



バージョン 13

基本的な統計分析

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

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 基本的な統計分析』 Cary, NC: SAS Institute Inc.

JMP® 13 基本的な統計分析

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、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.

2016年9月

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黙示的な保証を含め、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.
- 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、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、業務の中断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). 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、業務の中断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、業務の中断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、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、業務の中断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、サブライセンスする、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黙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から事前による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、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、業務の中断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. 事前による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、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、業務の中断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

基本的な統計分析

1 JMPの概要

マニュアルとその他のリソース	19
表記規則	20
JMPのマニュアル	21
JMP ドキュメンテーションライブラリ	21
JMP ヘルプ	27
JMPを習得するためのその他のリソース	27
チュートリアル	27
サンプルデータテーブル	28
統計用語とJSL用語の習得	28
JMPを使用するためのヒント	28
ツールヒント	29
JMP User Community	29
JMPer Cable	29
JMP 関連書籍	29
「JMP スターター」 ウィンドウ	30
テクニカルサポート	30

2 基本的な統計分析について

基本的な分析手法の概要	31

3 一変量の分布

「一変量の分布」プラットフォームを使用する	33
「一変量の分布」プラットフォームの概要	34
カテゴリカル変数	34
連続尺度の変数	34
「一変量の分布」プラットフォームの例	34
「一変量の分布」プラットフォームの起動	36
「一変量の分布」レポート	37
ヒストグラム	38
「度数」レポート	41
「分位点」レポート	42
「要約統計量」レポート	42
「一変量の分布」プラットフォームのオプション	44
カテゴリカル変数のオプション	45

カテゴリーカル変数の「表示オプション」サブメニュー	45
カテゴリーカル変数の「ヒストグラムオプション」サブメニュー	46
カテゴリーカル変数の「保存」サブメニュー	46
連続変数のオプション	47
連続変数の「表示オプション」サブメニュー	47
連続変数の「ヒストグラムオプション」サブメニュー	48
正規分位点プロット	49
外れ値の箱ひげ図	50
分位点の箱ひげ図	51
幹葉図	52
累積確率プロット	52
平均の検定	53
標準偏差の検定	54
同等性の検定	55
連続変数における信頼区間	56
連続変数における保存コマンド	56
予測区間	57
許容区間	57
工程能力分析	58
分布のあてはめ	60
連続分布のあてはめ	60
離散分布のあてはめ	61
あてはめた分布のオプション	61
「一変量の分布」プラットフォームのその他の例	64
複数のヒストグラムにおけるデータ選択の例	64
「割合の検定」オプションの例	65
予測区間の例	67
許容区間の例	69
工程能力分析の例	70
「一変量の分布」プラットフォームの統計的詳細	71
標準誤差のバー	71
分位点	72
要約統計量	72
正規分位点プロット	73
Wilcoxonの符号付順位検定の統計量	74
標準偏差の検定	76
正規分位点	76
標準化データの保存	76
予測区間	76
許容区間	77
工程能力分析	78
連続分布のあてはめ	82
離散分布のあてはめ	88

あてはめた分布の分位点	90
あてはめた分布のオプション	91
4 「二変量の関係」プラットフォームの概要	
2変数の関係を調べる	93
「二変量の関係」プラットフォームの概要	94
「二変量の関係」プラットフォームの起動	94
「JMP スターター」ウィンドウからの各分析の実行	95
5 二変量分析	
2つの連続変数の関係を調べる	97
二変量分析の例	98
「二変量」プラットフォームの起動	98
二変量に対する散布図	99
あてはめ用コマンドとオプション	100
あてはめのオプション	101
あてはめ用コマンドの種類	102
同じあてはめ用コマンドの繰り返し	103
ヒストグラム軸	103
平均のあてはめ	103
「平均のあてはめ」レポート	104
直線のあてはめと多項式のあてはめ	104
「直線のあてはめ」と「多項式のあてはめ」のレポート	105
その他のあてはめ	111
その他のあてはめのレポートとメニュー	111
ノンパラメトリックな曲線	112
スプライン曲線のあてはめ	112
カーネル平滑化	113
値ごとのあてはめ	114
直交のあてはめ	115
「直交のあてはめ」のオプション	115
「直交回帰」レポート	115
ロバスト	116
ロバストなあてはめ	116
Cauchy のあてはめ	116
確率楕円	117
「相関」レポート	117
ノンパラメトリック密度	118
「ノンパラメトリック二変量の密度」レポート	119
グループ別	119
あてはめ用メニュー	120
あてはめ用メニューのオプション	120
診断プロット	123

「二変量」プラットフォームのその他の例	123
「その他のあてはめ」 コマンドの例	123
[直交のあてはめ] コマンドの使用例	125
[ロバストなあてはめ] コマンドの使用例	127
グループ別確率楕円の例	129
グループ別回帰直線の例	130
「二変量」プラットフォームの統計的詳細	131
直線のあてはめ	131
スプライン曲線のあてはめ	131
直交のあてはめ	131
「あてはめの要約」 レポート	132
「あてはまりの悪さ (LOF)」 レポート	132
「パラメータ推定値」 レポート	133
「平滑化スプライン曲線によるあてはめ」 レポート	133
「相関」 レポート	134

6 一元配置分析

連続尺度のY変数とカテゴリカルなX変数の関係を調べる	135
一元配置分析の概要	136
一元配置分析の例	136
「一元配置」プラットフォームの起動	138
一元配置プロット	138
「一元配置」プラットフォームのオプション	139
表示オプション	142
分位点	144
外れ値の箱ひげ図	144
平均/ANOVA および平均/ANOVA/プーリングしたt検定	145
「あてはめの要約」 レポート	145
「t検定」 レポート	146
「分散分析」 レポート	147
「各水準の平均」 レポート	148
「ブロック平均」 レポート	148
平均のひし形と標本サイズに比例したX軸	148
平均線、平均誤差バー、標準偏差線	149
平均分析法	150
位置の指標に対する平均分析	150
尺度の指標に対する平均分析	151
平均分析のグラフ	152
平均分析のオプション	152
平均の比較	153
比較円の使用	155
各ペア, Studentのt検定	156
すべてのペア, TukeyのHSD検定	156

最適値との比較(HsuのMCB)	157
コントロール群との比較(Dunnett)	158
「平均の比較」のオプション	158
ノンパラメトリック検定	159
Wilcoxon検定、メディアン検定、Van der Waerden検定のレポート	160
Kolmogorov-Smirnov二標本検定のレポート	161
ノンパラメトリックな多重比較	162
等分散性の検定	165
「分散が等しいことを調べる検定」レポート	166
同等性の検定	167
ロバスト	167
ロバストなあてはめ	168
Cauchyのあてはめ	168
検出力	168
「検出力の詳細」ウィンドウとレポート	169
正規分位点プロット	170
累積確率プロット	170
密度	170
対応のある列を設定	171
「一元配置」プラットフォームのその他の例	171
平均分析法の例	171
分散に対する平均分析法の例	172
[各ペア, Studentのt検定]の例	173
[すべてのペア, TukeyのHSD検定]の例	175
[最適値との比較(HsuのMCB)]の例	177
[コントロール群との比較(Dunnett)]の例	178
「平均の比較」で用意されているすべての検定	180
ノンパラメトリックなWilcoxon検定の例	180
等分散性検定の例	183
同等性の検定の例	184
[ロバストなあてはめ] オプションの例	185
[検出力] オプションの例	186
正規分位点プロットの例	188
累積確率プロットの例	189
[密度]の各オプションの例	190
[対応のある列を設定] オプションの例	191
データを積み重ねて一元配置分析を実行する例	193
「一元配置」プラットフォームの統計的詳細	198
比較円	198
検出力	200
「あてはめの要約」レポート	200
「分散が等しいことを調べる検定」レポート	201
ノンパラメトリックな検定の統計量	202

7 分割表分析

2つのカテゴリカルな変数の関係を調べる	205
分割表分析の例	206
「分割表」プラットフォームの起動	207
「分割表」レポート	208
「分割表」プラットフォームのオプション	209
モザイク図	210
ポップアップメニュー	211
分割表	212
「分割表」レポートの説明	213
検定	214
「検定」レポートの説明	214
Fisherの正確検定	215
割合の平均分析	215
対応分析	215
対応分析プロットの理解	216
「対応分析」のオプション	216
「詳細」レポート	216
Cochran-Mantel-Haenszel 検定	216
一致性の統計量	217
相対リスク	218
割合の2標本検定	218
関連の指標	218
Cochran-Armitage の傾向検定	220
正確検定	220
「分割表」プラットフォームのその他の例	221
割合の平均分析の例	221
対応分析の例	222
Cochran-Mantel-Haenszel 検定の例	224
[一致性の統計量] オプションの例	226
[相対リスク] オプションの例	227
割合の2標本検定の例	228
[関連の指標] オプションの例	229
Cochran-Armitage の傾向検定の例	230
「分割表」プラットフォームの統計的詳細	231
[一致性の統計量] オプション	231
[オッズ比] オプション	231
「検定」レポート	232
対応分析の「詳細」レポート	232

8 ロジスティック分析

カテゴリーなY変数と連続尺度のX変数の関係を調べる	235
ロジスティック回帰の概要	236
名義ロジスティック回帰	236
順序ロジスティック回帰	236
名義ロジスティック回帰の例	237
「ロジスティック」プラットフォームの起動	238
データ構造	239
「ロジスティック」レポート	239
ロジスティックプロット	240
反復履歴	240
モデル全体の検定	240
パラメータ推定値	242
「ロジスティック」プラットフォームのオプション	243
ROC曲線	244
確率の計算式の保存	245
逆推定	245
ロジスティック回帰のその他の例	245
順序ロジスティック回帰の例	245
ロジスティックプロットの例2	247
ROC曲線の例	249
十字ツールを使用した逆推定の例	250
[逆推定] オプションを使用した逆推定の例	251
「ロジスティック」プラットフォームの統計的詳細	253
「モデル全体の検定」レポート	253

9 表の作成

対話的に要約テーブルを作成する	255
「表の作成」プラットフォームの使用例	256
「表の作成」プラットフォームの起動	260
ダイアログボックスの使用	261
統計量の追加	262
「表の作成」の出力	265
分析列	266
グループ列	266
列テーブルと行テーブル	267
表の編集	268
「表の作成」プラットフォームのオプション	269
テスト集計パネルの表示	270
列の右クリックメニュー	270
「表の作成」プラットフォームのその他の例	271
複数の列を持つテーブルの例	274

ページ列の使用例	276
10 シミュレーション	
パラメトリックな標本再抽出によるシミュレーション	279
シミュレーションの概要	280
シミュレーションの例	280
分散成分に対する信頼区間の計算	281
並び替え検定の実行	286
一般化回帰モデルに含める説明変数の検討	288
非線形モデルにおける検出力の事前計算	293
「シミュレーション」ウィンドウの起動	301
「シミュレーション」ウィンドウ	302
シミュレーション結果のテーブル	302
「シミュレーションの結果」レポート	303
「検出力のシミュレーション」レポート	303
11 ブートストラップ	
標本再抽出によって統計量の分布を近似する	305
ブートストラップの概要	306
ブートストラップの例	307
「ブートストラップ」ウィンドウのオプション	309
結果を積み重ねたデータテーブル	310
結果を積み重ねていないデータテーブル	311
作成した統計量データの分析	312
ブートストラップの別例	313
ブートストラップの統計的詳細	318
小数の重みの計算	318
バイアス修正済みの信頼限界	318
12 テキストエクスペローラ	
非構造化テキストを探索する	321
「テキストエクスペローラ」プラットフォームの概要	322
テキスト処理の手順	323
「テキストエクスペローラ」プラットフォームの例	324
「テキストエクスペローラ」プラットフォームの起動	327
Regexのカスタマイズ：正規表現エディタ	329
「テキストエクスペローラ」レポート	334
要約レポート	334
単語と句のリスト	335
「テキストエクスペローラ」プラットフォームのオプション	337
テキストの前処理オプション	337
テキストの分析オプション	342
保存オプション	344

レポートのオプション	346
潜在クラス分析	347
潜在意味分析（特異値分解）	348
「特異値分解プロット」レポート	349
トピックの分析	349
「トピック 単語」レポート	349
「トピックのスコア」レポート	350
「テキストエクスプローラ」プラットフォームの別例	350
A 統計的詳細	
基本的な統計分析	355
検証の機能を使用できるプラットフォーム	356
B 参考文献	
索引	
基本的な統計分析	361

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	[分析] メニューからアクセスできる以下のプラットフォームの説明： • 一変量の分布 • 二変量の関係 • 表の作成 • テキストエクスプローラ [分析] > [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。

マニュアル	目的	内容
『グラフ機能』	データに合った理想的なグラフを見つける	[グラフ] メニューからアクセスできる以下のプラットフォームの説明： • グラフビルダー • 重ね合わせプロット • 三次元散布図 • 等高線図 • バブルプロット • パラレルプロット • セルプロット • ツリーマップ • 散布図行列 • 三角図 • チャート このマニュアルには背景マップやカスタムマップの作成方法も記載されています。
『プロファイル機能』	対話式のプロファイルツールの使い方を学ぶ。任意の応答曲面の断面を表示できるようになります。	[グラフ] メニューに表示されるすべてのプロファイルについて。誤差因子の分析が、ランダム入力を使用したシミュレーションの実行とともに含まれています。
『実験計画(DOE)』	実験の計画方法と適切な標本サイズの決定方法を学ぶ	[実験計画 (DOE)] メニューと [分析] > [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	[分析] メニューの「モデルのあてはめ」プラットフォームで利用できる、以下の手法の説明： - 標準最小2乗 - ステップワイズ - 一般化回帰 - 混合モデル - MANOVA - 対数線形-分散 - 名義ロジスティック - 順序ロジスティック - 一般化線形モデル

マニュアル	目的	内容
『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	[分析] > [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： - モデル化ユーティリティ - ニューラル - パーティション - ブートストラップ森 - ブースティングツリー - K近傍法 - 単純Bayes - モデルの比較 - 計算式デポ [分析] > [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： - 曲線のあてはめ - 非線形回帰 - Gauss 過程 - 時系列分析 - 対応のあるペア [分析] > [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： - 応答のスクリーニング - 工程のスクリーニング - 説明変数のスクリーニング - アソシエーション分析 [分析] > [発展的なモデル] > [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。

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	[分析] > [多変量] メニューで利用できる以下のプラットフォームの説明： • 多変量の相関 • 主成分分析 • 判別分析 • PLS [分析] > [クラスター分析] メニューで利用できる以下のプラットフォームの説明： • 階層型クラスター分析 • K Means クラスター分析 • 正規混合 • 潜在クラス分析 • 変数のクラスタリング
『品質と工程』	工程を評価し、向上させるためのツールについて理解を深める	[分析] > [品質と工程] メニューで利用できる以下のプラットフォームの説明： • 管理図ビルダーと個々の管理図 • 測定システム分析 • 計量値/計数値ゲージチャート • 工程能力 • パレート図 • 特性要因図

マニュアル	目的	内容
『信頼性/生存時間分析』	製品やシステムにおける信頼性を評価し、向上させる方法、および人や製品の生存時間データを分析する方法について学ぶ	[分析] > [信頼性/生存時間分析] メニューで利用できる以下のプラットフォームの説明： • 寿命の一変量 • 寿命の二変量 • 累積損傷 • 再生モデルによる分析 • 劣化分析と破壊劣化 • 信頼性予測 • 信頼性成長 • 信頼性ブロック図 • 修理可能システムのシミュレーション • 生存時間分析 • 生存時間(パラメトリック)のあてはめ • 比例ハザードのあてはめ
『消費者調査』	消費者選好を調査し、その洞察を使用してより良い製品やサービスを作成するための方法を学ぶ	[分析] > [消費者調査] メニューで利用できる以下のプラットフォームの説明： • カテゴリカル • 多重対応分析 • 多次元尺度構成 • 因子分析 • 選択モデル • MaxDiff • アップリフト • 項目分析
『スクリプトガイド』	パワフルなJMPスクリプト言語 (JSL) の活用方法について学ぶ	スクリプトの作成やデバッグ、データテーブルの操作、ディスプレイボックスの構築、JMPアプリケーションの作成など。
『スクリプト構文リファレンス』	JSL 関数、その引数、およびオブジェクトやディスプレイボックスに送信するメッセージについて理解を深める	JSL コマンドの構文、例、および注意書き。

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」](#) (27 ページ) を参照)
- サンプルデータ ([「サンプルデータテーブル」](#) (28 ページ) を参照)
- 索引 ([「統計用語と JSL 用語の習得」](#) (28 ページ) を参照)
- 使い方ヒント ([「JMP を使用するためのヒント」](#) (28 ページ) を参照)
- Web リソース ([「JMP User Community」](#) (29 ページ) を参照)
- 専門誌『JMPer Cable』([「JMPer Cable」](#) (29 ページ) を参照)
- JMP に関する書籍 ([「JMP 関連書籍」](#) (29 ページ) を参照)
- JMP スターター ([「JMP スターター」 ウィンドウ](#) (30 ページ) を参照)
- 教育用リソース ([「サンプルデータテーブル」](#) (28 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://community.jmp.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

「JMP スターター」 ウィンドウ

JMP またはデータ分析にあまり慣れていないユーザは、「JMP スターター」ウィンドウから開始するとよいでしょう。カテゴリ分けされた項目には説明がついており、ボタンをクリックするだけで該当の機能を起動できます。「JMP スターター」ウィンドウには、[分析]、[グラフ]、[テーブル]、および [ファイル] メニュー内の多くのオプションがあります。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 スターター」ウィンドウを開くには、[表示] (Macintosh では [ウィンドウ]) > [JMP スターター] を選択します。
- Windows で JMP の起動時に自動的に「JMP スターター」を表示するには、[ファイル] > [環境設定] > [一般] を選び、「開始時の JMP ウィンドウ」リストから [JMP スターター] を選択します。Macintosh では、[JMP] > [環境設定] > [起動時に JMP スターターウィンドウを表示する] を選択します。

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

基本的な統計分析について 基本的な分析手法の概要

本書では、JMPでよく使用される基本的な分析機能について解説します。

- 「一変量の分布」プラットフォームでは、1変数の分布を分析できます。ヒストグラムを表示し、平均や標準偏差などをレポートします。データがどのように分布しているかわかれば、さらに具体的な分析を行い、検討を進めることができます。第3章「一変量の分布」を参照してください。
- 「二変量の関係」プラットフォームでは、X変数とY変数の尺度によって、自動的に分析が選択されます。第4章「二変量の関係」プラットフォームの概要」を参照してください。次の4種類の分析を実行できます。
 - 「二変量」プラットフォームでは、2つの連続変数における関係を分析します。第5章「二変量分析」を参照してください。
 - 「一元配置」プラットフォームでは、X変数のグループごとに、連続尺度のY変数の分布がどのように異なるかを分析します。第6章「一元配置分析」を参照してください。
 - 「分割表」プラットフォームでは、X変数のグループごとに、カテゴリカルなY変数の分布がどのように異なるかを分析します。第7章「分割表分析」を参照してください。
 - 「ロジスティック」プラットフォームでは、カテゴリカルな応答変数 (Y) の確率を、連続尺度の説明変数 (X) にあてはめます。第8章「ロジスティック分析」を参照してください。
- 「表の作成」プラットフォームでは、記述統計量のテーブルを対話形式で作成できます。第9章「表の作成」を参照してください。
- パラメトリックおよびノンパラメトリックの乱数シミュレーションを柔軟に行えるようになりました。第10章「シミュレーション」を参照してください。
- ブートストラップ分析は、データから何度もサンプリングすることによって統計量の標本分布を近似する手法です。反復ごとに、データは復元抽出（重複抽出）され、抽出されたデータから統計量が計算されます。この計算を反復して、統計量の分布を求めます。第11章「ブートストラップ」を参照してください。
- 「テキストエクスプローラ」プラットフォームでは、書式が設定されていないテキストデータの分類や分析を実行できます。分析前の単語や句の取り出しに、正規表現を使用することもできます。第12章「テキストエクスプローラ」を参照してください。

第3章

一変量の分布

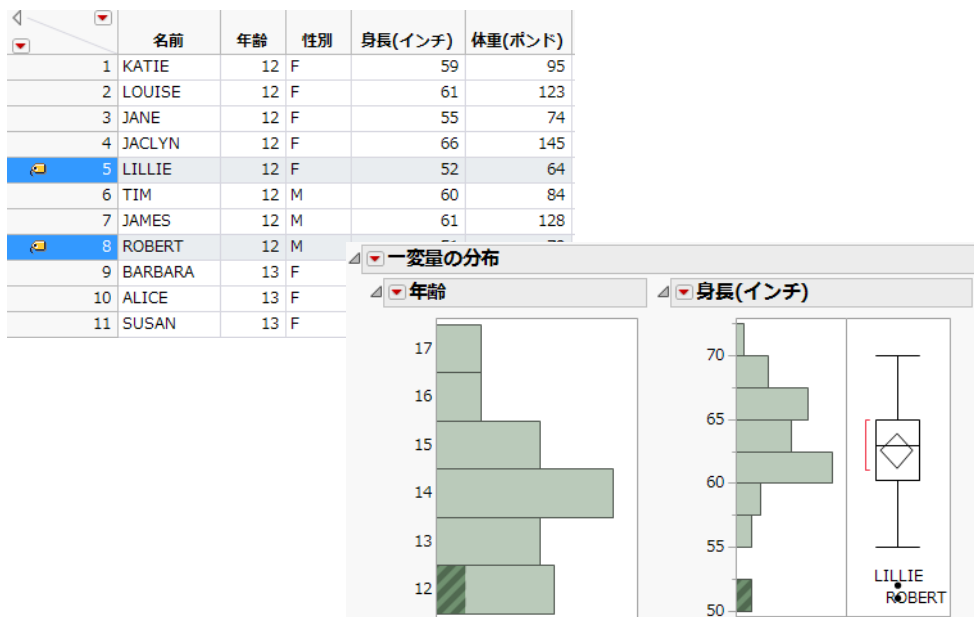
「一変量の分布」プラットフォームを使用する

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

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

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

図3.1 「一変量の分布」プラットフォームの例



「一変量の分布」プラットフォームの概要

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

カテゴリカル変数

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

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

連続尺度の変数

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

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

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

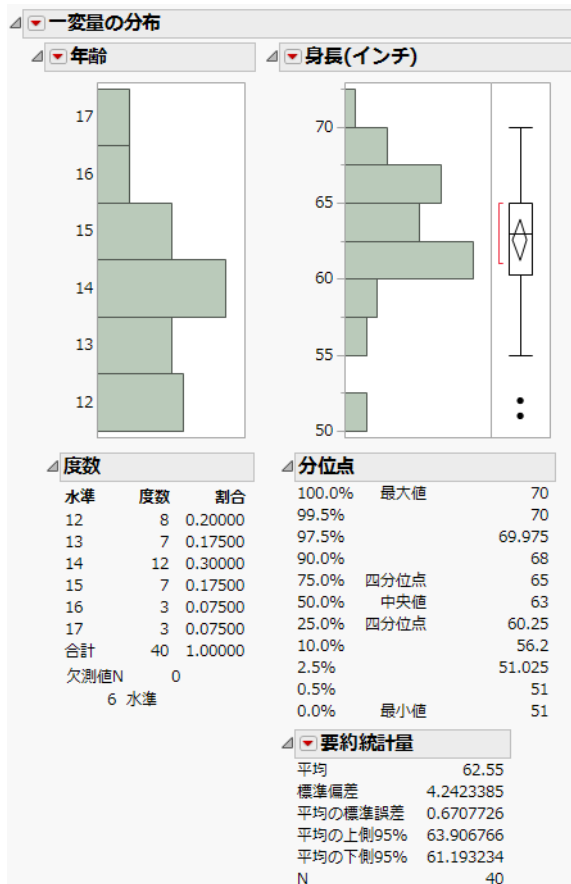
- 順位、確率スコア、正規分位点などをデータテーブルの新しい列に保存する。
- ある定数に対して、列の平均や標準偏差を検定する。
- さまざまな確率分布をあてはめたり、ノンパラメトリックな密度曲線を求めたりする。
- 品質管理を目的とした工程能力分析を実施する。
- 信頼区間、予測区間、許容区間。

「一変量の分布」プラットフォームの例

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

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

図3.2 「一変量の分布」プラットフォームの例



ヒストグラムを見ると、次のことがわかります。

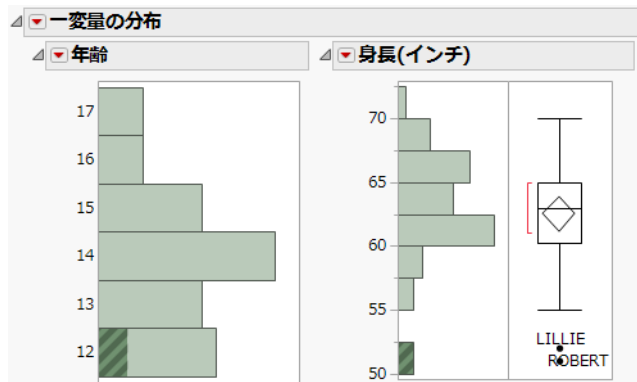
- 年齢の分布は一樣ではありません。
 - 「身長(インチ)」のヒストグラムには、離れている点が2つあります（外れ値である可能性があります）。
- 「身長(インチ)」のヒストグラムで「50」の棒をクリックし、外れ値の可能性が高い点について検討します。
- 「年齢」のヒストグラムで、対応する年齢が強調表示されます。外れ値の年齢は12であることがわかります。
 - データテーブル内では、選択した棒に対応する行が強調表示されます。外れ値の名前は Lillie と Robert であることがわかります。

「身長(インチ)」のヒストグラムの外れ値にラベルを追加します。

1. 両方の外れ値を選択します。
 2. いずれかの外れ値を右クリックして「行ラベル」を選択します。
- データテーブルでは、該当行にラベルアイコンが追加されます。

3. 必要に応じて、ラベル全体が表示されるように箱ひげ図の幅を広げてください。

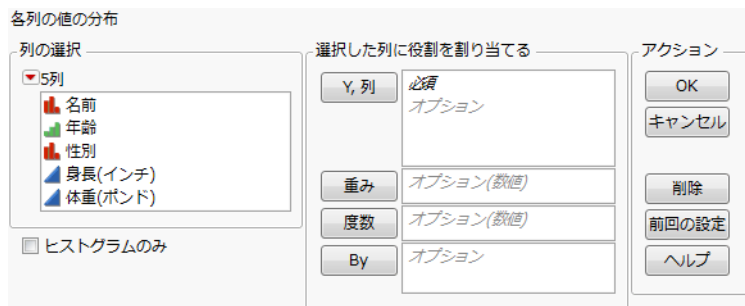
図3.3 ラベルを付けた外れ値



「一変量の分布」プラットフォームの起動

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

図3.4 「一変量の分布」の起動ウィンドウ



Y, 列 分析の対象となる変数を割り当てます。選択した変数ごとに、ヒストグラムと、いくつかのレポートが表示されます。

重み 連続尺度のYに適用する重み変数を割り当てます。Yがカテゴリカル変数である場合は、重みは適用されません。平均や標準偏差などの統計量は、重みを考慮して計算されます。

度数 度数の変数を指定します。データがすでに要約されている場合に便利です。度数の変数には、Y値の発生頻度を含む変数を指定してください。度数の合計は、「要約統計量」レポートに全体の度数（[N]）として表示されます。「要約統計量」レポートに表示されるすべての統計量（平均や標準偏差など）が、**度数**を考慮して計算されます。

By By変数の水準ごとに個別のレポートが作成されます。複数のBy変数を割り当てた場合、それらのBy変数の水準の組み合わせごとに個別のレポートが作成されます。

ヒストグラムのみ レポートウィンドウに、ヒストグラムだけを表示し、それ以外の情報は表示しません。

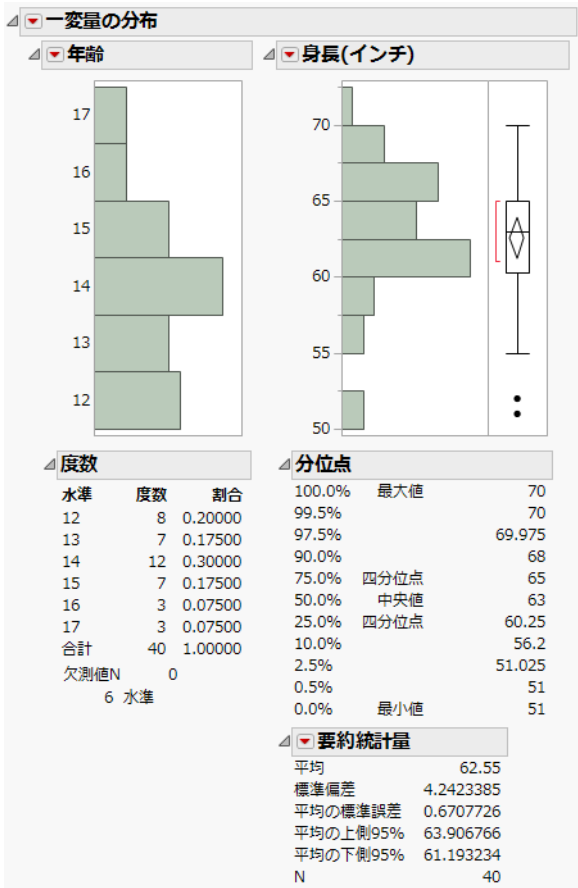
起動ウィンドウ全般については、『JMPの使用法』の「はじめに」章を参照してください。

[OK] をクリックすると、「一変量の分布」レポートウィンドウが開きます。[「一変量の分布」レポート](#) (37ページ) を参照してください。

「一変量の分布」レポート

[「一変量の分布」プラットフォームの例](#) (34ページ) の手順に従って作業すると、図3.5のレポートが作成されます。

図3.5 最初に表示される「一変量の分布」レポートウィンドウ



メモ: データテーブルの行を「表示しない」に設定すると、プロット上に点としては表示されませんが、ヒストグラムはこれらの行も含まれたままになります。ヒストグラムから除外し、分析の対象外にするには、行を「除外する」に設定したうえで、「一変量の分布」の横にある赤い三角ボタンのメニューから「やり直し」>「分析のやり直し」を選択してください。

「一変量の分布」レポートウィンドウには、変数ごとにヒストグラムとレポートが表示されます。次の点を念頭に置いてください。

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

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

- 「一変量の分布」の横にある赤い三角ボタンをクリックするとメニューが開き、レポートウィンドウ内のすべての変数に適用されるオプションが表示されます。「[「一変量の分布」プラットフォームのオプション](#)」(44 ページ) を参照してください。
- 各変数名の横にある赤い三角ボタンをクリックするとメニューが開き、該当する変数にのみ適用されるオプションが表示されます。「[カテゴリカル変数のオプション](#)」(45 ページ) または「[連続変数のオプション](#)」(47 ページ) を参照してください。Ctrl キーを押したまま変数のオプションを選択すると、尺度が同じすべての変数にそのオプションが適用されます。
- ヒストグラムはデータを視覚的に表示します。「[ヒストグラム](#)」(38 ページ) を参照してください。
- カテゴリカル変数のレポートには、「度数」レポートが含まれます。「[「度数」レポート](#)」(41 ページ) を参照してください。
- 連続変数のレポートには、「分位点」レポートと「要約統計量」レポートが含まれます。「[「分位点」レポート](#)」(42 ページ) および「[要約統計量 レポート](#)」(42 ページ) を参照してください。

ヒストグラム

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

データの強調表示 ヒストグラムの棒をクリックすると、対応する行がデータテーブル内で強調表示されます。他にもヒストグラムがある場合は、そのヒストグラムでも対応する部分が強調表示されます。「[棒の強調表示と行の選択](#)」(40 ページ) を参照してください。

サブセットの作成 ヒストグラムの棒をダブルクリックするか、ヒストグラムの棒を右クリックして「サブセット」を選択すると、該当するデータだけを抽出した新しいデータテーブルが作成されます。

ヒストグラム全体のサイズの変更 ヒストグラムの枠線の上にマウスを置くと、カーソルの形が双方向の矢印に変わります。カーソルの形が変化したら、枠線をクリックしてドラッグすると、グラフのサイズが変更されます。

軸の再スケール（連続尺度の変数のみ）軸をクリックしてドラッグすると、スケールが変更されます。

または、軸の上にマウスを置くと、手のひらの形にカーソルが変わります。カーソルの形が変化したら、軸をダブルクリックし、「軸の指定」ウィンドウでスケールを設定しても、スケールが変更されます。

ヒストグラムにおける棒の幅の変更（連続尺度の変数のみ）ヒストグラムの棒の幅は、複数の方法で変更できます。「[連続変数のヒストグラムにおける棒の幅の変更](#)」（39ページ）を参照してください。

棒の選択方法 複数のヒストグラムにおける棒を選択するには、いくつかの方法があります。「[複数のヒストグラムにまたがる選択](#)」（41ページ）を参照してください。

ヒストグラムに関するオプションを表示するには、次の操作を実行します。

- ヒストグラム上で右クリックします。詳細については、『JMPの使用法』を参照してください。
- 軸を右クリックします。軸ラベルを追加したり、軸を編集したりできます。詳細については、『JMPの使用法』の「JMPのレポート」章を参照してください。
- 変数名の横にある赤い三角ボタンをクリックし、[ヒストグラムオプション] を選択します。表示されるオプションは、変数の尺度によって若干異なります。「[カテゴリカル変数のオプション](#)」（45ページ）または「[連続変数のオプション](#)」（47ページ）を参照してください。

連続変数のヒストグラムにおける棒の幅の変更

連続変数に対するヒストグラムでは、次を使用して棒の幅を変更できます。

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

手のひらツールの使用

手のひらツールを使用すると、データをすばやく調べることができます。

1. [ツール] > [手のひらツール] を選択します。

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

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

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

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

【棒の幅の設定】オプションの使用

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

【目盛り間隔】オプションの使用

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

棒の強調表示と行の選択

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

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

図3.6 棒と行の強調表示



棒を選択すると、データテーブルの行と他の出力の対応部分が強調表示される。

複数のヒストグラムにまたがる選択

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

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

関連情報

- 「[複数のヒストグラムにおけるデータ選択の例](#)」（64 ページ）

「度数」レポート

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

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

ヒント：「度数」レポートの値をクリックすると、ヒストグラムとデータテーブルの対応するデータが選択状態になります。

水準 応答変数の値をリストしたもの。

度数 応答変数の水準ごとの行数をリストしたもの。ただし、度数変数を指定した場合、度数変数の値の合計が表示されます。

割合 応答変数の水準ごとに、その割合（確率）をリストしたもの。割合は、その水準の度数を全体の度数（表の下端に表示されている）で割ったものです。

割合の標準誤差 割合の標準誤差をリストしたもの。デフォルトでは、この列は非表示になっています。表示するには、度数表内を右クリックし、[列] > [割合の標準誤差] を選択します。

累積割合 割合の累積合計が表示されます。デフォルトでは、この列は非表示になっています。表示するには、度数表内を右クリックし、[列] > [累積割合] を選択します。

「分位点」レポート

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

関連情報

- 「[分位点](#)」(72 ページ)

「要約統計量」レポート

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

ヒント：「一変量の分布」レポートに表示されるデフォルトの要約統計量を設定するには、[ファイル] > [環境設定] > [プラットフォーム] > [一変量の要約統計量] を選択し、表示したい項目を選択します。

- 「[「要約統計量」レポートの説明](#)」に、デフォルトで表示される統計量を示します。
- 「[その他の要約統計量](#)」に、「[要約統計量のカスタマイズ](#)」ウィンドウで追加できる統計量を示します。

「要約統計量」レポートの説明

平均 算術平均。標本平均。平均は、応答変数の母集団分布における期待値の推定値です。非欠測値に関して、それらの合計を個数で割って求めます。

標準偏差 データのばらつきを示す指標。なお、正規分布は、平均と標準偏差だけによって定義されます。標本サイズが大きく、データが正規分布に従っている場合、これら2つのパラメータによって、次のようにデータを簡単に要約することができます。

- データの68%は、平均から ± 1 標準偏差の区間内にあります。
- データの95%は、平均から ± 2 標準偏差の区間内にあります。
- データの99.7%は、平均から ± 3 標準偏差の区間内にあります。

平均の標準誤差 平均（標本平均）の標準偏差の推定値。

平均の上側95%信頼限界と平均の下側95%信頼限界 平均に関する両側95%信頼区間。「この区間に、母平均が含まれている」という判断の信頼性が高い区間を表します。

N 非欠測値の個数。

その他の要約統計量

重みの合計 （起動ウィンドウで）[重み]の役割を割り当てられた列の合計。平均を計算する際、Nの代わりに分母として使用されます。

合計 応答の値の合計。

分散 不偏分散。標準偏差の2乗。

歪度 歪みや対称性に関する指標。

尖度 尖りや裾の重さに関する指標。式の詳細は、「**尖度**」（73ページ）を参照してください。

変動係数 単位がパーセントの変動係数。標準偏差を平均で割り、100を掛けた値です。変動係数は相対的な変動の評価に効果的です。たとえば、異なる単位で測定されたデータの変動を比較するときに使用できます。

欠測値N 欠測値の個数。

ゼロの個数 0の値の数。

一意な値の個数 一意（ユニーク）な値の個数。

無修正平方和 修正を加えていない平方和。データそのものの2乗和。

修正平方和 修正を加えた平方和。平均からの偏差の2乗和。

自己相関 （[度数]変数を指定していない場合のみ表示されます。）1次の自己相関。行と、その1つ前の行との間に、相関が見られるかどうかを示します。データにおける系列相関を検出するのに役立ちます。

最小値 0パーセント点。

最大値 100パーセント点。

中央値 メディアン。50パーセント点。

最頻値 モード。データ内で最も出現頻度が高い値。最頻値が複数ある場合、その中の最小値が表示されます。

トリム平均 データの大きい方から p パーセントと小さい方から p パーセントを除外して計算した平均値。
 p の値は、ウィンドウの最下部にある「トリム平均に用いるパーセントを入力」に入力される値です。「重み」変数を指定した場合は、「トリム平均」オプションは表示されません。

幾何平均 データの積の n 乗根。ゼロまたは負の数値は欠測値として扱われます。たとえば、異なる範囲・指標に基づく2社のデータを比較したり、歪んだ分布に大きな値が含まれる場合などに、便利です。

範囲 データの最大値と最小値の差。

四分位範囲 第3四分位点と第1四分位点の差。

中央絶対偏差 ([重み] 変数を指定した場合は表示されません。) 中央値からの偏差の絶対値の中央値。

ロバスト 平均 ロバストな平均値。外れ値の影響を受けにくいHuberのM推定で求められた平均値です。Huber and Ronchetti (2009) を参照してください。

ロバスト 標準偏差 ロバストな標準偏差。外れ値の影響を受けにくいHuberのM推定で求められた標準偏差です。Huber and Ronchetti (2009) を参照してください。

平均の信頼区間に対する $(1-\alpha)$ を入力 平均の信頼区間に対する α 水準を指定します。

トリム平均に用いるパーセントを入力 トリム平均に使用するパーセントを指定します。指定したパーセントのデータが、データの上下から除外されます。

要約統計量のオプション

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

要約統計量のカスタマイズ 表示したい統計量を、一覧から選択してください。すべての統計量を選択したり、逆に、すべての選択を解除したりもできます。

すべての最頻値を表示 複数の最頻値があるときに、そのすべてが表示されます。

関連情報

- 「要約統計量」(72 ページ)

「一変量の分布」プラットフォームのオプション

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

スケールの統一 すべての軸で最小値、最大値、間隔が同じになり、分布が簡単に比較できるようになります。

積み重ねて表示 ヒストグラムとレポートを横に並べて表示します。複数の変数の分布レポートがある場合は、縦に積み重ねて表示します。このオプションの選択を解除すると、レポートウィンドウのレイアウトは元に戻ります。

複数行に配置 横に並べて表示するプロットの数を入力します。複数のプロットを横一列に配列するのではなく、縦列に配置したい場合に便利です。

Flash(.SWF)形式で保存 ヒストグラムを Adobe Flash Player に対応した .SWF ファイル形式で保存します。保存したファイルはプレゼンテーションや Web ページで使用できます。SWF ファイルの使用に必要なコードを含む HTML ページも同時に保存されます。

このオプションの詳細については、<http://www.jmp.com/support/swfhelp/ja> を参照してください。

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。

カテゴリカル変数のオプション

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

連続尺度の変数に適用されるオプションについては、「[連続変数のオプション](#)」（47 ページ）を参照してください。

カテゴリカル変数の [表示オプション] サブメニュー

度数 「度数」レポートの表示／非表示を切り替えます。「[度数レポート](#)」（41 ページ）を参照してください。

横に並べる ヒストグラムとレポートの表示方向を縦／横に切り替えます。

軸を左に 「度数」、「割合」、「密度」の各軸を右側から左側へ移動させます。

このオプションは、[横に並べる] が選択されている場合のみ使用可能です。

カテゴリカル変数の「ヒストグラムオプション」サブメニュー

ヒストグラム ヒストグラムの表示／非表示を切り替えます。「[ヒストグラム](#)」(38 ページ) を参照してください。

縦に表示 ヒストグラムの表示方向を縦／横に切り替えます。

標準誤差バー ヒストグラムの水準ごとに標準誤差バーが描かれます。

棒の間を離す ヒストグラムの棒を離して表示します。

ヒストグラムの色 ヒストグラムの棒の色を変更します。

度数軸 それぞれの棒に含まれている値の度数を表す軸が表示されます。

割合軸 それぞれの棒に含まれている値の割合を表す軸が表示されます。

密度軸 ヒストグラムの棒の長さを示す軸が追加されます。

度数と割合は、それぞれ次のように計算されます。

$$\text{割合} = (\text{棒の幅}) \times \text{密度}$$

$$\text{度数} = (\text{棒の幅}) \times \text{密度} \times (\text{合計度数})$$

パーセントの表示 ヒストグラムの各棒に、その棒の度数が全体に占める割合が表示されます。

度数の表示 ヒストグラムの各棒に、その棒の度数が表示されます。

モザイク図 1変数に対するモザイク図が表示されます。モザイク図とは、長方形を縦に積み重ねた一種の棒グラフであり、長方形の面積がグループの度数に比例しています。

棒の表示順序 ヒストグラム、モザイク図、「度数」レポートを度数の大きい順または小さい順に並べ替えます。新しい順序を列プロパティとして保存するには、**[保存] > [値の順序]** オプションを使用します。

割合の検定 仮説の割合を検定したレポートが表示されます。詳細は、「[\[割合の検定\] オプションの例](#)」(65 ページ) を参照してください。

信頼区間 割合に対する信頼区間が計算されます。信頼区間の計算方法は、スコア法です。このメニューでは、信頼区間の α 水準を選択できます。メニューにある値を選択するか、**[その他]**を選択して独自の値を入力します。

カテゴリカル変数の「保存」サブメニュー

水準番号 データテーブルの中に、「水準<列名>」という新しい列が作成されます。各オブザベーションの水準番号は、ヒストグラム上でそのオブザベーションが属している棒の番号です。

値の順序 (このオプションは、**[棒の表示順序]** オプションと併用してお使いください。) データテーブルの列に、「値の順序」列プロパティを作成します。保存された「値の順序」列プロパティでは、現在、指定されている順序でデータが並びかえられています。

スクリプトをログに保存 現在のレポートを生成するためのスクリプトコマンドがログウィンドウに表示されます。[表示] > [ログ] を選択して、ログウィンドウを表示します。

削除 「一変量の分布」レポートから、変数とそのすべてのレポートが完全に削除されます。

連続変数のオプション

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

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

連続変数の「表示オプション」サブメニュー

分位点 「分位点」レポートの表示／非表示を切り替えます。「[「分位点」レポート](#)」(42ページ)を参照してください。

分位点の間隔の設定 分位点の間隔を変更するか、デフォルトの値に戻します。

カスタム分位点 計算する分位点を、値または増分によって指定できます。信頼区間の信頼水準を変更したり、平滑化された経験尤度分位点を計算することもできます（大きなデータでは、計算に時間がかかる場合があります）。

- 重み付き平均で求められる分位点の計算方法については、「[分位点](#)」(72ページ)を参照してください。
- 重み付き平均の分位点に対する信頼限界には、分布によらないノンパラメトリックな方法が使われていますが、その詳細については、Hahn and Meeker (1991) のセクション5.2を参照してください。
- 平滑化した経験尤度法による分位点は、カーネル密度推定に基づいて求められます。これらの分位点および信頼限界の推定方法の詳細については、Chen and Hall (1993) を参照してください。

要約統計量 「要約統計量」レポートの表示／非表示を切り替えます。「[「要約統計量」レポート](#)」(42ページ)を参照してください。

要約統計量のカスタマイズ 「要約統計量」レポートに表示される統計量を追加または削除します。「[「要約統計量」レポート](#)」(42ページ)を参照してください。

横に並べる ヒストグラムとレポートの表示方向を縦／横に切り替えます。

軸を左に 「度数」、「割合」、「密度」、「正規分位点プロット」の各軸を右側から左側へ移動させます。

このオプションは、[横に並べる] が選択されている場合のみ使用可能です。

連続変数の「ヒストグラムオプション」サブメニュー

ヒストグラム ヒストグラムの表示／非表示を切り替えます。「[ヒストグラム](#)」(38 ページ) を参照してください。

シャドウグラム ヒストグラムの代わりにシャドウグラムを表示します。ヒストグラムのバンド幅を変更すると、ヒストグラムの形状が変化します。シャドウグラムは、いくつかのバンド幅に対するカーネル密度推定の結果を合わせたものです。シャドウグラムにすると、分布の主要な特長は把握しにくくなります。

シャドウグラムでは次のオプションは使用できません。

- － 標準誤差バー
- － 度数の表示
- － パーセントの表示

縦に表示 ヒストグラムの表示方向を縦／横に切り替えます。

標準誤差バー 標準誤差を使って、ヒストグラムの水準ごとに標準誤差バーが作成されます。手のひらツールで棒の数を変更すると、標準誤差バーも自動的に調整されます。「[連続変数のヒストグラムにおける棒の幅の変更](#)」(39 ページ) および「[標準誤差のバー](#)」(71 ページ) を参照してください。

棒の幅の設定 ヒストグラムの棒の幅を変更します。「[連続変数のヒストグラムにおける棒の幅の変更](#)」(39 ページ) を参照してください。

ヒストグラムの色 ヒストグラムの棒の色を変更します。

度数軸 それぞれの棒に含まれている値の度数を表す軸が表示されます。

メモ：ヒストグラムの棒のサイズを変更すると、度数軸も変化します。

割合軸 それぞれの棒に含まれている値の割合を表す軸が表示されます。

メモ：ヒストグラムの棒のサイズを変更すると、割合軸も変化します。

密度軸 密度軸が追加されます。密度とは、ヒストグラムの棒の長さを指します。度数と割合は、それぞれ次のように計算されます。

$$\text{割合} = (\text{棒の幅}) \times \text{密度}$$

$$\text{度数} = (\text{棒の幅}) \times \text{密度} \times (\text{合計度数})$$

「分布のあてはめ」オプションで作成した密度曲線を検討する場合は、密度軸によって、密度曲線上の値を知ることができます。

メモ：ヒストグラムの棒のサイズを変更すると、密度軸も変化します。

パーセントの表示 ヒストグラムの各棒に、その棒の度数が全体に占める割合が表示されます。

度数の表示 ヒストグラムの各棒に、その棒の度数が表示されます。

正規分位点プロット 変数がどの程度正規分布に従っているかを示す正規分位点プロットが作成されます。
「[正規分位点プロット](#)」(49ページ)を参照してください。

外れ値の箱ひげ図 外れ値の箱ひげ図が表示されます。「外れ値の箱ひげ図」は、外れ値を探すのに役立ちます。
「[外れ値の箱ひげ図](#)」(50ページ)を参照してください。

分位点の箱ひげ図 分位点の箱ひげ図が表示されます。「分位点」レポートに示されている分位点が表示されます。
「[分位点の箱ひげ図](#)」(51ページ)を参照してください。

幹葉図 ヒストグラムの一種である幹葉図が作成されます。「[幹葉図](#)」(52ページ)を参照してください。

累積確率プロット 経験累積分布関数のプロットが作成されます。「[累積確率プロット](#)」(52ページ)を参照してください。

平均の検定 平均の一標本検定が実行されます。「[平均の検定](#)」(53ページ)を参照してください。

標準偏差の検定 標準偏差の一標本検定が実行されます。「[標準偏差の検定](#)」(54ページ)を参照してください。

同等性の検定 同等性の検定が実行されます。「[同等性の検定](#)」(55ページ)を参照してください。

信頼区間 平均と標準偏差に対する信頼区間が計算されます。「[連続変数における信頼区間](#)」(56ページ)を参照してください。

予測区間 次に無作為抽出される1つの観測値、もしくは、次に無作為抽出される標本の平均や標準偏差が含まれる区間が計算されます。「[予測区間](#)」(57ページ)を参照してください。

許容区間 少なくとも指定した割合だけのオブザベーションを含むと主張できる区間が計算されます。「[許容区間](#)」(57ページ)を参照してください。

工程能力分析 工程が特定の仕様限界に適合しているかどうかを測定します。「[工程能力分析](#)」(58ページ)を参照してください。

連続分布のあてはめ 連続分布をデータにあてはめます。「[分布のあてはめ](#)」(60ページ)を参照してください。

離散分布のあてはめ 離散分布をデータにあてはめます。「[分布のあてはめ](#)」(60ページ)を参照してください。

保存 連続変数またはカテゴリカル変数についての情報が保存されます。「[連続変数における保存コマンド](#)」(56ページ)を参照してください。

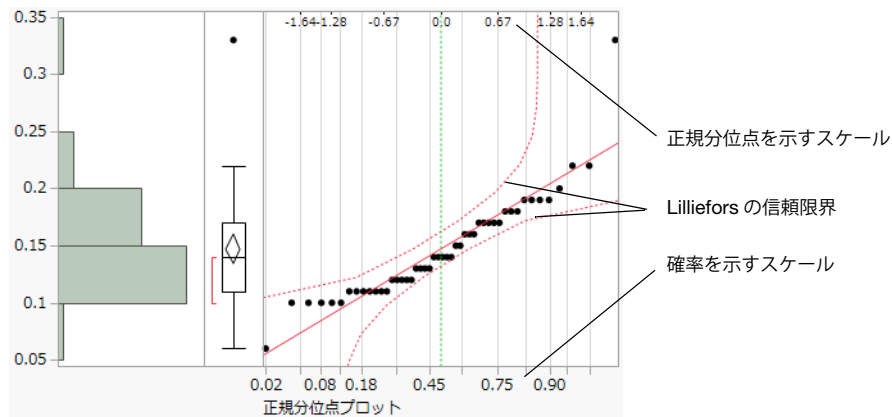
削除 「一変量の分布」レポートから、変数とそのすべてのレポートが完全に削除されます。

正規分位点プロット

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

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

図3.7 正規分位点プロット



次の点に注目してください。

- Y軸は、列の値を表します。
- X軸は、各値の経験累積確率を示します。

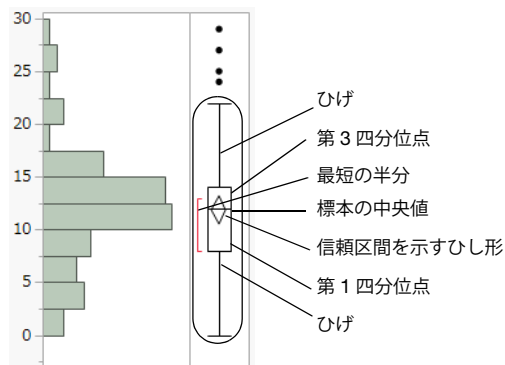
関連情報

- 「[正規分位点プロット](#)」(73 ページ)

外れ値の箱ひげ図

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

図3.8 外れ値の箱ひげ図



外れ値の箱ひげ図には、次のような情報が描かれます。

- 箱を横切る線は標本の中央値（メディアン）を示します。
- ひし形は、平均と、平均の95%信頼区間を表します。ひし形の中央にある水平線が平均を示し、ひし形の上下の点が、それぞれ平均の上側と下側の95%信頼限界を表します。
- 箱の両端は25%と75%の分位点で、それぞれ第1および第3四分位点とも呼ばれます。
- 第1四分位点と第3四分位点の差が四分位範囲です。
- 箱の両端にはひげと呼ばれる線が付いています。ひげは、箱の端から、次の式で計算された範囲内で最も遠くにある点まで伸びています。

第1四分位点 - 1.5*（四分位範囲）

第3四分位点 + 1.5*（四分位範囲）

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

- 箱の外側に表示されている括弧は、最短の半分（shortest half）を示します。これは、オブザベーションの50%が密集する最も短い範囲です（Rousseuw and Leroy 1987）。

箱ひげ図からのひし形や最短半分の削除

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

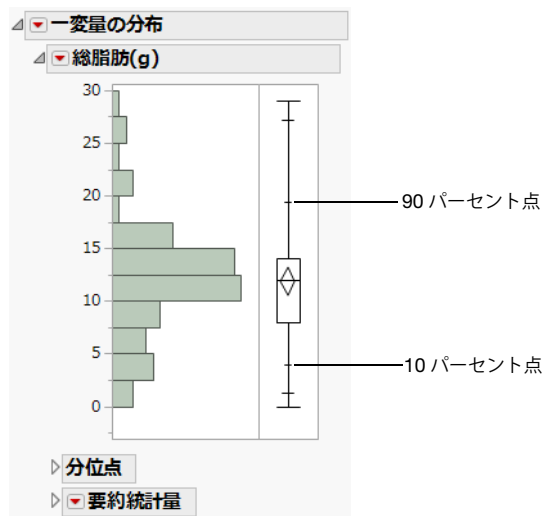
1. 外れ値の箱ひげ図を右クリックし、[カスタマイズ] を選択します。
2. 「箱ひげ図」を選択します。
3. [信頼のひし形] または [最短の半分] のチェックを外します。

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

分位点の箱ひげ図

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

図3.9 分位点の箱ひげ図



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

幹葉図

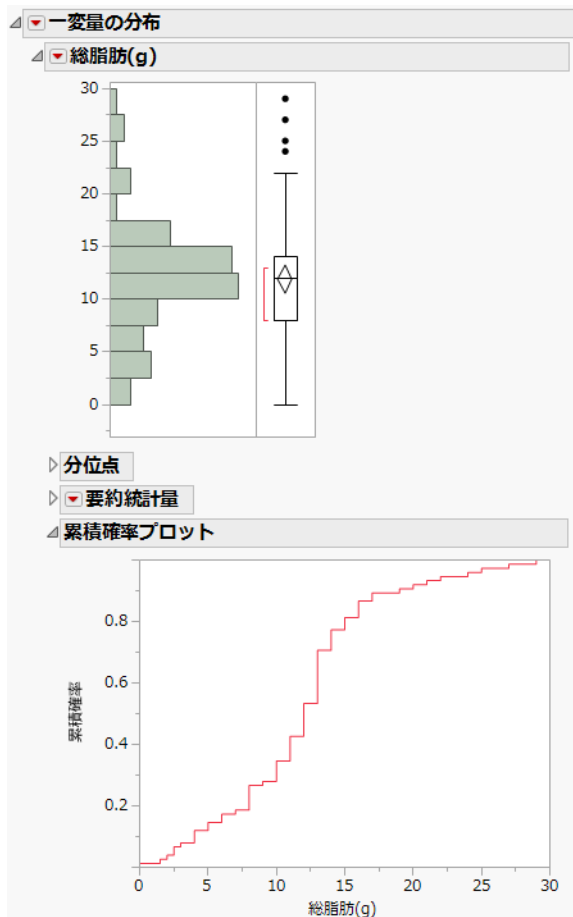
プロットの各行に表示された**幹**の値は、データ値の最初のいくつかの桁を表します。**葉**の値は、各データ値の残りの桁を示します。幹と葉を結合すると、データ値になります。幹葉図には、テーブル上の実測値を丸めた値が表示される場合もあります。幹葉図の値はマウスやブラシツールで選択できます。

メモ: 幹葉図は、度数が小数点であるデータをサポートしていません。

累積確率プロット

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

図3.10 累積確率プロット



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

平均の検定

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

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

「平均の検定」レポートの説明

平均の検定で計算される統計量

t検定（またはz検定） 検定統計量の値と、片側検定と両側検定の p 値がリストされます。

Wilcoxonの符号付順位検定（Wilcoxonの符号付順位検定を選択した場合のみ表示）Wilcoxonの符号付順位検定の統計量と、片側検定と両側検定の p 値がリストされます。この検定では、0の値はPratt法を使用して処理されます。「中央値が仮説値と等しい」という帰無仮説を検定するノンパラメトリックな検定の1つです。なお、「中央値が仮説値と等しい」を帰無仮説とする場合は、分布の対称性を仮定する必要があります。「[Wilcoxonの符号付順位検定の統計量](#)」（74ページ）を参照してください。

p値

p値(Prob>|t|) 母平均が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本の t 値の絶対値よりも、 t 値の絶対値が大きくなる確率です。これは両側 t 検定の p 値です。

p値(Prob>t) 母平均が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本の t 値より大きい t 値を得る確率です。これは上側の片側検定の p 値です。

p値(Prob<t) 母平均が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本の t 値より小さい t 値を得る確率です。これは下側の片側検定の p 値です。

「平均の検定」オプションの説明

p値アニメーション 対話的なグラフに、 p 値を描きます。平均の仮説値などを変更した場合に、 p 値にどのような影響が及ぶかを確認できます。

検出力アニメーション 対話的なグラフに、検出力とベータを描きます。真の平均や、平均の仮説値によって、検出力とベータがどのように変化するかを確認できます。

検定結果の削除 平均の検定結果を削除します。

標準偏差の検定

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

検定統計量 カイ2乗検定統計量の値。「標準偏差の検定」（76ページ）を参照してください。

最小p値 母標準偏差が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本のカイ2乗値から極端に異なるカイ2乗値を得る確率です。「標準偏差の検定」（76ページ）を参照してください。

p値(Prob>ChiSq) 母標準偏差が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本のカイ2乗値より大きいカイ2乗値を得る確率です。これは片側 t 検定の p 値です。

p 値 (Prob<ChiSq) 母標準偏差が仮説値と等しいと仮定したときに、まったくの偶然だけで、計算された標本のカイ2乗値より小さいカイ2乗値を得る確率です。これは片側 t 検定の p 値です。

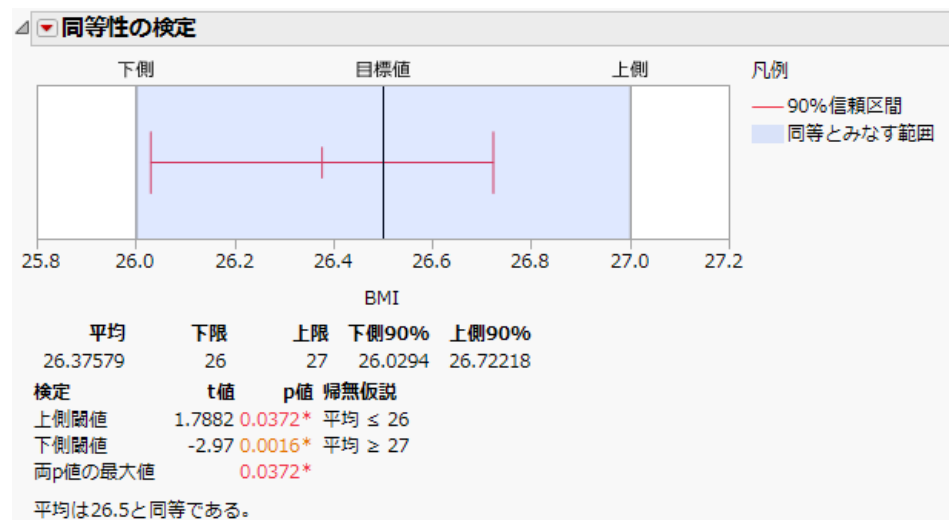
同等性の検定

同等性の検定は、母平均が仮説値と等しいかどうかを評価するものです。差の閾値を設定し、その閾値より小さければ差がないとみなします。JMP の [同等性の検定] オプションは、Two One-Sided Tests (TOST; 2つの片側検定) アプローチを採用しています。2つの片側 t 検定は、「真の平均と仮説値の差は、閾値を超える」という帰無仮説の検定です。両方の帰無仮説が棄却されれば、真の差は閾値を超えない、つまり、平均は仮説値に実質的に等しいとみなせるという結論が導かれます。

[同等性の検定] オプションを選択すると、仮説平均、差の閾値（実質的に 0 とみなす差）、信頼水準を指定する画面が表示されます。信頼水準は $(1 - \alpha)$ です。ここで、 α は 2つの片側検定の有意水準です。

図3.11の「同等性の検定」レポートは、「Diabetes.jmp」サンプルデータテーブルの「BMI」変数に対して実行したものです。仮説平均は 26.5、差の閾値（実質的に 0 とみなす差）は 0.5 と指定されています。

図3.11 「同等性の検定」レポート



レポートには次の情報が表示されます。

- 同等とみなす範囲のプロット。指定した目標値と範囲が縦線で示され、範囲の両端に「下側」、「上側」というラベルが付きます。
- 仮説平均の信頼区間。この信頼区間は、 $1 - 2 \times \alpha$ 水準の区間です。
- 平均、下限と上限、平均の信頼区間 ($1 - 2 \times \alpha$) を示す表。
- 2つの片側検定の結果を示す表。
- 結果の要約と、平均が目標値と等しいとみなせるかどうかを示す記述。

連続変数における信頼区間

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

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

連続変数における保存コマンド

〔保存〕メニューのコマンドを使用して、連続尺度の変数に関する情報を保存できます。各保存コマンドを選択すると、現在のデータテーブルに新しい列が作成されます。新しい列の名前は、選択した保存コマンド名の後ろに変数名（下の表で<列名>と表記）が付いた形になります。表3.1を参照してください。

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

表3.1 〔保存〕メニューのコマンドの説明

コマンド	データテーブルに追加される列	説明
水準番号	水準 <列名>	各オブザベーションの水準番号は、ヒストグラム上でそのオブザベーションが属している棒の番号です。棒に対して、小さいものから順に1から始まる通し番号が付けられます。 メモ：元の情報を維持するために、新しい列に値ラベルが追加されますが、デフォルトではオフになっています。
水準中間点	中間点 <列名>	各オブザベーションの中間点の値は、水準の下側の境界に、水準の幅の半分を足したものです。 メモ：元の情報を維持するために、新しい列に値ラベルが追加されますが、デフォルトではオフになっています。
順位	順位 <列名>	対応する列の値の1から始まる順位が保存されます。重複している値には、データテーブル内での順序に従って、連続した順位が割り当てられます。
平均順位	平均順位 <列名>	データに重複がない場合は、平均順位は、通常の順位と同じです。同じ値が k 回起こっている場合、それらの値の順位を合計し、 k で割ったものが平均順位です。

表 3.1 「保存」メニューのコマンドの説明（続き）

コマンド	データテーブルに 追加される列	説明
確率スコア	確率 <列名>	非欠測値の数を N としたとき、ある値の確率スコアは、その値の平均順位を $N + 1$ で割って算出されます。この列の値は、経験累積分布関数（CDF）に似ています。
正規分位点	正規分位点 <列名>	正規分位点がデータテーブルに保存されます。「 正規分位点プロット 」（73 ページ）を参照してください。
標準化	標準化 <列名>	標準化したデータがデータテーブルに保存されます。「 標準化データの保存 」（76 ページ）を参照してください。
中心化	中心化 <列名>	0 で中心化した値が保存されます。
仕様限界	（なし）	工程能力分析での仕様限界が、現在のデータテーブルの対応列の列プロパティとして保存されます。次に工程分析を再実行したとき、自動的にこの仕様限界が読み込まれ、表示されます。
スクリプトを ログに保存	（なし）	スクリプトをログウィンドウに出力します。スクリプトを実行すると、分析を再現できます。

予測区間

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

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

関連情報

- 「[予測区間](#)」（76 ページ）
- 「[予測区間の例](#)」（67 ページ）

許容区間

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

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

関連情報

- 「許容区間」(77ページ)
- 「許容区間の例」(69ページ)

工程能力分析

【工程能力分析】オプションでは、工程が特定の仕様限界に適合しているかどうかを測定します。工程能力は、デフォルトでは正規分布に基づいて計算されますが、連続分布または離散分布のあてはめを行った場合は、あてはめた分布に基づいて工程能力のレポートが生成されます。あてはめた分布に基づいて工程能力を求めるには、【仕様限界】を選択するか、または【仕様限界】列プロパティを設定します。変数の【工程能力分析】オプションを選択すると、「工程能力分析」ウィンドウが表示されます。このウィンドウで、仕様限界、分布の種類、シグマについての情報を入力します。

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

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

「工程能力分析」ウィンドウの説明

<分布の種類> デフォルトでは、工程能力の統計量や仕様限界から外れているものの割合は、正規分布の仮定に基づいて算出されます。非正規分布について工程能力分析を実行するには、「工程能力分析」(78ページ)の「仕様限界」の説明を参照してください。

<シグマの種類> シグマ(σ)の推定値が選択した手法で計算されます。「工程能力分析」(78ページ)を参照してください。

「工程能力分析」レポートの説明

仕様 仕様限界のリストです。

値 仕様限界と目標値のフィールドにユーザが入力した値のリストです。

比率、実測値 % 「比率」列には、「実測値 %」列の数値について説明する次のラベルが表示されます。

- 「LSL未満」は、下限仕様限界未満のデータの割合を示します。
- 「USL超え」は、上限仕様限界を超えたデータの割合を示します。
- 「限界外の合計数」は、下限仕様限界未満と上限仕様限界超の割合の合計を示します。

工程能力 工程能力指数の種類。「工程能力分析」のオプションの説明」(59ページ)を参照してください。

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

指数 工程能力指数の値。

上側信頼限界 上側信頼限界。

下側信頼限界 下側信頼限界。

比率、パーセント 「比率」列には、「パーセント」列の数値について説明するラベルが表示されます。

- 「LSL 未満」は、あてはめた分布から計算された、下限仕様限界未満の割合を示します。
- 「USL 以上」は、あてはめた分布から計算された、上限仕様限界を超える割合を示します。
- 「限界外の合計数」は、あてはめた分布から計算された、下限仕様限界未満と上限仕様限界超の割合を合計したものです。

PPM（100万分の1） 「PPM」列の値は、「パーセント」列の値に1万を掛けたものです。

シグマクオリティ 「シグマクオリティ」は、シックスシグマ分析の基準として使用するもので、**Process Sigma**（工程シグマ）とも言います。[「工程能力分析」](#)（78ページ）を参照してください。

「工程能力分析」のオプションの説明

Zベンチ ベンチマークZの値を、「**指数**」という列に表示します。この指標については、『AIAG Statistical Process Control』マニュアルで説明されています。平均と仕様限界との差を、標準偏差を単位として表した指標です。Z USL および Z LSL は、それぞれ上側仕様限界と下側仕様限界までの平均からの距離を表します。[「工程能力分析」](#)（78ページ）を参照してください。

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

関連情報

- [「工程能力分析」](#)（78ページ）
- [「工程能力分析の例」](#)（70ページ）

分布のあてはめ

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

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

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

連続分布のあてはめ

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

- 正規分布は、分布が左右対称で、ほとんどのデータが中央にある分布です。正規分布は、負の値や小数点などが、データにあってもかまいません。JMPは、不偏推定量を使用して正規分布のパラメータを求めます。
- 対数正規分布。0を下限值とします。右に大きく歪んでおり、非常に大きな値がいくつかある場合に使います。正規分布に従うデータの指数をとると、対数正規分布になります。JMPは、最尤推定を使用して対数正規分布のパラメータを求めます。
- Weibull分布、閾値つきWeibull分布、極値分布。これらの分布は、特に、工業製品の故障時間や、生物学の生存時間を分析するときに使われています。
- 指数分布は、工業製品の故障時間や、生物学分野の生存時間を分析するときに使われています。また、あるイベントが生じる間隔をモデル化するのにも使われます。たとえば、コンピュータのサーバーにクエリが送られてくる時間、お客様がサービスカウンターに来る間隔、交換機が着信する間隔などです。
- ガンマ分布。この分布も、下限値は0です。
- ベータ分布。(0, 1)の区間を範囲としています。標準ベータ分布は、割合などの0から1の範囲に収まるデータに対してよく使われています。
- 正規混合分布。正規分布を混合した分布です。多峰性の分布にも対応した柔軟な分布です。[二重正規混合]、[三重正規混合]、または[その他]オプションを選択すると、指定した数だけの分布を混合した正規混合分布があてはめられます。
- [平滑曲線]分布は、ノンパラメトリックな密度の推定値(カーネル密度推定)を計算し、平滑曲線をあてはめます。ヒストグラム上に平滑曲線が表示され、プロットの下にスライダが表示されます。スライダを使ってカーネル標準偏差を変更することで、平滑化の度合いを調節できます。[カーネル標準偏差]の初期値は、データの標準偏差に基づいて算出されます。
- Johnson Su分布、Johnson Sb分布、Johnson Sl分布。これら3つで構成される分布システムは、歪度と尖度のあらゆる組み合わせに対応しており、柔軟にデータにあてはまります。

- 一般化対数分布（Glog 分布）。生物検定法（バイオアッセイ）など、分散が一定でない非正規なデータに役立ちます。

すべての分布の比較

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

「分布」のリストは、AICc の値で昇順に並んでいます。

変数に負の値がある場合は、正の値のみを対象とする分布は除外されます。このコマンドの対象は連続分布のみです。すべての場合において、閾値パラメータがある分布（ベータ分布や Johnson Sb 分布など）は除外されます。

関連情報

- [「連続分布のあてはめ」](#) (82 ページ)
- [「あてはめた分布の分位点」](#) (90 ページ)
- [「あてはめた分布のオプション」](#) (91 ページ)

離散分布のあてはめ

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

- Poisson
- 二項
- ガンマ Poisson
- ベータ二項

関連情報

- [「離散分布のあてはめ」](#) (88 ページ)
- [「あてはめた分布の分位点」](#) (90 ページ)
- [「あてはめた分布のオプション」](#) (91 ページ)

あてはめた分布のオプション

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

診断プロット 分位点プロットまたは確率プロットが作成されます。「[診断プロット](#)」(62 ページ) を参照してください。

密度曲線 分布のパラメータ推定値を基に密度曲線が計算され、ヒストグラムに重ねて表示されます。

適合度 あてはめた分布に対して、適合度検定が計算されます。「[適合度](#)」(63 ページ) を参照してください。

分布パラメータの指定 一部のパラメータを任意の値に設定し、他のパラメータの推定値を計算することができます。また、入力したパラメータ値がデータにあてはまるかどうかを検定され、その結果が「妥当性の尤度比検定」レポートとして表示されます。

分位点 このオプションを選択すると、指定した下側累積確率に対して、尺度化も中心化もしていない分位点が戻されます。

Kシグマから仕様限界を設定 設定すべき仕様限界（規格限界）が決まっておらず、データから求めたい場合にこのオプションを使ってください。

仕様限界は、工程を工学的に検討して決めるのが普通です。工学的な検討が行えず、かつ、現在のデータが信頼に足るベンチマークである（工程能力が高い）場合には、現在のデータに分布をあてはめて、その分位点を仕様限界としてもよいでしょう。「[あてはめた分布のオプション](#)」(91 ページ) を参照してください。

仕様限界 指定した仕様限界と目標値を基に、一般化された工程能力指数が計算されます。「[仕様限界](#)」(64 ページ) を参照してください。

あてはめたモデルの分位点を保存 あてはめた分布の分位点が、データテーブルの列に保存されます。「[あてはめた分布の分位点](#)」(90 ページ) を参照してください。

密度関数の保存 あてはめた分布の密度関数が、データテーブルの列に保存されます。密度関数では、パラメータ推定値が使用されます。

仕様限界の保存 仕様限界を列プロパティとして保存します。「[あてはめた分布のオプション](#)」(91 ページ) を参照してください。

変換の計算式を保存 正規分布に変換するための計算式が保存されます。この計算式は、あてはめた分布に基づいて、データを正規分布に近づくように変換するものです。このオプションは、Johnson 分布か一般化対数分布があてはめられた場合のみ使用できます。

あてはめの削除 分布のあてはめに関する情報がレポートウィンドウから削除されます。

診断プロット

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

X 軸：データ、Y 軸：あてはめた分布の分位点

- 閾値つき Weibull
- ガンマ
- ベータ
- Poisson

- ガンマ Poisson
- 二項
- ベータ二項

X軸：データ、Y軸：あてはめた分布の確率

- 正規
- 正規混合分布法
- 指数

X軸：対数スケールのデータ、Y軸：あてはめた分布の確率

- Weibull 損失関数
- 対数正規
- 極値損失関数

X軸：標準正規分布の分位点、Y軸：あてはめた分布の確率

- Johnson SI
- Johnson Sb
- Johnson Su
- 一般化対数

「診断プロット」の赤い三角ボタンのメニューには、次のオプションが表示されます。

回転 X軸とY軸を入れ換えます。

信頼限界 正規分位点プロットでは Lilliefors の 95% 信頼限界、その他の分位点プロットでは $\alpha = 0.01$ 、 $b = 0.99$ のときの 95% 等精度信頼帯 (95% equal precision band) を表示します (Meeker and Escobar 1998)。

あてはめ線 参照線としてまっすぐな対角線を表示します。変数の値がこの参照線の付近にあるなら、選択した分布が変数にあてはまっていることがわかります。

中央値の参照線 応答変数の中央値を示す水平線を表示します。

適合度

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

- 正規分布については、標本サイズが 2000 以下のときは正規性に関する Shapiro-Wilk の検定が、標本サイズが 2000 より大きいときは KSL (Kolmogorov-Smirnov-Lilliefors) の検定が計算されます。

- Poisson 分布などの離散分布において標本サイズが30以下の場合は、2つの正確な Kolmogorov 片側検定を組み合わせる両側検定とした方法（Conover 1972）が使われます。この両側検定は、ほぼ正確です。また、離散分布において標本サイズが30を超える場合は、Pearson のカイ2乗検定が使われます。

関連情報

- [「あてはめた分布のオプション」](#) (91 ページ)

仕様限界

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

関連情報

- [「あてはめた分布のオプション」](#) (91 ページ)

「一変量の分布」プラットフォームのその他の例

ここでは、「一変量の分布」プラットフォームを使った例をさらに紹介します。

複数のヒストグラムにおけるデータ選択の例

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

小規模企業について、「タイプ」の分布を確認しましょう。

5. 「small」の横にある棒をクリックします。

製薬（Pharmaceutical）よりもコンピュータ関連（Computer）の会社において、小規模企業（small）が多いことがわかります。次に、選択範囲を拡げて、小規模企業だけでなく、中規模企業も選択しましょう。

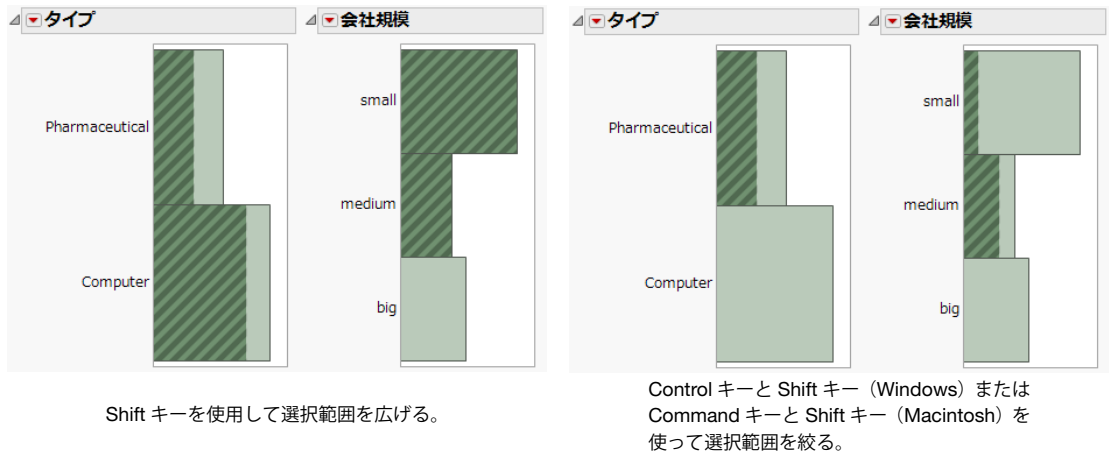
6. 「会社規模」のヒストグラム上で、Shift キーを押しながら、「medium」の棒をクリックします。

小規模（small）および中規模（medium）における「タイプ」の分布を確認できます。図3.12の左側を参照してください。次は、選択範囲を絞って、小規模および中規模の製薬会社だけを表示することにします。

7. Ctrl キーと Shift キー（Windows）または Command キーと Shift キー（Macintosh）を同時に押ししながら、「Computer」に対する選択を解除するために、「タイプ」のヒストグラム上で、Ctrl キーと Shift キーを押しながら、「Computer」の棒をクリックします。

小規模（small）、もしくは中規模（medium）の製薬会社（Pharmaceutical）が、どれぐらいデータにあるかが分かります。図3.12の右側を参照してください。

図3.12 複数のヒストグラムにおけるデータの選択



〔割合の検定〕 オプションの例

3水準以上のカテゴリカルな変数に対して、割合の検定を実行する例を述べます。

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

図3.13の左側を参照してください。

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

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

図3.13の右側を参照してください。

図3.13 「割合の検定」のオプションの例

一変量の分布

セル

度数

割合の検定

水準	推定割合	仮説割合
Adeno	0.19708	.
Large	0.19708	.
Small	0.35036	.
Squamous	0.25547	.

クリックして、仮説割合を入力してください。

割合の和が1になるようにスケール変更方法を選択。

☐ 省略された値を推定値に固定し、仮説値のスケールを変更
☒ 仮説値を固定し、省略された値のスケールを変更

完了 ヘルプ

水準が3つ以上の変数のレポート
オプション

一変量の分布

反応

度数

割合の検定

水準	推定割合	仮説割合
Cured	0.53704	.
Died	0.46296	.

クリックして、仮説割合を入力してください。

割合の検定に使う対立仮説を選択します。

☒ 割合が仮説値と等しくない(両側カイ2乗検定)
☐ 割合が仮説値より大きい(正確な片側二項検定)
☐ 割合が仮説値より小さい(正確な片側二項検定)

完了 ヘルプ

水準が2つの変数のレポート
オプション

「割合の検定」の実行例

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

1. 図3.13の左側における「仮説割合」のすべてのフィールドに「0.25」と入力します。
2. [仮説値を固定し、省略された値のスケールを変更] ボタンをクリックします。
3. [完了] をクリックします。

尤度比検定とPearsonのカイ2乗検定が計算されます。図3.14の左側を参照してください。

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

1. 図3.13の右側における「仮説割合」の両方のフィールドに「0.5」と入力します。
2. [割合が仮説値より小さい] ボタンをクリックします。
3. [完了] をクリックします。

二項検定の正確なp値が計算されます。図3.14の右側を参照してください。

図3.14 「割合の検定」レポートの例



予測区間の例

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

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

図3.15 「予測区間」 ウィンドウ

予測区間として(1- α)を入力

予測する標本サイズを入力

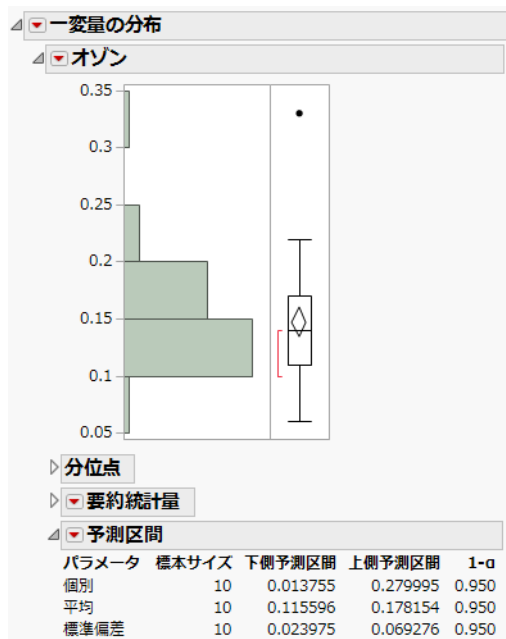
☒ 両側
☐ 片側-下側限界
☐ 片側-上側限界

OK キャンセル ヘルプ

6. 「予測区間」 ウィンドウで、「予測する標本サイズを入力」の横に「10」と入力します。

7. [OK] をクリックします。

図3.16 「予測区間」 レポートの例



この例では、95%の信頼水準で、次のように結論することができます。

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

許容区間の例

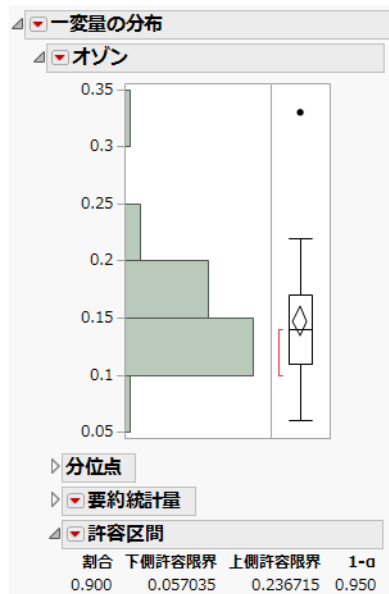
オゾン量の測定値の90%が含まれる区間を推定したいとしましょう。

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

図3.17 「許容区間」ウィンドウ

6. デフォルトの値をそのまま使用し、[OK] をクリックします。

図3.18 「許容区間」レポートの例



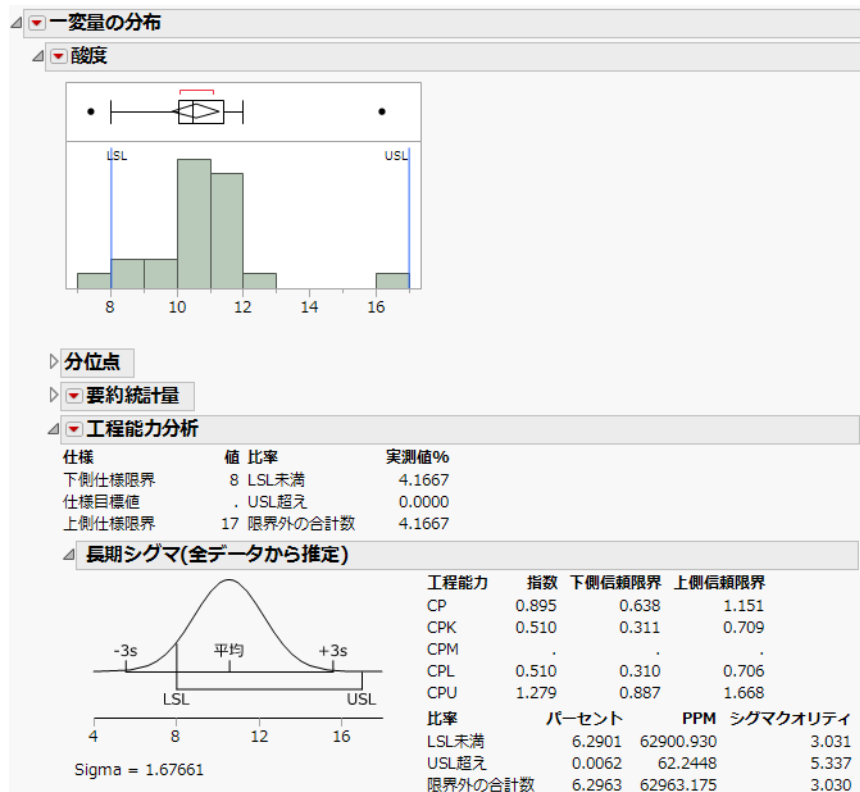
この例では、「下限許容限界」と「上限許容限界」の結果から、95%の信頼水準で、「母集団の少なくとも90%が0.057035～0.236715の区間にある」と結論できます。

工程能力分析の例

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

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

図3.19 「工程能力分析」レポートの例



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

「一変量の分布」プラットフォームの統計的詳細

ここでは、「一変量の分布」で使われている統計分析の詳細について説明します。

標準誤差のバー

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

分位点

ここでは、分位点の計算方法について説明します。

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

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

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

ここで、

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

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

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

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

要約統計量

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

平均

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

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

標準偏差

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

$$s = \sqrt{s^2} \quad \text{ここで、}$$

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

$$\bar{y}_w = (\text{重みつき}) \text{ 平均}$$

平均の標準誤差

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

歪度

平均周りの3次のモーメントを表す指標です。次式によって計算されます。

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

上の式で、 w_i は重みを表す係数（重み付けをしない場合は1）です。

尖度

平均周りの4次のモーメントを表す指標です。次式によって計算されます。

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

ここで、 w_i は重みです（重み付けをしない場合、すべての重みは1です）。この式によって求められた尖度は、正規分布の場合には0となります。この式は、過剰尖度と呼ばれることがあります。

正規分位点プロット

各値の経験累積確率は、次のような式で計算されます。

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

ここで、 r_i はデータ値の順位、 N は非欠測値の個数を表します。

正規分位点の値は次のような式で計算されます。

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

ここで、 Φ は正規分布の累積確率分布関数です。

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

Wilcoxon の符号付順位検定の統計量

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

Wilcoxon 符号付順位検定では、同順位も考慮されます。また、データ値にゼロがある場合は、Pratt の方法で計算されます。Lehman (2006)、Pratt (1959)、および Cureton (1967) を参照してください。

1 標本の中央値の検定

- N 個の観測値を、以下のように表すとします。

$$X_1, X_2, \dots, X_N$$

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

$$H_0: \text{中央値が } m \text{ に等しい}$$

- 内部的な計算では、次式による観測値と仮説値 m との差が使われます。

$$D_j = X_j - m$$

2つの母集団の中央値が等しいことの検定（対応のあるペアのデータ）

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

- 2つの母集団のそれぞれに、 N 個の観測値があるとします。

$$X_1, X_2, \dots, X_N \text{ および } Y_1, Y_2, \dots, Y_N$$

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

$$H_0: X - Y \text{ の中央値が } 0 \text{ である}$$

- 1 標本の問題とするには、次のような式で観測値間の差を計算します。

$$D_j = X_j - Y_j$$

Wilcoxon の符号付順位検定の統計量

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

- 差の絶対値 $|D_j|$ を、小さい値から順番に順位付けします。

- この時、差が0のものがあっても、順位付けにそれらの値も含め、順位は1から始めます。
- また、差の絶対値に同じ値があった場合には、それらには平均順位（中間順位）を与えます。

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

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

符号付順位の統計量は次のように計算されます。

$$W = \frac{1}{2} \sum_{j=1}^N \text{符号付順位}$$

次のように定義します。

d_0 は、0と等しい符号付順位の数。

R^+ は、正の符号付順位の合計。

このとき、次の式が成立します。

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

Wilcoxonの符号付順位検定のp値

$N \leq 20$ の場合は、正確なp値が計算されます。

$N > 20$ の場合は、以下のように、Studentのt分布に近似して計算されます。同順位のペアに対して、修正が行われることに注意してください。Iman（1974）およびLehmann（1998）を参照してください。

帰無仮説のもとでは、Wの平均はゼロです。Wの分散は次式により計算されます。

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

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

次式で求められた統計量 t が、自由度 $N - 1$ のt分布で近似されます。

$$t = W / \sqrt{Var(W)}$$

標準偏差の検定

検定統計量は次の式で計算されます。

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

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

「最小 p 値」は、両側検定の p 値で、次の式で計算されます。

$$2 \times \min(p_1, p_2)$$

ここで、 p_1 は下側の片側検定の p 値で、 p_2 は上側の片側検定の p 値です。

正規分位点

正規分位点の値は次のような式で計算されます。

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

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

標準化データの保存

標準化された値は、次の式で計算されます。

$$\frac{X - \bar{X}}{S_X} \text{ ここで、}$$

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

予測区間

JMP では、次式により予測区間を計算しています。

- 将来の m 個の観測値に対する予測区間

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

- 将来の m 個の観測値から計算される平均に対する予測区間

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

- 将来の m 個の観測値から計算される標準偏差に対する予測区間

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

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

- なお、予測区間が片側の場合には、 $1-\alpha$ の分位点を用いて計算します。

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

許容区間

ここでは、片側許容区間と両側許容区間の統計的詳細について説明します。

片側許容区間

許容区間は、片側区間の場合、次式によって計算されます。

$$\text{上側の場合} = \bar{x} + g's$$

$$\text{下側の場合} = \bar{x} - g's$$

この式で、

$$g' = t(1-\alpha, n-1, \Phi^{-1}(p) \cdot \sqrt{n}) / \sqrt{n}$$

t は非心 t 分布の分位点です。

Φ^{-1} は標準正規分布の分位点です。

両側区間

両側区間は次の式で計算されます。

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

この式で、

s = 標準偏差

$g(1-\alpha/2; p, n)$ は定数です。

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

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

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

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

$1 - \gamma$ が、母集団のうち、許容区間に含まれるものの割合となります。

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

工程能力分析

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

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

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

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

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

$$\sigma = \frac{\bar{R}}{d_2(n)} \text{ この式で、}$$

$\bar{R}$ は移動範囲の平均

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

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

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

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

工程能力指数の定義

工程能力指数を一般化した式は、次のようになります。次式において、 T は目標値、 LSL および USL は、それぞれ下側仕様限界および上側仕様限界、 P_α はあてはめた分布の分位点です。

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

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

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

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

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

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

分布が正規分布の場合には、これらは、通常の工程能力指数を表すよりシンプルな式になります。詳細については、表3.2を参照してください。

メモ：次の表の信頼区間は、有意水準0.05を使って計算されています。

表3.2 工程能力指数と計算式の説明

指数	指数名	計算式
CP	工程能力指数 C_p	$(USL - LSL)/6\sigma$ ここで • USL は上限仕様限界 • LSL は下限仕様限界

表 3.2 工程能力指数と計算式の説明（続き）

指数	指数名	計算式
CPの信頼区間	CPの下側信頼限界	$CP \sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}}$
	CPの上側信頼限界	$CP \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}}$
CPK (AIAGのPPK)	工程能力指数 C_{pk}	$\min(CPL, CPU)$
CPKの信頼区間 (Bissell (1990) を参照)	下側信頼限界	$\hat{C}_{pk} \left[1 - \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{1}{9n\hat{C}_{pk}^2} + \frac{1}{2(n-1)}} \right]$
	上側信頼限界	$\hat{C}_{pk} \left[1 + \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{1}{9n\hat{C}_{pk}^2} + \frac{1}{2(n-1)}} \right]$
CPM	工程能力指数 C_{pm}	$\frac{\min(\text{Target} - LSL, USL - \text{Target})}{3\sqrt{s^2 + (\bar{x} - \text{Target})^2}}$ メモ：CPM の信頼区間は、目標値が「下側信頼限界」から「上側信頼限界」までの範囲にない場合は表示されません。目標値がこの範囲に含まれる場合に限り表示されます。CPM の信頼区間が算出されない場合は、ログにメッセージが書き込まれます。
CPMの信頼区間	CPMの下側信頼限界	$CPM \sqrt{\frac{\chi_{\alpha/2, \gamma}^2}{\gamma}}, \text{ ここで } \gamma$ $= \frac{n \left(1 + \left(\frac{\bar{x} - \text{Target}}{s} \right)^2 \right)^2}{1 + 2 \left(\frac{\bar{x} - \text{Target}}{s} \right)^2}$
	CPMの上側信頼限界	$CPM \sqrt{\frac{\chi_{1-\alpha/2, \gamma}^2}{\gamma}}$ ここで $\gamma = \text{同上}$ 。
CPL	下側仕様の工程能力比	$(\text{平均} - LSL)/3s$

表 3.2 工程能力指数と計算式の説明（続き）

指数	指数名	計算式
CPU	上側仕様の工程能力比	$(USL - \text{平均}) / 3s$

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

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

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

- CPU の正確な $100(1 - \alpha)\%$ 下側および上側信頼限界も、Chou et al. (1990) による手法を一般化した式によって計算されます。Chou et al. (1990) によると、CPU の $100(1 - \alpha)\%$ の下側信頼限界（CPULCL）は、次式を満たします。

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

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

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

- シグマクオリティは次のように定義されます。

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

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

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

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

- ベンチマーク Z の値は次の式で計算されます。

$$Z_{USL} = (USL - \bar{X}) / \sigma = 3 * CPU$$

$$Z_{LSL} = (\bar{X} - LSL) / \sigma = 3 * CPL$$

$$Z_{Bench} = \left[1 - P(LSL) - P(USL)\right] \text{ に対する分位点}$$

ここで、

$P(\text{LSL}) = \text{Prob}(X < \text{LSL}) = Z \text{ LSL}$ における下側累積確率

$P(\text{USL}) = \text{Prob}(X > \text{USL}) = 1 - \text{「}Z \text{ LSL}$ における下側累積確率」

連続分布のあてはめ

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

正規

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

正規分布には、次のパラメータがあります。

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

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

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

$E(x) = \mu$

$\text{Var}(x) = \sigma^2$

対数正規

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

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

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

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

Weibull、閾値つき Weibull、極値

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

閾値つき Weibull 分布の確率密度関数は、次のような式で計算されます。

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

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

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

上の式で、 $\Gamma(\cdot)$ はガンマ関数です。

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

指数

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

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

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

$$E(x) = \sigma$$

$$\text{Var}(x) = \sigma^2$$

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

ガンマ

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

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

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

$$\text{Var}(x) = \alpha\sigma^2$$

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

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

ベータ

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

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

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

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

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

上の式で、 $B(\cdot)$ はベータ関数です。

正規混合分布法

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

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

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

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

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

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

平滑曲線

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

Johnson Su、Johnson Sb、Johnson Sl

Johnson の分布システムは、変換すると正規分布になる3つの分布で構成されています。3つの分布は、次のとおりです。

- Johnson Su（有界でない）
- Johnson Sb（推定可能なパラメータにより定義される有界を上下に持つ）
- Johnson Sl（推定可能なパラメータにより定義される有界を上下のいずれかに持つ対数正規分布）

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

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

Z を標準正規変量とすると、分布システムは次のように定義されます。

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

Johnson Su の場合は、

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

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

Johnson Sb の場合は、

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

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

Johnson SI の場合は、 $\sigma = \pm 1$ で

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

$$Y = \frac{X - \theta}{\sigma} \quad \begin{array}{ll} \theta < X < \infty & \text{if } \sigma = 1 \\ -\infty < X < \theta & \text{if } \sigma = -1 \end{array}$$

Johnson Su

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

Johnson Sb

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

Johnson SI

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

上の式で、 $\phi(\cdot)$ は標準正規分布の確率密度関数です。

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

一般化対数 (Glog)

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

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

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

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

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

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

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

すべて

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

AICc は、次のような式で計算されます。

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

ここで、

- $\log L$ は対数尤度
- n は標本サイズ
- v はパラメータの個数

列に負の値がある場合は、正の値のみを対象とする分布は除外されます。また、連続分布のみが表示されます。連続分布でも、閾値パラメータがある分布（ベータ分布や Johnson Sb 分布など）は除外されます。

離散分布のあてはめ

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

Poisson

Poisson 分布では、尺度パラメータ $\lambda > 0$ が推定されます。

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

$$E(x) = \lambda$$

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

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

ガンマ Poisson

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

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

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

$$E(x) = \lambda$$

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

上の式で、 $\Gamma(\cdot)$ はガンマ関数です。

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

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

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

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

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

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

ここで $r = \alpha$ および $(1-p)/p = \tau$

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

二項

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

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

$$E(x) = np$$

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

上の式で、 n は独立した試行の回数を示します。

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

ベータ二項

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

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

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

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

$$E(x) = np$$

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

上の式で、 $\Gamma(\cdot)$ はガンマ関数です。

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

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

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

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

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

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

あてはめた分布の分位点

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

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

あてはめた分布のオプション

ここでは、あてはめた分布に対する、適合度検定と仕様限界について説明します。

適合度

表 3.3 JMP の適合度検定の説明

分布	パラメータ	適合度検定
正規 ^a	μ と σ が未知	Shapiro-Wilk ($n \leq 2000$) KSL (Kolmogorov-Smirnov-Lilliefors) ($n > 2000$)
	μ と σ が既知	KSL (Kolmogorov-Smirnov-Lilliefors)
	μ と σ のどちらかが既知	(なし)
対数正規	μ と σ が未知または既知	Kolmogorov の D 検定
Weibull 損失関数	α と β が未知または既知	Cramér-von Mises の W^2 検定
閾値つき Weibull	α 、 β 、 θ が未知または既知	Cramér-von Mises の W^2 検定
極値損失関数	α と β が未知または既知	Cramér-von Mises の W^2 検定
指数	σ が既知または未知	Kolmogorov の D 検定
ガンマ	α と σ が既知	Cramér-von Mises の W^2 検定
	α と σ のどちらかが未知	(なし)
ベータ	α と β が既知	Kolmogorov の D 検定
	α と β のどちらかが未知	(なし)
二項	ρ が既知または未知、 n が既知	Kolmogorov の D 検定 ($n \leq 30$) Pearson のカイ 2 乗 (χ^2) ($n > 30$)
ベータ二項	ρ と δ が既知または未知	Kolmogorov の D 検定 ($n \leq 30$) Pearson のカイ 2 乗 (χ^2) ($n > 30$)
Poisson	λ が既知または未知	Kolmogorov の D 検定 ($n \leq 30$) Pearson のカイ 2 乗 (χ^2) ($n > 30$)
ガンマ Poisson	λ または σ が既知または未知	Kolmogorov の D 検定 ($n \leq 30$) Pearson のカイ 2 乗 (χ^2) ($n > 30$)

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

Kシグマから仕様限界を設定

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

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

第4章

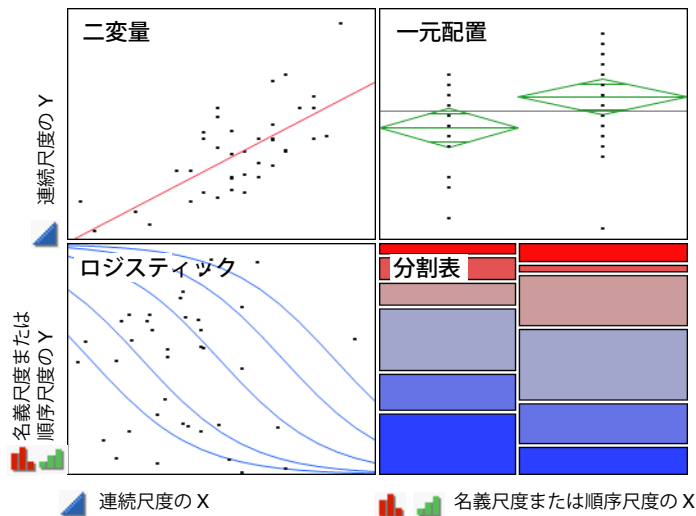
「二変量の関係」プラットフォームの概要 2変数の関係を調べる

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

次の4種類の分析を実行できます。

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

図4.1 4種類の分析



「二変量の関係」プラットフォームの概要

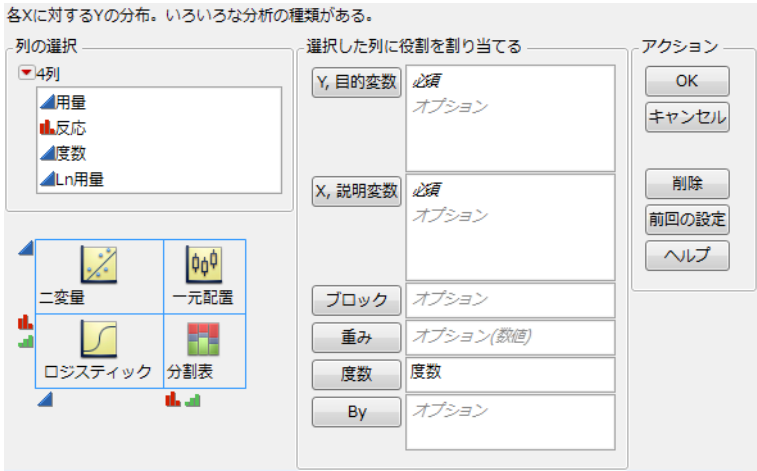
「二変量の関係」プラットフォームは、次の4つの分析プラットフォームで構成されています。

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

「二変量の関係」プラットフォームの起動

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

図 4.2 「二変量の関係」 起動ウィンドウ



二変量、一元配置、ロジスティック、分割表 このグリッドには、変数の尺度に応じて、どの分析が実行されるかが示されています。列に役割を割り当てると、実行される分析名がグリッドの上部に表示されます。

ブロック（オプション。一元配置分析または分割表分析の場合のみ）

- 分散分析の場合には、指定された因子を第2の因子として、交互作用のない二元配置分散分析が行われます。分散分析でブロック変数を指定する場合には、釣り合いがとれていなければいけません。つまり、ブロックごとの各グループのセル度数がすべて等しくないといけません。プロットでは、Y変数の値がブロック変数ごとに中心化されます。
- 分割表の場合には、指定されたブロックの列を、層別因子として、Cochran-Mantel-Haenszel 検定が行われます。

起動ウィンドウの詳細については、『JMPの使用法』の「はじめに」章を参照してください。

「JMP スターター」ウィンドウからの各分析の実行

「JMP スターター」ウィンドウから、特定の分析（「二変量」、「一元配置」、「ロジスティック」、または「分割表」）を実行できます。この方法では、列の尺度（Y変数とX変数）に応じた分析を自分自身で指定する必要があります。表4.1を参照してください。

「JMP スターター」ウィンドウから特定の分析を実行するには、**[基本統計]** カテゴリをクリックし、分析を選択します。

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

表4.1 プラットフォームとボタン

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

第5章

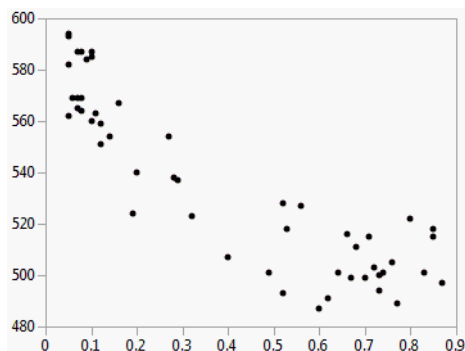
二変量分析

2つの連続変数の関係を調べる

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

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

図5.1 二変量分析の例

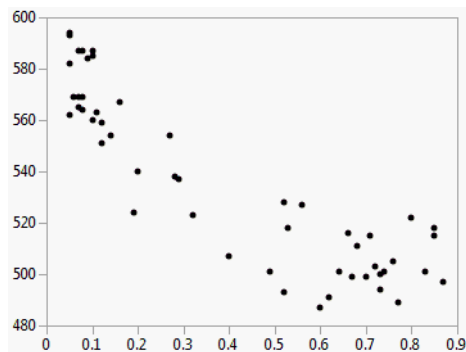


二変量分析の例

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

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

図5.2 受験率に対するテストスコアの例



受験率が低い州ほど、言語テストの点数が高いことがわかります。

「二変量」プラットフォームの起動

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

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

図5.3 「二変量」 起動ウィンドウ

Xが連続に変化する時の応答の分布

列の選択
▼ 23列

- 州
- 2004 受験率(%)
- 2004 言語
- 2004 数学
- 2003 言語
- 2003 数学
- 2002 言語
- 2002 数学
- 2001 言語
- 2001 数学
- 1999 言語
- 1999 数学
- 1994 言語
- 1994 数学

選択した列に役割を割り当てる

Y, 応答変数	必須 連続変数(数値) オプション 連続変数(数値)
X, 説明変数	必須 連続変数(数値) オプション 連続変数(数値)
重み	オプション(数値)
度数	オプション(数値)
By	オプション

アクション

OK
キャンセル
削除
前回の設定
ヘルプ

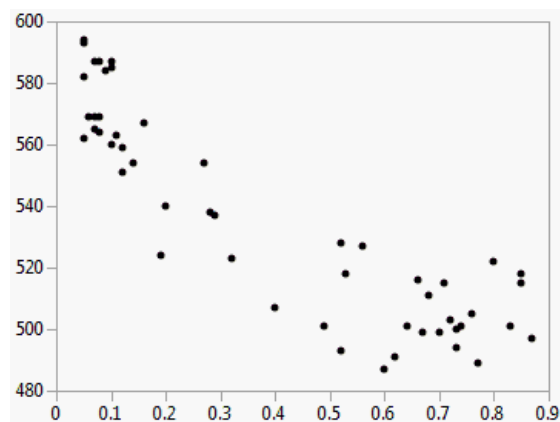
この起動ウィンドウの詳細については、「[「二変量の関係」プラットフォームの概要](#)」章（93ページ）を参照してください。

[OK] をクリックすると、二変量に対する散布図が表示されます。[「二変量に対する散布図」](#)（99ページ）を参照してください。

二変量に対する散布図

図5.4は、「[二変量分析の例](#)」（98ページ）の手順に従って作成したものです。

図5.4 二変量に対する散布図



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

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

赤い三角ボタンをクリックすると、メニューが表示されます。このメニューには、さまざまなあてはめのコマンドが用意されています。コマンドを選択すると、プロットに曲線をあてはめたり、統計量のレポートを表示したり、追加のメニューを表示したりできます。「あてはめ用コマンドとオプション」（100ページ）を参照してください。

あてはめ用コマンドとオプション

メモ: Y変数またはX変数を複数指定した場合は、「あてはめのグループ」メニューが表示されます。このメニューのオプションを使って、レポートの配置を変えたり、適合度の良い順に並べ替えたりできます。詳細については、『基本的な regres 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。

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

点の表示 散布図上の点の表示／非表示を切り替えます。チェックマークが付いている場合は、点が表示されています。

ヒストグラム軸 散布図のX軸とY軸にヒストグラムを追加表示します。チェックマークが付いている場合は、ヒストグラム軸が表示されています。「ヒストグラム軸」（103ページ）を参照してください。

メモ: データテーブルの行を「表示しない」に設定すると、散布図には表示されませんが、ヒストグラムには含められます。ヒストグラムから除外し、分析の対象外にするには、行を「除外する」に設定したうえで、赤い三角ボタンのメニューから「やり直し」>「分析のやり直し」を選択してください。

グループ別 分類変数（グループ変数）を選択できます。グループ変数の水準ごとに個別の分析が行われ、散布図に回帰曲線または楕円が重ねて表示されます。「グループ別」（119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。

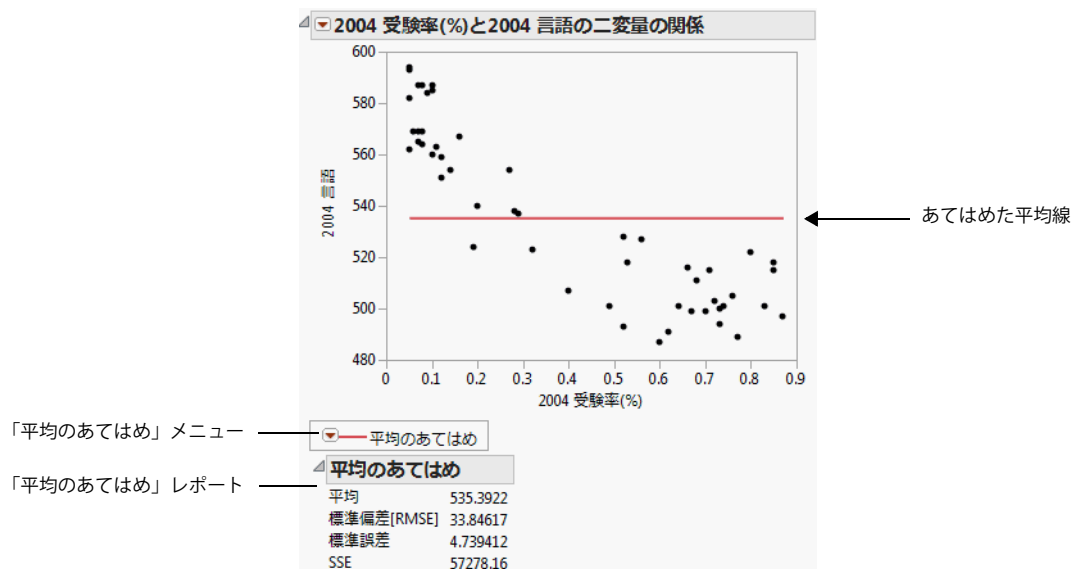
By グループのスクリプトを保存 By 変数の全水準に対するレポートを再現するためのスクリプトを保存するオプションが、保存先ごとに用意されています。起動ウィンドウで By 変数を指定した場合のみ使用可能です。

あてはめのオプション

あてはめのコマンドを選択すると、次の情報が追加で表示されます。

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

図5.5 [平均のあてはめ] コマンドの例



あてはめのオプションには次のものがあります。

平均のあてはめ 応答変数 Y の平均を示す水平線を、散布図に表示します。詳細は、「[平均のあてはめ](#)」(103 ページ)を参照してください。

直線のあてはめ 最小2乗法によって、散布図に回帰直線をあてはめます。「[直線のあてはめと多項式のあてはめ](#)」(104 ページ)を参照してください。

多項式のあてはめ 最小2乗法によって、指定された次数の多項式曲線をあてはめます。「[直線のあてはめと多項式のあてはめ](#)」(104 ページ)を参照してください。

その他のあてはめ YやXを変換した後に、直線や曲線をあてはめます。変換には、自然対数、平方根、2乗、逆数、指数があります。多項式の中心化をオフにしたり、切片と傾きの制約を指定したり、多項式モデルをあてはめることも可能です。「[その他のあてはめ](#)」(111 ページ) を参照してください。

ノンパラメトリックな曲線 滑らかな曲線をあてはめるためのオプションが用意されています。「[ノンパラメトリックな曲線](#)」(112 ページ) を参照してください。

直交のあてはめ Yだけではなく、Xの値もばらつくような状況で役立つ直交回帰をあてはめるオプションがあります。XとYの分散比に対して、いくつかのサブオプションがあります。「[直交のあてはめ](#)」(115 ページ) を参照してください。

ロバスト ここで用意されているオプションでは、データにおける外れ値の影響を少なくしながら、回帰モデルをあてはめます。「[ロバスト](#)」(116 ページ) を参照してください。

確率楕円 XとYから推定される二変量正規分布から計算された確率楕円をプロットします。「[確率楕円](#)」(117 ページ) を参照してください。

ノンパラメトリック密度 滑らかな密度等高線をプロットします。密度等高線はデータ点の密度を表します。「[ノンパラメトリック密度](#)」(118 ページ) を参照してください。

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

あてはめ用コマンドの種類

あてはめを行うコマンドは、回帰のあてはめと、密度推定の2つに大別されます。

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

同じあてはめ用コマンドの繰り返し

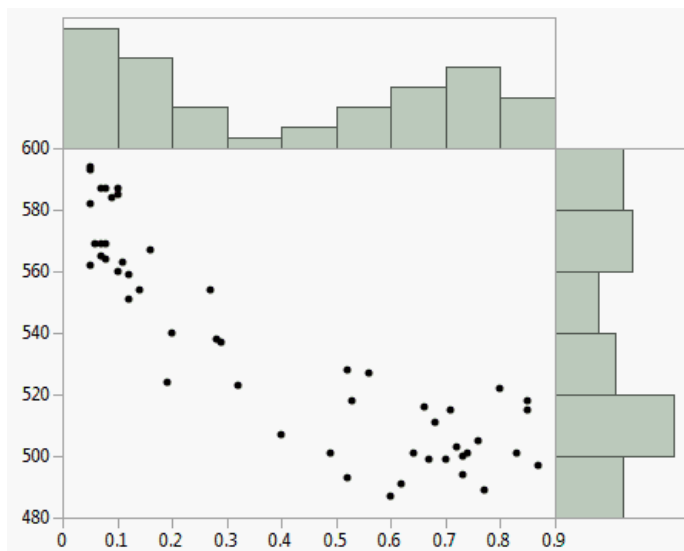
同じあてはめ用コマンドを、繰り返し選択できます。あてはめを行うたびに散布図に新しい線が追加されます。ある分析を試した後、点を除外してから同じ分析を実行し、それらの分析結果を散布図上で比較できます。

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

ヒストグラム軸

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

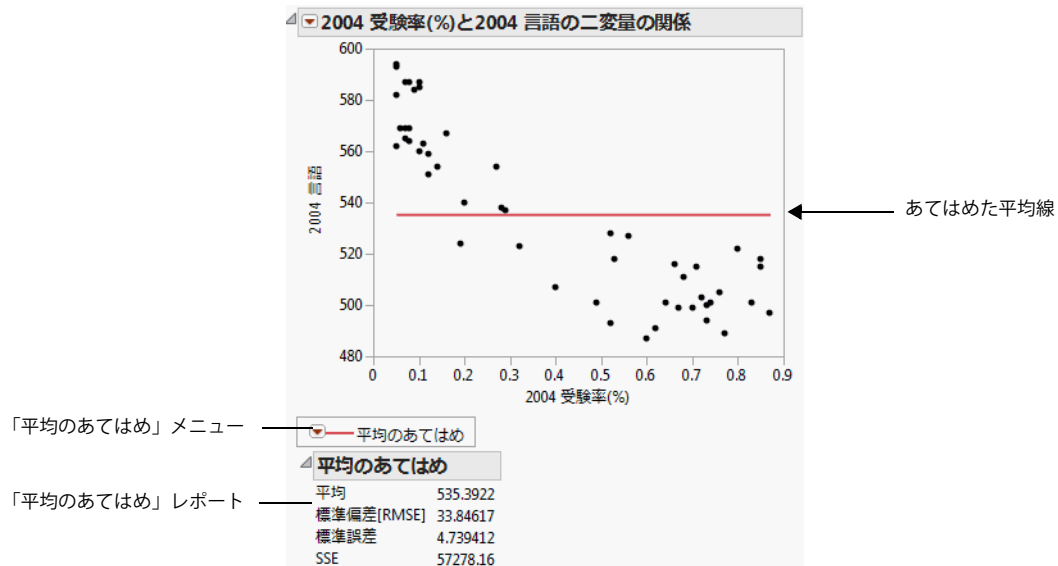
図5.6 ヒストグラム軸の例



平均のあてはめ

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

図 5.7 平均のあてはめの例



「平均のあてはめ」レポート

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

平均 応答変数の平均。モデル内で効果が指定されていないときの応答の予測値です。

標準偏差 [RMSE] 応答変数の標準偏差。誤差の平均平方の平方根で、「誤差の標準偏差 (RMSE)」と呼ばれます。

標準誤差 応答変数の平均の標準偏差。RMSE を標本サイズの平方根で割った値です。

SSE 単純な平均モデルの誤差平方和。モデルのあてはめでは、「分散分析」レポートに「誤差」の平方和として表示されます。

関連情報

- 「[あてはめ用メニュー](#)」(120 ページ)

直線のあてはめと多項式のあてはめ

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

図5.8 直線と多項式のあてはめ

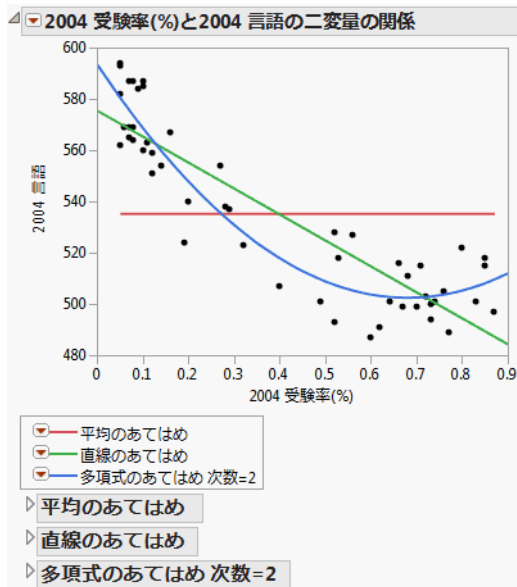


図5.8は、平均線、回帰直線、2次多項式を比較したものです。

次の点に注目してください。

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

「直線のあてはめ」と「多項式のあてはめ」のレポート

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

図5.9 あてはめの式の例



メモ: 式はクリックして編集できます。

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

「あてはめの要約」レポート

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

図5.10 直線と多項式の「あてはめの要約」レポート

直線のあてはめ		多項式のあてはめ次数=2	
2004 言語 = 575.62539 - 101.32814*2004 受験率(%)		2004 言語 = 562.47113 - 111.07875*2004 受験率(%) + 196.95725*(2004 受験率(%) - 0.39706)^2	
あてはめの要約		あてはめの要約	
R2乗	0.790274	R2乗	0.879413
自由度調整R2乗	0.785994	自由度調整R2乗	0.874389
誤差の標準偏差(RMSE)	15.65752	誤差の標準偏差(RMSE)	11.99566
Yの平均	535.3922	Yの平均	535.3922
オブザベーション(または重みの合計)	51	オブザベーション(または重みの合計)	51

「あてはめの要約」レポートには、次の情報が表示されます。

R2乗 モデルによって説明できる変動の割合を表します。残りの変動は、モデルで説明できないため、ランダムな誤差とみなされます。モデルが完全にあてはまっていると、R2乗は1になります。

メモ R2乗の値が小さい場合は、変動を説明する変数が、モデルに含まれていない可能性があります。ただし、扱っているデータ自体の変動が大きい場合は、意味のある回帰モデルであっても、R2乗の値が小さくなることがあります。研究分野により求められるR2乗値が異なる場合がありますので、その分野に関する文献を参考にしてください。

図5.10の「R2乗」の値から、2次多項式は、ほんの少しだけ直線よりもあてはまりが良いことがわかります。「[「あてはめの要約」レポート](#)」(132ページ)を参照してください。

自由度調整 R2乗 パラメータ数の異なるモデルでも比較できるように「R2乗」に調整を加えたもので、計算の際に自由度が考慮されます。「[「あてはめの要約」レポート](#)」(132ページ)を参照してください。

誤差の標準偏差 (RMSE) ランダムな誤差の標準偏差の推定値です。「分散分析」レポートにある「誤差」の平均平方の平方根です。図5.12を参照してください。

応答の平均 応答変数の標本平均 (算術平均)。これは、モデルで効果が指定されていないときのYの予測値です。

オブザベーション あてはめを推定するために使われたオブザベーションの数。ただし、重み変数が使われた場合は、重みの合計です。

「あてはまりの悪さ (LOF)」 レポート

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

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

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

直線のあてはめ				
2004 言語 = 575.62539 - 101.32814*2004 受験率(%)				
あてはめの要約				
あてはまりの悪さ (LOF)				
要因	自由度	平方和	平均平方	F値
あてはまりの悪さ (LOF)	36	9662.983	268.416	1.4850
純粋誤差	13	2349.750	180.750	p値(Prob>F)
合計誤差	49	12012.733		0.2252
				最大R2乗
				0.9590

多項式のあてはめ次数=2				
2004 言語 = 562.47113 - 111.07875*2004 受験率(%) + 196.95725*(2004 受験率(%) - 0.39706)^2				
あてはめの要約				
あてはまりの悪さ (LOF)				
要因	自由度	平方和	平均平方	F値
あてはまりの悪さ (LOF)	35	4557.2473	130.207	0.7204
純粋誤差	13	2349.7500	180.750	p値(Prob>F)
合計誤差	48	6906.9973		0.7862
				最大R2乗
				0.9590

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

「あてはまりの悪さ (LOF)」 レポートには、次の情報が表示されます。

要因 変動の3つの要因 (「あてはまりの悪さ (LOF)」、「純粋誤差」、「合計誤差」)。

自由度 誤差の各要因の自由度 (DF)。

- 「合計誤差」の自由度は、「分散分析」レポートの「誤差」の行に表示されている自由度です ([「分散分析」レポート](#)) (108ページ) を参照)。この自由度は、**データ全体**の自由度から、**モデル**の自由度を引いたものです。「誤差」の自由度は、あてはまりの悪さ (LOF) と純粋誤差の2つの自由度に分けることができます。
- 「純粋誤差」の自由度は、X 値が同じである行をグループにまとめ、それぞれのグループの自由度を合計したものです。[「あてはまりの悪さ \(LOF\)」 レポート](#) (132ページ) を参照してください。
- 「あてはまりの悪さ (LOF)」の自由度は、「合計誤差」と「純粋誤差」の自由度の差です。

平方和 誤差の各要因の平方和 (SS)。

- 「合計誤差」の平方和は、「分散分析」レポートの「誤差」の行に表示されている平方和です ([「分散分析」レポート](#)) (108ページ) を参照)。

- 「純粋誤差」の平方和は、 X 値が同じである行をグループにまとめ、それぞれのグループの平方和を合算したものです。この平方和は、モデルの X 効果ではまったく説明できない、純粋にランダムな誤差の大きさを推定したものです。「[「あてはまりの悪さ \(LOF\)」レポート](#)」(132 ページ) を参照してください。
- 「あてはまりの悪さ (LOF)」の平方和は、「合計誤差」と「純粋誤差」の平方和の差です。LOF の平方和が大きいつきは、モデルがデータに良くあてはまっていない可能性があります。この後で説明する「 F 値」によって、あてはまりの悪さによって生じる変動が十分に小さいかどうか、つまり、純粋誤差に対して無視できる大きさであるかが検定されます。

平均平方 平方和を関連する自由度で割った値。この計算により、平方和が平均（平均平方）に換算されます。統計的検定で使用する F 値は平均平方の比です。

F 値 「あてはまりの悪さ (LOF)」の平均平方の、「純粋誤差」の平均平方に対する比。「LOF 誤差は 0 である」という帰無仮説を検定します。

p 値 (Prob > F) LOF に起因する変動がない場合に、まったくの偶然だけで、現在の F 値よりも大きな F 値を得る確率。 p 値が大きい場合は、あてはまりの悪さの誤差は有意ではありません。

最大 R^2 乗 モデルに含まれている変数で達成できる最大の R^2 乗。「[「あてはまりの悪さ \(LOF\)」レポート](#)」(132 ページ) を参照してください。

「分散分析」レポート

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

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

図 5.12 直線および多項式のあてはめの「分散分析」レポートの例

直線のあてはめ					多項式のあてはめ次数=2				
2004 言語 = 575.62539 - 101.32814*2004 受験率(%)					2004 言語 = 562.47113 - 111.07875*2004 受験率(%) + 196.95725*(2004 受験率(%) - 0.39706)^2				
▷ あてはめの要約					▷ あてはめの要約				
▷ あてはまりの悪さ (LOF)					▷ あてはまりの悪さ (LOF)				
分散分析					分散分析				
要因	自由度	平方和	平均平方	F 値	要因	自由度	平方和	平均平方	F 値
モデル	1	45265.424	45265.4	184.6379	モデル	2	50371.160	25185.6	175.0265
誤差	49	12012.733	245.2	p 値 (Prob> F)	誤差	48	6906.997	143.9	p 値 (Prob> F)
全体(修正済み)	50	57278.157		<.0001*	全体(修正済み)	50	57278.157		<.0001*

「分散分析」レポートには、次の情報が表示されます。

要因 変動の 3 つの要因（「モデル」、「誤差」、「全体 (修正済み)」）。

自由度 変動の各要因の自由度 (DF)。

- 自由度は、非欠測値の標本サイズ (N) から、使用したパラメータ数を引いて求めます。標本全体の変動においては、全体平均を表す1つのパラメータだけが使用されるので、標本サイズから自由度が1つ引かれます。例では、49となっています。「全体(修正済み)」の自由度は、「モデル」項と「誤差」項に分かれます。
- 「直線のあてはめ」では、切片のパラメータに、傾きのパラメータを1つ追加したモデルが推定されます。分散分析表における「モデル」の自由度は、1です。また、2次の「多項式のあてはめ」では、切片のパラメータに、2つのパラメータ (β_1 と β_2) を追加したモデルが推定されます。分散分析表における「モデル」の自由度は2となります。
- 「誤差」の自由度は、「全体(修正済み)」と「モデル」の自由度の差です。

平方和 変動の各要因の平方和 (SS)。

- 図5.12の例では、各応答から標本平均までの距離の平方和（「全体(修正済み)」）は57,278.157です。これは、基本モデル（単純な平均モデル）の平方和で、その他のモデルとの比較に使用されます。
- 線形回帰においては、各点から直線までの距離の平方和は12,012.733に減少します。これが、線形モデルでは説明できない「誤差」（残差）の平方和です。2次多項式をあてはめると、残差の平方和は6,906.997になり、直線のときよりわずかに多く変動が説明されていることになります。言い換えると、2次多項式の方が、直線よりもモデル平方和が大きくなっており、より多くの変動を説明しています。「全体(修正済み)」の平方和から「誤差」の平方和を引くと、モデルで説明される平方和が求められます。

平均平方 平方和に関連する自由度で割った値。統計的検定で使用される「 F 値」は、次に示す2つの平均平方の比です。

- 「直線のあてはめ」の「モデル」平均平方は45,265.4です。この値は、モデルのパラメータが0であるという仮説のもとでの誤差分散の推定値です。
- 「誤差」平均平方である245.2は、誤差分散の推定値です。

F 値 モデルの平均平方を誤差の平均平方 (MSE) で割ったもの。これは、(切片を除く) すべての回帰パラメータが0であるという仮説を検定します。この仮説が真のとき、「誤差」と「モデル」の両方の平均平方が誤差分散の推定値となり、その比は F 分布に従います。0でないパラメータがある場合、「 F 値」は通常、偶然だけから期待される値よりも大きくなります。

p 値 (Prob > F) 真のモデルが、平均だけのモデルであった場合に、まったくの偶然だけで、現在より大きな F 値が得られる確率 (p 値) を表します。有意確率が0.05以下だと、効果がある証拠だと考えられています。

「パラメータ推定値」レポート

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

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

図5.13 直線および多項式のあてはめの「パラメータ推定値」レポートの例

直線のあてはめ 2004 言語 = 575.62539 - 101.32814*2004 受験率(%) ▶ あてはめの要約 ▶ あてはまりの悪さ(LOF) ▶ 分散分析 ▲ パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	575.62539	3.684288	156.24	<.0001*
2004 受験率(%)	-101.3281	7.457094	-13.59	<.0001*

多項式のあてはめ次数=2 2004 言語 = 562.47113 - 111.07875*2004 受験率(%) + 196.95725*(2004 受験率(%) - 0.39706)^2 ▶ あてはめの要約 ▶ あてはまりの悪さ(LOF) ▶ 分散分析 ▲ パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	562.47113	3.583843	156.95	<.0001*
2004 受験率(%)	-111.0788	5.942967	-18.69	<.0001*
(2004 受験率(%) - 0.39706)^2	196.95725	33.06487	5.96	<.0001*

「パラメータ推定値」レポートには、次の情報が表示されます。

項 モデルの各パラメータの名前。定数項である切片は、どのモデルにも含まれます。

推定値 線形モデルのパラメータ推定値。回帰モデルの予測式は、これらの推定値と、対応する変数との線形結合で表されます。

標準誤差 パラメータ推定値の標準誤差を推定した値。この標準誤差を使って、検定と信頼区間が計算されます。

t値 各パラメータが0であるという仮説を検定する統計量。t値は、パラメータ推定値とその標準誤差の比です。仮説が真のとき、t値はStudentのt分布に従います。

p値(Prob>|t|) それぞれのt値から計算した有意確率。これは、帰無仮説が真のときに、まったくの偶然だけで、計算されたt値より（絶対値が）大きいt値が得られる確率を示します。値が0.05（場合によっては0.01）より小さいときは、パラメータがゼロとは有意に異なる証拠と解釈されます。

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

下側95% パラメータ推定値の下側95%信頼限界。

上側95% パラメータ推定値の上側95%信頼限界。

標準β 標準化されたパラメータ推定値。異なるスケールで測定したX変数の効果を比較する場合に便利です。「[「パラメータ推定値」レポート](#)」（133ページ）を参照してください。

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

計画の標準誤差 計画行列に基づく、パラメータ推定値の標準誤差。「[「パラメータ推定値」レポート](#)」（133ページ）を参照してください。

関連情報

- [「直線のあてはめ」](#)（131ページ）
- [「あてはめ用メニュー」](#)（120ページ）

その他のあてはめ

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

「変換または制約を指定」ウィンドウには次のオプションがあります。

Yの変換 ここで選択したオプションに従ってY変数が変換されます。

Xの変換 ここで選択したオプションに従ってX変数が変換されます。

次数 指定した次数の多項式をあてはめます。

多項式の中心化 多項式の中心化をオフにするには、[多項式の中心化]のチェックを外します。図5.19を参照してください。なお、X変数を変換した場合、多項式の中心化は行われません。多項式の中心化オプションを用いると、多重共線性の影響が軽減され、回帰係数の数値計算が安定します。

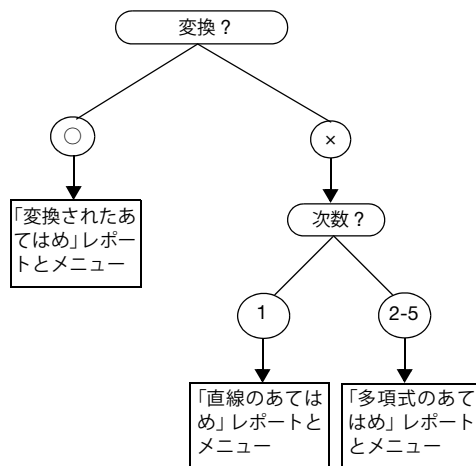
切片の制約 このチェックボックスをオンにすると、指定された数値にモデルの切片が固定されます。

傾きの制約 このチェックボックスをオンにすると、指定された数値にモデルの傾きが固定されます。

その他のあてはめのレポートとメニュー

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

図5.14 その他のあてはめのフローチャート



「変換されたあてはめ」レポート

「変換されたあてはめ」レポートには、「[「直線のあてはめ」と「多項式のあてはめ」のレポート](#)」(105ページ)で説明されているレポートが表示されます。また、Yを変換した場合は、このほかにも、「元のスケールで測定したあてはめ」レポートも表示されます。このレポートに表示される統計量は、Y変数に関して、変換前のスケールで計算されています。

関連情報

- 「[「その他のあてはめ」コマンドの例](#)」(123ページ)
- 「[「直線のあてはめ」と「多項式のあてはめ」のレポート](#)」(105ページ)
- 「[あてはめ用メニュー](#)」(120ページ)

ノンパラメトリックな曲線

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

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

スプライン曲線のあてはめ

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

次の点に注目してください。

- 平滑化スプライン曲線は、Xに対するYの期待値を調べるのに役立ちます。
- 曲線の各部分に近い点ほど、その各部分の形状に大きな影響を及ぼしています。 λ の値を小さくすると、局所的に点が曲線に与える影響が大きくなり、曲線が柔軟になります。
- メニューに表示されていないラムダ値を使用する場合は、[スプライン曲線のあてはめ] > [その他] を選択します。X変数のスケールが変わると、あてはめたモデルも変わります。これを避けるためには、「**Xの標準化**」オプションを選択します。そうすれば、X変数のスケールや単位をどのように変更しても、あてはめたモデルの結果はどれも同じになります。

- 異なる λ 値を試してみてください。「平滑化スプライン曲線によるあてはめ」レポートの下にある「 λ の変更」スライダを動かして λ の値を変更できます。ただし、 λ はデータのスケールによって変化するので注意が必要です。たとえば、X 軸がインチ単位の場合とセンチメートル単位の場合では、 λ の値が異なります。

「平滑化スプライン曲線によるあてはめ」レポート

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

R2 乗 データの全変動のうち、平滑化スプライン曲線モデルによって説明されている割合。詳細は、「[「平滑化スプライン曲線によるあてはめ」レポート](#)」(133 ページ) を参照してください。

誤差平方和 (SSE) 各点からスプライン曲線までの距離の平方和。これは、スプラインモデルをあてはめた後もまだ説明できていない「誤差」(残差)を表します。

λ の変更 λ 値を変更できます。値を入力するか、スライダを動かして変更します。

関連情報

- 「[あてはめ用メニュー](#)」(120 ページ)
- 「[スプライン曲線のあてはめ](#)」(131 ページ)

カーネル平滑化

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

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

「局所的な平滑化」レポート

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

R2 乗 データの全変動のうち、平滑化モデルによって説明されている割合。詳細は、「[「平滑化スプライン曲線によるあてはめ」レポート](#)」(133 ページ) を参照してください。

誤差平方和 (SSE) 各点からあてはめた平滑化曲線までの距離の平方和。これは、平滑化モデルをあてはめた後もまだ説明できていない「誤差」(残差)を表します。

局所的なあてはめ (λ) 個々の局所的なあてはめに対し、多項式の次数を選択します。2次多項式を使用すると、局所的なでこぼこが滑らかになります。 λ は、この手法であてはめる多項式の次数を表します。 λ には、0次（平均）、1次（線形）、または2次を指定できます。

重み関数 それぞれの局所的な回帰において、近傍のデータをどのように重み付けするかを指定します。デフォルトでは、重み付けの関数として、トリキューブ（3次式の3乗）が使われます。このような重み付け関数は、局所的な回帰をあてはめていくときに、各点 (x_i, y_i) が局所的な回帰に与える影響を決めます。重みは、 x から x_i までの距離が大きくなるにつれて小さくなり、やがてゼロになります。

滑らかさ (α) 個々の局所的なあてはめを構成する点の数を指定します。スライダーで値を調整するか、値を直接入力します。 α は、平滑化パラメータで、正の数値なら何でもかまいませんが、通常は0.25～1の値を使用します。 α を大きくすると、曲線の滑らかさが増します。

ロバスト性 モデルから離れた点に小さな重みを与えることにより、外れ値による影響を小さくします。ここでは、重みを再計算する計算過程の反復回数（パスの回数）を指定します。この反復計算の目的は、外れ値に対する重みを小さくすることで、自動的に外れ値をなくすことです。

関連情報

- 「[あてはめ用メニュー](#)」（120ページ）

値ごとのあてはめ

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

「値ごとのあてはめ」レポート

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

オブザベーションの数 オブザベーションの合計数。

異なるXの値の個数 X の異なる値の個数。

自由度 純粋誤差の自由度。

平方和 純粋誤差の平方和。

平均平方 純粋誤差の平均平方。

関連情報

- 「[あてはめ用メニュー](#)」（120ページ）

直交のあてはめ

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

「直交のあてはめ」のオプション

次のいずれかの分散比を選択します。

一変量分散, 主成分 X と Y の一変数ずつから求めた分散の推定値を使います。この結果は、標準化したデータに対する主成分分析の第1主成分です。このオプションは、測定システムの分析には向いていません。得られたデータにおける分散の比は、母集団における分散の比と異なる場合があります。

等分散性 誤差分散が等しいと仮定し、分散比を1とします。この結果は、標準化していないデータに対する主成分分析の第1主成分です。散布図において、 X 軸と Y 軸のスケールを同じにし、正規密度楕円を表示すると、得られた線が楕円の長軸になります。

$Y \rightarrow X$ のあてはめ Y の分散が0であるとし、分散比を0とします。

指定された分散比 このオプションでは、任意の分散比を入力できます。 X と Y の測定誤差の大きさが事前にわかっている場合に活用できます。

「直交回帰」レポート

「直交回帰」レポートは、直交回帰モデルの要約統計量をまとめたものです。

変数 直線のあてはめに使用された変数の名前。

平均 各変数の平均。

標準偏差 各変数の標準偏差。

分散比 直線のあてはめに使用された分散比。

相関 2つの変数間の相関。

切片 あてはめた直線の切片。

傾き あてはめた直線の傾き。

下側信頼限界 傾きの下側信頼限界。

上側信頼限界 傾きの上側信頼限界。

有意水準(α) 信頼区間の計算に使用する α 水準を入力します。

関連情報

- 「[あてはめ用メニュー](#)」(120ページ)
- 「[直交のあてはめ](#)」(131ページ)
- 「[\[直交のあてはめ\] コマンドの使用例](#)」(125ページ)

ロバスト

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

ロバストなあてはめ

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

関連情報

- 「[あてはめ用メニュー](#)」(120ページ)
- 「[\[ロバストなあてはめ\] コマンドの使用例](#)」(127ページ)

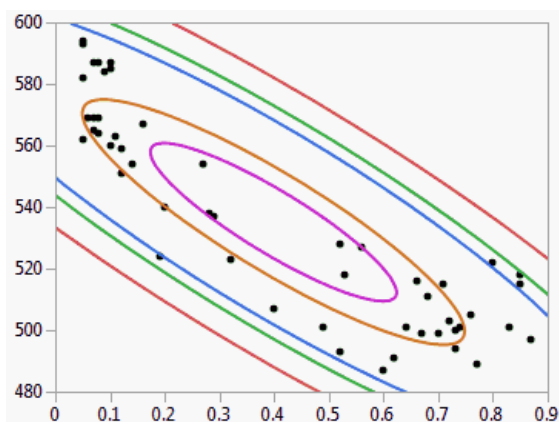
Cauchyのあてはめ

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

確率楕円

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

図5.15 確率楕円の例



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

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

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

「相関」レポート

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

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

変数 楕円の作成に使用された変数の名前。

平均 X 変数と Y 変数のそれぞれの平均。

標準偏差 X変数とY変数のそれぞれの標準偏差。

平均と標準偏差の詳細については、「一変量の分布」章の「[「要約統計量」レポート](#)」(42ページ)を参照してください。

相関 Pearsonの相関係数。二変量間に完全な線形関係がある場合、相関係数は1(正の相関)または-1(負の相関)になり、関係が存在しない場合、相関は0に近くなります。

詳細は、「[「相関」レポート](#)」(134ページ)を参照してください。

p値 X変数とY変数の間に相関関係がないときに、まったくの偶然だけで、計算された相関よりも絶対値の大きい相関が得られる確率。

数 計算に使用されたオブザベーションの数。

関連情報

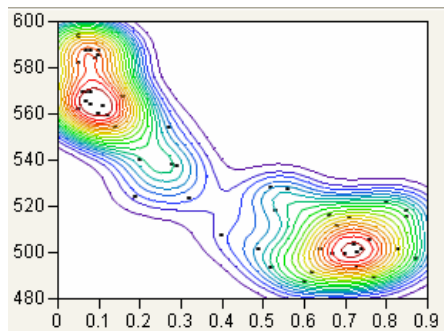
- [「あてはめ用メニュー」](#) (120ページ)
- [「グループ別確率楕円の例」](#) (129ページ)

ノンパラメトリック密度

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

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

図5.16 ノンパラメトリック密度の例



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

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

「ノンパラメトリック二変量の密度」レポート

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

関連情報

- [「あてはめ用メニュー」](#) (120ページ)

グループ別

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

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

[グループ別] オプションは、次のような用途で使用できます。

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

関連情報

- [「グループ別確率楕円の例」](#) (129ページ)
- [「グループ別回帰直線の例」](#) (130ページ)

あてはめ用メニュー

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

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

*X は、何らかの文字または数字です。

あてはめ用メニューのオプション

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

- 「ほとんどのあてはめに適用できるオプション」(121 ページ)。
- 「複数のあてはめに適用できるオプション」(121 ページ)。
- 「二変量正規楕円」に適用できるオプション」(122 ページ)。
- 「分位点密度等高線」に適用できるオプション」(122 ページ)。

ほとんどのあてはめに適用できるオプション

あてはめ線 モデルのあてはめを説明する直線または曲線の表示／非表示を切り替えます。「二変量正規楕円」レポートの場合、このオプションは等高線の境界を示す楕円の表示／非表示を切り替えます。このオプションは、「分位点密度等高線」にはありません。

線の色 パレットが開き、各あてはめ線に割り当てる色を選択できます。このオプションは、「分位点密度等高線」にはありません。

線種 パレットが開き、各あてはめ線に割り当てる線種を選択できます。このオプションは、「分位点密度等高線」にはありません。

線の幅 あてはめ線の幅を3つの中から選択します。デフォルトでは最も細い線が使われています。このオプションは、「分位点密度等高線」にはありません。

レポート あてはめのレポートの表示／非表示を切り替えます。二変量のプロットは変更されません。

あてはめの削除 グラフとレポートからあてはめ線を削除します。

複数のあてはめに適用できるオプション

回帰の信頼区間 期待値（平均）に対する信頼限界の表示／非表示を切り替えます。このオプションは、スプライン曲線のあてはめ、確率楕円、値ごとのあてはめ、および直交のあてはめではグレー表示になっていて使用できません。

個別の値に対する信頼区間 個別の予測値に対する信頼限界の表示／非表示を切り替えます。信頼区間は、誤差における変動とパラメータ推定値における変動の両方を考慮しています。このオプションは、平均のあてはめ、スプライン曲線のあてはめ、確率楕円、値ごとのあてはめ、直交のあてはめではグレー表示になっていて使用できません。

予測値の保存 データテーブル内に「**予測値 列名**」という新しい列を作成します。「**列名**」の部分にはY変数の名前が挿入されます。この列には、計算式の形式で予測式が保存されます。データテーブルに行を追加すると、保存された予測式によって、その行の予測値が自動的に計算されます。このオプションは、値ごとのあてはめ、確率楕円のあてはめではグレー表示になっていて使用できません。

【**予測値の保存**】と【**残差の保存**】のコマンドは、あてはめごとに適用できます。これらのコマンドを何度も実行した場合や、グループ変数と併用した場合は、データテーブル内に作成された列がどのあてはめのものかわかるように列名を変更してください。

残差の保存 データテーブル内に「**残差 列名**」という新しい列を作成します。「**列名**」の部分にはY変数の名前が挿入されます。この列には、各行の実測値（観測値）と予測値の差が保存されます。【**予測値の保存**】コマンドとは異なり、このコマンドでは新しい列に計算式は保存されません。このオプションは、値ごとのあてはめ、確率楕円のあてはめではグレー表示になっていて使用できません。

【**予測値の保存**】と【**残差の保存**】のコマンドは、あてはめごとに適用できます。これらのコマンドを何度も実行した場合や、グループ変数と併用した場合は、データテーブル内に作成された列がどのあてはめのものかわかるように列名を変更してください。

平均の信頼限界の計算式 平均の信頼区間の計算式が保存されます。

個別の信頼限界の計算式 個々の応答値に対する信頼区間の計算式が保存されます。

残差プロット（「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」のみ）「予測値と残差のプロット」、「予測値と実測値のプロット」、「行番号と残差のプロット」、「残差の正規分位点プロット」の4種類の診断プロットを表示します。「[診断プロット](#)」（123 ページ）を参照してください。

α 水準の設定 さまざまなあてはめの信頼区間を計算するための α 水準を指定します。

回帰の信頼区間内を塗る [回帰の信頼区間] コマンドと同じ曲線を描き、曲線間に陰影を付けます。

個別の値に対する信頼区間内を塗る [個別の値に対する信頼区間] コマンドと同じ曲線を描き、曲線間に陰影を付けます。

係数の保存 新しいデータテーブルが作成され、スプライン曲線の係数が、**X**、**A**、**B**、**C**、**D**という列に保存されます。**X**列の値は節点を指します。**A**、**B**、**C**、**D**はそれぞれ、現在の**X**値から次に大きな**X**値までの区間における3次スプライン曲線の切片、1次係数、2次係数、3次係数です。

「二変量正規楕円」に適用できるオプション

楕円内を塗る 確率楕円の内側の領域を塗ります。

内側の点を選択 確率楕円の内側の点を選択します。

外側の点を選択 確率楕円の外側の点を選択します。

「分位点密度等高線」に適用できるオプション

カーネルの制御 標準偏差を変更するためのスライダを各変数に表示します。ノンパラメトリック密度を計算するためのXとYの標準偏差を変更できます。

5%の等高線 5%の等高線の表示／非表示を切り替えます。

等高線 10%の等高線の表示／非表示を切り替えます。

等高線間を塗りつぶす 等高線と等高線の間の領域を塗りつぶします。

密度による点の選択 指定された分位点範囲内の点を選択します。

密度分位点による色分け 密度分位点に応じて点を色分けします。

密度分位点の保存 新しい列を作成し、各点の密度分位点を保存します。

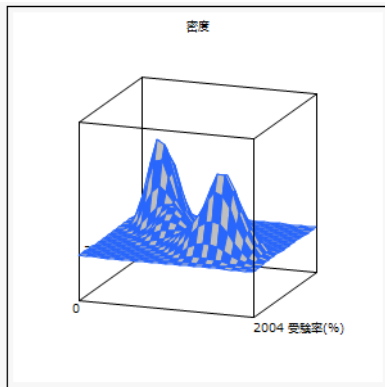
メッシュプロット 分析対象の2変数をグリッドにして、密度推定値を3次元でプロットします。図5.17を参照してください。

峰型クラスター データテーブル内に新しい列を作成し、クラスター値を保存します。

メモ： 峰型クラスター値を先に保存してから密度グリッドを保存すると、グリッドテーブルにもクラスター値が含まれます。クラスター値は、プロットの点に色やマークを付けるときに役立ちます。

密度グリッドの保存 密度推定値とその分位点を新しいデータテーブルに保存します。「三次元散布図」や「等高線図」プラットフォームなどで密度をグラフ化したいときは、このグリッドデータを用いてください。

図5.17 メッシュプロットの例



診断プロット

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

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

予測値と実測値のプロット 予測値と実測値をプロットしたものです。

行番号と残差のプロット X変数と残差をプロットしたものです。

Xと残差のプロット X変数と残差をプロットしたものです。

残差の正規分位点プロット 残差の正規分位点プロットです。

「二変量」プラットフォームのその他の例

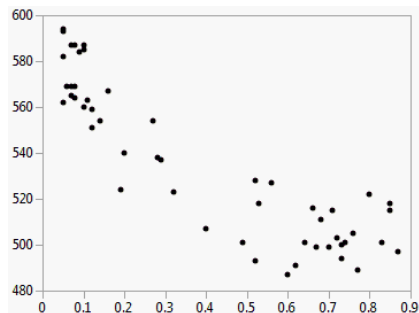
ここでは、「二変量」プラットフォームを使った例をさらに紹介します。

「その他のあてはめ」コマンドの例

Yを対数変換、Xを平方根変換するには、次の手順に従います。

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

図5.18 受験率に対するテストスコアの例



6. 「二変量の関係」の赤い三角ボタンをクリックし、メニューから[その他のあてはめ]を選択します。「変換または制約を指定」ウィンドウが表示されます。このウィンドウの説明については、[「その他のあてはめ」](#) (111 ページ) を参照してください。

図5.19 「変換または制約を指定」ウィンドウ

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

図5.20 その他のあてはめのレポートの例

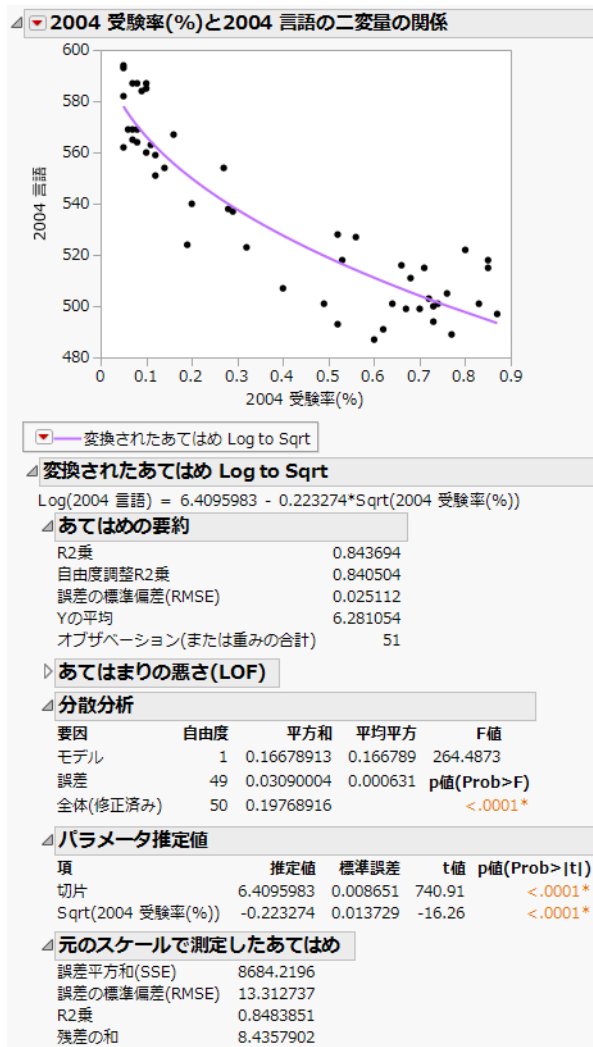


図5.20に示すように、あてはめられた線は元のスケールでプロットされます。プロットされた線の近くにデータが分布しているので、このモデルはデータに良くあてはまっているようです。

〔直交のあてはめ〕 コマンドの使用例

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

変数の標準化

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

2. [分析] > [一変量の分布] を選びます。
3. 「身長(インチ)」と「体重(ポンド)」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。
5. Ctrl キーを押したままにします。「身長 (インチ)」の赤い三角ボタンをクリックし、[保存] > [標準化] を選択します。

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

6. 「一変量の分布」レポートウィンドウを閉じます。

標準化された変数を使用した直交回帰モデルのあてはめ

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

図5.21 「直交のあてはめ」のオプション例

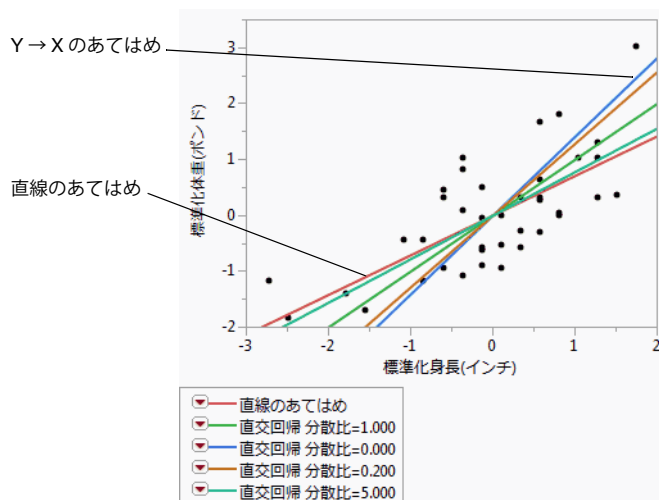


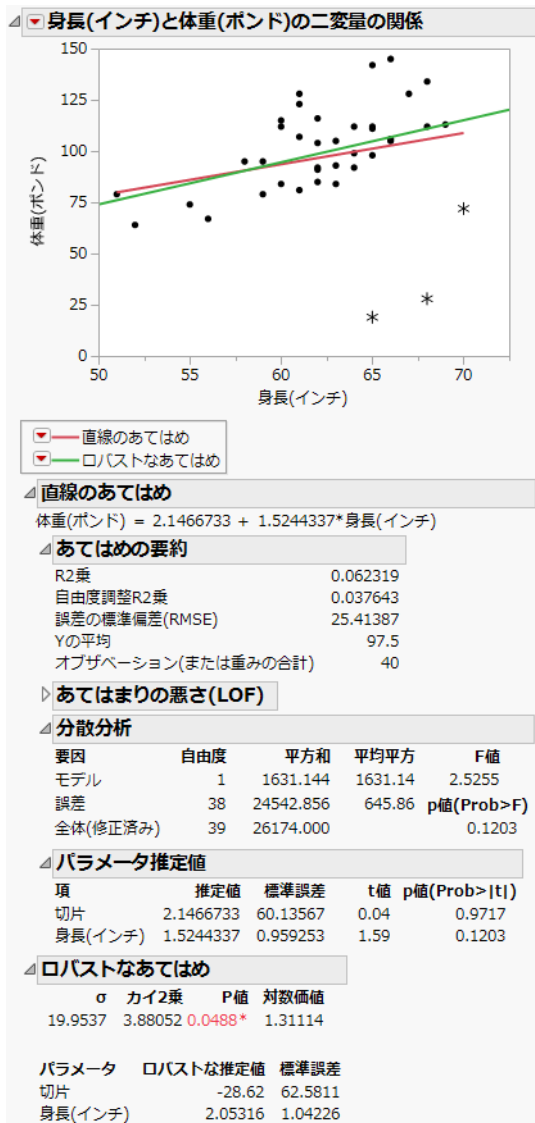
図5.21の散布図は、「身長(インチ)」と「体重(ポンド)」を標準化し、いろいろな直線をあてはめたものです。分散比を変化させたときの「直交のあてはめ」の様子がわかります。X変数の分散をゼロとした場合は、普通の回帰直線です。その対極にあるのがXのYへの回帰で、そこではY変数の分散をゼロとしています。他の直線は、これら2つの間に位置し、分散比に応じて少しずつ傾きが変わっています。分散比が大きくなるにつれて、Yの変動だけが考慮されるようになり、直線の傾きがYのXへの回帰直線（普通の回帰直線）に近付きます。逆に、分散比を小さくすると、直線の傾きはXのYへの回帰直線に近付きます。

「ロバストなあてはめ」 コマンドの使用例

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

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

図 5.22 ロバストなあてはめの例



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

グループ別確率楕円の例

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

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

点を「タイプ」別に色分けするには、次の手順に従います。

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

図5.23 グループ別の例

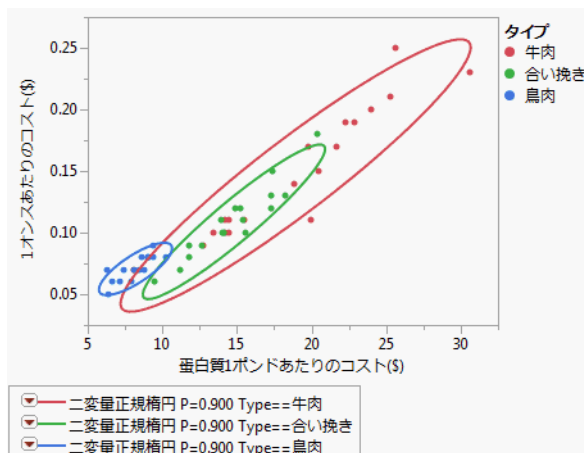


図5.23は、異なる種類のホットドッグの確率楕円を、コストを変数としてプロットしたものです。種類ごとに点がクラスターを形成しているのがわかります。

グループ別回帰直線の例

次の例では、グループ別に回帰直線を求めて、その傾きを比較します。

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

図5.24 左側の回帰直線を描画するには、次の手順に従います。

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

図5.24 右側の回帰直線を描画するには、次の手順に従います。

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

図5.24 データ全体とグループ別の回帰分析の例

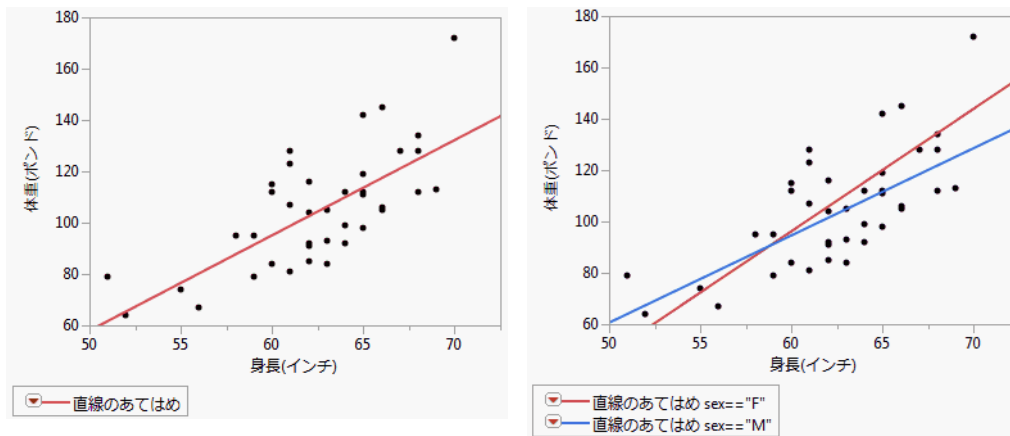


図5.24 右側の散布図に描かれている回帰直線は、それぞれ男子および女子のデータから推定されています。左側の散布図に描かれている回帰直線は、データ全体から推定されています。

「二変量」プラットフォームの統計的詳細

ここでは、「二変量」プラットフォームの統計的詳細について説明します。

直線のあてはめ

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

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

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

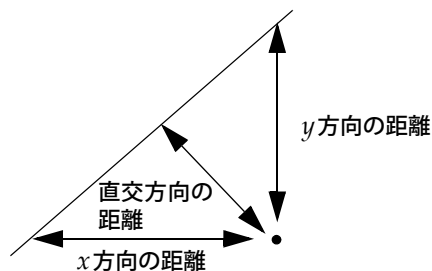
スプライン曲線のあてはめ

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

直交のあてはめ

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

図5.25 回帰直線への垂線



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

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

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

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

「あてはめの要約」レポート

R2乗

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

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

自由度調整 R2乗

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

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

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

「あてはまりの悪さ (LOF)」レポート

「純粋誤差」の自由度

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

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

ここで、 n_i は*i*番目のグループの被験者数を表します。

「純粋誤差」の平方和

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

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

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

最大 R2 乗

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

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

「パラメータ推定値」レポート

標準 β

標準 β は、次の式で計算されます。

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

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

計画の標準誤差

この値は、パラメータ推定値の標準誤差を RMSE で割って求めます。

「平滑化スプライン曲線によるあてはめ」レポート

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

「相関」レポート

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

$$r_{xy} = \frac{s_{xy}^2}{\sqrt{s_x^2 s_y^2}} \quad \text{この式で、} s_{xy}^2 = \frac{\sum w_i (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{df}$$

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

第6章

一元配置分析

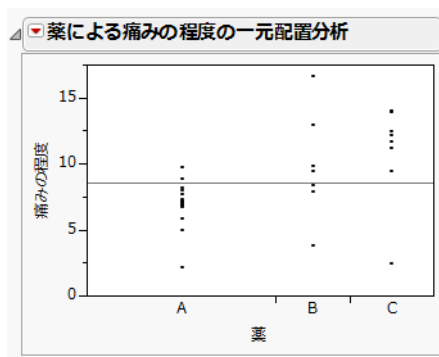
連続尺度のY変数とカテゴリカルなX変数の関係を調べる

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

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

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

図6.1 一元配置分析



一元配置分析の概要

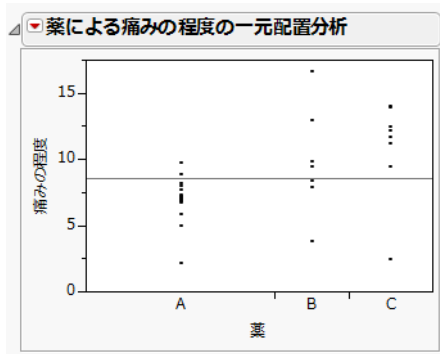
一元配置の分散分析では、群（グループ）ごとの平均に差があるかどうかを検定します。応答変数の総変動は、群内変動と群間変動に分割できます。群内変動に比べて群間変動が大きいときは、群ごとの平均に有意差があると判定します。

一元配置分析の例

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

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

図6.2 一元配置分析の例



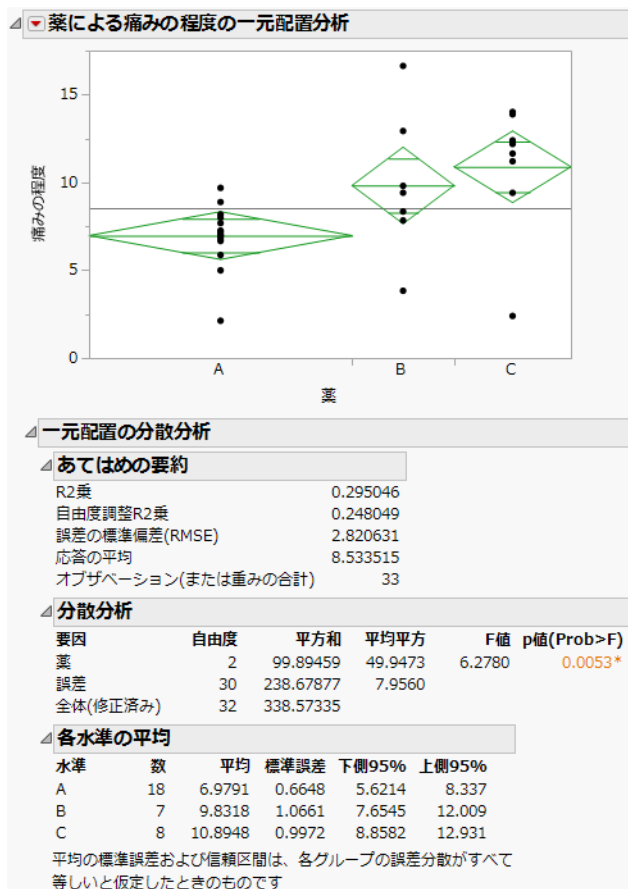
A薬のスコアが、他の薬と比べて全体的に低いことがわかります。なお、X軸の目盛りが等間隔ではありませんが、これは、各薬ごとのデータの個数に目盛りの幅が比例しているためです。

データの分散分析を行います。

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

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

図6.3 [平均/ANOVA] オプションの例



次の点を確認してください。

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

「平均のひし形と標本サイズに比例したX軸」(148ページ)を参照してください。

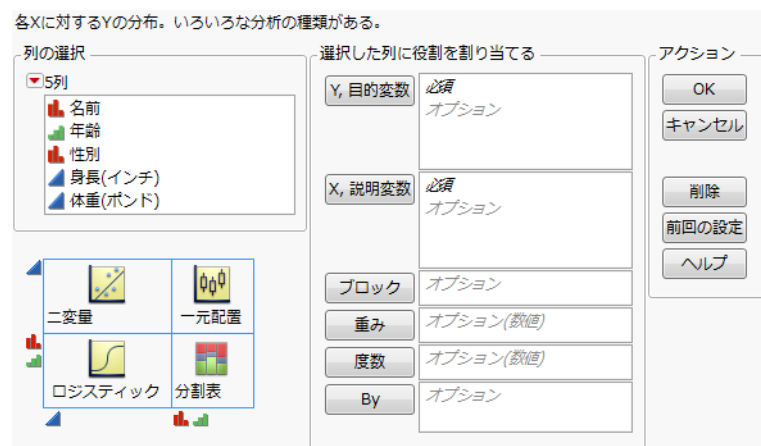
- 「あてはめの要約」レポートは、分析に関する全体的な要約情報をまとめたものです。
- 「分散分析」レポートには、分散分析(ANOVA; Analysis of Variance)の基本的な情報が表示されます。この例では「Prob>F」(p値)が0.0053です。これは、薬の平均には有意差がありそうだという、グラフからの直観的な判断が正しいことを裏付けています。
- 「各水準の平均」レポートには、カテゴリカルな因子の水準ごとに、平均、標本サイズ、標準誤差が表示されます。

「一元配置」プラットフォームの起動

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

- 「二変量の関係」プラットフォームを起動するには、[分析] > [二変量の関係] を選択します。
または
- 「一元配置」プラットフォームを起動するには、「JMP スターター」ウィンドウのカテゴリリストで[基本統計] をクリックし、[一元配置] をクリックします。

図6.4 「一元配置」起動ウィンドウ



この起動ウィンドウの詳細については、「[「二変量の関係」プラットフォームの概要](#)」章（93ページ）を参照してください。

[OK] をクリックすると、「一元配置分析」レポートウィンドウが開きます。「[一元配置プロット](#)」（138ページ）を参照してください。

一元配置プロット

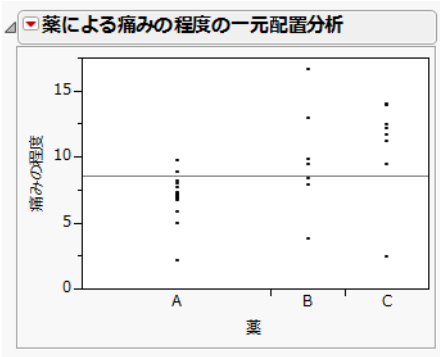
一元配置プロットは、説明変数（X）の値ごとに、応答変数（Y）の点の分布を示します。Yの分布を、Xの水準間で比較できます。Xにおける一意な各値は、「水準」とも言います。

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

「一元配置分析」の赤い三角ボタンをクリックし、メニューから各種オプションを選択して、レポート、プロット、検定結果をレポートウィンドウに追加できます。「[「一元配置」プラットフォームのオプション](#)」(139ページ)を参照してください。

図6.5は、「[一元配置分析の例](#)」(136ページ)の手順に従って作成したものです。

図6.5 一元配置プロット



メモ: データテーブルで非表示にした行は、一元配置プロットでも非表示になります。

「一元配置」プラットフォームのオプション

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

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

表6.1 オプションと要素の例

プラットフォーム オプション	プロットに追加されるグラフ要素	レポートウィンドウに追加される結果
分位点	箱ひげ図	「分位点」レポート
平均/ANOVA	平均のひし形	「一元配置の分散分析」(ANOVA) レポート
平均と標準偏差	平均線、平均誤差バー、標準偏差線	「平均と標準偏差」レポート

表 6.1 オプションと要素の例（続き）

プラットフォーム オプション	プロットに追加されるグラフ要素	レポートウィンドウに追加される結果
平均の比較	比較円 ([ノンパラメトリックな多重比較] オプションを除く)	「平均の比較」レポート

「一元配置分析」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。特定の条件下でなければ表示されないオプションもあります。

分位点 グループごとに次の分位点が表示されます。

- 0%（最小値）
- 10%
- 25%
- 50%（中央値）
- 75%
- 90%
- 100%（最大値）

[表示オプション] サブメニューの [箱ひげ図] オプションが選択状態になります。[「分位点」](#)（144 ページ）を参照してください。

平均/ANOVA 各グループの平均が求められ、一元配置分散分析によって平均間に差があるかどうかを検定されます。[「平均/ANOVA および平均/ANOVA/プーリングしたt検定」](#)（145 ページ）を参照してください。

X 変数の水準が2つの場合は、このオプションは [平均/ANOVA/プーリングしたt検定] という名前で表示されます。

平均と標準偏差 グループごとに要約統計量が表示されます。ここで計算される平均の標準誤差は、各グループの誤差分散がすべて等しいと仮定した推定値ではなく、グループごとに計算された標準偏差が使われます。

プロットには、平均線、平均誤差バー、標準偏差線が表示されます。これらの要素の簡単な説明については、[「表示オプション」](#)（142 ページ）を参照してください。さらに詳しい解説については、[「平均線、平均誤差バー、標準偏差線」](#)（149 ページ）を参照してください。

個々の分散を用いたt検定 分散が等しくないと仮定したときの t 検定が計算されます。[「t 検定」レポート](#)（146 ページ）を参照してください。

このオプションは、X 変数の水準数が2つの場合にのみ表示されます。

平均分析法 サブメニューが開き、平均分析（Analysis of Means）を実行するコマンドが5つ表示されます。これらは、平均、分散、および範囲を比較するためのコマンドです。「[平均分析法](#)」（150ページ）を参照してください。

平均の比較 サブメニューが開き、複数のグループ平均を比較するための多重比較法が表示されます。「[平均の比較](#)」（153ページ）を参照してください。

ノンパラメトリック サブメニューが開き、グループの位置をノンパラメトリックに比較するための方法が表示されます。「[ノンパラメトリック検定](#)」（159ページ）を参照してください。

等分散性の検定 4種類の等分散性の検定を実行します。さらに Welch 検定も行います。Welch 検定は、グループ内の分散が等しくないときに、平均を比較する分散分析です。「[等分散性の検定](#)」（165ページ）を参照してください。

同等性の検定 平均の差が、特定の閾値より小さいことを検定します。「[同等性の検定](#)」（167ページ）を参照してください。

ロバスト 外れ値の影響を受けにくい2つの手法が用意されています。「[ロバスト](#)」（167ページ）を参照してください。

検出力 特定の仮説検定に関して、検出力などを計算します。「[検出力](#)」（168ページ）を参照してください。

「検出力の詳細」ウィンドウとレポートは、「モデルのあてはめ」プラットフォームでも表示されます。検出力の計算の詳細および例については、『基本的な回帰モデル』の付録「統計的詳細」を参照してください。

α 水準の設定 サブメニューから、よく使われている α 水準を選択するか、[その他] を選択して任意の α 水準を入力します。 α 水準を変更すると、次の処理が行われます。

- 信頼限界が再計算されます。
- プロット上に平均のひし形が表示されている場合は調整されます。
- レポート内の上側信頼限界と下側信頼限界の値が変更されます。
- すべての「平均の比較」レポートの棄却値と比較円が変更されます。
- すべての「ノンパラメトリックな多重比較」レポートの棄却値が変更されます。

正規分位点プロット サブメニューが開き、各グループの分位点プロットに関する次のオプションが表示されます。

- [分位点-実測値プロット] を選択すると、X 軸が正規分布の分位点、Y 軸が実測値を表す分位点プロットが作成されます。カテゴリカルな X 因子の水準ごとの分位点プロットが表示されます。
- [実測値-分位点プロット] は、[分位点-実測値プロット] の X 軸と Y 軸を逆にしたものです。
- [あてはめ線] を選択すると、X 変数の水準ごとに、プロット上に参照線として直線が引かれます。このオプションは、プロットを作成（[分位点-実測値プロット] または [実測値-分位点プロット] を選択）してからでないと使用できません。

累積確率プロット グループごとの累積分布関数がプロットされます。「[累積確率プロット](#)」(170 ページ) を参照してください。

密度 グループ間の密度を比較します。「[密度](#)」(170 ページ) を参照してください。

対応のある列を設定 対応のあるモデルの分析を行うときに、対応のある変数を指定します。このオプションは、異なるグループのオブザベーションが同じ個体から取られた場合など、一元配置分析で使うデータが対応のあるデータから取り出されたものである場合に使用します。

プロット上では、対応のある点が線でつながれます。「[対応のある列を設定](#)」(171 ページ) を参照してください。

保存 現在のデータテーブルの新しい列に、次の数値を保存します。

- **【残差の保存】** では、X の各水準ごとに、Y から Y の平均を引いた値が保存されます。
- **【標準化データの保存】** では、X の各水準ごとに標準化された Y の値が保存されます。これは、各水準ごとに、中心化した Y を標準偏差で割ったものです。
- **【正規分位点の保存】** では、カテゴリカルな X 変数の各水準内で計算された正規分位点保存されます。
- **【予測値の保存】** では、X の水準ごとに応答変数の平均が保存されます。

表示オプション プロットに対して要素を追加または削除します。「[表示オプション](#)」(142 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。

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。

表示オプション

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

すべてのグラフ プロット全体の表示／非表示を切り替えます。

点 プロット上のデータ点の表示／非表示を切り替えます。

箱ひげ図 各グループの箱ひげ図の表示／非表示を切り替えます。例として、「[一元配置分析の実行](#)」(196 ページ) を参照してください。

平均のひし形 ひし形の中央にある水平線は、各グループの平均に位置しています。また、その長さは、X軸の長さに比例しています。ひし形の上下の点は、グループの上側と下側の95%信頼限界を表します。「[平均のひし形と標本サイズに比例したX軸](#)」(148ページ)を参照してください。

平均線 各グループの平均を表す線が引かれます。「[平均線、平均誤差バー、標準偏差線](#)」(149ページ)を参照してください。

平均の信頼区間 各グループ平均の上側と下側の95%信頼限界に線が引かれます。

平均誤差バー 各グループの平均と、平均から上下へ1標準誤差の区間が表示されます。「[平均線、平均誤差バー、標準偏差線](#)」(149ページ)を参照してください。

全体平均 プロット上に、Y変数の全体平均を示す線が引かれます。

標準偏差線 各グループの平均から上下へ1標準偏差だけ離れたところに、線が表示されます。「[平均線、平均誤差バー、標準偏差線](#)」(149ページ)を参照してください。

比較円 比較円の表示／非表示を切り替えます。このオプションは、[[平均の比較](#)]のオプションが選択されている場合にだけ使用できます。「[比較円](#)」(198ページ)を参照してください。例として、「[一元配置分析の実行](#)」(196ページ)を参照してください。

平均をつなぐ グループの平均をつなぐ直線が表示されます。

平均の平均 グループ平均の平均を表す線が引かれます。

標本サイズに比例したX軸 X軸が、各水準の標本サイズに比例した長さに区切られます。「[平均のひし形と標本サイズに比例したX軸](#)」(148ページ)を参照してください。

点の拡散 点がX軸方向に拡散して表示されます。

点をずらす Yの値が等しいために重なってしまっている点を少しずつずらして表示します。
 $4 * (\text{一様分布乱数} - 0.5)^5$ の乱数により、-0.125～0.125の範囲で点が水平方向に移動します。

対応を表す線 ([[対応のある列を設定](#)] オプションが選択されている場合にのみ表示されます)。対応のある点を線でつなぎます。

対応を表す点線 ([[対応のある列を設定](#)] オプションが選択されている場合にのみ表示されます)。テーブル内の欠損セルのセル平均を点線でつなぎます。線の終点は、二元配置の分散分析モデルで計算されます。

ヒストグラム 元のプロットの隣に、横に並べたヒストグラムが表示されます。

ロバスト平均の線 ([[ロバスト](#)] のオプションが選択されている場合にのみ表示されます)。各グループのロバスト平均を表す線が引かれます。

凡例 正規分位点プロット、累積確率プロット、密度のオプションに対して、凡例を表示します。

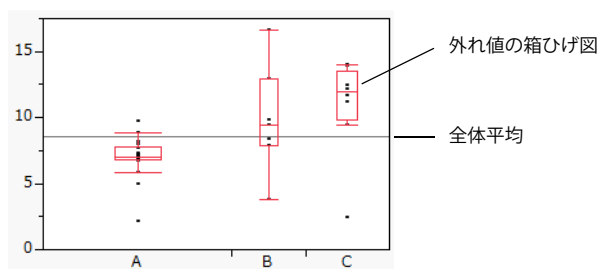
分位点

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

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

- Y 変数の全体平均を示す線
- X の水準別に点の分布を要約した箱ひげ図

図6.6 外れ値の箱ひげ図と全体平均



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

外れ値の箱ひげ図

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

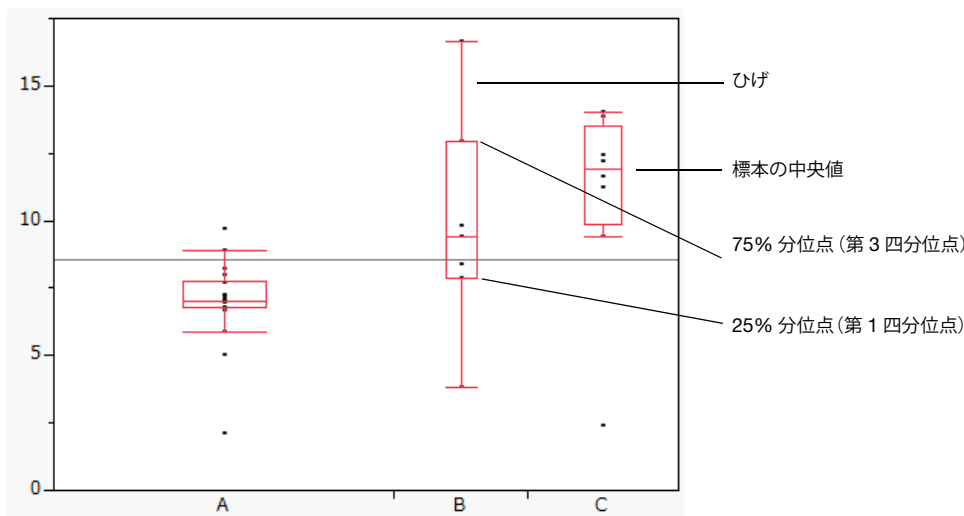
- 箱の中の横線は標本の中央値（メディアン）を示します。
- 箱の両端は25%と75%の分位点で、それぞれ第1および第3四分位点とも呼ばれます。
- 第1四分位点と第3四分位点の差が**四分位範囲**です。
- どの箱にも両端に**ひげ**と呼ばれる線が付いています。ひげは、箱の端から、次の式で計算された範囲内で最も遠くにある点まで伸びています。

$$\text{第3四分位点} + 1.5 * (\text{四分位範囲})$$

$$\text{第1四分位点} - 1.5 * (\text{四分位範囲})$$

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

図6.7 外れ値の箱ひげ図の例



平均/ANOVA および平均/ANOVA/ プーリングしたt検定

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

一元配置プロットの平均のひし形「表示オプション」（142ページ）および「平均のひし形と標本サイズに比例したX軸」（148ページ）を参照してください。

レポート 「「あてはめの要約」レポート」（145ページ）、「分散分析」レポート」（147ページ）、「各水準の平均」レポート」（148ページ）、「t検定」レポート」（146ページ）、および「ブロック平均」レポート」（148ページ）を参照してください。

- 「t検定」レポートは、[平均/ANOVA/ プーリングしたt検定] オプションを選択した場合にのみ表示されます。
- 「ブロック平均」レポートは、起動ウィンドウでブロック変数を指定した場合にのみ表示されます。

「あてはめの要約」レポート

「あてはめの要約」レポートは、一元配置の分散分析の結果をまとめたものです。

R²乗 Yの変動のうち、Xの各水準に平均をあてはめることにより説明される変動の割合を表します。残りの変動は、モデルで説明できないため、ランダムな誤差とみなされます。R²が1ということは、グループ

平均をあてはめることで変動のすべてが説明され、誤差がないことを意味します。 R^2 が0だと、このあてはめが予測モデルとして全体平均より優れていないことを示します。詳細は、「[「あてはめの要約」レポート](#)」(200ページ)を参照してください。

R^2 は**決定係数**とも言います。

メモ: R^2 乗の値が小さい場合は、変動を説明する変数が、モデルに含まれていない可能性があります。ただし、扱っているデータ自体の変動が大きい場合は、意味のある回帰モデルであっても、 R^2 乗の値が小さくなることがあります。研究分野により求められる R^2 乗値が異なる場合がありますので、その分野に関する文献を参考にしてください。

自由度調整 R^2 乗 パラメータ数異なるモデルでも比較できるように、自由度によって R^2 を調整したもの。詳細は、「[「あてはめの要約」レポート](#)」(200ページ)を参照してください。

誤差の標準偏差 (RMSE) ランダムな誤差の標準偏差の推定値です。「分散分析」レポートにある「誤差」の平均平方の平方根です。

応答の平均 Y 変数の全体平均 (算術平均)。

オブザベーション (または重みの合計) あてはめを推定するために使われたオブザベーションの数。ただし、重み変数が使われた場合は、重みの合計です。「[「あてはめの要約」レポート](#)」(200ページ)を参照してください。

「t検定」レポート

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

t検定には2種類あります。

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

「t検定」レポートには、次の情報が表示されます。

t検定プロット 帰無仮説が真であると仮定したときの、平均の差の標本分布を示します。赤色の縦線は平均の実際の差を表しています。陰影が付いた領域はp値に相当します。

差 2つのX水準に見られる差の推定値を表示します。プロットでは、2水準を比較した「差」の値が赤い線で示されます。

差の標準誤差 差の標準誤差を示します。

差の上側信頼限界 差の上側信頼限界を示します。

差の下側信頼限界 差の下側信頼限界を示します。

信頼率 信頼水準 ($1 - \alpha$) を示します。信頼水準を変更するには、プラットフォームの赤い三角ボタンをクリックし、[α 水準の設定] サブメニューから新しい α 水準を選択します。

t値 t 検定統計量の値。

自由度 t 検定で使用された自由度。

p値 (Prob>|t|) 両側検定の p 値。

p値 (Prob>t) 下側の片側検定の p 値。

p値 (Prob<t) 上側の片側検定の p 値。

「分散分析」レポート

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

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

要因 変動の3つの要因（モデル要因、「誤差」、「全体 (修正済み)」）。

自由度 各変動因の自由度 (DF)。

- 「**全体 (修正済み)**」の自由度は、 $N - 1$ で求められます。 N は分析に使用されたオブザベーションの数です。
- X 変数の水準数を k とすると、モデルの自由度は $k - 1$ になります。
- 「**誤差**」の自由度は、「**全体 (修正済み)**」とモデルの自由度の差 ($N - k$) です。

平方和 変動の各要因の平方和 (SS)。

- 全体（「**全体 (修正済み)**」）の平方和は、各応答から全体平均までの距離の平方和です。「**全体 (修正済み)**」の平方和は、基本モデルの平方和として、その他のモデルとの比較に使用されます。
- 「**誤差**」（残差）の平方和は、各点からそれぞれのグループ平均までの距離の平方和です。この平方和は、分散分析モデルをあてはめた後に残った、説明できない部分の平方和です。

全体平方和から誤差平方和を引いたものがモデルの平方和です。全変動のうちモデルによって説明されている部分を表します。

平均平方 平方和に関連する自由度で割った値。

- モデルの平均平方は、グループ平均が等しいという仮説のもとで推定した誤差の分散です。
- 「**誤差**」の平均平方は、モデルの平均平方とは無関係に推定した誤差の分散で、モデルの仮説による制約を受けません。

F 値 モデルの平均平方を「誤差」の平均平方 (MSE) で割ったもの。グループ平均が等しい (実際に差がない) という仮説が真のとき、誤差とモデルの平均平方が両方とも誤差分散の推定値となり、その比は F 分布に従います。分散分析モデルをあてはめた結果、誤差の変動が全体の変動よりも有意に減少した場合、「 F 値」は期待値よりも大きくなります。

p 値 (Prob>F) グループ平均の間に実際には差がないときに、まったくの偶然だけで現在の F 値よりも大きな F 値を得る確率。観測有意確率が 0.05 以下だと、グループ平均に有意差がある証拠だとみなされます。

「各水準の平均」レポート

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

水準 X 変数の水準。

数 各グループ内のオブザベーションの個数。

平均 各グループの平均。

標準誤差 グループ平均の標準偏差の推定値。標準誤差は、 Y の分散がどの水準でも等しいという仮定のもとに計算されます。この標準誤差は、「あてはめの要約」レポートにある誤差の標準偏差 (RMSE) を、グループごとの標本サイズの平方根で割ったものです。

下側 95% と上側 95% グループ平均の下側 95% 信頼限界と上側 95% 信頼限界。

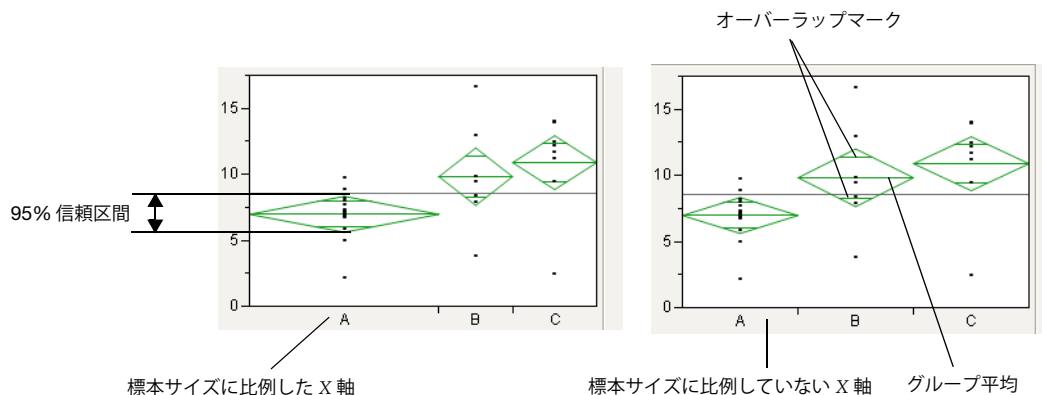
「ブロック平均」レポート

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

平均のひし形と標本サイズに比例した X 軸

平均のひし形は、標本平均とその信頼区間を表します。

図6.8 平均のひし形と標本サイズに比例したX軸の例



次の点を確認してください。

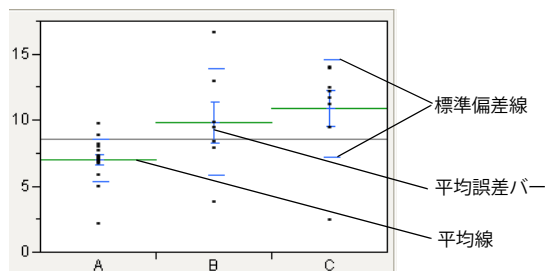
- 各ひし形の上下の点が、各グループの $(1 - \alpha) \times 100\%$ の信頼区間を表します。信頼区間は、分散がすべてのオブザベーションにわたって等しいという前提のもとに計算されるので、ひし形の高さは、グループ内にあるオブザベーション数の平方根の逆数に比例します。
- 【標準サイズに比例した X 軸】 オプションが選択されているときは、X 軸に沿った各グループの幅（ひし形の横幅）が、そのグループ（X 変数の各水準）の標本サイズに比例します。そのため、幅の狭いひし形は縦長になる傾向があります。それは、データ点の数が少ないと信頼区間が広くなるためです。
- 各ひし形の中央を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

平均線、平均誤差バー、標準偏差線

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

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

図6.9 平均線、平均誤差バー、標準偏差線



平均分析法

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

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

メモ：「分割表」プラットフォームでは、応答のカテゴリが2つの場合は「[割合の平均分析](#)」を使用できます。詳細については、「[分割表分析](#)」章（205ページ）を参照してください。

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

位置の指標に対する平均分析

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

- 平均分析
- 順位変換平均分析

平均分析

グループの平均を全体平均と比較するには、「平均分析」を使用します。この分析法では、データがほぼ正規分布に従っているものと仮定します。「[平均分析法の例](#)」（171ページ）を参照してください。

順位変換平均分析

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

順位変換

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

- すべての観測値に小さい値から順番に順位を与えます。この時、同順位も考慮します。同順位のものには、平均順位を割り当てます。
- 順位を $R_1, R_2, \dots, R_n$ とします。
- i 番目の観測値に対しての変換は、次のように求められます。

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

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

尺度の指標に対する平均分析

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

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

分散の平均分析

グループの標準偏差（または分散）を誤差の標準偏差（RMSE）（またはMSE）と比較するには、[分散の平均分析] を使用します。この分析法では、データがほぼ正規分布に従っているものと仮定します。この方法を使用するには、各グループのオブザベーション数が4個以上でなければなりません。分散の平均分析の詳細については、Wludyka and Nelson (1997) および Nelson et al. (2005) を参照してください。例として、[「分散に対する平均分析法の例」](#) (172 ページ) を参照してください。

分散の平均分析-Levene の ADM

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

範囲の平均分析

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

メモ：「範囲の平均分析」は、バランス（釣り合い）のとれたデータ、および、特定の標本サイズにのみ使用できます。「[「範囲の平均分析」の制限](#)」（152ページ）を参照してください。

「範囲の平均分析」の制限

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

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

平均分析のグラフ

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

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

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

平均分析のオプション

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

有意水準の設定 サブメニューから、よく使われている α 水準を選択するか、**【その他】**を選択して任意の α 水準を入力します。有意水準を変更すると、上側決定限界と下側決定限界が変更されます。

メモ：「範囲の平均分析」の場合、選択できるのは 0.10、0.05、および 0.01 のみです。

要約レポートの表示 これらのレポートは、平均分析法に基づいています。

- 「平均分析」の場合は、グループ平均と決定限界を示すレポートが作成されます。
- 「順位変換平均分析」の場合は、順位のグループ平均と決定限界を示すレポートが作成されます。
- 「分散の平均分析」の場合は、グループごとの標準偏差（もしくは、分散）と決定限界を示すレポートが作成されます。
- 「分散の平均分析-LeveneのADM」の場合は、ADMのグループ平均と決定限界を示すレポートが作成されます。
- 「範囲の平均分析」の場合は、グループ範囲と決定限界を示すレポートが作成されます。

Y軸スケールを分散に（「分散の平均分析」のみ）Y軸のスケールを標準偏差から分散に変更します。

表示オプション 次の表示オプションがあります。

- 「決定限界の表示」は、決定限界線の表示／非表示を切り替えます。
- 「決定限界の陰影の表示」は、決定限界の陰影の表示／非表示を切り替えます。
- 「中心線の表示」は、中心線の表示／非表示を切り替えます。
- 「点のオプション」の「垂線の表示」は、垂線を表示します。これがデフォルトのオプションです。「接続線の表示」は、各グループの平均をつなぐ線を表示します。「点のみを表示」は、各グループの平均を示す点のみを表示します。

平均の比較

メモ：平均の比較方法としては、平均分析（ANOM）法もあります。「[平均分析法](#)」（150ページ）を参照してください。

「平均の比較」の各オプションを使用して、グループ平均の多重比較を実行できます。これらのオプションの結果は、誤差分散が等しいと仮定して、プーリングした分散推定値を使っています。各オプションを選択すると、プロットの横に比較円が追加され、レポートウィンドウに特定のレポートが追加されます。比較円の詳細については、「[比較円の使用](#)」（155ページ）を参照してください。

オプション	説明	[ノンパラメトリック] メニューのオプション
各ペア, Studentのt検定	Studentのt検定を使用してペアごとの比較が計算されます。ペアごとの検定をたくさん行くと、仮説検定における実際の α 水準（第1種の誤り）が、検定に使用する名目的な α 水準よりも大きくなります。「各ペア, Studentのt検定」（156ページ）を参照してください。	[ノンパラメトリック] > [ノンパラメトリックな多重比較] > [ペアごと Wilcoxon 検定]
すべてのペア, TukeyのHSD検定	平均間のすべての差に対して検定が行われます。これはTukeyのHSD (honestly significant difference) 検定、またはTukey-KramerのHSD検定と呼ばれます (Tukey 1953, Kramer 1956)。この検定は、グループごとの標本サイズが同じときには、指定された有意水準の正確な検定です。グループごとの標本サイズが異なるときは、保守的な結果になります (Hayter 1984)。「すべてのペア, TukeyのHSD検定」（156ページ）を参照してください。	[ノンパラメトリック] > [ノンパラメトリックな多重比較] > [すべてのペア Steel-Dwass 検定]
最適値との比較 (HsuのMCB)	平均が未知の最大平均より小さいかどうか、未知の最小平均より大きいかが検定されます。これはHsuのMCB検定と呼ばれます (Hsu (1996) および Hsu (1981))。「最適値との比較 (HsuのMCB)」（157ページ）を参照してください。	なし
コントロール群との比較 (Dunnett)	平均がコントロール群の平均と異なるかどうか検定されます。これはDunnettの検定と呼ばれます (Dunnett 1955)。「コントロール群との比較 (Dunnett)」（158ページ）を参照してください。	[ノンパラメトリック] > [ノンパラメトリックな多重比較] > [コントロール群との比較 Steel 検定]

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

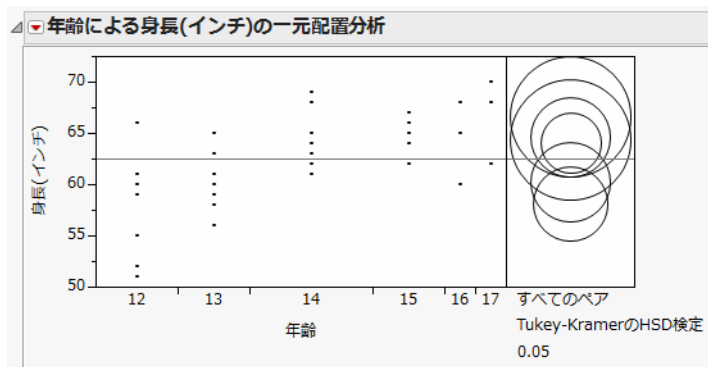
関連情報

- 「平均の比較」で用意されているすべての検定」（180ページ）

比較円の使用

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

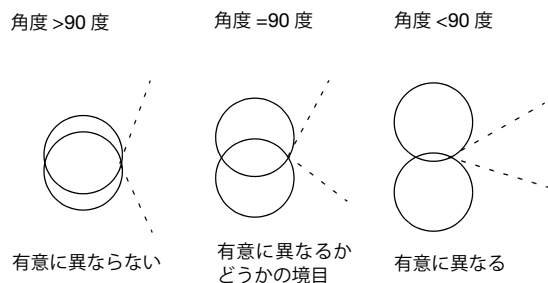
図6.10 グラフによるグループ平均の比較



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

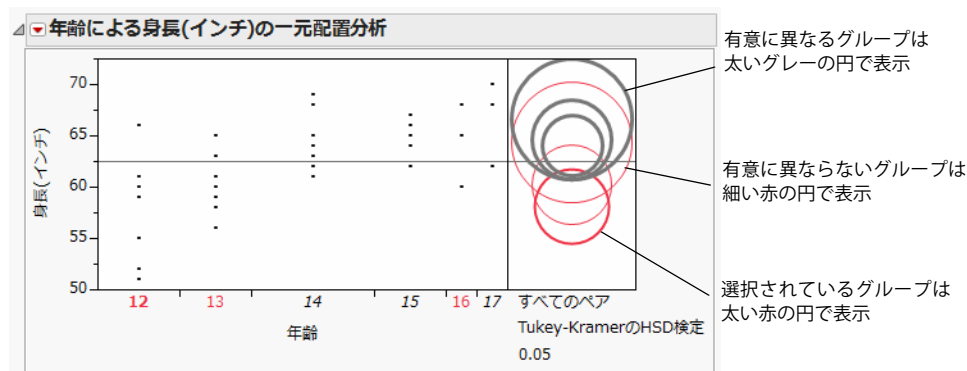
- 平均が有意に異なる場合、比較円は交わらないか交わりが浅く、外側の交角が90度以下になります。
- 比較円が90度以上の交角で交わっているか入れ子になっている場合は、平均が有意には異なることを意味します。

図6.11 交角と有意性



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

図6.12 比較円の選択



関連情報

- 「[比較円](#)」(198 ページ)

各ペア, Student の t 検定

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

関連情報

- 「[「各ペア, Student の \$t\$ 検定」の例](#)」(173 ページ)

すべてのペア, Tukey の HSD 検定

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

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

関連情報

- 「[「すべてのペア, Tukey の HSD 検定」の例](#)」(175 ページ)

最適値との比較 (Hsu の MCB)

【最適値との比較 (Hsu の MCB)】では、特定の水準の平均が、それ以外の水準の平均の最大値または最小値とどう異なるかを調べます。Hsu (1996) を参照してください。

Hsu の MCB 検定の分位点は、カテゴリカル変数の水準によって異なります。各水準の標本サイズが等しくない場合は、比較円のグラフは正確とはいえません。比較円の半径は、水準の標準誤差と、最大分位点を掛け合わせたものになります。正確に有意差を評価したい場合には、グラフではなく、検定の p 値を見てください。「最小/最大との比較」(157 ページ) を参照してください。

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

棄却限界値

このレポートは、カテゴリカル変数の各水準の分位点を示します。これらの値は、特定の Alpha の値に対応します。

最小/最大との比較

Hsu の MCB 検定における「最小/最大との比較」レポートの p 値は、片側 Dunnett 検定によって計算されています。標本平均が最適値 (最大値もしくは最小値) ではない水準の p 値は、その水準を Dunnett 検定のコントロール群とし、最適値となっている水準の平均と比較しています。最適値となっている水準の p 値は、その群をコントロール群とし、2 番目に最適な平均の水準と比較しています。

レポートには次の 3 つの情報が表示されます。

水準 カテゴリカル変数の水準。

p 値 (vs. 最大) カテゴリカル変数の各水準に対して、その水準と、その水準以外のなかで最も平均が大きい水準とを比較する p 値です。未知の) 真の平均の最大値以下に有意になっている平均を選ぶために使用できます。

p 値 (vs. 最小) カテゴリカル変数の各水準に対して、その水準と、それ以外の水準のなかで最も平均が小さい水準とを比較する p 値です。(未知の) 真の平均の最小値以上に有意になっている平均を選ぶために使用できます。

LSD 閾値行列

最大値に関するレポートと、最小値に関するレポートが順に表示されます。

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

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

関連情報

- 「[最適値との比較(HsuのMCB)]の例」(177ページ)

コントロール群との比較(Dunnett)

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

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

関連情報

- 「[コントロール群との比較(Dunnett)]の例」(178ページ)

「平均の比較」のオプション

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

差の行列 平均の差を、表にまとめて表示します。

棄却限界値 信頼区間に使用される t 値、または該当する他の分位点を表示します。

LSD閾値行列 それぞれの比較における差が最小有意差(LSD; Least Significant Difference)を上回るかどうかを示す行列を表示します。

文字の接続レポート 平均の有意差をアルファベットで示した、伝統的なレポートが作成されます。同じ文字でつながっていないグループの平均は、有意に異なります。

メモ: このオプションは、[最適値との比較(HsuのMCB)]と[コントロール群との比較(Dunnett)]では、表示されません。

差の順位レポート 正の値を取る差とその信頼区間、およびそれらのグラフが表示されます。差の信頼区間にゼロが含まれていない場合、平均が有意に異なることを示しています。

メモ: このオプションは、[最適値との比較(HsuのMCB)]と[コントロール群との比較(Dunnett)]では、表示されません。

詳細な比較レポート 各比較の詳細なレポートを表示します。比較ごとに、水準間の差、標準誤差と信頼区間、 t 値、 p 値、および自由度が表示されます。検定結果を図示したグラフが、各レポートの右側に表示されます。

メモ: このオプションは、[すべてのペア, Tukey の HSD 検定]、[最適値との比較 (Hsu の MCB)]、[コントロール群との比較 (Dunnett)] では、表示されません。

ノンパラメトリック検定

ノンパラメトリックな検定は、通常の分散分析で前提とされる正規性が成り立たない場合に役立ちます。[ノンパラメトリック] オプションにはいくつかの手法が用意されており、グループの平均または中央値がすべてのグループで同じ位置にあるかどうかを検定できます。

ノンパラメトリックな多重比較では、ペアの比較を繰り返すことによる全体的な誤りの確率も調整されます。ノンパラメトリックな検定では、「順位スコア」と呼ばれる、Y 変数の順位の関数を使用されます。Hajek (1969) および SAS Institute (2008) を参照してください。

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

Wilcoxon 検定 Wilcoxon の順位スコアに基づく検定を実行します。Wilcoxon の順位スコアは、データの順位そのものです。Wilcoxon の検定は、誤差がロジスティック分布に従っている場合に、最も検出力の高い順位検定となります。因子の水準が3つ以上の場合には、Kruskal-Wallis 検定が実行されます。このレポートについては、「[Wilcoxon 検定、メディアン検定、Van der Waerden 検定のレポート](#)」(160 ページ) を参照してください。例として、「[ノンパラメトリックな Wilcoxon 検定の例](#)」(180 ページ) を参照してください。

Wilcoxon 検定は Mann-Whitney 検定とも呼ばれます。

メディアン検定 メディアン順位スコアに基づく検定を実行します。メディアン順位スコアは、順位がメディアン（中央値）の順位の上か下かによって 1 または 0 の値を取ります。メディアン検定は、誤差が二重指数分布に従っている場合に、最も検出力の高い順位検定となります。このレポートについては、「[Wilcoxon 検定、メディアン検定、Van der Waerden 検定のレポート](#)」(160 ページ) を参照してください。

Van der Waerden の検定 Van der Waerden の順位スコアに基づく検定を実行します。Van der Waerden の順位スコアは、データの順位をオブザベーション数+1 で割り、正規分布関数の逆関数を使って正規スコアに変換したものです。Van der Waerden の検定は、誤差が正規分布に従っている場合に、最強力の順位検定となります。このレポートについては、「[Wilcoxon 検定、メディアン検定、Van der Waerden 検定のレポート](#)」(160 ページ) を参照してください。

Kolmogorov Smirnov 検定 経験分布関数 (EDF) に基づいて、応答の分布がグループ間で同じかどうかを検定します。近似検定と正確検定の両方が行われます。この検定は、X 変数の水準が2つの場合にだけ実行できます。このレポートについては、「[Kolmogorov-Smirnov 二標本検定のレポート](#)」(161 ページ) を参照してください。

JMP PRO **正確検定** サブメニューが開き、Wilcoxon、メディアン、Van der Waerden、Kolmogorov Smirnov の正確検定を実行するためのオプションが表示されます。これらのオプションは、X変数の水準が2つの場合にだけ実行できます。近似検定と正確検定の両方の結果が表示されます。

このレポートについては、「[二標本: 正確検定](#)」(161 ページ) を参照してください。正確 Wilcoxon 検定の例については、「[ノンパラメトリックな Wilcoxon 検定の例](#)」(180 ページ) を参照してください。

Wilcoxon 検定、メディアン検定、Van der Waerden 検定のレポート

どのレポートでも、記述統計量と検定結果が順に表示されます。検定結果は、「一元配置検定 (カイ2乗近似)」レポートに表示されます。X変数の水準が2つである場合は、「2標本検定 (正規近似)」レポートも表示されます。記述統計量は次のとおりです。

水準 Xの水準。

度数 各水準の度数。

スコア和 各水準の順位スコアの和。

スコアの期待値 水準間に差はないという帰無仮説のもとでのスコアの期待値。

スコア平均 各水準の平均順位スコア。

(平均 - 平均0)/標準偏差0 標準化したスコア。「平均0」は帰無仮説のもとでの平均スコアの期待値で、「標準偏差0」は帰無仮説のもとでの平均スコアの標準偏差です。帰無仮説は、グループの平均または中央値がすべてのグループで同じ位置にあるということです。

2標本検定 (正規近似)

Xの水準が2つである場合は、「2標本検定 (正規近似)」レポートが表示されます。このレポートには、以下の情報が表示されます。

S オブザベーション数が少ないほうの水準の順位スコアの和。

Z 正規近似検定の検定統計量。詳細は、「[2標本の場合の正規近似](#)」(203 ページ) を参照してください。

p値 (Prob>|Z|) 標準正規分布に基づく、正規近似検定の p 値。

一元配置検定 (カイ2乗近似)

このレポートは、位置に対するカイ2乗検定の結果を示します。詳細については、Conover (1999) を参照してください。

カイ2乗 カイ2乗検定統計量の値。詳細は、「[一元配置のカイ2乗近似](#)」(204 ページ) を参照してください。

自由度 検定の自由度。

p値 (Prob>ChiSq) 検定の p 値。この p 値は、自由度をXの水準数から1を引いた値としたカイ2乗分布に基づいています。

JMP PRO 二標本: 正確検定

データがまばらであったり、歪んでいたり、同順位が多い場合は、正確検定のほうが、漸近近似検定よりも適している可能性があります。Xが2水準である場合、正確検定の統計量が計算できます。[ノンパラメトリック] > [正確検定] を選択し、いずれかの正確検定を選択してください。「二標本: 正確検定」レポートが表示されます。このレポートには、以下の情報が表示されます。

S オブザベーション数が少ないほうの水準の順位スコアの和。両水準のオブザベーション数が等しい場合は、値の順序で並べて後になる水準を使用します。

p 値 (Prob ≥ S) 片側検定の *p* 値。

p 値 (Prob ≥ |S-Mean|) 両側検定の *p* 値。

Kolmogorov-Smirnov 二標本検定のレポート

Kolmogorov-Smirnov 検定は、Xの水準が2つの場合にだけ実行できます。レポートには、記述統計量と検定結果が順に表示されます。記述統計量は次のとおりです。

水準 Xの2つの水準。

度数 各水準の度数。

最大時の経験分布関数 2つの経験累積分布関数 (EDF) の差が最大になるときの、各水準の EDF の値。「合計」という行には、2つの EDF の差が最大になるときの、プーリングした EDF (データセット全体の EDF) の値が示されます。

最大時の平均からの乖離 各水準について、以下の手順に従って求めた値です。

- 該当の水準と、データセット全体 (合計) のプーリングした「最大時の経験分布関数」の差を求める。
- 上記の差に、その水準のオブザベーション数 (度数) の平方根を掛け合わせる。

漸近 Kolmogorov-Smirnov 検定

このレポートは検定の詳細を示します。

KS Kolmogorov-Smirnov 統計量は次の式で計算されます。

$$KS = \max_j \sqrt{\frac{1}{n} \sum_i n_i (F_i(x_j) - F(x_j))^2}$$

この式では、次のような表記を使用しています。

- x_j ($j = 1, \dots, n$) は、オブザベーション
- n_i は、Xの*i*番目の水準のオブザベーション数
- F は、プーリングした経験累積分布関数

– F_i は、 X の i 番目の水準の経験累積分布関数

この Kolmogorov-Smirnov 統計量は、 X の水準が3つ以上の場合でも求められます。ただし、JMP は、 X の水準が2つの場合にのみ、この分析を実行します。

KSa 漸近的な Kolmogorov-Smirnov 統計量。 $KS\sqrt{n}$ で求められます。ここで、 n は全体のオブザベーション数です。

D=max|F1-F2| 2 群の EDF の差で、その絶対値が最大のもの。2つの標本を比較するために使用される Kolmogorov-Smirnov 統計量です。

p 値(Prob > D) 検定の p 値。水準間に差はないという帰無仮説のもとで、 D が求められた値よりも大きくなる確率です。

D+ = max(F1-F2) 有意であれば、1つ目の水準が2つ目の水準を上回るという対立仮説が採択される片側検定の検定統計量。

p 値(Prob > D+) $D+$ の検定の p 値。

D- = max(F2-F1) 有意であれば、2つ目の水準が1つ目の水準を上回るという対立仮説が採択される片側検定の検定統計量。

p 値(Prob > D-) $D-$ の検定の p 値。

正確 Kolmogrov-Smirnov 検定

正確 Kolmogrov-Smirnov 検定のレポートには、漸近近似検定と同じ統計量が表示されますが、このレポートの p 値は、正確な確率です。

ノンパラメトリックな多重比較

このオプションには、ノンパラメトリックな多重比較のいくつかの検定法が用意されています。これらは順位を用いた検定で、[ペアごと Wilcoxon 検定] 以外では多重比較の調整が行われます。これらの検定の詳細については、Dunn (1964) および Hsu (1996) を参照してください。レポートについては、「[ノンパラメトリックな多重比較の手順](#)」(163 ページ) を参照してください。

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

ヒント： 反復測定データなどに対するノンパラメトリックな Friedman 検定は、JMP では直接サポートされていませんが、次のような方法で実行できます。まず、各ブロック内での順位を計算します。この新しい列を、順序尺度に設定します。そして、「二変量の関係」プラットフォームで、順位の列を Y に指定します。このとき、ブロック変数を指定してください。Cochran-Mantel-Haenszel 検定を実行し、統計量を取得します。

ノンパラメトリックな多重比較の手順

ペアごと Wilcoxon 検定 各ペアに対して Wilcoxon 検定を実行します。この方法では、多重比較の調整を行いません。これは、[平均の比較] メニューの [各ペア, Student の t 検定] オプションのノンパラメトリック版です。「ペアごと Wilcoxon 検定、すべてのペア Steel-Dwass 検定、コントロール群との比較 Steel 検定」(163 ページ) を参照してください。

すべてのペア Steel-Dwass 検定 すべてのペアに対して Steel-Dwass 検定を実行します。これは、[平均の比較] メニューの [すべてのペア, Tukey の HSD 検定] オプションのノンパラメトリック版です。「ペアごと Wilcoxon 検定、すべてのペア Steel-Dwass 検定、コントロール群との比較 Steel 検定」(163 ページ) を参照してください。

コントロール群との比較 Steel 検定 各水準をコントロール水準と比較します。これは、[平均の比較] メニューの [コントロール群との比較 (Dunnett)] オプションのノンパラメトリック版です。「ペアごと Wilcoxon 検定、すべてのペア Steel-Dwass 検定、コントロール群との比較 Steel 検定」(163 ページ) を参照してください。

コントロール群との比較 併合順位の Dunn 検定 [コントロール群との比較 Steel 検定] オプションと同様、各水準をコントロール水準と比較しますが、この Dunn 検定では、比較対象のペアにおける順位でなく、すべてのデータを通じた順位が計算に使われます。表内の p 値には、Bonferroni の調整が施されています。つまり、未調整の p 値に比較の数を掛けたものです。調整済み p 値が 1 を超えた場合は、1 と表示されます。「すべてのペア 併合順位の Dunn 検定、コントロール群との比較 併合順位の Dunn 検定」(164 ページ) を参照してください。

すべてのペア 併合順位の Dunn 検定 [すべてのペア Steel-Dwass 検定] オプションと同様に各ペアを比較しますが、この Dunn 検定では、比較対象のペアにおける順位でなく、すべてのデータを通じた順位が計算に使われます。表内の p 値には、Bonferroni の調整が施されています。つまり、未調整の p 値に比較の数を掛けたものです。調整済み p 値が 1 を超えた場合は、1 と表示されます。「すべてのペア 併合順位の Dunn 検定、コントロール群との比較 併合順位の Dunn 検定」(164 ページ) を参照してください。

ペアごと Wilcoxon 検定、すべてのペア Steel-Dwass 検定、コントロール群との比較 Steel 検定

これらの多重比較のレポートには、検定結果と信頼区間が表示されます。これらの検定では、オブザベーションは比較対象となる 2 つの水準の標本内での順位に変換されます。

q^* 信頼区間の計算に使用される分位点。

Alpha 信頼区間の計算に使用される α 水準。信頼水準は、一元配置分析のメニューから [α 水準の設定] を選択することによって変更できます。

水準 比較対象のペアのうち、1 つ目の水準。

- **水準** 比較対象のペアのうち、2 つ目の水準。

スコア平均の差 1つ目の水準（水準）のオブザベーションの順位スコアの平均から、2つ目の水準（-水準）のオブザベーションの順位スコアの平均を引いた値（連続修正が適用されます）。

1つ目の水準のオブザベーション数を n_1 、2つ目の水準のオブザベーション数を n_2 とします。これらのオブザベーションは、この2水準の標本内で順位付けされます。同順位の場合は、平均順位が使用されます。1つ目の水準の順位の合計を ScoreSum_1 、2つ目の水準の順位の合計を ScoreSum_2 とします。

平均スコアの差が正の数である場合、「スコア平均の差」は次の式で計算されます。

$$\text{スコア平均の差} = (\text{ScoreSum}_1 - 0.5)/n_1 - (\text{ScoreSum}_2 + 0.5)/n_2$$

平均スコアの差が負の数である場合、「スコア平均の差」は次の式で計算されます。

$$\text{スコア平均の差} = (\text{ScoreSum}_1 + 0.5)/n_1 - (\text{ScoreSum}_2 - 0.5)/n_2$$

差の標準誤差 スコア平均の差の標準誤差。

Z 標準化された検定統計量。平均に差がないという帰無仮説のもとで、漸近的に標準正規分布に従います。

p値 Zに基づく漸近検定のp値。

Hodges-Lehmann 位置の差に対するHodges-Lehmann推定値。1つ目の水準のオブザベーションと、2つ目の水準のオブザベーションのすべてのペアの差の中央値です。「上側信頼限界」の右側の棒グラフは、Hodges-Lehmann推定値の大きさを表します。

下側信頼限界 Hodges-Lehmann推定値の下側信頼限界。

メモ: 標本サイズが大きすぎてメモリの問題が発生する場合は、計算されません。

上側信頼限界 Hodges-Lehmann推定値の上側信頼限界。

メモ: 標本サイズが大きすぎてメモリの問題が発生する場合は、計算されません。

~差 各比較のHodges-Lehmann推定値の大きさを示す棒グラフ。

すべてのペア 併合順位のDunn検定、コントロール群との比較 併合順位のDunn検定

これらの比較は、データセットのすべてのオブザベーションの順位に基づいています。[コントロール群との比較 併合順位のDunn検定] では、コントロール群となる水準を選択する必要があります。

水準 比較対象のペアのうち、1つ目の水準。

-水準 比較対象のペアのうち、2つ目の水準。

スコア平均の差 1つ目の水準（水準）のオブザベーションの順位スコアの平均から、2つ目の水準（-水準）のオブザベーションの順位スコアの平均を引いた値（連続修正が適用されます）。ここでの順位は、データセット全体の中の順位です。同順位の場合は、平均順位が使用されます。連続修正については、「[スコア平均の差](#)」(164ページ)を参照してください。

差の標準誤差 スコア平均の差の標準誤差。

Z 標準化された検定統計量。平均に差がないという帰無仮説のもとで、漸近的に標準正規分布に従います。

p値 Zに基づく漸近検定の p 値。

等分散性の検定

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

メモ: 等分散性の検定には、分散の平均分析 (ANOMV) 法も使用できます。[「平均分析法」](#) (150 ページ) を参照してください。

等分散性の検定として、以下のものが用意されています。

O'Brien グループ平均が元の Y 変数のグループごとの不偏分散と等しくなるような従属変数が新しく作成されます。O'Brien 検定は、グループごとの分散に対する ANOVA であると言えます (O'Brien 1979、Olejnik and Algina 1987)。

Brown-Forsythe グループの中央値と各観測値との差の絶対値を応答として ANOVA を実行し、 F 検定を行います (Brown and Forsythe 1974)。

Levene グループの平均と各観測値との差の絶対値を応答として ANOVA を実行し、 F 検定を行います (Levene 1960)。ばらつきが (SAS のデフォルトが $z_{ij}^2 = (y_{ij} - \bar{y}_i)^2$ であるのとは対照的に) $z_{ij} = |y_{ij} - \bar{y}_i|$ として計算されます。

Bartlett 標本分散の重み付きの算術平均を、標本分散の重み付きの幾何平均と比べます。幾何平均は必ず算術平均以下になり、算術平均と幾何平均が等しくなるのは、すべての値が等しい場合です。すべての標本分散が等しいと、それらの算術平均と幾何平均は等しくなります。グループ分散の間に見られる変動が大きいほど、2つの平均の差が大きくなります。2つの平均から計算された値が、 χ^2 分布によって近似されます (χ^2 分布は、ある定式化における F 分布の一種です)。値が大きいということは、算術平均/幾何平均の比が大きい、つまりグループ分散の変動が大きいことを意味します。Bartlett のカイ2乗検定統計量を自由度で割ったものが、表の中の F 値です。Bartlett の検定は、正規性の仮定が満たされないケースに対して頑健ではありません (Bartlett and Kendall 1946)。

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

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

「分散が等しいことを調べる検定」レポート

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

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

「分散が等しいことを調べる検定」レポートの説明

水準 因子の水準。

度数 各水準の度数。

標準偏差 X変数の水準ごとに応答の標準偏差を表示したもの。O'Brien検定における変換された変数のグループ平均は、標準偏差の2乗（分散）です。データが1つしかない水準については、標準偏差は計算されません。

平均からの平均絶対偏差 グループ平均からのYの偏差の絶対値を平均した値。Levene検定における変換された変数のグループ平均は、この平均絶対偏差になります。

中央値からの平均絶対偏差 グループ中央値からのYの偏差の絶対値を平均した値。Brown-Forsythe検定における変換された変数のグループ平均は、この平均絶対偏差になります。

検定 実行された検定の種類。

F値 検定ごとの方法で計算されたF統計量。「[「分散が等しいことを調べる検定」レポート](#)」(201ページ)を参照してください。

分子自由度 各検定の分子で使用される自由度。因子の水準数を k とすると、分子には $k-1$ 個の自由度があります。O'Brien、Brown-Forsythe、およびLeveneの統計量の計算では、データが1つしかない水準は無視されます。その場合の分子自由度は、計算に使用された水準数から1を引いた数だけあります。

分母自由度 各検定の分母で使用される自由度。O'Brien、Brown-Forsythe、およびLeveneでは、検定統計量の計算に使用された因子水準の数だけ自由度が引かれます。因子の水準数を k とすると、分母自由度は $n - k$ となります。

p値 分散がすべての水準にわたって等しい場合に、まったくの偶然だけで現在のF値より大きなF値を得る確率。

「Welchの検定」レポートの説明

F値 等分散性の検定のF統計量。「[「分散が等しいことを調べる検定」レポート](#)」(201ページ)を参照してください。

分子自由度 検定の分子で使用される自由度。因子の水準数を k とすると、分子には $k-1$ 個の自由度があります。データ内で一度しか起こらなかった水準は、Welch の一元配置の計算に使用されません。その場合の分子自由度は、計算に使用された水準数から 1 を引いた数だけあります。

分母自由度 検定の分母で使用される自由度。「[分散が等しいことを調べる検定](#)」レポート」(201 ページ) を参照してください。

p 値 (Prob>F) 平均がすべての水準にわたって等しい場合に、まったくの偶然だけで現在の F 値より大きな F 値を得る確率。有意確率が 0.05 以下の場合には、いくつかの平均が他の平均と異なっているという証拠があるとみなされます。

t 検定 F 値と t 値の関係を示すために出力されます。 t 値は、 F 値の平方根です。この結果は、**X** 変数の水準が 2 つしかない時にだけ表示されます。

関連情報

- 「[等分散性検定の例](#)」(183 ページ)
- 「[分散が等しいことを調べる検定](#)」レポート」(201 ページ)

同等性の検定

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

関連情報

- 「[同等性の検定の例](#)」(184 ページ)

ロバスト

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

ロバストなあてはめ

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

関連情報

- 「〔ロバストなあてはめ〕 オプションの例」（185ページ）

Cauchyのあてはめ

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

検出力

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

- **最小有意値**（LSV: Least Significant Value）は、 p 値が特定の値になるようなパラメータ（もしくはパラメータ関数）の値です。つまり、ある値以下に p 値がなって有意と判断される最小の効果の大きさを表します。最小有意値（LSV）は、確率のスケールではなく、パラメータのスケール上で有意度を測定していると言えます。その値によって、計画とデータがどのぐらい敏感かがわかります。
- **最小有意数**（LSN: Least Significant Number）は、 p 値が特定の値になるような標本サイズです。つまり、Alpha（有意水準）、Sigma（誤差の標準偏差）、Delta（効果の大きさ）が与えられているときに、結果が有意となるほど推定値の分散が減少するために必要な標本サイズが、最小有意数（LSN）です。有意性を得るためにさらに多くのデータが必要な場合は、最小有意数（LSN）を参考にすれば、どれだけのデータが必要かがわかります。最小有意数（LSN）の値だけオブザベーションがあるとき、検出力は約50%になります。
- **検出力**とは、グループ間に実際に差がある場合に、検定が有意となる確率（ p 値 $< \alpha$ となる確率）を指します。これは、標本サイズ、効果の大きさ、誤差の標準偏差、および有意水準の関数です。検出力からは、その実験で、与えられた有意水準において差（効果の大きさ）をどれだけの確に検出できるかがわかります。

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

「検出力の詳細」ウィンドウとレポート

「検出力の詳細」ウィンドウとレポートは、「モデルのあてはめ」プラットフォームで実行される一般的な検出力計算と同じです。検出力の計算については、『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。

「 α 」、「 σ 」、「 δ 」、「数」の各列には値を1～2つ入力するか、連続した値の範囲として開始値、終了値、および間隔を入力します。図6.31を参照してください。入力した値のすべての組み合わせに対して、検出力が計算されます。

- a 0～1の値を取る有意水準（通常は0.05、0.01、0.10のどれかを使用します）。デフォルトでは0.05に設定されています。
- s モデル内の残差の標準誤差（誤差の標準偏差に対する推定値）。デフォルトとして、誤差の平均平方の平方根から推定されたRMSEが表示されています。
- d 標準化していない効果の大きさ（effect size; 効果量）。効果の大きさを計算する方法については、『基本的な回帰モデル』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。最初のフィールドには仮説の平方和を n の平方根で割ったものの平方根、つまり $\delta = \sqrt{SS/n}$ が表示されます。

数 (n) すべてのグループの標本サイズの合計。デフォルトでは、最初のフィールドに現在の標本サイズが表示されます。

検出力を求める 「 α 」、「 σ 」、「 δ 」、および「数」の関数として、検出力（有意な結果が出る確率）が計算されます。

最小有意数を求める 指定された「 α 」、「 σ 」、「 δ 」で、およそ50%の検出力を得るために必要なオブザベーションの個数。

最小有意値を求める p 値が α 以下となるパラメータまたは線形検定の値。これは「 α 」、「 σ 」、「数」、および推定値の標準誤差の関数です。この機能は、X変数の水準数が2つの場合にだけ使用でき、通常は個々のパラメータに適用します。

調整済み検出力と信頼区間 事後的に検出力を調べるときは、誤差の標準偏差と検定するパラメータに対して推定値を使用します

- 推定値を単純に代入してF分布の非心度パラメータを計算するとバイアスが生じるため、バイアスの少ない推定値を使って計算したものが、調整済み検出力です。
- また、調整済み検出力の信頼区間は、非心度推定値の信頼区間に基づいています。

調整済み検出力と信頼限界が計算されるのは、元の「 δ 」の推定値に対してだけで、 δ にその他の値を指定して計算することはできません（調整済み検出力は、 δ の推定値におけるランダムな変動を調整したものです）。

関連情報

- 「[\[検出力\] オプションの例](#)」(186ページ)
- 「[検出力](#)」(200ページ)

正規分位点プロット

2種類の正規分位点プロットを作成できます。

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

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

関連情報

- [「正規分位点プロットの例」](#) (188ページ)

累積確率プロット

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

関連情報

- [「累積確率プロットの例」](#) (189ページ)

密度

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

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

関連情報

- [「【密度】の各オプションの例」](#) (190ページ)

対応のある列を設定

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

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

【対応のある列を設定】オプションでは、主に2つの処理が行われます。

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

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

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

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

関連情報

- 「【対応のある列を設定】オプションの例」（191ページ）

「一元配置」プラットフォームのその他の例

ここでは、「一元配置」プラットフォームの例をさらに紹介します。

平均分析法の例

- 【ヘルプ】>【サンプルデータライブラリ】を選択し、「Analgesics.jmp」を開きます。
- 【分析】>【二変量の関係】を選びます。

3. 「痛みの程度」を選択し、[Y, 目的変数] をクリックします。
4. 「薬」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、メニューから [平均分析法] > [平均分析] を選択します。

図6.13 「平均分析」チャートの例

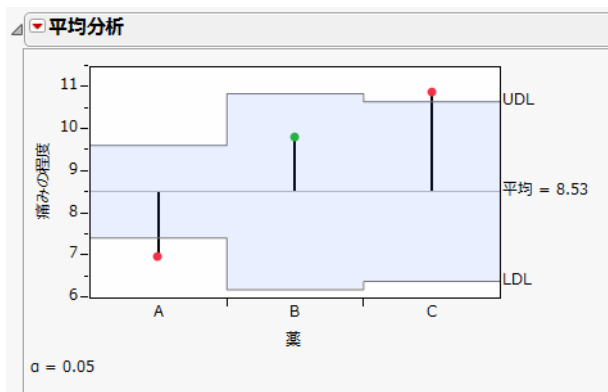


図6.13の例では、薬AとCの平均は、全体平均と有意に異なります。薬Aの平均は低く、薬Cの平均は高くなっています。各薬で標本サイズが異なっているため、薬ごとに決定限界が違ってきます。

分散に対する平均分析法の例

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

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

図6.14 「分散の平均分析」チャートの例

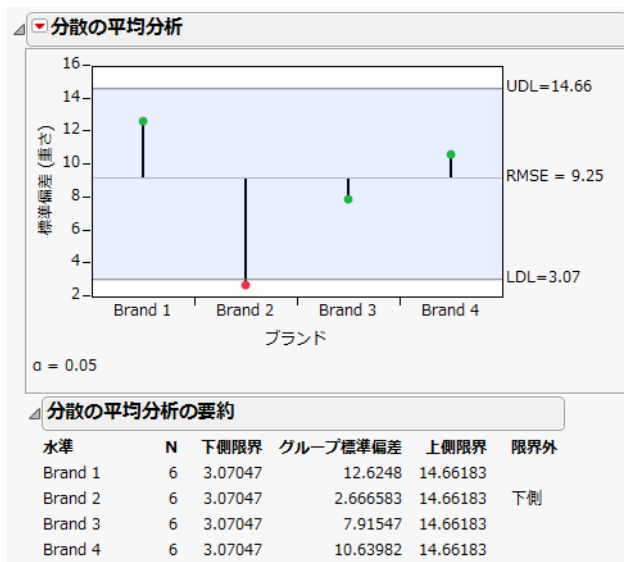


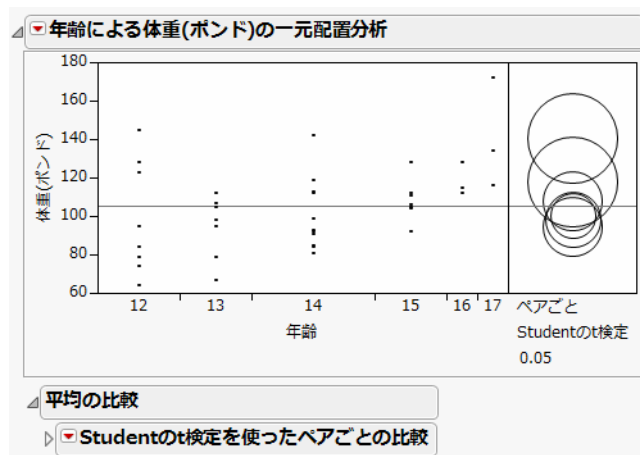
図6.14から、「Brand 2」の標準偏差が下側決定限界を下回っていることがわかります。つまり、「Brand 2」は、他のブランドと比較して、分散が有意に小さいことを示しています。

「各ペア, Studentのt検定」の例

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

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

図6.15 [各ペア, Studentのt検定] の比較円の例



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

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

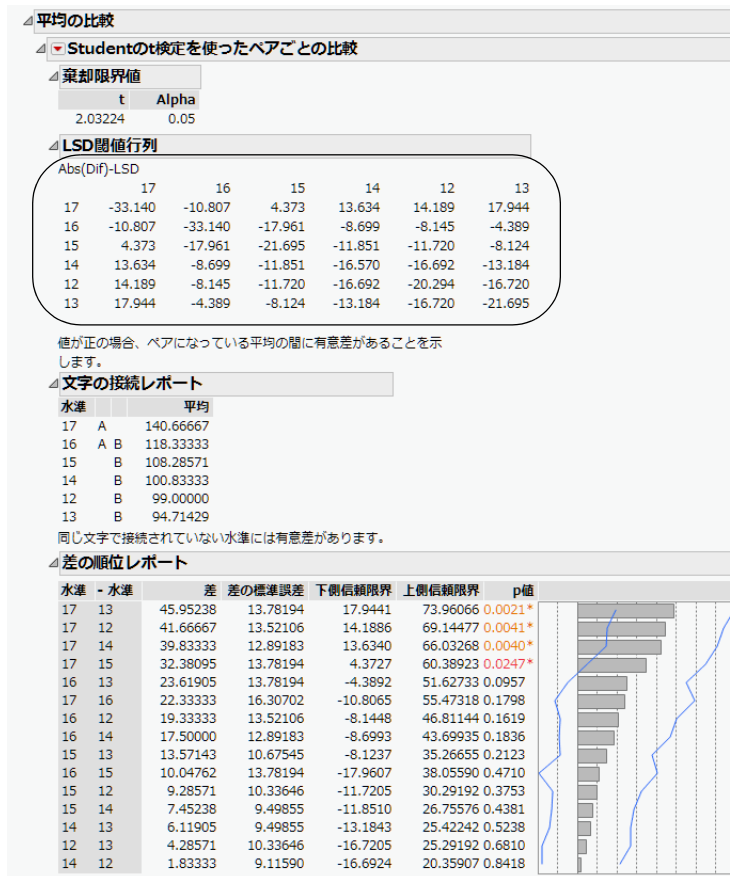


図6.16の先頭に表示されている行列は、観測された平均の差の絶対値から、LSD（最小有意差；検定が有意となる差）を引いたものです。値が正の場合、観測された平均の差がLSDよりも大きいことを意味していますので、2つの平均には有意差があると結論できます。

[すべてのペア, TukeyのHSD検定] の例

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

図 6.17 [すべてのペア,TukeyのHSD 検定] の比較円の例

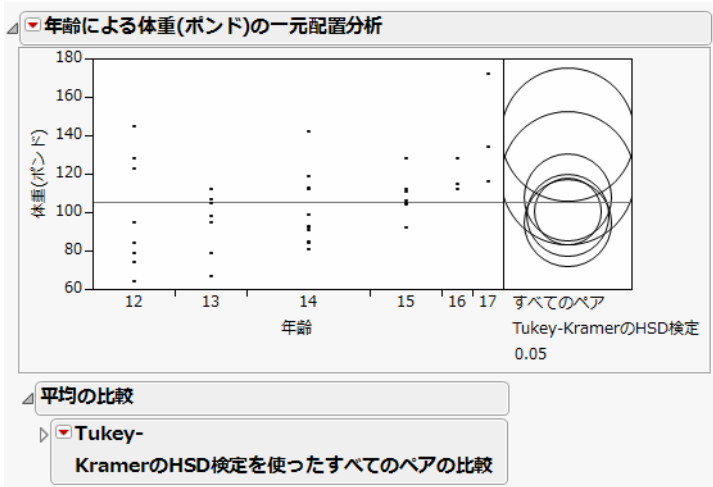


図 6.18 「Tukey-Kramer のHSD 検定を使ったすべてのペアの比較」 レポートの例

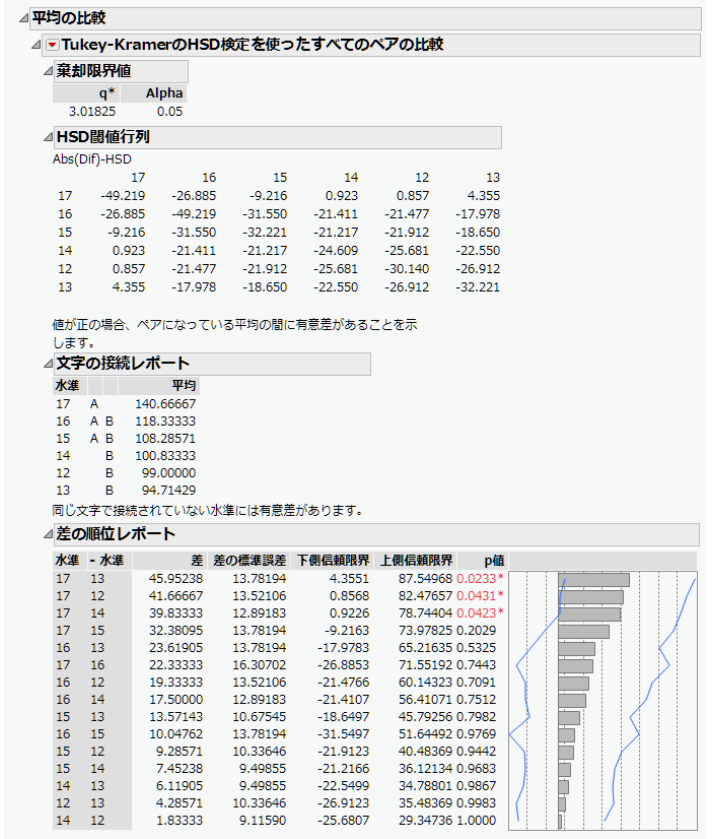
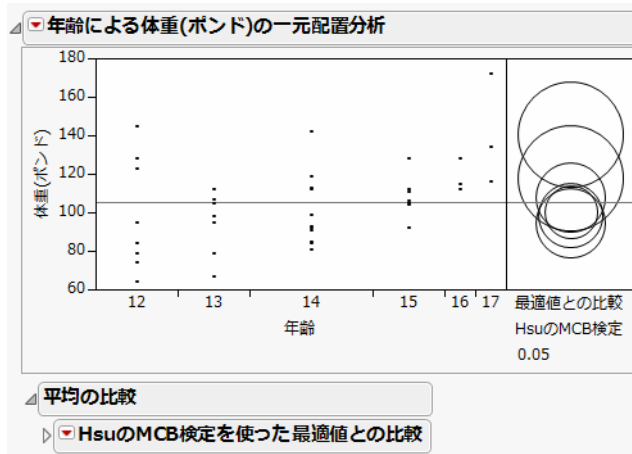


図6.18の先頭に表示されている行列は、観測された平均の差の絶対値から、HSD (Tukey 検定における最小有意差) を引いたものです。値が正の場合、観測された平均の差がHSD よりも大きいことを意味していますので、2つの平均には有意差があると結論できます。q* (レポートの先頭に表示されている値) は、Tukey 検定のHSDを計算するのに使用される分位点です。計算上では、ペアごとの比較における、Studentの t 統計量と同じような役割を果たします。

〔最適値との比較(HsuのMCB)〕の例

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

図6.19 〔最適値との比較(HsuのMCB)〕の比較円の例



4. 「年齢」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、メニューから [平均の比較] > [コントロール群との比較 (Dunnett)] を選択します。
7. コントロール群として使用するグループを選択します。この例では、年齢「12」を選択します。
または、行をクリックして散布図上で該当する点を強調表示してから、[平均の比較] > [コントロール群との比較 (Dunnett)] オプションを選択します。選択した行がコントロール群として使用されます。
8. [OK] をクリックします。

図6.21 [コントロール群との比較 (Dunnett)] の比較円の例

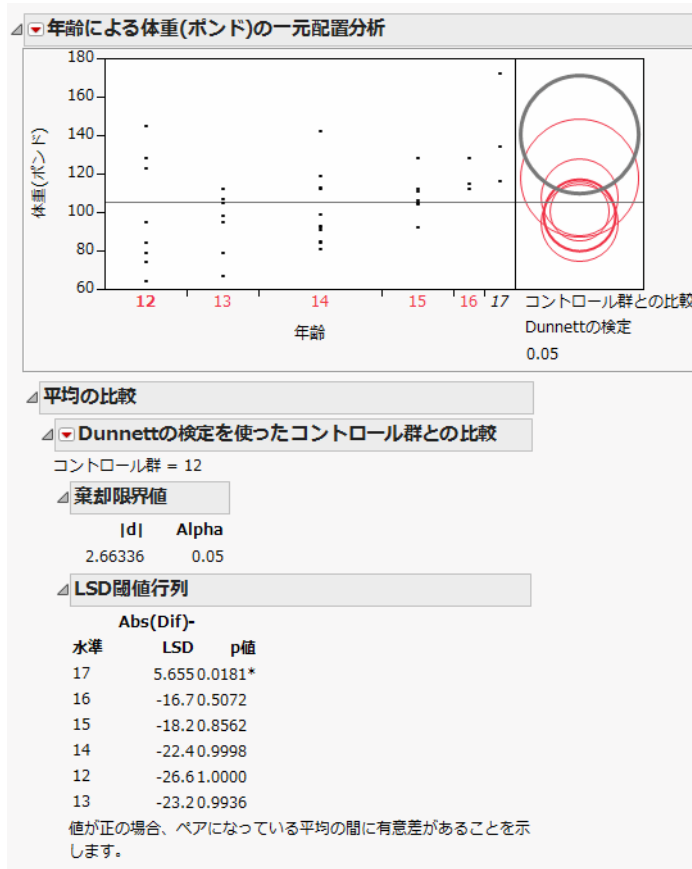


図6.21の比較円を見ると、水準17だけが、コントロール群である水準12と有意に異なることがわかります。

「平均の比較」で用意されているすべての検定

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

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

図6.22 4つの多重比較検定の比較円

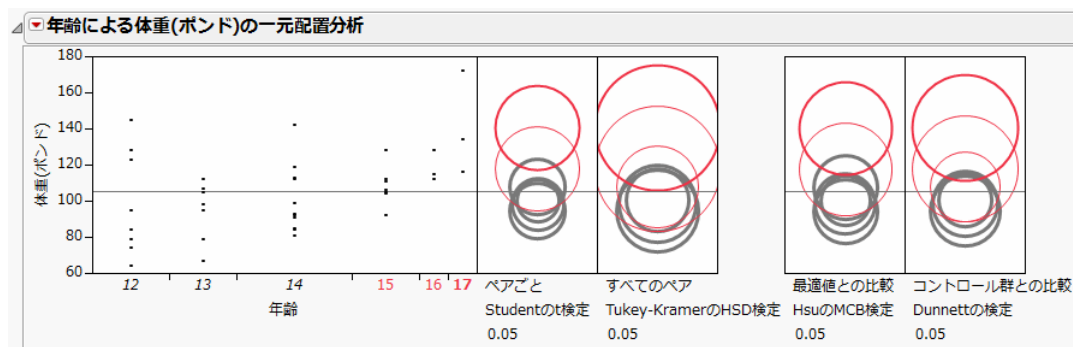


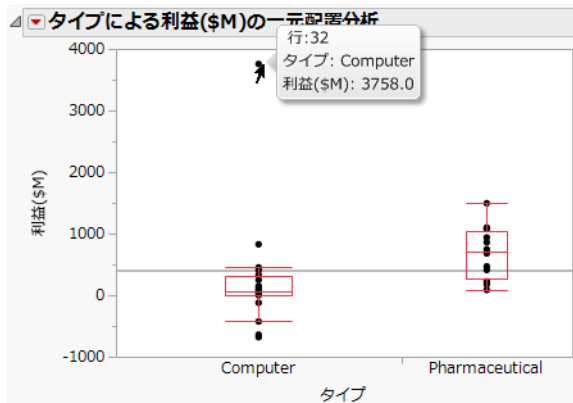
図6.22のStudentのt検定とHsu検定では、「年齢」が15才のグループ（上から3番目の円）が、17才のグループと有意に異なり、グレーで表示されています。しかし、Tukey検定とDunnnett検定では、15才と17才には有意差が認められず、赤で表示されています。

ノンパラメトリックなWilcoxon検定の例

会社の業種によって利益の平均に差があるかどうかを調べたいとしましょう。「Companies.jmp」には、2つの異なる業種の会社のデータが含まれています（製薬業界12社とコンピュータ業界20社）。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Companies.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「利益(\$M)」を選択し、[Y, 目的変数] をクリックします。
4. 「タイプ」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックして、メニューから[表示オプション] > [箱ひげ図] を選択します。

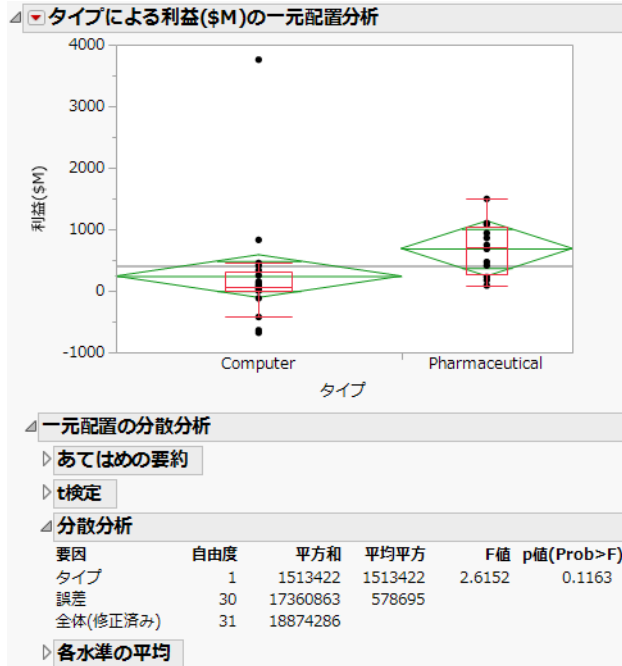
図6.23 「Computer」の利益の分布



箱ひげ図によると、データは正規分布に従わず、対称でもないようです。また、32行目に非常に大きな値があり、パラメトリックの検定に影響する恐れがあります。

7. 赤い三角ボタンのメニューから、[平均/ANOVA/プーリングしたt検定]を選択します。

図6.24 会社の分散分析



F検定では、 p 値が大きく ($p = 0.1163$)、有意ではないことが示されています。これは、大きな値をもつ32行目のデータや、正規性の仮定が満たされていないことに起因している可能性があります。

8. 赤い三角ボタンをクリックし、メニューから「個々の分散を用いたt検定」を選択します。

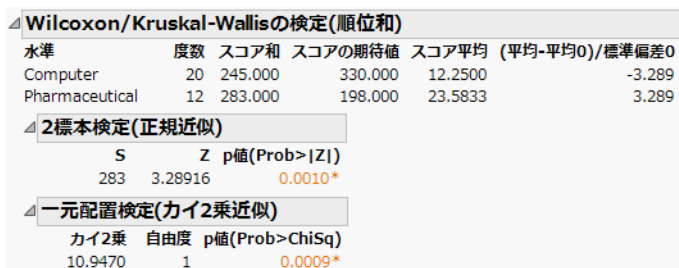
図6.25 t検定の結果



両側検定の $\text{Prob} > |t|$ は、0.0671 です。t 検定は等分散性を仮定していませんが、分散が異なる t 検定も、パラメトリックな検定です。

9. 赤い三角ボタンをクリックし、メニューから「ノンパラメトリック」>「Wilcoxon 検定」を選択します。

図6.26 Wilcoxon 検定の結果



Wilcoxon 検定は、ノンパラメトリックな検定です。順位に基づいているため、外れ値の影響を受けにくいという特徴があります。また、正規性も仮定していません。

正規近似とカイ2乗近似の両方で、 p 値が0.0010以下であり、有意となっています。これにより、分布の位置には有意な違いがあり、会社のタイプによって利益の平均に差があると結論付けることができます。

正規近似およびカイ2乗近似の検定は、いずれも検定統計量の漸近分布に基づいたものです。JMP Proでは、正確検定を行うことができます。

10. 赤い三角ボタンをクリックし、メニューから「ノンパラメトリック」>「正確検定」>「正確 Wilcoxon 検定」を選択します。

図 6.27 正確 Wilcoxon 検定の結果

Wilcoxon/Kruskal-Wallisの検定(順位和)					
水準	度数	スコア和	スコアの期待値	スコア平均	(平均-平均0)/標準偏差
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289
2標本検定(正規近似)					
S	Z	p値(Prob> Z)			
283	3.28916	0.0010*			
一元配置検定(カイ2乗近似)					
カイ2乗	自由度	p値(Prob>ChiSq)			
10.9470	1	0.0009*			
二標本: 正確検定					
S	p値(Prob≥S)	p値(Prob≥ S-Mean)			
283	0.0003*	0.0005*			

検定の結果では、 $S = 283$ となっています。この値は、「タイプ」のうち標本サイズが小さいほうの水準（「Pharmaceutical」）のオブザベーションの順位の和です。 S とその期待値との差の絶対値が、観測された値（観測された S とその期待値との差の絶対値）以上となる確率は、0.0005 です。これは、位置の差に対する両側検定の結果であり、会社の利益がタイプによって異なるという仮説を棄却する根拠となります。

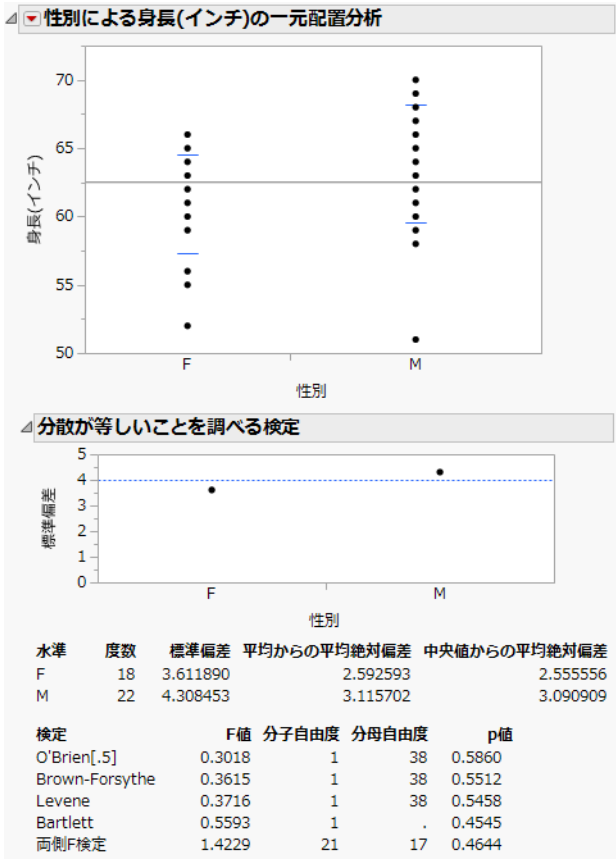
この例では、ノンパラメトリックな検定が正規性に基づく分散分析や分散が異なる t 検定よりも適切であると言えます。ノンパラメトリックな検定は、32 行目の大きな値の影響を受けにくく、各グループの「利益(\$M)」が正規分布に従っている必要もありません。

等分散性検定の例

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

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

図 6.28 「分散が等しいことを調べる検定」レポートの例



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

同等性の検定の例

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

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

8. [OK] をクリックします。

図6.29 同等性の検定の例

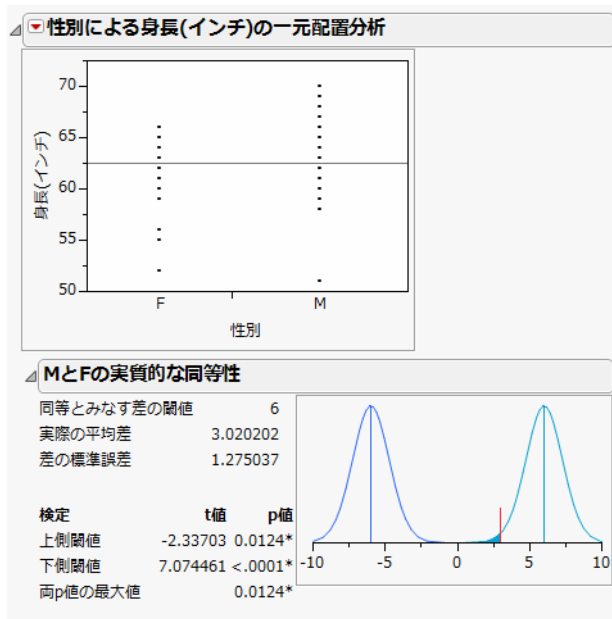


図6.29のような正規分位点プロットが描かれます。

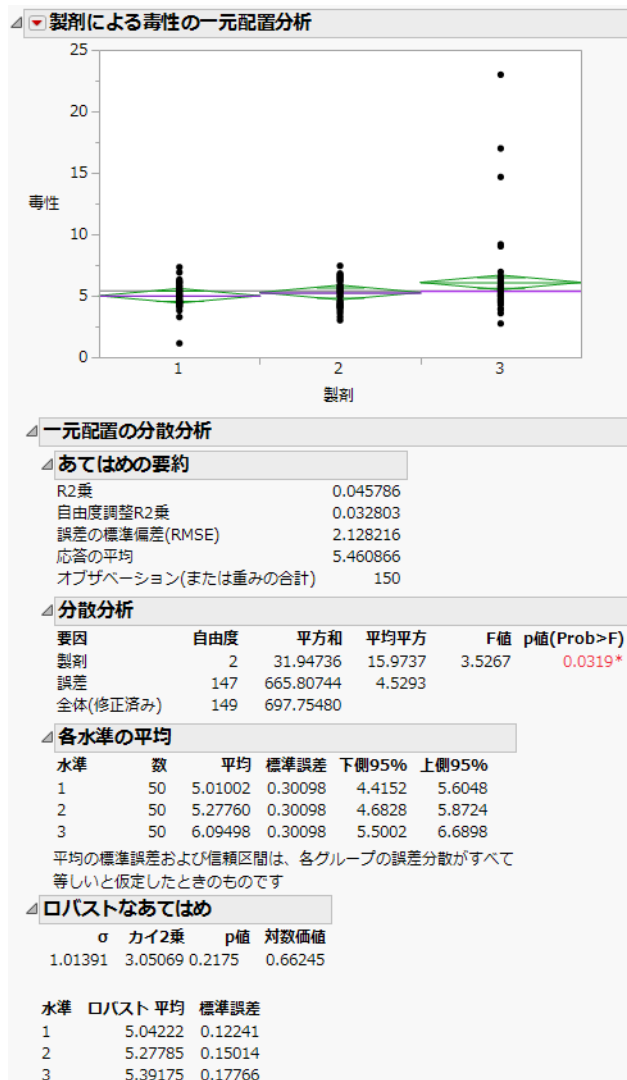
- 「下側閾値」の検定は、帰無仮説における平均の差を-6とした時の上側の片側検定です。
- 「上側閾値」の検定は、帰無仮説における平均の差を6とした時の下側の片側検定です。
- 2つの片側検定の両方でp値が小さいため、母平均の差は、-6から6までの間にあると結論付けることができます。したがって、平均は実質的に等しいということが出来ます。

[ロバストなあてはめ] オプションの例

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

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

図 6.30 ロバストなあてはめの例



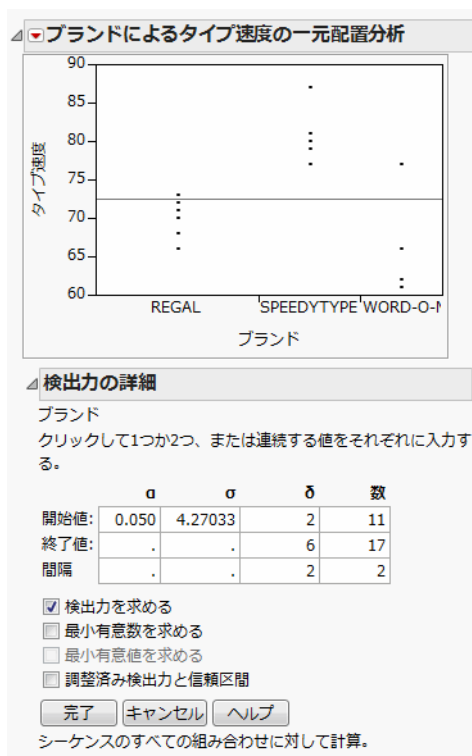
標準の「分散分析」レポートを見ると、p値が0.0319であるため、3つの製剤は有意に異なるという誤った結論に導かれてしまう可能性があります。「ロバストなあてはめ」レポートを確認すれば、p値が0.21755になっているため、3つの製剤に有意差があるという誤った結論には行きつかないでしょう。いくつかのオブザベーションで毒性が不自然に高く記録され、データに思わしくない影響を生んでいることが伺えます。

[検出力] オプションの例

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Typing Data.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。

3. 「タイプ速度」を選択し、[Y, 目的変数] をクリックします。
4. 「ブランド」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、メニューから[検出力]を選択します。
7. 「開始値」行の「 δ 」(3番目のボックス)に「2」、「数」に「11」と入力します。
8. 「終了値」行の「 δ 」に「6」、「数」に「17」と入力します。
9. 「間隔」行の「 δ 」と「数」に「2」と入力します。
10. [検出力を求める] チェックボックスをオンにします。

図6.31 「検出力の詳細」ウィンドウの例



11. [完了] をクリックします。

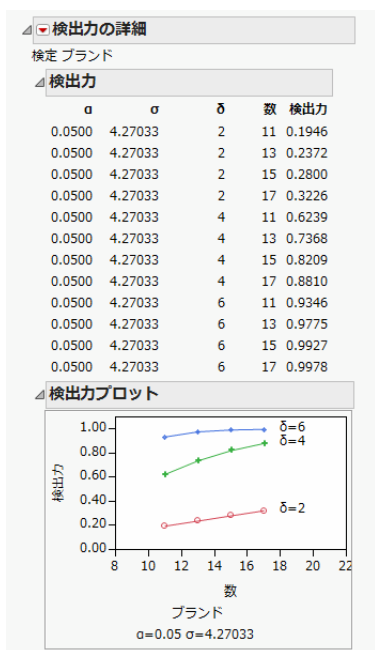
メモ: [完了] ボタンは、必要なオプションがすべて適用されるまでグレー表示のままです。

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

検出力の値をプロットにするには、次の手順に従います。

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

図6.32 「検出力」レポートの例



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

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

正規分位点プロットの例

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

図6.33 正規分位点プロットの例

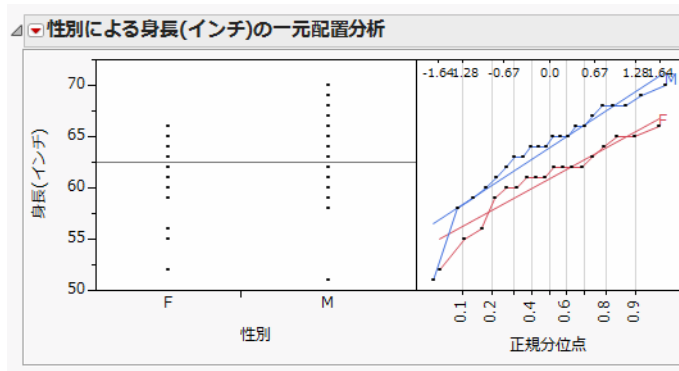


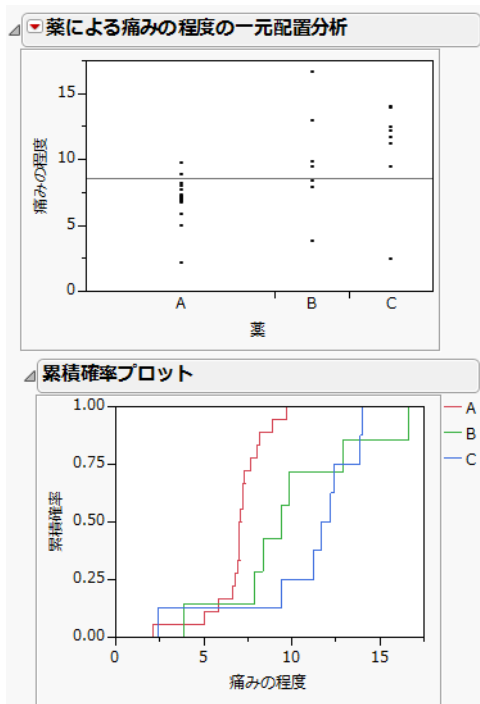
図6.33のような正規分位点プロットが描かれます。

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

累積確率プロットの例

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

図6.34 累積確率プロットの例

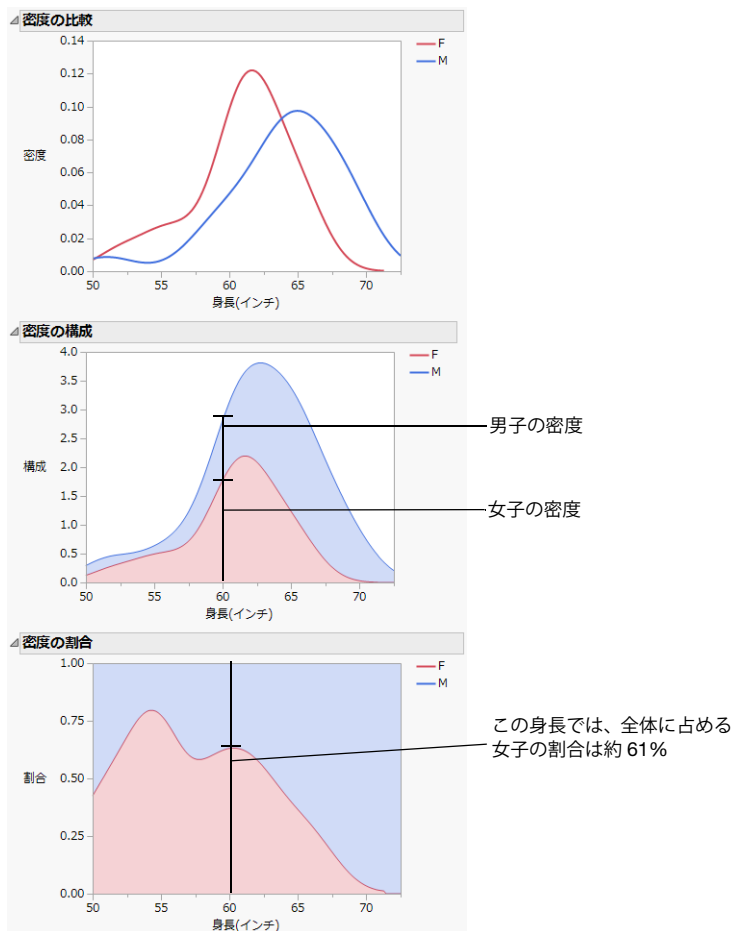


累積確率プロットでは、X変数の各水準が1本の線で表されています。また、横軸はYの値になっています。

〔密度〕の各オプションの例

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

図6.35 [密度] の各オプションの例

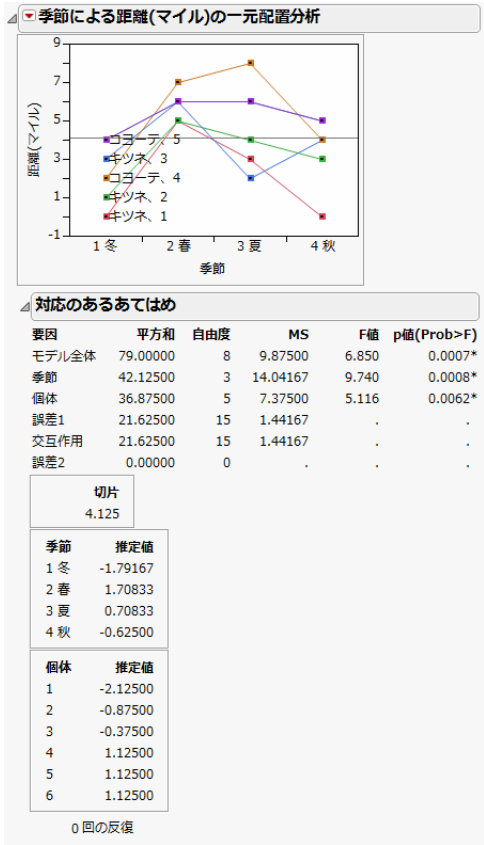


[対応のある列を設定] オプションの例

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

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

図 6.36 「対応のあるあてはめ」レポートの例



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

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

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

データを積み重ねて一元配置分析を実行する例

JMPのデータテーブル以外の形式では、1行に複数のオブザベーションが含まれていることがあります。JMPでこのようなデータを分析するには、データを読み込んでから、各行が1つのオブザベーションに対応するように構成し直す必要があります。たとえば、スプレッドシート形式のデータがあるとしましょう。このスプレッドシートには、3つの異なる製造ラインで作られたアイテムのデータが、3列に表示されています。これをJMPのデータテーブルに変更するには、各行が1つのアイテムのデータに対応するように、積み重ねた形式にする必要があります。

状況と目的

この例では、ファイル「Fill Weights.xlsx」を使用します。このファイルには、3つの異なる製造ラインからランダムに抽出された箱入りシリアルの重さが含まれています。データの形式は、図6.37のとおりです。

- 「ID」の列には、シリアルの箱のIDが含まれています。
- 「Line」の列には、各製造ラインから抽出された箱入りシリアルの重さ（単位はオンス）が含まれています。

図6.37 データの形式

Weights					
ID	Line A	ID	Line B	ID	Line C
215	12.42	705	13.63	254	11.73
287	12.49	670	12.56	282	11.40
381	12.80	715	12.87	938	12.78
		683	13.09	597	12.19
		514	13.31	179	12.25
		517	12.64		
		946	12.75		

箱入りシリアル目標の重さは、12.5オンスです。調べたいのは、この3つの製造ラインがすべて目標を満たしているかどうかということですが、まずは、各製造ラインの平均値を比較してみることにします。各ラインの重さの平均を比較するには、一元配置分析を使用できます。

「一元配置」プラットフォームを使用するには、次の手順に従います。

1. データをJMPに読み込みます。「[データの読み込み](#)」(193ページ)を参照してください。
2. 各行が1つのオブザベーションに対応するように、データを構成し直します。このためには、箱入りシリアルID、製造ラインの名前、重さの列を作成する必要があります。「[データの積み重ね](#)」(195ページ)を参照してください。

データの読み込み

この例では、Microsoft ExcelのデータをJMPに読み込む2つの方法を紹介합니다。どちらかを選択するか、または両方とも試してみてください。

- [ファイル] > [開く] から Microsoft Excel ファイルを選択し、Excel 読み込みウィザードを使ってデータを読み込みます。「[Excel 読み込みウィザードを使ったデータの読み込み](#)」(194ページ)を参照してください。この方法は、任意の Excel ファイルに使用でき、便利です。

- Microsoft Excelのデータを、JMPのデータテーブルにコピー＆ペーストします。「[Excelデータのコピー＆ペースト](#)」(194ページ)を参照してください。小さなデータファイルの場合は、この方法を使用できます。

Microsoft Excelからデータを読み込む方法については、『JMPの使用法』の「データの読み込み」章を参照してください。

Excel読み込みウィザードを使ったデータの読み込み

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Samples/Import Data」フォルダにある「Fill Weights.xlsx」を開きます。

Excel読み込みウィザードが表示されます。

2. 「列見出しの開始行」に「3」と入力します。

Excelファイルの1行目には表の説明が入力されており、2行目は空白です。列の見出しは3行目から始まります。

3. 「列見出しの行数」に「2」と入力します。

Excelファイルでは、3行目と4行目に列の見出しが含まれています。

4. [読み込み] をクリックします。

図6.38 Excel読み込みウィザードで作成したJMPのデータテーブル

	Weights-ID	Weights-Line A	Weights-ID 2	Weights-Line B	Weights-ID 3	Weights-Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.4
3	381	12.8	715	12.87	938	12.78
4	•	•	683	13.09	597	12.19
5	•	•	514	13.31	179	12.25
6	•	•	517	12.64	•	•
7	•	•	946	12.75	•	•

7行のデータが作成され、各行に3つのラインのデータが表示されています。また、3つの製造ラインそれぞれのIDと重みの列があり、全部で6列になっています。

IDの列名の「Weights」の部分は、不要であり、誤解を招く恐れがあります。ここで列名を変更することもできますが、データの積み重ねを行った後で列名を変更したほうが効率的です。

5. 「[データの積み重ね](#)」(195ページ)に進みます。

Excelデータのコピー＆ペースト

1. Microsoft Excelでファイル「Fill Weights.xlsx」を開きます。
2. 表内の、「Weights」という見出し行以外の部分を選択します。
3. 右クリックして[コピー]を選択します。
4. JMPで[ファイル] > [新規作成] > [データテーブル]を選択します。
5. [編集] > [列名とともに貼り付け]を選択します。

[編集] > [列名とともに貼り付け] は、コピーした範囲に列名も含まれる場合に使用するオプションです。

図 6.39 [列名とともに貼り付け] で作成した JMP のデータテーブル

	ID	Line A	ID 2	Line B	ID 3	Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.4
3	381	12.8	715	12.87	938	12.78
4	•	•	683	13.09	597	12.19
5	•	•	514	13.31	179	12.25
6	•	•	517	12.64	•	•
7	•	•	946	12.75	•	•

6. 「データの積み重ね」(195 ページ) に進みます。

データの積み重ね

1つのオブザベーションを1行に配置するには、[列の積み重ね] オプションを使用します。[列の積み重ね] オプションの詳細については、『JMPの使用法』の「データの再構成」章を参照してください。

1. JMP のデータテーブルで、[テーブル] > [列の積み重ね] を選択します。
2. 6列すべてを選択し、[積み重ねる列] をクリックします。
3. [複数系列の積み重ね] チェックボックスをオンにします。

ここで積み重ねるのは、「ID」と「Line」の2つの系列です。「系列の数」はデフォルトで「2」に設定されているので、変更する必要はありません。また、各系列は隣接しておらず、ID、Line A、ID、Line B、ID、Line Cと、交互になっています。このため、[隣接] はオフのままにしておきます。

4. [行による積み重ね] をオフにします。
5. [欠測値の行を除外] をオンにします。
6. 「出力テーブル名」に「積み重ねたテーブル」と入力します。
7. [OK] をクリックします。

この新しいデータテーブルで、「データ」および「データ2」には、それぞれIDと重さが含まれています。

8. 「ラベル」列の見出しを右クリックし、[列の削除] を選択します。

「ラベル」に表示されている項目は、元のデータテーブルで箱のIDの見出しとして表示されていたものであり、ここでは不要です。

9. 列の見出しをダブルクリックし、次のように名前を変更します。

- 「データ」→「ID」
- 「ラベル 2」→「ライン」
- 「データ 2」→「重さ」

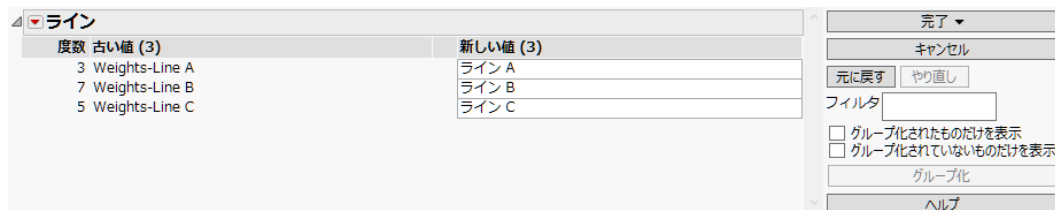
10. 「列」パネルで、「ID」の左側のアイコンをクリックし、[名義尺度] を選択します。

「ID」列の値は数値ですが、箱を識別する番号であるため名義尺度として扱う必要があります。この例では重大な支障はありませんが、列には常に適切な尺度を割り当てるようにしましょう。

11. ([ファイル] > [開く]) を使って Excel のデータを読み込んだ場合のみ) 以下の手順を実行してください。

1. 「ライン」列の見出しをクリックして列を選択し、[列] > [再コード化] を選択します。
2. 「新しい値」の値を、図 6.40 のように変更します。

図 6.40 列の値の再コード化



3. [完了] > [値の置換] をクリックします。

これで、JMPの分析に適した構造のデータテーブルが作成されました。1つの行は、1つの箱入りシリアルに対応しています。1列目には箱のID、2列目には製造ライン、3列目には重さが含まれています（図 6.41）。

図 6.41 再コード化したデータテーブル

	ID	ライン	重さ
1	215	ライン A	12.42
2	287	ライン A	12.49
3	381	ライン A	12.8
4	705	ライン B	13.63
5	670	ライン B	12.56
6	715	ライン B	12.87
7	683	ライン B	13.09
8	514	ライン B	13.31
9	517	ライン B	12.64
10	946	ライン B	12.75
11	254	ライン C	11.73
12	282	ライン C	11.4
13	938	ライン C	12.78
14	597	ライン C	12.19
15	179	ライン C	12.25

一元配置分析の実行

ここでは、以下のことを行います。

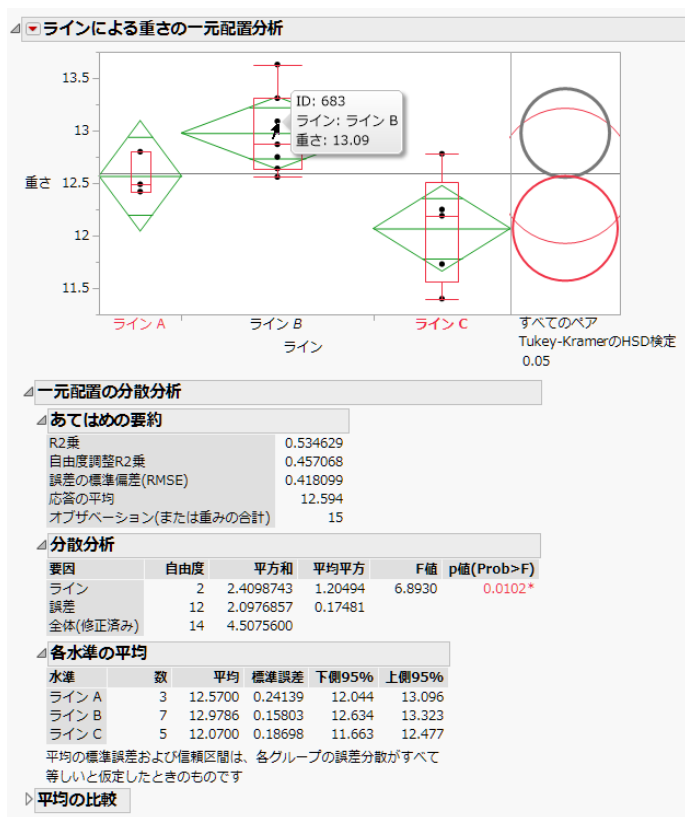
- 一元配置の分散分析を実行し、3つの製造ラインの重さの平均値に差があるかどうかを検定する。
- 比較円を作成し、どのラインに差があるかを調べる。
- 点にIDのラベルを付ける。これは、箱の重さを測り直したり、詳しく調査したいときに便利です。

まず、「積み重ねたテーブル」がアクティブになっていることを確認します。

1. [分析] > [二変量の関係] を選びます。
2. 「重さ」を選択し、[Y, 目的変数] をクリックします。
3. 「ライン」を選択し、[X, 説明変数] をクリックします。

4. [OK] をクリックします。
5. 赤い三角ボタンのメニューから、[平均/ANOVA] を選択します。
一元配置分析のプロットに表示されるひし形は、各製造ラインの平均値の95%信頼区間を示します。平均のひし形の外にある点は、外れ値のように見えますが、そうではありません。このことを確認するために、箱ひげ図を追加してみましょう。
6. 赤い三角ボタンをクリックして、メニューから [表示オプション] > [箱ひげ図] を選択します。
すべての点は、箱ひげ図の境界線内に収まっており、外れ値ではないことがわかります。
7. データテーブルの「列」パネルで「ID」を右クリックし、[ラベルあり/ラベルなし] を選択します。
8. プロットの点の上にマウスを置くと、その点の「ID」、「ライン」、「重さ」の値が表示されます。図6.42を参照してください。
9. 赤い三角ボタンをクリックし、メニューから [平均の比較] > [すべてのペア, Tukey の HSD 検定] を選択します。
プロットの右側に比較円が表示されます。
10. 下の比較円をクリックします。

図6.42 ラインごとの重さの一元配置分析



「分散分析」レポートによると、 p 値は 0.0102 となっており、すべての平均が等しいとは言えないという根拠になります。プロットにおいて、選択している Line C の比較円は、赤で表示されています (JMP のデフォルトの色を使用している場合)。Line B の円は太いグレーで示されており、Line C と Line B の平均は 0.05 の有意水準で違いがあると言えます。Line A と Line B の平均には、統計的に有意な差は見られません。

一元配置分析プロットに表示される平均のひし形は、各製造ラインの平均値の 95% 信頼区間を示しています。「一元配置の分散分析」レポートの「各水準の平均」に、95% 信頼区間の下限と上限が数値で示されています。どちらの結果からも、Line B と Line C の信頼区間に、目標値の 12.5 が含まれていないことがわかります (Line B は目標値を上回っており、Line C は下回っています)。この 2 つの製造ラインについては、重さが目標値と異なっている原因を調べ、対処する必要があります。

「一元配置」プラットフォームの統計的詳細

この節では、いくつかのオプションとレポートの統計的詳細を解説します。

比較円

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

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

2 つの独立した平均の差の標準誤差は、次の関係から計算されます。

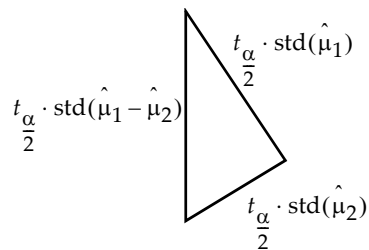
$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{std}(\hat{\mu}_1)]^2 + [\text{std}(\hat{\mu}_2)]^2$$

平均の間に相関がない場合、統計量の関係は次のようになります。

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

これらの平方値を図にすると、ピタゴラスの定理に基づく直角三角形になります (図 6.43)。

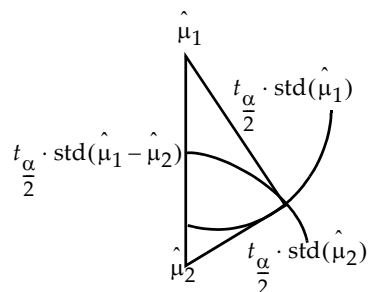
図6.43 2平均の差の関係



三角形の斜辺が、平均を比較する際の目安になります。観測された平均の差が、斜辺の長さ（最小有意差）より大きい場合に限り、平均は有意に異なります。

では、2つの平均がちょうど境界線上にあり、観測された平均の差が最小有意差に等しい場合にはどうなるでしょうか。縦軸に2つの平均を置き、それを頂点として三角形を描きます。また、それぞれの平均を中心として、それぞれの信頼区間の長さを直径とした円を描きます。

図6.44 t 検定統計量の幾何学的関係



各円の半径は対応する三角形の辺の長さと等しく、 $t_{\alpha/2} \text{std}(\hat{\mu}_i)$ で表されます。

円は三角形の2辺と同じ角度で交わるため、次のような関係が成り立ちます。

- 平均の差が最小有意差と等しい場合、平均を中心とした信頼区間の円は直角に交わります。つまり、接線の角度が直角になります。

では、平均の差が最小有意差より大きい場合や、逆に小さい場合に、円がどのように交わるかを考えてみましょう。

- 円の外側の交角が直角より大きいとき、平均は有意には**異なりません**。円の外側の交角が直角より小さいとき、平均は有意に**異なります**。外側の交角が90度より小さいということは、平均の差が最小有意差より大きいことを意味します。
- 円が交わらない場合、平均は有意に異なり、円が入れ子になる場合、平均は有意には**異なりません**。図6.11を参照してください。

Studentの t 分布の分位点ではなく、他の分位点を用いれば、いろいろな多重比較検定に対しても同様のグラフを描くことができます。

検出力

検出力を計算するには、非心度の F 分布を使用します。計算式は、次のとおりです (O'Brien and Lohr 1984)。

$$\text{Power} = \text{Prob}(F > F_{\text{crit}} \mid v_1, v_2, nc)$$

ここで、

- 上の式は、非心度パラメータ nc 、分子自由度 v_1 、分母自由度 v_2 の F 分布に従う確率変数が、 F_{crit} よりも大きくなる確率を指します。 $F(nc, v_1, v_2)$ は v_1 および v_2 の自由度を持つ F 分布の累積確率 $1 - \alpha$ における分位点です。
- $v_1 = r - 1$ は分子自由度。
- $v_2 = r(n - 1)$ は分母自由度。
- n は、各グループの標本サイズ。
- r はグループの数。
- $nc = n(\text{CSS})/\sigma^2$ は非心度パラメータ。

$$\text{CSS} = \sum_{g=1}^r (\mu_g - \mu)^2 \text{ は仮説の平方和。}$$

- μ_g は g 番目のグループの平均。
- μ は全体平均。
- σ^2 は平均2乗誤差 (MSE) による推定値。

「あてはめの要約」レポート

R2乗

連続尺度の応答のあてはめでは、 R^2 はモデルの「分散分析」レポートの統計量を使用して、次のように計算されます。

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

自由度調整 R2 乗

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

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

ここで、「誤差」の平均平方は「分散分析」レポートにある値です。分母の平均平方は、「全体（修正済み）」の平方和を、対応する自由度で割ったものです。「[「分散分析」レポート](#)」（147ページ）を参照してください。

「分散が等しいことを調べる検定」レポート

F値

O'Brienの検定では、元の応答変数を変換して、新しい応答変数を作成し、それに対して一元配置分散分析を行います。変換後の応答変数は、そのグループ平均が、元の応答変数のグループごとの不偏分散と等しくなるように作成されます。この変換後の応答変数は、次式で計算されます。

$$r_{ijk} = \frac{(n_{ij} - 1.5)n_{ij}(y_{ijk} - \bar{y}_{i.})^2 - 0.5s_{ij}^2(n_{ij} - 1)}{(n_{ij} - 1)(n_{ij} - 2)}$$

n は、観測値 y_{ijk} の個数を表します。

Brown-Forsythe 検定は、 $z_{ij} = |y_{ij} - \tilde{y}_i|$ に対する一元配置分散分析の F 検定です。ここで、 $\tilde{y}_i$ は i 番目の水準における Y の中央値です。

Leven 検定は、 $z_{ij} = |y_{ij} - \bar{y}_{i.}|$ に対する一元配置分散分析の F 検定です。ここで、 $\bar{y}_{i.}$ は i 番目の水準の応答平均です。

Bartlett 検定は、次式により計算されます。

$$T = \frac{v \log \left(\sum_i \frac{v_i}{v} s_i^2 \right) - \sum_i v_i \log(s_i^2)}{1 + \left(\frac{\sum_i \frac{1}{v_i} - \frac{1}{v}}{3(k-1)} \right)} \quad \text{この式で、} v_i = n_i - 1 \text{ および } v = \sum_i v_i$$

n_i は i 番目の水準における度数、 s_i^2 は i 番目の水準における Y の不偏分散です。この Bartlett 検定の検定統計量は、カイ2乗分布に従います。表中の Bartlett 検定の F 値に出力されている値は、この Bartlett のカイ2乗検定統計量を自由度で割ったものです。

Welch の検定の F 値

Welch の検定の F 値は、次の式で計算されます。

$$F = \frac{\left[\frac{\sum_i w_i (\bar{y}_i - \bar{y}_{..})^2}{k-1} \right]}{\left\{ 1 + \frac{2(k-2)}{k^2-1} \left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i-1} \right] \right\}}$$

この式で $w_i = \frac{n_i}{s_i^2}$ 、 $u = \sum_i w_i$ 、 $\bar{y}_{..} = \sum_i \frac{w_i \bar{y}_i}{u}$

n_i は i 番目の水準における度数、 $\bar{y}_i$ は i 番目の水準における Y の平均、 s_i^2 は i 番目の水準における Y の不偏分散です。

Welch 検定の分母自由度

Welch 検定の分母自由度は、次の近似式によって計算されます。

$$df = \frac{1}{\left(\frac{3}{k^2-1} \right) \left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i-1} \right]}$$

ここで、 w_i 、 n_i 、 u は F 値の式と同じ定義です。

ノンパラメトリックな検定の統計量

この節では、Wilcoxon、メディアン、Van der Waerden の各検定の統計量を求める式を紹介します。

表記

検定はスコアに基づいており、次の表記を使用するものとします。

$j = 1, \dots, n$ 標本全体に含まれるオブザベーション。

$n_1, n_2, \dots, n_k$ X の k 個の各水準のオブザベーション数。

R_j j 番目のオブザベーションの中間順位。**中間順位**は、同順位がない場合には、通常の順位です。同順位があった場合は、それらの同順位のものの中位値の平均（中間値）です。

α 中間順位の関数。各検定でスコアを定義するために使用します。

$\alpha(R_j)$ j 番目のオブザベーションの中間順位に対する関数 α 。

関数 α は、以下のようにスコアを定義します。

Wilcoxon のスコア

$$\alpha(R_j) = R_j$$

メディアン・スコア

$$\alpha(R_j) = \begin{cases} 1 & \text{if } R_j > (n+1)/2 \\ 0 & \text{if } R_j < (n+1)/2 \\ t & \text{if } R_j = \text{中央値} \end{cases}$$

メディアン（中央値）で同順位となるオブザベーションの数を n_t とすると、 n_t は次のように求められます。

$$n_t = \begin{cases} (n_t/2) & n_t \text{ が偶数の場合} \\ (n_t + 1)/(2n_t) & n_t \text{ が奇数の場合} \end{cases}$$

Van der Waerden のスコア

$$\alpha(R_j) = (R_j/(n+1)) \text{ における標準正規分布の分位点}$$

2 標本の場合の正規近似

正規近似に基づく検定は、X の水準が 2 つの場合にのみ行われます。この節で使用される表記については、「[表記](#)」(202 ページ) を参照してください。「2 標本検定（正規近似）」レポートに表示される統計量の定義は、以下のとおりです。

S 統計量 S は、オブザベーション数が少ないほうの水準の $\alpha(R_j)$ を合計した値です。両水準のオブザベーション数が等しい場合は、値の順序で並べて後になる水準を使用します。

Z Z は次の計算式で求められます。

$$Z = (S - E(S)) / \sqrt{\text{Var}(S)}$$

メモ： Wilcoxon 検定では、連続修正が行われます。 $(S - E(S))$ がゼロより大きい場合は、分子から 0.5 を引きます。 $(S - E(S))$ がゼロより小さい場合は、分子に 0.5 を足します。

$E(S)$ 帰無仮説のもとでの S の期待値。オブザベーション数が少ないほうの水準、または、両水準のオブザベーション数が等しい場合は値の順序で並べて後になる水準のオブザベーション数を、 n_l で表します。

$$E(S) = \frac{n_1}{n} \sum_{j=1}^n \alpha(R_j)$$

$Var(S)$ 全オブザベーションの平均スコアを ave とすると、 S の分散は次の計算式で求められます。

$$Var(S) = \frac{n_1 n_2}{n(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

一元配置のカイ2乗近似

メモ: Wilcoxon のスコアに基づくカイ2乗検定は、Kruskal-Wallis 検定と呼ばれます。

この節で使用される表記については、「[表記](#)」(202ページ)を参照してください。カイ2乗検定の統計量の計算には、次の値が使用されます。

T_i X の i 番目の水準のスコアの合計。

$E(T_i)$ i 番目の水準の合計スコアの期待値。水準間に差はないという帰無仮説のもとで、次の式で表されます。

$$E(T_i) = \frac{n_i}{n} \sum_{j=1}^n \alpha(R_j)$$

$Var(T)$ 全オブザベーションの平均スコアを ave とすると、 T の分散は次の計算式で求められます。

$$Var(T) = \frac{1}{(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

検定統計量は次の式で求められます。この統計量は、自由度 $k-1$ のカイ2乗分布に漸近的に従います。

$$C = \left(\sum_{i=1}^k (T_i - E(T_i))^2 / n_i \right) / Var(T)$$

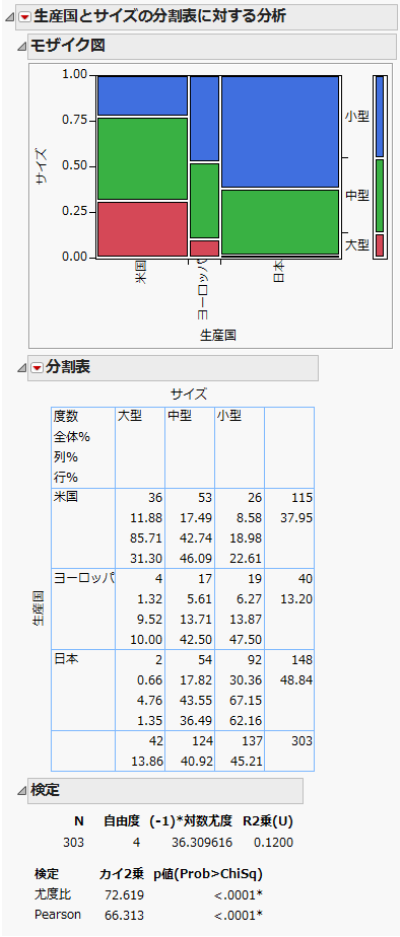
第7章

分割表分析

2つのカテゴリカルな変数の関係を調べる

「分割表」または「二変量の関係」プラットフォームは、カテゴリカルなY変数の分布を、カテゴリカルなX変数の水準ごとに比較します。カテゴリカル変数とは、名義尺度や順序尺度の変数を指します。分割表分析は、「二変量の関係」プラットフォームにおいて、XおよびYが、**両方ともカテゴリカル変数**のときに行われる分析です。分析結果として、モザイク図、度数、割合などが表示されます。また、割合の平均分析や対応分析なども行えます。

図7.1 分割表分析の例



分割表分析の例

この例では、車に関する調査の「Car Poll.jmp」サンプルデータを使用します。このデータには、調査対象者の属性（「性別」、「既婚／未婚」、「年齢」と、調査対象者の車の属性（「生産国」、「サイズ」、「タイプ」）の情報が含まれています。ここでは、車のサイズ（小型、中型、大型）と生産国の関係について調べます。

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

図 7.2 分割表分析の例

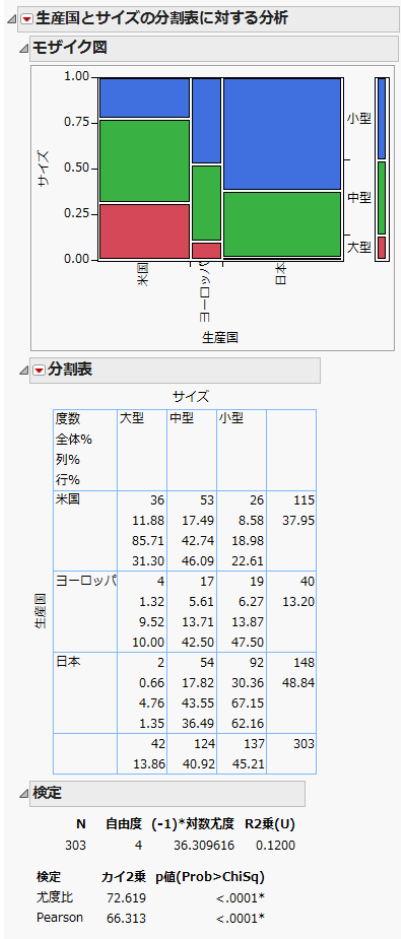


図7.2のモザイク図と凡例から、次のことがわかります。

- 日本車において、大型車に分類されるものはごくわずかです。
- ヨーロッパ車の大半は、小型車か中型車に分類されます。
- 米国車の大半は、大型車か中型車に分類されます。

「分割表」プラットフォームの起動

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

- 「二変量の関係」プラットフォームを起動するには、[分析] > [二変量の関係] を選択します。
または
- 「分割表」プラットフォームを起動するには、「JMP スターター」ウィンドウのカテゴリリストで [基本統計] をクリックし、[分割表] をクリックします。

図7.3 「分割表」起動ウィンドウ

グループ間の応答のカテゴリカル分布

列の選択

▼ 6列

- 性別
- 既婚/未婚
- 年齢
- 生産国
- サイズ
- タイプ

選択した列に役割を割り当てる

Y, 応答カテゴリ	必須 オプション
X, グループ化カテゴリ	必須 オプション
ブロック	オプション
重み	オプション(数値)
度数	オプション(数値)
By	オプション

アクション

OK

キャンセル

削除

前回の設定

ヘルプ

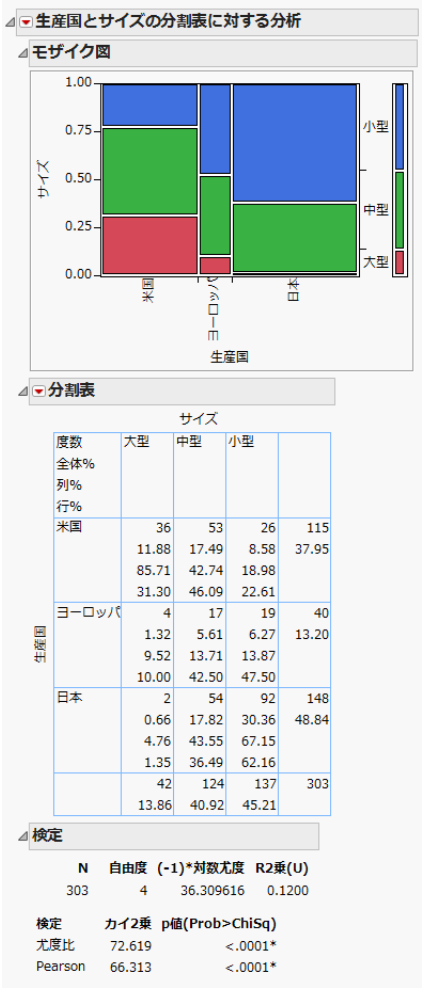
この起動ウィンドウの詳細については、「[「二変量の関係」プラットフォームの概要](#)」章（93 ページ）を参照してください。

[OK] をクリックすると、「分割表」レポートウィンドウが開きます。「[「分割表」レポート](#)」（208 ページ）を参照してください。

「分割表」レポート

図 7.4 は、「分割表分析の例」(206 ページ) の手順に従って作成したものです。

図 7.4 「分割表」レポートの例



メモ: データテーブルで除外した行は、モザイク図で非表示になります。

「分割表に対する分析」レポートには最初、「モザイク図」、「分割表」、「検定」レポートが表示されます。赤い三角ボタンをクリックし、表示されるメニューのオプションを使用して、他の分析や検定を追加できます。これらのレポートやオプションの詳細については、「[「分割表」プラットフォームのオプション](#)」(209 ページ) を参照してください。

「分割表」プラットフォームのオプション

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

「分割表に対する分析」の横にある赤い三角ボタンをクリックし、表示されるメニューのオプションを選択することにより、追加の分析や検定を実行できます。

モザイク図 「分割表」に表示されているデータをグラフで示します。[「モザイク図」](#) (210ページ) を参照してください。

分割表 2元度数表。列はYの水準と同数、行はXの水準と同数だけあります。[「分割表」](#) (212ページ) を参照してください。

検定 連続量のデータに対して作成される「分散分析」レポートに相当します。すべてのX水準においてY水準の割合が等しいかどうかを示す検定結果を表示します。[「検定」](#) (214ページ) を参照してください。

α 水準の設定 信頼区間の α 水準を変更します。サブメニューから、よく使われている有意水準（[0.10]、[0.05]、[0.01]）を選択するか、[「その他」](#)を選択して任意の有意水準を入力します。

割合の平均分析 応答変数の水準数が2つの場合にのみ表示されます。X変数の水準ごとの応答の割合を、割合の全体平均と比較します。[「割合の平均分析」](#) (215ページ) を参照してください。

対応分析 度数表において度数のパターンが似ている行や列をグラフに示します。対応分析のグラフには、分割表の各行と各列につき1つの点がプロットされます。[「対応分析」](#) (215ページ) を参照してください。

Cochran-Mantel-Haenszel 検定 3つ目の分類変数で層別したときに、2つのカテゴリカル変数の間に関係が見られるかどうかを検定します。[「Cochran-Mantel-Haenszel 検定」](#) (216ページ) を参照してください。

一致性の統計量 XとYにおける水準がまったく同じ場合にのみ表示されます。カッパ統計量 (Agresti 1990) とその標準誤差、信頼区間、仮説検定、Bowkerの対称性検定 (McNemarの検定) が表示されます。[「一致性の統計量」](#) (217ページ) を参照してください。

相対リスク リスク比を計算します。XとYが両方とも2水準である場合にのみ表示されます。[「相対リスク」](#) (218ページ) を参照してください。

オッズ比 オッズ比を計算します。XとYが両方とも2水準である場合にのみ表示されます。詳細は、[「\[オッズ比\] オプション」](#) (231ページ) を参照してください。

また、レポートには、オッズ比に対する信頼区間も表示されます。[「 \$\alpha\$ 水準の設定」](#) オプションを使用して α 水準を変更できます。

割合の2標本検定 2標本の割合に対する検定を実行します。X変数の2水準間で、Y変数の割合が異なるかどうかを比較します。XとYが両方とも2水準である場合にのみ表示されます。[「割合の2標本検定」](#) (218ページ) を参照してください。

関連の指標 分割表の変数間における関連度を表す指標を計算します。「[関連の指標](#)」(218ページ)を参照してください。

Cochran-Armitageの傾向検定 ある変数の水準間に見られる、二項割合の傾向を調べます。「2」の前後に比較的大きく間があいていますので、可能でしたら調整してください。「[Cochran-Armitageの傾向検定](#)」(220ページ)を参照してください。

JMP PRO 正確検定 次の正確検定を実行するためのオプションが表示されます。

- Fisherの正確検定
- Cochran-Armitageの傾向検定
- 一致性に対する検定

「[正確検定](#)」(220ページ)を参照してください。

モザイク図

モザイク図は、2元度数表(分割表)をグラフにしたものです。モザイク図は長方形に分割したプロットで、それぞれの長方形における縦の長さが、X変数の各水準内における、Y変数の各水準の割合を表します。モザイク図は1981年にHartiganとKleinerによって考案され、Friendly(1994)が改良を加えました。

図7.5は、「[分割表分析の例](#)」(206ページ)の手順に従って作成したものです。

図7.5 モザイク図の例

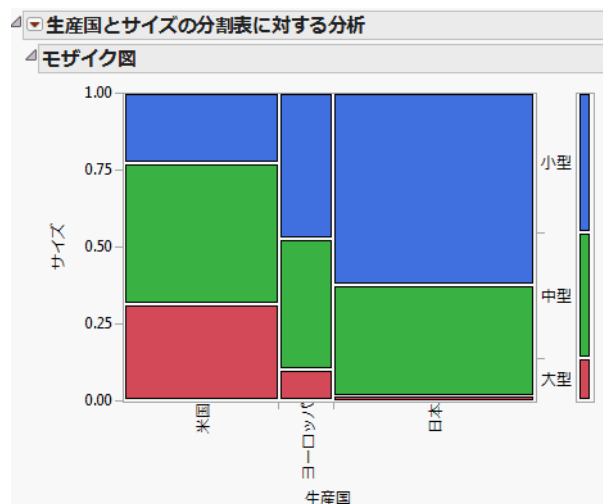


図7.5のモザイク図で、次の点に注目してください。

- X軸上の長さは、X変数(「生産国」)の各水準のオブザベーション数を示します。

- 右側のY軸は、全水準（「米国」、「ヨーロッパ」、「日本」）を組み合わせたときに、「小型」車、「中型」車、「大型」車が全体に占める割合を示します。
- 左側のY軸は、軸全体を1（標本合計）としたときの応答確率を表します。

モザイク図の長方形をクリックすると、その領域が強調表示され、関連するデータテーブルで対応するデータが選択されます。

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

ポップアップメニュー

モザイク図上を右クリックして、色を変更したり、セルのラベルを指定したりすることができます。

色の設定 水準に割り当てられている色を表示します。「色の設定」（211ページ）を参照してください。

セルのラベル モザイク図に表示するラベルを指定します。次のいずれかのオプションを選択します。

ラベルなし ラベルを表示しません。ラベルが表示されている場合は削除します。

度数の表示 各セルにオブザベーション数を表示します。

パーセントの表示 各セルにオブザベーションの割合を表示します。

ラベルを表示 各セルに対応するY変数の水準を表示します。

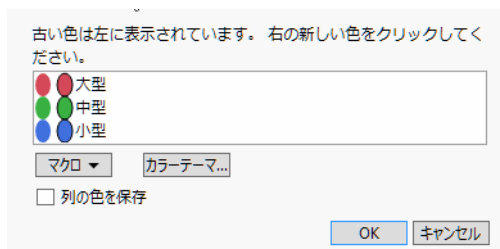
行ラベルの表示 セルが表しているすべての行の行番号を表示します。

メモ: 右クリックで表示されるその他のオプションの詳細については、『JMPの使用法』の「JMPのレポート」章を参照してください。

色の設定

「色の設定」オプションを選択すると、「値の色を選択」ウィンドウが開きます。

図7.6 「値の色を選択」ウィンドウ



デフォルトのモザイク図の色は、応答列が順序尺度か名義尺度か、「値の色」列プロパティが設定されているかどうかによって異なります。水準の色を変更するには、2列目の色の楕円形をクリックし、新しい色を選択します。

「値の色を選択」ウィンドウの説明

マクロ 2つの水準間における色のグラデーションを、次のような方法で設定できます。

- 2つの水準を選択し（それら2つの水準の範囲をドラッグするか、Shift キーを押しながら最初と最後の水準をクリックする）、[選択した点の間でグラデーション] を選択すると、それらの範囲にだけグラデーションが適用されます。
- [両端の間でグラデーション] オプションを選択すると、変数のすべての水準にグラデーションが適用されます。
- [元の色に戻す] を選択すると、変更内容を取り消すことができます。

カラーテーマ カラーテーマに従って各値の色を変更します。

列の色を保存 このチェックボックスをオンにすると、関連するデータテーブルの列に新しい列プロパティ（[値の色]）が追加されます。データテーブルからこのプロパティを編集する場合は、[列] > [列情報] を選択します。

分割表

「分割表」は2元度数表です。列はYの水準と同数、行はXの水準と同数だけあります。

図7.7は、「分割表分析の例」（206 ページ）の手順に従って作成したものです。

図7.7 分割表の例

分割表				
度数	サイズ			
	大型	中型	小型	合計
全体%				
列%				
行%				
米国	36	53	26	115
国別計	11.88	17.49	8.58	37.95
	85.71	42.74	18.98	
	31.30	46.09	22.61	
ヨーロッパ	4	17	19	40
地域計	1.32	5.61	6.27	13.20
	9.52	13.71	13.87	
	10.00	42.50	47.50	
日本	2	54	92	148
国別計	0.66	17.82	30.36	48.84
	4.76	43.55	67.15	
	1.35	36.49	62.16	
合計	42	124	137	303
列計	13.86	40.92	45.21	

この分割表について、次の点を確認してください。

- この分割表において、「度数」、「全体%」、「列%」、「行%」は、行と列が交わる各セル（たとえば、「米国」と「大型」が交差するセルなど）内の数値を指しています。
- 最後の列には、各行の合計度数と割合が表示されています。
- 最下行には、各列の合計度数と割合が表示されています。

たとえば、図7.7で、「米国」製の「大型」車に着目してください。次の表で、車の分割表から導き出せる結論について説明します。

表7.1 例の分割表による結果

数値	説明	表のラベル
36	米国製大型車の台数	度数
11.88%	全車両のうち米国製大型車が占める割合 (36/303) ^a	全体%
85.71%	大型車のうち米国車が占める割合 (36/42) ^b	列%
31.30%	米国車のうち大型車が占める割合 (36/115) ^c	行%
37.95%	全車両のうち米国車が占める割合 (115/303)	(なし)
13.86%	全車両のうち大型車が占める割合 (42/303)	(なし)

a. 303は、調査対象の全車両の台数です。

b. 42は、調査対象の大型車の台数です。

c. 115は、調査対象の米国車の台数です。

ヒント：分割表における「度数」、「全体%」、「列%」、「行%」の数値は、「分割表」の赤い三角ボタンをクリックし、その項目を選択すると、表示／非表示を切り替えられます。

「分割表」レポートの説明

度数 セルの度数、行／列の合計度数、全体の合計度数（全体の標本サイズ）。

全体% セルの度数と行／列の合計度数が全体の合計度数に占める割合。

行% セルの度数が行の合計度数に占める割合。

列% セルの度数が列の合計度数に占める割合。

期待値 行と列が独立していると仮定して計算された各セルの期待度数（ E ）。対応する行合計と列合計の積を全体合計で割った値です。

偏差 観測されたセル度数（ O ）から期待値（ E ）を引いたもの。

セルのカイ2乗 $(O - E)^2 / E$ で求めた各セルのカイ2乗。

累積列度数 列の累積合計。

累積列 % 列の累積割合。

累積行度数 行の累積合計。

累積行 % 行の累積割合。

検定

「検定」レポートには、すべてのX水準においてY水準の割合が等しいかどうかを調べる2つの検定の結果が表示されます。

図7.8のレポートを作成するには、「[分割表分析の例](#)」(206ページ)の手順に従います。

図7.8 「検定」レポートの例

検定			
N	自由度	(-1)*対数尤度	R2乗(U)
303	4	36.309616	0.1200
検定	カイ2乗	p値(Prob>ChiSq)	
尤度比	72.619	<.0001*	
Pearson	66.313	<.0001*	

カイ2乗統計量について、次の点を確認してください。

- これらのカイ2乗検定は、2つのカテゴリーカル変数が両方とも応答（Y変数）のとき、その2つの変数が独立であることを検定します。
- 2つの変数のうち、Xが所与である場合、カイ2乗統計量は、Y変数の分布がXの全水準で同じであることを検定します。

「検定」レポートの説明

N オブザベーションの合計数。

自由度 検定に関係する自由度。自由度は $(c-1)(r-1)$ に等しくなります。 c は列数、 r は行数です。

(-1)*対数尤度 対数尤度の符号を逆にしたもの。適合度や不確定性を表す指標です（連続尺度の応答変数における平方和と同じような指標です）。

R2乗(U) 全体の不確定性のうち、モデルによって説明される割合。

- R^2 が1のときは、カテゴリーカルな応答が因子によって完全に予測できることを意味します。
- R^2 が0ということは、標本全体で応答確率が一定であることに比べて、モデルを使っても予測が向上しないことを意味しています。

詳細は、「[「検定」レポート](#)」（232 ページ）を参照してください。

検定 「すべての標本グループにおいて、応答確率は等しい」という仮説を検定する2種類のカイ2乗検定。詳細は、「[「検定」レポート](#)」（232 ページ）を参照してください。

p 値（Prob>ChiSq） X変数とY変数の間に関係が存在しないときに、まったくの偶然だけで、現在のカイ2乗値より大きなカイ2乗値を得る確率。両方の変数が2水準しかない場合は、Fisherの正確検定も実行されます（両側検定と片側検定のp値が表示されます）。

Fisherの正確検定

このレポートには、2×2 分割表に対する Fisher の正確検定の結果が表示されます。2×2 分割表の場合は、Fisher の正確検定が自動的に実行されます。2×2 および $r \times c$ 分割表に対する Fisher の正確検定の詳細については、「[正確検定](#)」（220 ページ）を参照してください。

割合の平均分析

応答変数が2水準である場合、このオプションによって、X変数の水準ごとの応答の割合を、割合の全体平均と比較できます。この方法では、二項分布に対する正規近似を計算に用いているため、標本サイズが小さすぎる場合は、警告が表示されます。

メモ: 平均分析法の解説については、Nelson, Wludyka, and Copeland（2005）を参照してください。

有意水準の設定 分析に使用する α 水準を選択します。

要約レポートの表示 X変数の水準ごとに、応答変数の割合と決定限界を表示します。このレポートには、決定限界を超過しているかどうかとも示されます。

応答水準の切り替え 分析に使用する応答変数のカテゴリを変更します。

表示オプション [決定限界の表示]、[決定限界の陰影の表示]、[中心線の表示]、[点のオプション] の各オプションがあります。

関連情報

- 「[割合の平均分析の例](#)」（221 ページ）

対応分析

対応分析は、度数表の中から度数のパターンが似ている行や列を見つけるための手法です。対応分析のグラフでは、各行と各列につき1つの点がプロットされます。水準数が多く、モザイク図から有用な情報を引き出すのが難しい場合は、対応分析が有効です。

対応分析プロットの理解

行ごとの割合、つまり、行内にある度数をその行の合計度数で割ったものを、**行プロファイル**と言います。2つの行の行プロファイルが似ているとき、対応分析のグラフにおいて、それらの行を表す点は互いに近い位置にあります。プロットにおける2点間の平方距離は、「カイ2乗距離」という2行の等質性を示す距離を近似したものです。

列プロファイルも、行プロファイルと同じように定義され、同じように解釈できます。行の点と列の点の距離には意味がありませんが、原点からの方向には意味があり、その関係を見ることでプロットを解釈できます。

「対応分析」のオプション

「対応分析」の横にある赤い三角ボタンをクリックし、表示されるメニューのオプションを使用して、三次元散布図を作成したり、データテーブルに列プロパティを追加したりすることができます。

三次元対応分析 三次元散布図を作成します。

値の順序を各列に保存 対応分析における第1次元のスコアで値を並べた順序が、X列とY列の「値の順序」列プロパティに保存されます。

「詳細」レポート

「詳細」レポートは、対応分析の統計情報をまとめたもので、プロットに使われている値をリストしています。

特異値 分割表の特異値分解。式については、「[対応分析の「詳細」レポート](#)」(232ページ)を参照してください。

慣性 特異値の平方。各次元で説明できる相対的な変動を反映しています。

割合 合計慣性に対する各慣性の割合。

累積 慣性の割合の累積。最初の2つの特異値が慣性のほとんどを捉えているときは、2次元の対応分析プロットで、表内に存在する関係を十分に示すことができます。

X変数 c1、c2、c3 対応分析プロット上に表示される値（図7.11）。

Y変数 c1、c2、c3 対応分析プロット上に表示される値（図7.11）。

関連情報

- [「対応分析の例」](#)（222ページ）

Cochran-Mantel-Haenszel 検定

Cochran-Mantel-Haenszel 検定では、3つ目の分類変数で層別したときに、2つのカテゴリカル変数の間に関係が見られるかどうかを調べます。

スコアの相関 YとXの両方が順序尺度か間隔尺度であるときに使います。対立仮説は、層別変数の少なくとも1つの層において、XとYの間に線形関係があるというものです。

X間でのスコア比較 Yが順序尺度か間隔尺度であるときに使います。対立仮説は、層別変数の少なくとも1つの層において、ある列の行スコアの平均が、他の列の行スコアの平均と異なるというものです。

Y間でのスコア比較 Xが順序尺度か間隔尺度であるときに使います。対立仮説は、層別変数の少なくとも1つの層において、ある行の列スコアの平均が、他の行の列スコアの平均と異なるというものです。

カテゴリの一般連関 層別変数の少なくとも1つの水準で、XとYの間に何らかの関連があることを検定します。

関連情報

- ・ [「Cochran-Mantel-Haenszel 検定の例」](#) (224 ページ)

— 一致性の統計量

2変数の水準がまったく同じ場合には、**〔一致性の統計量〕** オプションが使用できます。このオプションは、カッパ統計量 (Agresti 1990) の算出と、Bowker の対称性検定を行います。カッパ統計量に対しては、標準誤差、信頼区間、仮説検定が計算されます。

この節で紹介しているカッパ統計量と p 値は近似値ですが、正確な値も計算できます。[「正確検定」](#) (220 ページ) を参照してください。

カッパ カッパ統計量。

標準誤差 カッパ統計量の標準誤差。

下側 95% カッパ統計量の下側信頼区間。

上側 95% カッパ統計量の上側信頼区間。

p 値 (Prob>Z) カッパ統計量の片側検定の p 値。「カッパ統計量が 0 である」という帰無仮説を検定します。

p 値 (Prob>|Z|) カッパ統計量の両側検定の p 値。

カイ 2 乗 分割表の対称性に対する Bowker 検定の検定統計量。Bowker 検定の帰無仮説は、「分割表における確率が対称性である (すべてのセルにおいて $p_{ij}=p_{ji}$ である)」というものです。なお、X と Y が 2 水準の場合、Bowker 検定は、McNemar 検定と等しくなります。

p 値 (Prob>ChiSq) Bowker 検定の p 値。

関連情報

- ・ [「〔一致性の統計量〕 オプションの例」](#) (226 ページ)
- ・ [「〔一致性の統計量〕 オプション」](#) (231 ページ)

相対リスク

分割表が2×2表である場合、**「相対リスク」** オプションを選択すると、リスク比が計算されます。レポートには信頼区間も表示されます。この手法の詳細は、Agresti（1990）の3.4.2節で説明されています。

「相対リスク」 オプションを選択すると、「相対リスクカテゴリの選択」ウィンドウが開きます。1対の応答と因子について、または、応答と因子のすべての組み合わせについて相対リスクを計算することができます。

関連情報

- [「相対リスク」オプションの例](#)（227ページ）

割合の2標本検定

分割表が2×2表である場合は、割合の差に対する信頼区間も計算できます。また、信頼区間に相当する検定も計算されます。

説明 行われている検定の種類。

割合の差 X変数の2水準間における割合の差。

下側95% 割合の差に対する両側信頼区間の下限。この信頼区間は、調整を伴ったWald法によって計算されています。

上側95% 割合の差に対する両側信頼区間の上限。この信頼区間は、調整を伴ったWald法によって計算されています。

調整済みWald検定 両側検定と片側検定を示します。

p値 検定のp値。

対象となる応答<変数>のカテゴリ 検定で使用する応答の水準を選択します。

関連情報

- [「割合の2標本検定の例」](#)（228ページ）

関連の指標

「関連の指標」 オプションを選択すると、分割表の変数間における関連を示す、複数の指標が計算されます。

ガンマ 順序関係が一致するペアと一致しないペアの個数から計算される相関係数です。同順位のペアは無視します。-1から1の値を取ります。

Kendallのタウ-b ガンマに似ていますが、同順位のペアに対して、修正が行われます。-1から1の値を取ります。

Stuartのタウ-c ガンマに似ていますが、表のサイズによる調整、および、同順位のペアに対する修正が行われます。-1から1の値を取ります。

SomersのD Kendallのタウ-bに、非対称性の修正を加えた指標です。

- 「C|R」は、行変数Xを独立変数、列変数Yを従属変数とすることを示します。
- 同様に、「R|C」は、列変数Yを独立変数、行変数Xを従属変数とすることを示します。

SomersのDは、同順位のペアに対する修正を、独立変数の同順位に対してだけ行うという点で、タウ-bとは異なります。-1から1の値を取ります。

非対称ラムダ C|RとR|Cで異なります。

- 「非対称ラムダ (C|R)」は、行変数Xで列変数Yを予測したときに、予測的中率がどれくらい向上するかを示します。
- 「非対称ラムダ (R|C)」は、列変数Yから行変数Xを予測したときに、予測的中率がどれくらい向上するかを示します。

0から1の値を取ります。

対称ラムダ 2つの非対称ラムダ係数の平均のような指標です。0から1の値を取ります。

不確実性係数 -「不確実性係数 (C|R)」は、行変数Xによって説明される列変数Yの不確実性の割合を示します。

- 「不確実性係数 (C|R)」は、列変数Yによって説明される列変数Xの不確実性の割合を示します。

0から1の値を取ります。

不確実性係数 (対称) 上記2つの不確実性係数を、対称性をもつように変えた指標です。0から1の値を取ります。

メモ:

各推定値、および、その標準誤差と信頼区間が表示されます。

- ガンマ、Kendallのタウ-b、Stuartのタウ-c、SomersのDは、X変数の増加に伴い、Y変数も増加する傾向にあるかどうかを示す順序的な関係を示す指標です。これらの統計量は、順序関係の一致するペアと一致しないペアの個数から計算されます。データのペアにおいて、X値がより大きい方が、Y値もより大きくなっているペアを、「順序関係が一致するペア」と言います。逆に、X値が大きい方が、Y値が小さくなっているペアを「順序関係が一致しないペア」と言います。これらの指標は、両変数が順序尺度の場合だけに適しています。
- ラムダおよび不確実性係数は、順序尺度と名義尺度の変数に適しています。

関連の指標の詳細については、次の文献を参考にしてください。

- Brown and Benedetti (1977)

- Goodman and Kruskal (1979)
- Kendall and Stuart (1979)
- Snedecor and Cochran (1980)
- Somers (1962)

関連情報

- 「[\[関連の指標\] オプションの例](#)」(229 ページ)

Cochran-Armitage の傾向検定

Cochran-Armitage の傾向検定は、順序変数の水準間における二項割合の傾向を調べます。この検定は、片方の変数が2水準の変数で、もう一方が順序変数の場合にだけ適切な検定です。2水準の変数が応答変数、順序尺度の変数が説明変数として扱われます。帰無仮説は「割合の傾向がない」、つまり、「説明変数の全水準において二項割合が同じ」というものです。

この節で紹介している検定統計量と p 値は近似値です。正確な傾向検定も実行できます。「[正確検定](#)」(220 ページ) を参照してください。

関連情報

- 「[Cochran-Armitage の傾向検定の例](#)」(230 ページ)

正確検定

JMP PRO 「分割表」プラットフォームでは、以下の正確検定を使用できます。

Fisher の正確検定 $r \times c$ 分割表に対する Fisher の正確検定を実行します。この検定は、2変数の関連を調べます。Fisher の正確検定は、行和および列和を固定して、超幾何分布を用いて p 値を求めます。

この検定は大標本近似ではありません。標本サイズが小さい場合や、度数が0のセルがある場合など、尤度比検定や Pearson 検定では妥当な結果が得られないときに適しています。

レポートには次の情報も表示されます。

表の確率 (P) 現在、観測されている表が生じる確率。検定の p 値ではありません。

両側 p 値 (表の生起確率 $\leq p$) 両側検定の p 値。

1つの行または1つの列のすべてが0ではない限り（この場合、検定を実行できません）、2x2 分割表に対して Fisher の正確検定が自動的に実行されます。「[検定](#)」(214 ページ) を参照してください。

正確 Cochran-Armitage 傾向検定 正確な Cochran-Armitage 傾向検定を実行します。この検定は、いずれかの変数が2水準である場合にのみ実行できます。傾向検定の詳細については、「[Cochran-Armitage の傾向検定](#)」(220 ページ) を参照してください。

正確一致性統計量 2変数間の一致性に対する正確検定を実行します。これはカッパ統計量に対する正確検定です。2変数の水準がまったく同じである場合にのみ実行できます。一致性検定の詳細については、「[一致性の統計量](#)」(217ページ)を参照してください。

「分割表」プラットフォームのその他の例

ここでは、「分割表」プラットフォームのオプションを使った例をさらに紹介します。

割合の平均分析の例

この例では、「Office Visits.jmp」サンプルデータを使用します。このデータには、ある地域の6つの診療所において、予約どおりの時間に患者が診察できたかどうかが記録されています。6つの診療所ごとに、1週間分のデータから、60の診察予約データを無作為抽出しました。予約時間から5分以内に診察室に通された場合には、定刻どおりの診察とみなします。予約時間どおりに診察を受けた患者の割合を調べてみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Office Visits.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「時間どおり」を選択し、[Y, 目的変数] をクリックします。
4. 「診療所」を選択し、[X, 説明変数] をクリックします。
5. 「度数」を選択し、[度数] をクリックします。
6. [OK] をクリックします。
7. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから[割合の平均分析]を選択します。
8. 「割合の平均分析」の横にある赤い三角ボタンをクリックし、メニューから[要約レポートの表示] および[応答水準の切り替え]を選択します。

図7.9 割合の平均分析の例



図7.9は予約時間どおりに診察を受けた患者の割合を、診療所ごとに示しています。図7.9から、以下のことが分かります。

- 予約時間どおりに受診できた割合が一番高いのは、診療所Fです。診療所Bがそれに次いでいます。
- 診療所Dは、予約時間どおりに受診できた割合が一番低く、診療所Aがそれに次いでいます。
- 診療所CとEは平均に近く、決定限界を超えていません。

対応分析の例

この例では、「Cheese.jmp」サンプルデータを使用します。このデータは、Newell によるチーズの試食試験の結果で、McCullagh and Nelder (1989) で取り上げられています。試験では、4 種類のチーズ添加物ごとに、9 段階評価の度数が記録されました。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Cheese.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「評価」を選択し、[Y, 目的変数] をクリックします。
「評価」は、1（最下位）から9（第1位）までを範囲としています。
4. 「チーズ」を選択し、[X, 説明変数] をクリックします。
「A」、「B」、「C」、「D」はチーズの4つの添加物を表します。
5. 「度数」を選択し、[度数] をクリックします。

6. [OK] をクリックします。

図7.10 チーズデータのモザイク図

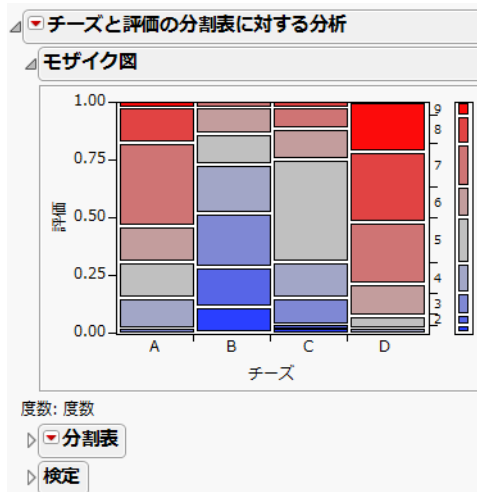


図7.10のモザイク図を見ると、分布はそれぞれ異なるようですが、9つも水準があるのでモザイク図だけでは傾向が読みません。このようなときは、対応分析が役立ちます。

7. 対応分析プロットを表示するには、「分割表に対する分析」の横にある赤い三角ボタンをクリックし、[対応分析] を選択します。

図7.11 対応分析プロットの例

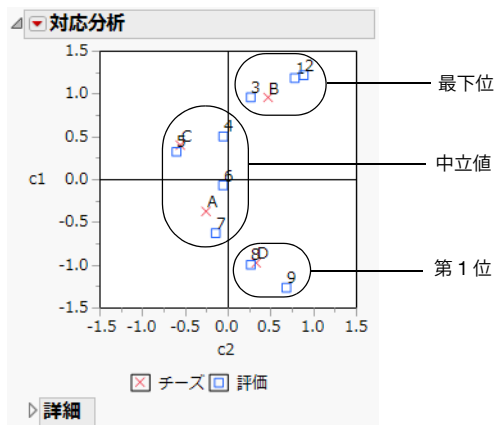


図 7.11 は、対応分析をグラフ化したもので、軸には c1 と c2 という名前が付いています。上の例で、次の点に注意してください。

- c1 は、全体的な満足度に対応しているようです。c1 軸の上部に 1（最下位）、下部に 9（第1位）という評価が分布しています。
- チーズ D は、評価が 8 と 9 である割合が、他よりも高くなっています。
- チーズ B は、評価が 1、2、3 である割合が、他よりも高くなっています。
- チーズ C と A は、4、5、6、7 という中間的な評価を得ています。

8. 「対応分析」の横にある赤い三角ボタンをクリックし、メニューから「三次元対応分析」を選択します。

図 7.12 三次元散布図の例

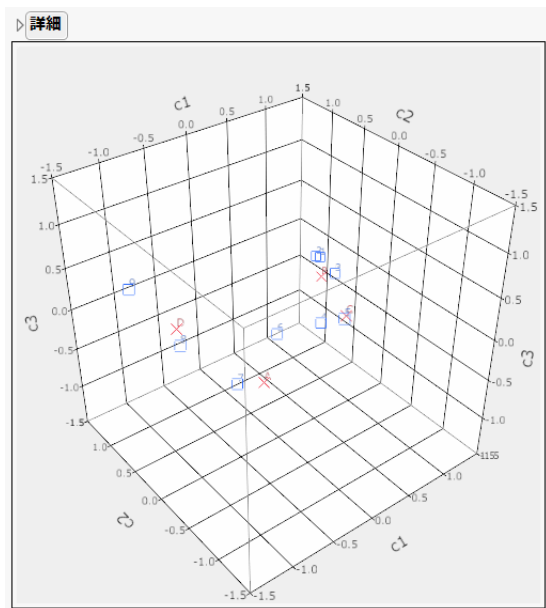


図 7.12 のような正規分位点プロットが描かれます。

- c1 軸を見ると、評価 1 から評価 9 までが、正から負に並んでいます。
- c2 軸を見ると、A と C が 0 の右側（正）、B と D が 0 の左側（負）に位置しています。
- c1 は全体的な満足度を表しています。

Cochran-Mantel-Haenszel 検定の例

この例では、「Hot Dogs.jmp」サンプルデータを使用します。ホットドッグに関して、タイプと味の関係を調べてみましょう。

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

2. [分析] > [二変量の関係] を選びます。
3. 「タイプ」を選択し、[Y, 目的変数] をクリックします。
4. 「味」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから [Cochran-Mantel-Haenszel 検定] を選択します。
7. 層別変数として「脂肪に対する蛋白質の比率」を選択し、[OK] をクリックします。

図7.13 Cochran-Mantel-Haenszel 検定の例

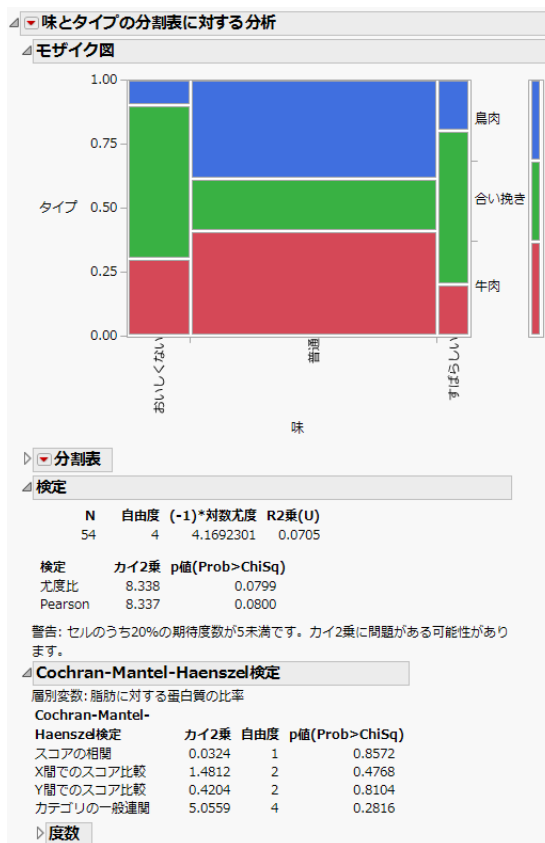


図7.13 からは次のようなことがわかります。

- 「検定」レポートを見ると、Pearsonのカイ2乗検定に対するp値は0.0800で、5%よりは少し大きいですが、ホットドッグの「味」と「タイプ」の間に何らかの関係があるかもしれないことを示唆しています。
- 一方、「Cochran-Mantel-Haenszel 検定」レポートを見ると、一般的な連関性に対するp値は0.2816であり、5%よりもかなり大きくなっています。

【一貫性の統計量】オプションの例

この例では、「Attribute Gauge.jmp」サンプルデータを使用します。3人の判定者が50個の部品をそれぞれ3回評価した結果が記録されています。判定者AとBの関係を調べてみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Attribute Gauge.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「A」を選択し、[Y, 目的変数] をクリックします。
4. 「B」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから【一貫性の統計量】を選択します。

図7.14 「一貫性の統計量」レポートの例

▼ BとAの分割表に対する分析

▷ モザイク図

▼ 分割表

		A		合計
		0	1	
B	度数			
	全体%			
	列%			
	行%			
	0	44	3	47
	29.33	2.00	31.33	
	88.00	3.00		
	93.62	6.38		
	1	6	97	103
	4.00	64.67	68.67	
12.00	97.00			
5.83	94.17			
合計	50	100	150	
	33.33	66.67		

▷ 検定

▼ 一貫性の統計量

▼ カッパ係数

一貫性の度合い	カッパ	標準誤差	下側95%	上側95%
	0.862944	0.044198	0.776317	0.949571

漸近検定

p値(Prob>z)	p値(Prob> Z)
<.0001*	<.0001*

▼ Bowkerの検定

対称性	カイ2乗	p値(Prob>ChiSq)
	1	0.3173

2x2表の場合、Bowker検定はMcNemar検定と等価です。

図7.14を見ると、一貫性の統計量は0.86と、1に近い大きな値となっています。また、そのp値は<.0001と小さくなっています。分割表を見ても、対角線上における度数が大きく、一貫性が高いことを裏付けています。2名の評価者が同じ評価をしている場合（つまり、ともに0と評価するか、または、ともに1と評価している場合）、一貫性の統計量は大きくなります。

「相対リスク」オプションの例

この例では、「Car Poll.jmp」サンプルデータを使用して、調査対象者における性別ごとの既婚の割合を調べてみます。

1. 「ヘルプ」 > 「サンプルデータライブラリ」を選択し、「Car Poll.jmp」を開きます。
 2. 「分析」 > 「二変量の関係」を選びます。
 3. 「既婚/未婚」を選択し、「Y, 目的変数」をクリックします。
 4. 「性別」を選択し、「X, 説明変数」をクリックします。
 5. 「OK」をクリックします。
 6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから「相対リスク」を選択します。
- 「相対リスクカテゴリの選択」ウィンドウが表示されます。

図7.15 「相対リスクカテゴリの選択」ウィンドウ

リスク比は $P(Y=y_j | X=x_i) / P(Y=y_j | X=x_k)$

応答変数既婚/未婚の対象カテゴリ: $Y=y_j$

☒ 既婚
☐ 未婚

分子となる性別の標本カテゴリ: $X=x_i$

☒ 女性
☐ 男性

☐ すべての組み合わせを計算

OK キャンセル ヘルプ

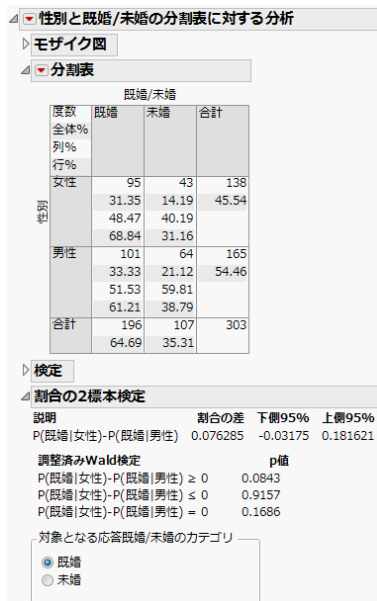
「相対リスクカテゴリの選択」ウィンドウについて、次の点を確認してください。

- 1 対の応答と因子についてのみ相対リスクを計算するには、このダイアログボックスで対象の変数を指定します。たとえば、図7.15のウィンドウで「OK」をクリックすると、次のように計算されます。

$$\frac{P(Y = \text{既婚} | X = \text{女性})}{P(Y = \text{既婚} | X = \text{男性})}$$

- 応答と因子のすべての組み合わせ（ $2 \times 2 = 4$ ）について相対リスクを計算するには、「すべての組み合わせを計算」チェックボックスをオンにします。図7.16を参照してください。
7. 「すべての組み合わせを計算」チェックボックスをオンにして、すべての組み合わせについて計算を行います。他のすべての設定はデフォルト値のままにしておきます。

図7.17 「割合の2標本検定」レポートの例



この例では、女性と男性で既婚者の割合を比較します。「分割表」の「行%」で次の値を確認します。

$$P(\text{既婚} | \text{女性}) = 0.6884$$

$$P(\text{既婚} | \text{男性}) = 0.6121$$

2つの割合の差は0.0763で、これがレポートの「割合の差」に表示されています。割合の差に対する両側信頼区間は、[-0.03175, 0.181621]となっています。この信頼区間に対応した修正Wald法による両側検定のp値は、0.1686となっています。これは、Pearsonのカイ2乗検定によるp値（0.1665）に近い値です。一般に、割合の差の検定には、修正Wald検定よりも、Pearsonのカイ2乗検定の方がよく使用されます。

〔関連の指標〕 オプションの例

この例では、「Car Poll.jmp」サンプルデータを使用して、女性と男性について既婚者の割合を調べてみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「既婚/未婚」を選択し、[Y, 目的変数] をクリックします。
4. 「性別」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから〔関連の指標〕を選択します。

図7.18 「関連の指標」レポートの例

性別と既婚/未婚の分割表に対する分析				
モザイク図				
分割表				
検定				
関連の指標				
指標	値	標準誤差	下側95%	上側95%
ガンマ	0.1667	0.1184	-0.0654	0.3987
Kendallのタウ-b	0.0795	0.0570	-0.0321	0.1911
Stuartのタウ-c	0.0757	0.0543	-0.0307	0.1821
SomersのD (C R)	0.0763	0.0547	-0.0309	0.1835
SomersのD (R C)	0.0828	0.0593	-0.0335	0.1991
非対称ラムダ(C R)	0.0000	0.0000	0.0000	0.0000
非対称ラムダ(R C)	0.0000	0.0000	0.0000	0.0000
対称ラムダ	0.0000	0.0000	0.0000	0.0000
不確実性係数(C R)	0.0049	0.0070	0.0000	0.0186
不確実性係数(R C)	0.0046	0.0066	0.0000	0.0176
不確実性係数(対称)	0.0047	0.0068	0.0000	0.0181

分析の対象である変数（性別と既婚/未婚）が名義尺度なので、ラムダと不確実性係数を使用します。いずれの値も小さくなっているので、関連性は弱いようです。

Cochran-Armitageの傾向検定の例

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

この検定の目的に合わせて、「サイズ」を順序尺度に変更します。

2. 「列」パネルで、「サイズ」の横にあるアイコンを右クリックし、[順序尺度] を選択します。
3. [分析] > [二変量の関係] を選びます。
4. 「性別」を選択し、[Y, 目的変数] をクリックします。
5. 「サイズ」を選択し、[X, 説明変数] をクリックします。
6. [OK] をクリックします。
7. 「分割表に対する分析」の横にある赤い三角ボタンをクリックし、メニューから[Cochran-Armitageの傾向検定] を選択します。

図7.19 「Cochran-Armitageの傾向検定」レポートの例

サイズと性別の分割表に対する分析			
モザイク図			
分割表			
検定			
Cochran-Armitageの傾向検定			
漸近検定	Z	p値(Prob<Z)	p値(Prob> Z)
	-0.37271	0.3547	0.7094

両側検定のp値(0.7094)は大きい値になっています。このことから、各サイズの車両を購入する男性と女性の割合に関係があるとは結論できません。

「分割表」プラットフォームの統計的詳細

ここでは、「分割表」プラットフォームのいくつかの統計手法の詳細について説明します。

【一貫性の統計量】 オプション

2名の評価者が、 n 個の対象を判定したとします。2人の判定者の判定が完全に一致している場合は、カッパ係数が1になります。観測された一致が、偶然によって一致する期待値を上回ると、カッパ係数は正の値になります。カッパ係数が大きいほど、一貫性の高いことを示します。実際にはほとんどあり得ませんが、観測された一致が、偶然によって一致する期待値を下回ると、カッパは負になります。カッパ係数の最小値は、周辺度数に応じて-1から0までの間の値を取ります。

カッパ統計量は、次式により計算されます。

$$\hat{\kappa} = \frac{P_0 - P_e}{1 - P_e} \text{ この式で、 } P_0 = \sum_i p_{ii} \text{ および } P_e = \sum_i p_{i.} p_{.i}$$

単純なカッパ係数の漸近分散は、次のような推定式で計算されます。

$$\text{var} = \frac{A + B - C}{(1 - P_e)^2 n} \text{ この式で } A = \sum_i p_{ii} [1 - (p_{i.} + p_{.i})(1 - \hat{\kappa})]^2, B = (1 - \hat{\kappa})^2 \sum_{i \neq j} \sum p_{ij} (p_{.i} + p_{.j})^2 \text{ および}$$

$$C = [\hat{\kappa} - P_e(1 - \hat{\kappa})]^2$$

Fleiss, Cohen, and Everitt (1969) を参照してください。

対称性に対する Bowker 検定は、「分割表における確率が対称である ($p_{ij}=p_{ji}$ である)」という帰無仮説を検定します。

【オッズ比】 オプション

オッズ比は次の式で計算されます。

$$\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}$$

上の式で、 p_{ij} は 2×2 分割表の第 i 行第 j 列の度数です。

「検定」レポート

R2 乗 (U)

R2 乗 (U) は次の式で計算されます。

$$\frac{\text{モデルの「(-1)* 対数尤度」}}{\text{全体の修正済み「(-1)* 対数尤度」}}$$

分母の「(-1)* 対数尤度」は、標本全体において応答確率が一定であるモデルのものです。

検定

次の2種類の検定が行われます。

「尤度比」カイ2乗検定の検定等計量は、「検定」レポートに表示されている「(-1)* 対数尤度」を2倍した値です。この「(-1)* 対数尤度」は、グループごとに応答確率をあてはめたモデルの対数尤度から、標本全体において応答確率が一定であるモデルの対数尤度を引いたものです。尤度比カイ2乗値は、 G^2 という記号で示されることがあり、次式で求めます。

$$G^2 = 2 \left[\sum_{ij} (-n_{ij}) \ln(p_j) - \sum_{ij} -n_{ij} \ln(p_{ij}) \right] \text{この式で、} p_{ij} = \frac{n_{ij}}{N} \text{ および } p_j = \frac{N_j}{N}$$

この式は簡略に次式のようにも表せます。

$$G^2 = 2 \sum_i \sum_j n_{ij} \ln \left(\frac{n_{ij}}{e_{ij}} \right)$$

「Pearson」カイ2乗検定の検定統計量は、各セルにおける観測度数と期待度数の差の平方和から計算されます。Pearson のカイ2乗検定は、大標本では度数が正規分布に従う傾向にあるという性質を用いています。Pearson のカイ2乗統計量を算出するには、次式がよく知られています。

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

上式において、O はセルの観測度数、E はセルの期待度数です。和記号は、すべてのセルの合計を表しています。2×2 分割表に対してときどき採用されている連続修正は、JMP では行っていません。

対応分析の「詳細」レポート

対応分析の「詳細」レポートに表示される特異値は、次の行列の特異値です。

$$D_r^{-0.5} (P - rc') D_c^{-0.5}$$

ここで、

- P は、各セルの度数を合計度数で割ったものを要素としてもつ行列。
- r は P の行和。 c は P の列和。
- D は、 r と c を対角要素とする対角行列。

第8章

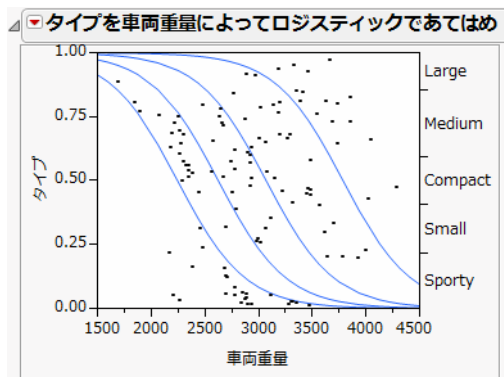
ロジスティック分析

カテゴリカルなY変数と連続尺度のX変数の関係を調べる

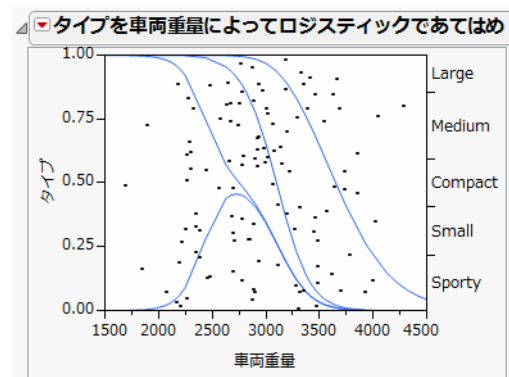
「ロジスティック」プラットフォームは、カテゴリカルな応答変数（Y）の確率を、連続尺度の説明変数（X）にあてはめます。モデルをあてはめた結果、各X値における応答変数の確率を予測する式が作成されます。「ロジスティック」プラットフォームは、「二変数の関係」プラットフォームのうち、応答変数が**名義尺度**もしくは**順序尺度**で、説明変数が**連続尺度**の場合に実行される分析です。応答変数が名義尺度のときには名義ロジスティック回帰、順序尺度のときには順序ロジスティック回帰が行われます。

- 名義ロジスティック回帰では、応答変数の値ごとに、各値が生じる確率のロジスティック曲線が計算されます。
- 順序ロジスティック回帰では、応答変数の値が与えられた値以下になる確率がモデル化されます。そのため、ロジスティック曲線は1つだけ計算され、それを水平に移動させることですべての順序カテゴリの確率が推定されます。このモデルは、名義ロジスティックよりも複雑でなく、順序尺度の応答変数に適しています。

図8.1 ロジスティック回帰の例



順序ロジスティック回帰



名義ロジスティック回帰

ロジスティック回帰の概要

ロジスティック回帰は古くから普及している手法で、医薬品開発における用量反応データや、マーケティングの購買選択など、いろいろな分野で応用されています。非常に単純な手法であるにもかかわらず、初歩的な統計の授業では残念ながらあまり取り上げられません。ロジスティック回帰の文献（Hosmer and Lemeshow 1989）のほかに、多くのカテゴリカル分析のテキスト（Agresti 1998）で取り上げられています。正規分布の分布関数を使用する手法もあり、その場合の分析は**プロビット分析**と言います。また、同じ状況でも、カテゴリカル変数を X 、連続変数を Y として考えて、ロジスティック回帰ではなく判別分析を行う人もいます。判別分析では、連続量のデータが固定された説明変数ではなく正規分布に従ったランダムな応答であると仮定されます。

単純ロジスティック回帰は、「モデルのあてはめ」プラットフォームで行われるカテゴリカル応答の分析を、単純かつグラフィカルにしたものです。より複雑なロジスティック回帰モデルについては、『基本的な回帰モデル』の「ロジスティック回帰」章を参照してください。

名義ロジスティック回帰

名義ロジスティック回帰では、ある水準に応答変数になる確率を、説明変数の滑らかな関数によって表します。あてはめた確率は $0 \sim 1$ の間をとります。また、説明変数のある値に対する確率を、すべての応答変数の水準で合計すると、1になります。

ロジスティック確率プロットでは Y 軸が確率を表します。応答変数の水準数を k とすると、 $k - 1$ 本の滑らかな曲線が作成され、それによって合計確率（1）が応答水準ごとに分割されます。ロジスティック回帰のあてはめでは、「観測された応答データが生じる確率の自然対数」を合計した値の符号を逆にしたものが最小化されます。この推定方法を「最尤法」と呼びます。

順序ロジスティック回帰

Y が順序尺度の場合、順序性を考慮したロジスティック回帰が行われます。応答変数が特定の水準以下になる累積確率がロジスティック曲線で表されます。どの水準に対しても曲線の形状は同じで、ただ水平方向に位置がずれているだけです。

順序ロジスティックモデルでは、応答変数の水準数を r としたとき、累積ロジスティック曲線の $r - 1$ 個の異なる切片を推定しますが、傾きは1つです。実際に必要になることはまずありませんが、各パラメータ推定値を個別に検定することも可能です。

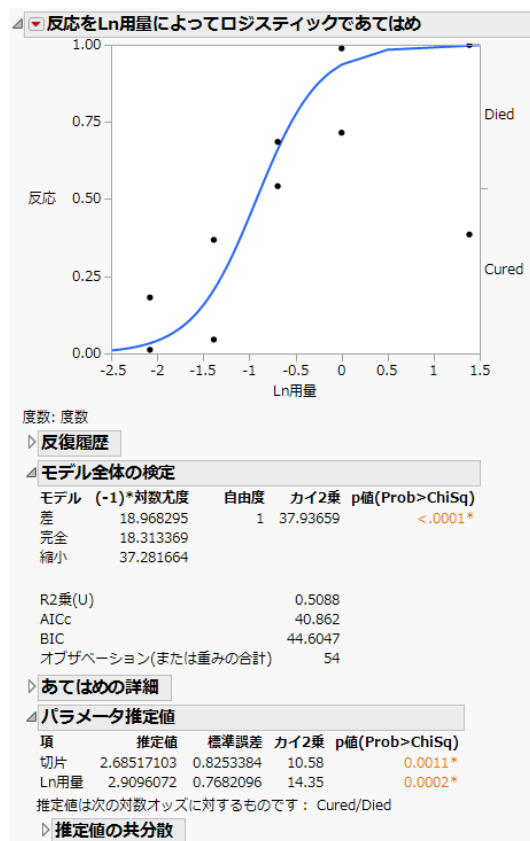
順序モデルは計算されるパラメータが名義モデルより少ないため、適切であれば順序モデルを使用した方が効率的です。特に、順序尺度の応答変数に水準が何百もあるときに適しています。

名義ロジスティック回帰の例

この例では、「Penicillin.jmp」サンプルデータを使用します。このデータは、ウサギを12匹ずつ5群に分け、連鎖球菌を注入した実験から得たものです。体内に連鎖球菌が認められたウサギに対して、用量を変えてペニシリンを投与します。ウサギの治癒率に対して、用量の自然対数（「Ln用量」）が効果を持つかどうかを調べます。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Penicillin.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「反応」を選択し、[Y, 目的変数] をクリックします。
4. 「Ln 用量」を選択し、[X, 説明変数] を選択します。
「度数」が自動的に「度数」に設定されます。これは、「度数」の役割が、あらかじめ「度数」列に割り当てられているためです。
5. [OK] をクリックします。

図8.2 名義ロジスティックのレポートの例



あてはめられたモデルを示すグラフが作成されます。治癒の予測確率が「Ln用量」の関数として表示されます。p値は統計的に有意で、「薬の用量は、ウサギが治癒する確率に影響している」と結論できます。

ヒント：分析された応答水準を変更するには、「値の順序」列プロパティを使用します。

「ロジスティック」プラットフォームの起動

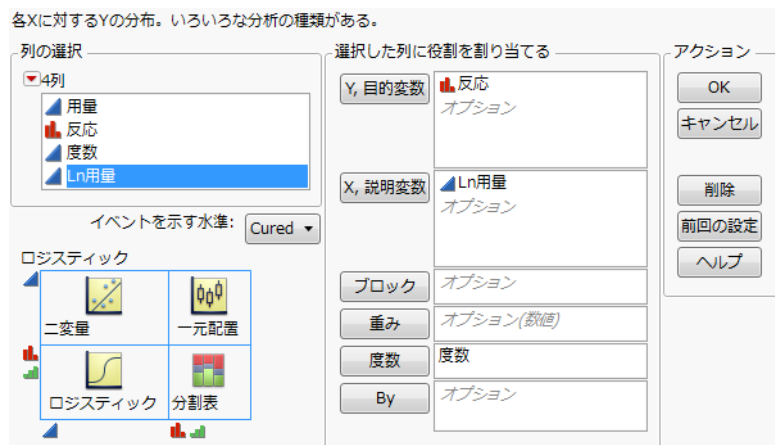
ロジスティック分析を実行するには、次の手順に従ってください。

1. [分析] > [二変量の関係] を選びます。
2. 名義尺度または順序尺度の列を選択し、[Y, 目的変数] をクリックします。
3. 連続尺度の列を選択し、[X, 説明変数] をクリックします。

模式図に、ロジスティック分析を行われようとしていることが示されます。

メモ：「JMPスターター」ウィンドウから、ロジスティック分析を起動することもできます。「[「JMPスターター」ウィンドウからの各分析の実行](#)」（95ページ）を参照してください。

図8.3 「二変量の関係」のロジスティック分析起動ウィンドウ



応答変数が2水準の名義尺度である場合、起動ダイアログにおいて「イベントを示す水準」というメニューが表示されます。このドロップダウンリストで、2水準のうちにどちらの水準に興味があるか（オッズの分子にどちらの水準をもって来るか）を指定できます。

「二変量の関係」の起動ウィンドウの詳細については、「[「二変量の関係」プラットフォームの概要](#)」章（93ページ）を参照してください。

[OK] をクリックすると、「ロジスティック」レポートウィンドウが開きます。「**「ロジスティック」レポート**」(239 ページ) を参照してください。

データ構造

要約されていないデータ、および要約されているデータのいずれも使用可能です。

要約されていないデータ 各行がそのX値とY値を持つ1つの観測値を表します。

要約されたデータ XとYの値の組み合わせが同じである複数の観測値が1行にまとめられているデータです。データテーブルには、各行の度数を示す列も必要となります。起動ウィンドウで、この列を [度数] に指定してください。

「ロジスティック」レポート

図8.4は、「**名義ロジスティック回帰の例**」(237 ページ) の手順に従って作成したものです。

図8.4 「ロジスティック」レポートの例



「ロジスティック」レポートウィンドウには、ロジスティックプロット、「反復回数」レポート、「モデル全体の検定」レポート、「パラメータ推定値」レポートが表示されます。

メモ: 赤い三角ボタンをクリックするとメニューが開き、最初に表示されるレポートウィンドウに情報を追加するための各種オプションが表示されます。「[「ロジスティック」プラットフォームのオプション](#)」(243ページ)を参照してください。

ロジスティックプロット

ロジスティックプロットを見ると、ロジスティック回帰の全体像を把握できます。Xのそれぞれの値において、応答変数のカテゴリごとの確率にY軸が区切られています。曲線と曲線の間の縦方向の間隔が確率を表し、すべてのYカテゴリにわたって確率を合計すると1になります。

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

関連情報

- [「ロジスティックプロットの例2」](#) (247ページ)

反復履歴

「反復履歴」レポートには、各反復計算とモデルの収束判定基準が表示されます。「反復履歴」は名義ロジスティック回帰の場合にのみ表示されます。

モデル全体の検定

「モデル全体の検定」レポートには、「データ全体で応答確率が一定である」と仮定したときよりも、現在のモデルがどれくらい良くあてはまっているかが示されます。このレポートは、連続尺度の応答に対する分散分析表に相当します。ここでは、モデルがデータを説明しているかどうか、カイ2乗近似の尤度比検定で検定されます。

観測された確率の自然対数を取り、その合計の符号を逆にしたものが「**(-1)*対数尤度**」です。カテゴリカルな応答における「**(-1)*対数尤度**」は、連続尺度の応答における平方和に相当します。現在のモデルの「**(-1)*対数尤度**」から、全体で確率が一定であると仮定したモデルの「**(-1)*対数尤度**」を引いた差を2倍したものが、ここでの尤度比検定のカイ2乗統計量です。この尤度比検定は、「X変数が応答変数に対して、まったく効果を持たない」という帰無仮説を検定します。

「**R2乗(U)**」(R^2)は0~1の値を取ります。モデルのあてはまりがよい場合は R^2 値が高い値になりますが、カテゴリカルなモデルでは高い値になることはほとんどありません。

「モデル全体の検定」レポートには、次の情報が表示されます。

モデル（「要因」ともいう）

- 「**縮小**」モデルは、切片だけを含むモデルです。
- 「**完全**」モデルは、すべての効果と切片を含むモデルです。
- 「**差**」は、完全モデルと縮小モデルの対数尤度の差です。

自由度 モデルに関係する自由度。

(-1)*対数尤度 データの変動やばらつきを表す指標。**不確定性**ともいう。

「**完全**」（完全モデル）は、あてはめたモデルの「(-1)*対数尤度」（不確定性）です。ロジスティック回帰では、線形式をロジスティック変換して、確率が予測されます。この予測された確率から計算される不確定性が最小になるように、モデルがあてはめられます。

「**縮小**」（縮小モデル）は、応答確率が全体で一定であると仮定し、応答全体の割合で応答確率を推定したときの「(-1)*対数尤度」（不確定性）です。これは、モデルに効果がないときの不確定性を表します。

上の2つの「(-1)*対数尤度」の差は、モデルをあてはめた結果生じた対数尤度の減少で、それに2を掛けた値が尤度比カイ2乗です。

詳細については、『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。

カイ2乗 尤度比カイ2乗。「応答確率が標本全体で一定であるとしたときよりも、モデルの適合度は良いわけではない」という帰無仮説を検定する値です。これは「差」の「(-1)*対数尤度」を、2倍したものです。つまり、あてはめたモデルの対数尤度から、標本全体で応答確率が一定であるとしたモデルの対数尤度を引いた差を、2倍したものです。詳細は、「[「モデル全体の検定」レポート](#)」(253ページ)を参照してください。

p値 (Prob>ChiSq) カイ2乗検定の観測有意確率 (p 値)。これは、まったくの偶然だけで、現在のカイ2乗値よりも大きいカイ2乗値を得る確率です。通常、p値が0.05より小さいとモデルが有意だと考えられます。

R2乗 (U) 不確定性のうち、あてはめたモデルに起因する部分。負の対数尤度の「差」を「縮小」で割った値で定義される。R2乗 (U) の値が1である場合、イベントが発生する予測確率が1に等しい（予測確率に不確定性はない）ということを意味します。ロジスティックモデルでは、予測確率を確定できることはまれであるため、R2乗 (U) は、通常、小さい値になります。詳細は、「[「モデル全体の検定」レポート](#)」(253ページ)を参照してください。

メモ : R2乗 (U) は、**McFadden の疑似 R2乗**と呼ばれることもあります。

AICc 修正済みの赤池の情報量規準 (Corrected AIC)。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。

BIC ベイズ情報量規準。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。

オブザベーション（または重みの合計）計算に使用したすべてのデータの標本サイズ。ただし、**重み変数**を指定した場合は、重みの合計です。

指標 次の適合度指標が表示されます。

エントロピー R2 乗 あてはめたモデルの対数尤度と、切片だけのモデルの対数尤度を比較します。

一般化 R2 乗 一般的な回帰モデルに適用可能な指標です。尤度関数 L に基づいており、最大値が 1 になるように尺度化されています。正規分布に従う連続尺度の応答変数に対する通常の R2 乗に相当します。一般化 R2 乗は、Nagelkerke の R^2 または Craig and Uhler の R^2 と呼ばれ、Cox and Snell の疑似 R^2 を正規化したものです。Nagelkerke (1991) を参照してください。

平均 -Log p $-\log(p)$ の平均。p は、実際に発生したイベントの確率の予測値です。

RMSE 誤差の標準偏差。応答と p (実際に発生したイベントの確率の予測値) の差を示します。

平均 絶対偏差 応答と p (実際に発生したイベントの確率の予測値) の差の絶対値の平均。

誤分類率 確率の予測値が最も大きい応答カテゴリが、観測されたカテゴリではない確率。

[エントロピー R2 乗] と [一般化 R2 乗] は、1 に近いほど適合度が良いことを示します。[平均 -Log p]、[RMSE]、[平均 絶対偏差]、[誤分類率] は、値が小さいほど、適合度が良いことを示します。

学習 適合度指標の値。

定義 適合度指標の定義式。

パラメータ推定値

名義ロジスティックモデルでは、応答水準の数を k としたとき、 $k-1$ 個の比較ごとに切片と傾きのパラメータがあてはめられます。「パラメータ推定値」レポートには、次のような推定値が表示されます。実際に必要になることはまずありませんが、各パラメータ推定値を個別に検定することも可能です。

項 ロジスティックモデル内のパラメータ。応答変数の各水準（最後の水準を除く）における因子ごとに、切片と傾きが表示されます。

推定値 ロジスティックモデルで計算されたパラメータ推定値。

標準誤差 各パラメータ推定値の標準誤差。各項を 0 と比較する統計量の計算に使用されます。

カイ 2 乗 各パラメータが 0 であるという仮説を検定する Wald 検定。Wald のカイ 2 乗は $(\text{推定値} / \text{標準誤差})^2$ という式で計算されます。

p 値 (Prob>ChiSq) カイ 2 乗検定の観測有意確率。

推定値の共分散

パラメータ推定値の分散および共分散の推定値が表示されます。分散推定値の平方根は、「標準誤差」で表示される数値に一致します。

「ロジスティック」プラットフォームのオプション

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

オッズ比 「パラメータ推定値」レポートにオッズ比を追加します。詳細については、『基本的な回帰モデル』の「ロジスティック回帰」章を参照してください。

このオプションは、応答の水準数が2つの場合のみ使用できます。

逆推定 与えられたYの値からXの値を推定します。詳細は、「[逆推定](#)」(245ページ)を参照してください。

このオプションは、応答の水準数が2つの場合のみ使用できます。

ロジスティックプロット ロジスティックプロットの表示／非表示を切り替えます。

プロットのオプション 「プロットのオプション」サブメニューには次のオプションが表示されます。

点の表示 点の表示／非表示を切り替えます。

応答率折れ線の表示 各X値に複数の点がある場合に役立つコマンドです。各X値のグループごとに応答の割合が計算されます。このように計算された応答の割合を、ロジスティック曲線と比較することができます。度数が少なく応答の割合が0か1になってしまうグループが多数生じるのを防ぐため、点が3つ以上あるX値のグループに対してだけ、応答の割合は表示されます。

線の色 プロットの曲線の色を指定できます。

ROC曲線 受診者動作特性(ROC)曲線は、Xの値ごとに、「1-特異度」に対して感度をプロットしたものです。「[ROC曲線](#)」(244ページ)を参照してください。

リフトチャート このコマンドを選択すると、リフトチャートが描画されます。リフトチャートは、ROC曲線と同じ情報を使って計算されますが、順序付けした情報の上位の部分に特に重点を置きます。Y軸には、予測確率が上位のグループにおいて、該当する応答水準をもつ割合がどれくらい高いかが、集団全体における割合に対する比として表示されます。リフトチャートの詳細については、『基本的な回帰モデル』の「ロジスティック回帰」章を参照してください。

確率の計算式の保存 データテーブル内の新しい列に計算式が保存されます。「[確率の計算式の保存](#)」(245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。

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。

ROC 曲線

ある疾患にかかっているかどうかを、次のように X の値によって診断するとします。

- X の値が特定の閾値より大きいときは、その疾患にかかっているとします。
- X の値が閾値より小さいときは、その疾患にかかっていないと診断するとします。

たとえば、特定の種類のガンを診断する検査では、ある血液成分の濃度を測定します。診断における閾値をいろいろな値に変更し、それに伴って変化した偽陽性率と偽陰性率をプロットしてみれば、診断の良し悪しが分かります。分析の目的は、真陰性と真陽性を最も高める X の基準値を求めることです。受診者動作特性 (ROC) 曲線を見ると、これらの値がどれだけ急激に変化するかがわかります。理想的な診断では、ROC 曲線における曲線下の面積 (AUC: Area Under Curve) が大きくなります。

次の2つは、医学の分野で使用されている概念です。

- **感度** ある状態が存在しているときに、その状態が存在していることを、与えられた X の値によって (診断や基準値によって) 正確に予測する確率。 X の値が与えられているときに、状態が本当は存在しているのに存在していないと予測される確率は、「1-感度」です。
- **特異度** ある状態が存在していないときに、その状態が存在しないことを、診断によって正確に予測する確率。

受診者動作特性 (ROC) 曲線は、 X の値ごとに、「1-特異度」に対して感度をプロットしたものです。ROC 曲線より下の面積 (AUC: Area Under Curve) は、曲線によって示唆された情報を要約するための指標として使用されます。

単純ロジスティック回帰の応答が2値の場合は、ROC 曲線コマンドがプラットフォームのポップアップメニューに表示されます。[ROC 曲線] オプションを選択すると、どちらの水準を「陽性」にするかを指定するウィンドウが開きます。

検査による予測が完全であれば、1つの値を境にして、その上にすべての異常な標本値が位置し、下にすべての正常な値が位置するようになります。その場合、曲線は完璧な感度を持ち、グリッド上の (0,1) の点を通ります。ROC 曲線がこの理想の点に近ければ近いほど、その判別能力は高くなります。逆に、予測することがまったくできない場合は、ROC 曲線はグリッドの対角線上を通ります (DeLong, et al. 1988)。

ROC 曲線は、偽陽性率と偽陰性率の関係をグラフ化したものです。関係は通常、レポートのプロットの下に表示される曲線下の面積 (AUC) に基づいて評価されます。プロット内には、ROC 曲線に接する45度の線が黄色で表示されます。黄色い線と曲線との交点は、偽陰性と偽陽性のコストが等しいと仮定したときに、コストが最小になる点を示しています。

関連情報

- 「ROC 曲線の例」(249 ページ)

確率の計算式の保存

【確率の計算式の保存】 オプションを選択するとデータテーブルに新しい列が作成されます。これらのデータテーブル列には、次の情報が保存されます。

- X因子の線形結合式（ロジット）
- 各水準が生じる確率の予測式
- 最も生じる確率が高い応答変数の値を計算する予測式

逆推定

逆推定は予測の反対で、与えられたYの値からXの値を推定するものです。ただし、ロジスティック回帰では、Yの値ではなくYの1水準の確率を使います。逆推定は、応答のカテゴリが2つ（2値応答）の場合にしか機能しません。

「モデルのあてはめ」プラットフォームにも、逆推定の信頼限界を計算するオプションがあります。逆推定の詳細については、『基本的な回帰モデル』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。

関連情報

- 「[十字ツールを使用した逆推定の例](#)」（250ページ）
- 「[\[逆推定\] オプションを使用した逆推定の例](#)」（251ページ）

ロジスティック回帰のその他の例

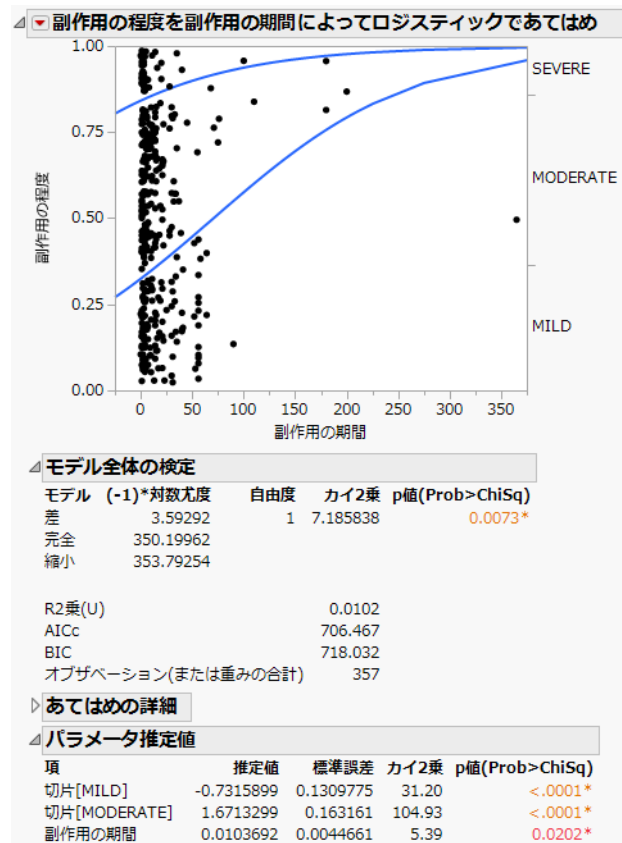
ここでは、ロジスティック回帰を使った例をさらに紹介します。

順序ロジスティック回帰の例

この例では、「AdverseR.jmp」サンプルデータを使用します。副作用の程度を、治療期間の関数として、順序ロジスティック回帰によりモデル化します。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「AdverseR.jmp」を開きます。
2. 「副作用の程度」の左側のアイコンを右クリックし、順序尺度に変更します。
3. [分析] > [二変量の関係] を選びます。
4. 「副作用の程度」を選択し、[Y, 目的変数] をクリックします。
5. 「副作用の期間」を選択し、[X, 説明変数] をクリックします。
6. [OK] をクリックします。

図8.5 順序ロジスティックのレポートの例



このレポートの内容は、名義尺度のレポートと同様です。「[「ロジスティック」レポート](#)」(239ページ)を参照してください。

プロットにおける点のX座標は、データ値をそのまま反映しています。一方、点のY軸は、同じY値をもつデータをランダムに散らばせて表示しています。ランダムに散らばせることにより、点の間に間隔をあげ、各点を見やすくしています。

点の数が多いところでは曲線と曲線の間隔が広くなり、点の数が少ないところでは曲線どうしが接近しています。データ点が曲線の間隔を押し広げる傾向にあるのは、確率の積が最大になるようにモデルが推定されるからです。つまり、混み合った点の確率が小さくならないように、曲線は求められています。計算の詳細については、『基本的な回帰モデル』を参照してください。

「モデル全体の検定」レポートと「パラメータ推定値」レポートの詳細については、「[「ロジスティック」レポート](#)」(239ページ)を参照してください。「パラメータ推定値」レポートを見ると、切片パラメータは最後の応答水準を除いたすべての水準について計算されていますが、傾きパラメータは1つしかありません。切片パラメータは応答水準間の間隔を示し、常に単調増加しています。

ロジスティックプロットの例2

この例では、「Car Physical Data.jmp」サンプルデータを使用して、ロジスティックプロットの例をもう1つ紹介します。「車両重量」によって、116台の車両のサイズ（「タイプ」）を予測するとしましょう。車両サイズは小さいものから順に、「Sporty」、「Small」、「Compact」、「Medium」、「Large」とします。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Physical Data.jmp」を開きます。
2. 「列」パネルで、「タイプ」の左側のアイコンを右クリックし、[順序尺度] を選択します。
3. 「タイプ」を右クリックし、[列情報] を選択します。
4. 「列プロパティ」メニューから [値の順序] を選択します。
5. 上から順に「Sporty」、「Small」、「Compact」、「Medium」、「Large」となるようにデータを移動します。
6. [OK] をクリックします。
7. [分析] > [二変量の関係] を選びます。
8. 「タイプ」を選択し、[Y, 目的変数] をクリックします。
9. 「車両重量」を選択し、[X, 説明変数] をクリックします。
10. [OK] をクリックします。

レポートウィンドウが表示されます。

図8.6 「タイプ」を「車両重量」によってロジスティックであてはめたプロットの例

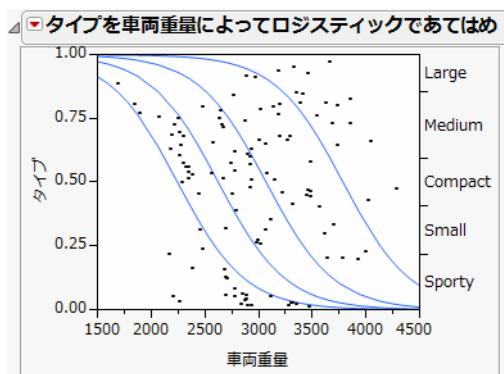
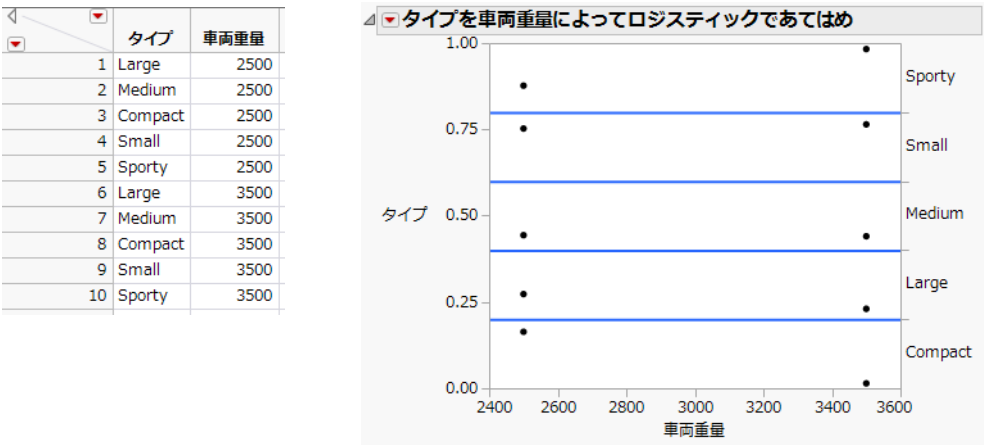


図8.6を見ると、次のことがわかります。

- 最初（一番下）の曲線は、所定の「車両重量」の車両が「Sporty」である確率を表しています。
- 2番目の曲線は、車両が「Small」または「Sporty」である確率を表しています。1番目と2番目の曲線の垂直距離が、車両が「Small」である確率です。
- プロットを見ると、予想どおり、重量が重い車両ほど「Large」である確率が高くなります。
- プロット上に描かれている点のX座標での位置はデータの値を反映していますが、Y軸方向では応答カテゴリの内側でランダムに散らばっています。

X変数が応答変数に対して効果を持たない場合はロジスティック曲線は水平になります。つまり、応答変数の各水準の確率が、説明変数の範囲全体にわたって一定になります。図8.7のロジスティックプロットが、ロジスティック曲線が水平となる例です。このような場合では、「タイプ」の確率を予測するのに、「車両重量」は役立ちません。

図8.7 予測が無意味なデータとロジスティックプロットの例

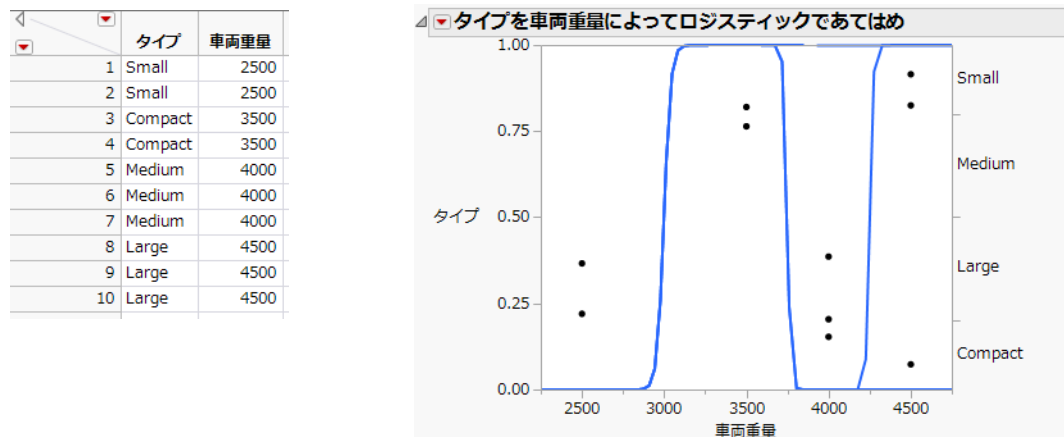


メモ: 図8.7および図8.8のプロットを再現するには、図中のデータテーブルを作成してから、この節の冒頭で紹介した手順7～10を実行します。

説明変数の値から応答変数の値を完全に予測できる場合、ロジスティック曲線は事実上垂直になります。応答の予測値はどの説明変数の値においてもほぼ確か（確率がほとんど1）です。図8.8のロジスティックプロットでは、「車両重量」が「タイプ」をほぼ完全に予測しています。

メモ: この場合、パラメータ推定値が非常に大きくなり、回帰レポート上に「不安定」と表示されます。

図8.8 予測が完全となるデータとロジスティックプロット



ROC 曲線の例

ROC 曲線を表示するには、次の手順に従います。

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

「度数」が自動的に「度数」に設定されます。これは、「度数」の役割が、あらかじめ「度数」列に割り当てられているためです。

5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、メニューから [ROC 曲線] を選択します。
7. 「Cured」を陽性として選択します。
8. [OK] をクリックします。

メモ: この例では、名義尺度の応答の ROC 曲線を示します。順序尺度の ROC 曲線の詳細については、『予測モデルおよび発展的なモデル』の「パーティション」章を参照してください。

この例では、「反応」を応答として、「Ln 用量」を説明変数としています。応答を予測する確率が、前述した説明によって計算され、それらの結果に基づき ROC 曲線が描かれます。なお、「ROC テーブル」では、「感度 - (1 - 特異度)」が最も大きい行にアスタリスクが付けられます。

図8.9 ROC 曲線と「ROC テーブル」の例



ROC 曲線が対角線より明らかに上にあるため、モデルの予測能力は高いと言えます。

十字ツールを使用した逆推定の例

ウサギを用いた試験をもとに、50%の確率でウサギが治癒する投与量（の自然対数）を求めるとします。治癒率が50%になるときの投与量は、**ED50**（Effective Dose 50）と呼ばれています。この例では、十字ツールを使用して、グラフ上で逆推定の大まかな値を求めてみます。

治癒率が50%となる「Ln 用量」を求めるには、次の手順に従います。

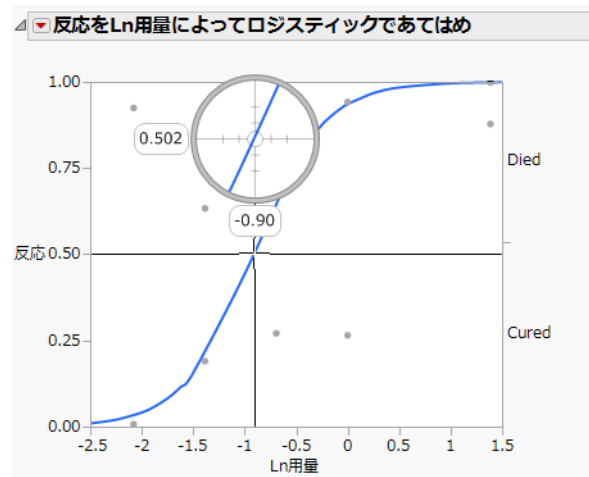
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Penicillin.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「反応」を選択し、[Y, 目的変数] をクリックします。
4. 「Ln 用量」を選択し、[X, 説明変数] を選択します。

「度数」が自動的に [度数] に設定されます。これは、[度数] の役割が、あらかじめ「度数」列に割り当てられているためです。

5. [OK] をクリックします。

6. 十字ツールをクリックします。
 7. まず、縦軸（「反応」）が0.5となる位置に十字ツールを合わせ、次に、ロジスティック曲線との交点まで、カーソルを動かします。
 8. 十字ツールのカーソルと、ロジスティック曲線が交った点で、横軸における「Ln用量」の値を読みます。
- この例では、「Ln用量」の値が約-0.9のときに、治癒率と致死率が等しくなります。

図8.10 ロジスティックプロット上の十字ツールの例



〔逆推定〕 オプションを使用した逆推定の例

応答の水準が2つのときは、〔逆推定〕オプションを選択すれば正確な逆推定が計算されます。低い方の応答カテゴリが生じる確率に対するX値と、そのX値の信頼区間が求められます。

〔逆推定〕オプションを使用するには、次の手順に従います。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Penicillin.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「反応」を選択し、[Y, 目的変数] をクリックします。
4. 「Ln用量」を選択し、[X, 説明変数] を選択します。
「度数」が自動的に「度数」に設定されます。これは、「度数」の役割が、あらかじめ「度数」列に割り当てられているためです。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、〔逆推定〕を選択します。図8.11を参照してください。
7. 「信頼水準」の入力ボックスに「0.95」と入力します。
8. 「信頼水準」のドロップダウンリストから「両側」を選択します。

9. 対象となる応答確率を指定します。この例では、「0.5」と「0.9」を入力します。つまり、治癒率が0.5と0.9であるときの「Ln用量」の値を求めます。
10. [OK] をクリックします。
- 逆推定プロットが表示されます。

図8.11 「逆推定」ウィンドウ

逆推定したい確率値を1つ以上、指定してください。

Ln用量 (予測対象)

信頼水準

0.95

両側

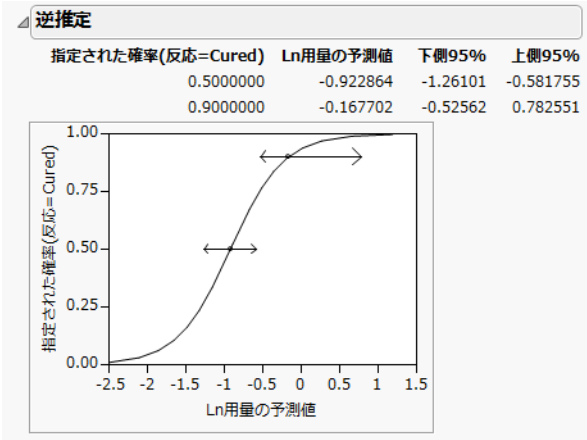
確率(反応=Cured)

OK

キャンセル

ヘルプ

図8.12 逆推定プロットの例



Xの推定値と信頼区間がレポートと確率プロットに示されます。たとえば、治癒率が90%になるときの「Ln用量」の両側信頼区間は、-0.526から0.783までとなっています。

「ロジスティック」プラットフォームの統計的詳細

「モデル全体の検定」レポート

カイ2乗

モデル全体に対する尤度比検定のカイ2乗統計量は、次式により計算されます。この統計量は、教科書では、ときどき、 G^2 という記号で表されています。

$$G^2 = 2(\sum -\ln p(\text{縮小モデル}) - \sum -\ln p(\text{モデル}))$$

ここでの総和は、すべてのオブザベーションに対する和を表しています。

R2乗(U)

縮小モデルの対数尤度に対するこの統計量の比が R^2 です。式では、次のように表されます。

$$\frac{\text{差の「(-1)*対数尤度」}}{\text{縮小モデルの「(-1)*対数尤度」}}$$

これらの統計量の値は、「モデル全体の検定」レポートに表示されています。

メモ: R2乗(U)は、**McFaddenの疑似R2乗**と呼ばれることもあります。

第9章

表の作成

対話的に要約テーブルを作成する

「表の作成」プラットフォームでは、記述統計量の表を対話的に作成できます。簡単かつ柔軟な方法で、要約データを表形式にできます。表は、グループ列、分析列、および各種統計量により構成されます。

図9.1 表の作成例

	性別											
	F						M					
	年齢						年齢					
	12	13	14	15	16	17	12	13	14	15	16	17
	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)
最小値	64	67	81	92	112	116	79	79	92	104	128	134
平均	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
最大値	145	112	142	112	115	116	128	105	119	128	128	172

		性別							
		女性				男性			
		既婚/未婚				既婚/未婚			
		既婚		未婚		既婚		未婚	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	サイズ	年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
米国	大型	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	中型	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	小型	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
ヨーロッパ	大型	34.0	7.071	28.0	.	.	.	26.0	.
	中型	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	小型	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
日本	大型	25.0	.	.	.	32.0	.	.	.
	中型	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	小型	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
生産国									
米国		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
ヨーロッパ		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
日本		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

タイプ		平均			
		会社規模	利益(\$M)	売上(\$M)	従業員一人あたりの利益
Computer	big		1089.9	20597.48	4530.478
	medium		-85.75	3018.85	-3462.51
	small		44.94	1758.06	7998.815
	すべて		240.87	5652.02	6159.015
Pharmaceutical	big		894.42	7474.04	17140.70
	medium		698.98	4261.06	24035.11
	small		156.95	1083.75	38337.19
	すべて		690.08	5070.25	23546.12
すべて		すべて	409.32	5433.86	12679.18
タイプ					
Computer			240.87	5652.02	6159.015
Pharmaceutical			690.08	5070.25	23546.12
すべて			409.32	5433.86	12679.18

「表の作成」プラットフォームの使用例

男子生徒と女子生徒の身長を測定したデータがあります。このデータを基に、男子生徒と女子生徒それぞれの平均身長と全体の平均を示す 図9.2のような表を作成します。

図9.2 平均身長を示す表

身長(インチ)	
性別	平均
F	60.9
M	63.9
すべて	62.6

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
身長が分析対象の変数であるため、これを表の一番上に表示します。
3. 「身長(インチ)」をクリックし、「列のドロップゾーン」にドラッグします。

図9.3 「身長」変数の追加後

表の作成

表に追加するには、列や統計量を、テーブルの列見出しや行ラベルの領域にドラッグ&ドロップしてください。

元に戻す

やり直し

完了

▼ 5列

名前

年齢

性別

身長(インチ)

体重(ポンド)

度数

重み

ページ列

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値 (メディアン)

幾何平均

四分位範囲

分位点

列%

行%

すべて

☐ グループ変数の欠測値を含める

☐ グループ変数の度数順

☐ 集計統計量の追加

デフォルトの統計量

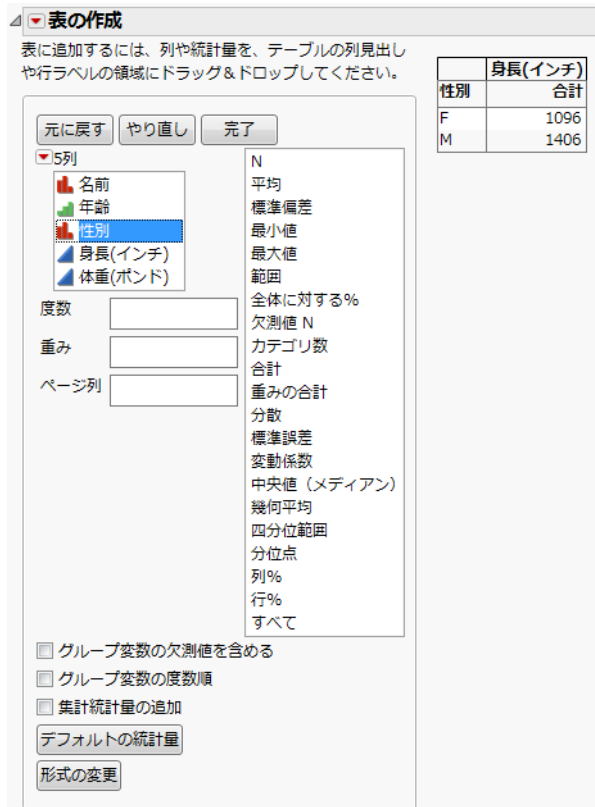
形式の変更

身長(インチ)	
	合計
	2502

性別の統計量が必要なため、性別を左側に表示します。

4. 「性別」をクリックしてドラッグし、数値「2502」の横にある空のセルにドロップします。

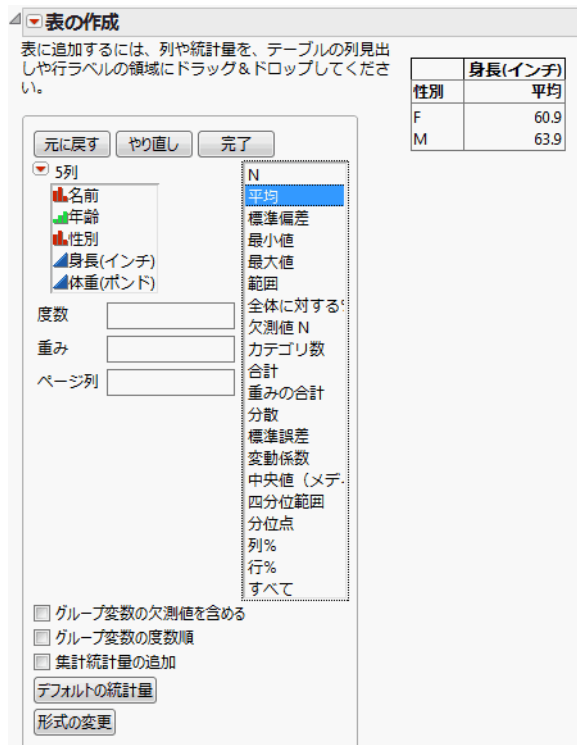
図9.4 「性別」変数の追加後



合計の代わりに平均を表示します。

5. 「平均」をクリックしてドラッグし、「合計」の上にドロップします。

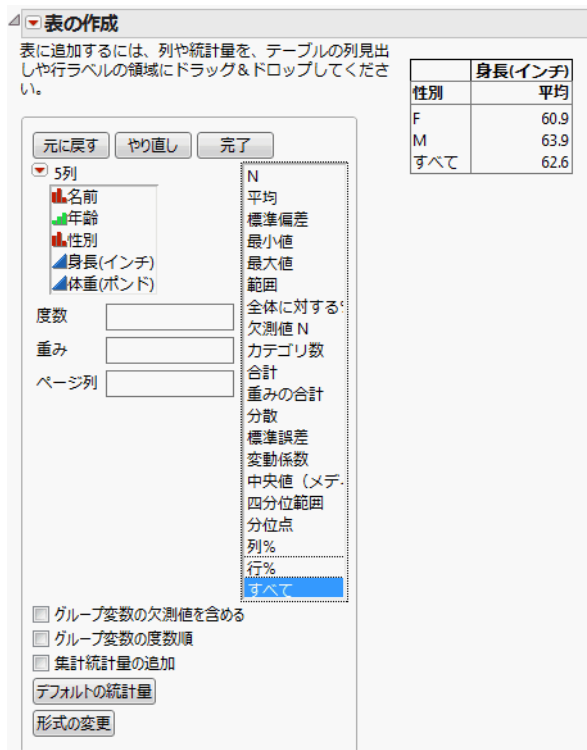
図9.5 「平均」の追加後



さらに、男子と女子を合わせた全体の平均も表示します。

6. 「すべて」をクリックしてドラッグし、「性別」の上にドロップします。または、「集計統計量の追加」チェックボックスをオンにします。

図9.6 「すべて」の追加後



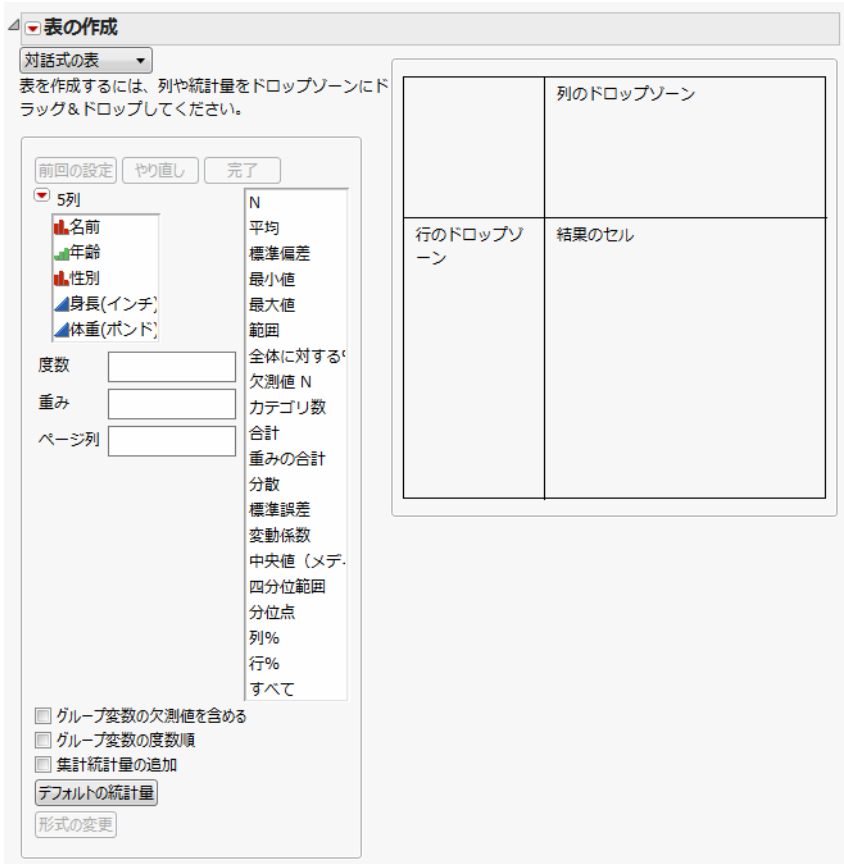
7. (オプション) 「完了」をクリックします。

完成した表に、女子と男子の平均身長と全体の平均身長が表示されます。

「表の作成」プラットフォームの起動

「表の作成」プラットフォームを起動するには、[分析] > [表の作成] を選択します。

図9.7 「表の作成」の対話的な表



メモ: 赤い三角ボタンのメニューのオプションについては、「[「表の作成」プラットフォームのオプション](#)」(269 ページ) を参照してください。

「表の作成」ウィンドウには、次のオプションがあります。

対話式の表／ダイアログボックス 2つのモード間で切り替えます。項目をドラッグ&ドロップしながら独自の表を作成していく場合は、[対話式の表] を、決まった形式の簡単な表を作成する場合は [ダイアログボックス] を選択します。「[ダイアログボックスの使用](#)」(261 ページ) を参照してください。

[統計量] オプション 標準的な統計量のリストが表示されます。統計量をリストから表にドラッグして表に組み込むことができます。「[統計量の追加](#)」(262 ページ) を参照してください。

列のドロップゾーン 列または統計量をドラッグしてこのゾーンにドロップし、列を作成します。

メモ: [統計量] オプションに表示されているものと同じ名前の列がデータテーブルにある場合は、必ず列リストから列名をドラッグ&ドロップしてください。そうしないと、表で同名の統計量が代用される場合があります。

行のドロップゾーン 列または統計量をドラッグしてこのゾーンにドロップし、行を作成します。

結果のセル ドラッグ&ドロップした列や統計量に基づいて結果のセルが表示されます。

度数 この役割を割り当てた列の値は、各行の度数（繰り返し数）を表します。このオプションは、分析対象のデータが要約されたものである場合に、度数を割り当てののに便利です。

重み データに重み（重要度、影響度など）をつけるのに使う変数の列を指定します。

ページ列 名義尺度または順序尺度の列のカテゴリごとに個別の表を作成します。「[ページ列の使用例](#)」（276ページ）を参照してください。

グループ変数の欠測値を含める グループ変数にある欠測値を1つのグループとして集計した結果を、表に追加します。これをチェックしない場合、欠測値はテーブルに含まれません。列プロパティとして定義した欠測値のコードが考慮されます。

グループ変数の度数順 度数の大きい順に並べ直してグループを表示します。

集計統計量の追加 すべての行と列の集計統計量を追加します。

デフォルトの統計量 分析列または非分析列（グループ変数など）をドラッグ&ドロップしたときにデフォルトで表示される統計量を変更できます。

形式の変更 特定の統計量の表示形式を変更できます。「[数値の表示形式の変更](#)」（264ページ）を参照してください。

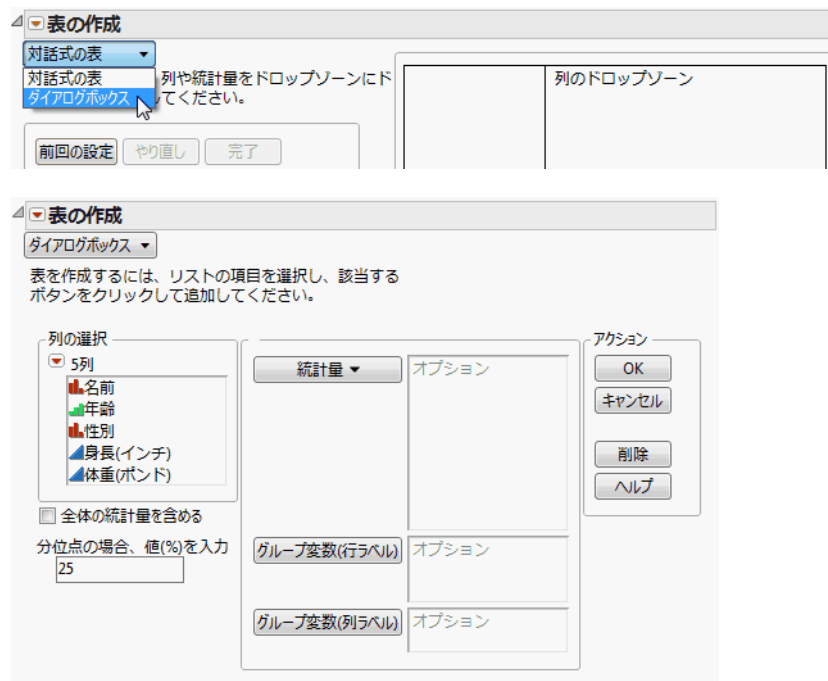
プロットスケールの変更 （赤い三角ボタンのメニューで「[チャートの表示](#)」が選択されている場合にのみ表示）統一したカスタムスケールを指定できます。

プロットスケールの統一 （赤い三角ボタンのメニューで「[チャートの表示](#)」が選択されている場合にのみ表示）このチェックボックスをオフにすると、各列のデータに基づき、それぞれの棒グラフのスケールが決められます。

ダイアログボックスの使用

ドラッグ&ドロップを使って対話式の表を作成する代わりに、ダイアログボックスを使って簡単な表を作成することもできます。この場合、[分析] > [表の作成] の選択後に、メニューから「[ダイアログボックス](#)」を選択してください（図9.8を参照）。表の作成後、必要であれば赤い三角ボタンのメニューから「[設定パネルの表示](#)」を選択し、新しい項目を表にドラッグ&ドロップして、より複雑な表にすることもできます。

図9.8 ダイアログボックスの使用



ダイアログボックスには、次のオプションがあります。

全体の統計量を含める グループ列のカテゴリごとに集計した要約統計量を表示します。

分位点の場合、値(%)を入力 指定した分位点を求めます。たとえば、75%の分位点を求めた場合、データの75%がその分位点より小さくなっています。すべてのグループ列に適用されます。

統計量 列を選択したら、その列に関して求めたい統計量をこのリストから選択します。「[統計量の追加](#)」(262ページ)を参照してください。

グループ変数(行ラベル) 行ラベルとして使用する列を選択します。

グループ変数(列ラベル) 列ラベルとして使用する列を選択します。

統計量の追加

ヒント： 列と統計量の両方を同時に選択して、表にドラッグできます。

表には標準的な統計量を含めることができます。使用できる統計量は、設定パネルに表示されます。列の操作と同じく、任意の統計量をリストから表にドラッグできます。次の点を念頭に置いてください。

- グループ列によって定義されるカテゴリに属する全オブザベーションから、分析列の値に基づいて統計量が計算されます。
- 統計量も、行テーブルと列テーブルのどちらにでも指定できますが、どちらかに統一しておく必要があります。
- 連続尺度の列をデータ領域にドラッグすると、分析列として扱われます。

「表の作成」では、次の統計量を使用できます。

N その列にある欠測値でないデータの数です。分析列がない場合のデフォルトの統計量です。

平均 列の値の算術平均です。つまり、欠測値以外の値の合計を **【重みの合計】** で割った値です（定義されている場合は、**【重み】** 変数を掛けます）。

標準偏差 欠測値以外の値について計算される標本の標準偏差です。つまり、分散の平方根です。

最小値 欠測値を除く、列内の最小値です。

最大値 欠測値を除く、列内の最大値です。

範囲 **【最大値】** と **【最小値】** の差です。

全体に対する % データ全体の合計に対する割合を計算します。分母はすべてのオブザベーションの合計で、分子はカテゴリごとの合計です。分析列がない場合、**【全体に対する %】** は度数の合計に対する割合になります。分析列がある場合、**【全体に対する %】** は分析列の全合計に対する割合になります。つまり、分母はオブザベーションすべての分析列の合計で、分子はカテゴリに属するオブザベーションの分析列の合計になります。テーブルに統計量のキーワードをドラッグして、異なる割合を求めることができます。

- **【全体に対する %】** の見出しに、表の中から1つまたは複数のグループ列をドロップすると、分母の定義が変わります。このとき、各グループごとの合計が分母に使われます。
- 行テーブルと列テーブルにグループ列を指定して2元表を作成したとき、列合計に対する割合を取得したい場合は、行テーブルのすべてのグループ列をドラッグして、**【全体に対する %】** の見出しにドロップします（**【列 %】** と同様）。同様に、行合計に対する割合を取得したい場合は、列テーブルのすべてのグループ列をドラッグして、**【全体に対する %】** の見出しにドロップします（**【行 %】** と同様）。

欠測値 N 欠測値の数です。

カテゴリ数 カテゴリの数です。

合計 列内のすべての値の合計です。表に他の統計量が表示されていない場合の、分析列のデフォルトの統計量です。

重みの合計 列内のすべての **【重み】** の合計です。**【重み】** に割り当てられた列がない場合、**【重みの合計】** は欠測値以外の値の個数になります。

分散 欠測値以外の値について計算される分散（不偏分散）です。つまり、平均からの偏差の2乗の合計を、欠測値以外の値の個数から1を引いた数で割った値です。

標準誤差 平均の標準誤差です。つまり、標準偏差を **N** の平方根で割った値です。列に「重み」役割が割り当てられている場合、分母は重みの和の平方根です。

変動係数 ばらつきの度合いを示す指標で、標準偏差を平均で割り、100を掛けて求められます。

中央値（メディアン） 50%のパーセント点です。下半分のデータが中央値より小さく、上半分のデータが中央値より大きくなっています。

幾何平均 n 個の数値の積の n 乗根です。ゼロまたは負の数値は欠測値として扱われます。たとえば、異なる範囲を持つ指標に基づいて2つの会社を比較したり、データに特に大きな値が含まれていて分布が歪んでいる場合などに、便利です。

四分位範囲 第3四分位点と第1四分位点の差です。

分位点 指定した分位点を求めます。たとえば、75%の分位点を求めた場合、データの75%がその分位点より小さくなっています。異なる分位点を指定するには、**[分位点]** をクリックして表にドラッグし、表示されるボックスに分位点を入力します。

列 % 分析列がない場合は、セルの度数が列の合計度数に占める割合を示します。分析列がある場合、**[列 %]** は分析列の値の列合計に対する割合になります。統計量を上部に配置した表の場合、複数の行テーブルがある表（縦に積み重ねて表示されているもの）に **[列 %]** を追加できます。

行 % 分析列がない場合は、セルの度数が行の合計度数に占める割合を示します。分析列がある場合、**[行 %]** は分析列の値の行合計に対する割合になります。統計量を左側に配置した表の場合、複数の列テーブルがある表（横に並べて表示されているもの）に **[行 %]** を追加できます。

すべて グループ列のカテゴリごとに集計した要約統計量を表示します。

数値の表示形式の変更

各セルの形式は、分析列と統計量によって決まります。度数の場合、デフォルトでは小数点以下は表示されません。統計量を含むセルの形式は、分析列および指定した統計量の形式に基づき、適切な形式に設定されます。デフォルトの形式を変更するには、次の手順を行います。

1. 「表の作成」ウィンドウの下部にある **[形式の変更]** ボタンをクリックします。
2. 表示されるパネルで、フィールドの幅、カンマ、小数点以下の桁数を入力します。図 9.9 を参照してください。
3. セル値をパーセント形式で表示するには、小数点以下の桁数の後にカンマを追加し、「**Percent**」と入力します。
4. （オプション）最適な形式を自動的に選択するには、テキストボックスに「**最適**」と入力します。
すると、各セル値の精度を考慮した上で、その値を表示するのに最適な表示形式が選ばれます。
5. 変更内容を確定し、「形式」セクションを閉じるには **[OK]** をクリックします。または、**[形式の設定]** をクリックすると、「形式」セクションが開いたまま、変更内容が反映されます。

図9.9 数値の表示形式の変更

形式

☐ 同じ小数形式を適用

特定の統計量の表示形式を変更できます。各形式は、フィールド幅と小数桁数の2つの整数で構成されます。「最適」形式を適用するには、2番目の整数の代わりに'Best'というキーワードを入力します。「パーセント」形式を指定するには、2番目の整数の後に'Percent'というキーワードを入力します。

N
9, 最適

形式の設定
キャンセル
OK

「表の作成」の出力

「表の作成」の出力ウィンドウには、行テーブルや列テーブルが上下左右に連結された表を表示できます。列テーブルや行テーブルが1つのみの場合も考えられます。

図9.10 「表の作成」の出力

	性別											
	F						M					
	年齢						年齢					
	12	13	14	15	16	17	12	13	14	15	16	17
	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)
最小値	64	67	81	92	112	116	79	79	92	104	128	134
平均	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
最大値	145	112	142	112	115	116	128	105	119	128	128	172

		性別							
		女性				男性			
		既婚/未婚		既婚/未婚		既婚/未婚		既婚/未婚	
		既婚		未婚		既婚		未婚	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	サイズ	年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
米国	大型	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	中型	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	小型	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
ヨーロッパ	大型	34.0	7.071	28.0	.	.	.	26.0	.
	中型	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	小型	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
日本	大型	25.0	.	.	.	32.0	.	.	.
	中型	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	小型	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
生産国									
米国		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
ヨーロッパ		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
日本		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

タイプ		平均		
		会社規模	利益(\$M)	売上(\$M)
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
Pharmaceutical	big	240.87	5652.02	6159.015
	medium	894.42	7474.04	17140.70
	small	698.98	4261.06	24035.11
すべて	big	156.95	1083.75	38337.19
	medium	690.08	5070.25	23546.12
	small	409.32	5433.86	12679.18
タイプ				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
すべて		409.32	5433.86	12679.18

表を対話的に作成する際は、次のような作業を繰り返します。

- 項目（列または統計量）を適切なリスト内でクリックし、（行または列の）ドロップゾーンまでドラッグします。「[表の編集](#)」（268ページ）および「[列テーブルと行テーブル](#)」（267ページ）を参照してください。
- ドラッグ&ドロップ操作を繰り返してテーブルに適宜項目を追加します。表は、項目を追加するたびに更新されます。すでに列見出しや行ラベルがある場合は、既存の項目を基準にして別の項目の追加場所を決めます。

クリック操作やドラッグ操作の際には、次の点に注意してください。

- JMP では、尺度を基に列の役割を判断します。連続尺度の列は分析列と想定されます。「[分析列](#)」（266ページ）を参照してください。順序尺度または名義尺度の列はグループ列と想定されます。「[グループ列](#)」（266ページ）を参照してください。
- 複数の列を表にドラッグ&ドロップした場合は、次のように処理されます。
 - 複数の列が共通の値のセットを持っている場合は、1つのテーブルに結合され、列名とそれらの列が持つカテゴリから成る集計表が生成されます。この時、各セルは、指定された各列と、それらの列の各値から構成されます。
 - 複数の列が共通の値のセットをもっていない場合は、個別のテーブルに分かれます。
 - このデフォルトの結果は、列を右クリックして「[テーブルの結合](#)」または「[テーブルを分離](#)」を選択すれば変更できます。詳細は、「[列の右クリックメニュー](#)」（270ページ）を参照してください。
- 列を入れ子状にする場合は、表に第1列を作成してから、追加の列をその第1列にドラッグします。
- 適切に作成された表では、グループ列、分析列、統計量がそれぞれまとまって配置されます。分析列のリストの中に統計量が配置されたり、グループ列のリストの中に分析列が挿入されたりすることはありません。
- 「表の作成」上の設定パネルだけではなく、データテーブルのテーブルパネルからも、「表の作成」テーブルに列をドラッグすることができます。

メモ: 開いているデータテーブルで、データを追加したり、行を削除したり、データを再コード化したりした場合は、「表の作成」テーブルも更新されます。

分析列

分析列は、統計量の計算に使用する任意の数値列であり、連続尺度です。グループ列によって区別されるカテゴリごとに、分析列の統計量が計算されます。

分析列は、行テーブルと列テーブルのどちらにでも指定できますが、どちらかに統一しておく必要があります。

グループ列

データをカテゴリ別に分類するために使う列を、グループ列と呼びます。文字、整数値、小数値などを使用できますが、カテゴリ数は制限する必要があります。グループ列は名義尺度または順序尺度です。

次の点を念頭に置いてください。

- グループ列が入れ子状になっている場合は、列の値の入れ子階層に基づいてカテゴリが作成されます。たとえば、「性別」と「既婚／未婚」の2つのグループ列がある場合、「女性で既婚」、「女性で未婚」、「男性で既婚」、「男性で未婚」の4つのカテゴリを持つ表が作成されます。
- グループ列は、列テーブルと行テーブルのどちらでも使用できます。これらのグループ列を組み合わせると、表のセルのカテゴリが決まります。
- グループ列の値に欠測値がある場合、そのオブザベーションはデフォルトでは表に含まれません。[グループ変数の欠測値を含める] オプションをクリックすると、それらのオブザベーションを含めることができます。
- 欠測値として扱うコードまたは値を指定するには、「欠測値のコード」列プロパティを使用します。[グループ変数の欠測値を含める] オプションをクリックすると、それらのオブザベーションを含めることができます。欠測値のコードの詳細については、『JMPの使用法』の「列情報ウィンドウ」章を参照してください。

列テーブルと行テーブル

表は、列見出しと行ラベルを基準として作成されます。これらのサブテーブルは、**行テーブル**および**列テーブル**と呼ばれます。図9.11を参照してください。

行テーブルと列テーブルの例

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
3. 「サイズ」をドラッグし、「行のドロップゾーン」にドロップします。
4. 「生産国」をドラッグし、「サイズ」見出しの左側にドロップします。
5. [平均] をドラッグし、「N」見出しの上にドロップします。
6. [標準偏差] をドラッグし、「平均」見出しの下にドロップします。
7. 「年齢」をドラッグし、「平均」見出しの上にドロップします。
8. 「タイプ」をドラッグし、表の右端にドロップします。
9. 「性別」をドラッグし、表の下にドロップします。

図9.11 行テーブルと列テーブル

2つの列テーブル

生産国	サイズ	年齢		タイプ		
		平均	標準偏差	ファミリー	スポーツ	ワーク
米国	大型	33.8	6.167	28	1	7
	中型	31.1	6.976	35	14	4
	小型	30.4	5.846	11	8	7
ヨーロッパ	大型	30.5	5.802	1	0	3
	中型	30.9	6.194	9	8	0
	小型	30.8	5.388	5	13	1
日本	大型	28.5	4.95	1	0	1
	中型	30.2	4.668	32	15	7
	小型	29.7	5.928	33	41	18
性別						
女性		30.6	6.386	76	41	21
男性		30.8	5.643	79	59	27

2つの行テーブル

複数の列テーブルがある場合は、左側のラベルを複数の列テーブルが共有します。この例では、「生産国」と「性別」が共有されています。同様に、複数の行テーブルがある場合は、上部の見出しを複数の行テーブルが共有します。この例では、「年齢」と「タイプ」の両方がテーブル間で共有されています。

表の編集

表に追加した項目を編集するにはいくつかの方法があります。

項目の削除

表に項目を追加した後で、次のいずれかの方法で項目を削除できます。

- 項目をドラッグし、表の外側にドロップします。
- 最後の項目を削除するには、[元に戻す] をクリックします。
- 項目を右クリックし、[削除] を選びます。

列ラベルの削除

グループ列の各カテゴリの上部には列名が表示されますが、列名が必要ない場合もあります。列テーブルから列名を削除するには、列名を右クリックし、[列ラベルの削除] を選びます。列ラベルを再度挿入するには、カテゴリのいずれかを右クリックし、[列ラベルの復元] を選びます。

統計量とラベルの編集

統計量の名前やラベルを編集できます。たとえば、「平均」の代わりに「平均値」とすることができます。編集する項目を右クリックし、[項目ラベルの変更] を選びます。表示されるボックスに、新しいラベルを入力します。または、名前やラベルをダブルクリックして編集ボックスへ直接入力することもできます。

統計量の名前を、別の統計量の名前に変更した場合、実際に計算される統計量も別のものになるので注意してください。別の統計量の名前にすることは、統計量をいったん表から削除し、その別の統計量を追加するのと同じです。

「表の作成」プラットフォームのオプション

「表の作成」の横にある赤い三角ボタンをクリックすると、メニューに次のオプションが表示されます。

表の表示 要約データを表形式で表示します。

チャートの表示 要約統計量の表のデータを棒グラフで表示します。簡単な棒グラフで、要約統計量の相対的な大きさを視覚的に比較できます。デフォルトでは、すべての列の棒グラフに同じスケールが適用されます。**【プロットスケールの統一】** チェックボックスをオフにすると、各列のデータに基づき、それぞれの棒グラフのスケールが決まります。**【プロットスケールの変更】** ボタンを使用して、統一したカスタムスケールを指定できます。チャートの0の線は、左端か中央に位置します。データがすべて負か正のどちらかに統一されている場合は、0を基準線としてチャートが作成され、それ以外の場合は、チャートの中央が0になります。

設定パネルの表示 インタラクティブに操作するための設定パネルを表示します。

濃淡の表示 表内に複数の行がある場合に、グレーの濃淡を適用します。

ツールヒントの表示 表上でマウスを動かすと、関連項目のヒントが表示されます。

テスト集計パネルの表示 元のテーブルから無作為に抽出した値を使ってテスト集計を行うための、テスト集計パネルを表示します。膨大な量のデータがある場合に特に便利です。**「テスト集計パネルの表示」** (270ページ) を参照してください。

データテーブルに出力 表からデータテーブルを作成します。行テーブルが複数ある場合は、それぞれのラベルの構成が異なるため、行テーブルごとにデータテーブルが作成されます。

フルパスの列名 作成されるデータテーブルの列名に、グループ列のフルパス（列名=水準名の形式）を使用します。

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 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。

「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』の「はじめに」章を参照してください。

テスト集計パネルの表示

大量のデータを含むデータテーブルを使用する場合、そのデータテーブルのサブセットを使ってさまざまな表のレイアウトを試し、最適なレイアウトを見つけることができます。そのような場合、JMPでは、標本サイズを指定して無作為抽出を行った後、表を作成します。テスト集計機能を使うには、次の手順を行います。

1. 「表の作成」の横にある赤い三角ボタンをクリックし、[テスト集計パネルの表示] を選びます。
2. 図9.12に示すように、「**標本サイズ(>1)または標本抽出率(<1)**」の下ボックスに、標本のサイズを入力します。標本のサイズは、アクティブなテーブルに対する割合、または新しいテーブルの行数のどちらかとして認識されます。

図9.12 テスト集計パネル

3. [リサンプリング] をクリックします。
4. JMP データテーブルでサンプリングされたデータを確認するには、[テストデータの表示] ボタンをクリックします。テスト集計パネルを非表示にすると、データテーブルのすべてのデータを使って表が再生成されます。

列の右クリックメニュー

「表の作成」で列を右クリックすると、次のオプションが表示されます。

削除 選択した列を削除します。

グループ変数として使用 分析列をグループ列に変更します。

分析列として使用 グループ列を分析列に変更します。

項目ラベルの変更 (個別の列または入れ子状の列の場合にのみ表示) 新しいラベルを入力できます。

テーブルの結合 (カテゴリごとの列) (個別の列または入れ子状の列の場合にのみ表示) 個々の列または入れ子状の列を結合します。「[複数の列を持つテーブルの例](#)」(274 ページ) を参照してください。

テーブルを分離 (結合したテーブルの場合にのみ表示) 列ごとに個別のテーブルを作成します。

グループ変数を入れ子にする グループ変数を縦または横に入れ子にします。

「表の作成」プラットフォームのその他の例

この例は、次の手順で進めます。

1. 「[度数表の作成](#)」(271 ページ)
2. 「[統計量の表の作成](#)」(272 ページ)
3. 「[表の配置の変更](#)」(273 ページ)

度数表の作成

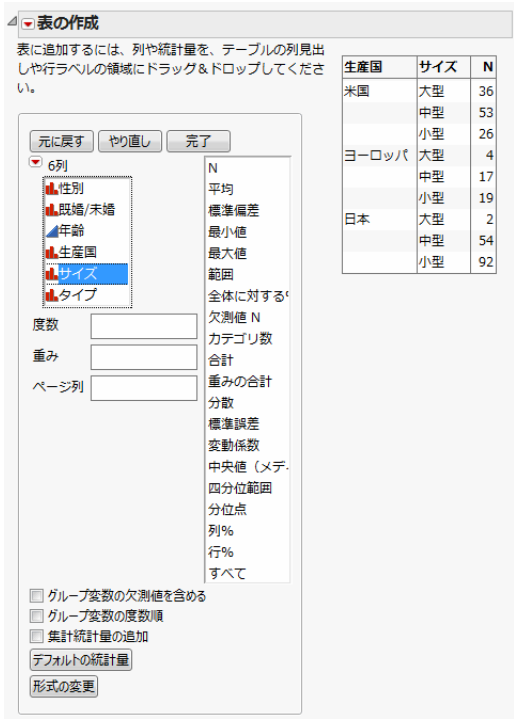
日本車、ヨーロッパ車、米国車の所有者数を車のサイズ別に表示する表を作成してみましょう。図9.3 のような表を作成します。

図9.13 車の所有者数を示す表

生産国	サイズ	N
米国	大型	36
	中型	53
	小型	26
ヨーロッパ	大型	4
	中型	17
	小型	19
日本	大型	2
	中型	54
	小型	92

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
3. 「生産国」をクリックし、「行のドロップゾーン」にドラッグします。
4. 「サイズ」をクリックしてドラッグし、「生産国」見出しの右側にドロップします。

図9.14 「生産国」と「サイズ」を表に追加



統計量の表の作成

次に、各サイズの車の所有者の平均年齢と標準偏差を表示してみましょう。図9.15のような表を作成します。

図9.15 年齢と標準偏差を年齢別に示した表

生産国	サイズ			
米国	大型	年齢	平均	33.8
			標準偏差	6.2
	中型	年齢	平均	31.1
			標準偏差	7.0
ヨーロッパ	小型	年齢	平均	30.4
			標準偏差	5.8
	大型	年齢	平均	30.5
			標準偏差	5.8
日本	中型	年齢	平均	30.9
			標準偏差	6.2
	小型	年齢	平均	30.8
			標準偏差	5.4
	大型	年齢	平均	28.5
			標準偏差	4.9
	中型	年齢	平均	30.2
			標準偏差	4.7
	小型	年齢	平均	29.7
			標準偏差	5.9

1. 図9.14の状態から開始します。「年齢」をクリックしてドラッグし、「サイズ」見出しの右側にドロップします。
2. [平均] をクリックしてドラッグし、「合計」の上にドロップします。
3. [標準偏差] をクリックしてドラッグし、「平均」の下側にドロップします。

表の「平均」の下に「標準偏差」が表示されます。[標準偏差] を「平均」の上にドロップすると、「平均」の上に「標準偏差」が表示されます。

図9.16 「年齢」、[平均]、[標準偏差] を表に追加

表の作成

表に追加するには、列や統計量を、テーブルの列見出しや行ラベルの領域にドラッグ&ドロップしてください。

元に戻す やり直し 完了

6列

性別

既婚/未婚

年齢

生産国

サイズ

タイプ

度数

重み

ページ列

☐ グループ変数の欠測値を含める

☐ グループ変数の度数順

☐ 集計統計量の追加

デフォルトの統計量

形式の変更

N

平均

標準偏差

最小値

最大値

範囲

全体に対する欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値 (×)

四分位範囲

分位点

列%

行%

生産国	サイズ			
米国	大型	年齢	平均	33.8
			標準偏差	6.167
	中型	年齢	平均	31.1
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
日本	大型	年齢	平均	28.5
			標準偏差	4.95
	中型	年齢	平均	30.2

表の配置の変更

「サイズ」を表の上部に移動し、クロス表形式にします。図9.17のような表を作成します。

図9.17 「サイズ」を上部に移動

	サイズ					
	大型		中型		小型	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	年齢	年齢	年齢	年齢	年齢	年齢
米国	33.8	6.167	31.1	6.976	30.4	5.846
ヨーロッパ	30.5	5.802	30.9	6.194	30.8	5.388
日本	28.5	4.95	30.2	4.668	29.7	5.928

表の配置を変更するには、次の手順を行います。

- 1. 図9.16の状態から開始します。「サイズ」見出しをクリックしてドラッグし、表見出しの一番右側にドロップします。図9.18を参照してください。

図9.18 「サイズ」の移動

生産国	サイズ			
米国	大型	年齢	平均	33.8
			標準偏差	6.167
	中型	年齢	平均	31.1
			標準偏差	6.976
	小型	年齢	平均	30.4
			標準偏差	5.846
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
			標準偏差	6.194
	小型	年齢	平均	30.8
			標準偏差	5.388
日本	大型	年齢	平均	28.5
			標準偏差	4.95
	中型	年齢	平均	30.2
			標準偏差	4.668
	小型	年齢	平均	29.7
			標準偏差	5.928

→

			サイズ		
生産国			大型	中型	小型
米国	年齢	平均	33.8	31.1	30.4
		標準偏差	6.167	6.976	5.846
ヨーロッパ	年齢	平均	30.5	30.9	30.8
		標準偏差	5.802	6.194	5.388
日本	年齢	平均	28.5	30.2	29.7
		標準偏差	4.95	4.668	5.928

- 2. 「年齢」をクリックしてドラッグし、「大型」、「中型」、「小型」見出しの下にドロップします。
 - 3. [平均] と [標準偏差] の両方を選択してドラッグし、「大型」見出しの下にドロップします。
- これで、見やすい表になりました。車のサイズと生産国別に、所有者の平均年齢と標準偏差がすぐにわかります。

複数の列を持つテーブルの例

人気者になるのに重要だと思う要素（成績、スポーツ、要望、お金持ち）を、生徒が評価した結果のデータがあります。これらの要素をすべて1つのテーブルにまとめ、統計量や他の因子もあわせて表示してみましょう。図9.19のような表を作成します。

図9.19 人口統計データの追加

		全体に対する%														
		都会/田舎														
		Rural					Suburban					Urban				
性別		1	2	3	4	すべて	1	2	3	4	すべて	1	2	3	4	すべて
boy	成績	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	スポーツ	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	容貌	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
girl	お金持ち	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
	成績	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	スポーツ	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	容貌	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	お金持ち	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Children's Popularity.jp」を開きます。
2. [分析] > [表の作成] を選択します。
3. 「成績」、「スポーツ」、「容貌」、「お金持ち」を選択してドラッグし、「行のドロップゾーン」にドロップします。

図9.20 カテゴリの列

表の作成

表に追加するには、列や統計量を、テーブルの列見出しや行ラベルの領域にドラッグ&ドロップしてください。

元に戻す やり直し 完了

11列

- 性別
- 学年
- 年齢
- 人種
- 都会/田舎
- 学校
- ゴール
- 成績
- スポーツ
- 容貌
- お金持ち

度数

重み

ページ列

☐ グループ変数の欠測値を含める

☐ グループ変数の度数順

☐ 集計統計量の追加

デフォルトの統計量

形式の変更

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値 (メデ)

四分位範囲

分位点

列%

行%

すべて

	1	2	3	4
成績	94	123	133	128
スポーツ	165	154	112	47
容貌	185	126	101	66
お金持ち	34	75	132	237

4つの列は同じ水準値を持つため、これらをまとめたテーブルが表示されます。

各要素に対する4段階評価の割合を表にします。

4. 「性別」をドラッグし、左側の空の見出しにドロップします。
5. [全体に対する %] をドラッグし、数字の見出しの上にドロップします。
6. [すべて] をドラッグし、数字の見出し「4」の右側にドロップします。

図9.21 「性別」、「全体に対する %」、「すべて」を表に追加

		全体に対する%				
性別		1	2	3	4	すべて
boy	成績	8.16%	12.76%	14.02%	12.55%	47.49%
	スポーツ	26.57%	12.55%	6.49%	1.88%	47.49%
	容貌	9.21%	15.48%	12.34%	10.46%	47.49%
	お金持ち	3.56%	6.69%	14.64%	22.59%	47.49%
girl	成績	11.51%	12.97%	13.81%	14.23%	52.51%
	スポーツ	7.95%	19.67%	16.95%	7.95%	52.51%
	容貌	29.50%	10.88%	8.79%	3.35%	52.51%
	お金持ち	3.56%	9.00%	12.97%	26.99%	52.51%

人の属性データを追加して、さらに細かく集計します。

7. 「都会/田舎」をドラッグし、「全体に対する %」見出しの下にドロップします。

図9.22 「都会/田舎」を表に追加

		全体に対する%														
		都会/田舎														
性別		Rural					Suburban					Urban				
		1	2	3	4	すべて	1	2	3	4	すべて	1	2	3	4	すべて
boy	成績	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	スポーツ	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	容貌	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
	お金持ち	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
girl	成績	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	スポーツ	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	容貌	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	お金持ち	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

これで、田舎（「Rural」）、郊外（「Suburban」）、都会（「Urban」）の男子（「boys」）にとって、人気者になるために一番重要な要素は「スポーツ」だということがわかります。田舎、郊外、都会の女子（「girls」）は、「容貌」を一番大事な要素だと考えているようです。

ページ列の使用例

男子生徒と女子生徒の身長を測定したデータがあります。平均身長を生徒の年齢別にまとめた表を作成してみましょう。さらにデータを性別で層別し、性別ごとの表を作成します。このとき、層別の列をページ列として追加します。ページ列を指定すると、グループごとにページが作成されます。図9.23のような表を作成します。

図9.23 生徒の性別ごとの平均身長

身長(インチ)		身長(インチ)	
年齢	合計	年齢	合計
12	293	12	172
13	177	13	245
14	313	14	457
15	126	15	326
16	125	16	68
17	62	17	138

女子 男子

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
身長が分析対象の変数であるため、これを表の一番上に表示します。
3. 「身長(インチ)」をクリックし、「列のドロップゾーン」にドラッグします。
年齢別の統計量が必要なため、年齢を左側に表示します。
4. 「年齢」をクリックしてドラッグし、数値「2502」の横にある空のセルにドロップします。
5. 「性別」をクリックしてドラッグし、「ページ列」のフィールドにドロップします。
6. 「ページ列」リストから「F」を選択すると、女子の平均身長だけが表示されます。
7. 「ページ列」リストから「M」を選択すると、男子の平均身長だけが表示されます。[選択行なし] をクリックして、性別を指定せず、全体の値を表示させることもできます。

図9.24 ページ列の使用例

表の作成

表に追加するには、列や統計量を、テーブルの列見出しや行ラベルの領域にドラッグ&ドロップしてください。

元に戻す

やり直し

完了

5列

名前

年齢

性別

身長(インチ)

体重(ポンド)

度数

重み

ページ列

M

☐ グループ変数の欠測値を含める
☐ グループ変数の度数順
☐ 集計統計量の追加

デフォルトの統計量

形式の変更

N
平均
標準偏差
最小値
最大値
範囲
全体に対する欠測値 N
カテゴリ数
合計
重みの合計
分散
標準誤差
変動係数
中央値 (×)
四分位範囲
分位点
列%
行%

性別-M

身長(インチ)	
年齢	合計
12	172
13	245
14	457
15	326
16	68
17	138

第10章

JMP PRO シミュレーション

パラメトリックな標本再抽出によるシミュレーション

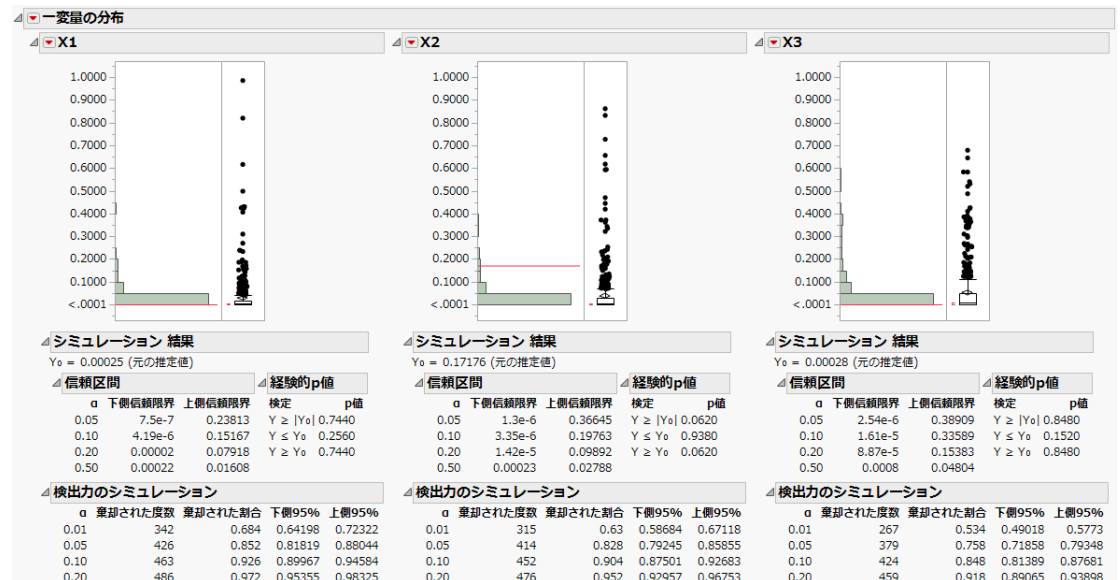
「シミュレーション」プラットフォームは、JMP Pro でのみ使用可能です。

この機能を用いると、パラメトリックおよびノンパラメトリックな乱数シミュレーションを柔軟に行えます。シミュレーションは、次のような場合に使用します。

- パラメトリックなブートストラップを行う。
- 複雑なモデルの検出力計算を行う。
- 複雑なモデルにおいて、予測値や信頼区間などの統計量の分布を近似する。
- 並び替え検定を行う。
- 説明変数に対する仮定が、モデルにおよぼす効果を調べる。
- 様々な仮定に基づいて、モデルを分析する。
- 新しい統計手法や、既存の統計手法を評価する。

このシミュレーション機能は、ブートストラップを行うときと同じ操作で実行できます。レポートの表において右クリックしてください。

図10.1 検出力のシミュレーション



JMP[®] PRO シミュレーションの概要

この章で説明する機能は、レポートに表示されている統計量をシミュレーションによって生成します。この機能を用いるには、統計量の列を右クリックして、[シミュレーション] を選択し、「シミュレーション」ウィンドウで、乱数データに置き換えたい列を指定します。この列を、「切り替え元の列」と呼びます。この列は、どんなものでもよく、応答変数であっても説明変数であっても構いません。その後、生成する乱数を示す計算式を含む列を指定します。この列を、「切り替え先の列」と呼びます。「切り替え元の列」に対して、「切り替え先の列」の乱数が生成されます。

メモ: このシミュレーションの機能を用いるには、データテーブルに、乱数を生成する計算式を含んだ列を準備しておく必要があります。

処理の流れは次のとおりです。まず、「切り替え先の列」の計算式に基づいて、乱数データが生成されます。次に、そこで生成された「切り替え先の列」の乱数を、「切り替え元の列」のデータと置き換えて、元のレポート全体が再計算されます。この処理を、 N 回繰り返します。ここで、 N はシミュレーション回数 (標本数) です。

シミュレーションが終了すると、次のようなデータテーブルが出力されます。

- データテーブルの各行は、各シミュレーションの結果です。
- 各列は、各統計量に対応しています。
- シミュレーション結果を見るためのスクリプトも作成されます。

ヒント: シミュレーションを実行すると、元のレポートに表示されているすべての分析が、再実行されます。そのため、特定の列を選択してシミュレーションを行っても、すべての分析が再実行されるために、時間がかかる場合があります。処理に時間がかかる場合は、プラットフォームのレポートから、無関係なオプションを除いてください。

JMP[®] PRO シミュレーションの例

この節では、次のような使用例を紹介します。なお、最後に挙げた例は、他のマニュアルで紹介されています。

- 「分散成分に対する信頼区間の計算」(281 ページ)
- 「並び替え検定の実行」(286 ページ)
- 「一般化回帰モデルに含める説明変数の検討」(288 ページ)
- 「非線形モデルにおける検出力の事前計算」(293 ページ)
- 正規分布に従わない変数について、PpK や不適合率の信頼区間をシミュレートする方法については、『品質と工程』の「工程能力」章を参照してください。

JMP PRO 分散成分に対する信頼区間の計算

この例では、温度、時間、触媒の量が、反応におよぼす効果を調べたいとします。温度は、変更が非常に困難な変数（一次単位因子）です。時間は、変更が困難な変数（二次単位因子）です。触媒の量は、変更が容易な変数です。一次単位因子と二次単位因子については、『実験計画(DOE)』の「カスタム計画」章を参照してください。

ここで乱数シミュレーションを行うのは、一次単位と二次単位の分散成分に対する信頼区間を求めるためです。過去の研究から、一次単位と二次単位の標準偏差は、それぞれ、誤差の標準偏差の約2倍と約1.5倍であることが分かっています。JMPがデフォルトで求めている信頼区間は、Wald法による近似によるもので、分散成分の推定量が漸近的に正規分布に従うものとして算出されています。このWald法による近似は精度が悪く、信頼区間の幅は狭くなる傾向があります。ここでは、分散成分の標本分布をシミュレートで求めて、信頼区間を計算してみます。

この例では、次の手順に従います。

- カスタム計画を使用して、2段分割計画を作成します。「計画の作成」(281ページ)を参照してください。
- REML法を使ってモデルをあてはめます。「モデルのあてはめ」(283ページ)を参照してください。
- 分散成分の推定値をシミュレートして、分散成分の信頼区間を求めます。「検出力の考察」(284ページ)を参照してください。

JMP PRO 計画の作成

この節の手順を省略するには、[ヘルプ] > [サンプルデータライブラリ] を選択し、「Design Experiment」フォルダの「Catalyst Design.jmp」を開きます。そして、「Catalyst Design.jmp」データテーブルの「DOEシミュレート」スクリプトの横にある緑の三角ボタンをクリックします。この操作が済んだら、「モデルのあてはめ」(283ページ)に進んでください。

1. [実験計画(DOE)] > [カスタム計画] を選択します。
2. 「因子」アウトラインで、「N個の因子を追加」の横のボックスに「3」と入力します。
3. [因子の追加] > [連続変数] を選択します。
4. 因子の名前をダブルクリックして、「温度」、「時間」、「触媒」に変更します。
値は、デフォルトの「-1」と「1」をそのまま使用します。
5. 「温度」の[容易]をクリックして、[非常に困難]に変更します。
これで、「温度」が一次単位因子になります。
6. 「時間」の[容易]をクリックして、[困難]に変更します。
これで、「時間」が二次単位因子になります。
7. [続行] をクリックします。
8. 「モデル」パネルで[交互作用] > [2次] を選択します。
すると、モデルにすべての2次交互作用が追加されます。
9. 「カスタム計画」の赤い三角ボタンをクリックし、メニューから[応答のシミュレート]を選択します。

この操作を行っておくと、計画テーブルを作成するために「テーブルの作成」をクリックしたときに、「応答をシミュレート」ウィンドウが開きます。

メモ: 乱数シード値 (手順 10) と開始点の数 (手順 11) を設定すると、以下の数値例と同じ結果が得られます。同じ結果でなくても良い場合は、これらの手順は不要です。

10. (オプション)「カスタム計画」の赤い三角ボタンをクリックして、「乱数シード値の設定」を選択します。「12345」と入力して [OK] をクリックします。
11. (オプション)「カスタム計画」の赤い三角ボタンをクリックして、「開始点の数」を選択します。「1000」と入力して [OK] をクリックします。
12. [計画の作成] をクリックします。
13. [テーブルの作成] をクリックします。

メモ: 乱数を用いているため、「Y」および「Yのシミュレーション」の値は、図 10.2 とは異なったものになります。

図10.2 計画テーブル

Catalyst Design			一次単位	二次単位	温度	時間	触媒	Y	Yのシミュレーション
計画	カスタム計画								
基準	D-最適計画		1	1	1	1	-1	-0.046768023	1.60918231
モデル			2	1	1	1	1	6.539690219	7.92020524
Yのシミュレーションのモデル			3	1	1	1	1	6.5397108257	6.13783855
計画の評価			4	1	2	1	-1	0.494801276	-0.5997359
DOEシミュレート			5	1	2	1	-1	-2.65841644	-1.1010741
DOEダイアログ			6	1	2	1	-1	-0.200740508	-3.2448733
			7	2	3	-1	1	3.1335699565	4.23469875
			8	2	3	-1	1	2.2739679213	2.3785821
列(7/0)			9	2	3	-1	1	1.1704395547	2.88111529
一次単位 *			10	2	4	-1	-1	-0.693941035	0.25146684
二次単位 *			11	2	4	-1	-1	2.3525055616	-1.7479193
温度 *			12	2	4	-1	-1	-1.545149968	-2.0878163
時間 *			13	3	5	-1	1	-1.4838427	-2.3160787
触媒 *			14	3	5	-1	1	-1.375362439	-1.4243049
Y *			15	3	5	-1	1	0.7434199911	-1.7073105
Yのシミュレーション			16	3	6	-1	-1	-1.11989643	-0.1160015
			17	3	6	-1	-1	3.4346315881	2.15367996
			18	3	6	-1	-1	0.2987179478	1.2809475
行			19	4	7	1	-1	-1.245681444	1.55842569
すべての行	24		20	4	7	1	-1	0.4951842993	0.68125318
選択されている行	0		21	4	7	1	-1	-0.6602484	-1.3744632
除外されている行	0		22	4	8	1	1	0.7779088461	0.97977409
表示しない行	0		23	4	8	1	1	7.9432230978	7.88267366
ラベルのついた行	0		24	4	8	1	1	1.0668195454	-0.9506064

図10.3 「応答をシミュレート」ウィンドウ

計画のデータテーブルと「応答をシミュレート」ウィンドウが表示されます。データテーブルには、「DOE シミュレート」スクリプトが含まれています。このスクリプトを実行すれば、いつでもパラメータ値を変更できます。

次の節では、一次誤差と二次誤差の標準偏差を指定します。その後、乱数シミュレーションによって生成したデータに対して、REMLモデルをあてはめます。

JMP PRO モデルのあてはめ

一次誤差と二次誤差は、正規分布に従うと仮定します。また、誤差の標準偏差を1.0とした場合、一次誤差の標準偏差は2.0、二次誤差の標準偏差は1.5になっていると仮定します。一次単位と二次単位のばらつきにのみ興味がある場合は、「応答をシミュレート」ウィンドウにおける「効果」の数値は特に変更しなくてかまいません。

1. 「分布」パネル（図10.3）の「一次単位 σ 」に「2」と入力します。

デフォルトで正規分布が選択されていることに注意してください。これは、正規分布に従う誤差が計算式に追加されることを示しています。

2. 「二次単位 σ 」に「1.5」と入力します。

3. [適用] をクリックします。

データテーブルの「Yのシミュレーション」の計算式が更新されます。計算式を表示するには、「列」パネルで列名の右側にある+記号をクリックします。

4. データテーブルの「モデル」スクリプトの横の緑の三角ボタンをクリックします。

5. [Y] ボタンの横の変数「Y」をクリックし、[削除] ボタンをクリックします。

6. 「Yのシミュレーション」をクリックし、[Y] ボタンをクリックします。

これにより、「Y」が、シミュレーションの計算式を含む列に置き換えられます。

7. [実行] をクリックします。

シミュレートした応答の値1組に基づき、モデルがあてはめられます。

メモ: 「Yのシミュレーション」は乱数によって生成された値であるため、ここまでの手順に沿って作成したレポートの結果は図 10.4 とは異なります。

図10.4 REML法のレポートとWald法の信頼区間

REML法による分散成分推定値							
変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	Wald p値	全体に対する百分率
一次単位	3.8107856	1.556364	1.8007472	-1.973036	5.0857636	0.3874	68.445
二次単位	0.7568606	0.3091097	0.4661521	-0.604532	1.222751	0.5073	13.594
残差		0.4084103	0.160228	0.2146172	1.060295		17.961
合計		2.2738839	1.8035082	0.7468087	28.115033		100.000
-2対数尤度 = 63.890021519							
注: 「合計」は、分散成分のうち、正のものだけを足した和です。							
負の推定値も含めた合計 = 2.2738839							

JMP PRO 検出力の考察

次に、分散成分をシミュレートして、どれぐらいでばらついているかを確認してみましょう。

1. 「REML 法による分散成分推定値」アウトラインで、「分散成分」列を右クリックして、[シミュレーション] を選択します。

図10.5 「シミュレーション」ウィンドウ

このシミュレーションでは、現モデルの推定に用いた「Yのシミュレーション」列のデータ値が、シミュレーションごとに「Yのシミュレーション」列の計算式で生成される乱数に置き換えられます。右クリックして選択された「分散成分」列がシミュレーションされます。右クリックした「分散成分」列が強調表示され、この列が「パラメータ推定値」テーブルに表示されている各効果に対してシミュレーションされることを示します。

- 「標本数」に「200」と入力します。
- (オプション)「乱数シード値」に「456」と入力します。
この操作を行うと、1行目を除き、図10.6と同じ値が生成されます。
- [OK] をクリックします。
1行目の値は、図10.6とは異なります。

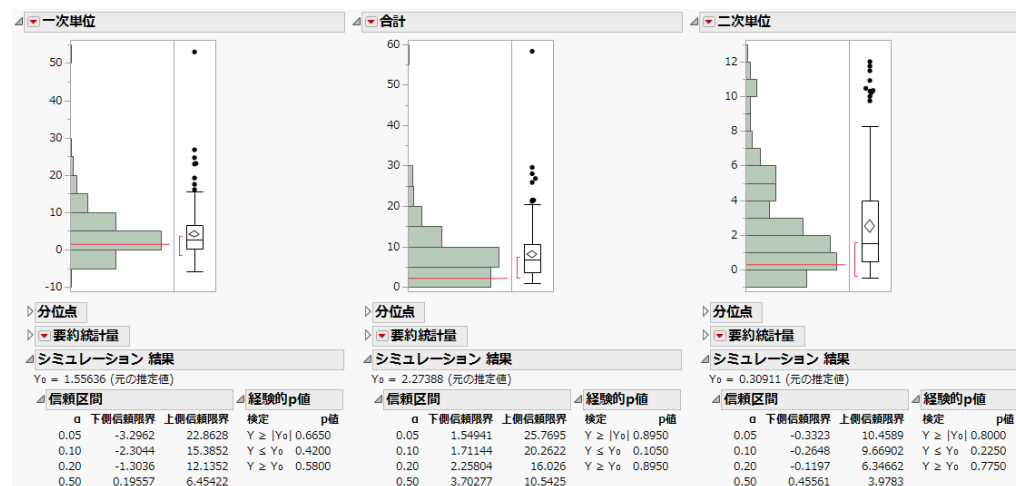
図10.6 分散成分のシミュレーション結果 (一部)

	Y	SimID*	一次単位	合計	残差	二次単位
1	Y Simulated	0	1.5563639995	2.2738839302	0.4084102799	0.3091096508
2	Y Simulated	1	0.1160090855	11.825684937	1.7118437836	9.9978320675
3	Y Simulated	2	3.3652848741	8.6529959081	0.7259348275	4.5617762065
4	Y Simulated	3	22.95631251	26.836360491	0.8416572537	3.0383907272
5	Y Simulated	4	12.156574946	13.143778421	0.9872034755	-0.176636748
6	Y Simulated	5	-0.060658009	1.3514412755	1.3514412755	-0.153702674
7	Y Simulated	6	11.804365807	13.894478258	0.9837455205	1.1063669307
8	Y Simulated	7	2.6824797817	3.7015003812	0.8559296173	0.1630909822
9	Y Simulated	8	26.765100099	28.022776455	1.0422128078	0.2154635485
10	Y Simulated	9	12.75877132	17.850598404	1.0302645404	4.0615625437
11	Y Simulated	10	-0.433818317	3.0738531832	1.2569403028	1.8169128804
12	Y Simulated	11	2.4634884406	11.105096973	1.6719518802	6.9696566519
13	Y Simulated	12	4.2166345958	7.2251377019	1.0326091926	1.9758939136
14	Y Simulated	13	2.8027318323	4.6661579836	0.7050109246	1.1584152266
15	Y Simulated	14	0.3194564512	6.53640214	0.9281404099	5.2888052789
16	Y Simulated	15	-0.270350632	3.1414624009	1.6586718874	1.4827905134
17	Y Simulated	16	-0.192580268	1.7108138759	0.9471174937	0.7636963822
18	Y Simulated	17	0.756815441	2.927052512	0.9782751744	1.1919618966
19	Y Simulated	18	4.8135329401	7.4052160139	1.5524955876	1.0391874862

「最小2乗法によるあてはめ シミュレーション結果 (分散成分)」データテーブルの最初の行は、元データから計算された「分散成分」の推定値であり、除外の状態になっています。2行目以降のデータは、シミュレートによって生成されたものです。

- 「一変量の分布」スクリプトを実行します。

図10.7 分散成分に対して「一変量の分布」を実行した結果 (一部)



各分散成分の95%信頼区間は、2.5%および97.5%の各分位点を使って求められます（図10.7の「シミュレーション結果」レポートを参照してください）。これらの信頼区間を、REMLレポート（図10.4）の信頼区間と比較してみましょう。

- 一次単位の分散成分の95%信頼区間は、シミュレーションでは-3.296～22.863です。REMLレポートのWald法では、-1.973～5.086です。
- 二次単位の分散成分の95%信頼区間は、シミュレーションでは-0.332～10.459です。REMLレポートのWald法では、-0.605～1.223です。

シミュレーションによって求めた信頼区間は、1組の値を使ってREML法で計算した信頼区間に比べ、はるかに広がっています。信頼区間の精度をさらに高めるには、シミュレーションの回数を増やしてみてください。

JMP PRO 並び替え検定の実行

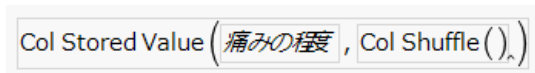
この例では、3種類の鎮痛剤の効果に違いがあるかどうかに興味があります。標本サイズが非常に小さく、通常の分散分析では妥当な結果とならない恐れがあります。そこで、シミュレーションに基づく並び替え検定を行ってみましょう。

まず、痛みの程度の測定値を、3種類の薬に対してランダムにシャッフルする計算式を作成します。「鎮痛剤の効果が無い」という帰無仮説が正しければ、測定値をランダムにシャッフルしたデータから計算されるF値と、実際に得られたデータから計算されるF値にはそれほど違いは生じないはずです。測定値をランダムにシャッフルして得られるF値は、標本サイズが大きければ近似的にF分布に従います。そこで、観測されたF値を、帰無仮説が正しいとした上でのシミュレーションによって得られたF値の分布と比較します。

JMP PRO シミュレーションの計算式の定義

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Analgesics.jmp」を開きます。
2. [列] > [列の新規作成] を選びます。
3. 「列名」に「痛みの程度（シャッフル）」と入力します。
4. 「列プロパティ」リストから [計算式] を選択します。
5. 関数のリストから [行] > [Col Stored Value] を選択します。
6. 「列」のリストから、[痛みの程度] をダブルクリックします。
7. エディタパネルの上部の記号リストから、挿入キー (^) をクリックします。
8. 関数のリストから、[乱数] > [Col Shuffle] を選択します。

図10.8 完成した式



この計算式は、「痛みの程度」列の値をランダムにシャッフルします。

9. 計算式エディタウィンドウで【OK】をクリックします。
10. 列情報ウィンドウで【OK】をクリックします。

並び替え検定の実行

1. 【分析】>【二変量の関係】を選びます。
2. 「痛みの程度」を選択し、【Y, 目的変数】をクリックします。
3. 「薬」を選択し、【X, 説明変数】をクリックします。
4. 【OK】をクリックします。
5. 赤い三角ボタンをクリックし、【平均/ANOVA】を選択します。

図10.9 「分散分析」レポート

分散分析					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
薬	2	99.89459	49.9473	6.2780	0.0053*
誤差	30	238.67877	7.9560		
全体(修正済み)	32	338.57335			

ここで、F値が6.2780であることに注目してください。

6. 「分散分析」アウトラインの「F値」列を右クリックし、【シミュレーション】を選択します。
7. 「切り替え元の列」のリストから、「痛みの程度」をクリックします。
8. 「切り替え先の列」のリストから、「痛みの程度 (シャッフル)」をクリックします。
9. 「標本数」に「1000」と入力します。
10. (オプション)「乱数シード値」に「456」と入力します。

こうすると、この例と同じ値が再現されます。

図10.10 設定後の「シミュレーション」ウィンドウ

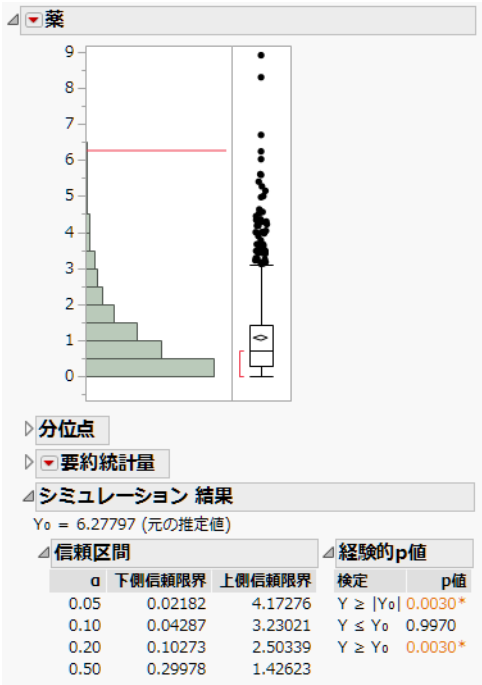
切り替え元の列		切り替え先の列	
痛みの程度		痛みの程度 (シャッフル)	
薬			
標本数	1000		
乱数シード値	456		
		OK キャンセル ヘルプ	

11. 【OK】をクリックします。

シミュレーションの結果を示すテーブルでは、「全体 (修正済み)」と「誤差」は空白になっています。F値だけが、「薬」の列に出力されています。

12. このテーブルで、「一変量の分布」スクリプトを実行します。

図10.11 シミュレーションによって得たF値の帰無分布



観測されたF値の6.2780は、ヒストグラムにおいて赤い線で表示されます。この値は、シミュレーションによって得たF値の帰無分布における上位0.5%点よりも大きくなっています。このことより、今回の実験結果は、「3種類の薬が痛みの程度に与える効果のいずれかは異なる」ことを示す強い証拠であると言えます。

JMP PRO 一般化回帰モデルに含める説明変数の検討

この例では、製薬会社が持つ過去のデータに基づき、錠剤の溶出速度とその速度に影響をおよぼす可能性のある要因について分析します。溶出速度が70に満たない錠剤は、不適合と見なされます。どの要因が溶出速度に影響するのかを調べてみましょう。

この例では、次の手順に従います。

- 一般化回帰モデルを作成する。
- セロでない項を使って、説明変数を減らしたモデル（縮小モデル）を作成する。
- 縮小モデルでシミュレーションを実行し、特定の説明変数をモデルに含まれるべきかどうかを検討する。

JMP PRO モデルのあてはめ

この節では、一般化回帰モデルのあてはめを行います。この節の手順を省略するには、「Tablet Production.jmp」データテーブルの「一般化回帰」スクリプトの横の緑の三角ボタンをクリックし、モデルを作成してください。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Tablet Production.jmp」を開きます。
2. [分析] > [モデルのあてはめ] を選択します。
3. 「溶出」を選択し、[Y] をクリックします。
4. 「粉碎時間」から「噴霧器圧力」までを選択し、[追加] をクリックします。
5. 「手法」のメニューから「一般化回帰」を選択します。
6. [実行] をクリックします。
7. 「モデルの設定」パネルで [実行] をクリックします。

図10.12 適応型Lassoに基づくモデル

元の説明変数に対する推定値						
項	推定値	標準誤差	Waldカイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
切片	110.35025	24.526587	20.242885	<.0001*	62.279021	158.42147
粉碎時間	0.1291212	0.0283186	20.789802	<.0001*	0.0736177	0.1846247
スクリーンサイズ[3-5]	4.1694529	0.5433106	58.89275	<.0001*	3.1045837	5.234322
スクリーンサイズ[4-5]	2.3600819	0.5712373	17.069508	<.0001*	1.2404774	3.4796865
ステアリン酸マグネシウム供給業者[Jones Inc-Smith Ind]	0	0	0	1.0000	0	0
乳糖供給業者[Bond Inc-James Ind]	0	0	0	1.0000	0	0
砂糖供給業者[Sour-Sweet]	0	0	0	1.0000	0	0
タルク供給業者[Rough-Smooth]	0	0	0	1.0000	0	0
攪拌時間	0.7177884	0.1384385	26.883081	<.0001*	0.4464539	0.9891228
攪拌速度	0.1821224	0.2418817	0.5669188	0.4515	-0.291957	0.6562018
圧縮機[Compress1-Compress2]	-0.495065	0.3993136	1.5370784	0.2151	-1.277705	0.2875754
圧力	0	0	0	1.0000	0	0
滑沢剤供給業者[Coat-Mac]	0	0	0	1.0000	0	0
滑沢剤供給業者[Down-Mac]	0	0	0	1.0000	0	0
滑沢剤の粘度	0.181504	0.0503564	12.991611	0.0003*	0.0828073	0.2802007
注入温度	0	0	0	1.0000	0	0
排出温度	0	0	0	1.0000	0	0
噴霧量	-0.203293	0.0421928	23.214889	<.0001*	-0.285989	-0.120596
噴霧器圧力	0	0	0	1.0000	0	0
尺度	2.0288847	0.1659965	149.38839	<.0001*	1.7035375	2.3542319

「適応法 Lasso による推定（検証法: AICc）」レポートのパラメータ推定値を見てみましょう。パラメータ推定値がゼロでないものは、「粉碎時間」、「スクリーンサイズ」、「攪拌時間」、「攪拌速度」、「圧縮機」、「滑沢剤の粘度」、および「噴霧量」で、これらが「溶出」に関係している可能性があります。

JMP PRO モデルの縮小

モデルの縮小に進む前に、「Tablet Production.jmp」データテーブルで、列が選択されていないことを確認します。列が選択されたままだと、後述の手順1でも選択された状態になります。列の選択を解除しておく、誤ってゼロの項の列を含めてしまうことが避けられます。

この節の手順を省略するには、「Tablet Production.jmp」データテーブルの「一般化回帰 縮小モデル」スクリプトの横の緑の三角ボタンをクリックし、縮小モデルを作成してください。

1. 「適応法 Lasso による推定（検証法: AICc）」レポートの赤い三角ボタンをクリックし、[非ゼロの効果を
使って再起動] を選択します。
「モデルのあてはめ」ウィンドウが開き、「モデル効果の構成」リストに、パラメータ推定値がゼロでない
項が表示されます。応答は「Y」にすでに指定されています。「手法」は「一般化回帰」が選択されています。
2. [実行] をクリックします。
3. 「モデルの設定」パネルで [実行] をクリックします。

図10.13 適応型 Lasso による縮小モデル

元の説明変数に対する推定値						
項	推定値	標準誤差	Waldカイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
切片	95.142391	23.430178	16.489099	<.0001*	49.220085	141.0647
粉碎時間	0.1394103	0.0275609	25.585971	<.0001*	0.0853919	0.1934288
スクリーンサイズ[3-5]	4.3323833	0.534237	65.763638	<.0001*	3.285298	5.3794685
スクリーンサイズ[4-5]	2.6331283	0.5457852	23.275584	<.0001*	1.563409	3.7028476
攪拌時間	0.7583048	0.1385246	29.966364	<.0001*	0.4868017	1.029808
攪拌速度	0.4575802	0.2289168	3.9955744	0.0456*	0.0089116	0.9062488
圧縮機[Compress1-Compress2]	-0.877986	0.4127286	4.5252866	0.0334*	-1.686919	-0.069053
滑沢剤の粘度	0.198673	0.0486798	16.656386	<.0001*	0.1032624	0.2940836
噴霧量	-0.212753	0.0410929	26.805238	<.0001*	-0.293294	-0.132213
尺度	1.9982636	0.1574951	160.97992	<.0001*	1.689579	2.3069483

「攪拌速度」の信頼区間（「下側95%」）が、ゼロに非常に近くなっていることに注目してください。次に、シミュレーションによって、現在の推定値が真値であった場合、適応型 Lasso を行うと「攪拌速度」の推定値がどれぐらいの割合でゼロになるかを調べてみましょう。

JMP PRO 攪拌速度の推定値がどれぐらいゼロになるかの考察

以下の手順では、縮小モデル（図10.13）のレポートを使用します。

1. 「適応法 Lasso による推定（検証法: AICc）」レポートの赤い三角ボタンをクリックし、[列の保存] > [シミュレーション計算式の保存] を選択します。
「Tablet Production.jmp」データテーブルに、「溶出 シミュレーション計算式」という新しい列が追加されます。
2. （オプション）データテーブルの「列」パネルで、「溶出 シミュレーション計算式」の右側にある＋記号をクリックします。

図10.14 シミュレーション計算式

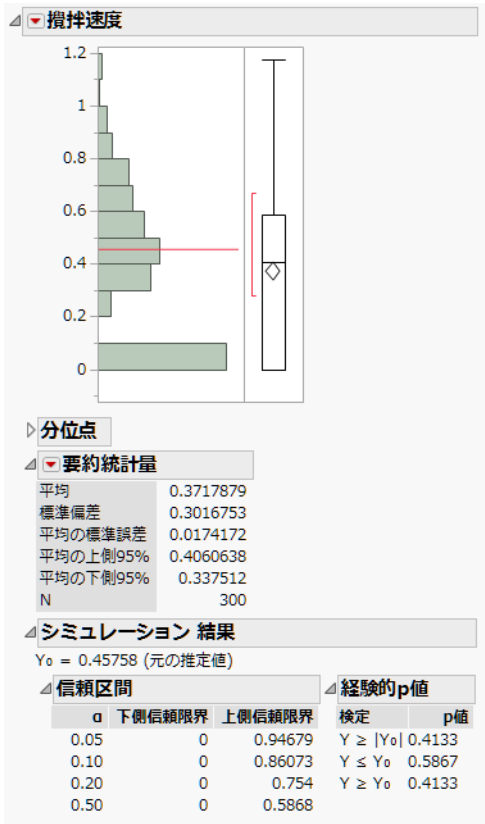
この計算式は、あてはめたモデルと「溶出」の分布に従って、各行の値をシミュレートします。「溶出」の分布は、標準偏差が約1.998の正規分布に従うと推定されています。

3. [キャンセル] をクリックします。
4. 縮小モデルのレポートウィンドウに戻ります。「元の説明変数に対する推定値」レポートで、「推定値」列を右クリックし、[シミュレーション] を選択します。
5. 「標本数」に「300」と入力します。
「溶出」列を「溶出 シミュレーション計算式」列のシミュレーション値に置き換えて、分析を300回実行します。
6. (オプション)「乱数シード値」に「123」と入力します。
こうすると、この例と同じ値が再現されます。

図10.15 設定後の「シミュレーション」ウィンドウ

- 7. [OK] をクリックします。
 テーブルの最初の行は、推定値の初期値であり、除外されます。その他の行は、シミュレートした値です。
- 8. 「一変量の分布」スクリプトを実行します。
- 9. Ctrl キーを押しながら、「切片」の赤い三角ボタンをクリックし、[表示オプション] > [要約統計量のカスタマイズ] を選択します。
- 10. [ゼロの個数] を選択します。
- 11. [OK] をクリックします。
- 12. 「攪拌速度」のレポートにスクロールします。

図10.16 シミュレーションによる攪拌速度の推定値のヒストグラム



- 「要約統計量」レポートによると、シミュレーションのうち $97/300 = 32.3\%$ で、「攪拌速度」の推定値がゼロになっていることがわかります。この情報は、「攪拌速度」をモデルに含めるかどうかを判断するときの材料となります。

JMP PRO 非線形モデルにおける検出力の事前計算

この例では、6つの連続尺度の因子が、部品検査の可否におよぼす主効果を調べたいとしましょう。応答は二項分布に従い、実験は60回行えるものとします。

以下の手順に従います。

1. 実験のカスタム計画を作成します。「[計画の作成](#)」(294ページ)を参照してください。

メモ: カスタム計画は、非線形な状況に最適とはいえませんが、この例では、単純化するために、「非線形計画」プラットフォームではなく「カスタム計画」プラットフォームを使用します。「非線形計画」プラットフォームで作成した計画のほうが直交計画より優れていることを示す例については、『実験計画(DOE)』の「非線形計画」章を参照してください。

2. 一般化線形モデルを使って、ロジスティックモデルをあてはめます。「[一般化線形モデルのあてはめ](#)」(297ページ)を参照してください。
3. 尤度比検定の p 値をシミュレートして、線形予測子によって異なる確率の差の検出力を調べます。「[検出力の考察](#)」(298ページ)を参照してください。

この例での計画

一般化線形モデルで、リンク関数としてロジット関数を使用して、成功確率をモデル化します。リンク関数としてロジット関数を使用すると、次のロジスティックモデルがあてはめられます。

$$\pi(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_6 X_6)}}$$

ここで、 $\pi(\mathbf{X})$ は、 $\mathbf{X} = (X_1, X_2, \dots, X_6)$ という因子設定のもと、部品が合格する確率を表します。

線形予測子 $L(\mathbf{X})$ は、次の式で表されます。

$$L(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \dots + \beta_6 X_6$$

線形予測子の係数が次のような場合だったときの検出力を求めてみましょう。

係数	値
β_0	0
β_1	1
β_2	.9
β_3	.8
β_4	.7

係数	値
β_5	.6
β_6	.5

線形予測子の切片は0なので、全因子が0に設定されている場合、部品が合格する確率は50%に等しくなります。 i 番目の因子における各水準に関連する確率は、その他の因子がすべて0である場合、次のようになります。

因子	$X_i = 1$ の場合の合格率	$X_i = -1$ の場合の合格率	差
X_1	73.11%	26.89%	46.2%
X_2	71.09%	28.91%	42.2%
X_3	69.00%	31.00%	38.0%
X_4	66.82%	33.18%	33.6%
X_5	64.56%	35.43%	29.1%
X_6	62.25%	37.75%	24.5%

たとえば、 X_1 以外の全因子が0に設定されている場、検出したい差は46.2%になります。検出したい差が最小になるのは、 X_6 以外の全因子が0に設定された場合で、その差は24.5%です。

JMP PRO 計画の作成

メモ: この節の手順を省略するには、[ヘルプ] > [サンプルデータライブラリ] を選択し、「Design Experiment」フォルダの「Binomial Experiment.jmp」を開きます。「DOE シミュレート」スクリプトの横にある緑色の三角形をクリックし、「[\[応答をシミュレート\] の設定](#)」(296 ページ) に進んでください。

1. [実験計画 (DOE)] > [カスタム計画] を選択します。
2. 「因子」アウトラインで、「N個の因子を追加」の横のボックスに「6」と入力します。
3. [因子の追加] > [連続変数] を選択します。
4. [続行] をクリックします。

主効果の計画を作成したいため、「モデル」アウトラインには変更を加えません。

5. 「実験の回数」の「ユーザ定義」テキストボックスに「60」と入力します。
6. 「カスタム計画」の赤い三角ボタンをクリックし、メニューから「[\[応答のシミュレート\]](#)」を選択します。

この操作を行っておくと、計画テーブルを作成するために「[\[テーブルの作成\]](#)」をクリックしたときに、「[\[応答をシミュレート\]](#)」ウィンドウが開きます。

メモ: 乱数シード値（手順7）と開始点の数（手順8）を設定すると、以下の数値例と同じ結果が得られます。同じ結果でなくても良い場合は、これらの手順は不要です。

7. (オプション)「カスタム計画」の赤い三角ボタンをクリックして、「乱数シード値の設定」を選択します。「12345」と入力して [OK] をクリックします。
8. (オプション)「カスタム計画」の赤い三角ボタンをクリックして、「開始点の数」を選択します。「1」と入力して [OK] をクリックします。
9. [計画の作成] をクリックします。
10. [テーブルの作成] をクリックします。

メモ：乱数を用いているため、「Y」および「Yのシミュレーション」の値は、図 10.17 とは異なったものになります。

図10.17 計画テーブル (一部)

カスタム計画			X1	X2	X3	X4	X5	X6	Y	Yのシミュレーション
計画	カスタム計画									
基準	D-最速計画		1	-1	-1	1	1	-1	1.2642334906	1.26423349
モデル			2	-1	1	-1	1	1	4.0747268604	4.07472686
計画の評価			3	-1	-1	-1	1	-1	-0.182075879	-0.1820759
シミュレーション			4	1	-1	1	-1	1	0.4472254964	0.4472255
DOEダイアログ			5	-1	1	1	1	-1	2.5401448778	2.54014488
列(8/0)			6	-1	-1	-1	1	1	-0.233821987	-0.233822
X1 *			7	-1	1	1	-1	-1	0.8584646156	0.85846462
X2 *			8	1	-1	-1	1	1	4.0420036424	4.04200364
X3 *			9	1	-1	-1	-1	-1	-2.934267782	-2.9342678
X4 *			10	1	-1	1	1	-1	4.2252590297	4.22525903
X5 *			11	1	1	-1	1	-1	2.8516958617	2.85169586
X6 *			12	-1	1	-1	-1	1	-0.709876725	-0.7098767
Y *			13	1	1	1	1	1	5.8506056173	5.85060562
Yのシミュレーション			14	1	-1	1	-1	-1	0.5174517568	0.51745176
			15	-1	1	-1	1	-1	-1.452974096	-1.4529741
			16	1	1	-1	1	-1	3.2738674886	3.27386749
			17	1	1	-1	1	-1	0.7758670811	0.77586708
すべての行	60		18	1	1	-1	-1	1	1.218553009	1.21855301
選択されている行	0		19	-1	1	1	-1	1	2.5799848241	2.57998482
除外されている行	0		20	-1	-1	-1	-1	-1	-4.753644005	-4.753644
表示しない行	0		21	1	1	1	1	1	7.9079430296	7.90794303
ラベルのついた行	0		22	-1	-1	-1	-1	1	-1.515031782	-1.5150318
			23	-1	1	-1	1	1	4.6632191757	4.66321918

図10.18 「応答をシミュレート」ウィンドウ

効果	Y
切片	1
X1	1
X2	1
X3	1
X4	1
X5	1
X6	1

係数のリセット

分布

☒ 正規 エラー σ: 1

☐ 二項

☐ Poisson

適用

計画テーブルと「応答をシミュレート」ウィンドウが表示されます。計画テーブルに次の2つの列が追加されます。

- － 「Y」列：「応答をシミュレート」ウィンドウの設定に従ってシミュレートされたデータ値。
- － 「Yのシミュレーション」列：計算式が含まれており、その計算結果が表示されている。「応答をシミュレート」ウィンドウでモデルを設定すると、そのモデルに従った乱数が生成されます。計算式を表示するには、「列」パネルで列名の右側にある＋記号をクリックしてください。

次の節では、二項分布に従う応答をシミュレートして、一般化線形モデルをあてはめます。

JMP PRO 「応答をシミュレート」の設定

ここでは、二項分布に従う応答データをシミュレートします。成功率はロジスティックモデルに従うものとします。[応答をシミュレート]の詳細については、『実験計画 (DOE)』の「カスタム計画」章を参照してください。

メモ： この節の手順を省略するには、「二項分布のシミュレート」スクリプトの横にある緑の三角ボタンをクリックします。この操作が済んだら、「[一般化線形モデルのあてはめ](#)」(297ページ)に進んでください。

1. 「応答をシミュレート」ウィンドウ (図10.18) で、「Y」の下に以下の値を入力します。

- － 「切片」に「0」と入力します。
- － 「X1」は、デフォルト値の「1」をそのまま使います。
- － 「X2」に「0.9」と入力します。
- － 「X3」に「0.8」と入力します。
- － 「X4」に「0.7」と入力します。
- － 「X5」に「0.6」と入力します。

- 「X6」に「0.5」と入力します。
- 2. 「分布」アウトラインから【二項】を選択します。
「N」の値は「1」のままにしておきます。これは、それぞれの試行回数が1回だけであることを示します。

図10.19 設定後の「応答をシミュレート」ウィンドウ

効果	Y
切片	0
X1	1
X2	0.9
X3	0.8
X4	0.7
X5	0.6
X6	0.5

係数のリセット

分布

☐ 正規

☒ 二項 N

☐ Poisson

適用

- 3. 【適用】をクリックします。
データテーブルの「Yのシミュレーション」列が、二項分布に従う乱数を生成する計算式に置き換えられます。「Y試行回数」の列には、各実験における試行回数が表示されます。
- 4. (オプション)「列」パネルで、「Yのシミュレーション」の右側にある+記号をクリックします。

図10.20 「Yのシミュレーション」に挿入される、二項分布に従う乱数の計算式

$$\text{Random Binomial} \left(1, \left(1 + \text{Exp} \left(-1 \cdot \left(0 + 1 \cdot X1 + 0.9 \cdot X2 + 0.8 \cdot X3 + 0.7 \cdot X4 + 0.6 \cdot X5 + 0.5 \cdot X6 \right) \right) \right) \right)$$

- 5. 【キャンセル】をクリックします。

JMP PRO 一般化線形モデルのあてはめ

- 1. データテーブルの「モデル」スクリプトの横の緑の三角ボタンをクリックします。
- 2. [Y] ボタンの横の変数「Y」をクリックし、[削除] ボタンをクリックします。
- 3. 「Yのシミュレーション」をクリックし、[Y] ボタンをクリックします。
応答変数Yとして、二項分布に従う乱数の列が設定されます。
- 4. 「手法」のメニューから【一般化線形モデル】を選択します。

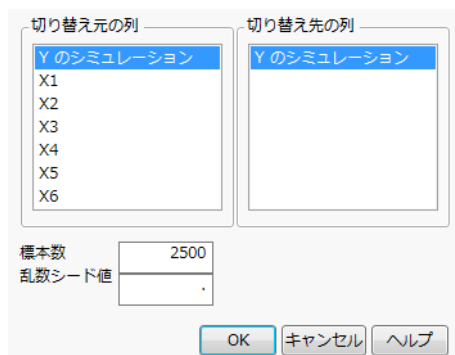
5. 「分布」メニューから「二項」を選択します。
「リンク関数」メニューに「ロジット」が表示されます。
6. 「実行」をクリックします。
二項分布のデータ1組に対して、モデルがあてはめられます。

JMP PRO 検出力の考察

「この例での計画」(293 ページ) では、線形予測子および係数と、係数によって決められる確率を示しました。ここでは、それらの係数に対する検出力を求めます。

1. 「効果の検定」アウトラインで、「p 値 (Prob>ChiSq)」列を右クリックして「シミュレーション」を選択します。

図10.21 「シミュレーション」ウィンドウ



「切り替え元の列」の「Y のシミュレーション」は、モデルの推定に使用されたデータ値を含んでいます。「切り替え先の列」でも「Y のシミュレーション」を選択すると、切り替え元の「Y のシミュレーション」データ値を、「Y のシミュレーション」の計算式で指定されている乱数に置き換えて、シミュレーションが実行されます。

レポートで選択した「p 値 (Prob>ChiSq)」列は、「係数がゼロである」という帰無仮説に対する尤度比検定の p 値です。「効果の検定」表にある、各効果に対するこの「p 値 (Prob>ChiSq)」をシミュレートしましょう。

2. 「標本数」に「500」と入力します。
3. (オプション)「乱数シード値」に「123」と入力し、テキストボックスの外をクリックします。
4. 「OK」をクリックします。

「一般化線形モデル シミュレーション結果」データテーブルが表示されます。

メモ: シミュレーションは乱数によって行われるため、結果は図 10.22 とは異なります。

図10.22 「シミュレーション結果」テーブル（一部）

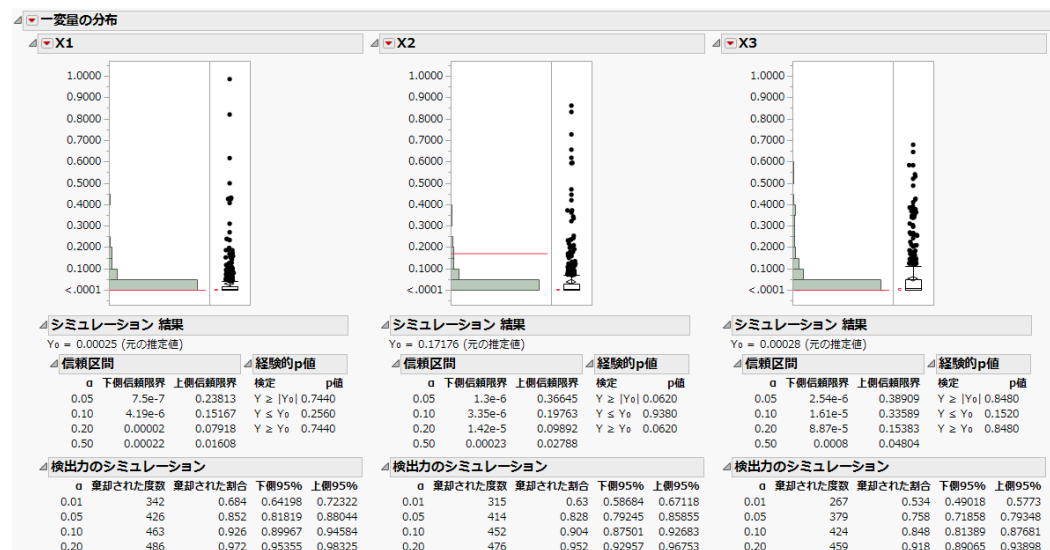
一般化線形モデル シミュレーシ...	乱数シード値	123	SimID	X1	X2	X3	X4	X5	X6
一変量の分布	検出力の分析	1	0	0.0053	0.0509	<.0001	0.1992	0.0055	0.2639
		2	1	0.0001	0.0065	0.1445	0.9470	0.0263	0.1742
		3	2	0.0152	0.0128	0.0129	0.2524	0.4177	0.0463
		4	3	0.0017	0.0330	0.0011	0.0200	<.0001	0.1135
		5	4	0.0002	0.0761	0.0003	0.1305	0.1993	0.6392
		6	5	0.0179	0.1074	0.0310	0.0314	0.0765	0.1103
		7	6	0.1678	0.0060	0.0004	0.0004	0.0015	0.1920
		8	7	0.3097	0.0003	0.0010	0.1448	0.0003	0.0157
		9	8	0.0183	0.0005	0.4225	0.0005	0.1902	0.3351
		10	9	0.0213	<.0001	0.0076	0.0039	0.3077	0.2824
		11	10	0.0873	0.0039	0.0107	0.0014	0.0006	0.4335
		12	11	0.1144	0.1976	0.1020	0.0002	0.0459	0.5622
		13	12	0.0371	0.0181	<.0001	0.2194	0.0032	0.2504
		14	13	0.0005	0.0002	<.0001	0.0032	0.0028	0.9486
		15	14	0.0022	0.0035	0.0529	0.0054	0.0016	0.1947
		16	15	0.0184	0.0009	<.0001	0.0073	0.0747	0.9105
		17	16	0.0012	<.0001	<.0001	0.0018	0.1585	0.0230
		18	17	0.0044	<.0001	0.0018	0.0051	0.7172	0.0007
		19	18	0.0049	0.0099	0.0009	0.2483	0.0822	0.4739

テーブルの最初の行は、観測されたデータから得られた「p 値(Prob>ChiSq)」であり、除外の状態になっています。2行目以降の500行は、シミュレーションしたデータから計算された結果です。

5. 「検出力の分析」スクリプトを実行します。

メモ：シミュレーションは乱数によって行われるため、検出力のシミュレーションの結果は、図 10.23 とは異なります。

図10.23 一変量の分布（最初の3つの効果）



各主効果のヒストグラムには、シミュレーションによって得られた500個の「 p 値(Prob>ChiSq)」が描かれています。また、「検出力のシミュレーション」アウトラインには、500回のシミュレーションのうち有意性検定で棄却されたものの割合が表示されます。

次に、見やすくするために、レポートを縦に積み重ねて、プロットを非表示にしてみましょう。

6. 「一変量の分布」の赤い三角ボタンをクリックし、[積み重ねて表示] を選択します。
7. Ctrl キーを押しながら、「X1」の赤い三角ボタンをクリックし、[外れ値の箱ひげ図] の選択を解除します。
8. Ctrl キーを押しながら「X1」の赤い三角ボタンをクリックし、[ヒストグラムオプション] を選択し、[ヒストグラム] の選択を解除します。

メモ: シミュレーションは乱数によって行われるため、検出力のシミュレーションの結果は、図 10.24 とは異なります。

図10.24 検出力（最初の3つの効果）

一変量の分布									
X1									
シミュレーション 結果					検出力のシミュレーション				
Y ₀ = 0.00025 (元の推定値)									
信頼区間			経験的p値						
α	下側信頼限界	上側信頼限界	検定	p値	α	棄却された度数	棄却された割合	下側95%	上側95%
0.05	7.5e-7	0.23813	Y ≥ Y ₀	0.7440	0.01	342	0.684	0.64198	0.72322
0.10	4.19e-6	0.15167	Y ≤ Y ₀	0.2560	0.05	426	0.852	0.81819	0.88044
0.20	0.00002	0.07918	Y ≥ Y ₀	0.7440	0.10	463	0.926	0.89967	0.94584
0.50	0.00022	0.01608			0.20	486	0.972	0.95355	0.98325
X2									
シミュレーション 結果					検出力のシミュレーション				
Y ₀ = 0.17176 (元の推定値)									
信頼区間			経験的p値						
α	下側信頼限界	上側信頼限界	検定	p値	α	棄却された度数	棄却された割合	下側95%	上側95%
0.05	1.3e-6	0.36645	Y ≥ Y ₀	0.0620	0.01	315	0.63	0.58684	0.67118
0.10	3.35e-6	0.19763	Y ≤ Y ₀	0.9380	0.05	414	0.828	0.79245	0.85855
0.20	1.42e-5	0.09892	Y ≥ Y ₀	0.0620	0.10	452	0.904	0.87501	0.92683
0.50	0.00023	0.02788			0.20	476	0.952	0.92957	0.96753
X3									
シミュレーション 結果					検出力のシミュレーション				
Y ₀ = 0.00028 (元の推定値)									
信頼区間			経験的p値						
α	下側信頼限界	上側信頼限界	検定	p値	α	棄却された度数	棄却された割合	下側95%	上側95%
0.05	2.54e-6	0.38909	Y ≥ Y ₀	0.8480	0.01	267	0.534	0.49018	0.5773
0.10	1.61e-5	0.33589	Y ≤ Y ₀	0.1520	0.05	379	0.758	0.71858	0.79348
0.20	8.87e-5	0.15383	Y ≥ Y ₀	0.8480	0.10	424	0.848	0.81389	0.87681
0.50	0.0008	0.04804			0.20	459	0.918	0.89065	0.93898

「検出力のシミュレーション」アウトラインの「棄却された割合」は、シミュレーションのうち、 p 値が α より小さくなったものの割合を示します。たとえば、係数が0.8、確率の差が38%である「X3」で、棄却された割合を見てみると、有意水準0.05の検定で $379/500 = 0.758$ となっています。表10.1シミュレーションで計算された、有意水準0.05の検定に対する検出力を、効果ごとに示したものです。「検出したい

差」が小さくなるにつれ、検出力も小さくなっています。たとえば、「検出したい差」が24.5% (X6) における検出力は、約0.37にとどまっています。

メモ: 応答値はシミュレートされた値であるため、検出力のシミュレーションの結果は、表 10.1 とは異なります。

表 10.1 検出力のシミュレーション (有意水準 0.05)

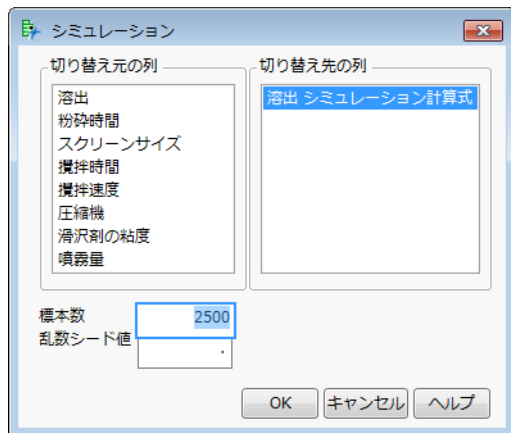
因子	$X_1 = 1$ の場合の 合格率	$X_1 = -1$ の場合の 合格率	検出したい差	シミュレーションで得られた検出力 ($\alpha=0.05$ で棄却された割合)
X_1	73.11%	26.89%	46.2%	0.852
X_2	71.09%	28.91%	42.2%	0.828
X_3	69.00%	31.00%	38.0%	0.758
X_4	66.82%	33.18%	33.6%	0.654
X_5	64.56%	35.43%	29.1%	0.488
X_6	62.25%	37.75%	24.5%	0.372

JMP PRO 「シミュレーション」ウィンドウの起動

「シミュレーション」ウィンドウを開くには、レポートウィンドウで計算値の列を右クリックし、[シミュレーション] を選択します。このシミュレーション機能は、ブートストラップを行うときと同じ操作で実行できます。なお、シミュレーション機能を使用するには、予め、乱数を生成する計算式の列をデータテーブルに準備しておく必要があります。

JMP PRO 「シミュレーション」ウィンドウ

図10.25 「シミュレーション」ウィンドウ (Tablet Production.jmp)



「シミュレーション」ウィンドウには、以下のパネルとオプションがあります。

切り替え元の列 切り替え先の列によって置き換えられる列。

切り替え先の列 切り替え元の列を置き換える列。この列の計算式に従ってシミュレーションした値を使って、分析が繰り返されます。このリストボックスには、計算式を含んだ列だけが表示されます。

標本数 データをシミュレーションし、レポートを再実行する回数。デフォルトの値は2500です。

乱数シード値 乱数を決めるためのシード値。この乱数シード値は、シミュレーション結果を再現したいときに使用します。

「シミュレーション」ウィンドウで [OK] をクリックすると、進行状況を示すバーと、[ここで停止] ボタンが表示されます。進行状況を示すバーの上には、シミュレーションが終了した標本数が表示されます。[ここで停止] をクリックすると、その時点までの計算結果が「シミュレーション結果」テーブルに表示されます。進行状況を示すウィンドウには、実行中の分析名も表示されます。

JMP PRO シミュレーション結果のテーブル

シミュレーションの結果は、テーブル形式で表示されます。次の点を念頭に置いてください。

- データテーブルにおける最初の行は、元のレポートの結果となっています。この行には、除外の属性が与えられます。
- その他の行には、シミュレーションの結果が示され、その行数は、「シミュレーション」ウィンドウで「標本数」に指定した数と同じです。

- 元のレポートにおける表の1列目が、結果のデータテーブルにおける列になります。結果のデータテーブルでは、これらの見出しが列名になります。
- 結果のデータテーブルには、「一変量の分布」レポートを作成する「**一変量の分布**」スクリプトが含まれます。結果のデータテーブルの各列について、ヒストグラム、分位点、要約統計量、シミュレーションの結果を表示します。「一変量の分布」レポートには、基本的な情報のほか、次の情報が表示されます。
 - ヒストグラムに表示される元の推定値を示す赤い線。
 - 元の推定値、信頼区間、経験 p 値を含む「シミュレーションの結果」レポート。「[シミュレーションの結果](#)」レポート」(303ページ)を参照してください。
 - 結果が p 値の形式である場合は、「検出力のシミュレーション」レポートも表示されます。「[検出力のシミュレーション](#)」レポート」(303ページ)を参照してください。
- p 値の列に対してシミュレーションを行った場合にのみ、「**検出力の分析**」スクリプトが表示されます。このスクリプトは「一変量の分布」レポートを生成し、 p 値のヒストグラムと、「検出力のシミュレーション」レポートを表示します。「[検出力のシミュレーション](#)」レポート」(303ページ)を参照してください。

「シミュレーションの結果」レポート

元の推定値 「シミュレーション結果」データテーブルの最初の行にも表示されている元の推定値。この値は Y_0 と記されています。

信頼区間 0.05、0.10、0.20、および0.50の有意水準における、分位点に基づく信頼区間の下限および上限を示します。

経験 p 値 両側検定と片側検定の経験 p 値は、元の推定値をシミュレートした値を比較します。これらの経験 p 値は、レポートのテスト列で指定された範囲内にあるシミュレートした値の割合で計算されます。

「検出力のシミュレーション」レポート

Alpha 有意水準。0.01、0.05、0.10、および0.20です。

棄却された度数 検定によって棄却されたシミュレーションの数を、有意水準別に表示します。

棄却された割合 検定によって棄却されたシミュレーションの割合を、有意水準別に表示します。

下側95%と上側95% 棄却されたシミュレーションの割合について、95%信頼区間の下限および上限を示します。信頼区間の計算には、Wilsonのスコア法を使用しています。Wilson (1927) を参照してください。

ヒント: 信頼区間を狭くするには、シミュレーションの標本数を増やしてください。

第11章

JMP Pro ブートストラップ

標本再抽出によって統計量の分布を近似する

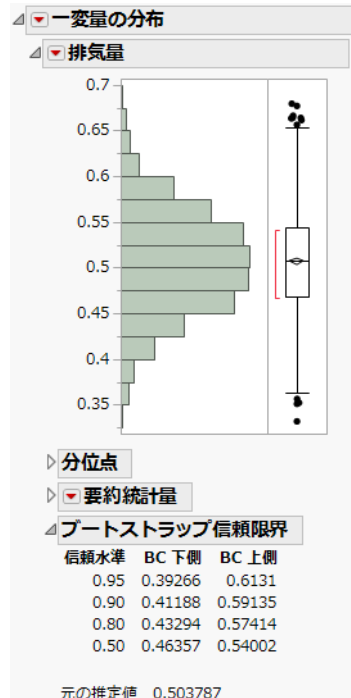
ブートストラップは、JMP Pro でのみ使用できます。

ブートストラップは、データから標本を何度も無作為抽出することにより、統計量の標本分布を近似する手法です。この手法を使えば、統計量（平均、バイアス、標準誤差、信頼区間など）やその分布を推測することができます。ブートストラップは、特に以下のような状況で役立ちます。

- 統計量の標本分布を理論的に導出するのが複雑であるか、不明である
- 推測を行うための仮定が満たされないため、パラメトリックな推測が行えない

メモ：[Bootstrap] オプションは、通常の赤い三角ボタンのコマンドではなく、レポートの表を右クリック（コンテキストクリック）すると呼び出されるメニューに用意されています。

図11.1 傾きパラメータに対して、ブートストラップを実行した結果



JMP PRO ブートストラップの概要

ブートストラップでは、得られたデータから何度も復元抽出することにより、統計量の分布を求めます。なお、JMPのブートストラップでは、「データの各行は互いに独立している」と仮定します。

単純なブートストラップでは、標本サイズが n である元データから、標本サイズが n であるブートストラップ標本が復元抽出されます。このブートストラップ標本には、同じデータ行が重複して抽出されたり、逆に、抽出されないデータ行があったりすることに注意してください。ブートストラップ標本において、同じデータ行が重複している回数を、「**ブートストラップの重み**」と呼びます。ブートストラップを実行するたびに、以下のような情報が求められ、分析全体が再実行され、統計量が再計算されます。

- ブートストラップの標本（ n 個のオブザベーションから成るデータセット）
- ブートストラップの重み（分析プラットフォームの度数変数）

この計算を反復して、目的の統計量の分布を求めます。

ただし、単純なブートストラップが適切でない状況もあります。たとえば、データセットが小さかったり、ロジスティック回帰で分離の問題がある場合などです。このような場合に対して、JMPでは、ベイズ流のブートストラップ（Bayesian bootstrap）も用意されています。ベイズ流のブートストラップでは、重みが小数点以下を含むものになっています。この小数の重みは、合計すると n になります。この小数の重みを、各分析用のプラットフォームで度数変数として用いることにより、目的の統計量の分布を求めます。小数の重みの詳細については、「[小数の重み](#)」（309ページ）および「[小数の重みの計算](#)」（318ページ）を参照してください。

レポートでブートストラップ分析を実行するには、テーブルから統計量の列を選択し、[Bootstrap] を選択します。

メモ：[Bootstrap] オプションは、通常の赤い三角ボタンのコマンドではなく、レポートの表を右クリック（コンテキストクリック）すると呼び出されるメニューに用意されています。

ブートストラップの機能は、JMPのほとんどの統計分析用プラットフォームで用いることができます。目的の統計量を計算するのに使われたデータから、標本抽出が行われます。なお、分析するときに「度数」の列が用いられている場合には、オブザベーションはその度数の回数だけ繰り返し発生するものとして処理されます。また、分析するときに「重み」の列が用いられている場合には、ブートストラップにも、その重みが反映されます。

ヒント：ブートストラップを実行すると、元のレポートに表示されているすべての分析が、再実行されます。そのため、特定の列を選択してブートストラップを行っても、すべての分析が再実行されるために、時間がかかる場合があります。処理に時間がかかる場合は、プラットフォームのレポートから、無関係なオプションを除いてください。

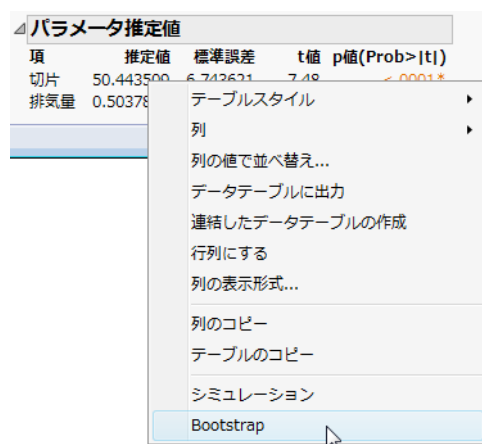
JMP PRO ブートストラップの例

この例では、「Car Physical Data.jmp」サンプルデータテーブルを使用します。あるタイヤメーカーが、エンジンの排気量 (in³) からエンジンの馬力を予測しようとしています。2つの変数には直線関係があるとし、傾きについて推測するとします。直線の傾きがわかれば、排気量の増加に伴う、馬力の増加が分かります。

この例では、回帰分析で信頼限界を求めるときの前提である等分散性が満たされていません。そのため、回帰モデルの傾きに対する信頼限界が、不適切である可能性があります。そこで、この会社では、ブートストラップ法を用いることにしました。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Car Physical Data.jmp」を開きます。
2. [分析] > [二変数の関係] を選びます。
3. 「馬力」を選択し、[Y, 目的変数] をクリックします。
4. 「排気量」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、メニューから「直線のあてはめ」を選択します。
傾きの推定値は0.503787 (約0.504) です。
7. (オプション)「パラメータ推定値」レポートを右クリックし、[列] > [下側95%] を選択します。
8. (オプション)「パラメータ推定値」レポートを右クリックし、[列] > [上側95%] を選択します。
回帰分析における傾きの信頼限界は、0.4249038 および0.5826711 となっています。
9. 「パラメータ推定値」レポートの「推定値」列を右クリックし、[Bootstrap] を選択します (図11.2)。

図11.2 [Bootstrap] オプション



[選択された列の分割] オプションを選択する場合は、どの列で右クリックしたかによって、分析対象となる列が変わる点に注意してください。詳細は、「「ブートストラップ」ウィンドウのオプション」(309ページ) を参照してください。

10. 「ブートストラップ標本数」に「1000」と入力します。

11. (オプション) 図11.3の結果と一致させるには、「乱数シード値」に「12345」と入力します。

12. [OK] をクリックします。

ブートストラップ処理が行われ、傾きと切片の結果を積み重ねていない「ブートストラップ結果」データテーブルが作成されます。

次に、ブートストラップによって生成された傾きの分布を分析します。

13. 「ブートストラップ結果」テーブルで、「一変量の分布」スクリプトを実行します。

「一変量の分布」レポートには、「ブートストラップ信頼限界」レポートが含まれます(図11.3)。

図11.3 ブートストラップのレポート

一変量の分布		
排気量		
分位点		
100.0%	最大値	0.6800107545
99.5%		0.6483289879
97.5%		0.6212501769
90.0%		0.5781520893
75.0%	四分位点	0.5423786474
50.0%	中央値	0.5051381996
25.0%	四分位点	0.4692123772
10.0%		0.4374744231
2.5%		0.4035615438
0.5%		0.3686834278
0.0%	最小値	0.3494461047
要約統計量		
ブートストラップ信頼限界		
信頼水準	BC 下側	BC 上側
0.95	0.40028	0.61892
0.90	0.41646	0.59849
0.80	0.43405	0.57489
0.50	0.46765	0.54007
元の推定値 0.503787		

傾きの推定値(手順6)は、0.504です。ブートストラップで求めた95%の信頼水準に基づくと、傾きは0.40028~0.61892であると推定できます。つまり、排気量が1単位増加すると、95%の信頼度で、馬力は0.40028~0.61892だけ増加と言えます。ブートストラップによる傾きの信頼区間(0.400~0.619)は、通常の回帰分析(手順7および手順8)による信頼区間(0.425~0.583)に比べ、わずかに広がっています。

メモ: 「ブートストラップ信頼限界」レポートの「BC 下側」列および「BC 上側」列は、**バイアス修正済み(bias-corrected)の信頼限界**です。[「バイアス修正済みの信頼限界」](#)(318ページ)を参照してください。

JMP PRO 「ブートストラップ」ウィンドウのオプション

ブートストラップを実行するには、レポートウィンドウ内のテーブルで、統計量の数値列を右クリックし、メニューから **[Bootstrap]** を選択します。選択した列が強調表示され、「ブートストラップ」ウィンドウが表示されます。このウィンドウでオプションを選択し、**[OK]** をクリックすると、選択した列のすべての統計値に対してブートストラップが実行され、デフォルトの結果テーブルが表示されます。

「ブートストラップ」ウィンドウには、次のオプションがあります。

ブートストラップ標本数 データからの無作為抽出と、そこからの統計量の計算といった一連の処理を反復する回数を設定します。回数を増やせば、それだけ統計量の状態をより正確に推定することができます。デフォルトでは、2500に設定されています。

乱数シード値 後でブートストラップ分析を再実行したときに、同じ結果を再現するには、同じ乱数シード値を入力します。デフォルトでは、シード値は設定されていません。

小数の重み バイズ流のブートストラップ分析を実行します。ブートストラップを実行するたびに、各データ行に対する重みが適用されます（重みの計算方法については、「[小数の重みの計算](#)」（318ページ）を参照してください）。この重みを使って、目的の統計量が計算されます。デフォルトでは、このオプションは選択されておらず、通常の単純なブートストラップが実行されます。

ヒント: 標本サイズが小さかったり、ロジスティック回帰で分離の可能性がある場合は、**[小数の重み]** オプションを使用してください。

[小数の重み] オプションを選択すると、ブートストラップにおける抽出ごとに、各データ行に対して重み（0より大きな、小数点を含む値）が乱数により生成されます。なお、これらの重みの合計は、目的の統計量を計算したデータの標本サイズ n になっています。重みの計算方法と使用方法については、「[小数の重みの計算](#)」（318ページ）を参照してください。

選択された列の分割 ブートストラップで得られた統計量を、結果のデータテーブルにおいて列ごとに分割します。このデータテーブルでは、各行が、ブートストラップ標本1組に対応しています（ただし、最初の1行目は、観測されたデータに対するものです）。

このオプションをオフにすると、積み重ねた形式のデータテーブルだけが作成されます。この形式では、ブートストラップにおける抽出ごとに、元の表全体が縦に積み重ねられていきます。この形式では、各行が1つの統計量になっています。そして、元の表における各列が、結果のデータテーブルにおける各列に対応しています。例として、「[結果を積み重ねたデータテーブル](#)」（310ページ）を参照してください。

分割が成功したら積み重ねたテーブルを破棄（**[選択された列の分割]** オプションを選択した場合のみ選択可能）生成される結果のデータテーブルの数が決まります。

このチェックボックスをオフにした場合は、以下の2つのデータテーブルが作成されます。

- 積み重ねた形式のデータテーブル。元の表を積み重ねた形式で、表のすべての列に対する結果を含んでいます。

- － 分割した形式のデータテーブル。積み重ねた形式のデータテーブルを分割し、元の表で選択した列に対する結果だけを含んでいます。

このチェックボックスをオンにすると、[選択された列の分割] の処理が成功した場合は、積み重ねた形式のデータテーブルは破棄されます。

JMP PRO 結果を積み重ねたデータテーブル

ブートストラップを行うと、結果が縦に積み重なったデータテーブルが作成されます (図11.4)。ただし、[分割が成功したら積み重ねたテーブルを破棄] オプションを選択した場合、このデータテーブルは多くの場合、破棄されます。図11.4は、「Car Physical Data.jmp」を「二変数の関係」プラットフォームで分析し、「パラメータ推定値」レポートに対してブートストラップを実行した結果です。「ブートストラップの概要」(306ページ)を参照してください。

図11.4 ブートストラップ結果テーブル (積み重ね)

	X	Y	項	～バイアス	推定値	標準誤差	t値	p値(Prob> t)	BootID●
x	1 排気量	馬力	切片		50.443509471	6.7436209999	7.48	<.0001	0
x	2 排気量	馬力	排気量		0.5037874592	0.0398202611	12.65	<.0001	0
	3 排気量	馬力	切片		45.385604758	5.4724273046	8.29	<.0001	1
	4 排気量	馬力	排気量		0.5451674092	0.0306151772	17.81	<.0001	1
	5 排気量	馬力	切片		40.843862813	7.0508187326	5.79	<.0001	2
	6 排気量	馬力	排気量		0.5854988173	0.0457041828	12.81	<.0001	2
	7 排気量	馬力	切片		47.765642104	5.4677610087	8.74	<.0001	3
	8 排気量	馬力	排気量		0.4943459579	0.0305765677	16.17	<.0001	3
	9 排気量	馬力	切片		59.758060908	7.0436720036	8.48	<.0001	4

縦に積み重なったデータテーブルには、次のような特徴があります。

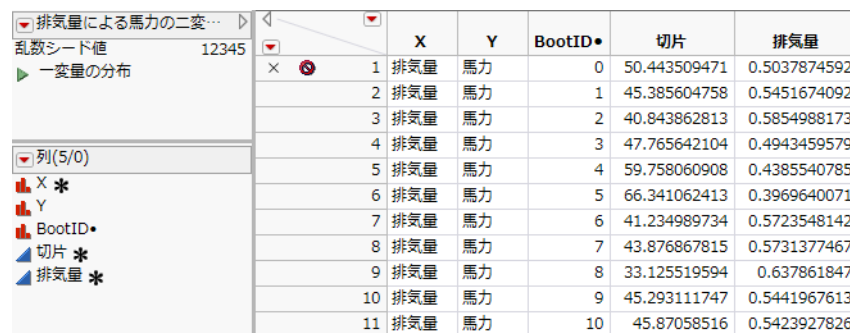
- ・ 各ブートストラップ標本に対して、元の表における1列目で示されている統計量ごとに、別々の行が生成されます。表におけるこの1列目の列は、結果のデータテーブルでも保存されます。この例では、各ブートストラップ標本に対し、元の表における「項」列における「切片」と「排気量」の結果が、別々の行に保存されています。
- ・ 分析で使ったデータテーブル列が含まれます。この例では、Xの排気量と、Yの馬力です。
- ・ 元の表にあるすべての列が、結果のデータテーブルにも保存されます。この例では、「～バイアス」、「推定値」、「標準誤差」、「t値」、「p値 (Prob>|t|)」といった列があります。ここで、「～バイアス」は表にある列なのですが、今回の設定ではバイアスはないために、非表示になっている点に注意してください。
- ・ 「BootID●」列は、ブートストラップ標本の識別IDです。「BootID●」が0の行には、元の推定値が含まれています。この行には、x印のマーカーと、「除外」の行属性が与えられます。この例では、各ブートストラップ標本につき、結果が2行ずつ(「切片」と「排気量」)生成されています。
- ・ このデータテーブル名の末尾には、「ブートストラップの結果 (積み重ね)」と付きます。

[選択された列の分割] オプションを選択した場合、結果を積み重ねていないデータテーブルも表示されます。「結果を積み重ねていないデータテーブル」(311ページ)を参照してください。

JMP PRO 結果を積み重ねていないデータテーブル

表で選択した列に含まれている値ごとに、別々のデータテーブル列にするには、[選択された列の分割] チェックボックスをオンにします。この例では、元の表における「項」の各値が、結果のデータテーブルにおける各列に対応しています。たとえば、図11.5では、図11.4の「推定値」列の値が、データテーブルでは2つの列（「排気量」と「切片」）に分割されています。この2列は、「項」の列に含まれている2つの統計量です。

図11.5 結果を積み重ねていないデータテーブル



	X	Y	BootID•	切片	排気量
1	排気量	馬力	0	50.443509471	0.5037874592
2	排気量	馬力	1	45.385604758	0.5451674092
3	排気量	馬力	2	40.843862813	0.5854988173
4	排気量	馬力	3	47.765642104	0.4943459579
5	排気量	馬力	4	59.758060908	0.4385540785
6	排気量	馬力	5	66.341062413	0.3969640071
7	排気量	馬力	6	41.234989734	0.5723548142
8	排気量	馬力	7	43.876867815	0.5731377467
9	排気量	馬力	8	33.125519594	0.637861847
10	排気量	馬力	9	45.293111747	0.5441967613
11	排気量	馬力	10	45.87058516	0.5423927826

積み重ねていない形式の結果のデータテーブルには、次のような特徴があります。

- ブートストラップ標本ごとに、1つの行が作成されます。
- 分析でを使用したデータテーブル列が含まれます。この例では、Xの「排気量」と、Yの「馬力」です。
- 元の表における各行が、結果のデータテーブルにおいて、別々の列になります。
- 「ブートストラップ」ウィンドウで乱数シード値を指定すると、結果のデータテーブルに、「乱数シード値」というテーブル変数とその値が表示されます。
- 結果を積み重ねていないテーブルには、「ソース」および「一変量の分布」というテーブルスクリプトがあります。「一変量の分布」を実行すると、このブートストラップ標本に基づく統計量（ブートストラップ信頼限界を含む）をすばやく得ることができます。
- 「BootID•」列は、ブートストラップ標本の識別IDです。「BootID•」が0の行には、元の推定値が含まれています。この行には、x印のマーカと、「除外」の行属性が与えられます。積み重ねていない形式のデータテーブルでは、各ブートストラップ標本につき、1つの行が作成されます。
- データテーブル名の末尾には、「ブートストラップ結果 (<列名>）」と付きます（ここで、<列名>は、ブートストラップを実行した列名です）。

JMP PRO 作成した統計量データの分析

ブートストラップの結果は、「一変量の分布」プラットフォームを使って分析します。

- 積み重ねていない形式のデータテーブルにおいては、生成されたデータテーブルに「一変量の分布」スク립トが保存されるので、それを実行できます。
- 積み重ねた形式のデータテーブルにおいては、[分析] > [一変量の分布] を選択し、列を選択して役割を割り当てます。ほとんどのケースでは、元の表で1列目になっているものを [BY] に割り当てればうまくいきます。

その「一変量の分布」レポートでは、2種類のブートストラップ信頼限界が計算されます。また、「BootID・」列があるテーブルについては、「ブートストラップ信頼限界」レポートも生成されます（図11.6）。

「一変量の分布」レポートを使うと、2種類のブートストラップ信頼限界を取得できます。

- 「分位点」レポートには、**分位点による信頼区間**が示されます。たとえば、分位点に基づいて95%信頼区間を求めるには、2.5%と97.5%の分位点を、下限および上限として使用します。
- 「ブートストラップ信頼限界」レポートには、**バイアス修正済みの信頼限界**が表示されます。信頼水準が95%、90%、80%、50%の信頼区間が示され、「BC 下側」列および「BC 上側」列には、それぞれ下限と上限が表示されます。バイアス修正済みの信頼限界の計算については、「[バイアス修正済みの信頼限界](#)」（318ページ）を参照してください。

図11.6 「ブートストラップ信頼限界」レポート

一変量の分布		
排気量		
分位点		
100.0%	最大値	0.6800107545
99.5%		0.6483289879
97.5%		0.6212501769
90.0%		0.5781520893
75.0%	四分位点	0.5423786474
50.0%	中央値	0.5051381996
25.0%	四分位点	0.4692123772
10.0%		0.4374744231
2.5%		0.4035615438
0.5%		0.3686834278
0.0%	最小値	0.3494461047
要約統計量		
ブートストラップ信頼限界		
信頼水準	BC 下側	BC 上側
0.95	0.40028	0.61892
0.90	0.41646	0.59849
0.80	0.43405	0.57489
0.50	0.46765	0.54007
元の推定値 0.503787		

「ブートストラップ信頼限界」レポートの下にある「元の推定値」は、観測された元データから算出された推定値です。

「ブートストラップ信頼限界」レポートの解釈の詳細については、「[ブートストラップの概要](#)」(306ページ)を参照してください。分位点に基づいて信頼区間を求める方法(バイアス未修正および修正済みのもの)については、Efron (1981) が説明しています。

ブートストラップの別例

この例では、標本サイズの小さいデータに対して、「小数の重み」オプション(ベイズ流のブートストラップ法)を使うメリットを説明します。使用するデータは、7種類の異なる土壌ごとに3つのサンプルを採取し、応答 Y を測定したものです。ここで、土壌「wabash」における平均に対する信頼区間を求めるとします。

各土壌について、3つしか測定値がないので、単純なブートストラップ法では、抽出されたブートストラップ標本に、土壌が「wabash」であるデータが1つも含まれない場合が出てきます。「小数の重み」オプションを使えば、各ブートストラップ標本において、常に「wabash」のデータが含まれます。

「wabash」の標本について、平均値の分布を2つのブートストラップ法で調べてみましょう。

- 「[単純なブートストラップ分析](#)」(313ページ)
- 「[ベイズ流のブートストラップ法](#)」(316ページ)

単純なブートストラップ分析

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Snapdragon.jmp」を開きます。
2. [分析] > [二変量の関係] を選びます。
3. 「 Y 」を選択し、[Y , 目的変数] をクリックします。
4. 「土壌」を選択し、[X , 説明変数] をクリックします。
5. [OK] をクリックします。
6. 赤い三角ボタンをクリックし、[平均/ANOVA] を選択します。
7. 「各水準の平均」レポートで、「平均」列を右クリックし、[Bootstrap] を選択します。
8. 「ブートストラップ標本数」に「1000」と入力します。
9. (オプション) 図11.7の結果と一致させるには、「乱数シード値」に「12345」と入力します。
10. [OK] をクリックします。

図11.7 単純なブートストラップの結果

	X	Y	BootID•	clarion	clinton	compost	knox	o'neill	wabash	webster
1	土壌	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	土壌	Y	1	32.2333	30.9000	28.0000	35.7000	34.3500	31.9000	31.1000
3	土壌	Y	2	32.4333	30.3000	30.2000	33.9667	31.2000	38.0000	•
4	土壌	Y	3	32.5000	30.7500	29.2000	34.0333	32.8000	38.0000	31.1000
5	土壌	Y	4	31.5000	29.4000	29.9000	35.7500	35.0400	34.9500	•
6	土壌	Y	5	32.5000	30.6000	31.8000	35.7000	34.3000	35.0500	31.5667
7	土壌	Y	6	32.4000	30.1000	29.8500	34.4500	36.0000	38.2000	30.4000
8	土壌	Y	7	31.9000	29.3400	• 34.4000	33.6000	35.9667	31.4500	
9	土壌	Y	8	32.1400	31.3500	29.6667	33.1000	35.1000	37.9333	30.6333
10	土壌	Y	9	32.7000	30.8000	30.2000	35.2600	31.2000	34.8500	32.5000
11	土壌	Y	10	32.1000	32.1000	28.0000	33.1000	34.1143	34.0000	31.8000
12	土壌	Y	11	31.5000	30.3000	• 33.1000	34.4000	37.2125	30.6333	
13	土壌	Y	12	32.7000	30.4200	31.8000	34.0333	35.1000	35.0500	•
14	土壌	Y	13	32.1667	• 30.1000	34.9000	33.9000	38.2000	31.8000	
15	土壌	Y	14	32.3000	• 29.2000	33.9667	34.4000	35.6000	31.9400	
16	土壌	Y	15	32.2500	29.1000	29.6667	33.1000	35.1000	31.9000	•
17	土壌	Y	16	31.5000	30.4200	30.0000	• 35.2800	37.8000	31.1000	
18	土壌	Y	17	32.4333	29.7000	29.2000	33.1000	34.4000	34.0000	31.1000
19	土壌	Y	18	32.3800	32.1000	29.9000	35.8000	34.6800	35.0500	31.8000
20	土壌	Y	19	32.0600	30.1000	30.2800	• 33.1500	31.9000	31.1000	

図11.7の欠測値は、そのときのブートストラップ標本の抽出において、該当する土壌タイプのデータ値が1つも選択されなかったことを意味します。

11. [分析] > [一変量の分布] を選びます。
12. 「wabash」を選択し、[Y, 列] をクリックします。
13. [OK] をクリックします。

図11.8 「wabash」の平均値の分布（ベイズ流ブートストラップ）

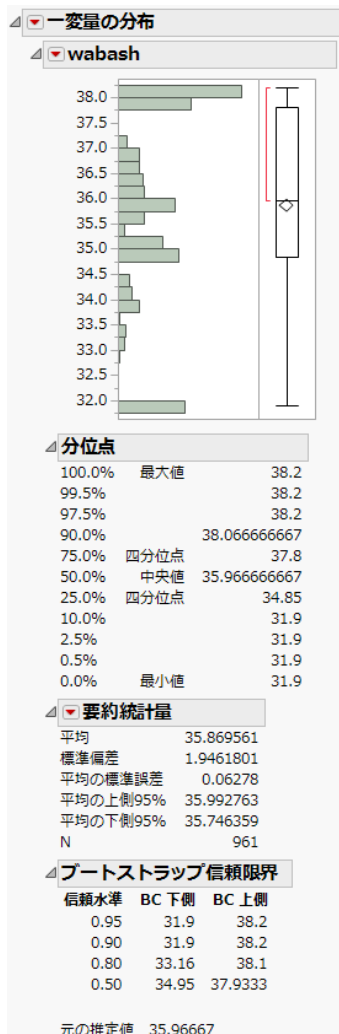


図11.8は、「wabash」の平均値の分布を、単純なブートストラップ分析で求めた結果です。上の例で、次の点に注意してください。

- 「要約統計量」レポートによると、ブートストラップを実行した結果、「wabash」の平均値を含む行数は、 $N = 961$ でした。これは、ブートストラップを1,000回実行し、そのうち39回では、「wabash」の3つの測定値がいずれもブートストラップ標本に選ばれなかったということを意味しています。
- 標本の平均値のヒストグラムは滑らかではなく、両端にピークの値があります。「wabash」の3つの値は、38.2、37.8、31.9です。分布の下端のピークは、31.9のみがブートストラップ標本に含まれた場合の結果です。分布の上端のピークは、38.2と37.8だけがブートストラップ標本に含まれた場合の結果です。

次に、ブートストラップ標本が無い状態になるのを回避するために、[小数の重み] オプション（ベイズ流のブートストラップ法）を使用してみましょう。

ベイズ流のブートストラップ法

1. 「一元配置分析」レポートの「各水準の平均」レポートで、「平均」を右クリックし、[ブートストラップ] を選択します。
2. 「ブートストラップ標本数」に「1000」と入力します。
3. (オプション) 図11.9の結果と一致させるには、「乱数シード値」に「12345」と入力します。
4. [小数の重み] チェックボックスをオンにします。
5. [OK] をクリックします。

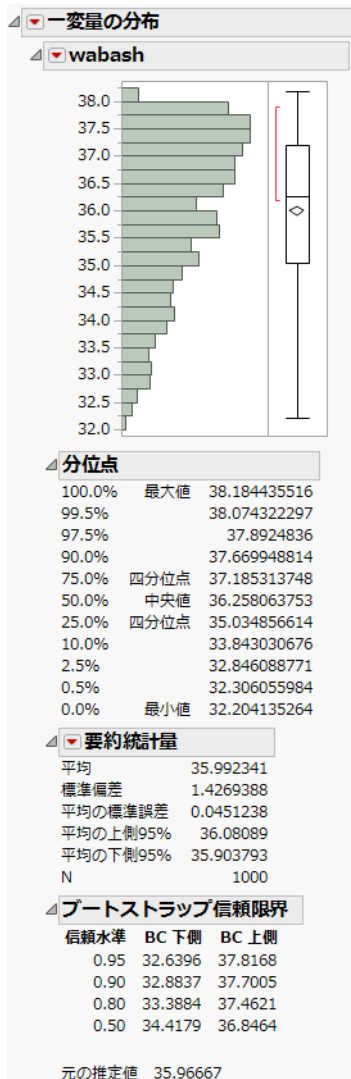
図11.9 ベイズ流ブートストラップによる結果

	X	Y	BootID	clarion	clinton	compost	knox	o'neill	wabash	webster
1	土壌	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	土壌	Y	1	31.9365	31.3493	29.3234	35.4497	34.7105	33.7270	31.4071
3	土壌	Y	2	32.4189	30.3474	29.5143	35.6281	32.7006	34.1027	32.1212
4	土壌	Y	3	32.2339	30.2001	31.4102	34.8758	33.6674	38.0389	31.5428
5	土壌	Y	4	32.4054	30.3242	30.6227	33.8495	32.9495	36.8344	31.9607
6	土壌	Y	5	32.2262	30.8672	29.7999	33.6759	32.2022	35.5792	31.9058
7	土壌	Y	6	32.3222	31.9732	28.8823	35.6307	34.8863	35.3014	30.1325
8	土壌	Y	7	31.9948	30.8828	29.2516	35.3424	33.2094	36.3367	30.8386
9	土壌	Y	8	31.6254	29.7677	28.6390	34.4697	33.0662	36.6183	31.4667
10	土壌	Y	9	32.3499	29.9416	29.5732	35.2564	31.9583	35.8246	29.9302
11	土壌	Y	10	32.5228	30.4506	28.4859	34.9088	35.8126	34.6317	31.6100
12	土壌	Y	11	31.9057	30.0711	29.0693	35.4018	34.4654	33.2086	31.0309
13	土壌	Y	12	31.7275	29.6189	29.1609	34.3984	33.6840	35.0815	31.5446
14	土壌	Y	13	32.5893	30.5210	28.6054	33.4594	34.0958	33.7692	31.7129
15	土壌	Y	14	32.2473	30.5199	31.6080	35.5617	34.1706	35.7146	31.8023
16	土壌	Y	15	32.2329	29.7275	30.1770	35.3286	32.7820	37.4866	30.9706
17	土壌	Y	16	31.5831	29.7508	28.7122	33.5304	34.5348	37.1092	31.2752
18	土壌	Y	17	32.3545	31.3237	29.1542	35.5890	32.2606	37.2005	30.8430
19	土壌	Y	18	32.3811	29.6241	30.6138	35.4308	33.2024	33.0787	31.2926
20	土壌	Y	19	31.7488	29.7763	28.7327	34.7007	33.8910	34.0573	29.9064

ベイズ流ブートストラップでは、結果のデータテーブルに欠測値はありません。各標本抽出において、「Snapdragon.jmp」データテーブルの21行がすべて含まれるように、ブートストラップの重みが設定されています。

6. [分析] > [一変量の分布] を選びます。
7. 「wabash」を選択し、[Y, 列] をクリックします。
8. [OK] をクリックします。

図11.10 「wabash」の平均値の分布（ベイズのブートストラップ）



ベイズ流のブートストラップ法による結果では、「wabash」の平均が、より滑らかに分布しています。1,000個のブートストラップ標本のすべてに、「wabash」の3つの測定値が含まれています。この方法では、各標本抽出において、異なる小数の重みを適用して、「wabash」の標本平均を求めています。「ブートストラップ信頼限界」レポートでは、95%水準の平均値の信頼区間は、32.6396~37.8168となっています。

JMP PRO ブートストラップの統計的詳細

この節では、ブートストラップ法の詳細を説明します。

JMP PRO 小数の重みの計算

「小数の重み」のオプションは、ベイズ流のブートストラップ法 (Bayesian bootstrap; Rubin 1981) に基づいています。ある1つのブートストラップ標本において、同じデータ行が重複して抽出される回数を、「ブートストラップの重み」と呼びます。通常のブートストラップ法では、各ブートストラップの重みは、単純な復元抽出法によって決められます。

ベイズ流のブートストラップ法では、ブートストラップ抽出における抽出確率を確率変数とみなし、無情報の事前分布を仮定して、その事後分布を求めます。そして、抽出確率の乱数を、その事後分布から生成します。その抽出確率から、ブートストラップの重みを求めます。具体的には、以下の手順に従います。

- 形状パラメータが $(n - 1)/n$ 、尺度パラメータが1である、独立なガンマ分布に従う n 個の乱数を生成する。

メモ: Rubin (1981) は、ガンマ分布の形状パラメータとして1を使用しています。JMP Pro で採用している形状パラメータは、重みの平均と分散が、単純なブートストラップの重みの平均と分散に等しくなるように設定されています。

- $S = (n \text{ 個の乱数の和})$ を計算する。
- n 個の乱数のそれぞれに n/S を掛けたものを、小数の重みとする。
- 分析において度数変数や重み変数を使用した場合には、行ごとに、小数の重みに、度数や重みを掛け合わせる。ただし、度数や重みの合計は、1より大きいものとします。度数変数または重み変数の値の合計は、1より大きいものとします。

この手順で計算された小数の重みは、その平均と分散が、単純なブートストラップの重みの平均と分散と等しくなっています。小数の重みは正の値であり、すべてを合計すると n 、その平均は1となっています。

JMP PRO バイアス修正済みの信頼限界

この節では、バイアス修正済み (bias-corrected、以下BC) の信頼限界について説明します。BC 信頼限界は、結果のデータテーブルで「一変量の分布」スクリプトを実行すると、「ブートストラップ信頼限界」レポートに示されます。BC 信頼限界は、ブートストラップ分布の非対称性に対処して、分位点の信頼限界を補正しています。Efron (1981) を参照してください。

記号の説明

- p^* は、ブートストラップ標本の推定値が、元の推定値以下となっている割合。
- z_0 は、標準正規分布の p^* 分位点。
- z_α は、標準正規分布の α 分位点。

バイアス修正済み信頼限界の両端

信頼水準が $(1 - \alpha)$ であるバイアス修正済み信頼限界は、ブートストラップ分布における以下の分位点です。

- 下限は、累積確率 $\Phi(2z_0 + z_\alpha)$ に対する分位点。
- 上限は、累積確率 $\Phi(2z_0 + z_{1-\alpha})$ に対する分位点。

第12章

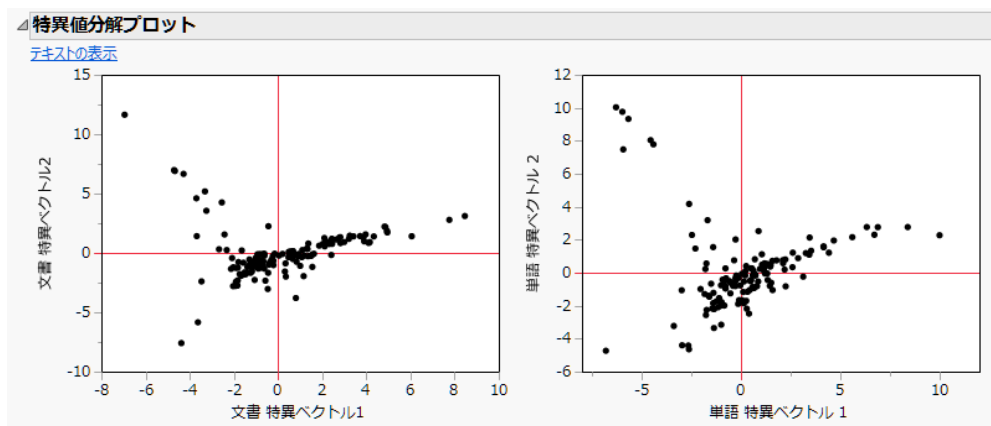
テキストエクスプローラ 非構造化テキストを探索する

JMP PRO このプラットフォームの機能の多くは、JMP Pro だけの機能です。JMP Pro だけで動作する機能の説明には、このマークが付いています。

「テキストエクスプローラ」は、アンケート調査の自由回答や、医療現場のインシデントレポートなど、構造化されてないテキストデータを分析するためのプラットフォームです。このツールを使って、似ている単語を1つにまとめたり、誤っている語句を再コード化したりでき、テキストデータに隠れているパターンを把握できます。

JMP PRO JMP Pro では、特異値分解に基づいて、似ている文書をトピックごとに分類することもできます。似ている文書を分類したり、複数の文書で一緒に使われている単語を分類することができます。また、潜在クラス分析によって、文書を分類することもできます。

図12.1 テキストエクスプローラの特異値分解プロット



「テキストエクスプローラ」プラットフォームの概要

構造化されていないテキストデータは珍しくありません。たとえば、アンケートの自由回答欄や、製品のレビューコメント、インシデントレポートといった例があります。「テキストエクスプローラ」プラットフォームを使えば、構造化されていないテキストを分析し、その意味の核心に迫ることができます。テキスト分析は、通常、単語の整理（キュレーティング）と、統計分析とを交互に行いながら進めていきます。

単語リストの調整

テキスト分析では、独自の用語が使用されます。**単語 (term, token)** はテキストを構成する最小単位で、文章におけるワード (word) のようなものです。単語を定義する方法は、正規表現の使用をはじめ、多数の方法があります。テキストを単語に分解することを、**トークン化 (tokenization)** と呼びます。

- **句**は少数の単語から構成されます。このプラットフォームには、句を単語として扱うオプションも用意されています。
- **文書**は単語の集合です。JMP のデータテーブルでは、テキスト列に含まれる各行の非構造化テキストが、1つの文書に対応します。
- **コーパス**は複数の文書の集合を指します。

通常は、頻出する一般的な単語を除外して分析したほうがよいでしょう。これらの除外する単語は、「**ストップワード**」と呼ばれます。「テキストエクスプローラ」プラットフォームには、いくつかのストップワードが予め定義されていますが、分析者が自分でストップワードを追加することもできます。ストップワードは、単語としては除外されますが、句の中には含まれていても構いません。

単語を再コード化することもできます。この機能は、同義語を1つにまとめるのに便利です。

語幹抽出とは、単語の語尾の変化を取り除き、**語幹**が共通する単語をまとめる処理のことを指します。たとえば、「jump」、「jumped」、「jumping」は、すべて「jump」として扱われます。語幹抽出の方法は、文字列処理言語の Snowball の方法と似ています。句に対しては、単語単体に対する語幹抽出と同じように、句に含まれる各単語を語幹抽出します。

単語リストの分析

「テキストエクスプローラ」プラットフォームでは、テキスト分析に **bag of words (単語の袋)** というアプローチを採用しています。このアプローチでは、句における構成を除き、語順は考慮されません。分析は、単語の出現頻度だけに基づいて行われます。

正規表現 (Regex)、ストップワード、再コード化、語幹抽出などの整理（キュレーティング）を行って単語リストを作成したら、その整理した単語リストに対して統計分析を行います。このプラットフォームでは、**文書単語行列 (DTM; Document-Term Matrix)** に対する統計分析を行えます。文書単語行列の各行は、1つの文書に対応しています (JMP データテーブルにおけるテキスト列の1セルが、1文書に対応しています)。文書単語行列の各列は、整理した後の各単語に対応しています。文書単語行列に対する統計分析では、語順を無視する **bag of words** アプローチを採用しています。もっとも単純な場合では、文書単語行列は、単語の出現頻度となっています。それ以外の重みを用いる方法もいくつかあり、これについては「[保存オプション](#)」(344 ページ) で説明します。

JMP PRO このプラットフォームの統計分析では、最初に文書単語行列に対して特異値分解(SVD; Singular Value Decomposition)を実行します。これにより、単語情報に関するデータの次元を減らすことができます。特異値分解については、『多変量分析』の付録「統計的詳細」を参照してください。また、階層型クラスター分析によって、単語や文書をクラスタリングできます。クラスター分析によって、似たような単語や文書をグループにまとめることができます。

プラットフォームの分析手順

「テキストエクスプローラ」プラットフォームの通常の手順は以下のとおりです。

1. どのようにトークン化を行うかを指定します。予め用意されている正規表現や、自分でカスタマイズした正規表現などを指定します。
2. レポート上で、ストップワードの追加や、単語リストへの句の追加、単語の再コード化、語幹抽出の例外指定などを行います。
3. 語幹抽出の方法を指定します。
4. 単語と句の出現頻度、特異値分解、クラスター分析などの手法を使って、重要な単語や句を見極めます。

メモ: 特異値分解とクラスター分析は、JMP Pro でのみ使用できます。

5. 単語テーブル、文書単語行列、特異値などの結果を、他の分析で使用できる形式で保存します。

メモ: 特異値を保存するオプションは、JMP Pro でのみ使用可能です。

6. 句、再コード化、ストップワードの設定を、他の似たようなテキストデータの分析で再利用できるよう保存します。

テキスト処理の手順

テキスト処理は、トークン化、句の抽出、単語の抽出という3つの段階に分けられます。

トークン化

この段階では次の処理が実行されます。

1. テキストを小文字に変換します。
2. [基本ワード] によるトークン化、または正規表現([Regex])に基づくトークン化を実行し、文字をトークンに分けます。
3. 指定された再コード化の定義に従って、トークンを再コード化します。再コード化は、語幹抽出の前に行われることに注意してください。

句の抽出

この段階では、コーパス（文書の集合）に出現する句を抽出します。抽出された句を1つの単語として単語リストに追加することもできます。句の先頭または末尾にストップワードがくることはありませんが、句の途中に含めることはできません。

単語の抽出

この段階では、前の段階で抽出されたトークンと句の情報に基づいて、単語リストを作成します。

各トークンについて、次の処理を行います。

1. 起動ウィンドウで指定された、単語あたりの最小文字数および最大文字数の条件を満たしているかどうかをチェックします。数字のみのトークンは、この処理から除外されます。
2. トークンを単語として見なせるかどうかをチェックします。トークン化の方法として「基本ワード」を選択した場合は、少なくとも1文字のアルファベットまたはユニコード文字が含まれていることが条件となります。数字のみのトークンは、この処理から除外されます。「Regex」を選択した場合は、正規表現を使用して、トークンを構成する文字を識別します。
3. トークンがストップワードでないかどうかをチェックします。
4. 語幹抽出およびその例外を適用します。

ある句を1つの単語として単語リストに追加すると、次の処理が行われます。

1. 句を単語リストに追加します。このとき、句に含まれる各単語に対して、単語に対して行われた語幹抽出と同じ語幹抽出を行います。その語幹抽出で同じになった句がある場合は、単語リストにて1つにまとめられます。
2. 追加された句に含まれる単語に関して、その単語単体での出現頻度を、該当の句に含まれている分だけ減らします。

「テキストエクスプローラ」プラットフォームの例

この例では、ペットに関するアンケートへの回答を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、「Pet Survey.jmp」を開きます。
2. 「分析」>「テキストエクスプローラ」を選択します。
3. 「質問への回答」を選択し、「テキスト列」ボタンをクリックします。
4. 「OK」をクリックします。

図12.2 「テキストエクスプローラ」レポートの最初の画面の例



このデータには、194文書、異なる単語が372単語あることが先頭の情報からわかります。また、トークンの総数は1921個です。もっとも頻繁に使用されている単語は「cat」で、55回出現しています。

- 「質問への回答」のテキストエクスプローラレポートの赤い三角ボタンをクリックし、[単語のオプション] > [語幹抽出] > [すべての単語] を選択します。
- 「句」のリストから「cat food」と「dog food」を選択して右クリックし、[句の追加] を選択します。
「cat food」と「dog food」が「単語」リストに追加されます。
- 「単語」リストをスクロールダウンして、「cat food」と「dog food」があることを確認します。
それぞれ4回ずつ出現していることがわかります。

図12.3 「単語」リストを編集後にスクロール

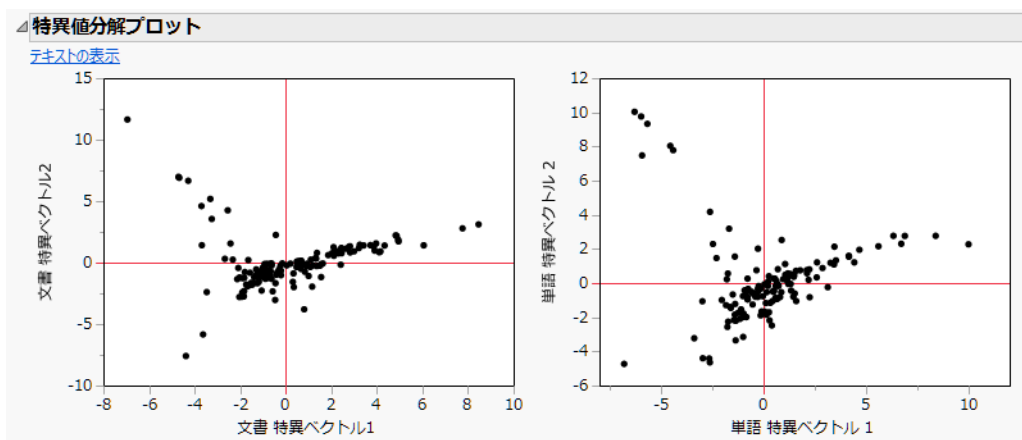
単語	度数
cat food	4
dog food	4
anymore	3
bath	3
shake	3
kid	3
see	3
box	3
stay	3
roll	3
fluffy	3
read	3
hilarious	3
best	3
leg	3

「句」リストの「cat food」と「dog food」は、このレポートでは単語として扱うことにしたため、グレーで表示されます。

8. **JMP PRO** 「「質問への回答」のテキストエクスプローラ」レポートの赤い三角ボタンをクリックし、**[潜在意味分析, 特異値分解]** を選択します。
9. **JMP PRO** デフォルト値をそのままにして、**[OK]** をクリックします。

図12.4のように、特異値分解プロットが表示されます。左のプロットには、文書空間の最初の2つの特異ベクトルが表示されています。右のプロットには、単語空間の最初の2つの特異ベクトルが表示されています。

図12.4 特異値分解プロット



10. **JMP PRO** 左の特異値分解プロットの、一番右側にある3つの点を選択します。
この3つの回答は、他の回答から離れてクラスターを形成しています。このクラスターを詳しく調べるために、質問への回答文を読んでみましょう。
11. **JMP PRO** 「特異値分解プロット」アウトラインのタイトルの下の**[テキストの表示]** をクリックします。

図12.5 選択した文書のテキスト

選択された行の内容:

There was this funny video of a cat trying to jump into someones lap, but fell into the pool instead. [56]

The funny cat video where the cats jumped through the window right into the bathtub was hilarious. [142]

We made this funny video of the cat trying to climb the wall to chase a laser pointer. [153]

ウィンドウが開き、選択した3つの文書のテキストが表示されます。これらの回答は、「funny」、「cat」、「video」という単語が組み合わせて使われているという点で、共通しています。また、これらの3つの文書の第1次元の特異ベクトルが、他の文書よりも大きい正の値をとっており、第1次元において他の文書と異なることが示唆されています。

特異ベクトルについてさらに詳しく分析すると、その次元が何を表しているのかがわかる可能性があります。たとえば、プロットの右端のほうにある回答の多くが、猫に関する内容であるのに対し、左端のほうにある回答の多くは、犬に関する内容です。したがって、第1次元の特異ベクトルは、回答における猫と犬の違いを抽出していると言えます。

「テキストエクスプローラ」プラットフォームの起動

「テキストエクスプローラ」プラットフォームを起動するには、[分析] > [テキストエクスプローラ] を選択します。

図12.6 「テキストエクスプローラ」起動ウィンドウ

「テキストエクスプローラ」起動ウィンドウには、以下のオプションがあります。

テキスト列 テキストデータを含む列を割り当てます。複数の列を指定した場合には、列ごとに別々の分析が実行されます。

ID [積み重ねた文書単語行列の保存] オプションによって保存されるデータテーブルにおいてIDとなる列を指定します。このデータテーブルは、アソシエーション分析を行うのに適しています。

By 指定した列の水準ごとにレポートが作成されます。複数のBy変数を割り当てた場合、それらのBy変数の水準の組み合わせごとに個別のレポートが作成されます。

メモ: By変数を指定した場合、[Regexのカスタマイズ]オプションの設定はBy変数の全水準に適用されます。

言語 テキスト処理に使用する言語を指定します。これは、語幹抽出と、ストップワード、再コード化、および句のビルトインリストに影響します。JMPを実行している言語と同じ言語でなくても構いません。特に設定を変更しない限り、JMPの「表示言語」と同じ言語に設定されますが、テキストエクスプローラの「言語」オプションは、中国語、日本語、韓国語には対応していません。JMPの表示言語が中国語、日本語、または韓国語の場合は、テキストエクスプローラの「言語」オプションはデフォルトで英語に設定されます。

句あたりの最大単語数 分析に含める句の、1句あたりの単語数の最大値を指定します。

句の最大数 句リストに表示する句の最大数を指定します。

単語あたりの最小文字数 分析に含める単語の最小文字数を指定します。

単語あたりの最大文字数 分析に含める単語の最大文字数（最大2000まで）を指定します。

語幹抽出 先頭が同じで末尾が異なる語を同じとみなすかどうかを指定します。次のオプションがあります。

しない 語幹が同じ語をまとめません。

同じ語幹の単語 1つの語幹から2つ以上の単語に派生しているもののみ、同じ語幹にまとめます。

すべての単語 派生している単語がないものを含め、すべての単語を語幹にまとめます。

メモ: 語幹抽出のオプションは、単語リストに追加された句にも適用されます。この処理では、句に含まれる各単語に語幹抽出を行い、その語幹抽出によって同じになる句は1つにまとめられます。たとえば、「dogs bark」、「dog barks」という句は、両方とも「dog·bark」という1つの句にまとめられます。

トークン化 句を単語(トークン)に分解する方法を指定します。トークン化には、次のオプションがあります。

Regex 予め用意されているデフォルトの正規表現(ビルトインの正規表現)によって、テキストを解析します。ビルトインの正規表現を追加、削除、または編集したい場合は、[Regexのカスタマイズ]オプションを選択してください。[「Regexのカスタマイズ: 正規表現エディタ」](#)(329ページ)を参照してください。

基本ワード 一般的によく使用される区切り文字に基づいて、単語を抽出します。この区切り文字には、スペース、タブ、改行文字のほか、句読点の多くが含まれています。単語として数字も含めたい場合は、[数値を単語として扱う]オプションを選択してください。このオプションを選択しなかった場合、区切り文字に基づいて抽出された単語のうち、数字のみのものは除外されます。

ヒント: デフォルトの区切り文字を表示するには、「トークン化」で[基本ワード]を指定してレポートを作成し、「テキストエクスプローラ」レポートの赤い三角ボタンから[表示オプション]>[区切り文字の表示]を選択します。

Regexのカスタマイズ（「トークン化」で[Regex]を指定したときのみ有効）「テキストエクスプローラ 正規表現エディタ」ウィンドウで、正規表現の設定を編集できます。特殊な単語に対応するには、このオプションを使用します。たとえば、電話番号や、文字と数字から構成される単語を抽出するように設定できます。[Regexのカスタマイズ] オプションを使用するのは、デフォルトの正規表現の結果が望ましくない場合のみにしてください。[「Regexのカスタマイズ：正規表現エディタ」](#)（329ページ）を参照してください。

数字を単語として扱う（「トークン化」で[基本ワード]を指定したときのみ有効）数字を単語として抽出します。このオプションを選択すると、数字の単語については、「単語あたりの最小文字数」の設定は無視されます。

起動ウィンドウで[OK]をクリックすると、[Regexのカスタマイズ]を選択している場合は、「テキストエクスプローラ 正規表現エディタ」ウィンドウが表示されます。そうでない場合は、「テキストエクスプローラ」レポートが表示されます。

メモ：テキスト処理では、入力テキストの大文字と小文字は区別されません。テキスト全体が内部で小文字に変換されてから、トークン化されます。正規表現処理や、「テキストエクスプローラ」レポートに出力される単語の集計処理にも、小文字に変換されたテキストが使用されます。

Regexのカスタマイズ：正規表現エディタ

[Regexのカスタマイズ] オプションを選択すると、「テキストエクスプローラ 正規表現エディタ」ウィンドウが表示されます。このウィンドウから、電話番号、時間、金額など、予め提供されている、さまざまなビルトインの正規表現を選択できます。また、独自の正規表現を定義することもできます。

メモ：[Regexのカスタマイズ] オプションを使用するのは、デフォルトの正規表現の結果が望ましくない場合のみにしてください。

図12.7 テキストエクスプローラ 正規表現エディタ



スクリプトエディタボックスでの解析

ウィンドウの上部には、スクリプトエディタボックスが表示され、サンプルテキストの解析結果が表示されます。解析結果の色は、「Regexエディタ」に示されている色に対応しています。

- 「行」の「**前の行**」ボタンまたは「**次の行**」ボタンをクリックすると、分析対象となっているテキストがスクリプトエディタボックスに表示されます。ここで、各行のテキストがどのように解析されるのかを確認できます。
- 「**列に保存**」ボタンをクリックすると、この正規表現によるトークン化の結果が、データテーブルに新しい列として追加されます。正規表現を変更する方法については、「[正規表現の編集](#)」(331 ページ)を参照してください。

メモ: 「**列に保存**」ボタンをクリックしたときは、正規表現のみがテキストのマッチに使用され、ストップワード、再コード化、語幹抽出、句、単語あたりの最小/最大文字数の設定は使用されません。

正規表現の追加

トークン化に使用する正規表現を独自に定義するには、まず、「**ライブラリからコピー**」をクリックします。すると、「正規表現ライブラリからの選択」ウィンドウが表示されます。このウィンドウには、予め用意されている正規表現(ビルトインの正規表現)がすべて表示されます。また、直前に「正規表現エディタ」で編集した正規表現も表示されます。予め用意されている正規表現は「ビルトイン」と記されています。分析者が独自に定義した正規表現(カスタムの正規表現)は、保存した名前で表示されます。同じ名前のカスタムの正規表現は、最新のものだけがライブラリに保存されます。ライブラリからカスタムの正規表現を削除するには、「**選択アイテムを削除**」ボタンを使用します。正規表現のライブラリは、ユーザごとに別のJSLファイルとして、「TextExplorer」というディレクトリに保存されます。このディレクトリの場所は、以下のとおりです。

- Windows: "%C:¥Users¥<ユーザ名>¥AppData¥Roaming¥SAS¥JMP¥TextExplorer¥"
- Macintosh: "/Users/<ユーザ名>/Library/Application Support/JMP/TextExplorer/"

このファイルは、他のユーザと共有しても構いませんが、直接編集しないでください。編集する場合は、「正規表現エディタ」を使用してください。

正規表現の編集

トークン化では、「Regexエディタ」パネルで指定された正規表現が順番に適用されます。適用される順序を変更するには、リストから正規表現を選択して、リストの下にある上矢印または下矢印をクリックしてください。または、リスト内で正規表現をドラッグ&ドロップしても、適用される順序を変更できます。現在選択している正規表現には、青い三角が付きます。正規表現をリストから削除するには、正規表現を選択して、リストの下にあるマイナス記号をクリックしてください。なお、「レフトオーバー」という正規表現は、リストから削除することはできず、常にこのリストの最後に表示されます。

「Regexエディタ」パネルのリストから正規表現を選択すると、編集可能なフィールドにその正規表現の属性が表示されます。これらのフィールドは、すべてクリックして編集できます。

正規表現には次のような属性があります。

タイトル 現在のウィンドウで各正規表現を識別するための名前を指定します（正規表現ライブラリにも、この名前で表示されます）。

Regex 正規表現の定義を指定します。正規表現にマッチする文字列を取得するため、最低1組の括弧が必要です。

結果 正規表現にマッチしたテキストを、どのように置き換えるかを指定します。置き換えるものとして、静的なテキスト、空白、正規表現にマッチしたテキストなどを指定できます。以下のように定義してください。

- 正規表現にマッチしたテキストを、特定の静的なテキストに置き換えるには、その静的なテキストをそのまま指定します。
- マッチした文字列を無視するには、このフィールドを空白にします。
- 正規表現にマッチした一番外側の括弧の結果を使用するには、このフィールドに「¥1」と指定します。
- 正規表現にマッチした結果全体を使用するには、このフィールドに「¥0」と指定します。

例（オプション）正規表現の処理の例を、テキストと色を使って記述します。

コメント（オプション）正規表現の説明や処理を記述します。

色 正規表現にマッチしたテキストに与える色を指定します。スクリプトエディタボックスおよび「例」フィールドにおいて、指定した色が使われます。

メモ：「Regex」フィールドの定義が無効の場合は、正規表現リストのその正規表現名の横に赤いXが表示されます。

独自の正規表現を作成する方法

自分自身で独自の正規表現を定義したい場合（カスタムの正規表現を定義したい場合）には、以下の手順に従います。

1. [ライブラリからコピー] をクリックします。
2. 「正規表現ライブラリからの選択」ウィンドウで、作成したい正規表現に近そうなものを選択します。
3. [OK] をクリックします。
4. 「Regex エディタ」パネルで、正規表現の定義を編集します。
5. 「タイトル」フィールドに、定義した正規表現に対する名前を指定します。

ヒント：正規表現の定義フィールドを編集するときは、ログウィンドウを開いて見える状態にしておくとう便利です。一部のエラーメッセージは、ログウィンドウにのみ表示されます。ログウィンドウを開くには、[表示] > [ログ] を選択します。正規表現のトラブルシューティングについては、www.regexpr.com をはじめ、インターネット上に多数のリソースがあります。

単語の区切り文字のリスト

[単語の区切り文字のリスト] をクリックすると、トークン化処理において単語の区切りと見なす文字のリストを指定できます。**区切り文字**は、ある1つの単語の先頭文字には使うことができません。ただし、正規表現で許可されていれば、単語の途中にある文字としては含めることはできます。[単語の区切り文字のリスト] をクリックするとウィンドウが表示され、このウィンドウで区切り文字の追加や削除を行えます。デフォルトでは、空白（ホワイトスペース）のみが区切り文字として指定されています。「区切り文字」ウィンドウで、区切り文字のリストに加えた変更を取り消すには、[リセット] ボタンをクリックします。区切り文字のリストに加えた変更は、そのときのトークン化処理にのみ適用されます。

指定した正規表現と、必須の「レフトオーバー」正規表現の処理方法は、以下のとおりです。

1. テキストストリームの現在の文字を、区切り文字のリストと比べます。
 - － 現在の文字が、区切り文字のリストのいずれかと一致したら、その文字を除外し、そこまでの時点で蓄積された一時的な「レフトオーバー」文字列を処理し、次の文字に移動して、手順1を繰り返します。
 - － 現在の文字が、区切り文字のリストと一致しない場合は、手順2に進みます。
2. 現在の文字から始まる文字列を、指定された正規表現とひとつずつ比較します（「レフトオーバー」の正規表現は除きます）。
 - － 現在の文字から始まる文字列が、いずれかの正規表現とマッチしたら、次の一連の処理を実行します。ここまでの一時的な「レフトオーバー」文字列に蓄積された文字を処理します。「結果」フィールドに定義されている文字列を、単語として保存します。そして、正規表現にマッチした文字列のすぐ後の1文字を、現在の文字に設定します。これらの処理が済んだら、手順1に戻ります。
 - － 現在の文字から始まる文字列が、「レフトオーバー」以外のどの正規表現にも一致しなかった場合は、手順3に進みます。
3. 現在の文字を一時的な「レフトオーバー」文字列に追加して、テキストストリーム内の次の文字を現在の文字に設定します。手順1に戻ります。

- 一時的な「レフトオーバー」文字列には、他の正規表現に一致するまで、1文字ずつ追加されます。
- 「レフトオーバー」正規表現の「結果」はデフォルトでは空白で、一時的に蓄積された「レフトオーバー」文字列を破棄する設定になっています。

ヒント:

- 「レフトオーバー」正規表現の「結果」を「¥1」に設定した場合は、句読点などを区切り文字に追加することを検討してください。指定した句読点は、結果から除かれます。
- 目的の単語を取得するために、「レフトオーバー」正規表現の結果を「¥1」に変更する代わりに、正規表現ライブラリに正規表現を追加したり、カスタム正規表現を作成したりできないかどうか検討してください。

上記のような処理は、データテーブルの各行において、テキストの最後まで実行されます。

結果をデータテーブルの列に保存する

【**列に保存**】 ボタンをクリックすると、正規表現によるトークン化の結果が、データテーブルに新しい列として追加されます。新しい列は、データタイプが「文字」で、「テキストエクスプローラ」起動ウィンドウで「テキスト列」に指定した名前に、番号が追加されたものになります。

メモ: 正規表現を独自に定義した場合、適用されるトークン化は、元のそのままのテキストに対して、その正規表現を適用したものです。小文字に変換したテキストに対する結果ではない点に注意してください。

「テキストエクスプローラ 正規表現エディタ」を閉じる

「テキストエクスプローラ 正規表現エディタ」ウィンドウで **[OK]** をクリックすると、次の一連の処理が行われます。

1. 「テキストエクスプローラ 正規表現エディタ」ウィンドウで独自に定義した、カスタムの正規表現が、正規表現ライブラリに保存されます。

注意: ユーザーが独自に定義した、カスタムの正規表現がライブラリに保存されるのは、正規表現を定義した後に **[OK]** をクリックしたときのみです。次にライブラリを使用できるのは、最後に保存した正規表現だけです。正規表現ライブラリに定義した正規表現を保存したい場合は、固有の名前を付けてください。以前の正規表現を後で使用したい場合は、「テキストエクスプローラ」レポートウィンドウで、スクリプトを保存することもできます。

2. 「テキストエクスプローラ」レポートが生成されます。このレポートに、指定した正規表現の設定を使ってテキストをトークン化した結果が表示されます。

「テキストエクスプローラ」レポート

「テキストエクスプローラ」レポートには、要約レポートと「単語と句のリスト」レポートが表示されます。

図12.8 「テキストエクスプローラ」レポートの例



要約レポート

「テキストエクスプローラ」レポートの上部には、以下の要約統計量が表示されます。

単語の数 単語リストに含まれる単語の数。

ケース数 コーパスに含まれる文書の数。

トークンの総数 コーパスに含まれる単語の総数。

ケースあたりのトークン トークンの総数を、ケースの数で割った値。

無記入でないケース数 少なくとも1つの単語を含む文書の数。

無記入でないケースの割合 コーパスの全文書数に対し、少なくとも1つの単語を含む文書数の割合。

単語と句のリスト

「単語と句のリスト」レポートには、トークン化が行われた後の単語と句の一覧が表示されます。このレポートの例については、図12.8を参照してください。「単語」リストの「度数」は、コーパス中のその単語の出現頻度です。「句」リストの「度数」は、コーパス中のその句の出現頻度です。「N」は、句を構成する単語数です。

デフォルトでは、「単語」リストは出現頻度の降順に並べられます。出現頻度が同じ単語は、アルファベット順に並べられます。「句」リストは、出現頻度の降順に並べられます。出現頻度が同じ句は、「N」の降順に並べられます。出現頻度とNの両方が同じ句は、アルファベット順に並べられます。両リストとも、アルファベット順に並べ替えるオプションが用意されています。

「句」リストには、起動ウィンドウで設定した【句あたりの最大単語数】と【句の最大数】が反映されています。なお、データテーブルに1度しか登場しない句は、このリストに表示されません。

いろいろな場所に、どの句を単語として扱うかを保存できます。単語として扱われる句は、それが指定された保存場所（スコープ）ごとに、「句」リストで色分けされます。表12.1を参照してください。句を単語として登録するときの保存場所については、「[\[単語のオプション\] から開く管理ウィンドウ](#)」（339ページ）を参照してください。

表12.1 単語に指定された句の色

保存場所	色
ビルトイン	赤
ユーザライブラリ	緑
列プロパティ	オレンジ
ローカル	グレー

単語と句のオプションメニュー

「単語と句のリスト」の各表で、一番左の列を右クリックすると、オプションが表示されます。各表をデータテーブルとして保存するには、「度数」列を右クリックして、[データテーブルに出力] を選択します。

「単語」リストのポップアップメニュー

「単語」リストを右クリックすると、以下のポップアップメニューが表示されます。

行の選択 データテーブルで、選択された単語を含む行を選択します。

テキストの表示 選択された単語を含む文書を表示します。

メモ: デフォルトでは、最初の 10,000 個の文書が表示されます。文書数が 10,000 個を超える場合は、文書数を増やすためのウィンドウが表示されます。

アルファベット順 単語の順序を、アルファベット順に切り替えます。オフにすると、度数の降順に戻ります。

検索 単語リストから、特定の文字列を検索します。検索結果が表示されている状態から単語リストに戻るには、再びこのオプションを選択します。

コピー 選択された単語をクリップボードにコピーします。

色 選択された単語に色を付けます。

ラベル 単語の特異値分解プロットで、選択された単語に対応する点にラベルを付けます。

この単語を含む句を選択 選択された単語を含む句を、「句」リストから選択します。

指示変数の保存 選択された単語ごとに指示変数列を作成し、データテーブルに保存します。指示変数列の値は、その単語を含む文書は1、含まない文書は0になります。

計算式の保存 選択された単語ごとに計算式列を作成し、データテーブルに保存します。計算式の値は、その単語を含む文書は1、含まない文書は0になります。新しい文書に対してこの式を使うと便利です。

再コード化 1つまたは複数の単語の値を変更します。単語を選択してから、このオプションを選択してください。このオプションを選択すると、「再コード化」ウィンドウが表示されます。『JMPの使用法』の「データの入力と編集」章を参照してください。

ストップワードの追加 選択された単語をストップワードに追加し、単語リストから削除します。この操作を行うと、句のリストも更新されます。

語幹の例外的追加 選択された単語を、語幹抽出の例外リストに追加します。

データテーブルに出力 レポートの表から、JMPデータテーブルを作成します。

連結したデータテーブルの作成 レポート上の選択された表と同じ種類の表を連結していき、1つのJMPデータテーブルを作成します。

「句」リストのポップアップメニュー

「句」リストを右クリックすると、以下のポップアップメニューが表示されます。

行の選択 データテーブルで、選択された句を含む行を選択します。

テキストの表示 選択された句を含む文書を表示します。

指示変数の保存 選択された句ごとに指示変数列を作成し、データテーブルに保存します。指示変数列の値は、その句を含む文書は1、含まない文書は0になります。

アルファベット順 句の順序を、アルファベット順に切り替えます。オフにすると、度数の降順に戻ります。

検索 句リストから、特定の文字列を検索します。検索結果が表示されている状態から句のリストに戻るには、再びこのオプションを選択します。

コピー 選択された句をクリップボードにコピーします。

この句を含むものを選択 選択された句を含む、それより長い句を、「句」リストから選択します。

この句に含まれたものを選択 選択された句に含まれる、それより短い句を、「句」リストから選択します。

句の追加 選択された句を「単語」リストに追加し、「度数」列とともに更新します。

ストップワードの追加 選択された句をストップワードに追加し、句リストから削除します。この操作を行うと、単語リストも更新されます。

データテーブルに出力 レポートの表から、JMPデータテーブルを作成します。

連結したデータテーブルの作成 レポート上の選択された表と同じ種類の表を連結していき、1つのJMPデータテーブルを作成します。

「テキストエクスプローラ」プラットフォームのオプション

ここでは、「テキストエクスプローラ」プラットフォームに用意されているオプションについて説明します。

テキストの前処理オプション

「テキストエクスプローラ」レポートの赤い三角ボタンで用意されているオプションのうち、テキストを前処理するものには以下のものがあります。

表示オプション レポートの表示内容を制御するサブメニューが表示されます。

ワードクラウドの表示 「ワードクラウド」レポートの表示／非表示を切り替えます。「ワードクラウド」レポートの赤い三角ボタンをクリックすると、ワードクラウドのレイアウトやフォントを変更するメニューが表示されます。[「ワードクラウド」のオプション](#)（339 ページ）を参照してください。

ワードクラウドの幅は、インタラクティブに変更できます。幅を変更すると、自動的に高さが調整されます。「単語」リストの単語と、ワードクラウドの単語はリンクしています。

単語リストの表示 単語リストの表示／非表示を切り替えます。

句リストの表示 句リストの表示／非表示を切り替えます。

単語と句のオプションの表示 単語と句のポップアップメニューのオプションを、「単語と句のリスト」レポート内に表示します。[「単語と句のリスト」](#)（335 ページ）を参照してください。

要約の表示 要約表の表示／非表示を切り替えます。[「要約レポート」](#)（334 ページ）を参照してください。

ストップワードの表示 分析に使用したストップワードリストの表示／非表示を切り替えます。プラットフォームを起動した最初には、予め用意されているストップワード（ビルトインのストップワード）が使用されます。ストップワードを追加するには、「単語」リストから単語を右クリックし、ポップアップメニューから[「ストップワードの追加」](#)を選択します。[「単語のオプション」から開く管理ウィンドウ](#)（339 ページ）を参照してください。

再コード化の表示 再コード化された単語リストの表示／非表示を切り替えます。[「単語のオプション」から開く管理ウィンドウ](#)（339 ページ）を参照してください。

指定した句の表示 ユーザによって単語に指定された句のリストの表示／非表示を切り替えます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」(339 ページ) を参照してください。

語幹の例外の表示 語幹抽出の例外リストの表示／非表示を切り替えます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」(339 ページ) を参照してください。

区切り文字の表示 トークン化の方法として「基本ワード」を使用したときの区切り文字の表示／非表示を切り替えます。区切り文字のセットを編集するには、JSL で `Add Delimiters()` または `Set Delimiters()` メッセージを使用する必要があります。このオプションは、トークン化の方法として「基本ワード」を選択したときのみ有効です。

語幹レポートの表示 「語幹抽出」レポートの表示／非表示を切り替えます。このレポートには、語幹抽出の結果が2つの形式で表示されます。左の表には、語幹に対して、その語幹に該当する単語が表示されます。右の表には、単語に対して、その単語の語幹が表示されます。このオプションは、「語幹抽出」のメニューで「しない」を選択したときには無効です。

選択した行の表示 ウィンドウが開き、現在選択している行の文書のテキストが表示されます。

単語のオプション 単語リストに関連するサブメニューが表示されます。

語幹抽出 語幹抽出のオプションについては、「[\[テキストエクスプローラ\] プラットフォームの起動](#)」(327 ページ) を参照してください。

ビルトインのストップワードを含める トークン化において、JMP で予め用意されているストップワードを用いるかどうかを指定します。

ビルトインの句を含める JMP で予め用意されている句を単語リストに含めるかどうかを指定します。

ストップワードの管理 ストップワードの追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」、「列」、または「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定されたストップワードをローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」(339 ページ) を参照してください。

再コード化の管理 再コード化する単語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」、「列」、または「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された再コード化をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」(339 ページ) を参照してください。

句の管理 単語として扱う句の追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」、「列」、または「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された句をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」(339 ページ) を参照してください。

語幹の例外の管理 語幹抽出の例外とする単語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」、「列」、または「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された語幹抽出の例外をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」(339 ページ) を参照してください。

解析のオプション 解析方法やトークン化に関連するサブメニューが表示されます。

トークン化 トークン化のオプションについては、「[「テキストエクスプローラ」プラットフォームの起動](#)」(327 ページ) を参照してください。

Regex のカスタマイズ 正規表現をカスタマイズするウィンドウを表示します。現在の「テキストエクスプローラ」レポートに適用する正規表現の設定を編集できます。このオプションは、トークン化の方法として「Regex」を選択したときのみ有効です。

メモ : プラットフォームの起動ウィンドウで By 変数を指定したときは、「Regex のカスタマイズ」オプションの設定は、By 変数の全水準に自動的に適用されます。

数字を単語として扱う 数字を単語としてトークン化します。このオプションを指定すると、数字については「単語あたりの最小文字数」の設定は無視されます。このオプションは、トークン化の方法として「基本ワード」を選択したときのみ有効です。

「ワードクラウド」のオプション

「ワードクラウド」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

レイアウト (デフォルトは「度数順」) ワードクラウド中の単語の並べ方を指定します。

度数順 単語の出現頻度順に、左から右へと表示します。

アルファベット順 単語のアルファベット順に、左から右へと表示します。

中心化 単語を雲状に表示します。表示されている単語の大きさは出現頻度を表しています。

色分け (デフォルトは「なし」) ワードクラウド中の単語の色を指定します。

なし 単語を黒で表示します。この色は、凡例から変更できます。

グレー濃淡 単語にグレーの濃淡を付けます。

任意の色 単語にさまざまな色を付けます。この色は、凡例から変更できます。

列の値による色分け ... 単語の色として、グラデーションがある色を用います。このグラデーションは、各単語のスコア平均と全体平均との差に基づいています ([列による単語のスコア] で、各単語のスコア平均は求めることができます)。色と濃淡は、凡例から変更できます。

フォント 単語のフォント、スタイル、サイズを指定します。

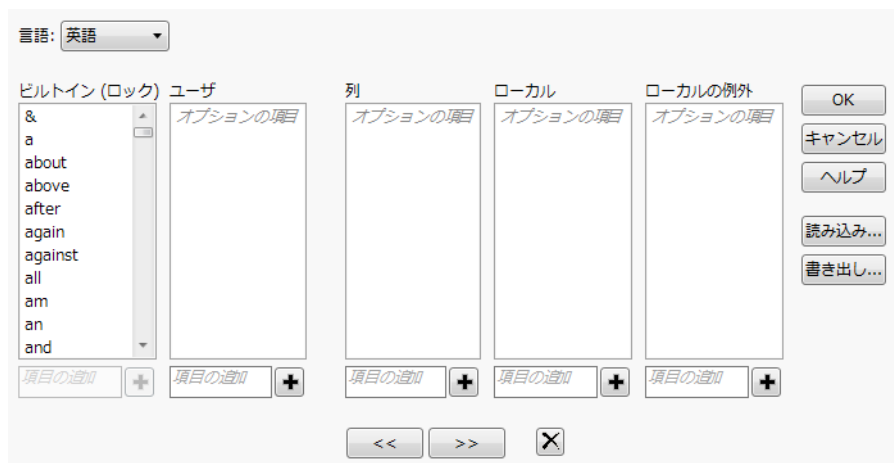
凡例の表示 (デフォルトでオン) 凡例の表示／非表示を切り替えます。

[単語のオプション] から開く管理ウィンドウ

句、ストップワード、再コード化、および語幹の例外に関する情報は、保存場所 (スコープ) を指定することができます。これらの情報が保存される場所は、それぞれ異なり、「ユーザ」はテキストエクスプローラのユーザライブラリに、「列」は分析対象列の列プロパティに、「ローカル」はプラットフォームのスクリプトに、それぞれ保存されます。特定の「テキストエクスプローラ」レポートのローカル定義やローカル例外を保存するには、そのレポートのスクリプトを保存してください。

[単語のオプション] から、ストップワード、再コード化、句、語幹の例外的管理ウィンドウを開くことができます。この4つのウィンドウの構成はほぼ同じです。図12.9に、「ストップワードの管理」ウィンドウを示します。「句の管理」と「語幹の例外的管理」ウィンドウは、このウィンドウと同じです。「再コード化の管理」ウィンドウは、少し異なります。「[再コード化の管理](#)」(341 ページ) を参照してください。

図12.9 「ストップワードの管理」ウィンドウ



ストップワードの管理

「ストップワードの管理」ウィンドウには、ストップワードのリストが保存場所（スコープ）ごとに表示されます。各リストの下には、テキストボックスと追加ボタンがあり、これらによって独自のストップワードを追加できます。また、ストップワードを別の保存場所にドラッグして移動したり、コピー＆ペーストしたりできます。ウィンドウの最下部にある2つのボタンを使えば、選択した項目を左隣りまたは右隣りの保存場所に移すことができます。選択した項目を現在の保存場所から削除するには、[X] ボタンをクリックします。既存の項目を編集するには、その項目をダブルクリックして、テキストを変更します。

言語 JMP で予め準備されているストップワード（ビルトインのストップワード）と、どのユーザライブラリに保存するかについて、言語を指定します。なお、[項目の適用] を選択した場合は、変更内容がマスターのユーザライブラリに保存されます。言語の設定は、「ビルトイン」および「ユーザ」のストップワードにのみ適用されます。

ビルトイン（ロック） 指定された言語について、予め準備されているストップワードが一覧表示されています。ここに表示されているストップワードをローカルでは除外したい場合は、「ローカルの例外」に移します。

ユーザ 指定された言語のユーザライブラリに含まれるストップワードが一覧表示されています。

列 テキスト列の列プロパティ「ストップワード」に含まれるストップワードが一覧表示されています。

ローカル ローカルのストップワードが一覧表示されています。テキストエクスプローラを起動するときに、JSLにおいてローカルのストップワードを指定することもできます。ローカルのストップワードは、現在の「テキストエクスプローラ」レポートでのみ適用されます。

ローカルの例外 現在の「テキストエクスプローラ」プラットフォームで、例外的に除外するストップワードが一覧表示されています。テキストエクスプローラを起動するときに、JSLにおいてローカルの例外を指定することもできます。「ローカルの除外」にリストされた単語は、他のいずれかの保存場所で指定されたストップワードを無効にします。

読み込み ストップワードをテキストファイルから読み込みます。読み込まれたストップワードは、クリップボードにコピーされます。クリップボードから、「ビルトイン」以外のいずれかのリストボックスに貼り付けてください。

書き出し ストップワードを、クリップボードやテキストファイルに書き出します。「書き出し」ウィンドウが表示されるので、書き出したいストップワードの保存場所を選択し、書き出し先を指定します。

ユーザライブラリファイルは、TextExplorerディレクトリに保存されます。このディレクトリの場所は、次のとおりです。

- Windows: "%C:%Users%<ユーザ名>%AppData%Roaming%SAS%JMP%TextExplorer%<言語>%"
- Macintosh: "/Users/<ユーザ名>/Library/Application Support/JMP/TextExplorer/<言語>/"

マスターユーザライブラリファイルは、TextExplorerディレクトリのすぐ下に保存されます。このファイルは、言語に関わらず共通です。

[OK] をクリックすると、「ユーザ」に加えた変更はユーザライブラリに、「列」に加えた変更は列プロパティに、それぞれ保存されます。「ローカル」と「ローカルの例外」に指定した項目は、「テキストエクスプローラ」レポートでスクリプトを保存したときのみ保存されます。

ストップワードをユーザライブラリに保存すると、「stopwords.txt」というファイルが作成されます。また、列プロパティに保存すると、「ストップワード」という列プロパティが作成されます。

再コード化の管理

「再コード化の管理」ウィンドウの構成は、「ストップワードの管理」ウィンドウと少し異なり、各リストの下に、2つのテキストボックスが表示されます。上のボックスに指定した古い値が、下のボックスに指定した新しい値に再コード化されます。

再コード化情報をユーザライブラリに保存すると、「recodes.txt」というファイルが作成されます。また、列プロパティに保存すると、「再コード化」という列プロパティが作成されます。

句の管理

句をユーザライブラリに保存すると、「phrases.txt」というファイルが作成されます。また、列プロパティに保存すると、「句」という列プロパティが作成されます。

語幹の例外の管理

語幹の例外をユーザライブラリに保存すると、「stemExceptions.txt」というファイルが作成されます。また、列プロパティに保存すると、「語幹の例外」という列プロパティが作成されます。

メモ:「語幹の例外の管理」ウィンドウで「ローカルの例外」に指定された単語は、語幹の例外から除外され語幹抽出処理の対象となります。

テキストの分析オプション

「テキストエクスプローラ」レポートの赤い三角ボタンで用意されているオプションのうち、統計分析に関するものには以下のものがあります。

潜在クラス分析 2値の文書単語行列に対して潜在クラス分析を行います。なお、分析の計算には、疎な行列に対する数値計算ルーチンが使用されています。

このオプションを選択すると、「設定」ウィンドウが開き、以下のオプションが表示されます。

単語の最大数 潜在クラス分析に含める単語の最大数。

単語の最小度数 潜在クラス分析に含める単語の最小出現頻度。

クラスターの数 潜在クラス分析のクラスター数。

詳細については、「[潜在クラス分析](#)」(347 ページ) を参照してください。

潜在意味分析, 特異値分解 文書単語行列に対する特異値分解を実行します。なお、分析の計算には、疎な行列に対する数値計算ルーチンが使用されています。「[潜在意味分析 \(特異値分解\)](#)」(348 ページ) を参照してください。

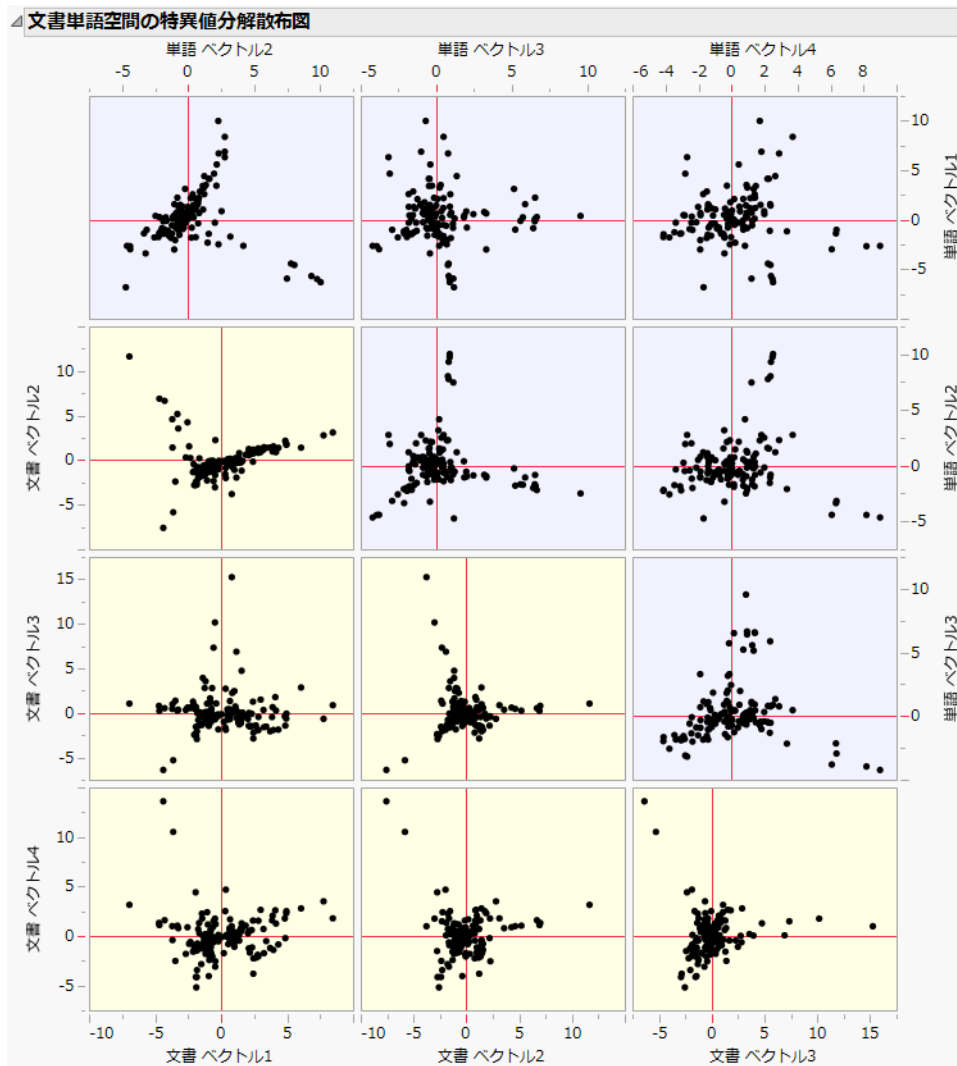
トピック分析, 特異値分解を回転 文書単語行列の特異値分解を Varimax 回転します。単語を「トピック」と呼ばれるグループにまとめます。「[トピックの分析](#)」(349 ページ) を参照してください。

クラスター 単語 単語に対する階層型クラスター分析の表示／非表示を切り替えます。樹形図の右側に、クラスター数を設定するオプションと、クラスターをデータテーブルに保存するオプションが表示されます。このデータテーブルには、単語ごとに、度数、単語を含む文書数、割り当てられたクラスターが保存されます。

クラスター 文書 文書に対する階層型クラスター分析の表示／非表示を切り替えます。樹形図の右側に、クラスター数を設定するオプション、クラスターをデータテーブルの列に保存するオプション、樹形図で選択した枝の文書を表示するオプションが表示されます。

特異値分解 散布図行列 単語と文書の特異ベクトルを描いた散布図行列の表示／非表示を切り替えます。このオプションを選択すると、次元を指定するウィンドウが呼び出されます。最初の2次元だけではなく、指定した次元までの特異値分解の散布図が描かれます。文書を選択して「テキストの表示」ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

図12.10 文書単語空間の特異値分散散布図



トピックの散布図行列（[トピック分析, 特異値分解を回転] を選択したときのみ有効）回転した特異 ベクトルを描いた散布図行列の表示／非表示を切り替えます。文書を選択して [テキストの表示] ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

JMP PRO 特異値分解の設定ウィンドウ

「テキストエクスプローラ」プラットフォームの統計分析は、文書単語行列 (DTM; Document-Term Matrix) に対して行われます。文書単語行列の各列は、単語リストに含まれている各単語に対応しています（ただし、列数の最大値は、「単語の最大数」で指定したものです）。文書単語行列の各行は、1つの文書に対応しています（JMP データテーブルにおけるテキスト列の1セルが、1文書に対応しています）。文書単語行列の各セルに含まれる数値をどのように算出するかは、「設定」ウィンドウにて指定します。

図12.11に、特異値分解の「設定」ウィンドウを示します。「テキストエクスプローラ」の赤い三角ボタンから、文書単語行列の特異値分解を実行するオプションを選択すると、「設定」ウィンドウが開き、以下のオプションが表示されます。

単語の最大数 特異値分解に含める単語の最大数。

単語の最小度数 特異値分解に含める単語の最小出現頻度。

重み 文書単語行列におけるセルの数値を算出する方法。この方法については、「[文書単語行列の設定ウィンドウ](#)」（346ページ）を参照してください。

特異ベクトルの数 特異値分解ベクトルの数。デフォルト値は100です。ただし、文書数または単語数が100を下回るときは、その値がデフォルト値になります。

中心化と尺度化 文書単語行列を中心化や尺度化するか否かを指定します。[中心化]、[中心化と尺度化]、または[中心化しない]のいずれかを選択します。デフォルトでは、中心化と尺度化の両方が適用されます。

図12.11 特異値分解の「設定」ウィンドウ

単語と重みの設定

単語の最大数	419
単語の最小度数	10
重み	TF IDF
特異ベクトルの数	100
中心化と尺度化	中心化と尺度化

OK キャンセル ヘルプ

保存オプション

「テキストエクスプローラ」レポートの赤い三角ボタンには、新たなデータテーブル、データテーブルの新たな列、または、列プロパティに情報を保存するための以下のオプションが含まれています。

文書単語行列の保存 文書単語行列の各列を、データテーブルの列に保存します（列数の最大値は、「単語の最大数」に指定します）。

JMP PRO **文書 特異ベクトルの保存** ユーザが指定した数の特異ベクトルを、データテーブルの列に保存します。

JMP PRO **文書 トピックベクトルの保存** ユーザが指定した数の、回転後の特異ベクトルを、データテーブルの列に保存します。

JMP PRO **積み重ねた文書単語行列の保存** 文書単語行列を、積み重ねた形式で JMP データテーブルに保存します。この積み重ねた形式のデータテーブルは、「アソシエーション分析」プラットフォームで分析を行いたい場合に適しています。詳細については、『予測モデルおよび発展的なモデル』の「アソシエーション分析」章を参照してください。「テキストエクスプローラ」の起動ウィンドウで ID 変数を指定した場合は、この ID によって、単語が元のテキストデータテーブルのどの行からきたのかを識別できます。積み重ねたテーブルには、アソシエーション分析を起動するスクリプトも含まれています。

文書単語行列の計算式の保存 ベクトルで記述された計算式を、データテーブルの列に保存します。ベクトルのサイズは、[単語の最大数]、[単語の最小度数]、および [重み] にユーザが指定した値によって決まります。列の計算式には、JSL 関数の `Text Score()` が使用されます。この関数の詳細については、[ヘルプ] > [スクリプトの索引] を参照してください。

JMP PRO **特異ベクトルの計算式の保存** （[潜在意味分析, 特異値分解] を選択したときのみ有効）文書の特異ベクトルを求める計算式を、データテーブルの列に保存します。列の計算式には、JSL 関数の `Text Score()` が使用されます。この関数の詳細については、[ヘルプ] > [スクリプトの索引] を参照してください。

JMP PRO **トピックベクトルの計算式の保存** （[トピック分析, 特異値分解を回転] を選択したときのみ有効）回転した特異値ベクトルを求める計算式を、データテーブルの列に保存します。列の計算式には、JSL 関数の `Text Score()` が使用されます。この関数の詳細については、[ヘルプ] > [スクリプトの索引] を参照してください。

単語テーブルの保存 単語リスト、出現頻度、単語を含む文書数から成る JMP データテーブルを作成します。[列による単語のスコア] を選択した場合は、スコアの列も作成されます。

JMP PRO **単語 特異ベクトルの保存** 新しいデータテーブルを作成し、行に単語を、列に単語の特異ベクトルをユーザが指定した次元だけ保存します。単語テーブルがすでに開かれている場合は、そのテーブルに列を追加します。

JMP PRO **単語 トピックベクトルの保存** （[トピック分析, 特異値分解を回転] を選択し、かつ [単語テーブルを保存] を選択して単語テーブルを作成済みのときのみ有効）単語テーブルに、トピックベクトルの列を追加します。

列による単語のスコア スコアを計算する列を指定し、単語リストとそのスコアを JMP データテーブルに保存します。各単語のスコアは、指定した列の各値に、単語の出現頻度に基づく重み付けをしたうえで、平均値をとったものです。

文書単語行列の設定ウィンドウ

「テキストエクスプローラ」レポートの赤い三角ボタンから「文書単語行列の保存」または「文書単語行列の計算式の保存」を選択すると、設定ウィンドウが開き、以下のオプションが表示されます。

単語の最大数 文書単語行列に含める単語の最大数。

単語の最小度数 文書単語行列に含める単語の最小出現頻度。

重み 文書単語行列の要素を算出する方法。

以下のオプションがあります。

2 値 各文書において、該当の単語が1つでもある場合は1を、1つもない場合は0を割り当てます。

3 値 各文書において、該当の単語が2つ以上ある場合には2、1つしかない場合は1、1つもない場合は0を割り当てます。

度数 各文書内の単語の出現頻度を割り当てます。

度数の対数 $\log_{10}(1 + x)$ を割り当てます (x は各文書内の単語の出現頻度)。

TF IDF $TF * \log(nDoc / nDocTerm)$ を割り当てます。TF IDF は、**単語度数 (term frequency)** と **書度数の逆数 (inverse document frequency)** の略。これがデフォルトの重みです。計算式の各項の定義は以下のとおりです。

- TF = 文書内の単語の出現頻度
- $nDoc$ = コーパスに含まれる文書の数
- $nDocTerm$ = その単語を含む文書の数

レポートのオプション

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。

JMP PRO 潜在クラス分析

潜在クラス分析を実行すると、内容の似たような文書が同じクラスターに属するように分類されます。「潜在クラス分析」レポートには、そのモデルのベイズ情報量規準値（BIC）と「テキストの表示」ボタンが表示されます。「クラスターの混合確率」表から1つまたは複数のクラスターを選択し、「テキストの表示」ボタンをクリックすると、そのクラスターに分類される確率の高い文書のテキストが表示されます。

「潜在クラス分析」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

表示オプション 「潜在クラス分析」レポートに表示する内容を指定します。デフォルトで、すべてのオプションが選択されています。

クラスターの混合確率 該当のクラスターに文書が含まれる事前確率を示した表の表示／非表示を切り替えます。

ヒント:「クラスターの混合確率」表で1つまたは複数の行を選択すると、そのクラスターに事後確率により割り当てられるデータ行も選択されます。

クラスターごとの単語確率 ある文書が特定のクラスターに属しているという条件のもとで、その文書に各単語が登場する条件付き確率の推定値を示します。単語は、デフォルトで、コーパス全体での出現頻度の降順で並べられます。

「最も特徴的なクラスター」には、その単語が最も高い割合で出現するクラスターが示されます。

「最も確率の高いクラスター」には、その単語を含む文書をランダムに選んだときに、その文書が入る確率が最も高いクラスターが示されます。

各クラスターの上位単語 各クラスターに最も頻繁に出現する、上位10個の単語の表示／非表示を切り替えます。

多次元尺度構成プロット 多次元尺度構成プロット（MDSプロット）の表示／非表示を切り替える。このプロットは、クラスターの類似度を2次元で表します。MDSプロットの詳細については、『消費者調査』の「多次元尺度構成」章を参照してください。文書を選択して「テキストの表示」ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

行ごとのクラスター確率 「混合確率」表の表示／非表示を切り替えます。この表には、各クラスターに属する事後確率が行ごとに示されます。「再尤クラスター」には、この事後確率が最も高いクラスターが示されます。

確率の保存 「混合確率」表の値を、データテーブルの対応する行に保存します。

確率の計算式の保存 データテーブルに、各クラスターのスコア計算式の列と、再尤クラスターの計算式の列を追加します。

スコア計算式は、JSL関数のText Score()を使用します。なお、この際、計算方法を示す引数として、“LCA”が指定されます。

クラスターによる色分け データテーブルの行に、再尤クラスターごとに色を付けます。

削除 「テキストエクスプローラ」レポートから、「潜在クラス分析」レポートを削除します。

潜在クラス分析の詳細については、『多変量分析』の「潜在クラス分析」章を参照してください。

メモ：「テキストエクスプローラ」プラットフォームの潜在クラス分析で使われているアルゴリズムは、文書単語行列の特定の構造に基づいているため、「潜在クラス分析」プラットフォームの結果と正確には一致しません。

JMP[®] PRO 潜在意味分析（特異値分解）

潜在意味分析では、文書単語行列（DTM; Document-Term Matrix）を特異値分解（SVD; Singular Value Decomposition）します。この手法では、テキストデータを特異値分解して、少数の次元まで縮小します。潜在意味分析は、主成分分析に似ています。

DTMの特異値分解では、 $\mathbf{U}$ 、 $\mathbf{S}$ 、 $\mathbf{V}'$ の3つの行列を使ってDTMを近似します。これらの行列の関係は、次のとおりです。

$$DTM \approx \mathbf{U} * \mathbf{S} * \mathbf{V}'$$

$nDoc$ をDTMの文書数（行数）、 $nTerm$ をDTMの単語数（列数）、 $nVec$ を指定された特異値ベクトルの数とします。なお、 $nVec$ は、 $nDoc$ と $nTerm$ のうちの小さいほうの値以下になります。この式において、 $\mathbf{U}$ は $nDoc \times nVec$ の行列です。 $\mathbf{S}$ は、次元数が $nVec$ の対角行列で、 $\mathbf{S}$ の対角成分が、特異値分解の特異値となります。 $\mathbf{V}'$ は、 $nVec \times nTerm$ の行列です。 $\mathbf{V}'$ の行は、特異値ベクトルとなります。

特異ベクトルは、似たような意味やトピックをもつ単語間の関連を捉えます。たとえば、3つの単語が同じ文書でよく一緒に使用される場合は、 $\mathbf{V}'$ の特異ベクトルにおいて、その3つの単語に対応する数値がどれも大きくなります。一方、 $\mathbf{U}$ の特異ベクトルは、特異値分解で得られた単語のベクトルに文書を射影したものです。

潜在意味分析では、間接的な関連も捉えることができます。たとえば、2つの単語が同じ文書で使用されていなくても、それぞれがまた別の同じ単語とよく一緒に使用されている場合、そのような間接的な関連を捉えることができます。なお、2つの文書に共通する単語がまったくなくても、それぞれの文書に含まれる単語が縮減された次元では近くに位置すれば、それらの文書も似たような特異ベクトルに射影されます。

テキストデータを特異値分解すると、数値のベクトルに変換されますので、その変換された数値データに基づいて、クラスターリング・分類・回帰分析といった統計分析を行えるようになります。他のJMPプラットフォームの統計分析を行いたい場合には、保存に関するオプションを使って、特異値分解の結果を保存してください。

文書単語行列は、デフォルトでは、特異値分解を行う前に中心化および尺度化されます。中心化と尺度化をオフにするには、「設定」ウィンドウでそのように指定します。なお、文書単語行列の特異値分解で使われるアルゴリズムでは、中心化されている場合を含め、この行列が疎であることを最大限に活用しています。

JMP PRO 「特異値分解プロット」レポート

[潜在意味分析, 特異値分解] オプションを選択すると、2つの特異値分解プロットが描かれます。1つ目のプロットには、文書に関する、最初の2次元の特異ベクトルが描かれます。これは、**U**行列の最初の2列です。2つ目のプロットには、単語に関する、最初の2次元の特異ベクトルが描かれます。これは、**V'**行列の最初の2行です。どちらのプロットでも、点を選択できます。文書の特異値分解プロットの各点は、データテーブルの各行に対応しています。単語の特異値分解プロットの各点は、「単語」リストの各行に対応しています。

2つの特異値分解プロットの下には、特異値の表が表示されます。この特異値は、文書単語行列の特異値分解における**S**行列の対角要素です。

JMP PRO トピックの分析

[トピック分析, 特異値分解を回転] オプションを選択すると、文書単語行列を特異値分解して得られた特異ベクトルがVarimax回転されます。この際、回転させる特異ベクトルの数を指定する必要があります。ここで指定した値は、トピック数に相当します。トピックごとに単語とスコアが表示されます。このトピック分析は、因子分析に似ています。

Varimax回転は、特異ベクトルのいくつかの座標がくっきりと大きくなるように（少数の単語における数値が大きくなるように）、回転します。回転後のベクトルは、少数の単語だけが大きな値となりますので、テキストの内容を解釈しやすくなります。負の値は、該当のトピックだけにはその単語が含まれていない傾向を表します。各トピックにおいて、負の値の単語は、正の値の単語よりも少なくなっています。

JMP PRO 「トピック 単語」レポート

「トピック 単語」レポートには、以下の行列も表示されます。また、回転した特異値分解に関連する他のレポートも表示されます。

「トピック 単語」レポートには、トピックごとに、スコアの絶対値が大きい単語の表が表示されます。単語は、スコアの絶対値の降順で並べられます。これらの表は、各トピックの概念を判断するために役立ちます。

「トピック 単語」レポートには、以下の行列レポートも表示されます。

変換 Varimax回転の回転行列が表示されます。

回転後のV行列 トピックと単語スコアの行列が表示されます。回転後の**V**行列は、特異値分解の定義の**V**行列をVarimax回転した結果です。「[潜在意味分析（特異値分解）](#)」（348ページ）を参照してください。

回転後のU行列 トピックと文書スコアの行列が表示されます。トピックに対する文書のスコアが高い場合は、そのトピックとの関連が強いことを示します。

トピック 割合 各トピックの割合の値が表示されます。

JMP PRO 「トピックのスコア」レポート

「トピックのスコア」レポートには、[テキストの表示] ボタンと、トピックのスコアのプロットが表示されます。文書を選択して [テキストの表示] ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

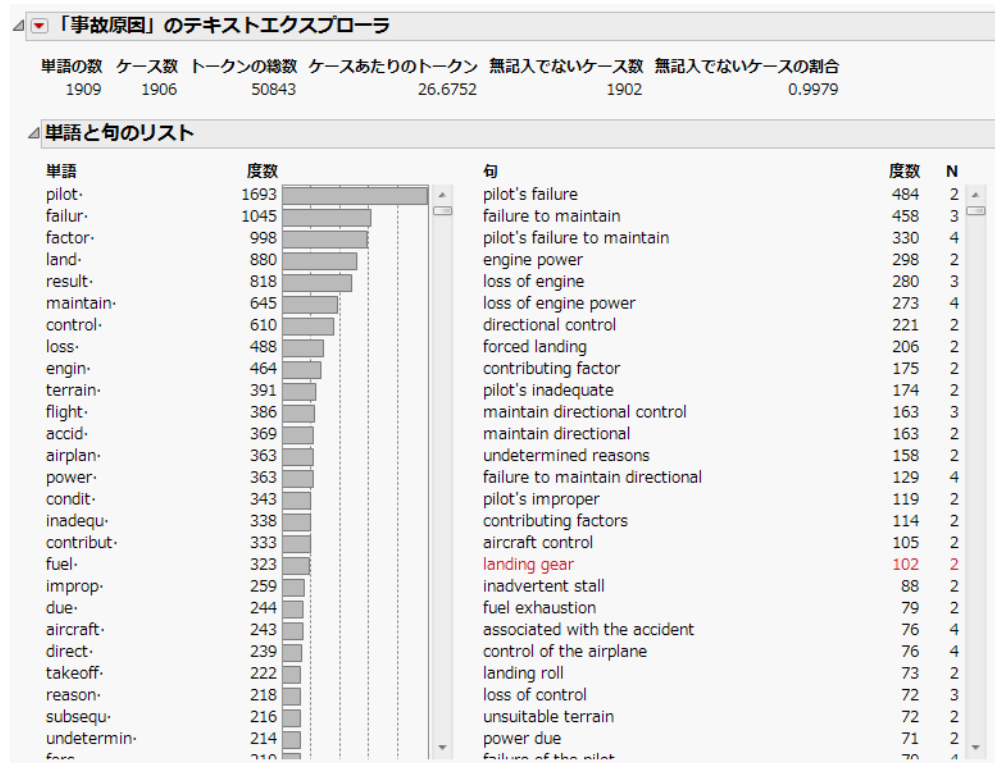
「トピックのスコア」レポートのプロットは、回転後の **U** 行列を視覚化したものです。各パネルは、トピック（回転後の **U** 行列の行）に対応しています。パネル中の各点は、コーパス中の文書（回転後の **U** 行列の列）に対応しています。

JMP PRO 「テキストエクスプローラ」プラットフォームの別例

この例では、国家運輸安全委員会による、2001年に米国で発生した航空機のインシデントレポートを使用します。まず、事故原因の調査結果に関する記述を分析します。次に、トピックごとに分類してみて、各トピックの内容を探ります。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Aircraft Incidents.jmp」を開きます。
2. [行] > [列の値による色/マーカー分け] を選択します。
3. 列のリストで「**死亡者**」を選択し、[OK] をクリックします。
致命的な事故の行が、赤で示されます。
4. [分析] > [テキストエクスプローラ] を選択します。
5. 「列の選択」リストで「**事故原因**」を選択し、[テキスト列] をクリックします。
6. 「語幹抽出」リストから [すべての単語] を選択します。
7. 「トークン化」リストから [基本ワード] を選択します。
8. [OK] をクリックします。

図12.12 「事故原因」のテキストエクスプローラ



レポートによると、約51,000個のトークンと、約1,900個の異なる単語が含まれていることがわかります。

- 「単語」リストの上位3つを選択します。

全部で約51,000個のトークン中、これらの3つの単語は約2%を占めています。

- 選択した単語を右クリックして、[ストップワードの追加]を選択します。

この3つの単語は、他の単語に比べ頻繁に出現しているため、文書間の違いを読み取るためには役立ちません。

- JMP PRO** 「「事故原因」のテキストエクスプローラ」レポートの赤い三角ボタンをクリックし、[潜在意味分析，特異値分解]を選択します。

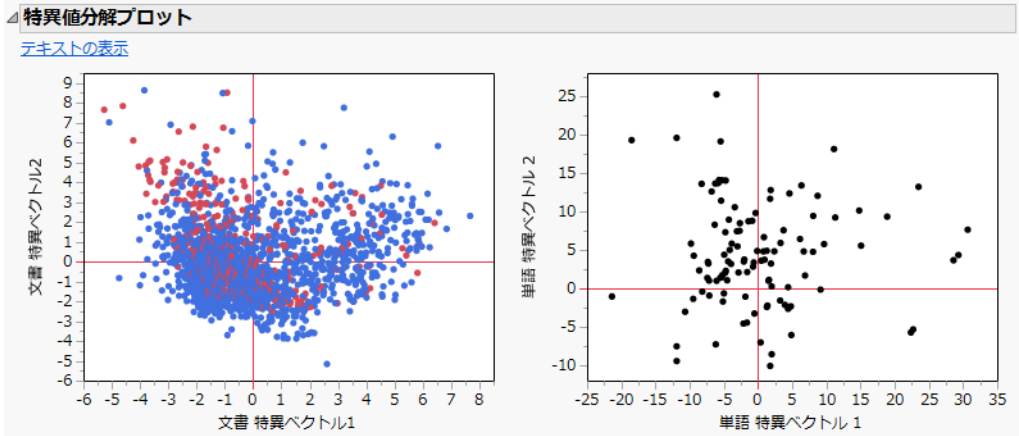
これは、特異値分解の結果を回転するトピック分析を行う前の、最初のステップです。

- JMP PRO** 「設定」ウィンドウで、「単語の最小度数」に「50」と入力します。

約51,000個の全トークン中、少なくとも0.1%の割合を占める単語を抽出します。

- JMP PRO** [OK] をクリックします。

図12.13 事故原因の特異値分解プロット



文書の特異値分解プロットでは、「fatal」と「non-fatal」の違いはあまり見られません。

14. **JMP PRO** 「テキストエクスプローラ」レポートの赤い三角ボタンから、[トピック分析, 特異値分解を回転]を選択します。

単語のグループから形成されるトピックを表示します。

15. **JMP PRO** 「トピックの個数」に「5」と入力します。
16. **JMP PRO** [OK] をクリックします。

図12.14 事故原因の「トピック 単語」レポート

▲トピック 単語

トピック1		トピック2		トピック3		トピック4		トピック5	
単語	スコア	単語	スコア	単語	スコア	単語	スコア	単語	スコア
power.	0.33885	factor.	0.2592	altitud.	0.2607	fuel.	0.2780	experi.	0.28340
loss.	0.33317	condit.	0.2371	low.	0.2427	personnel.	0.2558	control.	0.28041
engin.	0.30801	unsuit.	0.2303	dark.	0.2166	mainten.	0.2508	direct.	0.26311
forc.	0.30333	select.	0.2139	clearanc.	0.2107	result.	0.2289	lack.	0.25283
suitabl.	0.29414	associ.	0.2098	night.	0.2080	preflight.	0.2281	pilot.	0.23436
lack.	0.27680	failur.	-0.2048	instrument.	0.2033	exhaust.	0.2076	student.	0.21350
reason.	0.24828	area.	0.2008	pilot.	0.2031	reason.	-0.2065	maintain.	0.20836
undetermin.	0.24187	accid.	0.1954	maintain.	0.1968	undetermin.	-0.2046	factor.	0.18982
terrain.	0.18186	compens.	0.1899	airspe.	0.1739	inspect.	0.2042	aircraft.	0.18800
total.	0.16480	wind.	0.1755	stall.	0.1716	plan.	0.2041	failur.	0.18485
contribut.	0.14370	result.	-0.1685	maneuv.	0.1705	improp.	0.1938	instructor.	0.15134
		airspe.	-0.1655	continu.	0.1652	inadequ.	0.1917	crosswind.	0.14904
		inadequ.	0.1617	condit.	0.1597	subsequ.	0.1793	total.	0.14800
		stall.	-0.1613	land.	-0.1532	maintain.	-0.1683	airplan.	0.14734
		maintain.	-0.1524	weather.	0.1490				

各トピック中の最もスコアが高い単語は、インシデントレポートの主旨を見極める手がかりとなります。

たとえば、トピック1では、「power」、「loss」、「engine」という単語が高いスコアを記録しており、エンジンの出力低下に起因する事故であることが示唆されています。これは、「loss of engine power」という句が、インシデントレポート全体に273回出現しているのに対応しています。

トピック3では、「low」、「dark」といった単語が高いスコアを記録しており、暗さや低空と関連する事故であることが示唆されています。

この後にも、様々な分析が考えられます。テキスト分析は、反復的なプロセスです。トピック情報を参考に、ストップワードを追加したり、特定の句を単語に指定するなどして、単語リストを調整してもよいでしょう。また、文書単語行列・特異ベクトル・回転した特異ベクトルをデータテーブルに保存して、JMPの他のプラットフォームで分析を進めてもよいでしょう。そのとき、テキスト分析で得られた情報と、データにある他の情報との関係を調べてみるのもよいでしょう。

付録 **A**

統計的詳細 **基本的な統計分析**

この付録には、JMP Proの複数のプラットフォームに共通する統計的詳細を記載します。

検証の機能を使用できるプラットフォーム

次の表に、各プラットフォームで使用できる交差検証の種類を示します。

親プラットフォーム	プラットフォーム	行の除外による 検証	無作為抽出による 検証	K分割交差検証	検証列
モデルの あてはめ	最小2乗法による あてはめ	×	×	×	○（モデルの 評価のみ）
モデルの あてはめ	変数増加法による ステップワイズ 回帰	×	×	○（応答変数が 連続尺度のモデ ルのみ）	○
モデルの あてはめ	ロジスティック 回帰	×	×	×	○（モデルの 評価のみ）
モデルの あてはめ	一般化回帰	×	○	○	○
モデルの あてはめ	PLS回帰	×	○	○	○
パーティション	ディシジョン ツリー	○	○	○	○
パーティション	ブートストラップ 森	○	○	×	○
パーティション	ブースティング ツリー	○	○	×	○
パーティション	K近傍法	○	○	×	○
パーティション	単純ベイズ	○	○	×	○
ニューラル	ニューラル	○	○	○	○

-
- Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, New York: John Wiley and Sons, Inc.
- Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley and Sons, Inc.
- Agresti, A., and Coull, B. (1998), "Approximate is Better Than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119–126
- Akaike, H. (1974), "Factor Analysis and AIC," *Psychometrika*, 52, 317–332.
- Akaike, H. (1987), "A new Look at the Statistical Identification Model," *IEEE Transactions on Automatic Control*, 19, 716–723.
- American Society for Quality Statistics Division (2004), *Glossary and Tables for Statistical Quality Control*, Fourth Edition, Milwaukee: Quality Press. "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," *JRSS Suppl* 8, 128–138.
- Baglama, J. and Reichel, L. (2005), "Augmented Implicitly Restarted Lanczos Bidiagonalization Methods," *SIAM Journal on Scientific Computing* 27, 19–42.
- Bartlett, M.S. (1966), *An Introduction to Stochastic Processes*, Second Edition, Cambridge: Cambridge University Press.
- Bissell, A. F. (1990), "How Reliable is Your Capability Index?," *Applied Statistics*, 30, 331–340.
- Brown, M.B. and Benedetti, J.K. (1977), "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables," *Journal of the American Statistical Association* 72, 305–315.
- Brown, M.B. and Forsythe, A.B. (1974), "The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means," *Technometrics* 16:1, 129–132.
- Brown, M.B. and Forsythe, A.B. (1974), "Robust tests for the equality of variances," *Journal of the American Statistical Association*, 69, 364–367.
- Chen, Song Xi and Hall, Peter. (1993), "Empirical Likelihood Confidence Intervals for Quantiles," *The Annals of Statistics* 21(3): 1166–1181.
- Chou, Y.M., Owen, D.B., and Borrego, S.A. (1990), "Lower confidence limits on process capability indices," *Journal of Quality Technology*, 22(3): 223–229.
- Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368): 829–836.
- Cleveland, W.S. (1994). *Visualizing Data*, Summit, N.J: Hobart Press.
- Cohen, J. (1960), "A coefficient of agreement for nominal scales," *Education Psychological Measurement*, 20: 37–46.
- Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*, Second Edition, New York: John Wiley and Sons.

- Conover, W. J. (1972). "A Kolmogorov Goodness-of-fit Test for Discontinuous Distributions". *Journal of the American Statistical Association* 67: 591–596.
- Conover, W.J. (1980), *Practical Nonparametric Statistics*, New York: John Wiley and Sons, Inc.
- Conover, W.J. (1999), *Practical Nonparametric Statistics*, 3rd Edition, New York: John Wiley and Sons, Inc.
- Cureton, E.E. (1967), "The Normal Approximation to the Signed-Rank Sampling Distribution when Zero Differences are Present," *Journal of the American Statistical Association*, 62:319, 1068–1069.
- DeLong, E., Delong, D, and Clarke-Pearson, D.L. (1988), "Comparing the Areas Under Two or more Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics* 44, 837–845.
- Devore, J. L. (1995), *Probability and Statistics for Engineering and the Sciences*, Duxbury Press, CA.
- Dunn, O.J. (1964), "Multiple Comparisons Using Rank Sums," *Technometrics* 6, 241–252.
- Dunnnett, C.W. (1955), "A multiple comparison procedure for comparing several treatments with a control" *Journal of the American Statistical Association*, 50, 1096–1121.
- Dwass, M. (1955), "A Note on Simultaneous Confidence Intervals," *Annals of Mathematical Statistics* 26: 146–147.
- Efron, B. (1981), "Nonparametric Standard Errors and Confidence Intervals," *The Canadian Journal of Statistics* 9:2, 139–158.
- Eubank, R.L. (1999), *Nonparametric Regression and Spline Smoothing*, Second Edition, Boca Raton, Florida: CRC.
- Fisher, L. and Van Ness, J.W. (1971), "Admissible Clustering Procedures," *Biometrika*, 58, 91–104.
- Fleiss, J.L., Cohen J., and Everitt, B.S. (1969), "Large-Sample Standard Errors of Kappa and Weighted Kappa," *Psychological Bulletin*, 72: 323–327.
- Friendly, M. (1991), "Mosaic Displays for Multiway Contingency Tables," *New York University Department of Psychology Reports*: 195.
- Golub, G.H. and van der Vorst, H.A., (2000), "Eigenvalue Computation in the 20th Century," *Journal of Computational and Applied Mathematics* 123, 35–65.
- Goodman, L.A. and Kruskal, W.H. (1979), *Measures of Association for Cross Classification*, New York: Springer-Verlag (reprint of JASA articles).
- Gupta, S.S. (1965), On Some Multiple Decision (selection and ranking), Rules., *Technometrics* 7, 225–245.
- Hahn, G. J. and Meeker, W. Q. (1991) *Statistical Intervals: A Guide for Practitioners*. New York: Wiley.
- Hajek, J. (1969), *A Course in Nonparametric Statistics*, San Francisco: Holden-Day.
- Hartigan, J.A. and Kleiner, B. (1981), "Mosaics for Contingency Tables," *Proceedings of the 13th Symposium on the Interface between Computer Science and Statistics*, Ed. Eddy, W. F., New York: Springer-Verlag, 268–273.

- Hayter, A.J. (1984), "A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative," *Annals of Mathematical Statistics*, 12 61-75.
- Hosmer, D.W. and Lemeshow, S. (1989), *Applied Logistic Regression*, New York: John Wiley and Sons.
- "Hot Dogs," (1986), *Consumer Reports* (June), 364-367.
- Hsu, J. (1981), "Simultaneous confidence intervals for all distances from the 'best'," *Annals of Statistics*, 9, 1026-1034.
- Hsu, J. (1989), *Tutorial Notes on Multiple Comparisons*, American Statistical Association, Washington, DC.
- Hsu, J.C. (1996), *Multiple Comparisons: Theory and Methods*, Chapman and Hall.
- Huber, Peter J. (1973), "Robust Regression: Asymptotics, Conjecture, and Monte Carlo," *Annals of Statistics*, Volume 1, Number 5, 799-821.
- Huber, P.J. and Ronchetti, E.M. (2009), *Robust Statistics*, Second Edition, Wiley.
- Iman, R.L. (1974), "Use of a t-statistic as an Approximation to the Exact Distribution of Wilcoxon Signed Ranks Test Statistic," *Communications in Statistics—Simulation and Computation*, 795-806.
- Kendall, M. and Stuart, A. (1979), *The Advanced Theory of Statistics, Volume 2*, New York: Macmillan Publishing Company, Inc.
- Kramer, C.Y. (1956), "Extension of multiple range tests to group means with unequal numbers of replications," *Biometrics*, 12, 309-310.
- Lehmann, E.L. (2006), *Nonparametrics: Statistical Methods Based on Ranks*, 2nd edition, Springer.
- Levene, H. (1960), "Robust tests for the equality of variances" In I. Olkin (ed), *Contributions to probability and statistics*, Stanford Univ. Press.
- Mason, R.L. and Young, J.C. (2002), *Multivariate Statistical Process Control with Industrial Applications*, Philadelphia: ASA-SIAM.
- McCullagh, P. and Nelder, J.A. (1983), *Generalized Linear Models*, London: Chapman and Hall Ltd.
- Meeker, W.Q. and Escobar, L.A. (1998), *Statistical Methods for Reliability Data*, pp. 60-62, New York: John Wiley and Sons.
- Nagelkerke, N.J.D. (1991), "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, 78:3 691-692.
- Nelson, P.R., Wludyka, P.S., and Copeland, K.A.F. (2005), *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*, Philadelphia: Society for Industrial and Applied Mathematics.
- Neter, J., Wasserman, W. and Kutner, M.H. (1990), *Applied Linear Statistical Models*, Third Edition, Boston: Irwin, Inc.
- O'Brien, R.G. (1979), "A general ANOVA method for robust tests of additive models for variances," *Journal of the American Statistical Association*, 74, 877-880.

- O'Brien, R., and Lohr, V. (1984), "Power Analysis For Linear Models: The Time Has Come," *Proceedings of the Ninth Annual SAS User's Group International Conference*, 840–846.
- Olejnik, S.F. and Algina, J. (1987), "Type I Error Rates and Power Estimates of Selected Parametric and Nonparametric Tests of Scale," *Journal of Educational Statistics* 12, 45–61.
- Pratt, J.W. (1959), "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures", *Journal of the American Statistical Association*, 54:287, 655–667.
- Reinsch, C.H. (1967), *Smoothing by Spline Functions*, *Numerische Mathematik*, 10, 177–183.
- Rousseuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: John Wiley and Sons.
- Rubin, D. (1981), "The Bayesian Bootstrap", *The Annals of Statistics*, 9:1, 130–134.
- SAS Institute (2008), *SAS/STAT 9.2 User's Guide Introduction to Nonparametric Analysis*, Chapter 16, SAS Institute Inc., Cary NC. Retrieved May 27, 2016 from <https://support.sas.com/documentation/cdl/en/statugnonparam/61764/PDF/default/statugnonparam.pdf>.
- Schafer, J. and Strimmer, K. (2005), "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics," *Statistical Applications in Genetics and Molecular Biology*, 4:1, 1–30.
- Slifker, J. F. and Shapiro, S. S. (1980). *Technometrics*, 22, 239–246.
- Snedecor, G.W. and Cochran, W.G. (1980), *Statistical Methods*, 7th edition, Ames, Iowa: Iowa State University Press.
- Somers, R.H. (1962), "A New Asymmetric Measure of Association for Ordinal Variables," *American Sociological Review*, 27, 799–811.
- Tan, Charles Y., and Iglewicz, Boris (1999), "Measurement-methods Comparisons and Linear Statistical Relationship," *Technometrics*, 41:3, 192–201.
- Tamhane, A. C. and Dunlop, D. D. (2000) *Statistics and Data Analysis*, Prentice Hall.
- Tukey, J. (1953), "A problem of multiple comparisons," Dittoed manuscript of 396 pages, Princeton University.
- Welch, B.L. (1951), "On the comparison of several mean values: an alternative approach," *Biometrika* 38, 330–336.
- Wheeler, D.J. (2003), *Range Based Analysis of Means*, SPC Press, Knoxville, TN.
- Wilson, E. B. (1927), "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, 22, 209–212.
- Wludyka, P.S. and Nelson, P.R. (1997), "An Analysis-of-Means-Type Test for Variances From Normal Populations", *Technometrics*, 39:3, 274–285.
- Wludyka, P.S. and Nelson, P.R. (1999), "Two Non-Parametric, Analysis-of-Means-Type Tests for Homogeneity of Variances," *Journal of Applied Statistics*, 26:2, 243–256.

記号

α 水準の設定オプション
一元配置 [141](#), [152](#)
二変量 [122](#)
分割表 [209](#), [215](#)

数字

5%の等高線オプション [122](#)

A

ANOM
順位変換平均分析 [151](#)
平均分析も参照
ANOMV の表示オプション [152](#)

B

Bartlett の検定 [165](#)
Brown-Forsythe の検定 [165](#)

C

Cochran-Armitage の傾向検定 [210](#), [220](#)
Cochran-Mantel-Haenszel 検定 [209](#), [216–217](#)

F

Fisher の正確検定 [215](#)
Flash(.SWF) 形式で保存オプション [45](#)

G

GLog のあてはめ [61](#)

J

Johnson のあてはめ [60](#)
Johnson 分布のあてはめ [85–86](#)

K

Kolmogorov Smirnov 検定 [159](#)
Kruskal-Wallis 検定
Wilcoxon 検定を参照
K シグマから仕様限界を設定オプション [62](#)

L

Levene の検定 [165](#)
LSD 閾値行列オプション [158](#)

M

Mann-Whitney 検定
Wilcoxon 検定を参照

N

N と欠測値 N (要約統計量) [263](#)

O

O'Brien の検定 [165](#)

P

Poisson 分布のあてはめ [88](#)

R

ROC 曲線オプション [243](#)

S

Student の t 検定 [154](#), [156–175](#)

T

t 検定
オプション [140](#)
レポート [146–147](#)

V

van der Waerden の検定 159

W-Z

Weibull 分布のあてはめ 60, 83

Welch の検定 166

Wilcoxon

検定 159

ペアごとの検定 163

X, グループ化カテゴリボタン 95

X, グループ変数ボタン 95

X, 説明変数ボタン 95

X, 連続量の説明変数ボタン 95

Y, 応答カテゴリボタン 95

Y, カテゴリカル応答変数ボタン 95

Y→Xのあてはめオプション 115

Y軸スケールを分散にオプション 153

ア

値ごとのあてはめ

コマンド 114

メニュー 120-121

レポート 114

値の色を選択ウィンドウ 211

値の順序を各列に保存オプション 216

あてはまりの悪さ (LOF) レポート 107-108

あてはめ線オプション 63, 121, 170

あてはめたモデルの分位点を保存コマンド 62

あてはめの削除オプション 62, 121

あてはめの要約レポート 106, 145

イ

一元配置プラットフォーム 135-136

オプション 139-143

起動 138

プロット 138

例 136-137

一変量の分布の保存コマンド 56-57

一変量の分布プラットフォーム 33

オプション 44-57

カテゴリカル変数 34

起動 36

例 34-36

レポートウィンドウ 37-38

連続尺度の変数 34

一変量分散, 主成分オプション 115

一致性の統計量オプション 209, 217, 231

一般化対数分布のあてはめ 87

色の設定オプション 211

ウ

内側の点を選択オプション 122

オ

応答水準の切り替えオプション 215

オッズ比オプション 209, 243

カ

カーネルの制御オプション 122

回帰の信頼区間オプション 121

回帰の信頼区間内を塗るオプション 122

解析のオプション、テキストエクスプローラ 338

回転オプション 63

各水準の平均レポート 148

各ペア, Student の t 検定 154, 156-175

確率楕円オプション 117-118

確率の計算式の保存オプション 243

カテゴリ数 (要約統計量) 263

カラーテーマオプション 212

ガンマ Poisson 分布のあてはめ 88-89

ガンマ分布のあてはめ 60, 84

幹葉図 52

関連の指標オプション 210, 218-219

キ

逆推定オプション 243

行テーブル 267

行プロファイル 216

極値分布のあてはめ 83

許容区間オプション 57-70, 77-78

ク

区切り文字の表示、テキストエクスプローラ 338

クラスター 単語、テキストエクスプローラ 342

クラスター 文書、テキストエクスプローラ 342
グループ別オプション 100, 119
グループ変数の欠測値を含めるオプション 261
グループ変数の度数順オプション 261

ケ

ケースあたりのトークン、テキストエクスプローラ 334
ケース数、テキストエクスプローラ 334
形式の変更オプション 261
係数の保存オプション 122
決定限界の陰影の表示オプション 153
決定限界の表示オプション 153
言語処理、テキストエクスプローラ 328
検出力オプション 141, 168–169, 200
検定
 分割表のオプション 209
 レポート 214–215

コ

工程能力分析オプション 58–59, 78–81
項目ラベルの変更 268
語幹抽出、テキストエクスプローラ 322, 328
語幹レポートの表示、テキストエクスプローラ 338
個別の値に対する信頼区間オプション 121
個別の値に対する信頼区間内を塗るオプション 122
コントロール群との比較 (Dunnett) 検定 154, 158
コントロール群との比較 Steel 検定 163

サ

最小値 (要約統計量) 263
最大値 (要約統計量) 263
最適値との比較 (Hsu の MCB) 検定 154, 157–178
差の行列オプション 158
差の順位レポート 158
サブセットオプション 38
残差の保存オプション 121
残差プロットオプション 122
三次元対応分析オプション 216–224

シ

指数分布のあてはめ 60, 83

実測値-分位点プロットオプション 170
指定された分散比オプション 115
「シミュレーション」プラットフォーム
 データテーブルの結果 302
シャドウグラムオプション 48
順序オプション 46
仕様限界オプション 62, 64
仕様限界の保存コマンド 62
詳細な比較レポート 158
詳細レポート 216
診断プロットオプション 61–63
信頼区間オプション 46, 56
信頼区間の分位点オプション 158
信頼限界オプション 63

ス

推定値の共分散レポート 242
数字を単語として扱う、テキストエクスプローラ 329
スケールの統一オプション 44
ストップワード、テキストエクスプローラ 322, 340
スプライン曲線のあてはめコマンド 112–114, 131
すべてのグラフオプション 142–143
すべての分布のあてはめ 61–87
すべてのペア Steel-Dwass 検定 163
すべてのペア, Tukey の HSD 検定 154, 156–177

セ

正確検定 210, 220
 オプション 160
 Fisher の正確検定も参照
正規混合分布のあてはめ 60, 85
正規表現エディタ、テキストエクスプローラ 329
正規分位点プロット 49–50, 141, 170
正規分布のあてはめ 60, 82
設定パネルの表示オプション 269
セルのラベルオプション 211
潜在意味分析, 特異値分解、テキストエクスプローラ 342
潜在クラス分析、テキストエクスプローラ 342, 347
線種オプション 121
全体平均オプション 143
選択した点の間でグラデーションオプション 212

線の色オプション 121

線の幅オプション 121

ソ

関連レポート 117

相対リスクオプション 209, 218

外側の点を選択オプション 122

その他のあてはめコマンド 102, 111–112

タ

対応のある列を設定オプション 142, 171

対応分析オプション 209, 215

対応を表す線オプション 143, 171

対応を表す点線オプション 143

対数正規分布のあてはめ 60, 82

楕円内を塗るオプション 122

多項式のあてはめコマンド 101, 104–133

多項式のあてはめ次数メニュー 120

多項式のあてはめレポート 105–133

縦方向に表示オプション 46, 48

単語（トークン）、テキストエクスプローラ 322

単語のオプション、テキストエクスプローラ 338

単語の区切り文字のリスト、テキストエクスプローラ 332

チ

チャートの表示オプション 269

中央値の参照線オプション 63

中央値（メディアン）（要約統計量） 264

中心線の表示オプション 153

直線のあてはめ

メニュー 120–122

レポート 105–133

直線のあてはめコマンド 101, 104–110, 131

直交回帰 分散比メニュー 120–121

直交回帰レポート 115

直交のあてはめコマンド 115–116, 131

ツ

ツールヒントの表示オプション 269

積み重ねて表示オプション 44

テ

データテーブルに出力オプション 269

適合度

オプション 62

検定 63–91

ケースあたりのトークン、テキストエクスプローラ 334

テキストエクスプローラ

解析のオプション 338

区切り文字の表示 338

クラスター 単語 342

クラスター 文書 342

ケース数 334

言語処理 328

語幹抽出 322, 328

語幹レポートの表示 338

ストップワード 322, 340

正規表現の編集 329

潜在意味分析, 特異値分解 342

潜在クラス分析 342, 347

単語（トークン） 322

単語のオプション 338

単語の区切り文字のリスト 332

トークン化 328

トークンの総数 334

特異値分解 散布図行列 342

トピックの散布図行列 343

トピック分析, 特異値分解を回転 342

プラットフォームのオプション 337

無記入でないケース数 334

無記入でないケースの割合 334

テスト集計パネルの表示オプション 269

点オプション 142

点の拡散オプション 143

点の表示オプション

二変量 100

点をずらす 246

点をずらすオプション 143

ト

トークン化、テキストエクスプローラ 328

トークンの総数、テキストエクスプローラ 334

等高線オプション 122
等高線間を塗りつぶすオプション 122
同等性の検定オプション 55, 141, 167-185
等分散性オプション 115
等分散性の検定 141, 165
特異値分解 散布図行列、テキストエクスプローラ 342

度数

オプション 45
レポート 41-42
度数軸オプション 46, 48
度数の表示オプション 46, 49
トピックの散布図行列、テキストエクスプローラ 343
トピック分析、特異値分解を回転、テキストエクスプローラ 342

二

二項分布のあてはめ 89
二変量正規楕円メニュー 120-122
二変量の回帰のあてはめ 102
二変量の関係プラットフォーム 93
二変量プラットフォーム 97
オプション 100-103, 120-123
起動 98
例 98
レポートウィンドウ 99

ノ

濃淡の表示オプション 269
ノンパラメトリック
オプション 141
検定 159
ノンパラメトリック二変量の密度レポート 119
ノンパラメトリック密度コマンド 118-119

ハ

パーセントの表示オプション 46, 48
箱ひげ図 142
外れ値の箱ひげ図 50-51
パラメータ推定値レポート 109-133, 242
パラメトリックな標本再抽出、「シミュレーション」プラットフォーム 279

範囲、要約統計量 263
反復回数レポート 240

ヒ

比較円 143, 155-156, 198-200
ヒストグラム 38-65
一元配置のオプション 143
色オプション 46, 48
再スケール 39
サイズの変更 39
サブセットの作成 38
軸オプション 100, 103
データの強調表示 38, 40
データの指定 39
データの選択 41
連続尺度の変数のオプション 48-49
ヒストグラムオプション 48
表示オプション 142-143, 153, 215
カテゴリカル変数 45
連続尺度の変数 47
標準誤差 264
標準誤差バーオプション 46, 48
標準偏差 263
標準偏差線オプション 143, 149
標準偏差と標準誤差（要約統計量）263
標準偏差の検定オプション 54-55
表の作成 255-277
表の表示オプション 269
標本サイズに比例したX軸オプション 143, 148-149
標本サイズ、表 270

フ

ブートストラップ 305-319
ブロック平均レポート 148
ブロックボタン 95
プロットスケールの統一オプション 261
プロットスケールの変更オプション 261
プロットのオプションメニュー、ロジスティック 243
分位点
オプション 47, 62, 72, 140, 144-145
レポート 42
分位点-実測値プロットオプション 170

分位点の箱ひげ図 51–52
分位点密度等高線メニュー 120–123
分割表 209, 212–214
分割表プラットフォーム 205
 オプション 209–210
 起動 207
 例 206–207
 レポートウィンドウ 208
分散分析
 レポート 108–109, 147–148
 一元配置も参照
分散（要約統計量）263
文書単語行列、テキストエクスプローラ 322
分布パラメータの指定オプション 62

へ

ベータ二項分布のあてはめ 89–90
ベータ分布のあてはめ 60, 84
平滑化スプライン曲線によるあてはめ
 メニュー 120
 レポート 113
平滑曲線のあてはめ 60, 85
平均/ANOVA オプション 140, 145
平均/ANOVA/プーリングしたt検定オプション 140, 145
平均誤差バーオプション 143, 149
平均線オプション 143, 149
平均と標準偏差オプション 140
平均のあてはめ
 コマンド 101, 103–104
 メニュー 120–121
 レポート 104
平均の検定オプション 53–54
平均の信頼区間オプション 143
平均の比較のオプション 141, 153–159
平均のひし形オプション 143, 148–149
平均の平均オプション 143
平均分析
 オプション 152–153
 チャート 152–172
 分散 172–173
 割合オプション 209, 215
平均分析法 141, 150–153

平均をつなぐオプション 143–144
変換されたあてはめ
 メニュー 120
 レポート 112
変換の計算式を保存コマンド 62
変換または制約を指定ウィンドウ 111, 124
変動係数 264

ホ

棒の間を離すオプション 46
棒の幅の設定オプション 48
保存オプション
 一元配置 142

マ

マクロオプション 212

ミ

密度関数の保存コマンド 62
密度曲線オプション 62
密度グリッドの保存オプション 123
密度軸オプション 46, 48
密度推定 102
密度による点の選択オプション 122
密度のオプション 142, 170
密度の構成オプション 170
密度の比較オプション 170
密度の割合オプション 170
密度分位点による色分けオプション 122
密度分位点の保存オプション 122
峰型クラスターオプション 122

ム

無記入でないケース数、テキストエクスプローラ 334
無記入でないケースの割合、テキストエクスプローラ 334

メ

メッシュプロットオプション 122
メディアン検定 159

モ

モザイク図

一変量の分布のオプション 46

分割表 209–212

文字の接続レポート 158

モデル全体の検定レポート 240–242

元の色に戻すオプション 212

ヨ

要約統計量

オプション 44

カスタマイズ 47

レポート 42, 47

要約レポートの表示オプション 153, 215

横に並べるオプション 45, 47

予測区間オプション 57–68, 76

予測値の保存オプション 121

リ

離散分布のあてはめのオプション 61–90

リフトチャートオプション 243

両端の間でグラデーションオプション 212

ル

累積確率プロット 142, 170

レ

列テーブル 267

列の色を保存オプション 212

列ラベルの削除 268

列ラベルの復元 268

レポートオプション 121

連続分布のあてはめのオプション 60–87

ロ

ロジスティック回帰 235–236

ロジスティックプラットフォーム 235

オプション 243–245

起動 238

推定値の共分散レポート 242

パラメータ推定値レポート 242

反復回数レポート 240

モデル全体の検定レポート 240–242

例 237–238, 245, 249

レポートウィンドウ 239–242

ロジスティックプロット 240, 243

ロジスティック回帰も参照

ワ

割合軸オプション 46, 48

割合の2標本検定 209, 218

割合の検定オプション 46–67

