」プラットフォームは、項目反応理論（IRT）に基づいて、テストの正誤データを分析します。項目反応理論は、「能力」といった潜在的な変数を想定しているテストや、同じように、消費者が潜在的な嗜好をもつと仮定し、「好んでいるか？否か？」といったデータ形式のアンケートをモデル化するのに使われています。約 1 世紀にわたって古典的なテスト理論が広く使用されてきましたが、項目反応理論は、より精緻なモデルに基づいています。詳細については、第 7 章「項目分析」を参照してください。

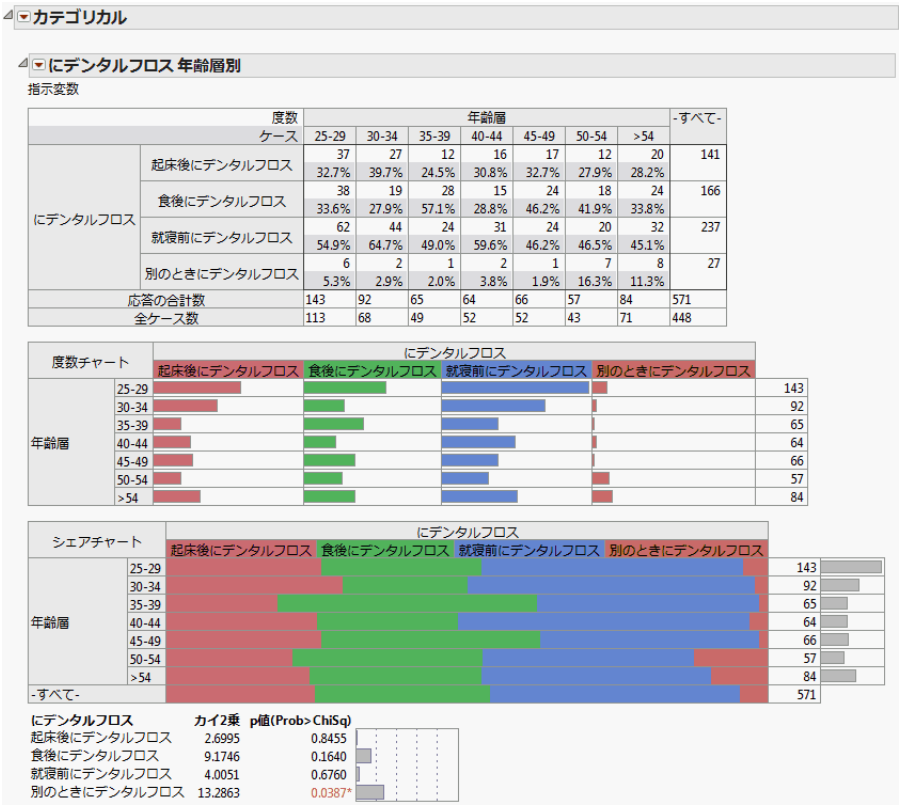
第 3 章

カテゴリカルな応答の分析

アンケート調査データなどの度数表

「カテゴリカル」プラットフォームは、カテゴリカルな応答データを集計し、検定統計量を計算するためのプラットフォームです。「カテゴリカル」プラットフォームの長所は、さまざまな形式の応答を、余分なデータ加工をせずにそのまま処理できる点です。アンケート調査、工業製品の不適合原因や故障原因、医薬品の副作用や有害事象といった多重応答データも扱うことができます。

図 3.1 カテゴリカルな応答データの分析例



カテゴリカル

にデンタルフロス 年齢層別

指示変数

度数		年齢層							-すべて-
ケース		25-29	30-34	35-39	40-44	45-49	50-54	>54	
にデンタルフロス	起床後にデンタルフロス	37	27	12	16	17	12	20	141
		32.7%	39.7%	24.5%	30.8%	32.7%	27.9%	28.2%	
	食後にデンタルフロス	38	19	28	15	24	18	24	166
		33.6%	27.9%	57.1%	28.8%	46.2%	41.9%	33.8%	
	就寝前にデンタルフロス	62	44	24	31	24	20	32	237
		54.9%	64.7%	49.0%	59.6%	46.2%	46.5%	45.1%	
	別のときにデンタルフロス	6	2	1	2	1	7	8	27
		5.3%	2.9%	2.0%	3.8%	1.9%	16.3%	11.3%	
応答の合計数		143	92	65	64	66	57	84	571
全ケース数		113	68	49	52	52	43	71	448

にデンタルフロス	カイ2乗	p値(Prob>ChiSq)
起床後にデンタルフロス	2.6995	0.8455
食後にデンタルフロス	9.1746	0.1640
就寝前にデンタルフロス	4.0051	0.6760
別のときにデンタルフロス	13.2863	0.0387*

目次

「カテゴリカル」プラットフォームの概要 25
「カテゴリカル」プラットフォームの使用例 25
「カテゴリカル」プラットフォームの起動 27
応答の役割 27
選択した列に役割を割り当てる 31
起動ウィンドウのその他のオプション 32
「カテゴリカル」レポート 33
シェアチャート 35
度数チャート 35
「カテゴリカル」プラットフォームのオプション 36
レポートオプション 36
統計量のオプション 37
自由回答のレポートオプション 50
表の構成レポートのオプション 52
「カテゴリカル」プラットフォームのその他の例 53
多重応答 54
応答の度数 55
指示変数 56
多重応答 区切り文字 57
多重応答 ID別 58
スコアの平均の例 59

「カテゴリカル」プラットフォームの概要

「カテゴリカル」プラットフォームでは、データをどのような構成で分析するかを、ユーザがタブ式パネルで指定します。「カテゴリカル」プラットフォームと同様の分析は、他のプラットフォームで行うこともできます。どのプラットフォームを選ぶかは、分析の種類や目的、およびデータの形式などによって決まります。「カテゴリカル」プラットフォームの長所は、さまざまな形式の応答を、余分なデータ加工をせずにそのまま処理できる点です。表3.1では、カテゴリカル分析を行えるプラットフォームと、その長所をまとめています。

表3.1 JMPでのカテゴリカル分析の比較

プラットフォーム	特徴
分布	カテゴリカルな応答を個別に分析
二変量の関係: 分割表	カイ2乗検定、対応分析、一致性など
パレート図	不適合の原因などの多重応答データをグラフで分析。［二変量の関係］とは異なり、ユニットあたりの発生率を比較
変動性図: 計数値	計数値用ゲージ調査。判定者の一致性において詳細な分析が可能
モデルのあてはめ	ロジスティックモデルと一般化線形モデル
パーティション、ニューラル	カテゴリカル応答に対するモデルの一種

「カテゴリカル」プラットフォームの使用例

この例では「Consumer Preferences.jmp」を使用します。このデータテーブルには、人々の意見や姿勢を示すデータ、口腔衛生に関する質問への回答が含まれています（出典: Rob Reul、Isometric Solutions）。

1. 「Consumer Preferences.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. **「仕事熱心である」**を選択して、［単純］タブの［**応答**］をクリックします。
4. **「年齢層」**を選択して［**X, グループ化カテゴリ**］をクリックします。
5. ［**OK**］をクリックします。
6. 「カテゴリカル」の赤い三角ボタンをクリックし、［**クロス表形式 転置**］を選択します。
7. 「カテゴリカル」の赤い三角ボタンをクリックし、［**応答の等質性に対する検定**］を選択します。

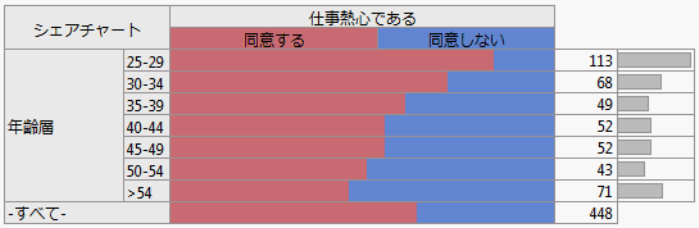
図3.2は、「仕事熱心であるか？」という質問に対する回答と、回答者の年齢層を示しています。この分析によれば、448名の回答者全体のうち、64.1%が「仕事熱心である」と回答しています。「仕事熱心である」という回答の割合が最も多かった年齢層は25～29歳で、その割合は84.1%でした。逆に、最も少なかった年齢層は54歳以上で、46.5%でした。

図 3.2 年齢層別のアンケート結果

カテゴリカル

仕事熱心である 年齢層別

度数		年齢層							-すべて-
シェア		25-29	30-34	35-39	40-44	45-49	50-54	>54	
仕事熱心である	同意する	95	49	30	29	29	22	33	287
		84.1%	72.1%	61.2%	55.8%	55.8%	51.2%	46.5%	64.1%
	同意しない	18	19	19	23	23	21	38	161
		15.9%	27.9%	38.8%	44.2%	44.2%	48.8%	53.5%	35.9%
応答の合計数		113	68	49	52	52	43	71	448



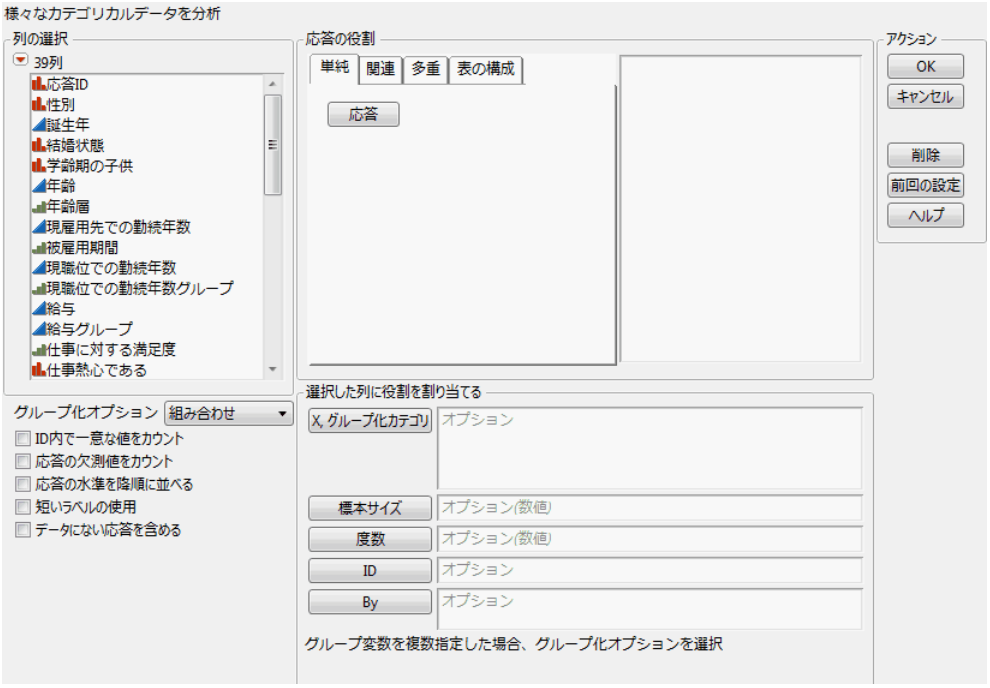
応答の等質性に対する検定

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	39.5874	<.0001*
Pearson	37.4593	<.0001*

「カテゴリカル」プラットフォームの起動

「カテゴリカル」プラットフォームを起動するには、［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。

図 3.3「カテゴリカル」プラットフォームの起動ウィンドウ



起動ウィンドウには、複数のタブがあります。そのうち、［単純］、［関連］、［多重］というタブは、応答がどのような種類であるかを示すものです。［表の構成］タブでは、表を自由に構成できます。これらのタブについて、以下の節で説明していきます。

応答の役割

集計する列を指定するには、その列を選択した後、各タブにある該当するボタンをクリックします。列名をリストボックスにドラッグしても、指定できます。各タブにおいて、どのような集計が行われるかについては、表3.2にまとめています。

［単純］タブ

［単純］タブで指定された列は、単純集計されます。データが特別な構造をもたず、指定された列ごとに単純集計したい場合には、このタブを使ってください。このタブには、［応答］ボタンがあるだけです。列を指定するには、列名を「列の選択」リストから応答リストにドラッグするか、列を選択して［**応答**］ボタンをクリックしてください。なお、「多重応答」の列プロパティが列に設定されている場合は、その列は自動的に多重応答（複数回答）の列だとして処理されます。

［関連］タブ

データにおける複数の列がまったく同じ値をもっており、関連性をもって集計したい場合には［関連］タブを使ってください。

共通の値をもつ応答　集計方法としては、［単純］タブと同じ単純集計が行われます。ただし、複数の列に対する結果が別々の表やグラフに表示されるのではなく、レポートをコンパクトにまとめるため、1つの表やグラフだけに並列して表示されます。

反復測定　このタブは、データが反復測定データであり、時間の経過とともに、どのように応答が変化しているのかを見たい場合に用いてください。データテーブルの各列が、時点ごとの応答であるときに用いてください。

判定の一致性　このタブは、ある対象に対する判定が、複数の判定者（個人）の間でどれぐらい異なっているかを見たい場合に用いてください。データテーブルの各列が、判定者一人ひとりに対応しており、各行が異なる対象に対する判定結果となっているときに用いてください。

［多重］タブ

［多重］タブは、データが多重応答（複数回答）である場合にお使いください。「多重応答」とは、ある 1 つの質問に対して、選択肢を複数選べる形式のデータを指します。多重応答の結果をデータで保持するには、いくつかの形式があります。このタブには、それらの形式に対応するため、複数のボタンが用意されています。

多重応答　複数の列にまたがって多重応答が保持されているデータ形式の場合に、用いてください。このデータ形式では、一部の列は欠測値となっています。

多重応答 ID 別　複数の行にまたがって多重応答が保持されているデータ形式の場合に、用いてください。このデータ形式では、個人が ID 列によって識別されており、その ID ごとに、複数行にわたって多重応答が保持されている形式です。

多重応答 区切り文字　1つの列だけにカンマ区切りで多重応答が保持されているデータ形式の場合に、用いてください。

指示変数　多重応答の応答ごとに 1 つの列が対応しており、それらの各列が指示変数（0 か 1 などの 2 値のみ）であるデータ形式の場合に用いてください。

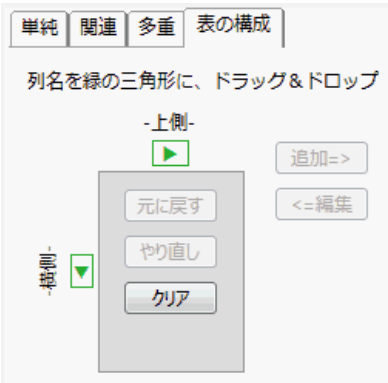
応答の度数　応答ごとに 1 つの列があり、その列が度数が含んでいるデータ形式の場合に用いてください。

自由回答　データが記述式の自由回答である場合に用いてください。なお、日本語には対応しておらず、英語のようなスペースで単語が区切られている言語のみに対応しています。実行すると、各単語の頻度が数えられ、その頻度が単語順と度数順で表示されます。また、無記入でないテキストの割合も表示されます。自由回答についての詳細は、「自由回答のレポートオプション」（50 ページ）を参照してください。

［表の構成］タブ

［表の構成］タブでは、記述統計量を複雑な形式のテーブルにまとめることができます。それには、列名を緑のドロップゾーンにドラッグし、横並びや枝分かれといった構成を作成します。テーブルの上側や横側をどのように構成するかによって、変数を別の変数の中で枝分かれさせたり、横に並べたりできます。列を次々に別の列の中または横にドラッグし、構成を仕上げます。レポートの構成ついての詳細は、「表の構成レポートのオプション」（52 ページ）を参照してください。

図 3.4 ［表の構成］タブ



次の手順によって、複雑な構成を指定していくことができます。

1. 列名を該当するドロップゾーンにドラッグすることで、他の変数の横や下に変数が追加されます。
2. 列名を該当するドロップゾーンにドラッグしていくことで、表の横側や下側に、次々と変数を追加していくことができます。

   なお、直前に追加した変数を削除するには、［**元に戻す**］をクリックします。

   また、直前に削除したものを再度追加するには、［**やり直し**］をクリックします。

   設定のすべて消去するには、［**クリア**］をクリックします。
3. 表の構成が終了したら、［**追加 =>**］をクリックして、右側に表示されているリストボックスに、その構成を追加します。

   追加した後に、もう一度、その表の構成を変更したい場合、変更したい表をクリックした後、［**<= 編集**］をクリックしてください。
4. 起動ウィンドウで、必要に応じて残りの設定を行い、［**OK**］をクリックします。

   「カテゴリカル」レポートウィンドウが表示されます。
5. いくつかの設定を少し変更して、もう一度、集計を行いたい場合は、「カテゴリカル」の赤い三角ボタンメニューから［**ダイアログの再起動**］を選択してください。起動ウィンドウが呼び出されて、項目や構成を変更できます。

［表の構成］での効果に関する注意

- ［表の構成］では、表の横側にある、最も内側の項目が応答（Y 変数）とみなされます。
- ［単純］タブにおいても、列のデータが区切り文字の多重応答の形式であり、かつ、分析対象の列に「多重応答」の列プロパティが設定されていれば、その列は多重応答として扱われます。この「多重応答」の列プロパティは、［**列情報**］ウィンドウで設定できます。

表 3.2 応答の役割

応答の役割	説明	データ例
［単純］タブ		
応答	それぞれ個別の列に入っている応答。個別に分析されます。	**ID** **飲み物** **メイン** John コーヒー チキン Jane 紅茶 野菜
［関連］タブ		
共通の値をもつ応答	全列でカテゴリが共通している応答。すっきりと整理されたレポートが作成されます。	**ID** **コーヒー** **紅茶** John 好き 嫌い Jane 嫌い 好き
反復測定	複数の時点や状況で測定された1つの個体からの応答。	**ID** **朝** **昼** **夜** John コーヒー コーヒー 水 Jane 紅茶 水 紅茶
判定の一致性	同じユニットを異なる判定者が評価したときの応答。判定者間に見られる一致性を調べます。	**飲み物** **John** **Jane** コーヒー 好き 嫌い 紅茶 嫌い 好き 水 好き 好き
［多重］タブ		
多重応答	複数の列に入力されているが、1つのグループとして扱われる多重応答。	**ID** **飲み物1** **飲み物2** **飲み物3** John コーヒー ミルク 水 Jane 紅茶 水
多重応答 ID 別	ID の値が同じである複数の行に入力された多重応答。	**ID** **飲み物** John コーヒー John ミルク John 水 Jane 紅茶 Jane 水

表 3.2 応答の役割（続き）

応答の役割	説明	データ例
多重応答 区切り文字	1つのセルに複数の応答を入力し、カンマで区切ったもの。	**ID** **飲み物** John コーヒー, ミルク, 水 Jane 紅茶、水
指示変数	同じ種類の複数の列に含まれている2値応答（「選択／選択解除」、「はい／いいえ」など）。	**ID** **コーヒー** **ミルク** **紅茶** **水** John Y Y N Y Jane N N Y Y
応答の度数	同じ種類の複数の列に含まれている応答の各水準の度数。	**グループ** **コーヒー** **ミルク** **紅茶** **水** A 12 15 8 19 B 9 20 6 22
自由回答	自由回答における各単語の頻度を数える。日本語はサポートされていません。英語のようにスペースで区切られている言語しかサポートしていません。	**ID** **コメント** John I liked the coffee. Jane The juice was too sweet.
［表の構成］タブ		
	変数を緑のドロップゾーンにドラッグして、独自の表を構成する。	

選択した列に役割を割り当てる

起動ウィンドウの右下では、次のような役割を列に指定できます。

X, グループ化カテゴリ　グループ分けに用いる変数（X変数）を指定します。なお、複数のX変数が指定された場合、デフォルトでは、それらすべての変数を組み合わせてグループが構成されます。1つのX変数ごとに1つの表を作成したい場合には、起動ウィンドウの左下にある［グループ化オプション］で指定します。

標本サイズ　多重応答のデータを集計する場合に、属するグループの大きさ（含まれるユニットの数）を定義します。たとえば、ある製品を100個、検査したときに、50箇所で不適合がみつかった場合、「標本サイズ」は「100個」で、多重応答の度数は「50箇所」となります。

度数　度数を含む列を指定します。これは、分析に用いるデータが、すでに集計された形式である場合に用いてください。

ID　［多重応答 ID別］を選択した場合のみ必要です。

By　ここで指定された列の値ごとに、別々に、分析が行われます。

起動ウィンドウのその他のオプション

起動ウィンドウの左下には、いくつかのオプションが用意されています。これらのオプションは、分析を実行した後に、「カテゴリカル」の赤い三角ボタンメニューから選択することもできます。また、これらのオプションのいくつかは、「カテゴリカル」の赤い三角ボタンメニューで、デフォルト値を変更できます。詳細については、「環境設定の変更」（49 ページ）を参照してください。

グループ化オプション　X 列を個別に使用するか、組み合わせで使用するかを指定します。このオプションは、複数の X（グループ化）列を指定する場合にのみ使用し、X 列をそれぞれ個別に扱うか、組み合わせて扱うか、それともその両方を行うかを指定します。たとえば、X 列が「宗教」と「年齢層」の場合なら、「宗教」と「年齢層」を別々に集計することも、宗教の各グループから年齢層が枝分かれした形式の表（つまり、「宗教」と「年齢層」とを組み合わせた表）を作成することも、また、その両方を作成することもできます。

［**組み合わせ**］は、X 変数の組み合わせに対して度数を求めます。

［**それぞれを個別に**］は、X 変数ごとに個別に度数を求めます。

［**両方**］は、X 変数の組み合わせと、個々の X 変数の両方に対して度数を求めます。

ID 内で一意な値をカウント　重複した応答の水準を、被験者ごとに 1 度だけカウントします。［ID］変数が指定されていなければなりません。

次のオプションは、起動ウィンドウまたは「カテゴリカル」の赤い三角ボタンメニューから指定できます。また、「環境設定」でも設定可能です。詳細については、「統計量のオプション」（37 ページ）と「環境設定の変更」（49 ページ）を参照してください。

応答の欠測値をカウント　このオプションをオンにすると、表において、欠測値のグループも、1 つのカテゴリとしてカウントされます。ただし、この欠測値のグループは、統計的検定の比較では除外されたままです。このオプションにより、欠測値をカウントして表に表示するか、完全に除外するかを指定できます。欠測値には、2 種類あります。1 つ目は、標準の欠測値で、数値の場合はピリオド（NaN）、文字列の場合は空文字で表されます。2 つ目は、列プロパティの「欠測値のコード」において欠測値として扱うように指定されたデータ値です。

応答の水準を降順に並べる　X 変数の順序は昇順のまま、応答を降順に変更します。デフォルトの順序は昇順です。順序は列プロパティ（**「値の順序」**）で制御できますが、常に降順で表示したい場合は、このオプションを選択してください。「評価」や「態度」といった順序のあるカテゴリは、低い順ではなく、高い順から並べたほうが見やすいかもしれません。

短いラベルの使用　共通する接頭部や接尾部を省略し、短いラベルで表示します。アンケートでは、共通の接頭部や接尾部を含んだ長いラベルが使われることがあります。たとえば、「昨年 1 年間で 5 ～ 10 回」というラベルがあった場合、「昨年 1 年間で」の部分はすべての値ラベルで繰り返し使用されており、レポートに表示する必要はありません。このオプションを使用すれば、共通の接頭部と接尾部が省略されます。また、複数の空白は 1 つの空白に置き換えられます。このオプションは、データ値や値ラベルだけに適用されるものであり、列名には適用されません。

データにない応答を含める　データ中には存在していない値も、度数を「0」として、表に表示します。アンケート調査においては、調査紙の選択肢には存在しているのに、誰にも選択されなかったために、データ

には存在しないものがあります。そのような選択肢も表に表示したい場合は、このオプションを使用してください。なお、データには登場していない値は、その列の「値ラベル」列プロパティから特定されます。この機能を用いる場合には、事前に「値ラベル」列プロパティで、値を設定しておいてください。

上位カテゴリ

たとえば、5段階評価の回答があった場合、『「1」と「2」のいずれかを答えた人』の人数や割合を求めたいことがあります。列プロパティの「上位カテゴリ」を使用すると、このようなグループも集計できます。

先ほどの例のように、**上位カテゴリ**は、複数のカテゴリをまとめて 1 つにしたい場合に用います。「上位カテゴリ」の列プロパティでは、「グループ」、「平均」、「標準偏差」、「すべて」という 4 つのオプションが用意されています。なお、このうち、「平均」や「標準偏差」は、値スコアから、平均や標準偏差を計算するためのものです。また、「すべて」は、全水準の度数を求めるためのものです。一方、「グループ」は、先ほどの例のように、『「1」と「2」のいずれかを答えた人』の人数のように、グループ分けします。なお、「平均」、「標準偏差」、「すべて」は使用をお勧めしません。これら 3 つは、「上位カテゴリ」を用いなくても、通常のコマンドでも計算できます。

上位カテゴリを作成するには、データテーブルにて列を選択した後、［**列情報**］>［**列プロパティ**］>［**上位カテゴリ**］を選択します。列プロパティのウィンドウで、列を選択してから、［**上位カテゴリ名**］に上位カテゴリ名を入力して［**追加**］をクリックします。上位カテゴリだけを表示するには、赤い三角ボタンのメニューから、［**オプション**］>［**非表示**］を選択します。このオプションを使うと、カテゴリの下位にあたるデータ値は表示されず、設定した上位カテゴリだけが表示されます。

上位カテゴリは、スクリプトでも設定できます。「上位カテゴリ」（supercategories）の列プロパティに、次のようなコードを送ると、上位カテゴリが設定されます。

```
{Group( "Top Two", {8, 9} ), All, Mean, StdDev}
```

上位カテゴリは、「カテゴリカル」の起動コマンドでも指定できます。その場合は、次のようにプロパティを列名の後の括弧内にリストします。

```
Categorical(Supercategories(Y({Group("Top Two",{4,5}),All})),…
```

上位カテゴリの機能は、［反復測定］と［判定の一致性］を除くすべての応答効果で使用できます。ただし、自然数のスコアが設定されていない列に対しては、たとえ上位カテゴリに「平均」や「標準偏差」を設定したとしても、それらの統計量は計算されません。

「カテゴリカル」レポート

「カテゴリカル」プラットフォームでは、選択した項目に応じて、度数表やグラフが作成されます。選択したオプションや応答の種類によっては、この節で紹介するオプションの一部は使えません。単純な度数表においては、デフォルトで出力される統計量は、「度数」と「シェア」です（多重応答の場合には、ケースあたりの比率も出力されます）。また、グラフとしては、デフォルトでは、シェアチャートが表示されます。「度数チャート」は、デフォルトでは表示されていません。表示したい場合には、該当するコマンドを赤い三角ボタンから選択してください。

各オプション（「**度数**」、「**応答のシェア**」、「**ケースあたりの比率**」、「**シェアチャート**」、「**度数チャート**」、「**度数チャート 転置**」）の表示／非表示は、「カテゴリカル」の赤い三角ボタンメニューで切り替えます。図3.5は、サンプルデータ「Consumer Preferences.jmp」に対するレポートの一例です。以下、この図を使って説明していきます。

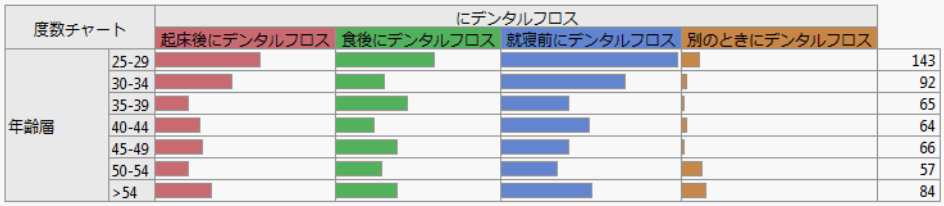
図3.5「カテゴリカル」レポート

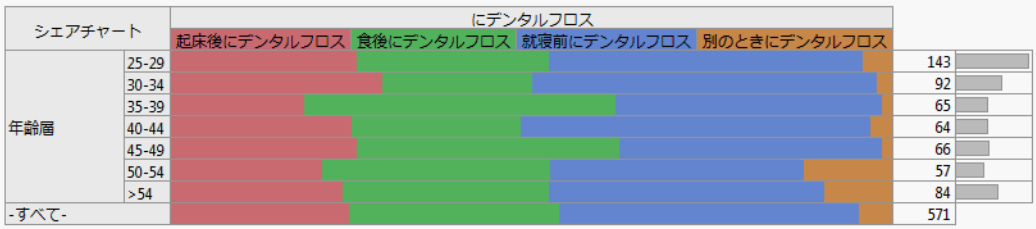
カテゴリカル

にデンタルフロス 年齢層別

指示変数

度数 / シェア / ケース		年齢層							-すべて-
		25-29	30-34	35-39	40-44	45-49	50-54	>54	
にデンタルフロス	起床後にデンタルフロス	37	27	12	16	17	12	20	141
		25.9%	29.3%	18.5%	25.0%	25.8%	21.1%	23.8%	24.7%
		32.7%	39.7%	24.5%	30.8%	32.7%	27.9%	28.2%	
	食後にデンタルフロス	38	19	28	15	24	18	24	166
		26.6%	20.7%	43.1%	23.4%	36.4%	31.6%	28.6%	29.1%
		33.6%	27.9%	57.1%	28.8%	46.2%	41.9%	33.8%	
	就寝前にデンタルフロス	62	44	24	31	24	20	32	237
		43.4%	47.8%	36.9%	48.4%	36.4%	35.1%	38.1%	41.5%
		54.9%	64.7%	49.0%	59.6%	46.2%	46.5%	45.1%	
	別のときにデンタルフロス	6	2	1	2	1	7	8	27
		4.2%	2.2%	1.5%	3.1%	1.5%	12.3%	9.5%	4.7%
		5.3%	2.9%	2.0%	3.8%	1.9%	16.3%	11.3%	
応答の合計数		143	92	65	64	66	57	84	571
全ケース数		113	68	49	52	52	43	71	448





にデンタルフロス	カイ2乗	p値(Prob>ChiSq)
起床後にデンタルフロス	2.6995	0.8455
食後にデンタルフロス	9.1746	0.1640
就寝前にデンタルフロス	4.0051	0.6760
別のときにデンタルフロス	13.2863	0.0387*

表は、「度数」を表しています。表の一番下側には、「応答の合計数」（各カテゴリの度数の合計）と、「全ケース数」（回答者の人数や、ユニットの個数などを示す統計量）が一番下にあります。

この例では、7つに分かれた年齢層の応答数とケース数が表示されています。最後の2行には、応答（571）とケース（448）の合計数が表示されています。

各セルの「シェア」は、応答の各度数を、その合計で割ったものです。これは、すべての応答に占める各応答の割合を示します。応答が表の横側にある場合には、縦方向に足すと100%になっています。

たとえば、表の2行目にある**「起床後にデンタルフロス」**を見てみましょう。起床後にデンタルフロスをすると答えた37人は、応答の合計数の25.9パーセントに相当します（37/143*100）。

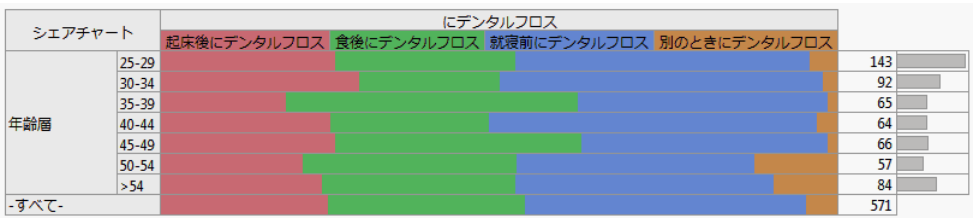
各セルの「ケース」には、1ケースあたりの比率が表示されています。これは、応答の各度数を、ケース数で割ったものです。多重応答（複数回答）の場合には、「シェア」と「ケース」という2種類の統計量が計算されます。「ケース」は、ケースの合計数に対する度数の比率です。一方、「シェア」は、応答の合計数に対する度数の割合です。「ケース」は、応答が多重応答の場合だけに計算されます。

たとえば、表の3行目（**「起床後にデンタルフロス」**）は、回答者113名に対して、37箇所（この例では、重複がないので、37名）が、この項目を選択しています。よって、1回答者あたりの比率（この例では、割合）は32.7パーセントとなります。

シェアチャート

「シェアチャート」には、分割した棒グラフが表示されます。棒の長さは、各種類に対する応答の割合を表します。右側の棒グラフは応答数を示しています。

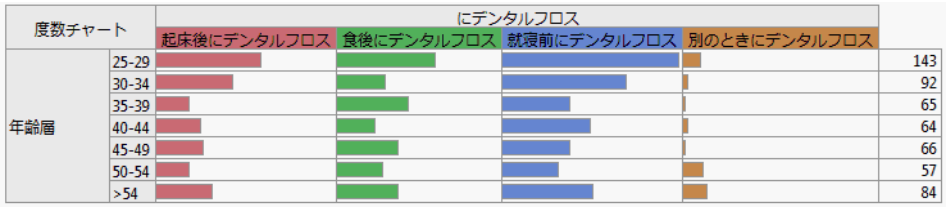
図3.6 シェアチャート



度数チャート

「度数チャート」は、応答の度数が棒グラフで示されています。すべての棒グラフには、同じスケールが使われています。チャートの右側に応答の度数が示されています。度数チャートを表示するには、「カテゴリカル」の赤い三角ボタンのメニューから［**度数チャート**］を選択してください。

図3.7 度数チャート



［度数チャート 転置］は、転置した度数チャートを表示します。行と列とが転置した度数チャートが描かれます。

「カテゴリカル」プラットフォームのオプション

「カテゴリカル」の赤い三角ボタンで用意されているコマンドには、レポートの見た目をカスタマイズするものや、統計的検定を行うものがあります。このメニューには、選択した応答の役割とオプションに応じて、次のようなコマンドが表示されます。選択した項目によっては、コマンドの一部は表示されません。

レポートオプション

デフォルトのレポート形式は［**クロス表形式**］です。これは、各標本と応答の3つの統計量をすべてまとめたものです。［クロス表形式］では、上側に応答、横側に標本カテゴリが配置されます。また、各セルには、複数の表の要素が表示されます。

図3.8 クロス表形式

度数 洗浄, 日付別

度数列

			度数 シェア ケース	度数							応答の合計数	全ケース数
				汚れ	腐食	ドーピング	金属被覆	その他	酸化物不良	シリコン不良		
洗浄	後	日付	10月1日	12	2	0	4	2	1	2	23	50
				52.2%	8.7%	0.0%	17.4%	8.7%	4.3%	8.7%		
				24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%		
			10月2日	10	1	1	5	1	2	3	23	50
				43.5%	4.3%	4.3%	21.7%	4.3%	8.7%	13.0%		
				20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%		
			10月3日	8	3	0	5	0	1	0	17	50
				47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%		
				16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%		
	前	日付	10月1日	14	2	1	2	3	8	1	31	50
				45.2%	6.5%	3.2%	6.5%	9.7%	25.8%	3.2%		
				28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%		
			10月2日	15	2	2	1	4	6	0	30	50
				50.0%	6.7%	6.7%	3.3%	13.3%	20.0%	0.0%		
				30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%		
			10月3日	22	2	3	4	0	3	2	36	50
				61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%		
				44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%		
-すべて-				81	12	7	21	10	21	8	160	300
				50.6%	7.5%	4.4%	13.1%	6.3%	13.1%	5.0%		

［クロス表形式］は、転置が可能です（［**クロス表形式 転置**］）。これは、応答カテゴリの数が多く、標本数はさほど多くない場合に便利です。［クロス表形式 転置］では、横側に応答、上側に標本カテゴリが配置されます。また、各セルには、いくつかの統計量が表示されます。

［表の構成］では、常に［クロス表形式 転置］が使用されますが、より複雑な配置になります。［自由回答］の場合は、特殊なレポートが作成されます。自由回答についての詳細は、「自由回答のレポートオプション」（50ページ）を参照してください。

［**凡例**］は、シェアチャートにおいて、応答列の凡例の表示／非表示を切り替えます。

統計量のオプション

どの形式の表でも、分析者の関心は、応答のシェアや1ケースあたりの比率が、標本グループによって異なるかどうかになるでしょう。特に、シャアや比率が、他のグループと異なっているグループを見つけることが重要でしょう。

統計的検定には、主に2種類のものがあります。単一の応答に対する検定と、多重応答（複数回答）に対応する検定です。

- 応答が単一式の場合（つまり、回答者は1つのカテゴリだけを選択する場合）、研究者の関心は、「応答のシェア」が標本グループ間で異なるかどうかにあるでしょう。
- 応答が多重応答の場合（つまり、回答者が複数のカテゴリを選択できる場合）、研究者の関心は、標本グループ間で「1ケースあたりの比率」が異なるかどうかにあるでしょう。

応答が単一式の場合、検定としては、等質性に対するカイ2乗検定が実行されます。この検定には、［尤度比カイ2乗］と［Pearsonカイ2乗］の2種類があります。これら2つの検定のうち、どちらの検定のほうが良いかは、分析者の好みと経験に左右されるでしょう。「カテゴリカル」の赤い三角ボタンのメニューにある［カイ2乗検定の選択］オプションで、いずれかの検定、もしくは両方を表示できます。詳細については、「検定オプション」（48ページ）を参照してください。

応答が多重応答の場合、イベント発生の度数がPoisson分布に従うと仮定した上で、各応答が個別に比較されます（この検定では、ある1名の回答者が、同じカテゴリを2度以上、選択することも許されます）。この検定では、応答の各カテゴリについて、パラメータがグループ間で同じかどうかが検定されます。

「カテゴリカル」の赤い三角ボタンのメニューには、状況に応じて次のようなオプションが表示されます。

表 3.3「カテゴリカル」プラットフォームのコマンド

	コマンド	応答の種類	調べる内容	詳細
多重応答データがある場合	応答の等質性に対する検定	• 応答 • 共通の値をもつ応答 • 反復測定 • 応答の度数（[標本サイズ] を指定しない場合） • 表の構成	応答の各カテゴリが生じる確率のパターンが、すべてのグループで同一のものかどうか。	周辺等質性（独立性）検定、Pearson のカイ 2 乗と尤度比カイ 2 乗 詳細については、「応答の等質性に対する検定」（40 ページ）を参照してください。
	各応答に対する検定	• 多重応答 • 多重応答 ID 別（[標本サイズ] を指定した場合） • 多重応答 区切り文字 • 標本サイズを使った応答の度数 • 表の構成	応答の特定のカテゴリにおけるケースあたりの比率（1 ケースあたりの平均発生度数）が、すべてのグループで同一のものかどうか。	度数を応答とし、グループを因子とした Poisson 回帰。詳細については、「多重応答の各応答に対する検定」（42 ページ）を参照してください。
	一致性の統計量	判定の一致性	判定者間にどの程度の一致性が見られるか。不一致性は対称か。	一致性を表すカッパ統計量、対称性を表す Bowker／McNemar 検定。詳細については、「判定の一致性」（45 ページ）を参照してください。
	遷移レポート	反復測定	時間が経過するにつれてカテゴリにどのような変化が見られるか。	遷移度数と遷移率行列 詳細については、「反復測定」（45 ページ）を参照してください。
	セルのカイ 2 乗	応答	より詳しい情報を得るためには、どのように分析を進めればよいか。	詳細については、「セルのカイ 2 乗」（43 ページ）を参照してください。

表 3.3「カテゴリカル」プラットフォームのコマンド（続き）

コマンド	応答の種類	調べる内容	詳細
標本の各ペア比較	• 応答 • 共通の値をもつ応答 • 反復測定 • 応答の度数（[標本サイズ] を指定しない場合） • 表の構成	応答確率全体を、標本グループ間でペア比較する。	詳細については、「標本の各ペア比較」（46 ページ）を参照してください。
セルの各ペア比較	• 単一応答および多重応答 • 表の構成	応答確率を、セルごとに、標本グループ間でペア比較する。	詳細については、「セルの各ペア比較」（47 ページ）を参照してください。
検定オプション	• カイ 2 乗検定の選択 • 警告の表示 • p 値の小さい順に並べ替え • 有意でない結果を非表示にする	より詳細な情報を得るために、結果をさらに分析するにはどうすればいいか。	詳細については、「検定オプション」（48 ページ）を参照してください。

各標本グループの詳細を表示するオプションが用意されています。使用できるオプションは、選択内容や分析の詳細によって異なります。

応答の合計数　各標本グループの度数の合計を表示します。

全ケース数　多重応答の場合に、ケースの度数（たとえば、回答者の人数や、ユニットの個数など）を表示します。全ケース数は、上述の「応答の合計数」とは異なる点に注意してください。

応答している全ケース数　表の構成で使用される多重応答列は、少なくとも 1 回応答のあった人全員の人数を表示します。全く応答しなかった人は含まれません。

スコアの平均　応答変数が数値カテゴリや値スコアの場合、それらの平均を計算します。このコマンドは、応答変数のデータが数値の場合、または、**「値スコア」**プロパティが設定されている場合に使えます。平均が意味あるものにするために、分析を行う前に、データテーブルの列に対する「列情報」ウィンドウの「値スコア」プロパティで、妥当な値スコアを割り当ててください。詳細と例については、「スコアの平均の例」（59 ページ）を参照してください。

スコア平均の比較　標本グループ間の平均を比較（t 検定を使用）し、どのグループが有意に異なるかを調べます。結果の見方については、「比較の結果を表すアルファベット文字の見方」（48 ページ）を参照してください。

スコアの標準偏差　値スコアの標準偏差を算出します。

スコア平均で並べ替え 各応答変数に対する結果を、スコア平均の大きい順に並べ替えます。このオプションは、分析にX列が含まれない場合にのみ表示されます。

テーブルの保存 レポートを新しいデータテーブルに保存します。詳細については、「テーブルの保存」（49ページ）を参照してください。

フィルタ データを特定のグループまたは範囲でフィルタします。［ローカルデータフィルタ］パネルが開き、さまざまなデータのサブセットを指定することができます。詳細については、『JMPの使用法』を参照してください。

要約の目次 レポートの冒頭に、検定とスコア平均の要約とリンクを表示します。

要素の表示形式 度数、シェア、比率、ゼロ度数に対して、それらの表示形式を指定します。デフォルトでは、度数は固定少数点（総桁数7、小数桁数0）、シェアと比率はパーセント値（総桁数6、小数桁数1）です。

複数行に配置 レポートを、縦方向ではなく横方向に並べます。横方向に並べるレポートの数を入力してください。

環境設定の変更 次回以降の起動やセッションに備えて、環境設定を変更できます。詳細については、「環境設定の変更」（49ページ）を参照してください。

カテゴリに関するオプション ［グループ化オプション］、［応答の欠測値をカウント］、［応答の水準を降順に並べる］、［短いラベルの使用］、［データにない応答を含める］のオプションがあります。これらのオプションは起動ウィンドウにも表示され、分析前に指定できます。このメニューから選択した場合は、分析が新しい設定でもう一度実行されます。詳細については、「起動ウィンドウのその他のオプション」（32ページ）を参照してください。

クロス表を濃淡表示 クロス表を濃淡表示にします。このオプションは、環境設定において、デフォルトに設定することもできます。

ダイアログの再起動 起動ウィンドウを呼び出します。起動ウィンドウに戻り、もう一度、分析をやり直すことができます。詳細については、「［表の構成］タブ」（29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。詳細については、『JMPの使用法』を参照してください。

応答の等質性に対する検定

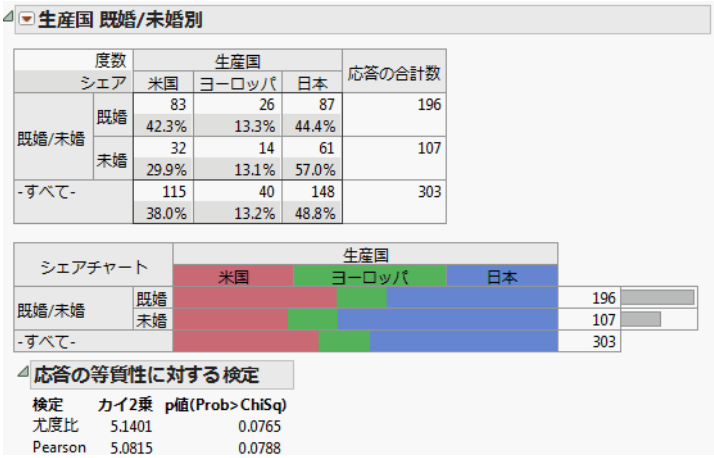
［応答の等質性に対する検定］は、2元表に対する標準的なカイ2乗検定を行います。この検定は、通常、カテゴリカルの応答変数（Y）が1つ、カテゴリカルのグループ変数（X）が1つある状況で使われている検定です。なお、グループ変数として複数の列が指定されている場合、「カテゴリカル」プラットフォームのデフォルトでは、それらの組み合わせでグループが構成されます。

カイ2乗検定により、応答確率のパターンが全グループにわたって同じかどうかが検定されます。この検定は、X変数間のYの違いを調べている検定ですが、同時に、2つのY変数の独立性に対する検定にもなっています。検定には、「Pearsonのカイ2乗検定」と「尤度比検定」の2つが用意されています。いずれの検定においても、カイ2乗分布で検定統計量を近似します。［検定オプション］メニュー（［カイ2乗検定の選択］）では、「Pearsonのカイ2乗検定」と「尤度比検定」の表示／非表示を切り替えることができます。［警告の表示］を設定すると、度数が少ないために、近似が悪くなっているだろうときには、レポートに警告が表示されます。

例：

1. 「Car Poll.jmp」データテーブルを開きます。
2. ［**分析**］＞［**消費者調査**］＞［**カテゴリカル**］を選択します。
3. **「生産国」**を選択して、［単純］タブの［**応答**］をクリックします。
4. **「既婚/未婚」**を選択して、［**X, グループ化カテゴリ**］をクリックします。
5. ［**OK**］をクリックします。
6. 「カテゴリカル」の赤い三角ボタンをクリックし、［**応答の等質性に対する検定**］を選択します。

図3.9 応答の等質性に対する検定



生産国 既婚/未婚別

度数 / シェア		生産国 米国	ヨーロッパ	日本	応答の合計数
既婚/未婚	既婚	83	26	87	196
		42.3%	13.3%	44.4%	
	未婚	32	14	61	107
		29.9%	13.1%	57.0%	
-すべて-		115	40	148	303
		38.0%	13.2%	48.8%	

応答の等質性に対する検定

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	5.1401	0.0765
Pearson	5.0815	0.0788

「シェアチャート」を見ると、既婚グループの方が米国製の車を買う確率が高く、未婚グループは日本車を買う確率が高いようですが、統計的検定の有意度は0.08に過ぎません。そのため、応答確率に見られる未婚と既婚の差は、α水準を0.05としたとき、統計的に有意ではありません。

多重応答の各応答に対する検定

多重応答の各応答に対する検定には、Poisson 分布を仮定した検定が行われます。応答のカテゴリごとに、1 つの検定統計量が計算されます。この検定統計量を、カイ 2 乗分布で近似して、p 値が計算されます。この検定の目的は、1 ケースあたりの比率（1 ケースあたりのカテゴリの度数）が、グループ間で同じかどうかを調べることです。度数が Poisson 分布に従うと仮定されています。この検定は、1 ケースあたりの度数を応答変数として、そして、グループ変数を説明変数とした Poisson 回帰モデルから計算されます。また、検定統計量は、尤度比検定に基づいて計算されています。

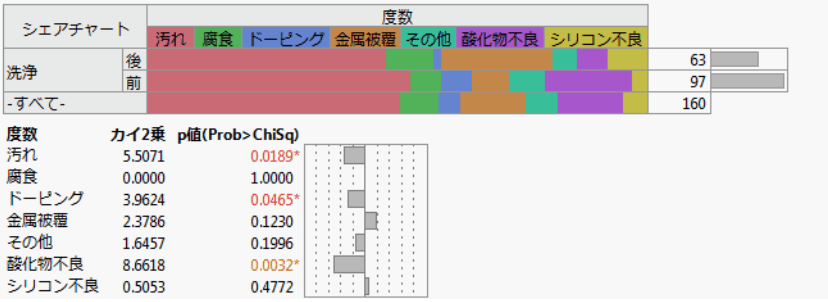
この検定は、「モデルのあてはめ」プラットフォームの［一般化線形モデル］手法や、「パレート図」プラットフォームでも実行できます。

例：

1. 「Quality Control」フォルダに入っている「Failure3Freq.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. 不適合の理由を表している列をすべて選択して、［多重］タブで［**応答の度数**］をクリックします。
4. 「**洗浄**」方法によって不適合の発生頻度がどのように異なっているかを見るため、「**洗浄**」を選択して［**X, グループ化カテゴリ**］をクリックします。
5. 「**標本サイズ**」を選択して、［**標本サイズ**］をクリックします。
6. ［**OK**］をクリックします。
7. 「カテゴリカル」の赤い三角ボタンをクリックし、［**各応答に対する検定**］を選択します。

図 3.10 各応答に対する検定

度数 シェア ケース		度数							応答の合計数	全ケース数
		汚れ	腐食	ドーピング	金属被覆	その他	酸化物不良	シリコン不良		
洗浄	後	30	6	1	14	3	4	5	63	150
		47.6%	9.5%	1.6%	22.2%	4.8%	6.3%	7.9%		
		20.0%	4.0%	0.7%	9.3%	2.0%	2.7%	3.3%		
	前	51	6	6	7	7	17	3	97	150
		52.6%	6.2%	6.2%	7.2%	7.2%	17.5%	3.1%		
		34.0%	4.0%	4.0%	4.7%	4.7%	11.3%	2.0%		
-すべて-		81	12	7	21	10	21	8	160	300
		50.6%	7.5%	4.4%	13.1%	6.3%	13.1%	5.0%		



度数	カイ2乗	p値(Prob>ChiSq)
汚れ	5.5071	0.0189*
腐食	0.0000	1.0000
ドーピング	3.9624	0.0465*
金属被覆	2.3786	0.1230
その他	1.6457	0.1996
酸化物不良	8.6618	0.0032*
シリコン不良	0.5053	0.4772

どの不適合の率が、**「洗浄」**処理間で有意に異なるでしょうか。*p* 値を見ると、**「酸化物不良」**が最も p 値が小さく、次に**「汚れ」**、**「ドーピング」**と続いています。データ量が少ないためかもしれませんが、その他の不適合は統計的に有意にはなっていません。

セルのカイ２乗

［**セルのカイ２乗**］は、セルごとのカイ２乗の p 値を表示します。このコマンドは、応答が多重応答ではなく、単一である場合に使えます。セルごとのカイ２乗値を合計すると、全体の Pearon のカイ２乗値になります。観測度数が期待度数よりも大きいセルは赤っぽい色で、小さいセルは青っぽい色でしめされます。表示されている値は、カイ２乗の p 値です。p 値が有意なら鮮やかな赤色や青色で、あまり有意でないときは灰色系の色で表示されるので、有意性が一目でわかります。

例：

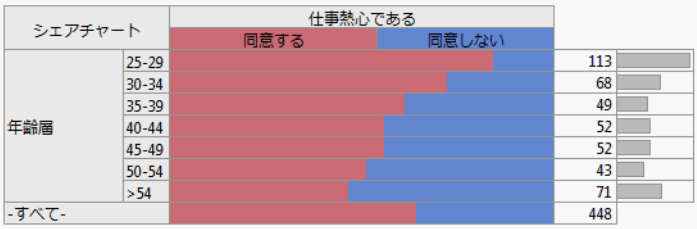
1. 「Consumer Preferences.jmp」データテーブルを開きます。
2. ［**分析**］＞［**消費者調査**］＞［**カテゴリカル**］を選択します。
3. **「仕事熱心である」**を選択して、［単純］タブの［**応答**］をクリックします。
4. **「年齢層」**を選択して［**X, グループ化カテゴリ**］をクリックします。
5. ［**OK**］をクリックします。
6. 赤い三角ボタンをクリックし、［**クロス表形式 転置**］を選択します。
7. 赤い三角ボタンをクリックし、［**セルのカイ２乗**］を選択します。

図 3.11 セルのカイ２乗

カテゴリカル

仕事熱心である 年齢層別

度数 / シェア / セルのカイ2乗p値		年齢層							-すべて-
		25-29	30-34	35-39	40-44	45-49	50-54	>54	
仕事熱心である	同意する	95	49	30	29	29	22	33	287
		84.1%	72.1%	61.2%	55.8%	55.8%	51.2%	46.5%	64.1%
		0.00788	0.41003	0.80398	0.45495	0.45495	0.29058	0.06415	
	同意しない	18	19	19	23	23	21	38	161
		15.9%	27.9%	38.8%	44.2%	44.2%	48.8%	53.5%	35.9%
		0.00039	0.27136	0.74035	0.31848	0.31848	0.15823	0.01345	
応答の合計数		113	68	49	52	52	43	71	448



相対リスク

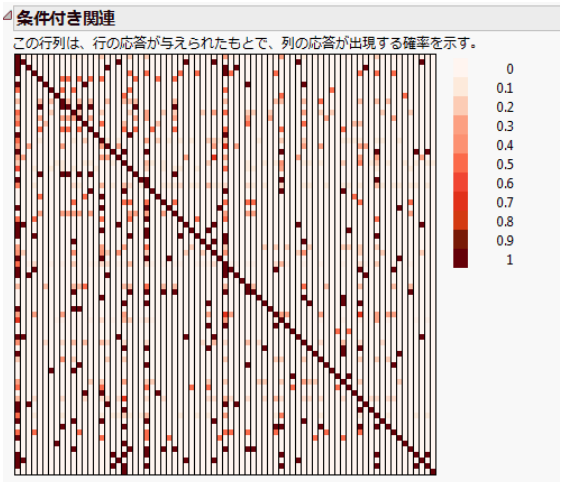
［相対リスク］オプションは、相対リスクを計算します。［X, グループ化カテゴリ］の各水準に対して、リスクが計算されます。相対リスクは、ある水準のリスクの、別の水準のリスクに対する比です。このオプションを使用するには、「カテゴリカル」の起動ウィンドウで**「ID 内で一意な値をカウント」**チェックボックスにマークをつける必要があります。

相対リスクは、一般に、応答が有害事象（副作用）の発生を表し、X 変数が処置（薬剤とプラシボなど）を表す場合などに使用します。各副作用が発生するリスクが、薬剤とプラシボの両方に対して計算されます。相対リスクは、2 つのリスクの比です。

条件付き関連

［**条件付き関連**］オプションは、応答に特定の値を指定したときの、別の値が発生する条件付き確率を計算します。条件付き確率の表とカラーマップが作成されます。このオプションを使用するには、「カテゴリカル」の起動ウィンドウで**「ID 内で一意な値をカウント」**チェックボックスにマークをつける必要があります。条件付き関連は、一般に、応答が有害事象（薬の副作用）の発生を表すときに使用します。ある副作用が発生したときの、別の副作用が発生する条件付き確率が計算されます。図 3.12 は、**「AdverseR.jmp」**の例に対して、（行に示されている）ある副作用が生じたという条件のもとで、（列に示されている）別の副作用が生じる、条件付き確率を示しています。なお、図 3.12 では、一部の出力を省略しています。

図 3.12 条件付き関連



条件付き関連

	ABDOMINAL PAIN	ABNORMAL VISION	ALOPECIA	AMBLYOPIA
ABDOMINAL PAIN	1.0000	0.0323	0.0323	0.0645
ABNORMAL VISION	1.0000	1.0000	0.0000	0.0000
ALOPECIA	1.0000	0.0000	1.0000	0.0000
AMBLYOPIA	1.0000	0.0000	0.0000	1.0000
ANEMIA	0.0000	0.0000	0.0000	0.0000
ANGINA PECTORIS	1.0000	0.0000	0.0000	0.0000
ANOREXIA	0.5000	0.0000	0.0000	0.0000
ANXIETY	0.0000	0.0000	0.0000	0.0000
ASTHENIA	0.2500	0.0000	0.0000	0.0000
ASTHMA	0.0000	0.0000	0.0000	0.0000
BACK PAIN	0.5000	0.0000	0.0000	0.0000
BRONCHITIS	0.3333	0.0000	0.0000	0.0000

判定の一致性

判定の一致性分析では、判定者の間にどの程度の一致性が見られるか、また、不一致性は対称か、などを調べることができます。例として「Attribute Gauge.jmp」を開いてみましょう。そこにある「計数値用チャート」スクリプトは、「計量値 / 計数値ゲージチャート」プラットフォームを実行し、判定者間の一致性を調べます。

図 3.13 一致性の比較

一致性の比較

判定者	比較対象の判定者	カッパ	.2 .4 .6 .8	標準誤差
A	B	0.8629		0.0442
A	C	0.7761		0.0547
B	C	0.7880		0.0537

「カテゴリカル」プラットフォームの起動ダイアログボックスで、3 人の判定者（A、B、C）を［**判定の一致性**］に指定します。結果のレポートには、一致性に対する指標であるκ係数とその標準誤差が表示されます。また、不一致のセルにおける度数が対称であることを検定する対称性の検定が表示されます。

図 3.14 一致性の統計量

一致性の統計量

一致性の度合い

応答1	応答2	カッパ	標準誤差	Bowkerの対称性	Bowkerのp値
A	B	0.862944	0.044198	1	0.3173
A	C	0.776119	0.054671	0.066667	0.7963
B	C	0.788007	0.053702	1.142857	0.2850

2x2表の場合、Bowker検定はMcNemar検定と等価です。

詳細

判定者の行	判定者の列	水準	0	1
B	A	0	44	3
B	A	1	6	97
C	A	0	43	8
C	A	1	7	92
C	B	0	42	9
C	B	1	5	94

反復測定

［反復測定］は、各列が、異なる時間に測定された応答のときに用います。このような反復測定データでは、時間ごとの変化に、分析者は興味があるでしょう。項目ごとに個別のレポートが表示され、最後に、遷移度数と遷移率行列を含む遷移レポートが表示されます。

例 :

1. 「Presidential Elections.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. 「**1980 年の勝者**」から「**2012 年の勝者**」までを選択して、［関連］タブで［**反復測定**］をクリックします。
4. 「**州**」を選択して［**X, グループ化カテゴリ**］をクリックします。
5. ［**OK**］をクリックします。

応答をスクロールして、この数年間の各州の投票結果を確認してみましょう。「New Mexico」では「Democratic」と「Republic」の間で投票結果が遷移していることがわかります。

図 3.15 反復測定

			Democrat	Republican		Democrat	Republican	1984年の勝者
New Mexico	1980年の勝者 1984年の勝者	Democrat	0	0	Democrat	.	0.0000	
		Republican	0	1	Republican	.	1.0000	
			Democrat	Republican		Democrat	Republican	1988年の勝者
New Mexico	1984年の勝者 1988年の勝者	Democrat	0	0	Democrat	.	0.0000	
		Republican	0	1	Republican	.	1.0000	
			Democrat	Republican		Democrat	Republican	1992年の勝者
New Mexico	1988年の勝者 1992年の勝者	Democrat	0	0	Democrat	0.0000	.	
		Republican	1	0	Republican	1.0000	.	
			Democrat	Republican		Democrat	Republican	1996年の勝者
New Mexico	1992年の勝者 1996年の勝者	Democrat	1	0	Democrat	1.0000	.	
		Republican	0	0	Republican	0.0000	.	
			Democrat	Republican		Democrat	Republican	2000年の勝者
New Mexico	1996年の勝者 2000年の勝者	Democrat	1	0	Democrat	1.0000	.	
		Republican	0	0	Republican	0.0000	.	
			Democrat	Republican		Democrat	Republican	2004年の勝者
New Mexico	2000年の勝者 2004年の勝者	Democrat	0	1	Democrat	.	1.0000	
		Republican	0	0	Republican	.	0.0000	
			Democrat	Republican		Democrat	Republican	2008年の勝者
New Mexico	2004年の勝者 2008年の勝者	Democrat	0	0	Democrat	0.0000	.	
		Republican	1	0	Republican	1.0000	.	
			Democrat	Republican		Democrat	Republican	2012年の勝者
New Mexico	2008年の勝者 2012年の勝者	Democrat	1	0	Democrat	1.0000	.	
		Republican	0	0	Republican	0.0000	.	

標本の各ペア比較

［標本の各ペア比較］は、標本グループの各ペアを比較します。各ペアで、応答確率全体のパターンが同じかどうかの検定を行います。このコマンドは、応答が多重応答ではなく、単一である場合のみに使えます。標本グループの各ペアに対して p 値の行列が作成されます。この行列は、図3.16のような対称行列です。

また、度数表に「比較」という名前の新しい行（または列）が追加されます。行の場合はテーブルの一番下に、列の場合は一番右に追加されます。（行か列かは、［クロス表形式］と［クロス表形式 転置］のどちらを指定したかによって異なります。）標本グループのどのペアが有意に異なるかがアルファベット文字で示されます。また、「標本の各ペア比較」という結果も下部に追加され、比較に関する詳細が表示されます。アルファベット文字の見方については、「比較の結果を表すアルファベット文字の見方」（48ページ）を参照してください。

図 3.16 標本の各ペア比較

標本の各ペア比較

ペアに対する尤度比カイ2乗のp値

	A	B	C	D	E	F	G
A	1.0000	0.0552	0.0020	0.0001	0.0001	<.0001	<.0001
B	0.0552	1.0000	0.2183	0.0641	0.0641	0.0261	0.0020
C	0.0020	0.2183	1.0000	0.5781	0.5781	0.3312	0.1108
D	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083
E	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083
F	<.0001	0.0261	0.3312	0.6540	0.6540	1.0000	0.6276
G	<.0001	0.0020	0.1108	0.3083	0.3083	0.6276	1.0000

ペアに対するPearsonカイ2乗のp値

	A	B	C	D	E	F	G
A	1.0000	0.0523	0.0015	<.0001	<.0001	<.0001	<.0001
B	0.0523	1.0000	0.2170	0.0638	0.0638	0.0255	0.0022
C	0.0015	0.2170	1.0000	0.5783	0.5783	0.3314	0.1119
D	<.0001	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
E	<.0001	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
F	<.0001	0.0255	0.3314	0.6540	0.6540	1.0000	0.6276
G	<.0001	0.0022	0.1119	0.3087	0.3087	0.6276	1.0000

セルの各ペア比較

［セルの各ペア比較］は、応答の 1 カテゴリごとに標本グループの各ペアを比較します。このコマンドは、単一応答と多重応答のどちらの場合でも使えます。「セルの各ペア比較」レポートには、*p* 値が表形式で表示されます。上部の文字は、X 変数の指定された水準の検定に使用された応答水準を示しています。図 3.17 はその一例です。

有意に異なるペアは、各セル内にアルファベット文字で表示されます。これらのアルファベット文字の見方については、「比較の結果を表すアルファベット文字の見方」（48 ページ）を参照してください。

図 3.17 セルの各ペア比較（「CG」以降の列は省略）

すべてのセルを比較 - 詳細

尤度比 ペア比較

	AA	AB	BB	AC	BC	CC	AD	BD	CD	DD	AE	BE	CE	DE	EE	AF	BF	CF
同意しない	1.0000	0.0552	1.0000	0.0020	0.2183	1.0000	0.0001	0.0641	0.5781	1.0000	0.0001	0.0641	0.5781	1.0000	1.0000	0.0000	0.0261	0.3312

Pearson ペア比較

	AA	AB	BB	AC	BC	CC	AD	BD	CD	DD	AE	BE	CE	DE	EE	AF	BF	CF
同意する	1.0000	0.0523	1.0000	0.0015	0.2170	1.0000	0.0001	0.0638	0.5783	1.0000	0.0001	0.0638	0.5783	1.0000	1.0000	0.0000	0.0255	0.3314
同意しない	1.0000	0.0523	1.0000	0.0015	0.2170	1.0000	0.0001	0.0638	0.5783	1.0000	0.0001	0.0638	0.5783	1.0000	1.0000	0.0000	0.0255	0.3314

Fisher正確 ペア比較

	AA	AB	BB	AC	BC	CC	AD	BD	CD	DD	AE	BE	CE	DE	EE	AF	BF	CF
同意する	1.0000	0.0591	1.0000	0.0022	0.2357	1.0000	0.0002	0.0825	0.6869	1.0000	0.0002	0.0825	0.6869	1.0000	1.0000	0.0001	0.0417	0.4009
同意しない	1.0000	0.0591	1.0000	0.0022	0.2357	1.0000	0.0002	0.0825	0.6869	1.0000	0.0002	0.0825	0.6869	1.0000	1.0000	0.0001	0.0417	0.4009

セルの比較

アルファベットの並びで後ろにあるカテゴリで、有意差があるものが表示される

* 星1つで警告する度数の基準	100	大文字で示す有意水準	0.05
** 星2つで警告する度数の基準	30	英小文字によって統計的有意性を示す際の有意水準を設定する。	0.1

比較の結果を表すアルファベット文字の見方

［セルの各ペア比較］、［標本の各ペア比較］、［スコア平均の比較］コマンドの結果では、各標本グループを A、B、C といったアルファベット文字で表しています。最初の標本を「A」、2 番目の標本を「B」、以下、同じようにアルファベット文字を与えています。27 個番目より後のグループは、「A1」、「B1」のように、アルファベット文字の後に数字が付きます。検定に関する前述の 3 つのコマンドを実行すると、比較の検定結果を表すのに、これらのアルファベット文字が使われます。

2 つの標本グループが有意に異なる場合、シェアが大きい方の標本グループのアルファベット文字が、もう一方の標本グループのセルに表示されます。つまり、たとえば、標本グループ「B」と有意に異なる標本グループを確かめるには、「B」のセルに他の文字がないかを確かめ、さらに他のすべての標本グループのセルに「B」がないかを見ます。

少しばかり有意でない個所（デフォルトでは p 値が 5 ～ 10% の時）には、大文字ではなく、小文字が使用されます（表 3.4 を参照）。なお、これらの比較は、標本グループの度数（ベースの度数）が小さいとうまくいきません。そのような場合、比較セルには警告を表すアスタリスクが表示されます。

比較機能は、［環境設定］またはスクリプトで設定した 4 つのオプションによって制御されます。詳細については、「環境設定の変更」（49 ページ）を参照してください。

表 3.4 比較の文字

大文字の有意水準	0.05	大文字が示す有意水準
小文字の有意水準	0.10	子文字が示す有意水準
ベースの度数 最小	≤ 29	星 2 つ（**）で警告を示す標本グループの度数
ベースの度数 警告	30 ～ 99	星 1 つ（*）で警告を示す標本グループの度数

検定オプション

「カテゴリカル」の赤い三角ボタンのメニューにある［検定オプション］をクリックすると、選択した内容に応じて、次のようなオプションが表示されます。

カイ 2 乗検定の選択　応答が多重応答（複数回答）ではなく、単一式である場合、尤度比検定や Pearson のカイ 2 乗検定によって等質性を調べます。オプションは、［**尤度比と Pearson**］、［**尤度比のみ**］、［**Pearson のみ**］です。これは［環境設定］で設定することもできます。

警告の表示　標本サイズが小さいときに警告を表示します。

p 値の小さい順に並べ替え　レポートを、統計的検定の有意確率が小さい順に並べ替えます。このオプションは、等質性検定が 1 つしかないレポートのみを対象とします。

有意でない結果を非表示にする　統計的に有意でないレポートを非表示にします。このオプションは、等質性検定が 1 つしかないレポートのみを対象とします。

テーブルの保存

「カテゴリカル」の赤い三角ボタンメニューにある［テーブルの保存］には、次のようなコマンドが用意されています。なお、応答の種類や状況に応じて、表示されないコマンドもあります。

度数の保存 「度数」レポートにおける度数表を、新しいデータテーブルに保存します。ただし、行や列の周辺度数は含まれません。

応答のシェアの保存 「応答のシェア」レポートを、新しいデータテーブルに保存します。ただし、行や列の周辺度数は含まれません。

ケースあたりの比率の保存 「ケースあたりの比率」レポートを、新しいデータテーブルに保存します。ただし、行や列の周辺度数は含まれません。

転置した度数の保存 「度数チャート 転置」レポートにおける度数表を、新しいデータテーブルに保存します。ただし、行や列の周辺度数は含まれません。

転置した応答のシェアの保存 「応答のシェア」を転置したものを、新しいデータテーブルに保存します。

転置したケースあたりの比率の保存 「ケースあたりの比率」を転置したものを、新しいデータテーブルに保存します。

比率検定の保存 ［各応答に対する検定］オプションの結果を、新しいデータテーブルに保存します。

等質性検定の保存 ［応答の等質性に対する検定］オプションの結果を、新しいデータテーブルに保存します。

Excel ファイルとして保存 Microsoft Excel スプレッドシートに、表を保存します。保存される表では、応答カテゴリを行として、標本グループを列として、すべての表を 1 つのシートにまとめています。複数の表にまたがる標本の見出しは共有されます。表の各セルに複数の要素がある場合は、それを Microsoft Excel ファイルに複数のセルとして表示するか、単一のセルとして表示するかを選択できます。

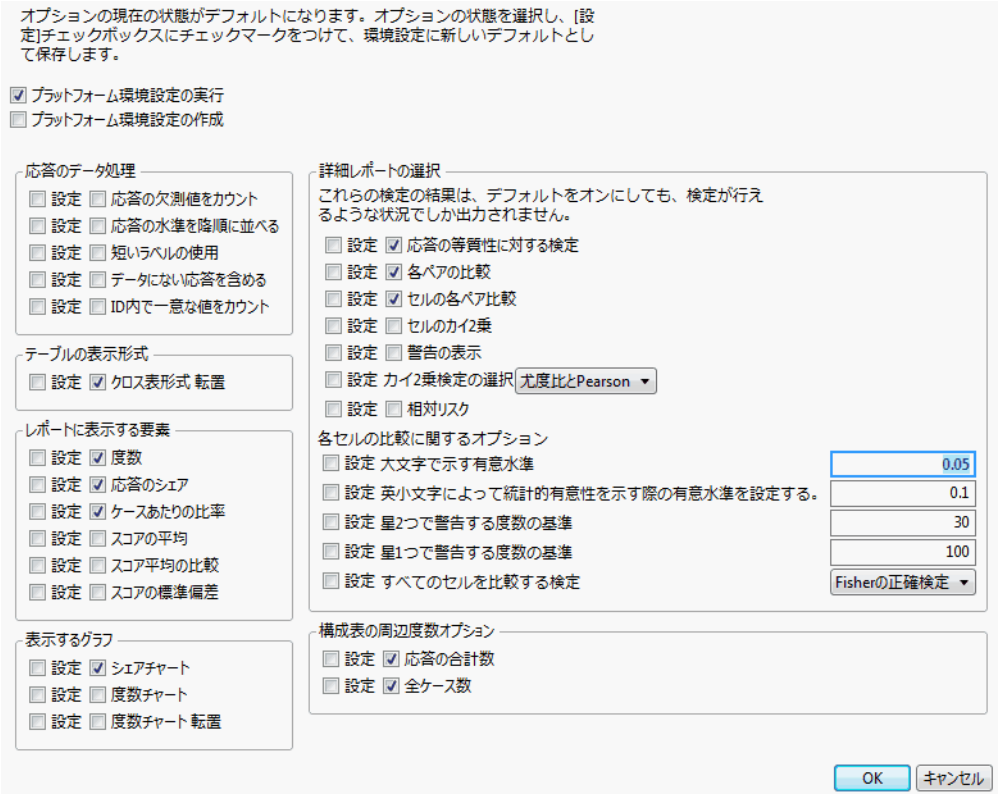
環境設定の変更

「カテゴリカル」プラットフォーム上で、環境設定を変更できます。環境設定のいくつかは起動ウィンドウにも表示されており、分析を行う前にも設定できます。また、「カテゴリカル」の赤い三角ボタンのメニューから選択できるオプションもあり、それらは選択すると、その設定でもう一度、分析が実行されます。

環境設定のダイアログは、現在の状態に初期化されています。変更したいオプションを選択し、［**プラットフォーム環境設定の実行**］や［**プラットフォーム環境設定の作成**］を選択すると、オプションが新しいデフォルトとして環境設定に保存されます。次回以降に、「カテゴリカル」プラットフォームを起動すると、保存した環境設定が有効になります。

複数の人々と環境設定を共有したい場合は、スクリプトを用いてください。アドインを作成して配布すれば、環境設定を共有しやすいです。環境設定の設定が初期状態にリセットされた場合でも、アドインがあれば設定を復元できます。

図 3.18 環境設定の設定ウィンドウ



自由回答のレポートオプション

コメント欄などの自由回答に、ここで説明する機能は使えます。ただし、この機能は日本語には対応していません。単語と単語がスペースで区切られている英語のような言語にしか使えません。分析を実行すると、各単語の使用頻度が数えられます。自由回答では、単語順と度数順で単語が表示されます。また、無記入でないテキストの割合も表示されます。次の操作例では、「Consumer Preferences.jmp」データテーブルを使用します。このデータは口腔衛生に関するアンケート調査で、回答者がデンタルフロスを使わない理由をコメント欄に記入しています。

1. サンプルデータのフォルダにある「Consumer Preferences.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. **「デンタルフロスを使わない理由」**を選択し、［多重］タブの［**自由回答**］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［**単語ごとに指定列の平均を求める**］を選択し、**「デンタルフロス」**を選択します。［**OK**］をクリックします。

図 3.19 自由回答のレポートの例

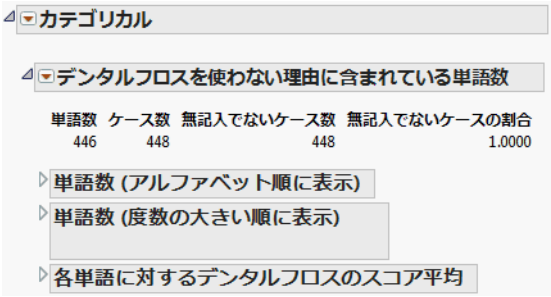


図 3.19 は、デンタルフロスを使わない理由として記入したコメントの分析結果です。「単語数」、「ケース数」、「無記入でないケース数」、「無記入でないケースの割合」が表示されています。この他にも、アルファベット順、度数の大きい順、スコア平均順に単語（とその度数）を表示することもできます。自由回答レポートの赤い三角ボタンのメニューから、次のような追加の分析を行うこともできます。

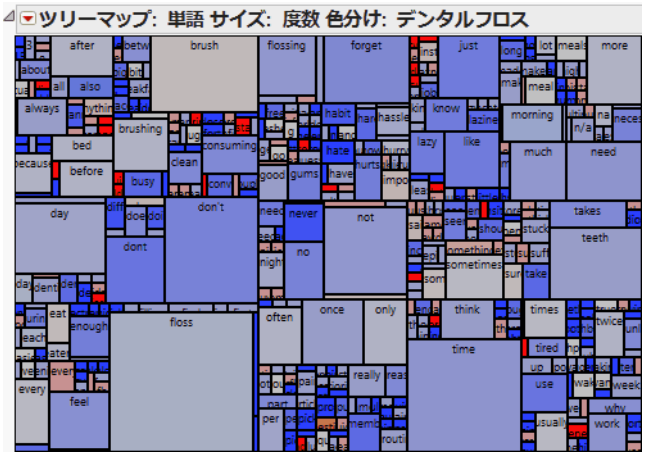
単語ごとに指定列の平均を求める　指定の列の平均値を、単語ごとに計算し、それを平均値で並べ替えます。後述するコマンドによって単語データとして保存すると、ツリーマップのスクリプト（図 3.20）が付随したデータテーブルが作成されます。

高頻度の単語に対する指示変数を保存　元のデータテーブルに、指示変数の列が追加されます。このコマンドを選択すると、まず、いくつの単語に対して指示変数を作成するかを指定するダイアログが呼び出されます。頻度が高い単語に関して、その単語が出現したかどうかを示す指示変数の列が追加されます。

単語データテーブルの保存　単語のリストを、度数やスコアと共に新しいデータテーブルに保存します。

削除　レポートから表を削除します。

図 3.20 ツリーマップの例



表の構成レポートのオプション

［表の構成］タブでは、複雑な表を構成できます。列名を緑のドロップゾーンにドラッグすることにより、並列や枝分かれといった構成を指定できます。次の例では「Consumer Preferences.jmp」データテーブルを使って、仕事に対する満足度と給与を、性別×年齢層、および、被雇用期間で比較したいとします。

1. 「Consumer Preferences.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. ［表の構成］タブを選択します。
4. **「性別」**を表の**「上側」**にある緑のドロップゾーンにドラッグします。
5. **「年齢層」**を**「性別」**の下にある緑のドロップゾーンにドラッグします。
6. **「被雇用期間」**を**「上側」**の**「性別」**の隣にある緑のドロップゾーンにドラッグします。
7. **「仕事に対する満足度」**を**「横側」**にある緑のドロップゾーンにドラッグします。
8. **「給与グループ」**を**「横側」**の**「仕事に対する満足度」**の下にある緑のドロップゾーンにドラッグします。

図 3.21 ［表の構成］タブのレポートの設定

9. ［**追加 =>**］をクリックします。
10. ［**OK**］をクリックします。

図 3.22［表の構成］タブのレポートの例

Job Satisfaction + Salary Group By Gender*Age Group + Position Tenure

	度数	性別																
	シェア	男性							女性									
		年齢層							年齢層							現職位での勤続年数グループ			
		25-29	30-34	35-39	40-44	45-49	50-54	>54	25-29	30-34	35-39	40-44	45-49	50-54	>54	0 - 5年	5 - 10年	10 - 20年	20年 -
仕事に対する満足度	非常に不満足	3	3	1	3	3	2	2	3	1	0	5	2	1	3	15	11	6	0
		5.5%	10.3%	3.3%	8.8%	7.9%	8.3%	4.1%	5.2%	2.6%	0.0%	27.8%	14.3%	5.3%	13.6%	7.2%	8.2%	6.8%	0.0%
	やや満足	32	14	15	19	19	10	26	34	30	12	6	5	11	9	122	70	40	10
		58.2%	48.3%	50.0%	55.9%	50.0%	41.7%	53.1%	58.6%	76.9%	63.2%	33.3%	35.7%	57.9%	40.9%	58.7%	52.2%	45.5%	55.6%
	非常に満足	20	12	14	12	16	12	21	21	8	7	7	7	7	10	71	53	42	8
		36.4%	41.4%	46.7%	35.3%	42.1%	50.0%	42.9%	36.2%	20.5%	36.8%	38.9%	50.0%	36.8%	45.5%	34.1%	39.6%	47.7%	44.4%
	応答の合計数	55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18
給与グループ	'0 - 40000	20	6	4	8	6	5	10	35	19	6	6	2	7	4	80	40	15	3
		36.4%	20.7%	13.3%	23.5%	15.8%	20.8%	20.4%	60.3%	48.7%	31.6%	33.3%	14.3%	36.8%	18.2%	38.5%	29.9%	17.0%	16.7%
	40000 - 60000	16	14	11	12	14	7	13	15	10	8	6	7	6	12	67	38	37	9
		29.1%	48.3%	36.7%	35.3%	36.8%	29.2%	26.5%	25.9%	25.6%	42.1%	33.3%	50.0%	31.6%	54.5%	32.2%	28.4%	42.0%	50.0%
	60000 - 80000	8	5	7	7	8	8	10	5	5	3	3	3	4	5	32	25	19	5
		14.5%	17.2%	23.3%	20.6%	21.1%	33.3%	20.4%	8.6%	12.8%	15.8%	16.7%	21.4%	21.1%	22.7%	15.4%	18.7%	21.6%	27.8%
	80000 - 120000	9	3	6	4	7	3	7	1	2	1	2	1	1	0	19	17	10	1
		16.4%	10.3%	20.0%	11.8%	18.4%	12.5%	14.3%	1.7%	5.1%	5.3%	11.1%	7.1%	5.3%	0.0%	9.1%	12.7%	11.4%	5.6%
	120000 -	2	1	2	3	3	1	9	2	3	1	1	1	1	1	10	14	7	0
		3.6%	3.4%	6.7%	8.8%	7.9%	4.2%	18.4%	3.4%	7.7%	5.3%	5.6%	7.1%	5.3%	4.5%	4.8%	10.4%	8.0%	0.0%
	応答の合計数	55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18

図 3.22 を見ると、男女とも、「やや満足」の割合が多いことが分かります。男性でその割合が最も大きいのは 25 ～ 29 歳の層で、女性は 30 ～ 34 歳です。現職位での勤続年数が 5 年未満において、仕事に「やや満足」している人が多いです。

レポートの赤い三角ボタンのメニューには、次のようなオプションがあります。

文字の表示　列名を識別するためのアルファベット文字を表示します。このアルファベット文字は、以下で説明する比較のコマンドを使用した場合は、自動的に表示されます。

比較の設定　デフォルトでは、最も内側にある組み合わせを標本グループとして比較が行われます。比較したいグループがそうでない場合に、比較する標本グループを変更できます。このオプションを使用するには、まずアルファベット文字を確認する必要があります。指定方法は、スラッシュ（/）やカンマ（,）で区切って指定します。たとえば、デフォルトでは「A/B/C, E/D/F」とグループ化されている設定を変更して、A を E と、B を D と、C を F と比較するとします。その場合は、「A/E, B/D, C/F」と指定します。これにより、新しい比較が行われます。なお、要約レポートには、グループ全体での検定結果が表示されます。

削除　レポートから表を削除します。

「カテゴリカル」プラットフォームのその他の例

この節では、異なる 2 つの条件下で、3 日間にわたり、生産ラインを検査した例を紹介します。各検査につき、50 個の製品を調べました。検査では、次の不適合が確認されました。

- **汚れ**
- **腐食**
- **ドーピング**
- **金属被覆**
- **その他**

- **酸化物不良**
- **シリコン不良**

1つのユニットに複数の種類の不適合が見られるケースや、同じ種類の不適合が複数見られるケースもあり得ます。ここでは、「カテゴリカル」プラットフォームを使い、いろいろな形でデータを表示してみましょう。

多重応答

各ユニットの不適合がWebページを通じて入力されたとしましょう。1つのユニットに3つを超える不適合が生じることはまずないので、入力フォームには不適合の種類を入力するフィールドが3つ用意されています（「Failure3MultipleField.jmp」）。

1. 「Quality Control」フォルダに入っている「Failure3MultipleField.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. 「**不適合1**」、「**不適合2**」、「**不適合3**」を選択し、［多重］タブの［**多重応答**］をクリックします。

   これらの列は、欠陥の種類を示す列、つまり分析の対象です。
4. 「**洗浄**」と「**日付**」を選択して［**X, グループ化カテゴリ**］をクリックします。
5. ［**OK**］をクリックします。

図3.23は、この操作によって作成された、不適合の種類とその度数のレポートが示されています。

図3.23 Failure3MultipleField.jmpの不適合（多重応答）

多重応答 洗浄, 日付別

多重応答

			度数 シェア ケース	汚れ	腐食	ドーピング	金属被覆	その他	酸化物不良	シリコン不良	応答の合計数	全ケース数
洗浄	後	日付	10月1日	12	2	0	4	2	1	2	23	50
				52.2%	8.7%	0.0%	17.4%	8.7%	4.3%	8.7%		
				24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%		
			10月2日	10	1	1	5	1	2	3	23	50
				43.5%	4.3%	4.3%	21.7%	4.3%	8.7%	13.0%		
				20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%		
			10月3日	8	3	0	5	0	1	0	17	50
				47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%		
				16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%		
	前	日付	10月1日	14	2	1	2	3	8	1	31	50
				45.2%	6.5%	3.2%	6.5%	9.7%	25.8%	3.2%		
				28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%		
			10月2日	15	2	2	1	4	6	0	30	50
				50.0%	6.7%	6.7%	3.3%	13.3%	20.0%	0.0%		
				30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%		
			10月3日	22	2	3	4	0	3	2	36	50
				61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%		
				44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%		
-すべて-				81	12	7	21	10	21	8	160	300
				50.6%	7.5%	4.4%	13.1%	6.3%	13.1%	5.0%		

応答の度数

データに、各バッチの不適合数を含んだ列と、バッチに含まれるユニットの総数を示した列があるとしましょう（「「Failure3Freq.jmp」）。

図 3.24「Failure3Freq.jmp」データテーブル

	洗浄	日付	汚れ	腐食	ドーピング	金属被覆	その他	酸化物不良	シリコン不良	標本サイズ
1	後	10月1日	12	2	0	4	2	1	2	50
2	後	10月2日	10	1	1	5	1	2	3	50
3	後	10月3日	8	3	0	5	0	1	0	50
4	前	10月1日	14	2	1	2	3	8	1	50
5	前	10月2日	15	2	2	1	4	6	0	50
6	前	10月3日	22	2	3	4	0	3	2	50

1. 「Quality Control」フォルダに入っている「Failure3Freq.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. 度数変数（「**汚れ**」、「**腐食**」、「**ドーピング**」、「**金属被覆**」、「**その他**」、「**酸化物不良**」、「**シリコン不良**」）を選択し、［多重］タブの［**応答の度数**］をクリックします。
4. 「**洗浄**」と「**日付**」を選択して［**X, グループ化カテゴリ**］をクリックします。
5. 「**標本サイズ**」を選択して、［**標本サイズ**］をクリックします。
6. ［**OK**］をクリックします。

図 3.25 のように、バッチごとに個別の列がある度数表が表示されます。最後の 2 列には、不適合の総数（「**応答の合計数**」）とケース数（「**全ケース数**」）が表示されます。

図 3.25 不適合数の表

度数 洗浄, 日付別

度数列

			度数 シェア ケース	度数							応答の合計数	全ケース数
				汚れ	腐食	ドーピング	金属被覆	その他	酸化物不良	シリコン不良		
洗浄	後	日付	10月1日	12	2	0	4	2	1	2	23	50
				52.2%	8.7%	0.0%	17.4%	8.7%	4.3%	8.7%		
				24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%		
			10月2日	10	1	1	5	1	2	3	23	50
				43.5%	4.3%	4.3%	21.7%	4.3%	8.7%	13.0%		
				20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%		
			10月3日	8	3	0	5	0	1	0	17	50
				47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%		
				16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%		
	前	日付	10月1日	14	2	1	2	3	8	1	31	50
				45.2%	6.5%	3.2%	6.5%	9.7%	25.8%	3.2%		
				28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%		
			10月2日	15	2	2	1	4	6	0	30	50
				50.0%	6.7%	6.7%	3.3%	13.3%	20.0%	0.0%		
				30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%		
			10月3日	22	2	3	4	0	3	2	36	50
				61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%		
				44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%		
-すべて-				81	12	7	21	10	21	8	160	300
				50.6%	7.5%	4.4%	13.1%	6.3%	13.1%	5.0%		

各度数グループには、次の情報が含まれます。

- 種類ごとの不適合の合計数。たとえば、10 月 1 日の洗浄後、汚れによる不適合が 12 件生じています。
- 応答のシェア。たとえば、10 月 1 日の洗浄後に生じた汚れによる不適合（12 件）は、すべての不適合の 52.2 パーセントを占めています（12/23）。
- ケースあたりの比率。たとえば、10 月 1 日の洗浄後では、50 ユニットに 12 件の汚れが生じているので、ユニットあたりの比率は 24 パーセントです。

指示変数

データが集計されておらず、ユニットごとに記録されている場合があります。そのような場合について、「Failures3Indicators.jmp」を例に見ていきましょう。

図 3.26「Failure3Indicators.jmp」データテーブル

	洗浄	日付	ID	IDラベル	汚れ	腐食	ドーピング	金属被覆	その他	酸化物不良	シリコン不良
1	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
2	前	10月1日	1	10月1日 前	0	0	0	0	0	1	0
3	前	10月1日	1	10月1日 前	1	0	0	0	0	1	0
4	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
5	前	10月1日	1	10月1日 前	1	0	0	0	0	0	0
6	前	10月1日	1	10月1日 前	0	0	0	0	0	1	0
7	前	10月1日	1	10月1日 前	1	0	0	0	0	0	0
8	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
9	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
10	前	10月1日	1	10月1日 前	1	0	0	1	0	0	0
11	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
12	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
13	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
14	前	10月1日	1	10月1日 前	1	0	0	0	0	0	0
15	前	10月1日	1	10月1日 前	0	0	0	0	0	0	0
16	前	10月1日	1	10月1日 前	0	0	0	0	1	0	0

1. 「Quality Control」フォルダに入っている「Failures3Indicators.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. 不適合の列（「**汚れ**」、「**腐食**」、「**ドーピング**」、「**金属被覆**」、「**その他**」、「**酸化物不良**」、「**シリコン不良**」）を選択し、［多重］タブの［**指示変数**］をクリックします。
4. 「**洗浄**」と「**日付**」を選択して［**X, グループ化カテゴリ**］をクリックします。
5. ［**OK**］をクリックします。

［**OK**］をクリックすると、応答グループの例（図 3.25）と同じ結果が出ます。

多重応答 区切り文字

各ユニットごとに観測した不適合が、1 つの列にカンマで区切って列挙されているとします（「Failures3Delimited.jmp」）。次図のデータテーブルを見ると、ユニットの一部は、不適合がまったく観測されず、「**不適合**」列が空白になっています。

図 3.27「Failure3Delimited.jmp」データテーブル

	不適合	洗浄	日付	ID	IDラベル
1		前	10月1日	1	10月1日 前
2	酸化物不良	前	10月1日	1	10月1日 前
3	汚れ,酸化物不良	前	10月1日	1	10月1日 前
4		前	10月1日	1	10月1日 前
5	汚れ	前	10月1日	1	10月1日 前
6	酸化物不良	前	10月1日	1	10月1日 前
7	汚れ	前	10月1日	1	10月1日 前
8		前	10月1日	1	10月1日 前
9		前	10月1日	1	10月1日 前
10	金属被覆,汚れ	前	10月1日	1	10月1日 前
11		前	10月1日	1	10月1日 前
12		前	10月1日	1	10月1日 前
13		前	10月1日	1	10月1日 前
14	汚れ	前	10月1日	1	10月1日 前

1. 「Quality Control」フォルダに入っている「Failures3Delimited.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. 「**不適合**」を選択し、［多重］タブの［**多重応答 区切り文字**］をクリックします。
4. 「**洗浄**」と「**日付**」を選択して［**X, グループ化カテゴリ**］をクリックします。
5. ［**ID**］を選択し、［**ID**］をクリックします。
6. ［**OK**］をクリックします。

［**OK**］をクリックすると、図 3.25 と同じ結果が出ます。

注：区切り文字を使った列を複数指定した場合、列ごとに分析が行われます。

多重応答 ID 別

不適合の種類ごとに個別に記録を取り、どのユニットに属するものかを識別できるよう ID 列を追加したとしましょう（「Failure3ID.jmp」）。

図 3.28「Failure3ID.jmp」データテーブル

	不適合	度数	洗浄	日付	標本サイズ	ID
1	汚れ	14	前	10月1日	50	10月1日 前
2	腐食	2	前	10月1日	50	10月1日 前
3	ドーピング	1	前	10月1日	50	10月1日 前
4	金属被覆	2	前	10月1日	50	10月1日 前
5	その他	3	前	10月1日	50	10月1日 前
6	酸化物不良	8	前	10月1日	50	10月1日 前
7	シリコン不良	1	前	10月1日	50	10月1日 前
8	ドーピング	0	後	10月1日	50	10月1日 後
9	腐食	2	後	10月1日	50	10月1日 後
10	金属被覆	4	後	10月1日	50	10月1日 後

1. 「Quality Control」フォルダに入っている「Failure3ID.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
3. **「不適合」**を選択し、［多重］タブの［**多重応答 ID 別**］をクリックします。
4. **「洗浄」**と**「日付」**を選択して［**X, グループ化カテゴリ**］をクリックします。
5. **「標本サイズ」**を選択して、［**標本サイズ**］をクリックします。
6. **「度数」**を選択し、［**度数**］をクリックします。
7. ［**ID**］を選択し、［**ID**］をクリックします。
8. ［**OK**］をクリックします。

［**OK**］をクリックすると、図3.25と同じ結果が出ます。

スコアの平均の例

応答の各カテゴリにスコアを割り当てて、その平均を計算するには、データテーブルの応答列に、「値スコア」列プロパティを設定してください。スコアの平均を解釈可能にするには、**「列情報」**ウィンドウで**「値スコア」**プロパティにおいて、適切な値スコアを割り当ててください。列プロパティについての詳細は、『JMPの使用法』を参照してください。

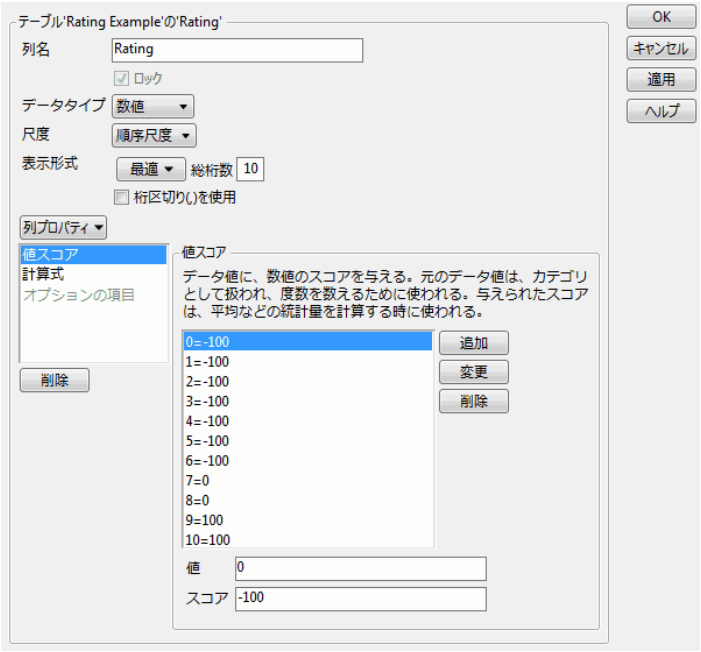
この例では、「値スコア」列プロパティを用いて、11段階の評価データに、-100から100までのNet Promoter Score（Reichheld, Harvard Business Review, 2003）を割り当てます。この例では、6以下のデータ値は、避難者（detractor）とみなし、スコアとして-100を割り当てます。

1. 次のスクリプトを実行します。

```
New Table("評価の例",
  Add Rows(300),
  New Script("Categorical",Categorical(Responses(:Rating),Mean Score(1))),
  New Column("Rating",Numeric,Ordinal,
    Set Property(
      "Value Scores",
{0=-100,1=-100,2=-100,3=-100,4=-100,5=-100,6=-100,7=0,8=0,9=100,10=100}),
    Formula(Random Category(
        0.05,0,0.05,1,0.05,2,0.05,3,0.05,4,
        0.05,5,0.05,6,0.05,7,0.05,8,0.3,9,0.25,10)),
    Set Selected
 )
);
```

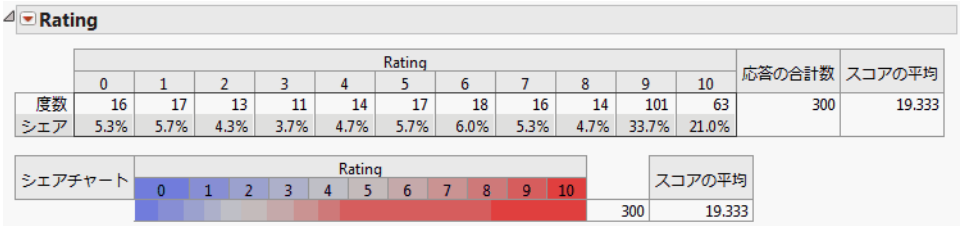
2. 疑似乱数によって生成された、300行からなるデータテーブルが作成されます。なお、「Rating」列には、「値スコア」列プロパティが定義されます。スコアを見るには、「Rating」列を右クリックし、［**列プロパティ**］>［**値スコア**］を選択します。

図 3.29 列プロパティ - 値スコア



3. ［**分析**］>［**消費者調査**］>［**カテゴリカル**］を選択します。
4. 「**Rating**」を選択して、［単純］タブの［**応答**］をクリックします。
5. ［**OK**］をクリックします。
6. 「カテゴリカル」の赤い三角ボタンをクリックし、［**スコアの平均**］を選択します。

図 3.30「評価の例」のレポート



Rating

	Rating											応答の合計数	スコアの平均
	0	1	2	3	4	5	6	7	8	9	10		
度数	16	17	13	11	14	17	18	16	14	101	63	300	19.333
シェア	5.3%	5.7%	4.3%	3.7%	4.7%	5.7%	6.0%	5.3%	4.7%	33.7%	21.0%		

シェアチャート	Rating												スコアの平均
	0	1	2	3	4	5	6	7	8	9	10	300	19.333

定義された値スコアをもとに、スコアの平均が 19.333 と計算されています。なお、「Rating」列の値を疑似乱数で生成していますので、結果は実行するたびに変わり、ここで示した結果とも異なります。

第 4 章

因子分析

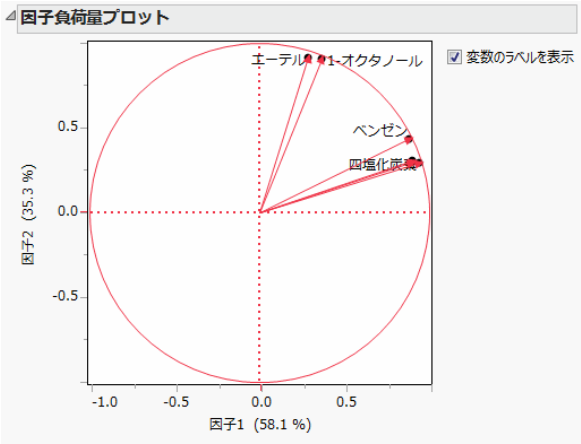
潜在的な因子によるモデル化

因子分析は、多数の観測変数を、少数の潜在変数で説明しようとする分析です。JMP でサポートされているのは、「探索的因子分析」と呼ばれている手法です。少数の潜在変数は、「因子」と呼ばれています。因子分析では、少数の因子によって、観測変数に**共通する**バラツキを説明しようとします。因子分析は、何らかの意味があるように因子を解釈し、多変量データを、それら少数の因子で説明しようとします。

因子分析は、さまざまな分野で利用されていますが、特に、心理学・社会学・教育学で使われています。因子分析では、観測された種々の行動を、潜在的な心理構造で解釈するために使用されています。たとえば、アウトドア・趣味・運動・旅行といった活動をどれぐらい行っているかという観測結果は、各個人の「活動性」という因子で説明できるかもしれません。因子分析は、観測変数に見られる相関を、少数の因子で説明しようとします。また、各観測変数の分散のうちどれぐらいが、それぞれの因子によって説明されているかも確認できます。子によって説明されるかも特定できます。

多変量データのパターンや構造を、少数の因子によって解釈や要約したい場合に、因子分析を利用してください。

図 4.1 回転後の因子負荷量



目次

「因子分析」プラットフォームの概要 63
「因子分析」プラットフォームの例 63
「因子分析」プラットフォームの起動 66
「因子分析」レポート 66
モデルの設定 67
回転方法 68
「因子分析」プラットフォームのオプション 70
因子分析の「モデルの設定」のオプション 71

「因子分析」プラットフォームの概要

因子分析は、観測された多変量データを、観測されない少数の潜在変数でモデル化します。これらの潜在変数は、「因子」と呼ばれています。因子分析では、因子によって、観測変数に見られる相関や共分散をモデル化します。因子分析では、まず因子を抽出します。そして、その後、意味のある解釈を得るために、因子を回転させます。

X_1、X_2、…、X_{10}という10個の観測変数があるとしましょう。この10個の変数を、F_1とF_2の2つの潜在変数で説明したいとします。ここでは説明を簡単にするため、因子は互いに相関がないものとします。なお、モデルを識別するため、通常、各因子は平均が0で、分散が1と仮定されます。求めたいモデルは、次のような式です。

$$X_i = \beta_{i0} + \beta_{i1}F_1 + \beta_{i2}F_2 + \varepsilon_i$$

これは、$Var(X_i) = \beta_{i1}{}^2 + \beta_{i2}{}^2 + Var(\varepsilon_i)$ に従います。X_iの分散のうち、因子に起因する部分、つまり**共通性**は、$\beta_{i1}{}^2 + \beta_{i2}{}^2$です。残りの分散$Var(\varepsilon_i)$は、$X_i$に特有のもので、独自性（特殊性）といいます。

「因子分析」プラットフォームでは、相関行列または共分散行列の固有値のスクリープロットが作成されます。このプロットは、抽出する因子の数を特定するための指針となります。プラットフォームのデフォルトでは、相関行列の固有値を計算して、1を超えている固有値の個数を、因子の個数とします。

手法のオプションとしては、「主成分分析」（主成分分析、もしくは、反復なしの主因子法）と「最尤法」の2つがあります。

また、共通性を推定するための、つまり、相関行列（または、共分散分析）の対角要素を設定するためのオプションが2つあります。ここで［主成分分析］を選択したときには、相関行列の対角要素に1が設定されます。一方、［共通因子分析］を選択したときは、相関行列の対角要素の値が、SMCに設定されます。

探索的因子分析では、因子を抽出した後、解釈しやすくするために、通常、分析者は、因子負荷量の行列を回転します。「因子分析」プラットフォームには、直交回転や斜交回転の様々な回転方法が用意されています。

主成分分析は、潜在変数の共通性に注目する因子分析とは異なり、観測変数から合成変数を作成することに主眼を置きます。『多変量分析』の「主成分分析」の章を参照してください。

「因子分析」プラットフォームの例

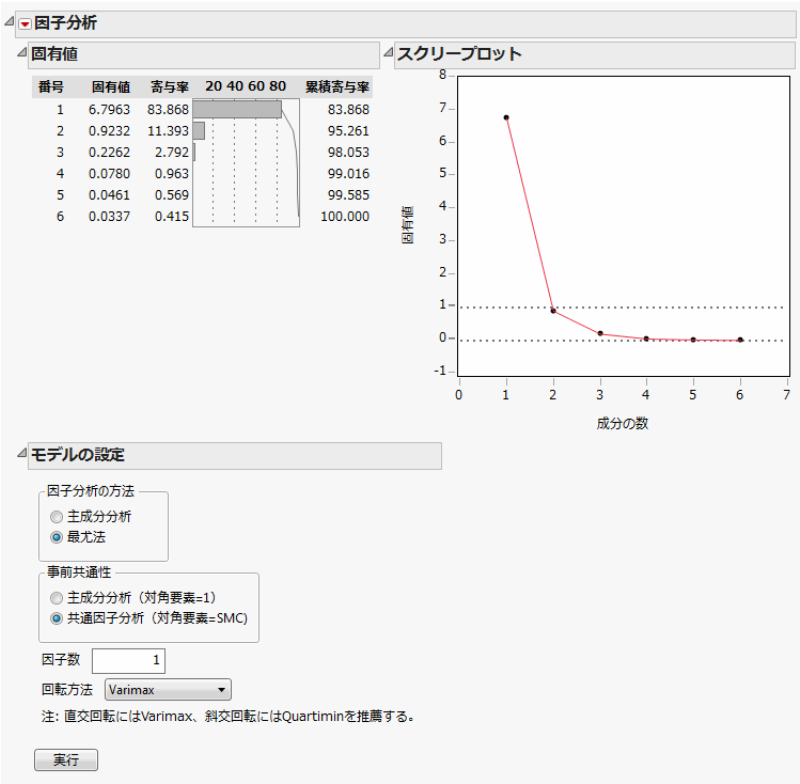
2因子のデータテーブルを使い、因子分析レポートの例を見ていきましょう。

1. 「Solubility.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**因子分析**］を選択します。

   「因子分析」起動ウィンドウが開きます。
3. 連続尺度の列をすべて選択し、［**Y, 列**］をクリックします。
4. **「推定法」**と**「分析対象」**をデフォルトのままにします。
5. ［**OK**］をクリックします。

最初の「因子分析」レポートが表示されます。

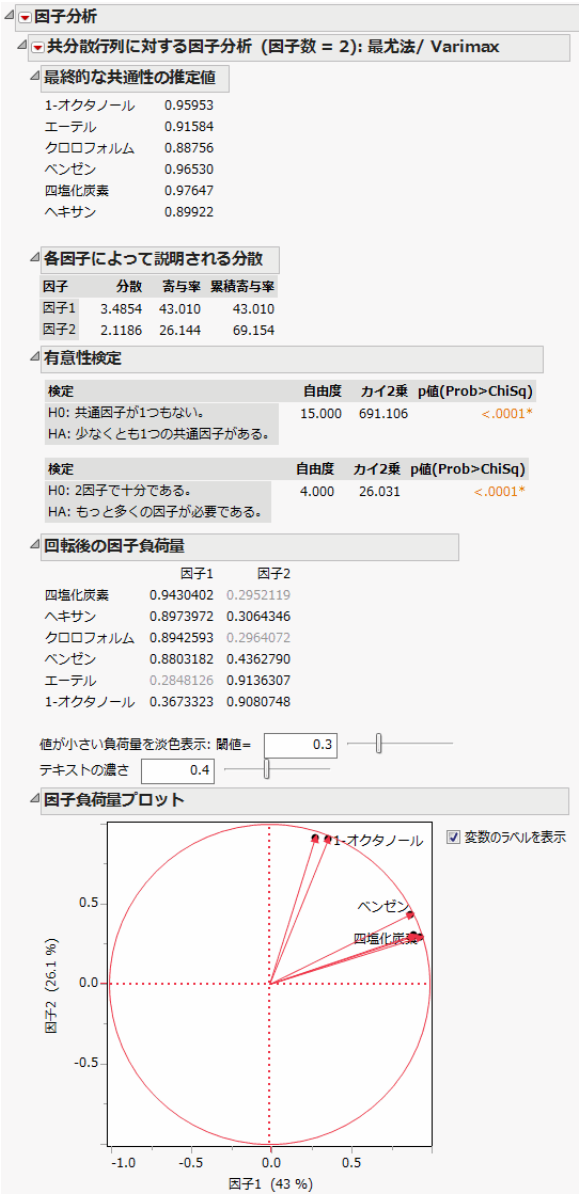
図 4.2 最初の「因子分析」レポート



6. 「モデルの設定」で、次のように設定します。
   - 「因子分析の方法」で［**最尤法**］を選択
   - 「事前共通性」で［**共通因子分析**］を選択
   - 因子数 = 2
   - 「回転方法」で［**Varimax**］を選択
7. 設定が終わったら、［**実行**］をクリックします。

   「因子分析」レポートが表示されます。

図 4.3「因子分析」レポートの例

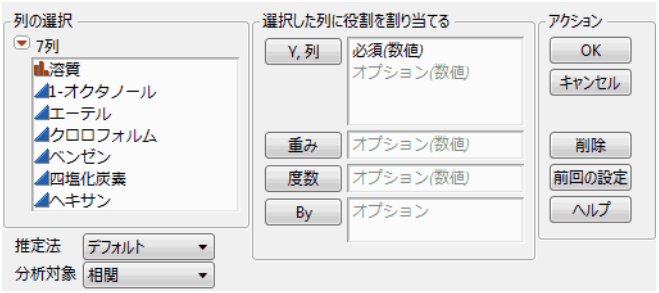


レポートには、共通性の推定値、分散、有意性検定、回転後の因子負荷量、因子負荷量プロットが表示されます。因子負荷量プロットでは、因子 1 が「**四塩化炭素 - クロロフォルム - ベンゼン - ヘキサン**」の変数クラスターに関連し、因子 2 が「**エーテル -1- オクタノール**」の変数クラスターに関連しています。レポートに表示される情報の詳細については、「因子分析の「モデルの設定」のオプション」(71 ページ) を参照してください。

「因子分析」プラットフォームの起動

「因子分析」プラットフォームを起動するには、［**分析**］>［**消費者調査**］>［**因子分析**］を選択します。この例では、「Solubility.jmp」データテーブルを使用します。

図 4.4「因子分析」起動ウィンドウ



Y, 列　分析対象となる連続尺度の列がリストされます。

重み　事前に要約されたデータを説明できるよう、分析に重みを付けます。

度数　この役割を割り当てた列の数値が、分析において各行の度数として使用されます。

By　By 列に指定した各値の因子分析レポートが作成され、グループごとに個別に分析できます。

推定法　モデルをあてはめるための各種手法がリストされます。手法の詳細については、『多変量分析』の「多変量分析」の章を参照してください。

分析対象　主成分分析と同じように、因子分析の分析対象とする行列を、［**相関**］、［**共分散**］、［**原点周りの積和**］の中から選択します。

「因子分析」レポート

最初の「因子分析」レポートには、「固有値」と「スクリープロット」が表示されます。「固有値」は、主成分分析から計算されます。「スクリープロット」は、固有値をグラフにしたものです。「モデルの設定」では、1.0 を超える固有値と同じ数が因子数として提案されます。

「固有値が 1 を超える」という基準ではなく、スクリープロットの減り具合で、最初の因子数を決めてもよいでしょう。その場合、スクリープロットが横ばいになる直前までの次元が、因子数としてよいでしょう。

図 4.5「因子分析」レポート

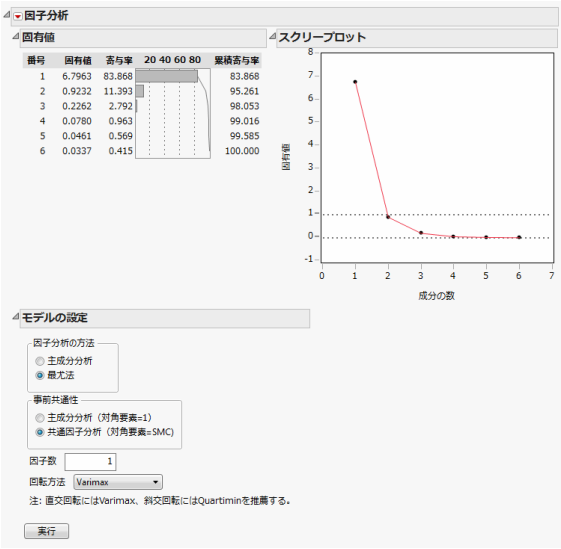
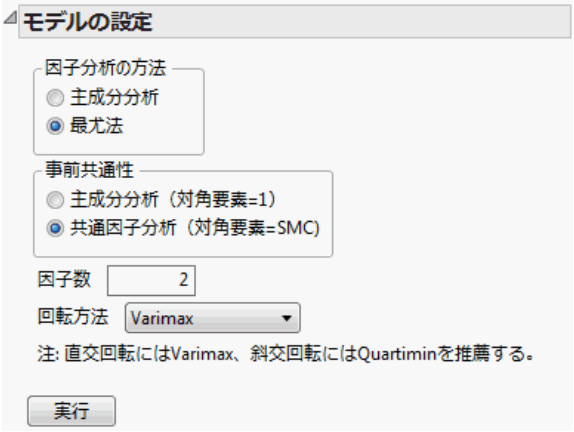


図 4.5 の例では、スクリープロットは 2 番目の固有値の後に横ばいになります。固有値の表を見ると、最初の固有値が変動の 79.75%、2 番目の固有値が 15.75% を説明しているため、この 2 つで全変動の 95.50% が説明されることがわかります。3 番目の固有値が説明する部分は 2.33% に過ぎず、残りの固有値による寄与は取るに足らない程度です。当初の「**因子数**」は 1 に設定されていますが、これらの結果を見る限りは、1 因子ではなく、2 因子とする方が良さそうです。

モデルの設定

因子分析モデルの設定は、「因子分析」レポートの下部の「モデルの設定」セクションで行います（図 4.6）。

図 4.6 モデルの設定



「モデルの設定」セクションでは、次のようなオプションを設定します。

1. **因子分析の方法** - ここでは、因子を抽出する方法を選択します。
   - 主成分分析や、反復なしの主因子法を選択する場合は、［**主成分分析**］を選択します。
   - 多変量正規分布の尤度に基づき、最尤推定を行いたい場合は、［**最尤法**］を選択します。［最尤法］では、統計的検定に基づいて、現在の因子数が有意かどうかも算出されます。

注：［**最尤法**］では、相関行列が正値定符号行列である必要があります。相関行列が正値定符号行列でない場合（いくつかの固有値が 0 以下の場合）には、［**主成分分析**］を選択してください。

2. **事前共通性** - 事前共通性の推定方法を選択します。
   - ［**主成分分析（対角要素 =1）**］は、共通性が 1 とされます。（方法として「**主成分分析**」が選択されている場合）主成分分析が実行されます。
   - ［**共通因子分析（対角要素 =SMC）**］は、共通性が SMC とされます。（方法として「主成分分析」が選択されている場合）反復なしの主因子法が実行されます。SMC（Squared Multiple Correlation）とは、Xi と、その他の p-1 個の変数との重相関を 2 乗した値のことです。
3. 「**因子数**」（主成分）は、デフォルトでは、固有値が 1.0 以上となっている次元が設定されます。スクリープロットを見て、グラフが横ばいになる直前までの次元を用いても良いでしょう。

注：相関行列の固有値のうち、1.0 より大きいものの個数を因子数とする基準は、**Kaiser 基準**などと呼ばれています。この章で用いている例で、Kaiser 基準に従えば、1 因子だけのモデルが選択されます。

4. 「**回転方法**」では、因子の回転方法を選択します。一般的に、回転することにより、因子が解釈しやすくなります。デフォルトの値は **Varimax**（バリマックス）です。各選択肢についての説明は、「回転方法」（68 ページ）を参照してください。
5. ［**実行**］をクリックすると、「因子分析」レポートが作成されます。

   選択した分析対象に合わせて、因子分析の結果が表示されます。レポートの内容の詳細については、「因子分析の「モデルの設定」のオプション」（71 ページ）を参照してください。相関行列、共分散行列、原点周りの積和行列といった分析対象に対して、出力の構成は同じです。

回転方法

因子を上手に回転させると、因子が解釈しやすくなります。回転後の因子負荷量において、一部の変数だけが特定の因子と強く関係していると、因子を解釈しやすくなります。

はじめに抽出された因子では、それらの間の相関はゼロとなっています。また、直交回転は、因子間の相関がゼロであるという制約のもとでの回転です。一方、斜交回転は、回転後の因子間に相関を許します。斜交回転では、直交回転より有益なパターンが見つかることがよくあります。しかし、因子間に相関を許すと、結果の解釈が難しくなります。各因子が変数に与えている大きさを、因子負荷量だけでは判断できなくなるからです。

ここでは、各回転方法について、「「因子分析」プラットフォームの例」（63 ページ）で説明した例を使用し、「回転後の因子負荷量」と「因子負荷量プロット」を表示しました。

直交回転方法

表 4.1 に、JMP で提供されている直交回転を示します。直交回転後の因子間には、相関がありません。

表 4.1 直交回転方法

方法	SAS の PROC FACTOR で同じ結果となるオプション
Varimax	ROTATE=ORTHOMAX、GAMMA = 1 **注：**これはデフォルトです。
Biquartimax	ROTATE=ORTHOMAX、GAMMA = 0.5
Equamax	ROTATE=ORTHOMAX、GAMMA = 因子の数 /2
Factorparsimax	ROTATE=ORTHOMAX、GAMMA = 変数の数
Orthomax	ROTATE=ORTHOMAX または ROTATE=ORTHOMAX(p)。p は直交回転の重み、または GAMMA=p。 **注：**p のデフォルト値は、GAMMA= オプションで指定されない限り、1 です。直交回転の重み付けの詳細については、SAS の PROC FACTOR のマニュアルにおける「Simplicity Functions for Rotations」を参照してください。
Parsimax	ROTATE=ORTHOMAX、GAMMA = $\left[\frac{(nvar(nfact-1))}{(nvar+nfact-2)}\right]$ $nvar$ は変数の数で、$nfact$ は因子の数。
Quartimax	ROTATE=ORTHOMAX、GAMMA=0

斜交回転方法

表 4.2 に、JMP で提供されている斜交回転を示します。斜交回転後の因子間には、相関があります。

表 4.2 斜交回転方法

方法	等価の SAS PROC FACTOR
Biquartimin	ROTATE=OBLIMIN(.5) または ROTATE=OBLIMIN、TAU=.5
Covarimin	ROTATE=OBLIMIN(1) または ROTATE=OBLIMIN、TAU=1

表 4.2 斜交回転方法（続き）

方法	等価の SAS PROC FACTOR
Obbiquartimax	ROTATE=OBBIQUARTIMAX
Obequamax	ROTATE=OBEQUAMAX
Obfactorparsimax	ROTATE=OBFACTORPARSIMAX
Oblimin	ROTATE=OBLIMIN。*p* のデフォルト値は、TAU= オプションで指定されない限り、1 です。 ROTATE=OBLIMIN(*p*)。*p* は斜交回転の重み、または TAU=*p*。 **注：**斜交回転の重み付けの詳細については、SAS の PROC FACTOR のマニュアルにおける「Simplicity Functions for Rotations」を参照してください。
Obparsimax	ROTATE=OBPARSIMAX
Obquartimax	ROTATE=OBQUARTIMAX
Obvarimax	ROTATE=OBVARIMAX
Quartimin	ROTATE=OBLIMIN(0) または ROTATE=OBLIMIN、TAU=0
Promax	ROTATE=PROMAX

「因子分析」プラットフォームのオプション

「因子分析」プラットフォームの赤い三角ボタンのメニューで、次のようなレポート要素の表示／非表示を切り替えることができます。

固有値　相関行列の固有値を示す表。固有値は、主成分によって説明される分散を表します。固有値の割合（主成分の寄与率）を示す棒グラフ、累積寄与率も表示されます。因子分析の因子数を決める 1 つの基準として、固有値が 1 以上のものを採用するという方法があります。

スクリープロット　主成分の次元に対して、固有値をプロットしたグラフです。因子分析の因子数を決める 1 つの基準として、グラフが横ばいになる直前までの因子数までが、十分な因子数と考える方法があります。

スクリプト　プラットフォームの［スクリプト］メニューのオプションをリストします。詳細については、『JMP の使用法』の「JMP のプラットフォーム」の章を参照してください。

例については、図 4.2（64 ページ）を参照してください。

因子分析の「モデルの設定」のオプション

「モデルの設定」を行い、［実行］をクリックすると、結果が表示されます。結果のレポートにある赤い三角ボタンのメニューには、次のような追加の処理が用意されています。

事前共通性　各変数のSMC（Squared Multiple Correlation）が表示されます。SMCは、各変数の変動のうち、すべての因子によって説明される変動の割合を、荒く推定したものです。

注：「事前共通性の推定値」表は、［**共通因子分析（対角要素=SMC)**］オプションが選択された場合にのみ表示されます。

図 4.7 事前共通性の推定値

事前共通性の推定値: SMC

1-オクタノール	0.89679
エーテル	0.88297
クロロフォルム	0.90228
ベンゼン	0.96385
四塩化炭素	0.96040
ヘキサン	0.90635

固有値　SMCを代入した相関行列の固有値と、それらが説明する共通分散を表示します。SMCを代入した相関行列は、対角要素に共通性の推定値を代入した相関行列です。固有値は因子によって説明される共通分散を示します。累積パーセントは100%を超える場合があります。SMCを代入した相関行列は正定値行列とは限らず、負の固有値を持つ場合もあるからです。

表の下には、分析で使われた因子数が表示されています。

［固有値］オプションは、「事前共通性」オプションで［［共通因子分析（対角要素=SMC)］オプションが選択された場合にのみ表示されます。共通性の推定値はSMC（Xiと、その他のp-1個の変数の重相関を2乗した値）です。

図4.8は、最初の2つの因子が、共通分散の100.731%を説明していることを示しています。このパターンによって、このデータをモデル化するのに、3つ以上の因子は必要ない可能性があることがわかります。

図 4.8 SMCを代入した相関行列の固有値

SMCを代入した相関行列の固有値

番号	固有値	寄与率	累積寄与率
1	4.7082	85.407	85.407
2	0.8448	15.324	100.731
3	0.0522	0.947	101.677
4	-0.0172	-0.311	101.366
5	-0.0239	-0.433	100.933
6	-0.0514	-0.933	100.000

2 因子が因子数の基準によって保持されます。

回転前の因子負荷量　回転前の因子負荷量を表示します。因子負荷量は、観測変数の分散に共通因子が与えている影響の大きさを示しています。回転前の因子は直交解であり、因子負荷量は、因子と観測変数との相関を示します。絶対値が1に近いほど、その因子が変数に強く影響することを示します。

「値が小さい負荷量を淡色表示: 閾値=」の値より小さい負荷量は、淡色表示になります。この値は、スライダを動かしたり、テキストボックスに入力したりして変更できます。淡色表示のテキストは、**「テキストの濃さ」**で指定した濃さで表示されます。

「テキストの濃さ」も、スライダを動かしたり、テキストボックスに入力したりして変更できます。

注：「値が小さい負荷量を淡色表示: 閾値=」の値と**「テキストの濃さ」**の値は、「回転後の因子負荷量」で使用されるものと同じです。一方の負荷量の表で設定を変更すると、それが他方の表にも反映されます。

図 4.9 回転前の因子負荷量

回転前の因子負荷量

	因子1	因子2
ベンゼン	0.977736	-0.096606
四塩化炭素	0.956158	-0.249473
ヘキサン	0.923401	-0.215765
クロロフォルム	0.915425	-0.222605
1-オクタノール	0.792777	0.575359
エーテル	0.725745	0.623805

値が小さい負荷量を淡色表示: 閾値= 0.3
テキストの濃さ 0.4

注：「回転前の因子負荷量」行列は、同じ因子に関連する変数同士が隣り合って表示されるように並べ替えられます。

回転行列　因子を回転するのに使われた行列を表示します。この行列に従って、因子負荷量が回転されています。

図 4.10 回転行列

回転行列

0.84800	0.53000
-0.53000	0.84800

最終的な共通性の推定値　共通性の最終的な推定値を表示します。共通性は、因子が直交解の場合、因子負荷量の平方和と等しくなります。

図 4.11 最終的な共通性の推定値

最終的な共通性の推定値

1-オクタノール	0.95953
エーテル	0.91584
クロロフォルム	0.88756
ベンゼン	0.96530
四塩化炭素	0.97647
ヘキサン	0.89922

標準化スコア係数　回転後の因子スコアに対する推定値を計算するときに使われる、元のデータ値に乗ずる係数を表示します。

図 4.12 標準化スコア係数

標準化スコア係数

	因子1	因子2
1-オクタノール	-0.269592	0.782996
エーテル	-0.153368	0.400176
クロロフォルム	0.125223	-0.054222
ベンゼン	0.311045	0.008094
四塩化炭素	0.646847	-0.305395
ヘキサン	0.138317	-0.056834

各因子によって説明される分散　回転後の各因子によって説明される分散、寄与率、累積寄与率を表示します。

図 4.13 各因子によって説明される分散

各因子によって説明される分散

因子	分散	寄与率	累積寄与率
因子1	3.4854	58.089	58.089
因子2	2.1186	35.309	93.399

有意性検定　統計的検定が実行されるのは、因子分析の方法として［**最尤法**］を選択した場合のみです。最尤推定を行うと、2つの統計的検定が実行されます。

１つは、「H_0: 共通因子が１つもない」という仮説を帰無仮説とします。この帰無仮説は、変数間の相互相関を有意に説明する共通因子がないことを示します。

もう１つは、「H_0: 現在の因子で十分である」という仮説を帰無仮説とします。この帰無仮説が棄却された場合、変数間の相互相関を説明するにはより多くの因子が必要であると結論付けられます。

図 4.14 の検定では、モデルに含まれている共通因子が相互相関の一部を説明しているものの、他の因子も必要であることが示されています。

注：「有意性検定」表は、因子分析の方法として［**最尤法**］を選択した場合にのみ表示されます。

図 4.14 有意性検定

有意性検定

検定	自由度	カイ2乗	p値(Prob>ChiSq)
H0: 共通因子が1つもない。 HA: 少なくとも1つの共通因子がある。	15.000	691.106	<.0001*
検定	**自由度**	**カイ2乗**	**p値(Prob>ChiSq)**
H0: 2因子で十分である。 HA: もっと多くの因子が必要である。	4.000	26.031	<.0001*

回転後の因子負荷量　回転後の因子負荷量が表示されます。なお、因子負荷量は、直交解の場合、因子と各変数との相関になっています。

「値が小さい負荷量を淡色表示: 閾値=」で指定されている値より小さい負荷量は、淡色表示になります。この値は、スライダを動かしたり、テキストボックスに入力したりして変更できます。

「テキストの濃さ」も、スライダを動かしたり、テキストボックスに入力したりして変更できます。

注：「値が小さい負荷量を淡色表示: 閾値=」の値と**「テキストの濃さ」**の値は、「回転前の因子負荷量」で使用されるものと同じです。一方の負荷量の表で設定を変更すると、それが他方の表にも反映されます。

図 4.15 回転後の因子負荷量

回転後の因子負荷量

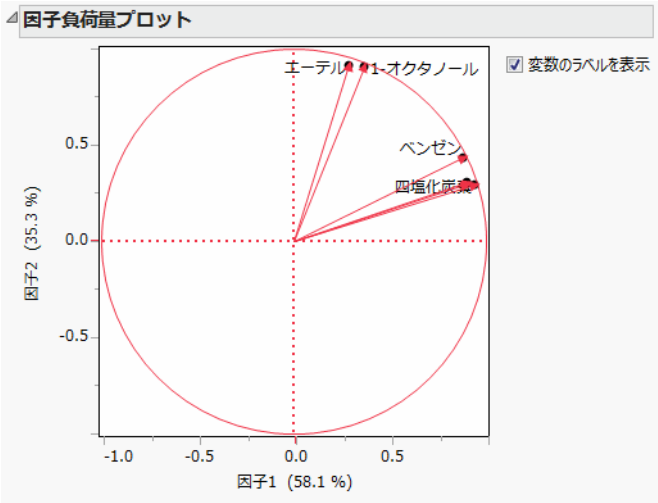
	因子1	因子2
四塩化炭素	0.9430402	0.2952119
ヘキサン	0.8973972	0.3064346
クロロフォルム	0.8942593	0.2964072
ベンゼン	0.8803182	0.4362790
エーテル	0.2848126	0.9136307
1-オクタノール	0.3673323	0.9080748

値が小さい負荷量を淡色表示: 閾値= 0.3
テキストの濃さ 0.4

注：「回転後の因子負荷量」行列は、同じ因子に関連する変数同士が隣り合って表示されるように並べ替えられます。

因子負荷量プロット　回転後の因子負荷量をプロットしたグラフが表示されます。

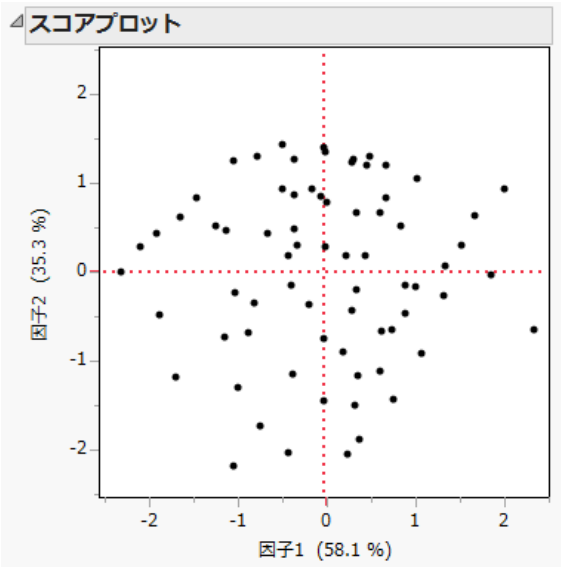
図 4.16 因子負荷量プロット



因子負荷量プロットを見ると、第1因子は**「四塩化炭素 - クロロフォルム - ベンゼン - ヘキサン」**といった変数に、第2因子は**「エーテル -1- オクタノール」**といった変数に強く関連しています。詳細は、「回転後の因子負荷量」（74ページ）の行列を参照してください。

スコアプロット　因子スコア（因子得点）がプロットされます。因子スコアは、平均が0、標準偏差が1になるよう標準化されています。

図4.17 スコアプロット



補完したスコアプロット　欠測値をすべて補完したスコアプロットが作成されます。このオプションは、欠測値がある場合にのみ使用できます。

回転後の成分を保存　回転後の成分の計算式がデータテーブルに保存されます。この計算式で計算される成分は、欠測値のあるデータに対しては、欠測値になります。

補完して回転後の成分を保存　欠測値を補完した後に算出した、回転後の成分がデータテーブルに保存されます。列には、欠測値を補完し、回転後の成分を算出する計算式が保存されます。このオプションは、欠測値がある場合に、［因子分析］オプションを使用すると表示されます。

あてはめの削除　「因子分析」レポートから、該当する因子分析の結果を削除します。このコマンドは、現在の内容を破棄し、「モデルの設定」の内容を変更し、新たに因子分析を行いたい場合に使ってください。

第 5 章

選択モデル

選択モデルのあてはめ

「選択モデル」は、市場調査（マーケットリサーチ）の実験データを分析するためのプラットフォームです。消費者の選好構造を明らかにし、消費者に好まれる製品やサービスを企画することができます。

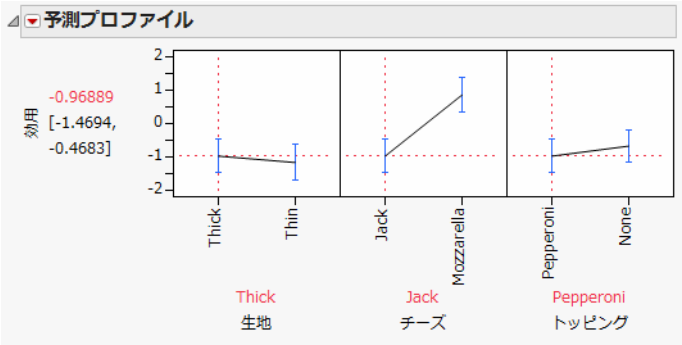
「選択モデル」プラットフォームには、次のような特長があります。

- 製品の属性だけでなく、消費者の特性に関する情報も利用できる
- さまざまな形式のデータを分析できる
- プロファイルによって、消費者の効用を視覚化・最適化できる
- 被験者ごとにスコアを求めることができ、求められたスコアを用いて、被験者をセグメントやクラスターに分けることができる
- Firth（1993）によるバイアス修正を伴う最尤推定が行えます。この推定法は、バイアス修正を行わない最尤法に比べ、推定と検定が優れていることが示されています。また、バイアス修正を伴う最尤推定法を用いることにより、ロジスティックモデルなどで生じる分離（separation）の問題を回避できます。ロジスティック回帰における分離問題については、Heinze and Schemper（2002）を参照してください。

「選択モデル」プラットフォームは、次のような要素を含むモデルには適していません。

- データが順位またはスコアの場合
- 枝分かれ階層の選択肢（このような分析には、SAS/ETS の PROC MDC を使用できます。）

図 5.1「選択モデル」プラットフォームの例 - 予測プロファイル



目次

「選択モデル」プラットフォームの概要 79
「選択モデル」プラットフォームの例 80
「選択モデル」プラットフォームの起動 81
選択モデルの出力 85
被験者効果 86
グリッドごとの効用最適化 88
「選択モデル」プラットフォームのオプション 90
例：トレードオフの評価 92
データテーブルが1つしかないデータ形式 97
例：ピザのデータを1つのテーブルにまとめた場合 98
セグメント化 100
データに関する特殊規則 105
デフォルトの選択肢集合 105
応答データを含んだ被験者データ 105
ロジスティック回帰 105
データの変換 109
データを2つの分析テーブルに変換する 109
データを1つの分析テーブルに変換する 113
対応のあるデータに対する条件付きロジスティック回帰 115
統計的詳細 116

「選択モデル」プラットフォームの概要

McFadden（1974）が先駆者となって考案された選択モデルは、提示された選択肢の中から個人がどのような選択を行うかを推定する、強力な分析方法です。選択モデルは、コンジョイントモデル、離散選択モデル、条件付きロジスティック回帰ともいいます。

JMP の「選択モデル」プラットフォームで使われているモデルは、条件付きロジスティック回帰の一形態です。選択モデルは、通常のロジスティック回帰とは異なり、消費者の特性だけでなく、応答の属性も考慮したモデルになっています。たとえば、A、B の 2 台の自動車のうち、どちらを消費者が購入するかを調査したとします。通常のロジスティック回帰では、年齢・性別・所得・学歴といった消費者の特性だけが説明変数（因子）として扱われ、応答はこれらの説明変数の関数として表わされます。一方、選択モデルでは、消費者の情報だけでなく、価格・乗車人数・カップホルダーの数・色・カーナビ・燃費・盗難防止システム・取り外し可能なシート・安全機能の数・保険料といった、応答（この例では自動車）の属性も考慮されます。

製品開発において、数百、数千という細かい製品仕様をエンジニアは決めていきます。そのとき、消費者の意見を取り入れずに製品を開発していくと、最終的な製品は最適にはなりません。消費者による商品テストがそれほどコストがかからず、見込み客の被験者を確保するのが難しくないのであれば、消費者選択実験（consumer choice experiment）を行って、その結果を製品開発に生かす価値があります。

製品の開発技術が進歩したおかげで、カスタマイズした製品を作成するコストが低くなってきています。デジタル・プロトタイピング（デジタルによる試作品作成）などの普及により、試作品の作成にあまりコストがかからなくなったため、複数のパターンの製品を試すことができるようになりました。また、インターネットの普及によって、選択実験を多くの見込み客に提供できるようになりました。「次回発表される製品に関して、あなたの意見を教えてください」と消費者に通知して、Web アンケートを実施することができます。

既存のシックス・シグマ・プログラムに「選択モデル」を加えることにより、製品をさらに改善できるでしょう。シックス・シグマは、製造工程を変更したり、製品の性能や信頼性を向上したりすることにより、製品を改善します。しかし、既存のシックス・シグマには重要な 1 つの側面が抜けています。それは、「人々が実際に望んでいる製品を作る」という点です。シックス・シグマ・プログラムでは、顧客満足度調査などによりお客様の声を調べます。満足度調査では、商品の何が悪いのかは知ることができますが、製品の属性に対する人々の好みは明らかになりません。消費者の実際の嗜好を調べるためのツールが選択実験です。選択モデル分析を行うことで、消費者の嗜好を特定することができます。

消費者選択を確かめるための市場調査実験は、昔から行われていますが、実施コストが高いという欠点がありました。また、価格の弾力性と他製品との競合状況に重点が置かれていました。従来の市場調査や選択モデルの手法は、製品設計のエンジニアリングの場面に用いてこそ、威力を発揮できるでしょう。

「選択モデル」プラットフォームの例

航空会社にピザを供給している会社が、飛行機の乗客に最も適したピザの属性を特定しようとしています。そのために、頻繁に飛行機を利用する乗客を対象に選択式のアンケート調査を行うことにしました。この調査の目的は、どの属性が重要なのか、また、属性の間に交互作用がないかどうかを知ることです。質問は、2つの選択肢から、好きなほうのピザを1つだけ選択してもらう形式にしました。比べられる2つのピザの中には、どちらのピザが良いのか甲乙付け難いものもあるでしょうから、二者択一の選択はそれほど簡単ではありません。

この例では、ピザの属性が3つ、各属性に水準が2つあるものとします。

- 生地（「Thick」（厚め）と「Thin」（薄め））
- チーズ（「Mozzarella」（モッツァレラ）と「Jack」（モンテレージャック））
- トッピング（「Pepperoni」（ペパロニ）と「None」（なし））

ある被験者は、薄めの生地にトッピングなしでモッツァレラだけをのせたピザが一番好きだとします。そして、この人に対して、「厚めでモッツァレラとペパロニをのせたピザ」と、「薄めでトッピングなしでモントレージャックをのせたピザ」のどちらが選択するか質問したとします。被験者は、どちらも自分が一番好きなピザではないため、どの属性がより重要かを考えることになります。

プロファイルデータテーブルに、被験者に提示する選択肢の組み合わせをすべて入力します。それぞれの組み合わせに **ID** を割り当てます。プロファイルデータテーブルは、図5.2のようになります。

図 5.2 ピザのプロファイルデータテーブル

	生地	チーズ	トッピング	ID
1	Thick	Mozzarella	Pepperoni	ThickOni
2	Thick	Mozzarella	None	ThickElla
3	Thick	Jack	Pepperoni	ThickJackoni
4	Thick	Jack	None	ThickJack
5	Thin	Mozzarella	Pepperoni	TrimOni
6	Thin	Mozzarella	None	Trimella
7	Thin	Jack	Pepperoni	TrimPepperjack
8	Thin	Jack	None	TrimJack

アンケート調査では、被験者1名につき4つの質問をするとします。被験者は、各質問において、2つの選択肢プロファイル（**「選択肢1」**と**「選択肢2」**）の中から自分の好みに合ったものを選択します。各質問において提示される選択肢プロファイルの集まりを、「選択肢集合」と呼びます。質問の内容（選択肢集合）は、各質問者で異なっていてもかまいません。選択モデルの最適計画を作成するには、実験計画（DOE）の「選択モデル計画」プラットフォームを参照してください。図5.3は、最初の3人の被験者による12の回答を示しています。

図 5.3 ピザの応答データテーブル（一部）

	被験者	選択肢1	選択肢2	選択
1	1	ThickJack	TrimPepperjack	TrimPepperjacl
2	1	TrimPepperjack	ThickElla	ThickElla
3	1	TrimOni	Trimella	TrimOni
4	1	ThickElla	ThickJack	ThickElla
5	2	Trimella	ThickJackoni	Trimella
6	2	TrimJack	ThickElla	ThickElla
7	2	Trimella	TrimPepperjack	Trimella
8	2	TrimPepperjack	TrimOni	TrimOni
9	3	TrimOni	ThickJackoni	TrimOni
10	3	TrimPepperjack	ThickElla	ThickElla
11	3	ThickJackoni	TrimPepperjack	ThickJackoni
12	3	ThickOni	Trimella	ThickOni
13	4	ThickElla	ThickOni	ThickElla
14	4	TrimPepperjack	ThickJack	ThickJack

このデータテーブルの「選択」の列には、被験者が2つの選択肢のうちの、どちらのピザを選択したかが入力されています。これらの値には、プロファイルデータテーブルの「**ID**」列に入力されている値と同じ値を使ってください。なお、プロファイルデータテーブルを見れば、該当するピザが、どのような属性をもっているかが分かります。

被験者に関するデータを使用する場合は、これとは別に被験者データテーブルが必要になります。被験者データテーブルには、被験者の**ID**と属性が含まれます。ピザの例では、「**被験者**」の属性として「**性別**」が入力されています。図5.4は、最初の4人の被験者に関する被験者データです。応答データとプロファイルデータにおいて、プロファイルの**ID**を対応させる必要があります。同様に、応答データと被験者データにおいて、被験者のIDを対応させる必要があります。

図 5.4 ピザの被験者データテーブル（一部）

	被験者	性別
1	1	M
2	2	F
3	3	M
4	4	F

「選択モデル」プラットフォームの起動

「選択モデル」プラットフォームの特徴は、次に示すような形式の、1つから3つまでのデータテーブルを使用する点です。

プロファイルデータ　各選択肢の属性を表すデータです。各選択肢は、多数の異なる属性をもちます。プロファイルデータの各列は、その属性の一つ一つを表します。各行は、各選択肢を表し、それぞれに一意の**ID**が割り当てられている必要があります。

応答データ 実験の結果を入力したデータテーブルです。各質問における複数の選択肢と、そのなかから被験者によって選択された選択肢を含みます。通常、一人の被験者が複数の質問に回答するため、一人の被験者に対するデータは複数の行に保存されます。たとえば、被験者の人数が100名で、各被験者が12回選択を行ったとすると、データテーブルは1,200行になります。応答データは、選択肢の集合を示す列と、実際に選択されたものを含む列により、プロファイルデータと関連付けられます。「選択肢の集合」とは、被験者が選択をする際に提示された選択肢の集まりを指します。応答データとプロファイルデータにおいて選択肢のIDを合わせるのにグループ化が必要な場合には、その変数をグループ変数として指定する必要があります。

被験者データ モデルに被験者効果を含める場合に使用します。被験者データを使わない場合には、指定する必要はありません。被験者データテーブルには、各被験者の属性とIDが含まれます。このデータテーブルは、被験者と同数の行で構成されます。応答データテーブルに、対応した被験者IDの列を用意する必要があります。被験者データは、応答データテーブルに一緒に含めておくこともできますが、その場合でも被験者データテーブルとしても指定する必要があります。

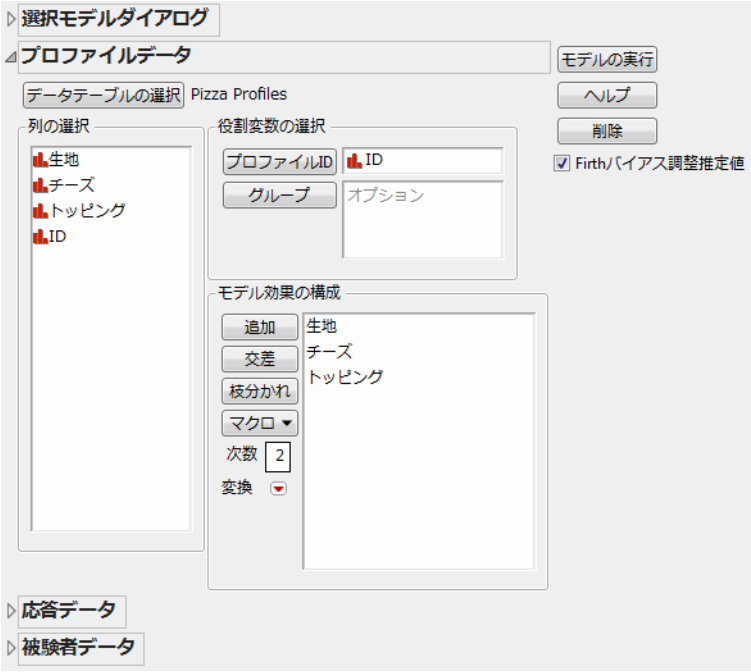
すべてのデータを1つのデータテーブルに入力したデータ形式でも、「選択モデル」プラットフォームを使用することはできますが、別々のデータテーブルを使うより手間がかかります。「データテーブルが1つしかないデータ形式」（97ページ）の節を参照してください。

「選択モデル」プラットフォームでは、複数のデータテーブルを扱います。そのため、他のプラットフォームのように、現在開いているデータテーブルが自動的に使用されるわけではありません。ユーザが選択する必要があります。どのデータテーブルを、プロファイルデータと応答データに使用するかを指定する必要があります。また、被験者属性の効果もモデルに含めたい場合は、被験者データも指定する必要があります。「選択モデル」ダイアログボックスの各セクションは、必要に応じて開いたり閉じたりできます。

ここでは、「選択モデル」プラットフォームの機能について学習するため、サンプルデータのフォルダにあるピザの3つのデータテーブルを使用します。最初に、「選択モデル」ダイアログボックスの「プロファイルデータ」セクションでデータを選択します（図5.5）。

1. ［**分析**］>［**消費者調査**］>［**選択モデル**］を選択し、起動ダイアログボックスを開きます。データソースを選択するセクションが3つあります。
2. 「プロファイルデータ」セクションにある［**データテーブルの選択**］をクリックします。新しいダイアログボックスが開いたら、プロファイルデータとして使うデータテーブルを指定します。現在開いているいずれかのデータテーブル、または［**その他**］を選択します。［**その他**］を選択した場合、まだ開いていないデータテーブルを開くことができます。
3. 「Pizza Profiles.jmp」を選択します。「選択モデル」ダイアログボックスの「**列の選択**」パネルにこのデータテーブル内の列が一覧されます。
4. 「**役割変数の選択**」パネルで「**ID**」を［**プロファイル ID**］に指定し、「**モデル効果の構成**」パネルで「**生地**］、［**チーズ**］、［**トッピング**］を［**追加**］します。［**プロファイルID**］列によってプロファイルのデータの各行を一意に識別できない場合は、［**グループ**］変数の列を追加し、［**グループ**］列と［**プロファイルID**］列の組み合わせによって行が一意に識別されるようにします。たとえば、［**プロファイルID**］の値が1である行が2つあったとします。一方の行の「**調査**」がAで、他方の行の「**調査**」がBであるなら、「**調査**」を［**グループ**］変数に指定します。この例では、2水準の因子が3個あり、8通りの組み合わせがありますが、プロファイルIDで一意に識別できています。

図 5.5「プロファイルデータ」セクション

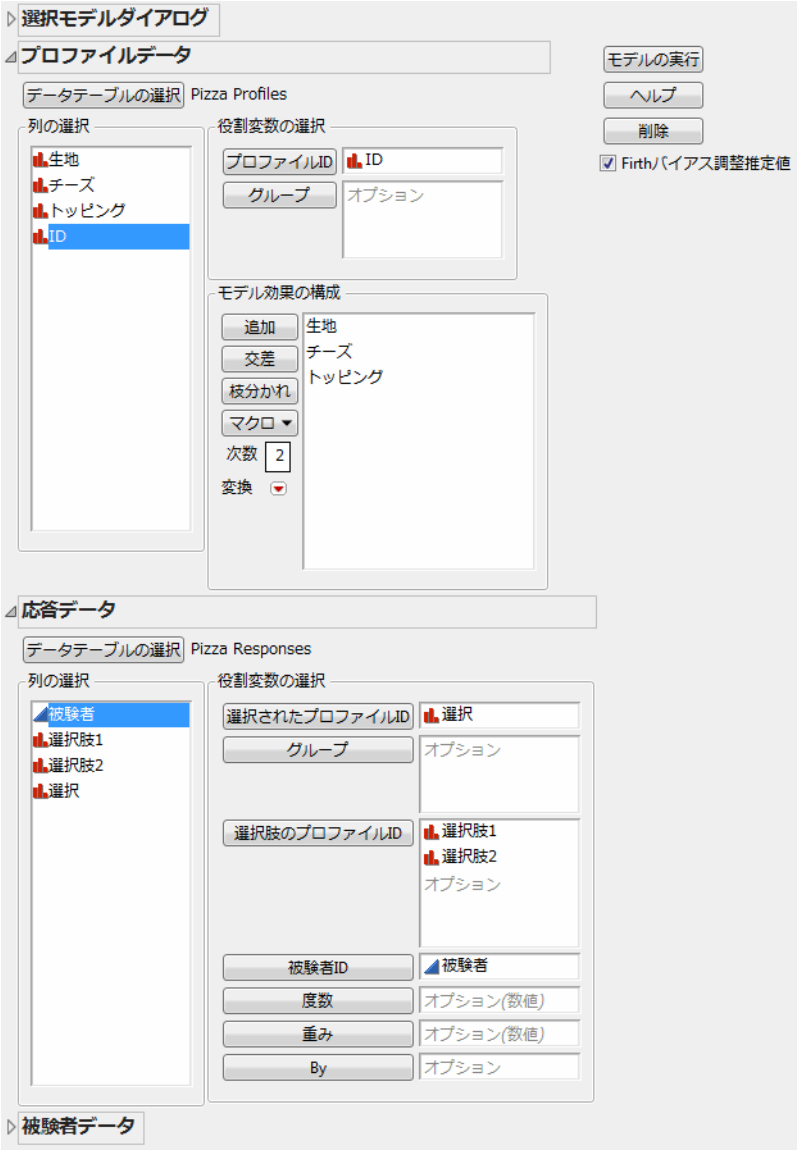


2つ目のデータテーブルは、実験結果を入力した［応答データ］です。ピザの実験では、二択式の質問が、1被験者あたりに4つ割り当てられました。応答データのサンプルデータでは、**「選択肢1」**と**「選択肢2」**という列とともに、実際に選択された値が**「選択」**という列に入力されています。

5. ダイアログボックスの「応答データ」セクションを開き、［**データテーブルの選択**］をクリックします。そして、「応答データテーブル」ダイアログボックスが開いたら、「Pizza Responses.jmp」を選択します。
6. ［**選択されたプロファイル ID**］に**「選択」**、［**選択肢のプロファイル ID**］に**「選択肢1」**と**「選択肢2」**を指定します。
7. 個々の被験者を区別できるよう［**被験者　ID**］として**「被験者」**を指定します。この指定はオプションです。被験者の効果を評価しないのであれば、被験者 ID の列を指定したり、被験者データを含めたりする必要はありません。
8. ［**度数**］と［**重み**］は、分析の重み付けに使います。たとえば、結果が同じである行を集計したデータの場合には［**度数**］が使用されます。

被験者データ以外のすべてを指定すると、ダイアログボックスは図5.6のようになります。

図 5.6「応答データ」セクション



「選択モデル」プラットフォームをスクリプトで実行する場合は、Choice() の指定に次のコマンドを追加することで、パラメータ推定時の収束基準を設定することができます。

```
Choice( ..., Convergence Criterion( fraction ), ... )
```

例については、[**ヘルプ**] メニューの［スクリプトの索引］を参照してください。

選択モデルの出力

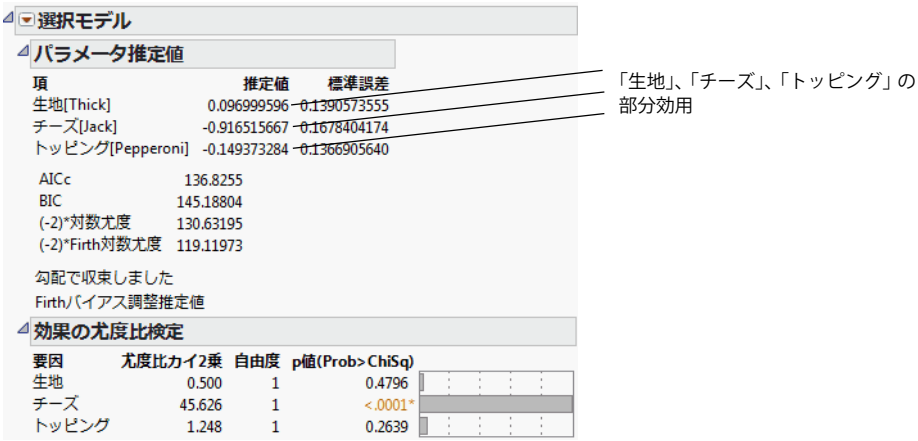
［**モデルの実行**］ボタンをクリックすると、分析結果が表示されます。結果は図5.7のようになります。

- 計算されたパラメータ推定値は、**部分効用（part-worths）**と呼ばれることがあります。パラメータ推定値（部分効用）は、各属性の効用に対する寄与を表す係数です。デフォルトでは、これらの推定値は、通常の最尤推定ではなく、Firthのバイアス調整最尤法によって推定されます。
- データに複数のモデルをあてはめて、それらのモデルを比較した場合には、比較規準によって適合度の良いモデルを選ぶことができます。規準値の小さいモデルが良いモデルです。「選択モデル」のレポートに、「AICc（修正された赤池情報量規準）」、「BIC」（ベイズの情報量規準）、「(-2) * 対数尤度」、「(-2) *Firth対数尤度」の4つの規準が表示されます。AICcの計算式は次のとおりです。

  $$\text{AICc} = -2\,\text{対数尤度} + 2k + \frac{2k(k+1)}{n-k-1}$$

  kは、モデルにおいて推定されたパラメータの数、nは、データのオブザベーション数を示します。BICの計算式は、次のとおりです。
  - 2 対数尤度 + k * ln(n)
  なお、［Firthバイアス調整推定値］にチェックマークをつけた場合（図5.6を参照）、この他にも、「(-2) *Firth対数尤度」が出力されます。このオプションは、デフォルトでオンになっています。なお、Firthバイアス調整による推定を行った場合も、「AICc」や「(-2) * 対数尤度」の計算式は同じです。
- モデルの各効果に対しては、統計的検定として尤度比検定が実行されます。モデル推定の処理時間が短い（5秒未満）場合には、これらの尤度比検定はデフォルトで表示されます。デフォルトで表示されていない場合でも、「選択モデル」のドロップダウンメニューにある［**尤度比検定**］をクリックすると表示されます。

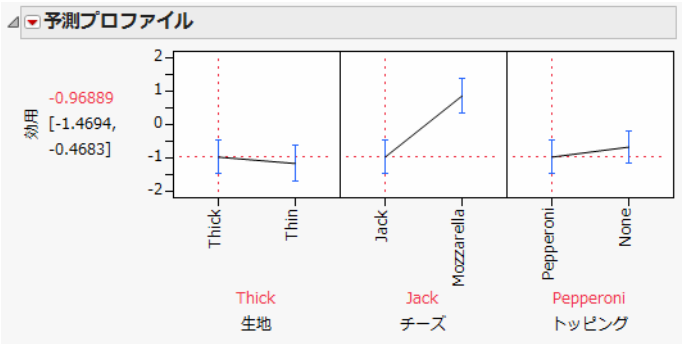
図5.7 ピザの例で被験者データを指定しなかったときの選択モデル



プロファイルは、モデルの理解に大いに役立ちます。プロファイルは、因子を1つずつ変化させた時の関数の変化を表したものです。選択モデルのプロファイルでは、効用（モデルの線形式部分）の値が表示されます。

- プロファイルを表示するには、プラットフォームメニューの［**プロファイル**］を選択します。図 5.8 は、この例におけるプロファイルです。
- 主効果しかないこの例で、最大の差が見られる因子は「**チーズ**」です。最も人気が高い種類はモッツァレラです。
- モデルに交互作用があるなら、プロファイルをより詳しく検討し、交互作用効果を理解する必要があります。

図 5.8 ピザの例で被験者データを指定しなかったときの予測プロファイル



被験者効果

モデルに被験者効果を含める場合は、「選択モデル」ダイアログボックスの「被験者データ」セクションを開きます。「**性別**」の主効果を詳しく調べたいとしましょう。

1. ダイアログボックスの「被験者データ」セクションを開きます。［**データテーブルの選択**］をクリックし、「被験者 データテーブル」が開いたら、「Pizza Subjects.jmp」を選択します。
2. ［**被験者 ID**］に「**被験者**」を指定し、「**モデル効果の構成**」に「**性別**」を追加します。図 5.9 は、ピザの例の「被験者データ」セクションです。
3. ［**モデルの実行**］をクリックします。

図 5.9 ピザの例の「被験者データ」セクション

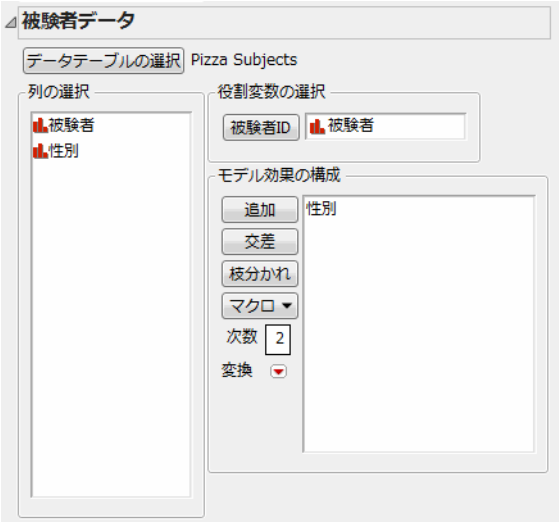
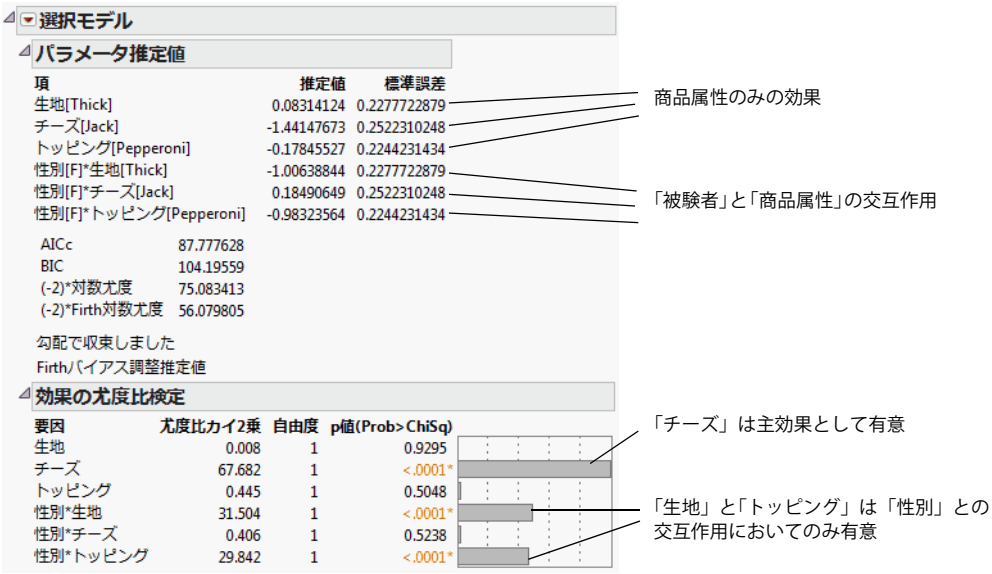


図 5.10 は、選択モデルに被験者効果を含めたときのパラメータ推定値と尤度比検定です。**「性別」**と**「生地」**、さらに**「性別」**と**「トッピング」**の間における交互作用は、統計的に有意となっています。**「生地」**と**「トッピング」**の主効果は、統計的に有意ではありません。しかし、**「性別」**との交互作用の効果を見た場合には、**「生地」**と**「トッピング」**の両方とも、高度に統計的に有意となっています。

図 5.10 ピザの例で選択モデルに被験者効果を含めたときの結果



各効果が効用に与えている影響を調べるには、プロファイルを使用します。「選択モデル」のドロップダウンメニューを開き、［**プロファイル**］を選択します。プロファイルにて水準をロックすると、縦の点線が実線に変わります。この例のデフォルトの状態では、性別の水準が、「F」にロックされています。被験者の因子水準をロックすると、それらの属性をもつ被験者に適した、製品の属性を評価しやすくなります。

- 図 5.11 のように「**性別**」の「F」（女性）をロックしてプロファイルを観察すると、女性が薄めの生地にモッツァレラをのせた、トッピングなしのピザを好むことがわかります。たとえば、「生地」を見ると、「Thin」（薄い）のときに「効用」の値が大きくなっています。これは、女性が薄めの生地を好むことを示します。
- 次に、「**性別**」を「M」（男性）に切り替えて男性の嗜好を見てみましょう。図5.12を見ると、男性は、厚めの生地にモッツァレラとペパロニをのせたピザを好むようです。

図 5.11 被験者データを含め、「性別」の水準を「F」（女性）に固定したときの予測プロファイル

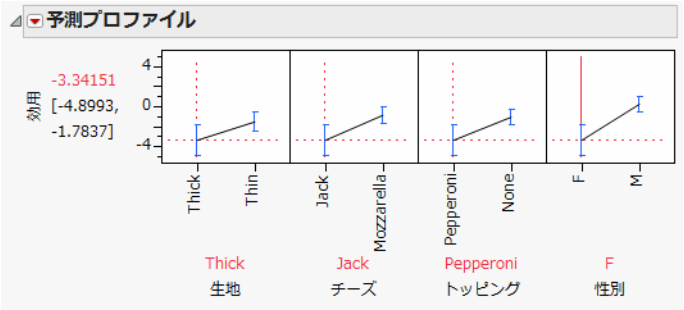
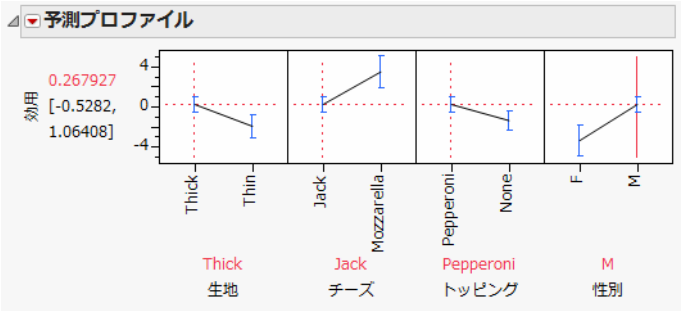


図 5.12 被験者データを含め、「性別」の水準を「M」（男性）に固定したときの予測プロファイル

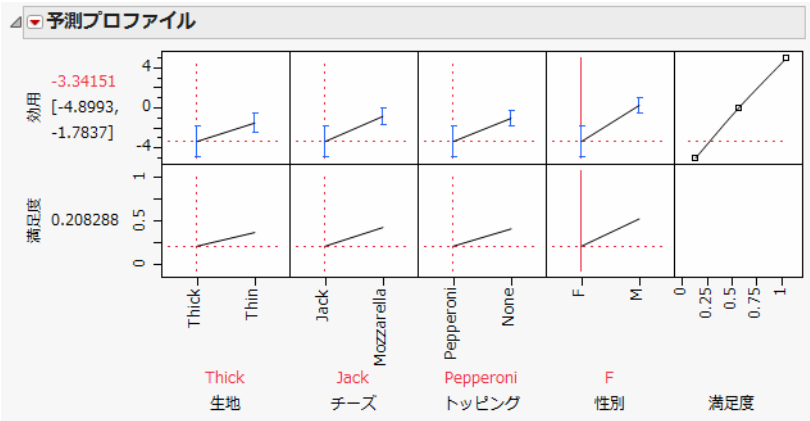


グリッドごとの効用最適化

予測プロファイルでは、手動で1つ1つ操作しなくても、被験者属性を固定したときの効用関数を最適化することができます。

1. プラットフォームの赤い三角ボタンのメニューから［**プロファイル**］を選択し、被験者因子の少なくとも1つがロックされていることを確認します。ロックされていない場合は、Altキーを押しながらプロファイル内をクリックし、任意の水準をロックします。次の操作によって、被験者の属性ごとに、「効用」関数が最大となる商品属性が求められます。
2. 予測プロファイルにある赤色の三角ボタンをクリックし、メニューから［**満足度関数**］を選択します。予測プロファイルに、満足度の行が追加され、全体の満足度を表すトレースが表示されます。図5.13に効用関数と満足度関数を示します。

図5.13 効用関数と満足度関数



3. 予測プロファイルにある赤色の三角ボタンをクリックし、メニューから［**各グリッド点で最大化**］を選択します。「設定の記録」表に、効用関数が最大になる設定がグリッド点ごとに表示され、もう1つのテーブルに、グリッド間における差が表示されます。図5.14を参照してください。ピザの例からわかるように、この機能を使えば、ある属性をもつ消費者に対し、最も満足度の高くなる商品属性の組み合わせをすばやく特定することができます。

図5.14 効用関数と満足度関数に基づいた設定

設定の記録

設定	生地	チーズ	トッピング	性別	効用	満足度
グリッド1	Thin	Mozzarella	None	F	3.3415083	0.820201
グリッド2	Thick	Mozzarella	Pepperoni	M	3.5206933	0.837771

差

設定A	- 設定B	効用	満足度
グリッド2	グリッド1	0.1791849	0.0175696

結果を見ると、女性の効用が最大になるのは、薄めの生地、モッツァレラ、トッピングなしの組み合わせです。一方、男性の効用が最大になるのは、厚めの生地、モッツァレラ、ペパロニの組み合わせです。

「選択モデル」プラットフォームのオプション

「選択モデル」プラットフォームには、多数のオプションがあります。これらのオプションは、プラットフォームのドロップダウンメニューに含まれています。

尤度比検定　モデルの各効果を検定します。CPU 時間が推定で 5 秒未満であれば、最初から自動的に計算されます。

複合因子検定　該当する因子が関連するすべての効果をまとめて検定（尤度比検定）します。

信頼区間　プロファイル尤度に基づき、各パラメータの信頼区間が計算されます。デフォルトの信頼率は 95%（有意水準 5%）です。有意水準を 0.05 以外にするには、Shift キーを押しながらプラットフォームのドロップダウンメニューを開いて［**信頼区間**］を選択します。

推定値の相関　パラメータ推定値の相関を示します。

全水準の推定値　該当する効果以外では中立的な値が取られた時の、効果の各水準がもつ効用の推定値を示します。

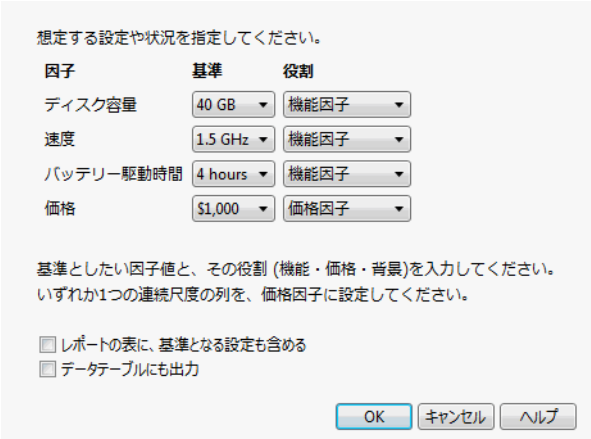
比較　設定された選択肢の比較を行います。比較したい因子の水準を設定できます。［すべて］チェックボックスにチェックを入れると、その因子のすべての水準との比較が行われます。なお、この［**すべて**］チェックボックスを複数の因子でチェックした場合は、それらの複数の因子をすべて組み合わせた比較ではなく、1 因子ごとのすべての水準が比較されます。［**いずれかを含む**］にチェックマークを入れた場合、全機能（特性）のあらゆる組み合わせが比較されるわけではなく、左側の設定を他の因子の設定とした上で、機能ごとに、すべての比較の組み合わせが行われます。

図 5.15 比較の例

支払意志額　製品の機能を新しいものに変更したときに、これまでの製品と同じ効用とするのに、価格をどれぐらい変更したら良いのかを計算します。結果は、各背景ごとに、「**基準**」に指定された値を基準に計算されます。また、次のような「**役割**」を変更すると、結果は再計算されます。

– 機能因子 - どれぐらいの価値があるか、価格を決めたい機能が含まれている列。
– 価格因子 - 価格を示す列。これは、連続尺度である必要があります。
– 背景因子 定数 - 基準を求める際に、計算に用いたい背景因子。
– 背景因子 変数 - すべての値で反復したい因子。

図 5.16 支払意志額の設定例



［**レポートの表に、基準となる設定も含める**］オプションを使うと、価格変更を加える前の基準設定がレポートに表示されます。結果の表に、新しい機能設定だけでなく基準設定も表示したい場合に便利です。

プロファイル　プロファイルを表示します。プロファイルは、一度に 1 因子ずつ変化させた時の効用の変化を表します。

効用計算式の保存　推定された効用（線形モデル）の計算式を新しい列に保存します。被験者効果がない場合は、効用計算式はプロファイルデータテーブルに保存されます。被験者効果がある場合は、計算式用に新しいデータテーブルが作成されます。この計算式は、各種プロファイルを始め、さらに分析を続ける際に使用できます。

被験者ごとの勾配を保存　各パラメータの変化量を被験者ごとに含んだデータテーブルを作成します。パラメータの変化量は、ヘッセ行列で尺度化した各パラメータの傾きを、被験者ごとに平均して求められています。これは、該当する被験者と他の被験者との差異に対するラグランジュ乗数検定に対応します。作成されたデータテーブルを、付随しているスクリプトでクラスター分析すれば、マーケットセグメント（同じ選好構造をもつ消費者集団）を特定することができます。

モデルダイアログ　「選択モデル」ダイアログボックスが開き、モデルを修正したり、別のモデルをあてはめたりすることができます。データテーブルや **ID**、モデル効果を新しく指定できます。

例：トレードオフの評価

「選択モデル」プラットフォームにより、製品がもつ属性の相対的な重要度を調べることもできます。製品について、いくつかの重要な属性は明らかになっていても、それらの属性の間にあるトレードオフの関係がわかっていない場合があります。選択モデルで分析することにより、最適なトレードオフを明らかにし、その情報を製品の企画・設計に反映させることができます。

トレードオフの関係を明らかにする利点について、次の例で解説しましょう。ノートパソコンの設計においては、ディスク容量、プロセッサの速度、バッテリーの寿命、販売価格の4つの属性が重要であることがわかっています。この調査のために収集したデータを使い、4つの属性（**「ディスク容量」**、**「速度」**、**「バッテリー駆動時間」**、**「価格」**）の中でどれが最も重要かを調べます。また、これらの属性に対する嗜好に**「性別」**や**「職種」**による違いがあるかどうかも調べます。

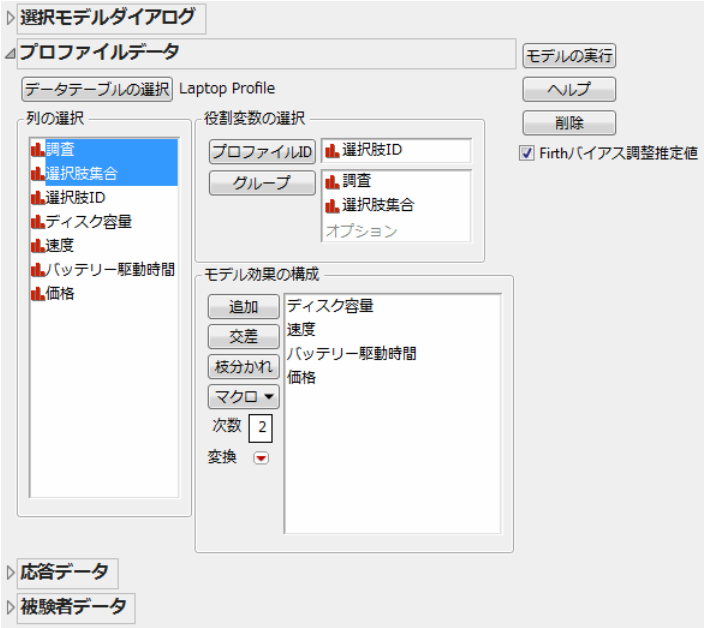
1. ［**分析**］>［**消費者調査**］>［**選択モデル**］を選択し、起動ダイアログボックスを開きます。
2. まず、サンプルデータのフォルダにある「Laptop Profile.jmp」を開きます。そして、「プロファイルデータ」セクションの［**データテーブルの選択**］をクリックし、「Laptop Profile.jmp」を選択します。図5.17はプロファイルデータテーブルの一部です。データは、**「調査」**1と2が各12行の計24行で構成されています。**「調査」**と**「選択肢集合」**はグループを表します。**「選択肢ID」**は、4つの属性（**「ディスク容量」**、**「速度」**、**「バッテリー駆動時間」**、**「価格」**）をもつノートパソコンを示すものです。

図5.17 ノートパソコンの例のプロファイルデータ

	調査	選択肢集合	選択肢ID	ディスク容量	速度	バッテリー駆動時間	価格
1	1	1	1	40 GB	1.5 GHz	6 hours	$1,000
2	1	1	2	40 GB	2.0 GHz	4 hours	$1,500
3	1	2	1	80 GB	1.5 GHz	6 hours	$1,500
4	1	2	2	80 GB	2.0 GHz	4 hours	$1,200
5	1	3	1	80 GB	1.5 GHz	6 hours	$1,200
6	1	3	2	40 GB	2.0 GHz	4 hours	$1,000
7	1	4	1	80 GB	2.0 GHz	4 hours	$1,500
8	1	4	2	40 GB	1.5 GHz	6 hours	$1,200
9	1	5	1	40 GB	2.0 GHz	6 hours	$1,200

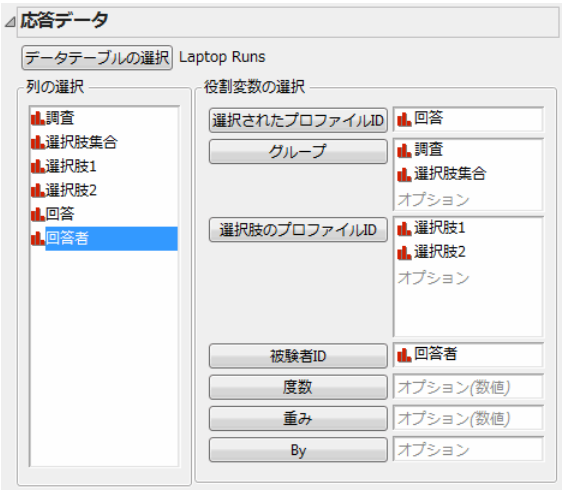
3. ［**プロファイルID**］として**「選択肢ID」**を指定し、**「ディスク容量」**、**「速度」**、**「バッテリー駆動時間」**、**「価格」**をモデル効果に［**追加**］します。
4. **「調査」**と**「選択肢集合」**を［**グループ**］変数に指定します。「プロファイルデータ」セクションは、図5.18のようになります。

図 5.18 ノートパソコン調査の「プロファイルデータ」セクション



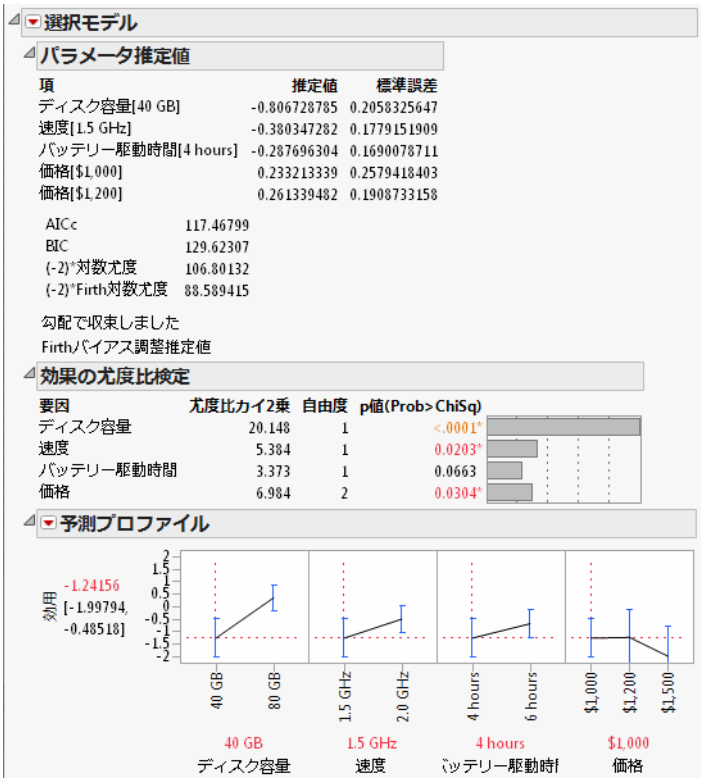
5. ［**応答データ**］>［**データテーブルの選択**］>［**その他**］>［**OK**］の順にクリックし、「Data」フォルダにある「Laptop Runs.jmp」を選択します。
6. 「**回答**」を［**選択されたプロファイル ID**］に指定し、「**選択肢 1**」と「**選択肢 2**」を［**選択肢のプロファイル ID**］、「**調査**」と「**選択肢集合**」を［**グループ**］変数列、「**回答者**」を［**被験者 ID**］に指定します。「応答データ」セクションは、図5.19のようになります。

図 5.19 ノートパソコン調査の「応答データ」セクション



7. 被験者効果を含めずにモデルを実行する場合は、［**モデルの実行**］をクリックします。
8. 赤い三角ボタンのメニューから［**プロファイル**］を選択します。

図 5.20 ノートパソコン調査に被験者効果を含めないときの結果



この結果から、すべての因子が重要で、中でも「**ディスク容量**」が最も重要であることがわかります。容量の大きいノートパソコンが好まれています。価格について言えば、被験者にとって 1000 ドルと 1200 ドルの差はそれほど重要ではありませんが、1200 ドルと 1500 ドルの差は大きいようです。この効果は、予測プロファイルを使ってインタラクティブに因子を検討すると、視覚的に把握できます。こういった事実は、価格を決める際に参考になるでしょう。

ノートパソコンの調査に被験者効果を含めるには、「選択モデル」ダイアログボックスを使います。

1. 「被験者データ」セクションで、［**データテーブルの選択**］>［**その他**］>［**OK**］の順にクリックし、「Data」フォルダにある「Laptop Subjects.jmp」を選択します。
2. 「**回答者**」を［**被験者 ID**］に指定し、「**性別**」と「**職種**」をモデル効果として追加します。「被験者データ」セクションは、図 5.21 のようになります。

図 5.21 ノートパソコン調査の「被験者データ」セクション

3. ［**モデルの実行**］をクリックします。

結果は、図 5.22、図 5.23、図 5.26 のようになります。

図 5.22 ノートパソコン調査に被験者データを含めたときのパラメータ推定値

パラメータ推定値

項	推定値	標準誤差
ディスク容量[40 GB]	-0.887027357	0.2421091374
速度[1.5 GHz]	-0.362097199	0.1907670046
バッテリー駆動時間[4 hours]	-0.394080806	0.2025662868
価格[$1,000]	0.344799986	0.3094758460
価格[$1,200]	0.305771815	0.2112382529
性別[F]*ディスク容量[40 GB]	-0.389344918	0.2594789312
性別[F]*速度[1.5 GHz]	0.017754116	0.2229164234
性別[F]*バッテリー駆動時間[4 hours]	-0.250255571	0.2306512647
性別[F]*価格[$1,000]	0.327081163	0.3293378134
性別[F]*価格[$1,200]	0.103980693	0.2219937705
職種[Development]*ディスク容量[40 GB]	0.083419819	0.2335285785
職種[Development]*速度[1.5 GHz]	-0.150191535	0.2167139358
職種[Development]*バッテリー駆動時間[4 hours]	-0.055689926	0.2115342604
職種[Development]*価格[$1,000]	0.042732902	0.2947913725
職種[Development]*価格[$1,200]	0.134544817	0.2100258497

AICc	129.35431
BIC	161.81953
(-2)*対数尤度	93.354306
(-2)*Firth対数尤度	41.678087

勾配で収束しました。

Firthバイアス調整推定値

図 5.23 ノートパソコン調査に被験者データを含めたときの尤度比検定

効果の尤度比検定

要因	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
ディスク容量	23.690	1	<.0001*
速度	4.535	1	0.0332*
バッテリー駆動時間	5.649	1	0.0175*
価格	9.896	2	0.0071*
性別*ディスク容量	3.183	1	0.0744
性別*速度	0.000	1	1.0000
性別*バッテリー駆動時間	1.763	1	0.1843
性別*価格	2.678	2	0.2621
職種*ディスク容量	0.035	1	0.8516
職種*速度	0.431	1	0.5116
職種*バッテリー駆動時間	0.020	1	0.8864
職種*価格	0.666	2	0.7167

4. プラットフォームメニューの［**複合因子検定**］を選択すると、図 5.24 のようなテーブルが作成されます。

図 5.24 ノートパソコン調査の複合因子検定

複合因子検定

要因	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
ディスク容量	25.027	3	<.0001*
速度	6.337	3	0.0963
バッテリー駆動時間	5.951	3	0.1140
価格	10.272	6	0.1136
性別	8.096	5	0.1510
職種	3.617	5	0.6058

5. プラットフォームメニューの［**全水準の推定値**］を選択すると、図 5.25 のような表が表示されます。各因子の水準ごとの推定値が表示されます。各因子の推定値は、全水準で合計したときに 0 になります。

図 5.25 ノートパソコン調査の全水準の推定値

全水準の推定値

推定値	ディスク容量
-0.83836	40 GB
0.83836	80 GB

推定値	速度
-0.36432	1.5 GHz
0.36432	2.0 GHz

推定値	バッテリー駆動時間
-0.36280	4 hours
0.36280	6 hours

推定値	価格
0.30391	$1,000
0.29277	$1,200
-0.59669	$1,500

図 5.26 ノートパソコン調査に被験者データを含めたときの「F」（女性）のプロファイル

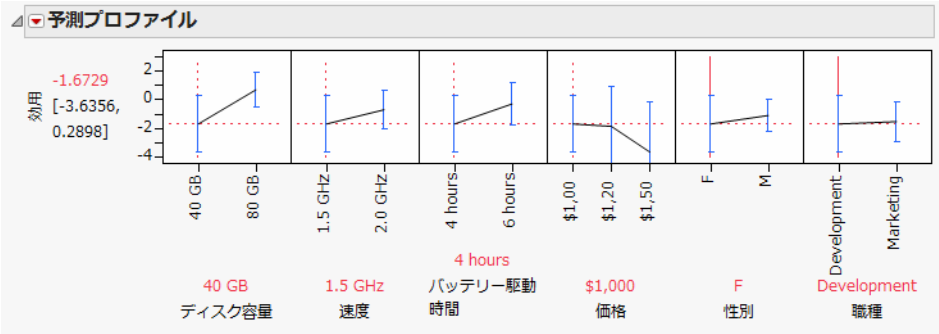
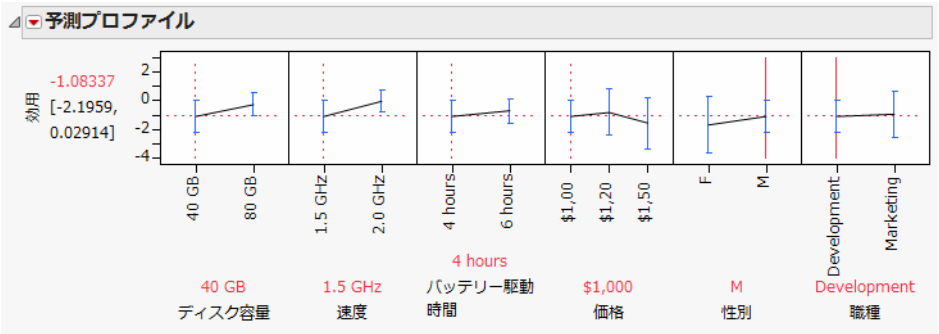


図 5.27 ノートパソコン調査に被験者データを含めたときの「M」（男性）のプロファイル



「**性別**」と「**ディスク容量**」との交互作用は、p 値が 0.0744 でかろうじて有意です（図 5.23（96 ページ）を参照）。「予測プロファイル」で「**性別**」の両水準の「**ディスク容量**」を確認すると、男性よりも女性のほうが傾きが大きいことがわかります。

データテーブルが 1 つしかないデータ形式

「選択モデル」プラットフォームは、データがすべて 1 つのテーブルに保存されている場合でも使用できます。テーブルが 1 つの場合は、「選択モデル」ダイアログボックスの「プロファイルデータ」セクションだけを使います。被験者の効果もモデルに含めることができますが、主効果として指定することはできません。テーブルを 1 つしか使用しない方法は、テーブルを 3 つ使用する方法に比べ、2 つの利点があります。

- 交互作用を柔軟に指定でき、特定の交互作用項だけをモデルに含めることができます。一方、テーブルを 3 つ使用した場合は、ノートパソコンの例からもわかるように、被験者効果とプロファイル効果の間で考えられるすべての交互作用が自動的にモデルに含まれます。
- 選択肢の組み合わせを柔軟に指定できます。たとえば、1 つ目の質問は二択、2 つ目の質問は三択、3 つ目の質問は五択、といったデータも分析できます。一方、複数のテーブルを用いるときは、どの質問も選択肢の数は同じでなければいけません。

テーブルを1つしか使用しない場合、［グループ］変数によって一意に識別される複数の行が、各質問の選択肢を表します。「選択モデル」ダイアログボックスで、指示変数（インジケータ変数）の列を［プロファイル ID］に指定します。指示変数には、選択されたプロファイルの行では1、選択されなかったプロファイルの行では0を保存しておきます。1の値は、［グループ］変数の組み合わせごとに1回だけ出現しなければいけません。

例：ピザのデータを1つのテーブルにまとめた場合

ここでは、ピザのデータを1つのテーブルにまとめた場合の分析手順を紹介します。図5.28は、ピザのデータをまとめたテーブル（一部）です。テーブル全体を見るには、サンプルデータのフォルダにある「Pizza Combined.jmp」を開きます。各被験者は、4つの二択問題に答えました。1つにまとめたテーブルには、被験者ごとに8つの行があります。指示変数は、各選択肢集合のなかから、どの選択肢が選ばれたかを示しています。**「被験者」**と**「テスト」**の組み合わせにより、各選択肢集合を識別することができるので、この2列を［**グループ**］変数とします。

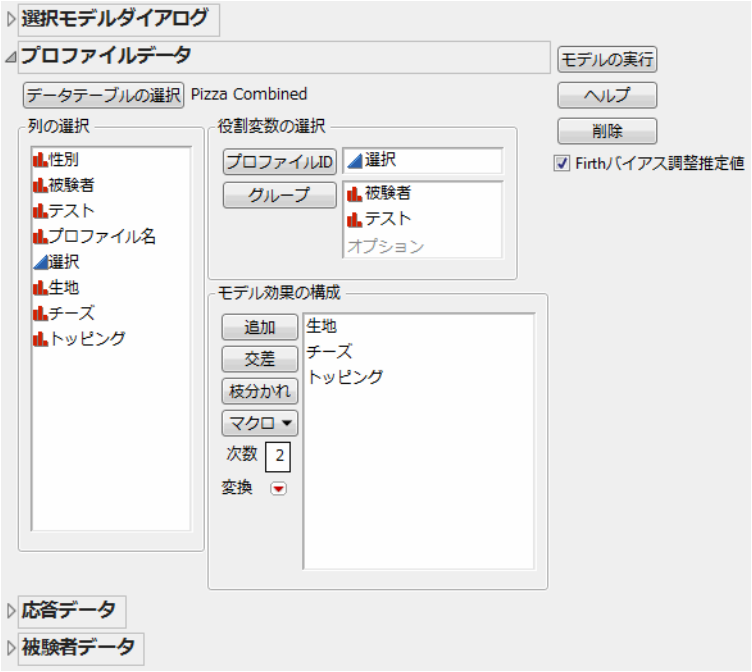
図5.28 ピザのデータを1つにまとめたテーブル（一部）

	性別	被験者	テスト	プロファイル名	選択	生地	チーズ	トッピング
1	M	1	1	ThickJack	0	Thick	Jack	None
2	M	1	1	TrimPepperjack	1	Thin	Jack	Pepperoni
3	M	1	2	TrimPepperjack	0	Thin	Jack	Pepperoni
4	M	1	2	ThickElla	1	Thick	Mozzarella	None
5	M	1	3	TrimOni	1	Thin	Mozzarella	Pepperoni
6	M	1	3	Trimella	0	Thin	Mozzarella	None
7	M	1	4	ThickElla	1	Thick	Mozzarella	None
8	M	1	4	ThickJack	0	Thick	Jack	None

1つにまとめられたデータを分析するには、「選択モデル」ダイアログボックスの「プロファイルデータ」セクションを開きます（図5.29）。

1. データテーブルとして「Pizza Combined.jmp」を指定します。
2. **「選択」**を［**プロファイル ID**］、**「被験者」**と**「テスト」**を［**グループ**］変数に指定し、**「生地」**、**「チーズ」**、**「トッピング」**を主効果として追加します。
3. ［**モデルの実行**］をクリックします。

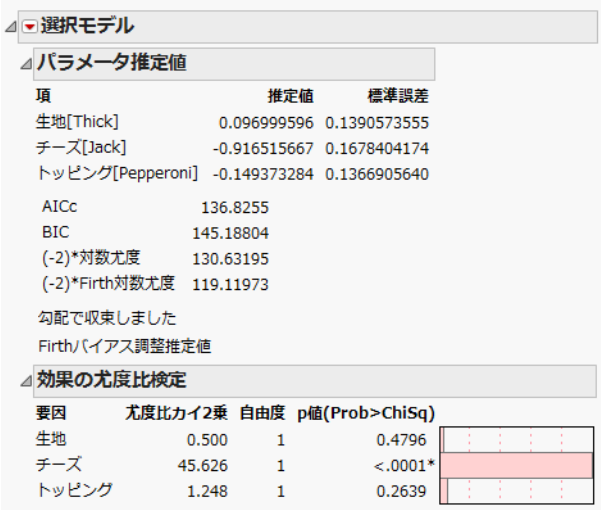
図 5.29 ピザのデータを 1 つのデータテーブルにまとめたときの「選択モデル」ダイアログボックス



新しいダイアログボックスが開き、プロファイルテーブルにすべてのデータが含まれた形式の分析かどうかを確認するメッセージが表示されます。

4. ［**はい**］をクリックしてモデルをあてはめます（図 5.30）。

図 5.30 ピザのデータを 1 つのデータテーブルにまとめたときの選択モデル



選択モデル

パラメータ推定値

項	推定値	標準誤差
生地[Thick]	0.096999596	0.1390573555
チーズ[Jack]	-0.916515667	0.1678404174
トッピング[Pepperoni]	-0.149373284	0.1366905640

AICc	136.8255
BIC	145.18804
(-2)*対数尤度	130.63195
(-2)*Firth対数尤度	119.11973

勾配で収束しました

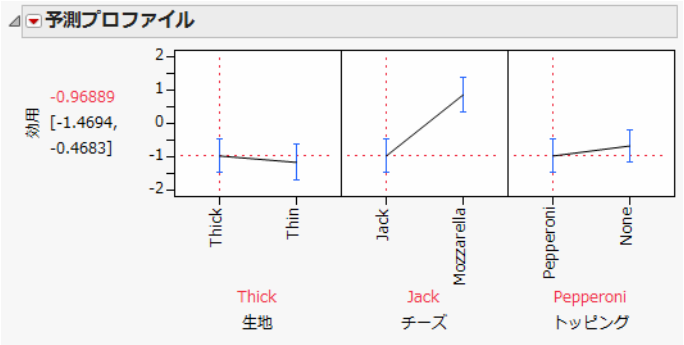
Firthバイアス調整推定値

効果の尤度比検定

要因	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
生地	0.500	1	0.4796
チーズ	45.626	1	<.0001*
トッピング	1.248	1	0.2639

5. ドロップダウンメニューから［**プロファイル**］を選択すると、図5.31のようなプロファイルが表示されます。パラメータ推定値と尤度比検定の結果は、テーブルを2つ使った場合（図5.7）と同じです。

図5.31 ピザのデータを1つのデータテーブルにまとめたときの予測プロファイル



セグメント化

マーケットリサーチ（市場調査）の分析では、消費者がもつ選好のパターンによって、グループに分けたい場合があります。しかし、通常の推定値を使ってグループ分けをするには、データが足りないのが普通です。もし、十分にデータがあるなら、「応答データ」セクションで［By］変数を指定するか、被験者IDをモデル項として含めることができます。ただし、この方法には、被験者の数が多いとうまく計算できません。それに代わる他の手法として、統計の文献では、Bayesモデルや混合モデルといった方法が紹介されています。

JMPでは、Bayesモデルや混合モデルといった手法とは別の方法で、被験者をクラスターに分ける機能が用意されています。たとえば、「Pizza Profiles.jmp」、「Pizza Responses.jmp」、そしてオプション指定の「Pizza Subjects.jmp」を使ってモデルを実行した後、「選択モデル」プラットフォームのドロップダウンメニューを開き、［**被験者ごとの勾配を保存**］をクリックします。新しいデータテーブルに被験者ごとの行が作成され、各パラメータの平均ステップ量が保存されます。

注：この機能は、試験的に導入されているものです。実用面での有用性について、ほとんど研究されていません。

JMPで算出される値は、Newton-Raphson法のステップ（反復計算で1回に進むベクトル）を被験者ごとに平均したものです。対数尤度の最大化が行われた段階での勾配の合計は0ベクトルになっています。また、この勾配の合計にヘッセ行列の逆行列をかけたもの（Newton-Raphson法のステップ幅）も0ベクトルです。

$\Delta = H^{-1}g = 0$。ここで、gは、最尤法の対数尤度から計算された勾配の合計です。

H^{-1}はヘッセ行列の逆行列です。ヘッセ行列とは、対数尤度の二階偏導関数（対数尤度を2回編微分したもの）を要素とする行列のことです。

ここで、Δを分解すると次のように表わされます。

$$\Delta = \Sigma_{ij}\Delta_{ij} = \Sigma H^{-1} g_{ij} = 0$$

上式において、iは被験者を示す番号、jは各被験者における選択を示す番号、

Δijは、Newton-Raphson法におけるステップ、

g_{ij}は、各選択における対数尤度の勾配を示します。

そこで、各被験者の平均ステップは、次のように計算できます。

$$\bar{\Delta}_i = \Sigma_j \frac{\Delta_{ij}}{n_i}$$ この式で、n_iは$_i$番目の被験者が行った選択の回数を示します。

$\bar{\Delta}_i$は、被験者iがパラメータに与える影響力に関係しています。

あるセグメントに属する被験者の選好構造が、他の被験者と異なる場合には、そのセグメントの被験者がパラメータに与える影響力は強くなります。よって、この平均ステップをクラスター分析に利用できると考えられます。

$\bar{\Delta}_i$はこの影響力を示し、データテーブルに保存されます。

この被験者ごとの勾配を含んだデータテーブルの一部を図5.32に示します。

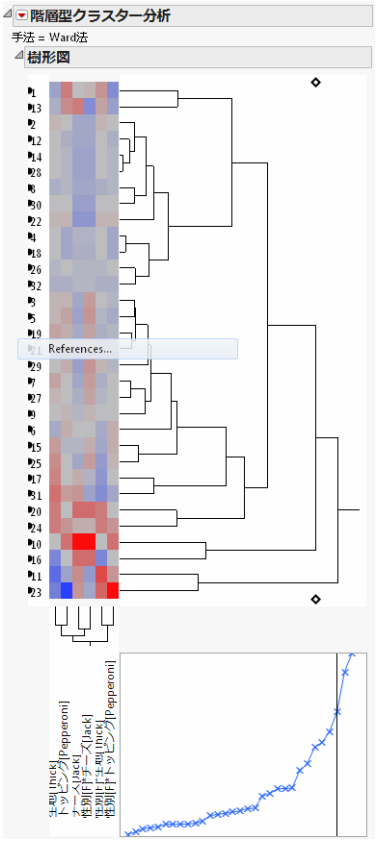
図5.32 ピザのデータから計算された被験者ごとの勾配

無題
階層型クラスター分析
元データに結果をマージ
列(7/0)
被験者
生地[Thick]

	被験者	生地[Thick]	チース[Jack]	トッピング [Pepperoni]	性別[F]*生地[Thick]	性別[F]*チース[Jack]	性別[F]*トッピング [Pepperoni]
1	1	-0.00959	-0.00168	0.014876	0.009585	0.001685	-0.01488
2	2	0.002373	-0.00758	-0.00239	0.002373	-0.00758	-0.00239
3	3	0.002129	-0.0079	0.003031	-0.00213	0.007899	-0.00303
4	4	-0.00106	-0.00485	-0.00901	-0.00106	-0.00485	-0.00901
5	5	0.002828	-0.00945	0.00725	-0.00283	0.009453	-0.00725
6	6	-0.0073	-0.00089	0.003761	-0.0073	-0.00089	0.003761

保存された勾配によって被験者をクラスターに分けるには、データに付随している「**階層型クラスター分析**」スクリプトのドロップダウンメニューを開き、［**スクリプトの実行**］をクリックします。図5.33のようなクラスターの樹形図が作成されます。

図 5.33 ピザのデータで作成された被験者クラスターの樹形図



ここで、樹形図の上または下にあるひし形のボタンを動かし、任意の数のクラスターを選択します。または、プラットフォームのドロップダウンメニューにある［**クラスターの数**］を選択し、数を入力することもできます。クラスター **ID** を保存するには、階層型クラスター分析のドロップダウンメニューを開き、［**クラスターの保存**］をクリックします。勾配を保存したデータテーブルに「**クラスター**」という列が追加されます。同じぐらいの勾配を持つ被験者が、同じクラスターに分けられます。階層型クラスター分析のその他のオプションについては、『多変量分析』の「クラスター分析」プラットフォームを参照してください。勾配の列は、クラスター分けのために作成したものなので、削除してかまいません。削除すると、データテーブルの列が「**被験者**」と「**クラスター**」だけになります。

図 5.34 クラスターを元のデータテーブルにマージ

	被験者	クラスター
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1

被験者ごとの勾配を保存したデータテーブル（図5.32）にある**「元データに結果をマージ」**のドロップダウンメニューを開き、［**スクリプトの実行**］をクリックすると、クラスター情報が「被験者」データテーブルに挿入されます。

これで、「被験者」データテーブルの列は**「被験者」**、**「性別」**、**「クラスター」**の3つになりました（図5.35）。

図 5.35「被験者」データテーブルに加わった「クラスター」列

	被験者	性別	クラスター
1	1	M	1
2	2	F	1
3	3	M	1
4	4	F	1
5	5	M	1
6	6	F	1

このデータテーブルを元に、分析を続けることができます。たとえば、［**分析**］>［**二変量の関係**］を選択します。次に、**「性別」**を［**Y, 応答変数**］、**「クラスター」**を［**X, 説明変数**］に指定します。ピザの例で作成した分割表は、図5.36のようになります。

図 5.36 ピザの例で行った、「性別」と「クラスター」の分割表に対する分析

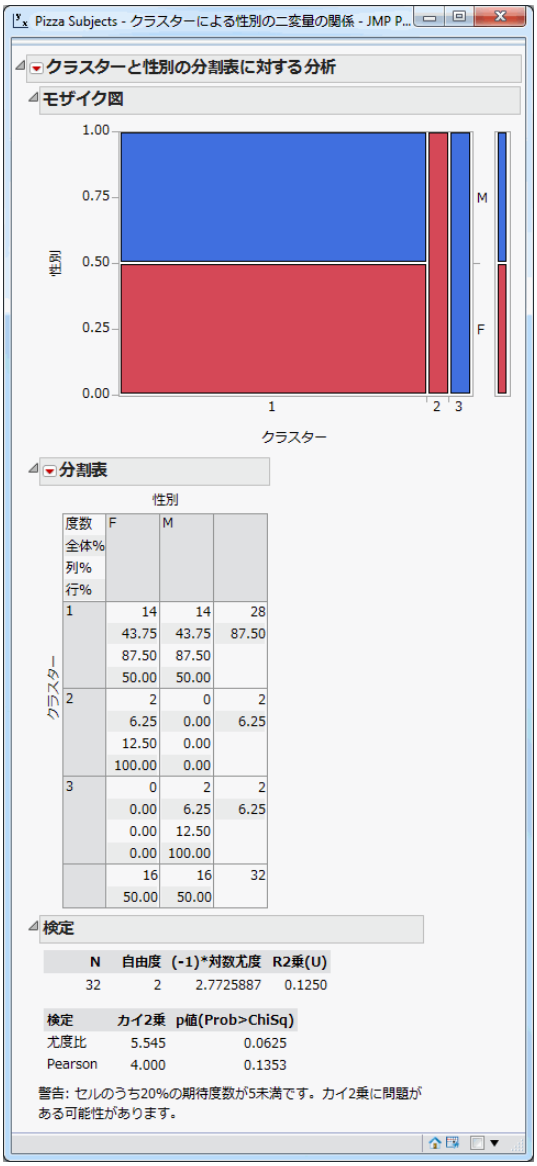


図5.36から、「クラスター」1は女性と男性が半々、「クラスター」2は女性のみ、「クラスター」3はすべて男性であることがわかります。場合によっては、**「クラスター」**変数を追加して、モデルを再分析してもよいでしょう。

データに関する特殊規則

デフォルトの選択肢集合

すべての質問において、プロファイルデータのすべてが選択肢に含まれているのであれば、「応答データ」セクションでの「**役割変数の選択**」パネルにある［**選択肢のプロファイル ID**］の指定を省略することができます。省略した場合、プロファイルデータに含まれているすべての選択肢が、すべての質問で被験者に提示されているとみなされます。

応答データを含んだ被験者データ

応答データテーブルに被験者の情報も一緒に保存している場合は、別々に保存している場合と同じように、「被験者データ」セクションの［**データテーブルの選択**］でそのデータテーブルを指定します。この場合、［**被験者 ID**］列は指定しません。被験者 ID が不要だからです。一般的には、被験者の属性データは、同一被験者であれば質問が変わっても変化しないでしょう。

ロジスティック回帰

「選択モデル」プラットフォームでは、通常のロジスティック回帰も実行できます。

注：ロジスティック回帰モデルをあてはめるには、「二変量の関係」または「モデルのあてはめ」プラットフォームの方が便利です。ただし、必要であれば「選択モデル」プラットフォームでロジスティック回帰を実行することもできるので、この節でその方法を紹介します。

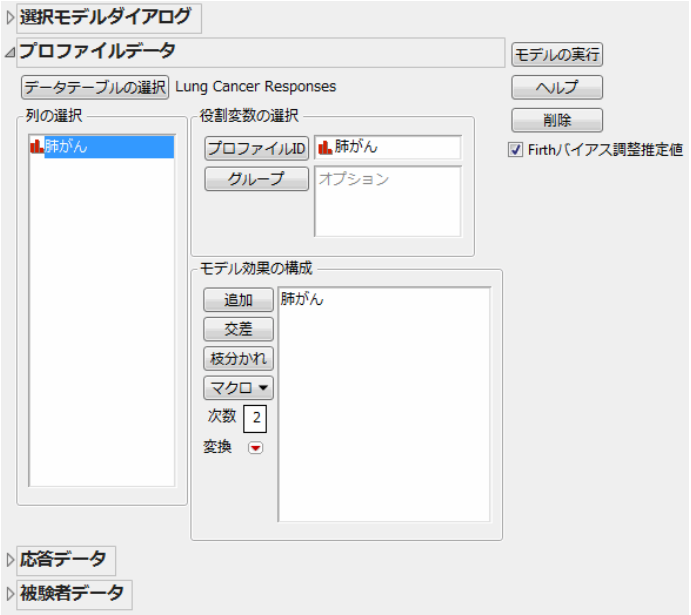
「選択モデル」によってロジスティック回帰分析を実行するには、まず、データが選択モデル用の形式になっている必要があります。データの準備として、次の 3 つの作業が必要となります。

- ロジスティック回帰における応答変数の各水準を、1 行ずつに収めたプロファイルデータテーブルを作成します。このデータテーブルには、応答変数の水準のみを含めます。
- 説明変数を含んだデータテーブルを保存し、それを応答データに指定します。
- 応答データテーブルを再び、今度は被験者データとして指定します。

「選択モデル」プラットフォームを使ってロジスティック回帰を実行する方法は、次のとおりです。

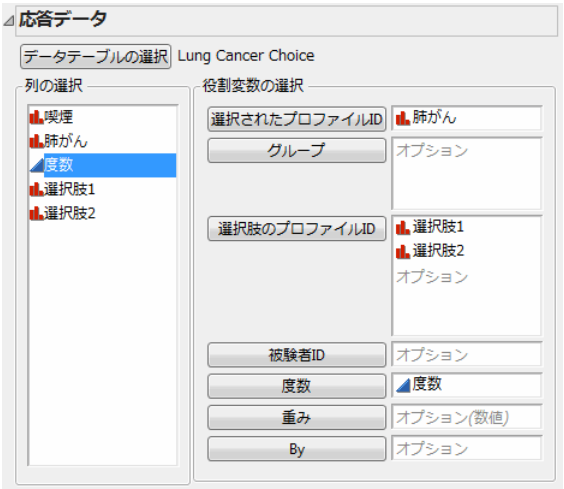
1. ［**分析**］>［**消費者調査**］>［**選択モデル**］>［**データテーブルの選択**］>［**その他**］>［**OK**］の順にクリックします。
2. 「Lung Cancer Responses.jmp」を開きます。このデータテーブルは、1 列（「**肺がん**」）、2 行（「**Cancer**」、「**NoCancer**」）しかありません。
3. 「**肺がん**」を［**プロファイル　ID**］に指定し、また、同じく「**肺がん**」をモデル効果にも［**追加**］します。「プロファイルデータ」セクションは、図 5.37 のようになります。

図 5.37 肺がんの例のプロファイルデータ



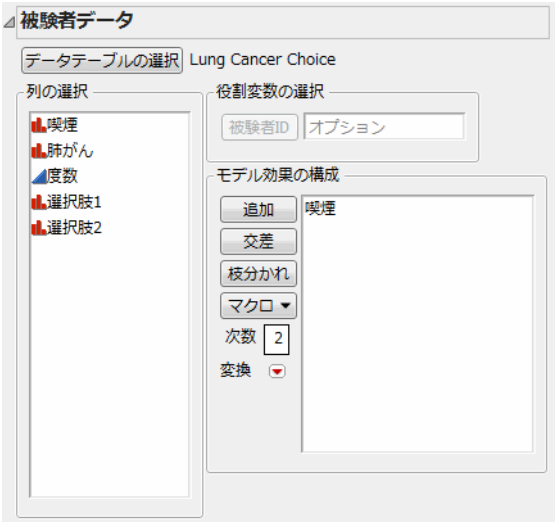
4. 「**応答データ**」の開閉アイコンをクリックし、［**データテーブルの選択**］>［**その他**］>［**OK**］の順にクリックします。
5. 「Lung Cancer Choice.jmp」を開きます。
6. 「**肺がん**」を［**選択されたプロファイル ID**］、「**選択肢 1**」と「**選択肢 2**」を［**選択肢のプロファイル ID**］、「**度数**」を［**度数**］に指定します。「応答データ」セクションは、図 5.38 のようになります。

図 5.38 肺がんの例の応答データ



7. 「**被験者データ**」の開閉アイコンをクリックし、［**データテーブルの選択**］> ［Lung Cancer Choice.jmp］> ［**OK**］の順にクリックします。
8. 「**喫煙**」をモデル効果として追加します。「被験者データ」セクションは、図 5.39 のようになります。

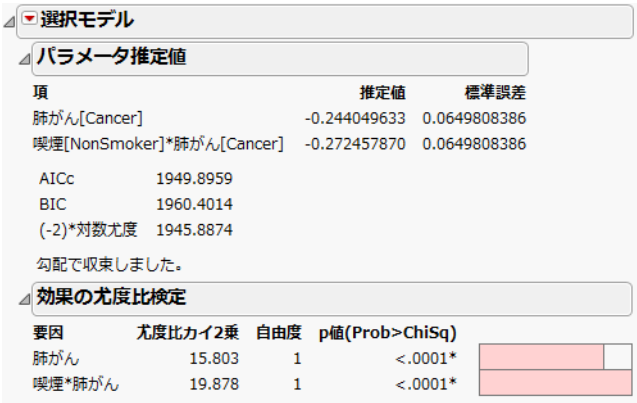
図 5.39 肺がんの例の被験者データ



9. ［**Firth バイアス調整推定値**］のチェックマークを外し、［**モデルの実行**］をクリックします。

   選択モデルは図 5.40 のようになります。

図 5.40 肺がんのデータと「選択モデル」を使ったロジスティック回帰



これらの結果を「モデルのあてはめ」プラットフォームで行ったロジスティック回帰と比べてみましょう。

1. サンプルデータのフォルダにある「Lung Cancer.jmp」を開きます。

2. ［**分析**］>［**モデルのあてはめ**］を選択します。自動的に、「**肺がん**」が［*Y*］、「**度数**」が［**度数**］に指定され、［**モデル効果の構成**］に「**喫煙**」が追加されます。また、［**名義ロジスティック**］の手法が自動的に指定されています。
3. ［**実行**］をクリックします。名義ロジスティックのあてはめは、図5.41のようになります。

図5.41 肺がんのデータと「モデルのあてはめ」を使ったロジスティック回帰

名義ロジスティックのあてはめ 肺がん

勾配で収束しました。4回の反復
度数: 度数

反復回数

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	9.93901	1	19.87802	<.0001*
完全	972.94369			
縮小	982.88270			

R2乗(U)	0.0101
AICc	1949.9
BIC	1960.4
オブザベーション(または重みの合計)	1418

指標	学習	定義
エントロピーR2乗	0.0101	1-Loglike(model)/Loglike(0)
一般化R2乗	0.0186	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
平均 -Log p	0.6861	Σ -Log(ρ[j])/n
RMSE	0.4966	√ Σ(y[j]-ρ[j])²/n
平均 絶対偏差	0.4933	Σ |y[j]-ρ[j]|/n
誤分類率	0.4732	Σ (ρ[j]≠ρMax)/n
N	1418	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	-0.4880993	0.1299617	14.11	0.0002*
喫煙[NonSmoker]	-0.5449157	0.1299617	17.58	<.0001*

推定値は次の対数オッズに対するものです： Cancer/NoCancer

推定値の共分散

効果の尤度比検定

要因	パラメータ数	自由度	尤度比カイ2乗	p値(Prob>ChiSq)
喫煙	1	1	19.878016	<.0001*

選択モデルにおける「喫煙*肺がん」の尤度比カイ2乗検定が、ロジスティックモデルにおける「喫煙」の尤度比カイ2乗検定と一致しています。図5.40と図5.41のレポートから、喫煙が肺がんの発病に対して強い効果を持つことがわかります。詳細については、『基本的な回帰モデル』の「ロジスティック回帰」の章を参照してください。

データの変換

データを 2 つの分析テーブルに変換する

データは、応答、プロファイル、被験者の 3 つに分かれているケースがほとんどですが、別の形式で保存されている場合は、選択モデル分析に適した形式に変換しなければなりません。たとえば、「Daganzo Trip.jmp」に保存されている Daganzo のデータを見てみましょう。このデータテーブルには、3 通りの交通手段のそれぞれにかかる時間と、各被験者が選択した交通手段が記録されています。図 5.42 はその一部です。

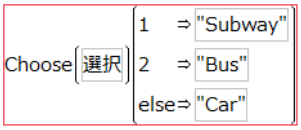
図 5.42 3 つの交通手段を使った Daganzo の移動時間（一部）

	Subway	Bus	Car	選択
1	16.481	16.196	23.89	2
2	15.123	11.373	14.182	2
3	19.469	8.822	20.819	2
4	18.847	15.649	21.28	2
5	12.578	10.671	18.335	2

まず、**「選択」**の番号を交通手段の名称に変換する必要があります。これは、計算式エディタにある **Choose** 関数を使えば、次のような手順で簡単にできます。

1. データテーブルに列を新規作成し、その列の名前を**「選択した交通手段」**とします。尺度として［**名義**］を指定します。**「選択した交通手段」**列を右クリックし、［**計算式**］を選択します。
2. ［**関数 (グループ別)**］の中から［**条件付き**］を選択し、［**Choose**］をクリックします。次に、カンマキーを 2 回押して関数に引数を追加します。
3. 「式」の枠をクリックした後、**「選択」**列をクリックして Choose の計算式（「式」）に挿入します。また、「節」入力ボックスをダブルクリックして「**"Subway"**」、「**"Bus"**」、「**"Car"**」と（引用符も含めて）入力します（図 5.43）。

図 5.43 Daganzo のデータの「選択した交通手段」列に Choose 関数を適用



これで、選択肢の応答が適切な形式になりました。

4. 各行には、各被験者によって選択された選択肢が含まれています。次に、各被験者の識別番号が含まれている列を作成します。それには、まず**「被験者」**という名前の列を作成します。次に、列の最初の行に「1」、第 2 行に「2」を入力します。最後に第 1 行と第 2 行を選択して強調表示し、右クリックして［**初期値を埋める / 挿入**］> ［**テーブルの最後までシーケンスを継続する**］を選択します。変更後のテーブルは、図 5.44 のようになります。

図 5.44 Daganzo のデータに「選択した交通手段」と「被験者」の列を追加

	Subway	Bus	Car	選択	選択した交通手段	被験者
1	16.481	16.196	23.89	2	Bus	1
2	15.123	11.373	14.182	2	Bus	2
3	19.469	8.822	20.819	2	Bus	3
4	18.847	15.649	21.28	2	Bus	4
5	12.578	10.671	18.335	2	Bus	5
6	11.513	20.582	27.838	1	Subway	6

プロファイルデータには、選択肢をそれぞれ 1 行ずつに収める必要があります。

5. ［**テーブル**］>［**列の積み重ね**］をクリックし、図 5.45 に従ってダイアログボックスへの入力を行います。変更後のデータテーブルを今後の分析に利用できるように、「Stacked Daganzo.jmp」のような名前を付けて保存します。［**OK**］をクリックします。最終的なデータテーブルは、図 5.46 のようになります。

図 5.45 Daganzo のデータで行う積み重ね

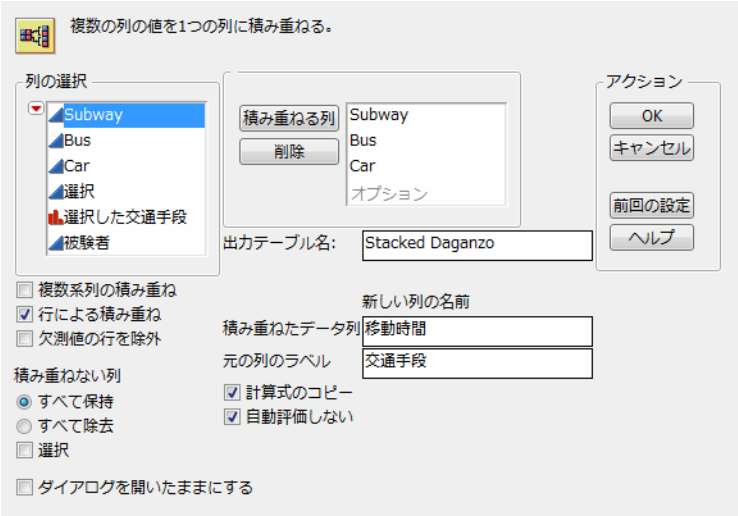


図 5.46「Stacked Daganzo」データテーブル（一部）

	選択	選択した交通手段	被験者	交通手段	移動時間
1	2	Bus	1	Subway	16.481
2	2	Bus	1	Bus	16.196
3	2	Bus	1	Car	23.89
4	2	Bus	2	Subway	15.123
5	2	Bus	2	Bus	11.373
6	2	Bus	2	Car	14.182
7	2	Bus	3	Subway	19.469

6. 積み重ねたデータから、「**被験者**」、「**交通手段**」、「**移動時間**」だけで構成されたサブセットを作成するため、これらの列を選択し、［**テーブル**］>［**サブセット**］の順にクリックします。［**選択されている列**］を選択し、［**OK**］をクリックします。図 5.47 は、データの一部です。

図 5.47 Stacked Daganzo のデータのサブセット（一部）

	被験者	交通手段	移動時間
1	1	Subway	16.481
2	1	Bus	16.196
3	1	Car	23.89
4	2	Subway	15.123
5	2	Bus	11.373
6	2	Car	14.182
7	3	Subway	19.469

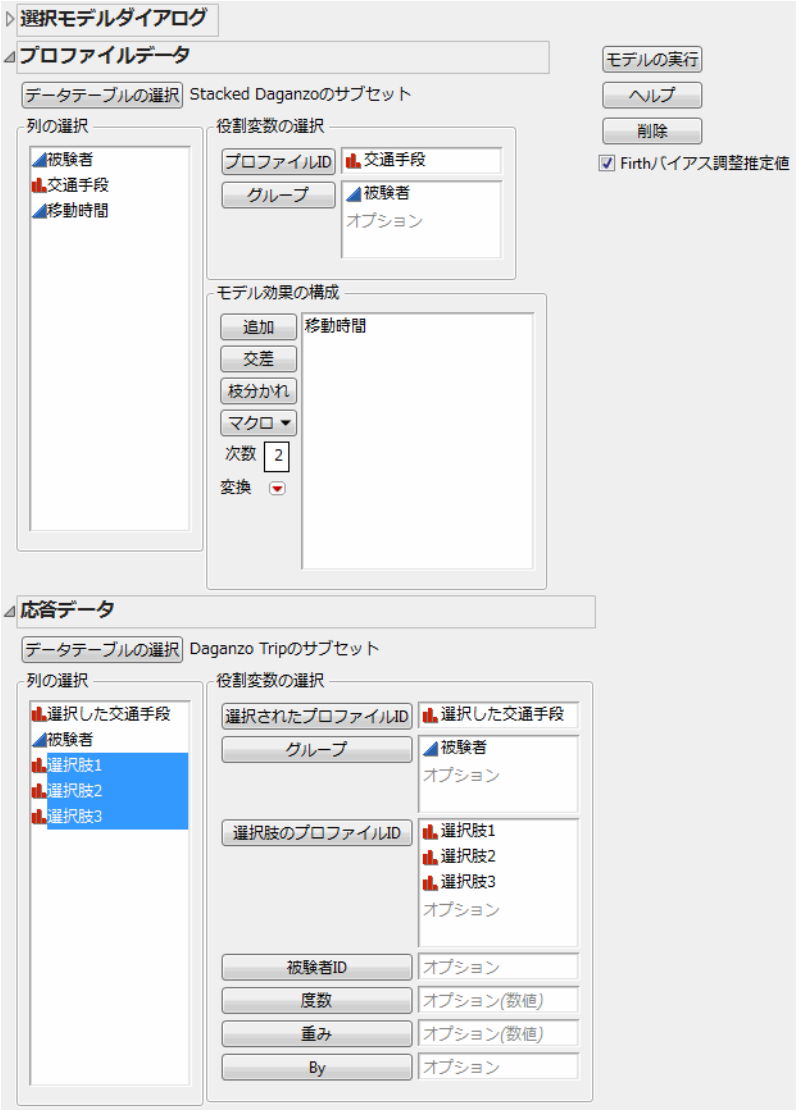
7. 今度は、元のデータ（「Daganzo Trip」）から、「**被験者**」と「**選択した交通手段**」だけで構成されたサブセットを作成します。次に、選択肢集合の列として「**選択肢 1**」、「**選択肢 2**」、「**選択肢 3**」の 3 つを追加し、図 5.48 に従ってそれぞれの初期値を定数に指定します。

図 5.48 Daganzo のデータのサブセットに選択肢集合を追加

	選択した交通手段	被験者	選択肢1	選択肢2	選択肢3
1	Bus	1	Bus	Subway	Car
2	Bus	2	Bus	Subway	Car
3	Bus	3	Bus	Subway	Car
4	Bus	4	Bus	Subway	Car
5	Bus	5	Bus	Subway	Car
6	Subway	6	Bus	Subway	Car
7	Subway	7	Bus	Subway	Car

8. 図 5.49 のとおりにモデルを指定します。

図 5.49 Daganzo のデータのサブセットを使ったときの「選択モデル」ダイアログボックス



9. モデルを実行します。計算されたパラメータ推定値は、効用関数における「**移動時間**」の係数を表します（図 5.50）。

図 5.50 Daganzo のデータの「移動時間」に対するパラメータ推定値

選択モデル

パラメータ推定値

項	推定値	標準誤差
移動時間	-0.341768586	0.0745222259

AICc	68.766653
BIC	70.595342
(-2)*対数尤度	66.683319
(-2)*Firth対数尤度	61.490004

勾配で収束しました

Firthバイアス調整推定値

効果の尤度比検定

要因	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
移動時間	43.178	1	<.0001*

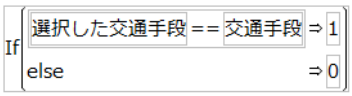
係数が負の値になっているのは、移動時間の増加が消費者の効用（満足度）に対して負の効果を持つことを表します。尤度比検定において、「移動時間」の効果は有意になっています。

データを 1 つの分析テーブルに変換する

2～3つのデータテーブルを作成するより、1つのデータテーブルにまとめた方が効率的な場合があります。1つのデータテーブルにする場合、被験者の列は先ほどの例と同じ手順で追加します。また、選択肢の集合として3つの列（選択肢1、選択肢2、選択肢3）を使用する代わりに、応答の指示変数（応答のインジケータの列）を作成します。1つのテーブルにまとめる場合の変換は、次のような手順で行います。

1. 「データを2つの分析テーブルに変換する」（109ページ）の手順に従って「Stacked Daganzo.jmp」を作成するか、すでに作成したものを開きます。
2. **「応答」**という名前の列を新しく作成し、列を右クリックします。［**計算式**］を選択します。
3. 計算式エディタで［**条件付き**］>［**If**］をクリックし、**「選択した交通手段」**列をクリックして式に挿入します。
4. 「=」を入力して**「交通手段」**を選択します。
5. ［**then 節**］に「1」を入力し、［**else 節**］に「0」を入力します。［**OK**］をクリックします。入力が完了した計算式は、図5.51のようになります。

図 5.51 指示変数を作成する計算式



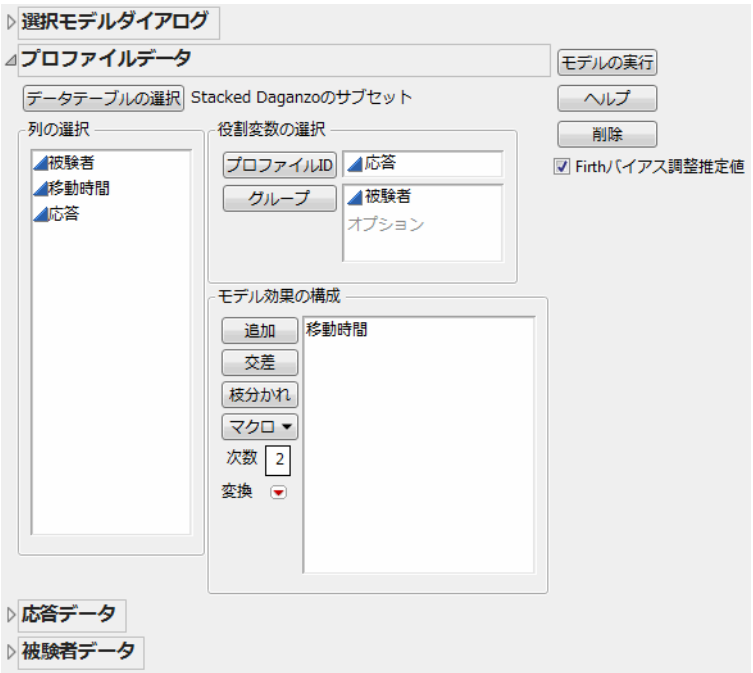
6. データテーブルのサブセットを作成するため、**「被験者」**、**「移動時間」**、**「応答」**を選択してから、［**テーブル**］>［**サブセット**］をクリックします。［**選択されている列**］を選択し、［**OK**］をクリックします。新しいデータテーブルは、図5.52のようになります。

図 5.52「Stacked Daganzo」データテーブルのサブセット（一部）

	被験者	移動時間	応答
1	1	16.481	0
2	1	16.196	1
3	1	23.89	0
4	2	15.123	0
5	2	11.373	1
6	2	14.182	0
7	3	19.469	0

7. ［**分析**］>［**消費者調査**］>［**選択モデル**］の順にクリックして起動ダイアログボックスを開き、図 5.53 のようにモデルを指定します。

図 5.53「Stacked Daganzo」データのサブセットを使い、テーブル 1 つで分析するときの「選択モデル」ダイアログボックス



8. ［**モデルの実行**］を選択します。ポップアップウィンドウが開き、1 つのプロファイルテーブルにすべてのデータが含まれた形式の分析かどうかを確認するメッセージが表示されます。［**はい**］をクリックすると、効用関数における「**移動時間**」の係数を表すパラメータ推定値が計算されます（図 5.54）。

図 5.54 1つのテーブルで分析したときの「移動時間」のパラメータ推定値

選択モデル

パラメータ推定値

項	推定値	標準誤差
移動時間	-0.341768586	0.0745222259

AICc	68.766653
BIC	70.595342
(-2)*対数尤度	66.683319
(-2)*Firth対数尤度	61.490004

勾配で収束しました

Firthバイアス調整推定値

効果の尤度比検定

要因	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
移動時間	43.178	1	<.0001*

結果は、テーブルを2つ使って分析した場合（図5.50）と同じです。

この章では、単純な例を使って「選択モデル」プラットフォームの機能を紹介しましたが、もっと複雑な変換や交互作用項を含む、より高度なモデルも実行できます。

対応のあるデータに対する条件付きロジスティック回帰

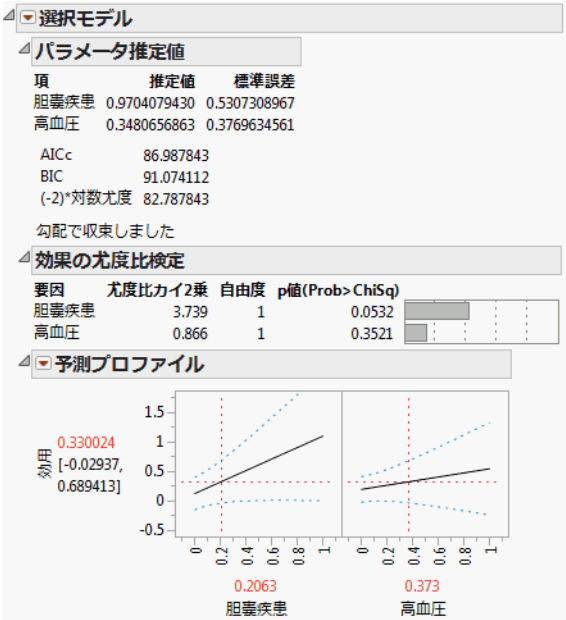
ここでは、「選択モデル」プラットフォームを使い、「ケース（症例）」と「コントロール（対照）」とでペアごとに対応している子宮内膜がんの調査結果に対して、条件付きロジスティック回帰分析を行う例を紹介します。使用するデータは、Breslow and Day（1980）およびSAS/STAT(R) 9.2 User's Guide, Second Edition（2006）で取り上げられているLos Angeles Study of the Endometrial Cancer Dataからの引用です。このケースコントロール分析の目標は、高血圧による影響を考慮しながら、胆嚢疾患の相対リスクを推定することです。「アウトカム」の1の値は、子宮内膜がんの発症（ケース群）を示し、0はコントロール群を示します。胆嚢疾患と高血圧も、同様に1と0で示されています。分析は次の手順に従って実行します。

1. 「Endometrial Cancer.jmp」を開きます。
2. ［**分析**］>［**消費者調査**］>［**選択モデル**］を選択します。
3. ［**データテーブルの選択**］ボタンをクリックします。
4. プロファイルデータテーブルとして「Endometrial Cancer」を選択します。
5. 「**アウトカム**」に［**プロファイル ID**］の役割を割り当てます。
6. 「**ペア**」に［**グループ**］の役割を割り当てます。
7. 「**胆嚢疾患**」と「**高血圧**」をモデル効果の構成として追加します。
8. ［**Firth バイアス調整推定値**］のチェックマークを外します。
9. ［**モデルの実行**］を選択します。
10. テーブル1つにすべてのデータが含まれているかを尋ねられたら［**はい**］を選択します。

11.「選択モデル」の赤い三角ボタンのメニューから［**プロファイル**］を選択します。

図 5.55 のようなレポートが作成されます。

図 5.55 子宮内膜がんデータのロジスティック回帰



要因ごとに尤比度検定が実行されています。**「胆嚢疾患」**が、α 水準を 0.05 としたとき、統計的にほぼ有意であることがわかります（p 値は 0.0532）。予測プロファイルを使い、応答に対する要因の影響を視覚的に確認しましょう。

統計的詳細

選択モデルでは、消費者の**効用**（utility）がパラメータ推定値から算出されます。「効用」とは、特定の属性をもつ製品から消費者が得る満足度のことで、モデルのパラメータ推定値から算出されます。なお、効用が線形関数で表されている場合、パラメータ推定値は限界効用（marginal utility）を示します。

選択モデルを式で表わす場合

$X[k]$ を、切片のある被験者属性の計画行とし、

$Z[j]$ を、切片のない選択属性の計画行とします。

この場合、第 k 被験者が、m 個の選択肢のなかから、j 番目のものを選択する確率は、次式で表されます。

$$P_i[jk] = \frac{\exp(\beta'(X[k] \otimes Z[j]))}{\sum_{l=1}^{m} \exp(\beta'(X[k] \otimes Z[l]))}$$

この式で、

– $\otimes$ は、Kronecker 積です。

– 分子は、実際に選択された j 番目の選択肢に対する値です。

– 分母は、その質問で被験者に提示された m 個の選択肢すべての合計を示します。

第 6 章

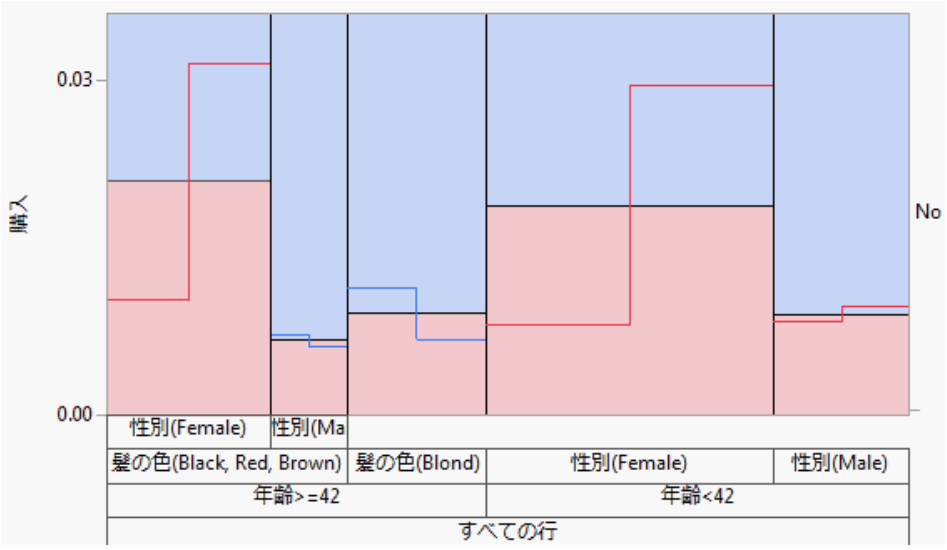
アップリフトモデル

アクションが消費者行動に及ぼす効果をモデル化

JMP PRO

アップリフトモデルは、特定のアクションに反応する可能性が高い個人の集団を特定するのに役立ちます。マーケティングキャンペーンの最適化や、個人の治療プロトコルの作成などでの応答が考えられます。アップリフトモデルは、増分モデル（incremental model）、真のリフトモデル（true lift model）、ネットモデル（net model）などとも呼ばれています。アップリフトモデルは、処置とその他の変数との間の交互作用を見つていくという点で従来の手法とは異なります。アクションや処置（treatment；処理、治療）に対し、ポジティブに反応すると予測される集団をデータから探し出します。

図 6.1 ヘアケア製品の販売促進のためのアップリフトモデル例



列の寄与

項	分岐数	寄与	割合
性別	2	42.4901339	0.4428
髪の色	1	39.236122	0.4089
年齢	1	14.2342338	0.1483
米国地域	0	0	0.0000
居住地	0	0	0.0000

目次

「アップリフト」プラットフォームの概要 121
「アップリフト」プラットフォームの使用例 122
「アップリフト」プラットフォームの起動 123
「アップリフトモデル」レポート 124
アップリフトモデルのグラフ. 124
アップリフトレポートのオプション 127

「アップリフト」プラットフォームの概要

「アップリフト」プラットフォームは、キャンペーンなどの**処置**（treatment；処理、治療）が消費者に及ぼす増分的な影響をモデル化します。アップリフトモデルは、処置に反応すると予測される消費者グループを特定するのに役立ちます。そのような消費者グループを特定できれば、リソースの割り当てや消費者に与える影響を最適化できるでしょう（Radcliffe and Surry, 2011 を参照）。

「アップリフト」プラットフォームでは、パーティションモデルをあてはめます。従来のパーティションモデルは予測が最良となるような分岐を見つけます。一方、アップリフトモデルは処置の差を最大化する分岐を見つけます。

パーティションによるアップリフトモデルでは、処置が強く影響するグループと、そうでないグループが存在していると仮定します。それらのグループを探し出すために、考えられるすべての二分岐で分けられたデータのそれぞれに、線形モデルやロジスティックモデルをあてはめます。応答が連続尺度の場合には、分岐、処置、および、分岐と処置の交互作用の線形関数としてモデル化されます。応答がカテゴリカルな場合には、同様に、分岐、処置、および、分岐と処置の交互作用のロジスティック関数としてモデル化されます。どちらの場合も、交互作用項は、分岐によって分かれた 2 つの消費者グループ間における、アップリフトの差に相当します。

「アップリフト」プラットフォームで分岐の定義に使用される基準は、可能なすべての分岐に対する交互作用に対する検定の p 値です。ただし、p 値のみに基づいて説明変数を選択すると、水準が多い説明変数が有意になりやすいという多重性のバイアスが生じます。このため、JMP は、p 値に水準数を考慮した調整を加えています（詳細については、JMP Web サイト「Monte Carlo Calibration of Distributions of Partition Statistics」を参照してください）。「アップリフト」プラットフォームの分岐は、交互作用効果の t 検定の調整済み p 値を最大化することで決定されます。そして、それぞれの分岐に対しては、調整済み p 値の対数価値（$-\log_{10}$）がレポートされます。

「アップリフト」プラットフォームの使用例

「Hair Care Product.jmp」データテーブルは、毛染め製品に関する、仮想的なデータです。ある大手美容雑貨チェーンが、126,184 名の男女のカード会員に対して、毛染め製品を購入するかどうかを追跡しました。まず、カード会員の中から約半数を無作為に抽出し、毛染め製品の広告を送付しました。そして、その後3か月にわたって、カード会員全員を対象に、製品を購入するかどうかを追跡しました。

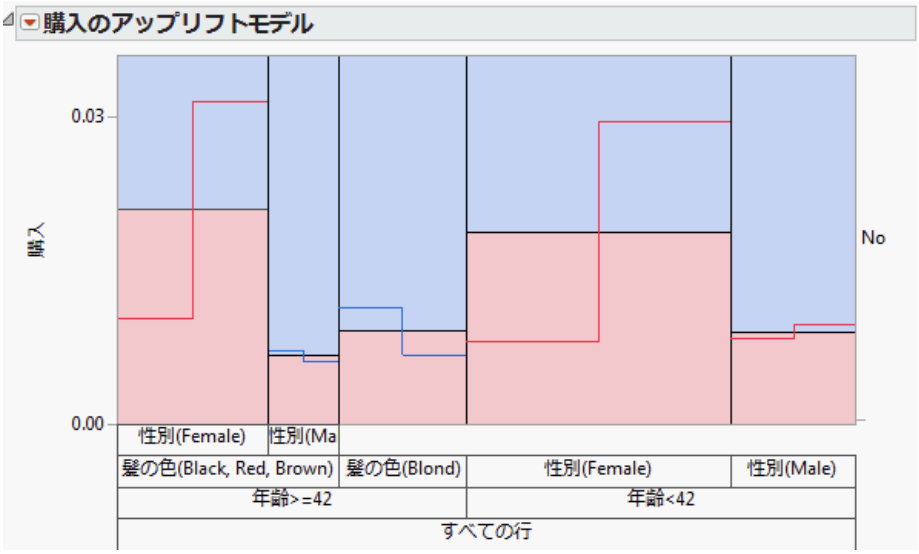
データテーブルの「**広告**」列は、広告が送付されたかどうかを示します。「**購入**」列は、製品を購入したかどうかを示します。各会員について、「**性別**」、「**年齢**」、「**髪の色**」、「**米国地域**」、「**居住地**」（都市部かどうか）の情報が収集されています。また、「**検証**」列によって、カード会員の約33%が検証データとして使われるように設定されています。

カテゴリカル応答の場合は、並び替えたときに順番が最初になっている水準が、関心のある水準として、JMP によって解釈されます。「**購入**」列に「値の順序」列プロパティが設定されているのは、「Yes」を最初の水準とするためです。この列プロパティを設定しないと、「No」が最初の水準となり、分析において関心のある水準として扱われてしまいます。

1. 「Hair Care Product.jmp」データテーブルを開きます。
2. ［**分析**］>［**消費者調査**］>［**アップリフト**］を選択します。
3. 「列の選択」リストから
   - 「**広告**」を選択し、「**処置**」をクリックします。
   - 「**購入**」を選択し、［**Y, 目的変数**］をクリックします。
   - 「**性別**」、「**年齢**」、「**髪の色**」、「**米国地域**」、「**居住地**」を選択し、［**X, 説明変数**］をクリックします。
   - 「**検証**」列を選択し、［**検証**］ボタンをクリックします。
4. ［**OK**］をクリックします。
5. レポートのグラフの下にある［**実行**］をクリックします。

   検証セットに基づき、最適な分岐の数が5と判断されます。図6.2のようなグラフが作成されます。わかりやすいように、縦軸のスケールを変更しています。

図 6.2 5 回の分岐後のグラフ

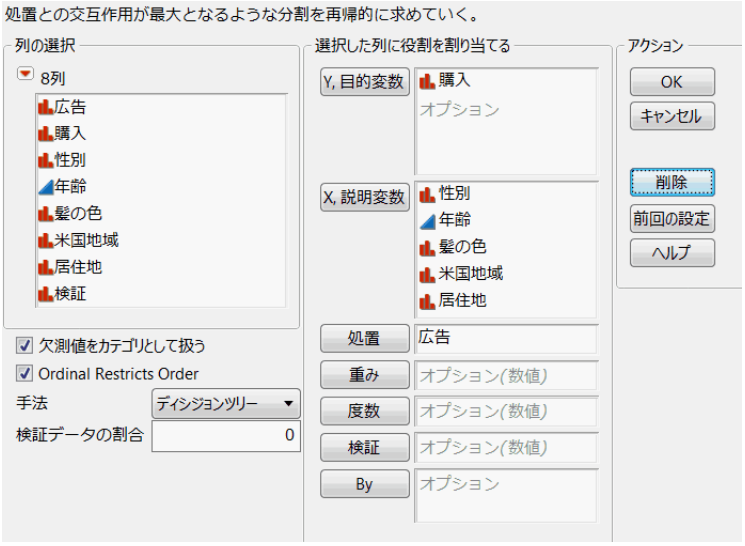


グラフでは、広告によって購入のアップリフトが生じているのは、年齢の低い層（**年齢** < 42）の男女と、年齢の高い層（**年齢** ≧ 42）の女性で、髪の色が黒（black）・赤（red）・茶色（brown）のいずれかであるグループであることがわかります。「**年齢** ≧ 42」のグループでは、ブロンド（blond）の男女とブロンドでない男性の両方で、広告の効果がマイナスとなっています。

「アップリフト」プラットフォームの起動

「アップリフト」プラットフォームを起動するには、［**分析**］>［**消費者調査**］>［**パーティション**］を選択します。図 6.3 は、「Hair Care Product.jmp」データテーブルの起動ウィンドウです。［Y, 目的変数］と［X, 説明変数］には、連続尺度の列とカテゴリカル列のどちらでも指定できます。［処置］には通常、2 水準のカテゴリカル列を指定します。［処置］列に 3 つ以上の水準がある場合、最初の水準が「処置 1」となり、残りの水準は「処置 2」としてまとめられます。

図 6.3 アップリフトの起動ウィンドウ



検証データを用いるには、予め検証データを示すフラグの列を用意しておき、その列を［検証］列に指定します。もしくは、「検証データの割合」に割合を指定すると、その割合だけ無作為にデータが抽出されて、検証データとして使われます。なお、「アップリフト」で現在サポートされている手法は「ディシジョンツリー」のみです。

「アップリフトモデル」レポート

最初の「アップリフトモデル」レポートには、グラフとツリーノード、そして、分岐などを行うためのボタンが表示されます。

アップリフトモデルのグラフ

アップリフトモデルのグラフにおける縦軸は、応答を示しています。そして、横軸は、分岐によって構成されたノードを表します。各ノードごとに描かれている黒色の横線は、そのノードの応答の平均を示しています。各ノードには、処置による副分岐を表す赤色または青色の線も描かれています。これらの赤もしくは青の線は、各ノード内での処置の副分岐によって構成される 2 つのグループの平均応答を示します。「処置」列におけるデータ値の並び順によって、赤色か、青色になるかは変化します。分岐を進めていくに従って、グラフも更新され、横軸の下に分岐の条件式が更新されていきます。縦の線は、分岐の分かれ目を表しています。

グラフの下に、［**分岐**］、［**剪定**］、［**実行**］というボタンがあります。なお、このうち、［実行］ボタンは、検証データを使用している場合にのみ表示されます。また、「処置」列の名前と、その列内の水準名も表示されます。「処置」列に 3 つ以上の水準がある場合、最初の水準以外はすべて「処置 2」という 1 つの水準にまとめられ、2 水準の応答として処理されます。

「処置」列情報の右側には、予測に関連する要約統計量がレポートされます（ただし、アップリフトモデルにおいては、最良の予測を行うことが分析目的ではない点に注意してください）。レポートは、分岐が実行されるたびに更新されます。検証データを使用している場合、学習データと検証データの両方に対する要約統計量が表示されます。

R2 乗　ツリーに関連する回帰モデルの R2 乗。この回帰モデルには、処置列との交互作用も含まれています。

N　オブザベーションの数。

分岐数　分岐が実行された回数。

AICc　回帰モデルを使って計算された、修正済みの赤池の情報量規準（AICc）。AICc は連続量の応答に対してのみ算出されます。

アップリフトのディシジョンツリー

アップリフトモデルで使用された分岐は、ツリーとして描画されます。「Hair Care Product.jmp」データテーブルを使った例は、図 6.4 を参照してください。描画されたツリーの各ノードには、次の情報が含まれます。

処置　「処置」列の列名と、その列に含まれている 2 つの水準が表示されます。

割合　応答変数が、2 水準のカテゴリカルの場合にのみ表示されます。各ノード内で、反応した被験者の割合が、処置の水準ごとに算出されています。

平均　応答変数が、連続尺度の場合にのみ表示されます。各ノード内で、応答の平均が、処置の水準ごとに算出されています。

度数　各ノードで、処置の水準ごとに、標本サイズ（たとえば、被験者の人数）が示されています。

t 値　処置の水準間における差を検定するときに使われる *t* 値。なお、応答がカテゴリカルな場合は、この統計量は、連続尺度（0 と 1）に変換された後に算出されます。

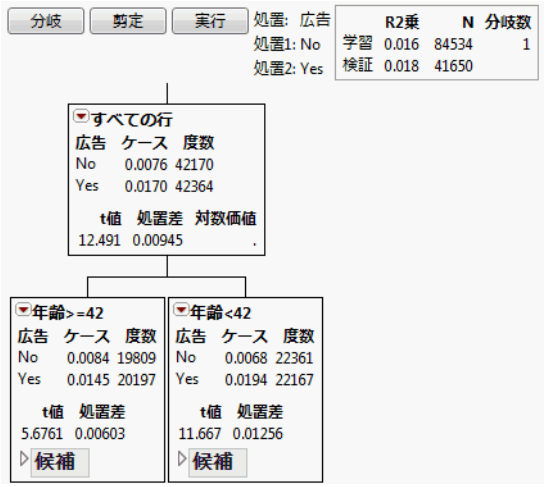
処置差　処置水準間の応答平均の差。これがアップリフトと呼ばれるものです。JMP の「アップリフト」プラットフォームでは、次のように仮定して計算されます。

– 「処置」列の値において、最初の水準が「処置」を表し、それ以外の水準は「対照」を表す。

– 応答変数の値は、値が大きいほど、ポジティブであることを示している。

対数価値　各ノードにおける、後続の分岐がもつ対数価値。

図 6.4 最初の分岐のノード



「候補」レポート

各ノードには「候補」レポートも含まれます。このレポートには、次のような情報が表示されます。

項 モデルの項。

対数価値 各項でのすべての分岐のうちで、分岐による統計量が最大となったときの対数価値。「対数価値」とは、$-\log_{10}$（p）で計算される（ここで、p は、調整済み p 値）

F 値 応答が連続尺度の場合は、線形回帰モデルの交互作用項に対する F 値。応答がカテゴリカルの場合は、名義ロジスティックモデルの交互作用項に対するカイ 2 乗値。回帰モデルは、処置、二分岐、および、それらの交互作用という 3 つの効果から構成されます。

ガンマ 応答が連続尺度の場合は、線形回帰モデルにおける交互作用の推定値。応答がカテゴリカルの場合は、名義ロジスティックモデルを Firth バイアス調整法で推定したときの、交互作用の推定値。

分岐点 項が連続尺度の場合は、分岐点が表示されます。項がカテゴリカルの場合は、最初（左）のノードに属するカテゴリが表示されます。

アップリフトレポートのオプション

「アップリフト」レポートの赤い三角ボタンのオプションについては、ここで紹介するオプションを除き、すべて「パーティション」プラットフォームのマニュアルで説明しています。これらのオプションについての詳細は、『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。

分岐の最小サイズ

許容する最小の標本サイズを、数または標本全体に占める割合として指定するダイアログボックスが開きます。度数（標本サイズ）として指定する場合は、1 以上の値を入力します。標本サイズ全体に対する割合として指定する場合は、1 未満の値を入力します。アップリフトでは、Max（25, Floor（n/2000））がデフォルト値として設定されます。ここで、n はデータの標本サイズです。Floor(x) は x 以下の一番大きな整数、Max(a, b) は a と b のうちの大きいほうの値です。

列の寄与

この表とプロットは、作成されたツリーに対する各列の寄与を表わしています。列の寄与は、分岐の F 値の合計として算出されます。F 値は、線形回帰モデルにおける、処置と分岐の交互作用項の有意性を示します。

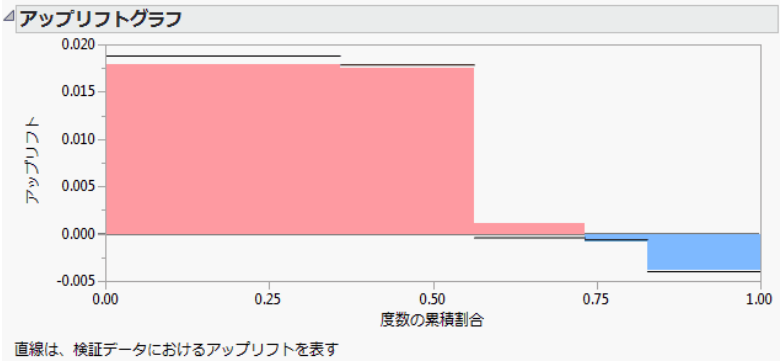
アップリフトグラフ

学習データから、各ノードのアップリフト値を求めます。アップリフト値は、応答がカテゴリカルな場合は、商品を購入する確率の処置間における差です。応答が連続尺度の場合は、処置間のおける平均差です。その予測値が、最終的なツリーに基づいて、各ノードに対して算出されます。各ノードごとのアップリフト値は、降順に並べて表示されます。アップリフトグラフでは、縦軸はアップリフト値です。横軸は、データの累積割合です。

図 6.5 は、「Hair Care Product.jmp」サンプルデータの、3 回の分岐の後、さらに「**髪の色**」グループの「**性別**」で分岐したアップリフトグラフの例です。2 つの被験者グループ（**年齢** ≥ 42 グループのブロンドでない男性とブロンドの男女）では、広告がマイナスの影響をもたらしていることがわかります。

アップリフトグラフにある横線は、検証データにおけるアップリフト値です。学習データから推定されたツリーが、検証データに対して評価され、アップリフト値の推定値が算出されます。

図 6.5 アップリフトグラフ



列の保存

差の保存　各ノードに対して計算された、処置の水準間における平均差の推定値を保存します。この値が、アップリフト値の推定値です。

差の計算式を保存　差（アップリフト）の計算式を保存します。

第 7 章

項目分析

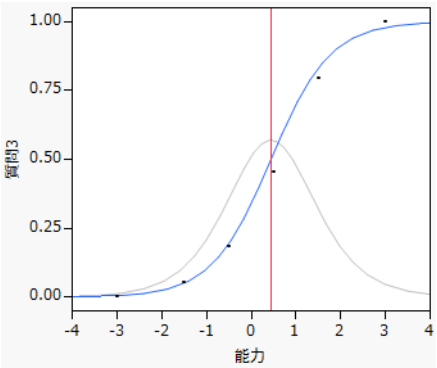
テスト結果に対する項目反応分析

項目反応理論（IRT; Item Response Theory）は、テスト（試験）を分析する方法です。およそ1世紀にわたって古典的なテスト理論が広く使用されてきましたが、項目反応理論は、より性質の良い統計的モデルに基づいて、各受験者や各テスト項目に対するスコア（得点）を求めます。

項目反応理論には、次のような利点があります。

- 各項目ごとにスコアを求めます。各項目が合計得点に占める割合が明らかになります。
- 受験者とテスト項目の両方に、同じスケールでスコアを求めます。
- 従来使用されていた線形式ではなく、非線形ロジスティック曲線がモデルで使われます。

図 7.1 項目分析の例



目次

「項目分析」プラットフォームの概要 131
「項目分析」プラットフォームの起動 134
「項目分析」レポート 136
特性曲線 136
情報曲線 137
デュアルプロット 137
「項目分析」プラットフォームのオプション 139
技術的詳細 139

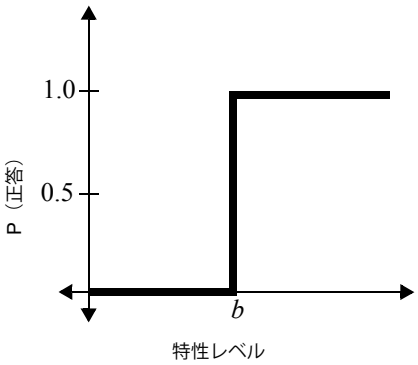
「項目分析」プラットフォームの概要

心理測定は、**心理的構成物**と呼ばれる個人または被験者の特性を、数値（量的な値）を割り当てて測定するプロセスです。その際、数値の割り当てに使用するルールの体系を**測定理論**と呼びます。項目反応理論（IRT）は、測定理論の1つです。

項目反応理論では、個人が項目に正しく反応する確率とその個人の特性との関係を、数学関数によって表します。この特性は、直接測定できない場合が多いことから、**潜在特性**と呼ばれます。

項目反応理論が特性をどのように確率と関連づけるかを調べるため、まず図7.2の「完全なガットマン尺度」に従ったテスト問題を検討してみましょう。横軸は、受験者が持つ理論上の特性の量を表します。縦軸は、受験者が項目に正答する確率を表します（テスト問題に答えなかった場合は、一般的に、誤答として扱われます）。図7.2の曲線は、**項目特性曲線**（ICC）といいます。

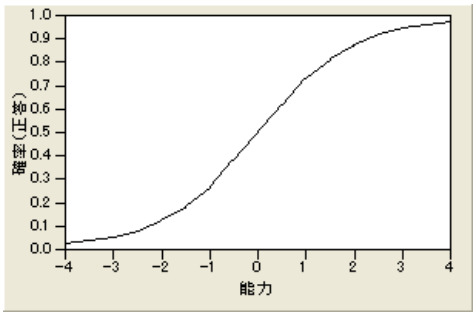
図7.2 完全なガットマン尺度に従う項目の項目特性曲線



この図から、特性の値がb未満である受験者がこの項目に正しく答える確率は、ゼロであることがわかります。特性がbを超えている受験者は、100パーセントの確率でこの項目に正しく答えられます。

もちろん、このような項目は現実的ではありませんが、特性とテスト項目の正答率がどのように関連するかを端的に表しています。より標準的な項目特性曲線では、確率が0～1の範囲で変化します。経験的に得られる典型的な曲線は、S字形のロジスティック関数で下方漸近線が0、上方漸近線が1になります。明らかに非線形であるのが特徴です。図7.3はその一例です。

図 7.3 項目反応曲線の例

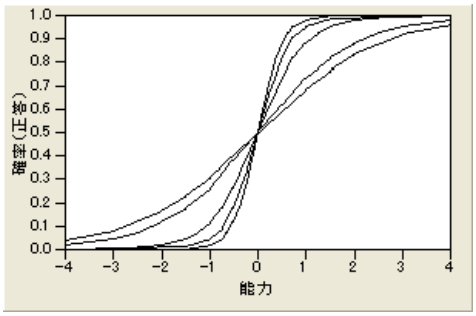


この曲線をモデル化するには、ロジスティックモデルが最適です。ロジスティックモデルは、好ましい漸近的性質を持ち、それでいて他に使用されるモデル（正規分布の累積分布関数など）よりも計算処理が簡単だという利点があります。モデル式は次のとおりです。

$$P(\theta) = c + \frac{1-c}{1+e^{-(a)(\theta-b)}}$$

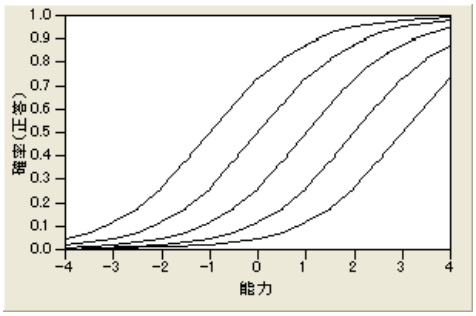
3パラメータロジスティック（**3PL**）モデルと呼ばれるこのモデルでは、変数 a が変曲点での曲線の角度を表します。a の値をさまざまに変化させた場合、曲線は、図7.4のようになります。このパラメータは、項目の識別力、つまり特性レベルの高い受験者と比べたとき、特性レベルの低い受験者にとってその項目がどれだけ難しいかを表すパラメータと解釈できます。a が非常に大きい値を取る場合、モデルの形状は、図7.2のステップ関数に近くなります。一般に、受験者の特性レベルが高ければ高いほど、項目に正答する確率は高くなります。そのため、a は正であると仮定され、項目特性曲線は単調に増加します。正の単調増加であるという属性によって、各テスト項目の適性を確認することもあります。その場合、曲線がS字形でない項目が、テストから除外する候補となります。

図 7.4 a の値を変化させたときのロジスティックモデル



b の値を変化させた場合、曲線は単に左右に移動するだけです（図7.5）。$P(\theta)=0.5$ の点における θ の値に相当します。そのため、b は、項目の困難度と解釈することができます。項目が難しければ難しいほど、変曲点はグラフの X 軸に沿って右方向に移動します。

図7.5 bの値を変化させたときのロジスティック曲線

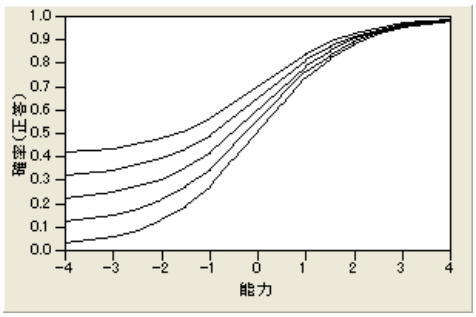


確率には、次のよう関係が成り立ちます。

$$\lim_{\theta \to -\infty} P(\theta) = c$$

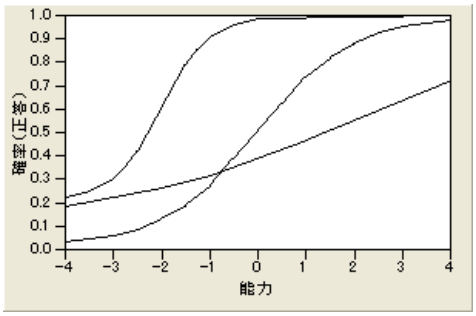
つまり、cは下方漸近線を表します。cは、ゼロでない場合もあります。cの値をさまざまに変化させたときの項目特性曲線は、図7.6のようになります。cは理論的には嬉しいパラメータと言うことができます。その特性の能力をまったく持たない人でも、項目に正しく答えられる確率がゼロでない場合があることを示しているからです。そのため、**当て推量パラメータ**と呼ばれることもあります。

図7.6 cの値を変化させたときのロジスティックモデル



モデル化の際、これら3つのパラメータを変化させることで、多様な確率曲線を作成できます。3つの異なる項目特性曲線の例を、図7.7に示します。下方漸近線は異なりますが、上方漸近線はどれも同じです。これは、当て推量パラメータは0ではない可能性があるのに対して、特性レベルが高くなれば、項目に正答する確率はいつか必ず100%に達すると仮定されるためです。

図 7.7 3 つの項目特性曲線



しかし、3PL モデルでは複雑すぎるケースも多くあります。たとえば *c* パラメータを 0 に限定（妥当な限定でしょう）すれば、推定するパラメータの数が 1 つ減ることになります。*a* と *b* の 2 つのパラメータだけが推定されるモデルを、2PL モデルといいます。

2PL モデルの（3PL よりも安定性が高いことの他に）もう一つの利点は、*b* を、受験者が項目に正答する確率が 50% である点と解釈できることです。この解釈は、3PL モデルでは成り立ちません。

テスト項目の識別力が等しいと仮定できる場合には、モデルをさらに制約できます。項目の識別力が等しいとき、*a* が 1 に固定されるので、推定パラメータは 1 つ（*b*）になります。この **1PL** モデルは、発明者であるデンマーク人数学者 Georg Rasch にちなんで、**Rasch モデル**とも呼ばれます。Rasch モデルは、非常にエレガントで計算コストの低いモデルです。

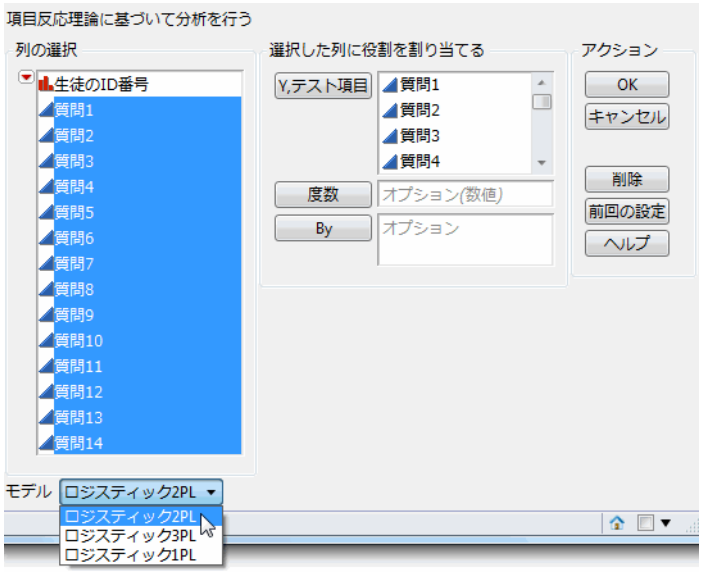
警告：3PL モデルで安定したパラメータ推定値を出すためには、多量のデータが必要となります。直感的に当て推量パラメータが必要だと思われるテストの場合でも、通常は 2PL モデルで十分です。このような背景から、デフォルトの推奨モデルとして 2PL モデルが設定されています。

「項目分析」プラットフォームの起動

例として、サンプルデータの「MathScienceTest.jmp」データテーブルを用います。このデータは、1996 年に実施された第 3 回国際数学・理科教育調査（TIMMS）のデータから採ったサブセットです。

「項目分析」プラットフォームを起動するには、［**分析**］>［**消費者調査**］>［**項目分析**］を選択します。すると、図 7.8 のようなダイアログボックスが表示されます。

図 7.8「項目分析」起動ダイアログボックス



Y, テスト項目　テストの質問に対する正誤が入った変数。

度数　オプションで指定できる変数。応答変数の各パターンが現れる回数を示します。

By　変数を指定すると、その変数の各水準ごとに個別に分析が行われます。

［**モデル**］ドロップダウンメニューから、使用するモデル（1PL、2PL、または3PL）を選択します。

この例では、14 の連続した質問（「**質問 1**」、「**質問 2**」、...「**質問 14**」）を［**Y, テスト項目**］に指定し、［**OK**］をクリックします。すると、デフォルトの 2PL モデルが作成されます。

3PL モデルに関するメモ

3PL モデルを選択した場合、*c* パラメータ（閾値）のペナルティを入力するためのダイアログボックスが開きます。閾値自体を入力しないように注意してください。ここでいうペナルティは、リッジ回帰やニューラルネットで使用するペナルティに似ています。

ペナルティの値が大きくなるほど閾値の推定値が互いに近くなるように、閾値の推定値の標本分散にペナルティを課します。その結果、計算が速くなり、閾値のばらつきが小さくなります（ただし、多少のバイアスが生じます）。

選択方式のテスト問題を項目とし、各設問に用意されている解答の選択肢が同数である場合、どの問題でも閾値パラメータが同じような値になるであろうことが予想されます。たとえば、20 個の問題から成る選択方式のテストの結果を分析する際、各設問に 4 つずつ選択肢があるときは、いずれの設問でも推量（閾値）パラメータが 0.25 に近いと考えるのが妥当です。つまり、適切なケースにこのようなペナルティを適用することは、計算上の利点をもたらすだけでなく、「物理的直感」にも沿った妥当な手法と言えます。

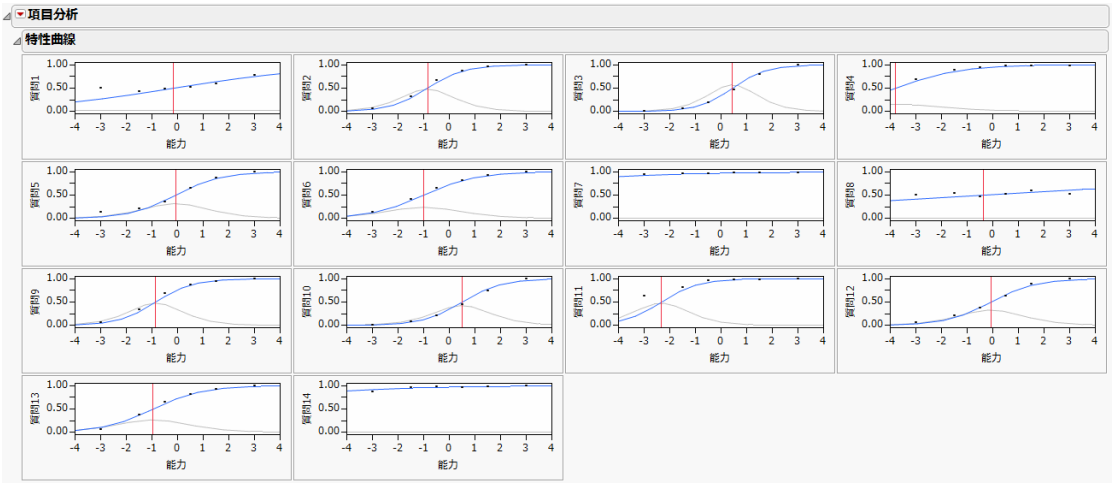
「項目分析」レポート

項目分析レポートには、以下のようなプロットが表示されます。

特性曲線

レポートの冒頭に各質問の項目特性曲線が表示されます。デフォルトでは、すべての曲線が縦 1 列に表示されます。何列かに配列し直すには、レポートのタイトルバーにあるドロップダウンメニューから［**横に並べるプロット数**］コマンドを選択します。図 7.9 では、4 列表示になっています。

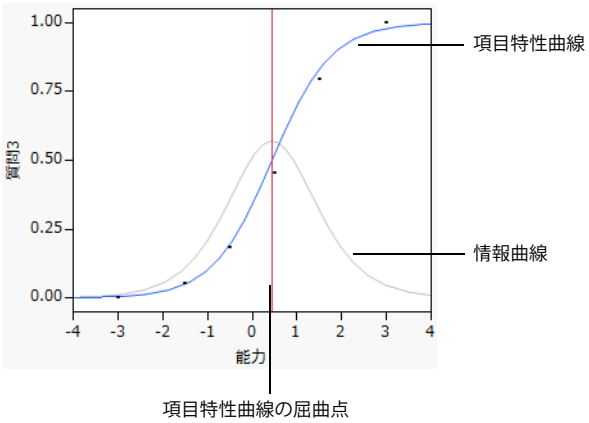
図 7.9 成分曲線



各プロットの赤い縦線は、各曲線の変曲点を示しています。さらに、各能力レベルにおける実際の正答率が点として表示されているため、モデルの適合度をグラフ上で確かめることができます。

グレーの情報曲線からは、テスト全体の情報量に各質問が寄与する量が分かります。情報曲線は、項目特性曲線の傾きなので、変曲点において最大の値を取ります。

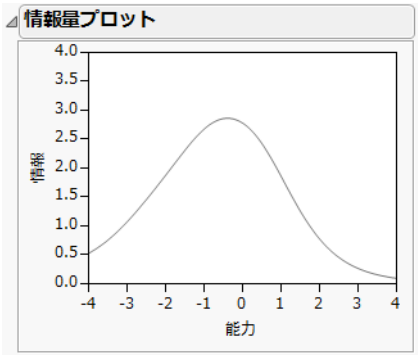
図 7.10 項目特性曲線の構成



情報曲線

同じ質問でも、能力レベルによって与える情報量が異なります。項目ごとに表示されるグレーの情報曲線は、各質問の情報量がテスト全体の情報量に占める割合を示しています。能力範囲全体でテストの全情報量がどのように変化するかは、レポートの「情報量プロット」を見るとわかります（図7.11）。

図 7.11 情報量プロット

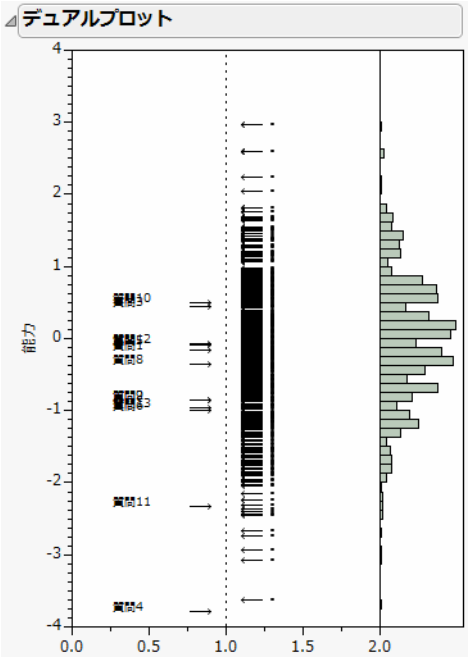


デュアルプロット

項目反応理論モデルの項目困難度パラメータから得られる情報を基に、受験者に適用するのと同じスケール上で、簡単な質問から難しい質問までの困難度スケールを作成することができます。その困難度スケールを見ると、高い特性レベルと低い特性レベルにそれぞれどの項目が結びついているかがわかります。

JMP では、この結びつきが**デュアルプロット**として表示されます。この例の「デュアルプロット」は図7.12のようになります。

図 7.12 デュアルプロット

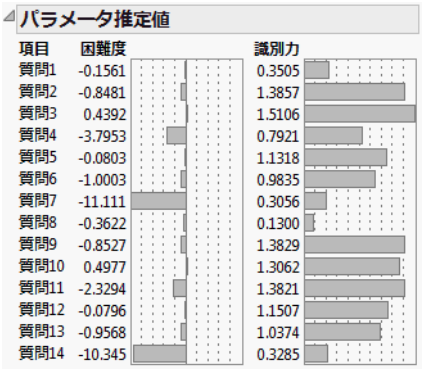


縦の点線の左側に質問、右側に受験者がプロットされています。さらにプロットの右側には、能力レベルのヒストグラムが表示されます。

この例では、能力の範囲がかなり広くなっています。**「質問 10」**は、能力が平均よりおよそ 0.5 標準偏差も高い受験者がやっと 50% の確率で正答できるという難しい質問です。他の質問は、それより低い能力レベルで散らばっており、グラフで表示されている中では、**「質問 11」**と**「質問 4」**が最も易しいようです。ただし、グラフの表示範囲から外れている質問（**「質問 7」**と**「質問 14」**）もあるので注意してください。

デュアルプロットの下にパラメータの推定値が表示されます（図 7.13）。

図 7.13 パラメータ推定値



パラメータ推定値

項目	困難度	識別力
質問1	-0.1561	0.3505
質問2	-0.8481	1.3857
質問3	0.4392	1.5106
質問4	-3.7953	0.7921
質問5	-0.0803	1.1318
質問6	-1.0003	0.9835
質問7	-11.111	0.3056
質問8	-0.3622	0.1300
質問9	-0.8527	1.3829
質問10	0.4977	1.3062
質問11	-2.3294	1.3821
質問12	-0.0796	1.1507
質問13	-0.9568	1.0374
質問14	-10.345	0.3285

項目　テスト項目の名前。

困難度　モデルの b パラメータの値。推定値の横にヒストグラムが表示されます。

識別力　モデルの a パラメータの値。2PL モデルと 3PL モデルの場合だけ計算されます。推定値の横にヒストグラムが表示されます。

閾値　モデルの c パラメータの値。3PL モデルの場合だけ計算されます。

「項目分析」プラットフォームのオプション

レポートのタイトルバーにあるドロップダウンメニューからは、以下の 3 つのコマンドが選択できます。

横に並べるプロット数　ダイアログボックスが開くので、プロットを何列に配列するかを指定します。デフォルトでは、縦 1 列に表示されます。図 7.9（136 ページ）では、4 列になっています。

能力計算式の保存　能力レベル（各受験者に対するスコア）を計算するための式がデータテーブル内の新しい列に保存されます。計算式として保存しておけば、データテーブルに行を追加したときも、保存されている計算式によって、スコアが自動的に算出されます。なお、同一のデータテーブルに対して複数の異なるモデルを実行し、それぞれの計算式をその同一のデータテーブルに保存することもできます。

能力は、`IRT Ability` 関数を使って計算されます。関数の形式は次のとおりです。

```
IRT Ability (Q1, Q2,...,Qn, [a1, a2,..., an, b1, b2,..., bn, c1, c2, ..., cn]);
```

上の式で、`Q1, Q2,...,Qn` はデータテーブル内の項目列、`a1, a2,..., an` は各項目の識別力パラメータ、`b1, b2,..., bn` は各項目の困難度パラメータ、`c1, c2, ..., cn` は各項目の閾値パラメータです。パラメータは、角括弧で囲み、行列として入力します。

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。詳細については、『JMP の使用法』を参照してください。

技術的詳細

$P(\theta)$ は、必ずしもある**特定の**個人が正しく答える確率を指すわけではありません。受験者一人一人で考えれば、特性レベル以外にも過去の受験経験や知識が影響し、間違った答えを選んだり正しく答えたりすることがあるためです。そのため $P(\theta)$ は、能力レベルが θ である受験者グループの正答率であると解釈した方が、定義としては正確です。言い換えれば、同じ特性レベルを持つ多数の受験者を集めたグループが、あるテスト項目に答えるとき、$P(\theta)$ は、グループがその項目に正しく答える確率を予測します。このことは、項目反応理論モデルの項目不変性を示唆しています。理論的には、受験グループが変わっても、パラメータは同じままです。

項目反応理論モデルでは、潜在特性が 1 次元である、つまりすべての質問が同じ潜在特性を測定し、その特性は理論的に連続量として測定できる、と仮定されます。この連続量を表すのが曲線プロットの横軸です。複数の特性が測定される場合は、それぞれが複雑な交互作用を持つため、1 次元モデルは適しません。

付録 A

参考文献

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.

Dwass, M. (1955), “A Note on Simultaneous Confidence Intervals,” *Annals of Mathematical Statistics* 26: 146–147.

Farebrother, R.W. (1981), “Mechanical Representations of the L1 and L2 Estimation Problems,” *Statistical Data Analysis*, 2nd Edition, Amsterdam, North Holland: edited by Y. Dodge.

Fieller, E.C. (1954), “Some Problems in Interval Estimation,” *Journal of the Royal Statistical Society, Series B*, 16, 175-185.

Firth, D. (1993), “Bias Reduction of Maximum Likelihood Estimates,” Biometrika 80:1, 27–38.

Goodnight, J.H. (1978), “Tests of Hypotheses in Fixed Effects Linear Models,” *SAS Technical Report R–101*, Cary: SAS Institute Inc, also in Communications in Statistics (1980), A9 167–180.

Goodnight, J.H. and W.R. Harvey (1978), “Least Square Means in the Fixed Effect General Linear Model,” *SAS Technical Report R–103*, Cary NC: SAS Institute Inc.

Heinze, G. and Schemper, M. (2002), “A Solution to the Problem of Separation in Logistic Regression,” *Statistics in Medicine* 21:16, 2409–2419.

Hocking, R.R. (1985), *The Analysis of Linear Models*, Monterey: Brooks–Cole.

Hosmer, D.W. and Lemeshow, S. (2000), *Applied Logistic Regression*, Second Edition, New York: John Wiley and Sons.

Kaiser, H.F. (1958), “The varimax criterion for analytic rotation in factor analysis” *Psychometrika*, 23, 187–200.

McFadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed., *Frontiers in Econometrics*, pp. 105–142.

Radcliffe, N. J., and Surry, P. D. (2011), “Real-World Uplift Modelling with Significance-Based Uplift Trees,” Stochastic Solutions White Paper, Portrait Technical Report TR-2011-1.

Reichheld, F. F. (2003) “The One Number You Need to Grow,” *Harvard Business Review*, Vol. 81 No. 12, 46-54.

Wright, S.P. and R.G. O’Brien (1988), “Power Analysis in an Enhanced GLM Procedure: What it Might Look Like,” *SUGI 1988, Proceedings of the Thirteenth Annual Conference*, 1097–1102, Cary NC: SAS Institute Inc.

索引
消費者調査

B
By 変数 66

D
Daganzo Trip.jmp 109

F
Firth バイアス調整推定値 107

I
IRT 129

J
JMP スターター 20
JMP スターターウィンドウを閉じる 20
JMP チュートリアル 18

L
Laptop Profile.jmp 92
Laptop Runs.jmp 93
Laptop Subjects.jmp 94
Lung Cancer.jmp 107
Lung Cancer Responses.jmp 105

P
Pizza Combined.jmp 98
Pizza Profiles.jmp 82, 100
Pizza Responses.jmp 83, 100
Pizza Subjects.jmp 86, 100

S
Stacked Daganzo.jmp 110

W-Z
応答 30
自由回答 31

ア
アップリフトモデル 119
　概要 121
　プラットフォーム 123
　レポート 124
　レポートオプション 127

イ
一致性の統計量 38
因子の数 68
因子分析プラットフォーム
　By 変数 66
　重み変数 66
　度数変数 66

オ
応答データ 82
応答の等質性に対する検定 38
応答の度数 31
重み変数 66

カ
回転方法 68
各応答に対する検定 38
各グリッド点で最大化 89
カテゴリカルプラットフォーム 23
　概要 25
　起動ウィンドウ 27
　自由回答 28
　上位カテゴリ 33

表の構成 29
レポート 33

キ
共通因子分析 61, 68
共通の値をもつ応答 30

ク
クロス表形式 36
クロス表形式 転置 36

コ
項目反応理論 129
「項目分析」プラットフォーム
オプション 139
項目分析プラットフォーム 134
概要 131
レポート 136
効用計算式の保存 91
固有値 70

シ
指示変数 31
斜交回転 68
自由回答 28
主成分分析 68
上位カテゴリ 33
信頼区間 90

ス
推定値の相関 90
スクリープロット 70

セ
遷移レポート 38
全水準の推定値 90, 96
選択モデル 77 115
選択モデルプラットフォーム
応答データ 83
セグメント化 100
データテーブルを1つ使った分析 97
被験者効果 86
プラットフォームの起動 82

タ
多重応答 30
多重応答 ID別 30
多重応答 区切り文字 31
探索的因子分析 61

チ
チュートリアル 18
直交回転 68

ツ
ツールヒント 19

ト
度数変数 66

ハ
判定の一致性 30
反復測定 30

ヒ
被験者ごとの勾配を保存 91, 100
被験者データ 82
ピザの選択の例 80 104
標本の各ペア比較 46 47
開く
JMP スターターウィンドウ 20

フ
複合因子検定 90, 96
プロファイルデータ 81

メ
メニューのヒント 19

モ
文字を使用した比較 48
モデルダイアログ 91

ユ

尤度比検定 85, 90

ヨ

横に並べるプロット数 136