



バージョン 13

# 消費者調査 第 2 版

「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

**13.1**

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. 2017.  
『JMP® 13 消費者調査 第2版』 Cary, NC: SAS Institute Inc.

**JMP® 13 消費者調査 第2版**

Copyright © 2017, SAS Institute Inc., Cary, NC, USA

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。

**米国 政府のライセンス権利、権利の制限:** 本ソフトウェアとそのマニュアルは、私的な費用負担の下に開発された商業的コンピュータソフトウェアであり、米国政府に対して権利を制限した上で提供されます。米国政府による本ソフトウェアの使用、複製または開示は、該当する範囲で FAR 12.212, DFAR 227.7202-1(a)、DFAR 227.7202-3(a)、DFAR 227.7202-4 に従った本合意書のライセンス条件に従うものとし、米国連邦法の下で求められる範囲において、FAR 52.227-19 (2007年12月) で規定されている制限された最小限の権利に従うものとします。FAR 52.227-19 が適用される場合、この条項は、その(c)項に基づく通告の役目を果たし、本ソフトウェアまたはマニュアルにその他の通告を添付する必要はありません。本ソフトウェアおよびマニュアルにおける政府の権利は、本合意書で規定されている権利に限られます。

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

2017年2月

SAS® と、SAS Institute Inc. の他の製品名およびサービス名は、米国および他の国における SAS Institute Inc. の登録商標または商標です。® は、米国において登録されていることを示します。

他のブランド名および製品名は、それぞれの会社の商標です。

SAS ソフトウェアは、オープンソースのソフトウェアを含むがそれに限らない、特定のサードパーティ製ソフトウェアと共に提供される場合があります。かかるソフトウェアは、適用されるサードパーティソフトウェアライセンス契約に基づいてライセンスを得たものです。SAS ソフトウェアと共に配布されるサードパーティ製ソフトウェアに関する情報は、<http://support.sas.com/thirdpartylicenses> を参照してください。

## テクノロジーライセンスに関する通知

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

All Rights Reserved.

何らかの目的でこのソフトウェアとそのマニュアルを手数料なしで使用、コピー、変更および配布することは、これをもって許可されます。ただし、すべてのコピーに上記の著作権に関する通知が記載されていること、および補助的なマニュアルに著作権に関する通知とこの許可に関する通知の両方が記載されていることを条件とします。

NEIL HODGSONは、商業性および適合性の默示的な保証を含め、このソフトウェアに関するすべての保証を放棄します。NEIL HODGSONは、いかなる場合においても、それが契約、過失、もしくは他の不法行為のどれであれ、このソフトウェアの使用もしくは性能から生じた、もしくはそれに関連して生じた使用、データ、もしくは利益の損失の結果として生じる特別損害、間接損害、もしくは付随的損害を始めとするいかなる損害に対しても責任を負いません。

- Telerik RadControls: Copyright © 2002-2012, Telerik. 含まれている Telerik RadControlsをJMP以外で使用することは許可されていません。
- ZLIB圧縮ライブラリ - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Natural Earthを使用して作成。無料のベクトルおよびラスター地図データ @ [naturalearthdata.com](http://naturalearthdata.com).
- パッケージ - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

ソースおよびバイナリの形で、そのまま、もしくは変更を加えて再配布および使用することは、次のような条件を満たす限り、許可されます。

再配布するソースコードには、上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。

バイナリ形式で再配布する場合は、共に提供されるマニュアルなどの資料に上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。

事前に書面による許可を得ることなく、このソフトウェアから派生した製品の推奨または宣伝のために WhiteBox の名前やその貢献者の名前を使用することはできません。

このソフトウェアは、著作権保有者および貢献者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する默示的な保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。いかなる場合においても、著作権保有者または貢献者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

- iODBC ソフトウェア - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin ([www.iodbc.org](http://www.iodbc.org)). All rights reserved.

ソースおよびバイナリの形で、そのまま、もしくは変更を加えて再配布および使用することは、次のような条件を満たす限り、許可されます。

- 再配布するソースコードには、上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
- バイナリ形式で再配布する場合は、共に提供されるマニュアルなどの資料に上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
- 事前に書面による許可を得ることなく、このソフトウェアから派生した製品の推奨または宣伝のために OpenLink Software Inc. の名前やその貢献者の名前を使用することはできません。

このソフトウェアは、著作権保有者および貢献者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する黙示的な保証を含むがそれに限らない、いかなる明示的もしくは黙示的な保証も行われません。いかなる場合においても、OPENLINKまたは貢献者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

- bzip2、関連ライブラリの「libbzip2」、およびすべてのマニュアル: Copyright © 1996-2010, Julian R Seward. All rights reserved.

ソースおよびバイナリの形で、そのまま、もしくは変更を加えて再配布および使用することは、次のような条件を満たす限り、許可されます。

再配布するソースコードには、上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。

このソフトウェアの供給源は正しく表記しなければならず、使用者が元のソフトウェアを記述したと主張することはできません。ある製品の中でこのソフトウェアを使用する場合は、その製品のマニュアルに謝辞を記載してもらえるとありがたいですが、必須ではありません。

ソースに変更を加えたバージョンには、その旨を明記しなければならず、元のソフトウェアとは違うものであることを明確にしてください。

事前に書面による許可を得ることなく、このソフトウェアから派生した製品の推奨または宣伝のために作成者の名前を使用することはできません。

このソフトウェアは、作成者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する黙示的な保証を含むがそれに限らない、いかなる明示的もしくは黙示的な保証も行われません。いかなる場合においても、作成者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可

能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

- R ソフトウェア: Copyright © 1999-2012, R Foundation for Statistical Computing.
- MATLAB ソフトウェア: Copyright © 1984-2012, The MathWorks, Inc. 米国特許法および国際特許法によって保護されています。www.mathworks.com/patents を参照してください。MATLAB および Simulink は、The MathWorks, Inc. の登録商標です。他の商標は、www.mathworks.com/trademarks に一覧されています。他の製品名やブランド名は、それぞれの所有者の商標または登録商標である可能性があります。
- libopc: Copyright © 2011, Florian Reuter. All rights reserved.

ソースおよびバイナリの形で、そのまま、もしくは変更を加えて再配布および使用することは、次のような条件を満たす限り、許可されます。

- 再配布するソースコードには、上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
- バイナリ形式で再配布する場合は、共に提供されるマニュアルなどの資料に上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
- 事前に書面による許可を得ることなく、このソフトウェアから派生した製品の推奨または宣伝のために Florian Reuter の名前やその貢献者の名前を使用することはできません。

このソフトウェアは、著作権保有者および貢献者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する默示的な保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。いかなる場合においても、著作権保有者または貢献者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

- libxml2 - ソースコードに特に記載がある場合を除く（たとえば、使用しているライセンスは類似しているが、著作権の通知が異なる hash.c、list.c ファイルや trio ファイル）、すべてのファイル:

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

これをもって、このソフトウェアのコピーと関連する文書ファイル（「本ソフトウェア」）を入手した人すべてに対し、無料で本ソフトウェアを使用、コピー、変更、マージ、パブリッシュ、配布、サプライセンスする、もしくはコピーを販売する権利を含むがそれに限定せず、本ソフトウェアを制限なく取り扱う権利、および本ソフトウェアの供給相手に対してそうすることを許可する権利が付与されます。ただし、以下の条件を満たさなければなりません。

上記の著作権に関する通知とこの許可に関する通知が、本ソフトウェアのコピーのすべてまたは大部分に記載されていること。

このソフトウェアは、「現状のままで」提供され、商業性および特定の目的に対する適合性、および非侵害の保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。DANIEL VEILLARDは、いかなる場合においても、それが契約、過失、もしくは他の不法行為のどれであれ、本ソフトウェアから、もしくは本ソフトウェアに関連して、または本ソフトウェアの使用もしくは他の取り扱いに関連して生じた申し立て、損害賠償もしくは他の義務に対し、責任を負いません。

この通知に含まれているものを除き、Daniel Veillardから事前に書面による許可を得ることなく、本ソフトウェアの広告、またはその他の手段による本ソフトウェアの販売、使用もしくは他の取り扱いの宣伝に Daniel Veillard の名前を使用することはできません。

- UNIX ファイルに使用された解凍アルゴリズムについて：

Copyright © 1985, 1986, 1992, 1993

カリフォルニア大学評議員。All rights reserved.

このソフトウェアは、評議員および貢献者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する默示的な保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。いかなる場合においても、評議員または貢献者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

1. 再配布するソースコードには、上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
  2. バイナリ形式で再配布する場合は、共に提供されるマニュアルなどの資料に上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
  3. 事前に書面による許可を得ることなく、このソフトウェアから派生した製品の推奨または宣伝のために大学の名前や貢献者の名前を使用することはできません。
- Snowball - Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.  
All rights reserved.

ソースおよびバイナリの形で、そのまま、もしくは変更を加えて再配布および使用することは、次のような条件を満たす限り、許可されます。

1. 再配布するソースコードには、上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。
2. バイナリ形式で再配布する場合は、共に提供されるマニュアルなどの資料に上記の著作権に関する通知、この条件リスト、これに続く放棄声明が記載されていなければなりません。

3. 事前に書面による許可を得ることなく、このソフトウェアから派生した製品の推奨または宣伝のために著作権保有者の名前や貢献者の名前を使用することはできません。

このソフトウェアは、著作権保有者および貢献者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する默示的な保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。いかなる場合においても、著作権保有者または貢献者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。



# 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	15
表記規則	16
JMPのマニュアル	17
JMP ドキュメンテーションライブラリ	17
JMP ヘルプ	23
JMPを習得するためのその他のリソース	23
チュートリアル	23
サンプルデータテーブル	24
統計用語とJSL用語の習得	24
JMPを使用するためのヒント	24
ツールヒント	25
JMP User Community	25
JMPer Cable	25
JMP関連書籍	25
「JMP スターター」 ウィンドウ	26
テクニカルサポート	26

### 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	27
-------------------	----

### 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	29
「カテゴリカル」 プラットフォームの使用例	30
「カテゴリカル」 プラットフォームの起動	31
応答の役割	31
列の役割	33
起動ウィンドウのその他のオプション	34
「カテゴリカル」 レポート	34
「カテゴリカル」 プラットフォームのオプション	36
レポートに関するオプション	36
統計的検定に関するオプション	37
「カテゴリカル」 プラットフォームのその他のオプション	38
クロス表の応答におけるオプション	41
比較を表すアルファベット文字	42

上位カテゴリ	43
〔上位カテゴリ〕 のオプション	43
環境設定の変更	44
「カテゴリカル」 プラットフォームの別例	45
応答の等質性に対する検定の例	45
複数の応答をもつ検定の例	46
セルのカイ2乗検定の例	48
比較を表すアルファベット文字による、標本の各ペア比較の例	49
比較を表すアルファベット文字による、セルの各ペア比較の例	50
比較を表すアルファベット文字による、ユーザ指定の比較の例	52
条件付き関連と相対リスクの例	53
判定の一致性の例	56
反復測定の例	57
〔多重応答〕 タブの例	58
比較のアルファベット文字を持つスコア平均の例	63
「表の構成」 レポートの例	64
「カテゴリカル」 プラットフォームの統計的詳細	66
Rao-Scott 修正	66

## 4 多重対応分析

カテゴリカルな変数の水準間にある関係の特定	69
多重対応分析の例	70
「多重対応分析」 プラットフォームの起動	72
「多重対応分析」 レポート	73
「多重対応分析」 プラットフォームのオプション	74
「対応分析」 のオプション	74
プロットの表示	75
詳細の表示	76
調整済み慣性の表示	76
座標の表示	76
要約統計量の表示	77
慣性への偏寄与率の表示	77
余弦2乗の表示	78
クロス表	78
追加行のクロス表	79
追加列のクロス表	79
「多重対応分析」 プラットフォームの別例	79
追加変数を使用した例	79
追加IDを使用した例	80
「多重対応分析」 プラットフォームの統計的詳細	83
「詳細」 レポート	83
調整済み慣性	83
要約統計量	84

慣性への偏寄与率 .....	84
<b>5 多次元尺度構成</b>	
対象間における類似度の視覚化 .....	85
「多次元尺度構成」プラットフォームの概要 .....	86
多次元尺度構成の例 .....	86
「多次元尺度構成」プラットフォームの起動 .....	88
「多次元尺度構成」レポート .....	90
多次元尺度構成プロット .....	90
Shepard図 .....	90
あてはめの詳細 .....	90
「多次元尺度構成」プラットフォームのオプション .....	91
Waernリンク .....	92
「多次元尺度構成」プラットフォームの別例 .....	93
「多次元尺度構成」プラットフォームの統計的詳細 .....	95
ストレス .....	95
変換 .....	95
属性の元データ .....	96
<b>6 因子分析</b>	
潜在的な因子によるモデル化 .....	97
「因子分析」プラットフォームの概要 .....	98
「因子分析」プラットフォームの例 .....	98
「因子分析」プラットフォームの起動 .....	101
「因子分析」レポート .....	101
モデルの設定 .....	102
回転方法 .....	104
「因子分析」プラットフォームのオプション .....	105
モデルをあてはめた後のオプション .....	106
<b>7 選択モデル</b>	
選択モデルのあてはめ .....	113
「選択モデル」プラットフォームの概要 .....	114
「選択モデル」プラットフォームの例 .....	115
「選択なし」がある1つのデータテーブル .....	115
複数のデータテーブル .....	118
「選択モデル」プラットフォームの起動 .....	124
[1つのデータテーブル, 積み重ね] の起動ウィンドウ .....	125
[複数のデータテーブル, 相互参照] の起動ウィンドウ .....	126
「選択モデル」レポート .....	131
効果の要約 .....	131
パラメータ推定値 .....	132
尤度比検定 .....	133

Bayesパラメータ推定値	133
「選択モデル」プラットフォームのオプション	135
支払意思額	137
その他の例	140
製品の使用を決める分析例	140
セグメント化の例	148
「選択モデル」プラットフォームを使用したロジスティック回帰の例	152
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の例	155
データを2つの分析テーブルに変換する例	157
データを1つの分析テーブルに変換する例	162
技術的詳細	164
分析対象のデータに対する特殊な規則	164
効用と確率	165
勾配	166

## 8 MaxDiff

MaxDiff モデルのあてはめ	167
「MaxDiff モデル」プラットフォームの概要	168
「MaxDiff」プラットフォームの例	168
1つのデータテーブル	169
複数のデータテーブル	172
「MaxDiff」プラットフォームの起動	175
[1つのデータテーブル, 積み重ね] の起動ウィンドウ	176
[複数のデータテーブル, 相互参照] の起動ウィンドウ	177
「MaxDiff モデル」レポート	181
効果の要約	181
MaxDiff 結果	183
パラメータ推定値	183
Bayesパラメータ推定値	185
尤度比検定	186
「MaxDiff」プラットフォームのオプション	186
「比較」レポート	188
Bayes チェーンの保存	189

## 9 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	191
「アップリフト」プラットフォームの概要	192
「アップリフト」プラットフォームの使用例	193
「アップリフト」プラットフォームの起動	194
「アップリフトモデル」レポート	195
アップリフトモデルのグラフ	196
アップリフトレポートのオプション	198

## 10 項目分析

テスト結果に対する項目反応分析	201
項目分析の例	202
「項目分析」プラットフォームの起動	204
「項目分析」レポート	205
特性曲線	205
情報量プロット	206
デュアルプロット	207
パラメータ推定値	207
「項目分析」プラットフォームのオプション	208
項目反応曲線	208
項目反応曲線のモデル	209
項目反応理論モデルの仮定	211
項目反応理論モデルの推定方法	212
能力計算式	213

## A 参考文献

### 索引

消費者調査	217
-------	-----



# 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
  - テクニカルサポートのオプション

---

## 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、次のような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
- 次の項目の表記規則は下記のとおりです。
  - 重要な単語や句、JMPに固有の定義を持つ単語や句は太字または「」で囲んで表記
  - マニュアルのタイトルは『』で囲んで表記
  - 変数名は斜体で表記
  - スクリプトの出力は斜体で表記
- JMP Proのみの機能には JMP Pro アイコン  がついています。JMP Proの機能の概要については [https://www.jmp.com/ja\\_jp/software/predictive-analytics-software.html](https://www.jmp.com/ja_jp/software/predictive-analytics-software.html) をご覧ください。

---

**メモ:** 特別な情報および制限事項には、この文のように「メモ」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では、PDF形式のマニュアルが用意されています。

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

### JMP ドキュメンテーションライブラリ

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

マニュアル	目的	内容
『はじめてのJMP』	JMPをあまりご存知ない方を対象とした入門ガイド	JMPの紹介と、データを作成および分析し始めるための情報
『JMPの使用法』	JMPのデータテーブルと、基本操作を理解する	一般的なJMPの概念と、データの読み込み、列プロパティの変更、データの並べ替え、SASへの接続など、JMP全体にわたる機能の説明
『基本的な統計分析』	このマニュアルを見ながら、基本的な分析を行う	<p>[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：</p> <ul style="list-style-type: none"><li>一変量の分布</li><li>二変量の関係</li><li>表の作成</li><li>テキストエクスプローラ</li></ul> <p>[分析] &gt; [二変量の関係] で二変量、一元配置分散分析、分割表に対する分析を実行する方法の説明。ブートストラップを使用した標本分布の近似方法やシミュレーションの機能を使用したパラメトリックな標本再抽出の実行方法の説明も含まれています。</p>

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	<p>【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：</p> <ul style="list-style-type: none"> <li>• グラフビルダー</li> <li>• 重ね合わせプロット</li> <li>• 三次元散布図</li> <li>• 等高線図</li> <li>• バブルプロット</li> <li>• パラレルプロット</li> <li>• セルプロット</li> <li>• ツリーマップ</li> <li>• 散布図行列</li> <li>• 三角図</li> <li>• チャート</li> </ul> <p>このマニュアルには背景マップやカスタムマップの作成方法も記載されています。</p>
『プロファイル機能』	対話式のプロファイルツールの使い方を学ぶ。任意の応答曲面の断面を表示できるようになります。	【グラフ】メニューに表示されるすべてのプロファイルについて。誤差因子の分析が、ランダム入力を使用したシミュレーションの実行とともに含まれています。
『実験計画(DOE)』	実験の計画方法と適切な標本サイズの決定方法を学ぶ	【実験計画 (DOE)】メニューと【分析】>【発展的なモデル】メニューの「発展的な実験計画モデル」に関するすべてのトピックについて。

マニュアル	目的	内容
『基本的な回帰モデル』	「モデルのあてはめ」プラットフォームとその多くの手法について学ぶ	<p>[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：</p> <ul style="list-style-type: none"><li>• 標準最小2乗</li><li>• ステップワイズ</li><li>• 一般化回帰</li><li>• 混合モデル</li><li>• MANOVA</li><li>• 対数線形-分散</li><li>• 名義ロジスティック</li><li>• 順序ロジスティック</li><li>• 一般化線形モデル</li></ul>

マニュアル	目的	内容
『予測モデルおよび発展的なモデル』	さらなるモデリング手法について学ぶ	<p>[分析] &gt; [予測モデル] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none"> <li>• モデル化ユーティリティ</li> <li>• ニューラル</li> <li>• パーティション</li> <li>• ブートストラップ森</li> <li>• ブースティングツリー</li> <li>• K近傍法</li> <li>• 単純Bayes</li> <li>• モデルの比較</li> <li>• 計算式デボ</li> </ul> <p>[分析] &gt; [発展的なモデル] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none"> <li>• 曲線のあてはめ</li> <li>• 非線形回帰</li> <li>• Gauss過程</li> <li>• 時系列分析</li> <li>• 対応のあるペア</li> </ul> <p>[分析] &gt; [スクリーニング] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none"> <li>• 応答のスクリーニング</li> <li>• 工程のスクリーニング</li> <li>• 説明変数のスクリーニング</li> <li>• アソシエーション分析</li> </ul> <p>[分析] &gt; [発展的なモデル] &gt; [発展的な実験計画モデル] で使用できるプラットフォームについては、『実験計画(DOE)』に説明があります。</p>

マニュアル	目的	内容
『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	<p>[分析] &gt; [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：</p> <ul style="list-style-type: none"><li>• 多変量の相関</li><li>• 主成分分析</li><li>• 判別分析</li><li>• PLS</li></ul> <p>[分析] &gt; [クラスター分析] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none"><li>• 階層型クラスター分析</li><li>• K Means クラスター分析</li><li>• 正規混合</li><li>• 潜在クラス分析</li><li>• 変数のクラスタリング</li></ul>
『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	<p>[分析] &gt; [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：</p> <ul style="list-style-type: none"><li>• 管理図ビルダーと個々の管理図</li><li>• 測定システム分析</li><li>• 計量値/計数値ゲージチャート</li><li>• 工程能力</li><li>• パレート図</li><li>• 特性要因図</li></ul>

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	<p>[分析] &gt; [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：</p> <ul style="list-style-type: none"> <li>• 寿命の一変量</li> <li>• 寿命の二変量</li> <li>• 累積損傷</li> <li>• 再生モデルによる分析</li> <li>• 劣化分析と破壊劣化</li> <li>• 信頼性予測</li> <li>• 信頼性成長</li> <li>• 信頼性ブロック図</li> <li>• 修理可能システムのシミュレーション</li> <li>• 生存時間分析</li> <li>• 生存時間(パラメトリック)のあてはめ</li> <li>• 比例ハザードのあてはめ</li> </ul>
『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	<p>[分析] &gt; [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：</p> <ul style="list-style-type: none"> <li>• カテゴリカル</li> <li>• 多重対応分析</li> <li>• 多次元尺度構成</li> <li>• 因子分析</li> <li>• 選択モデル</li> <li>• MaxDiff</li> <li>• アップリフト</li> <li>• 項目分析</li> </ul>
『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。

---

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

---

## JMPヘルプ

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

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

---

## JMPを習得するためのその他のリソース

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

- チュートリアル（「チュートリアル」（23ページ）を参照）
- サンプルデータ（「サンプルデータテーブル」（24ページ）を参照）
- 索引（「統計用語とJSL用語の習得」（24ページ）を参照）
- 使い方ヒント（「JMPを使用するためのヒント」（24ページ）を参照）
- Webリソース（「JMP User Community」（25ページ）を参照）
- 専門誌『JMPer Cable』（「JMPer Cable」（25ページ）を参照）
- JMPに関する書籍（「JMP関連書籍」（25ページ）を参照）
- JMPスターター（「「JMPスターター」ウィンドウ」（26ページ）を参照）
- 教育用リソース（「サンプルデータテーブル」（24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、実験の計画、標本平均と定数の比較など、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。サンプルデータディレクトリを開くには、【ヘルプ】>【サンプルデータライブラリ】を選択します。

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

サンプルデータテーブルは次のディレクトリにインストールされています。

Windowsの場合: C:\Program Files\SAS\JMP\13\Samples\Data

Macintoshの場合: \Library\Application Support\JMP\13\Samples\Data

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

サンプルデータの使用例を参照するには、【ヘルプ】>【サンプルデータ】を選択し、教育用セクションから検索してください。教育用リソースについては、<http://jmp.com/tools> にも情報があります。

## 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:** Windowsでは、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]を選択し、[メニューのヒントを表示]の選択を解除します。このオプションは、Macintoshでは使用できません。

## 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]を選択するか、<https://communityjmp.com/>をご覧ください。

## 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。

[https://www.jmp.com/ja\\_jp/academic/books-for-jmp-users.html](https://www.jmp.com/ja_jp/academic/books-for-jmp-users.html)

## 「JMPスター」 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」 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」 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。また、JMP Proの機能やプラットフォームのリストも含まれています。

- ・ 「JMPスター」 ウィンドウを開くには、[表示] (Macintoshでは [ウィンドウ]) > [JMPスター] を選択します。
- ・ WindowsでJMPの起動時に自動的に「JMPスター」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] を選択します。Macintoshでは、[JMP] > [環境設定] > [起動時に JMP スターウィンドウを表示する] を選択します。

---

## テクニカルサポート

JMPのテクニカルサポートは、JMPのエンジニアが担当し、その多くは、統計学などの技術的な分野の知識を有しています。

<http://www.jmp.com/japan/support> には、テクニカルサポートへの連絡方法などが記載されています。

# 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

JMP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。「どのように製品やサービスを消費者が利用しているか?」、「顧客の満足度はどのくらいか?」、また、「彼らはどんな新機能を望んでいるか?」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。このような消費者調査のための活動を分析するツールは、「消費者調査」メニューにあります。以下のようなプラットフォームが用意されています。

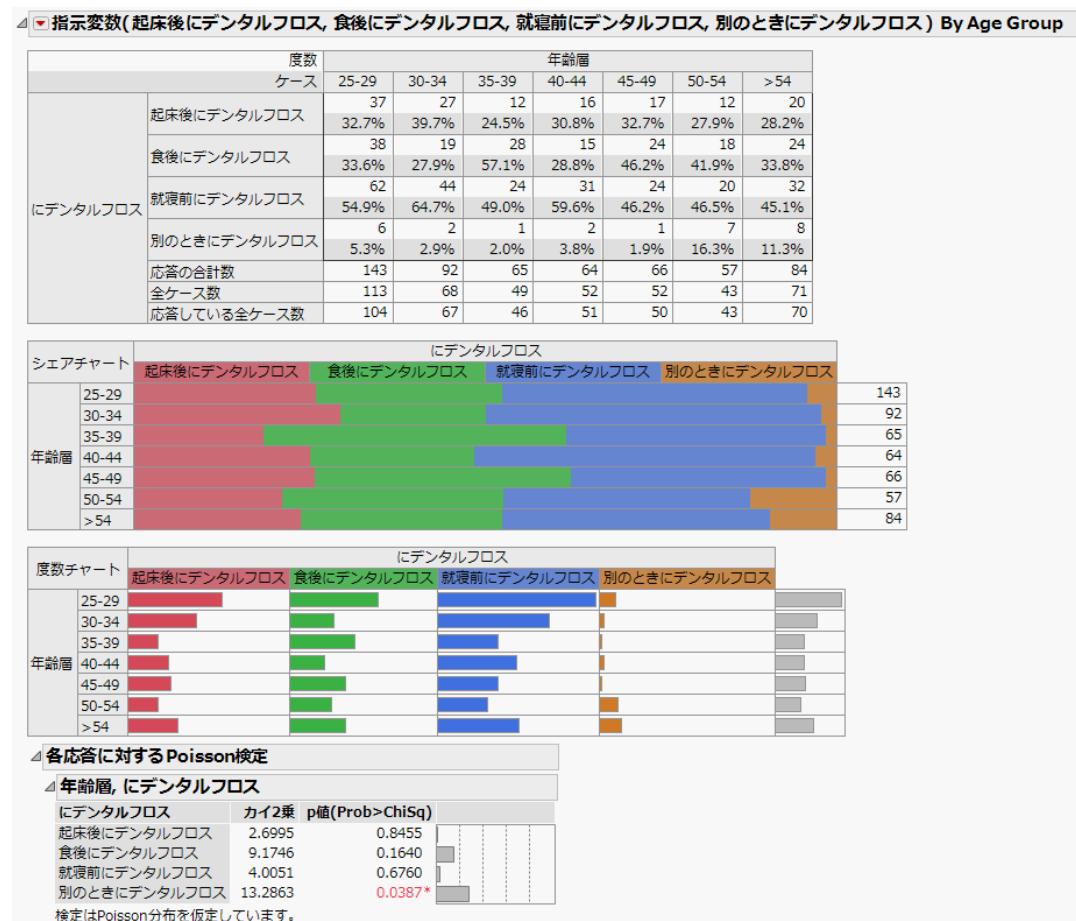
- ・「カテゴリカル」 プラットフォームでは、多重応答データを含むカテゴリカルな応答データを使って、表の作成、プロット、比較を行うことができます。このプラットフォームを使ってアンケート調査データや、工業製品の不適合原因や故障原因、研究の被験者統計などの他のカテゴリカルな応答データを分析することができます。「カテゴリカル」 プラットフォームを使って、さまざまな方法で構成されたデータの応答を分析することができます。詳細については、[第3章「カテゴリカルな応答の分析」](#)を参照してください。
- ・多重対応分析 (Multiple Correspondence Analysis; MCA) は、複数のカテゴリカルな変数を分析対象とし、それらのカテゴリカルな変数の水準間にある関係を調べます。多重対応分析は、特にフランスや日本の社会科学分野で広く利用されています。多重対応分析は、質問紙調査では、複数の質問における関係を調べるのに役立つでしょう。詳細については、[第4章「多重対応分析」](#)を参照してください。
- ・「多次元尺度構成」 (Multi-Dimensional Scaling; MDS) プラットフォームは、対象間の類似度（類似性、相違性、距離）を視覚化します。詳細については、[第5章「多次元尺度構成」](#)を参照してください。
- ・「因子分析」 プラットフォームでは、多数の観測変数から因子を構成することができます。このような因子は、観測変数のサブセットの線形結合で表されます。このプラットフォームでは、測定された観測変数で説明される因子の個数や、因子と変数間の関係の強さを調べることができます。詳細については、[第6章「因子分析」](#)を参照してください。
- ・「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。詳細については、[第7章「選択モデル」](#)を参照してください。
- ・「MaxDiff」 プラットフォームは、各商品に対する好ましさに関する質問から、それらの商品の相対的な重要度を算出します。MaxDiff法では、標準的な調査方法とは異なり、提示した選択肢を回答者に選んでもらうときに、最も好きな選択肢と最も嫌いな選択肢の2つを選んでもらいます。詳細については、[第8章「MaxDiff」](#)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9章「アップリフトモデル」](#)を参照してください。

- 「項目分析」プラットフォームでは、項目反応理論モデルをあてはめることができます。項目反応理論(IRT)法は、テストやアンケートなどの測定手段の分析とスコアリングに使われます。IRTは、モデルを利用して特性や能力を被験者の項目に対する推奨確率や正答率に結びつけます。IRTは、標準化された検定、認知発達、消費者の嗜好を調べるのに使うことができます。詳細については、[第10章「項目分析」](#)を参照してください。

## 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、カテゴリカルデータに対して、度数表の作成、グラフの作成、度数の比較を行います。多重応答（multiple responses; 複数回答）も扱えます。アンケート調査の回答、工業製品の不適合原因や故障原因、調査参加者のデモグラフィック情報など、さまざまな分野のカテゴリカルデータを分析できます。「カテゴリカル」プラットフォームでは、さまざまな形式のデータを分析でき、起動ウィンドウにおいてその形式を指定できます。

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」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。このデータには、一般的な事柄に対する意見や態度、口腔衛生に関する質問への回答が含まれています。「カテゴリカル」プラットフォームを使うと、年齢層別に応答を比較することができます。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Consumer Preferences.jmp」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て、[クロス表 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て、[応答の等質性に対する検定] を選択します。

図3.2 質問に対する年齢層別の応答

年齢層	性別	度数		年齢層						
		シェア	カウント	25-29	30-34	35-39	40-44	45-49	50-54	>54
仕事熱心である	Agree	84.1%	95	49	30	29	29	22	33	
	Disagree	15.9%	18	19	19	23	23	21	38	
	応答の合計数		113	68	49	52	52	43	71	

年齢層	性別	仕事熱心である								
		Agree	Disagree							
25-29		113								
30-34			68							
35-39				49						
40-44					52					
45-49					52					
50-54						43				
>54							71			

応答の列	標本の列	尤度比	カイ2乗	尤度比 p値	カイ2乗	Pearson p値
仕事熱心である	年齢層	39.5874	<.0001*	37.4593	<.0001*	

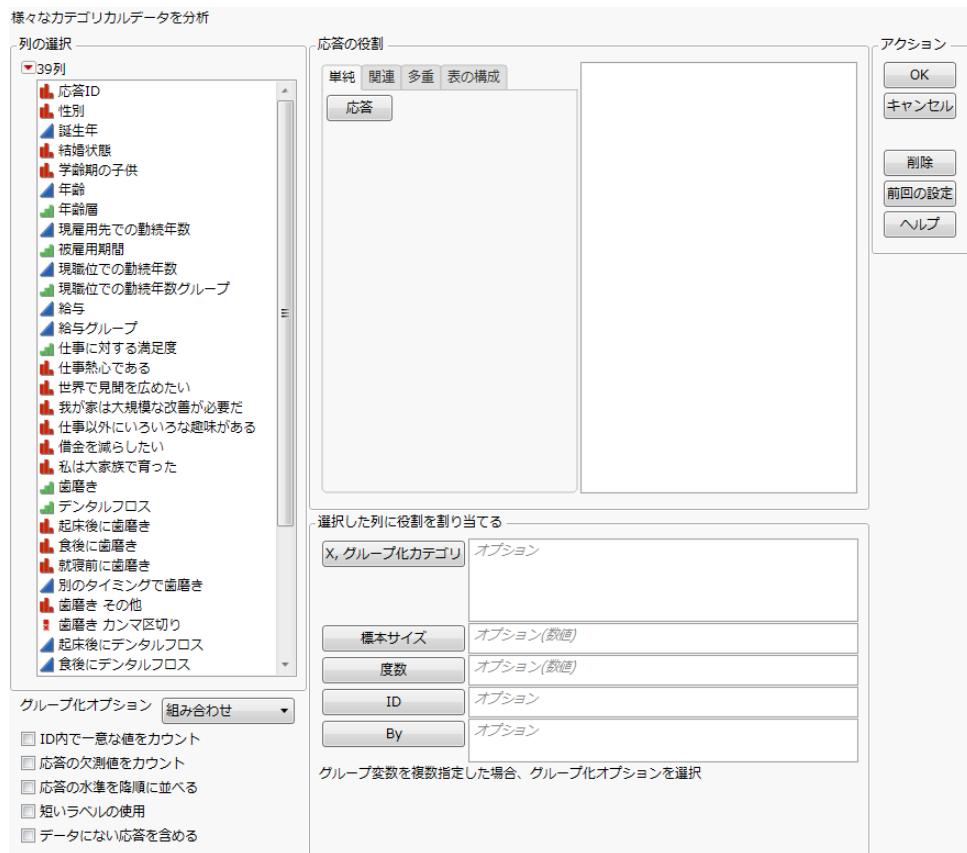
この質問に対する回答には、「Agree (同意)」と「Disagree (反対)」の2つの選択肢が用意されています。回答者は7つの年齢層に分けられ、各グループは43～113人です。最年少である年齢層（25～29）の回答者は113人であり、調査を行ったうちの最大グループでした。「Agree (同意)」の割合が最も高いのは、最年少の年齢層で、84.1%でした。シェアチャートには、同意／反対の割合が、年齢層ごとに描かれています。年齢が上がるにつれて、この質問に同意する回答者の割合は減少します。したがって、年長の回答者はより年少の回答者よりも仕事熱心ではない傾向にあると言えます。

ヒント: チャートの色は、「値の色」列プロパティで定義することができます。『JMPの使用法』の「列情報ウィンドウ」章を参照してください。

## 「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



## 応答の役割

起動ウィンドウには、応答の役割について、3種類のタブ（[単純]、[関連]、[多重]）があります。また、独自のクロス表を作成したいときに用いる [表の構成] タブがあります。これらの各タブは、分析対象とする応答の種類に対応しています。そして、それぞれのタブに用意されている各オプションは、データがどのような形式で保存されているかに対応しています。

### [単純] タブ

[単純] タブは、1列ずつ個別にデータを集計したいときに用います。[単純] タブには、次の1つのオプションしかありません。

**応答** 1つまたは複数の列を分析に追加します。複数の列を選択した場合、「カテゴリカル」レポートには、個々の列に対するレポートが表示されます。

### [関連] タブ

[関連] タブは、関係がある複数の列を集計したいときに用います。以下のオプションが用意されています。

**共通の値をもつ応答** 同じ水準を持つ複数列のデータを、1つの表に要約します。このオプションは、アンケート調査で、選択肢がまったく同じである質問を要約する場合に役立ちます。それらの質問を同時に要約・比較できます。

**反復測定** 異なる時点での応答が各列に含まれているデータを要約します。複数の時点で同一個体が応答している場合、標本が重複している（overlapping）と言います。標本に重複があるときは、検定を行うときにKish修正が使われます。Kish (1965; 12.4 節) を参照してください。

**判定の一致性** データが同一の対象物を判定した結果であり、各列が異なる回答者（判定者）に対応しているときに、そのデータを要約します。

### [多重] タブ

[多重] タブには、1つまたは複数の列に記録された多重応答（multiple responses; 複数回答）を分析するためのオプションが用意されています。「多重応答」は、アンケート調査において、複数の選択肢を選ぶことが許されている質問で見られます。ほかにも、「多重応答」は、製造業などで、1つの製品に複数の不適合が生じる場合にも見られます。[多重応答] タブの各オプションは、多重応答データがどのような形式で保存されているかに対応しています。

**多重応答** 多重応答データが複数列に含まれており、各列にそれぞれ1つの応答データが含まれている形式です。ここで選択された列の個数は、応答の最大数（選択肢の個数）に対応しています。実際のデータでは、セルの多くは空白になる可能性があります。

**多重応答 ID別** 回答者や部品のIDを含む列が別にあり、かつ、单一の列に積み重ねて多重応答データが保存されている形式です。

**多重応答 区切り文字** 多重応答データが、カンマ、セミコロン、タブで区切られて、单一の列に保存されている形式です。

**指示変数** 多重応答データが、指示変数で保存されている形式です。この形式は、各応答（質問の各選択肢）に対して1つの列があり、その列が0と1を含む指示変数となっているものです。

**応答の度数** 多重応答データが度数で集計されている形式です。この形式は、「指示変数」形式を集計したものです。

**自由回答** テキストデータを要約します。[自由回答] オプションを使用すると、「カテゴリカル」レポート ウィンドウの中にテキストエクスプローラが開きます。『基本的な統計分析』の「テキストエクスプローラ」章を参照してください。

## [表の構成] タブ

[表の構成] タブは、分析者が望むクロス表を柔軟に構成したいときに用います。

- 横並び、交差、枝分かれのクロス表を対話的に指定できます。[表の構成] タブで作成されるクロス表では、クロス表の横側に位置する最も内側の項目が応答と見なされ、その他のすべての項目は因子と見なされます。
- [表の構成] タブを使えば、1つの起動ウィンドウにて、複数の異なるクロス表を指定できます。
- 指定された列の尺度が[多重応答]に設定されている場合は、区切り文字で区切られた多重応答データとみなして分析を行います。列の尺度についての詳細は、『JMPの使用法』の「[列情報ウィンドウ](#)」章を参照してください。

詳細については、[「表の構成」レポートの例](#) (64ページ) を参照してください。

---

**ヒント:** 多重応答データで、応答の等質性に対する検定を行いたい場合は、[表の構成] タブを使用してください。

## 列の役割

次のような役割が用意されています。

**X, グループ化カテゴリ** グループを示す列を指定します。グループごとに度数がカウントされます。このオプションに複数の列が指定された場合に、デフォルト（[グループ化オプション] で「組み合わせ」が選択されている場合）では、枝分かれで集計されます。[グループ化オプション] によって、集計方法を変更できます。

**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）を定義する列を指定します。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。このオプションは、分析対象のデータが要約されたものである場合に用いてください。

**ID** 回答者を識別する列を指定します。このオプションは、[多重応答 ID別] が選択された場合にのみ必要で、他の応答の種類が入力された場合には使われません。

**By** By 変数の水準ごとに個別のレポートが作成されます。複数の By 変数を指定した場合は、By 変数の水準のすべての組み合わせごとにレポートが作成されます。

## 起動ウィンドウのその他のオプション

追加のオプションは、起動ウィンドウの左下にあります。また、これらのオプションは、起動ウィンドウで [OK] をクリックした後に、「カテゴリカル」の赤い三角ボタンのメニューから選択することもできます。

**グループ化オプション** [X, グループ化カテゴリ] に複数の列が指定されている場合に、どのようにそれらの列を用いるかを定義します。

**組み合わせ** 指定されたグループ列に含まれるグループのすべての組み合わせに対して分析します。最初に指定された列が、クロス表で最も外側に位置するグループとなります。

**それぞれを個別に** 1つ1つのグループ列に対して、個別に分析します。

**両方** グループ列を組み合わせた分析と、グループ列ごとの個別の分析が行われます。

**ID内で一意な値をカウント** 多重応答において、回答者内で重複している同一データを1つとしてカウントします。このオプションを用いるには、[ID] 変数が指定されていなければなりません。

**応答の欠測値をカウント** クロス表とグラフに欠測値をカテゴリとして含めます。欠測値は、統計的検定の比較から除外されます。欠測値は、空のセルまたは「欠測値のコード」列プロパティで定義された欠測コードのいずれかです。なお、欠測値しかない列の場合は、このオプションの状態とは無関係に欠測値がカウントされます。

**応答の水準を降順に並べる** 応答を高い水準から低い水準に並べます。(デフォルトの順序は昇順です。) このオプションは、[X, グループ化カテゴリ] の列には適用されず、応答にのみ適用されます。

**ヒント:** [値の順序] 列プロパティを使うと、表示されるカテゴリの順序を変更できます。『JMPの使用法』の「列情報ウィンドウ」章を参照してください。

**短いラベルの使用** すべてのラベルに共通する接頭部や接尾部を省略して、値ラベルを短くします。

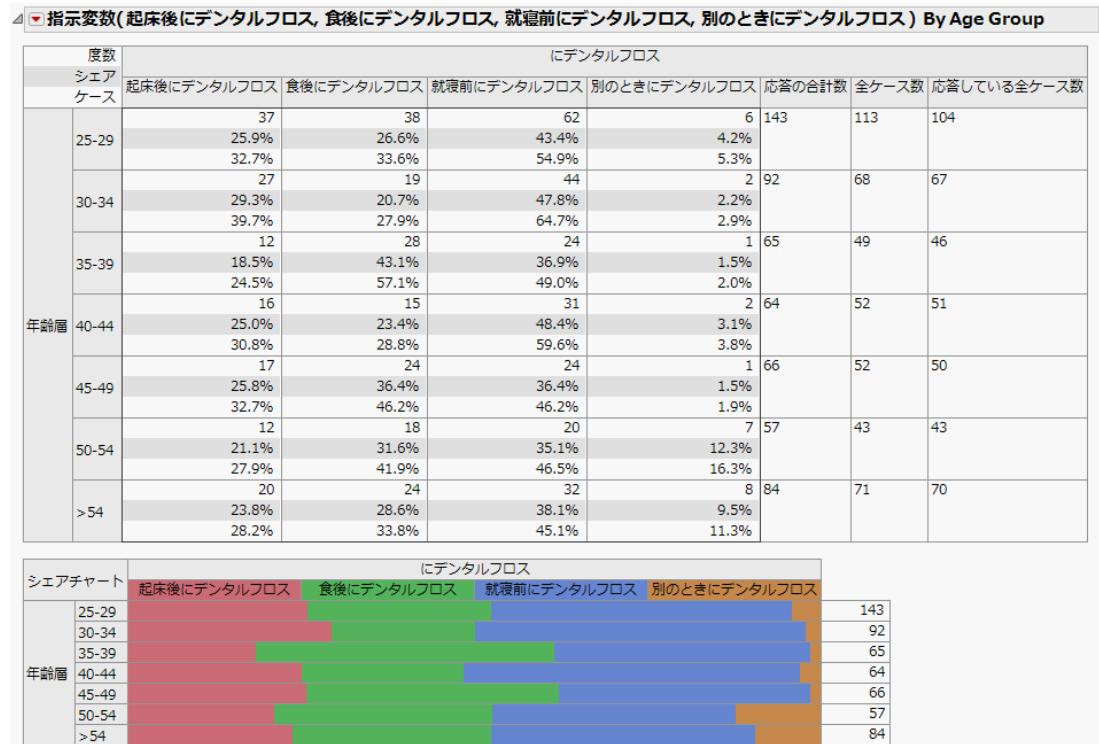
**メモ:** このオプションは、データ値や値ラベルだけに適用されるものであり、列名には適用されません。

**データがない応答を含める** データに存在していないカテゴリを、クロス表に含めます。そのカテゴリは、[値ラベル] 列プロパティには指定しておく必要があります。このオプションは、応答にのみ適用されます。グループ化カテゴリには、データに存在しているカテゴリのみが含まれます。

## 「カテゴリカル」レポート

最初の「カテゴリカル」レポートには、選択した応答列ごとに、二元クロス表（二元度数表）ヒシェアチャートが表示されます。[表の構成] タブを使用している場合、最初のレポートには、クロス表のみが表示されます。

図3.4 最初の「カテゴリカル」レポート



クロス表の左上隅には、クロス表の各セルに含まれている統計量の名前（「度数」、「シェア」、「ケース」）が記載されます。これらの項目は、赤い三角ボタンのメニューにあるオプションを使って削除することができます。

- クロス表の各セルには、「度数」が表示されます。また、クロス表の右端には、それら度数の合計（「応答の合計数」）が表示されます。多重応答がある場合、クロス表の右端に、「全ケース数」（すなわち、行数）、「応答の合計数」、「応答している全ケース数」も表示されます。
- 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。これは、クロス表における行パーセント（転置したクロス表では列パーセント）です。
- （欠測値を除外した）[比率]は、応答の度数を応答している全ケース数で割ったものです。

図3.4には、年齢層ごとの応答数とケース数が表示されています。クロス表の最初の行には、25～29歳の年齢層の結果が記載されています。この行をもとに統計量を説明いたします。

- 最初のセルには、37の応答があります。これは、25～29歳において、「起床後にデンタルフロス」と回答している回答数が、37あることを示しています。
- 25～29歳の年齢層における「応答の合計数」（回答総数）は143です。「起床後にデンタルフロス」の「シェア」は、143の応答のうちの25.9%（37/143）です。言い換えれば、25～29歳においては、デンタルフロスをするうちの25.9%が、起床後でのデンタルフロスであると言えます。

- 25～29歳の年齢層は全部で113人ですが、そのうち32.7% (37/113) が「起床後にデンタルフロス」と答えています。言い換えれば、25～29歳の32.7%が、起床後にデンタルフロスをしています。
- 「応答の合計数」(143) は、「応答しているケース数」(104) よりも多いので、25～29歳の年齢層においては、複数選択している回答者が存在していることが分かります。
- 「全ケース数」(113) は、「応答している全ケース数」(104) よりも多いです。この質問に回答していない25～29歳の回答者が9名いたということになります。

## 「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」の赤い三角ボタンのメニューには、必要に応じてレポートをカスタマイズできるオプションがあります。そこで使用できるオプションは、応答の役割、グループ列を指定したかどうか、起動ウィンドウで選択されたオプションによって異なります。

### レポートに関するオプション

**度数** クロス表において、度数の表示／非表示を切り替えます。この「度数」は、各カテゴリにおける応答の頻度です。

**応答のシェア（「シェア」）** クロス表において、「応答のシェア」の表示／非表示を切り替えます。「応答のシェア」は、全応答数に対して該当の応答数が占める割合です。

**ケースあたりの比率（「ケース」）**（多重応答の場合のみ使用可能。）クロス表において、「ケースあたりの比率」の表示／非表示を切り替えます。「ケースあたりの比率」は、全ケース数（応答しているかどうかに関わらない、すべてのケース数）で、該当の応答数を割ったものです。

**非欠測ケースあたりの比率**（多重応答の場合のみ使用可能。）クロス表において、「非欠測ケースあたりの比率」の表示／非表示を切り替えます。「非欠測ケースあたりの比率」は、応答数を、何らかの応答をしているケースの総数で割ったものです。

**シェアチャート** 分割された棒グラフの表示／非表示を切り替えます。棒の長さは、各グループ内での応答の割合を表しています。棒グラフの右側には、各グループでの全応答数が表示されます。[X, グループ化カテゴリ] に列が使われていない場合、右側には、総応答数が表示されます。

**ヒント：**「値の色」列プロパティを使用して、シェアチャートの色を変更することができます。『JMPの使用法』の「列情報ウィンドウ」章を参照してください。

**度数チャート** [度数チャート] の表示／非表示を切り替えます。各棒は、応答の度数を示しています。スケールはチャート内において同じものが使われています。右端にあるグレーの棒グラフは、各グループでの総応答数を表します。

**ヒント：**「値の色」列プロパティを使用して、度数チャートの色を変更することができます。『JMPの使用法』の「列情報ウィンドウ」章を参照してください。

**度数チャート 転置** 転置した度数チャートの表示／非表示を切り替えます。各棒は、応答の度数を示しています。応答が行で、グループ変数が列に描かれます。各グループでの総応答数が、最後の行にグレーの棒グラフで描かれます。

**クロス表** クロス表の表示／非表示を切り替えます。このクロス表では、列が応答、行が（指定されている場合は）グループとなっています。表の左上のセルに、表の各セルに表示されている統計量の名前（「度数」、「シェア」、「ケース」）が記載されています。なお、[クロス表 転置] を選択すると、この元のクロス表は削除されます。

**クロス表 転置** 転置したクロス表の表示／非表示を切り替えます。この転置したクロス表では、行が応答、列が（指定されている場合は）グループとなっています。表の左上のセルに、表の各セルに表示されている統計量の名前（「度数」、「シェア」、「ケース」）が記載されています。なお、[クロス表] オプションを選択すると、この転置したクロス表は削除されます。

## 統計的検定に関するオプション

統計的検定に関するオプションは、応答の役割、グループ変数があるかどうか、によって使用できるものが異なります。オプションには、応答の等質性に対する検定、関連性の指標、相対リスク、一致性の指標があります。

**多重応答の検定**（1つまたは複数のグループ列が指定されており、かつ、多重応答の場合にのみ使用可能。）  
「[複数の応答をもつ検定の例](#)」（46ページ）を参照してください。各グループ間での応答カテゴリの差について、次の検定が用意されています。

**Poisson 分布の度数検定** Poisson回帰の枠組みで、ケースあたりの比率の差を検定します。ケースあたりの比率が、グループごとに異なるかどうかが検定されます。1ケースあたりで計算される比率がグループ列の水準間で異なるかが、尤度比カイ2乗検定によって計算されます。

**二項分布の等質性検定** 割合の差について尤度比カイ2乗検定を行います。検定においては、各カテゴリに対して、「選択されている」／「選択されていない」の二項分布が仮定されます。

**応答の等質性に対する検定**（1つまたは複数のグループ列があり、かつ、单一選択の応答がある場合。もしくは、[表の構成] タブで、1つまたは複数のグループ列と多重応答がある場合。）応答の等質性に対する検定を行います。実行される検定は、次のように指定によって異なります。

- グループ間における応答の差に対する検定としては、尤度比検定と Pearson カイ2乗検定が実行されます。詳細については、「[応答の等質性に対する検定の例](#)」（45ページ）を参照してください。
- [表の構成] タブで多重応答が指定された場合には、Rao-Scott のカイ2乗検定が実行されます。

**セルのカイ2乗** クロス表の各セルにおける  $p$  値を求めます。これらの  $p$  値は、独立性に対するカイ2乗検定の枠組みで求められています。 $p$  値が小さい場合は、そのセルにおいて、「行と列とが独立である」という仮定のもとでの期待度数よりも、観測度数が大きいかまたは小さいことを示します。 $p$  値は、観測度数が期待度数よりも大きいかまたは小さいかに応じて、色分けされ、濃淡がつけられます。詳細については、「[セルのカイ2乗検定の例](#)」（48ページ）を参照してください。

**標本の各ペア比較** (1つまたは複数のグループ列があり、かつ、単一選択の応答がある場合にのみ使用可能。)

グループのペアごとに、応答の等質性に対する検定（行と列との独立性に対する検定）を行います。尤度比カイ2乗検定とPearsonのカイ2乗検定を行います。詳細については、「[比較を表すアルファベット文字による、標本の各ペア比較の例](#)」(49ページ) を参照してください。

**セルの各ペア比較** (1つまたは複数のグループ列があり、かつ、単一選択の応答または多重応答がある場合にのみ使用可能。) グループのペアごとに、応答の特定のカテゴリを比較します。このとき、特定のカテゴリ以外のカテゴリが、すべて併合されて検定が計算されます。尤度比カイ2乗検定、Pearsonのカイ2乗検定、Fisherの正確検定を行います。詳細については、「[比較を表すアルファベット文字による、セルの各ペア比較の例](#)」(50ページ) を参照してください。

**相対リスク** (グループ列が2水準であり、かつ、応答も2水準の場合。もしくは、グループ列が2水準であり、かつ、応答が多重応答であり、かつ、[ID内で一意な値をカウント] オプションが選択されている場合。) 相対リスクを計算します。詳細については、「[条件付き関連と相対リスクの例](#)」(53ページ) を参照してください。

**条件付き関連** ([ID内で一意な値をカウント] オプションが選択されている場合。) 特定の応答水準で条件付けたもとで、ある応答水準が出現する条件付き確率を求めます。詳細については、「[条件付き関連と相対リスクの例](#)」(53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] 応答の場合にのみ使用可能。) 一致性のカッパ係数を計算し、また、対称性に対するBowkerの検定を行います。詳細については、「[判定の一致性の例](#)」(56ページ) を参照してください。

**遷移レポート** ([反復測定] 応答の場合にのみ使用可能。) 時間経過に伴う応答の変化に関して、遷移度数と遷移率行列を求めます。詳細については、「[反復測定の例](#)」(57ページ) を参照してください。

**検定オプション** このメニューで使用できるオプションは、選択された分析によって異なります。

**カイ2乗検定の選択** 単一選択の応答に対するカイ2乗検定の種類を指定します。[尤度比とPearson]、[尤度比のみ]、[Pearsonのみ] から選択できます。

**警告の表示** カイ2乗検定の標本サイズが小さい場合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。

**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。

## 「カテゴリカル」プラットフォームのその他のオプション

**応答の合計数** クロス表とシェアチャートにおいて、総度数の表示／非表示を切り替えます。グループ列が使われている場合、この総度数はグループごとに集計されます。

**応答水準の表示** クロス表とシェアチャートにおいて、応答列におけるカテゴリの表示／非表示を切り替えます。

**上位カテゴリの表示** (1つまたは複数の「上位カテゴリ」が定義されている場合のみ使用可能。) クロス表と度数チャートにおいて、上位カテゴリの列の表示／非表示を切り替えます。上位カテゴリの詳細については、「[上位カテゴリ](#)」(43ページ) を参照してください。

---

**ヒント:** このオプションは、上位カテゴリの表示／非表示を切り替えるものです。上位カテゴリ内にある下位カテゴリを非表示にするには、「上位カテゴリ」列プロパティで「非表示」オプションを用いてください。または、「カテゴリカル」レポートの赤い三角ボタンメニューより「応答水準の表示」オプションをオフにすると、上位カテゴリだけが表示され、下位にあるすべての応答水準が非表示になります。

**全ケース数** (多重応答の場合のみ使用可能。) クロス表において、各グループにおけるケース数 (回答者数) の表示／非表示を切り替えます。

**応答している全ケース数** (多重応答列の場合のみ使用可能。) クロス表において、少なくとも1つは応答のあるケース (回答者) の度数の表示／非表示を切り替えます。全く応答しなかった人は、この数には含まれません。「応答している全ケース数」は、必ず、「全ケース数」以下です。

**スコアの平均** クロス表とシェアチャートにおいて、全体またはグループごとの平均の表示／非表示を切り替えます。この平均は、それぞれの応答変数のカテゴリに割り当てられた数値に基づいて計算されます。

- 数値の場合には、そのままスコアとして使われます。
- 数値以外の場合には、「値スコア」列プロパティで各カテゴリに割り当てられた数値が、スコアとして使われます。
- 「値スコア」列プロパティも指定されていない場合には、1から順に振られた通し番号がスコアとして使われます。

詳細については、「[比較のアルファベット文字を持つスコア平均の例](#)」(63ページ) を参照してください。

**スコア平均の比較** クロス表において、「平均の比較」列の表示／非表示を切り替えます。Satterthwaite の *t* 検定(プールしない誤差分散による *t* 検定)によって、グループ間のスコア平均を比較します。SAS Institute (2009) を参照してください。比較の結果は、アルファベット文字を使って表示されます。比較を表すアルファベット文字の詳細については、「[比較を表すアルファベット文字](#)」(42ページ) を参照してください。比較グループを指定する方法については、「[比較を表すアルファベット文字による、ユーザ指定の比較の例](#)」(52ページ) を参照してください。

**スコアの標準偏差** クロス表において、全体もしくはグループごとの標準偏差を含む列の表示／非表示を切り替えます。

**スコア平均で並べ替え** (結果に複数の応答があり、かつ、グループ列がない場合にのみ使用可能。) 複数の応答に対するレポートを、スコア平均順に並べます。

**テーブルの保存** レポートの特定の部分を、新しいデータテーブルに保存します。各オプションを実行すると、該当の情報を含んだデータテーブルが作成されます。このメニューで使用できるオプションは、選択された分析によって異なります。

---

**メモ:** 上位カテゴリは、新しいテーブルに含まれません。

**度数の保存** クロス表の「度数」を、新しいデータテーブルに保存します。

**応答のシェアの保存** クロス表の「応答のシェア」を、新しいデータテーブルに保存します。

**ケースあたりの比率の保存** クロス表の「ケースあたりの比率」を、新しいデータテーブルに保存します。

**転置した度数の保存** 転置したクロス表の「度数」を、新しいデータテーブルに保存します。

**転置した応答のシェアの保存** 転置したクロス表の「応答のシェア」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** 転置したクロス表の「ケースあたりの比率」を、新しいデータテーブルに保存します。

**比率検定の保存** [多重応答の検定] 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。

**スコア平均の保存** 各標本グループのスコアの平均を、新しいデータテーブルに保存します。

**t検定とp値の保存** 「スコア平均の比較」レポートのt検定統計量とp値を、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、標本水準を列として、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が共有されます。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。

**フィルタ** 特定のレポートのデータをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。このとき、応答を1つも含まない標本の水準は、表示されなくなります。そのような表示されなくなった水準を表示するには、起動ウィンドウで [データにない応答を含める] を選択します。または、「環境設定の変更」の赤い三角メニューを選択し、[データにない応答を含める] を選択します。

**要約の目次** 「カテゴリカル」レポート最上部の「要約の目次」レポートの表示／非表示を切り替えます。「要約の目次」では、すべての検定とスコア平均を要約にまとめて、関連レポートへのリンクとともに表示されます。

**レポートで使用された列を表示** 「レポートで使用された列」という情報の表示／非表示を切り替えます。このオプションは、「SPSS名」や「SAS名」、または「SPSSラベル」や「SASラベル」の列プロパティを持つ列のみに影響します。SASやSPSSからデータセットをインポートすると、データの変数名とラベル名が、自動的にJMPデータテーブルの列プロパティに追加されます。また、「その他」の列プロパティを使って、「SAS名」や「SPSS名」または「SASラベル」や「SPSSラベル」の列プロパティを追加することもできます。たとえば、「SAS名」や「SPSS名」の列プロパティを使って、短い列名を与えることができます。

**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、度数は固定小数点形式（総桁数7、小数桁数0）、シェアと比率はパーセント値形式（総桁数6、小数桁数1）です。

**複数行に配置** レポートを、ページの下に移動させずに横方向に配置します。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** 次回以降に「カテゴリカル」プラットフォームを起動したときのために、現在や次回以降の JMP セッションでの環境設定を変更することができます。詳細については、「環境設定の変更」(44 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ます。これらのオプションがここで選択されると、その新しい設定でプラットフォームが更新されます。「カテゴリに関するオプション」の詳細については、「起動ウィンドウのその他のオプション」(34 ページ) を参照してください。

**クロス表を濃淡表示** 環境設定で濃淡表示が指定されていなくても、クロス表を濃淡表示にします。このオプションが選択されていない場合、クロス表レポートは [テーブル行の濃淡表示] の現在の環境設定に合わせて濃淡表示されます。

**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。

以下のオプションについて詳しくは、『JMP の使用法』の「JMP のレポート」章を参照してください。

**ローカルデータフィルタ** 現在のレポートにおいてデータをフィルタリングするローカルデータフィルタを表示するか、非表示にします。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、[自動再計算] オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウで By 変数を指定した場合のみ使用可能です。

## クロス表の応答におけるオプション

**文字の表示** クロス表において、列名を識別するアルファベット文字の表示／非表示を切り替えます。これらのアルファベット文字は、統計的検定の結果を表すのに、どのグループとどのグループに差があるかを示すのに使われます。

**比較の設定** 検定において、どのグループとどのグループを比較するかを指定できます。スラッシュで区切られたグループが比較されます。各比較はカンマで区切ってください。グループは、アルファベットで指定してください。たとえば、A を E と、B を D と、C を F と、比較したい場合には、「A/E, B/D, C/F」のようにグループを指定します。[セルの各ペア比較] レポートに、指定された比較の結果が表示されます。詳細については、「比較を表すアルファベット文字による、ユーザ指定の比較の例」(52 ページ) を参照してください。

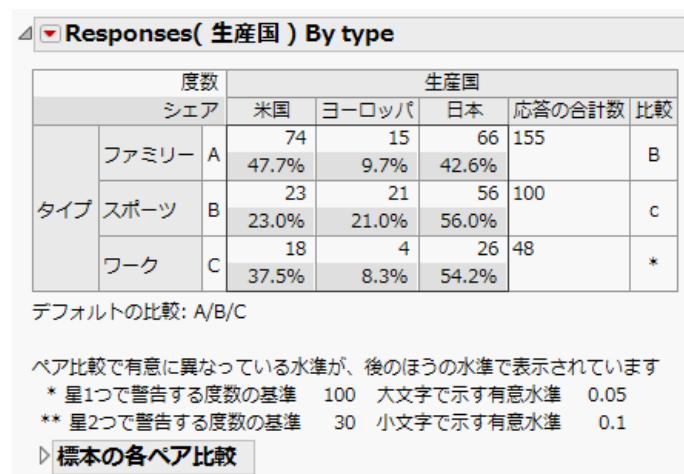
**削除** レポートウィンドウからレポートを削除します。

**注意:** [削除] オプションを元に戻すことはできません。

## 比較を表すアルファベット文字

[セルの各ペア比較]、[標本の各ペア比較]、[スコア平均の比較] オプションの結果では、それぞれの標本グループを A、B、C といったアルファベット文字で表しています。27 番目以降の標本グループには、「A1」、「B1」のように、アルファベット文字の後に数字が付けられます。比較に関するオプションが使われた場合、このアルファベット文字がクロス表の水準見出しに表示されます。

図3.5 比較の結果を表すアルファベット文字のあるクロス表



グループ間に有意差がある場合、シェアが小さいほうのグループのアルファベット文字が、もう一方のグループのセルに表示されます。アルファベットが大文字となっている場合、小文字よりも有意性が高いことを示します。デフォルトでは、p 値が 0.05 以下のときに大文字で、0.10 以下のときに小文字で表示されます。図3.5 における「B」は、生産国の割合における「スポーツ」と「ファミリー」との違いが、有意水準 5% で有意になっていることを示しています。「ファミリー」の合計数 (155) は、「スポーツ」の合計数 (100) よりも多いため、「ファミリー」の行に「B」と表示されています。「スポーツ」行の「c」は、「スポーツ」と「ワーク」を比較した場合、生産国の割合には有意水準 10% で有意差があることを示しています。「スポーツ」の応答数 (100) は、「ワーク」の応答数 (48) よりも多いため、この「c」は、「スポーツ」行に表示されます。

度数が小さい場合の警告も、比較のセルに表示されます。1つのアスタリスク (\*) は、その水準に含まれる応答が 100 より少ないことを示し、2つのアスタリスク (\*\*) は、30 よりも少ないことを示します。図3.5 では、「ワーク」行の合計数は 48 で、アスタリスクが 1 つ表示されています。有意水準や警告で使われる基準は、「カテゴリカル」プラットフォームの環境設定で変更できます。環境設定の変更についての詳細は、「環境設定の変更」(44 ページ) を参照してください。

「比較を表すアルファベット文字による、標本の各ペア比較の例」(49ページ)、「比較を表すアルファベット文字による、セルの各ペア比較の例」(50ページ)、および「比較のアルファベット文字を持つスコア平均の例」(63ページ)を参照してください。

## 上位カテゴリ

「上位カテゴリ」とは、いくつかのカテゴリをまとめたものです。たとえば、5段階評価を使用している場合に、上位2段階（上から2つの水準）の割合を調べたいとします。そのようなときは、「上位カテゴリ」列プロパティによって、それらのカテゴリを併合した上位カテゴリを定義することができます。

上位カテゴリは、クロス表と度数チャートには表示されます。しかし、上位カテゴリはシェアチャートには表示されません。また、グループ列には適用されません。

上位カテゴリを作成するには、次の手順に従います。

1. データテーブルの中で、1つにまとめたいカテゴリを含んでいる列を選択します。
2. [列] > [列情報] を選択します。
3. [列プロパティ] をクリックし、[上位カテゴリ] を選択します。
4. (オプション) デフォルトで表示される上位カテゴリの名前は、変更が可能です。
5. 列の「カテゴリ」リストから1つまたは複数のカテゴリを選択します。
6. [追加] をクリックします。
7. (オプション) 上位カテゴリを選択し、「上位カテゴリ」の赤い三角ボタンのメニューをクリックして追加のオプションを表示します。

### [上位カテゴリ] のオプション

列プロパティウィンドウにある「上位カテゴリ」の赤い三角ボタンのメニューには、次のオプションがあります。

**非表示** クロス表と度数チャートにおいて、上位カテゴリ内のカテゴリを非表示にします。

**ヒント:** レポート内で下位カテゴリの表示／非表示を切り替えたい場合には、この [非表示] オプションを使わないでください。「カテゴリカル」の赤い三角ボタンのメニューにある [応答水準の表示] オプションを使ってください。

**重複なし** (多重応答列の場合にのみ使用可能。) 下位のカテゴリが、複数の上位カテゴリで表示される場合に、重複して数えられるのを防ぎます。

**平均を追加** 平均を含む列がクロス表に追加されます。

**標準偏差を追加** 標準偏差を含む列がクロス表に追加されます。

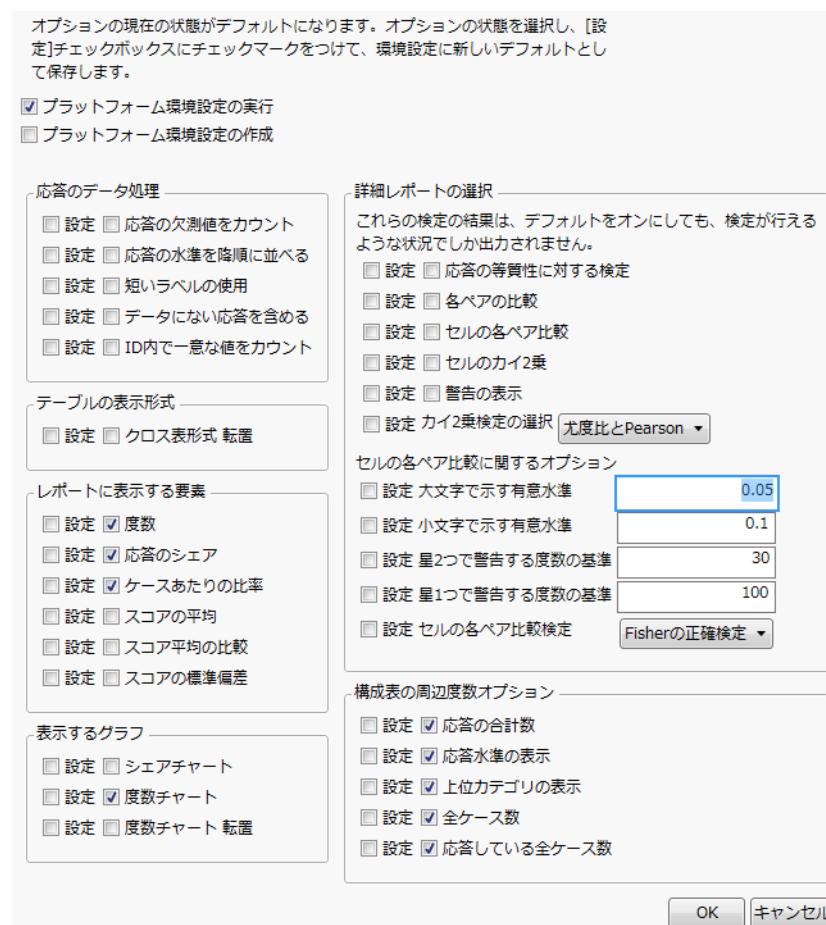
**すべて追加** 応答の合計数がクロス表に含まれます。なお、デフォルトでも、「応答の合計数」列は常に含まれています。

**メモ:** 「上位カテゴリ」の機能は、[反復測定] と [判定者の一致性] を除くすべての応答効果で使用できます。自然なスコアを応答に設定できない場合、[平均] と [標準偏差] のオプションはサポートされません。

## 環境設定の変更

「カテゴリカル」の赤い三角ボタンのメニューには、環境設定を指定できる [環境設定の変更] オプションがあります。

図3.6 環境設定の設定ウィンドウ



設定を変更したいオプションの【設定】チェックボックスを選択してください。このとき、デフォルトで実行したいオプションは、そのオプション名のすぐ左にあるチェックボックスを選択してください。逆に、デフォルトで実行たくないオプションは、そのチェックボックスの選択を解除してください。設定した変更を環境設定に適用したい場合には、【プラットフォーム環境設定の実行】ボックスを選択してください。また、環境設定を変更するスクリプトを生成したい場合には、【プラットフォーム環境設定の作成】ボックスを選択します。【プラットフォーム環境設定の実行】を選択すると、次回以降に、「カテゴリカル」プラットフォームを起動したときに、設定した環境設定が使われます。

---

**メモ:** 【プラットフォーム環境設定の作成】で生成されたスクリプトを実行すると、そこで設定がプラットフォーム環境設定に適用されます。このスクリプトを使えば、変更した環境設定を、複数のユーザーで共有したり、自分が使いたいときに利用したりできます。

---

## 「カテゴリカル」プラットフォームの別例

ここでは、次のような例を紹介します。

- ・「応答の等質性に対する検定の例」(45ページ)
- ・「複数の応答をもつ検定の例」(46ページ)
- ・「セルのカイ2乗検定の例」(48ページ)
- ・「比較を表すアルファベット文字による、標本の各ペア比較の例」(49ページ)
- ・「比較を表すアルファベット文字による、セルの各ペア比較の例」(50ページ)
- ・「比較を表すアルファベット文字による、ユーザ指定の比較の例」(52ページ)
- ・「条件付き関連と相対リスクの例」(53ページ)
- ・「判定の一致性の例」(56ページ)
- ・「反復測定の例」(57ページ)
- ・「[多重応答] タブの例」(58ページ)
- ・「比較のアルファベット文字を持つスコア平均の例」(63ページ)
- ・「「表の構成」レポートの例」(64ページ)

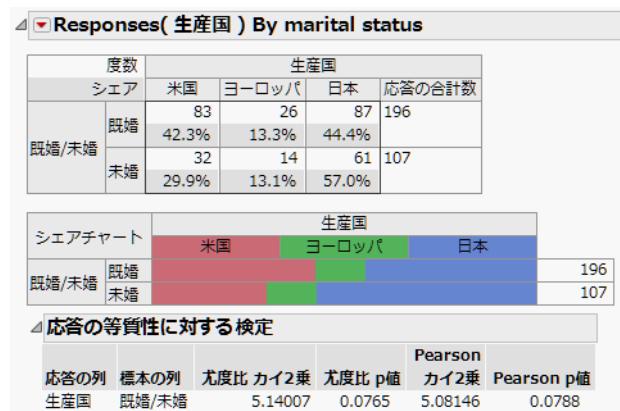
### 応答の等質性に対する検定の例

ここでは、自家用車に関する調査データである「Car Poll.jmp」データテーブルを使用します。このデータには、回答者の属性情報と所有する車についての情報が含まれています。既婚／未婚と車の生産国との間の関係を調べてみましょう。また、応答の等質性に対する検定も行ってみましょう。つまり、車の生産国分布が、既婚と未婚の回答者で同じなのかどうかを調べます。

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乗統計量とp値が計算されます。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [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て、[応答の等質性に対する検定] を選択します。

図3.7 応答の等質性に対する検定



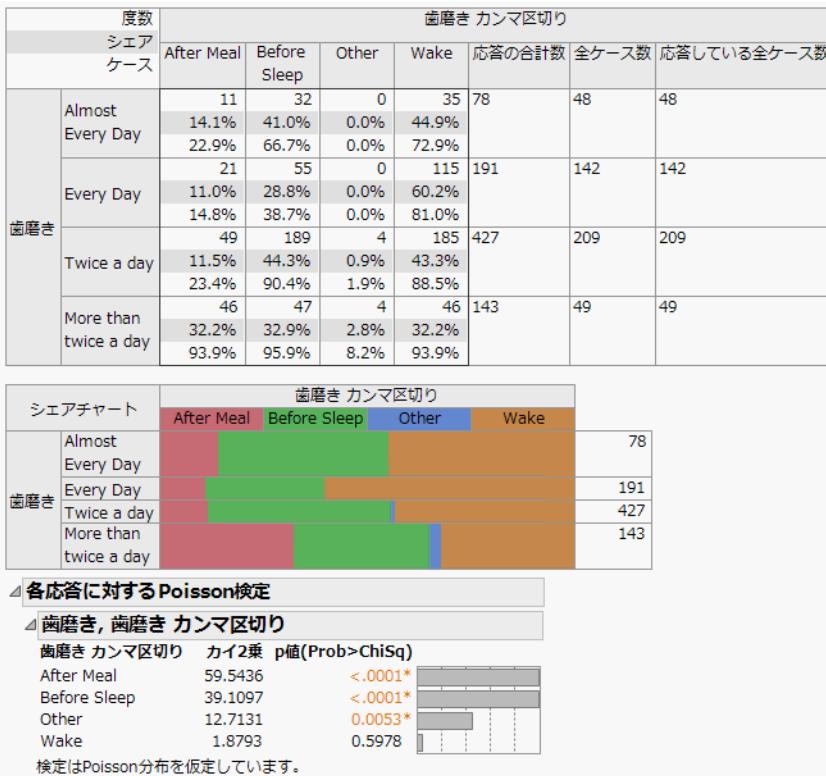
「シェアチャート」を見ると、既婚グループが保有する車は、米国車と日本車の間で均等に分かれています。未婚グループでは、日本車が最も多く保有されています。応答の等質性に対する検定において、p値は約0.08です。未婚と既婚との差は、有意水準を0.05としたとき、統計的に有意ではありません。

## 複数の応答をもつ検定の例

この例では「Consumer Preferences.jmp」を使用します。このデータには、口腔衛生に関する質問への回答が含まれています。それぞれの歯磨き時間帯（「歯磨き カンマ区切り」）に対する応答率が、「歯磨き」グループ間で同じであるかどうかを、[多重応答の検定] オプションを使って検定してみましょう。この「歯磨き」グループは、歯磨きを行う頻度で分類されているものです。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Consumer Preferences.jmp」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. 「歯磨き カンマ区切り」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. 「歯磨き」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て、[多重応答の検定] > [Poisson 分布の度数検定] を選択します。

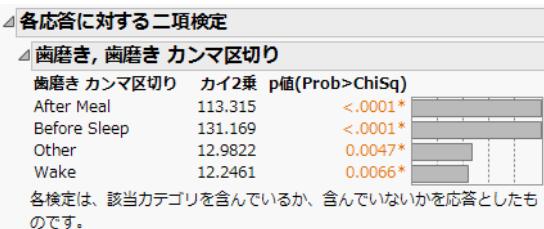
図3.8 多重応答の検定、Poisson分布



*p*を見ると、「After Meal (食後)」、「Before Sleep (就寝前)」、「Other (その他)」の応答率において、「歯磨き」グループ間で有意差があります。「Wake (起床後)」においては、有意差は見られません。クロス表を見ると、歯磨きの頻度とは関係なく、大半の人が起床後には歯磨きをしています。

7. 「カテゴリカル」の赤い三角ボタンをクリックして、[多重応答の検定] > [二項分布の等質性検定] を選択します。

図3.9 多重応答の検定、二項分布



【二項分布の等質性検定】オプションでは、[Poisson分布の度数検定]よりも、必ず検定統計量が大きくなります(つまり、p値が小さくなります)。【二項分布の等質性検定】は、応答が生じた回数(起床後に歯磨きをすると答えた人の数)だけを考慮するのではなく、応答が生じなかった回数(起床後に歯磨きをすると答えなかった人の数)も考慮します。

この例では、各応答(「After Meal(食後)」、「Before Sleep(就寝前)」、「Wake(起床後)」、「Other(その他)」)の割合は、年齢層間で異なっています。それらのp値は、0.05よりも小さくなっています。

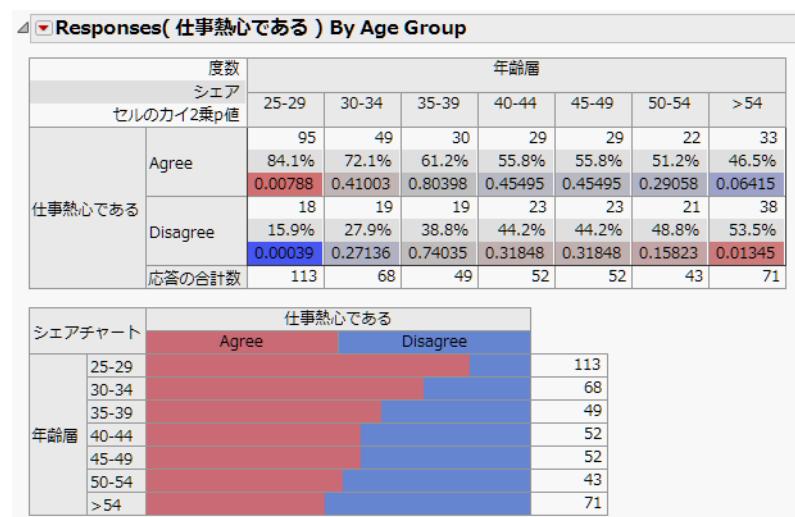
**メモ:** JMPは、「多重応答」の尺度または「多重応答」の列プロパティを持つ列を多重応答列として認識します。

## セルのカイ2乗検定の例

この例では、「Consumer Preferences.jmp」を使用します。このデータには、一般的な事柄に対する意見や態度についてが含まれています。年齢層によって「仕事熱心である」かどうかの分布が異なるかどうかを調べてみましょう。

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「Consumer Preferences.jmp」を開きます。
- [分析] > [消費者調査] > [カテゴリカル] を選択します。
- 「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。
- 「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。
- [OK] をクリックします。
- 「カテゴリカル」の赤い三角ボタンをクリックして、[クロス表 転置] を選択します。
- 「カテゴリカル」の赤い三角ボタンをクリックして、[セルのカイ2乗] を選択します。

図3.10 セルのカイ2乗



各セルにおいて、p値が小さい場合は、そのセルの観測度数と期待度数との間に有意差があることを示しています。また、各セルの色は、p値を基準にして、期待値よりも有意に度数が大きいほど濃い赤に、期待値よりも有意に度数が小さいほど濃い青に色分けされます。なお、各セルの期待度数は、観測された行和と列和から算出されます。

たとえば、25～29歳で「Agree（同意する）」と回答する人の期待度数は、 $(287 \times 113) / 448 = 72.4$ 人で、実際の観測度数は95人です。そのp値は0.00788ですので、観測度数は期待度数よりも有意に大きいと言えます。「仕事熱心である」に同意する25～29歳の人数は、独立であるとの仮定から算出された期待度数よりも有意に多い、ということが分かります。

比較を表すアルファベット文字による、標本の各ペア比較の例

この例では、「Consumer Preferences.jmp」を使用します。このデータには、一般的な事柄に対する意見や態度が含まれています。年齢層によって「仕事熱心である」という質問に対する回答の分布が異なるかどうかを調べてみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Consumer Preferences.jmp」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て、[標本の各ペア比較] を選択します。

図3.11 標本の各ペア比較

Responses( 仕事熱心である ) By Age Group						
年齢層	度数	仕事熱心である			応答の合計数	比較
		シェア	Agree	Disagree		
年齢層	25-29	A	95 84.1%	18 15.9%	113	b,C,D,E,F,G
	30-34	B	49 72.1%	19 27.9%	68	d,e,F*
	35-39	C	30 61.2%	19 38.8%	49	*
	40-44	D	29 55.8%	23 44.2%	52	*
	45-49	E	29 55.8%	23 44.2%	52	*
	50-54	F	22 51.2%	21 48.8%	43	*
	>54	G	33 46.5%	38 53.5%	71	B*

1996-03-27 10:00:00

ペア比較で有意に異なる水準が、後のほうの水準で表示されてい

\* 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

#### △標本の各ペア比較

英文字は、Pearson 力へ

ペアに対する尤度比カイ2乗のp値								ペアに対するPearsonカイ2乗のp値							
	A	B	C	D	E	F	G		A	B	C	D	E	F	G
A	1.0000	0.0552	0.0020	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	A	1.0000	0.0523	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>
B	0.0552	1.0000	0.2183	0.0641	0.0641	<b>&lt;.0001</b>	<b>&lt;.0001</b>	B	0.0523	1.0000	0.2170	0.0638	0.0638	<b>&lt;.0001</b>	<b>&lt;.0001</b>
C	0.0020	0.2183	1.0000	0.5781	0.5781	0.3312	0.1108	C	0.0015	0.0217	1.0000	0.5783	0.5783	0.3314	0.1109
D	<b>&lt;.0001</b>	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083	D	<b>&lt;.0001</b>	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
E	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083	E	<b>&lt;.0001</b>	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
F	<b>&lt;.0001</b>	0.0261	0.0312	0.6540	0.6540	1.0000	0.6276	F	<b>&lt;.0001</b>	0.0255	0.3314	0.6540	0.6540	1.0000	0.6276

クロス表には、各年齢層における「仕事熱心である」の回答が要約されています。クロス表の各セルには、年齢層ごとに、「Agree (同意する)」と「Disagree (反対する)」の度数とシェア（割合）が表示されています。また、比較の検定結果を示すアルファベット文字も表示されています。アルファベット文字がどのグループを示しているかは、グループ名のすぐ右側の列に表示されています。また、一番右側にある比較列では、それらのアルファベット文字によって、検定結果が表示されています。

〔標本の各ペア比較〕アウトラインには、ペアごとの比較に対して、尤度比検定と Pearson カイ2乗検定の **p** 値が表示されています。これらの **p** 値は、アルファベット文字のラベルを付加された行列の形式で表示されています。

この例の場合、次のようなことがわかります。

- 25～29歳の比較列には、すべてのアルファベット文字 (b～g) が含まれています。つまり、25～29歳層は、他の年齢層と比較して、「仕事熱心である」という質問に対する回答には有意差があります。アルファベット文字の **b** は小文字なので、25～29歳と30～34歳との違いにおける有意水準は、0.10です。他のアルファベット文字は大文字なので、有意水準0.05で差があります。
- 54歳以上の年齢層 (G) は、30～34歳 (B) と有意差があります。グループ G の応答数は、グループ B の応答数よりも多いので (71対68)、比較を表すアルファベット文字 (B) が、G の年齢層のほうに表示されています。
- 比較セルにある 1 つのアスタリスクは、標本サイズが小さいことを表す警告です。1 つのアスタリスクがある場合、そのグループの度数が、30 以上で、100 未満であることを示します。
- この例では見られませんが、2 つのアスタリスクがある場合、グループサイズが 30 よりも少ないことを示します。

## 比較を表すアルファベット文字による、セルの各ペア比較の例

この例では、「Consumer Preferences.jmp」を使用します。このデータには、一般的な事柄に対する意見や態度が含まれています。「被雇用期間」グループによって、仕事に対する満足度の分布が異なるかどうかを調べてみましょう。

- 〔ヘルプ〕 > 〔サンプルデータライブラリ〕を選択し、「Consumer Preferences.jmp」を開きます。
- 〔分析〕 > 〔消費者調査〕 > 〔カテゴリカル〕を選択します。
- 「仕事に対する満足度」を選択して、〔単純〕タブの「応答」をクリックします。
- 「被雇用期間」を選択して、〔X, グループ化カテゴリ〕をクリックします。
- 〔OK〕をクリックします。
- 「カテゴリカル」の赤い三角ボタンのメニューをクリックして、〔セルの各ペア比較〕を選択します。

図3.12 セルの各ペア比較

Responses( 仕事に対する満足度 ) By Employee Tenure

		仕事に対する満足度				応答の合計数
度数		Not at all satisfied	Somewhat satisfied	Extremely satisfied		
シェア比較						
被雇用期間	less than 5 years	A 11 6.9%	97 60.6%	52 32.5%	160	
	5 to 10 years	B 13 9.4%	72 51.8%	54 38.8%	139	
	10 to 20 years	C 4 4.7%	43 50.0%	39 45.3%	86	
	more than 20 years	D 4 6.3%	30 47.6%	29 46.0%	63	

デフォルトの比較: A/B/C/D

ペア比較で有意に異なる水準が、後のほうの水準で表示されています  
 \* 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 小文字で示す有意水準 0.1

仕事に対する満足度

シェアチャート		Not at all satisfied	Somewhat satisfied	Extremely satisfied	
被雇用期間	less than 5 years				160
	5 to 10 years				139
	10 to 20 years				86
	more than 20 years				63

セルの各ペア比較 - 詳細

英文字は、Fisherの正確検定の結果を表しています。

被雇用期間, 仕事に対する満足度

尤度比 ペア比較

	AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.4323	1.0000	0.4781	0.1806	1.0000	0.8872	0.4658	0.6515
Somewhat satisfied	1.0000	0.1246	1.0000	0.1092	0.7931	1.0000	0.0783	0.5820	0.7739
Extremely satisfied	1.0000	0.2526	1.0000	0.0476	0.3366	1.0000	0.0607	0.3377	0.9341

Pearson ペア比較

	AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.4316	1.0000	0.4870	0.1948	1.0000	0.8878	0.4763	0.6496
Somewhat satisfied	1.0000	0.1246	1.0000	0.1086	0.7931	1.0000	0.0774	0.5820	0.7740
Extremely satisfied	1.0000	0.2524	1.0000	0.0465	0.3360	1.0000	0.0585	0.3364	0.9341

Fisher正確 ペア比較

	AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.5235	1.0000	0.5853	0.2990	1.0000	1.0000	0.5912	0.7221
Somewhat satisfied	1.0000	0.1302	1.0000	0.1373	0.8909	1.0000	0.0983	0.6493	0.8685
Extremely satisfied	1.0000	0.2764	1.0000	0.0531	0.4034	1.0000	0.0649	0.3574	1.0000

ペア比較の **p** 値が、表形式で表示されます。これらの **p** 値は、尤度比検定、Pearson のカイ 2 乗検定、および Fisher の正確検定のものです。**p** 値の表には、アルファベット文字によって、どのグループとどのグループが比較されているかが示されています。クロス表のグループ名のすぐ右側に、そのグループを示すアルファベット文字が表示されています。また、クロス表のセルには、検定結果を示すアルファベット文字が表示されています。

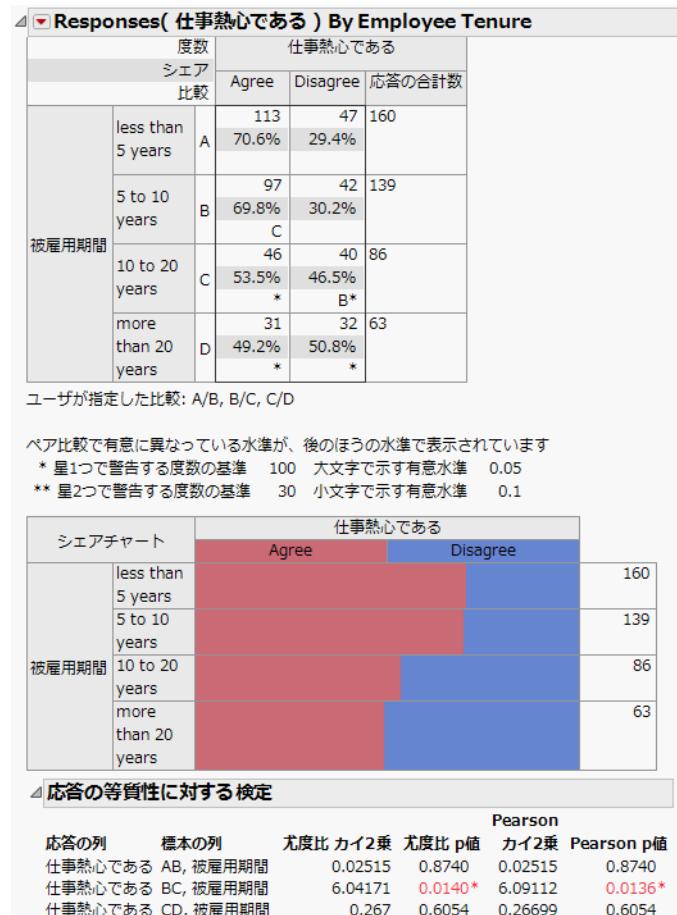
勤続年数 5 年未満の回答者は、勤続年数 20 年の回答者よりも、「Somewhat satisfied (やや満足)」の割合が高いです。このことは、クロス表の最初の行における「Somewhat satisfied」のセルに、「d」と示されています。このことから分かります。また、クロス表の最終行にある「Extremely satisfied (かなり満足)」のセルには、「a」と示されています。D グループは、A グループよりも、「Extremely satisfied」の割合が高いです。なお、検定結果を示すアルファベット文字は、応答のシェア (割合) が高いほうのセルに配置されます。

## 比較を表すアルファベット文字による、ユーザ指定の比較の例

この例では、「Consumer Preferences.jmp」を使用します。このデータには、一般的な事柄に対する意見や態度が含まれています。「被雇用期間」グループによって、「仕事熱心である」かどうかの分布が異なるかどうかを調べてみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Consumer Preferences.jmp」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. 「被雇用期間」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. 「Responses (応答)」の赤い三角ボタンのメニューをクリックし、[文字の表示] を選択します。
7. 「Responses (応答)」の赤い三角ボタンのメニューをクリックし、[比較の設定] を選択します。
8. 「A/B, B/C, C/D」と入力します。
9. 「カテゴリカル」の赤い三角ボタンのメニューをクリックして、[応答の等質性に対する検定] を選択します。

図3.13 比較の設定の例



応答の等質性に対する検定では、AとB、BとC、そしてCとDを比較します。グループB(5~10年)は、グループC(10~20年)よりも「仕事熱心である」という質問に同意する割合が高いです。Pearsonのp値が0.0136であるため、これらの割合には有意な差があると言えます。

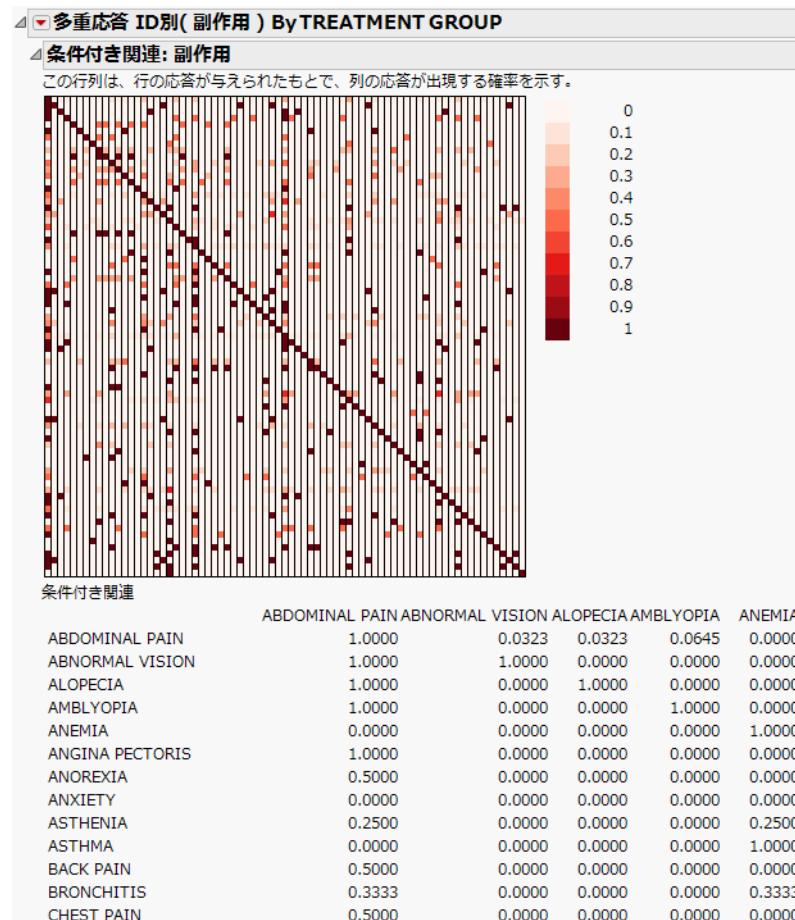
## 条件付き関連と相対リスクの例

この例では「AdverseR.jmp」を使用します。このデータには、臨床試験での副作用についての情報が含まれています。このデータを使用して、有害事象(副作用)の条件付き関連を調べ、処置群におけるそれらの事象の相対リスクを対照群と比較してみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「AdverseR.jmp」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. 「副作用」を選択して、[多重] タブの [多重応答 ID別] をクリックします。

4. 「処置グループ」を選択して [X, グループ化カテゴリ] をクリックします。
5. 「患者のID」を選択し、[ID] をクリックします。
6. [ID内で一意な値をカウント] を選択して、[OK] をクリックします。
7. 「カテゴリカル」の赤い三角ボタンのメニューをクリックして、[条件付き関連] を選択します。

図3.14 「条件付き関連」レポート（部分レポート）



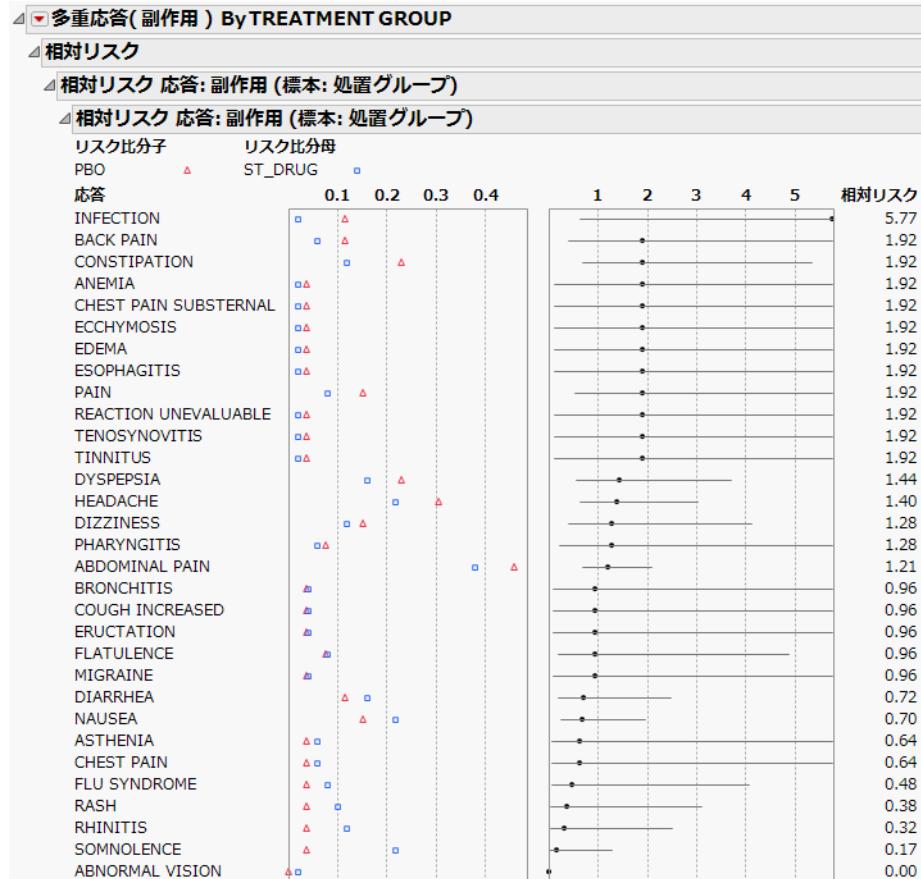
条件付き関連行列は、ある副作用が発生したときの、別の副作用が発生する条件付き確率を示します。これらの条件付き確率は、全グループにおけるものです。腹痛（ABDOMINAL PAIN）が報告された患者においては、0.0323の確率で視覚異常（ABNORMAL VISION）が報告されています。

**ヒント:** ヒートマップにポインタを合わせると、条件付き確率が表示されます。

8. 「カテゴリカル」の赤い三角ボタンのメニューをクリックして、[相対リスク] を選択します。
9. 表示されるウィンドウで [PBO] を選択して、[OK] をクリックします。

10. ウィンドウ内の「相対リスク」レポートを右クリックして [列の値で並べ替え] を選択します。
11. 「相対リスク」を選択して、[OK] をクリックします。

図3.15 「相対リスク」レポート（一部）



[相対リスク] オプションによって、各応答の相対リスクが計算されます。デフォルトの「相対リスク」レポートには、応答名、各グループのリスク（割合）、相対リスク、相対リスクの95%信頼区間が表示されます。この例では、副作用の相対リスクが計算されています。感染症（INFECTION）が生じる、「ST\_DRUG」に対する「PBO」の相対リスクは、5.7倍です。ただし、信頼区間が非常に広く、その信頼区間には1.0の相対リスクが含まれています。相対リスクが1.0となるのは、各グループのリスクが等しい場合です。

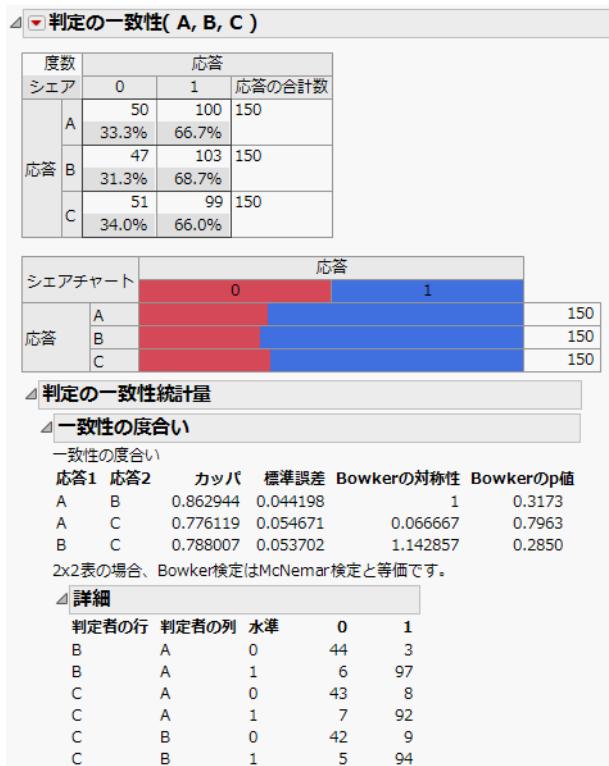
なお、[列] > [下側95%] および [列] > [上側95%] を右クリックして選択すると、相対リスク推定値の95%信頼区間がレポートに追加されます。

## 判定の一致性の例

この例では「Attribute Gauge.jmp」を使用します。このデータは、3名の判定者が、50個の部品を3回、0／1で判定したものです。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Attribute Gauge.jmp」を開きます。
2. [分析] > [消費者調査] > [カテゴリカル] を選択します。
3. A、B、およびCを選択します。
4. [関連] タブで、[判定の一致性] をクリックします。
5. [OK] をクリックします。

図3.16 「判定の一致性」レポート



カッパ値を見ると、3者の判定はかなり一致していることが分かります。カッパ値は、判定が独立のときに0となり、判定が完全に一致しているときに1.0となります。「詳細」アウトラインには、判定者の各ペアに対して2x2の分割表が表示されます。Bowkerの対称性検定は、「度数表が対角線に関して対称である」(つまり、「すべてのi, jに対して  $p_{ij} = p_{ji}$  である」という帰無仮説に対する検定を行います。ここでBowker検定のp値はすべて0.05よりも大きくなっています。対称性の帰無仮説は棄却されません。

## 反復測定の例

この例では「Presidential Elections.jmp」を使用します。このデータには、1980年～2012年の米国の大統領選挙の各州における結果が含まれています。このデータを反復測定データとみなして、分析してみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Presidential Elections.jmp」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」までを選択します。
4. [関連] タブで、[反復測定] をクリックします。
5. [OK] をクリックします。
6. レポートウィンドウの下のほうにあるグレーの「遷移レポート」の開閉アイコンをクリックして、遷移レポートを開きます。

図3.17 反復測定の「遷移レポート」



1980年～1984年において、「民主党」から「共和党」に遷移した州が5つありました。それらの5州は、1980年には「民主党」が選ばれていましたが、1984年には「共和党」が選ばれています。2008年～2012年において、「民主党」から「共和党」に遷移した州が2つありました。他のすべての州は、2008年と2012年の両方の選挙で同じ党が選ばれています。

## [多重応答] タブの例

次に示すいくつかの例では、同じ情報を異なる形式で保持している、いくつかのサンプルデータを分析します。それらのサンプルデータは、異なる2つの条件において、3日間にわたり、生産ラインを検査したものです。各条件において、50ユニットが検査されました。7種類の不適合について、検査されています。1つのユニットにおいて、不適合が1つもない場合もあれば、複数の不適合が見つかる場合もあります。また、1つのユニットで、同じ種類の不適合が複数、生じる場合もあります。

### 多重応答

「Failure3MultipleField.jmp」のデータには、1ユニットが1行に保存されており、また、不適合データを含む列が複数あります。この例では、それらの列は3列になっています。これは、どのユニットも、多くて3つまでの不適合しか生じていないことを意味しています。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Quality Control」フォルダ内の「Failure3MultipleField.jmp」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」、および「不適合3」を選択します。
4. [多重] タブの [多重応答] をクリックします。
5. 「洗浄」と「日付」を選択し、[X, グループ化カテゴリ] をクリックします。
6. [OK] をクリックします。

図3.18 「多重応答」レポート

△ 多重応答(不適合1, 不適合2, 不適合3) By clean\*date

度数 シェア ケース			不適合								
			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	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%		23
			24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%		
	後	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%		23
	前	10月3日	20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%		
			8	3	0	5	0	1	0	17	50
			47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%		17
			16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%		
洗浄	後	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
			28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%		
	前	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
	前	10月3日	30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%		
			22	2	3	4	0	3	2	36	50
			61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%		36
			44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%		

シェアチャート			不適合							
			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	10月1日								23
		10月2日								23
		10月3日								17
	前	10月1日								31
		10月2日								30
		10月3日								36

クロス表は、バッチごとに1つの行、不適合の種類ごとに1つの列で構成されています。各セルには、度数、シェア、1ユニットあたりの比率が表示されます。たとえば、「10月1日」の「洗浄後」におけるバッチでは、「汚れ」における問題が12個ありました。そのシェアは、このバッチ全体の12/23、つまり52.2%になっています。また、50個のユニットにおいて、「汚れ」の不適合は12だけ生じているので、1ユニットあたりの比率は24% (=12/50) となります。

### 多重応答 ID 別

「Failure3ID.jmp」のデータは、バッチごとに、かつ、不適合の種類ごとに1つの行となっており、各不適合の種類を示す列と、バッチのIDを示す列で構成されています。

図3.19 「Failure3ID」データテーブル（一部）

	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	OCT 1 before
2	腐食	2	前	10月1日	50	OCT 1 before
3	ドーピング	1	前	10月1日	50	OCT 1 before
4	金属被覆	2	前	10月1日	50	OCT 1 before
5	その他	3	前	10月1日	50	OCT 1 before
6	酸化物不良	8	前	10月1日	50	OCT 1 before
7	シリコン不良	1	前	10月1日	50	OCT 1 before
8	ドーピング	0	後	10月1日	50	OCT 1 after
9	腐食	2	後	10月1日	50	OCT 1 after
10	金属被覆	4	後	10月1日	50	OCT 1 after

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Quality Control」フォルダ内の「Failure3ID.jmp」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、[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、[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。

結果のレポートは、図3.18で表示されるレポートと同じです。

## 多重応答 区切り文字

「Failures3Delimited.jmp」のデータは、ユニットごとに1つの行があり、不適合の列には、不適合の種類がカンマで区切られて記録されています。図3.20のデータテーブルでわかるように、ユニットの一部は「不適合」列が空になっています。これは、不適合がまったく観測されなかったことを示します。

図3.20 「Failure3Delimited.jmp」データテーブル（一部）

	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

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Quality Control」フォルダ内の「Failures3Delimited.jmp」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、[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、図3.18と同じレポートが表示されます。

**メモ:** 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。

## 指示変数

「Failures3Indicators.jmp」のデータは、ユニットごとに1つの行と、各不適合の種類に対する指示変数の列で構成されています。各不適合列のデータは、該当のユニットに不適合が観測されなかった場合には「0」、不適合が観察された場合には「1」が入力されます。

図3.21 「Failures3Indicators.jmp」データテーブル（一部）

	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

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Quality Control」フォルダ内の「Failures3Indicators.jmp」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. 「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、[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、図3.18と同じレポートが表示されます。

## 応答の度数

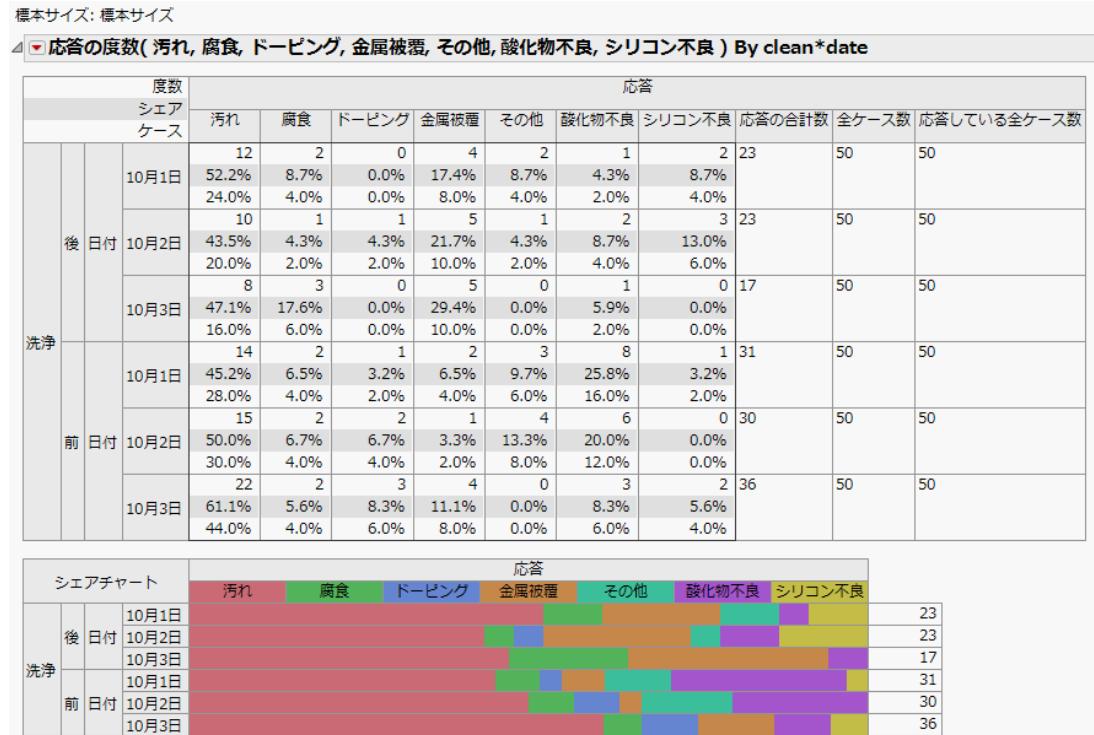
「Failure3Freq.jmp」のデータは、バッチごとに1つの行、各不適合の種類ごとに複数の列、バッチサイズについての1つの列で構成されています。各不適合列には、バッチにおける不適合の発生度数が入力されています。

図3.22 「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. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Quality Control」フォルダ内の「Failure3Freq.jmp」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」）を選択します。
4. [多重] タブの [応答の度数] をクリックします。
5. 「洗浄」と「日付」を選択し、[X, グループ化カテゴリ] をクリックします。
6. 「標本サイズ」を選択し、[標本サイズ] をクリックします。
7. [OK] をクリックします。

図3.23 不適合数の表



クロス表の「応答している全ケース数」列を除き、図3.18と同じ結果です。この例では、分析対象の不適合データがすでに要約されていました。この要約されたデータには、不適合が0であるユニット数の記録はありません。そのため、この例の「応答している全ケース数」には、バッチサイズである50が表示されています。

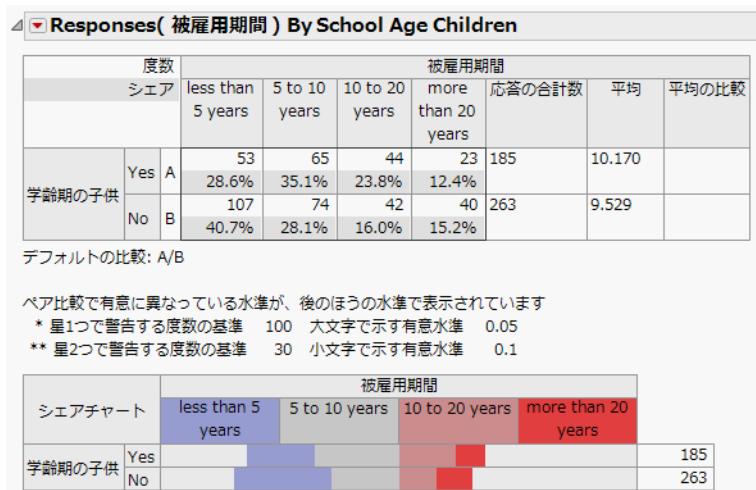
## 比較のアルファベット文字を持つスコア平均の例

この例では「Consumer Preferences.jmp」を使用します。このデータで、被雇用期間と、学齢期の子供がいることの関係を調べます。「被雇用期間」列には、便宜的な通し番号として、1、2、3、および4の数値が含まれています。そして、これらの値に対して、「値ラベル」列プロパティによって、ラベルが割り当てられています。「カテゴリカル」プラットフォームの「スコア平均」オプションで、被雇用期間の平均を求めるには、この列の値(1,2,3,4)に、被雇用期間を「値スコア」として割り当てる必要があります。列プロパティについての詳細は、『JMPの使用法』の「列情報ウィンドウ」章を参照してください。

1. 「ヘルプ」>「サンプルデータライブラリ」を選択し、「Consumer Preferences.jmp」を開きます。
2. データテーブルで、「被雇用期間」列見出しを右クリックし、[列プロパティ] > [値スコア] を選択します。
3. 「値」に1、「スコア」に3を入力してから、[追加] をクリックします。
4. 「値」に2、「スコア」に7.5を入力してから、[追加] をクリックします。
5. 「値」に3、「スコア」に15を入力してから、[追加] をクリックします。

6. 「値」に4、「スコア」に25を入力してから、[追加] をクリックします。
7. [OK] をクリックします。
8. [分析] > [消費者調査] > [カテゴリカル] を選択します。
9. 「被雇用期間」を選択して、[単純] タブの [応答] をクリックします。
10. 「学齢期の子供」を選択して [X, グループ化カテゴリ] をクリックします。
11. [OK] をクリックします。
12. 「カテゴリカル」の赤い三角ボタンをクリックして、[スコアの平均] を選択します。
13. 「カテゴリカル」の赤い三角ボタンをクリックして、[スコア平均の比較] を選択します。

図3.24 スコアの平均が表示された「カテゴリカル」レポート



学齢期の子供を持つ人の被雇用期間の平均は10.17、学齢期の子供がない人の平均は9.53です。クロス表における[平均の比較]列が空なので、統計的に有意差があるとは言えません。有意差があるグループには、アルファベット文字が表示されます。

**ヒント:** スコア平均オプションを使用している場合は、データの記録方法に注意してください。データが値レベルでコード化された数値で保存されている場合、平均値は、その数値データに基づいて計算されます。その数値データに意味がない場合、「値スコア」を使って、各水準に意味のあるスコア値を割り当ててください。

## 「表の構成」レポートの例

この例では「Consumer Preferences.jmp」を使用します。このデータで、仕事に対する満足度を、「表の構成」タブを使って、性別×年齢層、および現職での勤続年数で比較して調べます。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Consumer Preferences.jmp」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3.25 「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。
11. 「カテゴリカル」の赤い三角ボタンをクリックして、[応答の等質性に対する検定]を選択します。

図3.26 [表の構成] タブのレポートの例

仕事に対する満足度+給与グループ By 性別*年齢層+現職位での勤続年数グループ																						
度数 シェア	性別																現職位での勤続年数グループ					
	M 年齢層								F 年齢層													
	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	less than 5 years	5 to 10 years	10 to 20 years	more than 20 years				
仕事に対する満足度	Not at all satisfied	3	3	1	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%	6.6%	0.0%			
	Somewhat satisfied	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%			
	Extremely satisfied	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			
	less than 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 to 60000	16	14	11	12	14	7	13	15	10	8	6	7	6	12	67	38	37	9			
給与グループ	60000 to 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%			
	greater than 120000	2	1	2	3	3	1	9	2	3	1	1	1	1	1	10	14	7	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			
応答の等質性に対する検定																						
		Pearson																				
		応答の列	標本の列	尤度比 カイ2乗 尤度比 p値 カイ2乗 Pearson p値																		
		仕事に対する満足度	性別 + M, 年齢層	4.65556	0.95685	4.59049	0.9703															
		仕事に対する満足度	性別 + F, 年齢層	23.8406	0.0214*	24.9917	0.0149*															
		給与グループ	性別 + M, 年齢層	22.0707	0.5750	23.8074	0.4727															
		給与グループ	性別 + F, 年齢層	27.6209	0.2764	26.3992	0.3332															
		仕事に対する満足度	現職位での勤続年数グループ	8.00461	0.2378	6.76025	0.3436															
		給与グループ	現職位での勤続年数グループ	25.8336	0.0113*	24.0752	0.0199*															

「表の構成」タブのレポートには、「表の構成」タブで指定した表が含まれます。応答の等質性に対する検定は、グループ変数のそれぞれの組み合わせに対して行われます。男性の場合、年齢層間に仕事に対する満足度における違いがないことがわかります (Pearsonのp値 = 0.9703)。女性の場合、年齢層間で仕事に対する満足度に違いがあることがわかります (Pearsonのp値 = 0.0149)。中年女性は、仕事に対する満足度が最も低い傾向があります。これらの結果を視覚的に見るために、シェアチャートや度数チャートをレポートに追加してもよいでしょう。

## 「カテゴリカル」プラットフォームの統計的詳細

### Rao-Scott 修正

多重応答データ（複数回答データ）における「応答の等質性に対する検定」では、Rao-Scott修正が適用されます。Rao-Scott修正については、Lavassani et al. (2009) を参照してください。なお、多重応答データにおいては、「応答の等質性に対する検定」オプションは、「表の構成」タブでのみ使うことができます。

多重応答データでは、データの重複が許されています。つまり、1人の回答者から複数の応答を得ることが許されています。Pearsonのカイ2乗検定は、多重応答データに適していません。なぜなら、多重応答データは、Pearsonのカイ2乗検定の前提となっている、データの独立性を満たさないためです。また、周辺和は多重応答が許されない場合よりも大きくなるため、周辺和から計算される期待値も影響を受けます。

Rao-Scottのカイ2乗統計量は、次のように定義されます。

$$\chi_C^2 = \frac{\chi^2}{\delta}$$

この式で、

$\chi^2$  は、通常の Pearson カイ2乗検定統計量です。 $\delta$  は、次式で定義される補正係数です。

$$\delta = 1 - \frac{m_{++}}{n_+ C}$$

この式で、

$m_{++}$  は、多重応答の総度数です。

$n_+$  は、回答者（ケース）の総数です。

$C$  は、応答変数の水準数（クロス表における列数）です。

自由度は、 $(R-1) C$ （つまり、行数から1を引いたものに、列数を掛けたもの）です。



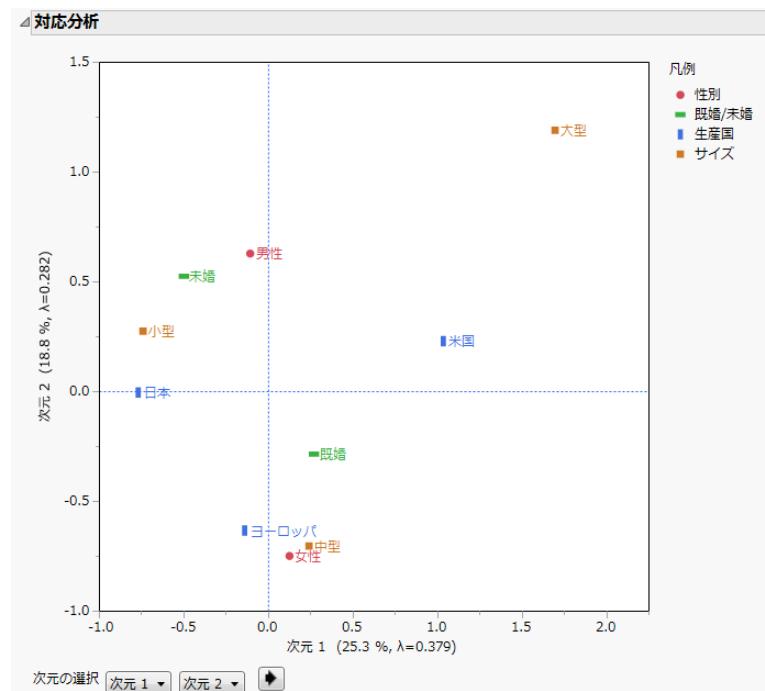
## 多重対応分析

### カテゴリカルな変数の水準間にある関係の特定

多重対応分析 (Multiple Correspondence Analysis; MCA) は、複数のカテゴリカルな変数を分析対象とし、それらのカテゴリカルな変数の水準間にある関係を調べます。対応分析が2変数を分析するのに対し、多重対応分析は3変数以上を扱えます。多重対応分析は、「カテゴリカルな変数に対する主成分分析」と喻えることができます。他の多変量分析と同様に、次元を減らす手法であり、カテゴリカルな多変量データを2次元や3次元の空間で表します。

多重対応分析は、特にフランスや日本の社会科学分野で広く利用されています。多重対応分析は、質問紙調査では、複数の質問における関係を調べるのに役立つでしょう。消費者調査では、潜在的な市場を探し出すのに使えるでしょう。また、遺伝子のマイクロアレイ解析では、遺伝子間の関係を調べるのに役立つでしょう。

図4.1 多重対応分析



## 多重対応分析の例

ここでは、自家用車に関する調査データである「Car Poll.jmp」データテーブルを使用します。このデータテーブルには、回答者の属性である性別や既婚／未婚、年齢が含まれています。また、回答者が所有する自動車の属性として、生産国、サイズ、タイプも含まれています。性別、既婚／未婚、車の生産国、車のサイズの間にある関係を調べ、消費者の嗜好を特定することを目的に分析を行います。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [消費者調査] > [多重対応分析] を選択します。
3. 「性別」、「既婚/未婚」、「生産国」、「サイズ」を選択し、[Y, 目的変数] をクリックします。
4. [OK] をクリックします。

図4.2 入力が完了した「多重対応分析」起動ウィンドウ

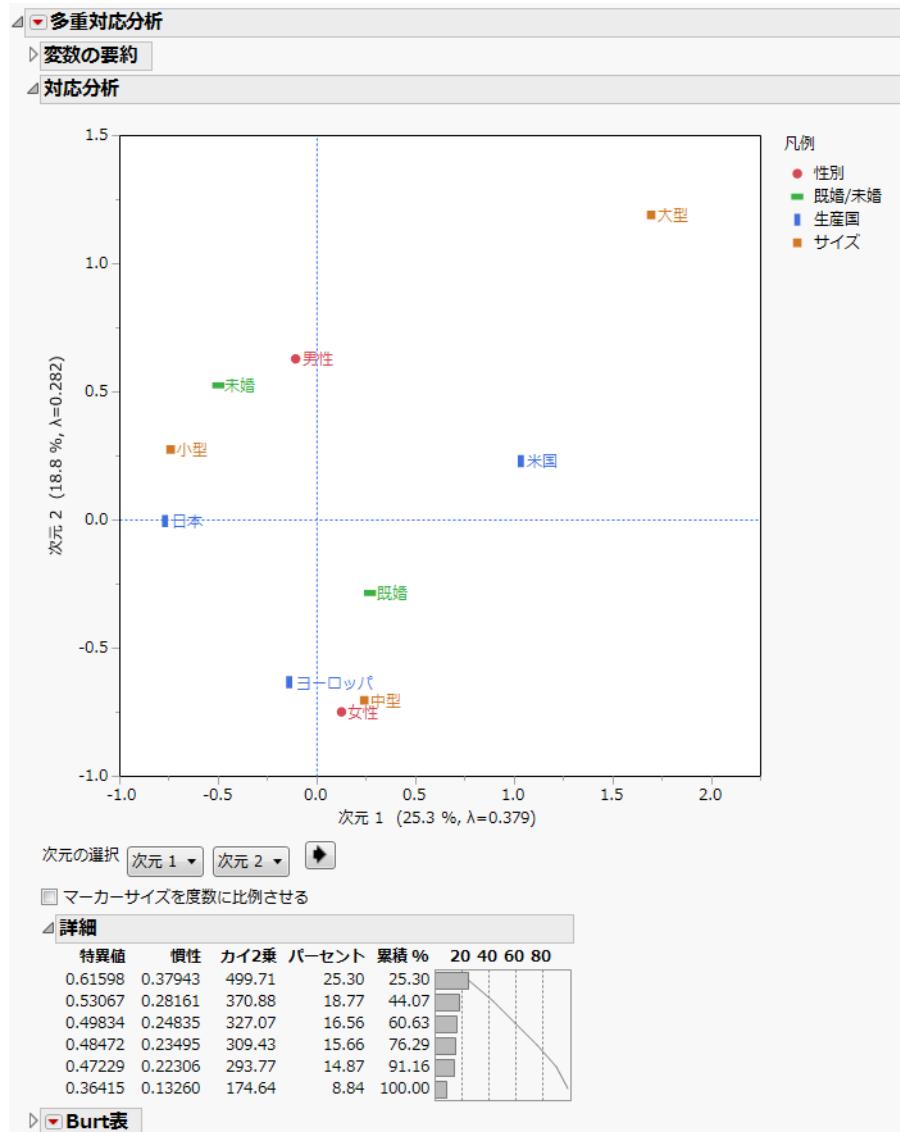


「多重対応分析」レポートは図4.3のようになります。スペースの関係上、一部のアウトラインが閉じた状態で表示されます。

「変数の要約」レポートは、分析に用いた変数を簡潔にまとめたものです。

「対応分析」レポートでは、4変数のカテゴリが2つの主軸の上に射影されています。このグラフを見ると、米国と大型車、日本と小型車の間に関係があるのが分かります。また、男性と小型車、女性と中型車にも関係が見られます。マーケティングにおいて、これらの情報を用い、広告のターゲット層を絞り込んでも良いでしょう。

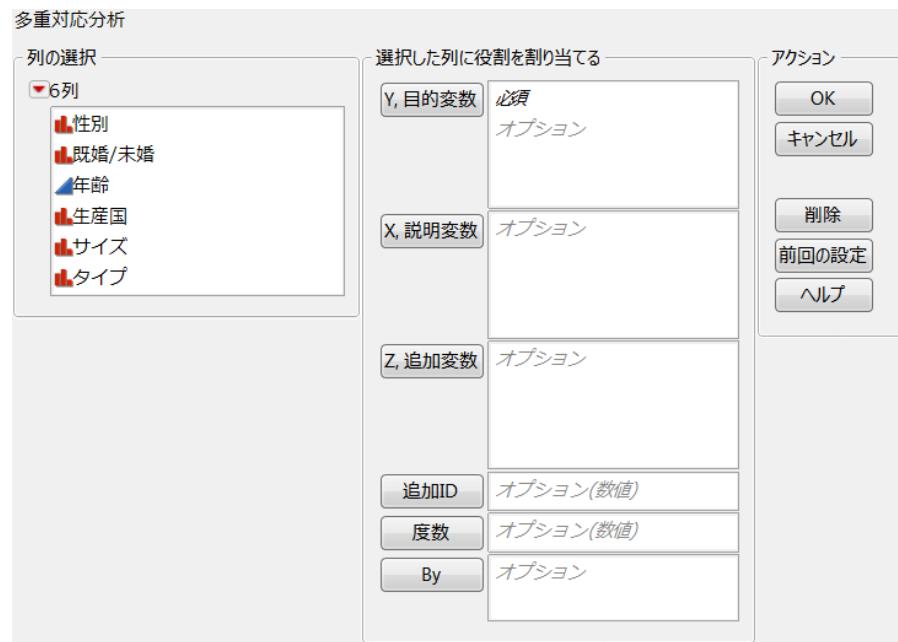
図4.3 「多重対応分析」レポート



## 「多重対応分析」プラットフォームの起動

「多重対応分析」プラットフォームを起動するには、[分析] > [消費者調査] > [多重対応分析] を選択します。

図4.4 「多重対応分析」起動ウィンドウ



**Y, 目的変数** 分析するカテゴリカル列を指定します。多重対応分析では、変数間の連関を調べるのが目的であり、説明変数と応答変数という明確な分類を使わないのが普通です。

**X, 説明変数** 因子、つまり説明変数として使用するカテゴリカル列を指定します。

**Z, 追加変数** 追加変数として使用する列を指定します。これらの変数は、関係を調べる上で関心がある変数ですが、空間を構成するときの計算には含めません。

**追加ID** 表の追加列として使用する行を識別するための変数です。この「追加ID」のデータには、通常、1と0の値を使用してください。[追加ID] が0である行が追加の行として、そして、0以外である行がアクティブな行として扱われます。XまたはYの水準に、アクティブな行に存在していないのに、追加の行には存在しているものがある場合は、[追加ID] 列は指定されていても無視されます。

**メモ:** [追加ID] と [Z, 追加変数] の役割は、いずれか一方しか指定できません。

**度数** 度数変数を指定します。このオプションは、データが要約データである場合に用いてください。

**By** By変数の水準ごとに個別のレポートが作成されます。複数のBy変数を指定した場合は、By変数の水準のすべての組み合わせごとにレポートが作成されます。

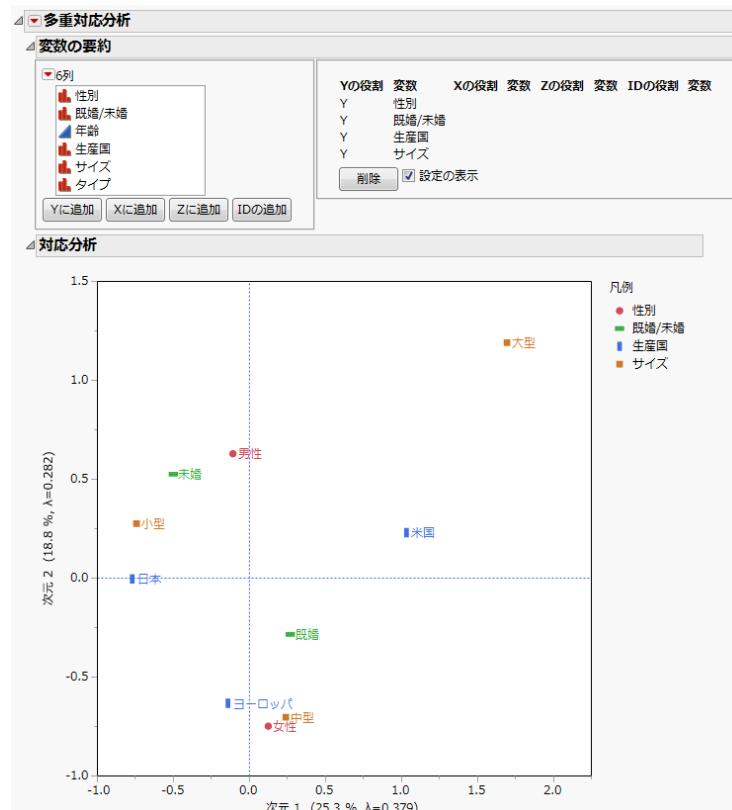
**メモ:** 「多重対応分析」プラットフォームでは、欠測値の扱いが他の多くのプラットフォームと異なります。あるデータ行において、いくつかの列が欠測値であっても、欠測値でないペアの度数はカウントされます。行全体が計算から除外されることはありません。

## 「多重対応分析」レポート

最初に作成される「多重対応分析」レポートには、変数の要約、対応分析プロット、重要度の大きい順で並べられた次元の詳細が表示されます。カテゴリがプロットされたグラフから、データに存在する関係を見ることができます。「詳細」は、データの関係を見るのに2次元だけで十分かどうかを判断する参考になります。

「変数の要約」には、分析に使った列と起動ウィンドウで指定した役割が表示されます。「設定の表示」チェックボックスをオンにすると、左側にデータテーブル内の列のリストが表示されます。列を選択して [Yに追加]、[Xに追加]、[Zに追加]、または [追加ID] をクリックすると、分析に使う列を変更できます。また、「変数の要約」表に列をドラッグしても、変更できます。これらの方法を使えば、起動ウィンドウに戻らなくても、分析に使う列を変更できます。

図4.5 「設定の表示」を選択した状態の「多重対応分析」レポート



## 「多重対応分析」プラットフォームのオプション

「多重対応分析」の赤い三角ボタンのメニューには、目的に合わせてレポートをカスタマイズするためのオプションがあります。出力できる結果は、分析の種類によって異なります。

**対応分析** 対応分析レポートを作成します。レポートには、プロット、詳細、座標、要約統計量が表示されます。詳細については、「[「多重対応分析」プラットフォームの別例](#)」(79ページ) を参照してください。

**クロス表** 変数の役割に応じて、「Burt表」または「分割表」を作成します。詳細については、「[クロス表](#)」(78ページ) を参照してください。

**追加行のクロス表** 追加変数と応答変数との分割表を作成します。この表は、起動ウィンドウで追加変数 (Z変数) を指定した場合にのみ、デフォルトで表示されます。

**追加列のクロス表** 追加変数と応答変数との分割表を作成します。この表は、起動ウィンドウで説明変数 (X変数) と追加変数 (Z変数) を指定した場合にのみ、デフォルトで表示されます。

**モザイク図** 名義尺度または順序尺度の応答変数に対し、モザイク図を作成します。モザイク図は、棒グラフを積み重ねたもので、各セグメントの大きさがそのグループの度数に比例します。このオプションを使用できるのは、Y変数とX変数を1つずつ指定した場合のみです。

**独立性の検定** 独立性の検定を行い、行変数と列変数の間に連関があるかどうかを調べます。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。このオプションを使用できるのは、Y変数とX変数を1つずつ指定した場合のみです。

以下のオプションについて詳しくは、『JMPの使用法』の「JMPのレポート」章を参照してください。

**ローカルデータフィルタ** 現在のレポートにおいてデータをフィルタリングするローカルデータフィルタを表示するか、非表示にします。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、[自動再計算] オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウでBy変数を指定した場合のみ使用可能です。

## 「対応分析」のオプション

「対応分析」メニューの下に用意されているレポートは、分析の種類によって異なります。これらのレポートのいくつかはデフォルトで表示されます。

**プロットの表示** 2次元上に（最初の2つの主軸で構成される平面上に）、カテゴリをプロットします。このグラフはデフォルトで表示されます。

**詳細の表示** 特異値、慣性、カイ2乗統計量、パーセント、累積パーセントを表示します。このレポートはデフォルトで表示されます。詳細については、「[詳細の表示](#)」（76ページ）を参照してください。

**調整済み慣性の表示** Benzécri と Greenacre の調整済み慣性を表示します。Benzécri (1979) と Greenacre (1984) を参照してください。X変数がある場合、このオプションは使用できません。詳細については、「[調整済み慣性の表示](#)」（76ページ）を参照してください。

**座標の表示** 各カテゴリ（各点）に対して、3次元までの主座標を表示します。詳細については、「[座標の表示](#)」（76ページ）を参照してください。

**要約統計量の表示** 各カテゴリ（各点）に対して、余弦2乗和（品質；quality）、質量、慣性を表示します。詳細については、「[要約統計量の表示](#)」（77ページ）を参照してください。

**慣性への偏寄与率の表示** 各カテゴリ（各点）に対して、各カテゴリが慣性にどれほど寄与しているかを表示します。詳細については、「[慣性への偏寄与率の表示](#)」（77ページ）を参照してください。

**余弦2乗の表示** 各カテゴリ（各点）に対して、3次元までの余弦2乗を表示します。レポートに含まれる棒グラフは、各Y変数の各水準に対し、最初の3つまでの次元の余弦の2乗を表示したものです。詳細については、「[余弦2乗の表示](#)」（78ページ）を参照してください。

**CochranのQ検定** （すべてのY変数が同じ2水準を持ち、X変数の値が各行で一意である場合のみ使用可能。）特定の応答の周辺確率がY変数すべてで変わらないことを検定するCochranのQ検定を表示します。CochranのQ検定は、2つのY変数に対する McNemar 検定を、3つ以上のY変数に一般化した検定です。Agresti (2002) を参照してください。

**三次元対応分析** 3次元上に（最初の3つの主軸で構成される空間上に）、カテゴリをプロットします。次元が2次元以下の場合は、このオプションは使用できません。

**座標の保存** 主座標をJMPデータテーブルに保存します。列の座標、行の座標、追加列の座標、追加行の座標がそれぞれ別のデータテーブルに保存されます。保存する次元数を選択することができます。

**座標計算式の保存** データテーブルの各行ごとの主座標を求める計算式を、元のデータテーブルに保存します。データテーブルの各行ごとの主座標は、Y変数のカテゴリースコアを平均したものを特異値で割った値です。保存する次元数を選択することができます。

## プロットの表示

プロットでは、最初の2つの主軸で構成される平面に、カテゴリを示す点が射影されています。このプロットでは、X軸とY軸の両軸が、1単位あたりの長さ（スケール）が同じになるように調整されています。プロットの下にある「次元の選択」コントロールを使ってプロットする次元を切り替えることができます。1つ目のコントロールがプロットの横軸、2つ目のコントロールが縦軸に該当します。プロットする次元を切り替えるには、矢印ボタンをクリックします。プロットに表示する次元を切り替えるには、矢印ボタンをクリックします。「次元の選択」コントロールの下にあるチェックボックスで、プロットにおける点の大きさを、カテゴリの度数に比例させるかどうかを指定できます。

---

**メモ:** 対応分析プロットで点を選択すると、レポートウィンドウ内の他の表でも対応する行が選択されます。ただし、データテーブル内の行は選択されません。プロット上で、特定の変数と関連がある点すべてを選択するには、プロットの凡例で変数の名前を選択します。

---

## 詳細の表示

特異値の表を表示します。

**特異値** 分割表またはBurt表の特異値分解の特異値を表示します。計算式については、「[「詳細」レポート](#)」(83ページ)を参照してください。

**慣性** 特異値の2乗。各次元で、データのどの程度の変動が説明されるかを示します。

**カイ2乗** Burt表や分割表のカイ2乗のうち、その次元によって説明される部分。

**パーセント** 全次元の慣性に対する、各次元の慣性が占める割合。

**累積パーセント** 慎性の累積割合。最初の2次元によってデータの大部分を説明できるようであれば、2次元だけのプロットでデータにおける関係を十分に見ることができます。

## 調整済み慣性の表示

多重対応分析のBurt表から求められる慣性は、固有値です。多重対応分析の慣性は、データが適合していても、過度に小さくなりすぎるという欠点があります。Benzécriは、このような欠点を解消するために、慣性を調整する方法を提案しました。その後、Greenacreは、Benzécriの調整では適合度が逆に過大評価されることを指摘し、別の調整法を提案しました。JMPでは、両方の調整を計算できます。詳細については、「[調整済み慣性](#)」(83ページ)を参照してください。

**慣性** 特異値の2乗。各次元で、データのどの程度の変動が説明されるかを示します。

**調整済み慣性** BenzécriやGreenacreが提案した方法で調整を加えた慣性。

**パーセント** 全次元の慣性に対する、各次元の調整済み慣性が占める割合。

**累積パーセント** 調整済み慣性の累積割合。最初の2次元によってデータの大部分を説明できるようであれば、2次元だけのプロットでデータにおける関係を十分に見ることができます。

## 座標の表示

「列座標」または「行座標と列座標」の表を表示します。

**X** [X, 説明変数] に指定した列。

**Y** [Y, 目的変数] に指定した列。

**Z** [Z, 追加変数] に指定した列。

カテゴリ X、Y、Zの各変数の水準。

**次元1、次元2、次元3** 各水準または各応答の、各次元の座標。デフォルトでは、最初3つの次元までの座標が表示され、その他の次元の座標列は非表示になっています。その他の次元を表示するには、表を右クリックし、[列] サブメニューから次元を選択してください。

**メモ:** [X, 説明変数] に列を指定した場合は、「座標」レポートに、Yの座標だけではなく、Xの座標も表示されます。また、[Z, 追加変数] に列を指定した場合は、XやYの座標の下に、Zの座標も表示されます。

## 要約統計量の表示

「列点の要約統計量」または「各点の要約統計量」の表が表示されます。「Y」の表には、Y変数の水準（列点）ごとに、余弦2乗和、質量、慣性がまとめられています。「X」の表には、X変数の水準（行点）ごとに余弦2乗和、質量、慣性がまとめられています。詳細については、「要約統計量」（84ページ）を参照してください。

X [X, 説明変数] に指定した列。

Y [Y, 目的変数] に指定した列。

カテゴリ X変数またはY変数の水準。

**余弦2乗和(2次元)** 2次元までによって各点がどの程度よく説明されているかを表す指標。

**質量** 該当カテゴリの度数を全度数で割った値。Burt表では、各行の合計のパーセントに相当します。

**慣性** 全体の慣性が各カテゴリによってどれだけ説明されるかを示す割合。この値をすべて合計すると1になります。

**メモ:** [X, 説明変数] を指定した場合は、「要約統計量」レポートに、Yだけではなく、Xに対する要約統計量も表示されます。

## 慣性への偏寄与率の表示

「慣性への各点の偏寄与率」表を表示します。「慣性への各点の偏寄与率プロット」も表示します。このプロットは、各カテゴリが各次元の慣性にどれぐらい寄与しているかを棒グラフにしたものです。

X [X, 説明変数] に指定した列。

Y [Y, 目的変数] に指定した列。

カテゴリ X変数またはY変数の水準。

**次元1、次元2、次元3** 各次元の慣性に対して、各カテゴリがどれぐらい寄与しているかを示す指標。デフォルトでは、最初の3つの次元までの列が表示され、他の列は非表示になっています。他の列を表示するには、表を右クリックし、[列] サブメニューから次元の列を選択してください。

この偏寄与率は、次元ごとの慣性に各カテゴリがどれくらい寄与しているかを示しています。この偏寄与率は、各次元において合計すると1になります。

**メモ:** **[X, 説明変数]** を指定した場合は、「慣性への各点の偏寄与率」レポートに、Yだけではなく、Xに対する偏寄与率も表示されます。詳細については、「[慣性への偏寄与率](#)」(84ページ) を参照してください。

## 余弦2乗の表示

「各点の余弦2乗」の表を表示します。「各点の余弦2乗プロット」も表示します。このプロットは、最初の3次元までに関して、各カテゴリの余弦2乗を棒グラフにしたものです。

**X** **[X, 説明変数]** に指定した列。

**Y** **[Y, 目的変数]** に指定した列。

**カテゴリ** **X**変数またはY変数の水準。

**次元1、次元2、次元3** 各次元によって、各カテゴリがどれほど良く表現されているのかの品質を示す指標。デフォルトでは、最初3つの次元までの結果が表示され、その他の列は非表示になっています。その他の列を表示するには、表を右クリックし、**[列]** サブメニューから次元の列を選択してください。

この指標は、各次元によって、各点がどれくらい表現されているかの品質を表します。余弦2乗は、点と次元との相関の2乗として解釈できます。最初の2つの次元の余弦2乗の和は、「要約統計量」レポートの「余弦2乗和(2次元)」と等しくなります。詳細については、「[要約統計量](#)」(84ページ) を参照してください。

**メモ:** **[X, 説明変数]** を指定した場合は、「余弦2乗」レポートに、Yだけではなく、Xに対する余弦2乗も表示されます。

## クロス表

多重対応分析は、Burt表に基づいて行われます。Burt表は、カテゴリカル変数のすべてのペアに対する分割表で構成されています。Burt表の対角線上に位置する分割表（変数とその変数自身との分割表）は、対角行列です。非対角線上に位置する分割表は、普通の分割表です。複数の **[Y, 目的変数]** 列を選択し、**[X, 説明変数]** 列を選択しなかった場合には、Burt表が作成されます。**[X, 説明変数]** 列を選択した場合は、Burt表ではなく、従来型の分割表が作成されます。

Burt表または分割表の赤い三角ボタンのメニューには、表に統計量を表示するためのオプションがあります。

**度数** セル度数、周辺度数、全体合計度数（標本サイズ）。これはデフォルトで表示されます。

**合計%** セル度数と周辺度数の全体合計度数に占める割合。これはデフォルトで表示されます。

**セルのカイ2乗** 各セルに対して、 $(O - E)^2 / E$  の式で計算したカイ2乗の値。

**列%** 各セル度数の列合計に占める割合。

**行%** 各セル度数の行合計に占める割合。

**期待** 独立性を仮定したときの各セルの期待度数 ( $E$ )。対応する行合計と列合計との積を、全体合計で割ったものです。

**偏差** 観測セル度数 ( $O$ ) から期待セル度数 ( $E$ ) を引いた値。

**累積列度数** 列合計の累積。

**累積列 %** 列パーセントの累積。

**累積行度数** 行合計の累積。

**累積行 %** 行パーセントの累積。

**データテーブルに出力** 表に含まれている各統計量につき、それぞれ1つのデータテーブルを作成します。

## 追加行のクロス表

[ $Z$ , 追加変数] に列を指定した場合、追加変数 ( $Z$ 変数) の水準を行とし、目的変数 (Y変数) の水準を列とする分割表が作成されます。赤い三角ボタンのメニューは、Burt表のものと同じです。

## 追加列のクロス表

[ $X$ , 説明変数] と [ $Z$ , 追加変数] に列を指定した場合、説明変数 ( $X$ 変数) の水準を行、追加変数 ( $Z$ 変数) の水準を列とする分割表が作成されます。赤い三角ボタンのメニューは、Burt表のものと同じです。

---

## 「多重対応分析」プラットフォームの別例

### 追加変数を使用した例

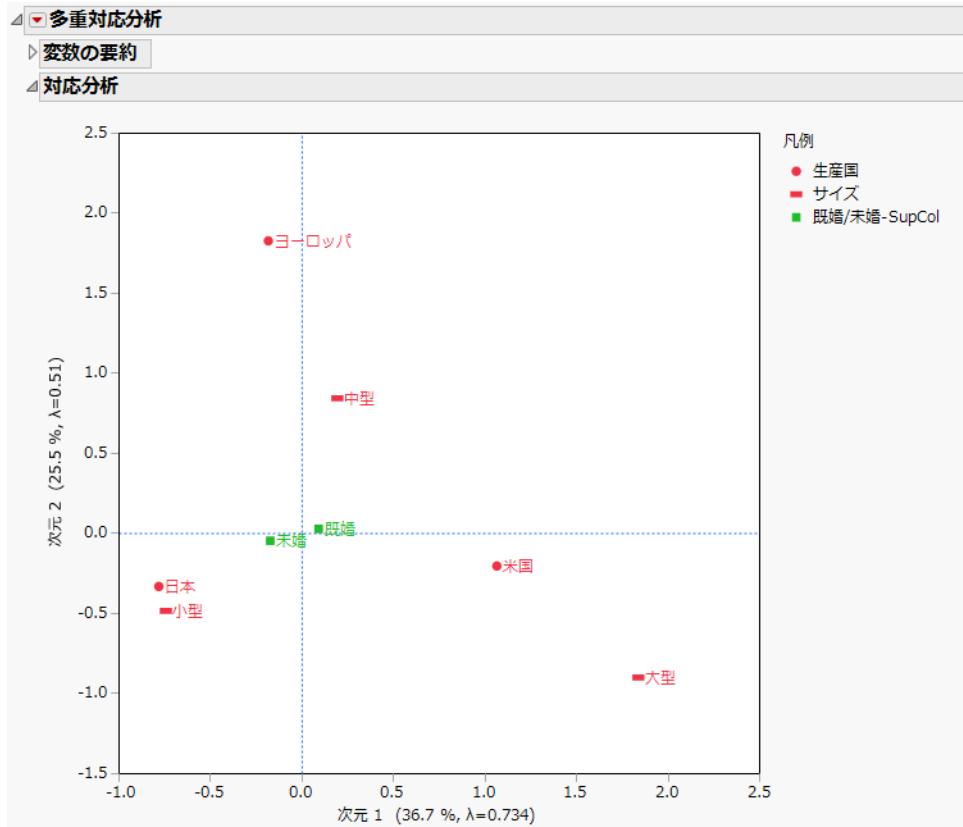
ここでは、自家用車に関する調査データである「Car Poll.jmp」データテーブルを使用します。このデータテーブルには、回答者の属性である性別や既婚／未婚、年齢が含まれています。また、回答者が所有する自動車の属性として、生産国、サイズ、タイプも含まれています。性別、車の生産国、車のサイズの間にある関係を調べ、消費者の嗜好を特定することを目的に分析を行います。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [消費者調査] > [多重対応分析] を選択します。
3. 「生産国」と「サイズ」を選択し、[Y, 目的変数] をクリックします。
4. 「既婚/未婚」を選択し、[Z, 追加変数] をクリックします。
5. [OK] をクリックします。

最初の例とは異なり、「既婚/未婚」は計算に使われません。「既婚/未婚」は、計算が完了してからプロットに追加されます。

プロットを見ると、日本と小型車、アメリカと大型車の間に強い関係があることがわかります。「既婚/未婚」の2つの値は、異なる色でプロットされています。独身の人たちは、既婚の人たちより小型車を好む傾向が多少強いようです。

図4.6 補助変数を使った多重対応分析のレポート



## 追加IDを使用した例

米国勢調査局のデータを使うと、過去1世紀にわたる人口の増加を調べることができます。「US Regional Population.jmp」データテーブルには、1920年から2010年までに調査された米国50州の人口が地域ごとにまとめられています。アラスカ州とハワイ州は、最初から米国に属していたわけではなく、陸続きでもないので、ここでは追加の地域として扱ってみます。この2州における人口増加が、米国の他の州と異なるかどうかを調べることにしましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「US Regional Population.jmp」を開きます。
2. [分析] > [消費者調査] > [多重対応分析] を選択します。
3. 「年」を選択し、[Y, 目的変数] をクリックします。
4. 「地域(大分類)」を選択し、[X, 説明変数] をクリックします。

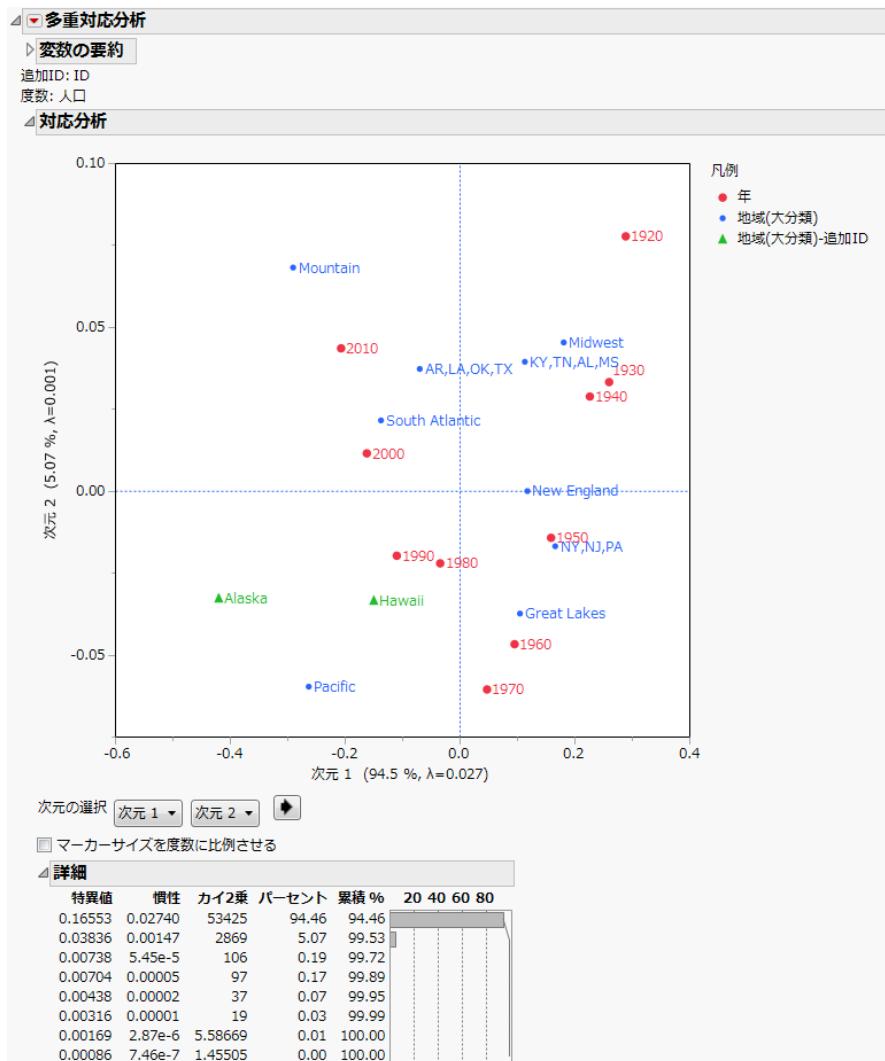
5. 「ID」を選択し、[追加ID] をクリックします。
6. 「人口」を選択し、[度数] をクリックします。
7. [OK] をクリックします。

「詳細」レポートから、年と地域との関係は、ほぼ完全に1次元で説明されることがわかります。プロットを見ると、第1次元において年が新しい順に並んでいます。対応分析では順序的な制約を結果に課していないので、年が新しい順に並んでいるのは分析結果から自然に生じたものです。

地域の順序からは、中西部から北東部、南部、さらに山岳部および西部へと人口が移動していることがわかります。

アラスカ州とハワイ州は、分析の計算には使われていませんが、結果に基づいてプロットに追加されています。これらの2州における増加パターンは、「Pacific」(太平洋) の地域とよく似ています。アラスカ州の人口は、「Pacific」地域よりさらに大きく増加しています。

図4.7 補助IDを使った多重対応分析のレポート



## 「多重対応分析」プラットフォームの統計的詳細

この節では、多重対応分析の統計的詳細について解説します。

### 「詳細」レポート

単純な対応分析においては、次式のような特異値分解が行われます。

$$\mathbf{D}_r^{-0.5}(\mathbf{P} - rc')\mathbf{D}_c^{-0.5} = \mathbf{UD}_u\mathbf{V}'$$

この式で、

- $\mathbf{P}$ は、度数を合計度数で割った値を要素とする行列
- $r$ および $c$ は、 $\mathbf{P}$ の行和と列和
- $\mathbf{D}_r$ と $\mathbf{D}_c$ は、 $r$ および $c$ を対角要素にもつ対角行列

多重対応分析においては、次式のような特異値分解が行われます。

$$\mathbf{D}^{-0.5} \left( \frac{\mathbf{C}}{Q^2 n} - \mathbf{D} \mathbf{1} \mathbf{1}' \mathbf{D} \right) \mathbf{D}^{-0.5} = \mathbf{UD}_u\mathbf{V}'$$

この式で、

$$\mathbf{D} = \left( \frac{1}{m} \right) \text{diag}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_Q)$$

- $\mathbf{C}$ はBurt表
- $Q$ は、カテゴリカルな変数の個数
- $n$ は、オブザベーション数
- $\mathbf{1}$ は、すべての要素が1である列ベクトル

### 調整済み慣性

$m$ 個のカテゴリカルな変数に対する多重対応分析における通常の慣性は、Burt表から計算される固有値（これを、 $u_k$ と記します）で、これは $\mathbf{D}_u^2$ の対角要素です。こうして計算された慣性は、データに適合していても、過度に小さくなってしまうという欠点があります。Benzécri (1979) は、この欠点を補うために、次のような慣性の調整を提案しました。この調整済み慣性は、Greenacre (1984, p. 145) でも説明されています。

$$\left( \frac{m}{m-1} \right)^2 \times \left( u_k - \frac{1}{m} \right)^2 \quad \text{ただし } u_k > \frac{1}{m}$$

この調整は、 $1/m$  より大きい慣性すべてについて、調整済み慣性の和に対する、各調整済み慣性の割合を計算します。

Greenacre (1994, p.156) は、Benzécri が提案した方法では、逆に、適合度を過大に評価してしまうと指摘しました。Greenacre は、次のような方法で計算した値を分母にして、調整済み慣性の割合を計算することを提案しました。

$$\frac{m}{m-1} \left( \text{trace}(\mathbf{D}_u^4) - \frac{n_c - m}{m^2} \right)$$

これを、 $1/m$  より大きいすべての慣性について計算します。この式で、 $\text{trace}(\mathbf{D}_u^4)$  は慣性の2乗の和、 $n_c$  は  $m$  個の変数すべてにおけるカテゴリの合計数です。

## 要約統計量

「要約統計量」における「余弦2乗和（2次元）」は、すべての次元上における原点からの距離に対する、2次元上における原点からの距離の比（つまり、すべての次元のベクトルと、2次元上のベクトルとのコサイン）を2乗したものです。ある次元における「余弦2乗」とは、すべての次元でのベクトルと、その次元でのベクトルとのコサインを2乗したものです。なお、「余弦2乗和（2次元）」は、その点の2次元までの慣性の和と、その点のすべての次元の慣性の和との比にもなっています。この「余弦2乗和（2次元）」は、2次元までの空間で、該当する点がどの程度、良く表現されているかを示します。

「質量」は、合計度数に対する、行または列の合計度数の比です。

「慣性」は、主成分分析の分散に相当します。データ全体の全慣性は、2元分割表全体における Pearson カイ2乗値を全標本サイズで割った値です。

「要約統計量」における相対的な「慣性」は、データ全体の全慣性への各点の寄与率を表しています。「要約統計量」の表では、相対的な慣性が「慣性」として表示されます。

## 慣性への偏寄与率

ある次元の慣性への行または列の寄与率は、次のように求められます。

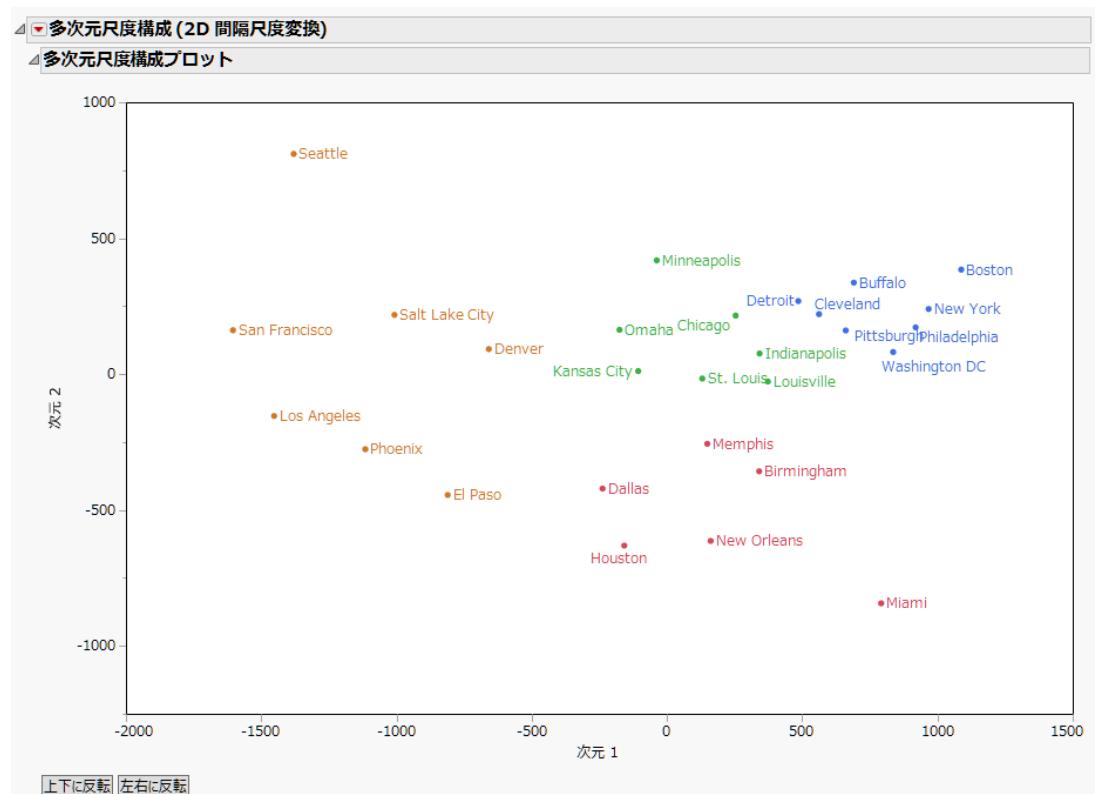
$$\text{偏寄与率} = \frac{\text{質量} \times \text{座標の2乗}}{\text{次元の慣性}}$$

## 多次元尺度構成 対象間における類似度の視覚化

多次元尺度構成 (Multi-Dimensional Scaling; MDS) は、対象間の類似度（類似性、相違性、距離）を視覚化する手法です。たとえば、都市間の地理的距離を示すデータがあった場合、多次元尺度構成を適用すると、都市の位置を示す、2次元の地図を復元できます。

多次元尺度構成は、ブランドや味などに対して消費者が主観的に評価した類似性を分析するために、消費者調査で広く利用されています。消費者調査以外の分野でも、対象の類似性を視覚化したい場合に多次元尺度構成は利用できます。

図 5.1 多次元尺度構成の例



---

## 「多次元尺度構成」プラットフォームの概要

「多次元尺度構成」プラットフォームは、対象間の距離や類似度から、それらの対象の座標を求めます。求められた座標のグラフから、どれとどれが近い位置にあるかを知ることができます。多次元尺度構成は、類似度 (proximity, similarity) や距離 (distance) などの距離行列を分析対象とし、低次元の空間に視覚化します。座標は、ストレス関数（類似度の実測値と予測値との差）を最小化することで求められます。

多次元尺度構成で「距離」といった場合は、都市間の地理的距離などの実際の距離を指すこともあります、主観的・心理的な距離を指す場合がほとんどです。たとえば、製品の類似度、犯罪率の相関、複数の国の経済的類似度といった、何らかの方法で評価された類似度を指します。このような「距離」は、「類似度」(proximity, similarity) や「相違性」(dissimilarity) と呼ばれることもあります。なお、分析対象のデータが、距離行列ではなくて、属性をリストした元データである場合、まずその元データから距離行列を計算したのちに、多次元尺度構成が実行されます。

多次元尺度構成の詳細については、Borg and Groenen (2005) または Jackson (2003) を参照してください。

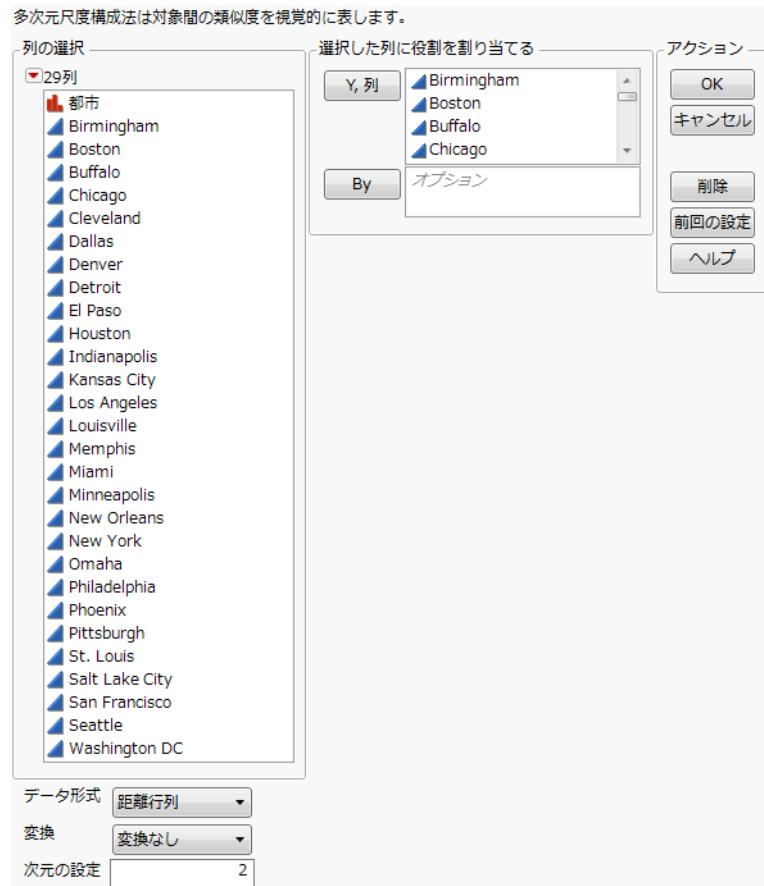
---

## 多次元尺度構成の例

この例では、「Flight Distances.jmp」を使います。このデータテーブルは、米国の28都市を結ぶフライトの飛行距離をまとめたものです。多次元尺度構成を使い、この距離データから2次元の地図を作成します。

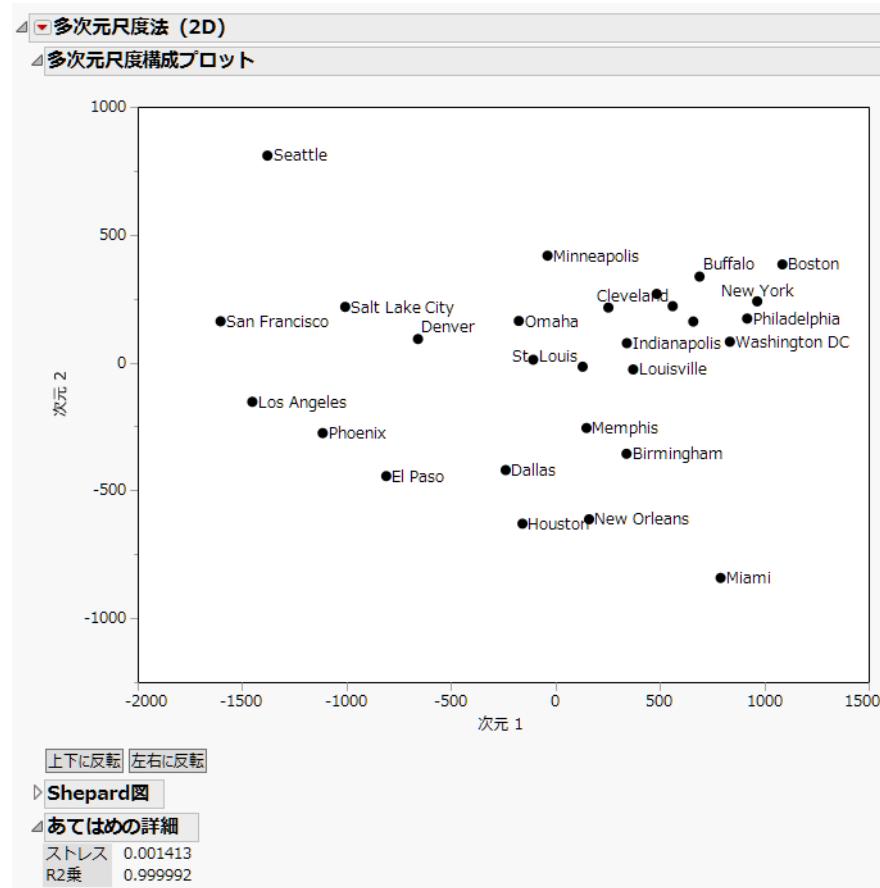
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Flight Distances.jmp」を開きます。
2. [分析] > [消費者調査] > [多次元尺度構成] を選択します。
3. 「Birmingham」から「Washington DC」までを選択し、[Y, 列] をクリックします。

図5.2 入力後の「多次元尺度構成」起動ウィンドウ



[上下に反転] ボタンと [左右に反転] ボタンを使うと、多次元尺度構成プロットの向きを変えることができます。多次元尺度構成の結果は、プロットの向きに影響されません。結果に、明らかな向き（物理的な位置など）がある場合は、プロットの向きを変えると見やすくなります。

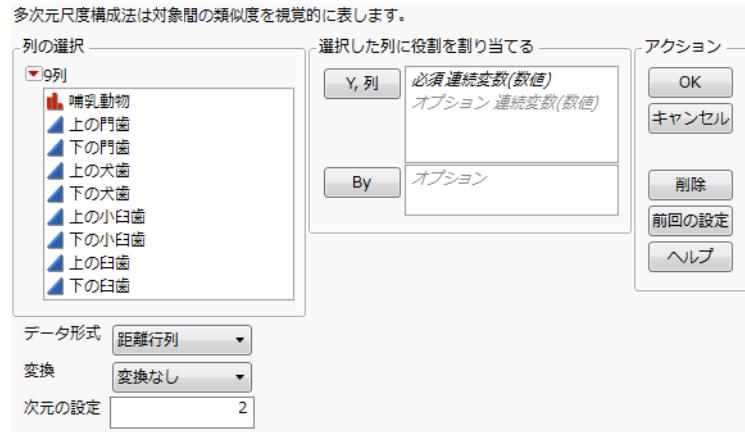
図5.3 多次元尺度構成プロット



## 「多次元尺度構成」プラットフォームの起動

「多次元尺度構成」プラットフォームを起動させるには、[分析] > [消費者調査] > [多次元尺度構成] を選択します。

図5.4 「多次元尺度構成」起動ウィンドウ



**Y, 列** 分析の対象となる列。列のデータタイプは「数値」でなければなりません。

**By** この列の水準に従ってデータがグループ化され、それぞれのグループごとに個別に分析されます。指定した列の各水準について、対応する行が分析されます。分析結果は、水準ごとに個別のレポートにまとめられます。複数の By 变数を指定した場合は、By 变数の水準のすべての組み合わせごとにレポートが作成されます。

---

**メモ:** 距離行列を用いる場合には、すべての対象の距離が含まれた行列が、By 变数の水準ごとに必要となります。

---

**データ形式** 多次元尺度構成では、2つのデータ形式がサポートされています。

**距離行列** 対称行列、または下三角行列。行列の列数と行数が等しく、また、対角要素がゼロか欠測値のいずれかでなければいけません。

**属性の元データ** 品質・特性・評価などの測定値の元データ。このデータには、各対象の名前や ID を含んだ列が含まれているのが普通です。そのような名前や ID の列は、分析自体では使用せず、プロットのデータ点に対するラベルとして使用してください。

**変換** 用意されている変換は、[比尺度]、[間隔尺度]、[順序尺度] の3つです。

**変換なし** 変換は行われません。

**比尺度** 類似度には大小関係があり、それらの差には意味があり、かつ、距離がゼロであることに意味があるときに用いる変換です。変換のときに、類似度が尺度化されます。

**間隔尺度** 類似度には大小関係があり、かつ、それらの差には意味があるときに用いる変換です。変換のときに、類似度がシフトされ、また、尺度化されます。

**順序尺度** 類似度の大小関係だけに意味があり、差などは意味がないときに用いる変換です。類似度が順序尺度であるときに用います。

**次元の設定** オブジェクト間の類似度を視覚化する際、何次元で表示するかを指定します。2次元または3次元が一般的です。3次元を超えると、視覚化が複雑になります。

**メモ:** 次元には、1～n-1 (nは対象の数) の数値を選択できますが、デフォルトでは2に設定されています。

## 「多次元尺度構成」レポート

最初に表示される「多次元尺度構成」レポートには、「多次元尺度構成プロット」、「Shepard図」、「あてはめの詳細」が表示されます。起動ウインドウで3次元以上を指定した場合は、「多次元尺度構成プロット」に次元を選択するコントロールが表示されます。

「多次元尺度構成プロット」でお互いに近くにある点は、属性が類似しています。ラベルや色を点に与えると、似たものどうしを特定するのに役立ちます。「Shepard図」と「あてはめの詳細」を見ると、あてはめたモデルが対象の類似度をどれほどよく表現しているかがわかります。

### 多次元尺度構成プロット

「多次元尺度構成プロット」は、各対象の位置関係を2次元に描いたものです。プロットの下にある2つのボタンを使うと、軸を中心に上下または左右に反転することができます。回転や変換を適用しても、点と点との間の距離は変わりません。特に向きが決まっている地理的な位置をプロットするさいには、回転や反転をしたほうが解釈しやすくなるときがあります。

分析に2次元以上を指定した場合は、プロットの下にある「次元の選択」コントロールを使って次元を切り替えることができます。1つ目のコントロールがプロットの横軸、2つ目のコントロールが縦軸に該当します。プロットする次元を切り替えるには、矢印ボタンをクリックします。

### Shepard図

「Shepard図」は、類似度の予測値と実測値（または変換値）をプロットしたもので、「多次元尺度構成プロット」が実際の類似度をどれほどよく反映しているかを示します。Shepard図は、「予測値と実測値のプロット」と似ています。点が、赤色の直線（YとXの値が等しいところを示す直線）上にあるのが理想です。

### あてはめの詳細

「あてはめの詳細」には、モデルから予測される類似度が実際の類似度にどの程度一致しているかを要約した統計量が表示されます。変換が行われた場合は変換についての詳細も表示されます。

**ストレス** 最適化計算において最小化されたストレス関数（ストレス1）の値。ストレスは、0～1の値で、小さいほど適合度が良いことを示します。

**R2乗** 類似度の予測値と実測値（または変換値）に線形モデルをあてはめた際の  $R^2$  の値。

**傾き** 比尺度または間隔尺度の変換を使用した場合、変換の傾きが表示されます。これは、類似度の実測値を変換値へと変換するときに使われた線形式の傾きです。

**切片** 間隔尺度の変換を使用した場合、変換の切片が表示されます。これは、類似度の実測値を変換値へと変換するときに使われた線形式の切片です。

---

## 「多次元尺度構成」プラットフォームのオプション

「多次元尺度構成」の赤い三角ボタンのメニューには、目的に合わせてレポートをカスタマイズするためのオプションがあります。使用できるオプションの種類は、データのタイプと分析で使用する次元の数によって異なります。

**MDS プロット** 多次元尺度構成プロットの表示／非表示を切り替えます。

**診断統計量** 多次元尺度構成の診断統計量を表示します。

**Shepard 図** 類似度の予測値と実測値（変換した場合は変換値）のプロットを表示します。このレポートはデフォルトで表示されます。詳細については、「[Shepard 図](#)」を参照してください。

**Waern リンク** 多次元尺度構成プロットに Waern リンクを表示します。このオプションを選択すると、最短距離または最長距離の割合を指定するコントロールが表示されます。詳細については、「[Waern リンク](#)」を参照してください。

**座標の表示** 解の座標のレポートを表示します。これは、多次元尺度構成プロットの点の座標です。レポートには、最大3次元までの座標が表示されます。レポート内を右クリックし、[列] を選択すると、次元を追加することができます。使用できる次元の数は、起動ウィンドウで設定した次元数までです。

**類似度の表示** 類似度のレポートを表示します。各ペアの類似度に対して、実測値と予測値（距離）がリストされます。どの対象とどの対象の類似度であるかを示すために、「対象1」と「対象2」の列に名前が表示されます。変換を行った場合は、類似度の変換値も表示されます。

**類似度の保存** (データ形式が「属性の元データ」である場合のみ表示。) 距離行列をデータテーブルに保存します。

**三次元プロット** (起動ウィンドウの「次元の設定」で3次元以上を指定した場合のみ表示。) 最初の3つの次元を使った3次元プロットを表示します。

**座標の保存** 解の座標をデータテーブルの個別の列に保存します。

以下のオプションについて詳しくは、『JMPの使用法』の「JMPのレポート」章を参照してください。

**ローカルデータフィルタ** 現在のレポートにおいてデータをフィルタリングするローカルデータフィルタを表示するか、非表示にします。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、「自動再計算」オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウでBy 変数を指定した場合のみ使用可能です。

## Waern リンク

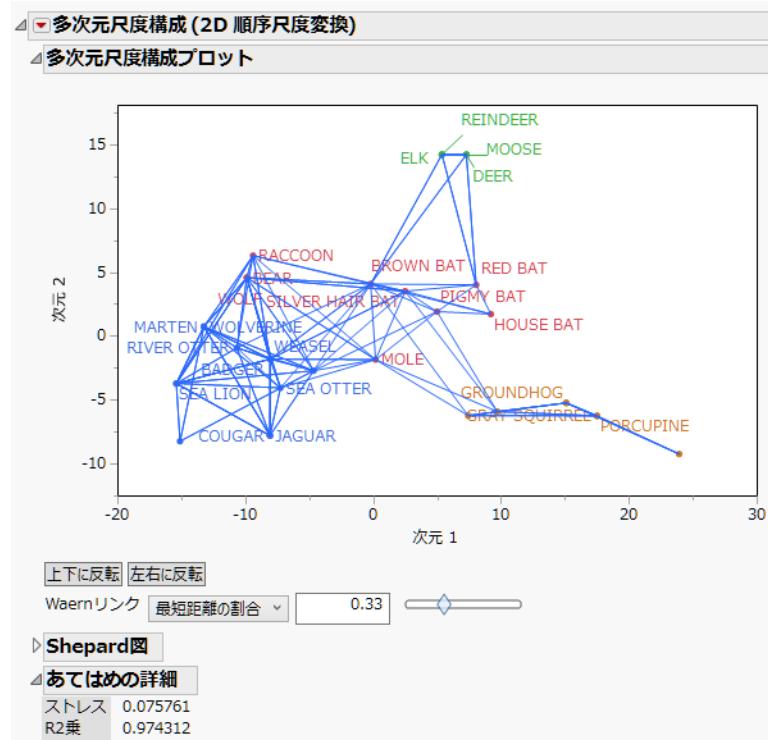
Waern リンクは、類似度の予測値と実測値を比較することで、多次元尺度構成の結果を視覚的に調べます。このリンクは、多次元尺度構成プロット上の点を、類似度の実測値にもとづいて直線で結びます。類似度が小さいもの（もしくは、類似度が大きいもの）を直線で結びます。デフォルトでは、類似度が小さいほうから 33% までを直線で結びます。類似度が小さいものを直線で結んだ場合、プロット上でも距離が近い点どうしが直線で繋がっている場合は、現在のモデルは実際の類似度をよく表していることを示しています。逆に、プロット上で離れた位置にある点が繋がっている場合は、現在のモデルはあてはまっていないことを示しています。

### Waern リンクのコントロール

プロット上に最短距離と最長距離のいずれを表示するかを、[最短距離の割合] と [最長距離の割合] で選ぶことができます。表示するリンクの割合は、テキストボックスに値を入力するか、スライダーを動かすことで調節できます。図5.5は、「Teeth.jmp」データテーブルで最短距離の割合を 33% にしたときの Waern リンクです。

Waern リンクの詳細については、Waern (1972) を参照してください。

図5.5 Waernリンクを表示した多次元尺度構成プロット

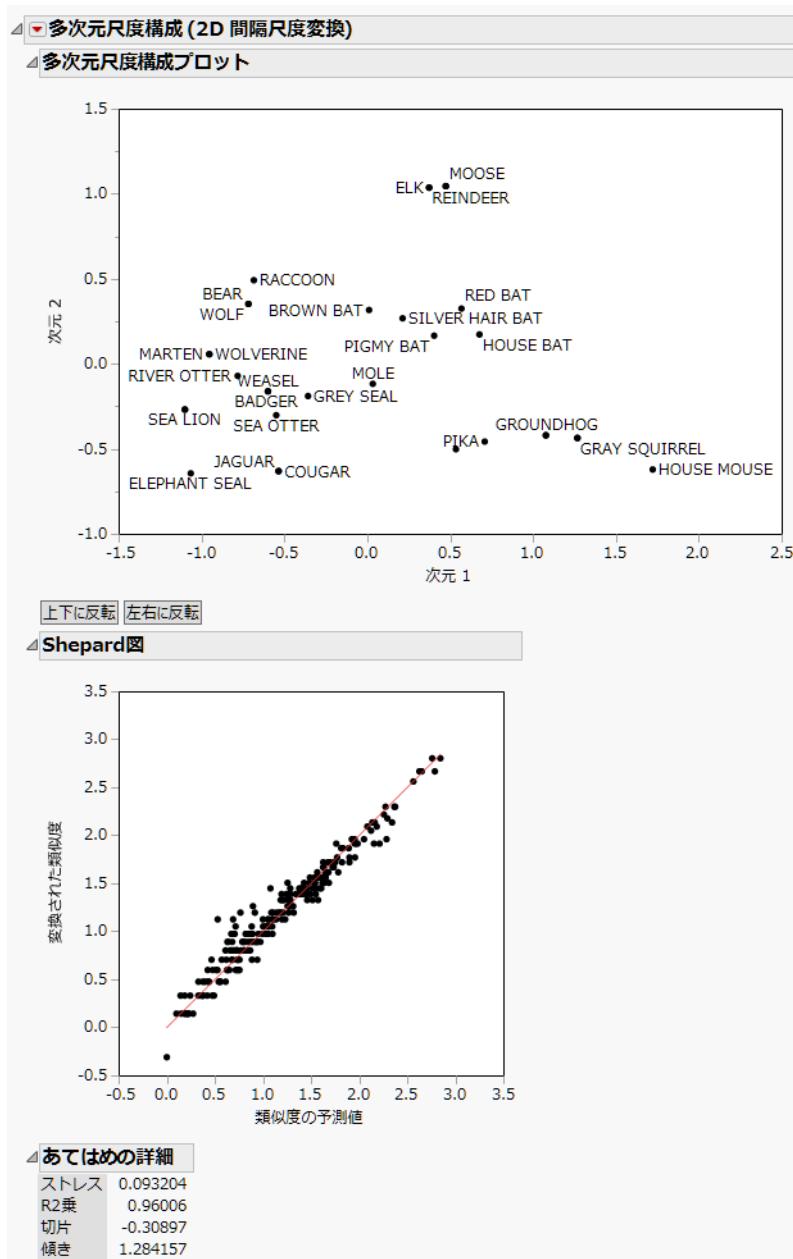


## 「多次元尺度構成」プラットフォームの別例

この例では、32種類の哺乳動物を対象に、8種類の歯の数を数えて記録した「Teeth.jmp」データテーブルを使います。「多次元尺度構成」プラットフォームで、歯に基づいた哺乳動物の類似度を調べます。この例では、変換として、間隔尺度の変換を用います。この元データには、順序に意味があります（4本の歯は、2本の歯よりも多いです）。間隔尺度以外の尺度（たとえば順序尺度）の変換を試してみてもいいでしょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Teeth.jmp」を開きます。
2. 「哺乳動物」列を右クリックし、[ラベルあり/ラベルなし] を選択します。
3. [行] > [行の選択] > [すべての行を選択] を選択します。
4. [分析] > [消費者調査] > [多次元尺度構成] を選択します。
5. 「上の門歯」から「下の臼歯」までを選択し、[Y, 列] を選択します。
6. 「データ形式」から [属性の元データ] を選択します。
7. 「変換」から [間隔尺度] を選択します。
8. [OK] をクリックします。

図5.6 「多次元尺度構成」レポート



「Shepard図」と「あてはめの詳細」を見ると、多次元尺度構成プロットが、歯の観点から見た哺乳動物の類似度をよく表していることがわかります。「ストレス」の値は0.075と低く、類似度の予測値と変換値のあてはめを示すR2乗は0.97と高い値です。「あてはめの詳細」には、類似度の実測値を変換したときに使われた切片と傾きが表示されています。

## 「多次元尺度構成」プラットフォームの統計的詳細

JMPは、準Newton法を使ってストレス関数を最小化し、多次元尺度構成の座標を求めます。この方法では、あらかじめ指定した次元数での座標から予測される距離と実際の類似度との差が小さくなるように、座標が求められます。変換方法が順序尺度の場合は、単調回帰が使用されます。それ以外の変換方法では、最小2乗法による回帰が使用されます。

### ストレス

ストレスの定義では、次のような表記が使用されます。

- $I$  - あてはめのために指定した次元数
- $i, j$  - 対象に対する添え字
- $d_{ij}$  - 対象  $i$  と対象  $j$  との距離の予測値
- $\delta_{ij}$  - 対象  $i$  と対象  $j$  との観測された類似度
- $f(\delta_{rs})$  - 類似度を変換する関数

ストレス関数の計算式は次のとおりです。

$$\text{Stress} = \left[ \frac{\sum_{i < j} [f(\delta_{ij}) - d_{ij}]^2}{\sum_{i < j} d_{ij}^2} \right]^{\frac{1}{2}}$$

この測度は、Kruskal のストレス、タイプI、ストレスIとも呼ばれます。

### 変換

ここでは、「[ストレス](#)」(95ページ) で説明した表記を使用します。変換は、類似度の実測値を尺度化するために行います。変換を通じてデータ内の特定の構造を考慮することで、多次元尺度構成による類似度の実測値の表現が改善されます。変換関数のパラメータが、最小化アルゴリズムにパラメータとして追加されます。

#### 比尺度

比尺度の変換式：

$$f(\delta_{rs}) = b\delta_{rs}$$

#### 間隔尺度

間隔尺度の変換式：

$$f(\delta_{rs}) = a + b\delta_{rs}$$

### 順序尺度

順序尺度のデータは、変換されるわけではなく、アルゴリズムに最小2乗法ではなく単調回帰が使用されます。

## 属性の元データ

データ形式が「属性の元データ」である場合、データをまず距離行列に変換してから多次元尺度構成を適用します。なお、変数間の距離行列は、それらの変数間の相関係数から算出することもできます。

多次元尺度構成の高度な例については、「San Francisco Crime Distances.jmp」データテーブルに保存されている「ソース」スクリプトを参照してください。

# 第 6 章

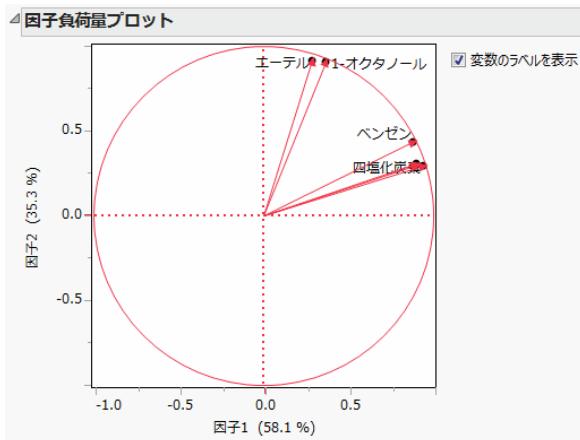
## 因子分析 潜在的な因子によるモデル化

因子分析は、多数の観測変数を、少数の潜在変数（因子）で説明しようとする分析です。JMPでサポートされている因子分析は、「共通因子分析」(common factor analysis) や「探索的因子分析」(exploratory factor analysis) とも呼ばれています。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。なお、各観測変数の分散のうちのどれぐらいが、それぞれの因子によって説明されているかを確認できます。また、観測変数全体が、それぞれの因子によってどれぐらい説明されているかも確認できます。

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6.1 回転後の因子負荷量



## 「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。

$X_1, X_2, \dots, X_{10}$ 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に、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。

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

このモデルでは、 $Var(X_i) = \beta_{i1}^2 + \beta_{i2}^2 + Var(\varepsilon_i)$ となります。 $X_i$ の分散のうち、因子に起因する部分（これを「共通性」と言います）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を超えていたる固有値の個数を、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」（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」章を参照してください。

## 「因子分析」プラットフォームの例

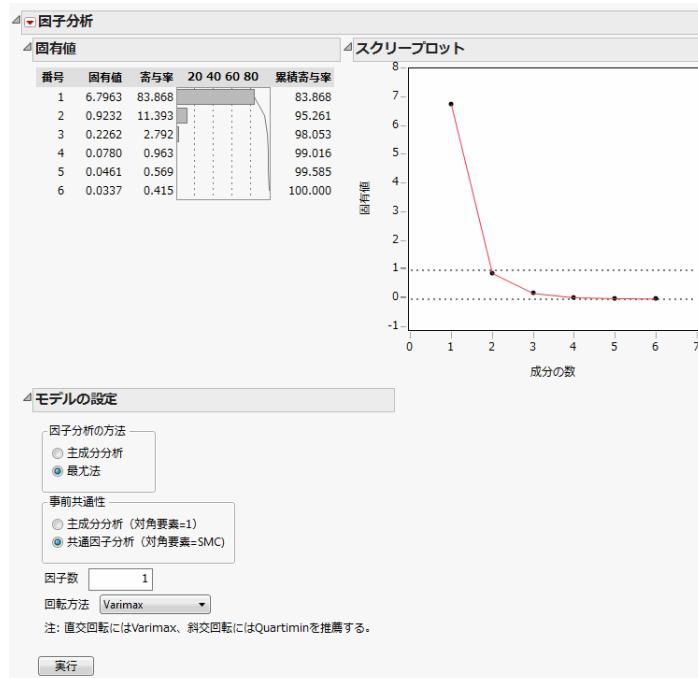
サンプルデータを用いて、2因子の因子分析の例を見ていきましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Solubility.jmp」を開きます。
2. [分析] > [消費者調査] > [因子分析] を選択します。
3. 「因子分析」起動ウィンドウが開きます。
4. 連続尺度の列をすべて選択し、[Y, 列] をクリックします。
5. 「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6.2 最初の「因子分析」レポート



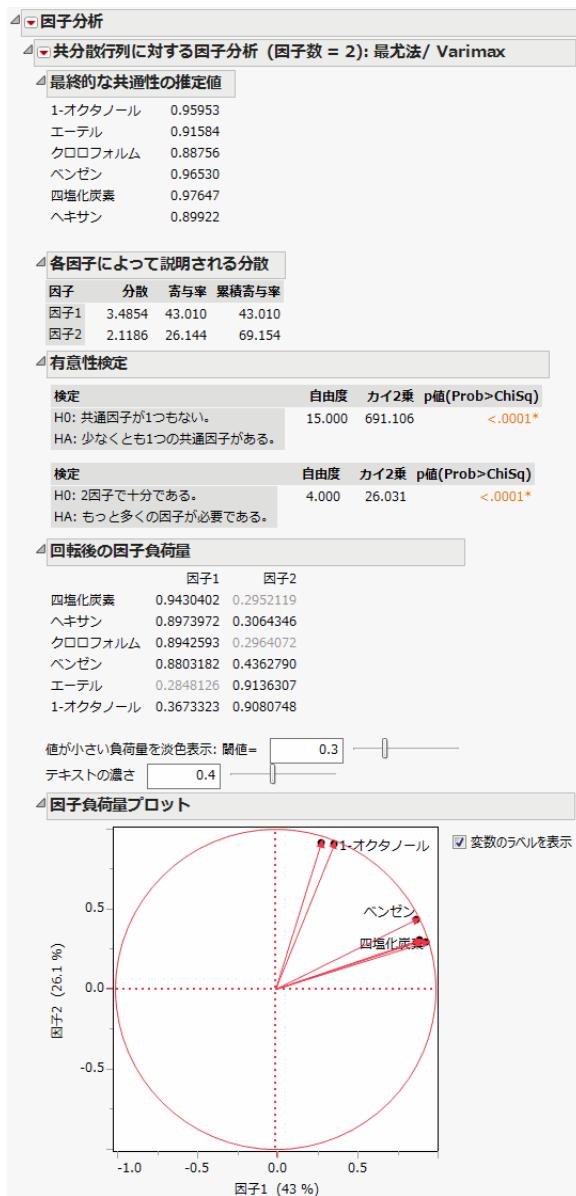
6. 「モデルの設定」で、次のように設定します。

- 「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6.3 「因子分析」レポートの例

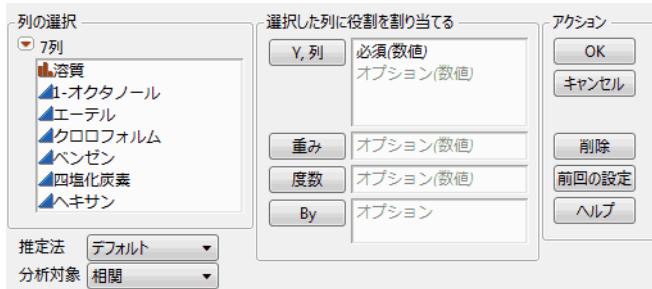


レポートには、共通性の推定値、分散、有意性検定、回転後の因子負荷量、因子負荷量プロットが表示されます。第1因子は「四塩化炭素」、「クロロフォルム」、「ベンゼン」、「ヘキサン」といった変数に、第2因子は「エーテル」、「1-オクタノール」といった変数に強く関連しています。レポートに表示される情報の詳細については、「[モデルをあてはめた後のオプション](#)」(106ページ) を参照してください。

## 「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。

図6.4 「因子分析」起動ウィンドウ



**Y, 列** 指定された連続尺度の列が分析に使用されます。

**重み** 要約されたデータなどを分析するときにおいて、各行に重みを付けます。

**度数** この役割に割り当てられた列に含まれる数値が、分析において各行の度数として使用されます。

**By** By列に含まれる値ごとに、因子分析のレポートが作成されます（By列のグループごとに個別に分析が行われます）。

**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」章を参照してください。

**分析対象** 主成分分析の場合と同じように、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を、因子数としてよいでしょう。

図6.5 「因子分析」レポート

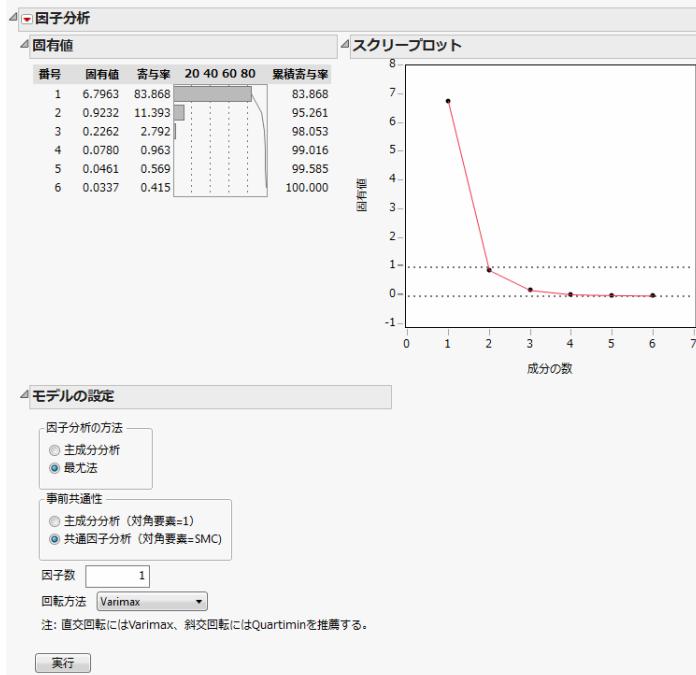
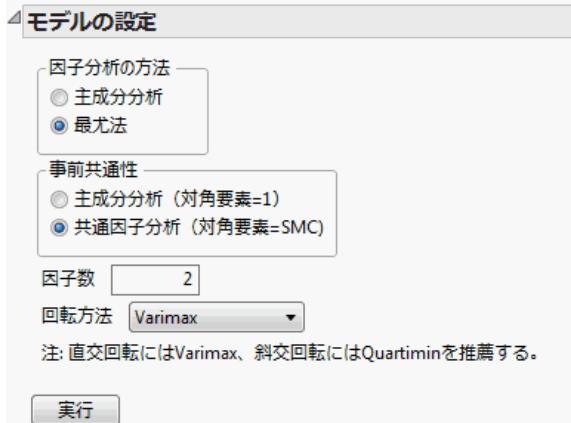


図6.5の例では、スクリープロットは2番目の固有値の後に横ばいになります。固有値の表を見ると、最初の固有値が変動の79.75%、2番目の固有値が15.75%を説明していることがわかります。つまり、この2つで全変動の95.50%が説明されます。3番目の固有値が説明する部分は2.33%に過ぎず、残りの固有値による寄与は取るに足らない程度です。当初の「因子数」は1に設定されていますが、これらの結果を見る限りは、1因子ではなく、2因子とする方が良さそうです。

## 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6.6）。

図6.6 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。

**メモ:** [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、 $X_i$  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。スクリーピットを見て、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。

**メモ:** 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」（104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、「[モデルをあてはめた後のオプション](#)」(106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。

### 直交回転方法

表6.1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6.1 直交回転方法

方法	SASのPROC FACTORでの等価なオプション
Varimax	ROTATE=ORTHOMAX、GAMMA = 1  メモ: 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( <i>p</i> )。 <i>p</i> は直交回転の重み、またはGAMMA= <i>p</i> 。  メモ: <i>p</i> 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 = $\left[ \frac{(nvar(nfact - 1))}{(nvar + nfact - 2)} \right]$ <i>nvar</i> は変数の数で、 <i>nfact</i> は因子の数。
Quartimax	ROTATE=ORTHOMAX、GAMMA=0

## 斜交回転方法

表6.2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6.2 斜交回転方法

方法	SASのPROC FACTORでの等価なオプション
<b>Biquartimin</b>	ROTATE=OBLIMIN(.5) または ROTATE=OBLIMIN、TAU=.5
<b>Covarimin</b>	ROTATE=OBLIMIN(1) または ROTATE=OBLIMIN、TAU=1
<b>Obbiquartimax</b>	ROTATE=OBBIQUARTIMAX
<b>Obequamax</b>	ROTATE=OBEQUAMAX
<b>Obfactorparsimax</b>	ROTATE=OBFATORPARSIMAX
<b>Oblimin</b>	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= $p$ 。 メモ: 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。
<b>Obparsimax</b>	ROTATE=OBPARSIMAX
<b>Obquartimax</b>	ROTATE=OBQUARTIMAX
<b>Obvarimax</b>	ROTATE=OBVARIMAX
<b>Quartimin</b>	ROTATE=OBLIMIN(0) または ROTATE=OBLIMIN、TAU=0
<b>Promax</b>	ROTATE=PROMAX

## 「因子分析」プラットフォームのオプション

「因子分析」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。

**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。スクリープロットの例は、[図6.2](#) (99ページ) を参照してください。

以下のオプションについて詳しくは、『JMPの使用法』の「JMPのレポート」章を参照してください。

**ローカルデータフィルタ** 現在のレポートにおいてデータをフィルタリングするローカルデータフィルタを表示するか、非表示にします。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、[自動再計算] オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウでBy 変数を指定した場合のみ使用可能です。

## モデルをあてはめた後のオプション

「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、粗く推定したものです。

**メモ:** 「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6.7 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、それらの割合を表示します。「SMCを代入した相関行列」とは、対角要素に共通性の推定値であるSMCを代入した相関行列です。その固有値は因子によって説明

される共通分散を示します。途中までの累積パーセントが100%を超える場合があります。なぜなら、「SMCを代入した相関行列」は正值定符号行列とは限らず、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 ( $X_i$  と、その他の  $p-1$  個の変数の重相関を2乗した値) です。

図6.8は、最初の2つの因子が、共通分散の100%を説明していることを示しています。この結果からは、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ことが示唆されます。

図6.8 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	104.1143	86.233	86.233
2	16.6212	13.767	100.000
3	0.7181	0.595	100.595
4	0.0236	0.020	100.614
5	-0.2677	-0.222	100.393
6	-0.4740	-0.393	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なので、回転前の因子負荷量は、因子と観測変数との相関にもなっています。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=」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。

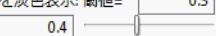
「テキストの濃さ」も、スライダを動かしたり、テキストボックスに入力したりして変更できます。

**メモ:** 「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」でも使われています。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6.9 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=  

テキストの濃さ  

**メモ:** 「回転前の因子負荷量」行列での変数の表示順序は、同じ因子に関連する変数同士が隣り合うように並べ替えられています。

**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6.10 回転行列

回転行列	
0.84800	0.53000
-0.53000	0.84800

**目標行列** Varimax回転のパターン行列を表示します。このオプションを使用できるのはPromax回転の場合のみです。

図6.11 目標行列

目標行列		
	因子1	因子2
1-オクタノール	0.060672	0.915571
エーテル	0.030328	1.000000
クロロフォルム	0.984004	0.035792
ベンゼン	0.827612	0.100628
四塩化炭素	1.000000	0.030643
ヘキサン	0.975109	0.038782

**因子構造** 变数と共通因子の間に見られる相関の行列を表示します。このオプションを使用できるのは斜交回転の場合のみです。

図6.12 因子構造

因子構造		
	因子1	因子2
1-オクタノール	0.6261049	0.9767863
エーテル	0.5491838	0.9569717
クロロフォルム	0.9420450	0.5544378
ベンゼン	0.9713036	0.6834367
四塩化炭素	0.9881517	0.5681409
ヘキサン	0.9480839	0.5649445

**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6.13 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6.14 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	-0.054226
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6.15 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、変数間に見られる相関を有意に説明する共通因子がないことを示します。この検定は、Bartlettの球面性の検定で、「因子の相関行列が単位行列である」という仮説を帰無仮説とします (Bartlett 1954)。

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、変数間に見られる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 (Bartlett 1954)。

図6.16の検定では、モデルに含まれている共通因子が変数間に見られる相関の一部を説明しているものの、より多くの因子が必要であることが示唆されています。

**メモ:** 「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6.16 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。	15.000	691.106	<.0001*
HA: 少なくとも1つの共通因子がある。			
検定	自由度	カイ2乗	p値(Prob>ChiSq)
H0: 2因子で十分である。	4.000	26.031	<.0001*
HA: もっと多くの因子が必要である。			

**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。

メモ: 「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」でも使われています。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6.17 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

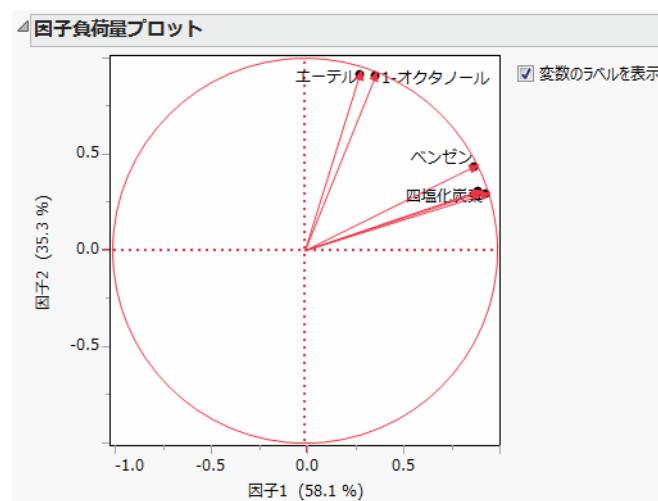
値が小さい負荷量を淡色表示: 閾値=

テキストの濃さ

メモ: 「回転後の因子負荷量」行列での変数の表示順序は、同じ因子に関連する変数同士が隣り合うように並べ替えられています。

**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。

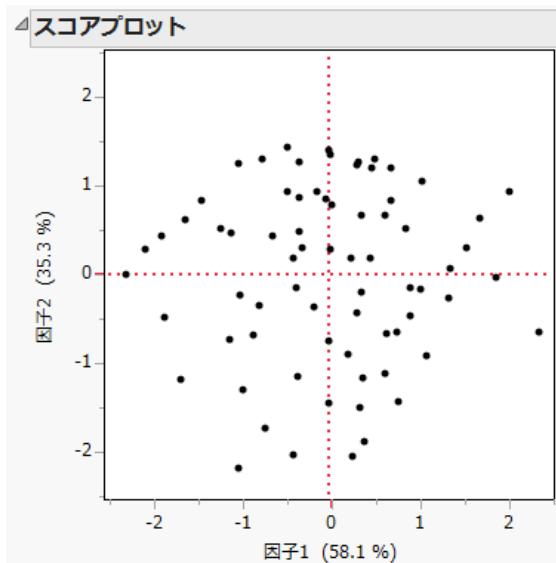
図6.18 因子負荷量プロット



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。具体的な数値は、「回転後の因子負荷量」(110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。因子スコアの値は、推定された因子分析モデルと観測された変数から推定されます。

図6.19 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。

**表示オプション** 矢印を表示できるプロットにおいて、矢印の表示／非表示を切り替えることができます。

**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。



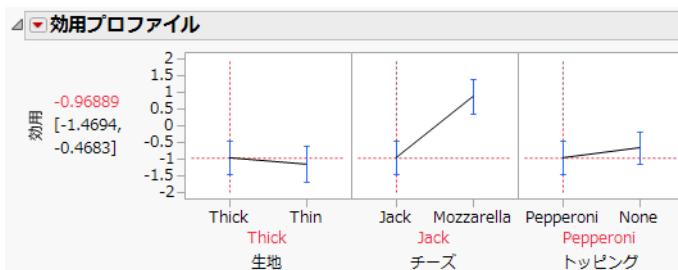
## 選択モデル 選択モデルのあてはめ

「選択モデル」プラットフォームは、マーケットリサーチ（市場調査）のひとつである選択実験（choice experiment）の結果を分析します。製品やサービスのどのような属性を消費者が好むかを特定します。分析結果から、消費者に好まれる製品やサービスを設計することができます。

「選択モデル」プラットフォームでは次のようなことができます。

- 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する。
- 「どの製品も選ばない」という選択肢が用意されている実験を分析する。
- 分析対象のデータとしていくつかの形式がサポートされている。
- プロファイルを使って消費者の効用を視覚化・最適化する。
- 被験者（消費者）ごとにスコアを求め、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。
- **JMP<sup>®</sup> PRO** Bayes（ベイズ）流の方法で、被験者に関する係数の推定値を求める。
- バイアス修正を伴う最尤推定（Firth 1993）を行う。

図7.1 「選択モデル」プラットフォームの効用プロファイル



## 「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」とも呼ばれています。

選択実験は、製品やサービスの属性を組み合わせたもの（プロファイル）を回答者に提示して、どの組み合わせを回答者が好むかを調べます。選択実験では、いくつかのプロファイルを回答者に提示します（回答者に提示する、いくつかのプロファイルを集めたものを、「選択肢集合」と言います）。そして、回答者は、選択肢集合のなかから、最も好きなプロファイルを選択します。各回答者には、選択肢集合を提示して選択を行ってもらうことを、複数回、行ってもらうのが普通です。「選択モデル」プラットフォームは、このような選択実験のデータを分析します。

**メモ:** 選択実験は、「選択モデル計画」プラットフォームで作成できます。『実験計画(DOE)』の「選択モデル計画」章を参照してください。

どの属性を重視するかは消費者によって異なるため、マーケットリサーチ関連の選択実験を分析する場合、消費者をセグメントに分けることが重要でしょう。セグメント化を行わないで製品やサービスを設計すると、実在しない「平均的な」消費者には好まれても、**実在する**消費者の嗜好を無視したものになるかもしれません。

選択モデルの背景については、Louviere et al. (2015)、Train (2009)、Rossi et al. (2006) を参照してください。

### 「選択モデル」プラットフォーム

「選択モデル」プラットフォームは、条件付きロジスティック回帰によって、ある製品が好まれる確率を推定します。「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、製品の属性も考慮したモデルになっています。たとえば自動車を例にすると、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、自動車の属性を選択モデルは考慮します。

「選択モデル」プラットフォームでは、選択肢の集合のいずれも回答者が選択しない場合も扱うことができます。この方法では、回答者がどれも選択しない状況を、単一の属性（「選択なし」という属性）をもつ1つの製品を選んだとみなしてモデル化されます。この「選択なし」に対するパラメータ推定値は、モデルの仮定に応じてさまざまな解釈があります。また、「選択モデル」プラットフォームでは、各回答者に関する情報を求める cũngできます。これは回答者の嗜好パターンをセグメント化するのに役立ちます。

「選択モデル」プラットフォームでは、モデルを推定する方法として、Firth (1993) によるバイアス修正を伴う最尤推定も行えます。この推定法は、通常の最尤法に比べ、推定や検定がより良い性質をもちます。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) の問題が改善できます。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。

**メモ:** 「選択モデル」プラットフォームは、順位やスコアに対するモデルや、枝分かれ階層の選択肢を含むモデルはサポートしていません。そういった分析には、SAS／ETSのPROC MDCを使用してください。

### 選択実験による製品・サービスの開発

満足度調査では、製品やサービスの評判が良いのか悪いのかを知ることはできますが、製品の各属性に対する人々の好みは明らかになりません。製品開発において、数百、数千という細かい仕様をエンジニアは決めていきます。その際、選択実験に協力してくれる消費者を確保できるのであれば、選択実験が製品開発の手助けとなるでしょう。

調査やプロトタイプ作成が安価に行えるのであれば、製品開発の段階で、多数の属性や代替案を消費者に評価してもらえるでしょう。選択実験をシックスシグマプログラムで実施することにより、多くの消費者に求められるように製品を改善できるでしょう。選択実験でデータを収集し、そのデータを分析することにより消費者の選好を明らかにできるでしょう。

### セグメント化

マーケットリサーチ（市場調査）では、選好のパターンによって、消費者をグループに分けたい場合があります。しかし、通常のモデルで各被験者のパラメータ推定値を求めるには、被験者1人あたりに多くのデータが必要です。被験者1人あたりに多くのデータがあるなら、被験者IDを「応答データ」セクションで [By] 変数に指定するか、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、後者の方法は、被験者数が多いと計算が難しくなります。

[By] 変数が指定できるほどのデータがない場合は、「被験者ごとの勾配を保存」オプションを使って被験者をクラスターに分け、セグメント化することができます。このオプションは、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、新しいデータテーブルに表示します。例として、「セグメント化の例」(148ページ) を参照してください。傾きの値についての詳細は、「勾配」(166ページ) を参照してください。

**JMP  
PRO** また、JMP Pro では、階層型Bayes モデルもサポートされています。この階層型Bayes モデルでは、選択モデルにおける製品のパラメータ（このパラメータを、「部分効用 part worth」と呼ぶこともあります）の推定値が被験者ごとに計算されます。その結果をクラスター分析などを行えば、いくつかのセグメントに消費者を分類できます。

---

## 「選択モデル」プラットフォームの例

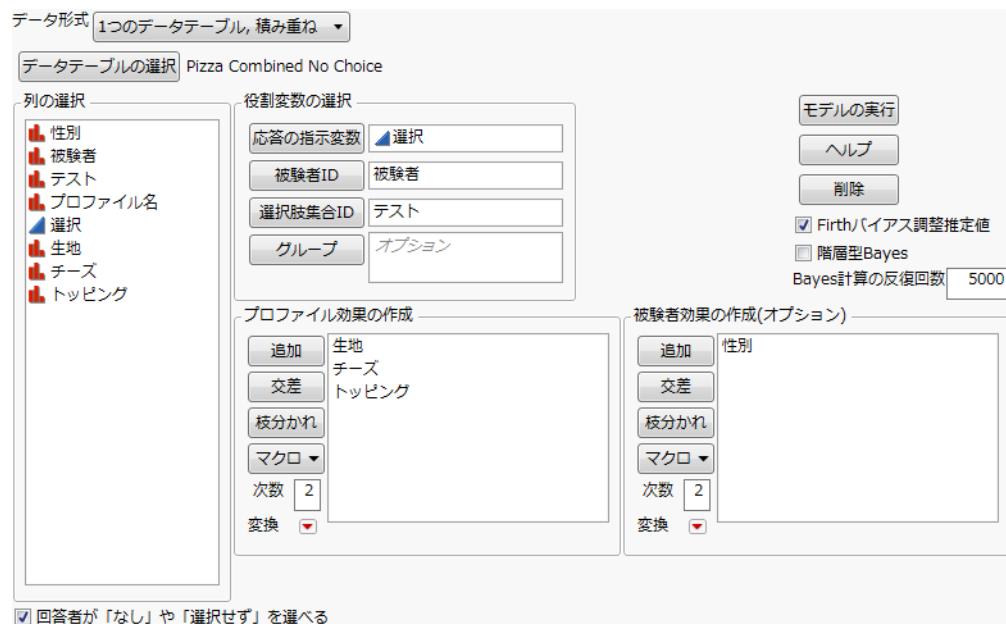
### 「選択なし」がある1つのデータテーブル

どのようなピザが好まれるかを調べた調査で、各回答者に2つのピザのどちらが好きかを、4回尋ねました。回答には、どちらのピザも選ばなかった場合もあります。収集したデータは、1つのデータテーブルにまとめられています。どちらのピザも選ばなかった場合は、指示変数の値が欠測値になっています。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Pizza Combined No Choice.jmp」を開きます。  
選択肢集合は、「被験者」と「テスト」の組み合わせで定義します。一部の選択肢集合で、「選択」列が欠測値になっていることに注目してください。
2. [分析] > [消費者調査] > [選択モデル] を選択します。  
[1つのデータテーブル、積み重ね] がデフォルトです。

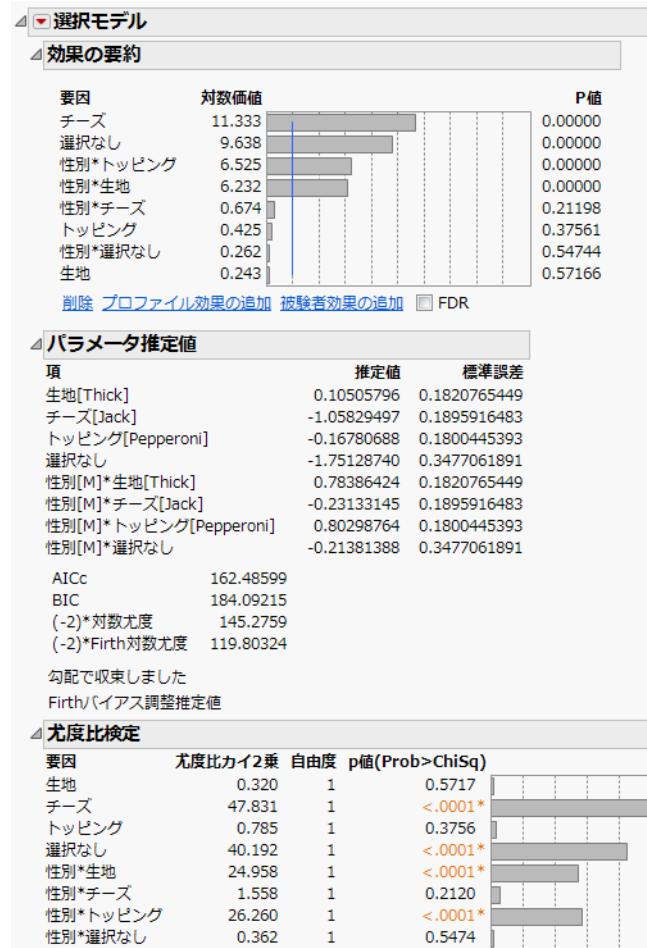
3. [データテーブルの選択] をクリックします。
4. 「Pizza Combined No Choice」を選択し、[OK] をクリックします。
5. 起動ウィンドウで次のように入力を完了します。
  - 「選択」を選択し、[応答の指示変数] をクリックします。
  - 「被験者」を選択し、[被験者ID] をクリックします。
  - 「テスト」を選択し、[選択肢集合ID] をクリックします。
  - 「生地」、「チーズ」、「トッピング」を選択し、「プロファイル効果の作成」パネルで [追加] をクリックします。
  - 「性別」を選択し、「被験者効果の作成(オプション)」パネルで [追加] をクリックします。

図7.2 入力が完了した起動ウィンドウ



6. [回答者が「なし」や「選択せず」を選べる] にチェックマークをつけます。
7. [モデルの実行] をクリックします。

図7.3 「選択なし」を効果として表示したレポート



「効果の要約」レポートは、効果を有意性の順に表示しています。「チーズ」が最も有意な効果で、次に、モデル効果として扱われている「選択なし」が続きます。被験者効果の交互作用である「性別\*トッピング」、「性別\*生地」も有意なことから、「トッピング」と「生地」に対する好みが「性別」というマーケットセグメントに依存すると考えられます。

どんなときにいずれのピザも選択されなかったかを調べるために、元データを確認してみましょう。

8. データテーブルの「選択」列で、応答がないセルを右クリックし、[一致するセルを選択] を選択します。
9. 「行」パネルで「選択されている行」を右クリックし、[データビュー] を選択します。

図7.4 「選択なし」の応答を含む選択肢集合

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 F	2	4	TrimPepperjack	• Thin	Jack	Pepperoni	
2 F	2	4	TrimOni	• Thin	Mozzarella	Pepperoni	
3 M	7	2	Trimella	• Thin	Mozzarella	None	
4 M	7	2	TrimJack	• Thin	Jack	None	
5 M	7	3	Trimella	• Thin	Mozzarella	None	
6 M	7	3	TrimJack	• Thin	Jack	None	
7 F	8	2	ThickElla	• Thick	Mozzarella	None	
8 F	8	2	ThickJack	• Thick	Jack	None	
9 M	11	4	ThickOni	• Thick	Mozzarella	Pepperoni	
10 M	11	4	ThickJackoni	• Thick	Jack	Pepperoni	
11 F	14	2	TrimOni	• Thin	Mozzarella	Pepperoni	
12 F	14	2	TrimPepperjack	• Thin	Jack	Pepperoni	
13 F	18	3	ThickJack	• Thick	Jack	None	
14 F	18	3	ThickElla	• Thick	Mozzarella	None	
15 F	24	3	ThickJack	• Thick	Jack	None	
16 F	24	3	TrimOni	• Thin	Mozzarella	Pepperoni	
17 M	29	1	TrimPepperjack	• Thin	Jack	Pepperoni	
18 M	29	1	ThickJackoni	• Thick	Jack	Pepperoni	

図7.4のデータテーブルにおいて、「被験者」と「テスト」の組み合わせで定義された最初の7つの選択肢集合（行1～14）を見てください。各選択肢集合において、2つのピザは、「チーズ」だけが異なっています。これらの回答者は、チーズの違いを見極められなかった可能性があります。しかし、分析結果においては、「選択なし」が考慮されるため、「チーズ」の効果が統計的に有意になっています。

このようなデータをさらに分析する方法については、「[最適なプロファイルを見つける](#)」（122ページ）を参照してください。

## 複数のデータテーブル

この節でも、3つの属性をもつピザを調べた実験を例にします。

この例では、「Pizza Profiles.jmp」、「Pizza Responses.jmp」、「Pizza Subjects.jmp」の3つのデータテーブルを使います。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Pizza Profiles.jmp」、「Pizza Responses.jmp」、「Pizza Subjects.jmp」を開きます。
  - プロファイルデータテーブルの「Pizza Profiles.jmp」は、被験者に提示する選択肢の組み合わせをリストしたものです。それぞれの組み合わせに、「ID」を割り当てています。
  - 応答のデータテーブルである「Pizza Responses.jmp」に実験計画と結果が含まれています。実験では、各被験者に4つの選択肢集合が提示されます。各集合は、選択肢である2つのプロファイル（「選択肢1」と「選択肢2」）から成ります。被験者は、各選択肢集合の中から好きな方（「選択」）を選びます。選択モデル計画の作成方法については、『実験計画 (DOE)』の「離散選択モデル計画」章を参照してください。このデータテーブルにおける「選択」列の値は、プロファイルのデータテーブルにおける「ID」列の値と対応しています。

- 被験者のデータテーブルである「Pizza Subjects.jmp」には、被験者 ID を値とする「被験者」列と、被験者の特性である「性別」が保存されています。「Pizza Subjects.jmp」の「被験者」列の値は、「Pizza Responses.jmp」の「被験者」列の値に対応しています。

2. [分析] > [消費者調査] > [選択モデル] を選択し、起動ウィンドウを開きます。

**メモ:** これは、開いている3つのデータテーブルのどれからでも実行できます。

3. 「データ形式」メニューから [複数のデータテーブル, 相互参照] を選択します。

各データソースに対応する3つのアウトラインがあります。

4. 「プロファイルデータ」アウトラインにある [データテーブルの選択] をクリックします。

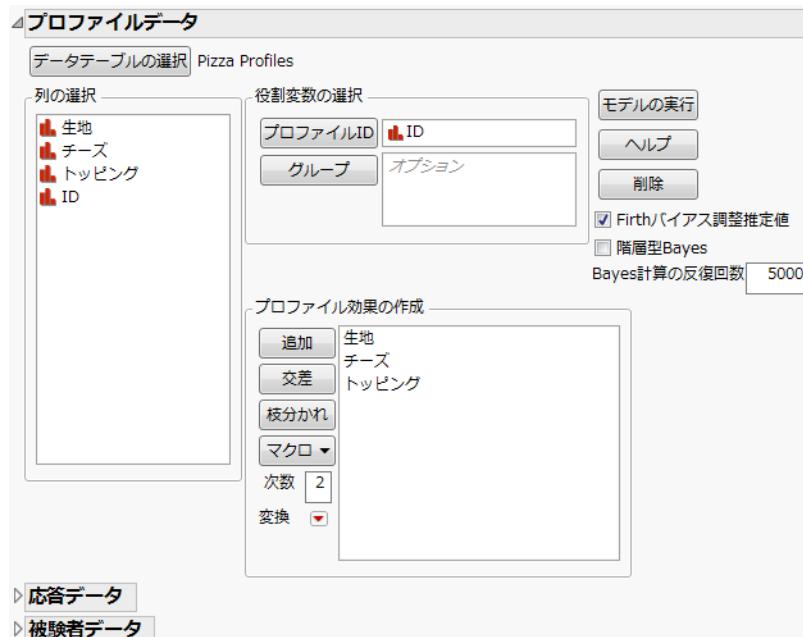
「プロファイルデータ」ウィンドウが開き、ここでプロファイルデータとして使うデータテーブルを指定します。

5. 「Pizza Profiles.jmp」を選択し、[OK] をクリックします。

6. 「ID」を選択し、[プロファイルID] をクリックします。

7. 「生地」、「チーズ」、「トッピング」を選択し、[追加] をクリックします。

図7.5 プロファイルデータ

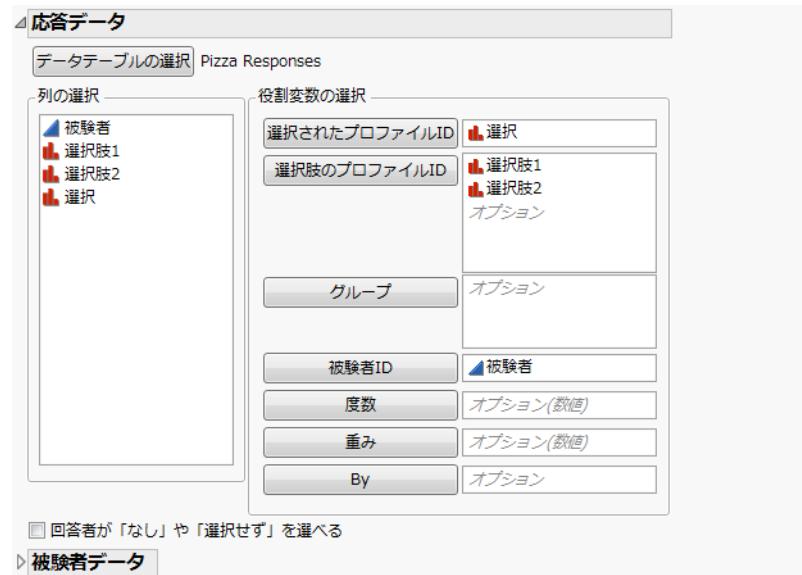


8. 「応答データ」の開閉アイコンをクリックしてアウトラインを開き、[データテーブルの選択] をクリックします。
9. 「Pizza Responses.jmp」を選択し、[OK] をクリックします。

10. 次の手順を行います。

- 「選択」を選択し、[選択されたプロファイルID] をクリックします。
- 「選択肢1」と「選択肢2」を選択し、[選択肢のプロファイルID] をクリックします。
- 「被験者」を選択し、[被験者ID] をクリックします。

図7.6 「応答データ」パネル



「選択肢1」と「選択肢2」という列は、各被験者に4回ずつ提示される2つの選択肢を表しています。また、「選択」という列には、「選択肢1」と「選択肢2」のどちらが選択されたかが入力されています。

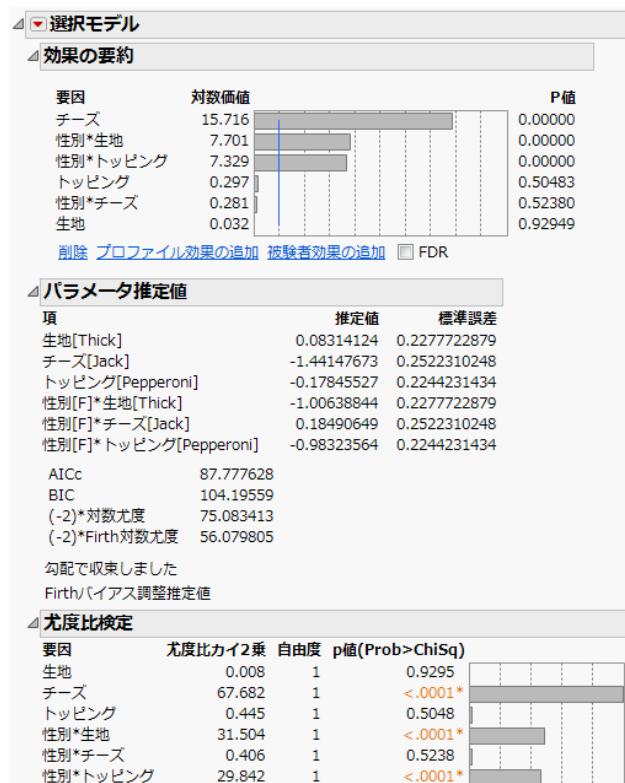
- 「被験者データ」の開閉アイコンをクリックしてアウトラインを開き、[データテーブルの選択] をクリックします。
- 「Pizza Subjects.jmp」を選択し、[OK] をクリックします。
- 「被験者」を選択し、[被験者ID] をクリックします。
- 「性別」を選択し、[追加] をクリックします。

図7.7 「被験者データ」パネル



15. [モデルの実行] をクリックします。

図7.8 「選択モデル」の結果



モデルに6つの効果が含まれています。「生地」、「チーズ」、「トッピング」は商品属性です。「性別\*生地」、「性別\*チーズ」、「性別\*トッピング」は、被験者効果と商品属性の交互作用です。これらの交互作用を検討することにより、各マーケットセグメントに好まれる商品を探すことができます。

**メモ:** 「選択モデル」の場合、被験者効果を主効果として含めることはできません。被験者効果は、交互作用項としてのみ表示されます。

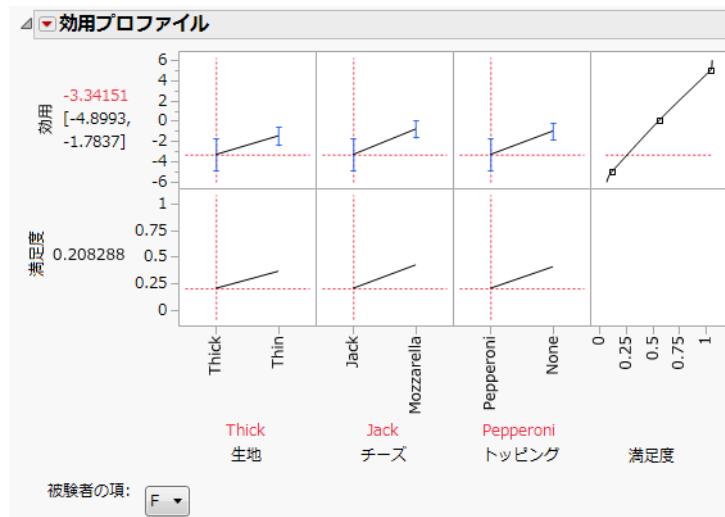
「効果の要約」と「尤度比検定」の各レポートから、「性別」と「生地」、「性別」と「トッピング」の間に強い交互作用があることがわかります。「生地」と「トッピング」の主効果は、有意ではありません。この例で被験者効果を含めなければ、マーケットのセグメント化に役立つ重要な情報を見逃したかもしれません。

### 最適なプロファイルを見つける

次に、「効用プロファイル」を使って結果を分析し、属性の最適設定を見つけましょう。

- 「選択モデル」の赤い三角ボタンをクリックし、[効用プロファイル]を選択します。  
プロファイルの下にある「被験者の項」は、女性の結果が表示されていることを示します。
- 「効用プロファイル」の赤い三角ボタンをクリックし、[最適化と満足度] > [満足度関数]を選択します。

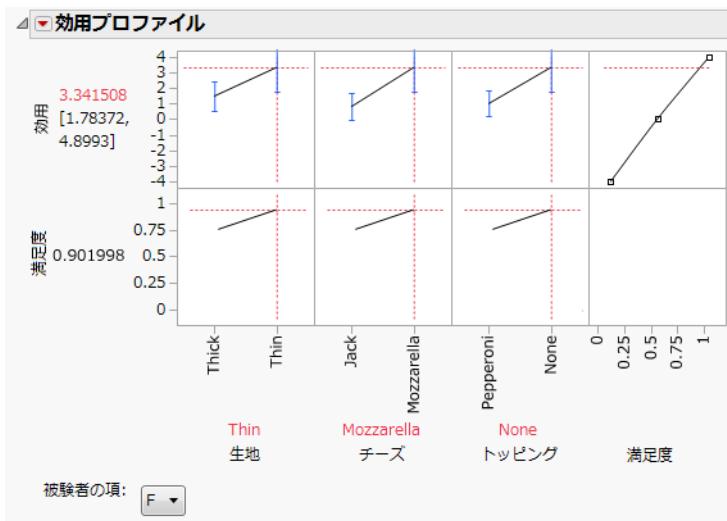
図7.9 満足度関数を表示した効用プロファイル



プロファイルに、効用を最大化する満足度関数が加わりました。『プロファイル機能』の「プロファイル」章を参照してください。

- 「効用プロファイル」の赤い三角ボタンをクリックし、[最適化と満足度] > [満足度の最大化]を選択します。

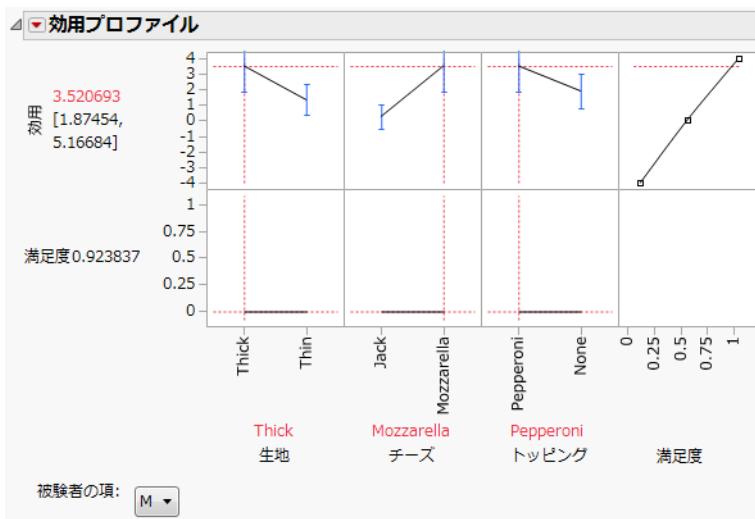
図7.10 女性に最適な設定の効用プロファイル



女性向けの最適な設定は、薄い生地、モツツアレラチーズ、トッピングなしです。

4. 「被験者の項」メニューから [M] を選択します。

図7.11 「性別」の水準を「M」(男性)に固定したときのユーティリティプロファイル



男性向けの最適な設定は、厚い生地、モツツアレラチーズ、ペパロニのトッピングです。

この例では、マーケットを性別で2つのセグメントに分け、その嗜好を理解して、それぞれに最適なピザを特定しました。

## 「選択モデル」プラットフォームの起動

「選択モデル」プラットフォームを起動するには、[分析] > [消費者調査] > [選択モデル] を選択します。

「選択モデル」プラットフォームで使用するデータは、1つのデータテーブルにまとめたものでも、2、3個のデータテーブルに分かれているものでもかまいません。「選択モデル」起動ウィンドウの冒頭で「データ形式」を指定します。

### 1つのデータテーブル、積み重ね

この形式では、データを1つのデータテーブルにまとめてください。被験者に提示されるプロファイルごとに1つの行を作成し、また、特定のプロファイルが選択されたかどうかを示す指示変数を作成してください。

データが1つのデータテーブルにまとめられている形式の例は、「「選択なし」がある1つのデータテーブル」(115ページ) を参照してください。詳細については、「[1つのデータテーブル、積み重ね] の起動ウィンドウ」(125ページ) を参照してください。

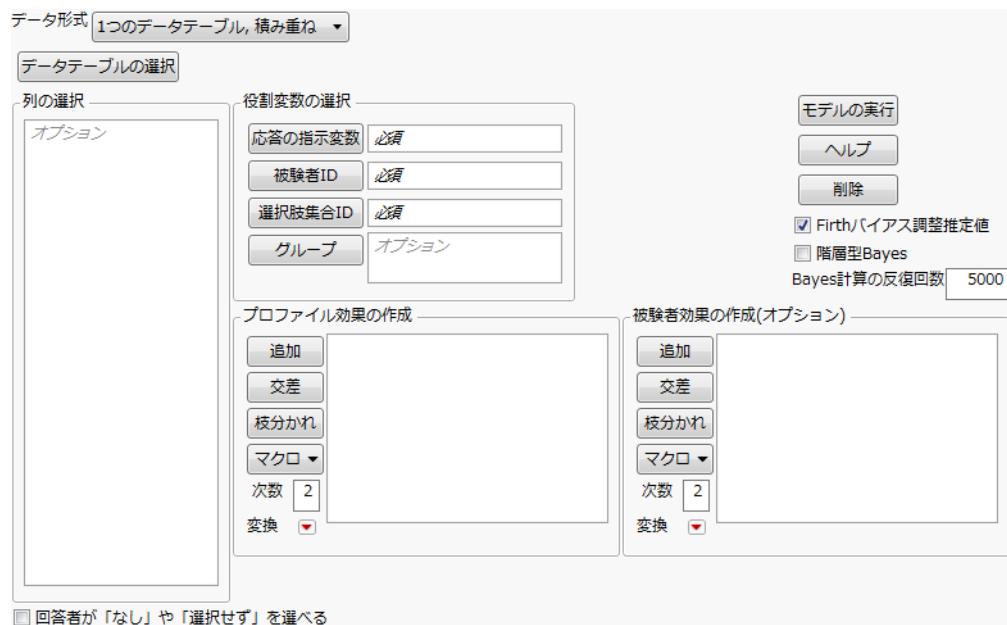
### 複数のデータテーブル、相互参照

この形式では、データが2つまたは3つのデータテーブルに分かれています。プロファイルデータと応答データは必須で、被験者データはオプションで使うことができます。「選択モデル」起動ウィンドウは、各データテーブルに対応する3つのセクションで構成されています。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。

データが複数のデータテーブルに分かれている形式については、「複数のデータテーブル」(118ページ) を参照してください。詳細については、「[複数のデータテーブル、相互参照] の起動ウィンドウ」(126ページ) を参照してください。

## [1つのデータテーブル, 積み重ね] の起動ウィンドウ

図7.12 [1つのデータテーブル, 積み重ね] データ形式の起動ウィンドウ



**データテーブルの選択** データが保存されているデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、[その他...]を選択します。

**応答の指示変数** その選択肢が選択されたかどうかを示す値を含んだ列です。「1」は選択されたプロファイル、「0」は選択されなかったプロファイルを示します。回答者が何も選択しないことを許している場合は、いずれの選択肢も選択されなかった集合には欠測値を入力します。詳細については、「回答者が「なし」や「選択せず」を選べる」(126ページ) を参照してください。

**被験者ID** 調査に参加した人のID。

**選択肢集合ID** 被験者に提示された選択肢集合のID。被験者は、この選択肢集合のなかから、最も好きなものを選択します。

**グループ** 「選択肢集合ID」列と共に使用したときに、各選択肢集合を一意に示すことができる列。たとえば、「選択肢集合ID」の値が1である選択肢集合が2つあったとします。一方の「調査」の値がAで、他方の選択肢集合の「調査」がBであるなら、「調査」を「グループ」列に指定します。

**プロファイル効果の作成** プロファイルの属性で構成された効果を追加します。

「プロファイル効果の作成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

**被験者効果の作成(オプション)** 被験者に関する効果を追加します。

「被験者効果の作成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

**Firthバイアス調整推定値** バイアス修正を伴う最尤推定は、通常の最尤推定に比べ、推定や検定がより良い性質をもたらします。さらに、ロジスティックモデルなどで生じる分離の問題も改善できます。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。

**JMP<sup>®</sup> PRO 階層型 Bayes** Bayes流の方法で被験者に関する係数の推定値を求める。詳細については、「[Bayes パラメータ推定値](#)」(133ページ) を参照してください。

**JMP<sup>®</sup> PRO Bayes計算の反復回数** ([階層型 Bayes] を選択した場合のみ有効。) 被験者に関する係数の推定値を計算する際に使う適応型 Bayes アルゴリズムの反復回数。ここで指定した反復回数には、推定からは破棄されるバーンインの反復回数も含まれています。バーンインの反復回数は、起動ウィンドウで指定した「Bayes 計算の反復回数」の半分です。

**回答者が「なし」や「選択せず」を選べる** 応答が欠測値である行を「選択なし」としてモデルに含めます。[1つのデータテーブル、積み重ね] 形式の場合、回答者がどれも選択しなかったときは、「応答の指示変数」列を(数値)欠測値にしてください。このオプションは、起動ウィンドウの一番下に表示されます。

## [複数のデータテーブル、相互参照] の起動ウィンドウ

図7.13 [複数のデータテーブル、相互参照] データ形式の起動ウィンドウ

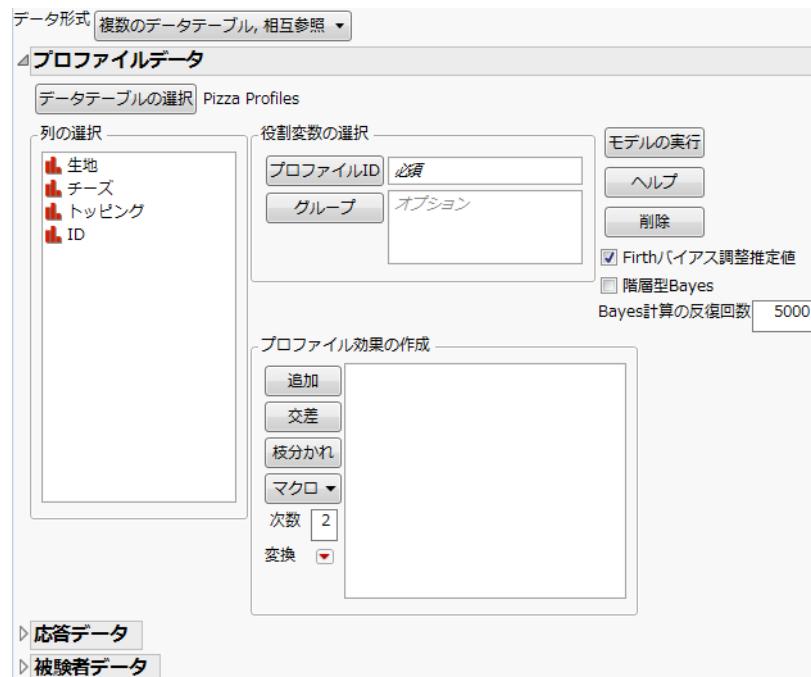


図7.13は、複数のデータテーブル用の起動ウィンドウで、「Pizza Profiles.jmp」をプロファイルテーブルとして使用しています。

[複数のデータテーブル, 相互参照] の場合、起動ウィンドウは3つのセクションに分かれています。

- 「プロファイルデータ」 (127ページ)
- 「応答データ」 (128ページ)
- 「被験者データ」 (130ページ)

### プロファイルデータ

プロファイルデータテーブルには、それぞれの選択肢の属性を示すデータを保存してください。データテーブルの各列が1つの属性に対応するように、また、各行が1つのプロファイルに対応するように、データを作成してください。さらに、各プロファイルのIDを含んだ列を設けてください。図7.14は、「Pizza Profiles.jmp」データテーブルと、列を指定した後の「プロファイルデータ」パネルです。

図7.14 プロファイルデータテーブルと設定後の「プロファイルデータ」アウトライン

	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	TrimPepperoni
8	Thin	Jack	None	TrimJack

△プロファイルデータ

データテーブルの選択 Pizza Profiles

列の選抲

- 生地
- チーズ
- トッピング
- ID

役割変数の選抲

- プロファイルID
- ID

グループ オプション

モデルの実行

ヘルプ

削除

Firth/バイアス調整推定値

階層型Bayes

Bayes計算の反復回数 5000

プロファイル効果の作成

追加 生地  
チーズ  
トッピング

交差

枝分かれ

マクロ ▾

次数 2

変換 ▾

△応答データ

△被験者データ

**データテーブルの選択** 応答のデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、[その他...] を選択します。

**プロファイルID** 属性の組み合わせ（プロファイル）を識別するためのID。[プロファイルID] によってプロファイルデータテーブルの各行を一意に識別できない場合は、[グループ] 変数も指定する必要があります。その場合、[グループ] 列と [プロファイルID] 列の組み合わせによって各行が一意に識別できるように、[グループ] 列を追加してください。

**グループ** 「選択肢集合ID」列と共に使用したときに、各選択肢集合を一意に示すことができる列。たとえば、[プロファイルID] の値が「1」である行が2行あったとします。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。

**プロファイル効果の作成** プロファイルの属性で構成された効果を追加します。

「プロファイル効果の作成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

**Firthバイアス調整推定値** バイアス修正を伴う最尤推定は、通常の最尤推定に比べ、推定や検定がより良い性質をもちます。さらに、ロジスティックモデルなどで生じる分離の問題も改善できます。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。

**JMP PRO 階層型 Bayes** Bayes流の方法で被験者に関する係数の推定値を求めます。詳細については、「[Bayes パラメータ推定値](#)」(133 ページ) を参照してください。

**JMP PRO Bayes計算の反復回数** ([階層型 Bayes] を選択した場合のみ有効。) 被験者に関する係数の推定値を計算する際に使う適応型 Bayes アルゴリズムの反復回数。ここで指定した反復回数には、推定からは破棄されるバーンインの反復回数も含まれています。バーンインの反復回数は、起動ウインドウで指定した「[Bayes計算の反復回数](#)」の半分です。

## 応答データ

応答データテーブルには、被験者ID・選択肢のプロファイルID・選択されたプロファイルIDの列を設けてください。そして、各被験者と選択肢集合ごとに1行ずつ保存してください。選択肢集合のグループが複数ある場合は、グループ変数を使って選択肢集合を区別してください。図7.15は、「[Pizza Responses.jmp](#)」データテーブルと入力後の「応答データ」パネルです。

なお、応答データとプロファイルデータにおけるプロファイルIDを対応させるのにグループ化が必要な場合には、対応がとれるようにグループ変数を指定してください。

図7.15 応答データテーブルと設定後の「応答データ」アウトライン

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

△応答データ

データテーブルの選択 Pizza Responses

列の選択

- 被験者
- 選択肢1
- 選択肢2
- 選択

役割変数の選択

選択されたプロファイルID	選択
選択肢のプロファイルID	選択肢1 選択肢2 オプション
グループ	オプション
被験者ID	被験者
度数	オプション(数値)
重み	オプション(数値)
By	オプション

□回答者が「なし」や「選択せず」を選べる

▷被験者データ

**データテーブルの選択** プロファイルのデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、[その他...]を選択します。

**選択されたプロファイルID** 被験者が選択したプロファイルを示すID。このIDは、プロファイルデータテーブルにおけるプロファイルIDに対応したもの。

**グループ** 「選択されたプロファイルID」列と共に使用したときに、各選択肢集合を一意に示すことができる列。

**選択肢のプロファイルID** 被験者に提示された複数のプロファイルを示すID。

**被験者ID** 調査に参加した人のID。

**度数** 度数を含んだ列。度数が  $n$  である行は、データに  $n$  回登場しているものとして計算に使用されます。度数が 1 未満である行や、欠測値である行は、分析に使用されません。

**重み** データテーブルの各行に対する重みを含んだ列。0 より大きい重みだけが分析に使われます。

**By** この列の水準に従ってデータがグループ化され、それぞれ個別に分析されます。指定した列の水準ごとに、対応する行が個別のテーブルとして個別に分析されます。分析結果は、個別のレポートにまとめられます。複数の By 変数を指定した場合は、By 変数の水準のすべての組み合わせごとに分析が行われます。

**回答者が「なし」や「選択せず」を選べる** 応答が欠測値である行を「選択なし」としてモデルに含めます。[複数のデータテーブル、相互参照] データ形式の場合、回答者がどれも選択しなかったときは、「選択されたプロファイル ID」列を欠測値にしてください。このオプションは、「応答データ」パネルの最後に表示されます。

## 被験者データ

被験者データテーブルは、必須ではなく、モデルに被験者効果を含めたいときに使用します。このデータテーブルの列には、被験者 ID（応答データテーブルでも使用されているもの）、および、被験者の属性を含めてください。被験者データは応答データテーブルに含めてもかまいませんが、その場合も、「被験者データ」アウトラインで被験者効果を指定する必要があります。図7.16は、「Pizza Subjects.jmp」データテーブルと入力後の「プロファイルデータ」パネルです。

図7.16 被験者データテーブルと設定後の「被験者データ」アウトライン

	被験者	性別
1	1	M
2	2	F
3	3	M
4	4	F

被験者データ

データテーブルの選択 Pizza Subjects

列の選択  
 被験者  
 性別

役割変数の選択  
 被験者ID  
 被験者

モデル効果の構成  
 性別  
  
  
 次数

**データテーブルの選択** 被験者のデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、[その他...]を選択します。

**被験者ID** 被験者を一意に識別する列。

**モデル効果の構成** 被験者データテーブルの列で構成した効果を追加します。

「モデル効果の構成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

---

## 「選択モデル」レポート

起動ウィンドウで [階層型 Bayes] を選択した場合を除き、「選択モデル」レポートは次のような構成です。

- 「効果の要約」 (131ページ)
- 「パラメータ推定値」 (132ページ)
- 「尤度比検定」 (133ページ)

**メモ:** 「効果の要約」と「尤度比検定」の各レポートは、データセットが比較的小さく、計算にあまり時間がかかるない場合のみ、デフォルトで作成されます。表示されない場合は、赤い三角ボタンのメニューから [尤度比検定] を選択すれば、両方のレポートが表示されます。

---

**JMP PRO** 起動ウィンドウで [階層型 Bayes] を選択した場合は、次のようなレポートが作成されます。

- 「Bayesパラメータ推定値」 (133ページ)

### 効果の要約

モデルに1つ以上の効果があり、計算にあまり時間がかかるない場合に、「効果の要約」レポートが表示されます。(表示されない場合は、赤い三角ボタンのメニューから [尤度比検定] を選択すると、両方のレポートが表示されます。) モデルにおいて推定された効果と、それらの効果の対数値 (またはFDR対数値) のプロットで構成されます。また、このレポートでは、モデルに効果を追加したり削除したりできます。このとき、モデルをあてはめた結果は、「効果の要約」レポートでの変更に合わせて自動的に更新されます。詳細については、『基本的な回帰モデル』の「標準最小2乗のレポートとオプション」章にある「効果の要約レポート」の節を参照してください。

起動ウィンドウで [Bayes流の被験者効果] チェックボックスをオンにすると、「効果の要約」レポートは表示されません。尤度比検定が行われないためです。

### 「効果の要約」表の列

「効果の要約」表には次の列があります。

**要因** モデル内の効果。 $p$ 値の小さい順に並べられます。

**対数値** 各効果の対数値 (LogWorth)。 $-\log_{10}(p\text{ 値})$  という式で計算されています。 $p$  値を対数値に変換すると、解釈がしやすくなります。対数値が 2 以上あるものは、有意水準 0.01 で有意です ( $-\log_{10}(0.01) = 2$ )。

**FDR 対数値** 各効果のFDR (False Discovery Rate; 偽発見率) の対数値で、 $-\log_{10}(\text{FDR 調整 } p\text{ 値})$  という式で計算されています。これは、検定の有意性をグラフに表すのに適している統計量です。「対数値」を「**FDR 対数値**」に変更するには、[FDR] チェックボックスをオンにします。

**棒グラフ** 対数値 (または FDR 対数値) の棒グラフ。このグラフには整数値に縦の点線が、2 の位置に青の参照線が引かれています。

**p値** 各モデル効果の  $p$  値。この  $p$  値は、「尤度比検定」レポートにある有意性検定に対応します。

**FDR p値** Benjamini-Hochberg 法で計算された、各効果の FDR 調整  $p$  値。FDR は、検定の多重性を考慮して、生の  $p$  値を調整したものです。「**P値**」列を「**FDR p値**」に変更するには、[FDR] チェックボックスをオンにします。

FDR については、Benjamini and Hochberg (1995) を参照してください。偽発見率の詳細については、『予測モデルおよび発展的なモデル』の「応答のスクリーニング」章、または Westfall et al. (2011) を参照してください。

## 「効果の要約」表のオプション

要約表の下にあるオプションによって、効果を追加や削除できます。

**削除** モデルから選択した効果を削除します。1つまたは複数の効果を削除するには、削除する効果に対応する列を選択し、[削除] ボタンをクリックします。

**プロファイル効果の追加** パネルが開き、[1つのデータテーブル, 積み重ね] データ形式ならデータテーブルのすべての列、[複数のデータテーブル, 相互参照] データ形式ならプロファイルデータテーブルの列がリストされます。モデルに追加する列を選択し、リストの下にある [追加] ボタンをクリックすると、その効果がモデルに追加されます。[閉じる] をクリックすると、パネルが閉じます。

**被験者効果の追加** パネルが開き、[1つのデータテーブル, 積み重ね] データ形式ならデータテーブルのすべての列、[複数のデータテーブル, 相互参照] データ形式なら被験者データテーブルの列がリストされます。モデルに追加する列を選択し、リストの下にある [追加] ボタンをクリックすると、その効果がモデルに追加されます。[閉じる] をクリックすると、パネルが閉じます。

## パラメータ推定値

「パラメータ推定値」レポートは、「項」としてリストされている効果ごとに、効用関数に関する係数の推定値と標準誤差が表示されます。製品の属性に対するパラメータは、「部分効用 (part-worth)」と呼ぶこともあります。起動ウィンドウで [Firthバイアス調整推定値] オプションをオンにした場合は、Firth のバイアス調整を伴う最尤推定が行われます。この推定値は、通常の最尤推定値に比べ、より良い性質をもちます。効用関数の詳細については、「[効用と確率](#)」(165 ページ) を参照してください。

### 比較規準

適合度統計量として、AICc（修正された赤池情報量規準）、BIC（ベイズの情報量規準）、 $(-2)^* \text{対数尤度}$ 、 $(-2)^* \text{Firth 対数尤度}$ が表示されます。これらの統計量は、モデルを比較するのに用いられます。詳細と計算式については、『基本的な回帰モデル』の付録「統計的詳細」の「尤度、AICc、BIC」の節を参照してください。

起動ウィンドウで [Firthバイアス調整推定値] オプションをオンにすると、レポートに  $(-2)^* \text{Firth 対数尤度}$  が表示されます。このオプションは、デフォルトではオフになっています。なお、Firthのバイアス調整を伴う最尤推定を行った場合も、行わなかった場合も、「AICc」や「 $(-2)^* \text{対数尤度}$ 」の計算式自体には同じものが使われます。

---

**メモ:** 適合度統計量は、いずれも値が小さいほどよくあてはまっていることを示します。

---

### 尤度比検定

「尤度比検定」レポートは、モデルが5秒以内にあてはめられるならデフォルトで表示されます。表示されない場合は、「選択モデル」の赤い三角ボタンのメニューから「尤度比検定」オプションを選択すると表示されます。レポートには次のものが表示されます。

**要因** モデルに含まれる効果。

**尤度比カイ2乗** 該当する効果に対する検定の尤度比カイ2乗の値。

**自由度** カイ2乗検定の自由度。

**p値(Prob>ChiSq)** カイ2乗検定に対する  $p$  値。

**棒グラフ** カイ2乗検定の値の棒グラフ。

### JMP PRO Bayes パラメータ推定値

（この結果は、起動ウィンドウで [階層型 Bayes] を選択した場合のみ表示されます。）「Bayes パラメータ推定値」レポートには、階層型 Bayes モデルで求められた、モデル効果に関する結果が表示されます。階層型 Bayes モデルでは、被験者による違いを尤度関数に組み込み、製品属性のパラメータにおける被験者の違いを推定します。Bayes 推定のアルゴリズムには、適応型 Metropolis-Hastings 法（Train 2001）の一種が使われています。各モデル効果に対し、事後分布の平均と分散が計算されます。また、各モデル効果に関して、被験者ごとの推定値も出力できます。詳細については、「[被験者推定値の保存](#)」（137 ページ）を参照してください。

この Bayes 流の方法では、被験者ごとの効果が確率変数（変量効果）とみなされます。製品の属性に関するパラメータが、被験者ごとに異なっており、特定の平均ベクトルと共分散行列を持つ多変量正規分布に従うと仮定されます。そして、ある被験者が複数の選択肢から選択することを表す尤度関数には、多項ロジットモデルが仮定されます。（被験者ごとの効果が従う）多変量正規分布の平均ベクトルに対する事前分布には、平均ベクトルが0で、共分散行列が対角要素がすべて等しい対角行列の多変量正規分布が仮定されます。また、（被験者ごとの効果が従う）多変量正規分布の共分散行列が従う事前分布は、対角要素がすべて等しい対角行列を尺度行列とする逆 Wishart 分布が仮定されます。

反復計算では、最初にバーンイン期間が設けられており、最初のほうの反復の結果は破棄されます。デフォルトでは、そのバーンインの反復回数は、起動ウィンドウで指定した「Bayes計算の反復回数」の半分です。

図7.17 Bayesパラメータ推定値レポート

項目	事後平均	事後標準偏差	被験者標準偏差
生地[Thick]	0.12241156	0.2270467796	0.7717269275
チーズ[Jack]	-2.34768104	0.5846517681	0.6746528415
トッピング[Pepperoni]	-0.40059962	0.2524944761	0.6955971534
全反復回数	5000		
バーンイン反復回数	2500		
回答者数	32		
バーンイン後の平均対数尤度	-31.72861		

**項目** モデルの項目。

**事後平均** 該当の項目に対する係数（パラメータ）の推定値。各反復において、各被験者に対する推定値が求められます。「事後平均」は、その推定値を各反復で被験者に関して平均し、バーンイン反復後のそれらの全被験者平均をさらに平均したものです。

**ヒント:** 赤い三角ボタンのメニューから [Bayesチェーンの保存] を選択すると、反復ごとの個々の推定値が確認できます。

**事後標準偏差** 各反復において、被験者ごとの推定値から平均が求められます。バーンイン反復後のそれらの全被験者平均から計算される標準偏差。

**被験者標準偏差** まず、被験者ごとに推定値の平均をバーンイン後のものから計算します。「被験者標準偏差」は、その標準偏差です。

**ヒント:** 赤い三角ボタンのメニューから [被験者推定値の保存] を選択すると、個々の推定値が確認できます。

**全反復回数** バーンイン反復を含め、実行された反復の合計数。

**バーンイン反復回数** バーンイン反復の回数。起動ウィンドウで指定した Bayes 反復回数の半分に相当します。

**回答者数** 被験者の数。

**バーンイン後の平均対数尤度** バーンイン反復後の生成された値から求めた対数尤度関数の平均。

## 「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を表示するには、「選択モデル」の赤い三角ボタンをクリックします。

**メモ：**〔階層型 Bayes〕を指定した場合、被験者レベルの推定値は、モンテカルロ標本を使って計算されます。そのため、以下のオプションの計算結果は、実行するたびに異なります。

**尤度比検定** 詳細については、「尤度比検定」(133 ページ) を参照してください。

**JMP PRO** **最尤推定値の表示** (起動ウィンドウで〔階層型 Bayes〕を選択した場合のみ表示。) モデル項の係数を、通常の最尤法で推定した結果 (パラメータ推定値と標準誤差)。これらの値は、階層型 Bayes アルゴリズムの開始値として使用されます。

**複合因子検定** (起動ウィンドウで〔階層型 Bayes〕を選択した場合は表示されません。) モデル内の各因子を検定する目的で、その因子が関連しているすべての効果の尤度比検定を行います。複合因子検定について詳しくは、『基本的な回帰モデル』の「標準最小2乗のレポートとオプション」章を参照してください。

**信頼区間** 〔階層型 Bayes〕が選択されていないときには、「パラメータ推定値」レポートにパラメータの信頼区間を表示します。

〔階層型 Bayes〕が選択されている場合、信頼区間は「Bayes パラメータ推定値」レポートに表示されます。この Bayes 流の信頼区間は、正規分布を仮定し、事後平均と事後標準偏差に基づいて求められます。

**推定値の相関** 〔階層型 Bayes〕が選択されていないときには、最尤法に基づいて、パラメータ推定値間の相関行列を表示します。

〔階層型 Bayes〕が選択されている場合は、パラメータ推定値の事後平均における相関行列を表示します。この相関行列は、バーンイン後の反復から求めたものです。バーンイン後、各反復ごとに事後平均が計算されます。これらの事後平均から求めた相関行列が、「推定値の相関」表に表示されます。

**全水準の推定値** モデル内の各効果の周辺確率と限界効用を表示します。周辺確率とは、他の属性をすべて平均またはデフォルトの水準に固定したときに、人が属性因子 B ではなく属性因子 A を選択する確率を指します。

図 7.18において、被験者が厚めの生地にモンテレージャックとペパロニをのせたピザではなく、厚めの生地にモッツアレラとペパロニをのせたピザを選ぶ確率は、0.9470 です。

図 7.18 全水準の推定値の例



**効用プロファイル** 異なる因子設定における効用の予測値を示します。効用とは、線形モデルによって予測された値です。「[最適なプロファイルを見つける](#)」(122ページ)に効用プロファイルの例があります。効用関数の詳細については、「[効用と確率](#)」(165ページ)を参照してください。[効用プロファイル]のオプションについては、『[プロファイル](#)』の「[プロファイル](#)」章にある「[予測プロファイルのオプション](#)」の節を参照してください。

**確率プロファイル** 2つの製品が提示されたときに一方の製品を選択する確率を求めることができます。この確率は、 $(\exp(U)) / (\exp(U) + \exp(U_b))$  で計算されます。Uは該当する製品の効用、U<sub>b</sub>は基準となる製品の効用です。なお、この式に基づくと、2つの製品がまったく同じ場合、各製品が選択される確率は50%ずつです。詳細については、「[効用と確率](#)」(165ページ)を参照してください。

確率プロファイルを使用する例は、「[基準との比較](#)」(144ページ)で紹介しています。[確率プロファイル]のオプションについては、『[プロファイル](#)』の「[プロファイル](#)」章にある「[予測プロファイルのオプション](#)」の節を参照してください。

**多選択プロファイル** 指定した数の確率プロファイルが表示されます。複数の製品が提示されたときに、各製品が選択される確率が求められます。各製品のプロファイルは自由に設定できます。多選択プロファイルを使用する例は、「[多選択プロファイルの比較](#)」(146ページ)で紹介しています。[多選択プロファイル]のオプションについては、『[プロファイル](#)』の「[プロファイル](#)」章にある「[予測プロファイルのオプション](#)」の節を参照してください。

**比較** 設定された選択肢の比較を行います。比較したい因子と値を選択することができます。なお、[すべて] チェックボックスをチェックした場合、すべての水準が比較されます。被験者効果がある場合は、製品の比較を行いたい、被験者の水準を指定できます。[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、それらすべての因子の水準での比較ではなく、該当する1因子のすべての水準が、その他の因子を左側で選択されている水準に固定した上で、比較されます。

図7.19 「効用の比較」 ウィンドウ



**支払意思額** このコマンドは、データテーブルに連続尺度の価格の列がある場合だけ使えます。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を与えるのに、価格をどれくらい変更できるかを計算します。結果は、「基準」に指定された値を基準にして、背景因子の各水準ごとに計算されます。

**効用計算式の保存** 効用の計算式列を含んだ、新しいデータテーブルを作成します。作成されたデータテーブルでは、各行がプロファイルと被験者の組み合わせに対応しており、プロファイルと被験者を示す列があります。

**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。詳細については、「[勾配](#)」（166ページ）を参照してください。例として、「[セグメント化の例](#)」（148ページ）を参照してください。

**JMP PRO 被験者推定値の保存** （起動ウィンドウで「階層型 Bayes」を選択した場合のみ表示。）各効果における被験者ごとのパラメータ推定値を含んだデータテーブルを新たに作成します。この被験者ごとのパラメータ推定値を平均した値は、「Bayes パラメータ推定値」レポートにある「事後平均」と等しいです。「被験者受容率」は、Metropolis-Hastings法のステップで求められた新しいパラメータ推定値の受容率を示します。一般に、0.20 ぐらいが良好な受容率とみなされています。詳細については、「[Bayes パラメータ推定値](#)」（133ページ）を参照してください。

**JMP PRO Bayes チェーンの保存** （起動ウィンドウで「階層型 Bayes」を選択した場合のみ表示。）Bayes 推定値を計算するために使用されたチェーンの情報を含んだデータテーブルを、新たに作成します。詳細については、「[Bayes チェーンの保存](#)」（139ページ）を参照してください。

**モデルダイアログ** 「選択モデル」起動ウィンドウが開き、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。

以下のオプションについて詳しくは、『JMPの使用法』の「[JMPのレポート](#)」章を参照してください。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、「自動再計算」オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウで By 変数を指定した場合のみ使用可能です。

## 支払意思額

「支払意思額」（willingness to pay）とは、製品の新しい機能に対して顧客が支払うであろう価格を指します。「支払意思額」は、基準機能の効用と等しくなる、新機能の価格です。たとえば、40GBのハードディスクを搭載したパソコンが、1000 ドルだったとします。80GBのハードディスクに対する支払意思額は、ハードディスクを80GBにした場合、1000 ドルである40GBのパソコンと同じ効用をもたらすのに、どれぐらいの価格にできるかを示します。

## 「支払意思額」起動ウィンドウのオプション

「支払意思額」オプションを選択すると、「支払意思額」起動ウィンドウが開きます。図7.20は、「Laptop Profile.jmp」データテーブルにある「選択モデル」スクリプトを実行した後、赤い三角ボタンメニューから「支払意思額」オプションを選択したときに表示されるウィンドウです。

**因子** 分析に含まれる変数。製品の機能や被験者の属性などです。

**基準** 各因子の基準設定。カテゴリカル因子の場合は、リストから基準値を選択します。数値因子の場合は、基準値を入力します。

**役割** 因子の種類。次の中から選択できます。

**機能因子** 支払意思額を求める製品やサービスの機能。

**価格因子** 選択実験で提示された価格の因子。価格因子は連続尺度でなければならず、1回の「支払意思額」分析において1つの価格因子しか指定できません。

**背景因子 定数** 「支払意思額」の計算で固定にする因子。通常は、被験者因子を指定します。

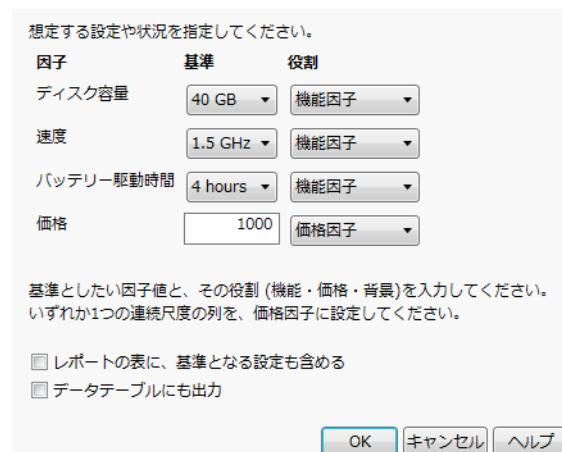
**背景因子 変数** 「支払意思額」の計算で各レベルごとに固定する因子。通常は、被験者レベルの因子を指定します。被験者因子を「背景因子 定数」ではなく「背景因子 変数」に指定すると、その変数の全水準に対して支払意思額が計算されます。

**レポートの表に、基準となる設定も含める** 「支払意思額」レポートに、価格変更を加える前の基準設定が表示されます。

**ヒント:** このチェックボックスをオンにしてデータテーブルに出力した場合、そのデータテーブルには基準となる設定も出力されます。

**データテーブルにも出力** 「支払意思額」レポートを含んだデータテーブルが作成されます。

図7.20 「支払意思額」起動ウィンドウ



1回目の「支払意思額」の計算が完了すると、選択した基準値と役割が記憶されます。そのため、基準情報を一度入力するだけで、「支払意思額」の比較を何度も行うことができます。「価格」という名前の因子はないが、分析に使われている連続尺度の因子が1つしかない場合は、「支払意思額」ウィンドウにて、その因子に自動的に「価格因子」の役割が割り当てられます。実際の金銭的な価格以外にも、旅行時間や距離などの金銭的ではないコストも価格因子と考えられます。

### 「支払意思額」レポート

「支払意思額」レポートには、各因子の基準値と、基準における効用が表示されます。そして、各因子について、機能設定、価格変更（価格がどれだけ変化するか）、新価格がリストされます。交互作用や2次の効果がない場合は、標準誤差と信頼区間も表示されます。標準誤差や信頼区間は、デルタ法で求められます。

図7.21 「支払意思額」レポート

支払意思額						
因子	基準における値					
ディスク容量	40 GB					
速度	1.5 GHz					
バッテリー駆動時間	4 hours					
価格	1000					
基準における効用						
	-3.4736					
因子	機能設定	価格変更	標準誤差	下側95%	上側95%	新価格
ディスク容量	80 GB	\$959.67	381.743	\$211.47	\$1,707.87	\$1,959.67
速度	2.0 GHz	\$496.86	218.515	\$68.58	\$925.14	\$1,496.86
バッテリー駆動時間	6 hours	\$387.88	191.088	\$13.35	\$762.40	\$1,387.88

価格変更の標準偏差はデルタ手法によって算出されました。

### JMP PRO Bayes チェーンの保存

Bayes チェーンのデータテーブルを使うと、生成される推定値の安定性を調べることができます。作成されるデータテーブルには、(起動ウィンドウで指定した)「Bayes 計算の反復回数」に1を足した数の行が含まれます。最初の行(反復1)は、開始値を示します。2行目以降には、生成された乱数が反復の順に保存されます。各列には、次のような情報が output されます。

**反復** 反復の番号。最初の行は、開始値です。

**対数尤度** その反復におけるモデルの対数尤度。各反復に対する対数尤度をプロットすれば、バーンイン期間(調整段階)での振る舞いを確認できます。

**<モデル効果> 適応型シグマ** 逆 Wishart 分布の尺度行列における対角要素の平方根。

**<モデル効果> 受容率** MCMC アルゴリズムでの受容率。

**<モデル効果> 平均** 被験者ごとに計算された推定値、の平均。

**<モデル効果> 分散** 被験者ごとに計算された推定値、の分散。

## その他の例

ここでは、次のような例を紹介します。

- ・「製品の使用を決める分析例」(140ページ)
- ・「セグメント化の例」(148ページ)
- ・「「選択モデル」プラットフォームを使用したロジスティック回帰の例」(152ページ)
- ・「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の例」(155ページ)
- ・「データを2つの分析テーブルに変換する例」(157ページ)
- ・「データを1つの分析テーブルに変換する例」(162ページ)

## 製品の使用を決める分析例

「選択モデル」プラットフォームを利用すると、製品がもつ属性の相対的な重要性を知ることができます。製品について、重要な属性がどれであるかは明らかになっていても、それらの属性におけるトレードオフの関係は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。このようなプロセスで製品を設計することの利点を、例を挙げて解説していきます。

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。

この例は、次のような節に分かれています。

- ・「起動ウインドウへの入力」(140ページ)
- ・「モデルの分析」(143ページ)
- ・「基準との比較」(144ページ)
- ・「多選択プロファイルの比較」(146ページ)

### 起動ウインドウへの入力

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Laptop Runs.jmp.」を開きます。

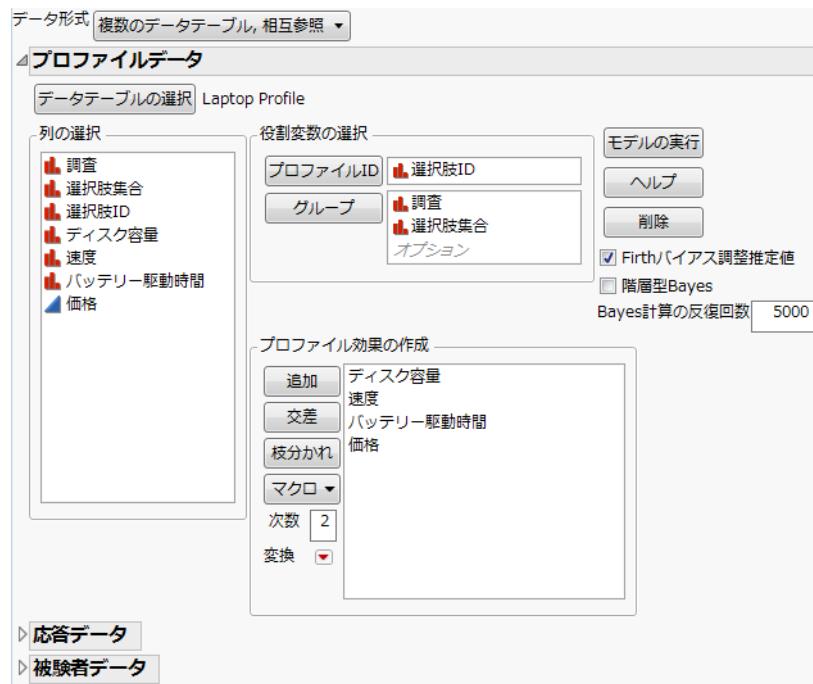
**メモ:** ここで説明している手順をスキップしたいときは、「性別を含めたモデル」スクリプトの隣にある緑色の三角ボタンをクリックしてモデルを実行し、「モデルの分析」(143ページ) に進んでください。

2. 「Open Profile and Subject Tables」スクリプトの隣にある緑色の三角ボタンをクリックします。  
「Laptop Profile.jmp.」と「Laptop Subjects.jmp.」のデータテーブルが開きます。
3. [分析] > [消費者調査] > [選択モデル] を選択します。

メモ: これは、開いている3つのデータテーブルのどれからでも実行できます。

4. 「データ形式」メニューから「複数のデータテーブル、相互参照」を選択します。
5. 「プロファイルデータ」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。「選択肢ID」を選択し、「プロファイルID」をクリックします。
6. 「ディスク容量」、「速度」、「バッテリー駆動時間」、「価格」を選択し、「追加」をクリックします。
7. 「調査」と「選択肢集合」を選択し、「グループ」をクリックします。

図7.22 ノートパソコン調査の「プロファイルデータ」パネル



8. 「応答データ」アウトラインを開きます。
9. 「データテーブルの選択」をクリックし、リストから「Laptop Runs.jmp」を選択します。
10. 「応答データ」パネルに次のように入力します。
  - 「回答」を選択し、「選択されたプロファイルID」をクリックします。
  - 「選択肢1」と「選択肢2」を選択し、「選択肢のプロファイルID」をクリックします。
  - 「調査」と「選択肢集合」を選択し、「グループ」をクリックします。
  - 「回答者」を選択し、「被験者ID」をクリックします。「応答データ」ウィンドウは、図7.23のようになります。

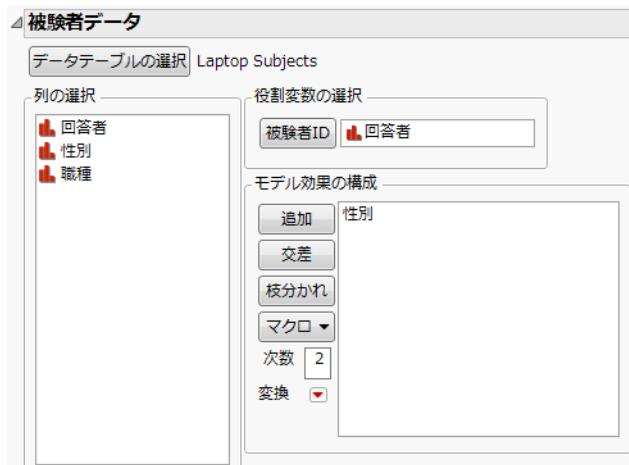
図7.23 ノートパソコン調査の「応答データ」パネル



11. 「被験者データ」アウトラインを開きます。
12. [データテーブルの選択] をクリックし、リストから「Laptop Subjects.jmp」を選択します。
13. 「回答者」を選択し、[被験者ID] をクリックします。
14. 「性別」を選択し、[追加] をクリックします。

「被験者データ」ウィンドウは、図7.24のようになります。

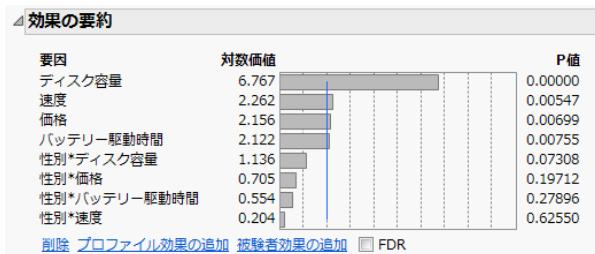
図7.24 ノートパソコン調査の「被験者データ」パネル



## モデルの分析

1. [モデルの実行] をクリックします。

図7.25 ノートパソコン調査の効果の要約



「効果の要約」レポートを見ると、「ディスク容量」が最も有意な効果であることがわかります。ここでは、 $p$ 値が0.15を超える効果を削除して、モデルをより単純なものにします。このとき、効果は1つずつ削除するほうがよいでしょう。この例では、「性別\*速度」の効果が最も有意性が低く、 $p$ 値が0.625となっています。

2. 「効果の要約」レポートで、「性別\*速度」を選択し、[削除] をクリックします。

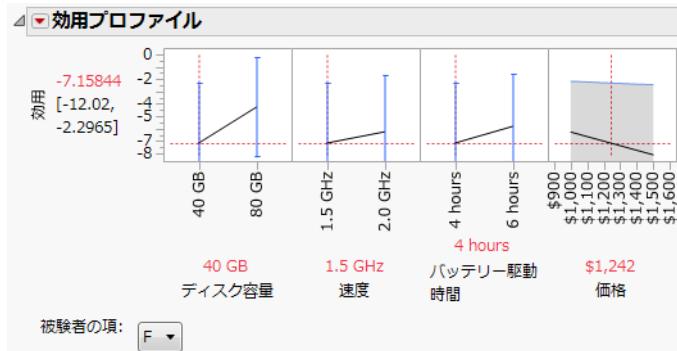
図7.26 ノートパソコン調査の結果



モデルから「性別\*速度」を削除すると、すべての効果の  $p$  値が 0.15 以下になります。そのため、これを最終モデルとして使用してみましょう。

3. 「選択モデル」の赤い三角ボタンをクリックし、[効用プロファイル] を選択します。

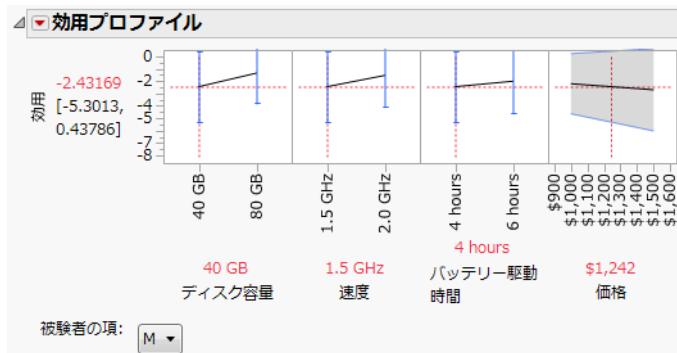
図7.27 ノートパソコン調査の「F」(女性) のプロファイル



ヒント：効用プロファイルが図7.27のようにならなかった場合は、効用プロファイルの赤い三角ボタンをクリックし、[表示方法] > [Y軸を自動調整] を選択します。

4. 「被験者の項」メニューから [M] を選択します。

図7.28 ノートパソコン調査の「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 0.0033 となっており、高度に有意です。詳細については、図7.26 (143ページ) を参照してください。「効用プロファイル」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。

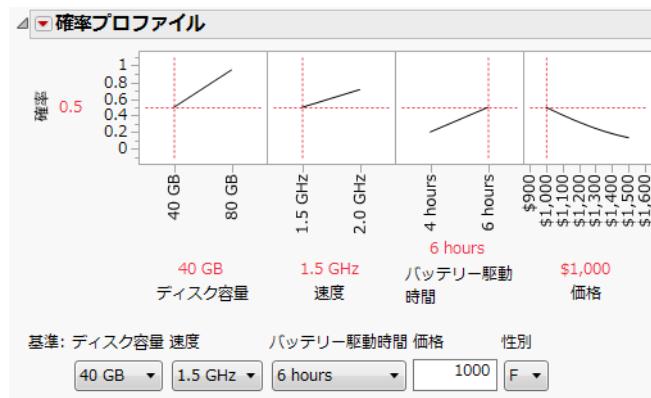
## 基準との比較

新しい製品を開発しているとしましょう。消費者が旧製品や競合製品ではなく新製品を選択するであろう確率を求めましょう。その場合、「確率プロファイル」を使ってプロファイルを基準プロファイルと比較します。

この例では、40 GBのハードディスク、1.5GHzのCPU、6時間のバッテリー駆動時間であるノートパソコンを1,000ドルで現在、販売しているとします。そして、現在の製品を基準として、機能のいくつかを変更して、製品の魅力を高めるとします。ここでの分析の計算では、まったく同じ製品が2つあった場合、各製品が選ばれる確率は50%ずつです。基準となる製品を変更することにより、消費者に選択される確率がどれくらい大きくなるかを調べます。

1. 次のいずれかを実行します。
  - 「起動ウインドウへの入力」(140ページ)の手順に従います。次に、「モデルの分析」(143ページ)のステップ1とステップ2を実行します。
  - 「Laptop Runs.jmp」データテーブルで、「変数を減らしたモデル」スクリプトの緑色の三角ボタンをクリックします。
2. 「選択モデル」の赤い三角ボタンのメニューから【確率プロファイル】を選択します。  
この確率プロファイルでは、「性別」が「F」に設定されています。これは、後で変更することができます。
3. プロファイルの下にある「基準」エリアのメニューとテキストボックスを使用し、基準構成を40 GB、1.5 GHz、6 hours、1000に設定します。
4. これらの値を「確率プロファイル」の値として指定します。価格を\$1000に設定するには、右端のプロファイルの下にある「価格」のすぐ上の「\$1,242」という値をクリックし、「1000」と入力します。次に、テキストボックスの外をクリックします。

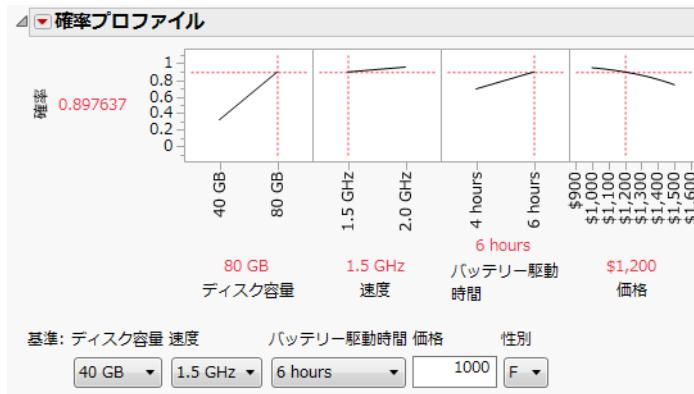
図7.29 確率プロファイルにある「価格」のテキスト入力エリア



この構成の確率は0.5です。

5. 確率プロファイルで、「ディスク容量」のスライダを80 GBまで動かします。  
変更された設定において、たとえ「価格」が増加しても、選択される確率はさほど変化しないことがわかります。
6. プロファイルの「価格」の上にある「\$1,000」というラベルをクリックし、「1200」と入力してテキストボックスの外をクリックします。

図7.30 基準的なノートパソコンに対する確率プロファイル



「ディスク容量」を40GBから80GBへと変更し、「価格」を\$1,200にすると、基準の製品とこの新製品の2つが選択肢にある場合、女性が新製品のほうを選択する確率は0.90です。なお、「性別」を【M】(男性)に変更して、男性におけるこの確率を見ると0.71です。この構成を選択する確率は、0.71です。

## 多選択プロファイルの比較

「多選択プロファイル」を使って製品プロファイルを比較します。

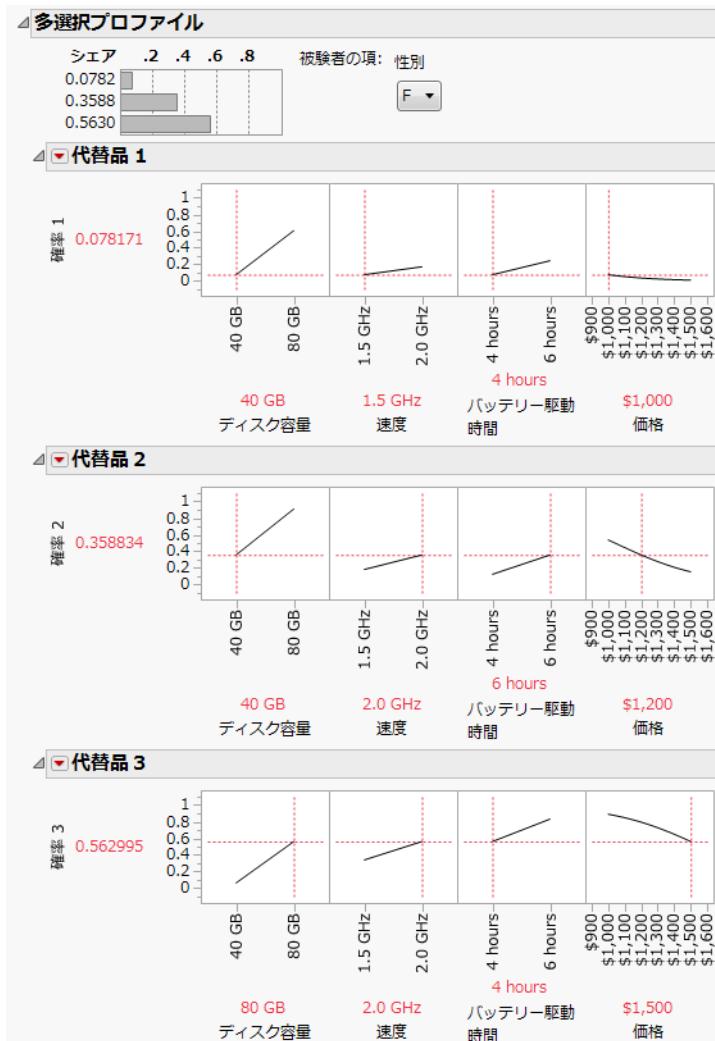
- 現在の自社製品は、ディスク容量が小さく、CPU速度が遅く、バッテリー駆動時間が短いですが、価格が\$1,000と最も安価になっています。
- 会社Aの製品は、CPUが高速で、バッテリー駆動時間が長く、価格は\$1,200と手頃です。
- 会社Bの製品は、ディスク容量が最大で、CPU速度も最速ですが、価格が\$1,500と高く、バッテリー駆動時間が短くなっています。

分析者は、性能や価格のいずれか1つを改善し、市場シェアを高めたいと考えています。

- 次のいずれかを実行します。
  - 「起動ウィンドウへの入力」(140ページ)の手順に従います。次に、「モデルの分析」(143ページ)のステップ1とステップ2を実行します。
  - 「Laptop Runs.jmp」データテーブルで、「変数を減らしたモデル」スクリプトの横にある緑色の三角ボタンをクリックします。
- 「選択モデル」の赤い三角ボタンのメニューから【多選択プロファイル】を選択します。
 ウィンドウが開いたら、プロファイルに含める選択肢の数を入力します。ここでは、デフォルトの「3」をそのまま使います。
- 【OK】をクリックします。
 3つの「代替品」のプロファイルが表示されます。このプロファイルでは、性別が「F」に設定されています。各プロファイルで、すべての因子がデフォルトの値に設定されています。「代替品1」は、自社がこれから開発する製品です。「代替品2」は、会社Aの製品です。「代替品3」は、会社Bの製品です。

4. 「代替品 1」の「ディスク容量」を40 GB、「速度」を1.5 GHz、「バッテリー駆動時間」を4 hours、「価格」を\$1,000に設定します。
5. 「代替品 2」の「ディスク容量」を40 GB、「速度」を2.0 GHz、「バッテリー駆動時間」を6 hours、「価格」を\$1,200に設定します。
6. 「代替品 3」の「ディスク容量」を80 GB、「速度」を2.0 GHz、「バッテリー駆動時間」を4 hours、「価格」を\$1,500に設定します。

図7.31 「F」(女性)の多選択プロファイル

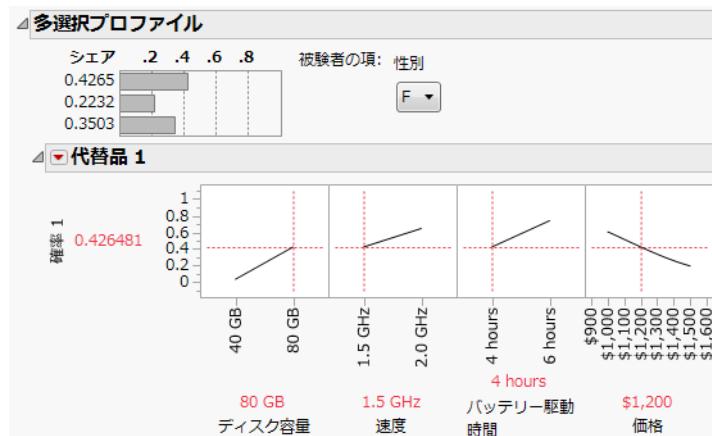


会社Bのシェアが最大で、0.5630です。今のままでは、自社製品を買ってくれる女性が非常に少ないとことがわかります。

価格を少し高めに変更しながらも、性能のいずれか1つを改善することにより、自社のシェアを増やすことを考えてみましょう。「代替品1」を見ると、「ディスク容量」の直線の傾きが大きいので、ディスク容量を増やすべば最も効果的にシェアを高められるようです。

7. そこで、「代替品1」の「ディスク容量」を80 GB、「価格」を\$1,200にしてみましょう。

図7.32 改善したノートパソコンの多選択プロファイル



ディスク容量を増やしたことで、ノートパソコンの価格を上げても女性客においておよそ43%のシェアが見込めます。この数字は、会社Bの高性能ノートパソコンをも上回り、図7.31にある現在の設定よりずっと高シェアです。

なお、女性ではなく男性におけるシェアの変化を調べてみると、「ディスク容量」と「速度」の両方を高めた場合に、44%のシェアが見込めます。

## セグメント化の例

この例では、ピザの選択実験データをもとに、消費者のセグメント化を行います。

起動ウィンドウの入力方法については、「複数のデータテーブル」(118ページ)のステップ1～ステップ15を参照してください。または、以下の指示に従います。

### クラスターの定義

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Pizza Responses.jmp」を開きます。
2. 「選択モデル」スクリプトの緑色の三角ボタンをクリックします。
3. 「選択モデル」の赤い三角ボタンをクリックし、[被験者ごとの勾配を保存] を選択します。

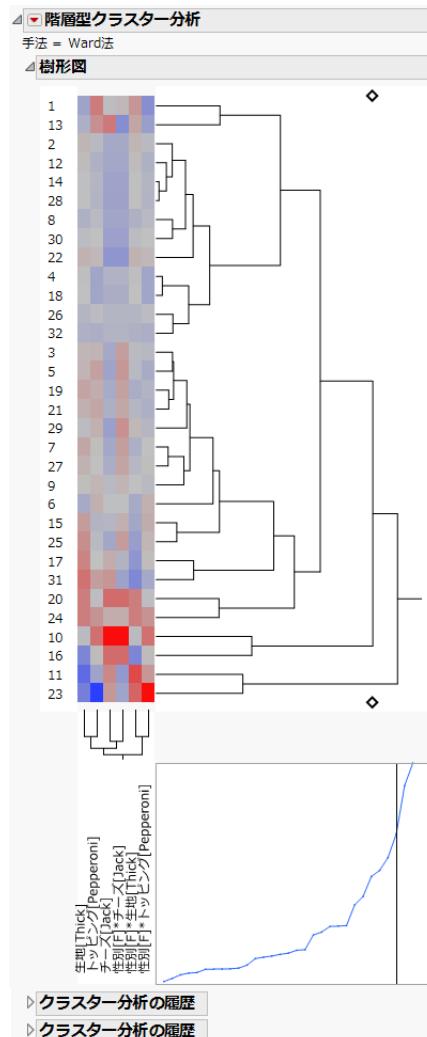
データテーブルが開き、主効果と被験者の各交互作用の勾配が保存されます。

図7.33 ピザのデータから計算された被験者ごとの勾配（一部）

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

4. 「階層型クラスター分析」スクリプトの緑色の三角ボタンをクリックします。

図7.34 ピザのデータで作成された被験者クラスターの樹形図



このスクリプトは、勾配のテーブルにある「被験者」列を除くすべての列を対象に、階層型クラスター分析を行います。いずれかのひし形をクリックすると、行が3つのクラスターに分かれていることがわかります。

5. 「階層型クラスター分析」の赤い三角ボタンをクリックし、[クラスターの保存] を選択します。

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「階層型クラスター分析」章を参照してください。

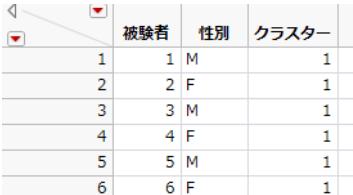
勾配の列は、クラスター分けのために作成したものなので、削除してかまいません。

6. 「被験者」と「クラスター」を除くすべての列を選択します。選択した列を右クリックし、[列の削除] を選択します。

7. 「元データに結果をマージ」スクリプトの横にある緑色の三角ボタンをクリックします（図7.33）。

クラスター情報が「被験者」データテーブルにマージされます。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7.35）。

図7.35 「被験者」データテーブルに加わった「クラスター」列



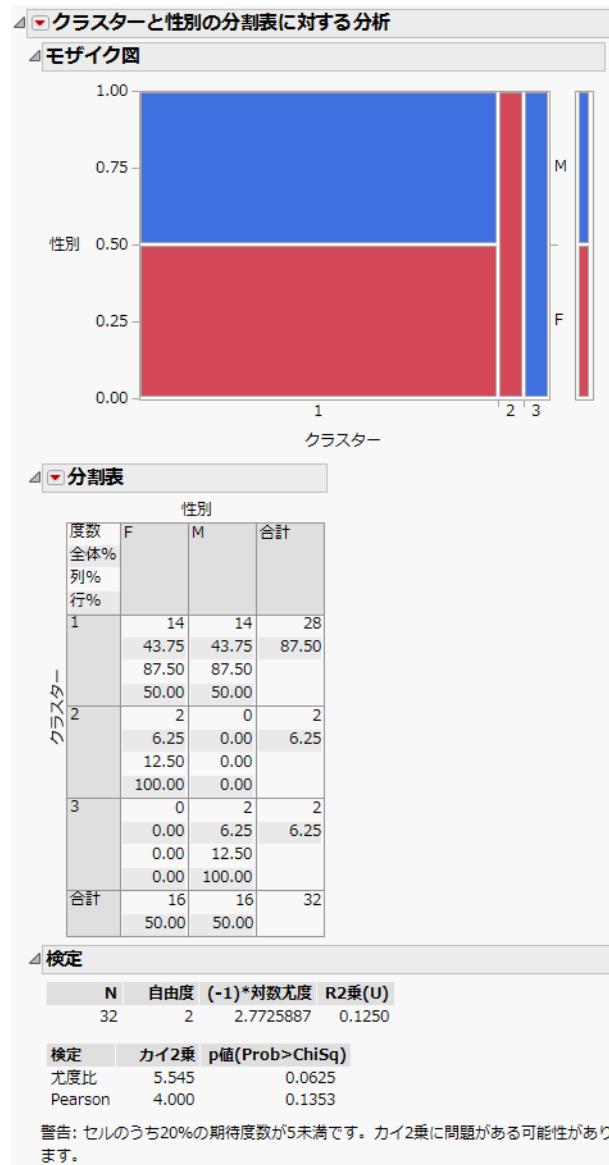
	被験者	性別	クラスター
1	1	M	1
2	2	F	1
3	3	M	1
4	4	F	1
5	5	M	1
6	6	F	1

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。

### クラスターの検討

1. 「列」パネルで、「クラスター」変数の左にあるアイコンをクリックし、[順序尺度] を選択します。
2. [分析] > [二変量の関係] を選択します。
3. 「性別」を選択し、[Y, 目的変数] をクリックします。
4. 「クラスター」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。

図7.36 クラスターと性別の分割表に対する分析



次のようなことがわかります。

- クラスター 1 は男性と女性の間で均等に分かれています。
- クラスター 2 には女性しか含まれていません。
- クラスター 3 には男性しか含まれていません。

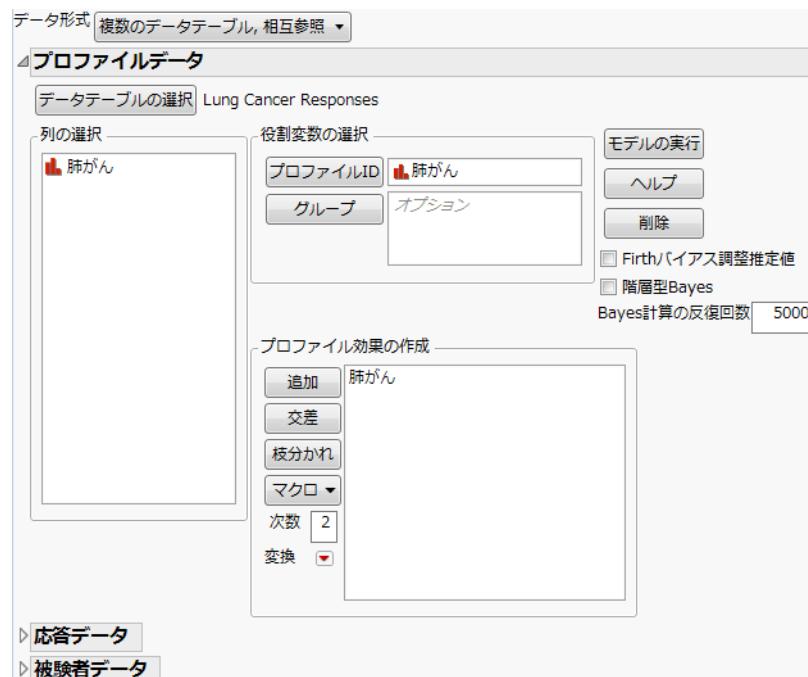
場合によっては、「クラスター」変数を含めた選択モデルをあてはめてみてもよいでしょう。

## 「選択モデル」プラットフォームを使用したロジスティック回帰の例

### 「選択モデル」プラットフォームを使用する

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「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。
2. [分析] > [消費者調査] > [選択モデル] を選択します。
3. 「データ形式」メニューから [複数のデータテーブル, 相互参照] を選択します。
4. [データテーブルの選択] をクリックし、「Lung Cancer Responses.jmp」を選択して [OK] をクリックします。
5. 「肺がん」を選択し、[プロファイルID] をクリックします。
6. 「肺がん」を選択し、[追加] をクリックします。
7. [Firthバイアス調整推定値] のチェックマークを外します。

図7.37 入力後の「プロファイルデータ」パネル

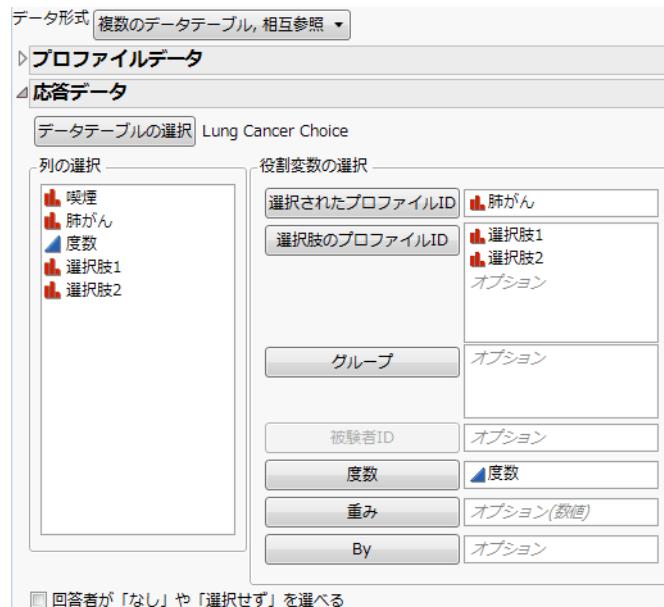


8. 「応答データ」アウトラインを開きます。
9. [データテーブルの選択] をクリックし、「Lung Cancer Choice.jmp」を選択して [OK] をクリックします。

10. 次の手順を行います。

- 「肺がん」を選択し、[選択されたプロファイルID] をクリックします。
- 「選択肢1」と「選択肢2」を選択し、[選択肢のプロファイルID] をクリックします。
- 「度数」を選択し、[度数] をクリックします。

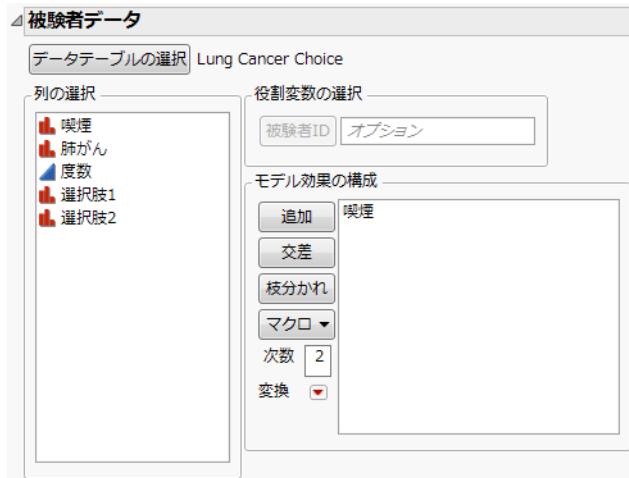
図7.38 入力後の「応答データ」パネル



11. 「被験者データ」アウトラインを開きます。

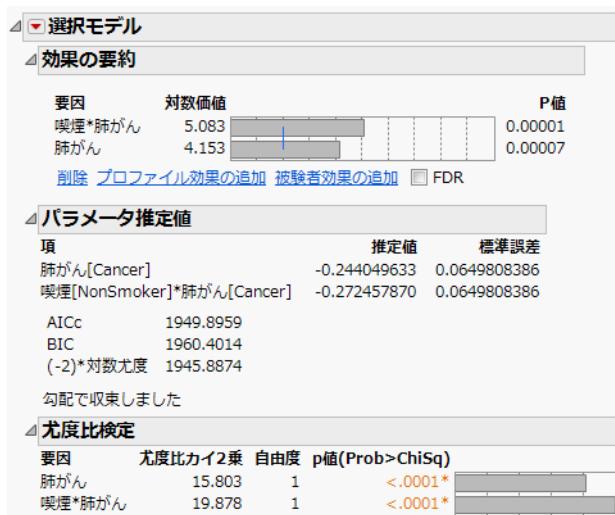
12. [データテーブルの選択] をクリックし、「Lung Cancer Choice.jmp」を選択して [OK] をクリックします。
13. 「喫煙」を選択し、[追加] をクリックします。

図7.39 入力後の「被験者データ」パネル



14. [モデルの実行] をクリックします。

図7.40 「選択モデル」を使ったロジスティック回帰



### 「モデルのあてはめ」プラットフォームを使用する

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「Lung Cancer.jmp」を開きます。
- [分析] > [モデルのあてはめ] を選択します。

データテーブルに「モデル」スクリプトが含まれているため、「モデルの指定」ウィンドウへの入力は自動的に行われます。[名義ロジスティック] の手法が指定されています。

3. [実行] をクリックします。

図7.41 「モデルのあてはめ」を使ったロジスティック回帰

図7.41は「モデルのあてはめ」を使ったロジスティック回帰の結果を示す。以下に主要な出力結果を要約する。

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

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

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7.40と図7.41のレポートから、喫煙が肺がんの発病に対して高度に有意であることが分かります。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」章を参照してください。

## 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の例

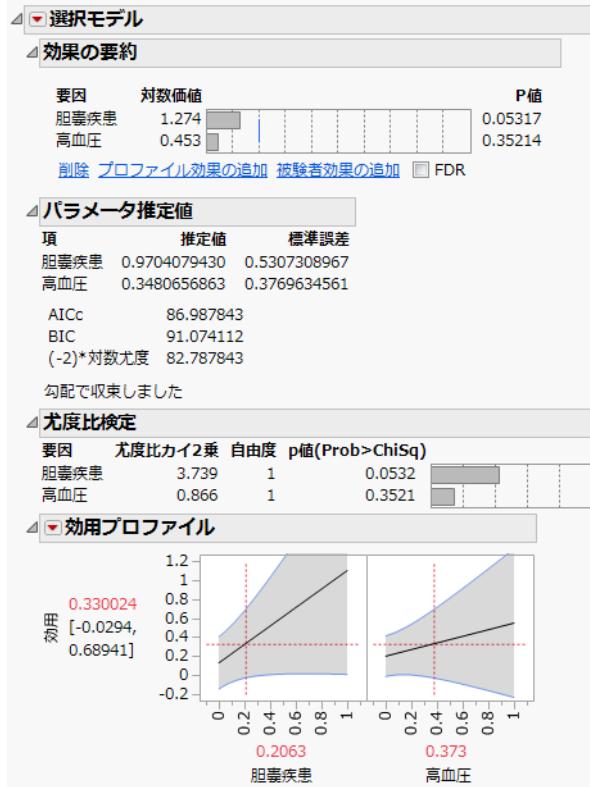
ここでは、「選択モデル」プラットフォームを使い、「ケース（症例）」と「コントロール（対照）」とでペアごとに対応している子宮内膜がんの調査結果に対して、条件付きロジスティック回帰分析を行う例を紹介します。使用するデータは、Breslow and Day (1980) およびSAS/STAT(R) 9.2 User's Guide, Second Edition (2006) で取り上げられているLos Angeles Study of the Endometrial Cancer Dataからの引用です。このケースコントロール分析の目標は、高血圧による影響を考慮しながら、胆囊疾患の相対リスクを推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

「選択モデル」プラットフォームによって、通常のロジスティック回帰を行う手順については、「ロジスティック回帰」(165ページ) を参照してください。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「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. 「データ形式」が [1つのデータテーブル, 積み重ね] になっていることを確認します。
4. [データテーブルの選択] ボタンをクリックします。
5. プロファイルデータテーブルとして「Endometrial Cancer」を選択します。[OK] をクリックします。
6. 「アウトカム」を選択し、[応答の指示変数] をクリックします。
7. 「ペア」を選択し、[グループ] をクリックします。
8. 「胆嚢疾患」と「高血圧」を選択し、「プロファイル効果の作成」パネルで [追加] をクリックします。
9. [Firthバイアス調整推定値] のチェックマークを外します。
10. [モデルの実行] をクリックします。
11. 「選択モデル」の赤い三角ボタンをクリックし、[効用プロファイル] を選択します。

図7.42のようなレポートが作成されます。

図7.42 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。「胆囊疾患」が、 $\alpha$ 水準を0.05としたとき、統計的にはほぼ有意であることがわかります（ $p$ 値は0.0532）。また、効用プロファイルによって、応答に対する要因の影響を視覚的に確認できます。

## データを2つの分析テーブルに変換する例

「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。

### 「選択した交通手段」と「被験者」の追加

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Daganzo Trip.jmp」データテーブルを開きます。

図7.43はその一部です。

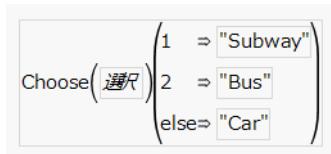
図7.43 3つの交通手段を使ったDaganzoの移動時間（一部）

	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、以下の手順で簡単に行えます。

2. [列] > [列の新規作成] を選択します。
3. 「列名」に「選択した交通手段」と入力し、「尺度」を [名義尺度] に設定します。
4. [列プロパティ] をクリックし、[計算式] を選択します。
5. [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。
6. 「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7.44）。

図7.44 Daganzoのデータの「選択した交通手段」列に Choose 関数を適用



```
Choose(選択, (1 => "Subway"), (2 => "Bus"), (else => "Car"))
```

7. 「計算式エディタ」ウィンドウで [OK] をクリックします。

8. 「列の新規作成」ウィンドウで [OK] をクリックします。

データテーブルに新しい「選択した交通手段」列が表示されます。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。

9. [列] > [列の新規作成] を選択します。

10. 「列名」に「被験者」と入力します。

11. 「データの初期化」のドロップダウンメニューを開き、[シケンステータ] を選択します。

12. [OK] をクリックします。

変更後のテーブルは、図7.45のようになります。

図7.45 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

### データの積み重ね

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。

1. [テーブル] > [列の積み重ね] を選択します。

2. 「Subway」、「Bus」、「Car」を選択し、[積み重ねる列] をクリックします。

3. 「出力テーブル名」に「Stacked Daganzo」と入力します。「積み重ねたデータ列」に「移動時間」、「元の列のラベル」に「交通手段」と入力します。

設定した後の「積み重ね」ウィンドウは、図7.46のようになります。

図7.46 Daganzoのデータで行う積み重ね



4. [OK] をクリックします。

最終的なデータテーブルは、図7.47のようになります。

図7.47 「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

#### プロファイルデータテーブルの作成

プロファイルデータテーブルには、「被験者」、「交通手段」、「移動時間」の各列が必要です。

- 「被験者」、「交通手段」、「移動時間」の各列を選択し、[テーブル] > [サブセット] を選択します。
- 【すべての行】と【選択されている列】を選択し、[OK] をクリックします。

図7.48は、データの一部です。デフォルトのテーブル名は、「Stacked Daganzoのサブセット」です。

図7.48 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	Subway	16.481
2	Bus	16.196
3	Car	23.89
4	Subway	15.123
5	Bus	11.373
6	Car	14.182
7	Subway	19.469

### 応答データテーブルの作成

応答データテーブルには、「被験者」、「選択した交通手段」の各列と、各選択肢のための列が必要です。

3. 「Daganzo Trip.jmp」を開いたまま、「被験者」と「選択した交通手段」の各列を選択します。

4. [テーブル] > [サブセット] を選択します。

5. [すべての行] と [選択されている列] を選択し、[OK] をクリックします。

デフォルトのテーブル名は、「Daganzo Trip のサブセット」です。

6. [列] > [列の新規作成] を選択します。

7. 「列名」に「選択肢」と入力します。

8. [文字] と [名義尺度] を選択します。

9. 「追加する列数」に「3」と入力します。

10. [OK] をクリックします。

「選択肢 1」、「選択肢 2」、「選択肢 3」という列が追加されています。

11. 「選択肢 1」の最初の行に「Bus」と入力します。セルを右クリックし、[初期値を埋める/挿入] > [テーブルの末尾まで入力] を選択します。

12. 「選択肢 2」の最初の行に「Subway」と入力します。セルを右クリックし、[初期値を埋める/挿入] > [テーブルの末尾まで入力] を選択します。

13. 「選択肢 3」の最初の行に「Car」と入力します。セルを右クリックし、[初期値を埋める/挿入] > [テーブルの末尾まで入力] を選択します。

図7.49のようなテーブルになります。

図7.49 Daganzoのデータのサブセットに選択肢集合を追加

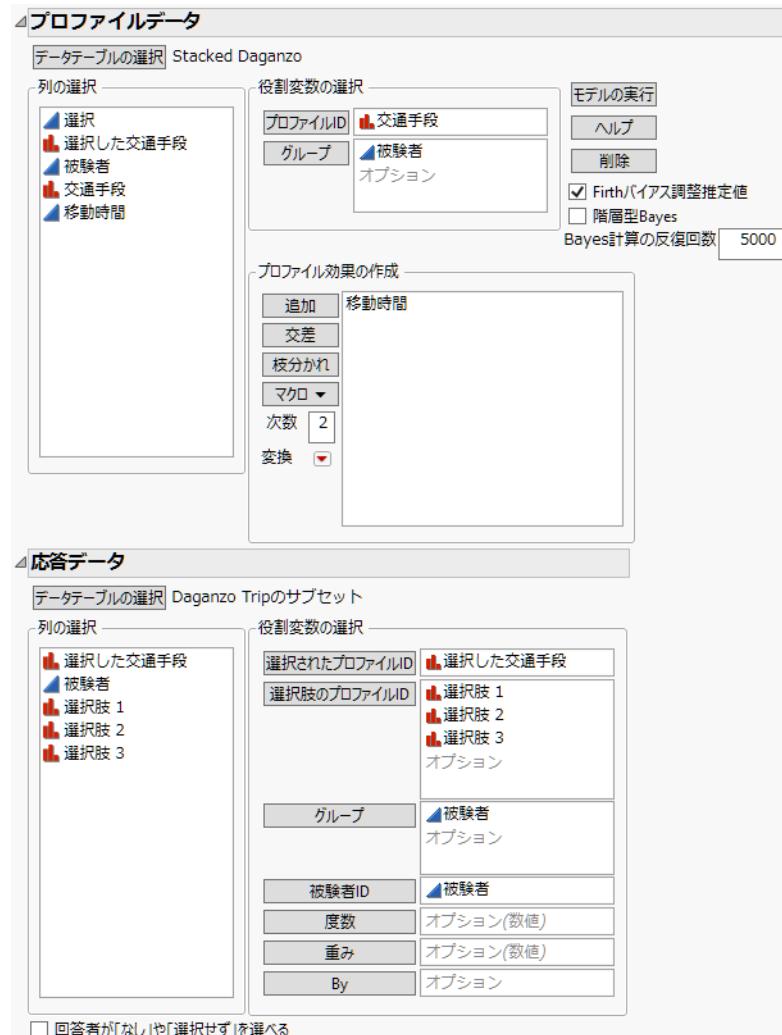
選択した交通手段	被験者	選択肢 1	選択肢 2	選択肢 3
Bus	1	Bus	Subway	Car
Bus	2	Bus	Subway	Car
Bus	3	Bus	Subway	Car
Bus	4	Bus	Subway	Car
Bus	5	Bus	Subway	Car
Subway	6	Bus	Subway	Car
Subway	7	Bus	Subway	Car

## モデルをあてはめる

元の「Daganzo Trip.jmp」テーブルを2つのテーブルに分けたところで、「選択モデル」プラットフォームを実行します。

1. [分析] > [消費者調査] > [選択モデル] を選択します。
2. 「データ形式」メニューから [複数のデータテーブル, 相互参照] を選択します。
3. 図7.50のとおりにモデルを指定します。

図7.50 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



4. [モデルの実行] をクリックします。

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7.51）。

図7.51 Daganzoのデータの「移動時間」に対するパラメータ推定値



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例

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つのテーブルに変換する場合は、以下の手順で操作します。

1. 「データを2つの分析テーブルに変換する例」(157ページ)の「データの積み重ね」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. [列] > [列の新規作成] を選択します。
  3. 「列名」に「応答」と入力します。
  4. [列プロパティ] をクリックし、[計算式] を選択します。
  5. 関数リストから [条件付き] を選択し、[If] を選択します。
  6. 「選択した交通手段」の列を選択し、計算式（「式」）に挿入します。
  7. 「=」を入力して「交通手段」を選択します。
  8. [then節] に「1」を入力し、[else節] に「0」を入力します。
  9. 「計算式エディタ」ウィンドウで [OK] をクリックします。「列の新規作成」ウィンドウで [OK] をクリックします。
- 入力が完了した計算式は、図7.52のようになります。

図7.52 指示変数を作成する計算式

$$\begin{aligned} \text{If} & \left( \begin{array}{l} \text{選択した交通手段} == \text{交通手段} \\ \Rightarrow 1 \end{array} \right) \\ \text{else} & \Rightarrow 0 \end{aligned}$$

10. 「被験者」、「移動時間」、「応答」の各列を選択し、[テーブル] > [サブセット] を選択します。

11. [すべての行] と [選択されている列] を選択し、[OK] をクリックします。

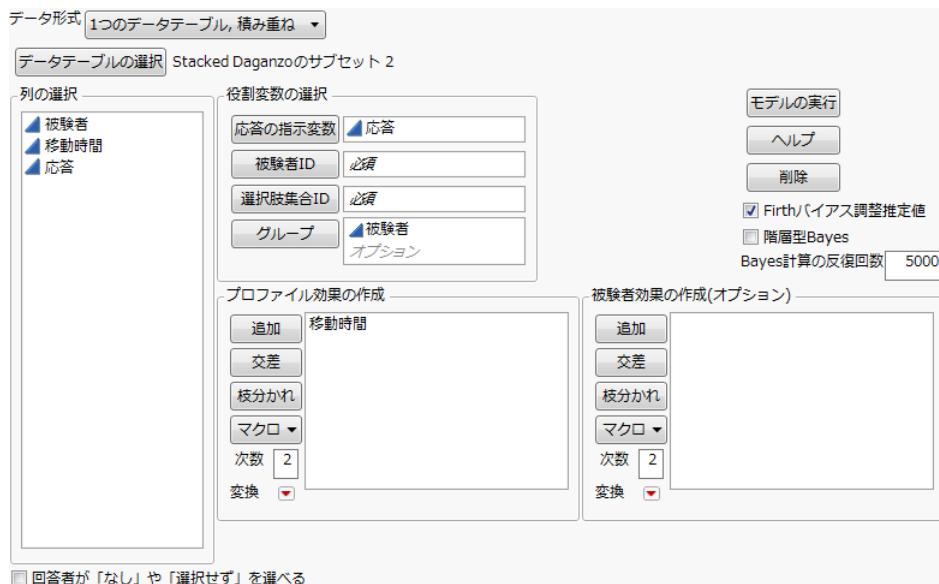
新しいデータテーブルは、図7.53のようになります。

図7.53 「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

12. [分析] > [消費者調査] > [選択モデル] の順にクリックして起動ウィンドウを開き、図7.54のようにモデルを指定します。

図7.54 「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



13. [モデルの実行] をクリックします。

図7.55 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



結果は、テーブルを2つ使って分析した場合（図7.51）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。

## 技術的詳細

この節では、次のようなトピックについて解説します。

- 「分析対象のデータに対する特殊な規則」（164ページ）
- 「効用と確率」（165ページ）
- 「勾配」（166ページ）

## 分析対象のデータに対する特殊な規則

### デフォルトの選択肢集合

プロファイルデータのプロファイルが、すべての質問で共通の選択肢になっている場合には、「選択モデル」起動ウィンドウの「応答データ」パネルにある「役割変数の選択」での【選択肢のプロファイルID】を指定しなくてもかまいません。省略した場合は、プロファイルデータに含まれているプロファイルが、すべての質問において共通して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】列は指定しなくともかまいません。たとえ被験者IDを指定しても、分析の計算には使われません。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。

**メモ:** 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施する例については、「[「選択モデル」プラットフォームを使用したロジスティック回帰の例](#)」(152ページ) および「[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の例](#)」(155ページ) を参照してください。

### 効用と確率

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、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。

選択モデルは、次のような式で表わされます。

$X[k]$  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、

- ⊗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。

## 勾配

〔被験者ごとの勾配を保存〕オプションで保存される勾配は、Newton-Raphson法のステップ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掛けたもの (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、 $H^{-1}$ 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回偏導したものです）を要素とする行列のことです。

ここで、 $\Delta$ を分解すると次のように表わされます。

$$\Delta = \sum_{ij} \Delta_{ij} = \sum H^{-1} g_{ij} = 0$$

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、 $\Delta_{ij}$ 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。

$$\bar{\Delta}_i = \sum_j \frac{\Delta_{ij}}{n_i}$$

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。 $\bar{\Delta}_i$ はそのような影響力を示します。この値は、データテーブルに保存することができますので、保存したあとに「クラスター分析」プラットフォームなどで分析できます。

# 第8章

## MaxDiff

### MaxDiff モデルのあてはめ

MaxDiff法（maximum difference scaling）は、消費者の選択データをもとに、製品の各属性がもつ相対的な重要度を求める方法のひとつです。MaxDiff法の実験では、回答者は最も好きな選択肢（最高）と、最も好きではない選択肢（最悪）を選びます。最高と最悪を選ぶので、最高だけを選ぶ実験よりも、順位付けに関する情報がデータに含まれています。

「MaxDiff」プラットフォームでは、次のようなことが実行できます。

- ・ 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する。
- ・ 分析対象のデータとしていくつかの形式がサポートされている。
- ・ 被験者（消費者）ごとにスコアを求め、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。
- ・ **JMP<sup>®</sup> PRO** Bayes（ベイズ）流の方法で、被験者に関する係数の推定値を求める。
- ・ バイアス修正を伴う最尤推定（Firth 1993）を行う。

図8.1 MaxDiff法による「全水準の比較レポート」

△全水準の比較レポート										
差(行 - 列)	Gyro	Sour Cream and Onion	Truffle Fries	Biscuits and Gravy	Reuben	Ketchup	All Dressed	Dill Pickle	Barbecue	Southern Barbecue
Gyro	0	-1.1747	-1.3082	-1.2688	-0.6992	-0.8229	-0.9467	-1.027	-2.4417	-1.9535
	0	0.33327	0.35477	0.34456	0.34758	0.32802	0.33571	0.35251	0.41499	0.38939
		0.00062	0.00035	0.00036	0.04668	0.01356	0.00569	0.00432	4.32e-8	2.01e-6
Sour Cream and Onion	1.17469	0	-0.1335	-0.0941	0.47548	0.35175	0.22804	0.14766	-1.267	-0.7788
	0.33327	0	0.31379	0.30716	0.30932	0.31915	0.28158	0.30707	0.37654	0.34896
		0.00062	0.67137	0.75983	0.12709	0.27279	0.41975	0.63156	0.00105	0.02764
Truffle Fries	1.30817	0.13348	0	0.03935	0.60896	0.48523	0.36152	0.28114	-1.1335	-0.6453
	0.35477	0.31379	0	0.31345	0.31817	0.31157	0.31136	0.30871	0.39213	0.34904
		0.00035	0.67137	0.75983	0.90032	0.05821	0.12223	0.24809	0.36444	0.00462
Biscuits and Gravy	1.26882	0.09413	-0.0393	0	0.56961	0.44588	0.32217	0.24179	-1.1729	-0.6847
	0.34456	0.30716	0.31345	0	0.30639	0.29752	0.29974	0.28249	0.36267	0.34038
		0.00036	0.75983	0.90032		0.06567	0.1368	0.28477	0.39388	0.00161
Reuben	0.69921	-0.47555	-0.609	-0.5696	0	-0.1237	-0.2474	-0.3278	-1.7425	-1.2543
	0.34758	0.30932	0.31817	0.30639	0	0.30893	0.30148	0.30328	0.37117	0.35324
		0.04668	0.12709	0.05821	0.06567		0.68956	0.41355	0.28208	7.69e-6
Ketchup	0.82294	-0.3517	-0.4852	-0.4459	0.12373	0	-0.1237	-0.2041	-1.6188	-1.1305
	0.32802	0.31915	0.31157	0.29752	0.30893	0	0.29728	0.27938	0.36911	0.34808
		0.01356	0.27279	0.12223	0.1368	0.68956		0.67812	0.46661	2.64e-5
All Dressed	0.94665	-0.228	-0.3615	-0.3222	0.24744	0.12371	0	-0.0804	-1.4951	-1.0068
	0.33571	0.28158	0.31136	0.29974	0.30148	0.29728	0	0.29447	0.36838	0.3418
		0.00569	0.41975	0.24809	0.28477	0.41355	0.67812		0.78538	9.22e-5
Dill Pickle	1.02703	-0.1477	-0.2811	-0.2418	0.32782	0.20409	0.08038	0	-1.4147	-0.9264
	0.35251	0.30707	0.30871	0.28249	0.30328	0.27938	0.29447	0	0.3539	0.35532
		0.00432	0.63156	0.36444	0.39388	0.28208	0.46661	0.78538		0.00012
Barbecue	2.44172	1.26703	1.13355	1.1729	1.7425	1.61877	1.49507	1.41468	0	0.48824
	0.41499	0.37654	0.39213	0.36267	0.37117	0.36911	0.36838	0.3539	0	0.39164
		4.32e-8	0.00105	0.00462	0.00161	7.69e-6	2.64e-5	9.22e-5	0.00012	0.21514
Southern Barbecue	1.95348	0.77879	0.6453	0.68465	1.25426	1.13053	1.00683	0.92644	-0.4882	0
	0.38939	0.34896	0.34904	0.34038	0.35324	0.34808	0.3418	0.35532	0.39164	0
		2.01e-6	0.02764	0.06715	0.0467	0.00056	0.00154	0.00393	0.01038	0.21514

## 「MaxDiff モデル」プラットフォームの概要

MaxDiff法は、選択実験データに対する統計手法です。MaxDiff法は、「最良-最悪スケーリング」Best-Worst Scaling (BWS)とも呼ばれています。MaxDiff法の実験では、複数の選択肢から最良のものと最悪のものを回答者に選んでもらいます。MaxDiff法では、好きなものだけを選んでもらった実験に比べ、選好についてより詳しい情報が得られます。MaxDiff法については、Louviere et al. (2015) を参照してください。また、選択モデルについては、Louviere et al. (2015)、Train (2009)、Rossi et al. (2006) を参照してください。

MaxDiffモデルは、「各自の効用に基づいて確率的に人々は選択している」と仮定しています。一つひとつの選択肢に、その人にとっての価値（効用）があると仮定しています。MaxDiff法は、その効用を推定します。推定された効用に基づいて、ある選択肢が他の選択肢より好まれる確率も推定できます。これらの推定には、条件付きロジスティック回帰が使用されます。McFadden (1974) を参照してください。

**メモ:** 1因子のMaxDiff調査は、「MaxDiff計画」プラットフォームで計画することができます。『実験計画(DOE)』の「MaxDiff計画」章を参照してください。

### セグメント化や被験者効果の Bayes 推定

マーケットリサーチ（市場調査）では、選好のパターンによって、消費者をグループに分けたい場合があります。しかし、通常のモデルで各被験者のパラメータ推定値を求めるには、被験者1人あたりに多くのデータが必要です。被験者1人あたりに多くのデータがあるなら、被験者IDを「応答データ」セクションで [By] 変数に指定するか、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、後者の方法は、被験者数が多いと計算が難しくなります。これに代わる統計手法として、統計学の文献では、Bayesモデルや混合モデルなどが提案されています。

[By] 変数が指定できるほどのデータがない場合は、応答データと [被験者ごとの勾配を保存] オプションを使って被験者をクラスターに分け、セグメント化することができます。このオプションは、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、新しいデータテーブルに表示します。例として、「選択モデル」章の「セグメント化の例」(148ページ) を参照してください。傾きの値についての詳細は、「選択モデル」章の「勾配」(166ページ) を参照してください。

**JMP PRO** MaxDiff法では、階層型 Bayesモデルを使って被験者効果を推定することもできます。この Bayes 法による統計分析は、マーケットのセグメント化に役立ちます。

## 「MaxDiff」プラットフォームの例

30人の回答者を対象に、MaxDiff法を使い、7種類の風味のポテトチップスを比較する調査を行いました。風味が異なるポテトチップスを3つずつ各回答者に提示しました。それら3つの選択肢のなかで回答者が最も好きなものを「1」、最も好きでないものを「-1」として記録しました。それ以外の選択肢は、「0」として記録しました。

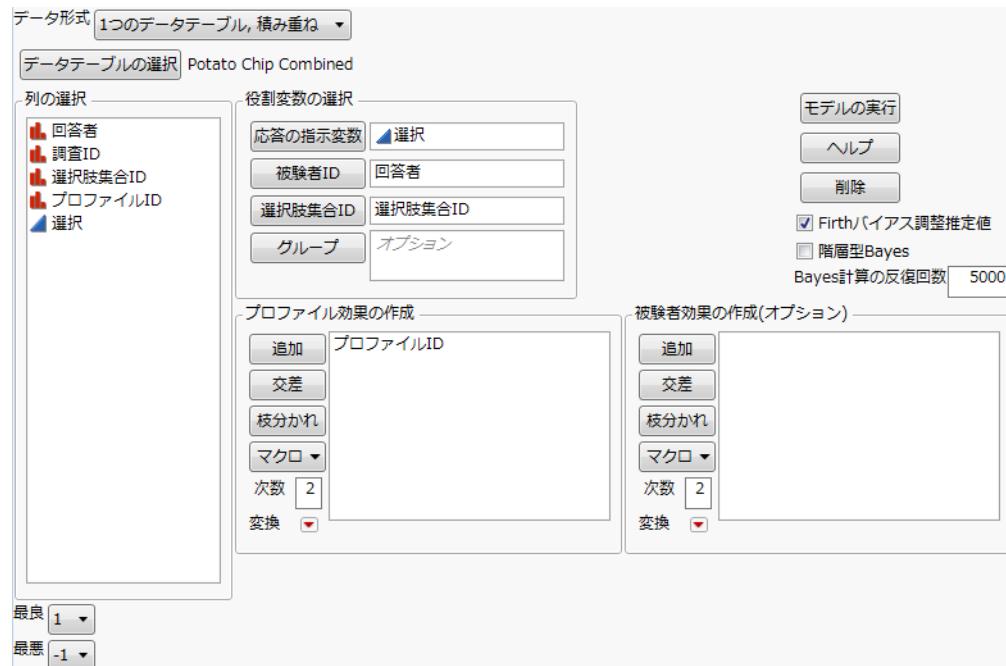
「MaxDiff」プラットフォームでは、1つのデータテーブルにまとめたデータでも、複数のデータテーブルに分かれたデータでも分析できます。複数のテーブルを使う場合は、応答、選択肢集合、被験者に関する情報をそれぞれ個別のデータテーブルに保存しておきます。1つのデータテーブルを使う場合は、すべての情報を1つのデータテーブルにまとめます。

- 「1つのデータテーブル」(169ページ) に、1つのデータテーブルにまとめたデータを分析する、簡単な例を説明しています。この例の分析を発展させて、プロファイルや被験者の情報を加えることもできます。
- 「複数のデータテーブル」(172ページ)に、複数のデータテーブルに記録されたデータに基づいて、MaxDiff分析を行う操作方法を説明しています。

## 1つのデータテーブル

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Potato Chip Combined.jmp」を開きます。
2. [分析] > [消費者調査] > [MaxDiff] を選択します。  
「データ形式」は、デフォルトでは [1つのデータテーブル, 積み重ね] に設定されています。
3. [データテーブルの選択] をクリックします。
4. 「Potato Chip Combined」を選択し、[OK] をクリックします。
5. 次の手順に従って、列に役割を割り当てます。入力後の起動ウィンドウは、図8.2のようになります。
  - 「選択」を選択し、[応答の指示変数] をクリックします。
  - 「回答者」を選択し、[被験者ID] をクリックします。
  - 「選択肢集合ID」を選択し、[選択肢集合ID] をクリックします。
  - 「プロファイルID」を選択し、「プロファイル効果の作成」パネルで [追加] をクリックします。

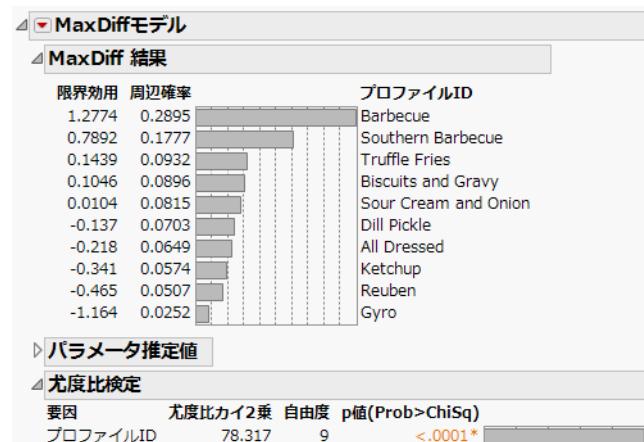
図 8.2 入力後の「MaxDiff」起動ウィンドウ



「最悪」の値は、応答の列を「応答の指示変数」に指定した時点で自動的に「-1」に変わります。ここでは、「最良」を示す値を「1」、「最悪」を示す値を「-1」にしたいので、自動的に設定された値そのまま使います。

6. [モデルの実行] をクリックします。

図8.3 「Potato Chip Combined.jpg」のMaxDiffレポート



レポートは、「プロファイルID」が有意であること、つまり異なる風味のポテトチップスに対する好みが有意に異なることを示しています。限界効用が最も高いのは、「Barbecue」風味です。「Barbecue」が他の風味よりも好まれる確率の推定値は、0.2895です。

7. 「MaxDiffモデル」の赤い三角ボタンをクリックし、[全水準の比較レポート] を選択します。

図8.4 全水準の比較レポート

全水準の比較レポート										
差(行 - 列) 差の標準誤差 Wald p値	Gyro	Sour Cream and Onion	Truffle Fries	Biscuits and Gravy	Reuben	Ketchup	All Dressed	Dill Pickle	Barbecue	Southern Barbecue
Gyro	0	-1.1747	-1.3082	-1.2688	-0.6992	-0.8229	-0.9467	-1.027	-2.4417	-1.9535
	0	0.33327	0.35477	0.34456	0.34758	0.32802	0.33571	0.35251	0.41499	0.38939
		0.00062	0.00035	0.00036	0.04668	0.01356	0.00569	0.00432	4.32e-8	2.01e-6
Sour Cream and Onion	1.17469	0	-0.1335	-0.0941	0.47548	0.35175	0.22804	0.14766	-1.267	-0.7788
	0.33327	0	0.31379	0.30716	0.30932	0.31915	0.28158	0.30707	0.37654	0.34896
		0.00062	0.67137	0.75983	0.12709	0.27279	0.41975	0.63156	0.00105	0.02764
Truffle Fries	1.30817	0.13348	0	0.03935	0.60896	0.48523	0.36152	0.28114	-1.1335	-0.6453
	0.35477	0.31379	0	0.29145	0.31817	0.31157	0.31136	0.30871	0.39213	0.34904
		0.00035	0.67137		0.90032	0.05821	0.12223	0.24809	0.36444	0.00462
Biscuits and Gravy	1.26882	0.09413	-0.0393	0	0.56961	0.44588	0.32217	0.24179	-1.1729	-0.6847
	0.34456	0.30716	0.31345	0	0.30639	0.29752	0.29974	0.28249	0.36267	0.34038
		0.00036	0.75983	0.90032		0.06567	0.1368	0.28477	0.39388	0.00161
Reuben	0.69921	-0.4755	-0.609	-0.5696	0	-0.1237	-0.2474	-0.3278	-1.7425	-1.2543
	0.34758	0.30932	0.31817	0.30639	0	0.30893	0.30148	0.30328	0.37117	0.35324
		0.04668	0.12709	0.05821	0.06567		0.68956	0.41355	0.28208	7.69e-6
Ketchup	0.82294	-0.3517	-0.4852	-0.4459	0.12373	0	-0.1237	-0.2041	-1.6188	-1.1305
	0.32802	0.31915	0.31157	0.29752	0.30893	0	0.29728	0.27938	0.36911	0.34808
		0.01356	0.27279	0.12223	0.1368	0.68956		0.67812	0.46661	2.64e-5
All Dressed	0.94665	-0.228	-0.3615	-0.3222	0.24744	0.12371	0	-0.0804	-1.4951	-1.0068
	0.33571	0.28158	0.31136	0.29974	0.30148	0.29728	0	0.29447	0.36838	0.3418
		0.00569	0.41975	0.24809	0.28477	0.41355	0.67812		0.78538	9.22e-5
Dill Pickle	1.02703	-0.1477	-0.2811	-0.2418	0.32782	0.20409	0.08038	0	-1.4147	-0.9264
	0.35251	0.30707	0.30871	0.28249	0.30328	0.27938	0.29447	0	0.3539	0.35532
		0.00432	0.63156	0.36444	0.39388	0.28208	0.46661	0.78538		0.00012
Barbecue	2.44172	1.26703	1.13355	1.1729	1.7425	1.61877	1.49507	1.41468	0	0.48824
	0.41499	0.37654	0.39213	0.36267	0.37117	0.36911	0.36838	0.3539	0	0.39164
		4.32e-8	0.00105	0.00462	0.00161	7.69e-6	2.64e-5	9.22e-5	0.00012	0.21514
Southern Barbecue	1.95348	0.77879	0.6453	0.68465	1.25426	1.13053	1.00683	0.92644	-0.4882	0
	0.38939	0.34896	0.34904	0.34038	0.35324	0.34808	0.3418	0.35532	0.39164	
		2.01e-6	0.02764	0.06715	0.0467	0.00056	0.00154	0.00393	0.01038	0.21514

この表は、行に記されている風味と列に記されている風味を比較し、効用の推定値の差を求めたものです。p値が小さいものは、濃い青色（マイナス）または赤色（プラス）で表示されています。たとえば、「Gyro」の行が全体に青色であることから、「Gyro」風味のポテトチップスの効用が他のすべての風味より有意に低いとわかります。「Barbecue」風味は、他のすべての味より効用が高いものの、「Southern Barbecue」風味との差は有意ではありません。

**メモ:** 「全水準の比較レポート」のp値は、多重比較に対する調整がまったく行われていません。これらのp値の結果は、大まかな目安として使用してください。

## 複数のデータテーブル

ここでは、「Potato Chip Profiles.jmp」、「Potato Chip Responses.jmp」、「Potato Chip Subjects.jmp」の3つのデータテーブルを使用します。調査のデータは、常に1つのデータテーブルにまとめることができますが、特に分析に含めたいプロファイル変数や被験者変数が他にもある場合は、複数のデータテーブルを使用する方が便利なことがあります。

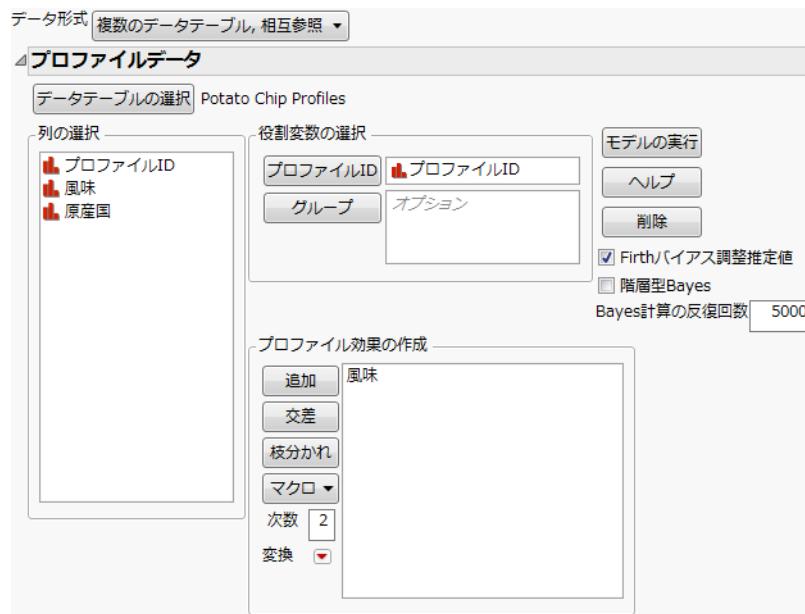
### 起動ウィンドウへの入力

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Potato Chip Responses.jmp」データテーブルを開きます。

**メモ:** 起動ウィンドウへの入力をスキップしたいときは、「風味による MaxDiff」スクリプトの緑色の三角ボタンをクリックします。その後、「モデルを調べる」(175ページ) に進んでください。

2. 「プロファイルと被験者のデータを開く」スクリプトの隣にある緑色の三角ボタンをクリックします。
  - プロファイルデータテーブルである「Potato Chip Profiles.jmp」には、調査の対象となったポテトチップスの種類（「風味」）と原産国（「原産国」）の情報がリストされています。各選択肢に「プロファイルID」がついています。
  - 被験者データテーブルである「Potato Chip Subjects.jmp」には、回答者がリストされています。回答者の「市民権」と「性別」も含まれています。
  - 応答データテーブルである「Potato Chip Responses.jmp」には、回答者がリストされています。回答者ごとに、各プロファイルセットの「調査ID」と「選択肢集合ID」、各選択肢集合の「プロファイルID」がまとめてあります。このデータテーブルには、応答データである「最良のプロファイル」と「最悪のプロファイル」の列もあります。
3. 3つのデータテーブルのいずれかで、[分析] > [消費者調査] > [MaxDiff] を選択します。
4. 「データ形式」メニューから [複数のデータテーブル, 相互参照] を選択します。  
各データソースに対応する3つのアウトラインがあります。
5. 「プロファイルデータ」アウトラインにある [データテーブルの選択] をクリックします。  
「プロファイルデータ」ウィンドウが開き、ここで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。
6. 「Potato Chip Profiles.jmp」を選択し、[OK] をクリックします。  
データテーブルの列が「列の選択」リストに表示されます。
7. 「列の選択」リストから「プロファイルID」を選択し、「役割変数の選択」にある [プロファイルID] をクリックします。
8. 「風味」を選択し、「モデル効果の作成」にある [追加] をクリックします。  
なお、「原産国」も効果に含めてもよいですが、この例では省きます。

図8.5 入力後の「プロファイルデータ」アウトライン



9. 「応答データ」アウトラインを開きます。[データテーブルの選択] をクリックします。
10. 「Potato Chip Responses.jmp」を選択し、[OK] をクリックします。
11. 次の手順に従って、列に役割を割り当てます。入力後の起動ウインドウは、図8.6のようになります。
  - 「最良のプロファイル」を選択し、[最良の選択] をクリックします。
  - 「最悪のプロファイル」を選択し、[最悪の選択] をクリックします。
  - 「選択肢 1」、「選択肢 2」、「選択肢 3」を選択し、[選択肢のプロファイルID] をクリックします。
  - 「回答者」を選択し、[被験者ID] をクリックします。

図8.6 入力後の「応答データ」アウトライン



12. 「被験者データ」アウトラインを開きます。[データテーブルの選択] をクリックします。
13. 「Potato Chip Subjects.jmp」を選択し、[OK] をクリックします。
14. 「回答者」を選択し、「被験者ID」をクリックします。
15. 「市民権」と「性別」を選択し、「モデル効果の作成」にある [追加] をクリックします。

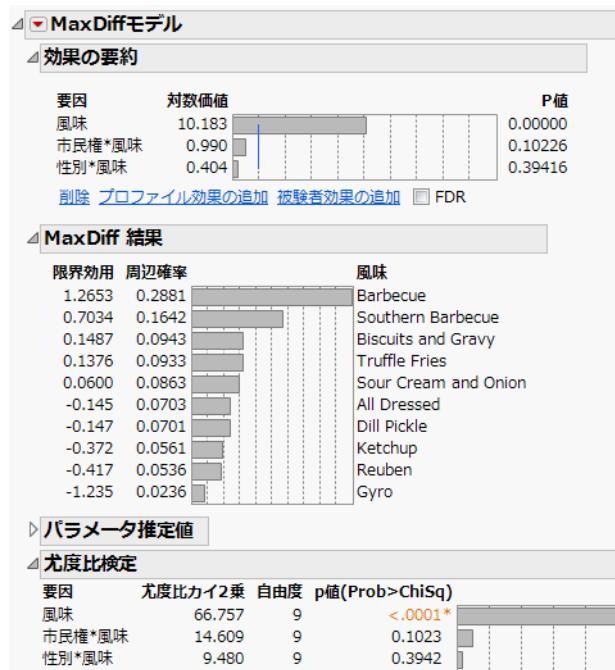
図8.7 入力後の「被験者データ」アウトライン



## モデルを調べる

1. [モデルの実行] をクリックします。

図8.8 「MaxDiff モデル」 レポート



「効果の要約」 レポートには、モデル項とその有意性を示す  $p$  値が表示されます（図8.8）。「風味」はプロファイル効果です。「市民権\*風味」と「性別\*風味」は、被験者効果とプロファイル効果の交互作用です。 「尤度比検定」 レポートは、「風味」が有意であることを示しています。

## 「MaxDiff」 プラットフォームの起動

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

「MaxDiff」 プラットフォームで使用するデータは、1つのデータテーブルにまとめたものでも、2、3個のデータテーブルに分かれているものでもかまいません。「選択モデル」 ウィンドウが開いたら、「データ形式」 リストで、1つのデータテーブルを使うか、複数のデータテーブルを使うかを指定します。

### 1つのデータテーブル、積み重ね

このデータ形式では、データが1つのデータテーブルにまとめられ、各行が被験者に提示される各プロファイルに対応します。また、選択された最良と最悪のものがどれであるかを示す指示変数の列も含めてください。

1つのデータテーブル形式の例については、「1つのデータテーブル」(169ページ)を参照してください。詳細については、「[1つのデータテーブル, 積み重ね] の起動ウィンドウ」(176ページ)を参照してください。

### 複数のデータテーブル, 相互参照

このデータ形式では、2つまたは3つのデータテーブルに分かれているデータを使用します。プロファイルデータと応答データは必須で、被験者データはオプションで使うことができます。「MaxDiff」起動ウィンドウは、各データテーブルに対応する3つのセクションで構成されています。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。

複数のデータテーブル形式の例については、「複数のデータテーブル」(172ページ)を参照してください。詳細については、「[複数のデータテーブル, 相互参照] の起動ウィンドウ」(177ページ)を参照してください。

## [1つのデータテーブル, 積み重ね] の起動ウィンドウ

図8.9は、「Potato Chip Combined.jmp」を使って入力した1つのデータテーブルの起動ウィンドウです。

図8.9 [1つのデータテーブル, 積み重ね] データ形式の起動ウィンドウ



**データテーブルの選択** データが1つにまとめて保存されているデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、[その他...] を選択します。

**応答の指示変数** どれが「最良」や「最悪」であるかを示す列。「最良」と「最悪」に「1」、「-1」、「0」のうちのいずれか2つを指定すると、残りの1つが「最良」でも「最悪」でもないプロファイルとみなされます。ウィンドウ左下で「最良」と「最悪」を指定しない限り、「最良」には「1」、「最悪」には「-1」が使用されます。

**被験者ID** 調査に参加した人のID。

**選択肢集合ID** 被験者に提示されたプロファイルの集合を識別するためのID。これらの集合のなかから、回答者はいずれかのプロファイルを選択します。

**グループ** 「選択肢集合ID」と共に使用したときに、各選択肢集合を一意に示すことができる列。たとえば、「選択肢集合ID」の値が「1」である、異なる選択肢集合があったとします。一方の選択肢集合は「調査」が「A」で、他方は「調査」が「B」である場合、その「調査」を【グループ】列に指定します。

**プロファイル効果の作成** プロファイルの属性で構成された効果を追加します。

「プロファイル効果の作成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

**被験者効果の作成(オプション)** 被験者に関する効果を追加します。

「被験者効果の作成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

**Firthバイアス調整推定値** バイアス修正を伴う最尤推定は、通常の最尤推定に比べ、推定や検定がより良い性質をもちます。さらに、ロジスティックモデルなどで生じる分離の問題も改善できます。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。

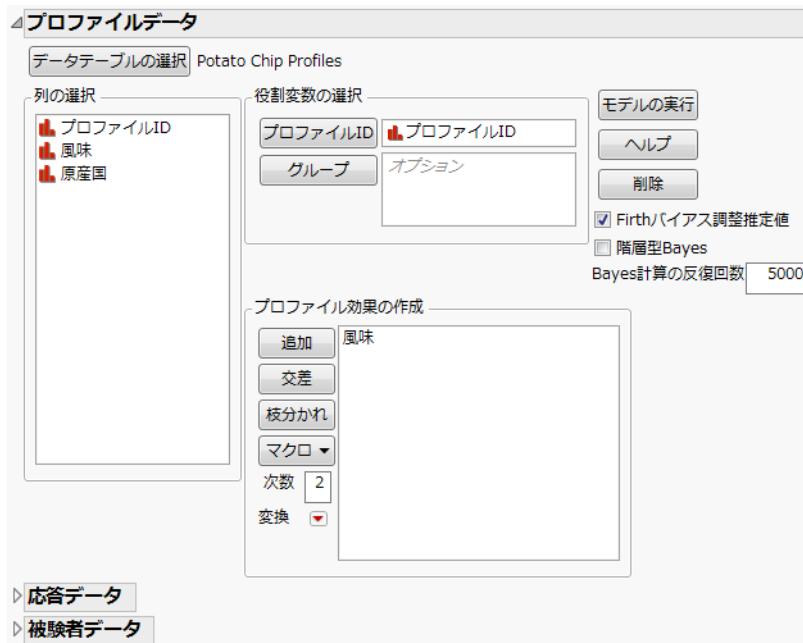
**JMP PRO 階層型Bayes** Bayes流の方法で被験者に関する係数の推定値を求めます。詳細については、「[Bayesパラメータ推定値](#)」(185ページ) を参照してください。

**JMP PRO Bayes計算の反復回数** ([階層型Bayes] を選択した場合のみ有効。) 被験者に関する係数の推定値を計算する際に使う適応型Bayesアルゴリズムの反復回数。ここで指定した値には、推定からは破棄されるバーンインの反復回数も含まれています。バーンインの反復回数は、起動ウィンドウで指定した「Bayes計算の反復回数」の半分です。

## [複数のデータテーブル, 相互参照] の起動ウィンドウ

図8.10は、複数のデータテーブルを用いる例です。起動ウィンドウにおいて、「プロファイルデータ」に「[Potato Chip Profile.jmp](#)」を用いています。

図8.10 [複数のデータテーブル, 相互参照] データ形式の起動ウィンドウ



〔複数のデータテーブル, 相互参照〕の場合、起動ウィンドウは3つのセクションに分かれています。

- ・「プロファイルデータ」(178ページ)
- ・「応答データ」(179ページ)
- ・「被験者データ」(180ページ)

## プロファイルデータ

プロファイルデータテーブルには、それぞれの選択肢の属性を示すデータを保存してください。データテーブルの各列が1つの属性に対応するように、また、各行が1つのプロファイルに対応するように、データを作成してください。さらに、各プロファイルのIDを含んだ列を設けてください。

**データテーブルの選択** プロファイルのデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、「その他...」を選択します。

**プロファイルID** 属性の組み合わせ（プロファイル）を識別するためのID。[プロファイルID] によってプロファイルデータテーブルの各行を一意に識別できない場合は、[グループ] 変数も指定する必要があります。その場合、[グループ] 列と [プロファイルID] 列の組み合わせによって各行が一意に識別できるように、[グループ] 列を追加してください。

**グループ** 「選択肢集合ID」列と共に使用したときに、各選択肢集合を一意に示すことができる列。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。一方の選択肢集合は「調査」が「A」で、他方は「調査」が「B」である場合、その「調査」を [グループ] 列に指定します。

**プロファイル効果の作成** プロファイルの属性で構成された効果を追加します。

「プロファイル効果の作成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

**Firthバイアス調整推定値** バイアス修正を伴う最尤推定法では、バイアス修正を行わない最尤推定に比べ、推定値と検定がより正確です。さらに、ロジスティックモデルなどで生じる分離の問題も改善できます。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。

**JMP<sup>®</sup> PRO 階層型 Bayes** Bayes流の方法で被験者の係数の推定値を求める。詳細については、「[Bayesパラメータ推定値](#)」(185ページ) を参照してください。

**JMP<sup>®</sup> PRO Bayes計算の反復回数** ([階層型 Bayes] を選択した場合のみ有効。) 被験者効果の推定値を計算する際に使う適応型 Bayes アルゴリズムの反復回数。この数値は、破棄されるバーンイン反復の回数も含みます。バーンイン反復回数は、起動ウィンドウで指定した「Bayes計算の反復回数」の半数に等しくなります。

## 応答データ

図8.11は、「Potato Chip Responses.jmp」を使って入力した「応答データ」アウトラインです。

図8.11 「応答データ」アウトライン



応答データテーブルは、調査の結果をまとめたものです。応答データテーブルには、各質問で提示された選択肢集合のIDと、回答者が選択した最良と最悪の選択肢を含めてください。このデータテーブルは、選択肢集合を示す列と、回答者が選択した選択肢を示す列により、プロファイルデータと関連付けられます。応答データとプロファイルデータにおける選択肢IDを対応さ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に指定してください。

**データテーブルの選択** プロファイルデータを含むデータテーブルを選択するか、開きます。まだ開いていないファイルを開くには、「その他...」を選択します。

**最良の選択** 被験者が「最良」であるとしたプロファイルのIDを含んだ列。

**最悪の選択** 被験者が「最悪」であるとしたプロファイルのIDを含んだ列。

**選択肢のプロファイルID** 各選択肢集合を構成する選択肢のプロファイルIDを含んだ列。

**グループ** 「選択されたプロファイルID」列と共に使用したときに、各選択肢集合を一意に示すことができる列。

**被験者ID** 調査に参加した人を一意に示すID。

**度数** 度数を含んだ列。度数が $n$ である行は、データに $n$ 回登場しているものとして計算に使用されます。

度数が1未満である行や、欠測値である行は、分析に使用されません。

**重み** データテーブルの各行に対する重みを含んだ列。0より大きい重みだけが分析に使われます。

**By** この列の水準に従ってデータがグループ化され、それぞれ個別に分析されます。指定した列の水準ごとに、対応する行が個別のテーブルとして個別に分析されます。分析結果は、個別のレポートにまとめられます。複数のBy変数を指定した場合は、By変数の水準のすべての組み合わせごとに分析が行われます。

## 被験者データ

図8.12は、「Potato Chip Subjects.jmp」を使って入力した「被験者データ」アウトラインです。

図8.12 「被験者データ」アウトライン



**メモ:** 被験者データテーブルは、被験者効果をモデルに含める場合のみ、必要になります。

このデータテーブルの列には、被験者ID、および、被験者の属性や特性を含めてください。このデータテーブルは、被験者数と同数の行で構成してください。また、応答データテーブルに、対応した被験者IDの列を用意してください。

**メモ:** 被験者データは応答データテーブルに含めてもかまいませんが、その場合も、「被験者データ」アウトライൻで被験者効果を指定する必要があります。

**データテーブルの選択** 被験者のデータテーブルを選択するか、まだデータテーブルが開いていない場合にはファイルを開きます。まだ開いていないデータテーブルを開くには、[その他...] を選択します。

**被験者 ID** 各被験者を一意に識別する ID。

**モデル効果の構成** 被験者に関する効果を追加します。

「モデル効果の構成」パネルについては、『基本的な回帰モデル』の「モデルの指定」章にある「モデル効果の構成」の節を参照してください。

## 「MaxDiff モデル」レポート

「MaxDiff モデル」ウィンドウには、起動ウィンドウで行った選択の内容に応じて以下のレポートのいくつかがデフォルトで表示されます。

- 「効果の要約」 (181 ページ)
- 「MaxDiff 結果」 (183 ページ)
- 「パラメータ推定値」 (183 ページ)
- 「Bayes パラメータ推定値」 (185 ページ)
- 「尤度比検定」 (186 ページ)

プラットフォームオプションについては、「「MaxDiff」プラットフォームのオプション」 (186 ページ) を参照してください。

### 効果の要約

モデルに少なくとも 1 つの効果が含まれている場合は、「効果の要約」レポートが表示されます。モデルによって推定された効果がリストされ、それらの効果の対数値 (または FDR 対数値) のプロットが表示されます。また、このレポートでは、モデルに効果を追加したり削除したりできます。このとき、モデルをあてはめた結果は、「効果の要約」レポートでの変更に合わせて自動的に更新されます。詳細については、『基本的な回帰モデル』の「標準最小2乗のレポートとオプション」章にある「効果の要約レポート」の節を参照してください。

起動ウィンドウで [Bayes 流の被験者効果] チェックボックスをオンにすると、「効果の要約」レポートは表示されません。尤度比検定が行われないためです。

図 8.13 は、「Potato Chip Responses.jmp」の「風味による MaxDiff」スクリプトを実行したときに作成される「効果の要約」レポートです。

図 8.13 「効果の要約」レポート



## 「効果の要約」表の列

「効果の要約」表には次の列があります。

**要因** モデル内の効果。 $p$  値の小さい順に並べられます。

**対数価値** 各効果の対数価値 (LogWorth)。 $-\log_{10}(p\text{ 値})$  という式で計算されています。 $p$  値を対数価値に変換すると、解釈がしやすくなります。対数価値が2以上あるものは、有意水準0.01で有意です ( $-\log_{10}(0.01) = 2$ )。

**FDR 対数価値** 各効果のFDR (False Discovery Rate; 偽発見率) の対数価値で、 $-\log_{10}(\text{FDR 調整 } p\text{ 値})$  という式で計算されています。これは、検定の有意性をグラフに表すのに適している統計量です。「対数価値」を「**FDR 対数価値**」に変更するには、【**FDR**】チェックボックスをオンにします。

**棒グラフ** 対数価値（またはFDR 対数価値）の棒グラフ。このグラフには整数値に縦の点線が、2の位置に青の参照線が引かれています。

**p 値** 各モデル効果の $p$  値。この $p$  値は、「尤度比検定」レポートにある有意性検定に対応します。

**FDR p 値** Benjamini-Hochberg法で計算された、各効果のFDR調整 $p$  値。FDRは、検定の多重性を考慮して、生の $p$  値を調整したものです。「**P 値**」列を「**FDR p 値**」に変更するには、【**FDR**】チェックボックスをオンにします。

FDRについては、Benjamini and Hochberg (1995) を参照してください。偽発見率の詳細については、『予測モデルおよび発展的なモデル』の「応答のスクリーニング」章、またはWestfall et al. (2011) を参照してください。

## 「効果の要約」表のオプション

要約表の下にあるオプションによって、効果を追加や削除できます。

**削除** モデルから選択した効果を削除します。1つまたは複数の効果を削除するには、削除する効果に対応する列を選択し、【削除】ボタンをクリックします。

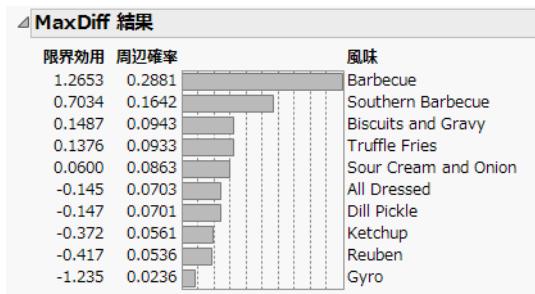
**プロファイル効果の追加** パネルが開き、[1つのデータテーブル、積み重ね] データ形式ならデータテーブルのすべての列、[複数のデータテーブル、相互参照] データ形式ならプロファイルデータテーブルの列がリストされます。モデルに追加する列を選択し、リストの下にある【追加】ボタンをクリックすると、その効果がモデルに追加されます。【閉じる】をクリックすると、パネルが閉じます。

**被験者効果の追加** パネルが開き、[1つのデータテーブル、積み重ね] データ形式ならデータテーブルのすべての列、[複数のデータテーブル、相互参照] データ形式なら被験者データテーブルの列がリストされます。モデルに追加する列を選択し、リストの下にある [追加] ボタンをクリックすると、その効果がモデルに追加されます。[閉じる] をクリックすると、パネルが閉じます。

## MaxDiff 結果

図8.14は、「Potato Chip Responses.jmp」の「風味による MaxDiff」スクリプトを実行したときに作成される「MaxDiff 結果」レポートです。

図8.14 「MaxDiff 結果」レポート



起動ウィンドウで指定したプロファイル効果に対し、次のような値が表示されます。

**限界効用** 効用に与える影響を示した指標。値が大きいほど、その機能に大きな価値があると考えられます。

**周辺確率** 被験者が、その効果の水準を他のすべての水準より良いと判断する確率の推定値。それぞれの効果において、周辺確率の和は1になります。

**棒グラフ** 周辺確率の棒グラフ。

**効果の列** 効果の名前の下に、水準がリストされます。各水準に、限界効用と周辺確率の推定値が対応しています。

## パラメータ推定値

このレポートには、パラメータ推定値、モデル選択の規準、反復計算の結果といった詳細が表示されます。

図8.15は、「Potato Chip Responses.jmp」の「風味による MaxDiff」スクリプトを実行したときに作成される「パラメータ推定値」レポートです。

図 8.15 「パラメータ推定値」レポート

△パラメータ推定値		
項	推定値	標準誤差
風味[All Dressed]	-0.15616234	0.2227718641
風味[Barbecue]	1.22210326	0.2978972284
風味[Biscuits and Gravy]	0.16398758	0.2249076601
風味[Dill Pickle]	-0.17356828	0.2151668801
風味[Gyro]	-1.11927509	0.2827016652
風味[Ketchup]	-0.47308393	0.2319900823
風味[Reuben]	-0.50927309	0.2294979936
風味[Sour Cream and Onion]	0.2111573	0.2450361368
風味[Southern Barbecue]	0.70149945	0.2695322771
市民権[Canadian]*風味[All Dressed]	-0.04368106	0.2239442590
市民権[Canadian]*風味[Barbecue]	-0.16180196	0.2978036763
市民権[Canadian]*風味[Biscuits and Gravy]	0.05734172	0.2233312021
市民権[Canadian]*風味[Dill Pickle]	-0.09824391	0.2188896544
市民権[Canadian]*風味[Gyro]	0.43257276	0.2907051874
市民権[Canadian]*風味[Ketchup]	-0.38035261	0.2349720398
市民権[Canadian]*風味[Reuben]	-0.34677939	0.2342643217
市民権[Canadian]*風味[Sour Cream and Onion]	0.56678250	0.2355493683
市民権[Canadian]*風味[Southern Barbecue]	-0.00712532	0.2720518372
性別[Female]*風味[All Dressed]	-0.26955535	0.2106170508
性別[Female]*風味[Barbecue]	0.39881410	0.2961368976
性別[Female]*風味[Biscuits and Gravy]	0.09184462	0.2245258406
性別[Female]*風味[Dill Pickle]	-0.11461679	0.2131955199
性別[Female]*風味[Gyro]	-0.40538827	0.2950786212
性別[Female]*風味[Ketchup]	-0.07317652	0.2135446256
性別[Female]*風味[Reuben]	0.20780054	0.2227994637
性別[Female]*風味[Sour Cream and Onion]	0.27353428	0.2337547848
性別[Female]*風味[Southern Barbecue]	0.07570279	0.2627935035
AICc	397.44423	
BIC	456.27172	
(-2)*対数尤度	327.00945	
(-2)*Firth対数尤度	244.39836	
勾配で収束しました		
Firthバイアス調整推定値		

項 モデル内の効果。

**推定値** 該当の項に対するパラメータの推定値。離散選択モデルでは、パラメータ推定値を「部分効用 (part-worth)」と呼ぶことがあります。パラメータ推定値は、各項が効用に対してどれくらい影響しているかを示す係数です。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で推定されます。

**標準誤差** パラメータ推定値の標準誤差の推定値。

### 比較規準

適合度統計量として、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対数尤度が表示されます。これらの統計量は、モデルを比較するのに用いられます。最初の3つの規準については、『基本的な回帰モデル』の付録「統計的詳細」に詳しい説明があります。

(-2)\*Firth対数尤度は、起動ウィンドウで [Firth バイアス調整推定値] オプションをオンにした場合のみ求められます。このオプションは、デフォルトではオフになっています。

適合度統計量は、いずれも値が小さいほどよくあてはまっていることを示します。

## JMP PRO Bayes パラメータ推定値

(この結果は、起動ウィンドウで〔階層型 Bayes〕を選択した場合のみ表示されます。)「Bayes パラメータ推定値」レポートには、階層型 Bayes モデルで求められた、モデル効果に関する結果が表示されます。階層型 Bayes モデルでは、被験者による違いを尤度関数に組み込み、製品属性のパラメータにおける被験者の違いを推定します。Bayes 推定のアルゴリズムには、適応型 Metropolis-Hastings 法 (Train 2001) の一種が使われています。各モデル効果に対し、事後分布の平均と分散が計算されます。また、各モデル効果に関して、被験者ごとの推定値も出力できます。詳細については、「[被験者推定値の保存](#)」(187 ページ) を参照してください。

この Bayes 流の方法では、被験者ごとの効果が確率変数（変量効果）とみなされます。製品の属性に関するパラメータが、被験者ごとに異なっており、特定の平均ベクトルと共に分散行列を持つ多変量正規分布に従うと仮定されます。そして、ある被験者が複数の選択肢から選択することを表す尤度関数には、多項ロジットモデルが仮定されます。（被験者ごとの効果が従う）多変量正規分布の平均ベクトルに対する事前分布には、平均ベクトルが 0 で、共分散行列が対角要素がすべて等しい対角行列の多変量正規分布が仮定されます。また、（被験者ごとの効果が従う）多変量正規分布の共分散行列が従う事前分布は、対角要素がすべて等しい対角行列を尺度行列とする逆 Wishart 分布が仮定されます。

反復計算では、最初にバーンイン期間が設けられており、最初のほうの反復の結果は破棄されます。デフォルトでは、そのバーンインの反復回数は、起動ウィンドウで指定した「Bayes 計算の反復回数」の半分です。

図 8.16 Bayes パラメータ推定値レポート

選択モデル				
Bayes パラメータ推定値				
項	事後 平均	事後 標準偏差	被験者	標準偏差
生地[Thick]	0.12241156	0.2270467796	0.7717269275	
チーズ[Jack]	-2.34768104	0.5846517681	0.6746528415	
トッピング[Pepperoni]	-0.40059962	0.2524944761	0.6955971534	
全反復回数	5000			
バーンイン反復回数	2500			
回答者数	32			
バーンイン後の平均対数尤度	-31.72861			

項 モデルの項。

**事後 平均** 該当の項に対する係数（パラメータ）の推定値。各反復において、各被験者に対する推定値が求められます。「事後 平均」は、その推定値を各反復で被験者に関して平均し、バーンイン反復後のそれらの全被験者平均をさらに平均したものです。

**ヒント:** 赤い三角ボタンのメニューから [Bayes チェーンの保存] を選択すると、反復ごとの個々の推定値が確認できます。

**事後 標準偏差** 各反復において、被験者ごとの推定値から平均が求められます。バーンイン反復後のそれらの全被験者平均から計算される標準偏差。

**被験者 標準偏差** まず、被験者ごとに推定値の平均をバーンイン後のものから計算します。「被験者 標準偏差」は、その標準偏差です。

**ヒント:** 赤い三角ボタンのメニューから「被験者推定値の保存」を選択すると、個々の推定値が確認できます。

**全反復回数** バーンイン反復を含め、実行された反復の合計数。

**バーンイン反復回数** 後で破棄されるバーンイン反復の回数。起動ウィンドウで指定した Bayes 反復回数の半分に相当します。

**回答者数** 被験者の数。

**バーンイン後の平均対数尤度** バーンイン反復後の生成された値から求めた対数尤度関数の平均。

## 尤度比検定

図8.17は、「Potato Chip Responses.jmp」の「風味による MaxDiff」スクリプトを実行したときに作成される「尤度比検定」レポートです。

図8.17 尤度比検定

尤度比検定			
要因	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
風味	66.757	9	<.0001*
市民権*風味	14.609	9	0.1023
性別*風味	9.480	9	0.3942

**要因** モデルに含まれる効果。

**尤度比カイ2乗** 該当する効果に対する検定の尤度比カイ2乗の値。

**自由度** カイ2乗検定の自由度。

**p値(Prob>ChiSq)** カイ2乗検定に対する *p* 値。

**棒グラフ** カイ2乗検定の値の棒グラフ。

## 「MaxDiff」プラットフォームのオプション

**JMP PRO** **最尤推定値の表示** (起動ウィンドウで「階層型 Bayes」を選択した場合のみ表示。) モデル項の係数を、通常の最尤法で推定した結果 (パラメータ推定値と標準誤差)。これらの値は、階層型 Bayes アルゴリズムの開始値として使用されます。

**複合因子検定** (起動ウィンドウで「階層型 Bayes」を選択しなかった場合のみ表示。) モデル内の各因子を検定する目的で、その因子が関連しているすべての効果の尤度比検定を行います。複合因子検定について詳しくは、『基本的な回帰モデル』の「標準最小2乗のレポートとオプション」章を参照してください。

**信頼区間** [階層型 Bayes] が選択されていないときには、「パラメータ推定値」レポートにパラメータの信頼区間を表示します。

[階層型 Bayes] が選択されている場合、信頼区間は「Bayes パラメータ推定値」レポートに表示されます。この Bayes 流の信頼区間は、正規分布を仮定し、事後平均と事後標準偏差に基づいて求められます。

**推定値の相関** [階層型 Bayes] が選択されていないときには、最尤法に基づいて、パラメータ推定値間の相関行列を表示します。

[階層型 Bayes] が選択されている場合は、パラメータ推定値の事後平均における相関行列を表示します。この相関行列は、バーンイン後の反復から求めたものです。バーンイン後、各反復ごとに事後平均が計算されます。これらの事後平均から求めた相関行列が、「推定値の相関」表に表示されます。

**比較** 設定された選択肢の比較を行います。比較したい因子と値を選択することができます。なお、[すべて] チェックボックスをチェックした場合、すべての水準が比較されます。[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、それらすべての因子の水準での比較ではなく、該当する 1 因子のすべての水準が、その他の因子を左側で選択されている水準に固定した上で、比較されます。詳細については、「[「比較」レポート](#)」(188 ページ) を参照してください。

**全水準の比較レポート** 「全水準の比較レポート」には、2 つのプロファイルを比較した結果が表示されます。モデルに被験者効果が含まれている場合は、被験者効果の水準を指定する必要があります。そのとき、被験者効果の指定された水準における比較だけが行われます。表の各セルには、効用の差、差の標準誤差、「差がない」ことを帰無仮説とする Wald 検定の  $p$  値が表示されます。

---

**注意:** 「全水準の比較レポート」の  $p$  値は、多重比較に対する調整がまったく行われていません。これらの  $p$  値の結果は、大まかな目安として使用してください。

---

Wald 検定の  $p$  値は、色付きで表示されます。行から列を引いた差が、正のときは青色、負のときは赤色で表示されます。青色または赤色の濃さは、有意性の高さを表します。濃い色ほど、差に対する  $p$  値が小さいことを示しています。

**効用計算式の保存** 効用の計算式列を含んだ、新しいデータテーブルを作成します。作成されたデータテーブルでは、各行がプロファイルと被験者の組み合わせに対応しており、プロファイルと被験者を示す列があります。

**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。詳細については、「選択モデル」章の「[セグメント化の例](#)」(148 ページ) を参照してください。

**JMP PRO 被験者推定値の保存** (起動ウィンドウで [階層型 Bayes] を選択した場合のみ表示。) 各効果における被験者ごとのパラメータ推定値を含んだデータテーブルを新たに作成します。この被験者ごとのパラメータ推定値を平均した値は、「Bayes パラメータ推定値」レポートにある「事後平均」と等しいです。「被験者受容率」は、Metropolis-Hastings 法のステップで求められた新しいパラメータ推定値の受容率を示

します。一般に、0.20ぐらいが良好な受容率とみなされています。詳細については、「[Bayesパラメータ推定値](#)」(185ページ) を参照してください。

**JMP<sup>®</sup> PRO** **Bayes チェーンの保存** (起動ウィンドウで [階層型 Bayes] を選択した場合のみ表示。) Bayes 推定値を計算するために使用されたチェーンの情報を含んだデータテーブルを、新たに作成します。詳細については、「[Bayes チェーンの保存](#)」(189ページ) を参照してください。

**モデルダイアログ** 現在の分析を実行した「MaxDiff」の起動ウィンドウを開きます。データテーブルや ID、モデル効果を新しく指定できます。たとえば、現在のモデルから設定を少し変更して、別のモデルをあてはめることができます。

以下のオプションについて詳しくは、『JMPの使用法』の「JMPのレポート」章を参照してください。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、[自動再計算] オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウで By 変数を指定した場合のみ使用可能です。

## 「比較」レポート

「比較」レポートは、ペアごとの比較を指定したときに表示されます。レポートには、次のような列が含まれます。

**因子** 指定した被験者因子の水準。

**比較対象1** 比較対象となっている1つ目のプロファイルの因子と水準。

**比較対象2** 比較対象となっている2つ目のプロファイルの因子と水準。

**効用1** 「因子」列の被験者における、1つ目のプロファイルの効用。

**効用2** 「因子」列の被験者における、2つ目のプロファイルの効用。

**確率1** 「因子」列の被験者において、2つ目のプロファイルより1つ目のプロファイルを好む確率。

**確率2** 「因子」列の被験者において、1つ目のプロファイルより2つ目のプロファイルを好む確率。

**オッズ比1** 「確率1」を「確率2」で割った値。

**オッズ比2** 「確率2」を「確率1」で割った値。

**比較の差** 「効用1」から「効用2」を引いた値。

**標準偏差** 「比較の差」における標準誤差。

## JMP PRO Bayes チェーンの保存

Bayes チェーンのデータテーブルを使うと、生成される推定値の安定性を調べることができます。作成されるデータテーブルには、(起動ウィンドウで指定した)「Bayes 計算の反復回数」に1を足した数の行が含まれます。最初の行(反復1)は、開始値を示します。2行目以降には、生成された乱数が反復の順に保存されます。各列には、次のような情報が outputされます。

**反復** 反復の番号。最初の行は、開始値です。

**対数尤度** その反復におけるモデルの対数尤度。各反復に対する対数尤度をプロットすれば、バーンイン期間(調整段階)での振る舞いを確認できます。

**<モデル効果> 適応型シグマ** 逆 Wishart 分布の尺度行列における対角要素の平方根。

**<モデル効果> 受容率** MCMC アルゴリズムでの受容率。

**<モデル効果> 平均** 被験者ごとに計算された推定値、の平均。

**<モデル効果> 分散** 被験者ごとに計算された推定値、の分散。

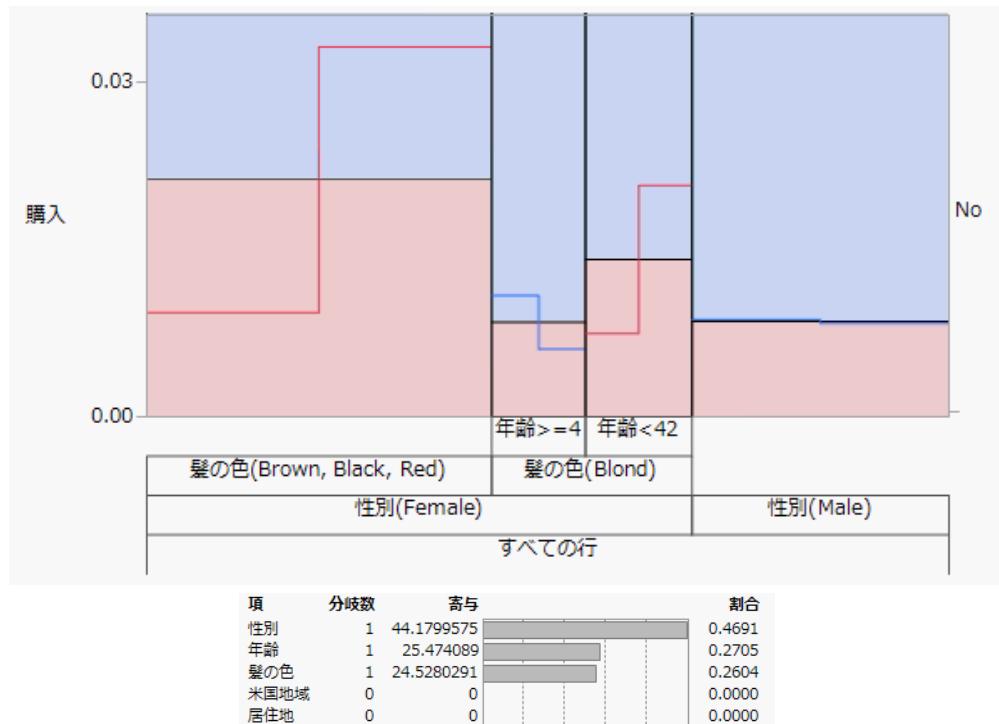


## JMP PRO アップリフトモデル アクションが消費者行動に及ぼす効果をモデル化

**JMP PRO** この左のマークが付いている機能は、JMP Proのみで提供されています。このプラットフォームは、JMP Proのみで提供されています。

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。たとえば、マーケティングキャンペーンを最適化したり、個別化医療の方法を探し出したりすることなどが考えられます。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け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9.1 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



## JMP PRO 「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に使用される基準は、各分岐の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  $p$  値の対数値 (-log<sub>10</sub>) がレポートされます。

## JMP PRO 「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。

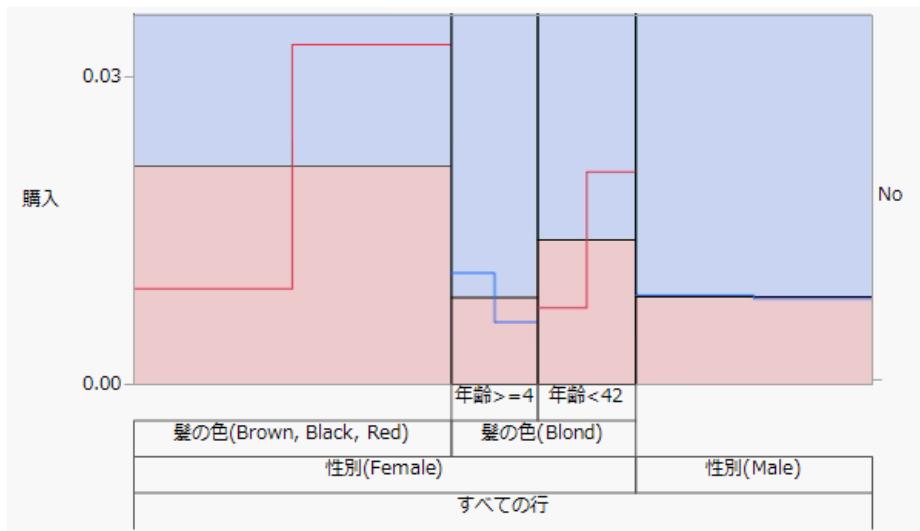
データテーブルの「広告」列は、広告が送付されたかどうかを示します。「購入」列は、製品を購入したかどうかを示します。各会員について、「性別」、「年齢」、「髪の色」、「米国 地域」、「居住地」（都市部かどうか）の情報が収集されています。また、「検証」列によって、カード会員の約33%が検証データとして使われるよう設定されています。

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」が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. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Hair Care Product.jmp」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3と判断されます。図9.2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9.2 3回分岐した後のグラフ

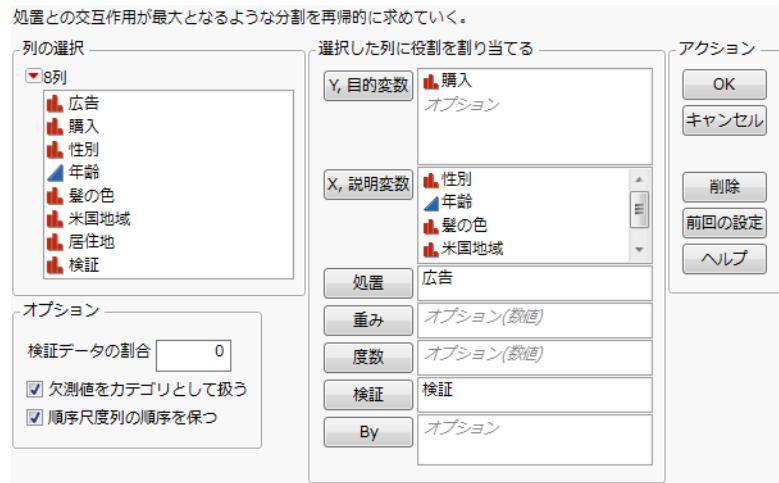


グラフから、広告によって購入にアップリフトが生じているのは、髪の色が黒 (black)・赤 (red)・茶色 (brown) の女性、または低年齢層 (年齢 < 42) で髪の色がブロンドである女性のいずれかのグループだとわかります。高年齢層 (年齢  $\geq 42$ ) でブロンドの女性の場合は、広告の効果がマイナスとなっています。

## JMP<sup>®</sup> 「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] > [アップリフト] を選択します。図9.3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9.3 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。「列の選択」リストで列を選択しない状態で「検証」ボタンをクリックすると、データテーブルに検証列を追加できます。「検証列の作成」ユーティリティについての詳細は、『予測モデルおよび発展的なモデル』の「モデル化ユーティリティ」章を参照してください。

次のようなオプションも用意されています。

**欠測値をカテゴリとして扱う** 説明変数がカテゴリカルな場合、このチェックボックスをオンにすると、欠測値を個別のカテゴリとして処理できます。説明変数が連続尺度の場合は、欠測値が同一の数値をもつものとして扱われます。

**順序尺度列の順序を保つ** このオプションを選択すると、順序が維持されるような分岐だけが考慮の対象となります。

---

## JMP PRO 「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。

## JMP<sup>®</sup> PRO 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、どちらを「処置1」とするかが決まり、描かれる線が赤色になるか、青色になるかが決まります。分岐を進めていくに従って、分岐の条件式（条件式は横軸の下に表示されています）やグラフ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。R2乗が1に近いほど、モデルのデータへのあてはまりが良いことを示します。

---

**メモ:** R2乗の値が0に近い場合、モデルに含まれていない変数の中に、説明されていない変動を説明するものがあると考えられます。ただし、データのばらつきがもともと大きなものであるときには、有効なアップリフトモデルであってもR2乗値が小さくなる場合もあります。

---

**RMSE** ツリーの回帰モデルにおける誤差の標準偏差 (RMSE)。RMSEは連続量の応答に対してのみ算出されます。詳細については、『基本的な回帰モデル』を参照してください。

**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。詳細については、『基本的な回帰モデル』の付録「統計的詳細」を参照してください。

## JMP<sup>®</sup> PRO 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9.4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。

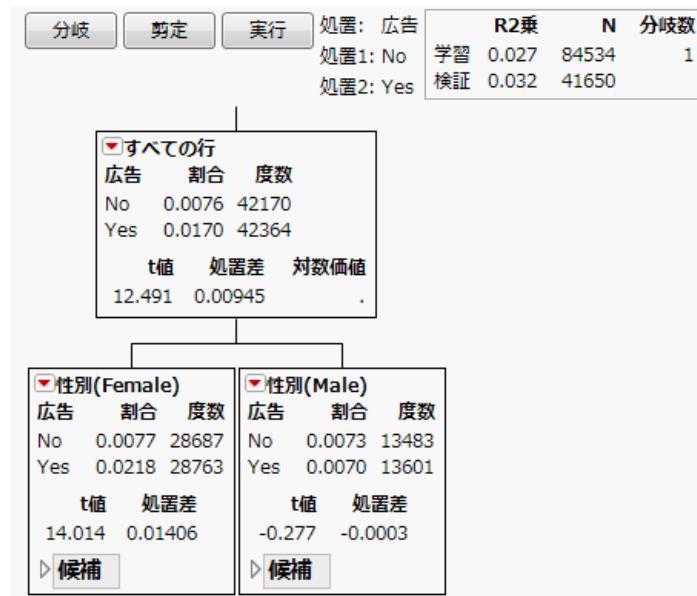
**t値** 処置の水準間における差を検定するときに使われるt値。なお、応答がカテゴリカルな場合は、この統計量は、連続尺度（0と1）に変換された後に算出されます。

**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9.4 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。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」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。

## JMP PRO 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、『予測モデルおよび発展的なモデル』の「パーティションモデル」章を参照してください。

## JMP PRO 分岐の最小サイズ

許容する最小の標本サイズを、度数または標本全体に占める割合として指定するウィンドウが開きます。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のうちの大きい方の値です。

## JMP PRO 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。

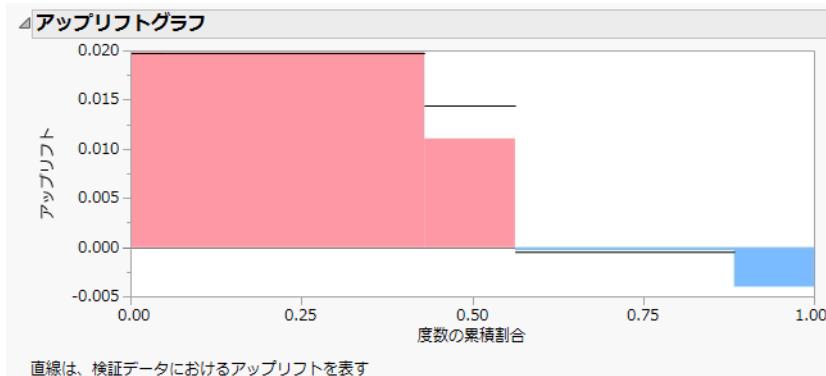
## JMP PRO アップリフトグラフ

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、処置間における平均差です。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9.5は、「Hair Care Product.jmp」サンプルデータで、3回分岐した後のアップリフトグラフの例です。被験者のうち、2つのグループ（男性と、年齢  $\geq 42$  のブロンドの女性）で、広告の効果がマイナスとなっています。

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9.5 アップリフトグラフ

**JMP PRO** 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。

**差の計算式を発行** 差の計算式を作成し、「計算式デポ」レポートに計算式列スクリプトとして保存します。「計算式デポ」レポートが開いていない場合は、このオプションを選択した時点でレポートが作成されます。『予測モデルおよび発展的なモデル』の「計算式デポ」章を参照してください。



# 第 10 章

## 項目分析 テスト結果に対する項目反応分析

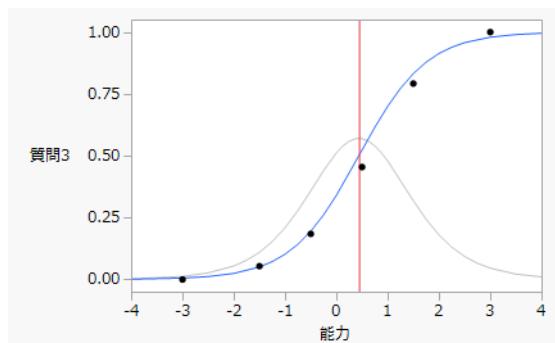
「項目分析」プラットフォームは、項目反応理論モデル (Item Response Theory model; IRT model) をあてはめます。項目反応理論は、テストや調査などを分析するのに使われています。項目反応理論は、各受験者の能力を、その受験者が問題を正答する確率と関連付けます。各受験者の能力は、直接測定できないために、「潜在特性 (latent)」とも呼ばれています。項目反応理論は、標準化テスト、認知テスト、消費者の選好などを調べるのに使用できます。古典的テスト理論 (Classical Test Theory; CTT) は、観測されたスコアの合計に注目した分析でした。項目反応理論は、そのような古典的テスト理論に代わる手法です。

「項目分析」プラットフォームでは、次のような結果を求めます。

- 項目ごとにスコアが求められます。これは、潜在特性から各項目への寄与を知るまでの手がかりになります。
- 受験者のスコアと、項目のスコアが、同じスケール上で求められます。
- 受験者のスコアと、項目のスコアが、同じスケール上に描かれます。
- 項目特性曲線 (item characteristic curve) が描かれます。この曲線は、受験者の潜在特性と、各項目との関係を描いたものです。

項目反応理論の詳細については、Ayala (2009) を参照してください。

図10.1 項目分析の特性プロット

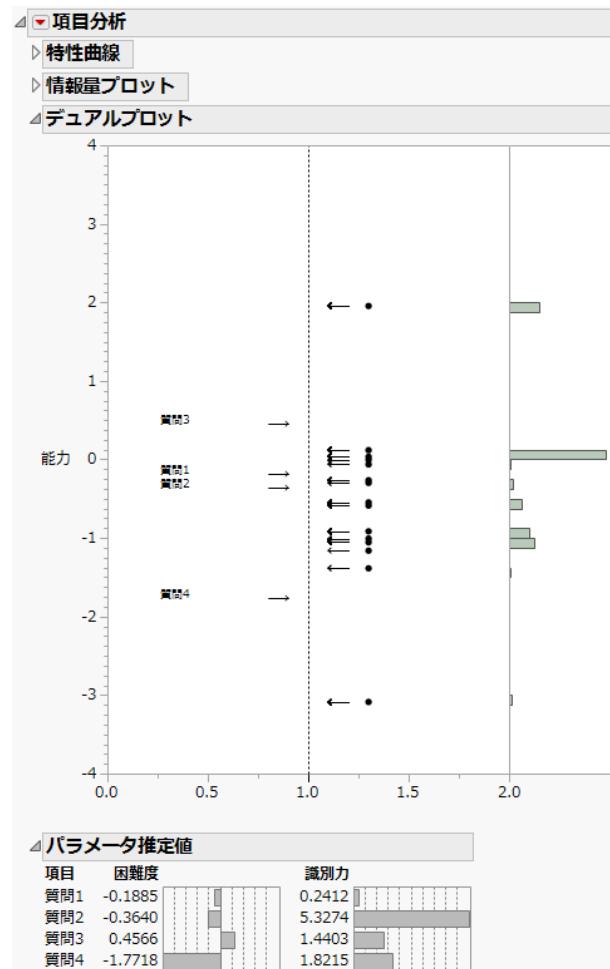


## 項目分析の例

この例では、「MathScienceTest.jmp」を使います。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 (TIMSS) のデータから採ったサブセットです。データテーブルには、14の問題に対する1263人の受験者のスコア (1 = 正解、または0 = 不正解) が含まれています。最初の4つの問題を調べて、問題と受験者の数学能力との関係を理解しましょう。テストの問題は、数学的な潜在能力を測るために使用される項目です。2PLモデルをこのデータにあてはめてみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「MathScienceTest.jmp」を開きます。
2. [分析] > [消費者調査] > [項目分析] を選択します。
3. 「質問1」から「質問4」までを選択し、[Y, テスト項目]をクリックします。そして、[OK]をクリックします。

図10.2 「項目反応」レポート

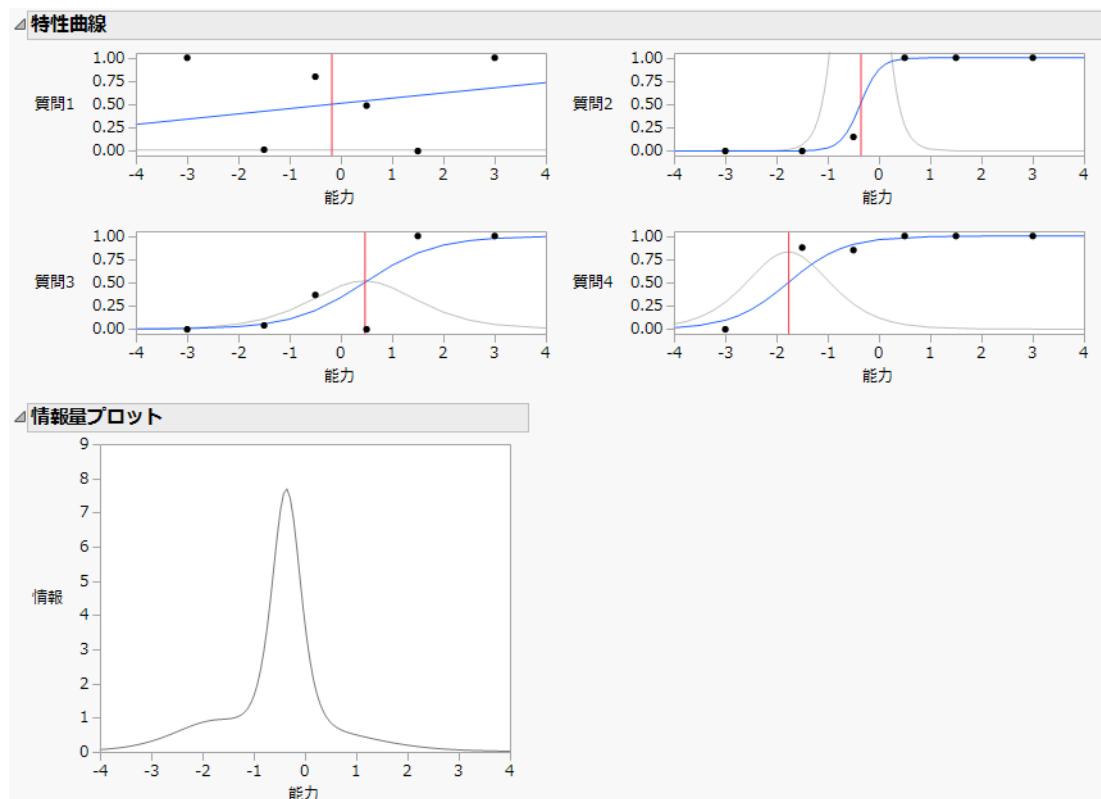


デュアルプロットから、質問4の困難度スコアが-1.78と最も低くなっているため、4つの問題のうち質問4が最も易しいことがわかります。質問3は、困難度スコアが0.46で、最も難しいことがわかります。受験者のほとんどは、グラフの中央部分にあるデータ点で示されているとおり、能力スケールの中央から下限までの範囲内に入っています。ヒストグラムでは、受験者の約40%が能力スケールにおいて、0のやや上にいることがわかります。

**メモ:** すべてが正解、またはすべてが不正解の受験者は、この分析には含まれていません。詳細については、「項目反応理論モデルの推定方法」(212ページ) を参照してください。

4. 「特性曲線」レポートの開閉アイコンをクリックして開きます。
5. 「項目分析」の赤い三角ボタンのメニューをクリックして、[横に並べるプロット数] を選択します。
6. 「2」と入力して [OK] をクリックします。
7. 「情報量プロット」レポートのグレーの開閉アイコンをクリックして開きます。

図10.3 項目反応の例



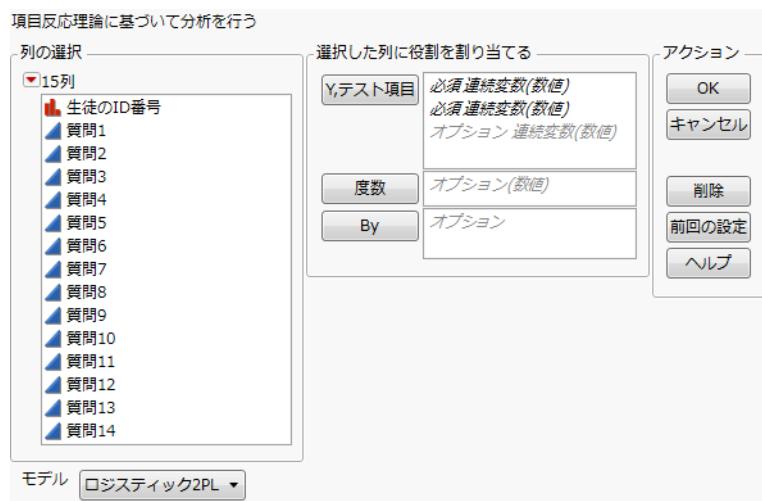
質問1のプロットを見ると、特性曲線も、情報量曲線も、平坦になっています。これは、質問1が受験者の数学的能力を識別するためにはそれほど多くの情報を有していないということを示しています。質問2の特性曲線は、傾斜が大きくなっています。これは、質問2は、受験者の能力を識別するのに役立つことを示しています。各プロットにおいて、特性曲線の変曲点の位置に、縦線が描かれています。この縦線は、該当の問題に受験者が正答する確率が50%となる能力レベルを示しています。

「情報量プロット」は、4つの問題の情報量を合わせたものです。おおよそ-1～0の能力レベルにおいて、最大の情報があることを示しています。困難度の高い問題を含めると、情報量曲線の幅が広がる可能性があることを示唆しています。

## 「項目分析」プラットフォームの起動

「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。

図10.4 「項目分析」起動ウィンドウ



**Y, テスト項目** 分析する列を2つ以上指定します。指定できる列は、数値型の連続尺度で、かつ、0と1のみを含んだ列です。

**ヒント:** [列] > [再コード化] を使えば、データを0と1に変換することができます。『はじめてのJMP』の「データの入力と編集」章を参照してください。

**度数** 度数変数を指定します。このオプションは、データが要約データである場合に用いてください。

**By** By変数の水準ごとに個別のレポートが作成されます。複数のBy変数を指定した場合は、By変数の水準のすべての組み合わせごとにレポートが作成されます。

**モデル** 次のオプションの中から必要なモデルを指定します。

**ロジスティック 2PL** 2パラメータのロジスティックモデル

**ロジスティック 3PL** 3パラメータのロジスティックモデル

**ロジスティック 1PL** Rasch型のパラメータ表現による1パラメータのロジスティックモデル

### ロジスティック 3PL モデルの詳細

[ロジスティック 3PL モデル] を選択した場合、[OK] をクリックした後に推量パラメータ  $c$  に対するペナルティ（罰則）を入力するウィンドウが表示されます。このペナルティのデフォルト値は0になっています。しかし、0以外の値を入力することもできます。このペナルティは、リッジ回帰のペナルティに似ています。このペナルティは、推量パラメータの推定値の分散に対するものです。ペナルティを使用すると、次のような利点があります。

- モデルパラメータの推定が安定する。
- 計算が速くなる。
- 少数のバイアスは生じるが、項目間の推量パラメータのばらつきが小さくなる。

ペナルティの値が大きいと、すべての推量パラメータが0になります。適度なペナルティは、推量パラメータのばらつきを軽減します。ペナルティ値をゼロにした場合は、ペナルティがまったく課せられません。

---

## 「項目分析」レポート

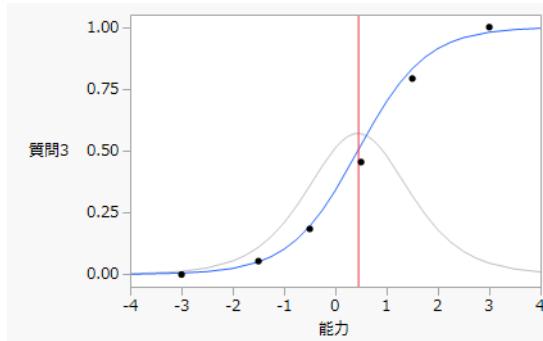
最初の「項目分析」レポートには、「デュアルプロット」レポートと「パラメータ推定値」レポートが表示されます。項目特性曲線と全体の情報量プロットは、レポート内にありますが、最初は閉じています。

### 特性曲線

「特性曲線」レポートには、起動ウィンドウで指定した各項目の項目特性曲線 (Item Characteristic Curve; ICC) が含まれます。

項目特性曲線は、項目への正答率と能力との関係をプロットしたものです。能力は標準化されているので、能力が0である受験者が平均的能力をもっていることを意味します。能力レベルに対して、観測された正答率が点でプロットされています。あてはめられた特性曲線と、その点とを比較すれば、モデルの適合度がよくわかります。また、特性プロットには、情報量曲線と、特性曲線の変曲点を示す縦線も描かれます。情報量曲線は、項目特性曲線の傾きをプロットしたものです。よって、情報量曲線は、変曲点において最大となります。

図10.5 項目特性曲線

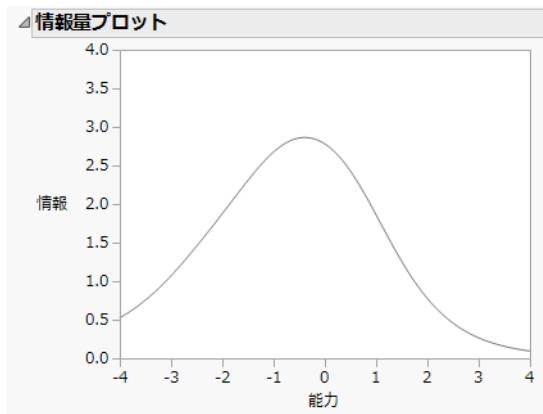


ヒント：「項目分析」の赤い三角ボタンのメニューにある「横に並べるプロット数」オプションを使って、1行あたりに表示されるプロットの数を変更できます。

## 情報量プロット

「情報量プロット」レポートには、全体の情報量曲線が描かれます。この曲線は、項目の情報量曲線を合計することで求められています。「情報量プロット」は、テストで測定できる能力レベルについて知る上での手がかりになります。図10.6は、高い能力レベルよりも、平均～低い能力レベルを評価するのに適した項目を含むテストであることを示しています。

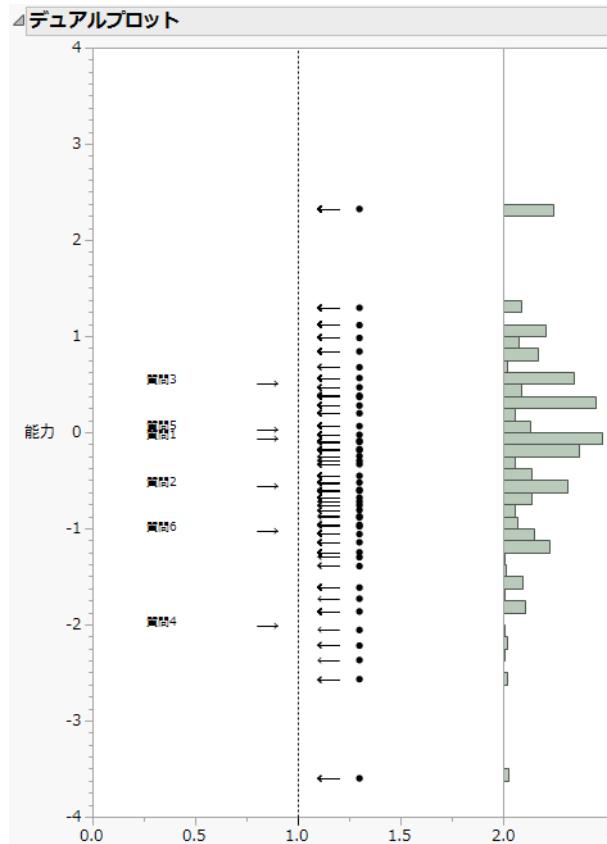
図10.6 情報量プロット



## デュアルプロット

「デュアルプロット」レポートには、項目の困難度と、受験者の能力が1つのプロットに描かれています。困難度と能力については、標準化された共通のスケールが使われています。各項目は、プロットの左側に困難度でプロットされます。各受験者は、プロットの右側に、能力でプロットされ、ヒストグラムも描かれます。デュアルプロットにより、各項目の困難度を各受験者の能力と関連付けることができます。

図10.7 デュアルプロット



## パラメータ推定値

「パラメータ推定値」レポートには、各項目のパラメータ推定値の表が表示されます。推定されるパラメータは、分析に使用したモデル（1PL、2PL、または3PL）によって異なります。

## 項目 テスト項目。

**困難度** 項目の困難度を表すパラメータ。推定値の横にヒストグラムが表示されます。

**識別力** (2PL と 3PL モデルのみ) 項目の識別力を表す  $a$  パラメータ。推定値の横にヒストグラムが表示されます。

**閾値** (3PL モデルのみ) 推量を表す  $c$  パラメータ。

## 「項目分析」プラットフォームのオプション

**横に並べるプロット数** 「特性曲線」レポートにおいて、各行に表示されるプロット数を指定することができます。デフォルトでは、行ごとに、1つのプロットが配置されます。

**能力計算式の保存** 能力の計算式を、データテーブルの新しい列に保存します。以下のオプションについて詳しくは、『JMPの使用法』の「JMPのレポート」章を参照してください。

**やり直し** 分析を繰り返すオプションや、やり直すオプションを含みます。また、[自動再計算] オプションを選択すると、このオプションに対応しているプラットフォームにおいて、データテーブルに加えた変更が、該当するレポートに即座に反映されるようになります。

**スクリプトの保存** レポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。

**By グループのスクリプトを保存** By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウで By 変数を指定した場合のみ使用可能です。

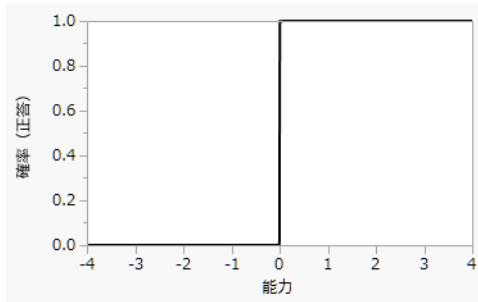
### 「項目分析」プラットフォームの統計的詳細

項目反応理論 (Item Response Theory; IRT) では、各受験者の（観測できない）能力を各項目に関連付ける数学モデルを仮定します。項目（テストの問題）に正解するかどうかは、直接観測することができない潜在的な能力から影響を受けていると仮定します。この数学モデルにおいて、各受験者の能力も、各項目の特性も、未知です。

## 項目反応曲線

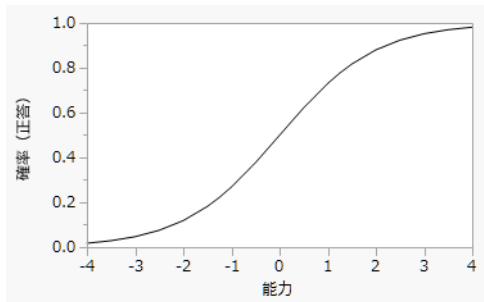
項目反応曲線; item response curve (項目特性曲線; item characteristic curve) は、能力と、各項目との関係を表したもので、項目反応曲線は、さまざまな能力レベルにおける項目の正答率をプロットしたもので、図 10.8 は、完全な識別力のある項目の項目反応曲線です。ある閾値を下回る能力を持つ受験者は正答率が 0% となっています。また、ある閾値を上回る能力を持つ受験者は正答率が 100% となっています。

図10.8 完全な識別力のある項目の特性曲線



より汎用的な項目反応曲線は、次図のようなS字型関数（シグモイド関数）で表されます。この関数では、受験者の能力が上がれば、その項目への正答率は100%まで上がります。この曲線の形状は、困難度や識別度によって変わってきます。

図10.9 一般的な項目反応曲線



## 項目反応曲線のモデル

項目反応理論モデルには、1、2、および3パラメータのロジスティックモデルがよく使われます。このうち、3パラメータのロジスティック（3PL）モデルは、次式のように定義されます。

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

- $P(\theta)$  は、能力レベルが  $\theta$  である受験者が、該当の項目に正答する確率です。項目反応理論モデルの推定に関する詳細は、「項目反応理論モデルの推定方法」（212ページ）を参照してください。
- $a$  パラメータは、変曲点での曲線の傾きです。項目の識別力（discriminatory power）を表しています。
- $b$  パラメータは、変曲点の位置です。項目の困難度（difficulty）を表しています。
- $c$  パラメータは下方漸近線です。これにより、推量による正答率の推定値が分かれます。

- 2PLモデルの場合、 $c$ パラメータは0に設定されます。

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

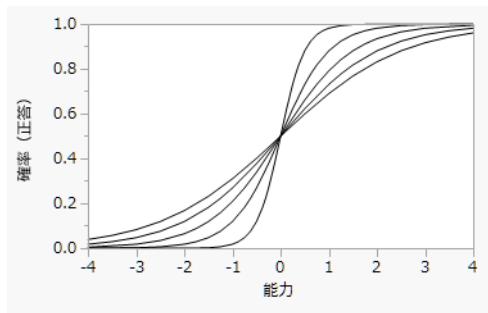
- 1PLモデルの場合、 $c$ パラメータは0に、 $a$ パラメータは1に固定されます。このパラメータ表現のモデルは、「Raschモデル」(Rasch 1980)とも呼ばれています。

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

### $a$ パラメータ：項目の識別力

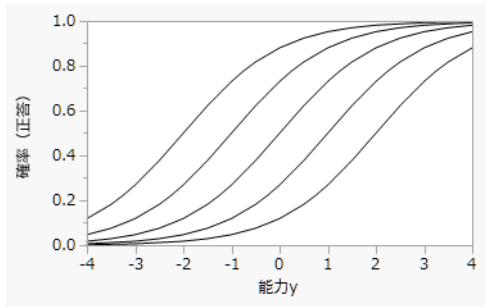
2PLモデルと3PLモデルでは、 $a$ パラメータ（変曲点での曲線の傾きを表すパラメータ）によって、項目の識別力（判別力）が分かれます。「識別力」とは、「能力レベルが低いものと、能力レベルが高いものとを、該当の項目がどれほどうまく識別できるか？」を示すものです。曲線の傾きが大きい項目は、強い識別力があることを示します。曲線の傾きが大きいと、能力レベルが低いとその項目に対する正答率が低くなり、逆に能力が高いと正答率は高くなる、ということを示しています。逆に、曲線が比較的平坦である項目は、識別力が低いです。識別力が低い項目は、テストから除外される候補になります。

図10.10  $a$ の値を変化させたときのロジスティックモデル



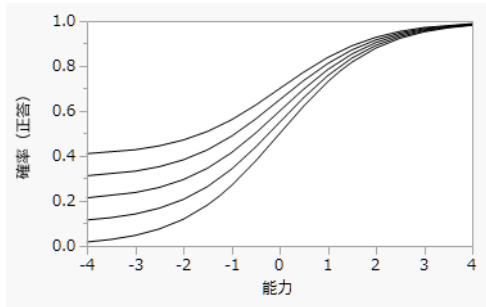
### $b$ パラメータ：項目の困難度

$b$ パラメータ（変曲点の位置を示すパラメータ）によって、項目の困難度が分かれます。項目反応曲線の変曲点が右側にある項目は、変曲点が左側にある項目よりも、難しいことを示します。1PLモデルと2PLモデルでは、 $b$ パラメータは、正答率が50%となる能力レベルを表しています。

図10.11  $b$ の値を変化させたときのロジスティック曲線

### $c$ パラメータ：項目の閾値

3PLモデルでは、 $c$ パラメータ（項目反応曲線の下方漸近線を示すパラメータ）は、あて推量で正答する確率を表します。下方漸近線が0ではない場合、能力レベルが非常に低い人が正答する確率が0ではないことを表しています。

図10.12  $c$ の値を変化させたときのロジスティックモデル

### 項目反応理論モデルの仮定

2PLモデルは、「項目分析」プラットフォームにおけるデフォルトのモデルです。1PLモデルは、すべての項目が等しい識別力を持つと仮定される場合に適しています。この仮定が適切でない場合は、2PLモデルまたは3PLモデルを使用する必要があります。2PLモデルは、データセットが小さい場合は特に、3PLモデルよりもかなり数値計算が安定しています。さらに、2PLモデルでは、 $b$ は、ある項目に対する正答率が50%となるのに必要な能力レベルとして解釈することができます。

項目反応理論モデルでは、能力は1次元であると仮定されます。つまり、潜在的特性は1つであると仮定されています。潜在的特性が複雑な交互作用を持つ複数のもので構成されている場合には、このような1次元のモデルは適しません。項目反応理論モデルで仮定される潜在的特性は、連続尺度です。潜在的特性としてカテゴリカルなものを仮定したい場合には、潜在クラスモデルを用いてください。潜在クラスモデルについては、『多変量分析』の「潜在クラス分析」章を参照してください。項目反応理論モデルは、項目不变性を仮定しています。「項目不变性」とは、 $P(\theta)$ を、「能力レベルが $\theta$ である受験者の正答率」と解釈することを意味します。言い換えれば、潜在的特性がまったく同じである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、その項目に正しく回答する受験者の割合を示します。また、項目反応理論モデルにおいては、受験者が変わっても、モデルのパラメータは同じままです。さらに、項目反応理論モデルでは、局所独立が仮定されます。「局所独立」とは、潜在的特性で条件付けた場合、各項目に正答するかどうかは、互いに独立であることを意味します。

## 項目反応理論モデルの推定方法

JMPでは、項目反応理論モデルの各パラメータは、周辺最尤推定 (Marginal Maximum Likelihood estimation; MMLE) によって推定されます。MMLEは、同時最尤推定 (Joint Maximum Likelihood estimation; JLE) の別の推定法です。MMLEでは、各受験者の違いを変量効果として扱います。項目と能力は、次式によって関連付けられます。

$$p(\mathbf{x}|\theta, \boldsymbol{\vartheta}) = \prod_{j=1}^L p_j(\theta)^{x_j} (1-p_j(\theta))^{1-x_j}$$

この式で、 $p(\mathbf{x}|\theta, \boldsymbol{\vartheta})$ は、受験者の能力を $\theta$ 、項目に関するパラメータのベクトルを $\boldsymbol{\vartheta}$ とした場合の、応答ベクトル $\mathbf{x}$ が生じる確率です。項目に関するパラメータの個数は、モデル (1PL, 2PL, 3PL) によって異なります。

MMLEでは、受験者の能力でこの尤度を積分します。JMPでは、積分において、ガウス求積法が使われています。そして、積分した尤度から、項目に関するパラメータを推定します。受験者の変量効果で積分した尤度は、次式のように表されます。

$$p(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{x}|\theta, \boldsymbol{\vartheta}) g(\theta|\mathbf{v}) d\theta$$

この式で、 $g(\theta|\mathbf{v})$ は、受験者の能力の分布を表しています。 $\mathbf{v}$ は、その分布の位置パラメータや尺度パラメータを表すベクトルです。JMPでは、この $g(\theta|\mathbf{v})$ に、平均が0で標準偏差が1の正規分布が仮定されています。

**メモ:** テスト問題にまったく答えなかった場合は、一般的に、すべての問題が誤答だったと考えられます。能力を推定する方法を次に述べますが、すべてが正解、または、すべてが不正解の受験者については、能力は推定できません。

JMPでの分析は、2段階で行われます。第1段階では、各受験者の能力が、平均0, 分散1の正規分布に従う変量効果として扱われます。この変量効果はガウス求積法によって積分されることにより尤度から消去されます。そして、項目に関するパラメータが、周辺尤度（受験者の能力が積分消去された尤度）において固定パラメータとして扱われます。第2段階において、能力パラメータが推定されます。この推定には、項目に関するパラメータを、第1段階で得られた推定値に固定し、同時最尤法で推定します。こうして得られた能力の推定値の基本的な考え方は、混合線形モデルにおけるBLUP (Best Linear Unbiased Prediction; 最良線形不偏予測値) に似ています。

$L$ 個の項目に対する応答には、 $2^L$ 個のパターンがあります。各応答パターンに対して、尤度が最大となるような能力が推定されます。それには、 $\theta$ が収束するまで次式を反復します。

$$\theta_i^{t+1} = \theta_i^t - \frac{X_i - \sum_{j=1}^L p_{ij}(t)}{\sum_{j=1}^L p_{ij}(t)(1 - p_{ij}(t))}$$

この式で、

$\theta$ は、推定したい受験者の能力（この値を、尤度が最大になるように推定する）

$t$ は、反復回数

$L$ は、項目数

$x_i$ は、観測された応答

$p_{ij}$ は、 $i$ 番目の受験者が $j$ 番目の項目に正答する確率

## 能力計算式

「項目分析」の赤い三角ボタンのメニューの〔確率の計算式の保存〕オプションを使うと、能力の計算式がデータテーブルの新しい列に保存されます。保存された計算式は、データテーブルに追加された新しい受験者の能力スコアを求めるのにも使用できます。また、計算式を新しい別のデータテーブルにコピーして、新たな受験者の能力スコアを求めることもできます。

データテーブルに保存される計算式は、IRT Abilityという関数が使われています。そして、項目に関するパラメータの推定値が、この関数の引数に、行列形式で保存されます。



# 付録 A

## 参考文献

---

- Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, New York: John Wiley and Sons, Inc.
- Akaike, H. (1974), "Factor Analysis and AIC," *Psychometrika*, 52, 317–332.
- Akaike, H. (1987), "A new Look at the Statistical Identification Model," *IEEE Transactions on Automatic Control*, 19, 716–723.
- Bartlett, M.S. (1954), "A Note on the Multiplying Factors for Various Chi Square Approximations," *Journal of the Royal Statistical Society, Series B*, 16 (Series B), 296–298.
- Benzécri, J. P. (1979), "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire, addendum et erratum à [BIN. MULT.]," *Cahiers de l'Analyse des Données*, 4, 377–378.
- Borg, I. and Groenen, P. J. F. (2005), *Modern Multidimensional Scaling: Theory and Applications*, Second Edition, New York: Springer.
- de Ayala, R. J. (2009), *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.
- 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.
- Greenacre, M. J. (1984), *Theory and Applications of Correspondence Analysis*, London: Academic Press.
- 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.

- Jackson, J.E. (2003), *A User's Guide to Principal Components*, New York: John Wiley and Sons, Inc.
- Kaiser, H.F. (1958), "The varimax criterion for analytic rotation in factor analysis" *Psychometrika*, 23, 187–200.
- Kish, L. (1965), *Survey Sampling*, New York: John Wiley and Sons, Inc.
- Lavassani, K.M., Movahedi, B., and Kumar, V. (2009). "Developments in Analysis of Multiple Response Survey Data in Categorical Data Analysis: The Case of Enterprise System Implementation in Large North American Firms," *Journal of Applied Quantitative Methods*, Vol. 4, No. 1, 45–53.
- Louviere, J.J., Flynn, T.N., and Marley, A.A. (2015), *Best-Worst Scaling: Theory, Methods and Applications*, Cambridge University Press.
- McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, pp. 105–142. See <http://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf>. Retrieved 4/25/2016.
- 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.
- Rasch, G. (1980), *Probabilistic models for some intelligence and attainment tests*, University of Chicago Press.
- Reichheld, F. F. (2003), "The One Number You Need to Grow," *Harvard Business Review*, Vol. 81 No. 12, 46–54.
- Rossi, P.E., Allenby, G.M., and McCulloch, R. (2006), *Bayesian Statistics and Marketing*, Wiley & Sons, Ltd.
- SAS Institute (2009), *SAS/STAT 9.2 User's Guide, Second Edition*, Chapter 92, SAS Institute Inc., Cary NC. See [https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\\_ttest\\_a0000000126.htm](https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_ttest_a0000000126.htm). Retrieved 7/14/2016.
- Train, K. E. (2001), "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," Department of Economics, University of California. See <https://eml.berkeley.edu/~train/compare.pdf>. Retrieved 11/7/2016.
- Train, K.E. (2009), *Discrete Choice Methods and Simulation*, Cambridge University Press.
- Waern, Y. (1972) "Structure in Similarity Matrices: A graphic approach," *Scandinavian Journal of Psychology*, Vol. 13, 5–16.
- 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

Burt表、多重対応分析 78

By変数 101

### I

IRT 28, 201

### M

MaxDiff

応答データ 179

効用計算式の保存 187

最良-最悪スケーリング 168

推定値の相関 187

全水準の比較レポート 187

被験者ごとの勾配を保存 187

ピザの選択の例 168–171

プロファイルデータ 178

MaxDiff、複合因子検定 186

### ア

アップリフトモデル 191

概要 192

プラットフォーム 194

レポート 195

レポートオプション 198

### イ

因子の数 103

因子分析プラットフォーム

By変数 101

重み変数 101

度数変数 101

### オ

応答データ、MaxDiff 179

応答の等質性に対する検定 37

重み変数 101

### 力

回転方法 103

各応答に対する検定 37

カテゴリカルプラットフォーム 29

起動ウィンドウ 31

自由回答 33

上位カテゴリ 45

表の構成 33

レポート 34

### ヰ

共通因子分析 97, 103

### ク

クロス表オプション、多重対応分析 74

### コ

項目特性曲線 205

項目反応理論 28, 201

「項目分析」 プラットフォーム

オプション 208

項目分析プラットフォーム 204

レポート 205

効用計算式の保存、MaxDiff 187

効用計算式の保存、選択モデルプラットフォーム 137

古典的テスト理論 201

固有値 105

**サ**

最良-最悪スケーリング 168

**シ**

支払意思額、選択モデルプラットフォーム 136  
 斜交回転 104  
 主成分分析 103  
 信頼区間、MaxDiff 187

**ス**

推定値の相関、MaxDiff 187  
 スクリープロット 106

**セ**

潜在特性 201  
 全水準の推定値、選択モデルプラットフォーム 135  
 全水準の比較レポート、MaxDiff 187  
 選択なしオプション、選択モデルプラットフォーム 126, 130  
 選択モデル 113-164  
 選択モデルプラットフォーム  
   効用計算式の保存 137  
   支払意思額 136  
   セグメント化 115, 168  
   全水準の推定値 135  
   選択なしオプション 126, 130  
   被験者ごとの勾配を保存 115, 137, 168  
   被験者データ 130  
   複合因子検定 135  
   プロファイルデータ 127

**タ**

対応分析のオプション 74  
 多重対応分析  
   Burt表 78  
   起動ウインドウオプション 72  
   クロス表オプション 74  
   追加ID変数 72  
   独立性の検定 74  
   プラットフォームのオプション 74  
 探索的因子分析 97

**チ**

直交回転 104

**ツ**

追加ID変数、多重対応分析 72

**ト**

独立性の検定、多重対応分析 74  
 度数変数 101

**ヒ**

比較、MaxDiff 187  
 被験者ごとの勾配を保存、MaxDiff 187  
 被験者ごとの勾配を保存、選択モデルプラットフォーム 115, 137, 168  
 被験者データ、選択モデルプラットフォーム 130  
 ピザの選択の例 148-151  
 ピザの選択の例、MaxDiff 168-171

**フ**

複合因子検定  
   maxdiff 186  
 複合因子検定、選択モデルプラットフォーム 135  
 プロファイルデータ、MaxDiff 178  
 プロファイルデータ、選択モデルプラットフォーム 127

**モ**

文字を使用した比較 38, 45

**ユ**

尤度比検定 135