



バージョン13

はじめてのJMP

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

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

このマニュアルを引用する場合は、次の正式表記を使用してください: SAS Institute Inc. 2016.『JMP 13[®] はじめての JMP』 Cary, NC: SAS Institute Inc.

JMP 13[®] はじめての JMP

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-62960-473-2 (印刷版)

ISBN 978-1-62960-570-8 (EPUB)

ISBN 978-1-62960-571-5 (MOBI)

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

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

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

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

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

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

2016年9月

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

他のブランド名および製品名は、それぞれの会社の商標です。

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

テクノロジーライセンスに関する通知

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

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

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

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

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

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

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

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

このソフトウェアは、著作権保有者および貢献者によって「現状のままで」提供され、商業性および特定の目的に対する適合性に関する默示的な保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。いかなる場合においても、著作権保有者または貢献者は、損害の原因が何で

あれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

- iODBC ソフトウェア - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). 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默示的な保証を含むがそれに限らない、いかなる明示的もしくは默示的な保証も行われません。いかなる場合においても、OPENLINKまたは貢献者は、損害の原因が何であれ、そして法的責任の根拠が何であれ、つまり、契約、厳格責任、不法行為（過失その他を含む）のどれであれ、かかる損害の発生する可能性を事前に知らされていたとしても、このソフトウェアをどのように使用して生じた損害であれ、いかなる直接損害、間接損害、付随的損害、特別損害、懲罰的損害、もしくは結果損害（代替品または代替サービスの調達、使用機会、データもしくは利益の損失、業務の中止を含むがそれに限らない）に対しても責任を負いません。

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

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

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

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

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

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

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

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

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

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

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

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

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

これをもって、このソフトウェアのコピーと関連する文書ファイル（「本ソフトウェア」）を入手した人すべてに対し、無料で本ソフトウェアを使用、コピー、変更、マージ、パブリッシュ、配布、サプライセンスする、もしくはコピーを販売する権利を含むがそれに限らず、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、業務の中止を含むがそれに限らない）に対しても責任を負いません。

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

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

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

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

JMP® を最大限に活用しよう

初めてのユーザも経験豊かなユーザも、JMPについて常に学べることができます。

JMP.com には次のものが用意されています。

- JMP の初心者に役立つライブまたは録画の Web キャスト
- 新機能および高度な技術に関するビデオデモと Web キャスト
- JMP トレーニングに登録するための詳細情報
- ユーザの地域でのセミナー開催スケジュール
- JMP の成功事例
- JMP スタッフによるヒントやコツ、ストーリーが満載のブログ
- JMP について他のユーザとディスカッションができるフォーラム

http://wwwjmp.com/ja_jp/getstarted/

目次

はじめてのJMP

このガイドについて	13
JMP グラフギャラリー	15
1 JMP の概要	
マニュアルとその他のリソース	35
表記規則	36
JMP のマニュアル	37
JMP ドキュメンテーションライブラリ	37
JMP ヘルプ	43
JMP を習得するためのその他のリソース	43
チュートリアル	43
サンプルデーターテーブル	44
統計用語とJSL用語の習得	44
JMPを使用するためのヒント	44
ツールヒント	45
JMP User Community	45
JMPer Cable	45
JMP関連書籍	45
「JMPスター」ウィンドウ	46
テクニカルサポート	46
2 JMP の基本	
基本概念	47
理解すべき基本操作	48
使い方入門	48
JMPの起動	49
サンプルデータの使用	51
データターブルの理解	51
JMP での作業手順の理解	53

ステップ1: プラットフォームを起動して結果を表示する	54
ステップ2: 箱ひげ図を削除する	56
ステップ3: 追加の出力を要求する	56
ステップ4: プラットフォームの結果を操作する	57
JMP と Excel の相違点	58

3 データの操作

グラフ作成や統計分析に向けたデータ準備	61
データの取り込み	62
データのコピーと貼り付け	62
データの読み込み	62
データの入力	65
Excelからのデータの転送	67
データテーブルの操作	68
データの編集	69
値の選択、選択解除、検索	70
列情報の表示または変更	74
計算式による値の計算	75
データのフィルタ処理	77
データの加工	79
要約統計量の表示	79
サブセットの作成	83
データテーブルの結合	85
テーブルの並べ替え	87

4 データの視覚化

共通のグラフ	89
一変量の分析	90
ヒストグラム	90
棒グラフ	92
多変量の考察	94
散布図	95
散布図行列	99
並列箱ひげ図	102
重ね合わせプロット	105

変動性図	108
グラフビルダー	111
バブルプロット	117
5 データの分析	
一変量の分布、二変量の関係、モデル	123
この章について	124
データからグラフを作成することの重要性	124
尺度の理解	127
例: 尺度に基づく結果	127
尺度の変更	129
「一変量の分布」による分析	130
連続変数の分布	130
カテゴリカル変数の分布	133
二変量の関係	136
1つの予測変数による回帰分析	136
1変数による平均の比較	141
割合の比較	144
複数の変数による平均の比較	146
複数の予測変数による回帰分析	150
6 全体像	
複数のプラットフォームを使ったデータの探索	157
はじめる前に	158
複数プラットフォームでのデータの探索	158
「一変量の分布」による分析	158
パターンと関係の分析	162
似たようなデータをグループにまとめる分析	165
7 作業結果の保存と共有	
結果の保存と再現	171
プラットフォーム結果のジャーナルへの保存	172
ジャーナルの作成例	172
ジャーナルへの分析の追加	173
プロジェクトの作成	173
スクリプトの保存と実行	175

スクリプトの保存と実行の例	175
スクリプトとJSLについて	176
インタラクティブHTMLとして保存	177
データを含むインタラクティブHTML	178
インタラクティブHTMLの作成例	178
レポートをPowerPointプレゼンテーションとして保存	179
プロファイル、バブルプロット、一変量の分布のAdobe Flashファイル作成	180
バブルプロットのAdobe Flashバージョンの保存例	180
ダッシュボードの作成	183
ウィンドウの結合例	183
2つのレポートのあるダッシュボードの作成例	184
ダッシュボードをアドインとして保存	185
8 特別な機能	
分析の自動更新やSASインテグレーション	187
分析とグラフの自動更新	188
自動再計算の使用例	188
環境設定の変更	192
環境設定の変更例	193
JMPとSASの連携	196
SASコードの作成例	196
SASコードのサブミット例	197
索引	
はじめてのJMP	199

このガイドについて

『はじめてのJMP』では、JMPソフトウェア全般について紹介します。このガイドは、JMPを初めて使う方を対象としています。アナリスト、研究者、学生、教授、統計学者など、さまざまな方を対象に、JMPのユーザインターフェースと機能全般について概説します。

このガイドで紹介する内容は次のとおりです。

- JMPの起動
- JMP ウィンドウの構造
- データの準備および操作方法
- 対話的なグラフを使用してデータから情報を得る方法
- 単純な分析を実行してグラフを補足する方法
- JMPのカスタマイズと特別な機能

このガイドは6つの章で構成されています。各章では、紹介する概念の理解を深めていただくために、例を挙げています。ここで扱う統計概念はいずれも入門レベルのものです。このガイドで採用しているサンプルデータは、ソフトウェアに付属しています。各章の内容を簡単に説明しておきましょう。

- 第2章「JMPの基本」では、JMPの概要を紹介します。ソフトウェアの構成と操作方法を学習します。
- 第3章「データの操作」では、各種のデータソースからデータを読み込み、データ分析の準備をする方法について説明します。データ操作用の各種ツールについても簡単に紹介します。
- 第4章「データの視覚化」では、データの視覚化と理解に役立つグラフやチャートについて説明します。1つの変数について検討する単純な分析から、複数の変数を取り扱い、それらの関係を示すグラフまで、多様な例を取り上げます。
- 第5章「データの分析」では、一般的に使用されているさまざまな分析手法について解説します。統計的手法を用いる必要のない単純なものから、統計の知識が役に立つ高度なものまで、多岐にわたる分析手法を紹介します。
- 第6章「全体像」では、いくつかのプラットフォームにおける、分布、パターンの分析や、似通った値のグループ化の方法を示します。
- 第7章「作業結果の保存と共有」では、PowerPointのプレゼンテーション、インタラクティブHTML、Adobe Flashを使って、JMPを持っていないユーザと分析結果を共有する方法について解説しています。また、JMPユーザのために、分析をスクリプトとして保存する方法や、作業をジャーナルやプロジェクトに保存する方法についても説明しています。

- 第8章「特別な機能」では、データの変更時に自動的にグラフや分析を更新する方法、環境設定を使用してレポートをカスタマイズする方法、JMPとSASの連動のしくみについて説明します。

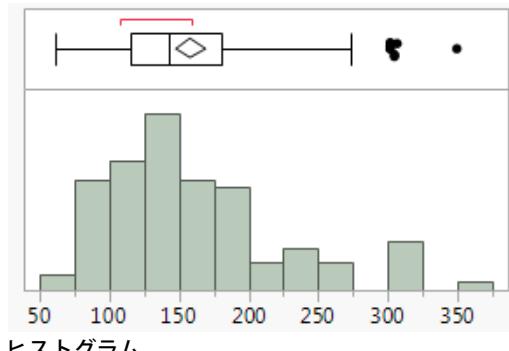
このガイドに目を通せば、JMPを快適に使いこなし、データを操作できるようになるでしょう。

JMPは、WindowsとMacintoshの両オペレーティングシステムに対応していますが、ここでは、Windowsオペレーティングシステムを前提としています。

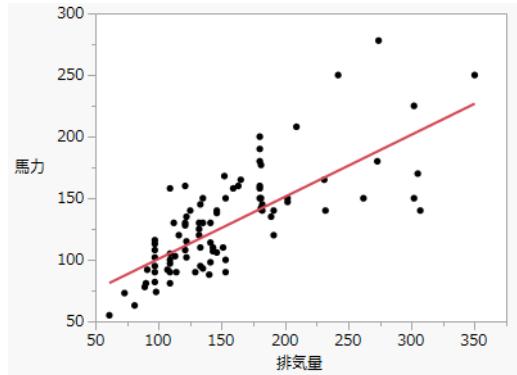
JMPグラフギャラリー

各種グラフおよび対応するプラットフォーム

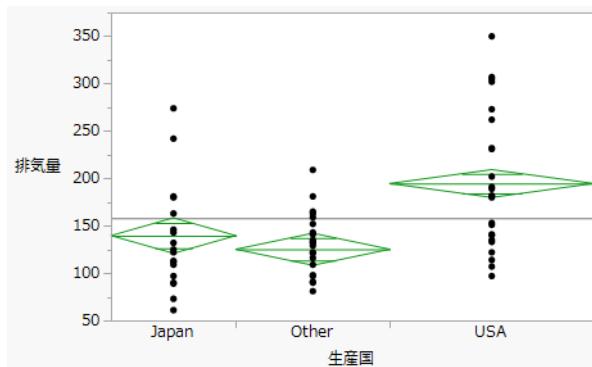
ここでは、JMPで作成できる多数のグラフを画像で紹介します。各画像の下には、グラフの作成に使用されるプラットフォームを記載しています。これらのプラットフォームや各種グラフの詳細については、[ヘルプ] > [ドキュメンテーション] メニューからマニュアルを参照してください。



ヒストグラム
[分析] > [一変量の分布]

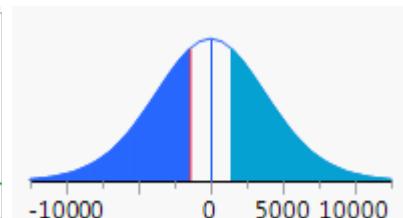


二変量の関係
[分析] > [二変量の関係]



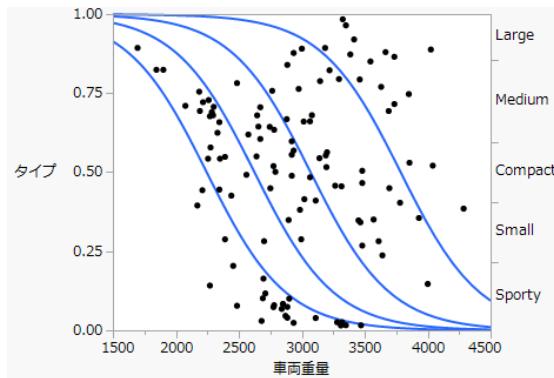
一元配置

[分析] > [二変量の関係]



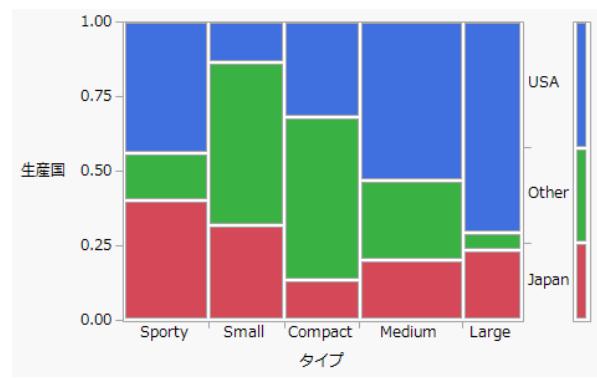
一元配置 t 検定

[分析] > [二変量の関係]



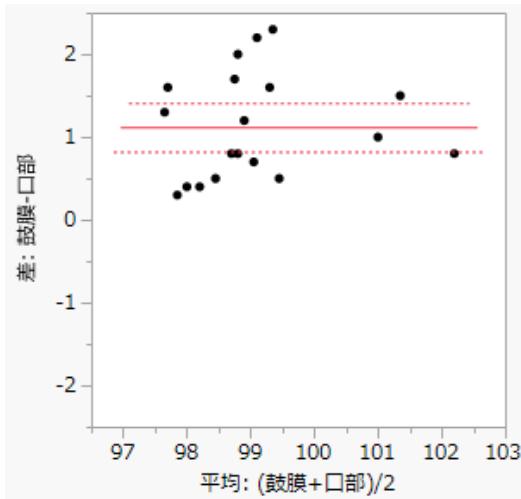
ロジスティック

[分析] > [二変量の関係]



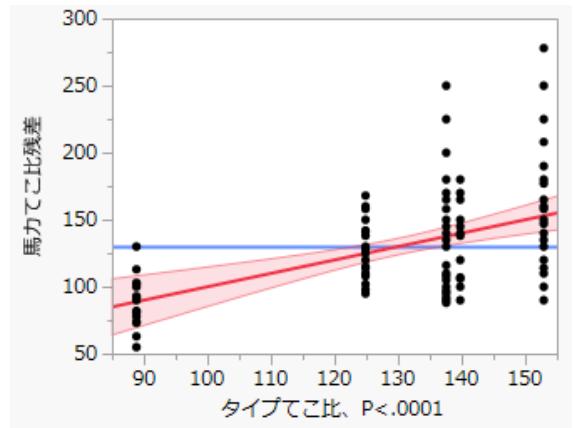
モザイク図

[分析] > [二変量の関係]



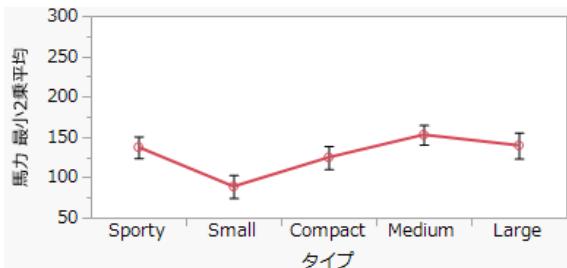
対応のあるペア

[分析] > [発展的なモデル] > [対応のあるペア]



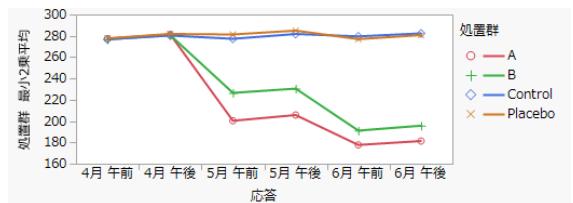
てこ比プロット

[分析] > [モデルのあてはめ]



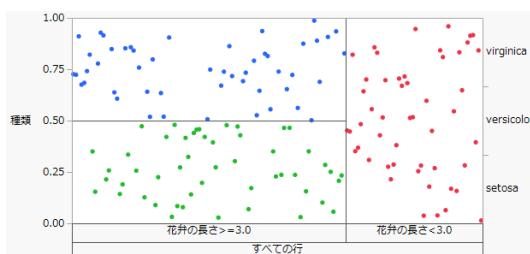
最小2乗平均プロット

[分析] > [モデルのあてはめ]



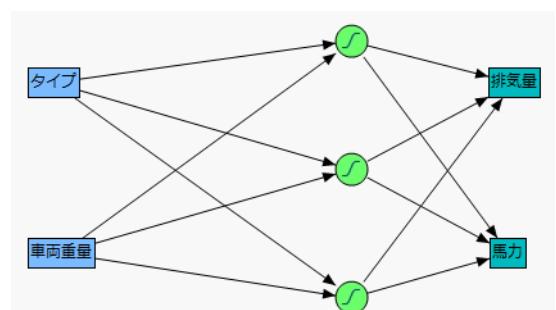
MANOVA

[分析] > [モデルのあてはめ]



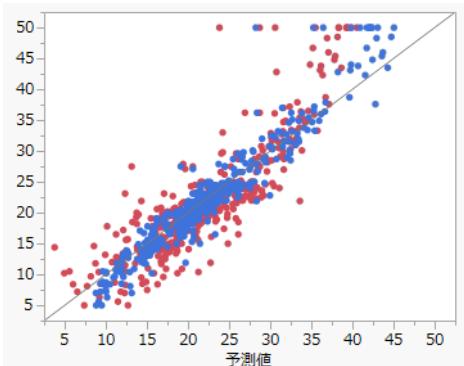
パーティション

[分析] > [予測モデル] > [パーティション]

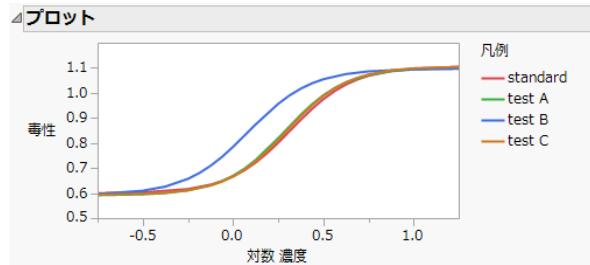


ニューラルネットダイヤグラム

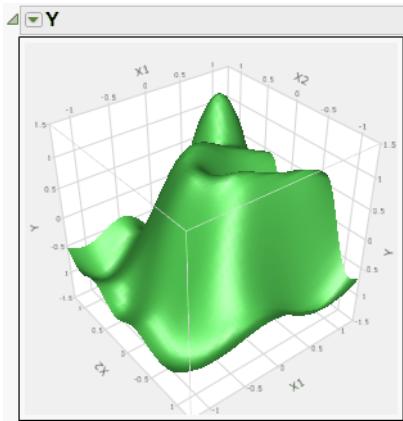
[分析] > [予測モデル] > [ニューラル]



[分析] > [予測モデル] > [モデルの比較]

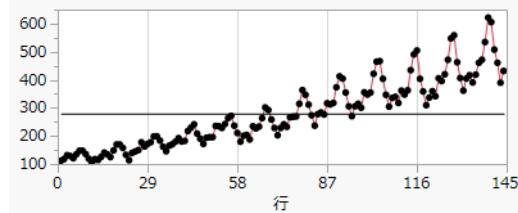


[分析] > [発展的なモデル] > [非線形回帰]



曲面プロファイル

[分析] > [発展的なモデル] > [Gauss 過程]



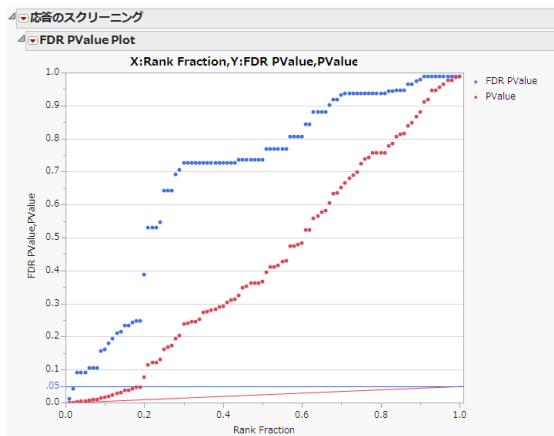
時系列分析

[分析] > [発展的なモデル] > [時系列分析]



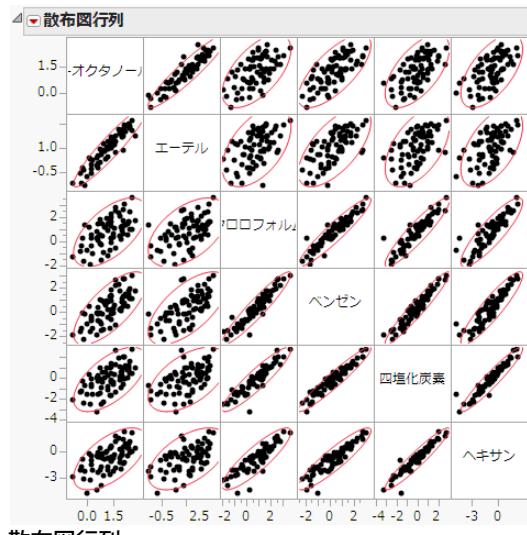
スクリーニング

[分析] > [発展的なモデル] > [発展的な実験計画モデル] > [2水準スクリーニングのあてはめ]



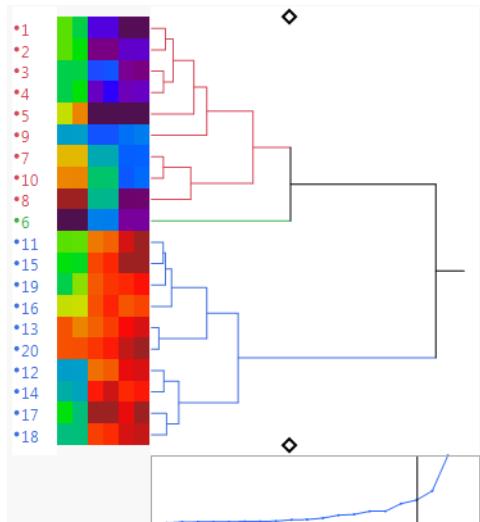
FDR p 値プロット

[分析] > [スクリーニング] > [応答のスクリーニング]



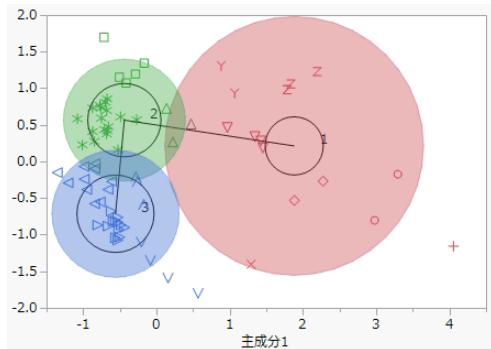
散布図行列

[分析] > [多変量] > [多変量の相関]



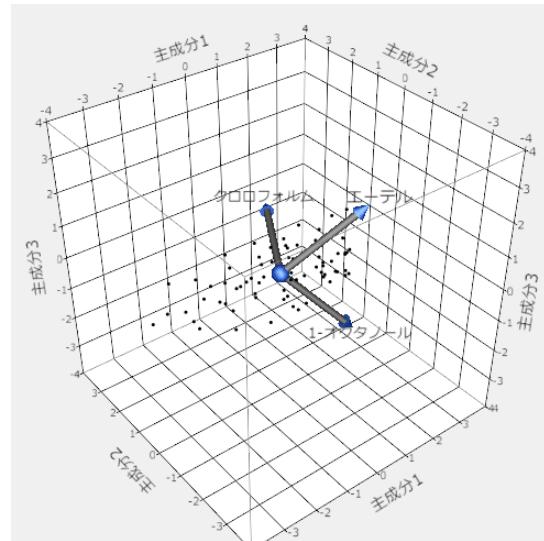
樹形図

[分析] > [クラスター分析] > [階層型クラスター分析]



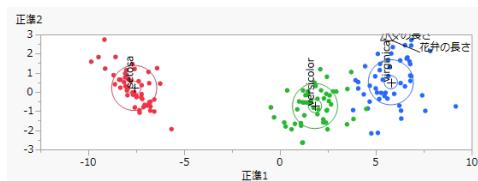
自己組織化マップ

[分析] > [クラスター分析] > [K Means クラスター分析]



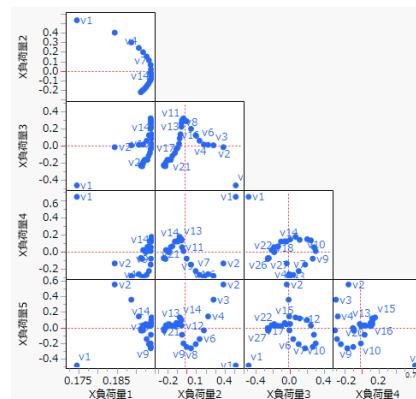
主成分分析

[分析] > [多变量] > [主成分分析]



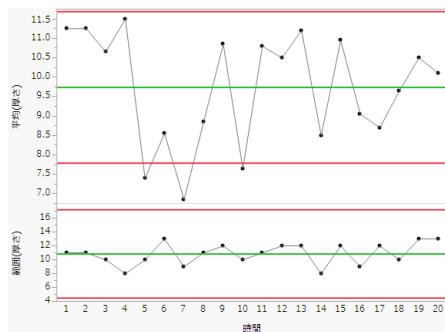
正準プロット

[分析] > [多变量] > [判別分析]

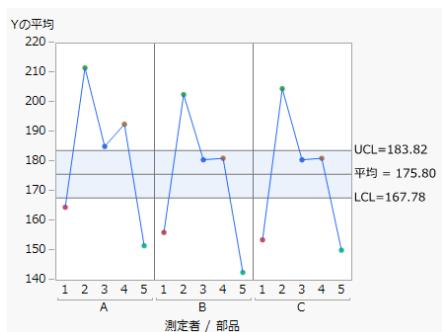


荷負量プロット

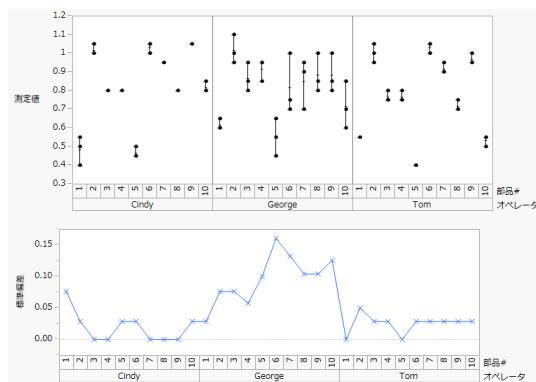
[分析] > [多变量] > [PLS回帰]



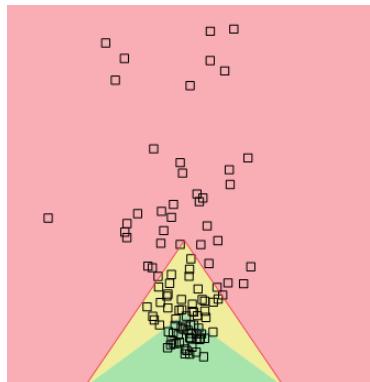
XBar管理図とR管理図
[分析] > [品質と工程] > [管理図ビルダー]



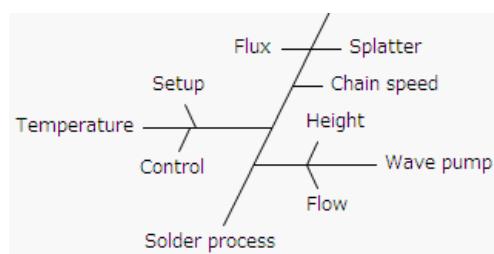
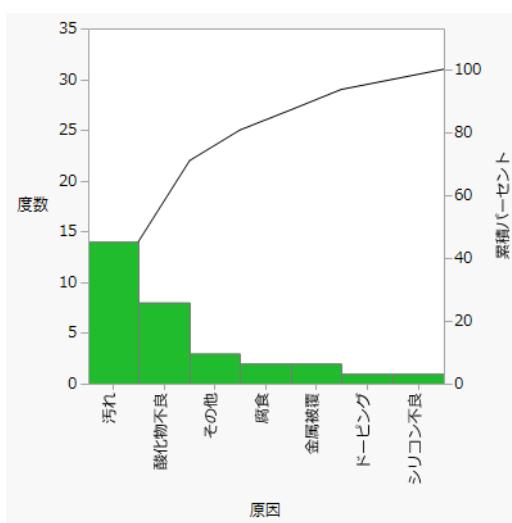
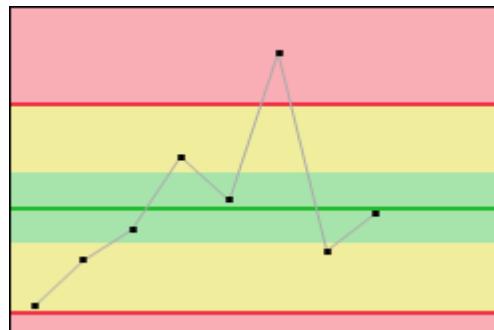
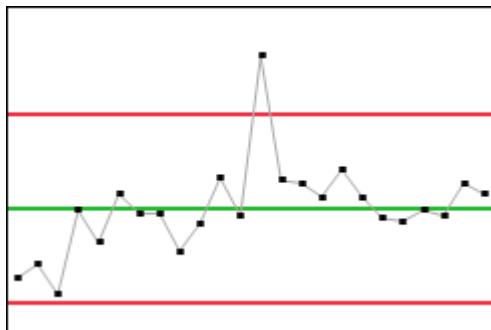
平均図
[分析] > [品質と工程] > [測定システム分析]

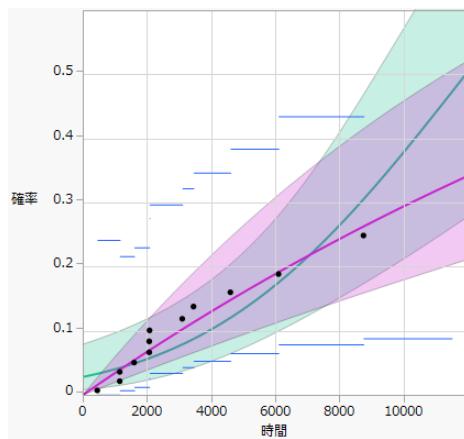


変動性図
[分析] > [品質と工程] > [計量値/計数値ゲージチャート]



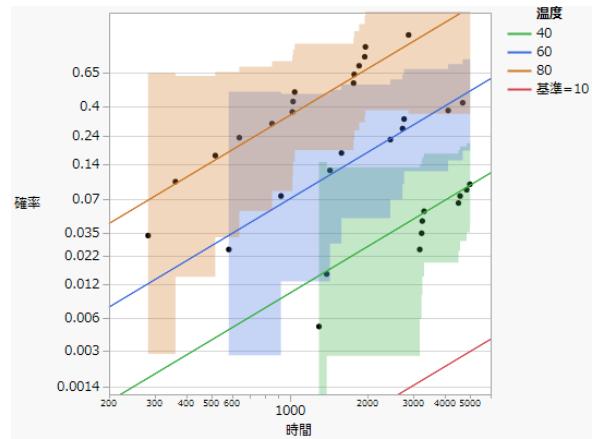
ゴールプロット
[分析] > [品質と工程] > [工程能力]





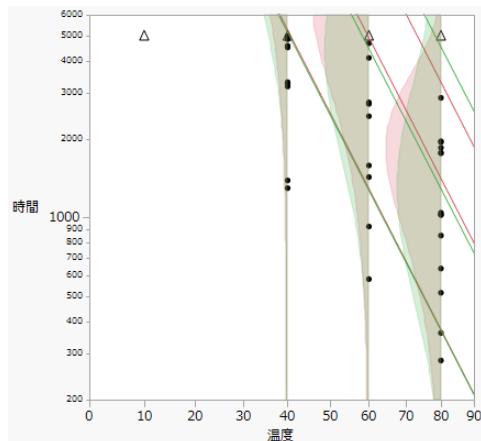
分布の比較

[分析] > [信頼性/生存時間分析] > [寿命の一変量]



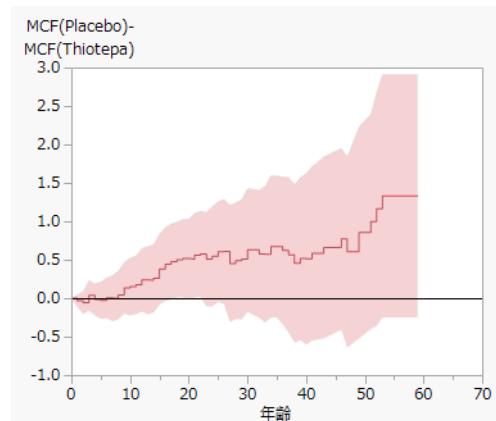
ノンパラメトリック

[分析] > [信頼性/生存時間分析] > [寿命の二変量]



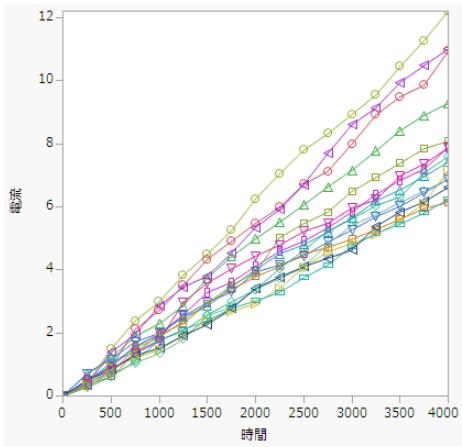
散布図

「分析」>「信頼性/生存時間分析」>「寿命の二変量」

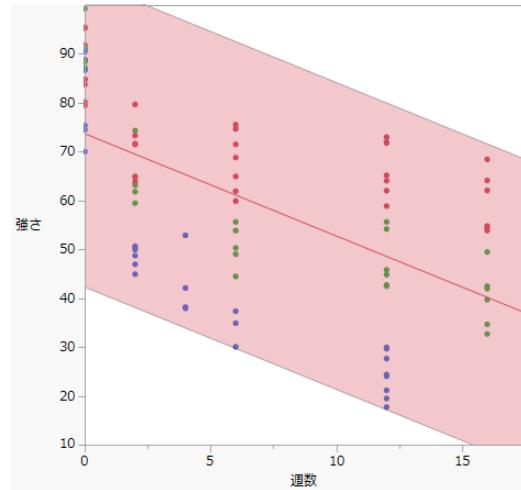


MCF プロット

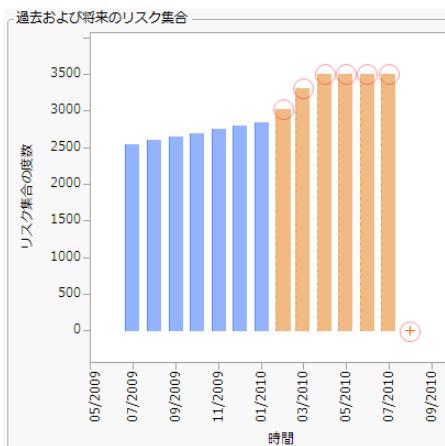
[分析] > [信頼性/生存時間分析] > [再生モデルによる分析]



[分析] > [信頼性/生存時間分析] > [劣化分析]

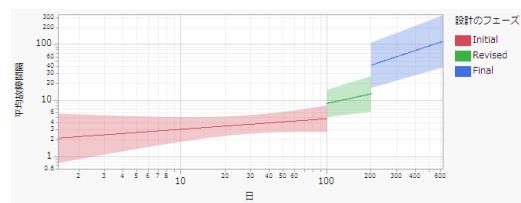


[分析] > [信頼性/生存時間分析] > [破壊劣化]

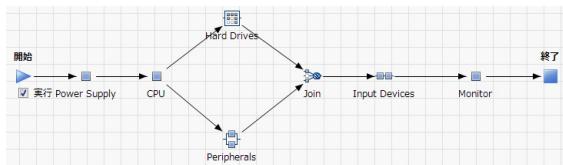


予測

[分析] > [信頼性/生存時間分析] > [信頼性予測]

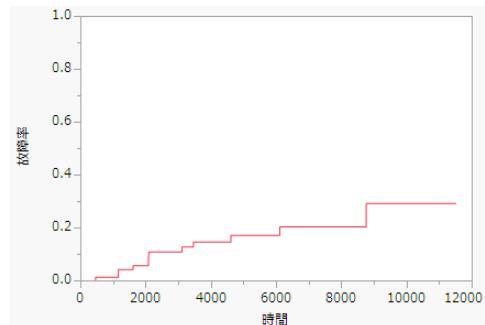


[分析] > [信頼性/生存時間分析] > [信頼性成長]



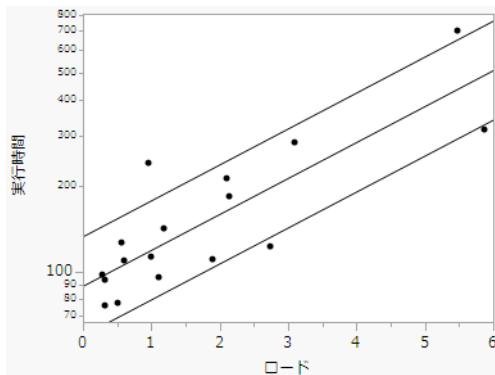
信頼性ブロック図

[分析] > [信頼性/生存時間分析] > [信頼性ブロック図]



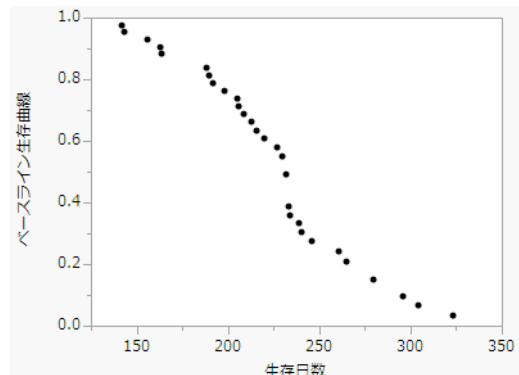
故障率プロット

[分析] > [信頼性/生存時間分析] > [生存時間分析]



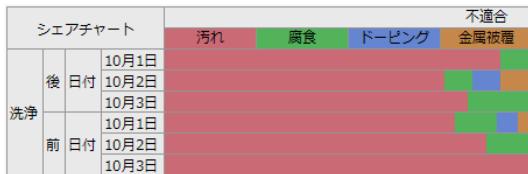
生存分位点

[分析] > [信頼性/生存時間分析] > [生存時間 (パラメトリック) のあてはめ]



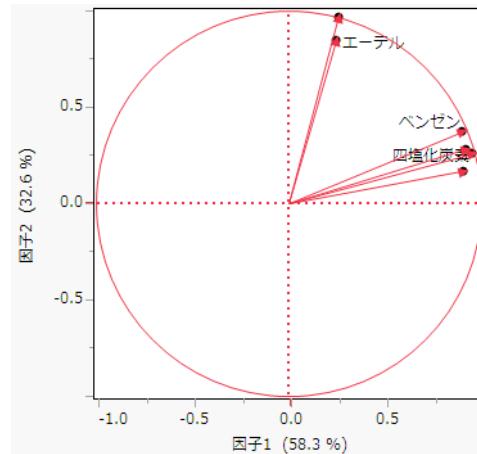
ベースライン生存曲線

[分析] > [信頼性/生存時間分析] > [比例ハザードのあてはめ]

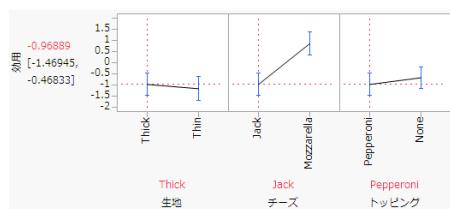


配合プロファイル

[分析] > [消費者調査] > [カテゴリカル]

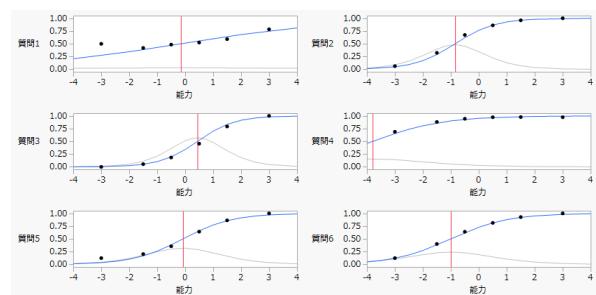


[分析] > [消費者調査] > [因子分析]

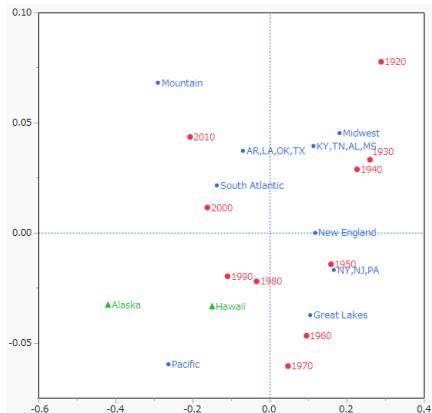


予測プロファイル

[分析] > [消費者調査] > [選択モデル]

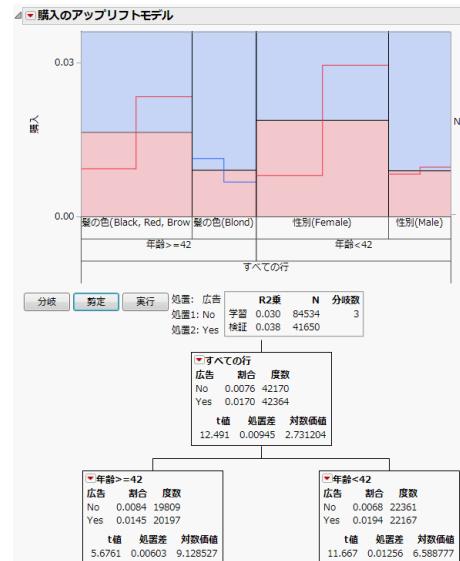


[分析] > [消費者調査] > [項目分析]



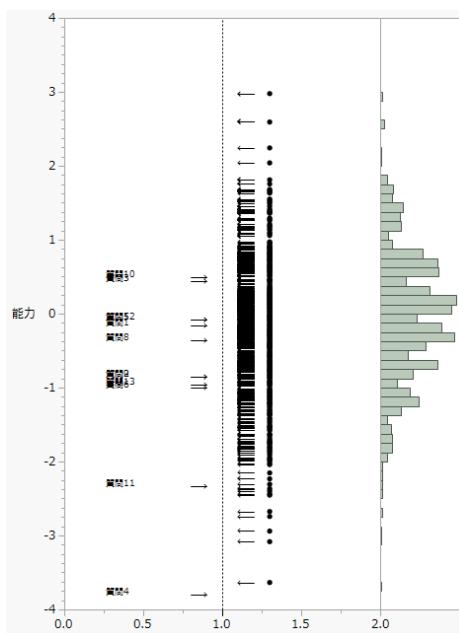
多重対応分析

[分析] > [消費者調査] > [多重対応分析]



アップリフトモデル

[分析] > [消費者調査] > [アップリフト]



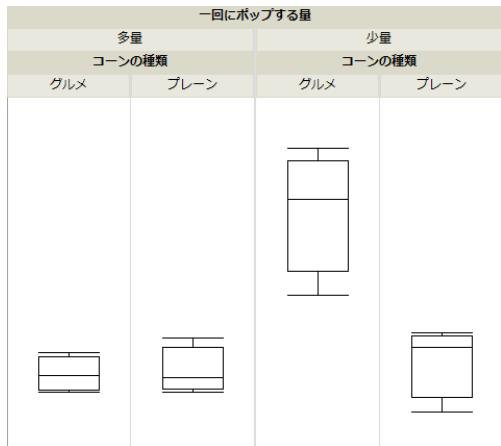
デュアルプロット

[分析] > [消費者調査] > [項目分析]



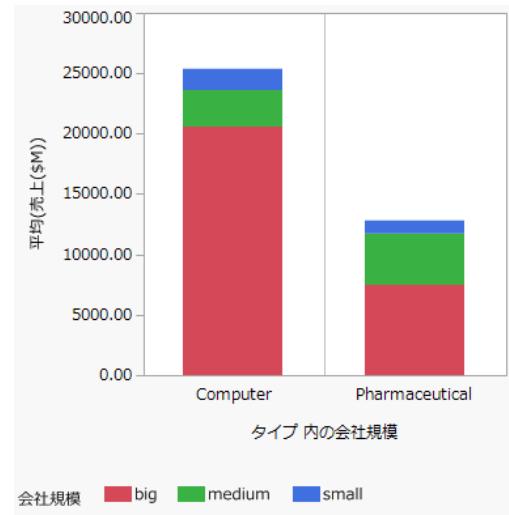
折れ線グラフ

[グラフ] > [グラフビルダー]



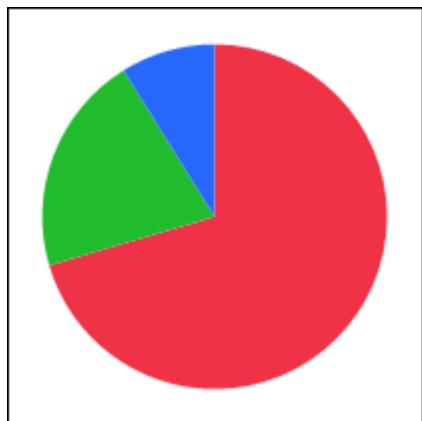
箱ひげ図

[グラフ] > [グラフビルダー]



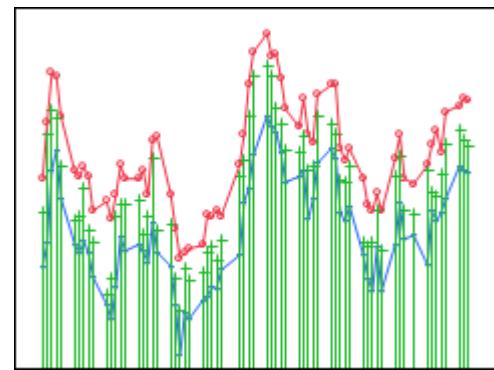
積み重ねた棒グラフ

[グラフ] > [チャート]



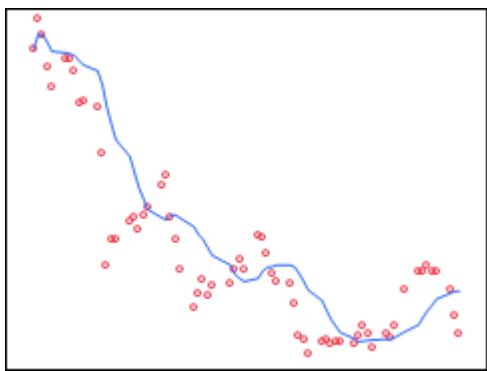
円グラフ

[グラフ] > [チャート]

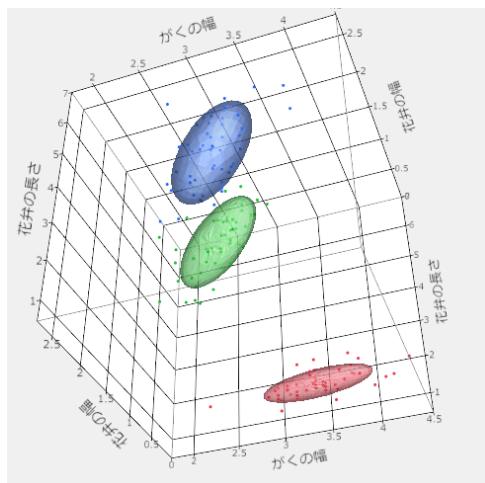


垂線および折れ線グラフ

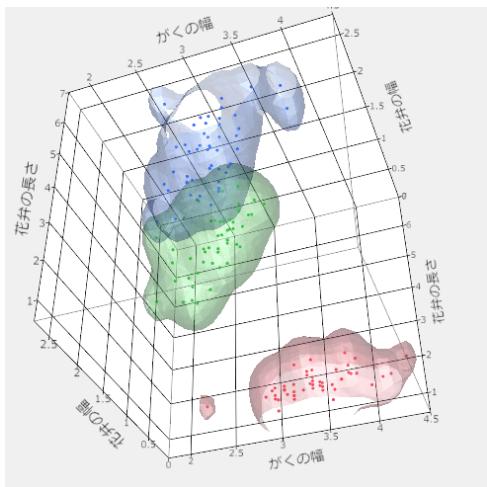
[グラフ] > [重ね合わせプロット]



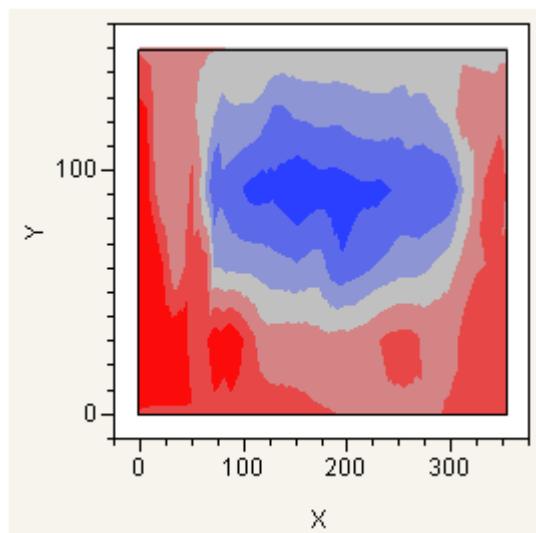
点および折れ線グラフ
[グラフ] > [重ね合わせプロット]



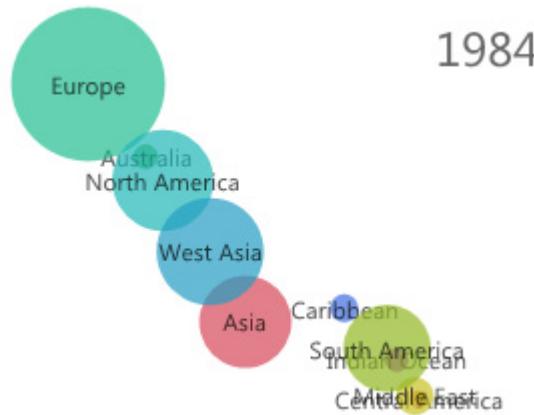
三次元散布図
[グラフ] > [三次元散布図]



三次元散布図
[グラフ] > [三次元散布図]

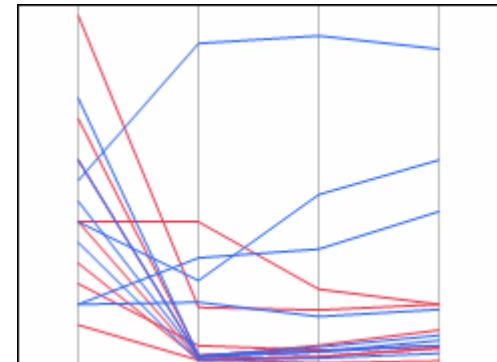
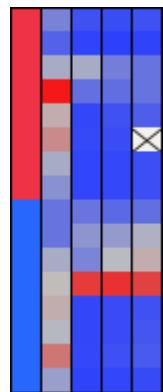


等高線図
[グラフ] > [等高線図]



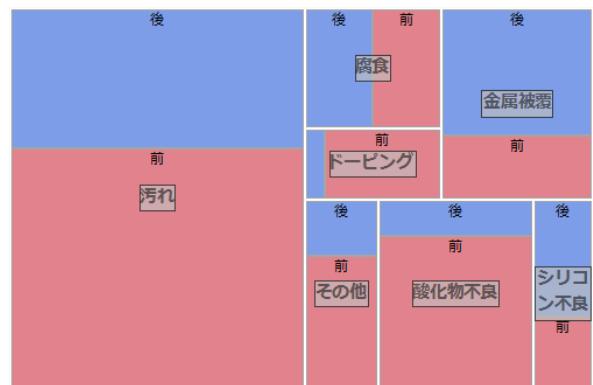
バブルプロット

[グラフ] > [バブルプロット]

パラレルプロット
[グラフ] > [パラレルプロット]

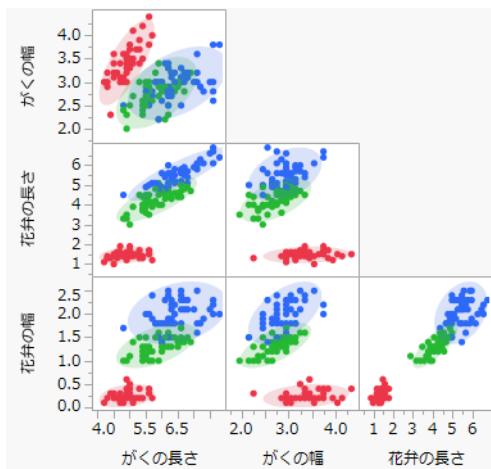
セルプロット

[グラフ] > [セルプロット]

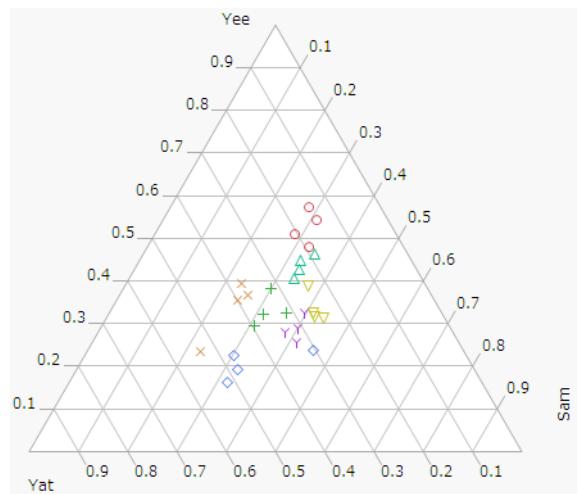


ツリーマップ

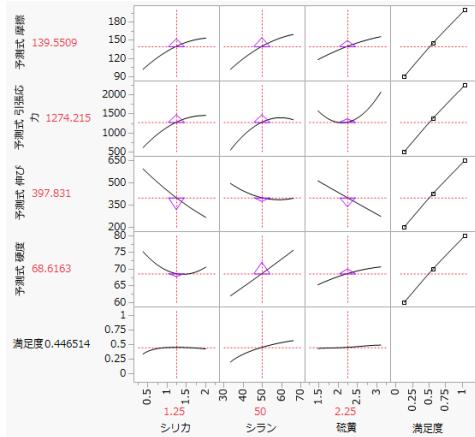
[グラフ] > [ツリーマップ]



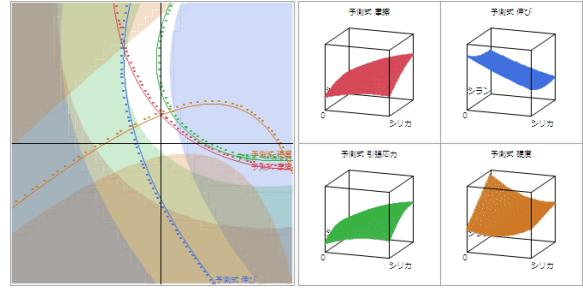
散布図行列
[グラフ] > [散布図行列]



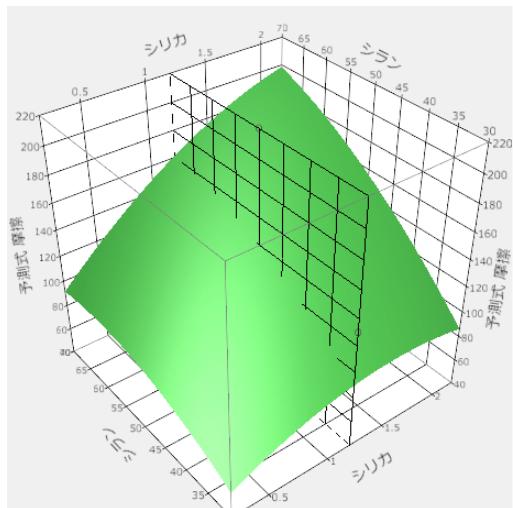
三角図
[グラフ] > [三角図]



予測プロファイ尔
[グラフ] > [プロファイル]

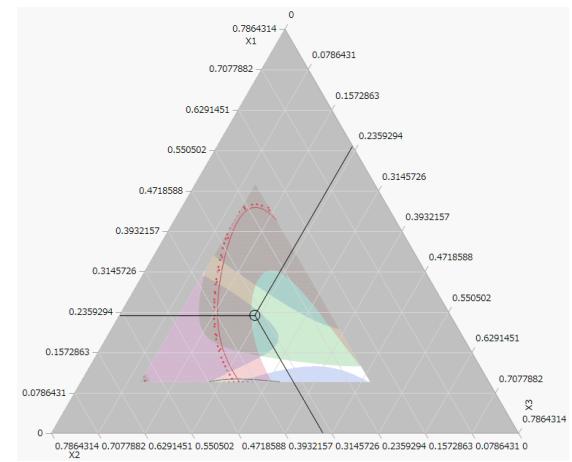


等高線プロファイ尔
[グラフ] > [等高線プロファイ尔]



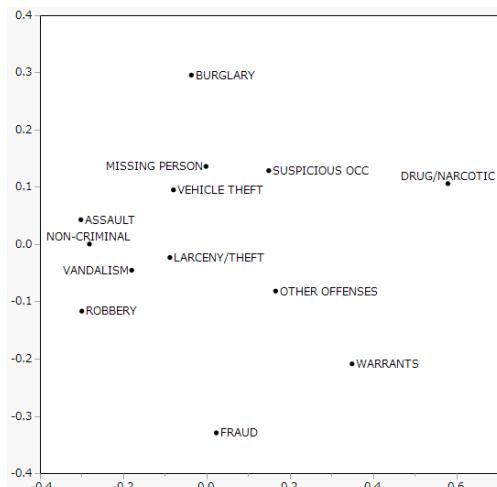
曲面プロット

[グラフ] > [曲面プロット]



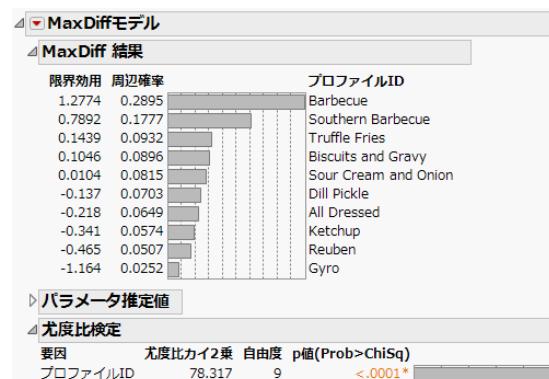
配合プロファイル

[グラフ] > [配合プロファイル]



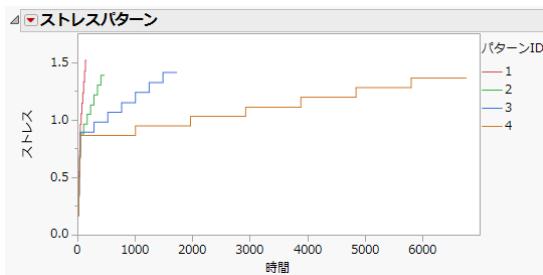
多次元尺度構成

[分析] > [消費者調査] > [多次元尺度構成]



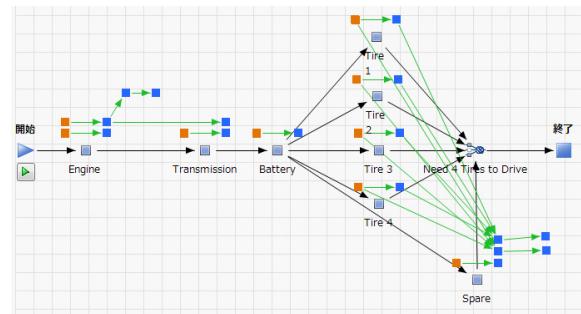
MaxDiff

[分析] > [消費者調査] > [MaxDiff]



ストレスパターンプロット

[分析] > [信賴性/生存時間分析] > [累積損傷]



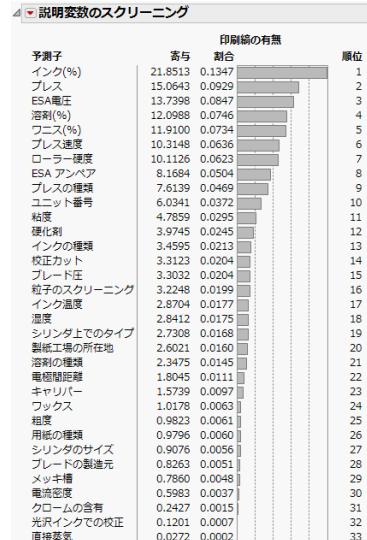
修理可能システムのシミュレーション

[分析] > [信頼性/生存時間分析] > [修理可能システムのシミュレーション]



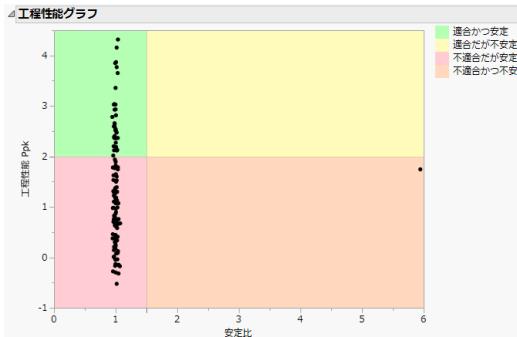
潜在クラス分析

[分析] > [クラスター分析] > [潜在クラス分析]



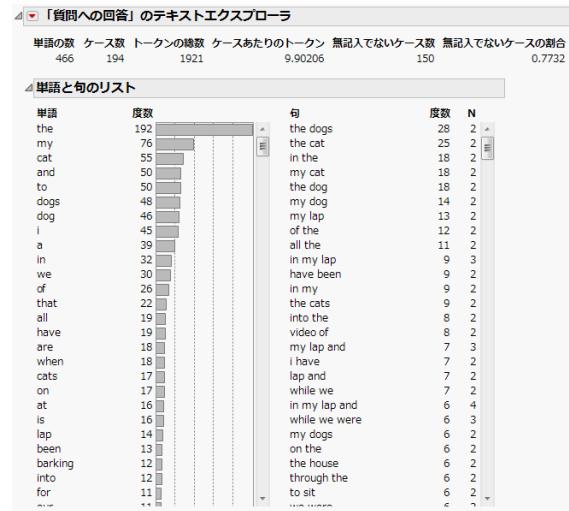
説明変数のスクリーニング

[分析] > [スクリーニング] > [説明変数のスクリーニング]



工程のスクリーニング

[分析] > [スクリーニング] > [工程のスクリーニング]



テキストエクスプローラ

[分析] > [テキストエクスプローラ]

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 http://wwwjmp.com/ja_jp/japan/pro/ をご覧ください。

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

ヒント: 役に立つ情報には「ヒント」という見出しがついています。

JMP のマニュアル

JMP 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： <ul style="list-style-type: none">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	<p>[グラフ] メニューからアクセスできる以下のプラットフォームの説明：</p> <ul style="list-style-type: none"> • 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 <p>このマニュアルには背景マップやカスタムマップの作成方法も記載されています。</p>
プロファイル機能	対話式のプロファイルツールの使い方を学ぶ。任意の応答曲面の断面を表示できるようになります。	[グラフ] メニューに表示されるすべてのプロファイルについて。誤差因子の分析が、ランダム入力を使用したシミュレーションの実行とともに含まれています。
実験計画	実験の計画方法と適切な標本サイズの決定方法を学ぶ	[実験計画 (DOE)] メニューと [分析] > [発展的なモデル] メニューの「発展的な実験計画モデル」に関するすべてのトピックについて。

マニュアル	目的	内容
基本的な回帰モデル	「モデルのあてはめ」プラットフォームとその多くの手法について学ぶ	<p>[分析] メニューの「モデルのあてはめ」プラットフォームで使用できる、以下の手法の説明：</p> <ul style="list-style-type: none">• 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	<p>[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：</p> <ul style="list-style-type: none"> • 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 • 計算式デボ <p>[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：</p> <ul style="list-style-type: none"> • 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 <p>[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：</p> <ul style="list-style-type: none"> • 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 <p>[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』に説明があります。</p>

マニュアル	目的	内容
多変量分析	複数の変数を同時に分析するための手法について理解を深める	<p>[分析] > [多変量] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none">• 多変量の相関• 主成分分析• 判別分析• PLS <p>[分析] > [クラスター分析] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none">• 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	<p>[分析] > [品質と工程] メニューで使用できる以下のプラットフォームの説明：</p> <ul style="list-style-type: none">• 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] メニューで使用できる以下のプラットフォームの説明： <ul style="list-style-type: none"> • 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] メニューで使用できる以下のプラットフォームの説明： <ul style="list-style-type: none"> • 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。

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

JMP ヘルプ

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

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

JMP を習得するためのその他のリソース

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

- チュートリアル（「チュートリアル」（43 ページ）を参照）
- サンプルデータ（「サンプルデータテーブル」（44 ページ）を参照）
- 索引（「統計用語と JSL 用語の習得」（44 ページ）を参照）
- 使い方ヒント（「JMP を使用するためのヒント」（44 ページ）を参照）
- Web リソース（「JMP User Community」（45 ページ）を参照）
- 専門誌『JMPer Cable』（「JMPer Cable」（45 ページ）を参照）
- JMP に関する書籍（「JMP 関連書籍」（45 ページ）を参照）
- JMP スターター（「「JMP スターター」 ウィンドウ」（46 ページ）を参照）
- 教育用リソース（「サンプルデータテーブル」（44 ページ）を参照）

チュートリアル

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

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

他のチュートリアルでは、実験の計画、標本平均と定数の比較など、JMPの具体的な活用法を学習できます。

サンプルデータテーブル

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

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

サンプルデータテーブルは次のディレクトリにインストールされています。

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

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

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

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

統計用語とJSL用語の習得

【ヘルプ】メニューには、次の索引が用意されています。

統計の索引 統計用語が説明されています。

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

JMPを使用するためのヒント

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

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

ツールヒント

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

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

ヒント：Windows では、JMP 環境設定でツールヒントを表示しないよう設定できます。[ファイル] > [環境設定] > [一般] を選択し、[メニューのヒントを表示] の選択を解除します。このオプションは、Macintosh では使用できません。

JMP User Community

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

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

Web サイトの JMP リソースにアクセスするには、[ヘルプ] > [JMP User Community] を選択するか、
<https://communityjmp.com/> をご覧ください。

JMPer Cable

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

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

JMP 関連書籍

JMP 関連書籍は、次の JMP Web ページで紹介されています。

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

「JMP スターター」 ウィンドウ

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

JMP の基本 基本概念

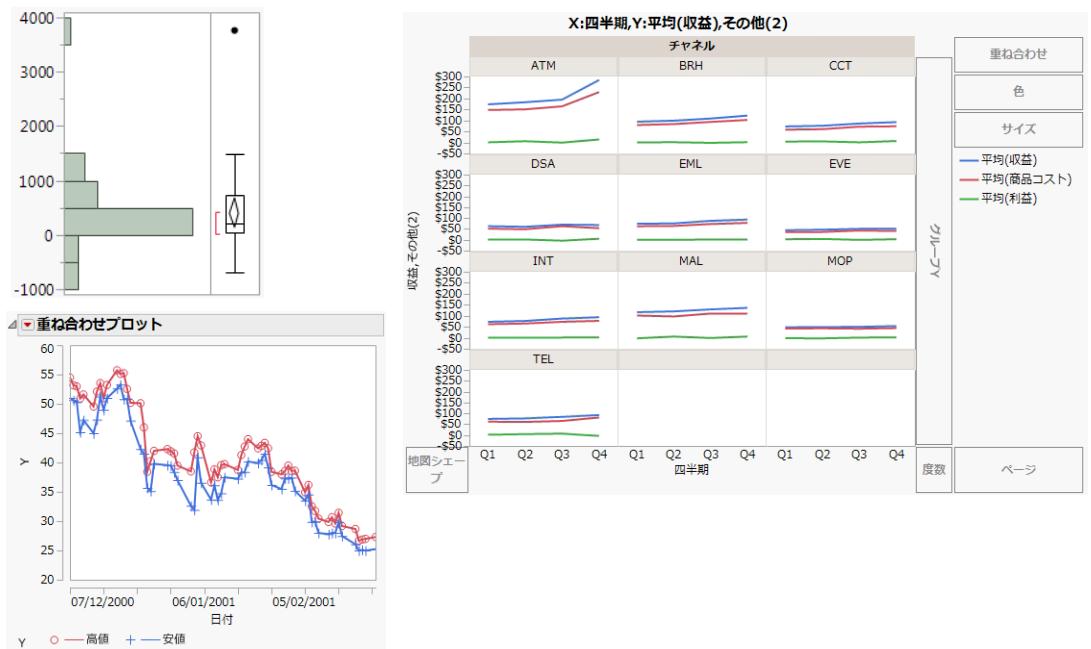
JMP（ジャンプと発音します）は、データを視覚化し、統計分析を行う高性能な対話型ツールです。JMPでは、データテーブル、グラフ、チャート、レポートを活用しながら、分析を実行し、データと対話することで、データから多くの情報を引き出すことができます。

さまざまな統計分析やモデル化を行う研究者にとって、JMPはとても便利です。また、データの傾向やパターンをすばやく発見する必要があるビジネスアナリストにとっても有益です。JMPを使用すれば、統計の専門家でなくても、データから情報を導き出すことができます。

たとえば、JMPを使用して、次の処理を実行できます。

- 対話型のグラフとチャートを作成してデータについて検討し、関係を明らかにする。
- 複数の変数間の変動パターンを一度に明らかにする。
- 大量のデータについて検討し、要約する。
- 高性能な統計モデルを開発し、将来を予測する。

図2.1 JMP レポートの例



理解すべき基本操作

JMP を用いるにあたって、次の基本操作を頭に入れておく必要があります。

- JMP データテーブルにおいて、データを入力、表示、編集、および操作します。
- [分析]、[グラフ]、[実験計画(DOE)] メニューからプラットフォームを選択します。プラットフォームは、データ分析やグラフ操作を行う対話型のウィンドウで構成されています。
- プラットフォームでは次のウィンドウを使用します。
 - 起動ウィンドウでは、分析を設定して実行します。
 - レポートウィンドウには、分析結果が表示されます。
- レポートウィンドウには通常、次の項目が表示されます。
 - ある種のグラフ（散布図やチャートなど）。
 - 特定のレポート。開閉ボタン  で表示したり非表示にしたりできます。
 - プラットフォームのオプション。赤い三角ボタン  をクリックすると表示されるメニューにあります。

使い方入門

JMP での大まかな作業手順は、とても簡単です。

1. データを JMP に読み込みます。
2. プラットフォームを選択し、起動ウィンドウで必要事項をすべて指定します。
3. 結果を検討し、データから導き出された情報を確認します。

この作業手順については、「[JMP での作業手順の理解](#)」(53 ページ) で詳しく解説しています。

一般的には、JMP で作業する際は、まずグラフを使用して個々の変数を視覚化し、それらの変数間の関係を明らかにします。グラフを作成すると、こうした情報がわかりやすく示され、さらに検討すべき点が明らかになります。その上で、分析プラットフォームを使用して、問題をさらに検討し、解決策を見つけます。

- 「[データの操作](#)」の章 (61 ページ) では、データを JMP に読み込む方法を紹介します。
- 「[データの視覚化](#)」の章 (89 ページ) では、JMP に用意されている便利なグラフの一部を使用して、データを詳しく考察する方法を紹介します。
- 「[データの分析](#)」の章 (123 ページ) では、一部の分析プラットフォームの使用方法を紹介します。
- 「[全体像](#)」の章 (157 ページ) では、いくつかのプラットフォームにおける、分布、パターン、および類似した値の分析方法を示します。

各章では、例を挙げて説明します。この章の残りの節では、データテーブルと、JMP での作業に関するごく基本的な操作について説明していきます。

JMP の起動

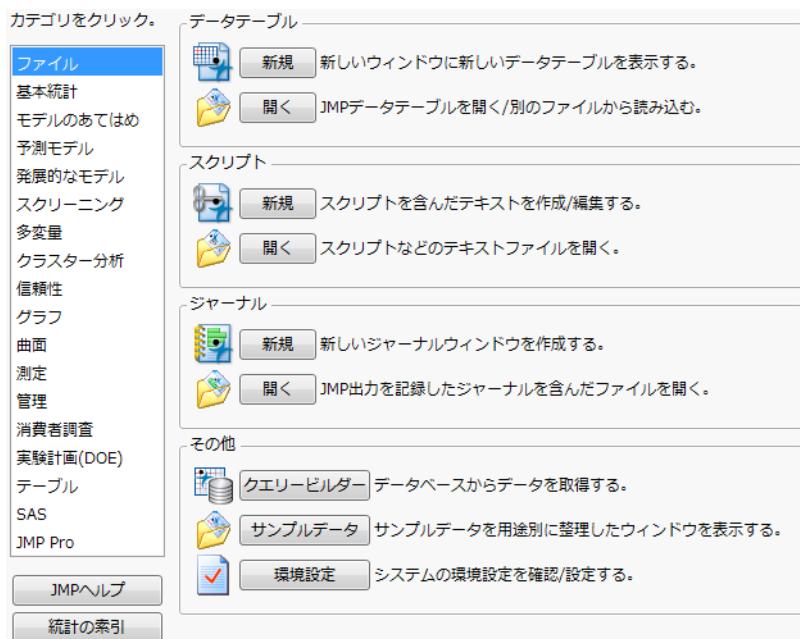
JMP は 2 通りの方法で起動できます。

- JMP アイコンをダブルクリックします。このアイコンは、通常はデスクトップ上にあります。この場合、JMP が起動しますが、既存の JMP ファイルは開きません。
- 既存の JMP ファイルをダブルクリックします。JMP が起動し、該当するファイルが開きます。

Windows 上で JMP を起動すると、「使い方ヒント」ウィンドウとホームウィンドウが表示されます。Macintosh では、「使い方ヒント」ウィンドウ、「JMP スターター」、およびホームウィンドウが表示されます。

「JMP スターター」ウィンドウには、操作やプラットフォームがカテゴリ別に整理して表示されます。

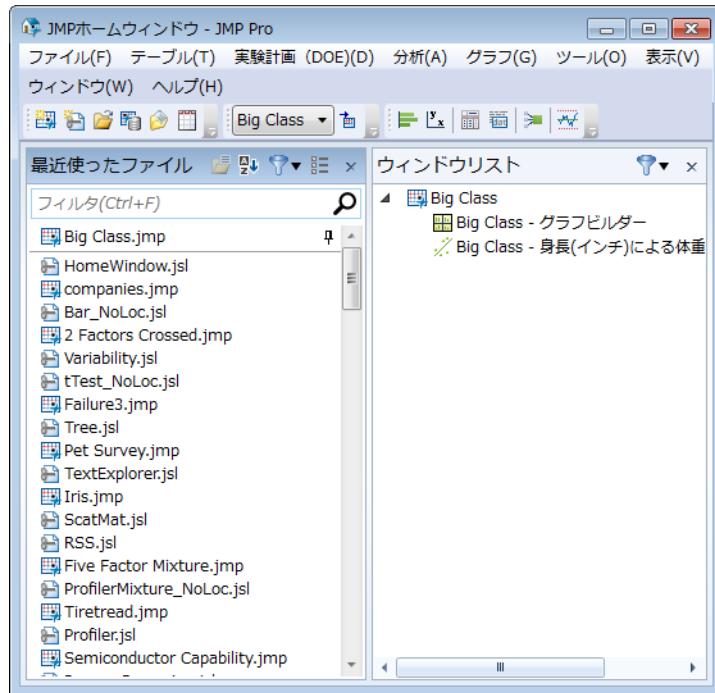
図2.2 「JMP スターター」 ウィンドウ



左側には、カテゴリの一覧が表示されます。カテゴリをクリックすると、そのカテゴリに関連する機能とコマンドが表示されます。「JMP スターター」には、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。

ホームウィンドウは、JMP 内でファイルを整理したり、それらにアクセスしたりするのに便利です。

図2.3 Windows上のホームウィンドウ



Windowsでホームウィンドウを開くには、[表示] > [ホームウィンドウ]を選択します。Macintoshでは、[ウィンドウ] > [JMPホーム]を選択します。ホームウィンドウには、次の項目へのリンクが表示されます。

- 現在開いているデータテーブルとレポートウィンドウ
- 最近開いたファイル

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

JMPのほぼすべてのウィンドウには、メニューバーとツールバーがあります。JMPの機能の大半は次の3つの方法で使用できます。

- メニューバーの使用
- ツールバーのボタンの使用
- 「JMPスターター」ウィンドウのボタンの使用

メニューバーとツールバーについて

メニューとツールバーは、多くのウィンドウで非表示になっています。これらを表示するには、ウィンドウのタイトルバーのすぐ下の青いバーの上にマウスのポインタを置きます。なお、「JMPスターター」ウィンドウ、ホームウィンドウ、およびすべてのデータテーブルでは、メニューは常に表示されています。

サンプルデータの使用

このガイドやその他の JMP マニュアルで取り上げる例では、サンプルデータテーブルを使用しています。これらのサンプルデータは、Windows ではデフォルトで次の場所に保存されています。

C:/Program Files/SAS/JMP/13/Samples/Data
C:/Program Files/SAS/JMPPro/13/Samples/Data
C:/Program Files/SAS/JMPSW/13/Samples/Data

[ヘルプ] メニューの [サンプルデータ] では、データテーブルがカテゴリ別にまとめて表示されます。開閉ボタンをクリックすると、該当するカテゴリのデータテーブルの一覧が表示されます。リンクをクリックすると、該当するデータテーブルが開きます。

Macintosh のサンプルデータは、/Library/Application Support/JMP/13/Samples/Data にインストールされます。

JMP サンプルデータテーブルを開く

1. [ヘルプ] メニューから [サンプルデータ] を選択します。
2. 「[はじめての JMP] で使うデータ」の横にある開閉ボタンをクリックして、データテーブルの一覧を表示します。
3. 適切なデータテーブルの名前をクリックして、このガイドの例と併せて使用します。適切なデータテーブルの名前をクリックすると、このマニュアルの例で使われているデータが開きます。

読み込み用サンプルデータ

JMP 以外のアプリケーションのデータファイルを JMP に読み込む方法を学習するために、読み込み用のサンプルデータも準備されています。

読み込み用のサンプルデータは、Windows ではデフォルトで次の場所に保存されています。

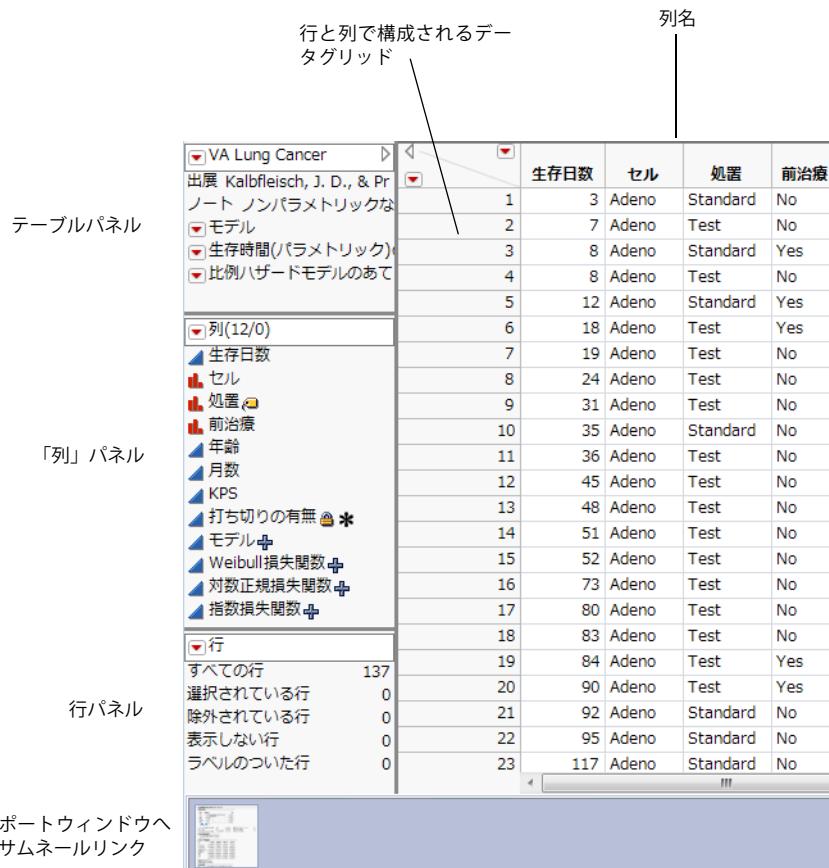
C:/Program Files/SAS/JMP/13/Samples/Import Data
C:/Program Files/SAS/JMPPro/13/Samples/Import Data
C:/Program Files/SAS/JMPSW/13/Samples/Import Data

データターブルの理解

データテーブルとは、行と列で構成されたデータの集合です。Microsoft® Excel® のスプレッドシートに似ていますが、一部重要な相違点もあります。これらの相違点については、「[JMP と Excel の相違点](#)」(58 ページ) で説明します。データテーブルには、ノート、変数、スクリプトなどの他の情報が格納されている場合もあります。これらの追加項目については、後述の章で説明します。

退役軍人に対して行った、肺がんに関する臨床試験のサンプルデータ 「VA Lung Cancer.jmp」を開いてみましょう。ここでは、このデータテーブルを基に説明していきます。

図2.4 データテーブル



データテーブルは次の各部で構成されています。

データグリッド データグリッドは、行列構成でデータを格納します。一般には、データグリッドの各行がオブザーバーションであり、列（変数とも呼びます）がオブザーバーションに関する情報を提供します。図2.4では、各行が被験者に対応し、12列分の情報があります。データグリッドでは12列の一部しか表示されませんが、列パネルではすべての列が一覧表示されています。被験者ごとに、「生存日数」、「セル」、「処置」などの情報が記録されています。各列には見出し（列名）があります。列名は、テーブルの行数の合計には含まれません。

テーブルパネル テーブルパネルには、テーブルの変数またはスクリプトが表示されます。図2.4では、「モデル」という名前で保存されたスクリプトが1つ表示されています。このスクリプトは、分析を自動的に再実行します。このテーブルには、「ノート」という名前の変数もあり、データに関する情報を格納しています。テーブル変数とテーブルスクリプトについては、後述の章で説明します。

「列」パネル 「列」パネルには、すべての列の数と、選択されている列の数、および、すべての列の名前が表示されます。たとえば、先ほどの図における括弧内の数字「(12/0)」は、12の列があり、どの列も選択

されていないことを示しています。各列名の左側にあるアイコンは、その列の尺度を示します。尺度については、「データの分析」の章の「[尺度の理解](#)」(127ページ)で説明します。右側のアイコンは、列に割り当てられている属性を示します。これらのアイコンの詳細については、「データの操作」の章の「[列情報の表示または変更](#)」(74ページ)を参照してください。

行パネル 「行」パネルには、データテーブル内の行の数、選択されている行の数、除外されている行の数、非表示の行の数、ラベルのついた行の数が表示されます。図2.4では、データテーブル内に137の行があります。

レポートウィンドウへのサムネールリンク この領域には、当該のデータテーブルをもとに作成されたレポートに対するサムネールが表示されます。サムネールの上にカーソルを置くと、該当するレポートウィンドウのプレビューが拡大表示されます。サムネールをダブルクリックすると、該当するレポートウィンドウが前面に表示されます。

行や列の追加、データの入力、データの編集など、データグリッドの操作については、「[データの操作](#)」の章(61ページ)で説明します。複数のデータテーブルを開くと、各データテーブルが個別のウィンドウに表示されます。

JMP での作業手順の理解

データをデータテーブルに表示したら、グラフを作成したり、分析を実行したりできます。分析やグラフの機能は、「プラットフォーム」で提供されています。各プラットフォームは、[分析] と [グラフ] メニューから呼び出されます。これらを「プラットフォーム」と呼んでいるのは、単に静的な結果を出力するだけにとどまらないからです。プラットフォームの結果はレポートウィンドウに表示されますが、それらの結果は、元のデータテーブルや他の結果と相互にリンクしており、かつ、対話的に操作できます。

[分析] と [グラフ] メニューにあるプラットフォームには、分析やデータ探索のための様々な機能が用意されています。

グラフ作成や分析実行のおおまかな操作手順は次のとおりです。

1. データテーブルを開きます。
2. [グラフ] または [分析] メニューからプラットフォームを選択します。
3. プラットフォームの起動ウィンドウで必要事項を指定し、分析を設定します。
4. [OK] をクリックすると、グラフや分析結果がレポートウィンドウに表示されます。
5. レポートのオプションを使用してレポートをカスタマイズします。
6. 結果を保存および書き出して、他のユーザと共有します。

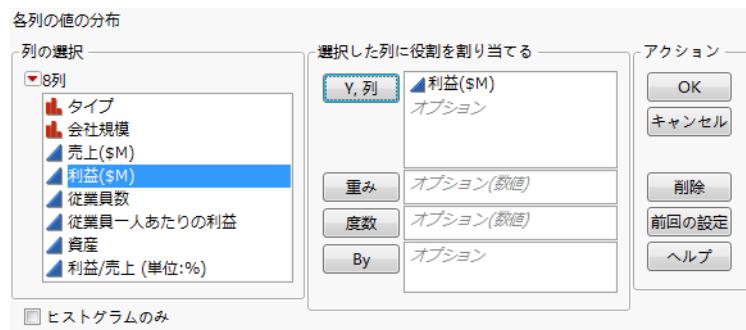
後述の各章では、これらの概念について詳しく解説していきます。

次の例では、单变量の分析を実行し、カスタマイズする方法を4つの手順で紹介します。ここでは、会社情報のサンプルデータ「Companies.jmp」を用いて、「**利益(\$M)**」の初步的な分析を行ってみましょう。

ステップ1: プラットフォームを起動して結果を表示する

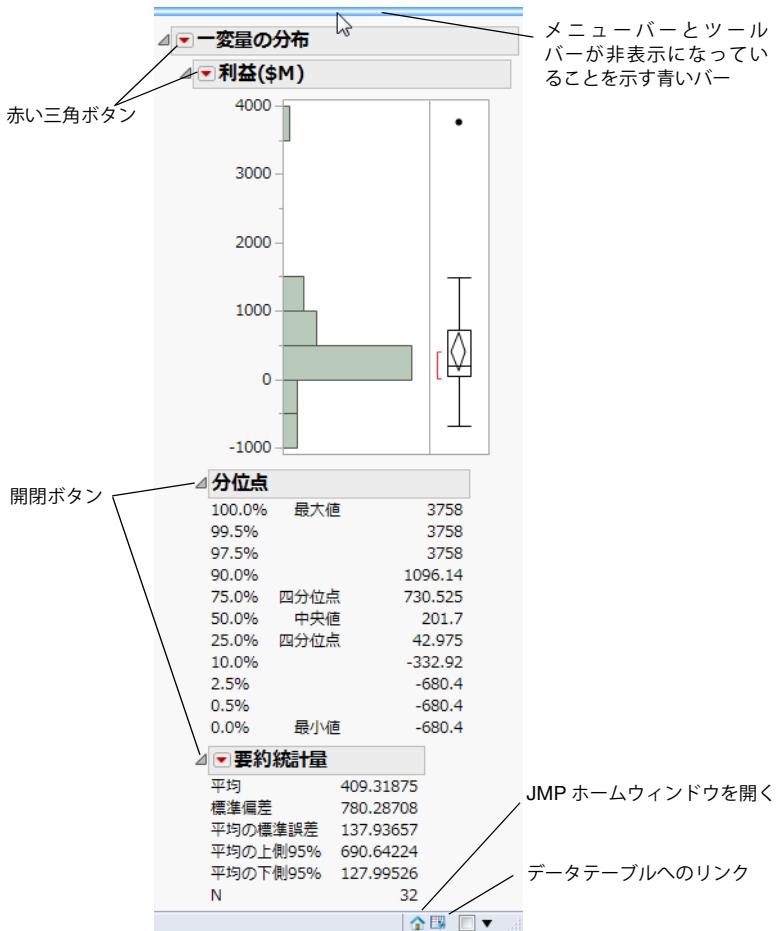
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Companies.jmp」を開きます。
2. [分析] > [一変量の分布] を選択して、「一変量の分布」起動ウィンドウを表示します。
3. 「列の選択」ボックスで「利益(\$M)」を選択し、[Y, 列] ボタンをクリックします。
 「利益(\$M)」変数が [Y, 列] 役割に表示されます。指定後のウィンドウは、図2.5のようになります。
 「列の選択」ボックスで列をクリックし、いずれかの役割のボックスにドラッグするという操作でも、変数に役割を割り当てることができます。

図2.5 「利益(\$M)」の割り当て



4. [OK] をクリックします。
 「一変量の分布」レポートウィンドウが表示されます。

図2.6 「一変量の分布」レポートウィンドウ (Windows)



レポートウィンドウには、基本的なグラフと分析レポートが表示されます。結果は「アウトライン」と呼ばれる形式で表示されます。アウトラインの開閉ボタンをクリックすることにより、各レポートの表示／非表示を切り替えることができます。

赤い三角ボタンをクリックして表示されるメニューからオプションやコマンドを選択し、隨時、追加のグラフを作成したり、さらなる分析を実行したりできます。

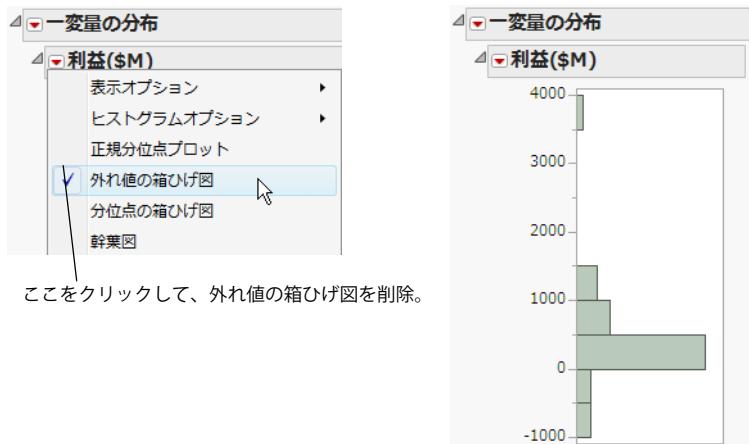
- Windowsの場合、ウィンドウ上部の青いバーの上にカーソルを置くと、メニューバーとツールバーが表示されます。
- Windowsの場合、右下隅のデータテーブルボタンをクリックすると、現在のレポートを作成するのに使用されたデータテーブルが表示されます。Macintoshの場合、レポートウィンドウの右上隅の【データテーブルの表示】ボタンをクリックします。
- Windowsの場合、右下隅のJMPホームウィンドウボタンをクリックすると、ホームウィンドウが表示されます。Macintoshでは、【ウィンドウ】 > 【JMPホーム】を選択します。

ステップ2: 箱ひげ図を削除する

前の手順で作成した「一変量の分布」レポートをそのまま使用します。

1. 「利益(\$M)」の横にある赤い三角ボタンをクリックして、レポートオプションのメニューを表示します。
2. [外れ値の箱ひげ図] の選択を解除して、該当オプションをオフにします。
外れ値の箱ひげ図がレポートウィンドウから削除されます。

図2.7 外れ値の箱ひげ図の削除



ステップ3: 追加の出力を要求する

引き続き同じレポートウィンドウを使用します。

1. 「利益(\$M)」の横にある赤い三角ボタンをクリックして、メニューから [平均の検定] を選択します。
「平均の検定」ウィンドウが表示されます。
2. 「仮説平均を指定」ボックスに「500」と入力します。
3. [OK] をクリックします。

平均に対する検定の結果がレポートウィンドウに追加されます。

図2.8 平均の検定



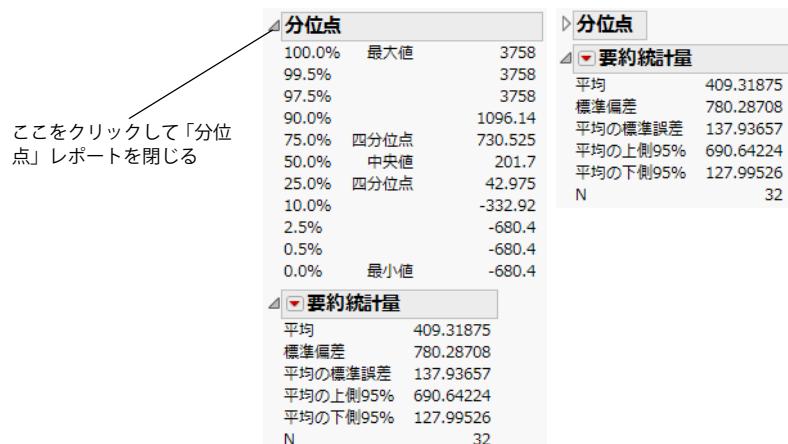
ステップ4: プラットフォームの結果を操作する

すべてのプラットフォームに表示される結果は、対話的に操作できます。たとえば、次の操作が可能です。

- レポートの表示／非表示を切り替えることができます。
- 目的に合わせて、グラフや統計分析を追加または削除できます。
- プラットフォームの結果は、元のデータテーブルや、他の結果と相互にリンクしています。

たとえば、「分位点」レポートを閉じるには、「分位点」の横にある開閉ボタンをクリックします。

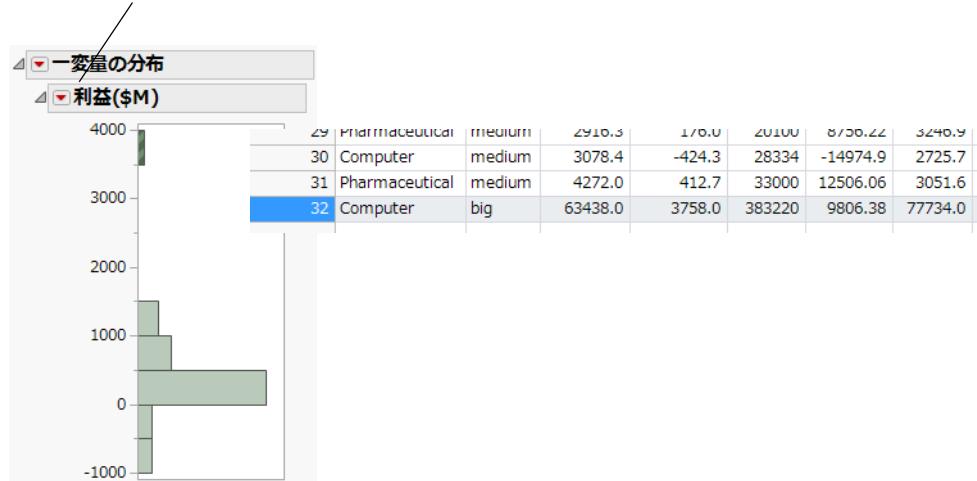
図2.9 「分位点」レポートを閉じる



プラットフォームの結果は、元のデータテーブルにリンクしています。図2.10のヒストグラムは、1つの会社グループが、他社よりもはるかに利益を上げていることを示しています。この会社グループをすばやく特定するには、ヒストグラム上の対応する棒をクリックします。すると、データテーブル内の対応する行が選択されます。

図2.10 プラットフォームの結果とデータテーブルとの対応

バーをクリックして、対応する行を表示。



この場合、該当グループには1社しか含まれていないため、1行だけが選択されています。

JMP と Excel の相違点

JMPには、Excelなどのスプレッドシートアプリケーションと重要な相違点がいくつかあります。

表2.1 JMP と Excel の違い

計算式	
Excel	計算式は個々のセルに対して適用されます。
JMP	計算式は列全体にのみ適用されます。「データの操作」の章の「 計算式による値の計算 」(75ページ)で、計算式の使用方法について説明します。
列名	
Excel	列名はグリッドの一部です。データが入力されていない部分にも、番号が振られた行やラベルの付いた列が続けて表示されます。数値データと文字データが同じ列に混在できます。
JMP	列名はグリッドの一部には含まれません。データがない部分には行や列は存在しません。グリッドの大きさは、データが入力されている部分とまったく同じ大きさになります。列には数値または文字のいずれかのみ入力できます。文字と数値データの両方が列に含まれている場合、列全体のデータタイプは文字になり、数値も文字データとして扱われます。

表2.1 JMP と Excel の違い（続き）

「データの分析」の章の「尺度の理解」（127ページ）では、プラットフォームの結果に対するデータタイプの影響について説明します。	
テーブルとワークシート	
Excel	1つのスプレッドシートに複数のテーブル（ワークシート）があります。
JMP	JMP には、ワークシートという概念がありません。各データテーブルは個別の .jmp ファイルであり、個別のウィンドウに表示されます。
データグリッド	
Excel	データはデータグリッドの任意の位置に入力できます。
JMP	データは、1行目の1列目から開始します。
分析とグラフのレポート	
Excel	データ、分析、グラフはすべてデータグリッド内に配置されます。
JMP	結果は個別のウィンドウに表示されます。

第3章

データの操作 グラフ作成や統計分析に向けたデータ準備

グラフを作成したり、統計分析したりする前に、対象のデータをデータテーブルに適切な形式で用意しておかなければなりません。この章では、次のような基本的なデータ処理作業を紹介します。

- データテーブルの新規作成
- 既存のデータテーブルを開く
- 他のアプリケーションからJMPへのデータの読み込み
- データの加工

図3.1 データテーブルの例

The screenshot shows the JMP software interface with a data table titled "Companies". The table contains 21 rows of data with the following columns: タイプ (Type), 会社規模 (Company Size), 売上(\$M) (Sales (\$M)), 利益(\$M) (Profit (\$M)), 従業員数 (Number of Employees), 従業員一人あたりの利益 (Profit per Employee), 資産 (Assets), and 利益/売上 (単位:%) (Profit/Sales (Unit: %)). The data includes various companies like Computer and Pharmaceutical, categorized by size (small, big, medium). The table also shows summary statistics at the bottom: 行 (Rows) 32, 選択されている行 (Selected Rows) 0, 除外されている行 (Excluded Rows) 0, 表示しない行 (Hidden Rows) 0, and ラベルのついた行 (Labeled Rows) 0.

タイプ	会社規模	売上(\$M)	利益(\$M)	従業員数	従業員一人あたりの…	資産	利益/売上 (単位:%)
1 Computer	small	855.1	31.0	7523	4120.70	615.2	3.63
2 Pharmaceutical	big	5453.5	859.8	40929	21007.11	4851.6	15.77
3 Computer	small	2153.7	153.0	8200	18658.54	2233.7	7.10
4 Pharmaceutical	big	6747.0	1102.2	50816	21690.02	5681.5	16.34
5 Computer	small	5284.0	454.0	12068	37620.15	2743.9	8.59
6 Pharmaceutical	big	9422.0	747.0	54100	13807.76	8497.0	7.93
7 Computer	small	2876.1	333.3	9500	35084.21	2090.4	11.59
8 Computer	small	709.3	41.4	5000	8280.00	468.1	5.84
9 Computer	small	2952.1	-680.4	18000	-37800.0	1860.7	-23.05
10 Computer	small	784.7	89.0	4708	18903.99	955.8	11.34
11 Computer	small	1324.3	-119.7	13740	-8711.79	1040.2	-9.04
12 Pharmaceutical	medium	4175.6	939.5	28200	33315.60	5848.0	22.50
13 Computer	big	11899.0	829.0	95000	8726.32	1007…	6.97
14 Computer	small	873.6	79.5	8200	9695.12	808.0	9.10
15 Pharmaceutical	big	9844.0	1082.0	83100	13020.46	7919.0	10.99
16 Pharmaceutical	small	969.2	227.4	3418	66530.13	784.0	23.46
17 Pharmaceutical	medium	6698.4	1495.4	34400	43470.93	6756.7	22.32
18 Computer	big	5956.0	412.0	56000	7357.14	4500.0	6.92
19 Pharmaceutical	big	5903.7	681.1	42100	16178.15	8324.8	11.54
20 Computer	medium	2959.3	252.8	31404	8049.93	5611.1	8.54
21 Pharmaceutical	small	1198.3	86.5	8527	10144.25	1791.7	7.22

データの取り込み

- 別のアプリケーションからデータをコピーし、貼り付ける場合は、「データのコピーと貼り付け」(62ページ) を参照してください。
- 別のアプリケーションからデータを読み込む場合は、「データの読み込み」(62ページ) を参照してください。
- データテーブルにデータを直接入力する場合は、「データの入力」(65ページ) を参照してください。
- データテーブルを開くには、目的のファイルをダブルクリックするか、[ファイル] > [開く] コマンドを使用します。

データベースから JMP にデータを読み込むこともできます。詳細については、『JMP の使用法』の「データの読み込み」の章を参照してください。

この章では、サンプルのデータテーブルと読み込み用データを使用します。これらのファイルは、JMPと一緒にインストールされます。これらのファイルについては、「JMP の基本」の章の「サンプルデータの使用」(51ページ) を参照してください。

データのコピーと貼り付け

Excel やテキストファイルなどの別のアプリケーションから、コピーして貼り付ける方法で、データを JMP に移動できます。

- Excel で「VA Lung Cancer.xls」ファイルを開きます。このファイルは、「Sample Import Data」フォルダにあります。
- 列名も含め、すべての行と列を選択します。138 行 12 列のデータがあります。
- 選択したデータをコピーします。
- JMP で、[ファイル] > [新規作成] > [データテーブル] を選択して、空のテーブルを作成します。
- [編集] > [列名とともに貼り付け] を選択して、データと列見出しを貼り付けます。

JMP に貼り付けるデータに列名が含まれていない場合は、[編集] > [貼り付け] を使用します。

データの読み込み

Excel、SAS、テキストファイルなどの別のアプリケーションから読み込む方法で、データを JMP に移動できます。データの基本的な読み込み手順は次のとおりです。

- [ファイル] > [開く] を選択します。
- 読み込み対象ファイルの場所に移動します。
- 目的のファイルが「データファイルを開く」ウィンドウの一覧に表示されない場合は、[ファイルの種類] メニューから適切なファイル形式を選択します。
- [開く] をクリックします。

Microsoft Excel ファイルの読み込み例

1. [ファイル] > [開く] を選択します。
2. 「Samples」フォルダ配下の「Import Data」フォルダに移動します。
3. 「Team Results.xls」を選択します。

データの開始行および開始列を書き留めておきます。このスプレッドシートには2つのワークシートが含まれています。この例では、Ungrouped Team Results ワークシートを読み込みます

4. [開く] をクリックします。

このスプレッドシートがExcel読み込みウィザード内に開き、データのプレビューと読み込みオプションが表示されます。

スプレッドシートの1行目のテキストが列見出しになっていますが、列見出しとして使用したいテキストは3行目になります。

5. 「列見出しの開始行」フィールドに「3」と入力し、Enterキーを押します。データプレビューの列見出しが更新されます。データの開始行の値は4に更新されます。
6. 設定をこのワークシートに対してのみ保存します。
 - ウィンドウの左下にある【すべてのワークシートに使用】チェックボックスをオフにします。
 - ウィンドウの右上で「**Ungrouped Team Results**」を選択します。
7. [読み込み] をクリックします。スプレッドシートが指定した形式に変換されます。

Excel ファイルをJMPに読み込む際、JMPは列名が存在するかどうかと、その列名が1行目にあるかどうかを推測します。次のような場合は、コピーして貼り付ける方法を推奨します。

- 列名が1行目以外の行にある場合
- ファイルに列名が含まれておらず、データが1行目から始まっていない場合
- ファイルに列名が含まれていて、データが2行目から始まっていない場合

Excel ファイルの読み込みの詳細については、「[データのコピーと貼り付け](#)」(62ページ)、および『JMPの使用法』の「データの読み込み」の章を参照してください。

テキストファイルの読み込み例

テキストファイルを読み込む方法の1つは、対象となるデータのレイアウトをあらかじめ設定しておいて、その後、テキストデータを読み込む方法です。この方法であらかじめ設定しておく項目は、「環境設定」で指定できます。テキスト読み込みの環境設定の詳細については、『JMPの使用法』の「JMP環境設定」の章を参照してください。

テキストファイルを読み込むもう1つの方法は、テキストのプレビューウィンドウを使用して、データテーブルが読み込み後にどうなるのかを確認しながら調整する方法です。次の例では、テキスト読み込みのプレビューウィンドウの使用方法を紹介します。

1. [ファイル] > [開く] を選択します。
2. 「Samples」フォルダ配下の「Import Data」フォルダに移動します。

3. 「Animals_line3.txt」を選択します。
4. 「データファイルを開く」ウィンドウの下部で、[データとしてプレビューで開く] を選択します。
5. [開く] をクリックします。

図3.2 最初に表示されるプレビューウィンドウ



このテキストファイルには、1行目にタイトル、3行目に列名があり、データ自体は4行目から始まっています。このファイルをJMPで直接開くと、「Animals Data」という行が最初の列名になり、すべての列名とその後に続くデータは適切に表示されません。プレビューウィンドウでは、ファイルを開く前に設定を調整することができ、また、調整後の最終的なデータテーブルがどのようになるかを確認できます。

6. 「次の行に列名を含む」フィールドに「3」と入力します。

7. 「データの開始行」フィールドに「4」と入力します。

8. [次へ] をクリックします。

2番目のウィンドウで、読み込み対象から除外する列を選択し、列の尺度を変更できます。この例では、デフォルトの設定をそのまま使用します。

9. [読み込み] をクリックします。

新しいデータテーブルには、「species」、「subject」、「miles」、「season」という名前の列があります。「species」と「season」は文字データの列です。「subject」と「miles」は連続尺度の数値データの列です。

データの入力

データテーブルにデータを直接入力することもできます。次の例では、数か月にわたって収集したデータをデータテーブルに入力する方法を紹介します。

シナリオ

表3.1は、高血圧の新薬に対する臨床試験データです。各被験者の血圧を6か月間にわたり測定しました。2つの用量（300mgと450mg）の投与群と、対照群および偽薬群に分けて観察されています。データは、各群の平均血圧を示しています。

表3.1 血圧データ

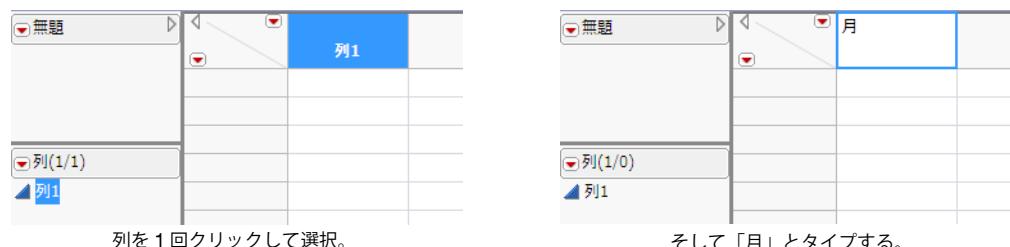
月	対照群	偽薬群	300mg	450mg
三月	165	163	166	168
四月	162	159	165	163
五月	164	158	161	153
六月	162	161	158	151
七月	166	158	160	148
八月	163	158	157	150

新規データテーブルへのデータの入力

- [ファイル] > [新規作成] > [データテーブル] を選択して、空のデータテーブルを作成します。
新しいデータテーブルには列が1つあり、行はありません。
- 列名を選択して、それを「月」に変更します。図3.3を参照してください。

列名をダブルクリックするか、または列を選択して Enter キーを押すことによっても、列名を変更することができます。

図3.3 列名の入力



- [行] > [行の追加] を選択します。
「行の追加」ウィンドウが表示されます。

4. 6行追加する必要があるため、「6」と入力します。
5. 【OK】をクリックします。空の行が6つ、データテーブルに追加されます。
6. 各セルをダブルクリックして、「月」のデータを入力します。

図3.4 完成した「月」列

	月
1	三月
2	四月
3	五月
4	六月
5	七月
6	八月

「列」パネルで、列名の左側にある尺度アイコンを確認します。「月」が名義尺度になったことを反映してアイコンが変化しています（変化前は連続尺度）。図3.3で「列1」に対して表示されている尺度と、図3.4で「月」に対して表示されている尺度を比較してください。この違いは重要です。詳細については、「[列情報の表示または変更](#)」(74ページ)で説明します。

7. 「月」列の右側のスペースをダブルクリックし、「**対照群**」用の列を追加します。
8. 名前を「**対照群**」に変更します。
9. 「**対照群**」のデータを表3.1に示されているとおりに入力します。これで、データテーブルは6行2列構成になりました。
10. 引き続き列を追加し、表3.1に示されているとおりにデータを入力して、最終的に6行5列から成るデータテーブルを作成します。

データテーブル名の変更

1. テーブルパネルでデータテーブル名（「無題」）をダブルクリックします。
2. 新しい名前（「血圧」）を入力します。

図3.5 データテーブル名の変更



Excel からのデータの転送

Excel 用の JMP アドインを使って、Excel シートを JMP の次のものへ転送することができます。

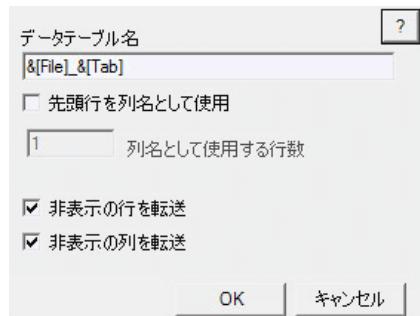
- データテーブル
- グラフビルダー
- 「一変量の分布」 プラットフォーム
- 「二変量の関係」 プラットフォーム
- 「モデルのあてはめ」 プラットフォーム
- 「時系列分析」 プラットフォーム
- 「管理図」 プラットフォーム

Excel で JMP アドインの環境設定を設定する

JMP アドインの環境設定を設定するには、次の手順に従ってください。

1. Excel で、[JMP] > [環境設定] を選択します。
「JMP 環境設定」 ウィンドウが開きます。

図3.6 JMP アドインの環境設定



2. 「データテーブル名」 ボックスにデフォルトで表示されている名前（「ファイル名_ワークシート名」）をそのまま使うか、別の名前を入力します。
3. ワークシートの最初の行に列見出しが含まれている場合は、[先頭行を列名として使用] をオンにします。
4. その場合、列見出しとして使用する行の数も指定します。
5. ワークシートに非表示の行があり、それも JMP データテーブルに含めたい場合は、[非表示の行を転送] をオンにします。
6. ワークシートに非表示の列があり、それも JMP データテーブルに含めたい場合は、[非表示の列を転送] をオンにします。
7. [OK] をクリックして設定を保存します。

JMPへの転送

Excel ワークシートを JMP に転送するには、次の手順に従ってください。

1. Excel ファイルを開きます。
2. 転送するワークシートを選択します。
3. [JMP] を選択し、続いて転送先を選択します。
 - データテーブル
 - グラフビルダー
 - 「一変量の分布」 プラットフォーム
 - 「二変量の関係」 プラットフォーム
 - 「モデルのあてはめ」 プラットフォーム
 - 「時系列分析」 プラットフォーム
 - 「管理図」 プラットフォーム

Excel ワークシートが JMP でデータテーブルとして開き、指定したプラットフォームの起動ウィンドウが表示されます。

データテーブルの操作

ここでは、次の情報を取り上げます。

- [「データの編集」 \(69ページ\)](#)
- [「値の選択、選択解除、検索」 \(70ページ\)](#)
- [「列情報の表示または変更」 \(74ページ\)](#)
- [「計算式による値の計算」 \(75ページ\)](#)
- [「データのフィルタ処理」 \(77ページ\)](#)

ヒント：開いたデータテーブルを一定の時間ごとに自動保存したいときは、[一般] 環境設定で [自動保存の間隔(分)] を設定します。この自動保存の間隔は、ジャーナル、スクリプト、プロジェクト、レポートにも適用されます。

データの編集

一度にいくつかのセル、または列全体に対して、データを入力または変更できます。ここでは、次の情報を取り上げます。

- ・「[値の変更](#)」(69ページ)
- ・「[値の再コード化](#)」(69ページ)
- ・「[パターンを持つデータの作成](#)」(70ページ)

値の変更

値を変更するには、セルを選択した後、変更後の値を入力します。セルをダブルクリックして編集することもできます。

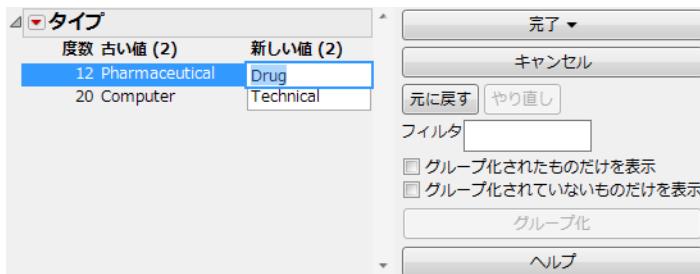
注: セルをダブルクリックした場合の動作は、セルを選択した場合とは異なります。1回だけクリックした場合は、セルが選択されます。セルの選択においては、複数のセルを同時に選択し、選択したセル範囲に対して特定の操作を実行することもできます。ダブルクリックした場合は、そのセルを編集できるようになるだけです。行、列、セルの選択方法の詳細については、「[値の選択、選択解除、検索](#)」(70ページ) を参照してください。

値の再コード化

再コード化を用いると、列の値を一度に変更できます。たとえば、コンピュータ関連企業と製薬会社の売上を比較するとしましょう。現在、業種のデータは「Computer」と「Pharmaceutical」になっています。これらを「Technical」と「Drug」に変更します。32行のデータをすべて調べて、値を1つ1つ変更していくのは非効率で単調である上、データが多い場合には誤入力する可能性もあります。値を1つ1つ変更するよりも、再コード化した方が得策です。

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. 「タイプ」列の見出しを1回クリックして列を選択します。
3. [列] > [ユーティリティ] > [再コード化] を選択します。
4. 「再コード化」ウィンドウの「新しい値」列で、「Computer」の行に「Technical」、「Pharmaceutical」の行に「Drug」と入力します。
5. [完了] をクリックし、表示されたリストから [値の置換] を選択します。

図3.7 「再コード化」 ウィンドウ



すべてのセルが自動的に新しい値に更新されます。

パターンを持つデータの作成

[初期値を埋める/挿入] を用いると、特定のパターンをもつデータを入力できます。[初期値を埋める/挿入] の各オプションは、データテーブルが大きく、各行への値の入力に手間がかかる場合に特に便利です。

列へのパターンの入力例

- 新しい列を追加します。
 - 最初のセルに「1」、2番目のセルに「2」、3番目のセルに「3」と入力します。
 - 3つのセルを選択し、選択したセル範囲内の任意の位置を右クリックしてメニューを表示します。
 - [初期値を埋める/挿入] > [テーブルの最後までシーケンスを繰り返す] を選択します。
- 列の残りのセルに同じ連番（1、2、3、1、2、3...）が繰り返し挿入されます。

連番を繰り返す代わりに、連番を継続する場合は（1、2、3、4、5、6...）、[テーブルの最後までシーケンスを継続する] を選択します。このコマンドは、（1、1、1、2、2、2、3、3、3...）のようなパターンを作る場合にも使用できます。

[初期値を埋める/挿入] のオプションでは、単純な等差数列や等比数列が認識されます。ただし、文字データに対しては、値を繰り返すことだけしか行えません。

値の選択、選択解除、検索

データテーブル内の行、列、またはセルは選択できます。たとえば、既存のデータテーブルのサブセットを作成したい場合（テーブルの一部分を抽出したい場合）は、まず抽出したい一部分を選択する必要があります。また、行を選択して、グラフ上で該当するデータ点を強調表示することも可能です。行や列をクリックして手動で選択したり、特定の検索条件に一致する行を選択したりできます。ここでは、次の情報を取り上げます。

- 「行の選択と選択解除」 (71ページ)
- 「列の選択と選択解除」 (71ページ)
- 「セルの選択と選択解除」 (72ページ)
- 「値の検索」 (72ページ)

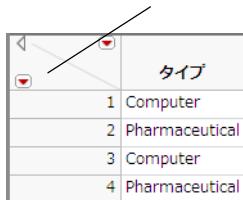
行の選択と選択解除

表3.2 行の選択と選択解除

目的	操作
行を一度に1つずつ選択する	目的の行番号をクリックします。
隣接する複数の行を選択する	開始行の行番号をクリックし、最終行の行番号までドラッグします。 または 開始行を選択してから、Shiftキーを押しながら最終行の行番号をクリックします。
隣接していない複数の行を選択する	開始行を選択してから、Ctrlキーを押しながら他の行番号を選択していきます。
行を一度に1つずつ選択解除する	Ctrlキーを押しながら行番号を選択していきます。
すべての行を選択解除する	テーブルの左上隅にある下側の三角形のスペースをクリックします。図3.8を参照してください。

図3.8 行の選択解除

すべての行を一度に選択解除するには、ここをクリック。



列の選択と選択解除

表3.3 列の選択と選択解除

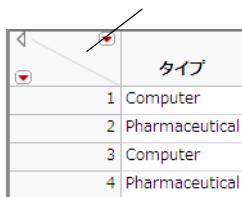
目的	操作
列を一度に1つずつ選択する	目的の列見出しをクリックします。
隣接する複数の列を選択する	開始列の列見出しをクリックし、最終列の見出しまでドラッグします。 または 開始列を選択してから、Shiftキーを押しながら最終列の見出しをクリックします。

表3.3 列の選択と選択解除（続き）

目的	操作
隣接していない複数の列を選択する	開始列を選択してから、Ctrlキーを押しながら他の列見出しを選択していきます。
列を一度に1つずつ選択解除する	Ctrlキーを押しながら列見出しを選択していきます。
すべての列を選択解除する	テーブルの左上隅にある上側の三角形のスペースをクリックします。図3.9を参照してください。

図3.9 列の選択解除

すべての列を一度に選択解除するには、ここをクリック。



タイプ
Computer
Pharmaceutical
Computer
Pharmaceutical

セルの選択と選択解除

表3.4 セルの選択と選択解除

目的	操作
セルを一度に1つずつ選択する	各セルを個別にクリックします。
隣接する複数のセルを選択する	開始セルをクリックし、最終セルまでドラッグします。 または 開始セルを選択してからShiftキーを押しながら最終セルをクリックします。
隣接していない複数のセルを選択する	開始セルを選択してから、Ctrlキーを押しながら他のセルを選択していきます。
すべてのセルを選択解除する	テーブルの左上隅にある上側および下側の三角形のスペースをクリックします。

値の検索

行数が非常に多いデータテーブルでは、テーブルをスクロールして特定のセルを見つけるのは至難の業です。特定のデータを探す場合は、検索機能で見つけることができます。データが検索条件に一致する場合は、該当セルが選択され、データグリッドがスクロールしてそのセルがウィンドウに表示されます。たとえば、会社情報のサンプルデータ「Companies.jmp」には、総売上が11,899ドルの会社の情報が含まれています。そのセルを検索するには、検索機能を使用します。

値の検索例

1. 「[編集] > [検索] > [検索]」を選択して、検索ウィンドウを開きます。
2. 「検索する文字列」ボックスに「11899」と入力します。
3. 「[検索]」をクリックします。「11899」が数値として入力された最初のセルが検索され、選択されます。

検索条件に一致するセルが複数ある場合は、「[検索]」を再度クリックすると、検索文字列に一致する次のセルが検索されます。

検索条件を満たす複数の行を一度に検索することもできます。

中規模企業の行をすべて選択する例

1. 「[行] > [行の選択] > [Where 条件で選択]」を選択して、「行の選択」ウィンドウを開きます。
2. 左側の列のリストボックスで、「会社規模」を選択します。
3. 右側のテキストボックスに「medium」と入力します。
4. 「[OK]」をクリックします。

図3.10 「行の選択」 ウィンドウ



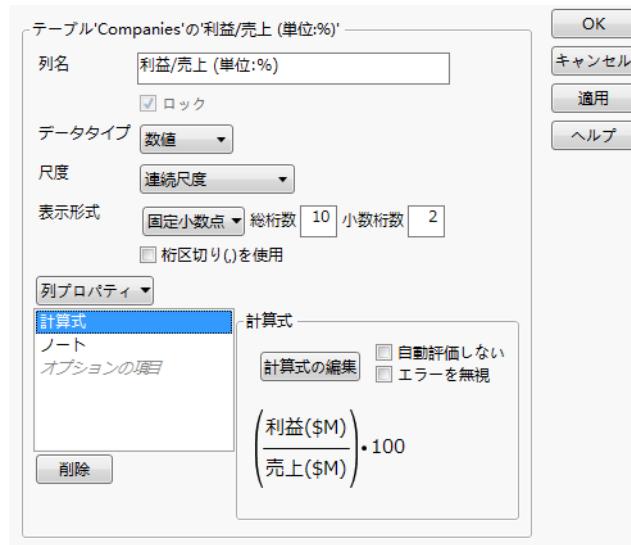
「会社規模」が「medium」に等しい行がすべて選択されます。全部で7行です。

列情報の表示または変更

列に関する情報は、列内のデータに限りません。データタイプ、尺度、表示形式、計算式も設定できます。

列の特性を表示または変更するには、列見出しをダブルクリックします。または、列見出しを右クリックし、[列情報] を選択します。列情報のウィンドウが表示されます。

図3.11 列情報のウィンドウ



列名 列名を入力または変更します。2つの列に同じ名前を指定することはできません。

データタイプ 次のデータタイプのいずれか1つを選択します。

- [数值] は、列の値を数値として指定します。
- [文字] は、列の値を数値以外の値（文字や記号など）として指定します。
- [行の属性] は、列の値を行の属性として指定します。これは上級ユーザ向けの内容です。『JMP の使用法』の「列情報ウィンドウ」の章を参照してください。

尺度 尺度は、分析時に値をどのように使用するかを定義します。次のいずれかの尺度を選択します。

- [連續尺度] の値は、数値のみです。
- [順序尺度] の値は、順序に意味のある数値または文字の質的変数です。
- [名義尺度] の値は、順序に意味のない数値または文字の質的変数です。

表示形式 数値の形式を選択します。このオプションは、文字データでは使用できません。以下は、一般的な表示形式です。

- [最適] を指定すると、最適な表示形式が自動的に選択されます。

- [固定小数点] では、小数点以下の表示桁数を指定します。
- [日付] では、日付値に対する表示形式を指定します。
- [時間] では、時間値に対する表示形式を指定します。
- [通貨] では、通貨の種類と、通貨値の小数点以下の表示桁数を指定します。

列プロパティ 計算式、ノート、値の順序など、列の特別なプロパティを設定します。『JMP の使用法』の「列情報ウィンドウ」の章を参照してください。

ロック 列の値が変更されないように、該当列をロックします。

計算式による値の計算

計算式エディタを使用して、計算された値を含む列を作成できます。

シナリオ

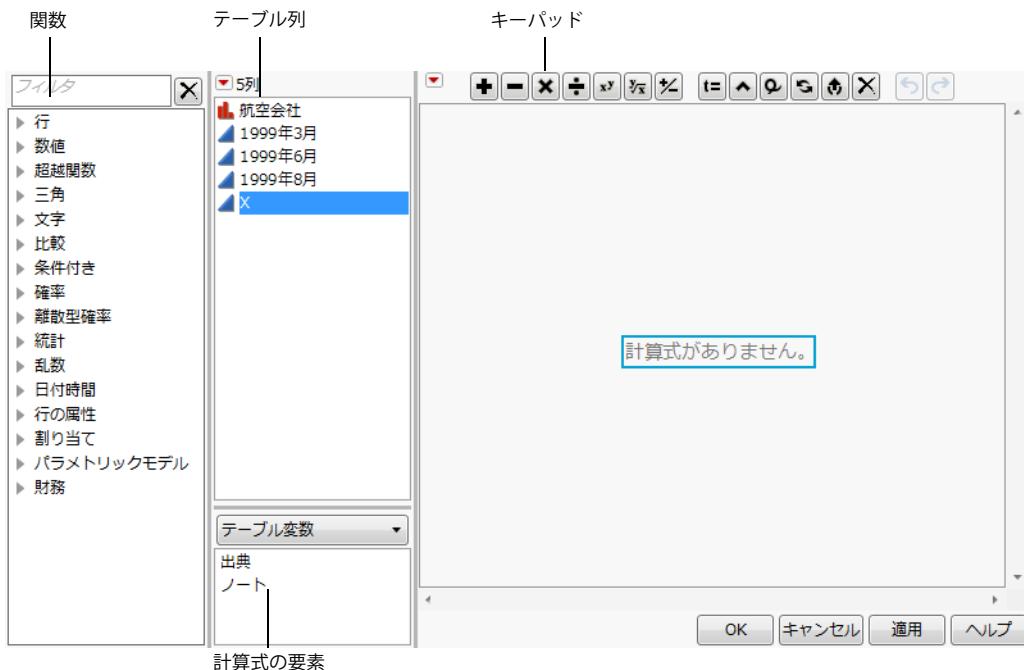
サンプルデータ「On-Time Arrivals.jmp」には、航空会社数社の定刻到着率が記録されています。1999年の3月、6月、および8月のデータが収集されました。

式の作成

各航空会社の平均定刻到着率を表示する列を新規作成しましょう。

1. 新しい列を追加します。
2. 新しい列の列見出しを右クリックし、[計算式] を選択します。計算式エディタウィンドウが表示されます。

図3.12 計算式エディタ



各航空会社の平均定刻到着率を算出する式を作成します。

3. 「列」リストから「1999年3月」を選択します。
4. キーパッドの $+$ ボタンをクリックします。
5. 「1999年6月」を選択し、もう一度 $+$ 記号をクリックします。
6. 「1999年8月」を選択します。

図3.13 月の和

$1999\text{年}3\text{月} + 1999\text{年}6\text{月} + 1999\text{年}8\text{月}$

「1999年8月」だけが選択されている（赤いボックスで囲まれている）点に注意してください。

7. 式全体を囲んでいるボックスをクリックします。

図3.14 式全体が選択された状態

$1999\text{年}3\text{月} + 1999\text{年}6\text{月} + 1999\text{年}8\text{月}$

8. $\frac{\oplus}{\ominus}$ ボタンをクリックします。

9. 分母ボックスに「3」と入力してから、式の外側の白いスペースをクリックします。

図3.15 完成した式

$$\frac{[1999年3月] + [1999年6月] + [1999年8月]}{3}$$

10. [OK] をクリックします。

新しい列に平均値が表示されます。

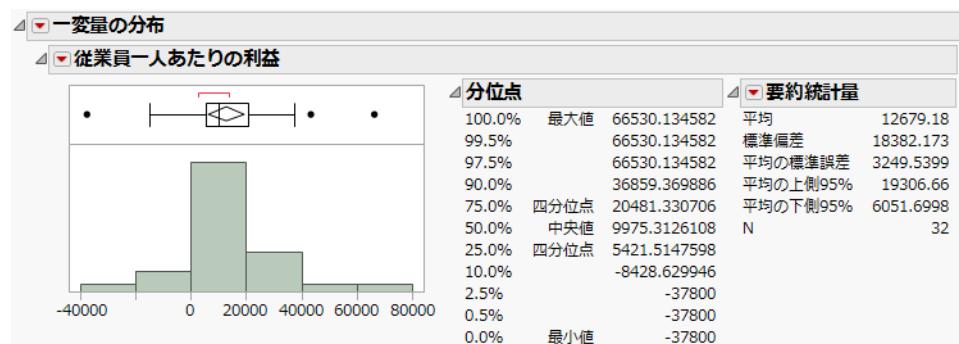
計算式エディタには、多くの算術関数と統計関数が用意されています。たとえば、平均定期到着率を計算する別の方法として、「統計」関数リストの「Mean」関数を使用することもできます。計算式エディタの関数についての詳細は、『JMPの使用法』の「計算式エディタ」「計算式エディタ」の章を参照してください。

データのフィルタ処理

[データフィルタ] を使用すると、データの様々なサブセット（一部分）を対話的に選択し、これらのサブセットをプロット上で非表示にしたり、分析から除外したりできます。たとえば、コンピュータ関連企業（「Computer」）と製薬会社（「Pharmaceutical」）の従業員一人あたりの利益を検討してみましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Companies.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「従業員一人あたりの利益」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. 「従業員一人あたりの利益」の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] > [横に並べる] を選択します。

図3.16 「従業員一人あたりの利益」の一変量の分布



6. 「一変量の分布」の赤い三角ボタンをクリックし、メニューから [やり直し] > [自動再計算] を選択して、自動再計算をオンにします。

このオプションがオンの場合、変更（データ点の非表示または除外など）を加えるたびに、レポートウィンドウが自動的に更新されます。

7. データテーブル内で、[行] > [データフィルタ] を選択します。

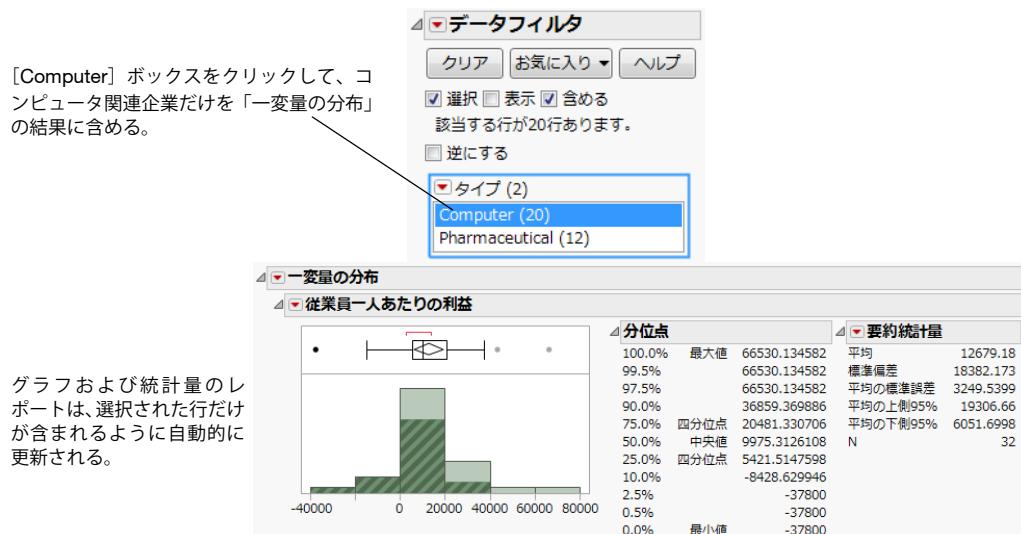
8. 「タイプ」を選択し、[追加] をクリックします。

9. [選択] と [含める] が選択されていることを確認します。

10. 製薬会社を「一変量の分布」の結果から除外し、コンピュータ関連企業だけを含める場合は、「データフィルタ」ウィンドウで [Computer] ボタンをクリックします。

一変量の分布の結果において、コンピュータ関連企業だけが含まれるように更新されます。

図3.17 コンピュータ関連企業だけを表示するフィルタ



反対に、「一変量の分布」の結果に、製薬会社だけを含める場合は、「データフィルタ」ウィンドウで [Pharmaceutical] ボタンをクリックします。

データの加工

「テーブル」メニューのコマンド群（および【分析】メニューの「表の作成」）には、要約統計量を求めたり、データテーブルを加工してグラフ作成や分析に必要な形式に変換したりするものがあります。ここでは、これらのコマンドの中から5つを紹介します。

要約 要約統計量（記述統計量）を含むテーブルを作成します。

表の作成 ドラッグ＆ドロップのマウス操作によって、要約統計量の表を作成できます。

サブセット データのサブセット（一部分）だけを抽出したテーブルを作成します。

結合(Join) 2つのデータテーブルのデータを結合し、新しい1つのデータテーブルを作成します。

並べ替え 1つまたは複数の列を基準に、データを並べ替えます。

上記を含む「テーブル」メニューのコマンドの詳細については、『JMPの使用法』の「データの再構成」「データの再構成」の章を参照してください。

要約統計量の表示

合計や平均などの要約統計量を求めるとき、データに関する有用な情報をすばやく確認できます。たとえば、32社それぞれの利益を見るだけでは、会社規模による利益の違いを把握するのは困難です。要約統計量を求めれば、会社規模による違いをすぐに確認できます。

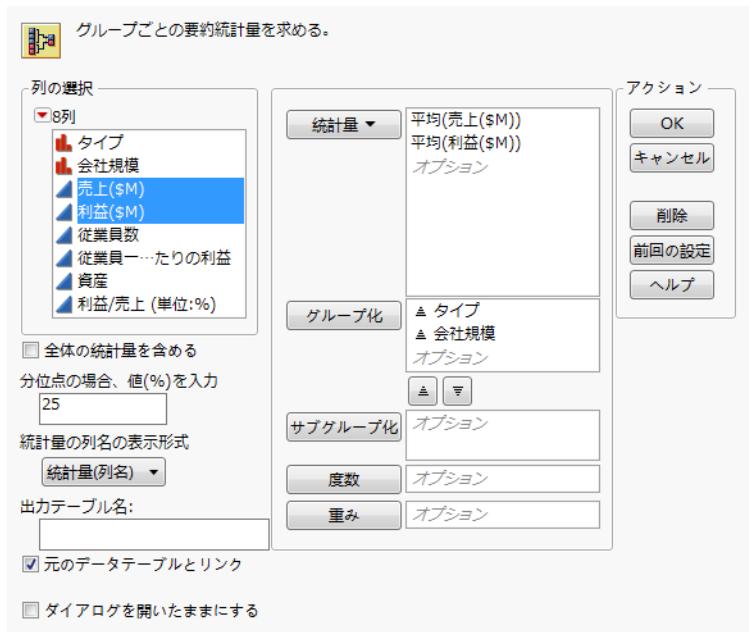
[要約] コマンドまたは【表の作成】コマンドのいずれかによって、要約表を作成できます。【要約】コマンドでは、新しいデータテーブルが作成されます。他のデータテーブルと同様に、要約データテーブルに対しても分析を実行したり、グラフを作成できたりします。【表の作成】コマンドでは、要約表が表示されたレポートウィンドウが作成されます。「表の作成」レポートからテーブルを作成することもできます。

要約

要約テーブルには、グループ変数の水準ごとに要約統計量が表示されます。たとえば、コンピュータ関連企業と製薬会社の財務データを検討してみます。会社のタイプと規模の組み合わせごとに、平均売上と平均利益を求めるものとします。

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. [テーブル] > [要約] を選択します。
3. 「タイプ」と「会社規模」を選択し、[グループ化] をクリックします。
4. 「売上(\$M)」と「利益(\$M)」を選択し、[統計量] > [平均] をクリックします。

図3.18 設定後の「要約」ウィンドウ



5. [OK] をクリックします。

「タイプ」と「会社規模」の組み合わせごとに、「売上(\$M)」の平均と「利益(\$M)」の平均が計算されます。

図3.19 要約テーブル

The screenshot shows the 'Companies' summary table. The left pane shows the summary structure with 'Companies' as the source, 'Type' and 'Company Size' as grouping variables, and 'Sales(\$M)' and 'Profit(\$M)' as average statistics. The right pane displays the resulting table with 6 rows, showing the average sales and profit for each combination of Type and Company Size.

	タイプ	会社規模	行数	平均(売上(\$M))	平均(利益(\$M))
1	Computer	big	4	20597.48	1089.93
2	Computer	medium	2	3018.85	-85.75
3	Computer	small	14	1758.06	44.94
4	Pharmaceutical	big	5	7474.04	894.42
5	Pharmaceutical	medium	5	4261.06	698.98
6	Pharmaceutical	small	2	1083.75	156.95
すべての行					
選択されている行			6		
除外されている行			0		
表示しない行			0		
ラベルのついた行			0		

要約テーブルには、次の情報が表示されます。

- 各グループ変数（この例では、「タイプ」と「会社規模」）の列があります。
- 「行数」列には、グループ変数の各組み合わせに対応する、元のテーブルの行数が表示されます。たとえば、元のデータテーブルには、小規模（small）なコンピュータ関連企業（Computer）に対応する行が14行あります。

- 要求された要約統計量ごとに1つずつ列があります。この例では、「売上(\$M)」の平均の列と、「利益(\$M)」の平均の列があります。

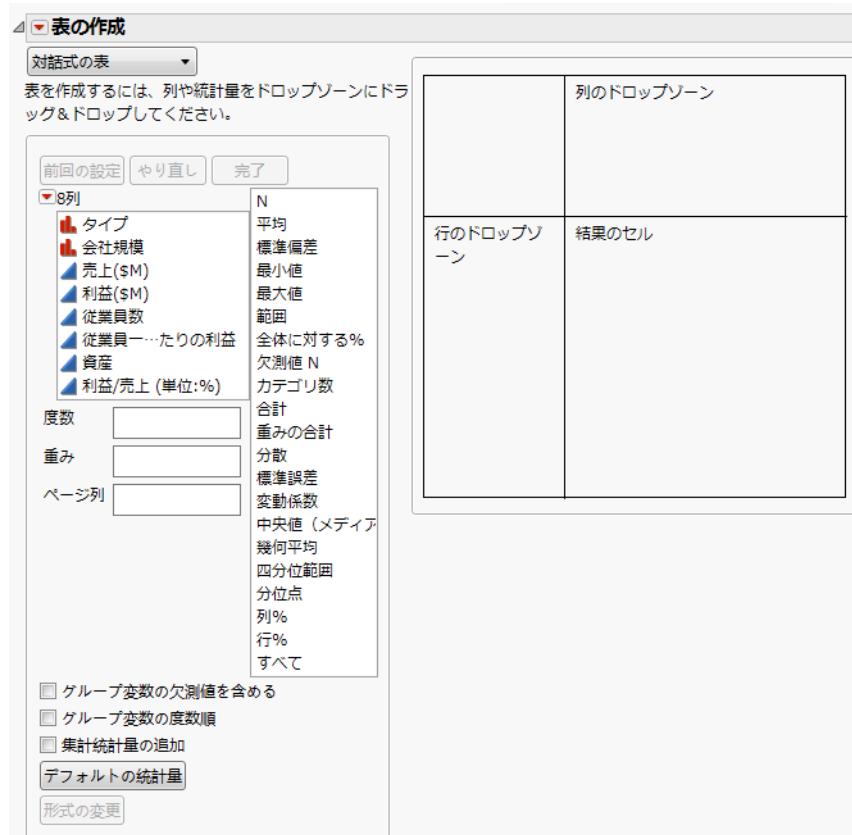
要約テーブルは、元のテーブルとリンクしています。要約テーブルで行を選択すると、元のテーブルでも対応する行が選択されます。

表の作成

「表の作成」コマンドでは、作業スペースに列をドラッグする操作方法により、グループ変数の組み合わせごとの要約統計量を求めることができます。この例では、先ほど「要約」を使用して作成した要約情報と同じものを、「表の作成」を使用して求める方法を紹介します。

- [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。
- [分析] > [表の作成] を選択します。

図3.20 「表の作成」作業スペース



- 「タイプ」と「会社規模」の両方を選択します。

4. 両者を「行のドロップゾーン」にドラッグ＆ドロップします。

図3.21 列を行ゾーンにドラッグ



5. 見出しを右クリックし、[グループ変数を入れ子にする]を選択します。

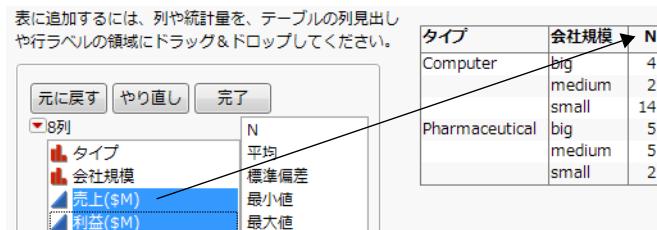
最初に表示される表には、グループごとの行数が示されます。

図3.22 最初に表示される表

タイプ	会社規模	N
Computer	big	4
	medium	2
	small	14
Pharmaceutical	big	5
	medium	5
	small	2

6. 「売上(\$M)」と「利益(\$M)」の両方を選択し、表の「N」の上にドラッグ＆ドロップします。

図3.23 「売上」と「利益」の追加



表には、グループごとの「売上(\$M)」の合計と、「利益(\$M)」の合計が表示されます。

図3.24 合計を示す表

タイプ	会社規模	売上(\$M)	利益(\$M)
		合計	合計
Computer	big	82389.9	4359.7
	medium	6037.7	-171.5
	small	24612.8	629.1
Pharmaceutical	big	37370.2	4472.1
	medium	21305.3	3494.9
	small	2167.5	313.9

7. 最後に、合計値を平均値に変更します。「合計」(いずれか)を右クリックし、[統計量]>[平均]を選択します。

図3.25 最終的な表

タイプ	会社規模	売上(\$M)	利益(\$M)
		平均	平均
Computer	big	20597.48	1089.9
	medium	3018.85	-85.75
	small	1758.06	44.94
Pharmaceutical	big	7474.04	894.42
	medium	4261.06	698.98
	small	1083.75	156.95

平均値は、[要約] コマンドで得た値と同じです。図3.25と図3.19を比較してください。

サブセットの作成

サブセットを作成することにより、データテーブルの一部分だけを綿密に検討することができます。たとえば、コンピュータ関連企業と製薬会社の売上と利益に関して、大中小の会社規模での比較はすでに行つたとしましょう。次に、中企業（「medium」）だけの売上と利益を検討したいとします。

サブセットは2段階で作成します。まず、対象となるデータを選択してから、そのデータを新しいテーブルに抽出します。

[サブセット] コマンドによるサブセットの作成

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。

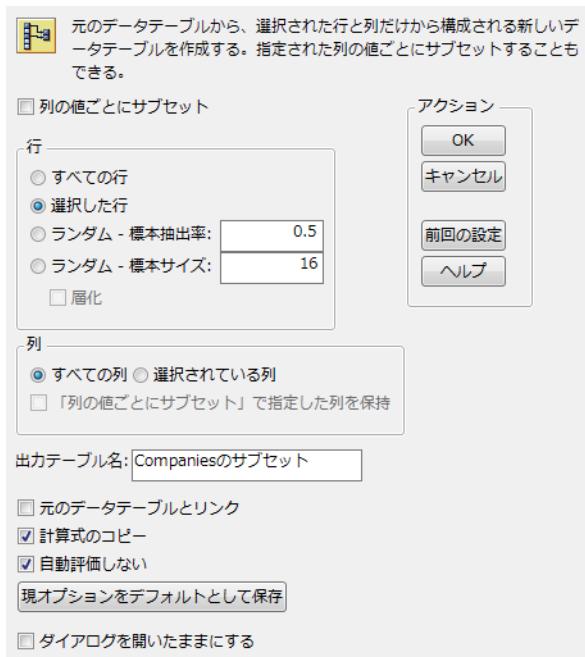
サブセットにする行と列を選択する

2. [行] > [行の選択] > [Where条件で選択] を選択します。
3. 左側の列のリストボックスで、「会社規模」を選択します。
4. テキスト入力ボックスに「medium」と入力します。
5. [OK] をクリックします。
6. Ctrl キーを押したまま、「タイプ」列、「売上(\$M)」列、「利益(\$M)」列を選択します。

サブセットテーブルを作成する

7. [テーブル] > [サブセット] を選択して、「サブセット」 ウィンドウを開きます。

図3.26 「サブセット」 ウィンドウ



8. 選択した列だけをサブセットするには、[選択されている列] を選択します。さらにオプションを選択して、サブセットテーブルをカスタマイズすることもできます。

9. [OK] をクリックします。

サブセットした結果のデータテーブルは、7行3列になります。[サブセット] コマンドの詳細については、『JMPの使用法』の「データの再構成」「データの再構成」の章を参照してください。

「一変量の分布」 プラットフォームでのサブセットの作成

プラットフォームの結果がデータテーブルとリンクしていることを利用して、サブセットを作成する方法もあります。

[一変量の分布] コマンドを使用したサブセットの作成例

- [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。
- [分析] > [一変量の分布] を選択します。
- 「タイプ」を選択し、[Y, 列] をクリックします。
- [OK] をクリックします。

5. 「Computer」を表すヒストグラムの棒をダブルクリックすると、コンピュータ関連企業だけを含むテーブルが作成されます。

警告: この方法でサブセットされた結果のデータテーブルは、元のテーブルとリンクしています。つまり、サブセットされたテーブルでデータを変更した場合は、元のテーブルでも対応する値が変更されます。

データテーブルの結合

[結合(Join)] オプションを使用して、複数のデータテーブルを1つのデータテーブルに結合できます。たとえば、ポップコーン収率に関する実験結果を記録したデータテーブルがあるとしましょう。また、別のデータテーブルには、ポップコーン収率の2回目の実験結果が記録されているとします。2つの実験を比較したり、組み合わせて分析したりする場合は、これらのデータを同じテーブルにまとめる必要があります。また、実験データは、2つのデータテーブルに同じ規則で入力されているわけではありません。列の1つは両テーブルで名前が異なっており、第2実験についてはデータの一部が欠測値になっています。つまり、片方のテーブルからもう片方のテーブルに単純にコピー&ペーストすることはできません。

2つのデータテーブルの結合例

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Trial1.jmp」と「Little.jmp」を開きます。
2. 「Trial1.jmp」をクリックして、このデータテーブルをアクティブにします。
3. [テーブル] > [結合(Join)] を選択します。
4. 「'Trial1' と結合するテーブル」ボックスで、「Little」を選択します。
5. [対応の指定] メニューから [対応する列の値で結合] を選択します。
6. 「元の列」にある両方のボックスで「コーンの種類」を選択し、[対応] をクリックします。
7. 同じ要領で、両方のボックスで「一回にポップする量」、および「油の量」と「油」を選択し、対応させます。
対応する列が同じ名前である必要はありません。
8. 両方のテーブルに対して [一致しない行も含める] チェックボックスをオンにします。
片方の実験データは不完全であるため、不足データも含め、すべての行を含める必要があります。
9. 重複列がないように、[テーブル結合のために列を選択] チェックボックスをオンにします。
10. 「Trial1」から4列すべてを選択し、[選択] をクリックします。
11. 「Little」からは「収率」だけを選択して [選択] をクリックします。

図3.27 設定後の「結合(Join)」ウィンドウ



12. [OK] をクリックします。

図3.28 結合後のテーブル

	コーンの種類	油の量	一回にポップする量	Trial1の収率	Littleの収率
1	グルメ	少なめ	多量	8.6	8.2
2	グルメ	多め	多量	9.2	*
3	グルメ	少なめ	少量	12.1	15.9
4	グルメ	多め	少量	18.0	*
5	プレーン	少なめ	多量	8.2	8.8
6	プレーン	多め	多量	10.4	*
7	プレーン	少なめ	少量	9.9	10.1
8	プレーン	多め	少量	10.6	*

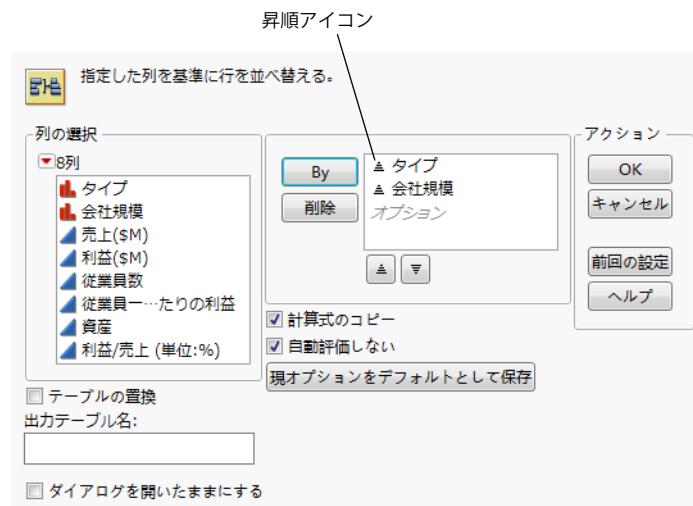
テーブルの並べ替え

[並べ替え] コマンドを使用して、データテーブルの 1 つまたは複数の列を基準に、データを並べ替えることができます。たとえば、コンピュータ関連企業と製薬会社の財務データを検討してみます。第 1 基準を「タイプ」、第 2 基準を「利益 (\$M)」として、データテーブルを並べ替えてみます。ここでは、「タイプ」内における「利益 (\$M)」の並び順を、降順にします。

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. [テーブル] > [並べ替え] を選択します。
3. 「タイプ」を選択して [By] をクリックし、「タイプ」を並べ替えに用いる変数に割り当てます。
4. 「利益 (\$M)」を選択して [By] をクリックします。

この時点で、両方の変数は昇順で並べ替える設定になっています。図3.29で、変数の横に表示されている昇順アイコンを確認してください。

図3.29 昇順アイコン



5. 「利益 (\$M)」の降順で並べ替えるように変更するには、「利益 (\$M)」を選択して、降順ボタンをクリックします。

図3.30 「利益」を降順に変更



「利益(\$M)」の横にあるアイコンが降順アイコンに変化します。

6. [テーブルの置換] チェックボックスをオンにします。

[テーブルの置換] チェックボックスをオンにすると、並べ替え後の結果が新しいテーブルとして作成されるのではなく、元のデータテーブル自体が並べ替えられます。元のデータテーブルから作成されたレポートウィンドウが開いている場合は、このオプションを使用できません。レポートウィンドウが開いているデータテーブルを並べ替えると、それらのレポートウィンドウで表示されている結果（特にグラフ）に影響する可能性があるからです。

7. [OK] をクリックします。

データテーブルが「タイプ」のアルファベット順、「タイプ」の中では総利益の降順で並べ替えられます。

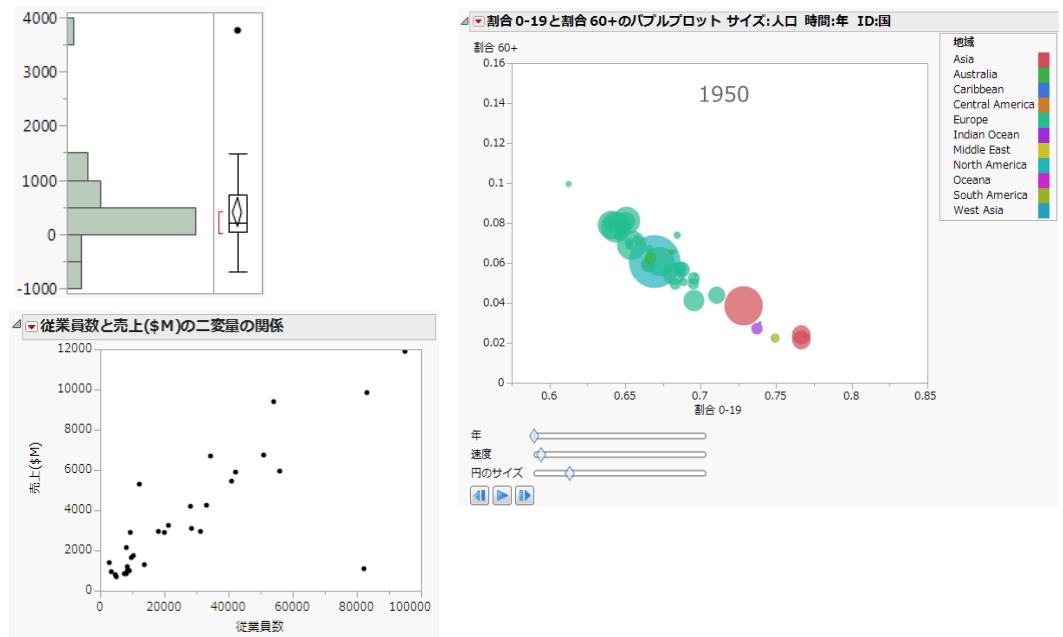
第4章

データの視覚化 共通のグラフ

データを視覚化することは、大事な最初の一歩です。この章で説明するグラフは、データから重要な情報を導き出すのに役立ちます。たとえば、ヒストグラムを見れば、データの分布や範囲が分かり、また、異常なデータ点を簡単に発見できます。

この章では、データを視覚化するのに、よく使われるグラフの一部を紹介しています。この章は、JMPで用意されているグラフィカルツールやプラットフォームの一部を入門者向けに説明しています。単変量の分布や多変量間の関係を、JMPによって視覚化できます。

図4.1 JMPでのデータの視覚化



一変量の分析

一変数（单一変量）グラフは、1つの変数について1つずつ詳しく検討するときに用います。データの検討を始めるにあたっては、変数間の関係を調べる前に、各変数1つずつについて理解しておくことが重要です。一変量のグラフは、1変数ごとに視覚化します。

このセクションでは、一変量の分布を示す2つのグラフを紹介します。

- ・「ヒストグラム」(90ページ) では、連続変数を扱います。
- ・「棒グラフ」(92ページ) では、カテゴリカル変数を扱います。

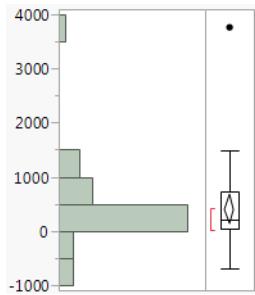
これら両方のグラフは、「一変量の分布」プラットフォームで作成します。「一変量の分布」は、各変数をグラフで示し、各変数の理解に役立つ統計量を表示します。

ヒストグラム

ヒストグラムは、連続変数の分布を理解するのに最も役に立つグラフの一つです。ヒストグラムでは、次の情報を確認できます。

- ・ 平均とばらつき
- ・ 最小値や最大値

図4.2 ヒストグラムの例



シナリオ

この例では、企業数社の利益に関するデータを記録した「Companies.jmp」データテーブルを使用します。

証券アナリストは、次の点を調査する必要があります。

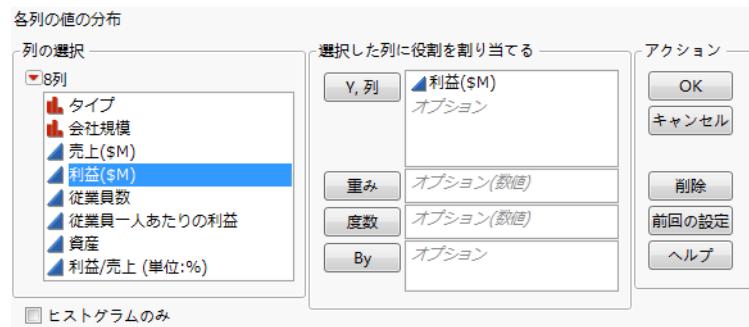
- ・ 概ね、各企業の利益はどれくらいか
- ・ 平均利益はどれくらいか
- ・ 他社と比較して利益が極端に多いまたは少ない企業があるか

回答を導き出すには、「利益(\$M)」のヒストグラムを使用します。

ヒストグラムの作成

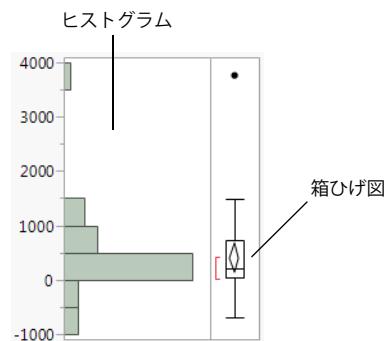
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Companies.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「利益(\$M)」を選択し、[Y, 列] をクリックします。

図4.3 「利益(\$M)」の「一変量の分布」 ウィンドウ



4. [OK] をクリックします。

図4.4 「利益(\$M)」のヒストグラム



ヒストグラムの解釈

このヒストグラムから、次のことがわかります。

- 大半の企業の利益は -1000 ドルから 1500 ドルの範囲に収まっている。
1つを除くすべての棒がこの範囲に収まっています。また、多くの企業の利益は、0 ドルから 500 ドルの範囲に集中しています。この範囲を表す棒が、他の棒よりも長くなっています。
 - 平均利益は 500 ドルより若干少ない程度である。
- 箱ひげ図のなかに表示されているひし形の中央が、平均を示します。この例では、平均は 500 ドルの目盛りの少し下になっています。

- 1社が他社と比べて著しく高い利益を上げており、外れ値である可能性が高い。外れ値とは、他のデータ点が示す一般的なパターンから飛び離れたデータ点のことです。

この例では、外れ値は、ヒストグラム上部の非常に短い1つの棒で示されています。この棒は短く、小さなグループ（この場合は1社）を表しており、ヒストグラムの他の棒から大きく離れています。

ヒストグラムの他に、このレポートには次の情報が表示されています。

- 箱ひげ図。同じデータを別の形でグラフィカルに示した要約プロットです。箱ひげ図の詳細については、『グラフ機能』の「グラフビルダー」の章を参照してください。
- 「分位点」レポートと「要約統計量」レポート これらのレポートについては、「データの分析」の章の「[一変量の分布](#)による分析」（130ページ）で説明します。

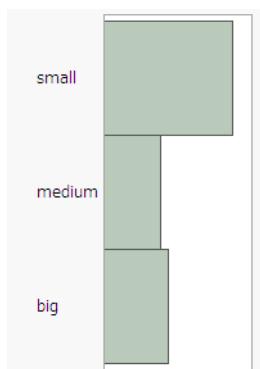
ヒストグラムの操作

JMPのデータテーブルとレポートはすべてリンクしています。ヒストグラムの棒をクリックすると、データテーブル内の対応する行が選択されます。

棒グラフ

棒グラフによって、カテゴリカル変数の分布を視覚化できます。棒グラフはヒストグラムに似ています。両者とも、変数の水準ごとに棒を描きます。ただし、棒グラフは変数の水準ごとに棒を描くのに対して、ヒストグラムは特定の範囲ごとに棒を描きます。

図4.5 棒グラフの例



シナリオ

この例では、会社情報のサンプルデータ「Companies.jmp」を使用します。このデータには、数社の企業に関する、会社規模とタイプの情報が含まれています。

証券アナリストは、次の点を調査する必要があります。

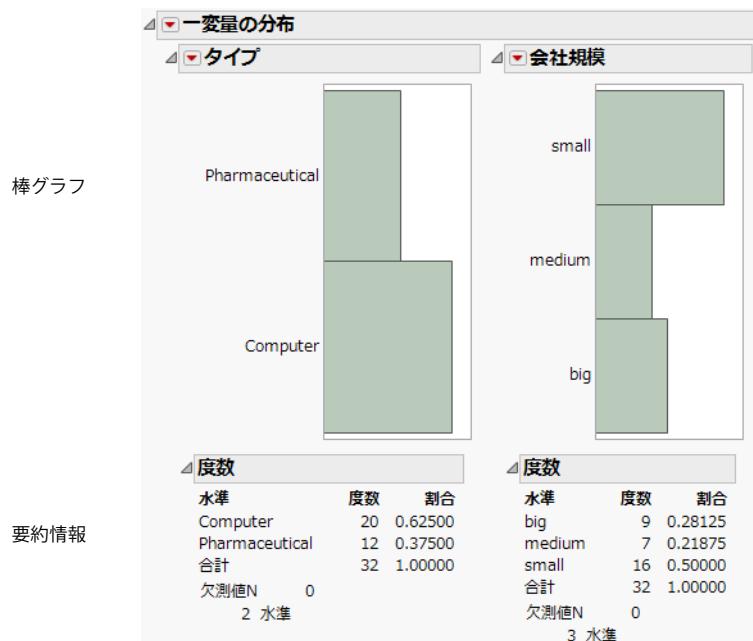
- どの企業タイプが最も一般的か
- どのくらいの規模の企業が一番多いか

回答を導き出すには、「タイプ」と「会社規模」の棒グラフを使用します。

棒グラフの作成

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] を選択します。
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。

図4.6 「タイプ」と「会社規模」の棒グラフ



棒グラフの解釈

この棒グラフから、次のことがわかります。

- 製薬会社よりもコンピュータ関連企業の方が多い。

製薬会社（「Pharmaceutical」）の棒よりも、コンピュータ関連企業（「Computer」）の棒のほうが長くなっています。

- 小規模企業が多い。

小規模（「small」）の棒が、中規模（「medium」）もしくは大規模（「big」）の棒よりも長くなっています。

一緒に出力されている要約のレポートを見ると、詳細な度数がわかります。このレポートについては、「データの分析」の章の「カテゴリカル変数の分布」（133ページ）で説明します。

棒グラフの操作

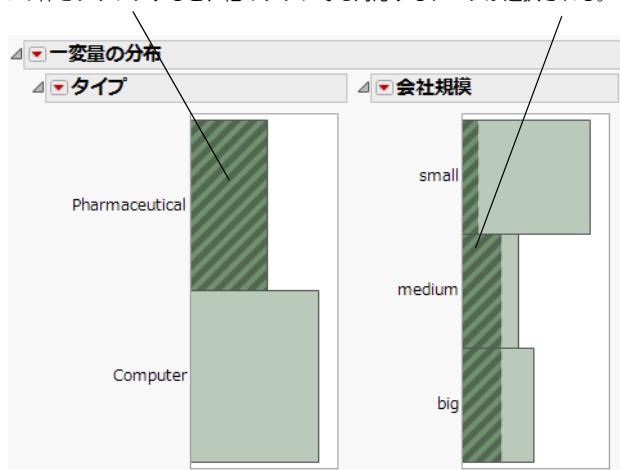
ヒストグラムの場合と同様に、個々の棒をクリックすると、データテーブル内で該当行が強調表示されます。複数のグラフが作成された場合、1つの棒グラフで棒をクリックすると、他の棒グラフでも対応する個所が強調表示されます。

たとえば、製薬会社の会社規模の分布を確認しましょう。「タイプ」の棒グラフの「Pharmaceutical」棒をクリックすると、「会社規模」の棒グラフで製薬会社に相当する部分が強調表示されます。図4.7から、このデータ全体では小規模企業が多いものの、製薬会社の多くは中規模または大規模であることがわかります。

なお、データテーブルでも対応する行が選択されます。

図4.7 棒のクリック

この棒をクリックすると、他のグラフでも対応するデータが選択される。



多変量の考察

2変数以上の関係やパターンを視覚化するには、多変量のグラフを使用します。この節では、次のグラフを取り上げます。

表4.1 複数個の変数を扱うグラフ

「散布図」(95ページ)	散布図を使用して、2つの連続変数を比較します。
「散布図行列」(99ページ)	散布図行列は、複数の連続変数のペアにおける関係を調べるときに用います。
「並列箱ひげ図」(102ページ)	並列箱ひげ図は、1つのカテゴリカル変数と1つの連続変数との関係を調べるときに用います。

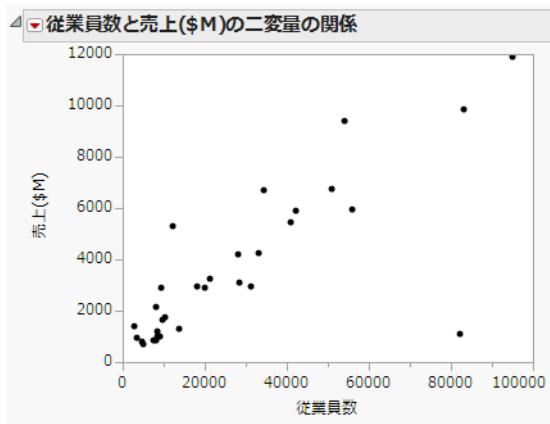
表4.1 複数個の変数を扱うグラフ（続き）

「重ね合わせプロット」(105ページ)	重ね合わせプロットは、X軸上のある1変数に対して、Y軸上に1つまたは複数の変数をプロットします。重ね合わせプロットは、時間経過に対する複数の変数の変化を比較できるため、X変数が時間である場合に特に役立ちます。
「変動性図」(108ページ)	変動性図は、1つのY連続変数と、複数のカテゴリカルなX変数の関係を調べるときに用います。変動性図は、複数のX変数から構成される組み合わせに対して、平均の差やばらつきを示します。
「グラフビルダー」(111ページ)	グラフビルダーでは、グラフを対話的に作成および変更できます。
「バブルプロット」(117ページ)	バブルプロットは、バブルの色やサイズなどにより、同時に最大5つの変数を表現する特殊な散布図です。時間の変数がある場合は、プロットを動画にして、他の変数の時間的変化を確認できます。

散布図

散布図は、多変量のグラフの中で一番シンプルなものです。散布図によって、2つの連続変数の関係を判断したり、2つの連続変数に相関があるかどうかを検討できます。相関とは、2つの変数がどれくらい関係しているかを示す指標です。2変数間の相関が強い場合、片方の変数が他方の変数に影響を与えている可能性があります。もしくは、第3の変数によって両変数が同じように影響を受けている可能性もあります。

図4.8 散布図の例



シナリオ

この例では、会社情報のサンプルデータ「Companies.jmp」を使用します。このデータには、数社の企業に関する、売上げと従業員数の情報が含まれています。

証券アナリストは、次の点を調査する必要があります。

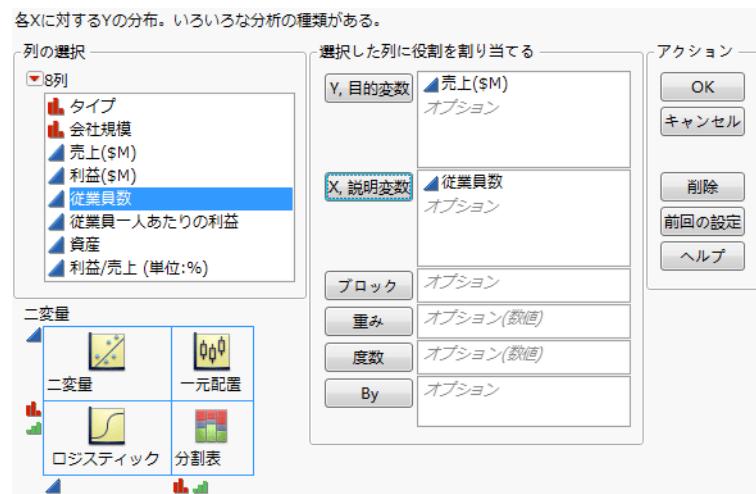
- ・ 売上と従業員数の間にどのような関係があるか
- ・ 売上高は従業員数に比例して伸びるか
- ・ 従業員数から平均売上高を予測できるか

回答を導き出すには、「売上(\$M)」と「従業員数」の散布図を使用します。

散布図の作成

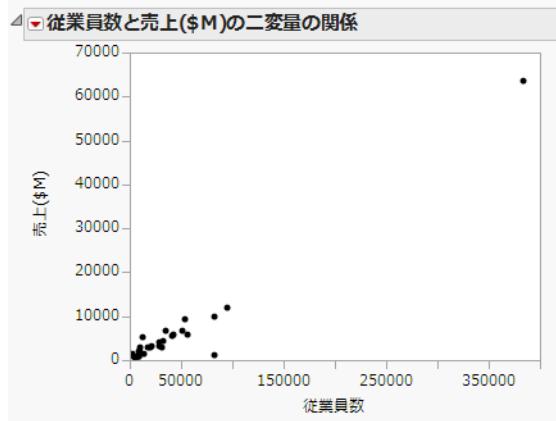
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Companies.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「売上(\$M)」を選択し、[Y, 目的変数] をクリックします。
4. 「従業員数」を選択し、[X, 説明変数] をクリックします。

図4.9 「二変量の関係」 ウィンドウ



5. [OK] をクリックします。

図4.10 「売上(\$M)」と「従業員数」の散布図



散布図行列の解釈

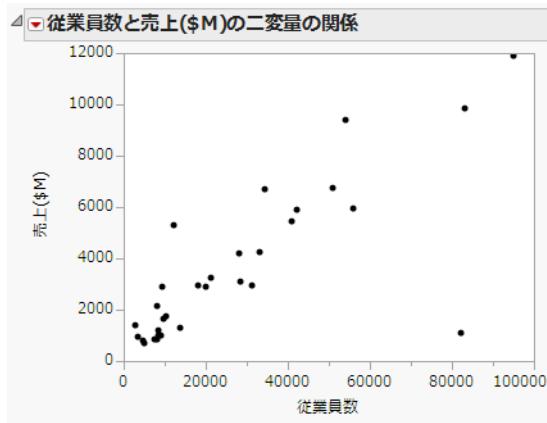
図の右上に点がぽつんと1つあることから、1社だけ従業員数が多く、売上も抜きん出ていることがわかります。この1点が他のデータ点から離れているせいで、残りの企業における関係が分かりづらくなっています。次の手順に従って、図からこの点を削除し、図を作成し直します。

1. 該当する点をクリックして選択します。
2. [行] > [非表示かつ除外] を選択します。該当するデータ点が非表示になり、計算の対象から外されます。

非表示と除外の違いは非常に重要です。点を非表示にした場合、その点はグラフから削除されますが、統計計算では引き続き使用されます。点を除外した場合は、この点は統計計算の対象から外されますが、グラフからは削除されません。点を除外し、さらに非表示にすると、すべての計算で対象外となり、すべてのグラフ上からも削除できます。

3. 外れ値を除いて図を再作成するには、「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 [やり直し] > [分析のやり直し] を選択します。元のレポートウィンドウは閉じてもかまいません。

図4.11 外れ値を削除した散布図



更新後の散布図から、次のことがわかります。

- 売上と従業員数の間には関係がある。

データ点は特定のパターンを示しています。データ点は、グラフ上で無関係に散らばっているわけではありません。ほとんどのデータ点の近くを通る直線を対角線上に引くことができます。

- 売上高は従業員数に伴って増え、直線関係が成り立っている。

データ点の近くを通るような直線を引くと、左下から右上に伸びる直線になります。直線の傾きから、従業員数が増加すると(X軸の左から右に移動すると)、売上も増加する(Y軸の下から上に移動する)ことがわかります。ほとんどのデータ点の近くを直線が通っているため、これら2変数の関係は直線関係であると言えます。なお、データ点の近くを通る線が曲線になった場合でも、点にパターンがあるならば「関係がある」と言えますが、その関係は「直線関係」ではありません。

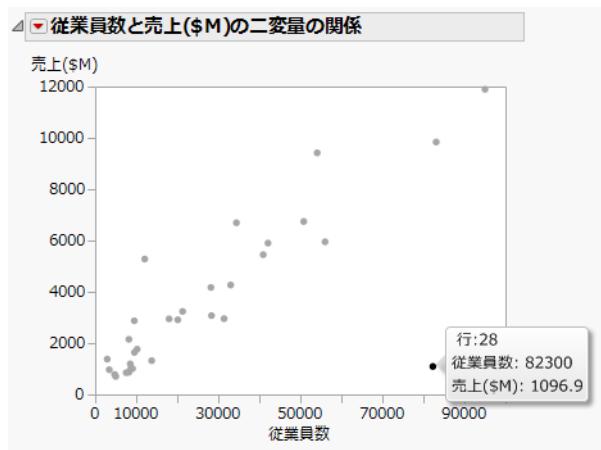
- 従業員数から平均売上を予測できる。

散布図から、従業員数が増加すると、売上も概ね伸びていることがわかります。つまり、従業員数さえ分かれば、その企業の売上を予測できるでしょう。また、先ほどの直線に基づいて予測を行えるでしょう。直線による予測は正確ではありませんが、本当の売上を近似したものになるでしょう。

散布図の操作

JMPの他のグラフと同様、散布図も対話型です。右下隅の点の上にマウスを置くと、該当する行の番号と、X、Yの値が表示されます。

図4.12 点の上にマウスを配置



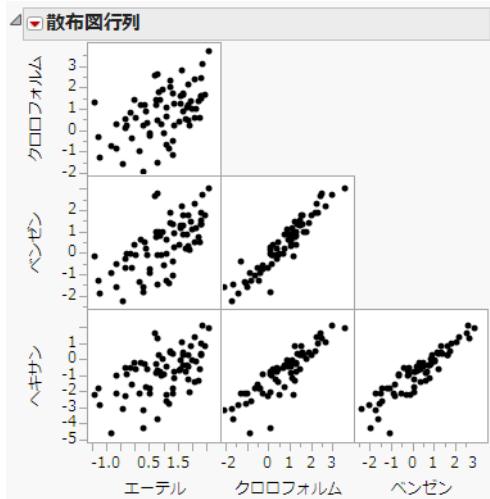
点をクリックすると、データテーブルでも対応する行が強調表示されます。複数の点を選択するには、次のいずれかを実行します。

- マウスをクリックし、目的の点の周りを囲むようにドラッグします。四角形の選択範囲内の点がすべて選択されます。
- なげなわツールを選択してから、マウスをクリックし、複数の点の周りを囲むようにドラッグします。なげなわツールでは、不規則な形状の領域を選択できます。

散布図行列

散布図行列は、複数の散布図をグリッド（行列）に整列して表示したものです。各々の散布図は、2変数の関係を示します。

図4.13 散布図行列の例



シナリオ

この例では、溶解度のサンプルデータ「Solubility.jmp」を使用します。このテーブルには、72種類の溶質の溶解度が記録されています。

研究所の技術者は、次の点を調査する必要があります。

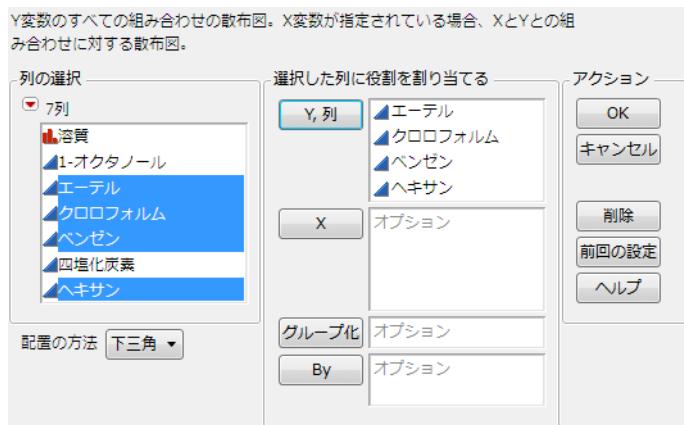
- ・ 溶媒のどのペアの間で関係があるか（全部で6ペアある）
- ・ どのペアの関係が一番強いか

回答を導き出すには、4つの溶剤の散布図行列を使用します。

散布図行列の作成

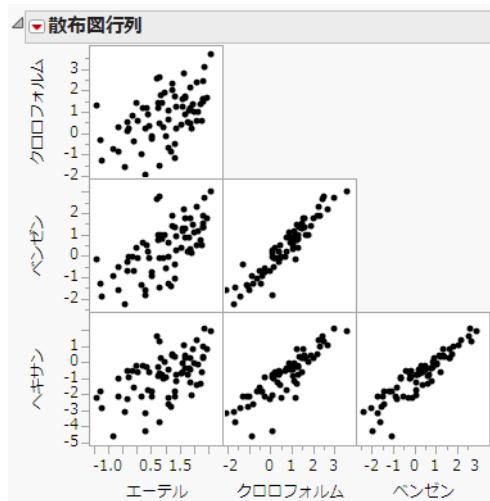
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Solubility.jmp」を開きます。
2. [グラフ] > [散布図行列] を選択します。
3. 「エーテル」、「クロロフォルム」、「ベンゼン」、「ヘキサン」を選択し、[Y, 列] をクリックします。

図4.14 「散布図行列」 ウィンドウ



4. [OK] をクリックします。

図4.15 散布図行列



散布図行列の解釈

この散布図行列から、次のことがわかります。

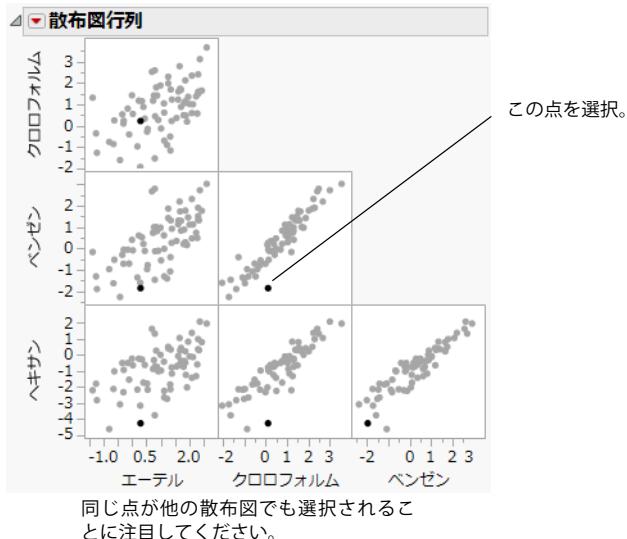
- 6つのペアすべてについて、正の相関がある。
片方の変数が増加すると、他方の変数も増加しています。
- 一番強い関係は、「ベンゼン」と「クロロフォルム」の間に成立している。
「ベンゼン」と「クロロフォルム」の散布図のデータ点が一番、直線に沿って密に分布しています。

散布図行列の操作

1つの散布図で点を選択すると、他のすべての散布図でもその点が選択されます。

たとえば、「ベンゼン」と「クロロフォルム」の散布図で点を選択すると、同じ点が他の5つの散布図でも選択されます。

図4.16 選択された点

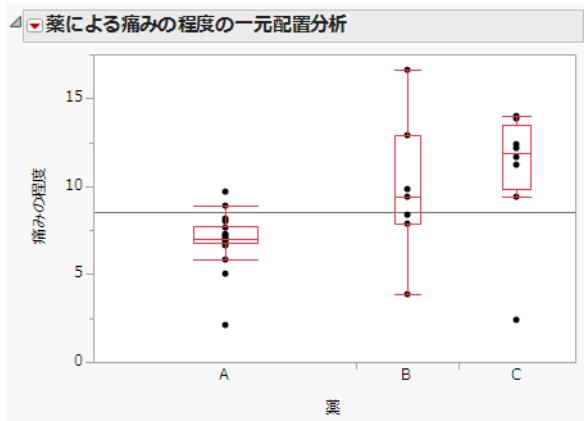


並列箱ひげ図

並列箱ひげ図(並べて表示した箱ひげ図)には、次の情報が示されます。

- 1つのカテゴリカル変数と1つの連続変数の関係
- カテゴリカル変数の水準ごとの連続変数の違い

図4.17 並列箱ひげ図の例



シナリオ

この例では、鎮痛剤に関するサンプルデータ「Analgesics.jmp」を使用します。このテーブルには、3つの薬のいずれかを処方された患者の痛みの程度が記録されています。

研究者は、次の点を調査する必要があります。

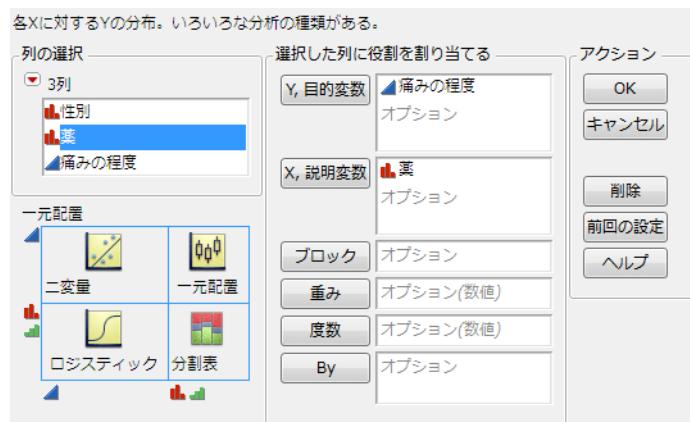
- 薬によって鎮痛作用の平均に違いはあるか。
- 各薬の鎮痛作用のばらつきに違いはあるか。ばらつきが大きい薬は、ばらつきが小さい薬より信頼性が低くなります。

これらの回答を導き出すには、「薬」と「痛みの程度」との並列箱ひげ図を表示します。

並列箱ひげ図の作成

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] を選択します。
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。

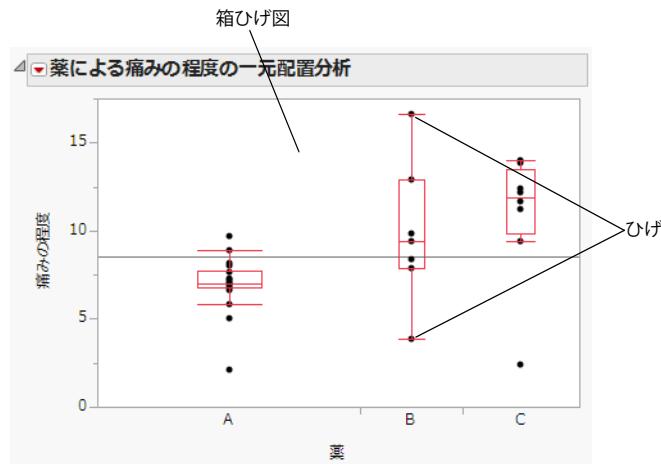
図4.18 「二変量の関係」 ウィンドウ



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4.19 並列箱ひげ図



並列箱ひげ図の解釈

箱ひげ図は、次の原則に従っています。

- 箱のなかを横切る水平線は、中央値を表す。
- データの半分は箱内に収まる。
- 大半のデータは、ひげの端から端までの間に収まる。
- ひげの外側のデータ点は、外れ値の可能性がある。

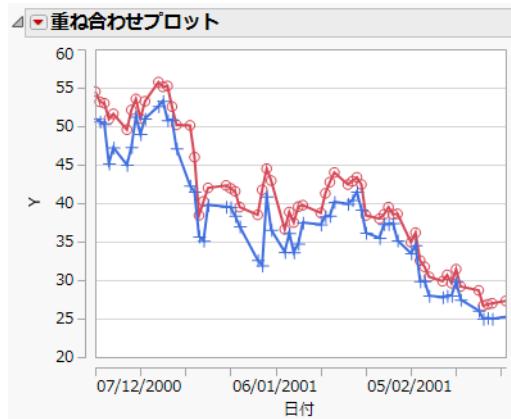
図4.19の箱ひげ図から、次のことがわかります。

- A薬を処方された患者は、他の薬よりも痛みの程度が小さくなっています。A薬の箱ひげ図は、他の箱ひげ図と比べて痛みの程度が低くなっています。
- B薬は、A薬やC薬と比べてばらつきが大きい。B薬の箱ひげ図は、他よりも縦幅が長くなっています。C薬には、他のデータと比べて著しく低い位置にある点が1つあります。この点にマウスを置くと、データテーブルの26行目に該当することがわかります。この点は、A薬やB薬のデータの方に似通っているように見受けられます。26行目のデータは調べたほうがよいでしょう。ひょっとすると、データの記録時に誤入力があったのかもしれません。

重ね合わせプロット

散布図と同様、重ね合わせプロットは、変数間の関係を示します。ただし、時間変数によって時間経過に伴う傾向を示したい場合には、散布図よりも有効です。

図4.20 重ね合わせプロットの例



注:時間経過に伴うデータをプロットに示すには、グラフビルダー、バブルプロット、管理図、変動性図も使用できます。グラフビルダーとバブルプロットの詳細については、『グラフ機能』の「[グラフビルダー](#)」「[グラフビルダー](#)」の章を参照してください。管理図と変動性図の詳細については、『品質と工程』の「[管理図ビルダー](#)」「[管理図](#)」の章と「[変動性ゲージチャート](#)」の章を参照してください。

シナリオ

この例では、3か月間にわたる株価を記録した「Stock Prices.jmp」データテーブルを使用します。

投資家は、次の点を調査する必要があります。

- 過去3か月間に株の終値に変動があるか
これを明らかにするには、時間に対する終値の重ね合わせプロットを描きます。
- 株の高値と安値の間に関係があるか

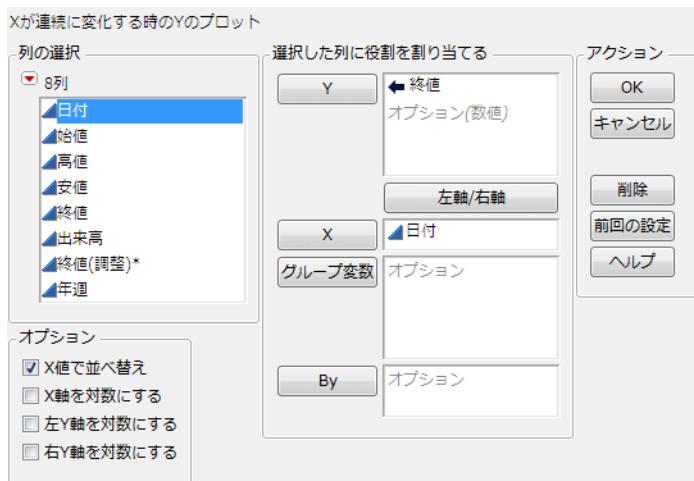
これを明らかにするには、時間に対する、高値および安値の重ね合わせプロットを描きます。

まず、最初の疑問に対する重ね合わせプロットを作成し、次に、2つ目の疑問に対する重ね合わせプロットを作成します。

時間に対する株価の重ね合わせプロットの作成

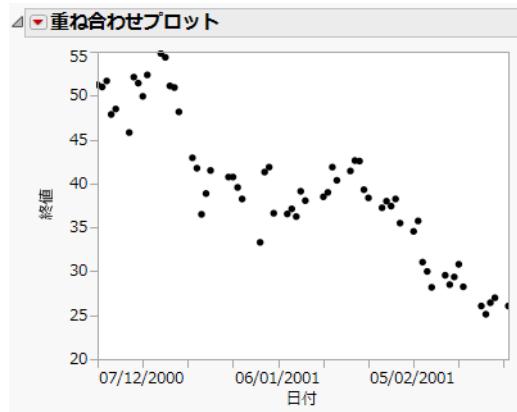
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Stock Prices.jmp」を開きます。
2. [グラフ] > [重ね合わせプロット] を選択します。
3. 「終値」を選択し、[Y] をクリックします。
4. 「日付」を選択し、[X] をクリックします。

図4.21 「重ね合わせプロット」 ウィンドウ



5. [OK] をクリックします。

図4.22 時間経過に伴う終値の重ね合わせプロット

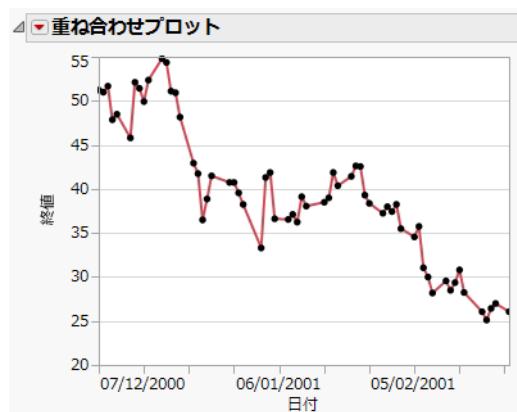


重ね合わせプロットの解釈と操作

この重ね合わせプロットは、過去数か月間において、株の終値が下がっていることを示しています。この傾向をより詳しく検討するには、点をつないでグリッド線を追加します。

1. 赤い三角ボタンをクリックして、メニューから [欠測値もつなぐ] を選択します。
2. Y軸をダブルクリックします。
3. 目盛り [大] の [グリッド線] チェックボックスを選択します。
4. [OK] をクリックします。

図4.23 つないだ点とグリッド線



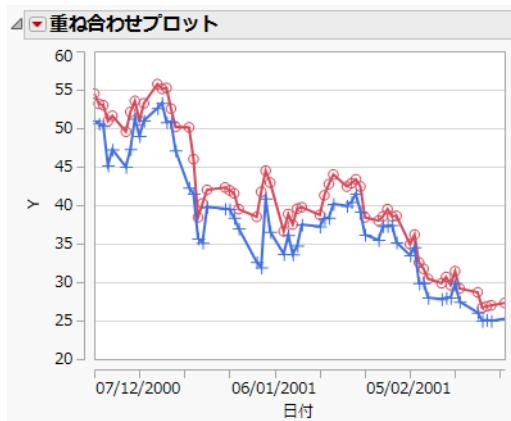
過去3か月において、株価は上下動しているものの、全体的には下降傾向にあることがわかります。

高値と安値の重ね合わせプロットの作成

重ね合わせプロットを使用して、複数の Y 变数をプロットできます。たとえば、同じプロットに高値と安値の両方を表示したいとします。

1. 「時間に対する株価の重ね合わせプロットの作成」(106 ページ) の手順に従い、今回は「高値」と「安値」の両方を [Y] の役割に割り当てます。
2. 「重ね合わせプロットの解釈と操作」(107 ページ) に示すように、点をつないでグリッド線を追加します。

図 4.24 2つのY变数



プロットの下部にある凡例は、グラフの「高値」と「安値」に使用されている色とマーカーを示しています。この重ね合わせプロットから、「高値」と「安値」の変動が酷似していることがわかります。

疑問に対する回答

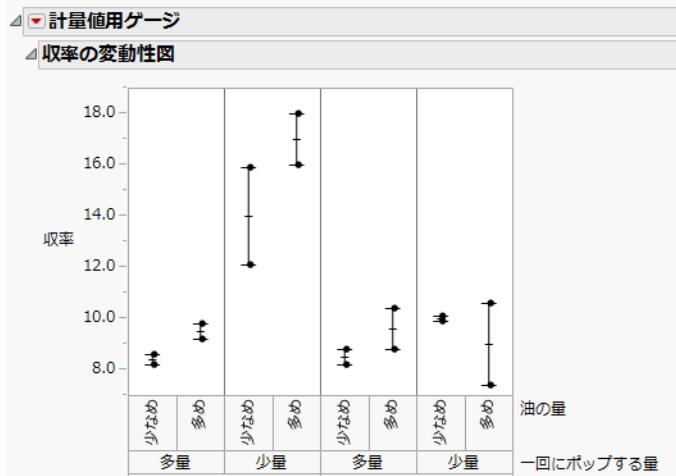
2つの重ね合わせプロットから、この例の最初に挙げた2つの疑問に対する回答が得られます。

- 最初のプロットから、株価は一定ではなく、下降傾向にあることがわかります。
- 2番目のプロットから、高値と安値にはさほど大きな差がないことがわかります。つまり、1日の間では株価にさほど大きな変動はありません。

変動性図

ここまで説明したグラフでは、X 变数を1つだけ指定しました。変動性図を使用すると、複数のX变数に対して、平均の差、および、ばらつき（変動性）を一度に確認できます。

図4.25 変動性図の例



シナリオ

この例では、ポップコーン製造業者から得たデータを記録した「Popcorn.jmp」データテーブルを使用します。コーンの種類、一回にポップする量、油の量の各組み合わせに対して、収率（同量のコーンから製造可能なポップコーンの量）が計測されています。

ポップコーン製造業者は、次の点を調査する必要があります。

- 各要因のどの組み合わせにおいて、ポップコーンの収率が高くなるか

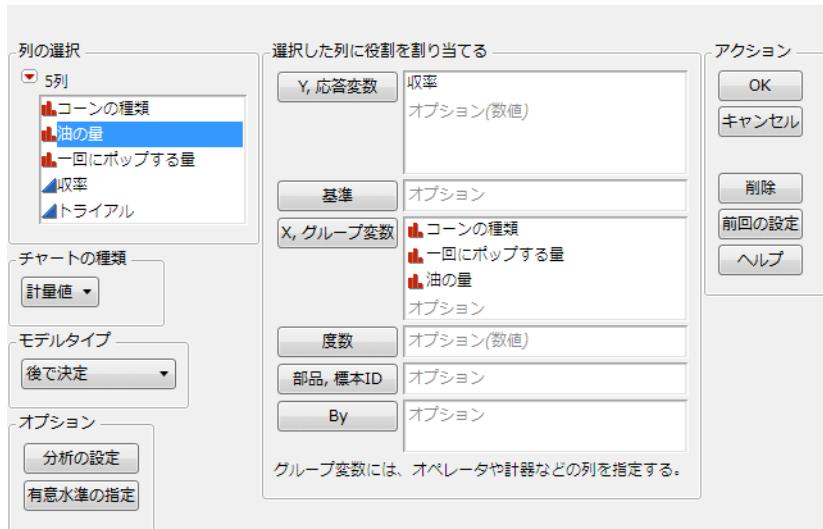
回答を導き出すには、コーンの種類、一回にポップする量、油の量に対する収率の変動性図を使用します。

変動性図の作成

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「Popcorn.jmp」を開きます。
- [分析] > [品質と工程] > [計量値/計数値ゲージチャート] を選択します。
- 「収率」を選択し、[Y, 目的変数] をクリックします。
- 「コーンの種類」を選択し、[X, グループ変数] をクリックします。
- 「一回にポップする量」を選択し、[X, グループ変数] をクリックします。
- 「油の量」を選択し、[X, グループ変数] をクリックします。

[X, グループ変数] の役割に変数を割り当てる順序は重要です。このウィンドウで指定した順序によって、変動性図の入れ子の順序が決まります。

図4.26 「変動性図」 ウィンドウ

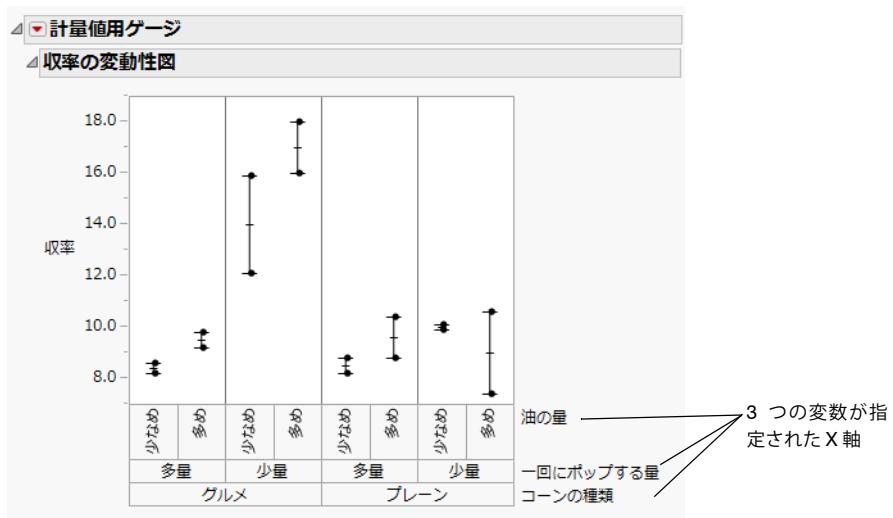


7. [OK] をクリックします。

結果において、先頭のグラフは、3つの変数の組み合わせに対する収率を示す変動性図です。それに続く図は、3つの変数の組み合わせに対する標準偏差を示しています。2番目の図には収率が示されていないため、ここでは非表示にします。

8. 赤い三角ボタンをクリックして、表示されたメニューで [標準偏差図] の選択を解除します。

図4.27 結果ウィンドウ



変動性図の解釈

収率の変動性図は、グルメコーンを使用して、かつ、一回に少量ポップした場合、収率が高くなることを示しています。

さらに掘り下げて検討するには、一回にポップする量が少量だから収率が高いのか、それともグルメコーンだから収率が高いのかを確認する必要があるでしょう。

この変動性図から、次の情報が得られます。

- プレーンコーンを使用して、かつ、一回に少量ポップした場合、収率は低い。
- グルメコーンを使用して、かつ、一回に多量ポップした場合、収率は低い。

この情報を踏まえると、グルメコーンを一回に少量ポップした場合だけが高収率になると結論付けることができます。単一の変数しか扱えない図では、この結論に達するのは困難でしょう。

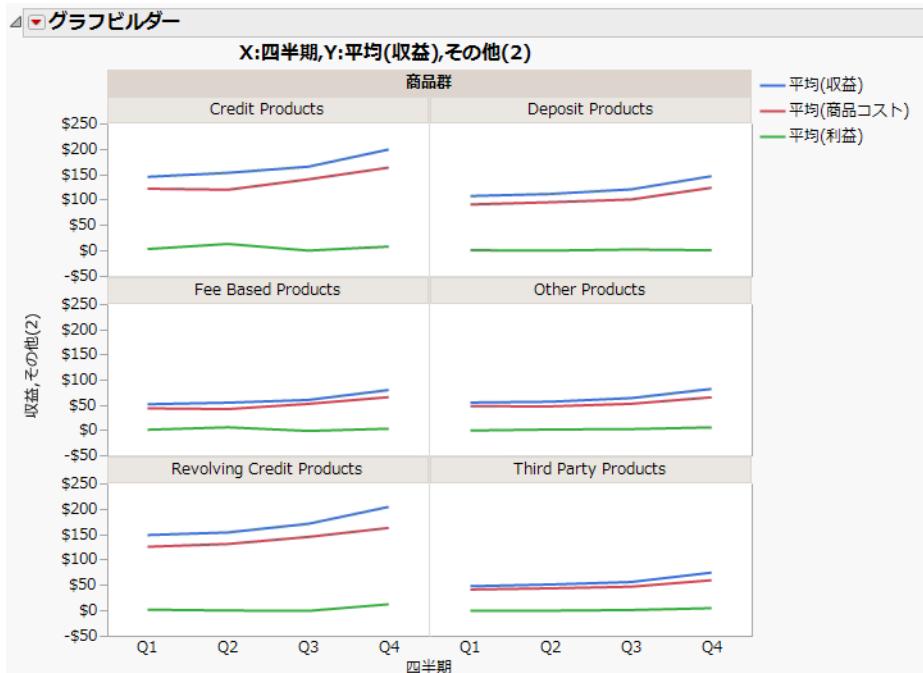
グラフビルダー

グラフビルダーでは、グラフを対話的に作成および変更できます。これまでに紹介したグラフは、プラットフォームを起動して変数を指定する操作方法によって作成しました。また、別の種類のグラフを作成したい場合には、もう一度、別のプラットフォームを起動しなければいけません。グラフビルダーでは、変数とグラフを隨時変更できます。

グラフビルダーを使用して、次の作業を実行できます。

- 変数をグラフの内外にドラッグ & ドロップして変更する。
- マウスを数回クリックするだけで異なる種類のグラフを作成する。
- グラフを横方向または縦方向に分割する。

図4.28 グラフビルダーで作成されたグラフの例



注: ここでは、グラフビルダーの一部の機能だけを取り上げています。詳細については、『グラフ機能』の「グラフビルダー」の章を参照してください。

シナリオ

この例では、銀行が提供している何種類かの商品群に対する利益を記録した「Profit by Product.jmp」データテーブルを使用します。

ビジネスアナリストは、次の点を調査する必要があります。

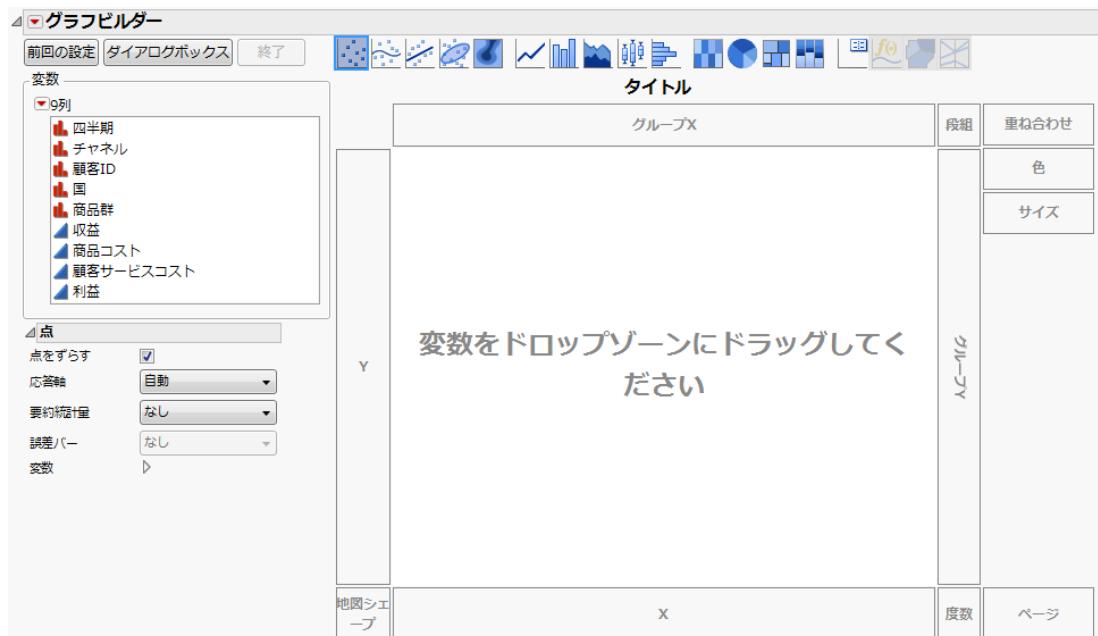
- 商品群で利益率がどのように異なるか

これを明らかにするために、商品群ごとに収益、商品コスト、利益の折れ線グラフを描きます。

グラフの作成

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「Profit by Product.jmp」を開きます。
- [グラフ] > [グラフビルダー] を選択します。

図4.29 グラフビルダーの作業スペース

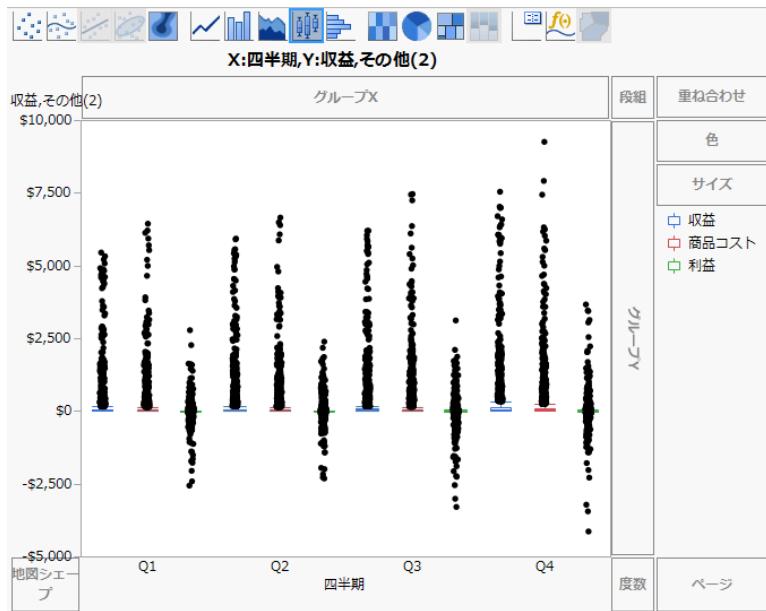


3. 「四半期」をX変数として指定するため、「四半期」をクリックし、Xゾーンにドラッグ&ドロップします。
4. 「収益」、「商品コスト」、「利益」をクリックし、Yゾーンにドラッグ&ドロップして、これら3つの変数をすべてY変数に指定します。

XゾーンとYゾーンが軸になります。

変数をクリックし、続いてゾーンをクリックすることでゾーン（役割）を割り当てることもできます。ただし、ゾーンが軸になった後で変数を追加する場合は、変数と軸をクリックするのではなく、変数を軸の上にドラッグ&ドロップしなければなりません。

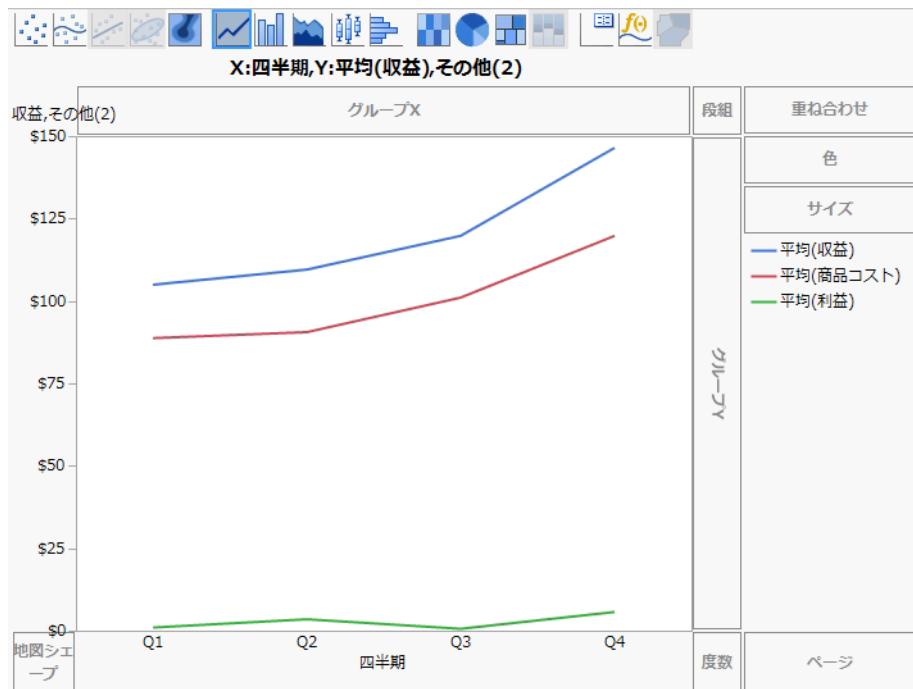
図 4.30 Y 变数と X 变数の追加後



設定された变数に基づいて、グラフビルダーに並列箱ひげ図が表示されます。

5. 箱ひげ図を折れ線グラフに変更するには、折れ線 アイコンをクリックします。

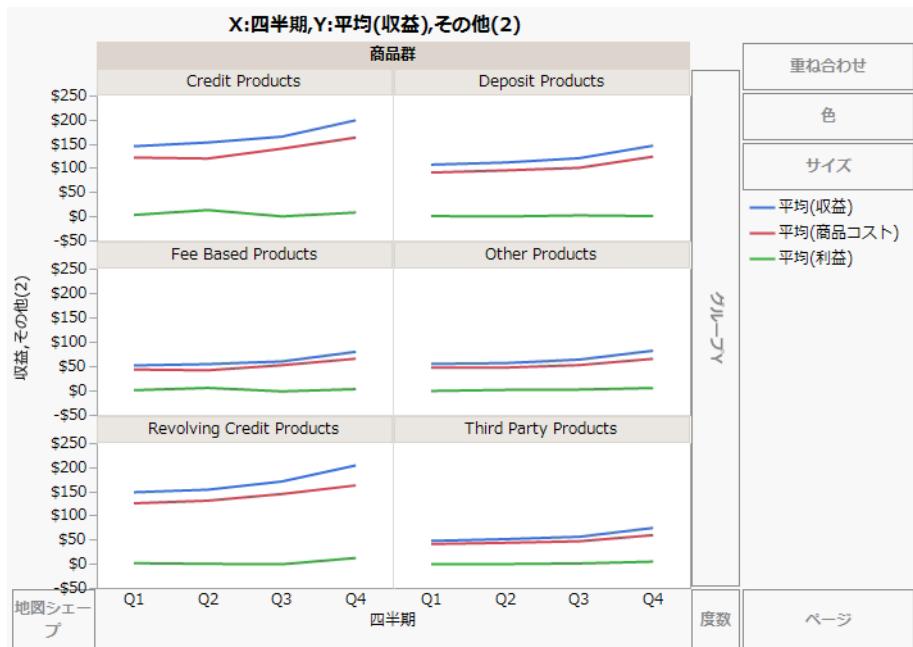
図4.31 折れ線グラフ



6. 商品ごとに別々のグラフを作成するには、「商品群」をクリックし、「段組」ゾーンにドラッグ & ドロップします。

商品ごとの折れ線グラフが表示されます。

図4.32 最終的な折れ線グラフ



グラフの解釈

図4.32には、収益、商品コスト、利益が商品群別に示されています。ビジネスアナリストは、商品群によって利益率に差がある点に興味を抱きました。図4.32の折れ線グラフから、次のような回答を導き出せます。

- 「Credit Products」、「Deposit Products」、「Revolving Credit Products」は、「Fee Based Products」、「Third Party Products」、「Other Products」に比べて収益が高い。
- ただし、すべての商品群の利益は似通っている。

データテーブルには、営業チャネルのデータも含まれています。ビジネスアナリストは、営業チャネルによって収益、商品コスト、利益にどのような差が生じているかも検討したいと思っています。

- グラフから「商品群」を削除するには、グラフのタイトル（「商品群」）をクリックし、グラフビルダー内の空の領域にドラッグ＆ドロップします。
- 「チャネル」を段組変数として追加するには、「チャネル」をクリックし、「段組」ゾーンにドラッグ＆ドロップします。

図4.33 営業チャネルごとの折れ線グラフ

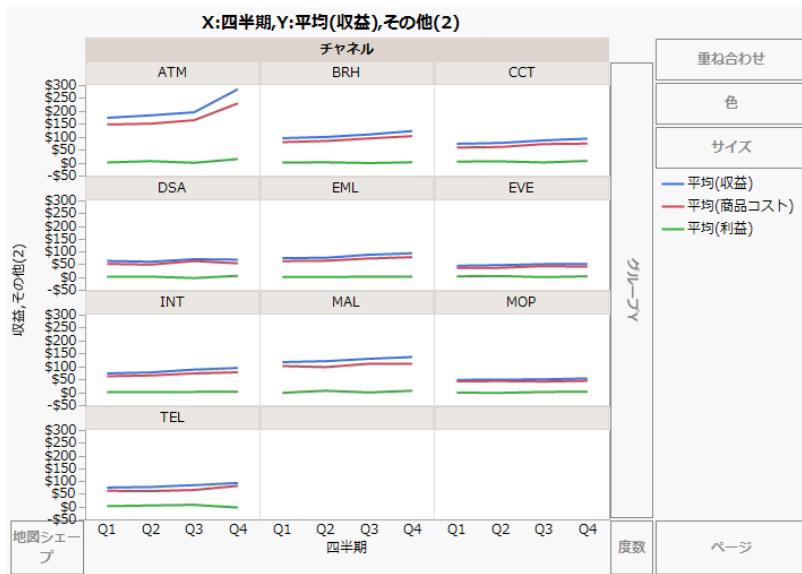
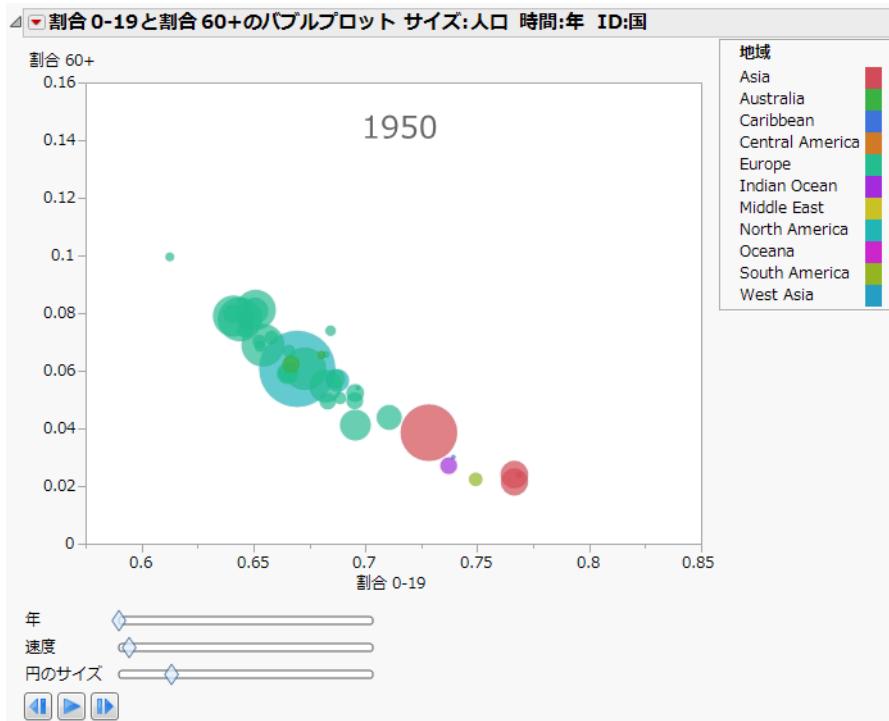


図4.33から、ATMの収益と商品コストが一番高く、急速に伸びていることがわかります。

バブルプロット

バブルプロットは、点をバブルで表現する散布図です。バブルのサイズと色は変更可能です。プロットを動画表示にして、時間的変化を見ることもできます。バブルプロットでは、最大5つの次元（ x 位置、 y 位置、サイズ、色分け、時間）を表現できるため、データを効果的に視覚化できます。

図4.34 バブルプロットの例



シナリオ

この例では「PopAgeGroup.jmp」データテーブルを使用します。このデータテーブルには、1950年から2004年までの116ヶ国（地域）の人口統計が記録されています。総人口とともに、年齢別人口も示されています。なお、一部の年のデータが欠落している国もあります。.

社会学者は、次の点を調査する必要があります。

- 世界の年齢別人口は変化しているか

回答を導き出すために、高齢グループ（60歳以上）と年少グループ（20歳未満）の人口の関係に着目します。バブルプロットを使用して、時間経過に伴いこの関係がどのように変化しているかを確認します。

バブルプロットの作成

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「PopAgeGroup.jmp」を開きます。
- [グラフ] > [バブルプロット] を選択します。
- 「割合 60+」を選択し、[Y] をクリックします。
これは、バブルプロット上のY変数に対応します。
- 「割合 0-19」を選択し、[X] をクリックします。
これは、バブルプロット上のX変数に対応します。

5. 「国」を選択し、[ID] をクリックします。

ID変数の水準ごとに、プロット上にバブルが描かれます。

6. 「年」を選択し、[時間] をクリックします。

これにより、バブルプロットの動画表示における時間軸が決まります。

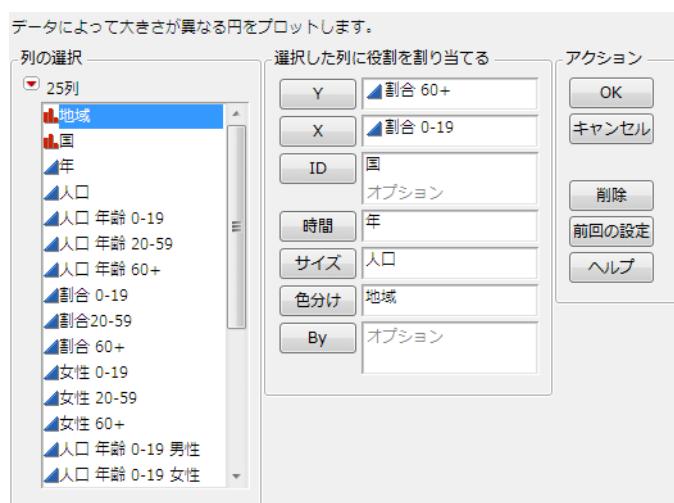
7. 「人口」を選択し、[サイズ] をクリックします。

これにより、バブルのサイズが決まります。

8. 「地域」を選択し、[色分け] をクリックします。

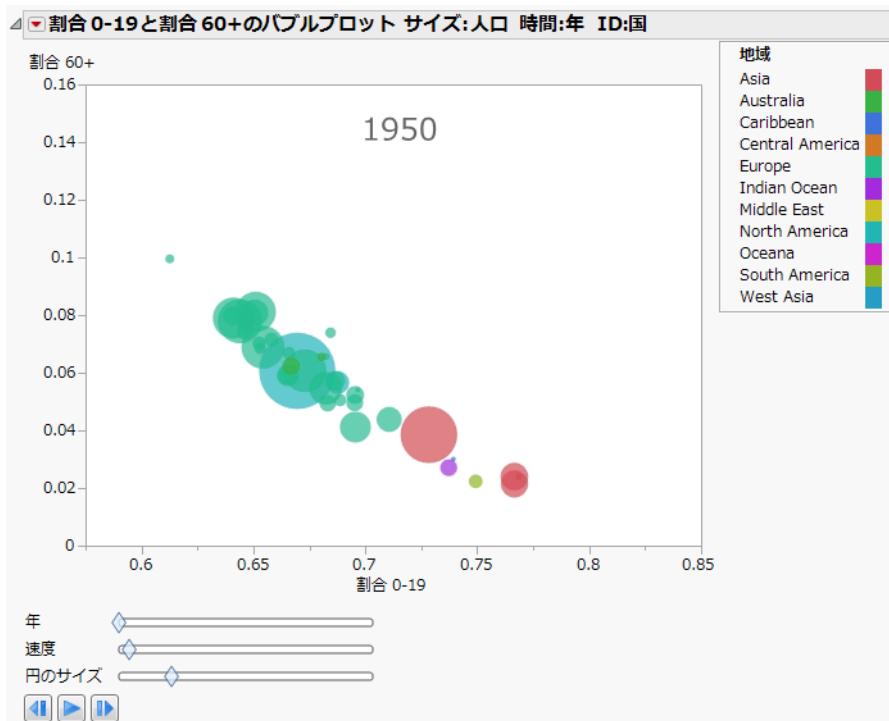
色分け変数の同じ水準が、同じ色のバブルで表されます。そのため、この例では、同じ地域内の国のバブルがすべて同じ色で表示されます。図4.36のバブルプロットで使用されている色は、JMPのデフォルト設定の色です。

図4.35 「バブルプロット」 ウィンドウ



9. [OK] をクリックします。

図4.36 最初に表示されるバブルプロット



バブルプロットの解釈

時間変数（この場合は「年」）の開始値が1950年であるため、最初のバブルプロットには1950年のデータが表示されます。再生／一時停止ボタンをクリックすると、動画が開始し、すべての年のデータが順番に表示されます。各バブルプロットには、該当年のデータが表示されます。各年のデータは、バブルプロットにおいて、次の要素で表されます。

- X座標とY座標
- バブルのサイズ
- バブルの色
- バブルごとの集計

注: バブルプロットにおいて、複数行のデータがどのように集計されているかの詳細については、『グラフ機能』の「バブルプロット」の章を参照してください。

1950年のバブルプロットでは、国の20歳未満の人口の割合が高い場合は、60歳以上の人口の割合が低いことがわかります。

再生／一時停止ボタンをクリックして、動画を開始し、残りの年についても順番に見ていきます。時間の経過に伴い、「割合 0-19」が減少し、「割合 60+」が増加します。

 動画を再生します。クリックした後、一時停止ボタンに変わります。

 動画を一時停止します。

 動画の時間を、1 単位ずつ手動で前に戻します。

 動画の時間を、1 単位ずつ手動で先に送ります。

年 手動で時間を指定する場合に使用します。

速度 動画の時間が進む速度を指定します。

バブルのサイズ バブルの絶対的な大きさを指定します。相対的な大きさは維持されます。

社会学者は、世界の年齢別人口がどのように変化しているかを知る必要がありました。このバブルプロットは、世界人口が高齢化していることを示しています。

バブルプロットの操作

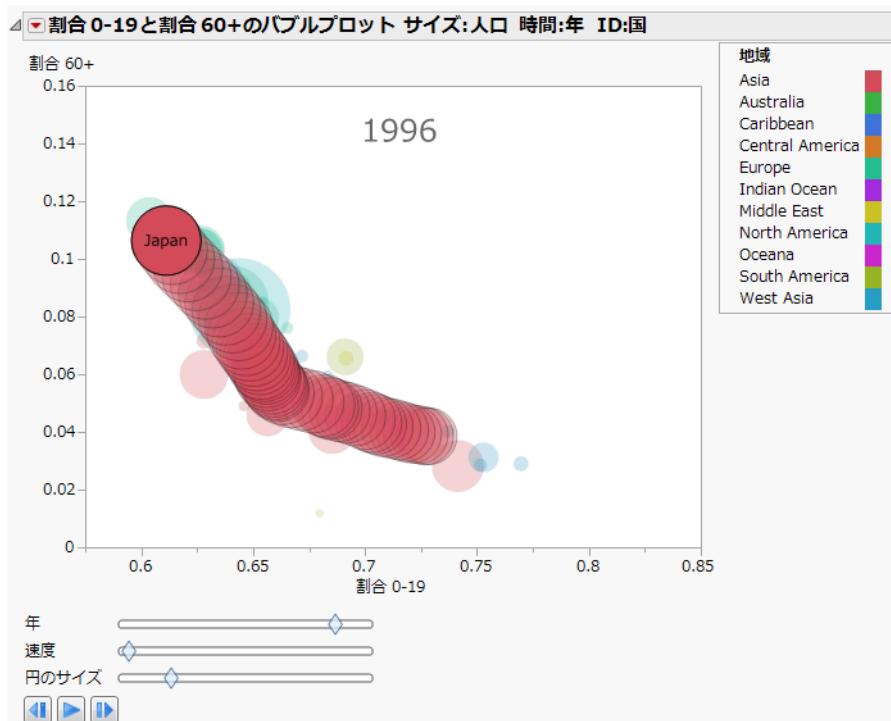
特定のバブルをクリックして選択すると、そのバブルの時間経過に伴う変化を確認できます。たとえば、1950 年のプロットの中央にある大きなバブルは日本を表しています。

日本における人口変化の傾向を確認するには

1. 日本を表すバブルの中央をクリックして選択します。
2. 赤い三角ボタンをクリックして、メニューから **【軌跡 バブル】 > 【選択されたもの】** を選択します。
3. 再生ボタンをクリックします。

動画で時間を進めると、日本のバブルには軌跡が描かれ、その変化が分かりやすくなっています。

図4.37 日本の人口変化



日本のバブルに着目すると、時間の流れの中で次の点が明らかになります。

- 20歳未満の人口の割合は減少している。
- 60歳以上の人口の割合は増加している。

第5章

データの分析 一変量の分布、二変量の関係、モデル

データ分析は、情報に基づく意思決定に役立ちます。多くの場合、データ分析には次の作業が伴います。

- 分布の確認
- 関係の洗い出し
- 仮説検定
- モデルの構築

図5.1 分析の例



この章について

実際にデータ分析に着手する前に、以下の事項を確認してください。

- ・「データからグラフを作成することの重要性」(124ページ)
- ・「尺度の理解」(127ページ)

この章の残りの部分では、いくつかの基本的な統計分析のJMPでの実行方法を紹介します。

- ・「一変量の分布」による分析」(130ページ)
- ・「二変量の関係」(136ページ)

高度なモデル化や分析手法については、以下のJMPマニュアルを参照してください。

- ・ 基本的な回帰モデル
- ・ 多変量分析
- ・ 予測モデルおよび発展的なモデル
- ・ 消費者調査
- ・ 信頼性/生存時間分析
- ・ 品質と工程

データからグラフを作成することの重要性

データからグラフを作成する（視覚化する）ことは、あらゆるデータ分析において重要であり、統計的検定やモデル構築の前に必ず行う必要があります。データ分析のはじめにデータを視覚化する必要がある理由を、次の例によって明らかにしましょう。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Anscombe.jmp」(F. J. Anscombe (1973), American Statistician, 27, 17-21) を開きます。
4組のX変数とY変数のデータが記録されています。
2. テーブルパネルの「カルテット」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。

「二変量の関係」によって、変数の各組に対して単回帰直線が描かれます。[点の表示] オプションはオフになっているため、散布図にデータは表示されていません。図5.2には、回帰直線ごとに、モデルのあてはめ結果とその他の要約情報が示されています。

図5.2 4つのモデル

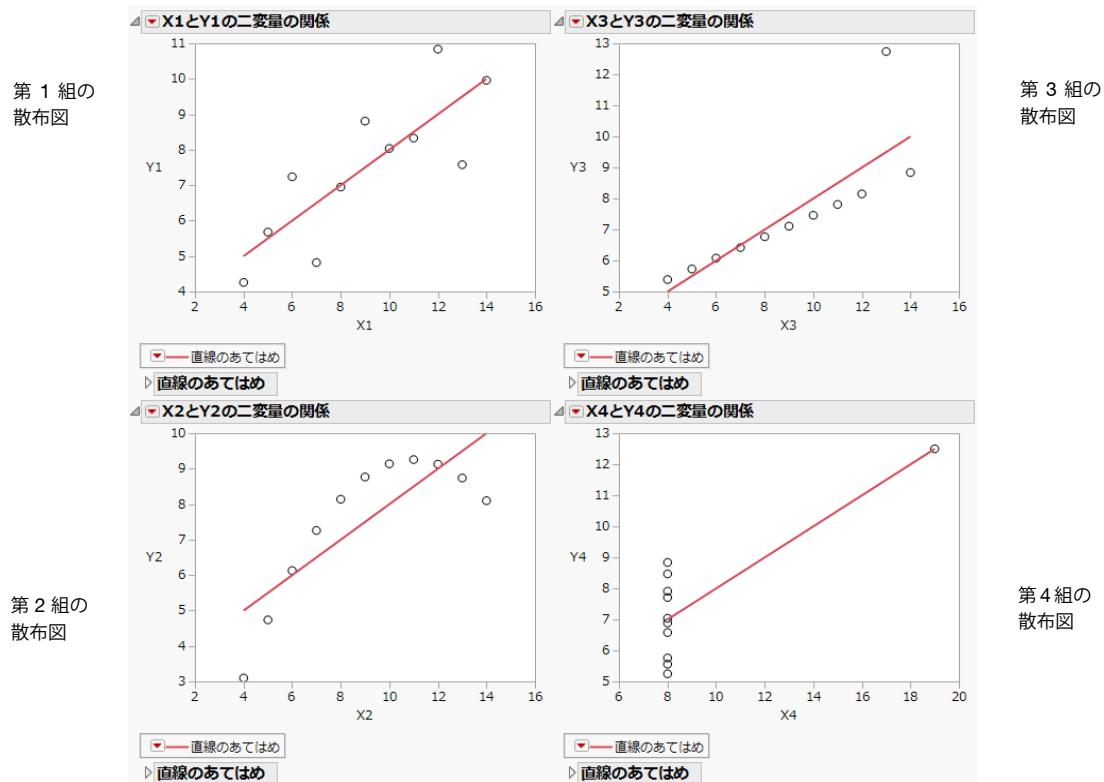
第1組のモデル		第3組のモデル																																																																																											
<p>直線のあてはめ $Y_1 = 3.0000909 + 0.5000909 \times X_1$</p> <p>あてはめの要約</p> <table border="1"> <tr><td>R2乗</td><td>0.666542</td></tr> <tr><td>自由度調整R2乗</td><td>0.629492</td></tr> <tr><td>誤差の標準偏差(RMSE)</td><td>1.236603</td></tr> <tr><td>Yの平均</td><td>7.500909</td></tr> <tr><td>オブザベーション(または重みの合計)</td><td>11</td></tr> </table> <p>分散分析</p> <table border="1"> <thead> <tr><th>要因</th><th>自由度</th><th>平方和</th><th>平均平方</th><th>F値</th></tr> </thead> <tbody> <tr><td>モデル</td><td>1</td><td>27.510001</td><td>27.5100</td><td>17.9899</td></tr> <tr><td>誤差</td><td>9</td><td>13.762690</td><td>1.5292</td><td>p値(Prob>F)</td></tr> <tr><td>全体(修正済み)</td><td>10</td><td>41.272691</td><td></td><td>0.0022*</td></tr> </tbody> </table> <p>パラメータ推定値</p> <table border="1"> <thead> <tr><th>項</th><th>推定値</th><th>標準誤差</th><th>t値</th><th>p値(Prob> t)</th></tr> </thead> <tbody> <tr><td>切片</td><td>3.0000909</td><td>1.124747</td><td>2.67</td><td>0.0257*</td></tr> <tr><td>X1</td><td>0.5000909</td><td>0.117906</td><td>4.24</td><td>0.0022*</td></tr> </tbody> </table>		R2乗	0.666542	自由度調整R2乗	0.629492	誤差の標準偏差(RMSE)	1.236603	Yの平均	7.500909	オブザベーション(または重みの合計)	11	要因	自由度	平方和	平均平方	F値	モデル	1	27.510001	27.5100	17.9899	誤差	9	13.762690	1.5292	p値(Prob>F)	全体(修正済み)	10	41.272691		0.0022*	項	推定値	標準誤差	t値	p値(Prob> t)	切片	3.0000909	1.124747	2.67	0.0257*	X1	0.5000909	0.117906	4.24	0.0022*	<p>直線のあてはめ $Y_3 = 3.0024545 + 0.4997273 \times X_3$</p> <p>あてはめの要約</p> <table border="1"> <tr><td>R2乗</td><td>0.666324</td></tr> <tr><td>自由度調整R2乗</td><td>0.629249</td></tr> <tr><td>誤差の標準偏差(RMSE)</td><td>1.236311</td></tr> <tr><td>Yの平均</td><td>7.5</td></tr> <tr><td>オブザベーション(または重みの合計)</td><td>11</td></tr> </table> <p>分散分析</p> <table border="1"> <thead> <tr><th>要因</th><th>自由度</th><th>平方和</th><th>平均平方</th><th>F値</th></tr> </thead> <tbody> <tr><td>モデル</td><td>1</td><td>27.470008</td><td>27.4700</td><td>17.9723</td></tr> <tr><td>誤差</td><td>9</td><td>13.756192</td><td>1.5285</td><td>p値(Prob>F)</td></tr> <tr><td>全体(修正済み)</td><td>10</td><td>41.226200</td><td></td><td>0.0022*</td></tr> </tbody> </table> <p>パラメータ推定値</p> <table border="1"> <thead> <tr><th>項</th><th>推定値</th><th>標準誤差</th><th>t値</th><th>p値(Prob> t)</th></tr> </thead> <tbody> <tr><td>切片</td><td>3.0024545</td><td>1.124481</td><td>2.67</td><td>0.0256*</td></tr> <tr><td>X3</td><td>0.4997273</td><td>0.117878</td><td>4.24</td><td>0.0022*</td></tr> </tbody> </table>		R2乗	0.666324	自由度調整R2乗	0.629249	誤差の標準偏差(RMSE)	1.236311	Yの平均	7.5	オブザベーション(または重みの合計)	11	要因	自由度	平方和	平均平方	F値	モデル	1	27.470008	27.4700	17.9723	誤差	9	13.756192	1.5285	p値(Prob>F)	全体(修正済み)	10	41.226200		0.0022*	項	推定値	標準誤差	t値	p値(Prob> t)	切片	3.0024545	1.124481	2.67	0.0256*	X3	0.4997273	0.117878	4.24	0.0022*
R2乗	0.666542																																																																																												
自由度調整R2乗	0.629492																																																																																												
誤差の標準偏差(RMSE)	1.236603																																																																																												
Yの平均	7.500909																																																																																												
オブザベーション(または重みの合計)	11																																																																																												
要因	自由度	平方和	平均平方	F値																																																																																									
モデル	1	27.510001	27.5100	17.9899																																																																																									
誤差	9	13.762690	1.5292	p値(Prob>F)																																																																																									
全体(修正済み)	10	41.272691		0.0022*																																																																																									
項	推定値	標準誤差	t値	p値(Prob> t)																																																																																									
切片	3.0000909	1.124747	2.67	0.0257*																																																																																									
X1	0.5000909	0.117906	4.24	0.0022*																																																																																									
R2乗	0.666324																																																																																												
自由度調整R2乗	0.629249																																																																																												
誤差の標準偏差(RMSE)	1.236311																																																																																												
Yの平均	7.5																																																																																												
オブザベーション(または重みの合計)	11																																																																																												
要因	自由度	平方和	平均平方	F値																																																																																									
モデル	1	27.470008	27.4700	17.9723																																																																																									
誤差	9	13.756192	1.5285	p値(Prob>F)																																																																																									
全体(修正済み)	10	41.226200		0.0022*																																																																																									
項	推定値	標準誤差	t値	p値(Prob> t)																																																																																									
切片	3.0024545	1.124481	2.67	0.0256*																																																																																									
X3	0.4997273	0.117878	4.24	0.0022*																																																																																									
第2組のモデル		第4組のモデル																																																																																											
<p>直線のあてはめ $Y_2 = 3.0009091 + 0.5 \times X_2$</p> <p>あてはめの要約</p> <table border="1"> <tr><td>R2乗</td><td>0.666242</td></tr> <tr><td>自由度調整R2乗</td><td>0.629158</td></tr> <tr><td>誤差の標準偏差(RMSE)</td><td>1.237214</td></tr> <tr><td>Yの平均</td><td>7.500909</td></tr> <tr><td>オブザベーション(または重みの合計)</td><td>11</td></tr> </table> <p>分散分析</p> <table border="1"> <thead> <tr><th>要因</th><th>自由度</th><th>平方和</th><th>平均平方</th><th>F値</th></tr> </thead> <tbody> <tr><td>モデル</td><td>1</td><td>27.500000</td><td>27.5000</td><td>17.9656</td></tr> <tr><td>誤差</td><td>9</td><td>13.776291</td><td>1.5307</td><td>p値(Prob>F)</td></tr> <tr><td>全体(修正済み)</td><td>10</td><td>41.276291</td><td></td><td>0.0022*</td></tr> </tbody> </table> <p>パラメータ推定値</p> <table border="1"> <thead> <tr><th>項</th><th>推定値</th><th>標準誤差</th><th>t値</th><th>p値(Prob> t)</th></tr> </thead> <tbody> <tr><td>切片</td><td>3.0009091</td><td>1.125302</td><td>2.67</td><td>0.0258*</td></tr> <tr><td>X2</td><td>0.5</td><td>0.117964</td><td>4.24</td><td>0.0022*</td></tr> </tbody> </table>		R2乗	0.666242	自由度調整R2乗	0.629158	誤差の標準偏差(RMSE)	1.237214	Yの平均	7.500909	オブザベーション(または重みの合計)	11	要因	自由度	平方和	平均平方	F値	モデル	1	27.500000	27.5000	17.9656	誤差	9	13.776291	1.5307	p値(Prob>F)	全体(修正済み)	10	41.276291		0.0022*	項	推定値	標準誤差	t値	p値(Prob> t)	切片	3.0009091	1.125302	2.67	0.0258*	X2	0.5	0.117964	4.24	0.0022*	<p>直線のあてはめ $Y_4 = 3.0017273 + 0.4999091 \times X_4$</p> <p>あてはめの要約</p> <table border="1"> <tr><td>R2乗</td><td>0.666707</td></tr> <tr><td>自由度調整R2乗</td><td>0.629675</td></tr> <tr><td>誤差の標準偏差(RMSE)</td><td>1.235695</td></tr> <tr><td>Yの平均</td><td>7.500909</td></tr> <tr><td>オブザベーション(または重みの合計)</td><td>11</td></tr> </table> <p>分散分析</p> <table border="1"> <thead> <tr><th>要因</th><th>自由度</th><th>平方和</th><th>平均平方</th><th>F値</th></tr> </thead> <tbody> <tr><td>モデル</td><td>1</td><td>27.490001</td><td>27.4900</td><td>18.0033</td></tr> <tr><td>誤差</td><td>9</td><td>13.742490</td><td>1.5269</td><td>p値(Prob>F)</td></tr> <tr><td>全体(修正済み)</td><td>10</td><td>41.232491</td><td></td><td>0.0022*</td></tr> </tbody> </table> <p>パラメータ推定値</p> <table border="1"> <thead> <tr><th>項</th><th>推定値</th><th>標準誤差</th><th>t値</th><th>p値(Prob> t)</th></tr> </thead> <tbody> <tr><td>切片</td><td>3.0017273</td><td>1.123921</td><td>2.67</td><td>0.0256*</td></tr> <tr><td>X4</td><td>0.4999091</td><td>0.117819</td><td>4.24</td><td>0.0022*</td></tr> </tbody> </table>		R2乗	0.666707	自由度調整R2乗	0.629675	誤差の標準偏差(RMSE)	1.235695	Yの平均	7.500909	オブザベーション(または重みの合計)	11	要因	自由度	平方和	平均平方	F値	モデル	1	27.490001	27.4900	18.0033	誤差	9	13.742490	1.5269	p値(Prob>F)	全体(修正済み)	10	41.232491		0.0022*	項	推定値	標準誤差	t値	p値(Prob> t)	切片	3.0017273	1.123921	2.67	0.0256*	X4	0.4999091	0.117819	4.24	0.0022*
R2乗	0.666242																																																																																												
自由度調整R2乗	0.629158																																																																																												
誤差の標準偏差(RMSE)	1.237214																																																																																												
Yの平均	7.500909																																																																																												
オブザベーション(または重みの合計)	11																																																																																												
要因	自由度	平方和	平均平方	F値																																																																																									
モデル	1	27.500000	27.5000	17.9656																																																																																									
誤差	9	13.776291	1.5307	p値(Prob>F)																																																																																									
全体(修正済み)	10	41.276291		0.0022*																																																																																									
項	推定値	標準誤差	t値	p値(Prob> t)																																																																																									
切片	3.0009091	1.125302	2.67	0.0258*																																																																																									
X2	0.5	0.117964	4.24	0.0022*																																																																																									
R2乗	0.666707																																																																																												
自由度調整R2乗	0.629675																																																																																												
誤差の標準偏差(RMSE)	1.235695																																																																																												
Yの平均	7.500909																																																																																												
オブザベーション(または重みの合計)	11																																																																																												
要因	自由度	平方和	平均平方	F値																																																																																									
モデル	1	27.490001	27.4900	18.0033																																																																																									
誤差	9	13.742490	1.5269	p値(Prob>F)																																																																																									
全体(修正済み)	10	41.232491		0.0022*																																																																																									
項	推定値	標準誤差	t値	p値(Prob> t)																																																																																									
切片	3.0017273	1.123921	2.67	0.0256*																																																																																									
X4	0.4999091	0.117819	4.24	0.0022*																																																																																									

4つの全モデルにおいて、パラメータ推定値およびR2乗値がほぼ一致している点に注目してください。各組であてはめたモデルは、いずれも、ほぼ $Y = 3 + 0.5X$ であり、R2乗値はどれも約0.66となっています。データを分析する時に上記のような要約情報だけしか調べなかつたら、おそらく、XとYの関係はどの組でも同じという結論にたどりつくでしょう。しかし、この時点では、まだデータを視覚化していません。その結論は誤っているかもしれません。

データを視覚化するために4つの散布図すべてに点を表示する

1. Ctrlキーを押したままにします。
2. いずれかの「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【点の表示】を選択します。

図5.3 点を表示した散布図



4本の直線だけを見ると、XとYの関係はすべての組で同じように見えます。しかし、データ点を見ると関係は各組で異なっていることがわかります。

- 第1組は、直線関係を表しています。
- 第2組は、非線形関係を表しています。
- 第3組は、直線関係を表していますが、外れ値が1つあります。
- 第4組は、1つの点を除く全データが、 $x = 8$ の位置に集中しています。

この例から、要約統計量だけから下した結論は不適切になる可能性があることが分かります。データを視覚化して検討する作業は、どのようなデータ分析においても、初期の段階で行う必要があります。

尺度の理解

JMPでは、データはさまざまな種類に分類されます。JMPでは、この種類をデータの尺度と呼んでいます。表5.1では、JMPの3つの尺度について説明しています。

表5.1 尺度

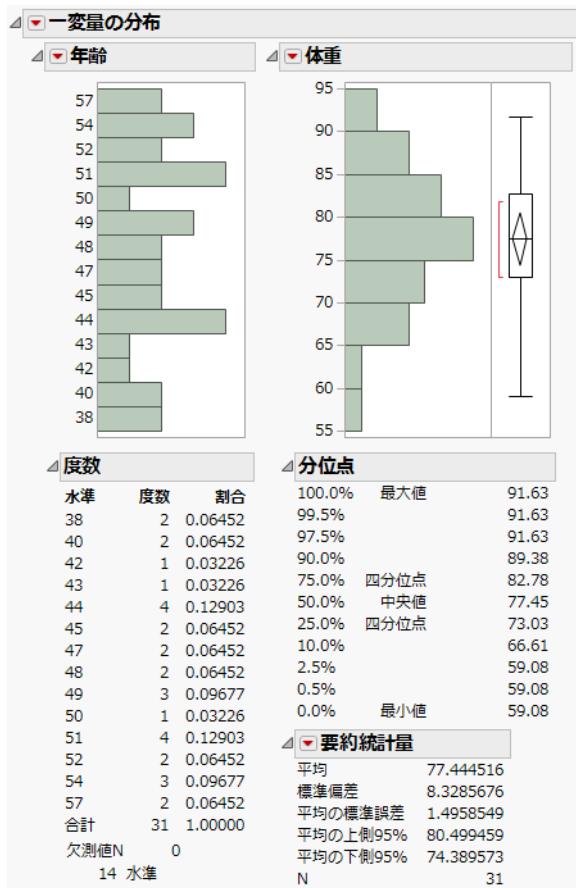
尺度と説明	例	具体的な例
連続尺度 数値データのみ。合計や平均など の演算を行うことができます。	高さ 温度 時間	試験時間は、2時間や2.13時間になる 場合があります。
順序尺度 数値データまたは文字データ。順 序に意味のあるカテゴリ。	月（1、2、…12） 文字の等級（A、B、…F） サイズ（大、中、小）	月は2（2月）または3（3月）になり 得ますが、2.13にはなりません。2月 は3月の前に訪れます。
名義尺度 数値データまたは文字データ。値 はカテゴリですが、順序に意味は ありません。	性別（男性、女性） 色 試験の結果（合格、不合格）	性別は男性または女性になります が、順序は関係ありません。性別カ テゴリは数値（男性=1、女性=2）で も表現できます。

例：尺度に基づく結果

尺度が異なれば、JMPで導き出される結果も異なります。この相違を実例で確認するには、次の手順に従ってください。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Linnerud.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「年齢」と「体重」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5.4 「年齢」と「体重」の一変量の分布結果



「年齢」と「体重」はいずれも数値変数ですが、扱われ方は異なっています。表5.2では、「体重」と「年齢」の分布結果の違いを比較しています。

表5.2 「体重」と「年齢」の結果

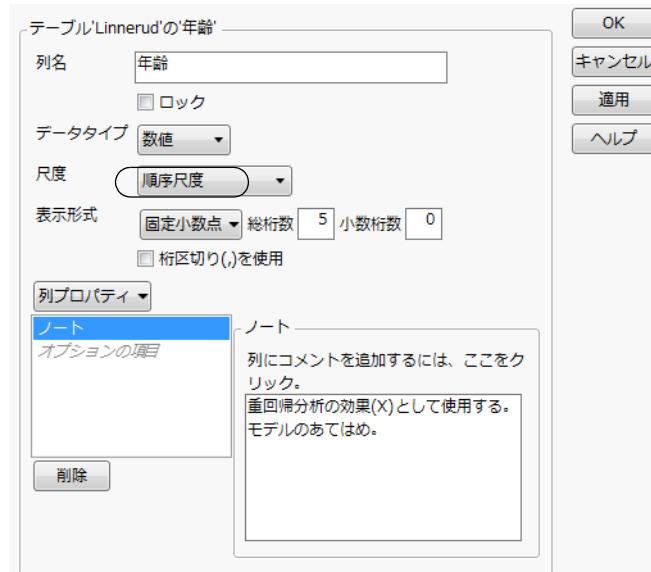
変数	尺度	結果
体重	連続尺度	ヒストグラム、分位点、要約統計量
年齢	順序尺度	棒グラフ、度数

尺度の変更

変数の処理方法を変更する場合は、尺度を変更します。たとえば、図5.4の場合、「年齢」の尺度は順序尺度になっています。JMPでは、順序尺度の変数については度数が算出されます。度数の代わりに平均年齢を算出する必要があるとしましょう。その場合は、尺度を連続尺度に変更すると、平均年齢が表示されます。

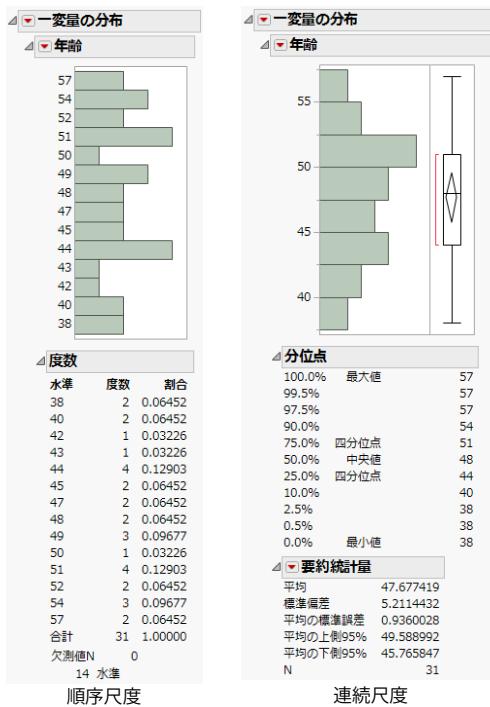
1. 「年齢」列見出しをダブルクリックします。列情報のウィンドウが表示されます。
2. 「尺度」を「連続尺度」に変更します。

図5.5 列情報のウィンドウ



3. [OK] をクリックします。
4. 「例」の手順（「例：尺度に基づく結果」（127ページ）を参照）を繰り返して、一変量の分布を作成します。
図5.6は、「年齢」が順序尺度と連続尺度のときの分布結果の違いを示しています。

図5.6 「年齢」に対する異なる尺度



「年齢」が順序尺度の場合、各年齢の度数が表示されます。たとえば、48歳は2回登場しています。「年齢」が連続尺度の場合は、平均年齢である約48（47.677）が表示されます。

「一変量の分布」による分析

単一の変数を分析する場合は、「一変量の分布」プラットフォームによって、その変数の分布を確認できます。各変数に対するレポートの内容は、カテゴリカル（名義または順序）変数か連続変数かによって異なります。

注: 「一変量の分布」プラットフォームの詳細については、『基本的な統計分析』における「一変量の分布」の章を参照してください。

連続変数の分布

連続変数を分析するときには、必要に応じて、次のような点を検討します。

- データの分布形状が既知の分布に一致しているか
- 外れ値の有無

- データの平均はどれぐらいか
- 平均が、目標値（もしくは過去の値）と統計的に異なっているか
- データはどれぐらいばらついているか。標準偏差はどれぐらいか
- 最小値と最大値はどれぐらいか

上記のような疑問については、グラフ、要約統計量、統計的検定によって回答を導き出せます。

シナリオ

この例では、自動車のサンプルデータ「Car Physical Data.jmp」を使用します。このデータには、116車種の自動車の車両情報が記録されています。

プランニング専門家が鉄道会社から依頼を受け、鉄道による自動車輸送時の問題について検討することになりました。データに基づいて、次の点を調査する必要があります。

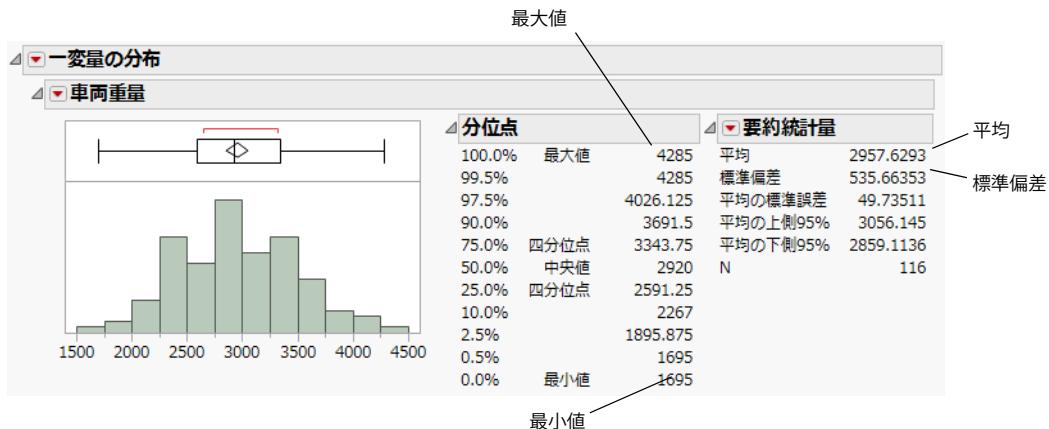
- 車両の平均重量
- 車両重量のばらつき（標準偏差）
- 車両の最小重量と最大重量
- 外れ値の有無

回答を導き出すために、車両重量のヒストグラムを描きます。

ヒストグラムの作成

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] > [一変量の分布] を選択します。
3. 「車両重量」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. レポートウィンドウの内容を横に並べるには、「車両重量」の横にある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] > [横に並べる] を選択します。

図5.7 「車両重量」の一変量の分布



レポートウィンドウには次の結果があります。

- ヒストグラムと箱ひげ図でデータが視覚化されます。
- 「分位点」レポートには、分布のパーセント点が示されます。
- 「要約統計量」レポートには、平均や標準偏差などの統計量が表示されます。

結果の解釈

図5.7に示されている結果から、次のことが分かります。

車両の平均重量 ヒストグラムから、車両の平均重量が約3,000ポンドであることがわかります。

車両重量のばらつき（標準偏差） 「要約統計量」レポートから、車両の平均重量が約2,958ポンドであることが分かります。また、同じ「要約統計量」レポートから、標準偏差が約536ポンドであることが分かります。

最小重量と最大重量 ヒストグラムから、最小重量が約1,500ポンドで、最大重量が約4,500ポンドであることがわかります。「分位点」レポートから、最小重量が約1,695ポンドで、最大重量が約4,285ポンドであることが分かります。

外れ値の有無 外れ値はありません。

デフォルトのレポートウィンドウ（図5.7を参照）には、最小限のグラフと統計量しか表示されません。赤い三角ボタンで表示するメニューを使用して、他のグラフや統計量を追加できます。

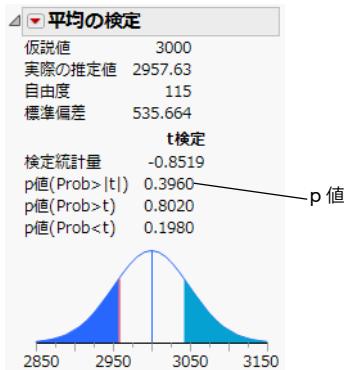
統計的検定

他の研究結果から、平均重量が3000ポンドでない場合には輸送効率が低いと鉄道会社は判断しました。そこで、輸送される車両の母集団において、平均重量が3000ポンドかどうかを確認する必要が生じました。*t*検定によって、現在の標本から母集団について推論します。

検定の結果

- 「車両重量」の赤い三角ボタンをクリックし、メニューから [平均の検定] を選択します。
- 表示されたウィンドウで、「仮説平均を指定」ボックスに「3000」と入力します。
- [OK] をクリックします。

図 5.8 平均の検定結果



t-検定の解釈

t-検定の主な結果はp値です。この例では、p値は0.396です。有意水準を0.05とした場合、0.396は0.05よりも大きいため、平均重量と3000ポンドの間に統計的な有意差は認められませんでした。p値が有意水準よりも小さいときには、「母集団の平均車両重量は3000ポンドと異なる」という結論が導き出されます。

カテゴリカル変数の分布

カテゴリカル（順序または名義）変数の分析では、次のような点を検討します。

- 変数の水準数
- 水準ごとの度数
- データが一様に分布しているか
- 全度数に対する各水準の割合

シナリオ

「連続変数の分布」（130ページ）のシナリオを参照してください。

車両の平均重量と目標重量には、統計的な有意差は見られませんでした。鉄道会社は、車両についてさらに詳しく検討することにしました。

プランニング専門家は鉄道会社の依頼を受け、次の点を調べます。

- 車種はどのようにになっているか

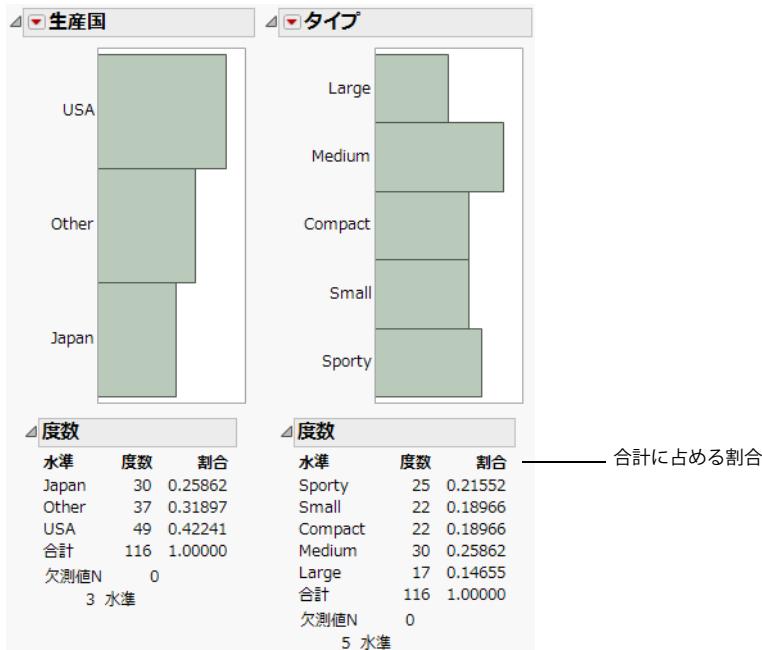
- 生産国はどのようにになっているか

これらを調べるために、「タイプ」と「生産国」の分布を検討します。

「一変量の分布」の実行

- [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。
- [分析] > [一変量の分布] を選択します。
- 「生産国」と「タイプ」を選択し、[Y, 列] をクリックします。
- [OK] をクリックします。

図5.9 「生産国」と「タイプ」の一変量の分布



一変量の分布結果の解釈

レポートウィンドウには、「生産国」と「タイプ」について、棒グラフと度数レポートが表示されます。棒グラフは、「度数」レポートに示されている度数情報をグラフで示しています。「度数」レポートには、次の情報が表示されます。

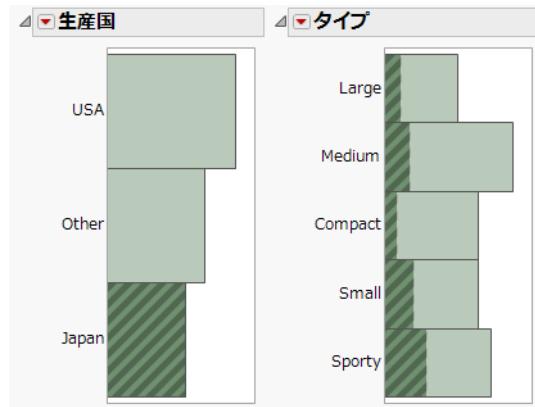
- データのカテゴリ。たとえば、「Japan」は「生産国」の一分類、「Sporty」は「タイプ」の一分類です。
- カテゴリごとの度数。
- 各カテゴリが合計に占める割合。

たとえば、コンパクトカー (Compact) は22台で、全体の台数116の19%を占めています。

一変量の分布結果の操作

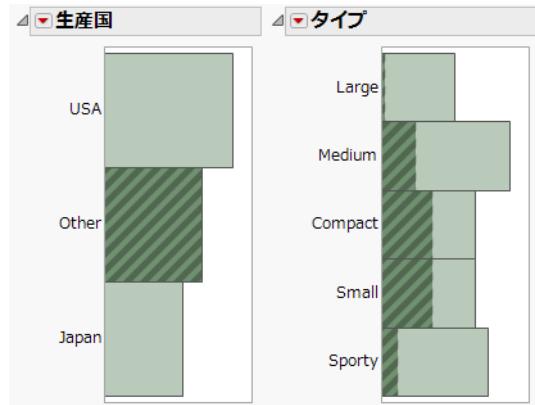
片方のグラフで棒を選択すると、別のグラフでも対応するデータが選択されます。たとえば、「生産国」の棒グラフで「Japan」の棒を選択すると、日本車にはスポーツカー（Sporty）が多いことがわかります。

図5.10 日本車



「Other」カテゴリを選択すると、これらの車両の大半が小型車またはコンパクトカーであり、大型車はほとんどないことがわかります。

図5.11 その他の車両



二変量の関係

散布図などのグラフは、変数間の関係を視覚化する上で役立ちます。関係を視覚化した次には、その関係を数値的に表現できるように、詳しく分析します。変数間の関係を数値的に説明したものを「モデル」と呼びます。変数（Y）の平均を別の変数（X）から予測するモデルは、「回帰モデル」と呼ばれています。また、回帰モデルのXは、「予測変数」、「説明変数」、「独立変数」などと呼ばれています。

JMPでは、「二変量の関係」と「モデルのあてはめ」プラットフォームで、回帰モデルの分析が実行できます。

注: ここでは、基本的なプラットフォームとオプションだけを取り上げています。全プラットフォームにおけるオプションの詳細および説明については、『基本的な統計分析』と『グラフ機能』、および「[この章について](#)」(124ページ)に一覧されているマニュアルを参照してください。

表5.3では、XとYの尺度に対して行われる統計分析を挙げています。

表5.3 関係の種類

X	Y	節
連続尺度	連続尺度	<ul style="list-style-type: none">「1つの予測変数による回帰分析」(136ページ)「複数の予測変数による回帰分析」(150ページ)
カテゴリカル	連続尺度	<ul style="list-style-type: none">「1変数による平均の比較」(141ページ)「複数の変数による平均の比較」(146ページ)
カテゴリカル	カテゴリカル	「割合の比較」(144ページ)
連続尺度	カテゴリカル	ロジスティック回帰分析は上級ユーザ向けの内容です。 『基本的な統計分析』の「ロジスティック分析」の章を参考してください。

1つの予測変数による回帰分析

シナリオ

この例では、会社情報のサンプルデータ「Companies.jmp」を用います。このデータには、製薬業界とコンピュータ業界における32社の財務データが記録されています。

従業員数が多い会社の方が、従業員数が少ない会社よりも売上が多いことは、直感的にもうなづけます。データアナリストは、各社の総売上を従業員数に基づいて予測しようと考えています。

そのためには、次の作業を行います。

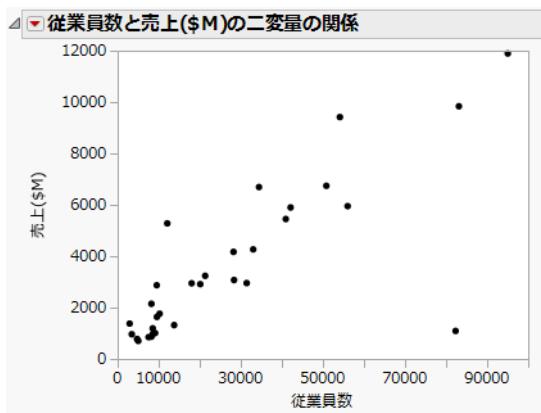
- 「関係の洗い出し」(137ページ)
- 「回帰モデルをあてはめる」(137ページ)

- 「平均売上の予測」(139ページ)

関係の洗い出し

まず、散布図を作成し、従業員数と売上の関係を確認します。この散布図は、「データの視覚化」の章の「[散布図の作成](#)」(96ページ) すでに作成しました。1つの外れ値（従業員数と売上が他社より抜きん出ている会社）を非表示にし、かつ除外した後の散布図を図5.12に示します。

図5.12 「売上(\$M)」と「従業員数」の散布図

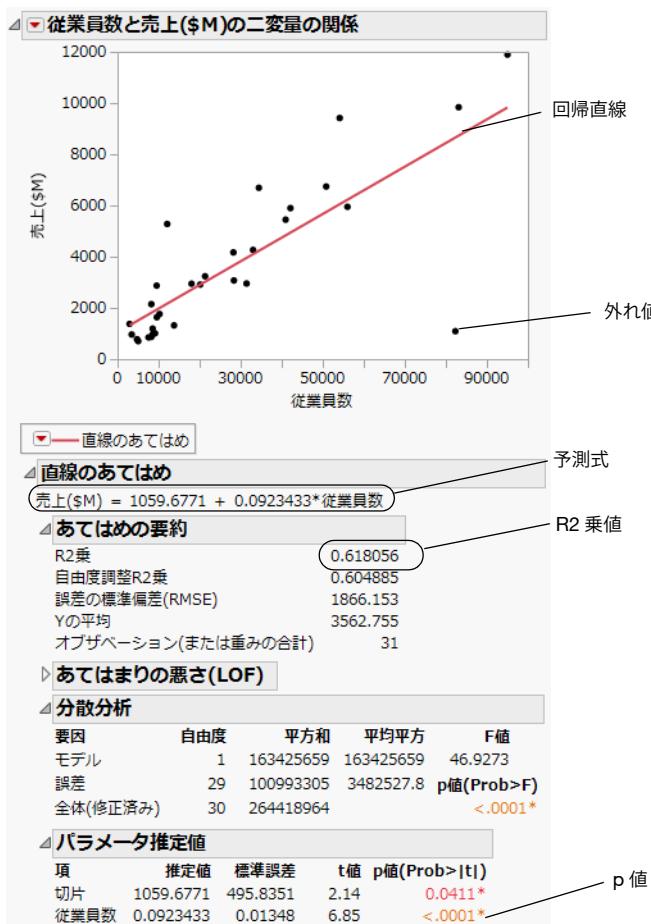


この散布図は、売上と従業員数間の関係を明確に示しています。直感どおりに、従業員数が多い会社ほど売上が高くなっています。この散布図は、データアナリストの直感を視覚的に裏付けています。しかし、この散布図だけからは、ある従業員数に対する売上高の予測値を決めることができません。

回帰モデルをあてはめる

従業員数から売上を予測するには、回帰モデルをあてはめます。「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 [直線のあてはめ] を選択します。散布図に回帰直線が追加され、新しいレポートがレポートウィンドウに追加されます。

図 5.13 回帰直線



レポート内で次の結果に着目してください。

- p 値 < .0001
- R2 乗値 = 0.618

これらの結果から、次の結果を導き出せます。

- p 値は有意水準 0.05 より低い値です。つまり、予測モデルに従業員数を含めることで、平均売上を予測する度合いが有意に向上了します。
- この例では R2 乗値が大きく、従業員数に基づく予測モデルで売上を予測できるという裏付けとなっています。R2 乗値は、変数間の関係の強さ（「相関」と呼ばれています）を示します。相関が 0 の場合は変数間に関係がないことを示唆し、1 の場合には完全な直線関係が成立しています。

平均売上の予測

回帰モデルによって、所定数の従業員を擁する会社の平均売上を予測します。レポートには、次のモデルの予測式が示されています。

$$\text{平均売上} = 1059.68 + 0.092 \times \text{従業員数}$$

たとえば、従業員が70,000人の会社の場合、式は次のようにになります。

$$7,499.68 \text{ ドル} = 1059.68 + 0.092 \times 70,000$$

現在の散布図の右下部分に、他社が示す一般的なパターンに合致しない外れ値があります。この外れ値を除外した場合に予測モデルが変化するかどうかを確認する必要があります。

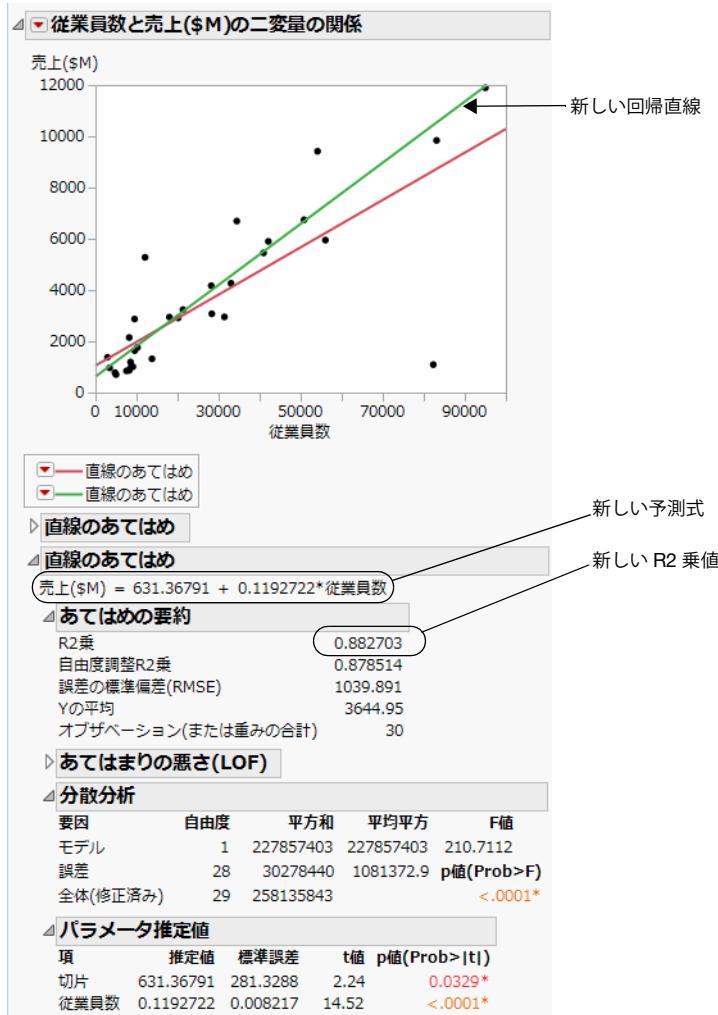
外れ値の除外

1. 外れ値をクリックします。
2. [行] > [除外する/除外しない] を選択します。
3. 「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 [直線のあてはめ] を選択して、モデルをあてはめます。

レポートウィンドウに次の情報が追加されます（図5.14を参照）。

- 新しい回帰直線。
- 新しい「直線のあてはめ」レポート。次の情報が表示されます。
 - 新しい予測式
 - 新しいR2乗値

図5.14 モデルの比較



結果の解釈

図5.14の結果を参考に、次の結論を導き出すことができます。

- 回帰直線は、外れ値によって大規模企業側で引き下げられ、小規模企業側で引き上げられていました。
- 新しいR2乗値（0.88）の方が最初のR2乗値（0.618）よりも1に近いため、新しいモデルの方がデータへのあてはまりが良好です。

結論の導き出し

新しい予測式を使用すると、従業員が 70,000 人の会社の平均売上は、次のように予測されます。

$$8961.37 \text{ ドル} = 631.37 + 0.119 * 70,000$$

最初のモデルによる予測値は 7499.68 ドルでした。つまり、このモデルで予測した売上の方が 1461.69 ドル高くなっています。

外れ値を除外した後の 2 番目のモデルの方が、最初のモデルと比べて、より良く売上を説明および予測しているようです。これで、使用すべき良いモデルを求めることができました。

1 変数による平均の比較

1 つの連続尺度の Y 变数に対して、1 つのカテゴリカルな X 变数がある場合、X 变数の水準間で平均を比較できます。

シナリオ

この例では、会社情報のサンプルデータ「Companies.jmp」を用います。このデータには、製薬業界とコンピュータ業界における 32 社の財務データが記録されています。

証券アナリストは、次の点を調査する必要があります。

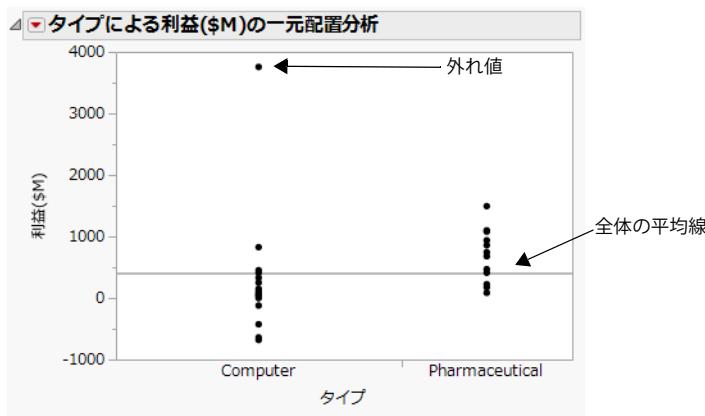
- コンピュータ関連企業と製薬会社において、利益は異なっているか

この比較を行うには、「利益(\$M)」を「タイプ」にあてはめます。

関係の洗い出し

- [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。
- 「Companies.jmp」データテーブルが開いたままになっている場合は、行が除外されたり非表示になっている可能性があります。行をデフォルトの状態（すべての行を分析に含め、かつ、表示する状態）に戻すには、[行] > [行の属性をクリア] を選択します。
- [分析] > [二変量の関係] を選択します。
- 「利益(\$M)」を選択し、[Y, 目的変数] をクリックします。
- 「タイプ」を選択し、[X, 説明変数] をクリックします。
- [OK] をクリックします。

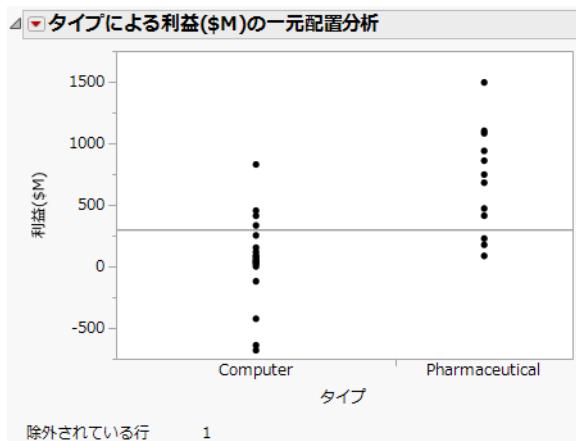
図5.15 会社のタイプ別の利益



「タイプ」が「Computer」のグループには、外れ値が1つあります。この外れ値があるせいで、散布図におけるY軸の範囲が広くなり、利益を比較しにくくなっています。この外れ値を除外し、非表示にします。

1. 外れ値をクリックします。
2. [行] > [除外する/除外しない] を選択します。該当するデータ点が計算の対象から外されます。
3. [行] > [表示しない/再表示] を選択します。該当するデータ点がすべてのグラフで非表示になります。
4. 外れ値を除いて図を再作成するには、「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 [やり直し] > [分析のやり直し] を選択します。元の散布図ウィンドウは閉じてもかまいません。

図5.16 更新後の図

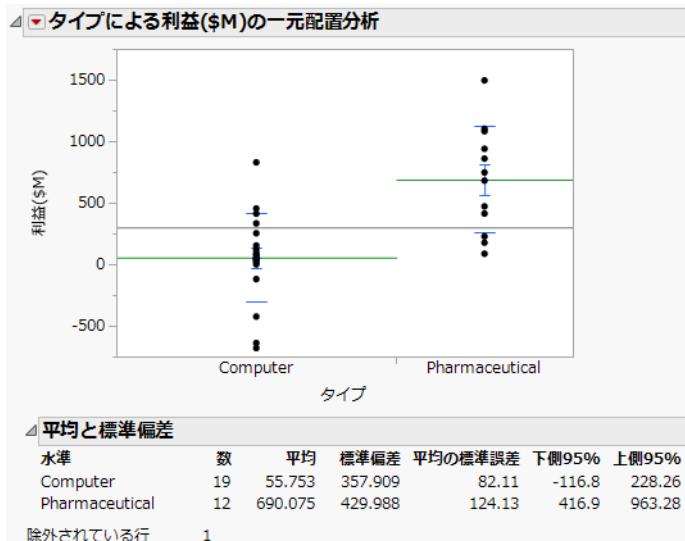


外れ値を削除した結果、データがより把握しやすくなりました。

5. 関係の分析を続けるには、「一元配置分析」の赤い三角ボタンをクリックし、メニューから次のオプションを選択します。

- [表示オプション] > [平均線]。散布図に平均線が追加されます。
- [平均と標準偏差]。平均と標準偏差を示すレポートが表示されます。

図 5.17 平均線とレポート



結果の解釈

証券アナリストは、コンピュータ関連企業と製薬会社の利益を比較する必要がありました。更新後の散布図から、製薬会社の方がコンピュータ関連企業よりも平均利益が高いことがわかります。レポートに示されている2つの平均値の差を計算すると、利益の差は約6億3,500万ドルになります。散布図からは、コンピュータ関連企業の中には赤字の会社があるのに対し、製薬会社は全社とも黒字であることも判明します。

t検定の実行

ここまででは、限られた標本（データテーブル内の会社）に基づいて考察してきました。ここで、次の点を検討するとします。

- 母集団において差はあるのか、それとも6億3,500万ドルの差は偶然だけで起きたものなのか
- 差がある場合には、それはどれぐらいなのか

これらの回答を導き出すには、二標本t検定を実行します。t検定の結果は、標本データをもとに、母集団について推論したものになっています。

t検定を実行するには、「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/プリングしたt検定] を選択します。

図5.18 t検定の結果



p値0.0001は有意水準0.05よりも小さいので、統計的に有意です。したがって、標本データの平均利益の差は偶然だけの産物ではないと結論付けることができます。つまり、製薬会社の平均利益とコンピュータ関連企業の平均利益は、母集団においても異なると言えます。

結論の導き出し

信頼限界の値によって、2業種の平均利益にどの程度の差があるかを判断します。図5.18の「差の上限信頼限界」と「差の下限信頼限界」の値を見てください。これらの結果から、製薬会社の平均利益はコンピュータ関連企業の平均利益よりも3億4,300万ドルから9億2,600万ドル高いという結論が導き出されます。

割合の比較

X変数およびY変数がカテゴリカルな変数の場合、Y変数の各水準の割合が、X変数の水準間で異なるかどうか比較できます。

シナリオ

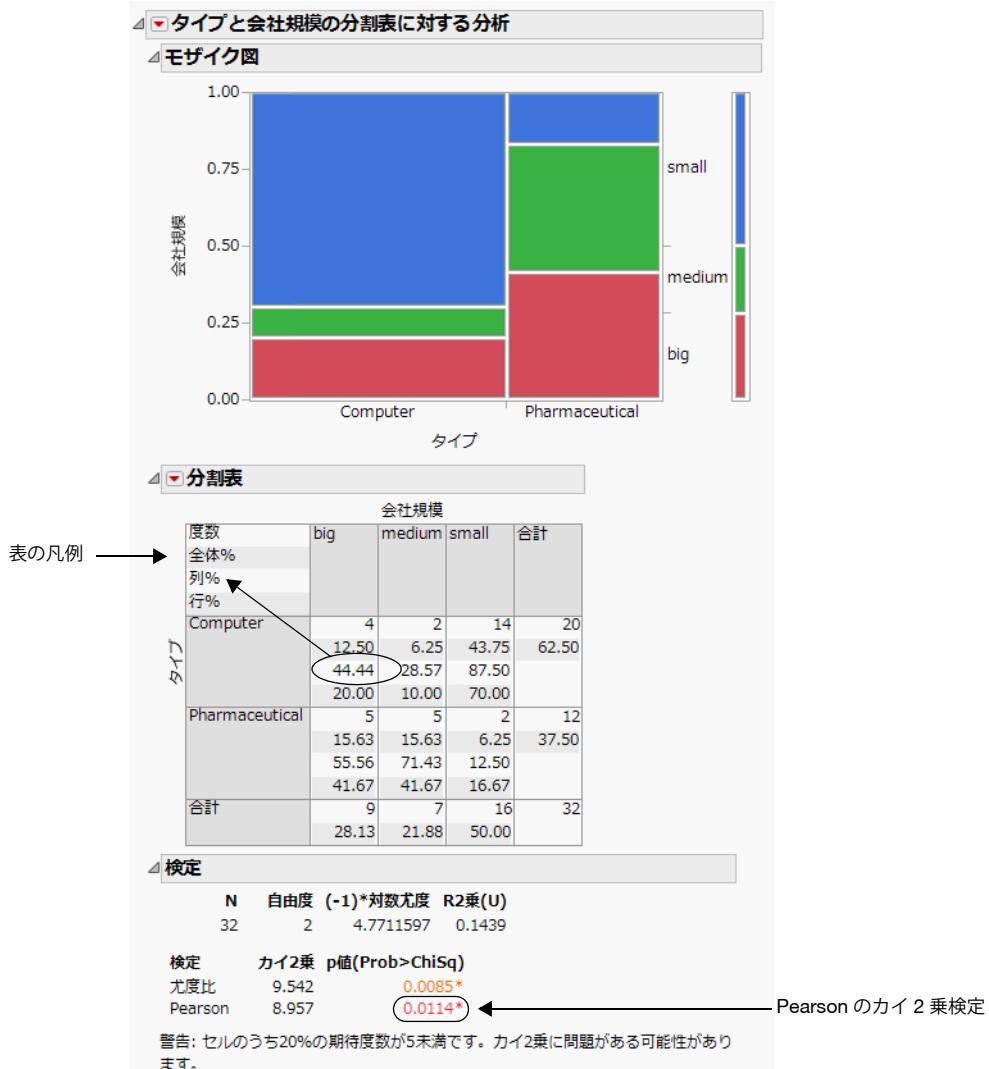
この例でも、会社情報のサンプルデータ「Companies.jmp」を使用します。「[1変数による平均の比較](#)」(141ページ)では、製薬会社の方がコンピュータ関連企業よりも平均利益が高いことを確認しました。

証券アナリストは、会社のタイプによって、会社規模が利益に与える影響に違いがあるかどうかに興味があります。ただし、この点を検討する前に、母集団において、コンピュータ関連企業と製薬会社で会社規模（大、中、小）の割合が同じかどうか調べてみます。

関係の洗い出し

- 【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。
- 「Companies.jmp」データファイルが、前の例で使われたままで開いた状態になっている場合は、行が除外されたり非表示になっている可能性があります。行をデフォルトの状態に戻すには（すべての行を対象に含め、いずれも非表示にしない）、【行】>【行の属性をクリア】を選択します。
- 【分析】>【二変量の関係】を選択します。
- 「会社規模」を選択し、【Y, 目的変数】をクリックします。
- 「タイプ」を選択し、【X, 説明変数】をクリックします。
- 【OK】をクリックします。

図5.19 会社タイプ別の会社規模



「分割表」には、この例では使用しない情報が含まれています。「分割表」の赤い三角ボタンをクリックし、表示されたメニューで [全体 %] と [列 %] の選択を解除して、これらの情報を削除します。図5.20は、更新後の表を示しています。

図5.20 更新後の分割表



分割表

会社規模				
度数	big	medium	small	合計
行%				
Computer	4	2	14	20
	20.00	10.00	70.00	
Pharmaceutical	5	5	2	12
	41.67	41.67	16.67	
合計	9	7	16	32

結果の解釈

「モザイク図」には、「分割表」に示されている統計量が視覚的に示されています。「モザイク図」と「分割表」は両方とも、2つの業界で大、中、小規模の会社が占める割合を比較しています。たとえば、「モザイク図」から、小規模企業の占める割合はコンピュータ業界の方が製薬業界に比べて高いことがわかります。「分割表」には具体的な統計量が示されています。小規模企業の占める割合はコンピュータ関連企業では70%、製薬会社では約17%となっています。

検定の解釈

ここまででは、限られた標本（データテーブル内の会社）に基づいて考察してきました。ここで、すべてのコンピュータ関連企業と製薬会社を含む母集団においても、割合に差があるかどうかを調べるとします。

差があることを確認するには、「検定」レポートにおける Pearson のカイ 2 乗検定の p 値を見ます。図5.19 を参照してください。p 値 0.011 は有意水準 0.05 よりも小さいため、次の結論を導き出せます。

- ・ 標本データの差は偶然だけで起きたものではない。
- ・ 母集団において、割合に差はある。

これで、大、中、小規模の割合に差があることが判明しました。次は、会社のタイプによって、会社規模が利益に与える影響に違いがあるかどうかを検討します。

複数の変数による平均の比較

「1変数による平均の比較」(141ページ) では、1つのカテゴリカル変数の水準間で平均を比較しました。複数の変数の水準間で平均値を一度に比較する場合には、分散分析 (ANOVA) を用います。

シナリオ

証券アナリストは、「割合の比較」節で挙げた問題「会社のタイプ（製薬会社またはコンピュータ関連企業）によって、会社規模が利益に与える影響に違いがあるかどうか」について検討します。

回答を導き出すには、次の2つの変数の組み合わせによって会社の利益を比較します。

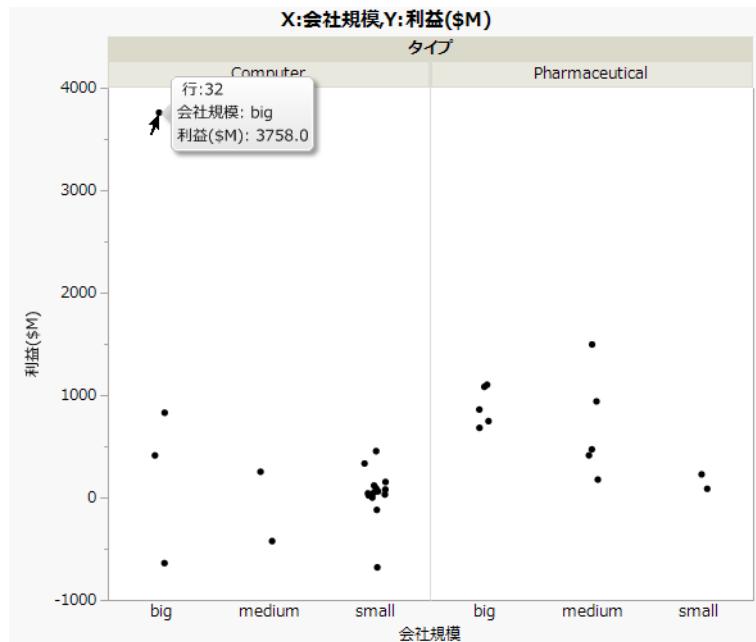
- ・ 「タイプ」(Pharmaceutical または Computer)
- ・ 「会社規模」(small、medium、big)

関係の洗い出し

まず、タイプと会社規模のすべての組み合わせについて、利益がどのように異なっているかをグラフで見てみます。

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. [グラフ] > [グラフビルダー] を選択します。「グラフビルダー」ウィンドウが表示されます。
3. 「利益(\$M)」をクリックし、「Y」ゾーンにドラッグ&ドロップします。
4. 「会社規模」をクリックし、「X」ゾーンにドラッグ&ドロップします。
5. 「タイプ」をクリックし、「グループX」ゾーンにドラッグ&ドロップします。

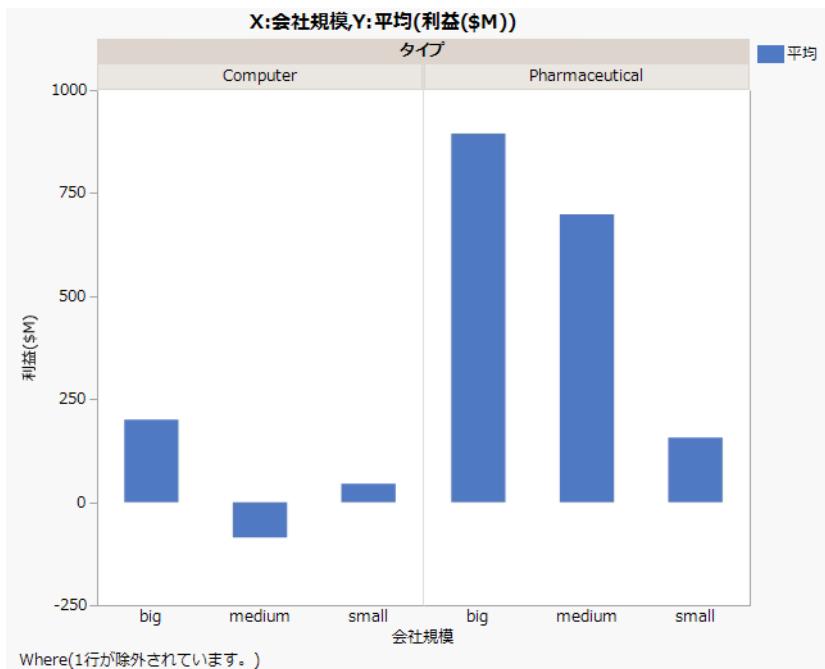
図5.21 会社の利益を表したグラフ



このグラフから、大手コンピュータ企業1社が極めて高い利益を上げていることがわかります。この外れ値があるせいで、グラフのY軸の範囲が広くなり、他のデータ点を比較しにくくなっています。

6. 外れ値をクリックして選択し、グラフ内を右クリックして [行] > [行の除外] を選択します。該当する点が削除され、グラフの範囲が自動的に更新されます。
7. 棒グラフアイコン をクリックします。平均だけに興味がある場合には、点よりも棒グラフのほうが分かりやすいです。

図5.22 外れ値を除外したグラフ



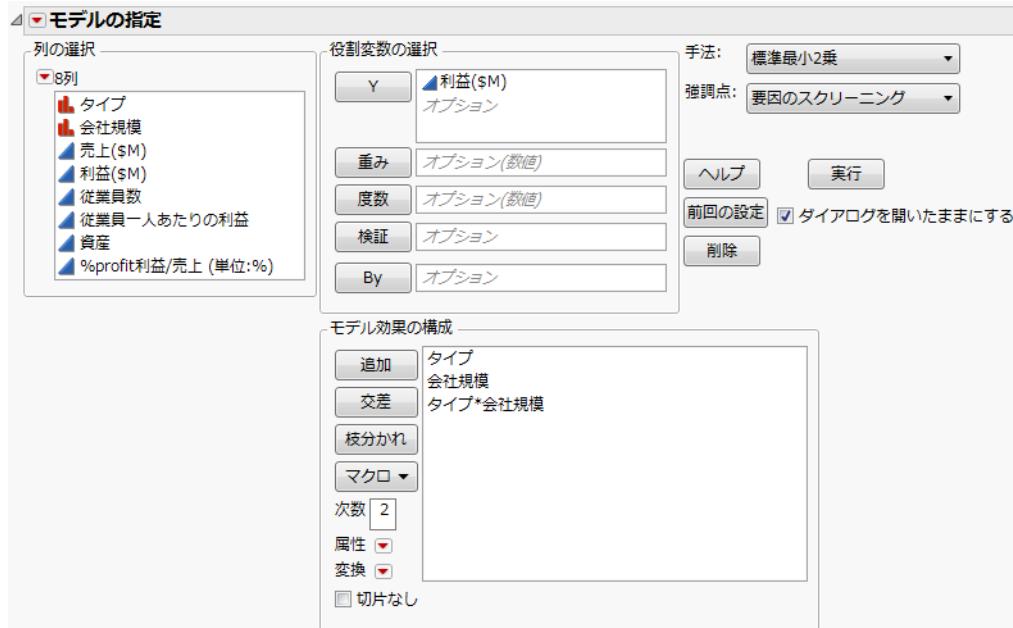
更新後のグラフから、製薬会社の方が平均利益が高いことがわかります。また、製薬会社についてのみ、会社規模によって利益に差が出ていることもうかがえます。1つの変数（会社規模）の効果が別の変数（会社タイプ）の水準に応じて変化する場合、これを**交互作用**と呼びます。

関係の定量化

このデータは1つの標本に過ぎないため、次の点を確認するとします。

- この標本で見られる差は、偶然だけで生じたものなのか
または
 - 母集団においても差があるといえるのか
1. 1 つのデータ点が除外された状態の「Companies.jmp」データテーブルに戻ります。「関係の洗い出し」(147ページ) を参照してください。
 2. [分析] > [モデルのあてはめ] を選択します。
 3. 「利益(\$M)」を選択し、[Y] をクリックします。
 4. 「タイプ」と「会社規模」の両方を選択します。
 5. [マクロ] ボタンをクリックして [完全実施要因] を選択します。
 6. 「強調点」メニューから [要因のスクリーニング] を選択します。
 7. [ダイアログを開いたままにする] チェックボックスをオンにします。

図5.23 設定後の「モデルのあてはめ」ウィンドウ



8. [実行] をクリックします。レポートウィンドウにモデルの結果が表示されます。

平均利益の差は、偶然だけで起きたものではなく、有意であることを確認するには、「効果の検定」レポートを見ます。

注: 「モデルのあてはめ」のすべての結果についての詳細は、『基本的な回帰モデル』の「モデルの指定」の章を参照してください。

効果の検定の表示

「効果の検定」レポート（図5.24を参照）には、統計的検定の結果が表示されます。「モデルのあてはめ」ウィンドウでモデルに含めた各効果（「タイプ」、「会社規模」、「タイプ*会社規模」）の検定結果が示されています。

図5.24 「効果の検定」レポート

要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
タイプ	1	1	1401847.4	10.1368	0.0039*
会社規模	2	2	724616.2	2.6198	0.0927
タイプ*会社規模	2	2	448061.5	1.6200	0.2180

まず、モデルの交互作用の検定（「タイプ * 会社規模」効果）に着目します。図5.22は、製薬会社の利益が会社規模によって異なることを示唆していました。しかし、効果の検定では、タイプと会社規模の交互作用の有

意性を認めることができません。p 値は 0.218 であり、有意水準 0.05 よりも大きくなっています。次に、モデルから交互効果を削除し、モデルを再実行してみます。

1. 「モデルのあてはめ」 ウィンドウに戻ります。
2. 「モデル効果の構成」 ボックスで、「タイプ*会社規模」 効果を選択して [削除] をクリックします。
3. [実行] をクリックします。

図5.25 更新後の「効果の検定」レポート

効果の検定					
要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
タイプ	1	1	1356297.9	9.3768	0.0049*
会社規模	2	2	434161.3	1.5008	0.2410

会社規模の p 値は 5% よりも大きく、会社規模による有意差は認められません。「タイプ」効果の p 値は小さいので、コンピュータ関連企業と製薬会社間のデータに見られた差は偶然だけで生じたものではないことを示しています。

結論の導き出し

証券アナリストは、会社のタイプ（製薬会社またはコンピュータ関連企業）によって、会社規模が利益に与える影響に違いがあるかどうかを確認する必要がありました。これに対し、次のような回答を得ることができました。

- 母集団において、コンピュータ関連企業と製薬会社の平均利益には差がある。
- 会社規模とタイプとの交互作用は、統計的に有意ではなかった。

複数の予測変数による回帰分析

「1つの予測変数による回帰分析」（136 ページ）では、予測変数 1 つと応答変数 1 つから成る単回帰モデルを紹介しました。重回帰では、複数の予測変数を用いて応答変数の平均を予測します。

シナリオ

この例では、キャンディバーの栄養価情報を記録した「Candy Bars.jmp」データテーブルを使用します。

栄養士は、次の情報からカロリーを予測したいと考えています。

- 総脂肪
- 炭水化物
- タンパク質

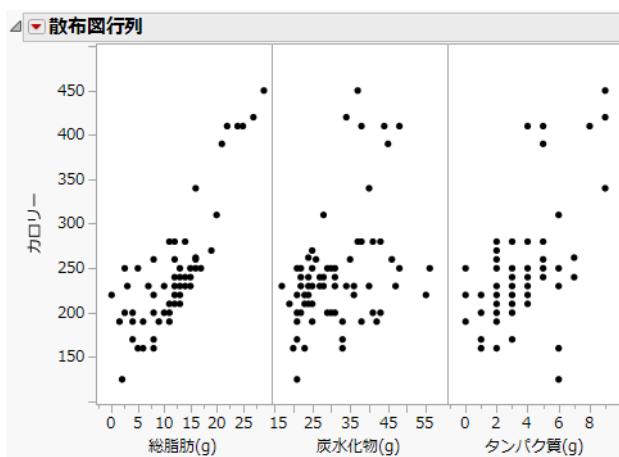
重回帰を用いて、これら 3 つの予測変数から応答変数の平均を予測します。

関係の洗い出し

総脂肪、炭水化物、タンパク質に対するカロリーの関係を視覚化するには、散布図行列を作成します。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Candy Bars.jmp」を開きます。
2. [グラフ] > [散布図行列] を選択します。
3. 「カロリー」を選択し、[Y, 列] をクリックします。
4. 「総脂肪(g)」、「炭水化物(g)」、および「タンパク質(g)」を選択し、[X] をクリックします。
5. [OK] をクリックします。

図 5.26 散布図行列の結果



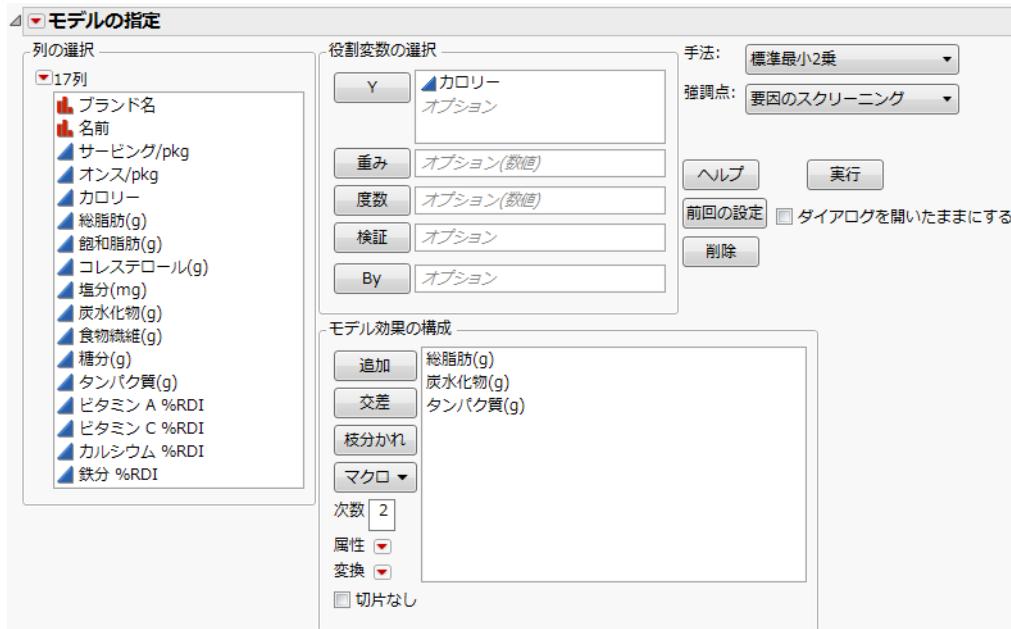
この散布図行列から、3変数とカロリーとの間に正の相関があることが分かります。相関が一番大きいのは、カロリーと総脂肪です。相関があることが分かったので、次に、平均カロリーを予測する重回帰モデルを構築します。

重回帰モデルの構築

引き続き、キャンディバーのサンプルデータ「Candy Bars.jmp」を使用します。

1. [分析] > [モデルのあてはめ] を選択します。
2. 「カロリー」を選択し、[Y] をクリックします。
3. 「総脂肪(g)」、「炭水化物(g)」、および「タンパク質(g)」を選択し、[追加] をクリックします。
4. 「強調点」メニューから [要因のスクリーニング] を選択します。

図5.27 「モデルのあてはめ」 ウィンドウ



5. [実行] をクリックします。

レポートウィンドウにモデルの結果が表示されます。結果を解釈する上では、次の情報に注目してください。

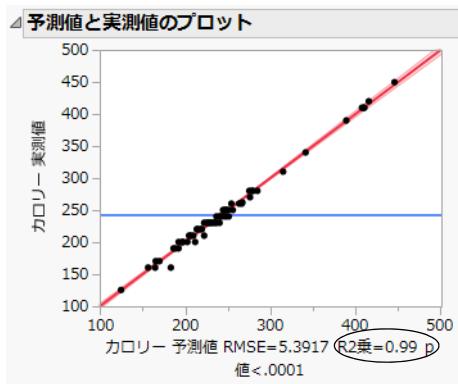
- 「「予測値と実測値のプロット」の表示」(152ページ)
- 「「パラメータ推定値」の解釈」(153ページ)
- 「「予測プロファイル」の使用」(154ページ)

注: すべてのモデルの結果についての詳細は、『基本的な回帰モデル』の「モデルの指定」の章を参照してください。

「予測値と実測値のプロット」の表示

「予測値と実測値のプロット」には、カロリーの予測値と実測値が表示されます。予測値が実測値に近いほど、散布図上の点が赤い線の近くに分布します。図5.28を参照してください。すべての点が赤い線の非常に近くに分布しているため、モデルはカロリーを良く予測していることがわかります。

図5.28 予測値と実測値のプロット



モデルの予測力を示す別の指標として、R2乗値（図5.28の図の下部に表示されています）があります。R2乗値は、カロリーの変動のうち、モデルによって説明できる割合を表します。この値が1に近いほど、モデルの予測力は高くなります。この例では、R2乗値は0.99です。

「パラメータ推定値」の解釈

「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。

- モデル係数
- 各パラメータのp値

図5.29 「パラメータ推定値」レポート

項	モデル係数		p 値	
	推定値	標準誤差	t値	p値($\text{Prob} > t $)
切片	-5.964301	2.899986	-2.06	0.0434*
総脂肪(g)	8.9899516	0.144981	62.01	<.0001*
炭水化物(g)	4.097505	0.071025	57.69	<.0001*
タンパク質(g)	4.4013313	0.39785	11.06	<.0001*

この例では、p値はすべて非常に小さな値になっています(<.0001)。つまり、3つの効果（総脂肪、炭水化物、タンパク質）すべてがカロリーの予測に有意に寄与しているということです。

モデル係数を使用して、特定量の総脂肪、炭水化物、タンパク質に対するカロリー値を予測できます。たとえば、次の特性を持つキャンディバーの平均カロリーを予測してみましょう。

- 総脂肪 = 11 g
- 炭水化物 = 43 g
- タンパク質 = 2 g

これらの値に対する平均カロリーの予測値は、次式で求められます。

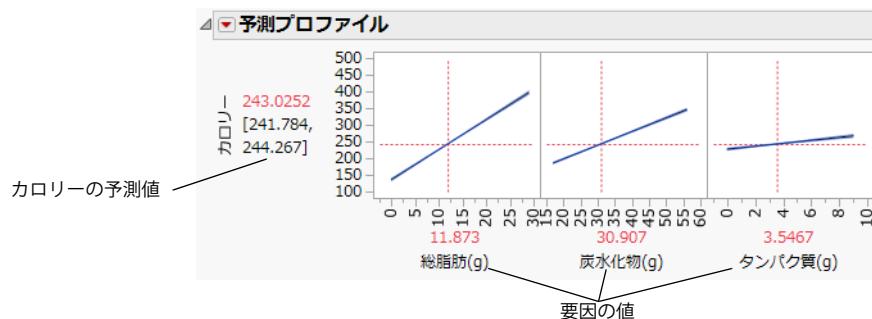
$$277.92 = -5.9643 + 8.99*11 + 4.0975*43 + 4.4013*2$$

この例で用いた特性は、Milky Way キャンディバー（データテーブルの59行目）と同じです。Milky Way のカロリーの実測値は280であり、モデルは良く予測していることを示しています。

「予測プロファイル」の使用

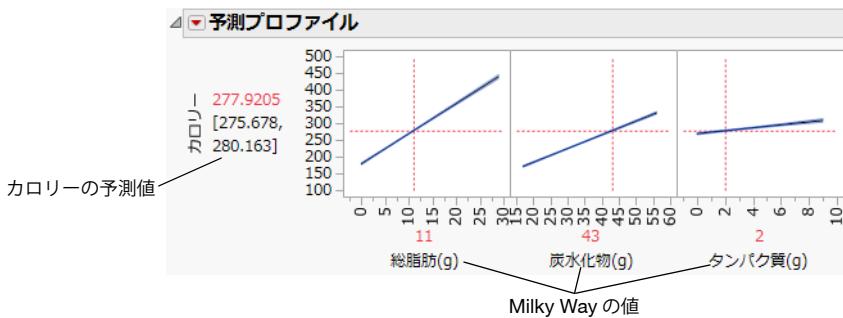
「予測プロファイル」では、要因の変化が予測値に与える影響を確認できます。プロファイルに示されている直線は、各要因の変化に伴うカロリーの変化の程度を示しています。「総脂肪(g)」の直線が一番急勾配になっています。つまり、総脂肪の変化がカロリーに一番影響を与えます。

図5.30 予測プロファイル



各要因の縦の線をクリックしてドラッグし、予測値の変化を確認できます。要因の現在値をクリックして変更することもできます。各要因の現在値をクリックして、たとえば、Milky Way キャンディバー（59行目）の値を入力すると、その予測値を知ることができます。

図5.31 Milky Way の要因値



注:「予測プロファイル」の詳細については、『プロファイル機能』の「プロファイル」の章を参照してください。

結論の導き出し

栄養士は、総脂肪、炭水化物、タンパク質に基づいてキャンディバーのカロリーを予測する優れたモデルを手に入れました。

第6章

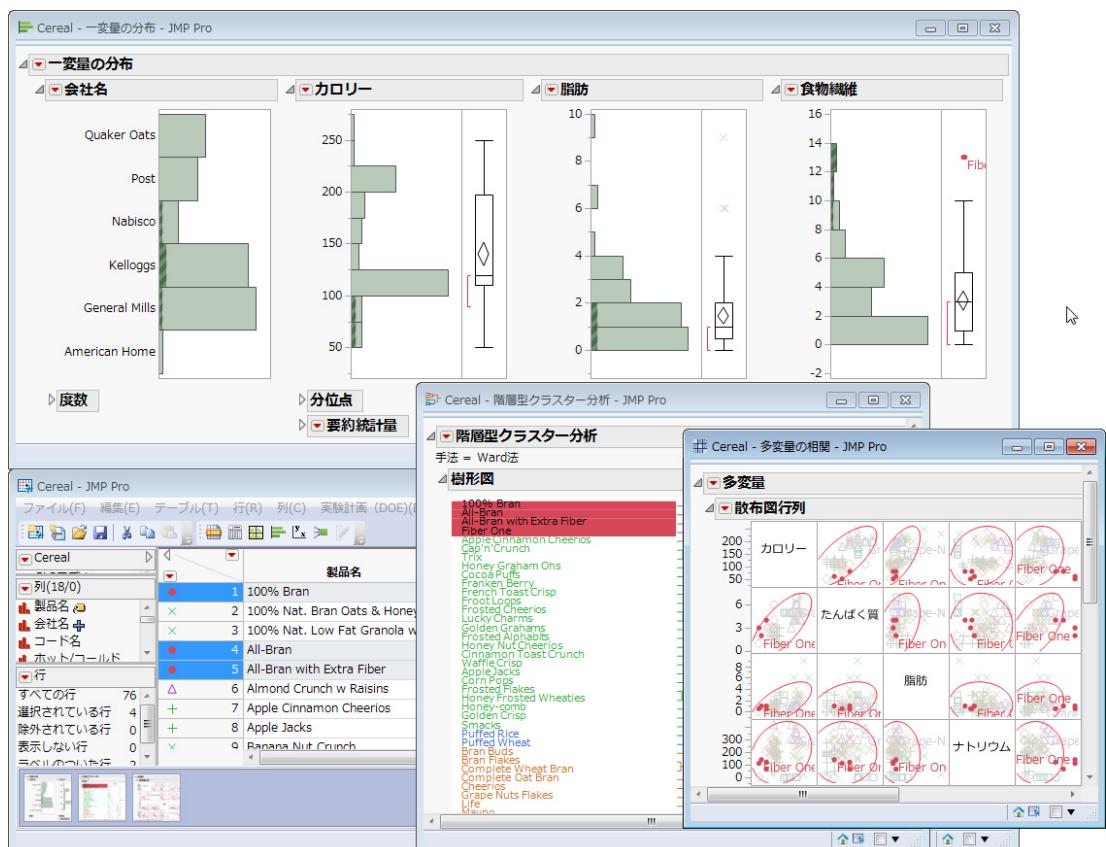
全体像

複数のプラットフォームを使ったデータの探索

JMPでは、複数のプラットフォームを用いることにより、データのさまざまな側面を探索できます。たとえば、まずはヒストグラムで個々の変数について簡単に把握し、その後、多変量分析やクラスター分析に進んで、より深く調べていくことができます。各ステップを実行するたびに、データについてより多くのことがわかつてきます。

この章では、JMPとともにインストールされる「Cereal.jmp」サンプルデータテーブルを使って、統計分析を段階的に行っていきます。[一変量の分布]、[多変量]、[階層型クラスター] の各プラットフォームでデータを探索する方法を学びます。

図6.1 JMPでの多変量分析



はじめる前に

JMP の便利な機能の 1 つは、データや分析結果が互いにリンクしていることです。作成したグラフやレポートは、データテーブルを通じて互いにリンクしています。図 6.1 で示しているように、データテーブル内でデータを選択すると、そのデータはレポートでも選択されます。この章の例を試すときは、リンクの機能を確認できるように、複数の JMP ウィンドウを開いたままにしておいてください。

複数プラットフォームでのデータの探索

どのシリアルが健康に良いのでしょうか？「Cereal.jmp」サンプルデータ（人気の高いシリアルに関して実際に収集されたデータ）には、食物繊維含有量、カロリー、およびその他の栄養素に関する情報の統計量が含まれています。最も健康に良いシリアルを特定するために、ヒストグラムと統計量の解釈、相関と外れ値の検出、散布図の作成、クラスター分析を段階的に行っていきましょう。

「一変量の分布」による分析

「一変量の分布」プラットフォームは、一変量に対する分析（ヒストグラムや要約統計量など）を行います。一変量という言葉は、2 つ（二変量）または 3 つ以上（多変量）の変数ではなく、1 つの変数だけが関与することを意味します。ただし、1 つのウィンドウ内で、複数の変数のにおける分布を調べることができます。各変数に対するレポートの内容は、カテゴリカル（名義または順序）変数か連続変数かによって異なります。

- カテゴリカル変数の場合、棒グラフが先頭に表示されます。この棒グラフには、順序変数や名義水準の各水準における度数を、棒で表示しています。レポートには、度数と割合が表示されます。
- 連続変数の場合、ヒストグラムと、外れ値の箱ひげ図が先頭に表示されます。ヒストグラムには、連続変数の値をグループ化して、そのグループごとの度数を棒で表示しています。レポートには、いくつかの分位点と要約統計量が表示されます。

データの分布を把握したら、続いてどのような分析を行うかを計画できます。

注: 「一変量の分布」プラットフォームの詳細については、『基本的な統計分析』の「一変量の分布」の章を参照してください。

シナリオ

ここでは、健康的な食生活を送るため、シリアルの栄養価を調べることにしましょう。シリアルデータの一変量の分布を分析すれば、次のような疑問が解決できます。

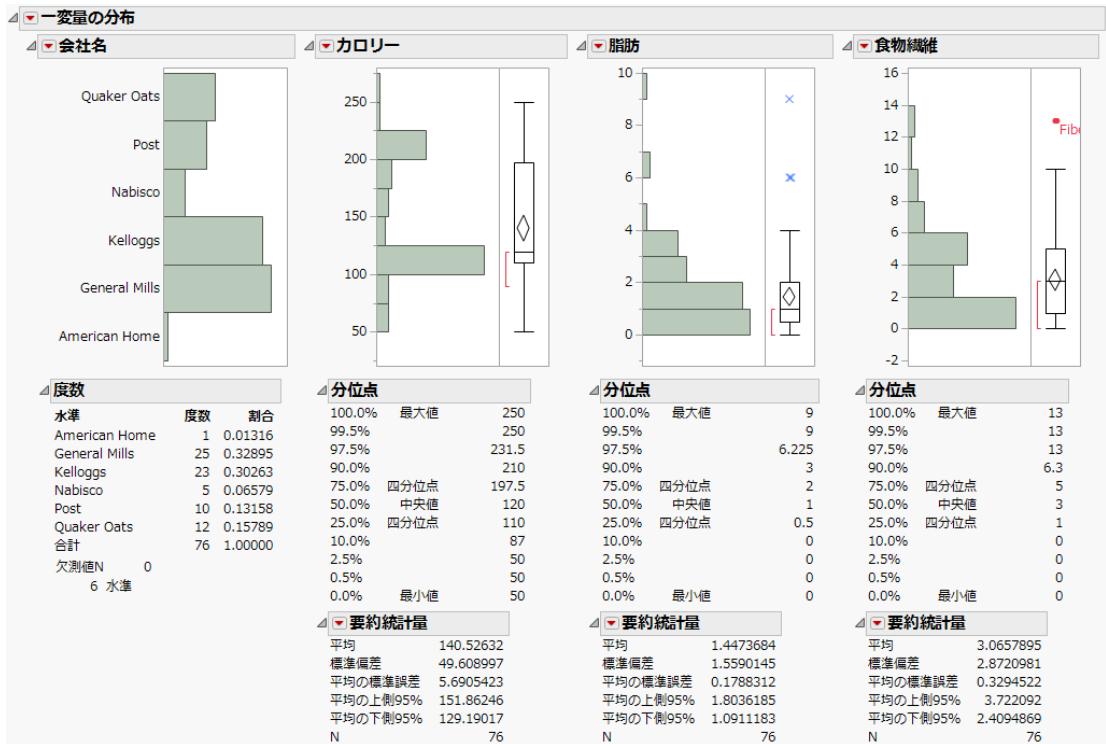
- 食物繊維が最も多く含まれているシリアルはどれか
- カロリーの中央値、最大値、最小値はどれくらいか
- 脂肪量の中央値はどれくらいか
- どのシリアルが最も脂肪量が多いか

- データに外れ値はあるか

一変量の分布の作成

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「Cereal.jmp」を開きます。
- [分析] > [一変量の分布] を選択します。
- Ctrlキーを押しながら「会社名」、「カロリー」、「脂肪」、「食物繊維」をクリックします。
- [Y, 列] をクリックし、[OK] をクリックします。

図6.2 「会社」、「カロリー」、「脂肪」、「食物繊維」の分布



「食物繊維」の分布で、次のことを確認してください。

- 「食物繊維」の箱ひげ図を見ると、「Fiber One」と「All-Bran with Extra Fiber」に最も多くの食物繊維が含まれていることがわかります。これらのシリアルは、食物繊維に関して外れ値です。

グラフ内のデータ点の横にシリアル名が表示されるよう、「Cereal.jmp」の「Fiber One」の行にはラベルが付けられています。ラベル全体を見るには、一番右側の縦の境界線を右にドラッグします。ラベルのないデータ点の上にカーソルを置くと、「All Bran with Extra Fiber」のラベルが表示されます。

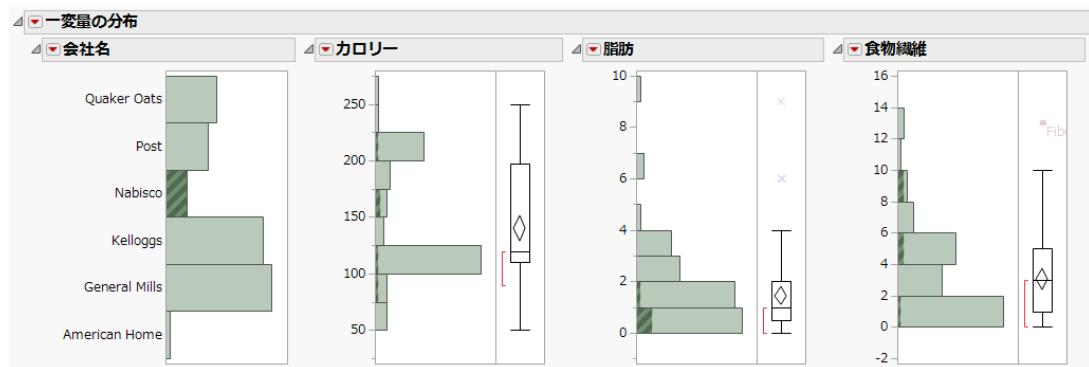
「脂肪」の分布で、次のことを確認してください。

- 「脂肪」の箱ひげ図で一番上のデータ点（xマーカー）の上にカーソルを置くと、「100% Nat. Bran Oats & Honey」が最も脂肪が多いことがわかります。

- 脂肪の「分位点」レポートを見ると、脂肪の中央値は1グラムです。
- カロリーの「分位点」レポートで、次のことを確認してください。
- カロリーの最大値は250です。
- カロリーの最小値は50です。

5. 「製品名」のヒストグラムで、「Nabisco」の棒をクリックします。

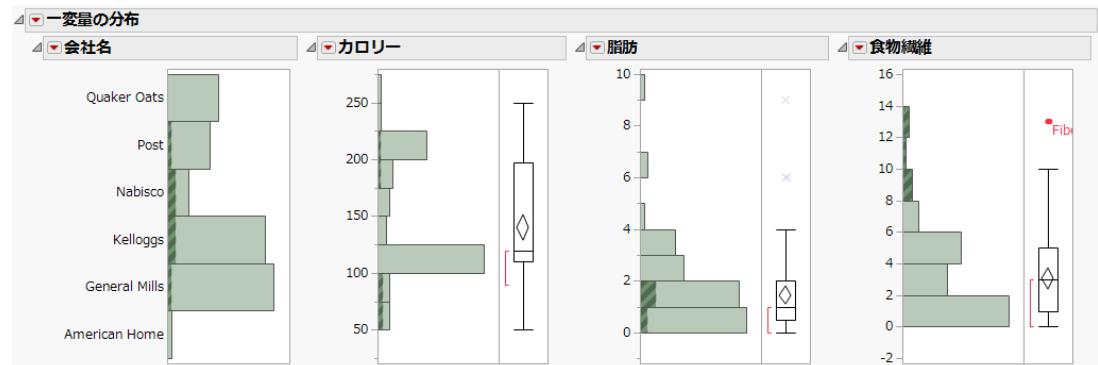
図6.3 Nabiscoシリアルの一変量の分布



Nabiscoシリアルのカロリー、脂肪、および食物繊維がそれぞれのヒストグラムでも強調表示されます。これにより、Nabiscoシリアルのカロリー、脂肪、および食物繊維の分布をデータ全体で相対的に比較できます。たとえば、脂肪に関して見ると、Nabiscoのシリアルはデータ全体と比べて少な目になっていることがわかります。

6. 「食物繊維」の最後の棒の下をクリックして、すべての棒の選択を解除します。
7. Shiftキーを押しながら「食物繊維」の値が8以上の棒をすべて選択します。

図6.4 食物繊維の多いシリアル

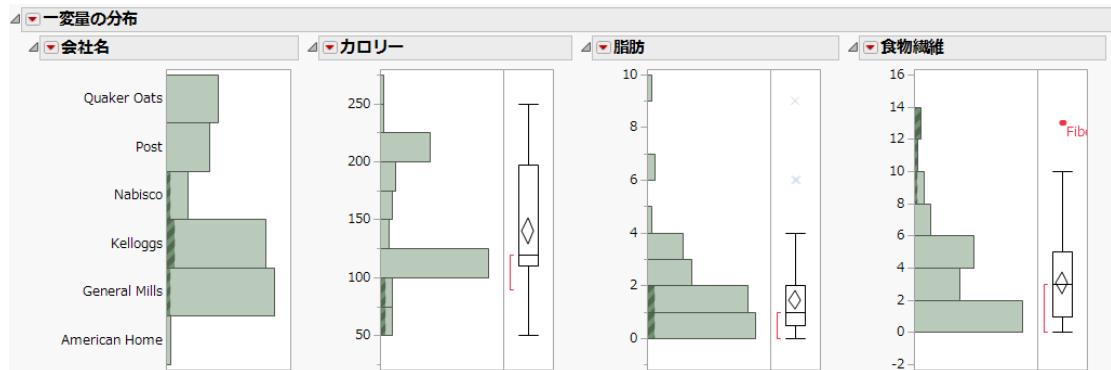


食物繊維の多いシリアルが「カロリー」と「脂肪」のヒストグラムでも強調表示されます。ヒストグラムはリンクされており、食物繊維の多いシリアルは脂肪が少ない傾向であることがわかります。

8. 「カロリー」ヒストグラムで、CtrlキーとShiftキーを押しながら200に近い棒をクリックして、それらの選択を解除します。

カロリーの高いシリアルがヒストグラムから除外されます。

図6.5 食物繊維が多くカロリーの低いシリアル



ヒント：「一変量の分布」レポートを開いたままにしておきます。これは、後ほどクラスター分析で使用します。[「似たようなデータをグループにまとめる分析」\(165 ページ\)](#) を参照してください。

結果の解釈

結果を見ると、次の疑問に対する回答が得られます。

どのシリアルが最も食物繊維が多いか 「食物繊維」の箱ひげ図を見ると、「Fiber One」と「All-Bran with Extra Fiber」に最も多くの食物繊維が含まれていることがわかります。これら2つのシリアルは外れ値です。

カロリーの中央値、最大値、最小値はどれくらいか 「カロリー」のヒストグラムから、カロリーの範囲は50～275であることがわかります。カロリーの分位点から、カロリーの範囲は50～250であり、中央値は120であることがわかります。分布は一様ではありません。

脂肪量の中央値はどれくらいか 脂肪の「分位点」レポートを見ると、脂肪の中央値は1グラムであることがわかります。

どのシリアルが最も脂肪量が多いか 脂肪の箱ひげ図から、100% Nat. Bran Oats & Honeyの脂肪量が最も多いことがわかります。このシリアルは外れ値です。

「一変量の分布」の結論

食物繊維を多く摂りたいのであれば、All-Bran with Extra Fiber や Fiber One を選ぶとよいでしょう。この2つのシリアルは、カロリーや脂肪も少ないです。ほとんどのシリアルでは食事の脂肪量が大幅に増えてしまうことはありませんが、100% Nat. Bran Oats & Honey は避けたほうが良さそうです。また、ほとんどのシリアルは脂肪が比較的少ないですが、カロリーが低いとは限らないこともわかりました。

パターンと関係の分析

どのシリアルを食べて、どのシリアルを避けるべきかがわかったので、次に、シリアルの成分がお互いにどのように関係しているかを見てみることにしましょう。「多変量の相関」プラットフォームでは、変数間のパターンや関係を観察することができます。「多変量の相関」レポートでは、次のことが実行できます。

- 相関係数行列を見て、変数の各ペアに関して、線形関係がどれくらい強いかを確かめる。
- 散布図行列を使って、従属性、外れ値、クラスターを特定する。
- 偏相関、逆相関、ペアごとの相関、共分散行列、主成分など、その他の手法を使って複数の変数を調べる。

注：「多変量の相関」プラットフォームの詳細については、『多変量分析』の「多変量の相関」の章を参照してください。

シナリオ

ここで、脂肪とカロリーなど、変数間の関係を調べてみましょう。「多変量の相関」プラットフォームでシリアルデータを分析すると、次のような疑問が解決できます。

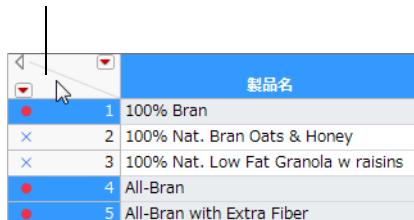
- どの変数ペアに強い相関があるか
- どの変数ペアに相関がないか

多変量レポートの作成

1. 「Cereal.jmp」データテーブル内で、列パネルの対角線より左下側部分をクリックして行の選択を解除します（図6.6）。

図6.6 行の選択解除

ここをクリックして行の選択を解除する。



2. [分析] > [多変量] > [多変量の相関] を選択します。

3. 「カロリー」から「カリウム」までをすべて選択し、[Y, 列] をクリックしてから [OK] をクリックします。

「多変量」レポートが表示されます。このレポートには、デフォルトで「相関」レポートと「散布図行列」が含まれています。「相関」レポートは、応答変数 (Y) の各ペアの線形関係の強さを要約する相関係数の行列です。濃い数字は、相関が高いことを示しています。

図6.7 「相関」 レポート

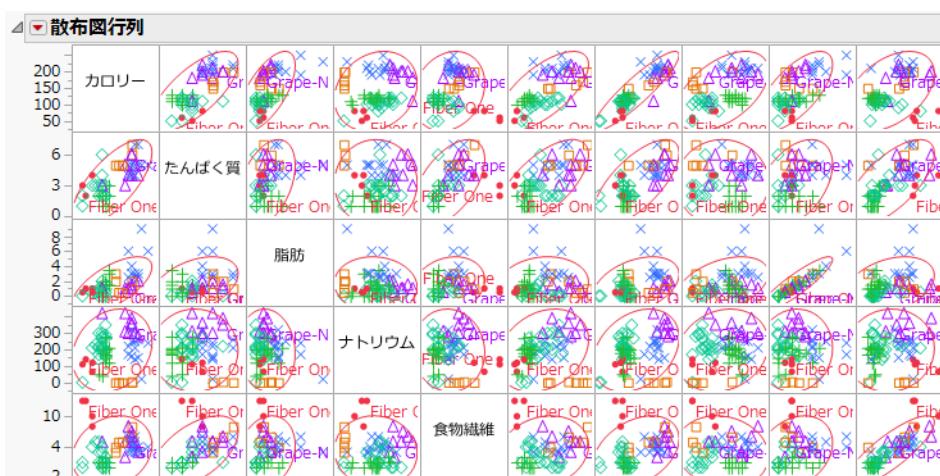
	カロリー	たんぱく質	脂肪	ナトリウム	食物繊維	複合炭水化物	炭水化物全体	砂糖	脂肪のカロリー	カリウム
カロリー	1.0000	0.7041	0.6460	0.1996	0.1953	0.6688	0.9076	0.5060	0.6709	0.4451
たんぱく質	0.7041	1.0000	0.4080	0.0050	0.5470	0.6486	0.6937	-0.0010	0.4288	0.6782
脂肪	0.6460	0.4080	1.0000	-0.0768	0.1824	0.1037	0.3860	0.4148	0.9013	0.3420
ナトリウム	0.1996	0.0050	-0.0768	1.0000	-0.0448	0.2619	0.3066	0.1767	0.0572	0.0459
食物繊維	0.1953	0.5470	0.1824	-0.0448	1.0000	0.1769	0.3668	-0.1264	0.2553	0.8326
複合炭水化物	0.6688	0.6486	0.1037	0.2619	0.1769	1.0000	0.7773	-0.1601	0.1558	0.2693
炭水化物全体	0.9076	0.6937	0.3860	0.3066	0.3668	0.7773	1.0000	0.4263	0.4636	0.5375
砂糖	0.5060	-0.0010	0.4148	0.1767	-0.1264	-0.1601	0.4263	1.0000	0.4369	0.1166
脂肪のカロリー	0.6709	0.4288	0.9013	0.0572	0.2553	0.1558	0.4636	0.4369	1.0000	0.3694
カリウム	0.4451	0.6782	0.3420	0.0459	0.8326	0.2693	0.5375	0.1166	0.3694	1.0000

次のことを確認してください。

- 「カロリー」の列を見ると、カロリーの値はナトリウムと食物繊維を除くすべての変数と相関が高いことがわかります。
- 「食物繊維」の列を見ると、食物繊維とカリウムは相関が高いようです。
- 「ナトリウム」の列を見ると、ナトリウムは他の変数とあまり相関がないことがわかります。

散布図行列上の確率楕円を見ると、変数間の関係をさらに深く理解できます。図6.8は、散布図行列の一部です。

図6.8 散布図行列（一部）



デフォルトでは、各散布図に95%の二変量正規楕円が表示されます。この楕円は、変数の各ペアが二変量正規分布に従うと仮定し、楕円のなかに約95%のデータが含まれるように描かれています。変数のペアが無相関の場合は、楕円が対角線上に伸びておらず真円に近くなります。一方、相関がある場合は、楕円が対角線上に伸び、細長くなります。

次のことを確認してください。

- 「ナトリウム」の列の楕円は真円に近いです。この形状は、ナトリウムが他の変数と無相関であることを示します。
- 「脂肪」の列の青いxマーカー (Nat. Bran Oats & Honey、Cracklin' Oat Bran、Banana Nut Crunch) は、楕円の外側にあります。つまり、これらは外れ値です (シリアルに含まれる脂肪の量に関する)。

散布団行列については、後でさらに詳しく見ていきます。

4. 「多変量」の赤い三角ボタンのメニューから、[ペアごとの相関係数] を選択して、「ペアごとの相関」レポートを表示します。

図6.9 「ペアごとの相関」レポート（一部）

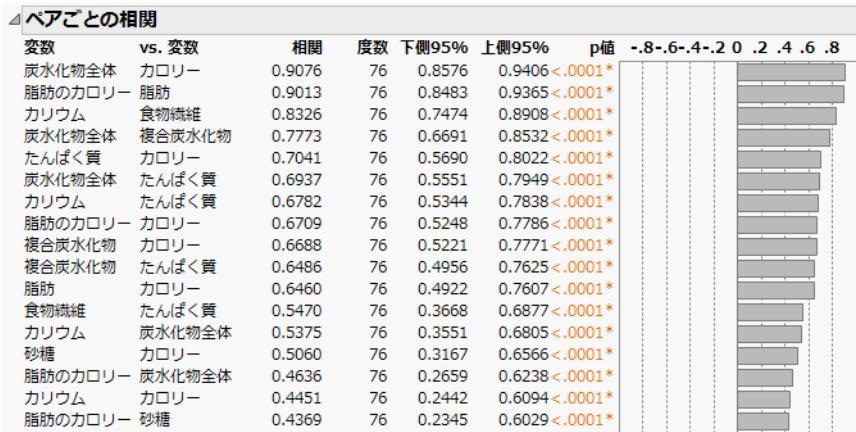


「ペアごとの相関」レポートには、Pearsonの積率相関係数が、Y変数のペアごとに表示されます。このレポートでは、有意確率も計算されます。また、相関の大きさが棒グラフで描かれます。

5. どのペアの統計的有意性が強いかを簡単に見るには、どのペアに相関性が高いかを簡単に見るには、レポートを右クリックし、[列の値で並べ替え] を選択します。そして、[p値] を選択し、[昇順] チェックボックスをオンにし、[OK] をクリックします。

これで、p値の小さいペアから順に表示されます。p値が小さいときは、「相関がある」という証拠をデータが示していることを意味します。最もp値が小さいものは、「炭水化物全体」と「カロリー」です。

図6.10 p値が小さいペア



結果の解釈

結果を見ると、次の疑問に対する回答が得られます。

どの変数ペアに強い相関があるか 「相関」 レポートと散布図行列を見ると、カロリーの値はナトリウムと食物繊維を除くすべての変数と相関が高いことがわかります。「ペアごとの相関」 レポートを見ると、炭水化物全体とカロリーの間の相関が最も強いことがわかります。

どの変数ペアに相関がないか 「相関」 レポートと散布図行列を見ると、ナトリウムには他の変数と相関がないことがわかります。

結論

今回の結果からも、100% Nat. Bran Oats & Honey を避けて、All-Bran with Extra Fiber や Fiber One を選ぶのがよいだろうという結論が導き出せます。なぜなら、食物纖維の多いシリアルは、カロリー、脂肪、砂糖が少なく、カリウムの量が多い傾向があるからです。また、炭水化物の多いシリアルは避けたほうがよいということも示唆されます。なぜなら、炭水化物が多いシリアルは、カロリーが高い傾向があるからです。

似たようなデータをグループにまとめる分析

クラスター分析は、多変量データにおいて近い値となっている行をグループにまとめる手法です。階層型クラスター分析を実行すると、データ行を結合していった結果が樹形図で示されます。クラスター分析を行うと、「食物繊維が多い」などの似た特徴をもつシリアルが、同じクラスターに分類されます。

注: 階層型クラスター分析の詳細については、『多変量分析』の「階層型クラスター分析」の章を参照してください。

シナリオ

お互いに似通っている、または似通っていないシリアルを調べてみることにしましょう。シリアルのデータに対してクラスター分析を実行すると、次の疑問に対する回答が得られます。

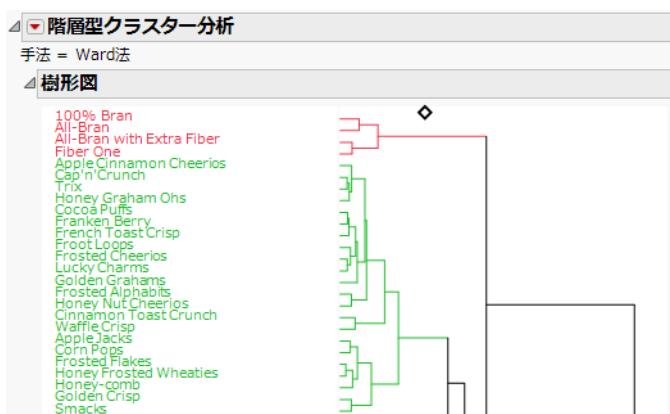
- 栄養価が少ないシリアルのクラスターはどれか
- ビタミンとミネラルが多く、砂糖と脂肪が少ないシリアルのクラスターはどれか
- 食物繊維が多く、カロリーが少ないシリアルのクラスターはどれか

階層型クラスタークラフの作成

1. 「Cereal.jmp」が表示された状態で、[分析] > [クラスター分析] > [階層型クラスター分析] を選択します。
2. 「カロリー」から「ビタミン強化」までを選択し、[Y, Columns] をクリックしてから [OK] をクリックします。

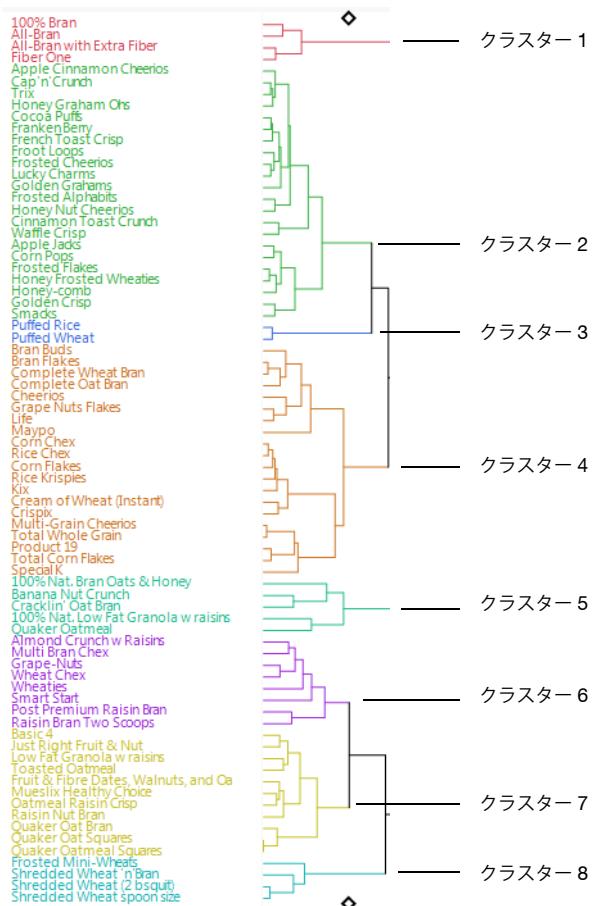
「階層型クラスター分析」レポートが表示されます。図6.11は、「階層型クラスター分析」レポートの一部です。クラスターは、データテーブルの行の状態に合わせて色分けされています。

図6.11 「階層型クラスター分析」レポート（一部）



3. 「階層型クラスター分析」レポートの赤い三角ボタンのメニューから、[クラスターの色分け] を選択します。
クラスターは、樹形図内の関係に従って色分けされています。

図6.12 色分けされたクラスター



各クラスター内のシリアルには似通った特徴があります。たとえば、クラスター1のシリアルの名前を見ると、これらのシリアルには多くの食物繊維が含まれていると察せられます。

All-Bran with Extra FiberとFiber Oneがどのようにグループ化されているかを確認してください。これらのシリアルは、クラスター内の他の2つのシリアルよりもより類似性が高いことがわかります。

図6.13 クラスター1内の似通ったシリアル



4. クラスター1を選択するために、右側の赤い横線上をクリックします。

4つのシリアルが赤く強調表示されます。

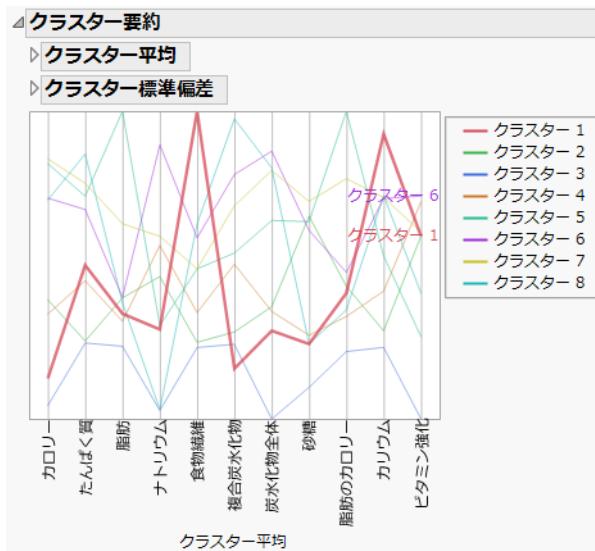
図6.14 クラスターの選択



5. クラスター内の似通った特徴を確認するために、赤い三角ボタンのメニューから【クラスターの要約】を選択します。

レポートの下部に、各クラスターの平均値を示した「クラスター要約」グラフが表示されます。たとえば、このクラスターのシリアルには、他のクラスターのシリアルよりも多くの食物繊維とカリウムが含まれています。

図6.15 クラスター要約

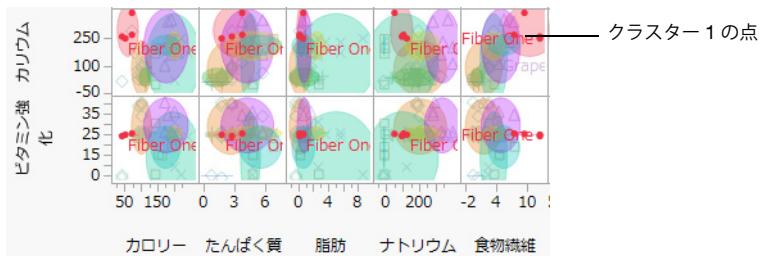


6. 赤い三角ボタンメニューから【散布図行列】を選択します。

このオプションは、「多変量の相関」プラットフォームで散布図行列を作成するのと同じです。

「カリウム」と「食物繊維」の散布図を見てください。選択されたシリアルは、プロットの右手、8~13グラムの間に位置しています。クラスター1のシリアルは、食物繊維とカリウムが豊富に含まれています。

図6.16 クラスター1の特徴



注: 前回作成した散布図行列がまだ開いたままの場合は、その図でも対応する点が選択されます。

結果の解釈

クラスターをクリックしたり「クラスター要約」レポートを見たりして、次のような特徴が把握できました。

- クラスター1のシリアル (Fiber One や All-Bran など) には食物繊維とカリウムが多く含まれており、カロリーは少なめである。
- クラスター2のシリアル (子供に人気のシリアルが多い) には砂糖が多く含まれており、食物繊維、複合炭水化物、たんぱく質の含有量は少ない。
- クラスター3のシリアル (Puffed Rice と Puffed Wheat) はカロリーは少ないが、栄養価も低い。
- クラスター4のシリアル (Total Corn Flakes や Multi-Grain Cheeriosなど) には一日に必要なビタミンとミネラルの100%が含まれている。これらのシリアルの脂肪、食物繊維、砂糖の含有量は少ない。
- クラスター5のシリアルにはたんぱく質と脂肪が多く含まれており、ナトリウムの含有量は少ない。このクラスターには Banana Nut Crunch や Quaker Oatmeal が含まれる。
- クラスター6のシリアルは脂肪が少なく、炭水化物が多い。このクラスターには Wheatus や Grape-Nuts といった昔ながらのシリアルが含まれている。
- クラスター7のシリアルはカロリーが高く、食物繊維が少ない。ドライフルーツが混ざっているシリアルの多くがこのクラスターに含まれる (Mueslix Healthy Choice, Low Fat Granola w Raisins, Oatmeal Raisin Crisp, Raisin Nut Bran, Just Right Fruit & Nut)。
- クラスター8のシリアルはナトリウムと砂糖が少なく、複合炭水化物、たんぱく質、カリウムが多い。Shredded Wheat と Mini-Wheat はこのクラスターに含まれる。

樹形図の結合を見ることで、各クラスター内でどのシリアルが似通った特徴を持っているかを調べることができます。

- クラスター1では、Fiber One と All-Bran with Extra Fiber の栄養価が似通っています。また、100% Bran と All-Bran も似通っています。これらの似通ったペアの各シリアルは、異なるメーカーによって製造されており、お互いに競合しています。
- クラスター2では、Frosted Flakes と Honey Frosted Wheatus が、一方はコーンフレークでもう一方は小麦フレークですが似通った特徴を持っています。Lucky Charms と Frosted Cheerios は似通っています。また、Cap' n' Crunch と Trix も似通っています。

結論

食物繊維が多く、カロリーの低いものを食べたいのなら、クラスター1のシリアルを試すのが妥当でしょう。クラスター3のシリアルは、ポップさせた小麦や米ですが、栄養価がよくありません。クラスター4のシリアルは栄養価が高いです。

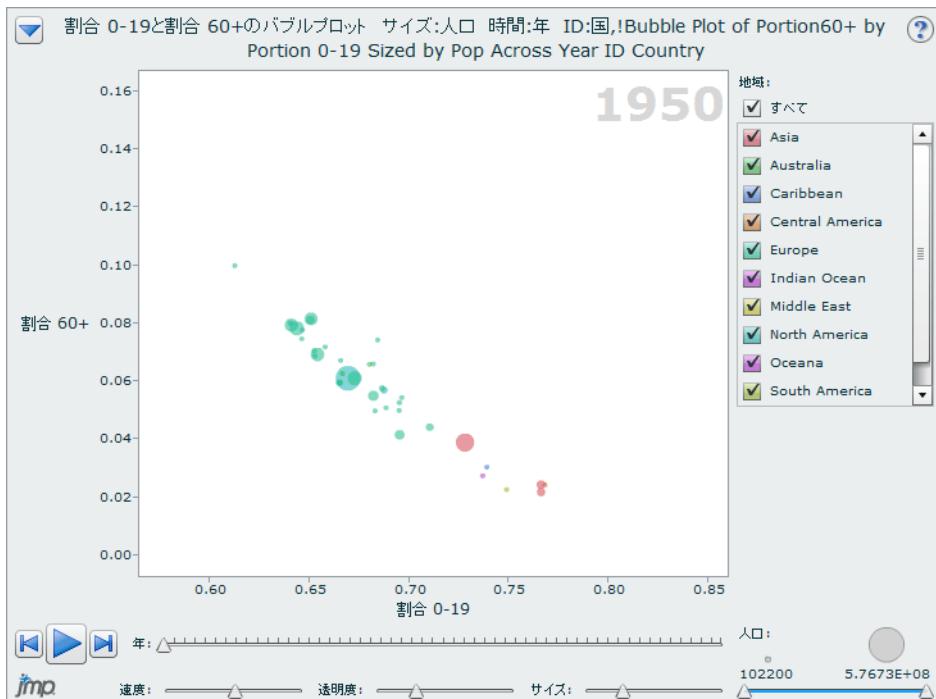
第7章

作業結果の保存と共有 結果の保存と再現

JMPでは、データから結果を求めた後、さまざまな方法で他のユーザと作業結果を共有できます。この章では、作業結果の共有方法をいくつか紹介します。

- プラットフォームの結果をジャーナルまたはプロジェクトとして保存する
- 結果、データテーブル、その他のファイルをプロジェクトに保存する
- 結果を再現するスクリプトをデータテーブルに保存する
- プラットフォームの結果からAdobe Flash (SWF) ファイルを作成する
- 結果をインタラクティブHTML (.htm, html) として保存する
- 結果をPowerPoint プрезентーション (.pptx) として保存する
- 結果をダッシュボードで共有する

図7.1 Adobe Flash (SWF) のバブルプロットの例



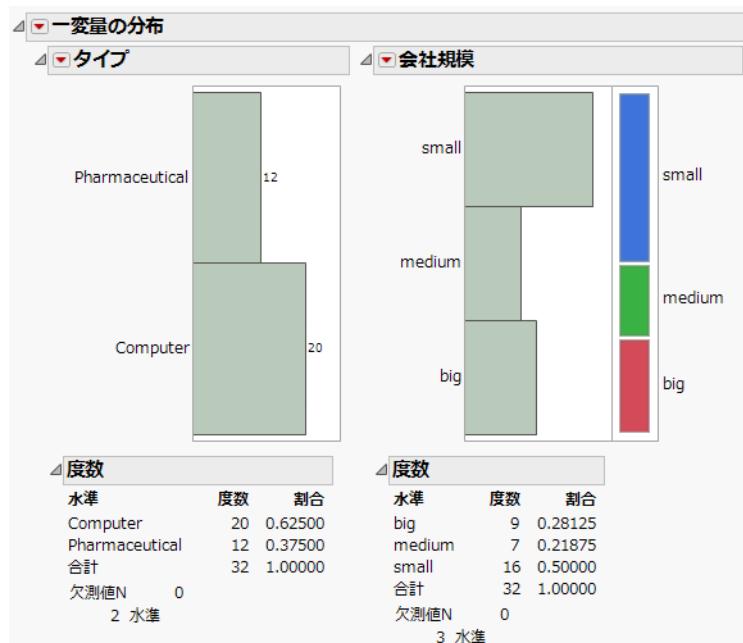
プラットフォーム結果のジャーナルへの保存

プラットフォームの結果を後から参照できるように保存するには、レポートウィンドウから「ジャーナル」を作成します。ジャーナルには、レポートウィンドウの内容がそのまま複製されます。既存のジャーナルを編集したり、既存のジャーナルにレポートを追加したりできます。ジャーナルはデータテーブルにはリンクしていません。ジャーナルを作成すれば、複数のレポートウィンドウの結果を、1つのレポートウィンドウにまとめて保存することもできます。ジャーナルを用いると、簡単に他のユーザと結果を共有できます。

ジャーナルの作成例

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] を選択します。
3. 「タイプ」と「会社規模」の両方を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。
5. 「タイプ」の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] > [度数の表示] を選択します。
6. 「会社規模」の赤い三角ボタンをクリックし、メニューから [モザイク図] を選択します。
7. [編集] > [ジャーナル] を選択して、これらの結果をジャーナルに追加します。結果がジャーナルウィンドウに複製されます。

図7.2 「一変量の分布」結果のジャーナル



ジャーナルに表示されている結果は、データテーブルにはリンクしていません。「タイプ」の棒グラフで、「Computer」の棒をクリックしても、データテーブルではどの行も選択されません。

ジャーナルは結果の単なる複製であり、赤い三角ボタンのメニューの大半は表示されません。ジャーナルにおいては、新たに追加されたレポートごとに赤い三角ボタンのメニューがあります。このメニューには次のオプションがあります。

新しいウィンドウで再実行 元のレポートの作成に使用した元のデータテーブルがある場合、このオプションで分析を再実行できます。結果は新しいレポートウィンドウに表示されます。

スクリプトの編集 分析を再実行するためのJSLスクリプトが記述されたスクリプトウィンドウが開きます。JSLは上級ユーザ向けの内容です。詳細については、『スクリプトガイド』と『スクリプト構文リファレンス』を参照してください。

ジャーナルへの分析の追加

別の分析を実行した場合、その分析結果を既存のジャーナルに追加できます。

1. ジャーナルを開いた状態で、[分析] > [一変量の分布] を選択します。
2. 「従業員一人あたりの利益」を選択し、[Y, 列] をクリックします。
3. [OK] をクリックします。
4. [編集] > [ジャーナル] を選択します。結果がジャーナルの下部に追加されます。

プロジェクトの作成

プロジェクトを作成して、複数の種類のJMPファイル（データテーブル、レポート、ジャーナル、スクリプトなど）を1つのファイルに保存できます。プロジェクトファイルには、その中のすべてのファイルを再度開くために必要な情報がすべて格納されます。

プロジェクトの作成例

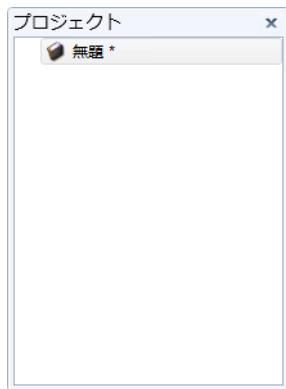
レポートを作成する

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] を選択します。
3. 「利益(\$M)」と「従業員一人あたりの利益」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。

プロジェクトを作成してデータテーブルとレポートを追加する

1. 新しいプロジェクトを作成するには、[ファイル] > [新規作成] > [プロジェクト] を選択します。新しいウィンドウが開いて、名称未設定のプロジェクトが表示されます。

図7.3 最初に表示されるプロジェクトウィンドウ（Windows）



2. プロジェクト（「無題」）を右クリックして [名前の変更] を選択し、新しい名前（「財務」）を入力します。
3. [一変量の分布] の結果をプロジェクトに追加するには、プロジェクト名を右クリックして [ウィンドウを追加] を選択します。
4. 「プロジェクトにウィンドウを追加」 ウィンドウで、「一変量の分布」の結果を選択します。
5. [OK] をクリックします。[一変量の分布] の結果がプロジェクトに追加されます。
6. 「Companies.jmp」データテーブルをプロジェクトに追加する場合は、ステップ3を繰り返して、ウィンドウから「Companies」データテーブルを選択します。
7. [OK] をクリックします。データテーブルがプロジェクトに追加されます。

図7.4 最終的なプロジェクト



プロジェクト内でリンクをダブルクリックして、データテーブルを開いたり、[一変量の分布] の結果を再現できます。

スクリプトの保存と実行

JMPのほとんどのプラットフォームオプションはスクリプトに記述できます。つまり、実行したほとんどの分析をJMPスクリプト言語 (JSL) のスクリプトとして保存できます。スクリプトを使用して、いつでも分析や結果を再現できます。

スクリプトの保存と実行の例

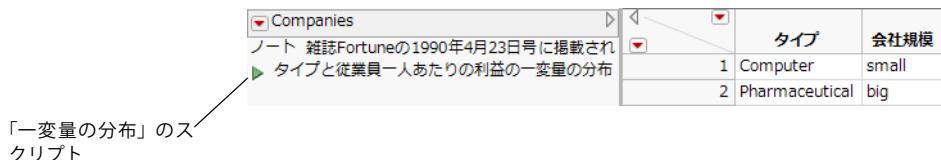
レポートを作成する

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] を選択します。
3. 「タイプ」と「従業員一人あたりの利益」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. 「タイプ」の赤い三角ボタンをクリックし、メニューから次のオプションを選択します。
 - [ヒストグラムオプション] > [度数の表示]
 - [信頼区間] > [0.95]
6. 「従業員一人あたりの利益」の赤い三角ボタンをクリックし、メニューから次のオプションを選択します。
 - [外れ値の箱ひげ図]（外れ値の箱ひげ図を削除）
 - [累積確率プロット]
7. 「一変量の分布」の赤い三角ボタンをクリックし、メニューから [積み重ねて表示] を選択します。

スクリプトをデータテーブルに保存して実行する

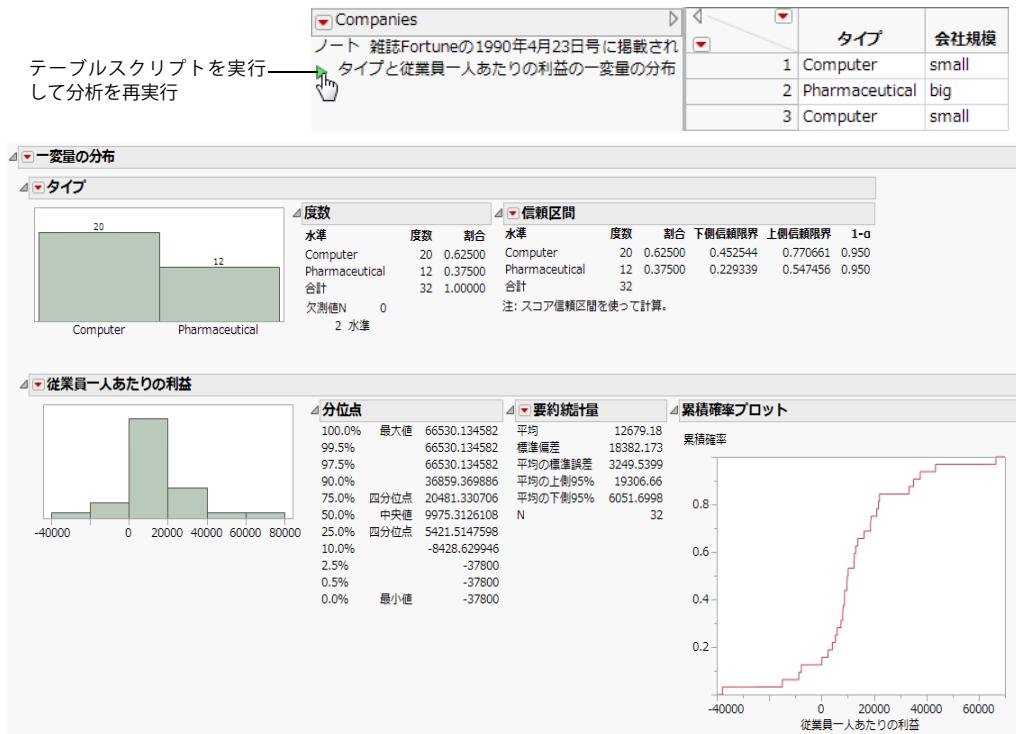
1. この分析を保存するには、「一変量の分布」の赤い三角ボタンをクリックし、メニューから [スクリプトの保存] > [データテーブルへ] を選択します。「一変量の分布」のスクリプトがテーブルパネルに表示されます。

図7.5 「一変量の分布」のスクリプト



2. 「一変量の分布」レポートウィンドウを閉じます。
3. 分析を再実行するには、「一変量の分布」の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。

図7.6 「一変量の分布」のスクリプトの実行



ヒント：テーブルスクリプトを右クリックすると、その他のオプションが表示されます。

スクリプトとJSLについて

この節で保存したスクリプトには、JMPスクリプト言語 (JSL) のコマンドが記述されています。JSLは上級ユーザ向けの内容です。詳細については、『スクリプトガイド』と『スクリプト構文リファレンス』を参照してください。

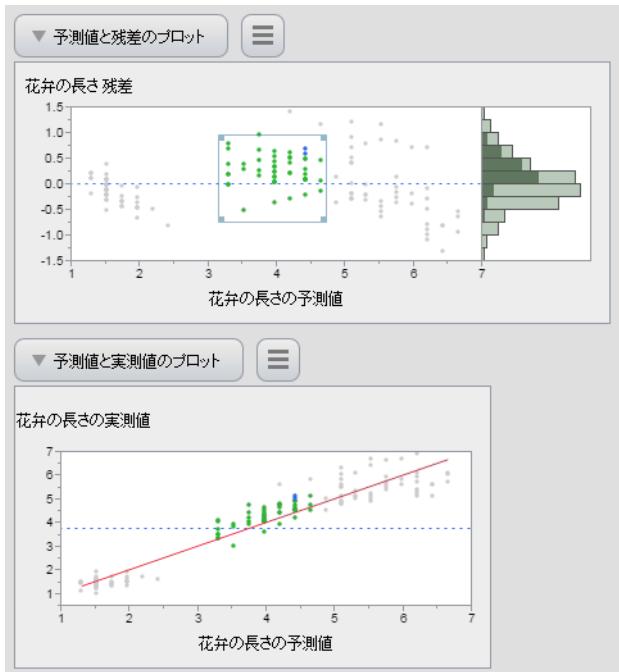
インタラクティブ HTML として保存

インタラクティブ HTML を使うと、動的なグラフが含まれたレポートを、JMPを使用していない人とも共有し、一緒にデータを探索することができます。JMP レポートは HTML 5 形式の Web ページとして保存でき、電子メールでの送信や、Web サイトへの掲載が可能です。ユーザはそれらのデータを、JMP で開いたときと同様に探索できます。

インタラクティブ HTML は、JMP の機能のうち次のような一部のものを備えています。

- ヒストグラムの棒を選択する、データの値を表示するなど対話式のグラフを探索する。
- 矩形をドラッグしてデータを選択する。
- レポートセクションの表示／非表示を切り替える。
- カーソルをレポートの上に置いてツールチップを表示する。
- マーカーのサイズを大きくする。

図7.7 インタラクティブ HTML でデータを選択



変数の水準値の順序、ヒストグラムの向き、背景色、データ点の色の指定など、グラフ内でカスタマイズした設定の多くは Web ページに保存されます。保存時に閉じていたグラフやレポートは、ユーザが開かない限り、Web ページ上でも閉じた状態で表示されます。

データを含むインタラクティブ HTML

JMP レポートをインタラクティブ HTML として保存すると、データも一緒に HTML に埋め込まれます。Web ブラウザは暗号化されたデータを読み込めないため、コンテンツは暗号化されません。機密データが含まれている場合は、結果をインタラクティブでない Web ページとして保存してください。(Windows の場合は [ファイル] > [名前を付けて保存] > [HTML ファイル]、Macintosh の場合は [ファイル] > [書き出し] > [HTML] を選択します。)

インタラクティブ HTML の作成例

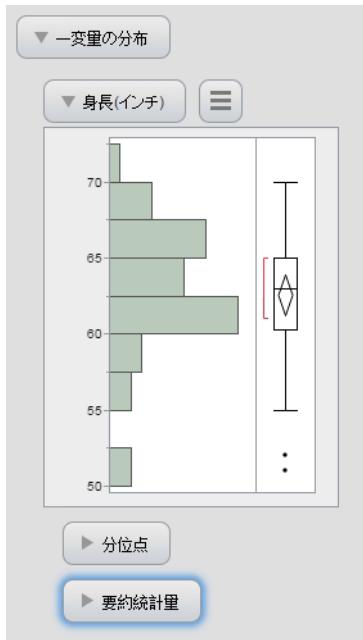
レポートを作成する

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] を選択します。
3. 「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。

インタラクティブ HTML として保存する

1. Windows の場合は、[ファイル] > [名前を付けて保存] を選択し、「ファイルの種類」のリストから [データを含むインタラクティブ HTML] を選択します。Macintosh の場合は、[ファイル] > [書き出し] > [データを含むインタラクティブ HTML] を選択します。
2. ファイルに名前を付けて保存します (Macintosh の場合は書き出します)。
デフォルトのブラウザに出力が表示されます。

図7.8 インタラクティブHTMLの出力



インタラクティブHTMLにおける操作方法については、
<http://www.jmp.com/support/help/InteractiveHTML/13/ShareJMPReports.shtml>（英語）をご覧ください。

レポートを PowerPoint プрезентーションとして保存

JMPの結果でプレゼンテーションを作成したいときは、Microsoft PowerPoint プрезентーション (.pptx) として保存します。.pptx ファイルとして保存した後、PowerPoint 内で JMP のコンテンツを並べ替えたり、テキストを編集したりできます。JMP レポートの各セクションは、PowerPoint に異なる形式で書き出されます。

- レポートの見出しは、編集可能なテキストボックスとして書き出されます。
- グラフは、イメージとして書き出されます。凡例など、グラフの要素によっては個別のイメージとして書き出されるものもあります。イメージのサイズは PowerPoint のスライドに合わせて調整されます。

プレゼンテーションに保存したいセクションを選択するには、選択ツールを使用してください。PowerPoint に保存してからファイルを開き、不要なものを削除することもできます。

注: Windows の場合、JMP で作成した .pptx ファイルを開くには PowerPoint 2007 バージョン以降が必要です。Macintosh の場合、PowerPoint 2011 以降が必要です。

1. JMP でレポートを作成します。

2. (Windows) [ファイル] > [名前を付けて保存] を選択し、「ファイルの種類」のリストから [PowerPoint プrezentation] を選択します。

3. (Macintosh) [ファイル] > [書き出し] > [Microsoft PowerPoint] を選択し、[次へ] をクリックします。

4. グラフィックファイル形式をリストから選択します。

Windows では、EMFがデフォルトの形式です。Macintosh では、PDFがデフォルトの形式です。

5. ファイルに名前を付けて保存します (Macintosh の場合は書き出します)。

[保存後にプレゼンテーションを開く] がデフォルトで選択されているため、ファイルが Microsoft PowerPoint で開きます。

注: Windows 版 JMP ではデフォルトのグラフィックファイル形式は EMF ですが、EMF は Macintosh 版 JMP ではサポートされていません。一方、Macintosh 版 JMP では PDF ですが、PDF は Windows 版 JMP ではサポートされていません。異なる OS の JMP で同一のグラフィックファイル形式を用いたい場合には、デフォルトのグラフィックファイル形式を変更する必要があります。それには、まず、[ファイル] > [環境設定] > [一般] を選択します。次に、[PowerPoint のイメージ形式] を PNG または JPEG に変更します。

プロファイル、バブルプロット、一変量の分布の Adobe Flash ファイル作成

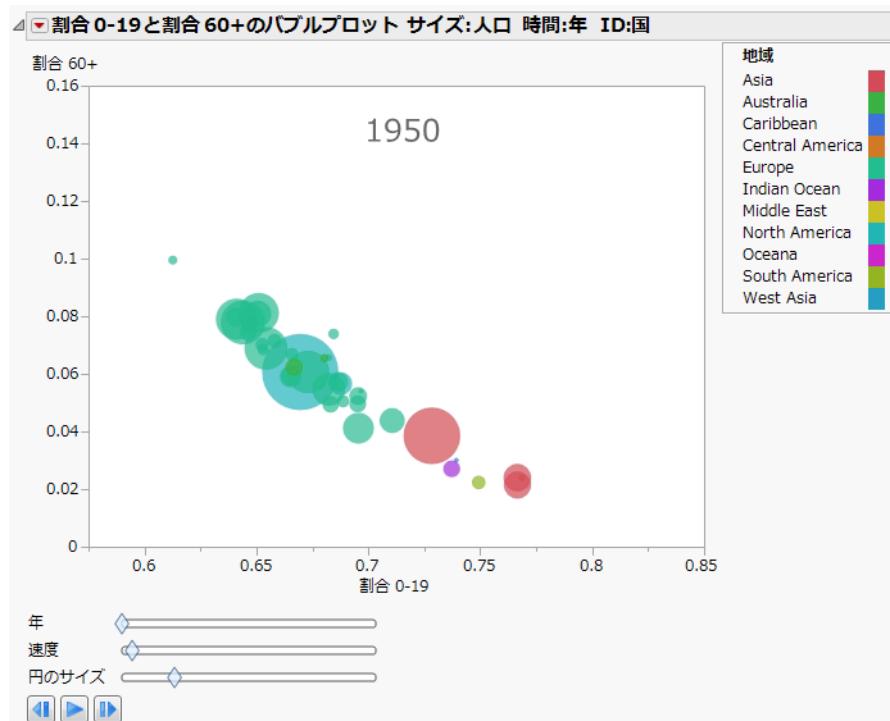
プロファイル、バブルプロット、または一変量の分布の対話型な結果は、SWF ファイルとして書き出すことができます。SWF ファイルは Adobe Flash Player によって表示できるので、JMP 以外でも共有することができます。SWF ファイルは、プレゼンテーションやアプリケーションに読み込むこともできます。また、SWF 出力を含む HTML ページとして、結果を保存することもできます。

バブルプロットの Adobe Flash バージョンの保存例

JMP でバブルプロットを作成する

- [ヘルプ] > [サンプルデータライブラリ] を選択し、「PopAgeGroup.jmp」を開きます。
- [グラフ] > [バブルプロット] を選択します。
- 「割合 60+」を選択し、[Y] をクリックします。
- 「割合 0-19」を選択し、[X] をクリックします。
- 「国」を選択し、[ID] をクリックします。
- 「年」を選択し、[時間] をクリックします。
- 「人口」を選択し、[サイズ] をクリックします。
- 「地域」を選択し、[色分け] をクリックします。
- [OK] をクリックします。

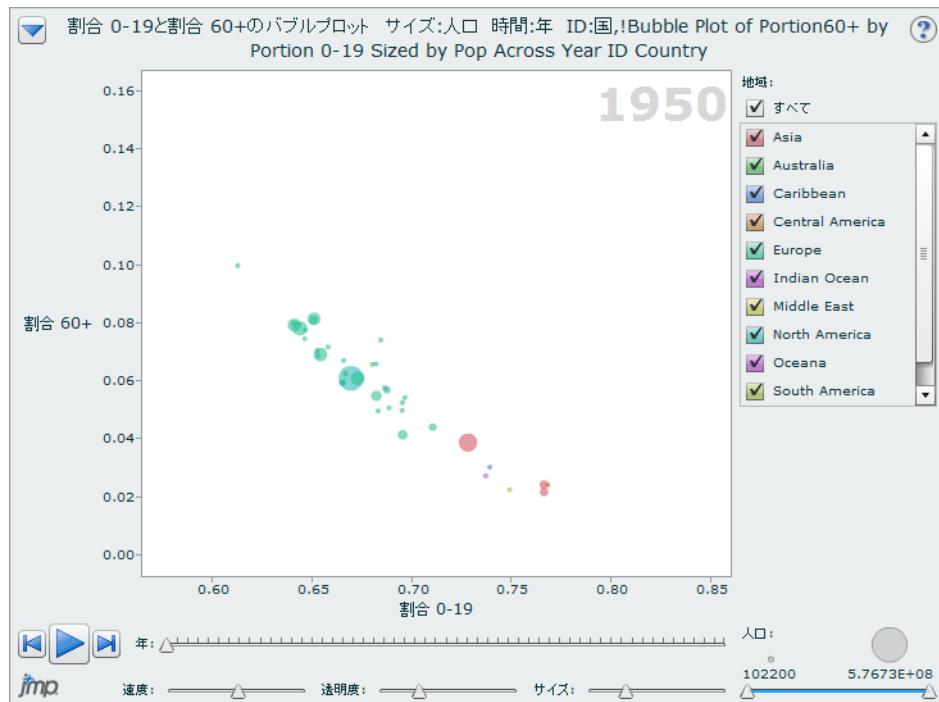
図7.9 最初に表示されるバブルプロット



バブルプロットをSWFファイルとして保存する

1. 赤い三角ボタンをクリックし、メニューから **[Flash(.SWF)形式で保存]** を選択します。
2. 「SWFで保存」ウィンドウで、ファイルの保存先を選択します。
3. **[保存]** をクリックします。Adobe Flash版のバブルプロットがHTML形式で保存され、Webブラウザに表示されます。

図7.10 Adobe Flash 版のバブルプロット



補足情報

JMP の Web サイトには、次のような追加情報が記載されています。

- プロファイル、バブルプロット、および一変量の分布の Adobe Flash ファイルの詳細
- Adobe Flash ファイルの Microsoft PowerPoint への読み込み手順

www.jmp.com/support/swfhelp/jp/ をご覧ください。

ダッシュボードの作成

ダッシュボードは、定期的にレポートを作成して発表するための視覚的なツールです。ダッシュボードには、レポート、データフィルタ、選択フィルタ、データテーブル、グラフィックを表示できます。ダッシュボードに表示される結果は、ダッシュボードを開くたびに更新されます。

ウィンドウの結合例

JMPで開いている複数のウィンドウを結合して、簡単にダッシュボードを作成できます。ウィンドウを結合するときには、レポートのグラフだけを表示したり、選択フィルタを含めたり、といったオプションを選択できます。

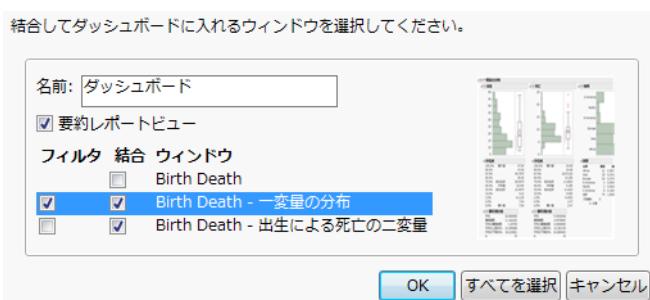
1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Birth Death.jmp」を開きます。
2. 「一変量の分布」および「二変量」のテーブルスクリプトを実行します。
3. [ウィンドウ] > [ウィンドウの結合] を選択します。

「ウィンドウの結合」ウィンドウが表示されます。

ヒント: Windowsの場合、JMP ウィンドウの右下隅の [整列メニュー] オプションから [ウィンドウの結合] を選択することもできます。

4. グラフだけを表示して、統計量の表を省略するために、[要約レポートビュー] を選択します。
5. 「結合」列で、「Birth Death - 出生による死亡の二変量」と「Birth Death - 一変量の分布」を選択します。
6. 「フィルタ」列で、「Birth Death - 一変量の分布」を選択します。

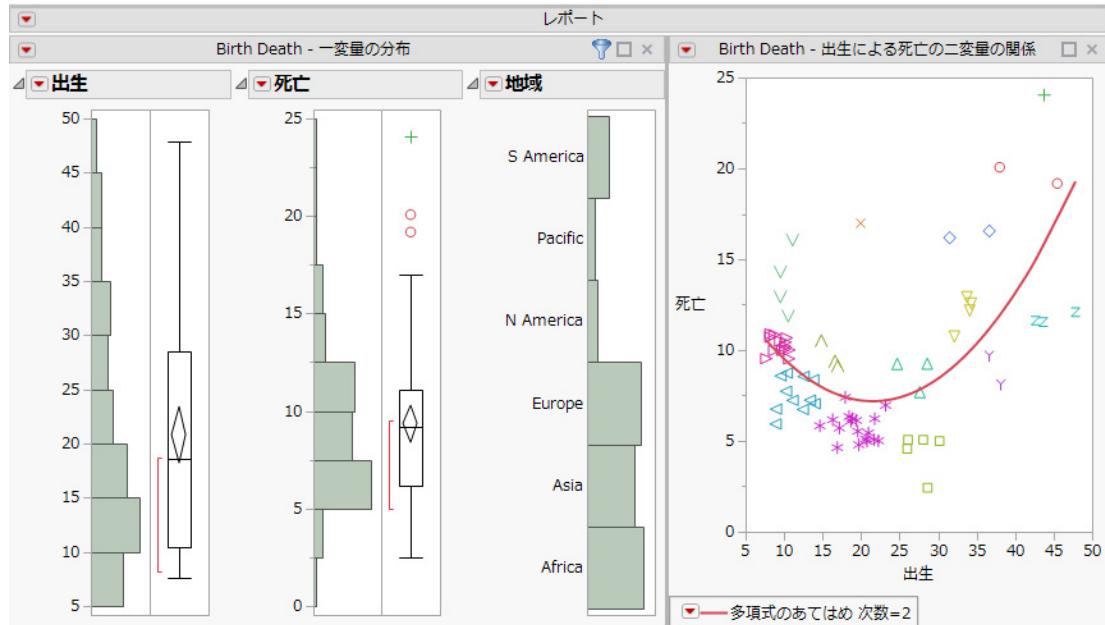
図7.11 ウィンドウの結合のオプション



7. [OK] をクリックします。

これで、2つのレポートが1つのウィンドウに結合されました。「一変量の分布」レポートの最上部にフィルタアイコン があることに注目してください。ヒストグラムの1つでいずれかの棒を選択すると、「二変量」のグラフ内の対応するデータが選択されます。

図7.12 結合されたウィンドウ

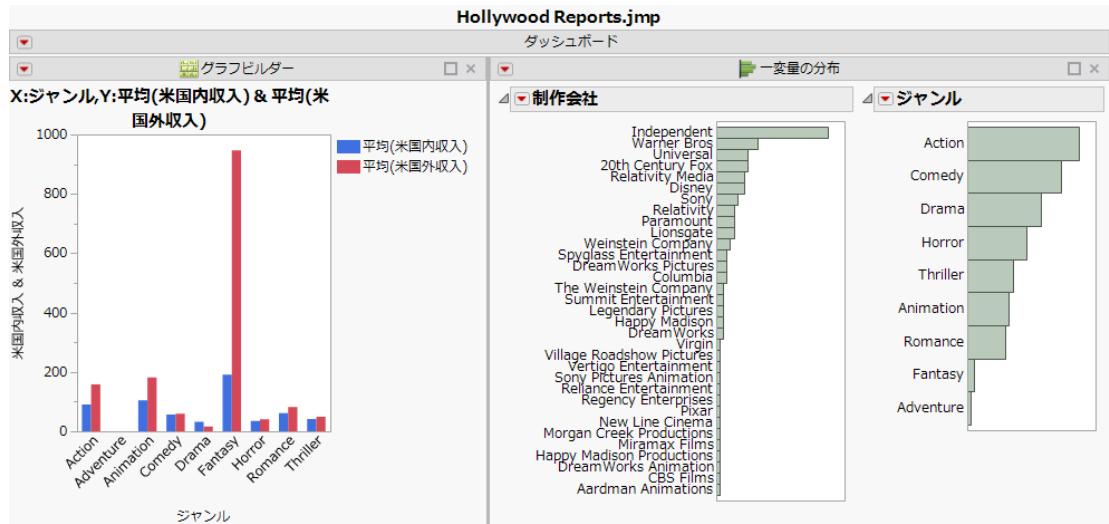


2つのレポートのあるダッシュボードの作成例

2つのレポートを作成し、翌日に、更新されたデータに対して再度レポートを作成したいとします。この例では、ダッシュボードビルダーによってダッシュボードを作成する方法を説明します。

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Hollywood Movies.jmp」を開きます。
2. 「一変量の分布: 制作会社とジャンルによる収益性」と「グラフビルダー: ジャンルごとの収入」のテーブルスクリプトを実行します。
3. いずれかのウィンドウから、[ファイル] > [新規作成] > [ダッシュボード] を選択します。
一般的なレイアウトのテンプレートが表示されます。
4. **2x1 ダッシュボード**のテンプレートを選択します。
ワークスペースに2つのレポートを配置するためのボックスが表示されます。
5. 「レポート」リストから、レポートのサムネールをダブルクリックして、それらをダッシュボードに配置します。
6. ダッシュボードビルダーの赤い三角ボタンメニューから [プレビューモード] を選択します。
ダッシュボードのプレビューが表示されます。グラフが、データテーブルとだけではなく、2つのグラフのあいだでもリンクしていることに注目してください。また、グラフには「一変量の分布」プラットフォームおよび「グラフビルダー」プラットフォームと同じ赤い三角ボタンのメニューオプションがあります。
7. [プレビューを閉じる] をクリックします。

図7.13 2つのレポートが表示されたダッシュボード



ダッシュボードの作成方法の詳細については、『JMPの使用法』の「JMPの拡張」の章を参照してください。

ダッシュボードをアドインとして保存

JMPアドインは、JSLスクリプトの1つで、[アドイン]メニューからいつでも実行できます。ダッシュボードをアドインとして保存すると、そのダッシュボードを他のJMPユーザと共有できます。

「[2つのレポートのあるダッシュボードの作成例](#)」(184ページ)に示すようにダッシュボードを作成した後、アドインとして保存するには、次の手順に従います。

1. ダッシュボードビルダーの赤い三角ボタンメニューから、[スクリプトの保存] > [アドインへ]を選択します。

アドインビルダーが開きます。

2. 「アドイン名」の横に、「ハリウッド映画ダッシュボード」とタイプします（アドインのファイル名）。
3. [メニュー項目]タブをクリックし、「メニュー項目名」の横に「ハリウッド映画ダッシュボード」とタイプします（アドインメニュー項目の名前）。

4. [保存]をクリックし、アドインをデスクトップに保存します。

アドインが保存され、[アドイン]メニューの中にインストールされます。

5. 「アドインビルダー」ウィンドウの[閉じる]をクリックします。
6. JMPのメインメニューから、[アドイン]を選択し、次に「ハリウッド映画ダッシュボード」を選択します。

「Hollywood Movies.jmp」から、グラフビルダーと「一変量の分布」レポートが生成されます。

アドインビルダーの詳細については、『スクリプトガイド』の「アプリケーションの作成」の章を参照してください。

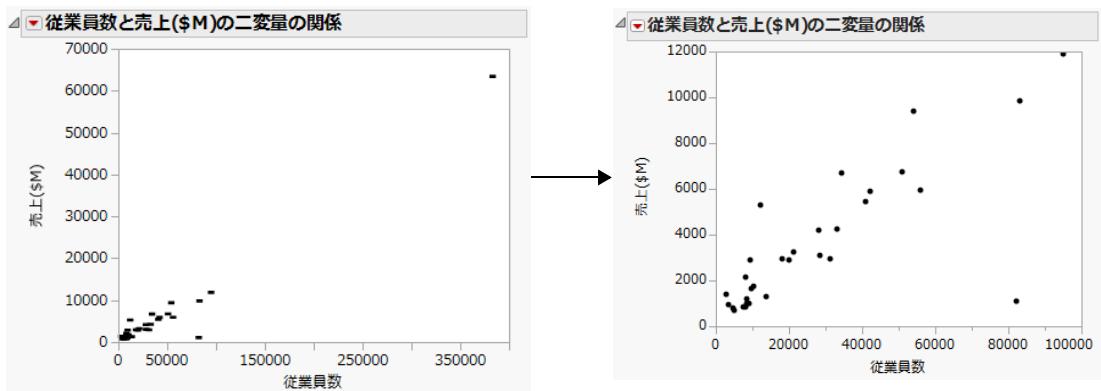
第8章

特別な機能 分析の自動更新やSASインテグレーション

JMPの特別な機能の一部を使用して、次の操作を実行できます。

- 分析やグラフを自動的に更新する
- プラットフォームの結果をカスタマイズする
- SASと連携して高度な分析機能を使用する

図8.1 特別な機能の例



```
DATA Candy_Bars; INPUT カロリー 総脂肪(g) 炭水化物(g) タンパク質(g); Lines;
310 20 28 6
230 12 27 4
220 12 24 3
170 8 21 3
200 2.5 43 1
260 16 26 5
190 1.5 42 2
190 11 21 2
230 12 28 3
230 13 25 2
;
RUN;
PROC GLM DATA=Candy_Bars ALPHA=0.05;
MODEL カロリー=総脂肪(g) 炭水化物(g) タンパク質(g);
RUN;
```

分析とグラフの自動更新

自動再計算の機能を用いると、データテーブルを変更した場合、そのデータテーブルに関連付けられている分析やグラフが自動的に更新されます。たとえば、データテーブルで値を除外、除外解除、または削除した場合、その変更内容が瞬時に、関連する分析やグラフに反映されます。次の点に注意してください。

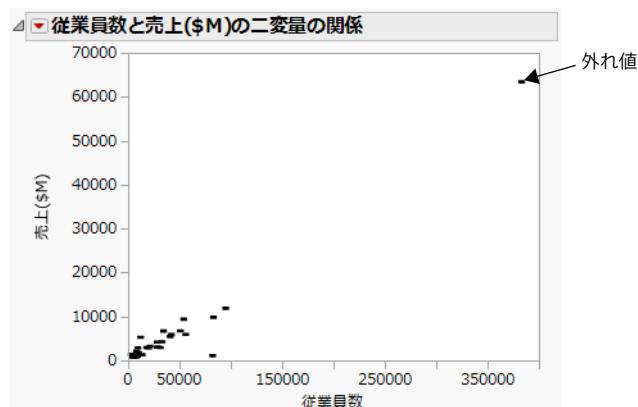
- 一部のプラットフォームは自動再計算に対応していない。詳細については、『JMPの使用法』の「JMPのレポート」の章を参照してください。
- [分析] メニューでサポートされているプラットフォームでは、自動再計算がデフォルトでオフになっている。ただし、[品質と工程] メニューでサポートされているプラットフォームでは、自動再計算がデフォルトでオンになっています。例外として、計量値/計数値ゲージチャート、工程能力、管理図では、デフォルトで自動再計算がオフになっています。
- [グラフ] メニューでサポートされているプラットフォームでは、自動再計算がデフォルトでオンになっている。

自動再計算の使用例

この例では、会社情報のサンプルデータ「Companies.jmp」を使用します。このデータには、製薬業界とコンピュータ業界の32社の財務データが記録されています。

- [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。
- [分析] > [二変量の関係] を選択します。
- 「売上(\$M)」を選択し、[Y, 目的変数] をクリックします。
- 「従業員数」を選択し、[X, 説明変数] をクリックします。
- [OK] をクリックします。

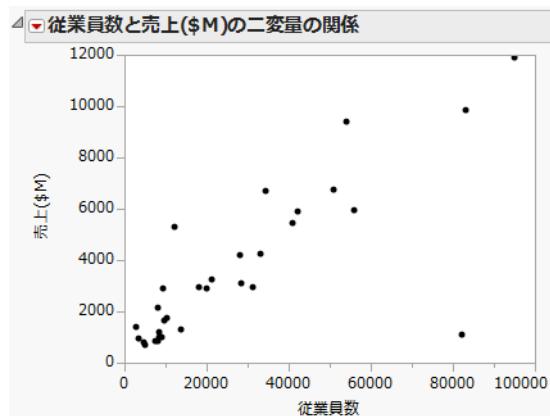
図8.2 最初に表示される散布図



最初に表示される散布図から、他社と比べて従業員数も売上も群を抜いている企業が1社あることがわかります。この企業は外れ値と判断し、除外することにします。点を除外する前に、除外したと同時に散布図が更新されるように、自動再計算をオンにしておきます。

6. 赤い三角ボタンをクリックし、メニューから [やり直し] > [自動再計算] を選択して、自動再計算をオンにします。
7. 外れ値をクリックして選択します。
8. [行] > [除外する / 除外しない] を選択します。該当する点が分析対象から除外され、散布図が自動的に更新されます。

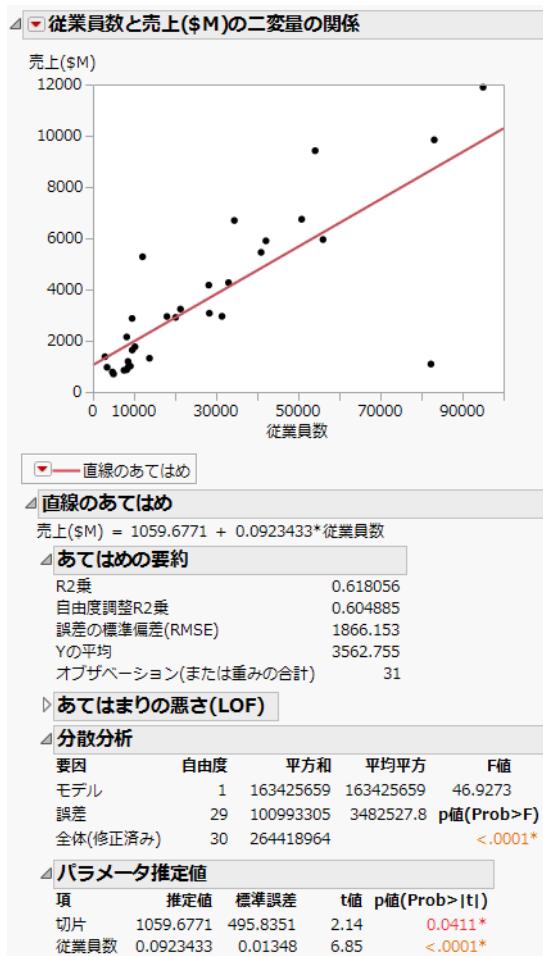
図8.3 更新後の散布図



データに回帰直線をあてはめると、右下隅の点が外れ値であり、直線の傾斜を引き下げていることがわかります。この外れ値を自動再計算をオンにした状態で除外すると、直線の傾斜が変化する様子を確認できます。

9. 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] を選択して、回帰直線をあてはめます。図8.4は、レポートウィンドウに追加された回帰直線と分析結果を示しています。

図 8.4 回帰直線と分析結果

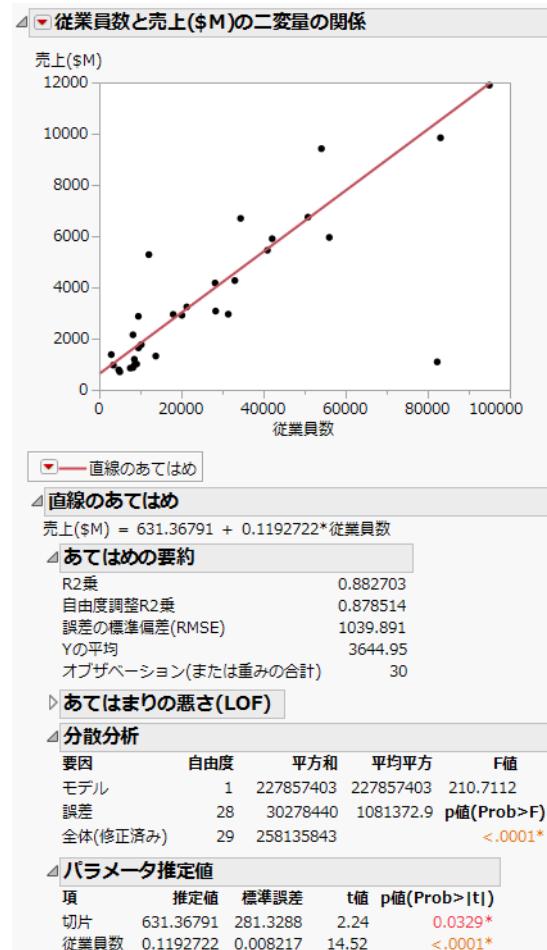


10. 外れ値をクリックして選択します。

11. [行] > [除外する / 除外しない] を選択します。点が除外されたことを受けて、回帰直線と分析結果が自動的に更新されます。

ヒント: 点を除外すると、そのデータ点を含めずに分析が再計算されますが、散布図上の該当データ点は表示されたままです。散布図上にも表示しない場合は、点を選択し、[行]>[非表示かつ除外]を選択します。

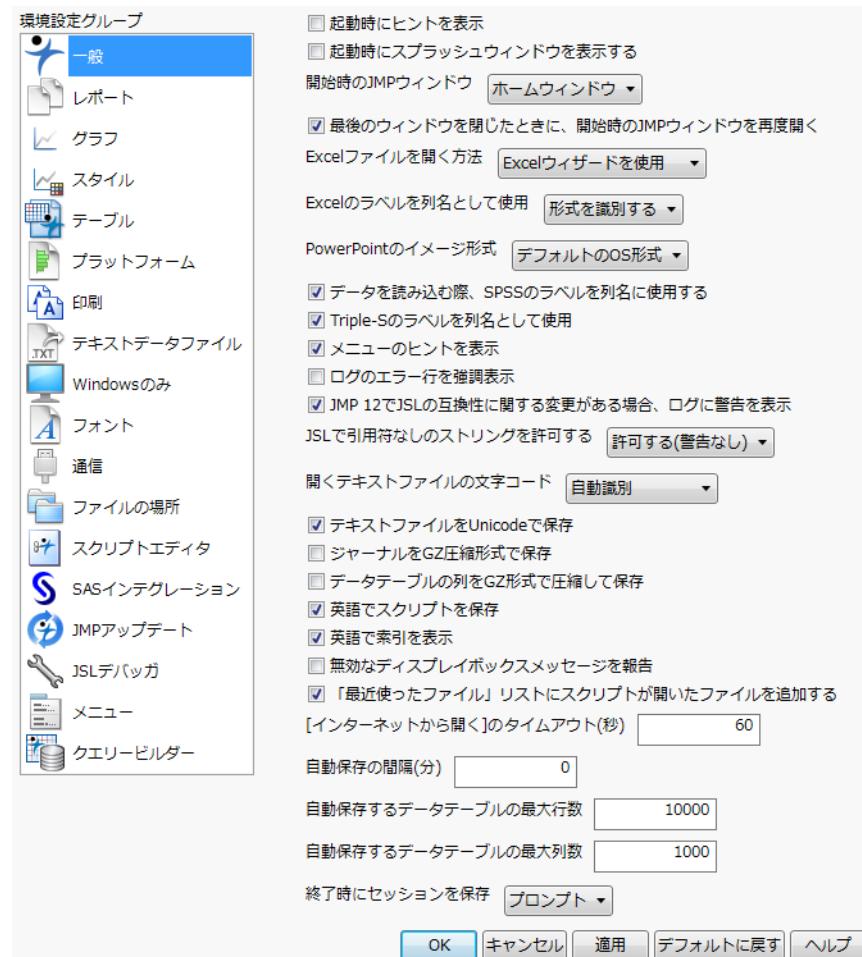
図8.5 更新後の回帰直線と分析結果



環境設定の変更

JMPの環境設定は「環境設定」ウィンドウで変更できます。「環境設定」ウィンドウを開くには、[ファイル] > [環境設定] を選択します。

図8.6 「環境設定」 ウィンドウ



「環境設定」 ウィンドウの左側には、環境設定グループの一覧が表示されます。ウィンドウの右側には、選択したカテゴリに属する変更可能な環境設定がすべて表示されます。

環境設定の変更例

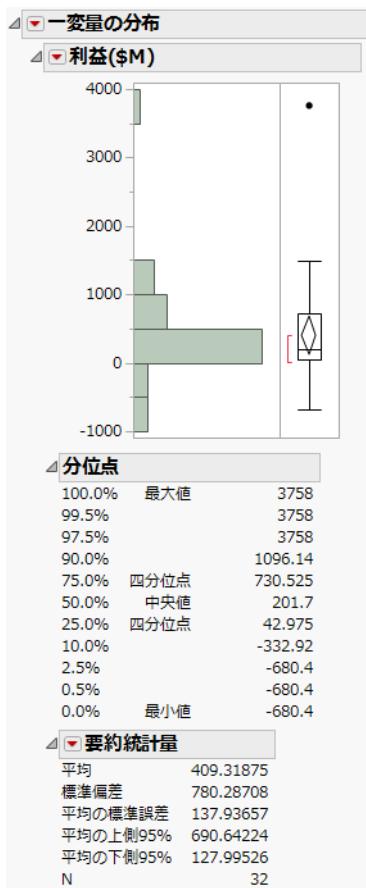
プラットフォームのオプションは、各レポートウィンドウでオンとオフを切り替えることができます。しかし、各レポートウィンドウでオプションを適用した場合、同じプラットフォームをもう一度呼び出した時に設定は元に戻っています。特定のオプションを常に適用したい場合は、これらのオプションを「環境設定」ウィンドウで設定します。

この例では、「一変量の分布」プラットフォームで最初に表示されるレポートに、外れ値の箱ひげ図を追加しないように設定します。

デフォルトの環境設定を使用して「一変量の分布」を作成する

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] を選択します。
3. 「利益(\$M)」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8.7 「一変量の分布」レポートウィンドウ

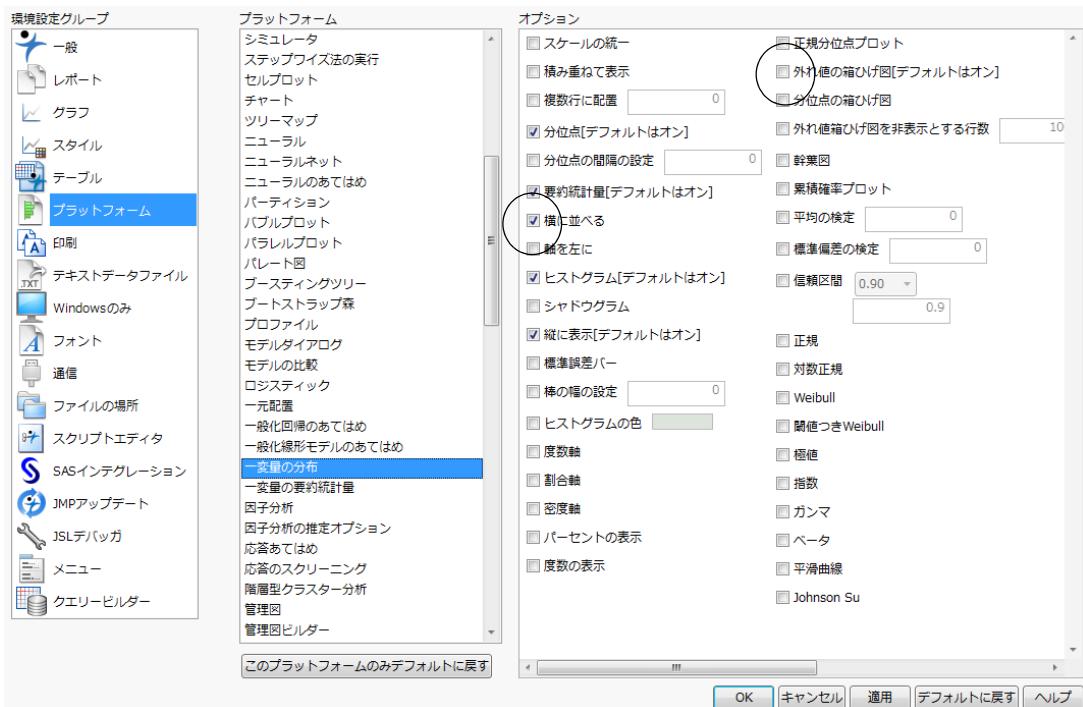


ヒストグラムは縦方向に表示され、グラフには外れ値の箱ひげ図が表示されています。ヒストグラムを横に並べて表示し、外れ値の箱ひげ図を表示しないように変更する場合は、「利益(\$M)」の赤い三角ボタンをクリックし、メニューから適切なオプションを選択します。しかし、これらの設定を常に適用したい場合は、「環境設定」ウィンドウで変更します。

外れ値の箱ひげ図の環境設定を変更して「一変量の分布」を再実行する

1. [ファイル] > [環境設定] を選択します。
2. 環境設定グループから「プラットフォーム」を選択します。
3. 「プラットフォーム」リストから「一変量の分布」を選択します。
4. [横に並べる] オプションを選択してオンにします。
5. [外れ値の箱ひげ図] オプションの選択を解除してオフにします。

図8.8 一変量の分布の環境設定



6. [OK] をクリックします。
7. 「一変量の分布」をもう一度、実行します。「デフォルトの環境設定を使用して「一変量の分布」を作成する」(193ページ)を参照してください。

ヒストグラムが横に並べて表示され、外れ値の箱ひげ図は表示されなくなりました。これらの環境設定は変更しない限りそのまま維持されます。

すべての環境設定の詳細については、『JMPの使用法』の「JMPの環境設定」の章を参照してください。

JMPとSASの連携

注: JMP 経由で SAS を使用するには、ローカルコンピュータまたはサーバーのいずれかの SAS に接続できる状態でなければいけません。

JMP から SAS に接続し、次のような作業を実行できます。

- JMP で SAS コードを記述または作成する
- SAS コードをサブミットし、結果を JMP で表示する
- リモートコンピュータ上の SAS メタデータサーバーまたは SAS サーバーに接続する
- ローカルコンピュータ上の SAS に接続する
- SAS データセットを開いて参照する
- SAS で生成されたデータセットを取得および表示する

JMP と SAS との連携についての詳細は、『JMP の使用法』の「データの読み込み」の章を参照してください。

SAS コードの作成例

この例では、キャンディバーの栄養価データを記録したサンプルデータ「Candy Bars.jmp」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、「Candy Bars.jmp」を開きます。
2. [分析] > [モデルのあてはめ] を選択します。
3. 「カロリー」を選択し、[Y] をクリックします。
4. 「総脂肪(g)」、「炭水化物(g)」、および「タンパク質(g)」を選択し、[追加] をクリックします。
5. 「モデルの指定」の赤い三角ボタンをクリックし、メニューから [SASジョブの作成] を選択します。

図8.9にSASコードを示します（データの一部は省略しています）。

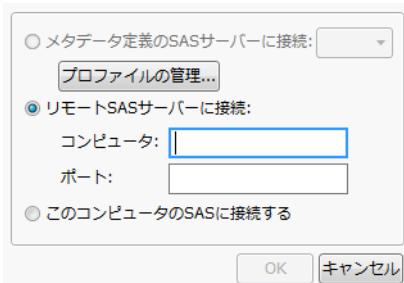
図8.9 SASコード

```
DATA Candy_Bars; INPUT カロリー 総脂肪(g) 炭水化物(g) タンパク質(g); Lines;
310 20 28 8
230 12 27 4
220 15 24 4
170 8 21 3
200 9 25 4
260 16 26 5
190 15 42 2
190 11 21 2
230 12 28 3
230 13 25 2
;
RUN;
PROC GLM DATA=Candy_Bars ALPHA=0.05;
MODEL カロリー =総脂肪(g) 炭水化物(g) タンパク質(g);
RUN;
```

SAS コードのサブミット例

1. [ヘルプ] > [サンプルデータライブラリ] を選択し、「Candy Bars.jmp」を開きます。
2. [分析] > [モデルのあてはめ] を選択します。
3. 「カロリー」を選択し、[Y] をクリックします。
4. 「総脂肪(g)」、「炭水化物(g)」、および「タンパク質(g)」を選択し、[追加] をクリックします。
5. 「モデルの指定」の赤い三角ボタンをクリックし、メニューから [SAS でサブミット] を選択します。
6. 「SAS サーバーに接続」ウィンドウ（図 8.10 を参照）で、SASへの接続方法を選択します（接続していない場合）。この例では、[このコンピュータの SAS に接続する] を選択します。

図 8.10 SAS サーバーに接続



7. [OK] をクリックします。

JMPがSASに接続します。SASでモデルが実行され、結果がJMPに戻されます。SASアウトプット、HTML、RTF、PDF、JMPレポートのいずれかの形式で結果を表示できます（出力形式はJMPの環境設定を使用して選択できます）。図8.11は、JMPレポート形式で出力された結果です。詳細については、『JMPの使用法』の「データの読み込み」の章を参照してください。

図 8.11 JMP レポート形式で出力された SAS 結果

SAS システム
GLM プロシージャ

GLM プロシージャ

データ

オブザベーション数

読み込んだオブザベーション数 75
使用されたオブザベーション数 75

従属変数: _3

分散分析

_3

ANOVA

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	3	282358	94119.3	3237.58	<.0001*
Error	71	2064.03	29.0709	.	.
Corrected Total	74	284422	.	.	.

適合度統計量

R2 乗 変動係数 Root MSE _3 の平均
0.99274 2.21858 5.39174 243.027

Type I モデル ANOVA

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
g	1	185260	185260	6372.68	<.0001*
_g_4	1	93540.4	93540.4	3217.67	<.0001*
_g_7	1	3557.86	3557.86	122.386	<.0001*

Type III モデル ANOVA

要因	自由度	III 平方和	平均平方	F 値	Pr > F
g	1	111777	111777	3844.97	<.0001*
_g_4	1	96756.1	96756.1	3328.28	<.0001*
_g_7	1	3557.86	3557.86	122.386	<.0001*

解

パラメータ	推定値	標準誤差	t 値	Pr > t
Intercept	-5.9643	2.89999	-2.0567	0.0434*
g	8.98995	0.14498	62.0078	<.0001*
_g_4	4.0975	0.07102	57.6913	<.0001*
_g_7	4.40133	0.39785	11.0628	<.0001*

索引

はじめてのJMP

A

Adobe Flash

バブルプロットの作成 [180](#)

プロファイルの結果の作成 [180](#)

Anscombe, F.J. [124](#)

H

HTML 5、出力の保存 [177](#)

J

JMP [47](#)

Microsoft Excelとの比較 [58](#)

SASコードの作成 [196](#)

SASコードのサブミット [197](#)

SASとの連携 [196](#)

起動、JMP [49](#)

使い方入門 [48–58](#)

データのコピーと貼り付け [62](#)

読み込み、データ [62](#)

JMPスクリプト言語 [176](#)

JSL (JMPスクリプト言語) [176](#)

M

Microsoft Excel

JMPとの比較 [58](#)

読み込み、データ [63](#)

P

PowerPoint、レポートの保存 [179](#)

S

SAS

JMPとの連携 [196](#)

コードの作成 [196](#)

コードのサブミット [197](#)

T

t検定

「一変量の分布」プラットフォーム [132](#)

二標本 [143](#)

「二変量の関係」プラットフォーム [143](#)

ア

アイコン、尺度 [66](#)

赤い三角ボタンで表示するメニュー [48, 55](#)

値

検索 [72](#)

数値の表示形式 [74](#)

単一値の変更 [69](#)

複数値の変更 [69](#)

値の検索 [72](#)

アドインビルダー [185](#)

イ

「一変量の分布」プラットフォーム

t検定 [132](#)

直線のあてはめレポート [139](#)

箱ひげ図 [132](#)

ヒストグラム [132](#)

分位点レポート [132](#)

要約統計量レポート [132](#)

例、カテゴリカル変数 [133–135](#)

一変量の分布プラットフォーム [130](#)

横に並べる、レポートウィンドウ [131](#)

例、連続変数 [131](#)

インタラクティブHTML、レポートの保存 [177](#)

ウ

ウィンドウ

JMPスタートーウィンドウ [49](#)使い方ヒントウィンドウ [49](#)**オ**折れ線グラフ、箱ひげ図からの変更 [114](#)**力**

回帰

1つの予測変数 [136](#)

回帰分析

2つの予測変数 [150](#)重回帰 [150–151](#)階層型クラスター分析 [165](#)開閉ボタン [48, 55](#)重ね合わせプロット [105–106](#)解釈 [107](#)グリッド線の追加 [107](#)操作 [107](#)点をつなぐ [107](#)凡例 [108](#)複数の変数 [108](#)例 [105](#)

カテゴロカル変数

関係 [136](#)平均の比較 [141](#)例 [133–135](#)割合の比較 [144](#)環境設定、変更 [192](#)**キ**軌跡 バブル [121](#)起動、JMP [49](#)起動ウィンドウ [48](#)

行

行の属性をクリア [144](#)除外する/除外しない [97](#)選択または選択解除 [71](#)追加 [65](#)デフォルトの状態に戻す [144](#)行の属性データタイプ [74](#)行パネル [53](#)**ク**クラスター分析 [165](#)

グラフ

1つの変数 [90](#)

「グラフビルダー」も参照

個別のチャートの作成 [115](#)作成 [112](#)自動更新 [188–191](#)複数の変数 [94](#)変動性図 [108](#)棒グラフ [92–94](#)グラフの作成、データ [124](#)グラフビルダー [111](#)起動 [147](#)グラフの解釈 [116](#)個別のチャートの作成 [115](#)作成、グラフ [112](#)箱ひげ図から折れ線グラフへの変更 [114](#)例 [112](#)グリッド線、重ね合わせプロットへの追加 [107](#)**ケ**計算式エディタ [75](#)結果、プラットフォームの保存 [172](#)結合、データテーブル [85](#)検定、差 [133](#)検定レポート [146](#)**コ**

コード、SAS

作成 [196](#)サブミット [197](#)効果の検定レポート [149](#)交互作用 [148](#)**サ**差、検定 [133](#)サブセットの作成 [83](#)散布図 [95–96](#)解釈 [97](#)

操作 98
例 95
散布図行列 99, 163
解釈 101
作成 100
操作 102
例 100
「散布図行列」プラットフォーム
起動 151
例 151
サンプルデータテーブル 51

シ

自動再計算 188–191
ジャーナル
作成 172
プラットフォーム結果の保存 172
分析の追加 173
ジャーナルへの結果の保存 172
尺度
アイコン 66
「カテゴリカル変数」も参照
順序 74, 127
変更 129
名義 74, 127
例 127
連続尺度 127
連続変数 74
重回帰分析 150–151
順序尺度 74, 127
「カテゴリカル変数」も参照
除外、データ 97
初期値を埋める/挿入オプション 70

ス

数値、表示形式 74
数値データタイプ 74
スクリプト
JSL 176
作成 175
実行 175

セ
セル、選択または選択解除 72

ソ
相関関係 95

タ
ダッシュボード、作成 183
ダッシュボードでのレポートの結合 183
多変量解析 162

チ
直線
あてはめ 137
平均線の追加 143
直線のあてはめ 137
直線のあてはめレポート 139

ツ
データテーブル
追加
プロジェクトへ 174
使い方ヒントウィンドウ 49

テ
データ
JMPへの移動 62
値の検索 72
グラフの作成 124
コピーと貼り付け 62
除外しない 97
選択
行 71
セル 72
データ点 99
列 71
選択解除
行 71
セル 72
列 71
データ点を除外しない 97

データ点を除外する 97
 パターンの作成 70
 フィルタ処理 77
 分析 123
 平均の検索 83
 変更
 尺度 129
 単一値 69
 複数値 69
 読み込み 62–63
 データグリッド 52
 データタイプ 74
 データテーブル 48
 行パネル 53
 結合 85
 作成 65
 サブセットの作成 83
 サンプルを開く 51
 も参照「尺度」 66
 数値の表示形式 74
 追加
 行 65
 列 66
 データグリッド 52
 データタイプ 74
 データの入力 65
 テーブルパネル 52
 名前の変更 66
 開く 62
 例 61
 列パネル 52
 列プロパティメニュー 75
 列を基準に並べ替え 87
 データ点
 除外しない 97
 除外する 97
 選択 99
 データのコピーと貼り付け 62
 データの再コード化 69
 データのフィルタ処理 77
 データの分析 123
 データフィルタ 77
 テーブルパネル 52

ト
 統計量、要約 79
 ナ
 なげなわツール 99
 名前
 データテーブル 66
 列 74
 ニ
 二変量の関係
 変数の関係の分析 136
 「二変量の関係」プラットフォーム
 検定レポート 146
 モザイク図 146
 t検定 143
 起動 141
 分割表 145
 平均と標準偏差レポート 143
 ハ
 箱ひげ図 102
 作成 103
 「一変量の分布」プラットフォーム 132
 折れ線グラフへの変更 114
 解釈 104
 外れ値の箱ひげ図 56
 例 103
 外れ値 92, 104
 影響 142, 147
 削除 97
 除外 139
 例 137, 139
 外れ値の箱ひげ図 56
 パターンデータ、作成 70
 バブルプロット 117
 Adobe Flashバージョンの作成 180
 オプション 121
 解釈 120
 軌跡 バブル 121
 作成 118
 操作 121

- 例 118
ばらつき 103
パラメータ推定値レポート 153
- ヒ**
ヒストグラム 90
「一変量の分布」プラットフォームと 132
解釈 91
作成 91
操作 92, 135
例 90
標準偏差 143
- フ**
プラットフォーム 48, 53
起動 54
起動ウィンドウ 48
結果の保存 172
結果を表示する 54
例 54–58
レポートウィンドウ 48, 55, 57
プロジェクト
作成 173
追加
結果 174
データテーブル 174
- プロット
特定の種類のプロットを参照
プロファイルプラットフォーム、Adobe Flash の結果
の作成 180
分位点レポート 132
分割表 145
分析
自動更新 188–191
ジャーナルへの追加 173
二変量の関係 136
- ヘ**
平均、比較
1つの変数 141
複数の変数 146
平均線 143
- 平均と標準偏差レポート 143
並列箱ひげ図
「箱ひげ図」を参照
変数
1つの変数の平均の比較 141
カテゴリカル 133, 136, 141, 144
関係の種類 136
関係の分析 136
尺度 127
単一の変数を扱うグラフ 90
複数の変数による平均の比較 146
複数の変数を扱う重ね合わせプロット 108
複数の変数を扱うグラフ 94
予測変数 136
連続 136
連続変数 127, 130, 136–137, 141
変動性図 108
解釈 111
グループ変数の順序 109
作成 109
非表示 110
例 109
- ホ**
ホームウィンドウ 49
棒グラフ 92
解釈 93
作成 93
操作 94
例 92–94
- メ**
名義尺度 74
「カテゴリカル変数」も参照
- モ**
モザイク図 146
文字データタイプ 74
「モデルのあてはめ」プラットフォーム
起動 151
効果の検定レポート 149
パラメータ推定値レポート 153

変数の関係の分析 136
 予測値と実測値のプロット 152
 予測プロファイル 154

☰

要約統計量 79
 要約統計量レポート 132
 横に並べる、レポートウィンドウ 131
 予測値と実測値のプロット 152
 予測プロファイル 154
 予測変数 136
 読み込み、データ 62

レ

例、分析 123
 列
 選択または選択解除 71
 追加 66
 名前の指定 74
 ロック 75
 列情報ウィンドウ 74
 列パネル 52
 列プロパティメニュー 75
 列名 74
 一意 74
 入力 74
 変更 74
 列を基準にデータテーブルを並べ替え 87

レポート
 検定 146
 効果の検定 149
 直線のあてはめ 139
 パラメータ推定値 153
 分位点 132
 平均と標準偏差 143
 要約統計量 132
 レポートウィンドウ 55
 赤い三角ボタンで表示するメニュー 48, 55
 開閉 57
 開閉ボタン 48, 55
 説明 48
 操作 57

プラットフォームオプション 48
 横に並べる 131
 レポート 48
 連続尺度 74, 127
 連続変数 127
 1つの予測変数による回帰分析 136
 関係 136
 尺度 127
 直線のあてはめ 137
 分析の例 131
 平均の比較 141

□

ロック、列 75

ワ

割合、比較 144