



バージョン 18

## 基本的な統計分析

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

マルセル・ブルースト

JMP Statistical Discovery LLC  
920 SAS Campus Drive  
Cary, North Carolina 27513-2414

The correct bibliographic citation for this manual is as follows: JMP Statistical Discovery LLC 2024. *JMP® 18 Basic Analysis*. Cary, NC: JMP Statistical Discovery LLC

## **JMP® 18 Basic Analysis**

Copyright © 2024, JMP Statistical Discovery LLC, Cary, NC, USA

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

JMP Statistical Discovery LLC, 920 SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2024

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

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

JMP software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For more information about third-party software distributed with JMP software, refer to <https://www.jmp.com/thirdpartysoftware>.

## **JMP を最大限に活用する**

JMPの初心者でも経験豊かなユーザでも、JMPについての新しい情報をいろいろなリソースから得られます。

JMP.com のサイトでは以下のリソースをご利用いただけます。

- JMP を使い始めるためのライブおよび収録済みの Webcast セミナー
- 新機能や高度なテクニックを紹介している動画デモや Webcast セミナー
- JMP トレーニングの申し込みに関する詳細
- 各地で開催されるセミナーのスケジュール
- お客様の JMP 利用事例
- JMP User Community、アドインやスクリプトの例を始めとするユーザ向けリソース、フォーラム、ブログ、イベント情報など

<https://www.jmp.com/getstarted>





# 目次

## 基本的な統計分析

---

<b>1</b>	<b>JMP の概要</b>	15
	マニュアルとその他のリソース	
	JMP Pro	17
	JMP オンラインヘルプ	17
	ドキュメンテーションPDF アドイン	17
	JMP ヘルプメニュー	25
	JMP を習得するためのその他のリソース	26
	JMP を使い始める	26
	JMP の検索	26
	サンプルデータテーブル	27
	JSL の概要	27
	JMP のツールヒント	27
	JMP User Community	28
	オンラインの統計的思考コース（無料）	28
	JMP New User Welcome Kit	28
	統計ナレッジポータル	28
	JMP トレーニング	28
	JMP 関連書籍	29
	「JMP スターター」ウィンドウ	29
	JMP テクニカルサポート	29
<b>2</b>	<b>基本的な統計分析について</b>	31
	基本的な分析手法の概要	
<b>3</b>	<b>一変量の分布</b>	33
	「一変量の分布」プラットフォームを使用する	
	「一変量の分布」プラットフォームの概要	36
	「一変量の分布」プラットフォームの例	36
	「一変量の分布」プラットフォームの起動	38
	「一変量の分布」レポート	40
	ヒストグラム	41
	「度数」レポート	45
	「分位点」レポート	46

「要約統計量」レポート	46
「一変量の分布」プラットフォームのオプション	49
カテゴリカル変数のオプション	50
カテゴリカル変数の [ヒストグラムオプション]	51
カテゴリカル変数の [保存] オプション	51
連続変数のオプション	52
連続変数の [表示オプション]	53
連続変数の [ヒストグラムオプション]	53
正規分位点プロット	55
外れ値の箱ひげ図	55
分位点の箱ひげ図	56
幹葉図	58
累積確率プロット	58
平均の検定	59
標準偏差の検定	60
同等性の検定	60
信頼区間	61
予測区間	62
許容区間	62
工程能力	63
分布のあてはめ	65
連続変数の [保存] オプション	70
「一変量の分布」プラットフォームの別例	72
複数のヒストグラムにおけるデータ選択の例	72
By 変数の例	73
2水準の割合の検定例	74
3水準以上の割合の検定例	76
予測区間の例	78
許容区間の例	80
工程能力分析の例	80
検出限界の例	82
「一変量の分布」プラットフォームの統計的詳細	85
標準誤差バーの統計的詳細	85
分位点の統計的詳細	85
要約統計量の統計的詳細	86
正規分位点プロットの統計的詳細	87
Wilcoxon の符号付順位検定の統計的詳細	88
標準偏差の検定の統計的詳細	90
正規分位点の統計的詳細	90
[標準化されたデータの保存] の統計的詳細	90

スコア信頼区間の統計的詳細 .....	91
予測区間の統計的詳細 .....	91
許容区間の統計的詳細 .....	92
連続分布のあてはめの統計的詳細 .....	94
離散分布のあてはめの統計的詳細 .....	101
<b>4 「二変量の関係」 プラットフォームの概要</b> .....	<b>107</b>
<b>2変数の関係を調べる</b>	
<b>5 二変量分析</b> .....	<b>109</b>
<b>2つの連続変数の関係を調べる</b>	
「二変量」プラットフォームの概要 .....	111
二変量分析の例 .....	111
「二変量」プラットフォームの起動 .....	113
データ形式 .....	114
「二変量の関係」のレポート .....	114
「二変量」プラットフォームのオプション .....	115
二変量の関係のオプション .....	118
「二変量の関係」レポート .....	121
「平均のあてはめ」レポート .....	122
「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」レポート .....	123
「その他のあてはめ」のウィンドウ .....	128
「スプライン曲線のあてはめ」レポート .....	129
「局所的な平滑化」レポート .....	129
「値ごとのあてはめ」レポート .....	130
「直交のあてはめ」レポート .....	130
「ロバストなあてはめ」レポート .....	131
「Passing-Bablokのあてはめ」レポート .....	132
「確率楕円」レポート .....	133
ノンパラメトリック密度 .....	135
「二変量」プラットフォームの別例 .....	136
[その他のあてはめ] オプションの例 .....	136
[直交のあてはめ] オプションの例 .....	138
[ロバストなあてはめ] オプションの例 .....	139
グループ別確率楕円の例 .....	141
グループ別回帰直線の例 .....	142
By 変数を使ってグループ別に分析する例 .....	143
「二変量」プラットフォームの統計的詳細 .....	144
[直線のあてはめ] オプションの統計的詳細 .....	144
[多項式のあてはめ] オプションの統計的詳細 .....	145

「スプライン曲線のあてはめ」オプションの統計的詳細	145
「直交のあてはめ」オプションの統計的詳細	146
「ロバスト」オプションの統計的詳細	147
「Passing Bablokのあてはめ」オプションの統計的詳細	148
「あてはめの要約」レポートの統計的詳細	148
「あてはまりの悪さ (LOF)」レポートの統計的詳細	149
「パラメータ推定値」レポートの統計的詳細	149
「平滑化スプライン曲線によるあてはめ」レポートの統計的詳細	150
「二変量正規楕円」レポートの統計的詳細	150

## 6 一元配置分析 151

### 連続尺度のY変数とカテゴリカルなX変数の関係を調べる

「一元配置」プラットフォームの概要	154
一元配置分析の例	154
「一元配置」プラットフォームの起動	157
データ形式	158
「一元配置」レポート	158
「一元配置」プラットフォームのオプション	159
「一元配置分析」レポート	168
「分位点」レポート	169
「平均/ANOVA/プーリングしたt検定」レポート	169
「平均と標準偏差」レポート	171
「t検定」レポート	172
平均分析レポート	173
平均の比較レポート	174
ノンパラメトリックな検定のレポート	177
ノンパラメトリックな多重比較のレポート	181
等分散性の検定レポート	183
同等性検定のレポート	185
ロバストなあてはめのレポート	188
検出力のレポート	189
「対応のあるあてはめ」レポート	190
一元配置プロットの要素	191
平均のひし形と標本サイズに比例したX軸	191
平均線・平均誤差バー・標準偏差線	192
比較円	193
「一元配置」プラットフォームの別例	195
平均分析の例	195
分散の平均分析の例	196
「各ペア, Studentのt検定」の例	197



[すべてのペア,TukeyのHSD検定]の例	199
[最適値との比較,HsuのMCB検定]の例	201
[コントロール群との比較,Dunnett検定]の例	202
[各ペアのステップワイズ,Newman-Keuls検定]の例	204
「平均の比較」で用意されている4種類の検定	204
Wilcoxon検定の例	205
等分散性検定の例	207
同等性の検定の例	208
[ロバストなあてはめ] オプションの例	210
[検出力] オプションの例	212
正規分位点プロットの例	213
累積確率プロットの例	214
[密度]の各オプションの例	215
[対応のある列を設定] オプションの例	216
データを積み重ねて一元配置分析を実行する例	218
「一元配置」プラットフォームの統計的詳細	224
比較円の統計的詳細	224
検出力の統計的詳細	226
平均分析の統計的詳細	226
「あてはめの要約」レポートの統計的詳細	227
等分散性検定の統計的詳細	227
ノンパラメトリックな検定の統計量の統計的詳細	228
[ロバストなあてはめ]の統計的詳細	231

## 7 分割表分析 ..... 233

### 2つのカテゴリカルな変数の関係を調べる

分割表分析の例	235
「分割表」プラットフォームの起動	236
データ形式	237
「分割表」レポート	237
モザイク図	239
分割表	241
「検定」レポート	242
「分割表」プラットフォームのオプション	243
「分割表分析」レポート	246
「割合の平均分析」レポート	246
「対応分析」レポート	247
Cochran-Mantel-Haenszel検定	248
[一致性の統計量] オプション	249
「相対リスク」レポート	249

割合の2標本検定	250
「関連の指標」レポート	250
Cochran-Armitageの傾向検定	252
「Fisherの正確検定」レポート	252
同等性検定のレポート	252
「分割表」プラットフォームの別例	254
割合の平均分析の例	255
対応分析の例	256
Cochran-Mantel-Haenszel 検定の例	257
「一致性の統計量」オプションの例	259
「相対リスク」オプションの例	260
割合の2標本検定の例	261
「関連の指標」オプションの例	262
Cochran-Armitageの傾向検定の例	263
相対リスクの同等性検定の例	264
「分割表」プラットフォームの統計的詳細	265
一致性の統計量の統計的詳細	265
オッズ比の統計的詳細	266
「検定」レポートの統計的詳細	266
対応分析の統計的詳細	267

## 8 ロジスティック分析 ..... 269

### カテゴリカルなY変数と連続尺度のX変数の関係を調べる

「ロジスティック」プラットフォームの概要	271
名義ロジスティック回帰の例	271
「ロジスティック」プラットフォームの起動	273
データ形式	274
「ロジスティック」レポート	274
ロジスティックプロット	275
「反復履歴」レポート	276
「モデル全体の検定」レポート	276
「あてはめの詳細」レポート	277
「パラメータ推定値」レポート	278
「ロジスティック」プラットフォームのオプション	278
ロジスティック分析レポート	280
ROC曲線	280
逆推定	281
ロジスティック回帰の別例	281
順序ロジスティック回帰の例	281
ロジスティックプロットの例	283

ROC 曲線の例	285
逆推定の例	286
「ロジスティック」プラットフォームの統計的詳細	289
<b>9 表の作成</b>	<b>291</b>
対話的に集計表を作成する	
「表の作成」プラットフォームの使用例	293
「表の作成」プラットフォームの起動	298
「表の作成」プラットフォームのダイアログウィンドウ	300
統計量の追加	301
「表の作成」プラットフォームのレポート	304
分析列	305
グループ列	305
列テーブルと行テーブル	306
表の編集	307
「表の作成」プラットフォームのオプション	307
テスト集計パネルの表示	308
列のポップアップメニュー	309
「表の作成」プラットフォームの別例	310
さまざまな表を作成して配置を変更する例	310
複数のデータ列に対する表の例	315
ページ列の使用例	317
グループ列を積み重ねる例	319
一意の ID 列を使用した例	320
グループ列をパッキングする例	322
<b>10 シミュレーション</b>	<b>325</b>
パラメトリックなモデルの乱数シミュレーション	
シミュレーション機能の概要	327
シミュレーション機能の例	327
シミュレーション機能の起動	333
「シミュレーション」ウィンドウ	333
シミュレーション結果のテーブル	334
「シミュレーション 結果」レポート	335
「検出力のシミュレーション」レポート	335
シミュレーション機能の別例	336
並び替え検定の例	336
一般化回帰モデルに含める説明変数の検討例	338
検出力の事前計算の例	343

<b>11 ブートストラップ</b>	353
<b>標本再抽出によって統計量の分布を近似する</b>	
ブートストラップの概要	355
ブートストラップを利用できる JMP プラットフォーム	356
ブートストラップの例	357
「ブートストラップ」ウィンドウのオプション	359
ブートストラップ結果テーブル（積み重ね）	360
結果を積み重ねていないデータテーブル	361
ブートストラップで生成された統計量の分析	362
ブートストラップの別例	363
ブートストラップの統計的詳細	368
「小数の重み」の統計的詳細	368
バイアス修正済みの信頼限界の統計的詳細	368
<b>12 テキストエクスプローラ</b>	371
<b>非構造化テキストを探索する</b>	
「テキストエクスプローラ」プラットフォームの概要	373
テキスト処理の手順	375
「テキストエクスプローラ」プラットフォームの例	376
「テキストエクスプローラ」プラットフォームの起動	380
正規表現エディタでの Regex のカスタマイズ	382
「テキストエクスプローラ」レポート	387
要約レポート	388
単語と句のリスト	388
「テキストエクスプローラ」プラットフォームのオプション	392
テキストを前処理するオプション	392
テキストの統計分析に関するオプション	398
保存に関するオプション	400
テキストエクスプローラのレポートオプション	401
潜在クラス分析	402
潜在意味分析（特異値分解）	403
特異値分解レポート	404
特異値分解レポートのオプション	405
トピックの分析	407
「トピック分析」レポート	408
「トピック分析」レポートのオプション	408
線形判別分析	409
「判別分析」レポート	410
「判別分析」レポートのオプション	410
単語選択分析	411

単語選択分析の設定 .....	411
「単語選択分析」レポート .....	412
「単語選択分析」レポートのオプション .....	414
感情分析 .....	415
「感情分析」レポート .....	416
「感情分析」レポートのオプション .....	418
「テキストエクスプローラ」プラットフォームの別例 .....	420
<b>A 参考文献</b> .....	425



# 第 1 章

## JMP の概要 マニュアルとその他のリソース

---

「JMP の概要」マニュアル。JMP Pro 限定機能、JMP ドキュメンテーションアドイン、各 JMP ドキュメントの説明、ヘルプメニューのオプション、その他のサポートの場所などを紹介しています。

## 目次

JMP Pro .....	17
JMP オンラインヘルプ .....	17
ドキュメンテーションPDF アドイン .....	17
JMP ヘルプメニュー .....	25
JMP を習得するためのその他のリソース .....	26
JMP を使い始める .....	26
JMP の検索 .....	26
サンプルデータテーブル .....	27
JSL の概要 .....	27
JMP のツールヒント .....	27
JMP User Community .....	28
オンラインの統計的思考コース（無料） .....	28
JMP New User Welcome Kit .....	28
統計ナレッジポータル .....	28
JMP トレーニング .....	28
JMP 関連書籍 .....	29
「JMP スターター」 ウィンドウ .....	29
JMP テクニカルサポート .....	29



---

## JMP Pro

JMP Proの限定機能には、JMP Proアイコンが付いています。JMP Proの機能の概要については<https://www.jmp.com/software/pro/>をご覧ください。

---

## JMP オンラインヘルプ

JMP オンラインヘルプでは、JMPの機能や統計手法、JMP スクリプト言語（JSL）に関する情報が検索できます。JMP オンラインヘルプは、次のような方法で開きます。

- Windowsで、[ヘルプ] > [JMP オンラインヘルプ] を選択します。
- macOSで、[ヘルプ] > [JMP ヘルプ] を選択します。
- WindowsでF1キーを押します。
- データテーブルやレポートウィンドウにおける特定の部分に関するヘルプを表示するには、[ヘルプ] > [ヘルプツール] を選択します。その後、データテーブルやレポートウィンドウの任意の箇所をクリックします。ヘルプツールを無効にするには、Escキーを押します。
- JMP ウィンドウ内で [ヘルプ] ボタンをクリックします。

---

**メモ:** JMP ヘルプを使用するにはインターネット接続が必要です。インターネット接続がない場合は、ドキュメンテーションアドインをインストールする方法があります。詳細については、「[ドキュメンテーションPDF アドイン](#)」を参照してください。

---

---

## ドキュメンテーションPDF アドイン

JMP ドキュメンテーションアドインをダウンロードし、インストールすることができます。このドキュメンテーションアドインには、JMP ライブラリの個々のドキュメントをPDFファイルとして保存したものと、『JMP ドキュメンテーションライブラリ』ファイルが含まれています。『JMP ドキュメンテーションライブラリ』ファイルは、個々のドキュメントのPDFファイルを1つにまとめたPDFファイルです。JMP オンラインヘルプのように、1つのPDFファイルで全マニュアル内を検索することができます。

このアドインをインストールすると、PDFファイルがコンピュータにインストールされ、[ヘルプ] メニューに [ドキュメンテーションPDF] オプションが表示されます。[ヘルプ] > [ドキュメンテーションPDF] を選択して、ローカルコンピュータ上のドキュメンテーションを使用することが可能になります。ドキュメンテーションアドインは、<https://www.jmp.com/doc-addin>でダウンロードしてください。

以下の表は、ドキュメンテーションアドインに含まれている各マニュアルの目的および内容を要約したものです。

マニュアル	目的	内容
『JMP ドキュメンテーションライブラリ』	個々のマニュアルの PDF ファイルを1つにまとめた PDF ファイルです。	1つの PDF にすべてのJMP ドキュメンテーションが含まれています。
『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を紹介し、データを作成・分析する方法や、結果を共有する方法を解説します。
『JMP の使用法』	JMP のデータテーブルと、基本操作を理解する	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、ワークフロービルダーの使い方など、JMP 全体で使用する一般的な概念や機能を説明しています。
『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	<div>[分析] メニューにある以下のプラットフォームについて説明しています。<ul style="list-style-type: none"><li>一変量の分布</li><li>二変量の関係</li><li>表の作成</li><li>テキストエクスプローラ</li></ul></div> <div>[分析] &gt; [二変量の関係] で二変量分析、一元配置分散分析、分割表分析を実行する方法を説明しています。ブートストラップによる標本分布の近似方法や、乱数シミュレーションによるパラメトリックな標本再抽出の実行方法も取り上げています。</div>

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] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"> <li>• グラフビルダー</li> <li>• 三次元散布図</li> <li>• 等高線図</li> <li>• バブルプロット</li> <li>• パラレルプロット</li> <li>• セルプロット</li> <li>• 散布図行列</li> <li>• 三角図</li> <li>• ツリーマップ</li> <li>• チャート</li> <li>• 重ね合わせプロット</li> </ul> <p>このマニュアルには背景地図や独自の地図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] メニューで用意されているすべてのプロファイルについて。誤差因子の分析が、ランダムなインプットを使用したシミュレーションの実行とともに含まれています。
『実験計画 (DOE)』	実験の計画方法と適切な標本サイズの決定方法を学ぶ	[実験計画 (DOE)] メニューで用意されているすべての機能について。

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	<div> <div>[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、以下の手法について説明しています。</div> <ul style="list-style-type: none"> <li>標準最小2乗</li> <li>ステップワイズ法</li> <li>一般化回帰</li> <li>混合モデル</li> <li>一般化線形混合モデル</li> <li>MANOVA</li> <li>対数線形-分散</li> <li>名義ロジスティック</li> <li>順序ロジスティック</li> <li>一般化線形モデル</li> </ul> </div>

マニュアル	目的	内容
『予測モデルおよび発展的なモデル』	さらなるモデリング手法について学ぶ	<p>[分析] &gt; [予測モデル] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>ニューラル</li><li>パーティション</li><li>ブートストラップ森</li><li>ブースティングツリー</li><li>K近傍法</li><li>単純Bayes</li><li>サポートベクトルマシン</li><li>モデルの比較</li><li>モデルのスクリーニング</li><li>検証列の作成</li><li>計算式デボ</li></ul> <p>[分析] &gt; [発展的なモデル] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>曲線のあてはめ</li><li>非線形回帰</li><li>関数データエクスプローラ</li><li>Gauss 過程</li><li>時系列分析</li><li>時系列予測</li><li>対応のあるペア</li></ul> <p>[分析] &gt; [スクリーニング] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>外れ値を調べる</li><li>欠測値を調べる</li><li>パターンを調べる</li><li>応答のスクリーニング</li><li>説明変数のスクリーニング</li><li>アソシエーション分析</li><li>工程履歴エクスプローラ</li></ul>

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方法。	<p>[分析] &gt; [多変量] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>• 多変量の相関</li><li>• 主成分分析</li><li>• 判別分析</li><li>• PLS回帰</li><li>• 多重対応分析</li><li>• 構造方程式モデル</li><li>• 因子分析</li><li>• 多次元尺度構成</li><li>• 多変量埋め込み</li><li>• 項目分析</li></ul> <p>[分析] &gt; [クラスター分析] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>• 階層型クラスター分析</li><li>• K Means クラスター分析</li><li>• 正規混合</li><li>• 潜在クラス分析</li><li>• 変数のクラスタリング</li></ul>

マニュアル	目的	内容
『品質と工程』	工程を評価し、向上させるためのツールについて学びます。	<p>[分析] &gt; [品質と工程] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>• 管理図ビルダーと個々の管理図</li><li>• 測定システム分析 (EMP、タイプ1 ゲージ)</li><li>• 計量値/計数値ゲージチャート</li><li>• 工程のスクリーニング</li><li>• 工程能力</li><li>• モデルに基づく多変量管理図</li><li>• 旧機能の管理図</li><li>• パレート図</li><li>• 特性要因図</li><li>• 品質に関する限界の設定</li><li>• OC 曲線</li></ul>
『信頼性/生存時間分析』	製品やシステムにおける信頼性を評価し、向上させる方法、および人や製品の生存時間データを分析する方法について学ぶ	<p>[分析] &gt; [信頼性/生存時間分析] メニューにある以下のプラットフォームについて説明しています。</p> <ul style="list-style-type: none"><li>• 寿命の一変量</li><li>• 寿命の二変量</li><li>• 累積損傷</li><li>• 疲労モデル</li><li>• 再生モデルによる分析</li><li>• 反復測定劣化</li><li>• 破壊劣化</li><li>• 信頼性予測</li><li>• 信頼性成長</li><li>• 信頼性ブロック図</li><li>• 修理可能システムのシミュレーション</li><li>• 生存時間分析</li><li>• 生存時間 (パラメトリック) のあてはめ</li><li>• 劣化分析</li><li>• 比例ハザードのあてはめ</li></ul>

マニュアル	目的	内容
『消費者調査』	消費者の嗜好を調査して製品・サービスの改良につなげる方法について学びます。	[分析] > [消費者調査] メニューにある以下のプラットフォームについて説明しています。 <ul style="list-style-type: none"><li>• カテゴリカル</li><li>• 選択モデル</li><li>• MaxDiff</li><li>• アップリフト</li><li>• 多重因子分析</li></ul>
『遺伝学』	遺伝子データを分析して育種計画をシミュレートし、最適な遺伝子交雑を予測する方法を学びます。	[分析] > [遺伝学] メニューにある以下のプラットフォームについて説明しています。 <ul style="list-style-type: none"><li>• マーカーの統計量</li><li>• マーカーのシミュレーション</li></ul>
『スクリプトガイド』	パワフルなJMPスクリプト言語 (JSL) について学びます。	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関数の引数やメッセージについて学びます。	JSLコマンドの構文、例、および注意書き。
『キーボードショートカット』	キーボードを使ってJMP内を移動し、タスクを実行する方法を学びます。	コマンドとそれに対応するWindowsとmacOSのショートカットを紹介しています。
『メニューの説明』	JMPのメニュー項目について学びます。	WindowsとmacOSのメニューオプションについて説明しています。



## JMP ヘルプメニュー

JMP 18 では、[ヘルプ] メニューが一新されています。

メニュー項目	説明
JMP の検索	統計検定などの機能を JMP 内で検索する。詳細については、 <a href="#">「JMP の検索」</a> を参照してください。
JMP オンラインヘルプ	Web ブラウザ内に最新バージョンのヘルプが表示される。
ヘルプツール	データテーブルやレポートウィンドウの任意の箇所をクリックしてヘルプを表示する。
クイックスタート	(「使い方ヒント」が「クイックスタート」に生まれ変わりました。) JMP の基本がすばやく学べるように、便利なヒントを表示する。詳細については、 <a href="#">「JMP を使い始める」</a> を参照してください。
ドキュメンテーション PDF	コンピュータ上にあるドキュメンテーション PDF ファイルにアクセスできる。詳細については、 <a href="#">「ドキュメンテーション PDF アドイン」</a> を参照してください。  <b>メモ:</b> このメニューオプションは、ドキュメンテーションアドインをダウンロードしてインストールした場合のみ表示されます。
JMP の機能	Web ブラウザを開いて JMP にあるツールや機能を一覧表示する。オンラインヘルプのリンクも表示され、詳細を確認することができます。
JMP の学習	Web ブラウザを開いて、JMP について学習するためのリソースを表示する。JMP の使い方を説明した短いビデオなどの教材が用意されています。
JMP User Community	Web ブラウザが開き、他の JMP ユーザとつながって問題の解決方法や JMP の改良案などを交換できる。詳細については、 <a href="#">「JMP User Community」</a> を参照してください。
JMP の新機能	Web ブラウザが開き、JMP の最新リリースで導入された新機能を確認できる。
サンプルデータフォルダ	サンプルデータを使って、JMP での分析について学ぶ。サンプルデータ ファイルを開いてスクリプトを実行すると、サンプルの分析を見ることができます。詳細については、 <a href="#">「サンプルデータテーブル」</a> を参照してください。
サンプルの索引	サンプルデータを分析の種類や産業別に検索できる。学習用サンプルなどへのリンクも用意されています。詳細については、 <a href="#">「サンプルデータテーブル」</a> を参照してください。
スクリプトの索引	JMP のスクリプトコマンドを検索し、その使い方を学ぶ。詳細については、 <a href="#">「JSL の概要」</a> を参照してください。

メニュー項目	説明
My JMP	my.jmp.com サイトを開く。
バージョン情報	使用している JMP のバージョンが表示される。ソフトウェアアップデートの有無も確認できます。このオプションは Windows でのみ利用できます。

## JMPを習得するためのその他のリソース

JMP ヘルプに加え、次のリソースも JMP を学習するのに役立ちます。

- [「JMPを使い始める」](#)
- [「JMPの検索」](#)
- [「サンプルデータテーブル」](#)
- [「JSLの概要」](#)
- [「JMPのツールヒント」](#)
- [「JMP User Community」](#)
- [「オンラインの統計的思考コース（無料）」](#)
- [「JMP New User Welcome Kit」](#)
- [「統計ナレッジポータル」](#)
- [「JMPトレーニング」](#)
- [「JMP関連書籍」](#)
- [「「JMPスターター」ウィンドウ」](#)

## JMPを使い始める

JMPを起動すると、最初に JMP の使い方を説明する「クイックスタート」ウィンドウが表示されます。クイックスタートが表示されないようにするには、**[起動時にクイックスタートを表示]** チェックボックスをオフにします。再び表示されるようにするには、**[ヘルプ] > [クイックスタート]** を選択します。また、「環境設定」ウィンドウでも表示するかどうかを設定できます。

## JMPの検索

ある統計手法をどこから実行すればよいかわからない場合は、JMP 全体で検索しましょう。検索を起動したウィンドウ（データテーブル、レポートなど）に合わせて、結果が表示されます。

1. **[ヘルプ] > [JMPの検索]** をクリックします。または、Ctrl キーを押しながらカンマキーを押します。
2. 検索したいテキストを入力します。

3. 探していた統計手法が含まれている結果をクリックします。  
右側にその手法の説明と場所が表示されます。
4. 該当するボタンをクリックして分析を起動するか、ヘルプを開きます。

## 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。サンプルデータのディレクトリを開くには、[ヘルプ] > [サンプルデータフォルダ] 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] > [サンプルの索引] 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\JMP\JMP18\Samples\Data

macOS の場合: /Library/Application Support/JMP/18/Samples/Data

JMP Pro では、サンプルデータが (JMP ではなく) JMP PRO 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見るには、[ヘルプ] > [サンプルの索引] を選択して「Teaching Examples」に移動します。

## JSL の概要

JSL スクリプトに関するヘルプやスクリプトの例を表示するには、[ヘルプ] > [スクリプトの索引] を選択します。[スクリプトの索引] では、JMP スクリプト言語 (JSL) の関数、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。サンプルのスクリプトに編集を加えて実行したり、コマンドに関するヘルプを表示したりできます。

## JMP のツールヒント

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 (または **ホバーラベル**) 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の「環境設定」ウィンドウ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、macOS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をさらに学習したり、他のSASユーザとのコミュニケーションを図ったりできます。ラーニングライブラリには1ページガイド、チュートリアル、デモなどが用意されており、JMPを使い始める上でとても便利です。また、JMPのさまざまなトレーニングコースに登録して、自己教育を進めることも可能です。

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

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

## オンラインの統計的思考コース（無料）

この無料のオンラインコースでは、「探索的データ分析」、「品質手法」、「相関と回帰」といったトピックについて、実用的な統計技術を学習できます。各コースは、短い動画、デモ、練習問題などで構成されています。<https://www.jmp.com/statisticalthinking> をご覧ください。

## JMP New User Welcome Kit

JMP New User Welcome Kitは、JMPの基本的な使用方法をすばやく習得するのに役立ちます。30の短いデモ動画と実習をこなし、ソフトウェアの使い方を身につけ、全世界のJMPユーザが集まった最大規模のオンラインコミュニティに参加できます。<https://www.jmp.com/welcome> をご覧ください。

## 統計ナレッジポータル

統計ナレッジポータルは、わかりやすい例とグラフを使って統計を簡潔に説明し、統計の基礎知識を提供しています。<https://www.jmp.com/skp> をご覧ください。

## JMP トレーニング

JMPでは、経験豊かなJMPのエキスパートを講師とした各種トレーニングを提供しています。パブリックコース、ライブWebセミナー、オンサイトコースをご利用いただけます。また柔軟に学べるeラーニングも用意しています。<https://www.jmp.com/training> をご覧ください。

## JMP 関連書籍

JMP関連書籍は、次のJMP Webページで紹介されています。<https://www.jmp.com/books> をご覧ください。

## 「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」ウィンドウから始めるとよいでしょう。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] メニューで用意されている多くの項目があります。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] (macOSでは [ウィンドウ]) > [JMP スターター] を選びます。
- 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] を選びます。macOSでは、[JMP] > [環境設定] > [全般] > [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。

---

## JMP 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。

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



# 第2章

## 基本的な統計分析について 基本的な分析手法の概要

---

『基本的な統計分析』では、JMPでよく使用される基本的な分析機能について解説します。

- 「一変量の分布」プラットフォームでは、1変数の分布を分析できます。ヒストグラムを表示し、平均や標準偏差などをレポートします。1変数の分析が終わった後は、さらに詳細な分析へと進むでしょう。「[一変量の分布](#)」を参照してください。
- 「二変量の関係」プラットフォームでは、X変数とY変数の尺度によって、自動的に分析が選択されます。「[二変量の関係](#)」プラットフォームの概要」を参照してください。次の4種類の分析を実行できます。
  - 「二変量」プラットフォームでは、2つの連続変数における関係を分析します。「[二変量分析](#)」を参照してください。
  - 「一元配置」プラットフォームでは、X変数のグループごとに、連続尺度のY変数の分布がどのように異なるかを分析します。「[一元配置分析](#)」を参照してください。
  - 「分割表」プラットフォームでは、X変数のグループごとに、カテゴリカルなY変数の分布がどのように異なるかを分析します。「[分割表分析](#)」を参照してください。
  - 「ロジスティック」プラットフォームでは、カテゴリカルな応答変数(Y)の確率を、連続尺度の説明変数(X)にあてはめます。「[ロジスティック分析](#)」を参照してください。
- 「表の作成」プラットフォームでは、記述統計量の表を対話形式で作成できます。「[表の作成](#)」を参照してください。
- シミュレーション機能によって、パラメトリックおよびノンパラメトリックの乱数シミュレーションを柔軟に行えます。「[シミュレーション](#)」を参照してください。
- ブートストラップ分析は、データから何度もサンプリングすることによって統計量の標本分布を近似する手法です。反復ごとに、データは復元抽出（重複抽出）され、抽出されたデータから統計量が計算されます。この計算を反復して、統計量の分布を求めます。「[ブートストラップ](#)」を参照してください。
- 「テキストエクスプローラ」プラットフォームでは、自由回答のテキストデータを集計したり、分析したりできます。分析前における単語や句の取り出しに、正規表現を使用することもできます。「[テキストエクスプローラ](#)」を参照してください。





# 第3章

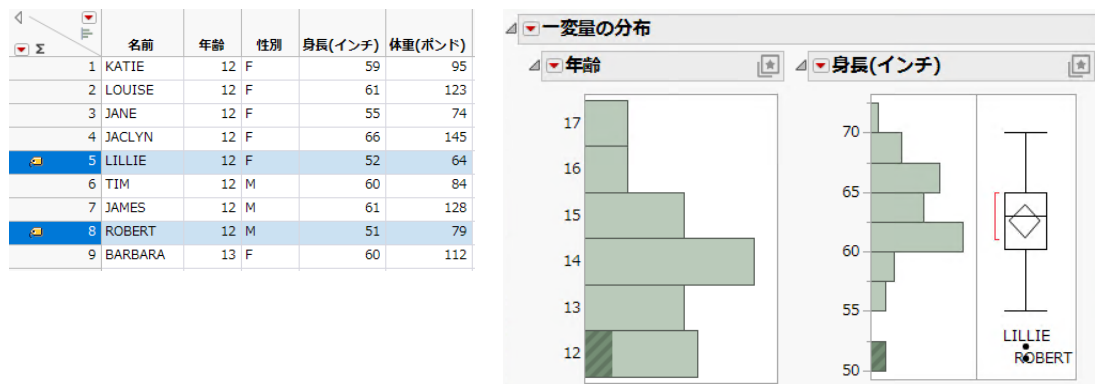
## 一変量の分布

### 「一変量の分布」プラットフォームを使用する

「一変量の分布」プラットフォームでは、ヒストグラム・箱ひげ図・要約統計量などを使って、1変数の分布を調べることができます。 $t$ 検定、 $z$ 検定、カイ2乗検定、同等性の検定などの仮説検定を実行できます。また、信頼区間・許容区間 予測区間を求めたり、工程能力指数を推定したりすることもできます。ここでの「**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、複数の変数についての一変量の分布を1つのレポート内に表示することもできます。各変数に対するレポートの内容は、カテゴリカル変数（名義変数または順序変数）か、連続変数かによって異なります。

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

図3.1 「一変量の分布」プラットフォームの例



## 目次

「一変量の分布」プラットフォームの概要	36
「一変量の分布」プラットフォームの例	36
「一変量の分布」プラットフォームの起動	38
「一変量の分布」レポート	40
ヒストグラム	41
「度数」レポート	45
「分位点」レポート	46
「要約統計量」レポート	46
「一変量の分布」プラットフォームのオプション	49
カテゴリカル変数のオプション	50
カテゴリカル変数の [ヒストグラムオプション]	51
カテゴリカル変数の [保存] オプション	51
連続変数のオプション	52
連続変数の [表示オプション]	53
連続変数の [ヒストグラムオプション]	53
正規分位点プロット	55
外れ値の箱ひげ図	55
分位点の箱ひげ図	56
幹葉図	58
累積確率プロット	58
平均の検定	59
標準偏差の検定	60
同等性の検定	60
信頼区間	61
予測区間	62
許容区間	62
工程能力	63
分布のあてはめ	65
連続変数の [保存] オプション	70
「一変量の分布」プラットフォームの別例	72
複数のヒストグラムにおけるデータ選択の例	72
By 変数の例	73
2水準の割合の検定例	74
3水準以上の割合の検定例	76
予測区間の例	78
許容区間の例	80
工程能力分析の例	80
検出限界の例	82

「一変量の分布」プラットフォームの統計的詳細 .....	85
標準誤差バーの統計的詳細 .....	85
分位点の統計的詳細 .....	85
要約統計量の統計的詳細 .....	86
正規分位点プロットの統計的詳細 .....	87
Wilcoxon の符号付順位検定の統計的詳細 .....	88
標準偏差の検定の統計的詳細 .....	90
正規分位点の統計的詳細 .....	90
[標準化されたデータの保存] の統計的詳細 .....	90
スコア信頼区間の統計的詳細 .....	91
予測区間の統計的詳細 .....	91
許容区間の統計的詳細 .....	92
連続分布のあてはめの統計的詳細 .....	94
離散分布のあてはめの統計的詳細 .....	101

---

## 「一変量の分布」プラットフォームの概要

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

### カテゴリカル変数

カテゴリカル変数の場合、最初に表示されるグラフは棒グラフです。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。モザイク図（分割された棒グラフ）を追加することもできます。

「度数」レポートには度数と割合が表示されます。赤い三角ボタンのメニューを使用して、割合に対する信頼区間や検定を追加することもできます。

### 連続尺度の変数

分析対象の変数が数値で連続尺度の場合は、ヒストグラムと外れ値の箱ひげ図が最初に表示されます。ヒストグラムには、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。次のプロットも必要に応じて表示できます。

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

レポートには、分位点と要約統計量が表示されます。赤い三角ボタンのメニューには、他にも次のようなレポートオプションがあります。

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

---

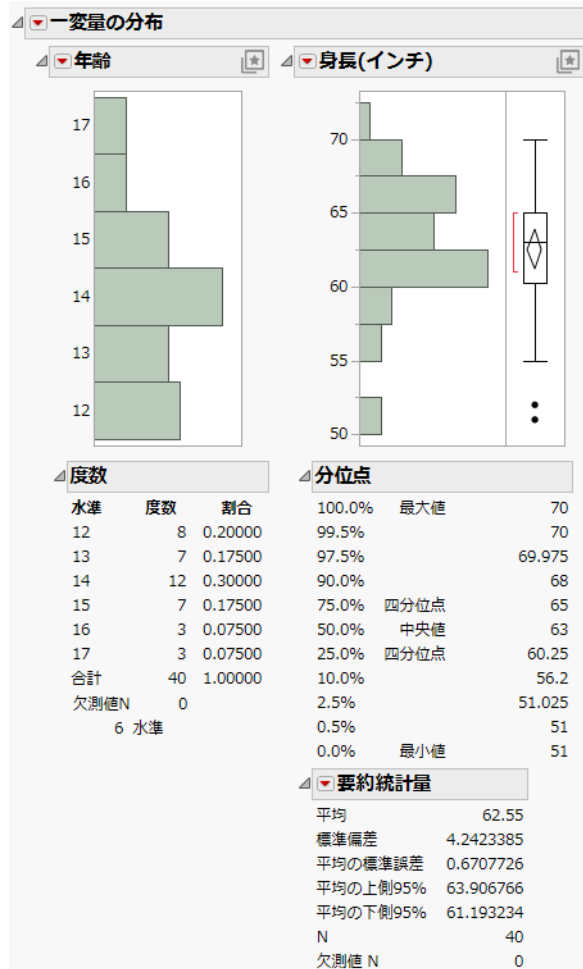
## 「一変量の分布」プラットフォームの例

この例では、40 人の生徒の年齢と身長に関するデータを使用します。変数の分布をグラフにして、分布の外れ値を特定しましょう。

1. [ヘルプ] > [サンプルデータフォルダ] 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. 「年齢」と「身長(インチ)」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。

図3.2 「一変量の分布」プラットフォームの例



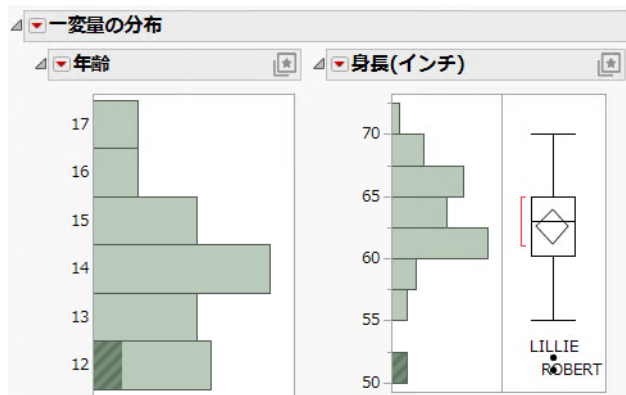
ヒストグラムを見ると、次のことがわかります。

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

「身長(インチ)」のヒストグラムの外れ値にラベルを追加します。

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

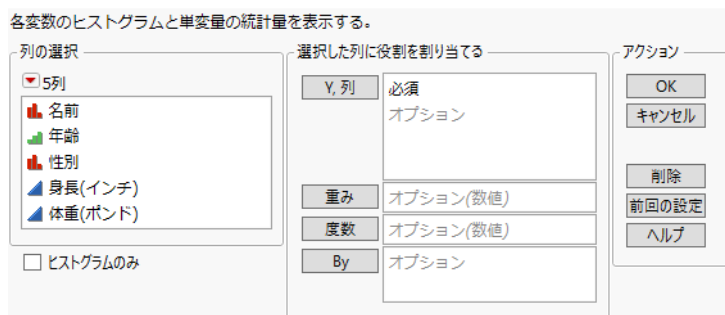
図3.3 ラベルを付けた外れ値



## 「一変量の分布」プラットフォームの起動

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

図3.4 「一変量の分布」の起動ウィンドウ



「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。

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

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

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

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

**工程能力分析の実行** （列に「仕様限界」列プロパティがある場合のみ表示されます。）「仕様限界」列プロパティを含む分析列の「工程能力」レポートを追加します。

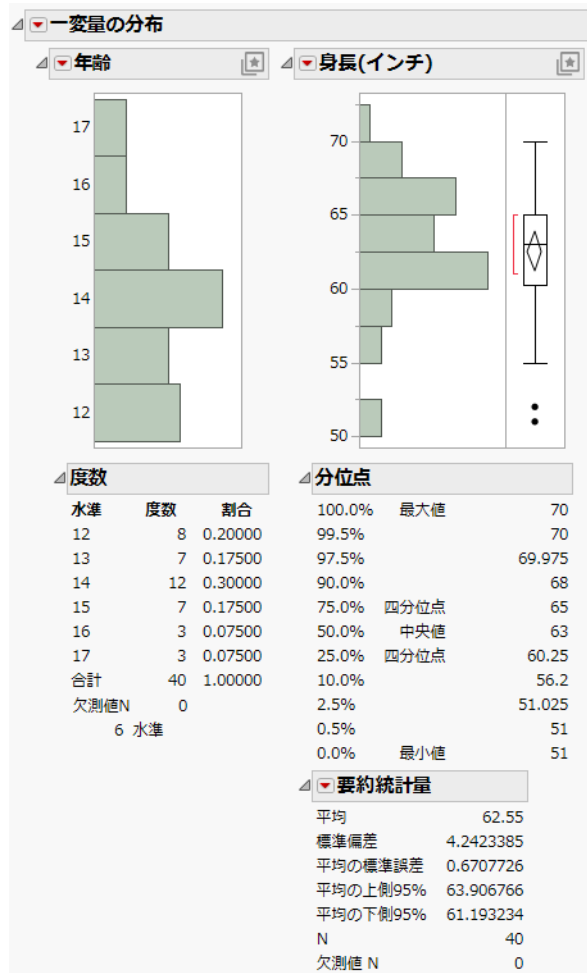
**ヒストグラムのみ** レポートウィンドウに、ヒストグラムだけを表示し、それ以外の情報は表示しません。

起動ウィンドウの詳細については、『JMPの使用法』を参照してください。

## 「一変量の分布」レポート

「一変量の分布」レポートウィンドウには、変数ごとにヒストグラムと統計量のレポートが表示されます。

図3.5 最初に表示される「一変量の分布」レポートウィンドウ



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

### 赤い三角ボタンのメニュー

「一変量の分布」の横にある赤い三角ボタンをクリックするとメニューが開き、レポートウィンドウ内のすべての変数に適用されるオプションが表示されます。「一変量の分布」プラットフォームのオプションを参照してください。



各変数名の横にある赤い三角ボタンをクリックするとメニューが開き、該当する変数にのみ適用されるオプションが表示されます。「[カテゴリカル変数のオプション](#)」または「[連続変数のオプション](#)」を参照してください。

---

**ヒント:** Ctrl キーを押したまま変数のオプションを選択すると、レポート内の尺度が同じすべての変数にそのオプションが適用されます。

---

## ヒストグラムとレポート

ヒストグラムはデータを視覚的に表示します。「[ヒストグラム](#)」を参照してください。

カテゴリカル変数のレポートには、「度数」レポートが含まれます。「[度数](#)レポート」を参照してください。

連続変数のレポートには、「分位点」レポートと「要約統計量」レポートが含まれます。「[分位点](#)レポート」および「[要約統計量](#)レポート」を参照してください。

## ヒストグラムの変数の置き換え、追加、削除

現在ヒストグラムに表示されている変数を、別の変数に置き換えることができます。データテーブルの「列」パネルから変数をドラッグし、ヒストグラムの軸の上にドロップすると、変数が置き換えられます。

別の変数を使って新しいヒストグラムを作成するには、変数をドラッグして、既存のヒストグラムの外にドロップします。新しい変数およびヒストグラムは、既存のヒストグラムの前、間、または後ろに追加できます。

ある変数のヒストグラムを削除するには、赤い三角ボタンのメニューから「**削除**」を選択します。

## 非表示の行と除外されている行

データテーブルの行を「表示しない」に設定すると、プロット上に点としては表示されませんが、ヒストグラムにはこれらの行も含まれたままになります。

ヒストグラムから除外し、分析の対象外にするには、行を「除外する」に設定したうえで、「一変量の分布」の横にある赤い三角ボタンのメニューから「**やり直し**」>「**分析のやり直し**」を選択してください。データテーブルで除外した行は、点のプロットでも非表示になります。

## ヒストグラム

「一変量の分布」プラットフォームでは、ヒストグラムを使ってデータを視覚的に表示します。変数がカテゴリカル（名義尺度もしくは順序尺度）の場合には、棒グラフが表示されます。棒グラフは、変数の水準ごとに1本の棒で度数を表します。変数が連続尺度である場合は、ヒストグラムが表示されます。ヒストグラムは、変数の数値をいくつかのグループに分け、各グループそれぞれ1つの棒で度数を表します。

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

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

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

**軸の再スケール** 軸をクリックしてドラッグすると、スケールが変更されます。

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

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

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

ヒストグラムに関するオプションを表示するには、次の操作を実行します。

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

## 連続変数のヒストグラムにおける棒の幅の変更

連続変数に対するヒストグラムでは、次を使用して棒の幅を変更できます。

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

### 手のひらツールの使用

手のひらツールを使用すると、データをすばやく調べることができます。

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

ヒストグラムはデータをカテゴリ化したものです。棒の高さは、その棒が表すカテゴリにどれだけの観測値（オブザベーション）が含まれるかを全体に占める割合として表しています。横棒を縦に並べたデフォルトのヒストグラムでは

- 手のひらツールを左に動かすと棒の幅が広くなり、棒の数が少なくなります。棒の幅が広くなると、棒の数が少なくなります。
- 手のひらツールを右に動かすと棒の幅が狭くなり、棒の数が多くなります。棒の幅が狭くなると、棒の数が多くなります。
- 手のひらツールを上下に動かすと、棒の開始値がシフトするため、その棒に含まれるデータ点もシフトします。

---

**メモ:** ヒストグラムの形式を変え、縦棒を左右に並べた場合は、操作の方向を置き換えて考えてください。手のひらツールを下に動かすと棒の幅が広くなり、上に動かすと棒の幅が狭くなります。左右へ動かすと、棒の開始値がシフトします。

---

### 【棒の幅の設定】 オプションの使用

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

### 【目盛り間隔】 オプションの使用

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

## 棒の強調表示と行の選択

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

---

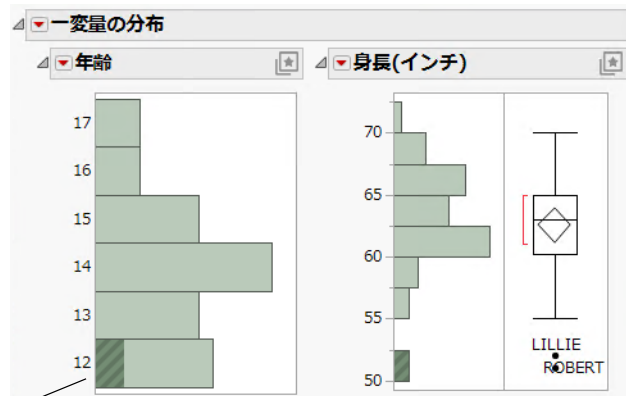
**ヒント:** 特定のヒストグラムの棒の選択を解除するには、Ctrlキーを押しながら、強調表示されている棒をクリックします。

---

図 3.6 棒と行の強調表示

Σ	名前	年齢	性別	身長(インチ)	体重(ポンド)
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112

棒を選択すると、データテーブルの行と他の結果における対応する部分が強調表示される。



## 複数のヒストグラムにまたがる選択

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

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

例として、「[複数のヒストグラムにおけるデータ選択の例](#)」を参照してください。

## 「度数」レポート

「一変量の分布」プラットフォームの「度数」レポートは、変数の尺度が名義・順序・多重応答のいずれかの場合に表示されます。レポートの内容は、変数の尺度によって異なります。

### 名義・順序尺度の変数

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

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

---

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

---

**水準** 応答変数の値をリストしたもの。

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

**割合** 応答変数の水準ごとの割合。割合は、その水準の度数を全体の度数(表の下端に表示)で割ったものです。

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

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

---

**メモ:** なお、欠測値である観測数と、一意な水準の個数が、表の下に表示されます。

---

### 多重応答の変数

多重応答の場合、「度数」レポートには、その変数の分布を要約した3つの表が表示されます。

最初の表には、次のような情報が含まれます。欠測値は分析から除外されます。

**水準** 応答変数の値をリストしたもの。

**度数** 応答変数の各水準を含む行の数。ただし、度数変数を指定した場合、応答変数の各水準を含む行の、度数変数の値の合計が表示されます。

**応答のシェア** 応答変数の水準ごとの割合。度数を応答の合計数で割った値です。

**ケースあたりの比率** 応答変数の各水準を含む観測（つまり、データ行）の割合。度数を全ケース数（つまり、標本サイズ）で割った値です。

その他の表には、次のような統計量が表示されます。

**全ケース数** 測定値の個数。

**応答の合計数** すべての行における応答の数。多重応答変数の応答の個数をすべて合計したものです。

**水準** 応答変数における一意な水準の数。

**応答なし** 応答変数に値がない観測の数。度数変数を指定した場合は、度数変数の値を含んだ「度数」の列も表示されます。

**応答あり** 応答変数に少なくとも1つの値がある観測の数。度数変数を指定した場合は、度数変数の値を含んだ「度数」の列も表示されます。

**単一の応答** 応答変数に値が1つだけある観測の数。度数変数を指定した場合は、度数変数の値を含んだ「度数」の列も表示されます。

**複数の応答** 応答変数に値が2つ以上ある観測の数。度数変数を指定した場合は、度数変数の値を含んだ「度数」の列も表示されます。

## 「分位点」レポート

連続変数の場合、「一変量の分布」プラットフォームの「分位点」レポートには、分位点（パーセント点）が表示されます。詳細については、「[分位点の統計的詳細](#)」を参照してください。

## 「要約統計量」レポート

連続変数の場合、「一変量の分布」プラットフォームの「要約統計量」レポートには、平均や標準偏差などが表示されます。表示する要約統計量は、自由に選べます。「要約統計量」の赤い三角ボタンのメニューから「**要約統計量のカスタマイズ**」を選択して、表示したい統計量を選んでください。

---

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

---

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

### 「要約統計量」レポートの説明

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

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

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

**平均の標準誤差** 平均（標本平均）の標準偏差の推定値。

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

**N** 非欠測値の個数。

**欠測値 N** 欠測値の個数。

## その他の要約統計量

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

**合計** 応答の値の合計。

**分散** 不偏分散。標準偏差の2乗。

**歪度** 歪みや対称性に関する指標。

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

**変動係数** 単位がパーセントの変動係数。標準偏差を平均で割り、100を掛けた値です。変動係数は、ばらつき（変動）の大きさを相対的に評価するときに役立ちます。たとえば、異なる単位で測定されたデータのばらつきを比較するときに使用できます。

**ゼロの個数** 0である値の個数。

**一意な値の個数** 一意（ユニーク）な値の個数。

**無修正平方和** 修正を加えていない平方和。データそのものの2乗和。

**修正平方和** 修正を加えた平方和。平均からの偏差の2乗和。

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

**最小値** 0パーセント点。

**最大値** 100パーセント点。

**中央値** メディアン。50パーセント点。

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

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

**幾何平均** データの総積の  $n$  乗根。利子の計算などによく使われます。データに特に大きな値が含まれていて分布が歪んでいる場合などに、便利です。

---

**メモ:** データに負の値がある場合、幾何平均は欠測値になります。また、データに0の値がある場合（かつ負の値がない場合）は0になります。

---

**範囲** データの最大値と最小値の差。

**四分位範囲** 第3四分位点と第1四分位点の差。

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

**ゼロの割合** 非欠測値に含まれる0の割合。

**非ゼロの割合** 非欠測値に含まれる0以外の値の割合。

**3x 標準偏差** 標準偏差に3を掛けた値。乗数は、「環境設定」>「プラットフォーム」>「一変量の分布」>「一変量の要約統計量」にある「**K 乗数**」で設定できます。

**平均 + 3x 標準偏差** 平均の値に標準偏差の3倍を足したものの。乗数は、「環境設定」>「プラットフォーム」>「一変量の分布」>「一変量の要約統計量」にある「**K 乗数**」で設定できます。

**平均 - 3x 標準偏差** 平均の値から標準偏差の3倍を引いたものの。乗数は、「環境設定」>「プラットフォーム」>「一変量の分布」>「一変量の要約統計量」にある「**K 乗数**」で設定できます。

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

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

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

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



## 要約統計量のオプション

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

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

**すべての最頻値を表示** 複数の最頻値があるときに、そのすべてが表示されます。

詳細については、「[要約統計量の統計的詳細](#)」を参照してください。

---

## 「一変量の分布」プラットフォームのオプション

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

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

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

**複数行に配置** 1行において表示するプロットの数を入力します。複数のプロットを横一列に配置するのではなく、縦方向にも配置したい場合に便利です。

以下のオプションの詳細については、『[JMPの使用法](#)』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMPレポートの設定に合わせて環境設定を変更したりできます。

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

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

---

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの[スクリプトの索引]を開いてください。また、[スクリプトの索引]には、この節で紹介されているオプションのスクリプト例もあります。

---

## カテゴリカル変数のオプション

「一変量の分布」レポートで各変数名の横にある赤い三角ボタンをクリックするとメニューが開きます。このメニューで用意されているオプションは、該当する変数にのみ適用されます。ここでは、カテゴリカル変数（名義尺度または順序尺度の変数）に適用されるオプションについて説明します。

**表示オプション** 以下のような、レポートの表示内容に関するオプションがあります。

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

**横に並べる** ヒストグラムとレポートの表示方向を縦／横に切り替えます。

**軸を左に** （[横に並べる] が選択されている場合のみ使用できます。）「度数」・「割合」・「密度」の各軸を右側から左側へ移動させます。

**ヒストグラムオプション** 「[カテゴリカル変数の \[ヒストグラムオプション\]](#)」を参照してください。

**モザイク図** 名義尺度または順序尺度の変数に対して、モザイク図の表示／非表示を切り替えます。モザイク図とは、長方形を縦に積み重ねた一種の棒グラフであり、長方形の面積がグループの度数に比例しています。

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

**割合の検定** 仮説割合を検定するためのレポートを表示します。この検定の帰無仮説は、「応答の各水準における真の割合は、仮説割合と等しい」というものです。また、対立仮説は、「真の割合の少なくとも2つが、仮説割合と異なる」というものです。「割合の検定」レポートに表示されるオプションは、応答変数の水準数によって異なります。

2水準の変数の場合、両側カイ2乗検定と正確な片側二項検定のいずれかを選べます。どちらの検定でも、変数の一方または両方の水準に対して仮説割合を指定することができます。両方の水準に対して仮説割合を入力し、その和が1でない場合、1になるように値が調整されます。片側検定の場合、観測された割合の方が仮説割合より高いことを検定するか、観測された割合の方が仮説割合より低いことを検定するかを選べます。片側検定で、両方の仮説割合を指定した場合、応答の2つ目の水準に対して検定が行われます。[「2水準の割合の検定例」](#)を参照してください。

水準が3つ以上ある応答の場合は、複数の仮説割合を入力できます。応答の全水準に対して仮説割合を入力し、その和が1でない場合、1になるように値が調整されます。応答の一部の水準に対してのみ仮説割合を入力した場合、仮説割合の調整方法を2つの中から選びます。1つは、指定していない仮説割合を観測された値に固定して、指定した仮説割合を調整する方法です。もう1つは、指定した仮説割合を固定して、指定していない仮説割合を調整する方法です。[「3水準以上の割合の検定例」](#)を参照してください。

**信頼区間** 応答の各水準におけるスコア信頼区間（スコア法によって近似した信頼区間）を計算します。メニューで用意されている信頼水準から1つを選ぶか、**[その他]** を選択して数値で信頼水準を入力します。[「スコア信頼区間の統計的詳細」](#)を参照してください。

**保存** 「カテゴリカル変数の [保存] オプション」を参照してください。

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

## カテゴリカル変数の [ヒストグラムオプション]

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

**縦に表示** ヒストグラムの表示方向を縦／横に切り替えます。

**標準誤差バー** ヒストグラムの水準ごとに標準誤差バーが描かれます。

**棒の間を離す** ヒストグラムの棒を離して表示します。

**ヒストグラムの色** ヒストグラムの棒の色を変更します。

**度数軸** それぞれの棒に含まれている値の度数を表す軸が表示されます。

**割合軸** それぞれの棒に含まれている値の割合を表す軸が表示されます。

**密度軸** ヒストグラムの棒の長さを示す軸が追加されます。

度数と割合は、それぞれ次のように計算されます。

割合 = (棒の幅) × 密度

度数 = (棒の幅) × 密度 × (合計度数)

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

---

**メモ:** 小数点以下の桁数を指定するには、ヒストグラムを右クリックし、[カスタマイズ] > [ヒストグラム] を選択してください。

---

**度数の表示** ヒストグラムの各棒に、その棒の度数が表示されます。

## カテゴリカル変数の [保存] オプション

**水準番号** データテーブルの中に、「水準<列名>」という新しい列が作成されます。各データ行の水準番号は、ヒストグラム上でそのデータ行が属している棒の番号です。

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

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

## 連続変数のオプション

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

**表示オプション** 「[連続変数の\[表示オプション\]](#)」を参照してください。

**ヒストグラムオプション** 「[連続変数の\[ヒストグラムオプション\]](#)」を参照してください。

**正規分位点プロット** 変数がどの程度、正規分布に従っているかを確認できます。「[正規分位点プロット](#)」を参照してください。

**外れ値の箱ひげ図** 分布を観察したり、外れ値を見つけたりするための箱ひげ図が描かれます。「[外れ値の箱ひげ図](#)」を参照してください。

**分位点の箱ひげ図** 「分位点」レポートに示されている分位点を用いた箱ひげ図が描かれます。「[分位点の箱ひげ図](#)」を参照してください。

**幹葉図** 「[幹葉図](#)」を参照してください。

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

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

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

**同等性の検定** 母平均が仮説値と等しいかどうかを評価します。「[同等性の検定](#)」を参照してください。

**信頼区間** 平均と標準偏差に対する信頼区間が計算されます。「[信頼区間](#)」を参照してください。

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

**許容区間** 少なくとも指定した割合だけの観測値を含むと主張できる区間が計算されます。「[許容区間](#)」を参照してください。

**工程能力** (Y列に「検出限界」列プロパティがある場合は使用できません。) 工程が特定の仕様限界(規格限界)に適合しているかどうかを測定します。「[工程能力](#)」を参照してください。

---

**メモ:** Y列に「検出限界」列プロパティがある場合に工程能力分析を行うには、あてはめた分布の赤い三角ボタンのメニューを使います。「[工程能力](#)」を参照してください。

---

**連続分布のあてはめ** 連続分布をデータにあてはめます。Y列に「検出限界」列プロパティがある場合に「連続分布のあてはめ」のオプションを使用すると、打ち切りのある分布があてはめられるため、全種類の分布が使用できるわけではありません。「[分布のあてはめ](#)」を参照してください。

**離散分布のあてはめ** (データの値がすべて整数のときのみ使用可能) 離散分布をデータにあてはめます。「[分布のあてはめ](#)」を参照してください。

**保存** 連続変数またはカテゴリカル変数についての情報が保存されます。「[連続変数の\[保存\] オプション](#)」を参照してください。

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

## 連続変数の [表示オプション]

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

**分位点の間隔の設定** 分位点の間隔を変更するか、デフォルトの値に戻します。

**カスタム分位点** 計算する分位点を、値または増分によって指定できます。信頼区間の信頼水準を変更できます。また、平滑化された経験尤度分位点も計算できます (大きなデータでは、平滑化された経験尤度分位点の計算には時間がかかる場合があります)。

- 分位点の計算方法については、「[分位点の統計的詳細](#)」を参照してください。
- 分位点の信頼限界には、分布によらないノンパラメトリックな方法が使われています。詳細については、Meeker et al. (2017) を参照してください。
- 平滑化した経験尤度法による分位点は、カーネル密度推定に基づいて求められます。これらの分位点および信頼限界の推定方法の詳細については、Chen and Hall (1993) を参照してください。
- 度数に小数点を使用しているときは、信頼限界および平滑化された経験尤度法による分位点は使用できません。

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

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

**横に並べる** ヒストグラムとレポートの表示方向を縦／横に切り替えます。

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

このオプションは、[\[横に並べる\]](#) が選択されている場合のみ使用可能です。

## 連続変数の [ヒストグラムオプション]

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

**シャドウグラム** ヒストグラムの代わりにシャドウグラムを表示します。シャドウグラムは、棒の幅を変更したヒストグラムを重ね合わせたようなものです。シャドウグラムは、いくつかのバンド幅に対するカーネ

ル密度推定の結果を合わせたものです。シャドウグラムでは、いくつかのバンド幅の密度推定において重なり合っている部分が濃く描かれます。

シャドウグラムでは次のオプションは使用できません。

- 標準誤差バー
- 度数の表示
- パーセントの表示

**縦に表示** ヒストグラムの表示方向を縦／横に切り替えます。

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

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

**ヒストグラムの色** ヒストグラムの棒の色を変更します。

**度数軸** それぞれの棒に含まれている値の度数を表す軸が表示されます。

---

**メモ:** ヒストグラムの棒のサイズを変更すると、度数軸も変化します。

---

**割合軸** それぞれの棒に含まれている値の割合を表す軸が表示されます。

---

**メモ:** ヒストグラムの棒のサイズを変更すると、割合軸も変化します。

---

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

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

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

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

---

**メモ:** ヒストグラムの棒のサイズを変更すると、密度軸も変化します。

---

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

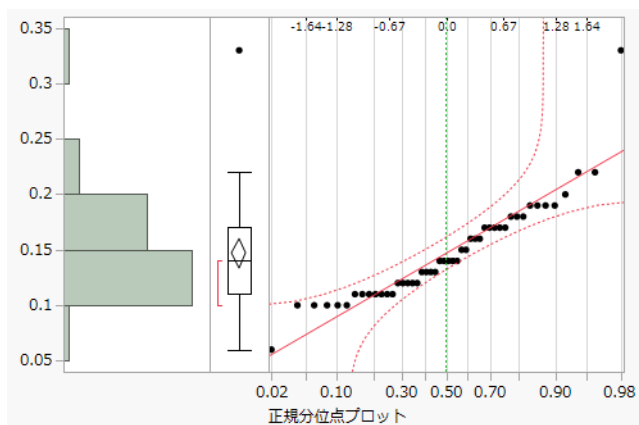
**度数の表示** ヒストグラムの各棒に、その棒の度数が表示されます。

## 正規分位点プロット

「一変量の分布」プラットフォームで「正規分位点プロット」オプションを選択すると、変数がどの程度、正規分布に従っているかを確認できます。ある変数が正規分布に従っている場合、正規分位点プロットはほぼまっすぐな対角線になります。このプロットは、「分位点一分位点プロット」または「Q-Qプロット」とも言います。

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

図3.7 正規分位点プロット



次の点に注意してください。

- 縦軸は、列の値を表します。
- 上の横軸は、正規分位点スケールを示します。
- 下の横軸は、各値の経験累積確率を示します。
- 赤の点線は、Lillieforsの信頼限界を示します。

詳細については、「[正規分位点プロットの統計的詳細](#)」を参照してください。

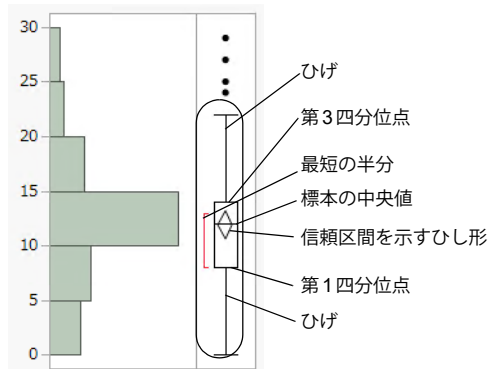
## 外れ値の箱ひげ図

「一変量の分布」プラットフォームの「外れ値の箱ひげ図」オプションを使用すると、分布を見て外れ値であるかも知れない点を特定できます。一般に、箱ひげ図には、いくつかの分位点が表されています。行数が10万未満であれば、デフォルトで外れ値の箱ひげ図が表示されます。この閾値は、[環境設定] > [プラットフォーム] > [一変量の分布] にある「外れ値箱ひげ図を非表示とする行数」で変更できます。

**メモ:** 外れ値の箱ひげ図は、「Tukeyの箱ひげ図」や「図式箱ひげ図 (schematic box plot)」とも呼ばれます。



図 3.8 外れ値の箱ひげ図



外れ値の箱ひげ図には、次のような情報が描かれます。

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

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

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

上の式で計算された範囲内にあるデータ点の一番大きい値と小さい値（つまり、外れ値を除いたもののなかでの最大値と最小値）が、ひげの両端になります。

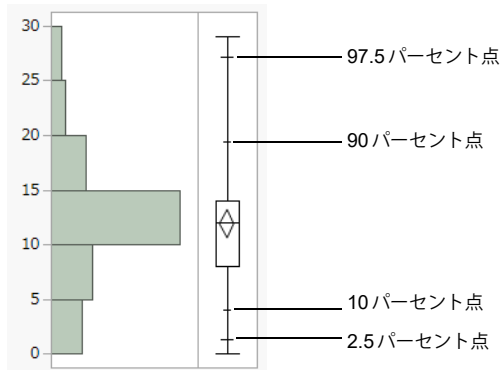
- 箱の外側に表示されている括弧は、最短の半分（**shortest half**）を示します。これは、データの 50% が密集する最も短い範囲です（Rousseeuw and Leroy 1987）。
- 外れ値の箱ひげ図からひし形や最短半分を削除する方法については、「[箱ひげ図からのひし形や最短半分の削除](#)」を参照してください。

## 分位点の箱ひげ図

「一変量の分布」プラットフォームの「分位点の箱ひげ図」オプションを使うと、「分位点」レポートにある特定の分位点を視覚的に確認できます。これを目安にすれば、分布が対称かは一目で判断できます。分布が対称であれば、箱ひげ図に表示される分位点も対称に並んでいるはずです。たとえば、分位点を示すマークの間隔が一方では狭く、もう一方では広い場合、その分布は間隔の広い側に歪んでいることがわかります。



図 3.9 分位点の箱ひげ図



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

### 箱ひげ図からのひし形や最短半分の削除

外れ値または分位点の箱ひげ図から、信頼区間を示すひし形や、「最短の半分」を示す括弧を削除できます。個々の箱ひげ図から削除するほか、常に表示しないようにすることも可能です。

#### 個々の箱ひげ図からひし形や最短半分の削除する手順

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

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

#### 箱ひげ図にひし形や最短の半分が常に表示されないようにする手順

1. **【ファイル】 > 【環境設定】 > 【プラットフォーム】 > 【一変量の分布】** を選択します。
2. 以下のオプションの選択を解除します。
  - 箱ひげ図に信頼のひし形を表示
  - 外れ値の箱ひげ図に最短の半分を表示
3. **【OK】** をクリックします。

これで、「一変量の分布」で箱ひげ図を追加したときに、信頼のひし形や最短の半分が表示されなくなります。

## 幹葉図

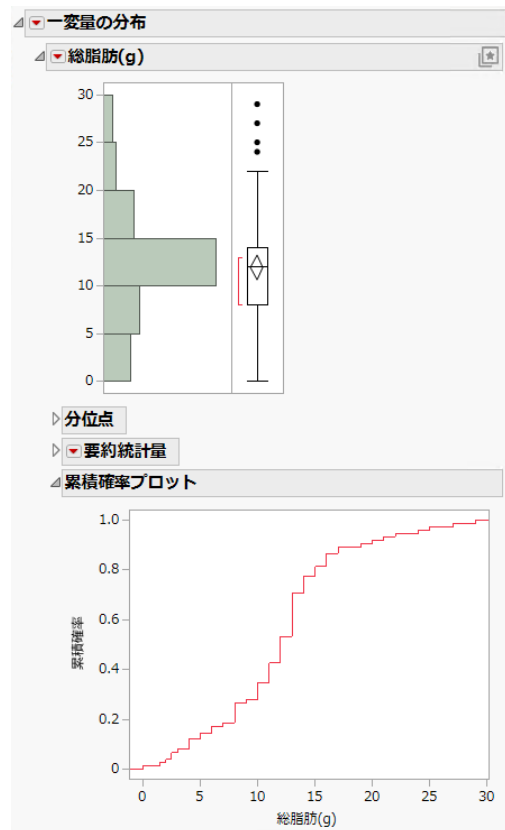
「一変量の分布」プラットフォームの「幹葉図」オプションを使うと、データをテキストで表示した図が作成されます。幹葉図の各行に表示された幹の値は、データ値の最初のいくつかの桁を表します。葉の値は、各データ値の残りの桁を表します。幹と葉を結合すると、データ値になります。幹葉図には、テーブル上の実測値を丸めた値が表示される場合もあります。幹葉図の値はマウスやブラシツールで選択できます。

**メモ:** 幹葉図では、小数点を含む度数は、整数に切り上げられます。

## 累積確率プロット

「一変量の分布」プラットフォームの「累積確率プロット」オプションは、経験累積分布関数のプロットを作成します。累積確率プロットでは、横軸上の所定値以下に収まるデータの割合を確認できます。データに確率分布をあてはめると、そのあてはめた確率分布の累積分布関数が累積確率プロットに重ね合わせて描かれます。

図3.10 累積確率プロット



たとえば、この累積確率プロットでは、データのおよそ34%で総脂肪の値が10グラムを下回っています。

## 平均の検定

「一変量の分布」プラットフォームで「平均の検定」オプションを選択すると、検定を設定するためのウィンドウが開きます。そこで指定された検定方法によって、一標本の平均に対する検定が実行されます。標準偏差の値を指定すると、 $z$ 検定が実行されます。標準偏差を入力しなかった場合は、標本から推定された標準偏差を使って、 $t$ 検定が実行されます。ノンパラメトリックな Wilcoxon の符号付順位検定を実行することもできます。

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

### 「平均の検定」レポートの説明

#### 平均の検定で計算される統計量

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

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

#### $p$ 値

**$p$  値 (Prob>| $t$ |)** 母平均が仮説値と等しいという仮定のもとで、計算された標本の  $t$  値の絶対値よりも、 $t$  値の絶対値が大きくなる確率です。これは両側  $t$  検定の  $p$  値です。

**$p$  値 (Prob> $t$ )** 母平均が仮説値と等しいという仮定のもとで、計算された標本の  $t$  値より大きい  $t$  値を得る確率です。これは上側の片側検定の  $p$  値です。

**$p$  値 (Prob< $t$ )** 母平均が仮説値と等しいという仮定のもとで、計算された標本の  $t$  値より小さい値  $t$  を得る確率です。これは下側の片側検定の  $p$  値です。

### 「平均の検定」オプションの説明

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

**検出力アニメーション** 対話的なグラフに、検出力とベータ（第2種の誤りの確率）を描きます。真の平均や、平均の仮説値によって、検出力とベータがどのように変化するを確認できます。

**検定結果の削除** 平均に対する検定の結果を削除します。

## 標準偏差の検定

「一変量の分布」プラットフォームで「標準偏差の検定」オプションを選択すると、一標本の標準偏差に対する検定が実行されます。このオプションは、値を変えて何度でも実行できます。実行するたびに、新しい「標準偏差の検定」レポートが追加されます。

**検定統計量** カイ2乗検定統計量の値。「標準偏差の検定の統計的詳細」を参照してください。

**最小p値** 母標準偏差が仮説値と等しいという仮定のもとで、計算された標本のカイ2乗値よりも極端な値となるカイ2乗値を得る確率です。「標準偏差の検定の統計的詳細」を参照してください。

**p値(Prob>ChiSq)** 母標準偏差が仮説値と等しいという仮定のもとで、計算された標本のカイ2乗値より大きいカイ2乗値を得る確率です。これは片側 $t$ 検定の $p$ 値です。

**p値(Prob<ChiSq)** 母標準偏差が仮説値と等しいという仮定のもとで、計算された標本のカイ2乗値より小さいカイ2乗値を得る確率です。これは片側 $t$ 検定の $p$ 値です。

## 同等性の検定

「一変量の分布」プラットフォームの「同等性の検定」オプションを使うと、母集団の平均が仮説値と等しいかどうかを評価することができます。差の閾値を設定し、その閾値より小さければ差がないとみなします。JMPの「同等性の検定」オプションは、Two One-Sided Tests (TOST; 2つの片側検定) アプローチを採用しています。2つの片側 $t$ 検定は、「真の平均と仮説値の差は、閾値を超える」という帰無仮説の検定です。両方の帰無仮説が棄却されれば、真の差は閾値を超えない、つまり、平均は仮説値に実質的に等しいとみなせるという結論が導かれます。

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

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

図3.11 「同等性の検定」レポート



レポートには次の情報が表示されます。

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

## 信頼区間

「一変量の分布」プラットフォームの「信頼区間」オプションは、平均と標準偏差の信頼区間を表示します。**[0.90]**、**[0.95]**、**[0.99]** の各オプションを選択した場合、両側信頼区間が計算されます。**[信頼区間] > [その他]** オプションを選択した場合、信頼水準を入力し、片側か両側のどちらかを選択できます。また、既知の  $\sigma$  を入力することもできます。既知の  $\sigma$  を入力した場合は、 $t$  値ではなく  $z$  値によって、平均の信頼区間が計算されます。

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

## 予測区間

「一変量の分布」プラットフォームの「予測区間」オプションを使うと、1つの観測値、もしくは次に無作為抽出される標本の平均や標準偏差が含まれる区間が計算されます。これらの予測区間は、「正規分布に従う母集団から、標本が無作為抽出されている」という仮定の下で計算されます。片側予測区間と両側予測区間のいずれかを選択できます。

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

### 関連情報

- 詳細については、「[予測区間の統計的詳細](#)」を参照してください。
- 例として、「[予測区間の例](#)」を参照してください。

## 許容区間

「一変量の分布」プラットフォームの「許容区間」オプションを使うと、少なくとも指定した割合だけの観測値を含むと主張できる区間が計算されます。これは、指定の割合の観測値を含む区間を推定したものです。許容区間は、Meeker et al. (2017) と Tamhane and Dunlop (2000) で詳細に論述されています。

変数の「許容区間」オプションを選択すると、「許容区間」ウィンドウが表示されます。このウィンドウで、信頼水準、観測値（オブザベーション）の割合、片側または両側限界、手法を指定します。手法には、「正規分布を仮定」と「ノンパラメトリック」があります。「正規分布を仮定」を選択すると、「正規分布に従う母集団から、標本が無作為抽出されている」という仮定の下で許容区間が計算されます。「ノンパラメトリック」を選択すると、分布によらない許容区間が計算されます。

---

**メモ：**「正規分布を仮定」オプションを指定した場合、区間の計算に使用した $k$ の値がレポートの表に含まれます。この値を表示するには、「許容区間」レポートの表を右クリックし、「列」>「 $K$ 係数」を選択します。

---

「許容区間」の赤い三角ボタンのメニューには、許容区間を「仕様限界」列プロパティとして保存するオプションと、「許容区間」レポートを削除するオプションがあります。

### 関連情報

- 詳細については、「[許容区間の統計的詳細](#)」を参照してください。
- 例として、「[許容区間の例](#)」を参照してください。

## 工程能力

「一変量の分布」プラットフォームの「工程能力」オプションを使うと、指定の仕様限界と比較したときの工程能力が測定できます。優れた工程とは、安定した状態にあり、かつ、仕様限界内にほどよく収まる製品を一貫して生産できる工程です。工程能力指数（capability index）は、工程のパフォーマンスを示す指標で、工程の中心とばらつき（変動）を、仕様限界に照らして要約した指標です。

### 仕様限界

列に「仕様限界」列プロパティが含まれている場合、起動ウィンドウで「工程能力分析の実行」オプションを選択すると、「工程能力」レポートが自動的に生成されます。このレポートは正規分布に基づいています。ただし、その列に「分布」列プロパティも含まれている場合は、そこで指定されている分布に基づきます。

---

**ヒント：**複数の列に対して同時に仕様限界を追加する方法については、『品質と工程』を参照してください。

---

列に仕様限界が含まれていない場合は、分析対象の変数名の横にある赤い三角ボタンをクリックして「**工程能力**」を選択した後に、「工程能力分析」ウィンドウで仕様限界を設定してください。

レポートの仕様限界をデータテーブルに列プロパティとして保存するには、「工程能力分析」の赤い三角ボタンをクリックし、「**仕様限界を列プロパティとして保存**」を選択します。次に工程能力分析を再実行したとき、その保存した仕様限界が自動的に読み込まれます。

### 「工程能力分析」ウィンドウ

分析対象の変数名の横にある赤い三角ボタンをクリックして「工程能力」を選択すると、「工程能力分析」ウィンドウが開きます。

このウィンドウで、仕様限界、前提とする分布、シグマの推定手法などを指定します。工程能力分析では、シグマ、つまり、（短期間の）群内変動を推定する方法を指定する必要があります。また、選択したオプションによって、異なるサブオプションが表示されます。

図 3.12 「工程能力分析」 ウィンドウ

仕様限界の入力

LSL	目標値	USL	限界の表示
			<input type="checkbox"/>

工程能力分析オプション

工程能力分析オプションの選択

- ☒ サブグループのサイズ=1
- ☐ サブグループID列を使用する
- ☐ 一定のサブグループサイズを使用する
- ☐ 履歴シグマの使用
- ☐ 非正規分布を使用する

移動範囲オプション

非正規分布のオプション

☒ 群内工程能力の表示

$\alpha$ 水準の指定

OK キャンセル

**仕様限界の入力** 下側仕様限界・目標値・上側仕様限界を指定します。これらのうち、少なくとも1つの値を指定する必要があります。[限界の表示] オプションを選択すると、「一変量の分布」レポートのヒストグラムに、仕様限界が表示されます。

**工程能力分析オプション** 選択したオプションに応じて、サブオプションが表示されます。次のいずれかのオプションを選択します。

**サブグループのサイズ=1** サブグループのサイズを1に設定します。また、「移動範囲オプション」が表示されます。『品質と工程』を参照してください。

**サブグループID列を使用する** サブグループID列を選択できます。また、「サブグループのオプション」と「移動範囲オプション」が表示されます。『品質と工程』を参照してください。

**一定のサブグループサイズを使用する** 一定のサブグループサイズを設定できます。また、「サブグループのオプション」と「移動範囲オプション」が表示されます。『品質と工程』を参照してください。

**履歴シグマを使用する** 過去に受け入れられたシグマの値を割り当てます。『品質と工程』を参照してください。

**非正規分布を使用する** 非正規分布を選択できます。また、「非正規分布のオプション」が表示されます。『品質と工程』を参照してください。

**群内工程能力の表示** (「工程能力分析オプション」で、[サブグループのサイズ=1] または [非正規分布を使用する] を選択している場合にのみ使用可能) 群内シグマの推定値をレポートに表示します。

**$\alpha$ 水準の指定** 信頼限界の有意水準を指定します。



## 「工程能力分析」レポート

「工程能力分析」ウィンドウで [OK] をクリックすると、「工程能力分析」レポートが表示されます。このレポートには、選択した変数の工程能力レポートが含まれます。このレポートの詳細については、『品質と工程』を参照してください。

- 統計的詳細については、『品質と工程』を参照してください。
- 例として、「[工程能力分析の例](#)」を参照してください。
- 「工程能力」プラットフォームについては、『品質と工程』を参照してください。

---

**メモ:** 「一変量の分布」プラットフォームの「工程能力」レポートの主な設定は、[ファイル] > [環境設定] > [プラットフォーム] > [工程能力] で変更できます。

---

## 分布のあてはめ

分析対象の列の尺度が「連続尺度」である場合には、[連続分布のあてはめ] または [離散分布のあてはめ] の各オプションのサブメニューを使用することにより、特定の確率分布をあてはめることができます。「連続尺度」の列に分布をあてはめると、ヒストグラムに曲線が追加されます。また、「分布の比較」レポートと、「分布のあてはめ」レポートがレポートウィンドウに追加されます。「分布のあてはめ」レポートには赤い三角ボタンがあり、そこで追加のオプションを選択できます。「[あてはめた分布のオプション](#)」を参照してください。「一変量の分布」レポートでは、列に [分布] 列プロパティが含まれている場合、そこで指定されている確率分布がデフォルトであてはめられます。

---

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

---

## 連続分布のあてはめ

[連続分布のあてはめ] のサブメニューには、連続分布をあてはめるためのオプションが含まれています。確率分布の詳細については、「[連続分布のあてはめの統計的詳細](#)」を参照してください。

**正規のあてはめ** データに正規分布をあてはめます。正規分布は、分布が左右対称で、ほとんどのデータが中央にある分布です。正規分布は、負の値や小数点などが、データにあってもかまいません。ここでの標準偏差推定値は、不偏分散推定値の平方根です。

**Cauchyのあてはめ** データにCauchy分布をあてはめます。Cauchy分布には、平均と標準偏差がありません。ほとんどのデータは本来Cauchy分布には従いませんが、外れ値の割合が大きいデータ（最大50%のものが外れ値となっているようなデータ）に対してロバストな位置と尺度を推定するときに役立ちます。

**Studentのtのあてはめ** Studentのt分布をデータにあてはめます。Studentのt分布は、正規分布とCauchy分布の間の空間に広がるロバストな分布です。Studentのt分布は、自由度が無限に近づくとき正規分布と

等しくなり、自由度が1になると、Cauchy分布と等しくなります。「一変量の分布」プラットフォームは、自由度の値も推定します。

**SHASHのあてはめ** データにSHASH分布（sinh-arcsinh分布）をあてはめます。Johnson分布と同じく、正規分布へと変換できる分布です。SHASH分布は、特殊な場合として正規分布も含んでいます。また、SHASH分布は、対称な分布と非対称な分布を含んでいます。

**ゼロ強調 SHASHのあてはめ** データにゼロ強調のSHASH分布（sinh-arcsinh分布）をあてはめます。ゼロ強調のSHASH分布とは、観測値が0において確率質量をもつSHASH分布です。また、SHASH分布は、対称な分布と非対称な分布を含んでいます。

**指数のあてはめ** （負の観測値がないときのみ使用できます。）データに指数分布をあてはめます。指数分布は右に裾をひいた形状です。生存時間や連続するイベント間の時間をモデル化するのによく使用されます。

**ExGaussのあてはめ** ExGauss分布（指数修正 Gauss分布）をデータにあてはめます。ExGauss分布（指数修正 Gauss分布）は、正規分布に従う確率変数と指数分布に従う確率変数の和が従う分布です。また、この分布は、対称な分布と非対称な分布を含んでいます。

**ガンマのあてはめ** （すべての観測値が正の値である場合にのみ使用できます。）データにガンマ分布をあてはめます。ガンマ分布は、正の値を柔軟にモデル化できる分布です。

**対数正規のあてはめ** （すべての観測値が正の値である場合にのみ使用できます。）データに対数正規分布をあてはめます。対数正規分布は右に裾をひいた形状です。生存時間や連続するイベントが発生するまでの時間をモデル化するのによく使用されます。対数正規分布のパラメータ推定値は、最尤推定値です。

**Weibullのあてはめ** （すべての観測値が正の値である場合にのみ使用できます。）データにWeibull分布をあてはめます。Weibull分布は柔軟で、生存時間やイベントが発生するまでの時間をモデル化するのによく使用されます。

**二重正規混合のあてはめ** 2重正規混合分布(2つの正規分布を混合した分布)をあてはめます。二峰性の分布にも対応した柔軟な分布です。

**三重正規混合のあてはめ** 3重正規混合分布(3つの正規分布を混合した分布)をあてはめます。多峰性の分布にも対応した柔軟な分布です。

**平滑化曲線のあてはめ** ノンパラメトリックな密度の推定値を計算し、平滑曲線をあてはめます。「ノンパラメトリック密度」レポートに表示されるスライドを使って**カーネルのバンド幅**を変更し、平滑化の度合いを調節できます。デフォルトにおけるカーネルのバンド幅は、次の式で求められます。 $n$ は標本サイズ（観測値の個数）、 $S$ は $n$ で割って求めた標本標準偏差です。

$$\text{バンド幅} = \frac{0.9S}{n^{1/5}}$$

**Johnsonのあてはめ** データにJohnson分布をあてはめます。3種類のJohnson分布（Su、Sb、Sl）のうち、最適なものがあてはめられ、レポートが作成されます。Johnson分布族は、歪度と尖度のあらゆる組み合わせに対応しており、柔軟にデータにあてはまります。Johnson分布の選択プロセスとパラメータ推定については、Slifker and Shapiro（1980）を参照してください。

**ベータのあてはめ** （すべての観測値が0～1の範囲内に収まる場合にのみ使用できます。）データにベータ分布をあてはめます。0より大きく1より小さいデータをモデル化するのに便利で、割合や比率を表すのによく使用されます。

**すべてをあてはめ** 変数に対してあてはめることが可能な連続分布をすべてあてはめます。「分布の比較」レポートに、あてはめられた各分布の統計量が表示されます。デフォルトでは、データに最も適合している分布のチェックボックスがオンになり、ヒストグラムにも表示されます。チェックボックスで分布を選択することによって、その分布のレポート、およびヒストグラム上に描かれる密度曲線の表示／非表示を切り替えることができます。「分布の比較」のリストは初期状態で、AICcの値で昇順に並べられます。

---

**ヒント：**「分布の比較」の「分布」列の分布名をダブルクリックすると、その分布をリストからすぐに削除できます。この操作を行うと、その分布のあてはめレポートも削除されます。

---

**旧版のメニューを有効にする** [旧版のあてはめメニュー] の表示／非表示を切り替えます。分布のあてはめの一部の機能はJMP 15で一新されました。このオプションを選択すると、JMPの旧版における機能が使用できるようになります（互換性のため、旧版の機能も残されています）。旧版の機能の使い方については、JMP 16.2のヘルプにある「旧版のあてはめの詳細」をご覧ください。

## 離散分布のあてはめ

[離散分布のあてはめ] には、離散分布をあてはめるためのオプションが含まれており、データの値がすべて整数の場合に使用できます。確率分布の詳細については、「[離散分布のあてはめの統計の詳細](#)」を参照してください。

**Poissonのあてはめ** データにPoisson分布をあてはめます。Poisson分布は、一定期間内に生じているイベントの回数などの度数データをモデル化するのに便利です。

**負の二項のあてはめ** データに負の二項分布をあてはめます。負の二項分布は、特定の失敗回数に達するまでの成功回数をモデル化するのに便利です。また、負の二項分布はガンマPoisson分布を別のパラメータ表現で表したものです。

**ゼロ強調 Poissonのあてはめ** （データ値にゼロがある場合のみ使用できます。）データにゼロ強調 Poisson分布をあてはめます。この分布は、データがゼロである割合が通常のPoisson分布よりも大きいと仮定しています。

**ゼロ強調 負の二項のあてはめ** （データ値にゼロがある場合のみ使用できます。）データにゼロ強調 負の二項分布をあてはめます。この分布は、データがゼロである割合が通常の負の二項分布よりも大きいと仮定しています。

**二項のあてはめ** データに二項分布をあてはめます。この分布は、 $n$ 回の独立した試行を行ったときの成功回数をモデル化するのに便利です。各試行内においては、成功確率 $p$ は一定です。標本サイズは、データのすべての行に対して一定の値を指定するか、または各行での標本サイズを含む列を指定できます。

---

**メモ:** 標本サイズが一定でない場合、密度曲線・診断プロット・プロファイルは使用できません。

---

**ベータ二項のあてはめ** データにベータ二項分布をあてはめます。この分布は、二項分布で過分散が生じたときの分布です。この分布では、データの各行における標本サイズが2以上でなければいけません。標本サイズは、データのすべての行に対して一定の値を指定するか、または各行での標本サイズを含む列を指定できます。

---

**メモ:** 標本サイズが一定でない場合、密度曲線・診断プロット・プロファイルは使用できません。

---

**ゼロ強調 二項のあてはめ** (データ値にゼロがある場合のみ使用できます。) データにゼロ強調の二項分布をあてはめます。この分布は、データがゼロである割合が通常の二項分布よりも大きいと仮定しています。

---

**メモ:** 標本サイズが一定でない場合、密度曲線・診断プロット・プロファイルは使用できません。

---

**ゼロ強調 ベータ二項のあてはめ** (データ値にゼロがある場合のみ使用できます。) データにゼロ強調のベータ二項分布をあてはめます。この分布は、データがゼロである割合が通常のベータ二項分布よりも大きいと仮定しています。

---

**メモ:** 標本サイズが一定でない場合、密度曲線・診断プロット・プロファイルは使用できません。

---

## あてはめた分布のオプション

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

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

**診断プロット** あてはめた分布の適合度を視覚的に調べるためのオプションがあります。診断プロットでは、それぞれの観測されたデータ点が、仮定した確率分布から推定される値と比較されます。以下のオプションがあります。

**分位点プロット** 分位点プロットの表示／非表示を切り替えます。このプロットは、観測値と分位点との関係を示します。この時に使われる分位点は、あてはめた分布に基づいた確率分布から算出されます。直線的な関係になっている場合は、データがあてはめた分布に従っていることを示唆しています。分位点プロットは、**確率プロット**とも呼ばれます。

---

**メモ:** 正規分位点プロットは、正規分布の分位点プロットです。分位点プロットの構成については、「[正規分位点プロット](#)」を参照してください。

---

**確率プロット** 確率プロットの表示／非表示を切り替えます。このプロットは、経験累積分布関数と、確率分布から算出された累積分布関数との関係を示します。

**プロファイル** 以下のオプションがあります。

**分布プロファイル** 累積分布関数を描いた予測プロファイルの表示／非表示を切り替えます。

**分位点プロファイル** 分位点関数を描いた予測プロファイルの表示／非表示を切り替えます。

**列の保存** 以下のオプションがあります。

**密度関数の保存** パラメータ推定値を代入した密度関数が、データテーブルの列に保存されます。

**分布関数の保存** パラメータ推定値を代入した累積分布関数が、データテーブルの列に保存されます。

**シミュレーション計算式の保存** パラメータ推定値を代入した確率分布に基づいて、値をシミュレーションする計算式が、データテーブルの列に保存されます。このオプションによって保存した列は、「シミュレーション」ユーティリティで切り替え先の列として使用できます。「[シミュレーション](#)」を参照してください。

**変換の計算式を保存** (Johnson 分布と SHASH 分布のあてはめでのみ使用できます。) 変換の計算式が、データテーブルの列に保存されます。この計算式は、あてはめた分布に基づいて、Y 列を正規分布に従う列へと変換します。

**適合度** (Johnson 分布・平滑化曲線・正規混合分布では使用できません。)  
「適合度検定」レポートの表示／非表示を切り替えます。このレポートには、あてはめた分布の適合度検定が表示されます。

連続分布のあてはめでは Anderson-Darling 検定が行われます。 $p$  値は、パラメトリックなブートストラップを使用したシミュレーションにより求められます (Stephens (1974) の節 4.1 で説明されている手順と似ています)。正規分布については、標本サイズが 2000 以下の場合で固定パラメータがないときは、正規性に関する Shapiro-Wilk 検定も行われます。

離散分布では Pearson カイ 2 乗検定が行われます。二項分布とベータ二項分布のあてはめでは、試行回数が一定である場合のみ、適合度検定を使用できます。

**分布パラメータの指定** (Johnson 分布または滑らかな曲線のあてはめでは使用できません。) 一部のパラメータを任意の値に固定し、他のパラメータを推定することができます。また、入力したパラメータ値がデータにあてはまるかどうかを検定され、その結果が「妥当性の尤度比検定」レポートとして表示されます。

**工程能力** (Cauchy 分布、Student の  $t$  分布、ゼロ強調 SHASH 分布、離散分布のあてはめでは使用できません。) あてはめた分布を使って、工程能力分析を実行します。工程能力分析とは、仕様限界に基づいて工程がどれほど優れているかを調べる方法です。あてはめた分布の赤い三角ボタンをクリックして「工程能力」を選択すると、以下のセクションを含むウィンドウが開きます。

**仕様限界の入力** 仕様限界を入力します。あてはめた分布に基づいて仕様限界を計算するには、ここを空白にしたまま、「分位点に基づく仕様限界の計算オプション」の下にあるオプションを使用してください。

**分位点に基づく仕様限界の計算オプション** あてはめた分布に基づいて仕様限界を計算します。計算方法は2種類あります。

1つ目は、分位点の確率を指定して、仕様限界を計算する方法です。

2つ目は、Kシグマ乗数を指定して、仕様限界を計算する方法です。この方法では、両側限界または片側限界のいずれかを選択できます。

確率またはシグマ乗数を入力して、**「仕様限界の計算」**をクリックすると、仕様限界が計算されて、「仕様限界の入力」パネル内にこれらの限界が自動的に入力されます。これらの限界を使って「工程能力」レポートを作成するには、**「OK」**をクリックしてください。[仕様限界と分布をレポートなしで列プロパティに保存] オプションが選択されている状態で **「OK」** をクリックした場合、該当する列プロパティがデータテーブルに保存され、「一変量の分布」レポートウィンドウには何も追加されません。

**工程能力分析オプション** 以下のオプションがあります。

「移動範囲オプション」アウトラインでは、移動範囲の種類を選択できます。『品質と工程』を参照してください。

「非正規分布のオプション」アウトラインでは、非正規分布に基づく工程能力の計算方法を選択できます。『品質と工程』を参照してください。

工程能力分析オプションとレポートの詳細については、『品質と工程』を参照してください。

---

**メモ:** 「一変量の分布」プラットフォームの「工程能力」レポートの主な設定は、**「ファイル」 > 「環境設定」 > 「プラットフォーム」 > 「工程能力」** で変更できます。

---

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

## 連続変数の「保存」オプション

「一変量の分布」プラットフォームの「保存」メニューのオプションを使用して、連続尺度の変数に関する情報を保存できます。保存に関する各オプションを選択すると、現在のデータテーブルに新しい列が作成されます。新しい列の名前は、選択した保存コマンド名の後ろに変数名（下の表で<列名>と表記）が付いた形になります（表3.1）。

「保存」メニューのオプションを繰り返し選択し、ヒストグラムの棒を組み合わせる前と後など、異なる状況で同じ情報を何度でも保存できます。なお、「保存」メニューの同じオプションを2回以上実行した場合には、列の名前には「列名1」、「列名2」、「列名3」というように連番が振られていきます。

表 3.1 「保存」メニューのオプションの説明

オプション	データテーブルに追加される列	説明
水準番号	水準 <列名>	各データ行の水準番号は、ヒストグラム上でそのデータ行が属している棒の番号です。棒に対して、小さいものから順に1から始まる通し番号が付けられます。  <b>メモ：</b> 元の情報を維持するために、新しい列に値ラベルが追加されますが、デフォルトではオフになっています。
水準中間点	中間点 <列名>	各データ行の中間点の値は、水準の下側の境界に、水準の幅の半分を足したものです。  <b>メモ：</b> 元の情報を維持するために、新しい列に値ラベルが追加されますが、デフォルトではオフになっています。
順位	順位 <列名>	対応する列の値の1から始まる順位が保存されます。重複している値には、データテーブル内での順序に従って、連続した順位が割り当てられます。
平均順位	平均順位 <列名>	データに重複がない場合は、平均順位は、通常の順位と同じです。同じ値が $k$ 回起こっている場合、それらの値の順位を合計し、 $k$ で割ったものが平均順位です。
確率スコア	確率 <列名>	非欠測値の個数を $N$ としたとき、ある値の確率スコアは、その値の平均順位を $N + 1$ で割って算出されます。この列の値は、経験累積分布関数（CDF）に似ています。
正規分位点	正規分位点 <列名>	正規分位点を保存します。「 <a href="#">正規分位点プロットの統計的詳細</a> 」を参照してください。
標準化	標準化 <列名>	標準化したデータを保存します。「 <a href="#">標準化されたデータの保存</a> 」の統計的詳細」を参照してください。
中心化	中心化 <列名>	平均を0に中心化した値を保存します。
ロバスト 標準化	ロバスト 標準化 <列名>	ロバストな平均値を引いて中心化したものを、ロバストな標準偏差で割って標準化した値を保存します。
ロバスト 中心化	ロバスト 中心化 <列名>	ロバストな平均値を引いて中心化した値を保存します。
スクリプトをログに保存	(なし)	スクリプトをログウィンドウに出力します。スクリプトを実行すると、分析を再現できます。



## 「一変量の分布」プラットフォームの別例

ここでは、「一変量の分布」プラットフォームを使った例をご紹介します。

- 「複数のヒストグラムにおけるデータ選択の例」
- 「By 変数の例」
- 「2水準の割合の検定例」
- 「3水準以上の割合の検定例」
- 「予測区間の例」
- 「許容区間の例」
- 「工程能力分析の例」
- 「検出限界の例」

### 複数のヒストグラムにおけるデータ選択の例

ここでは、複数のヒストグラムでデータを選択する方法を紹介します。

1. [ヘルプ] > [サンプルデータフォルダ] 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。

小規模企業について、「タイプ」の分布を確認しましょう。

5. 「会社規模」において、「small」の棒をクリックします。

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

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

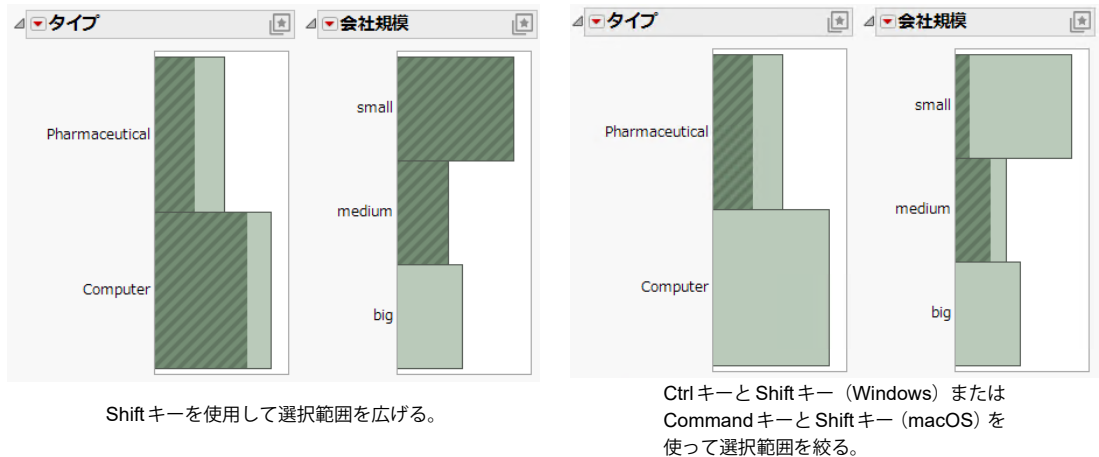
小規模（small）および中規模（medium）における「タイプ」の分布を確認できます。[図3.13](#)の左側を参照してください。次は、選択範囲を絞って、小規模および中規模の製薬会社だけを表示することにします。

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

小規模（small）、もしくは中規模（medium）の製薬会社（Pharmaceutical）が、どれぐらいデータにあるかが分かります。[図3.13](#)の右側を参照してください。



図 3.13 複数のヒストグラムにおけるデータの選択



## By 変数の例

ここでは、By 変数を使ってカテゴリカル変数の水準ごとに分析を行う方法を説明します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Lipid Data.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「コレステロール」を選択し、[Y, 列] をクリックします。
4. 「性別」を選択し、[By] をクリックします。

このように指定すると、性別（female と male）ごとに分析結果が表示されます。

5. [OK] をクリックします。

ヒストグラムとレポートの表示方向を切り替えてみましょう。

6. 「一変量の分布」の赤い三角ボタンをクリックし、[積み重ねて表示] を選択します。

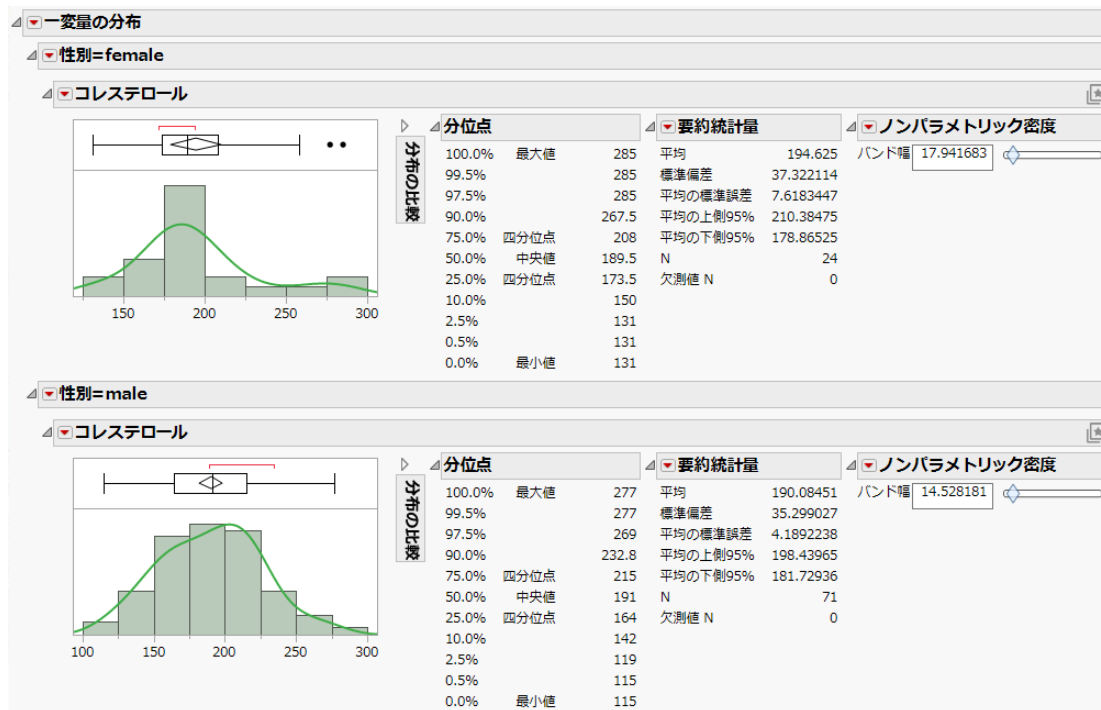
両方のヒストグラムに平滑曲線を追加してみましょう。

7. Ctrl キーを押したままにします。「コレステロール」の赤い三角ボタンをクリックし、[連続分布のあてはめ] > [平滑化曲線のあてはめ] を選択します。

「分布の比較」レポートを非表示にします。

8. Ctrl キーを押したままにします。「分布の比較」の横にあるグレーの開閉アイコンをクリックします。

図 3.14 性別ごとの一変量の分布



## 2水準の割合の検定例

この例では、「一変量の分布」プラットフォームを使い、水準数が2の変数に対して「割合の検定」レポートを作成します。

### 「割合の検定」レポートの開始

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Penicillin.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「反応」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。
5. 「反応」の赤い三角ボタンをクリックし、[割合の検定] を選択します。

図3.15 変数が2水準のときの「割合の検定」レポートオプション

一変量の分布

反応

Died

Cured

度数

割合の検定

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

クリックして、仮説割合を入力してください。

割合の検定に使う対立仮説を選択します。

☒ 割合が仮説値と等しくない(両側カイ2乗検定)

☐ 割合が仮説値より大きい(正確な片側二項検定)

☐ 割合が仮説値より小さい(正確な片側二項検定)

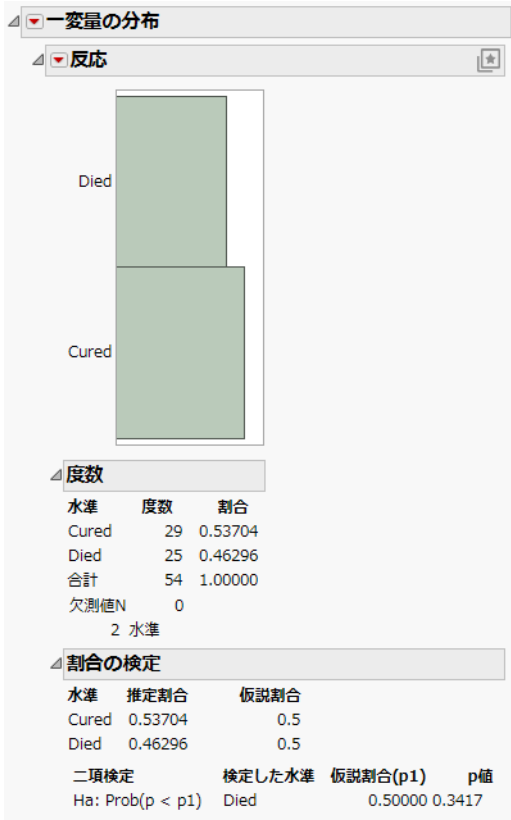
完了 ヘルプ

### 「割合の検定」レポートの作成

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

二項検定の正確な p 値が計算されます。

図 3.16 変数が2水準のときの「割合の検定」レポートの例



3 水準以上の割合の検定例

この例では、「一変量の分布」プラットフォームを使い、水準数が3以上の変数に対して「割合の検定」レポートを作成します。

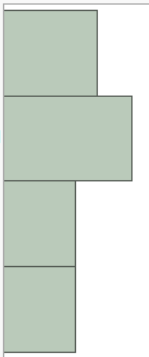
「割合の検定」レポートの開始

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

図 3.17 変数が3水準以上のときの「割合の検定」レポートオプション

一変量の分布

セル



度数

割合の検定

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

クリックして、仮説割合を入力してください。

割合の和が1になるようにスケール変更方法を選択。

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

完了 ヘルプ

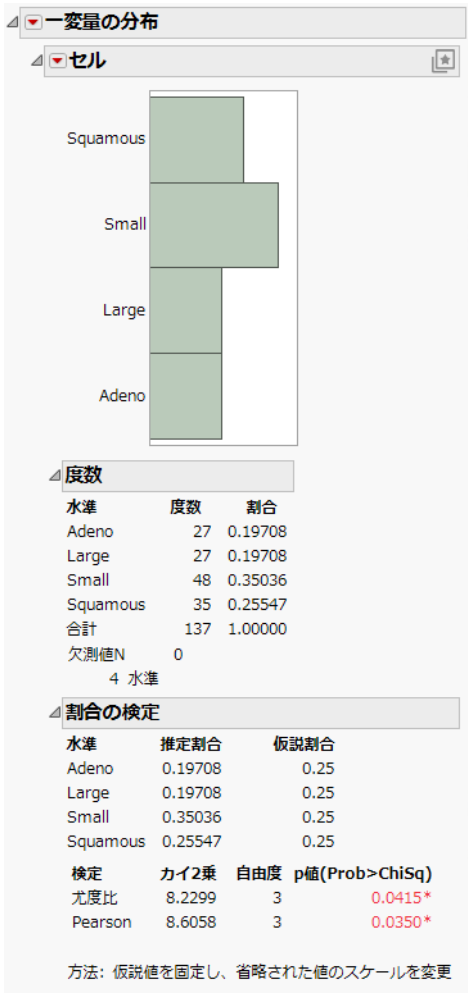
### 「割合の検定」レポートの作成

4水準の場合には、先ほど述べた操作例のあとに、次のような操作を行ってください。

1. 「仮説割合」のすべてのフィールドに「0.25」と入力します。
2. 「仮説値を固定し、省略された値のスケールを変更」を選択します。
3. 「完了」をクリックします。

尤度比検定とPearsonのカイ2乗検定が計算されます。

図 3.18 変数が3水準以上のときの「割合の検定」レポート



予測区間の例

この例では、「一変量の分布」レポートに予測区間を追加してみます。将来、観測される10個のオゾン量の観測値に関して、予測区間を計算したいとしましょう。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Cities.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「オゾン」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。
5. 「オゾン」の赤い三角ボタンをクリックし、[予測区間] を選択します。

図3.19 「予測区間」 ウィンドウ

予測区間として(1- $\alpha$ )を入力

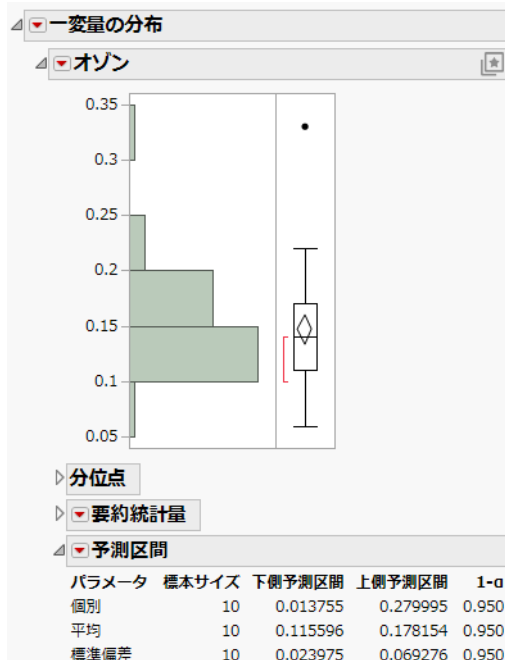
予測する標本サイズを入力

☒ 両側  
☐ 片側-下側限界  
☐ 片側-上側限界

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

7. [OK] をクリックします。

図3.20 「予測区間」 レポートの例



この例では、95%の信頼水準で、次のように結論することができます。

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

## 許容区間の例

この例では、「一変量の分布」レポートに許容区間を追加してみます。オゾン量の測定値の90%が含まれる区間を推定したいとしましょう。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Cities.jmp」を開きます。
2. [分析] > [一変量の分布] を選択します。
3. 「オゾン」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。
5. 「オゾン」の赤い三角ボタンをクリックし、[許容区間] を選択します。

図3.21 「許容区間」ウィンドウ

母集団のうち少なくとも指定の割合を含む (1- $\alpha$ )信頼区間を計算します。

(1- $\alpha$ )信頼区間を指定:

含める割合を指定:

☒ 両側  
☐ 片側－下側限界  
☐ 片側－上側限界

手法

☒ 正規分布を仮定  
☐ ノンパラメトリック

OK キャンセル ヘルプ

6. デフォルトの値をそのまま使用し、[OK] をクリックします。

図3.22 「許容区間」レポートの例

許容区間			
割合	下側許容限界	上側許容限界	1- $\alpha$
0.900	0.057035	0.236715	0.950

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

## 工程能力分析の例

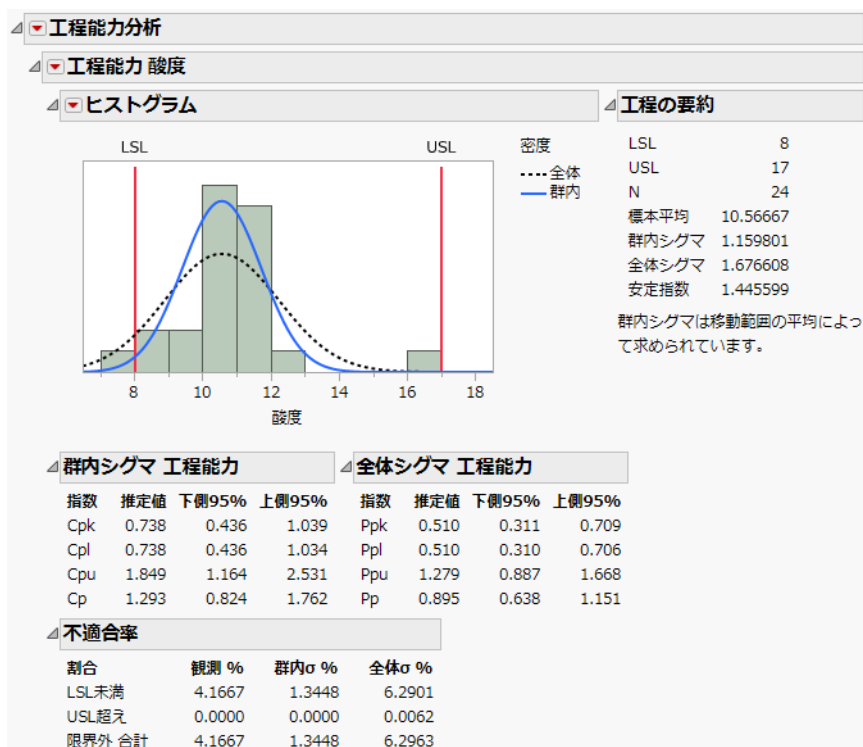
この例では、「一変量の分布」レポートに工程能力分析の結果を追加してみます。ピクルスの酸度を調べたいとしましょう。この例では、下限仕様限界は8、上限仕様限界は17とします。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Quality Control」フォルダ内の「Pickles.jmp」を開きます。



2. [分析] > [一変量の分布] を選択します。
3. 「酸度」を選択し、[Y, 列] をクリックします。
4. [OK] をクリックします。
5. 「酸度」の赤い三角ボタンをクリックし、[工程能力] を選択します。
6. 「LSL」(下側仕様限界)に「8」と入力します。
7. 「USL」(上側仕様限界)に「17」と入力します。
8. [OK] をクリックします。

図 3.23 「工程能力分析」レポートの例



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

## 検出限界の例

この例では、異なる検出限界が「一変量の分布」レポートの工程能力分析にどのように影響するかを説明します。この例で使用するデータテーブルには、実際の値が含まれている列と、その実際の値に対して異なる検出下限がある4つの列があります。

**メモ:** 現実の場面では、測定値の列が1列だけあり、そこに決まった検出限界が1つだけ設定されているのが普通ですが、この例では、検出限界を説明するために、異なる検出限界を設定した複数の測定値の列を用意しています。

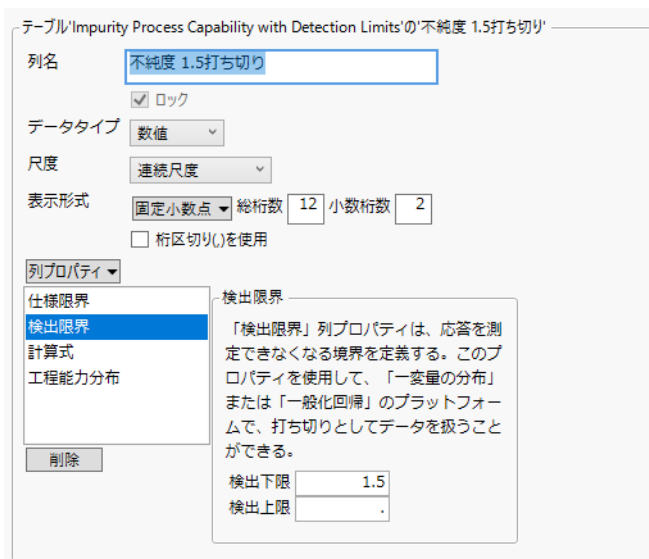
1. [ヘルプ]>[サンプルデータフォルダ]を選択し、「Impurity Process Capability with Detection Limits.jmp」を開きます。

このデータテーブルには、異なる検出下限を含む列が4つあります。たとえば、「不純度 1.5 打ち切り」列には、「不純度 実際の値」列の値が入っていますが、1.5 未満の値はすべて 1.5 に変更されています。また、この列には「検出限界」列プロパティがあり、そこで検出下限（測定が可能な応答の最小値）が 1.5 と指定されています。

2. (オプション)「不純度 1.5 打ち切り」列の「検出限界」列プロパティを見てみましょう。

1. データテーブルウィンドウの「列」パネルにおいて、「不純度 1.5 打ち切り」列の横にあるアスタリスクをクリックします。
2. [検出限界] を選択します。

図 3.24 「不純度 1.5 打ち切り」列の「列情報」ウィンドウ



3. [OK] をクリックします。
3. [分析] > [一変量の分布] を選択します。

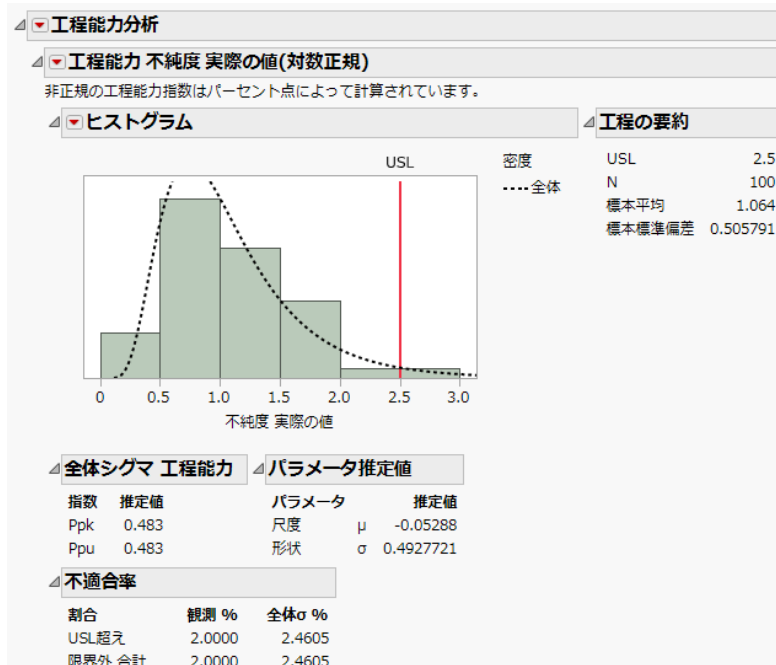
4. 「不純度 実際の値」を選択して【Y, 列】をクリックします。
5. 【工程能力分析の実行】オプションの選択を取り消します。

**メモ:**「検出限界」列プロパティがある列には、デフォルトの工程能力分析は使用できません。列に「検出限界」列プロパティがある場合は、まず連続変数の分布をあてはめ、あてはめのレポートに表示される赤い三角ボタンのメニューから【工程能力】を選択します。この例では、列スイッチャーを使って検出限界のある列を入力します。

6. 【ヒストグラムのみ】オプションを選択します。
7. 【OK】をクリックします。
8. 「不純度 実際の値」の赤い三角ボタンをクリックし、【連続分布のあてはめ】 > 【対数正規のあてはめ】を選択します。
9. 「対数正規分布のあてはめ」の赤い三角ボタンをクリックし、【工程能力】を選択します。  
「工程能力分析」ウィンドウが開き、「USL」が2.5と表示されます。これは、「不純度 実際の値」列の「仕様限界」列プロパティから取った値です。
10. 【OK】をクリックします。
11. (オプション)「不純度 実際の値」の赤い三角ボタンをクリックし、【ヒストグラムオプション】 > 【ヒストグラム】を選択します。

先頭のヒストグラムを非表示にすると、工程能力分析を表示するスペースができます。

図 3.25 「不純度 実際の値」の「工程能力分析」レポート



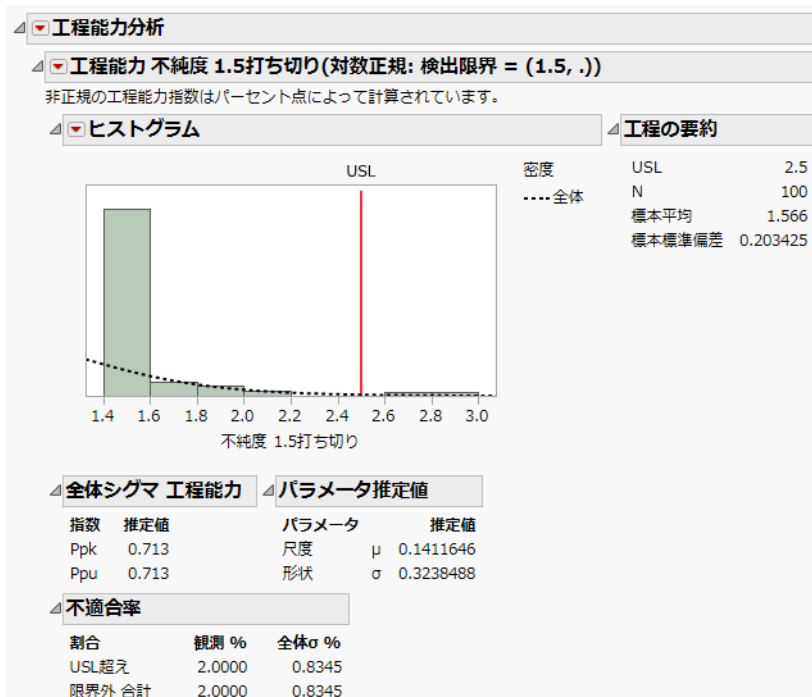
12. 「一変量の分布」の赤い三角ボタンをクリックし、[やり直し] > [列スイッチャー] を選択します。

列スイッチャーを使うと、検出限界としていろいろな値を設定した場合に、工程能力分析にどう影響するかを調べることができます。

13. 5列すべてを選択し、[OK] をクリックします。

14. 列スイッチャーで「不純度 1.5 打ち切り」を選択します。

図3.26 検出限界を 1.5 としたときの不純度の「工程能力」レポート



工程能力分析は、1.5 以下の観測値が打ち切られていることを考慮し、打ち切りのある対数正規分布をあてはめます。「工程能力」の下にあるレポートのタイトルには、分析で使われている検出限界の値が含まれています。

15. (オプション) 列スイッチャーで他の列を選択し、検出下限の値によって工程能力レポートがどのように変化するかを調べましょう。

この例では、検出下限を変化させて打ち切りの頻度を高めることで、工程能力分析にどのような影響が及ぶかを調べることができます。なお、前述したように、現実の場面では、測定値の列や検出限界は 1 つしかないのが普通です。

## 「一変量の分布」プラットフォームの統計的詳細

この節では、「一変量の分布」プラットフォームで使われている統計手法について説明します。

- [「標準誤差バーの統計的詳細」](#)
- [「分位点の統計的詳細」](#)
- [「要約統計量の統計的詳細」](#)
- [「正規分位点プロットの統計的詳細」](#)
- [「Wilcoxonの符号付順位検定の統計的詳細」](#)
- [「標準偏差の検定の統計的詳細」](#)
- [「正規分位点の統計的詳細」](#)
- [「\[標準化されたデータの保存\] の統計的詳細」](#)
- [「スコア信頼区間の統計的詳細」](#)
- [「予測区間の統計的詳細」](#)
- [「許容区間の統計的詳細」](#)
- [「連続分布のあてはめの統計的詳細」](#)
- [「離散分布のあてはめの統計的詳細」](#)

### 標準誤差バーの統計的詳細

「一変量の分布」プラットフォームのヒストグラムに表示される標準誤差バーは、 $\sqrt{np_i(1-p_i)}$  で求めた標準誤差を表しています。この式で、 $p_i = n_i/n$  です。

### 分位点の統計的詳細

ここでは、「一変量の分布」プラットフォームでの分位点の計算方法について説明します。

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

- $r \geq n$  の場合、最大値を  $p$  パーセント点とします。
- $r \leq 1$  の場合、最小値を  $p$  パーセント点とします。
- $r$  が整数のときは、その順位に相当するデータ値を  $p$  パーセント点とします。
- $r$  が整数でないときは、補間を行って  $p$  パーセント点を求めます。 $p$  パーセント点 ( $q_p$ ) は次の式で計算されます。

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

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

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

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

## 要約統計量の統計的詳細

ここでは、一変量の分布の「要約統計量」レポートに表示される個々の統計量について説明します。平均・分散・歪度・尖度は、データを表す確率分布の最初の4つのモーメント（積率）です。**モーメント**（積率）は、分布の位置・尺度・形状などを表す統計量です。

### 平均

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

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

### 標準偏差

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

$$s = \sqrt{s^2}$$

ここで、

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

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

### 平均の標準誤差

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

## 歪度

平均周りの3次のモーメントを表す指標です。次式によって計算されます。

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

上の式で、 $w_i$ は重みを表す係数（重み付けをしない場合は1）です。

## 尖度

平均周りの4次のモーメントを表す指標です。次式によって計算されます。

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

ここで、 $w_i$ は重みです（重み付けをしない場合、すべての重みは1です）。この式によって求められた尖度は、正規分布の場合には0となります。このように正規分布のときに0となるように求められた尖度は、「過剰尖度」（excess kurtosis）と呼ばれることがあります。

## 正規分位点プロットの統計的詳細

「一変量の分布」プラットフォームの正規分位点プロットでは、各値の経験累積確率が次のような式で計算されます。

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

ここで、 $r_i$ はデータ値の順位、 $N$ は非欠測値の個数を表します。

正規分位点の値は次のような式で計算されます。

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

ここで、 $\Phi$ は正規分布の累積確率分布関数です。

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

---

**メモ:** 正規分布の分位点ではなく、あてはめた確率分布の分位点を用いれば、その確率分布の分位点プロットとなります。

---

## Wilcoxon の符号付順位検定の統計的詳細

「一変量の分布」プラットフォームの Wilcoxon の符号付順位検定は、1 母集団、または対応のあるデータに対して用いることができ、分布が対称な場合には中央値に対する検定となります。対応のあるデータに対する検定は、それらの差に対する検定と等価です。この検定が中央値に対する検定となるには、母集団分布が対称でなければいけません。

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

### 1 標本の Wilcoxon 符号付順位検定

- $N$  個の観測値を、以下のように表すとしてします。

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

- Wilcoxon 符号付き順位検定の帰無仮説は以下ようになります（なお、母集団分布が対称であるという前提のもとでは、中央値に対する検定になります）。

$$H_0: X \text{ の分布は } m \text{ について対称である}$$

- 内部的な計算では、次式による観測値と仮説値  $m$  との差が使われます。

$$D_j = X_j - m$$

### 対応のあるデータに対する Wilcoxon 符号付順位検定

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

- 2 つの変数のそれぞれに、 $N$  個の観測値があるとします。

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

- Wilcoxon 符号付き順位検定の帰無仮説は以下ようになります（なお、母集団分布が対称であるという前提のもとでは、中央値に対する検定になります）。

$$H_0: X - Y \text{ の分布は } 0 \text{ について対称である}$$

- 1 標本の問題とするには、次のような計算式により観測値間の差を計算します。

$$D_j = X_j - Y_j$$

### Wilcoxon 符号付順位検定の統計量

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

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



以下において、差  $D_j$  の順位（中間順位）を  $R_j$  と記します。 $D_j$  の符号付順位を、次のように定義します。

- 差  $D_j$  が正の場合、符号付順位を  $R_j$  とします。
- 差  $D_j$  が 0 の場合、符号付順位を 0 とします。
- 差  $D_j$  が負の場合、符号付順位を  $-R_j$  とします。

符号付順位の統計量は次のように計算されます。

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

次のように定義します。

$d_0$  は、0 と等しい観測値の個数。

$R^+$  は、正の符号付順位の合計。

このとき、次式が成立します。

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

### Wilcoxon の符号付順位検定の p 値

$N \leq 20$  の場合は、正確な  $p$  値が計算されます。

$N > 20$  の場合は、以下のように、Student の  $t$  分布に近似して計算されます。同順位のペアに対して、修正が行われることに注意してください。Iman (1974) および Lehmann and D'Abrera (2006) を参照してください。

帰無仮説のもとでは、 $S$  の平均はゼロです。 $S$  の分散は次式により計算されます。

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

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

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

$$t = \frac{S}{\sqrt{\frac{N \cdot \text{Var}(S) - S^2}{N - 1}}}$$

## 標準偏差の検定の統計的詳細

「一変量の分布」プラットフォームの標準偏差の検定では、検定統計量が次の式で計算されます。

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

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

「最小  $p$  値」は、両側検定の  $p$  値で、次式で計算されます。

$$2 \times \min(p1, p2)$$

ここで、 $p1$  は下側の片側検定の  $p$  値で、 $p2$  は上側の片側検定の  $p$  値です。

## 正規分位点の統計的詳細

「一変量の分布」プラットフォームの正規分位点は、次の式で計算されます。

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

ここで

$\Phi$  は、正規分布の累積確率分布関数

$r_i$  は、 $i$  番目の観測値の順位

$N$  は、欠測値でない観測値の個数（標本サイズ）

## 【標準化されたデータの保存】の統計的詳細

「一変量の分布」プラットフォームの標準化されたデータは、次の式で計算されます。

$$\frac{X - \bar{X}}{S_X}$$

ここで

$X$  は、元のデータ値

$\bar{X}$  は、平均

$S_X$  は、列  $X$  の標準偏差

## スコア信頼区間の統計的詳細

「一変量の分布」プラットフォームは、カテゴリカルな応答変数の各水準におけるスコア信頼区間（スコア法に基づいて近似した信頼区間）を計算できます。応答の各水準について、次式によってスコア信頼区間は計算されます。

$$\frac{y_i + \frac{z^2}{2} \pm z \sqrt{\frac{y_i(n-y_i)}{n} + \frac{z^2}{4}}}{n + z^2}$$

ここで

$y_i$ は、応答の  $i$  番目の水準における標本サイズ

$n$ は、全体の標本サイズ

$z$ は、標準正規分布の  $(\alpha/2)$  の分位点

Agresti and Coull (1998) を参照してください。

## 予測区間の統計的詳細

「一変量の分布」プラットフォームの予測区間は、次のように定義されています。

- 将来の  $m$  個の観測値に対する予測区間

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

- 将来の  $m$  個の観測値から計算される平均に対する予測区間

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

- 将来の  $m$  個の観測値から計算される標準偏差に対する予測区間

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

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

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

Meeker et al. (2017, ch. 4) を参照してください。

## 許容区間の統計的詳細

ここでは、「一変量の分布」プラットフォームにおける片側許容区間と両側許容区間の統計的詳細について説明します。

### 正規分布に基づく許容区間

#### 片側許容区間

許容区間は、片側区間の場合、次式によって計算されます。

$$\text{下側の場合} = \bar{x} - k's$$

$$\text{上側の場合} = \bar{x} + k's$$

ここで、

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

$s$ は、標準偏差。

$t$ は、非心 $t$ 分布の分位点です。

$\Phi^{-1}$ は、標準正規分布の分位点です。

#### 両側許容区間

両側許容区間は、次の式で計算されます。

$$[T_{p_L}, T_{p_U}] = [\bar{x} - k_{(1-\alpha/2;p,n)}s, \bar{x} + k_{(1-\alpha/2;p,n)}s]$$

上の式で、 $s$ は標準偏差、 $k_{(1-\alpha/2;p,n)}$ は定数です。

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

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

$\Phi$ は標準正規分布の累積分布関数を表します。

そのため、 $k$ は次式を解くことにより計算されます。

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

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

正規分布に基づく許容区間の詳細については、Meeker et al. の表 J.1a、J.1b、J.6a、J.6b (2017) を参照してください。

## ノンパラメトリックな許容区間

### 下方片側の許容限界

信頼水準  $100(1 - \alpha)\%$  の下方片側許容限界 (標本サイズ  $n$  の標本分布において割合  $\beta$  以上のデータを含むもの) は、順序統計量  $x_{(l)}$  です。この順位  $l$  は、次のように計算されます。

$$l = n - \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$$

ここで、 $\Phi_{bin}^{-1}(1 - \alpha, n, \beta)$  は、試行  $n$  回、成功確率  $\beta$  の二項分布の第  $(1 - \alpha)$  分位点です。

実際の信頼水準は、 $\Phi_{bin}(n - l, n, \beta)$  で計算されます。ここで、 $\Phi_{bin}(x, n, \beta)$  は、試行  $n$  回、成功確率  $\beta$  の二項分布において、確率変数が  $x$  以下になる確率です。

なお、分布によらない下方片側許容限界を計算するには、標本サイズ  $n$  が  $(\log \alpha) / (\log \beta)$  以上であることが前提となります。

### 上方片側の許容区間

信頼水準  $100(1 - \alpha)\%$  である上方片側許容限界 (標本サイズ  $n$  の標本分布において割合  $\beta$  以上のデータを含むもの) は、順序統計量  $x_{(u)}$  です。この順位  $u$  は、次のように計算されます。

$$u = 1 + \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$$

ここで、 $\Phi_{bin}^{-1}(1 - \alpha, n, \beta)$  は、試行  $n$  回、成功確率  $\beta$  の二項分布の第  $(1 - \alpha)$  分位点です。

実際の信頼水準は、 $\Phi_{bin}(u - 1, n, \beta)$  で計算されます。ここで、 $\Phi_{bin}(x, n, \beta)$  は、試行  $n$  回、成功確率  $\beta$  の二項分布において、確率変数が  $x$  以下になる確率です。

なお、分布によらない上方片側許容限界を計算するには、標本サイズ  $n$  が  $(\log \alpha) / (\log \beta)$  以上であることが前提となります。

### 両側許容区間

信頼水準  $100(1 - \alpha)\%$  の両側許容区間 (標本サイズ  $n$  の標本分布において割合  $\beta$  以上のデータを含むもの) は、次のように計算されます。

$$[\tilde{T}_{p_L}, \tilde{T}_{p_U}] = [x_{(l)}, x_{(u)}]$$

ここで、 $x_{(i)}$  は  $i$  番目の順序統計量で、 $l$  と  $u$  は次のように計算されます。

$v = n - \Phi_{\text{bin}}^{-1}(1 - \alpha, n, \beta)$  としましょう。ここで、 $\Phi_{\text{bin}}^{-1}(1 - \alpha, n, \beta)$  は、試行  $n$  回、成功確率  $\beta$  の二項分布の第  $(1 - \alpha)$  分位点です。 $v$  が 2 未満の場合、分布によらない両側許容区間は計算できません。 $v$  が 2 以上の場合、 $l = \text{floor}(v/2)$ 、 $u = \text{floor}(n + 1 - v/2)$  です。

実際の信頼水準は、 $\Phi_{\text{bin}}(u - l - 1, n, \beta)$  で計算されます。ここで、 $\Phi_{\text{bin}}(x, n, \beta)$  は、試行  $n$  回、成功確率  $\beta$  の二項分布において、確率変数が  $x$  以下になる確率です。

なお、分布によらない両側許容区間を計算するには、標本サイズが  $n$  が、次式を満たすような  $n$  以上であることが前提となります。

$$1 - \alpha = 1 - n\beta^{n-1} + (n-1)\beta^n$$

分布によらない許容区間については、Meeker et al. (2017, sec. 5.3) を参照してください。

## 連続分布のあてはめの統計的詳細

ここでは、「一変量の分布」プラットフォームの「連続分布のあてはめ」メニューにあるオプションについて説明します。特に指定をしない場合、パラメータ推定値の信頼区間は最尤法を用いて計算されます。最尤法による信頼区間の詳細については、『予測モデルおよび発展的なモデル』を参照してください。Y 列に「検出限界」列プロパティがある場合に「連続分布のあてはめ」のオプションを使用すると、打ち切りを考慮して確率分布があてはめられます。この時、JMP で用意されている確率分布のいくつかだけしか、打ち切りのあるデータをサポートしていません。打ち切りのあるデータへの分布のあてはめについて詳しくは、Meeker and Escobar (1998) を参照してください。

## 正規のあてはめ

「正規のあてはめ」オプションは、正規分布の 2 つのパラメータを推定します。

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

標準正規分布では、 $\mu = 0$ 、 $\sigma = 1$  です。

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

$$E(x) = \mu$$

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

---

**メモ:** 「正規分布のあてはめ」での平均の信頼区間も、通常の方法である  $t$  分布に基づく方法に戻されました。「正規分布のあてはめ」での標準偏差の信頼区間も、通常の方法である  $\chi^2$  分布に基づく方法に戻されました。

---

## Cauchyのあてはめ

〔Cauchyのあてはめ〕 オプションは、位置 $\mu$ 、尺度 $\sigma$ のCauchy分布をあてはめます。

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

$$E(x) = \text{未定義}$$

$$\text{Var}(x) = \text{未定義}$$

## Studentのtのあてはめ

〔Studentのtのあてはめ〕 オプションは、位置 $\mu$ 、尺度 $\sigma$ 、自由度 $\nu$ のStudentのt分布をあてはめます。

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

$$\text{ただし } -\infty < x < \infty; -\infty < \mu < \infty; 0 < \sigma; 1 \leq \nu$$

$$E(x) = \mu \quad \text{ただし } 1 < \nu \text{ に限る}$$

$$\text{Var}(x) = \sigma^2 \nu / (\nu - 2) \quad \text{ただし } 2 < \nu \text{ に限る}$$

---

✖モ:  $\nu = 1$  の場合、Studentのt分布はCauchy分布と等しくなります。

---

## SHASHのあてはめ

〔SHASHのあてはめ〕 オプションは、sinh-arcsinh分布（正弦-逆正弦分布；SHASH分布）をあてはめます。SHASH分布は、正規分布に簡単に変換することができ、また正規分布を特殊なケースとして含んでいます。SHASH分布は、対称な分布と非対称な分布を含んでいます。SHASH分布の形状は、2つの形状パラメータ（ $\gamma$ と $\delta$ ）によって決まります。SHASH分布の詳細については、Jones and Pewsey（2009）を参照してください。

$$\text{確率密度関数: } f(x) = \frac{\delta \cosh(w)}{\sqrt{\sigma^2 + (x - \theta)^2}} \phi[\sinh(w)] \quad \text{ただし } -\infty < \gamma, x, \theta < \infty; 0 < \delta, \sigma$$

ここで、

$\phi(\cdot)$  は、標準正規分布の確率密度関数です。

$$w = \gamma + \delta \sinh^{-1} \left( \frac{x - \theta}{\sigma} \right)$$

- $\gamma = 0$ 、 $\delta = 1$  の SHASH 分布は、位置、尺度がそれぞれ  $\theta$ 、 $\sigma$  の正規分布と同じです。
- 変換  $\sinh(w)$  は通常、 $\mu = 0$ 、 $\sigma = 1$  の正規分布に従います。

## ゼロ強調 SHASH のあてはめ

[ゼロ強調 SHASH のあてはめ] オプションは、ゼロ強調の  $\sinh$ -arcsinh 分布（正弦-逆正弦分布；SHASH 分布）をあてはめます。ゼロ強調の SHASH 分布とは、観測値が 0 において確率質量をもつ SHASH 分布です。SHASH 分布は、対称な分布と非対称な分布を含んでいます。

$$\text{確率密度関数: } f(x) = \begin{cases} \pi + (1 - \pi) \frac{\delta \cosh(w)}{\sqrt{\sigma^2 - \theta^2}} \phi[\sinh(w)] & \text{この式で } x = 0 \\ (1 - \pi) \frac{\delta \cosh(w)}{\sqrt{\sigma^2 + x - \theta^2}} \phi[\sinh(w)] & \text{この式で } x \neq 0 \end{cases}$$

ただし  $-\infty < \gamma, x, \theta < \infty$ ;  $0 < \delta, \sigma$

ここで、

$\phi(\cdot)$  は、標準正規分布の確率密度関数です。

$$w = \gamma + \delta \sinh^{-1} \left( \frac{x - \theta}{\sigma} \right)$$

---

**メモ:** ゼロ強調の SHASH 分布のパラメータ推定値の信頼区間は、Wald 法によって計算されます。

---

## 指数のあてはめ

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

$\beta$  が 1 である Weibull 分布は、指数分布です（このとき、 $\alpha = \sigma$  と対応しています）。また、 $\alpha$  が 1 であるガンマ分布も、指数分布です。

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

$$E(x) = \sigma$$

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

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



## ExGauss のあてはめ

[ExGauss のあてはめ] オプションは、正規分布の確率変数と指数分布の確率変数を足した和が従う確率分布をあてはめます。ExGauss 分布（指数修正 Gauss 分布）の推定では、正規分布部分の位置パラメータ（ $\mu$ ）、尺度パラメータ（ $\sigma$ ）、そして指数分布部分のパラメータ  $\lambda$  の推定値が求められます。

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

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

上の式で、 $\Phi(\cdot)$  は標準正規分布の累積分布関数です。

$$E(x) = \mu + 1/\lambda$$

$$\text{Var}(x) = \sigma^2 + 1/\lambda^2$$

ExGauss 分布の詳細については、Ament et al. (2019) と Palmer et al. (2011) を参照してください。指数分布部分のパラメータ表現は、文献によって異なります。[ExGauss のあてはめ] オプションで用いているパラメータは、「一変量の分布」プラットフォームの [指数のあてはめ] オプションで使用されているパラメータの逆数になっている点に注意してください。

## ガンマのあてはめ

[ガンマのあてはめ] オプションを選択すると、ガンマ分布があてはめられ、 $\alpha > 0$ 、 $\sigma > 0$  というパラメータが推定されます。 $\alpha$  は分布の形状を、 $\sigma$  は分布の尺度を表します。データはゼロより大きくなければなりません。

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

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

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

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

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

## 対数正規のあてはめ

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

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

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

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

## Weibull のあてはめ

Weibull 分布は、 $\alpha$ （尺度パラメータ）と  $\beta$ （形状パラメータ）によって決められる分布です。特に機械装置や生物学上の寿命を推定するときのモデルとしてよく使用されます。

Weibull 分布の確率密度関数は、次のように定義されます。

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

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

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

ここで、 $\Gamma(\cdot)$  はガンマ関数です。

## 二重正規混合のあてはめと三重正規混合のあてはめ

〔二重正規混合のあてはめ〕 は2つの正規分布の混合分布を、〔三重正規混合のあてはめ〕 は3つの正規分布の混合分布をあてはめます。二峰性や多峰性のデータにも対応した柔軟な分布です。グループごとに1つずつ個別の平均、標準偏差、全体に占める割合が推定されます。以下の式で、 $k$  は混合分布に含まれる正規分布の数です。

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

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

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

ここで、 $\mu_i$ ,  $\sigma_i$ ,  $\pi_i$ は、 $i$ 番目のグループのそれぞれ平均・標準偏差・割合であり、 $\phi(\cdot)$ は、標準正規分布の確率密度関数です。

---

**メモ:** 正規混合分布のパラメータ推定値の信頼区間は、Wald法によって計算されます。

---

## Johnsonのあてはめ

[Johnsonのあてはめ] オプションは、Johnsonの分布システムのうち最適な分布を選んであてはめます。Johnsonの分布システムは、変換すると正規分布になる3つの分布で構成されています。3つの分布は、次のとおりです。

- Johnson Su (この分布は、有界ではありません)
- Johnson Sb (この分布は、パラメータにより定義される有界を上下に持ちます)
- Johnson Sl (この分布は、パラメータにより定義される有界を上下のいずれかに持つ対数正規分布です)

選択された分布のあてはめに関するレポートのみが作成されます。Johnson分布の選択方法とパラメータ推定については、Slifker and Shapiro (1980)を参照してください。パラメータ推定では、最尤法を使用しません。

Johnson分布システムは、柔軟性が高いことで人気があります。これら3つで構成される分布システムは、歪度と尖度のあらゆる組み合わせに対応しており、柔軟にデータにあてはまります。しかし、SHASH分布も非常に柔軟性が高いので、Johnson分布よりもSHASH分布をお勧めします。

Zを標準正規変量とすると、分布システムは次のように定義されます。

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

Johnson Suの場合は、

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

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

Johnson Sbの場合は、

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

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

Johnson SI の場合は、 $\sigma = \pm 1$  で

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

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

### Johnson Su

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

### Johnson Sb

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

### Johnson SI

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

ここで、 $\phi(\cdot)$  は標準正規分布の確率密度関数です。

---

**メモ:** Johnson 分布のパラメータ推定値の信頼区間は、Wald 法によって計算されます。

---

## ベータのあてはめ

(0, 1) の区間を範囲とする標準ベータ分布は、割合などの0から1の範囲に収まるデータに対してよく使われています。[ベータのあてはめ] オプションを選択すると、ベータ分布の2つの形状パラメータ  $\alpha > 0$ 、 $\beta > 0$  が推定されます。ベータ分布は、0から1の範囲外の値はとりません。

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

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

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

ここで、 $B(\cdot)$  はベータ関数です。

## すべてをあてはめ

このオプションを選択すると、用意されているほとんどすべての分布があてはめられ、「分布の比較」レポートにおいて、AICcが小さい順に、それらの分布の名前が表示されます。チェックボックスで分布を選択することによって、その分布のレポート、およびヒストグラム上に描かれる密度曲線の表示／非表示を切り替えることができます。

AICcとBICの計算式は以下のとおりです。

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

$$\text{BIC} = -2\log L + k\ln(n)$$

ここで

- $\log L$  は、対数尤度。
- $n$  は、標本サイズ。
- $k$  は、パラメータの個数。

「AICc 重み」列には、AICcの合計が1になるように正規化された値が示されます。AICc 重みは、あてはめた複数の分布のいずれかが真である場合に特定の分布が真である確率と解釈できます。そのため、AICc 重みが1に最も近い分布が良い分布となります。AICc 重みは、複数のモデルのAICcから、次のように算出されます。

$$\text{AICcの重み} = \exp[-0.5(\text{AICc} - \min(\text{AICc}))] / \sum(\exp[-0.5(\text{AICc} - \min(\text{AICc}))])$$

上の式で、 $\min(\text{AICc})$  は、あてはめた分布の中で最も小さいAICc値です。

「分布の比較」レポートに示される指標については、『基本的な回帰モデル』を参照してください。

## 離散分布のあてはめの統計的詳細

ここでは、「一変量の分布」プラットフォームの「離散分布のあてはめ」メニューにあるオプションについて説明します。特に指定をしない場合、パラメータ推定値の信頼区間是最尤法を用いて計算されます。最尤法による信頼区間の詳細については、『予測モデルおよび発展的なモデル』を参照してください。

### Poissonのあてはめ

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

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

$$E(x) = \lambda$$

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

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

## 負の二項のあてはめ

負の二項分布は、特定の失敗回数に達するまでの成功回数をモデル化するのに便利です。以下のパラメータ表現では、平均パラメータ $\lambda$ および過分散パラメータ $\sigma$ が含まれています。

$$\text{確率関数: } \frac{\Gamma[x + (1/\sigma)]}{\Gamma[x + 1]\Gamma[1/\sigma]} \left[ \frac{(\lambda\sigma)^x}{(1 + \lambda\sigma)^{x + (1/\sigma)}} \right], x = 0, 1, 2, \dots$$

$$E(x) = \lambda$$

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

ここで、 $\Gamma(\cdot)$ は、ガンマ関数です。

## 負の二項分布とガンマ Poisson 分布の関係

負の二項分布はガンマ Poisson 分布と同等です。ガンマ Poisson 分布は、平均 $\mu$ が異なる複数の Poisson 分布から、データが生成されているときに役立ちます。

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

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

$$E(x) = \lambda$$

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

ここで、 $\Gamma(\cdot)$ は、ガンマ関数です。

ガンマ Poisson 分布は、 $\sigma_{\text{negbin}} = (\sigma_{\text{gp}} - 1) / \lambda_{\text{gp}}$ の負の二項分布と同等です。

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

## ゼロ強調 Poisson のあてはめ

ゼロ強調 Poisson 分布には、尺度パラメータ ( $\lambda > 0$ ) と、ゼロ強調パラメータ ( $\pi$ ) があります。

$$\text{確率関数: } \begin{cases} \pi + (1 - \pi)\exp[-\lambda] & \text{この式で } x = 0 \\ (1 - \pi)\frac{\lambda^x}{x!}\exp[-\lambda] & \text{この式で } x = 1, 2, \dots \end{cases}$$

$$E(x) = (1 - \pi)\lambda$$

$$\text{Var}(x) = \lambda(1 - \pi)(1 + \lambda\pi)$$

### ゼロ強調 負の二項のあてはめ

ゼロ強調 負の二項分布には、尺度パラメータ ( $\lambda > 0$ )、過分散パラメータ ( $\sigma > 0$ )、ゼロ強調パラメータ ( $\pi$ ) があります。

$$\text{確率関数: } \begin{cases} \pi + (1 - \pi)(1 + \lambda\sigma)^{-(1/\sigma)}, & \text{この式で } x = 0 \\ (1 - \pi) \frac{\Gamma[x + (1/\sigma)]}{\Gamma[x + 1]\Gamma[1/\sigma]} \left[ \frac{(\lambda\sigma)^x}{(1 + \lambda\sigma)^{x + (1/\sigma)}} \right], & \text{この式で } x = 1, 2, \dots \end{cases}$$

$$E(x) = (1 - \pi)\lambda$$

$$\text{Var}(x) = \lambda(1 - \pi)[1 + \lambda(\sigma + \pi)]$$

### 二項のあてはめ

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

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

$$E(x) = np$$

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

上の式で、 $n$  は、独立した試行の回数を示します。

---

**メモ:** 二項分布のパラメータに対する信頼区間の計算には、プロファイル尤度法が使用されています。

---

### ベータ二項のあてはめ

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

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

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

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

$$E(x) = np$$

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

ここで、 $\Gamma(\cdot)$ は、ガンマ関数です。

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

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

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

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

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

「Samples/Scripts」フォルダにある「demoBetaBinomial.jsl」を実行すると、 $n$ を20に固定しパラメータ $p$ を可変として、二項分布と過分散パラメータ $\delta$ のベータ二項分布を比較できます。

## ゼロ強調 二項のあてはめ

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

$$\text{確率関数: } \begin{cases} \pi + (1 - \pi)(1 - p)^n & \text{この式で } x = 0 \\ (1 - \pi) \binom{n}{x} p^x (1 - p)^{n - x} & \text{この式で } x = 1, 2, \dots, n \end{cases}$$



$$E(x) = (1-\pi)np$$

$$\text{Var}(x) = (1-\pi)[np(1-p) + n^2p^2] - [(1-\pi)np]^2$$

上の式で、 $n$  は、独立した試行の回数を示します。

### ゼロ強調 ベータ二項のあてはめ

[ゼロ強調 ベータ二項のあてはめ] オプションでは、一定の標本サイズか、標本サイズを含む列のいずれかを指定することができます。ゼロ強調のベータ二項分布は、 $n \geq 2$  の場合にのみ、あてはめることができます。

確率関数:

$$\left\{ \begin{array}{ll} \pi + (1-\pi) \frac{\Gamma\left(\frac{1}{\delta}-1\right)\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left(n+\frac{1}{\delta}-1\right)} & \text{この式で } x = 0 \\ (1-\pi)\binom{n}{x} \frac{\Gamma\left(\frac{1}{\delta}-1\right)\Gamma\left[x+p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-x+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left(n+\frac{1}{\delta}-1\right)} & \text{この式で } x = 1, 2, \dots, n \end{array} \right.$$

$$E(x) = (1-\pi)np$$

$$\text{Var}(x) = (1-\pi)np[(1-p)(1+(n-1)\delta) + np] - [(1-\pi)np]^2$$

上の式で、 $n$  は、独立した試行の回数を示します。

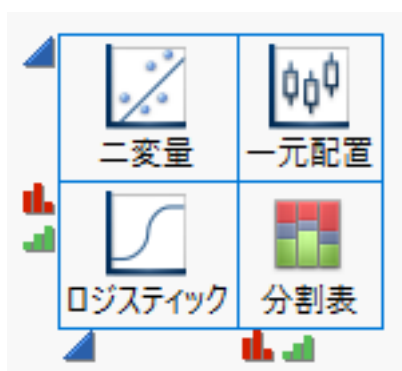


# 第4章

## 「二変量の関係」プラットフォームの概要 2変数の関係を調べる

変数のペアについての分析には、「二変量の関係」プラットフォームを使用してください。「二変量の関係」プラットフォームでは、散布図・線形回帰・分散分析・多重比較・ロジスティック回帰・分割表などのグラフや統計分析が利用できます。実行できる分析の種類は、変数の尺度によって異なります。

図4.1 「二変量の関係」プラットフォーム



「二変量の関係」起動ウィンドウは、次の4種類のプラットフォームのいずれかを開きます。

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



# 第5章

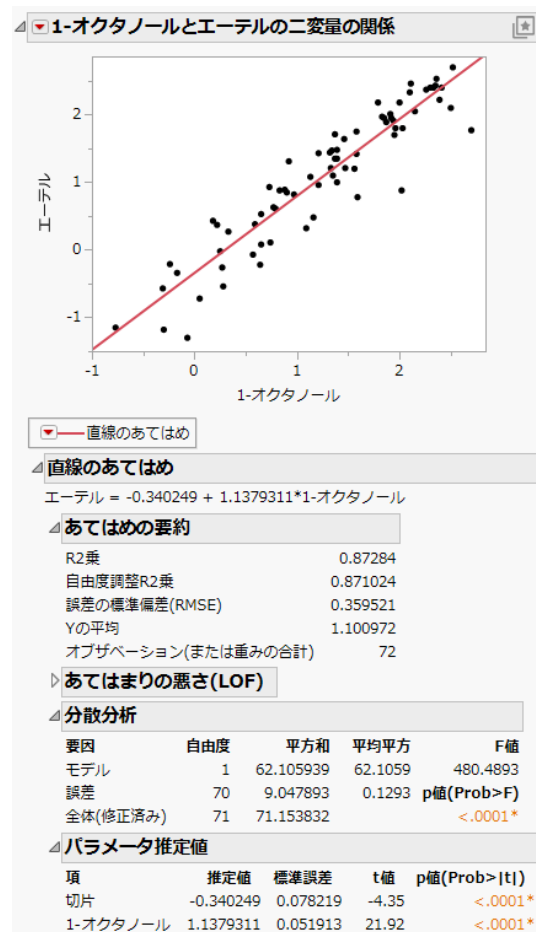
## 二変量分析

### 2つの連続変数の関係を調べる

「二変量」プラットフォームでは、2つの連続変数の関係を調べられます。散布図は、データを点でプロットしたグラフです。オプションを使って、単回帰直線や多項式回帰曲線、滑らかな曲線を図に加えることができます。

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

図 5.1 二変量分析の例



## 目次

「二変量」プラットフォームの概要	111
二変量分析の例	111
「二変量」プラットフォームの起動	113
データ形式	114
「二変量の関係」のレポート	114
「二変量」プラットフォームのオプション	115
二変量の関係のオプション	118
「二変量の関係」レポート	121
「平均のあてはめ」レポート	122
「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」レポート	123
「その他のあてはめ」のウィンドウ	128
「スプライン曲線のあてはめ」レポート	129
「局所的な平滑化」レポート	129
「値ごとのあてはめ」レポート	130
「直交のあてはめ」レポート	130
「ロバストなあてはめ」レポート	131
「Passing-Bablokのあてはめ」レポート	132
「確率楕円」レポート	133
ノンパラメトリック密度	135
「二変量」プラットフォームの別例	136
[その他のあてはめ] オプションの例	136
[直交のあてはめ] オプションの例	138
[ロバストなあてはめ] オプションの例	139
グループ別確率楕円の例	141
グループ別回帰直線の例	142
By 変数を使ってグループ別に分析する例	143
「二変量」プラットフォームの統計的詳細	144
[直線のあてはめ] オプションの統計的詳細	144
[多項式のあてはめ] オプションの統計的詳細	145
[スプライン曲線のあてはめ] オプションの統計的詳細	145
[直交のあてはめ] オプションの統計的詳細	146
[ロバスト] オプションの統計的詳細	147
[Passing Bablokのあてはめ] オプションの統計的詳細	148
「あてはめの要約」レポートの統計的詳細	148
「あてはまりの悪さ (LOF)」レポートの統計的詳細	149
「パラメータ推定値」レポートの統計的詳細	149
「平滑化スプライン曲線によるあてはめ」レポートの統計的詳細	150
「二変量正規楕円」レポートの統計的詳細	150

## 「二変量」プラットフォームの概要

「二変量」プラットフォームでは、モデルを対話的にあてはめ、散布図上で確認します。1つの図に複数のモデルを表示して比較することができます。

[二変量の関係] を選択して連続尺度のY変数と連続尺度のX変数を指定すると、「二変量」プラットフォームが開きます。このプラットフォームには、初め、X変数とY変数の各組み合わせの散布図が表示されます。散布図と赤い三角ボタンのオプションを使って、2つの連続変数の関係を対話的にモデル化しましょう。データに単回帰直線や複雑な回帰モデルをあてはめることができます。密度推定値を調べることもできます。

「二変量」プラットフォームのあてはめオプションは、回帰のあてはめと密度推定に分類されます。

表5.1 「二変量」プラットフォームのあてはめのカテゴリ

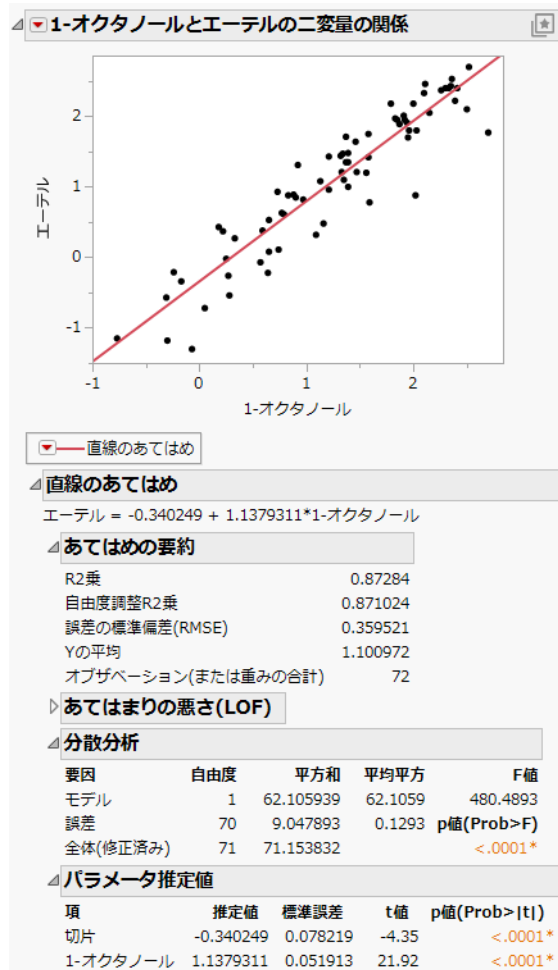
カテゴリ	説明	あてはめを行うオプション
回帰のあてはめ	回帰分析は、観測されたデータ点にモデルをあてはめます。この手法には、最小2乗のあてはめ、スプライン曲線のあてはめ、カーネル平滑化、直交のあてはめ、変換、ロバストなあてはめがあります。	<b>平均のあてはめ</b> <b>直線のあてはめ</b> <b>多項式のあてはめ</b> <b>その他のあてはめ</b> <b>ノンパラメトリックな曲線</b> <b>直交のあてはめ</b> <b>Passing Bablokのあてはめ</b> <b>ロバスト</b>
密度推定	密度推定では、二変量の分布が点にあてはめられます。楕円形の等高線で表される二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

## 二変量分析の例

この例では、2つの連続尺度の変数に対し、「二変量」プラットフォームを使って散布図を作成し、回帰直線をあてはめます。この例では、異なる溶剤で化合物の溶解度を測定したデータを使用し、2種類の溶剤の結果を比較します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Solubility.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「エーテル」を選択し、[Y, 目的変数] をクリックします。
4. 「1-オクタノール」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「二変量の関係」の赤い三角ボタンをクリックし、[直線のあてはめ] を選択します。

図 5.2 二変量の関係レポート



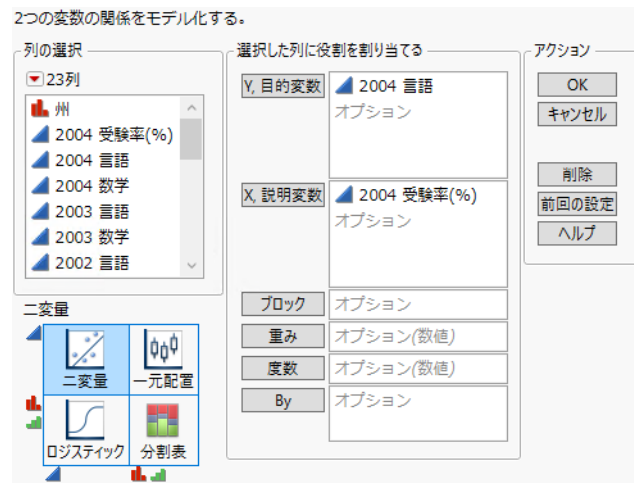
2つの溶剤の結果は類似性を示唆しています。線形モデルのR2乗値の0.87という数値は、応答のばらつきの87%がモデルによって説明されることを示します。なお、この例では、X変数とY変数の両方が、測定誤差をもつ測定値です。そのため、[直交のあてはめ] オプションまたは [Passing Bablokのあてはめ] オプションが適しているかもしれません。



## 「二変量」プラットフォームの起動

「二変量」プラットフォームを起動するには、[分析] > [二変量の関係] を選択します。「二変量の関係」の起動ウィンドウでは、4種類の分析が可能です。連続尺度のY変数と連続尺度のX変数を指定すると、「二変量」プラットフォームが開きます。

図5.3 「二変量の関係」起動ウィンドウ



「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。「二変量の関係」起動ウィンドウには、次のようなオプションがあります。

**Y, 目的変数** 分析したい目的変数（応答変数）。目的変数は従属変数とも呼ばれます。「二変量」の分析を行いたい場合は、連続尺度でなければなりません。

**X, 説明変数** 分析したい説明変数（予測変数）。説明変数は独立変数とも呼ばれます。連続尺度でなければなりません。

**ブロック** （二変量の分析には適用されません。）ブロック変数とする列。

**重み** データテーブルの各行に対する重みを含んだ列。値が0より大きい行だけが分析に用いられます。

**度数** 分析において各データ行に度数を割り当てます。分析対象のデータが要約されたものであるときに役立ちます。

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

## データ形式

「二変量」プラットフォームでは、要約されていないデータ、および要約されているデータのいずれも分析できます。

**要約されていないデータ** 観測値ごとに1行ずつあり、Xの値が1列、Yの値が1列に入力されています。

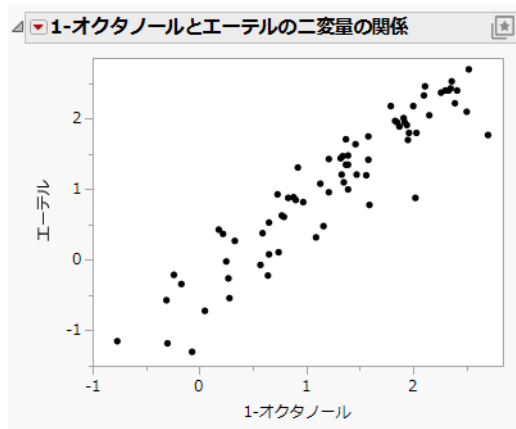
**要約されたデータ** XとYの値の組み合わせが同じである複数の観測値が1行にまとめられているデータです。データテーブルには、各行の観測値の数を示す度数列が必要です。起動ウィンドウで、この列を「度数」に指定してください。

**メモ:**「二変量の関係」起動ウィンドウでは、連続尺度・順序尺度・名義尺度の列が使用できます。[Y, 目的変数]の列と[X, 説明変数]の列がどちらも連続尺度である場合、「二変量」プラットフォームが起動します。それ以外の組み合わせでは、「一元配置」・「分割表」・「ロジスティック」のいずれかのプラットフォームが起動します。

## 「二変量の関係」のレポート

「二変量の関係」のレポートには、まず、XとYの組み合わせごとに散布図が描かれます。赤い三角ボタンのオプションを使うと、データにモデルがあてはめられ、統計的な結果のレポートが表示されます。「[「二変量」プラットフォームのオプション](#)」を参照してください。

図5.4 二変量に対する散布図



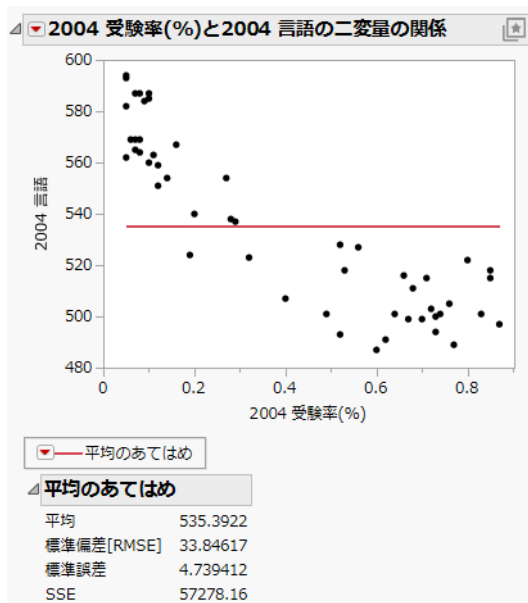
### 変数を対話的に置換する

プロット内で変数を対話的に置き換えるには、一方の軸の変数をドラッグしてもう一方の軸にドロップします。データテーブルの「列」パネルで該当する変数をクリックし、プロットの軸上までドラッグして変数を置換することもできます。

## 「二変量」プラットフォームのオプション

「二変量の関係」の赤い三角ボタンのメニューには、表示に関するオプション、あてはめを行うオプション、コントロールに関するオプションがあります。あてはめを行うオプションを選択すると、直線、曲線、分布が散布図に追加され、プロットの下には、あてはめごとに赤い三角ボタンのメニューが表示されます。レポートウィンドウには、各あてはめに固有のレポートが追加されます。

図5.5 [平均のあてはめ] オプションの例



**メモ:** Y変数またはX変数を複数指定した場合は、「あてはめのグループ」メニューが表示されます。「あてはめのグループ」メニューのオプションを使うと、レポートの配置を変更したり、適合度の順に並べたりできます。『基本的な回帰モデル』を参照してください。

「二変量の関係」の赤い三角ボタンをクリックすると、次のようなオプションが表示されます。

**点の表示** 散布図内の点の表示／非表示を切り替えます。

**ヒストグラム軸** 散布図の横軸と縦軸におけるヒストグラムの表示／非表示を切り替えます。

**メモ:** 非表示の行のデータ点は、散布図には表示されず、ヒストグラムには表示されます。行をヒストグラムから除外し、分析の対象外にするには、行の属性を「非表示かつ除外」に設定したうえで、赤い三角ボタンのメニューから「やり直し」>「分析のやり直し」を選択してください。

**要約統計量** プロットされている変数の要約統計量の表示／非表示を切り替えます。表示される要約統計量は、変数間の相関係数と共分散、各変数の平均と標準偏差です。

**平均のあてはめ** 目的変数の平均をあてはめます。傾きをゼロに制約した単回帰モデルということができます。「[「平均のあてはめ」レポート](#)」を参照してください。

**直線のあてはめ** 最小2乗法によって、回帰直線をあてはめます。プロット上に直線が表示され、あてはめのレポートが作成されます。「[「直線のあてはめ」](#)」、「[「多項式のあてはめ」](#)」、「[「その他のあてはめ」レポート](#)」を参照してください。

**多項式のあてはめ** 最小2乗法によって、指定された次数の多項式曲線をあてはめます。「[「直線のあてはめ」](#)」、「[「多項式のあてはめ」](#)」、「[「その他のあてはめ」レポート](#)」を参照してください。

**その他のあてはめ** YやXを変換した後で回帰モデルをあてはめることができます。変換には、自然対数・平方根・2乗・逆数・指数があります。変換後の変数を使い、**多項式の中心化**をオフにしたり、切片と傾きの制約を指定したり、多項式モデルをあてはめることも可能です。「[「その他のあてはめ」のウィンドウ](#)」を参照してください。

**ノンパラメトリックな曲線** ノンパラメトリックなモデルをあてはめます。モデルの種類には、スプライン曲線、カーネル平滑化、値ごとのあてはめがあります。

**スプライン曲線のあてはめ** 罰則付き最小2乗法による平滑化曲線をあてはめます。平滑化パラメータ入を使って、モデルの滑らかさを調整できます。「[「スプライン曲線のあてはめ」レポート](#)」を参照してください。

**カーネル平滑化** 局所的な重み付けを行う最小2乗法でモデルをあてはめます。このモデルは、LOWESS (locally weighted scatterplot smoothing) とも呼ばれます。平滑化パラメータ $\alpha$ を使って、モデルの滑らかさを調整できます。「[「局所的な平滑化」レポート](#)」を参照してください。

**値ごとのあてはめ** 各X値の応答平均を線でつなぎます。「[「値ごとのあてはめ」レポート](#)」を参照してください。

**直交のあてはめ** 直交回帰モデルをあてはめます。直交回帰モデルは、X変数とY変数の両方が誤差を伴って測定される場合に使います。「[「直交のあてはめ」レポート](#)」を参照してください。

**一変量分散, 主成分** 標準化したデータに対する主成分分析の第1主成分をあてはめます。

**等分散性** 分散比を1とした直交回帰モデルをあてはめます。このモデルでは、XとYの誤差分散が等しいと仮定します。この方法は「**Deming 回帰**」という名前でも知られています。

**Y→Xのあてはめ** 分散比を0、すなわちYの分散を0とした直交回帰モデルをあてはめます。

**指定された分散比** 分散比の値を指定して直交回帰モデルをあてはめます。

**Passing Bablokのあてはめ** Passing-Bablok法によって回帰モデルをあてはめます。Passing-Bablok法を行った後、赤い三角ボタンよりBland-Altman分析を実行することもできます。Passing-Bablok法、X変数とY変数の両方が誤差を伴って測定される場合に使います。「[「Passing-Bablokのあてはめ」レポート](#)」および「[「Passing Bablokのあてはめ」オプションの統計的詳細](#)」を参照してください。

**ロバスト** ロバスト（頑健）な推定方法による回帰モデルをあてはめます。ロバストな推定方法は、外れ値の影響を少なくします。「[「ロバストなあてはめ」レポート](#)」を参照してください。

**ロバストなあてはめ** HuberのM推定を使用してロバストな回帰モデルをあてはめます。

**Cauchyのあてはめ** Cauchy分布を仮定した最尤法を用いてロバストな回帰モデルをあてはめます。

**確率楕円** プロットに指定されたパーセントの二変量正規楕円を描画します。楕円の中に、指定の割合のデータ点が含まれることになります。このオプションを使うと相関が推定できます。「[「確率楕円」レポート](#)」を参照してください。

**ノンパラメトリック密度** プロットにノンパラメトリック密度の等高線を描画します。密度等高線はデータ点の密度を表します。「[ノンパラメトリック密度](#)」を参照してください。

**グループ別** グループ変数を指定することができます。その後は、指定したグループ変数の水準ごとにあてはめが計算されるようになります。グループごとにあてはめられた直線・曲線・楕円などが散布図に重ねて表示されます。こうすると、グループごとにあてはめたモデルを、視覚的に比較することができます。「[グループ別確率楕円の例](#)」および「[グループ別回帰直線の例](#)」を参照してください。

---

**メモ:** [グループ別] を使うと、グループごとにあてはめたモデルが1つのプロットに表示され、結果も1つのレポートに並べて表示されます。起動ウィンドウの [By] オプションを使った場合は、グループごとの結果が個別のレポートに表示されます。

---

以下のオプションの詳細については、『JMPの使用法』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMPレポートの設定に合わせて環境設定を変更したりできます。

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

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

---

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの [スクリプトの索引] を開いてください。また、[スクリプトの索引] には、この節で紹介されているオプションのスクリプト例もあります。

---

## 二変量の関係のオプション

「二変量」プラットフォームでは、指定したあてはめごとに赤い三角ボタンのメニューが用意されます。このメニューは、散布図の下にあるあてはめのラベルの隣に表示されます。

- 「ほとんどのあてはめで使えるオプション」
- 「二変量正規楕円」に適用できるオプション」
- 「分位点密度等高線」に適用できるオプション」

### ほとんどのあてはめで使えるオプション

「二変量」プラットフォームでは、指定したあてはめごとに赤い三角ボタンのメニューが用意されます。あてはめの種類によっては使えないオプションもあります。

**あてはめ線** あてはめたモデルを表す直線・曲線・等高線の表示／非表示を切り替えます。このオプションは、「分位点密度等高線」にはありません。

**回帰の信頼区間** 期待値（平均）に対する信頼限界（信頼区間の曲線）の表示／非表示を切り替えます。

**個別の値に対する信頼区間** 個別の応答値に対する信頼限界の表示／非表示を切り替えます。この信頼区間は、誤差における変動とパラメータ推定値における変動の両方を考慮しています。

**線の色** あてはめ線の色を指定できます。

**線種** 各あてはめの線種を選択できます。

**線の幅** 各あてはめの線の幅を選択できます。

**レポート** 各あてはめのレポートの表示／非表示を切り替えます。

---

**メモ:** 散布図には影響しません。

---

**プロファイル** X変数の予測トレースの表示／非表示を切り替えます。他にも、シミュレーションをしたり、複数の応答変数を最適化したりできます。『プロファイル機能』を参照してください。

**予測値の保存** データテーブル内に「予測値 <列名>」という新しい列を作成します。「列名」の部分にはY変数の名前が挿入されます。この列には、予測式と予測値が含まれます。

---

**ヒント:** データテーブルに新しい行を追加すると、保存された予測式によって、その行の予測値が自動的に計算されます。

---

**残差の保存** データテーブル内に「残差 <列名>」という新しい列を作成します。「列名」の部分にはY変数の名前が挿入されます。この列には、応答の観測値から予測値を引いた値が含まれます。

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

**スチューデント化残差の保存** データテーブル内に「スチューデント化残差 <列名>」という新しい列を作成します。「列名」の部分にはY変数の名前が挿入されます。この列には、残差を標準誤差で割った値が含まれます。

**平均の信頼限界の計算式** データテーブル内に「平均 <列名>の下側95%」と「平均 <列名>の上側95%」という2つの新しい列を作成します。「列名」の部分にはY変数の名前が挿入されます。列には、応答の平均に対する両側95%信頼区間の下限と上限を求める計算式、およびその結果が保存されます。

**個別の信頼限界の計算式** データテーブル内に「個別 <列名>の下側95%」と「個別 <列名>の上側95%」という2つの新しい列を作成します。「列名」の部分にはY変数の名前が挿入されます。列には、応答の個々の値に対する両側95%信頼区間の下限と上限を求める計算式、およびその結果が保存されます。

**残差プロット**（直線、多項式、Passing Bablok、その他のあてはめでのみ使用できます。）「予測値と残差のプロット」、「予測値と実測値のプロット」、「行番号と残差のプロット」、「Xと残差のプロット」、「残差の正規分位点プロット」の5種類の診断プロットの表示／非表示を切り替えます。

**Bland Altman 分析**（Passing Bablokでのみ使用できます。）新しいレポートウィンドウに対応のあるペアとBland Altman分析を表示します。

**$\alpha$  水準の設定** 信頼曲線の計算に使用する有意水準を指定できます。

**回帰の信頼区間内を塗る**（一部のあてはめでは使用できません。）応答の期待値（平均）に対する信頼区間を陰影つきの領域として描きます。

**個別の値に対する信頼区間内を塗る**（一部のあてはめでは使用できません。）応答の個々の値に対する信頼区間を陰影つきの領域として描きます。

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

**あてはめの削除** 散布図からあてはめ線を削除し、レポートから該当の結果を削除します。

## 「二変量正規楕円」に適用できるオプション

「二変量」プラットフォームで確率楕円をあてはめると、「二変量正規楕円」の赤い三角ボタンが表示されます。

**楕円内を塗る** 確率楕円の内側の領域の塗りを切り替えます。

**内側の点を選択** 確率楕円の内側の点を選択します。

**外側の点を選択** 確率楕円の外側の点を選択します。

## 「分位点密度等高線」に適用できるオプション

「二変量」プラットフォームでノンパラメトリック密度をあてはめると、「分位点密度等高線」の赤い三角ボタンが表示されます。

**カーネルの制御** 各変数の標準偏差を変更するスライダーの表示／非表示を切り替えます。標準偏差は、ノンパラメトリックなカーネル密度推定の滑らかさを左右します。

**5%の等高線** 5%の等高線の表示／非表示を切り替えます。

**等高線** 等高線の表示／非表示を切り替えます。

**等高線間を塗りつぶす** 等高線間の塗りつぶしの表示／非表示を切り替えます。

**カラーテーマ** 等高線のカラーテーマを変更することができます。

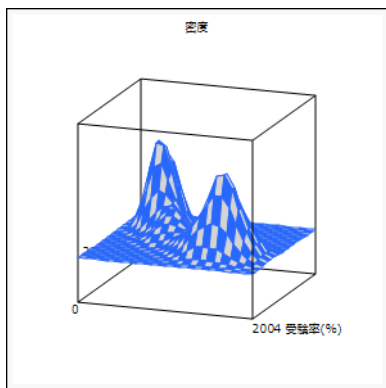
**密度による点の選択** 指定された分位点範囲内の点を選択します。

**密度分位点による色分け** 密度分位点に応じて点を色分けします。

**密度分位点の保存** データテーブル内に各点の密度分位点を含む新しい列を作成します。

**メッシュプロット** 2変量に対する密度推定値の3次元プロットの表示／非表示を切り替えます。

図5.6 メッシュプロットの例



**峰型クラスター** 峰型クラスターの結果の表示／非表示を切り替えます。データテーブル内に各データペアのクラスター番号を含む新しい列を作成します。峰型クラスターは、密度推定値に基づいて計算されます。JMPが生成するグリッドは、10,404個の密度推定値で構成されます。密度推定値の分布に見られる最頻値（モード）の数によってクラスターの数が決まります。最頻値は**クラスターの中心**となります。その他の点は、距離と密度推定値に従って各最頻値へ反復的にクラスタリングされます。

---

**メモ:** JMPのその他のクラスタリング手法については、『多変量分析』を参照してください。

---

**密度グリッドの保存** 密度推定値とその分位点を新しいデータテーブルに保存します。



---

**メモ:** 峰型クラスター値を先に保存してから密度グリッドを保存すると、グリッドテーブルにもクラスター値が含まれます。

---

---

## 「二変量の関係」レポート

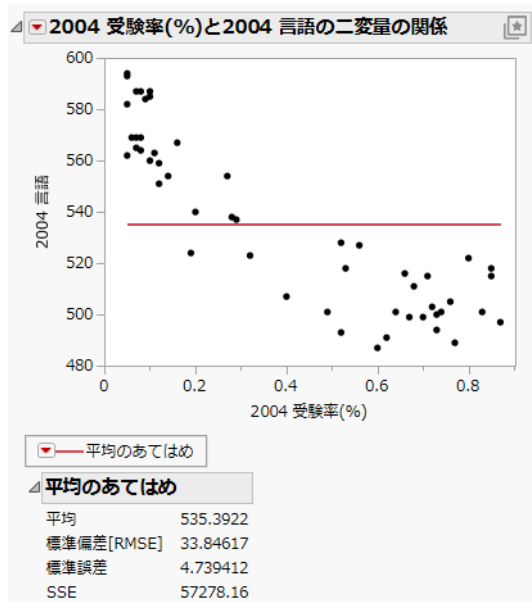
「二変量」プラットフォームでは、1つのX変数からY変数の値を予測する多彩なモデルがあてはめられます。この節では、モデルをあてはめたときに作成されるレポートの内容を詳しく説明します。

- 「「平均のあてはめ」レポート」
- 「「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」レポート」
- 「「その他のあてはめ」のウィンドウ」
- 「「スプライン曲線のあてはめ」レポート」
- 「「局所的な平滑化」レポート」
- 「「値ごとのあてはめ」レポート」
- 「「直交のあてはめ」レポート」
- 「「ロバストなあてはめ」レポート」
- 「「Passing-Bablokのあてはめ」レポート」
- 「「確率楕円」レポート」
- 「「ノンパラメトリック密度」

## 「平均のあてはめ」レポート

「二変量」プラットフォームの「平均のあてはめ」は、応答（Y）の平均をあてはめます。平均を表す直線を参照線として、他のあてはめと比較することができます。

図5.7 平均のあてはめの例



「二変量」プラットフォームの「平均のあてはめ」レポートには、要約統計量の表が表示されます。

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

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

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

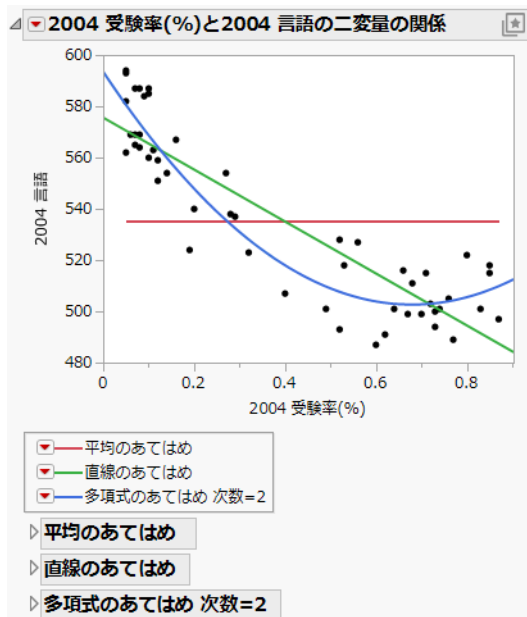
**SSE** 単純な平均モデルの誤差平方和。モデルのあてはめでは、「分散分析」レポートに「誤差」の平方和として表示されます。

「平均のあてはめ」メニューのオプションについては、[「二変量の関係のオプション」](#)を参照してください。

## 「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」レポート

「二変量」プラットフォームの「直線のあてはめ」、「多項式のあてはめ」、「その他のあてはめ」は、回帰モデルをあてはめます。複数のモデルをあてはめて散布図上で比較することができます。

図5.8 直線や多項式のあてはめ



「直線のあてはめ」メニューと「多項式のあてはめ」メニューのオプションについては、「[二変量の関係のオプション](#)」を参照してください。詳細については、「[\[直線のあてはめ\] オプションの統計的詳細](#)」を参照してください。

「二変量」プラットフォームでは、選択したあてはめのそれぞれにレポートが作成されます。「直線のあてはめ」、「多項式のあてはめ」、「変換されたあてはめ」の各レポートでは、テキストボックスの中にあてはめられた式が表示されます。各レポートは、あてはめの要約、分散分析（ANOVA）、パラメータ推定値で構成されます。データに反復がある場合は、あてはまりの悪さ（LOF; Lack Of Fit）の表も表示されます。変換したY変数のあてはめには、元のデータから計算した適合度指標の要約が含まれます。

## あてはめの要約

「二変量」プラットフォームのあてはめレポートには、モデルのあてはめの要約統計量をまとめた「あてはめの要約」表が表示されます。「あてはめの要約」表の上にあるのは、あてはめられた式です。

図 5.9 「あてはめの要約」表

直線のあてはめ	
2004 言語 = 575.62539 - 101.32814*2004 受験率(%)	
あてはめの要約	
R2乗	0.790274
自由度調整R2乗	0.785994
誤差の標準偏差(RMSE)	15.65752
Yの平均	535.3922
オブザベーション(または重みの合計)	51

「あてはめの要約」表には、次のような統計量が表示されます。

**R2乗** モデルによって説明できる変動の割合を表します。残りの変動は、モデルで説明できないため、ランダムな誤差とみなされます。モデルが完全にあてはまっていると、R2乗は1になります。「[「あてはめの要約」レポートの統計的詳細](#)」を参照してください。

**メモ:** R2乗が小さい場合は、必要な説明変数がモデルに含まれていない可能性があります。ただし、扱っているデータの誤差やランダムな変動が大きい場合は、意味のある回帰モデルであっても、R2乗の値が小さくなることがあります。自分の研究分野の文献を読み、R2乗の大きさが通常、どれぐらいであるかを把握してください。

**自由度調整 R2 乗** 「自由度調整済み R2 乗」、「自由度調整済み決定係数」などと呼ばれている指標であり、R2乗をモデルのパラメータ数で調整した指標です。自由度調整済み R2 乗は、パラメータ数の異なるモデルを比較するのに役立ちます。「[「あてはめの要約」レポートの統計的詳細](#)」を参照してください。

**誤差の標準偏差 (RMSE)** 誤差の標準偏差の推定値。これは、「分散分析」レポートにおける「誤差」行の「平均平方」列にある値の平方根です (図 5.11)。

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

**オブザベーション (または重みの合計)** モデルを推定するために使われた観測 (オブザベーション) の個数。ただし、重み変数が使われた場合は、重みの合計です。

## あてはまりの悪さ (LOF)

「二変量の関係」レポートの「あてはまりの悪さ (LOF)」表は、あてはまりの悪さ (Lack Of Fit) の検定結果を示します。あてはまりの悪さの検定は、X 値に反復があり、モデルが飽和モデルでない場合のみ、可能です。反復から計算される平方和を、**純粋誤差**と呼びます。純粋誤差は、データの誤差のうち、どのようなモデルを構築しても説明や予測ができない変動を表します。

図 5.10 「直線のあてはめ」の「あてはまりの悪さ (LOF)」表

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

モデルの残差誤差から、純粋誤差を引いたものを、**あてはまりの悪さ (LOF; Lack Of Fit)** の誤差といいます。指定したモデルが不適切だと、LOF 誤差が純粋誤差よりも有意に大きくなります。不適切なモデルとは、データを適切に描写していないモデルを指します。あてはまりの悪さ (LOF) 検定の帰無仮説は、「母集団において LOF 誤差は 0 である」というものです。そのため、 $p$  値が小さいときは、モデルがデータにあてはまっていないことを意味します。

「あてはまりの悪さ (LOF)」表には、次のような列があります。

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

**自由度** 誤差の各要因がもつ自由度 (DF)。

- 「合計誤差」の自由度は、「分散分析」表の「誤差」の行に表示されている自由度です。[「分散分析」](#)を参照してください。「合計誤差」の自由度は、「分散分析」表の「全体」の自由度と「モデル」の自由度の差です。「誤差」の自由度は、あてはまりの悪さ (LOF) と純粋誤差の 2 つの自由度に分けることができます。
- 「純粋誤差」の自由度は、反復されている観測値の各グループにおける自由度を足し合わせた値です。[「あてはまりの悪さ \(LOF\)」レポートの統計的詳細](#)を参照してください。
- 「あてはまりの悪さ (LOF)」の自由度は、「合計誤差」と「純粋誤差」の自由度の差です。

**平方和** 誤差の各要因の平方和 (SS)。

- 「合計誤差」の平方和は、「分散分析」表の「誤差」の行にある平方和です。[「分散分析」](#)を参照してください。
- 「純粋誤差」の平方和は、反復されている観測値の各グループにおける平方和を足し合わせた値です。この平方和を自由度で割った値は、モデルに依存しない、誤差分散の推定値になります。[「あてはまりの悪さ \(LOF\)」レポートの統計的詳細](#)を参照してください。
- 「あてはまりの悪さ (LOF)」の平方和は、「合計誤差」と「純粋誤差」の平方和の差です。LOF の平方和が大きいときは、モデルがデータに良くあてはまっていない可能性があります。

**平均平方** 要因の平均平方。平方和を自由度で割った値。「純粋誤差」の平均平方を基準にして、「あてはまりの悪さ (LOF)」の平均平方が相対的に大きくなっている場合、モデルがうまくあてはまっていないことを示します。F 値を使って仮説検定を行うことができます。

**F 値** 「あてはまりの悪さ (LOF)」の平均平方の、「純粋誤差」の平均平方に対する比。F 値の値が大きいほど、「母集団において LOF 誤差は 0 である」という帰無仮説を否定する証拠が強いことを意味します。

**p 値(Prob > F)** あてはまりの悪さ (LOF) に対する検定の  $p$  値。検定の帰無仮説は、「母集団において LOF 誤差は 0 である」というものです。 $p$  値が小さいときは、現在のモデルがデータにあてはまっていないことを意味します。

**最大 R2 乗** モデルに含まれている変数で達成できる最大の  $R^2$  値。「[「あてはまりの悪さ \(LOF\)」レポートの統計的詳細](#)」を参照してください。

## 分散分析

「二変量の関係」レポートの「分散分析」表には、分散分析表が表示されます。分散分析表は、現在のモデルと切片だけのモデルとを比較した統計量で構成されています。「分散分析」表の値は、モデルの有効性を評価する  $F$  比の計算に使われます。「 $F$  値」に関連する確率 ( $p$  値) が小さいとき、「そのモデルは、 $Y$  の平均だけのモデルよりも、良くあてはまっている」と見なすことができます。

図 5.11 「直線のあてはめ」の「分散分析」表

直線のあてはめ				
2004 言語 = 575.62539 - 101.32814*2004 受験率(%)				
▷ あてはめの要約				
▷ あてはまりの悪さ(LOF)				
分散分析				
要因	自由度	平方和	平均平方	F 値
モデル	1	45265.424	45265.4	184.6379
誤差	49	12012.733	245.2	p 値(Prob>F)
全体(修正済み)	50	57278.157		<.0001*

「分散分析」表には、次のような列があります。

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

**自由度** 変動の各要因に関連した自由度 (DF; Degrees of Freedom)。「全体 (修正済み)」の自由度は、標本サイズより常に 1 つ少なく、次のように「モデル」と「誤差」の自由度に分割されます。

- 「モデル」の自由度は、モデルのあてはめで使用されるパラメータの数 (切片を除く) に相当します。
- 「誤差」の自由度は、「全体 (修正済み)」と「モデル」の自由度の差です。

**平方和** 変動の各要因に関連した平方和 (SS; Sum of Squares)。

- 全体 (「全体 (修正済み)」) の平方和は、各応答値から平均までの距離を 2 乗して合計したものです。これは、応答値の全体変動を示します。
- 「誤差」の平方和は、予測値から実測値までの距離の平方和です。これは、あてはめたモデルでは説明のつかない変動を示します。
- 「モデル」の平方和は、「全体 (修正済み)」の平方和と「誤差」の平方和の差です。これは、モデルで説明される変動を示します。

**平均平方** 「誤差」と「モデル」における平均平方。平均平方は、平方和をその自由度で割った値です。

**メモ:** 「誤差」の平均平方の平方根は、「あてはめの要約」表にある「誤差の標準偏差 (RMSE)」と同じです。

**F 値** モデルの平均平方を、誤差の平均平方 (MSE; Mean Square for Error) で割った値。F 値は、「現在のモデルが、切片だけのモデル (つまり、予測値が応答変数の平均値であるモデル) と有意に異なるかどうか」を検定するのに使われる検定統計量です。これは、「(切片を除く) すべての回帰パラメータが0である」という帰無仮説を検定します。この仮説が真のとき、「誤差」と「モデル」の平均平方は両方とも誤差分散の推定値となり、その比は  $F$  分布に従います。

**p 値 (Prob > F)** 検定の観測有意確率 ( $p$  値)。 $p$  値が小さいと、回帰効果があると考えられます。

## パラメータ推定値

「二変量の関係」レポートの「パラメータ推定値」表には、モデルのパラメータ推定値が表示されます。

図 5.12 「直線のあてはめ」の「パラメータ推定値」表

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

「パラメータ推定値」表には、以下の列があります。

**項** パラメータ推定値に対応するモデル項。最初の項は切片です。

**推定値** 各項のパラメータ推定値。これらは、モデル係数の推定値です。

**標準誤差** パラメータ推定値の標準誤差を推定した値。

**t 値** 各パラメータがゼロであるという帰無仮説の検定統計量。t 値は、パラメータ推定値とその標準誤差の比です。モデルに関する仮定が満たされるときは、この  $t$  値は Student の  $t$  分布に従います。

**p 値 (Prob>|t|)** 「真のパラメータ値は0である」という帰無仮説、「真のパラメータ値は0ではない」という対立仮説の両側検定に対する  $p$  値。

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

**下側 95%** パラメータ推定値の両側 95% 信頼区間の下限。

**上側 95%** パラメータ推定値の両側 95% 信頼区間の上限。

**標準  $\beta$**  すべての項を平均0、分散1に標準化したときの回帰モデルから得られるパラメータ推定値 (標準化偏回帰係数)。「[パラメータ推定値](#) レポートの統計的詳細」を参照してください。

**VIF** モデルの各項に対する VIF (Variance Inflation Factor)。この値が大きい場合、モデル内の項間に共線性があると考えられます。

**計画の標準誤差** パラメータ推定値の相対分散の平方根。「パラメータ推定値」レポートの統計的詳細を参照してください。

## 元のスケールで測定したあてはめ

「二変量の関係」レポートの「元のスケールで測定したあてはめ」表には、変換前のスケールで測定した適合度の要約統計量が表示されます。この表は、Y変数を変換した場合のみ作成されます。

## 「その他のあてはめ」のウィンドウ

「二変量」プラットフォームの[その他のあてはめ] オプションを使うと、変数を変換して回帰モデルをあてはめることができます。切片と傾きの制約を指定したり、指定した次数の多項式をあてはめたり、多項式の中心化をオンにすることも可能です。

[その他のあてはめ] オプションを選択すると、「変換または制約を指定」ウィンドウが開き、次のようなオプションが表示されます。

**Yの変換、Xの変換** Y変数またはX変数の変換を指定します。変換の種類には、自然対数・平方根・2乗・逆数・指数があります。

**次数** あてはめる多項式の次数を指定します。

**多項式の中心化** 多項式を中心化できます。中心化を行う場合はチェックを入れ、行わない場合はチェックを外します。多項式の中心化オプションを用いると、多重共線性の影響が軽減され、回帰係数の数値計算が安定します。

---

**メモ:** なお、X変数を変換した場合、多項式の中心化は行われません。

---

**切片の制約** モデルの切片を指定の値に固定します。

**傾きの制約** モデルの傾きを指定の値に固定します。

---

**ヒント:** 図に $Y=X$ の直線を追加するには、[その他のあてはめ] 選択し、「切片の制約: 0」と「傾きの制約: 1」を指定します。

---

[変換されたあてはめ] メニューのオプションについては、「二変量の関係のオプション」を参照してください。例として、「[その他のあてはめ] オプションの例」を参照してください。



## 「スプライン曲線のあてはめ」レポート

「二変量」プラットフォームの[ノンパラメトリックな曲線] > [スプライン曲線のあてはめ] オプションでは、ラムダ ( $\lambda$ ) の値を指定してさまざまな滑らかさ（柔軟さ）のスプライン曲線をあてはめることができます。 $\lambda$  は、スプライン曲線を調整するパラメータです。 $\lambda$  を小さくすると、誤差項に大きな重みが置かれ、柔軟なモデルがあてはめられます。 $\lambda$  を大きくすると、誤差項に小さな重みが置かれ、直線に近い硬いモデルがあてはめられます。

[スプライン曲線のあてはめ] のメニューから  $\lambda$  の値を選択するか、[スプライン曲線のあてはめ] > [その他...] を選択して  $\lambda$  の値を入力します。また、「平滑化スプライン曲線によるあてはめ」レポートで対話的に  $\lambda$  を調整することもできます。

---

**メモ:** X 変数を標準化するには、[スプライン曲線のあてはめ] > [その他...] を選択し、[X の標準化] オプションをオンにします。

---

[平滑化スプライン曲線によるあてはめ] メニューのオプションについては、「二変量の関係のオプション」を参照してください。このあてはめの統計的詳細については、「[スプライン曲線のあてはめ オプションの統計的詳細](#)」を参照してください。

「平滑化スプライン曲線によるあてはめ」表には、次のような列があります。

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

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

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

## 「局所的な平滑化」レポート

「二変量」プラットフォームの[ノンパラメトリックな曲線] > [カーネル平滑化] オプションは、局所的な重み付けを行う最小2乗法でモデルをあてはめます。いくつかの抽出された点に関して、局所的に重み付けて直線や2次曲線といった単純なモデルを次々にあてはめていくことにより、滑らかな曲線を求めます。最大で合計512個の局所的な回帰モデルが組み合わせられ、データ全体にわたって滑らかな曲線が作成されます。この手法は、**LOWESS** (locally weighted scatterplot smoothing) と呼ばれます。[カーネル平滑化] オプションでは、Cleveland (1979) の手法を使用しています。ただし、ほぼ完全なあてはめの場合は微調整を加えており、Cleveland の双加重関数の引数  $6 * q50$  を、 $\max(6 * q50, 2 * q90)$  で置き換えています。ここで、 $q50$  は50パーセント点、 $q90$  は90パーセント点です。[局所的な平滑化] メニューのオプションについては、「[二変量の関係のオプション](#)」を参照してください。

「局所的な平滑化」レポートには、次のような列があります。

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

**誤差平方和 (SSE)** 各点からあてはめた曲線までの距離の平方和。これは、あてはめたモデルで説明できていない「誤差」(残差)を表します。

**局所的なあてはめ( $\lambda$ )** 局所的なあてはめに使われる多項式の次数( $\lambda$ )を指定します。

**重み関数** 重み関数を指定します。LOWESSモデルはトリキューブ(3次式の3乗)を使用します。このような重み付け関数は、局所的な回帰をあてはめていくときに、各点 $(x_i, y_i)$ が局所的な回帰に与える影響を決めます。

**滑らかさ( $\alpha$ )** 局所的なあてはめで対象となる点の数を指定します。値を入力するか、スライダーで調整します。 $\alpha$ は平滑化パラメータと呼ばれ、0~1の値を取ります。 $\alpha$ を大きくすると、曲線の滑らかさが増します。

**標本抽出デルタ** 計算に使われる標本の抽出率を指定することができます。デフォルトでは、標本抽出デルタはゼロであり、すべての点があてはめに使用されます。標本抽出デルタの値を大きくすると、直前の抽出点からデルタの範囲内にあるような点は抽出されなくなります。データの密度が高く、あてはめに使用する点の数を減らしたいような場合に、このオプションを使用してください。

**ロバスト性** あてはめのロバスト性を指定することができます。反復のたびに重みを調整し、モデルから離れた点に小さな重みを与えることにより、外れ値による影響を小さくします。

## 「値ごとのあてはめ」レポート

「二変量」プラットフォームの「値ごとのあてはめ」オプションは、X変数の一意の値ごとに応答の平均を計算し、それを折れ線でつなぎます。

「値ごとのあてはめ」メニューのオプションについては、「[二変量の関係のオプション](#)」を参照してください。

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

**オブザベーションの数** 観測(オブザベーション)の合計個数。

**異なるXの値の個数** Xの異なる値の個数。

**自由度** 純粋誤差の自由度。

**平方和** 純粋誤差の平方和。

**平均平方** 純粋誤差の平均平方。

## 「直交のあてはめ」レポート

「二変量」プラットフォームの「直交のあてはめ」オプションは、Deming回帰を含む直交回帰モデルをあてはめます。直交回帰モデルでは、Y変数におけるばらつきだけではなく、X変数におけるばらつきも考慮します。

「直交回帰 分散比」表には、直交回帰モデルの要約統計量が表示されます。

**変数** 変数の名前。

**平均** 各変数の平均。

**標準偏差** 各変数の標準偏差。

**分散比** 直線のあてはめに使用された分散比。

**相関** 2つの変数間の相関。

**切片** あてはめた直線の切片。

**傾き** あてはめた直線の傾き。

**下側信頼限界** 傾きの下側信頼限界。

**上側信頼限界** 傾きの上側信頼限界。

**アルファ** 信頼区間の計算に使用される $\alpha$ 水準。 $\alpha$ 水準を変更する場合は、テキストボックスに新しい値を入力します。

[直交回帰 分散比] メニューのオプションについては、[「二変量の関係のオプション」](#)を参照してください。このあてはめの統計的詳細については、[「\[直交のあてはめ\] オプションの統計的詳細」](#)を参照してください。このオプションの例については、[「\[直交のあてはめ\] オプションの例」](#)を参照してください。

## 「ロバストなあてはめ」レポート

「二変量」プラットフォームの[ロバスト] オプションは、ロバストな回帰モデルをあてはめます。ロバストな手法を使用すると、外れ値がモデルに与える影響が少なくなります。

「ロバストなあてはめ」と「Cauchyのあてはめ」のレポートは、次のような要約統計量を含む2つの表で構成されます。

**$\sigma$**  通常の最小2乗推定における「誤差の標準偏差」(RMSE)に相当する $\sigma$ 。

**カイ2乗** モデルが応答変数の単なる平均より予測モデルとして優れているという仮説の検定統計量。

**$p$  値** 「傾きが0である」に対するカイ2乗検定の $p$ 値。

**対数値** 対数値は、 $p$ 値を変換したもので、 $-\log_{10}(p \text{ 値})$ と定義されます。

**パラメータ** モデルパラメータの名前。

**ロバストな推定値** 各項のパラメータ推定値。これらは、モデル係数の推定値です。

**標準誤差** パラメータ推定値の標準誤差を推定した値。

[ロバストなあてはめ] メニューと[Cauchyのあてはめ] メニューのオプションについては、[「二変量の関係のオプション」](#)を参照してください。このオプションの例については、[「\[ロバストなあてはめ\] オプションの例」](#)を参照してください。

## 「Passing-Bablokのあてはめ」レポート

「二変量」プラットフォームの「Passing Bablokのあてはめ」オプションは、Passing-Bablok手法を使って回帰モデルをあてはめます。Passing-Bablok回帰は、同じ対象を2つの異なる測定方法で測定した場合に、それら2つの測定方法を比較する目的で開発されました。この手法は、2つの変数の間の関係は1次式であり、かつ、強い相関があると仮定しています。散布図に、あてはめ線と $Y = X$ の点線が追加されます。「Passing-Bablokのあてはめ」の赤い三角ボタンのメニューにある「Bland Altman分析」オプションは、対応のある $t$ 検定とBland-Altman分析を行います。

「Passing-Bablokのあてはめ」レポートは、3つの表で構成されます。

### ノンパラメトリック：Kendallの順位相関係数( $\tau$ )

Kendallの $\tau$ を使ってX変数とY変数の間の相関を評価します。

**X** X変数。

**Y** Y変数。

**Kendallの $\tau$**  X変数とY変数の相関を示すノンパラメトリックな指標。 $-1 \sim 1$ の値を取り、0に近い値は、X変数とY変数が独立であることを示唆します。

**p値(Prob>| $\tau$ |)** 「X変数とY変数は独立である」という帰無仮説に対する仮説検定の $p$ 値。小さい $p$ 値は、X変数とY変数は独立ではないことを示唆しています。独立でない場合にのみ、Passing-Bablok法は適しています。

### 直線性の累積和検定

「直線性の累積和検定」表は、直線性の検定結果をまとめたものです。 $p$ 値が小さい場合、直線性の帰無仮説は棄却されます。Passing-Bablok回帰は直線性を仮定しているので、直線性の帰無仮説が棄却された場合、Passing-Bablok回帰が適さないことを示唆します。

**最大累積和** 各点からPassing-Bablok回帰直線までの垂直距離に応じて並べられた $\sqrt{L/I}$ と $-\sqrt{I/L}$ の値の累積和の最大絶対値。 $I$ は、正の残差を持つ観測の個数、 $L$ は、負の残差を持つ観測の個数です。また、プラスとマイナスのいずれにするかは、残差の符号で決められます。手法間に強い相関がある場合、 $I$ と $L$ は等しくなります。そのため、多くの場合、累積和は $+1$ と $-1$ の和になります。最大累積和の値が小さいときは、Passing-Bablok回帰直線の上側と下側に均等に点が散らばっていることを示し、直線性の仮説が支持されます。

**H** 累積和検定の検定統計量。この検定統計量は、最大累積和を、残差が負になっている観測数に1を足した値の平方根で割ったものです。この検定統計量は、Kolmogorov-Smirnov分布に従います。

**p値(Prob > H)** 累積和検定の $p$ 値。小さい $p$ 値は、Passing-Bablok法が適切でない可能性を示唆します。

## パラメータ推定値

「パラメータ推定値」表には、Passing-Bablok 法による切片と傾きの推定値、およびその 95% 信頼区間が表示されます。

## 「対応のあるペア」レポート

「Passing-Bablok のあてはめ」の赤い三角ボタンのメニューにある [Bland Altman 分析] オプションは、対応のある  $t$  検定と Bland-Altman 分析を行います。対応のあるペアの詳細については、『予測モデルおよび発展的なモデル』を参照してください。

## Bland-Altman 分析

「Bland-Altman 分析」表には、次の「値」・「標準偏差」・「信頼限界」が表示されます。

**バイアス** X 変数と Y 変数の差。「値」には差の平均が、「標準偏差」には差の標準偏差が表示されます。

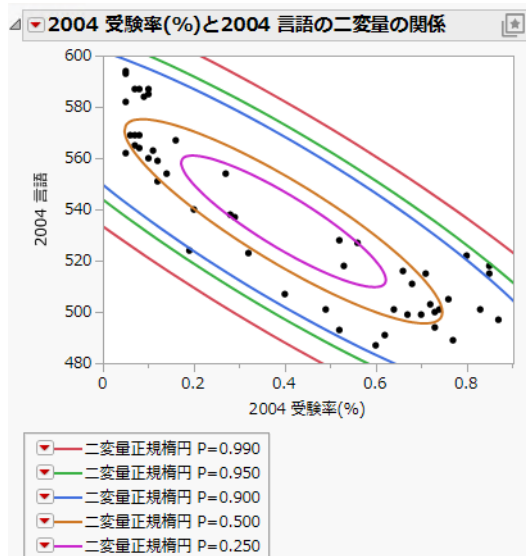
**許容範囲の限界** 許容範囲 (limits of agreement) の上限と下限です。これは、「差の平均」 $\pm z_{1-\alpha/2} \times$ 「差の標準偏差」です。

## 「確率楕円」レポート

「二変量」プラットフォームの [確率楕円] オプションは、一定の割合の点を含む楕円を描画し、X 変数と Y 変数の間の相関を調べます。楕円内に含まれる点の数（これは母集団が二変量正規分布に従っているとして、その母集団にて含まれる観測の個数）は、[確率楕円] メニューから選択した確率の値によって決まります。

[二変量正規楕円] メニューのオプションについては、「[二変量の関係のオプション](#)」を参照してください。このオプションの例については、「[グループ別確率楕円の例](#)」を参照してください。

図 5.13 確率楕円の例



確率楕円は、XとYから推定された二変量正規分布から計算されます。二変量正規分布は、X変数とY変数の平均と標準偏差、および、その相関によって決まります。

楕円は、二変量正規分布を仮定した上で、指定された割合だけ母集団においてデータが含まれる領域を表しています。確率楕円の形状からは、2変数間の相関が読み取れます。幅の狭い楕円は相関が強いことを示し、円に近い楕円は相関が弱いことを示します。

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

「二変量正規楕円」レポートには、確率楕円の要約統計量が表示されます。

**変数** 楕円の作成に使用された変数の名前。

**平均** 各変数の平均。

**標準偏差** 各変数の標準偏差。

**相関** Pearsonの相関係数。「二変量正規楕円」レポートの統計的詳細」を参照してください。

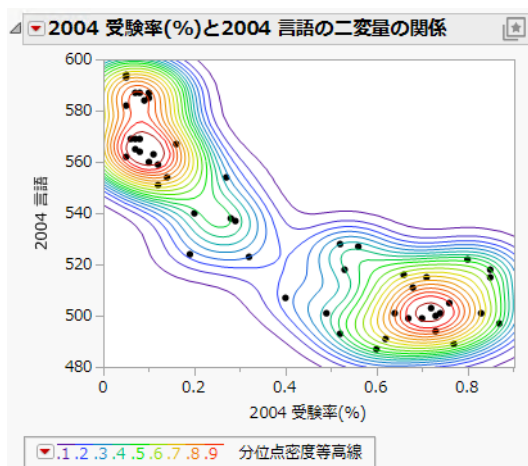
**p値** 母集団においてX変数とY変数の間に相関関係がないという仮定のもとで、観測された相関よりも絶対値の大きい相関が得られる確率。

**数** 計算に使用された標本サイズ（オブザベーションの個数）。

## ノンパラメトリック密度

「二変量」プラットフォームの[ノンパラメトリック密度]オプションは、データに滑らかな二変量のノンパラメトリック密度面をあてはめます。ノンパラメトリック密度面は、等高線を使って作成されます。等高線は、5% 間隔で描画されます。ノンパラメトリックな密度推定から得られた点のうち、最も低い等高線より下に位置する点は全体の5%、その次の等高線より下に位置するのは10%、というように続いていきます。一番外側の等高線には点の約95%が含まれます。

図 5.14 ノンパラメトリック密度の例



ノンパラメトリック密度等高線を描くグリッド点の個数を変更するには、Shift キーを押しながら「二変量の関係」の赤い三角ボタンをクリックし、メニューから[ノンパラメトリック密度]を選択してください。デフォルトの102よりも大きい値を入力できます。

[分位点密度等高線] メニューのオプションについては、「[二変量の関係のオプション](#)」を参照してください。

**ヒント:** 大量のデータ点があり、データ点が重なって見えにくくなっている場合は、[ノンパラメトリック密度]オプションを使うと散布図上のパターンが見やすくなります。

「分位点密度等高線」レポートには、ノンパラメトリック密度の計算に使用された標準偏差が表示されます。

**変数** 等高線の作成に使用された変数の名前。

**カーネル標準偏差** カーネル密度推定のバンド幅。最初に表示される値は、変数の標準偏差に基づいて計算されたものです。スライダを使って等高線の密度を調整することができます。

## 「二変量」プラットフォームの別例

この節では、「二変量」プラットフォームを使った例をご紹介します。

- 「[その他のあてはめ] オプションの例」
- 「[直交のあてはめ] オプションの例」
- 「[ロバストなあてはめ] オプションの例」
- 「グループ別確率楕円の例」
- 「グループ別回帰直線の例」
- 「By 変数を使ってグループ別に分析する例」

### [その他のあてはめ] オプションの例

この例では、「二変量」プラットフォームで変数を変換し、回帰モデルをあてはめる方法を解説します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「SAT.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「2004 言語」を選択し、[Y, 目的変数] をクリックします。
4. 「2004 受験率(%)」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「二変量の関係」の赤い三角ボタンをクリックし、[その他のあてはめ] を選択します。「変換または制約を指定」ウィンドウが表示されます。このウィンドウの説明については、「[\[その他のあてはめ\] のウィンドウ](#)」を参照してください。
7. 「Yの変換」で、[自然対数:  $\log(y)$ ] を選択します。
8. 「Xの変換」で、[平方根:  $\sqrt{x}$ ] を選択します。

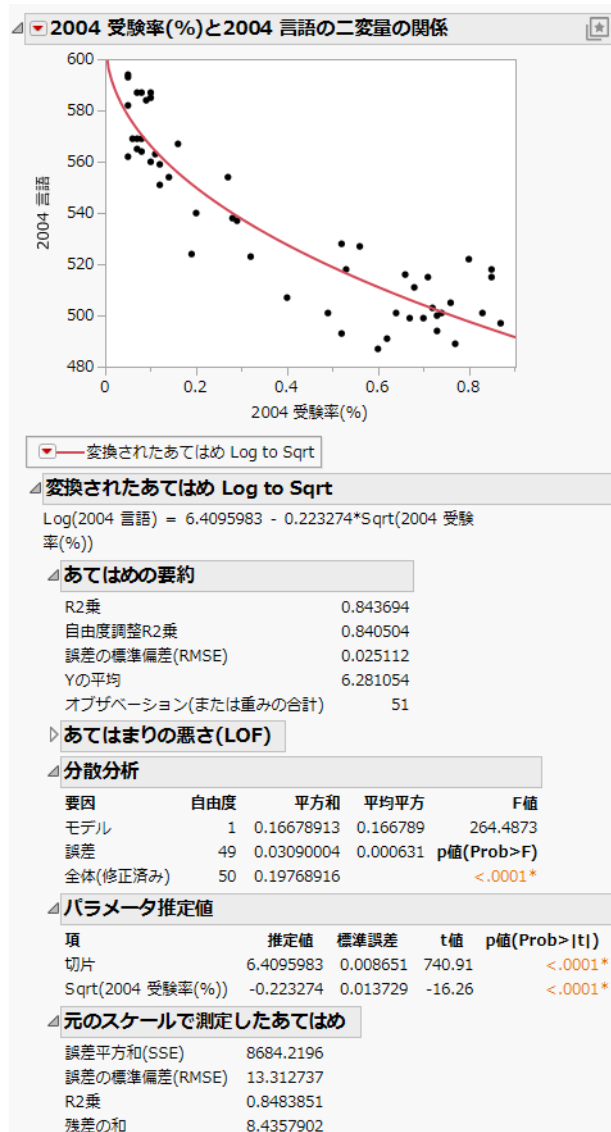
図5.15 「変換または制約を指定」ウィンドウ



9. [OK] をクリックします。



図 5.16 その他のあてはめのレポートの例



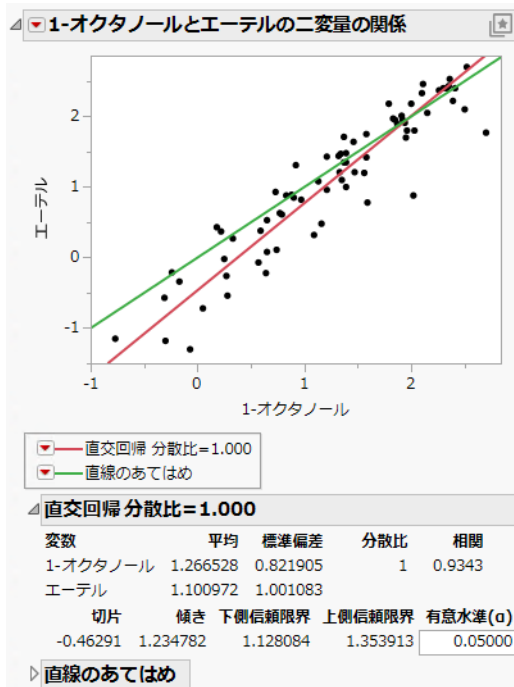
モデルはデータに良くあてはまっているようです。元のスケールでプロットを見ると、あてはめ線の近くにデータが分布しています。また、モデルのR2乗は0.84となっています。

## 「直交のあてはめ」オプションの例

この例では、「二変量」プラットフォームにて、直交回帰モデルをあてはめます。切片0で傾き1のモデルと比較することにより、直交回帰モデルは、同じ対象を2つの異なる測定方法で測定した場合にも使えます。そのような場合には、Passing-Bablok回帰も使えます。「[「Passing-Bablokのあてはめ」レポート](#)」を参照してください。この例では、いくつかの化合物の溶解度を異なる溶剤で測定しました。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Solubility.jmp」を開きます。
  2. [分析] > [二変量の関係] を選択します。
  3. 「エーテル」を選択し、[Y, 目的変数] をクリックします。
  4. 「1-オクタノール」を選択し、[X, 説明変数] をクリックします。
  5. [OK] をクリックします。
  6. 「二変量の関係」の赤い三角ボタンをクリックし、[直交のあてはめ] > [等分散性] を選択します。
  7. 参考のために、「二変量の関係」の赤い三角ボタンをクリックし、[その他のあてはめ...] を選択します。
  8. [切片の制約] をオンにして値を0に設定します。
  9. [傾きの制約] をオンにして値を1に設定します。
- これで、切片が0、傾きが1の  $Y = X$  の直線があてはめられます。
10. [OK] をクリックします。

図 5.17 直交のあてはめの例



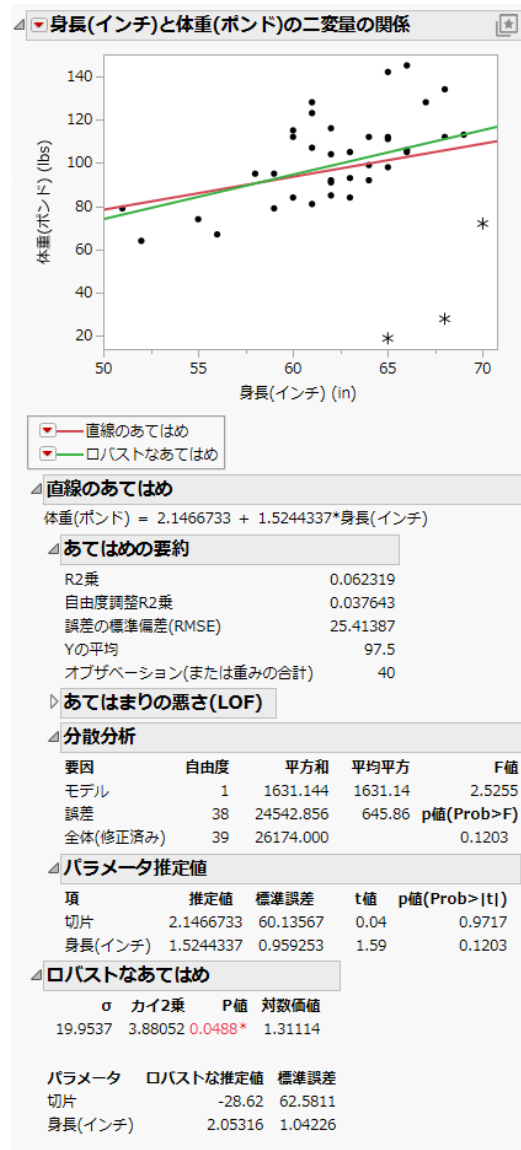
得られたデータの大部分の範囲で、直交回帰の予測値を示す赤い直線が、緑の  $Y = X$  線を下回っています。なお、直交回帰の表を見ると、傾きの95%信頼区間（下側限界と上側限界の間）には1が含まれていません。2つの直線を見てみると、得られたデータの大部分の範囲において、溶剤がエーテルでの溶解度は、1-オクタノールでの溶解度より低い傾向にあると言えます。

## 【ロバストなあてはめ】オプションの例

ここでは、「二変量」プラットフォームでHuberのM推定を使用して、ロバストなモデルをあてはめる方法を紹介します。

1. **【ヘルプ】 > 【サンプルデータフォルダ】** を選択し、「Weight Measurements.jmp」を開きます。
2. **【分析】 > 【二変量の関係】** を選択します。
3. 「体重(ポンド)」を選択し、**【Y, 目的変数】** をクリックします。
4. 「身長(インチ)」を選択し、**【X, 説明変数】** をクリックします。
5. **【OK】** をクリックします。
6. 「二変量の関係」の赤い三角ボタンをクリックし、**【直線のあてはめ】** を選択します。
7. 「二変量の関係」の赤い三角ボタンをクリックし、**【ロバスト】 > 【ロバストなあてはめ】** を選択します。

図 5.18 ロバストなあてはめの例



散布図を見ると、3つの測定値が星印で表示されています。この3つの測定値は、体重が回帰直線よりかなり低くなっており、また、回帰直線（赤色の直線）の推定に強く影響を与えています。「分散分析」表の $p$ 値が0.1203であることから、「身長(インチ)」と「体重(ポンド)」の間に線形関係がある」と主張することは難しいです。しかし、ロバストなあてはめ（緑色の直線）では、「身長(インチ)」と「体重(ポンド)」の間に、通常の回帰分析の時よりも、強い線形関係が見られます。「ロバストなあてはめ」の $p$ 値は0.0488であり、「身長(インチ)」と「体重(ポンド)」の間に線形関係がある」という仮説を支持します。つまり、体重が予測値よりも非常に軽いいくつかのデータ点が分析に影響を与えていたことがわかります。

## グループ別確率楕円の例

ここでは、「二変量」プラットフォームでグループ変数（By 変数）を使って確率楕円を追加する方法を紹介します。この例で使用するデータテーブルは、beef（牛肉）、meat（牛・豚肉）、poultry（鳥肉）を食材とした3種類のホットドッグについて調べたデータです。3種類のホットドッグでグループ分けして、コストを分析します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Hot Dogs.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「1オンスあたりのコスト(\$)」を選択し、[Y, 目的変数] をクリックします。
4. 「蛋白質1ポンドあたりのコスト(\$)」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「二変量の関係」の赤い三角ボタンをクリックし、[グループ別...] を選択します。
7. リストから「タイプ」を選択します。
8. [OK] をクリックします。

もう一度 [グループ別...] オプションを確認すると、横にチェックマークが付いていることがわかります。

9. 「二変量の関係」の赤い三角ボタンをクリックし、[確率楕円] > [0.90] を選択します。

点を「タイプ」別に色分けするには、次の手順に従います。

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

図5.19 [グループ別] オプションの例

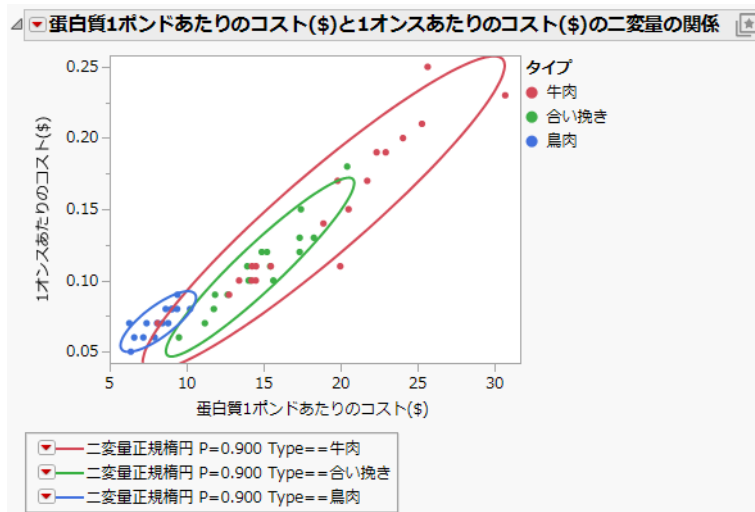


図5.19は、異なる種類のホットドッグの確率楕円を、コストを変数としてプロットしたものです。種類ごとに点がクラスターを形成しているのがわかります。

## グループ別回帰直線の例

この例では、「二変量」プラットフォームでグループ変数を使ってグループ別の回帰直線を重ねて表示し、各グループの傾きを比較します。

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

図5.20 左側の回帰直線を描画するには、次の手順に従います。

6. 「二変量の関係」の赤い三角ボタンをクリックし、[直線のあてはめ] を選択します。

図5.20 右側の回帰直線を描画するには、次の手順に従います。

7. [直線のあてはめ] メニューから [あてはめの削除] を選択します。
8. 「二変量の関係」の赤い三角ボタンをクリックし、[グループ別...] を選択します。
9. リストから「性別」を選択します。
10. [OK] をクリックします。
11. 「二変量の関係」の赤い三角ボタンをクリックし、[直線のあてはめ] を選択します。

図5.20 データ全体とグループ別の回帰分析の例

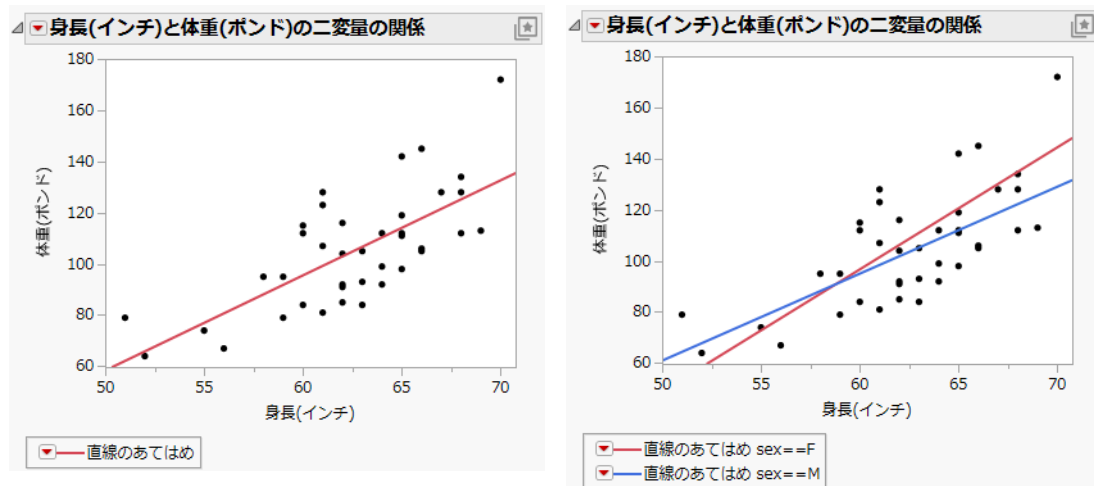


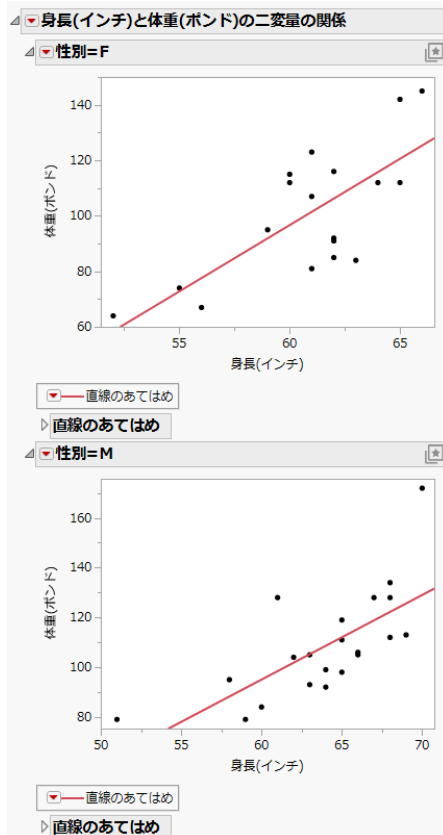
図5.20の左側の散布図には、データ全体から推定された回帰直線が描かれています。右側の散布図に描かれている回帰直線は、それぞれ女子および男子のデータから推定されています。

## By 変数を使ってグループ別に分析する例

グループ別に分析を実行するには、「二変量の関係」起動ウィンドウで By 変数を指定するという方法もあります。この方法では、By 変数の水準ごと（または By 変数の水準の組み合わせごと）に、レポートと散布図が生成されます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「体重(ポンド)」を選択し、[Y, 目的変数] をクリックします。
4. 「身長(インチ)」を選択し、[X, 説明変数] をクリックします。
5. 「性別」を選択し、[By] をクリックします。
6. [OK] をクリックします。
7. 「二変量の関係」の赤い三角ボタンをクリックし、[直線のあてはめ] を選択します。

図 5.21 By 変数を指定した散布図の例



By 変数（性別）の水準ごとに分析が実行され、女子生徒と男子生徒それぞれの散布図が表示されます。

## 「二変量」プラットフォームの統計的詳細

この節では、「二変量」プラットフォームで使われている統計手法の詳細を説明します。

- 「[直線のあてはめ] オプションの統計的詳細」
- 「[多項式のあてはめ] オプションの統計的詳細」
- 「[スプライン曲線のあてはめ] オプションの統計的詳細」
- 「[直交のあてはめ] オプションの統計的詳細」
- 「[ロバスト] オプションの統計的詳細」
- 「[Passing Bablokのあてはめ] オプションの統計的詳細」
- 「[あてはめの要約] レポートの統計的詳細」
- 「[あてはまりの悪さ (LOF)] レポートの統計的詳細」
- 「[パラメータ推定値] レポートの統計的詳細」
- 「[平滑化スプライン曲線によるあてはめ] レポートの統計的詳細」
- 「[二変量正規楕円] レポートの統計的詳細」

### [直線のあてはめ] オプションの統計的詳細

「二変量」プラットフォームの「直線のあてはめ」オプションは、残差平方和（誤差平方和）がなるべく小さくなるようなパラメータ  $\beta_0$  および  $\beta_1$  を推定することで、データ点に直線をあてはめます。なお、 $i$  番目の行に対して、直線のモデルは  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  と表されます。

直線のあてはめにおける平均の信頼区間は、次のように定義されます。

$$\hat{y}_i \pm t_{\alpha/2, n-2} \left[ \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}} \right] \hat{\sigma}$$

直線のあてはめにおける個々の応答値の信頼区間は、次のように定義されます。

$$\hat{y}_i \pm t_{\alpha/2, n-2} \left[ \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}} \right] \hat{\sigma}$$

ここで、

$\hat{y}_i$  は、 $i$  番目の観測（オブザベーション）である  $x_i$  における予測値

$\bar{x}$  は、 $x$  変数の平均



$$\hat{\sigma} = \sqrt{\frac{S_{yy} - \beta_1 S_{xy}}{n - 2}}$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$S_{xy} = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

$t_{\alpha/2, n}$  は、 $n-2$  個の自由度を持つ  $t$  分布の上側  $100\alpha/2\%$  点

$n$  は、標本サイズ。

---

**メモ：** [平均の信頼限界の計算式] と [個別の信頼限界の計算式] オプションを使うと、モデルのあてはめの信頼限界の計算式をデータテーブルの新しい列に保存できます。計算式は行列形式で保存されます。

---

## [多項式のあてはめ] オプションの統計的詳細

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

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

ここで、

$y_i$  は、 $i$  番目の観測（オブザベーション）である  $x_i$  における応答値

$\beta_j$  は、次数  $j$  である項に対するパラメータ

$k$  は、多項式の次数

$\varepsilon_i$  は、モデルの誤差項

## [スプライン曲線のあてはめ] オプションの統計的詳細

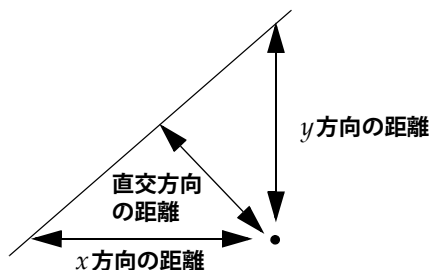
「二変量」プラットフォームの3次の平滑化スプライン法では、複数の3次多項式を、つなぎ目（節点）においても滑らかで連続な曲線になるように、つなぎ合わせます。また、ペナルティ（曲線の曲がり具合を、曲線全体の範囲で積分したもの）と誤差平方和の合計が最小になるような推定値を求めます。この手法の詳細については、Reinsch（1967）の論文か Eubank（1999）の本を参照してください。

## 「直交のあてはめ」オプションの統計的詳細

「二変量」プラットフォームの「直交のあてはめ」オプションは、直交方向の距離の平方和を最小にするような直線をあてはめます。これは、Xの測定値にランダムなばらつきがある場合に標準的な最小2乗法より有効です。標準的な最小2乗法では、X変数が所与で、Xの関数に誤差を加えたものがY変数であると仮定されます。

**メモ:** 直交方向に引かれた垂線の長さは、XとYのスケールに左右されます。XとYのスケールをどのように決めるべきかということは、単なるグラフ上の問題ではなく、統計的な問題です。

図 5.22 回帰直線への垂線



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

この直交のあてはめが最も使われるのは、2つの測定システム（測定器）で、同じ対象を測定したときです。この場合、XとYの測定値には、同じような種類の測定誤差があります。誤差の比は、1（等分散性）のように仮定される値であるか、測定システムに関する知識に基づいて計算されます。

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

## 【ロバスト】オプションの統計的詳細

この節では、「二変量」プラットフォームの「ロバストなあてはめ」と「Cauchyのあてはめ」の各オプションについて詳しく解説します。

### ロバストなあてはめ

「二変量」プラットフォームの「ロバストなあてはめ」オプションは、応答変数の外れ値がモデルのあてはめに与える影響を少なくします。HuberのM推定が使用されます。HuberのM推定では、次式によって定義されているHuberの損失関数を最小にするパラメータ推定値を求めます。

$$l(e) = \sum_i \rho(e_i)$$

この式で、

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & |e| < k \text{ の場合} \\ k|e| - \frac{1}{2}k^2 & |e| \geq k \text{ の場合} \end{cases}$$

$e_i$ は残差を表します

Huberの損失関数は、誤差が小さい場合は2次曲線として増加し、誤差が大きい場合は線形に増加します。JMPのアルゴリズムでは、 $k=2$ としています。ロバスト推定（頑健推定）の詳細については、Huber（1973）およびHuber and Ronchetti（2009）を参照してください。

### Cauchyのあてはめ

「二変量」プラットフォームの「Cauchyのあてはめ」オプションは、最尤推定によってパラメータ推定値を算出します。このオプションでは、誤差がCauchy分布に従うと仮定されます。Cauchy分布は正規分布よりも裾が広く、その結果、外れ値が推定に与える影響が小さくなります。

## 「Passing Bablok のあてはめ」 オプションの統計的詳細

「二変量」プラットフォームの「Passing Bablok のあてはめ」オプションは、 $y = \beta_0 + \beta_1 x$  という線形式をあてはめます。このモデルでは、 $x$  と  $y$  がどちらも誤差を伴って測定されていると仮定されます。この手法は外れ値の影響をあまり受けず、 $x$  と  $y$  の間に線形関係（1 次式の直線関係）がある場合に適しています。傾き（ $\beta_1$ ）の推定値は、データ点の可能な組み合わせが形成するすべての傾きの中央値です。ただし、傾きが 0/0 または -1 となるペアは計算から除外されます（そのようなペアは、データに稀にしかないと考えられます）。これらの傾きから得られる推定値には、独立性がない場合には推定バイアスが生じます（不偏推定値でなくなります）。そのバイアスを修正するため、-1 より小さい傾きの個数  $K$  に基づく係数だけ中央値がシフトされます。その結果、 $\beta_1$  に対するほぼ不偏な推定値が求められます。切片（ $\beta_0$ ）は、 $\{y_i - \beta_1 x_i\}$  の中央値で推定されます。Passing and Bablok (1983)、Passing and Bablok (1984)、Bablok et. al. (1988) を参照してください。

Passing-Bablok 法は、2 つの異なる測定方法を比較する場合によく利用されます。切片は 2 つの測定方法の「体系的なバイアス」（systematic bias）、傾きは 2 つの測定方法の「比例的なバイアス」と解釈できます。

## 「あてはめの要約」 レポートの統計的詳細

この節では、「二変量」プラットフォームの「あてはめの要約」レポートについて詳しく解説します。

### R2 乗

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

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

### 自由度調整 R2 乗

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

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

「誤差」の平均平方は、「分散分析」レポートに表示されています（図 5.11）。分母の平均平方は、「全体（修正済み）」の平方和を、対応する自由度で割ったものです。

---

**メモ：**切片と傾きのどちらかに制約がある場合、「R2 乗」と「自由度調整 R2 乗」は計算されません。

---

## 「あてはまりの悪さ (LOF)」レポートの統計的詳細

この節では、「二変量」プラットフォームの「あてはまりの悪さ (LOF)」レポートについて詳しく解説します。

### 「純粋誤差」の自由度

たとえば、身長が同じ生徒が複数名いるとします。一般に、X 値が同じであるグループが  $g$  個あるとき、純粋誤差の自由度 ( $DF_p$ ) は、次式のように各グループの自由度を足し合わせることで求められます。

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

ここで、 $n_i$  は  $i$  番目のグループのオブザベーション数を表します。

### 「純粋誤差」の平方和

一般に、X 値が同じであるグループが  $g$  個あるとき、純粋誤差の平方和 ( $SS_p$ ) は、次式のように各グループの平方和を足し合わせることで求められます。

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

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

### 最大 R2 乗

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

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

## 「パラメータ推定値」レポートの統計的詳細

この節では、「二変量」プラットフォームの「パラメータ推定値」レポートについて詳しく解説します。

### 標準 $\beta$

標準  $\beta$ （標準化回帰係数）は、次の式で計算されます。

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

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

## 計画の標準誤差

「計画の標準誤差」は、パラメータ推定値の標準誤差をRMSEで割ったものです。

## 「平滑化スプライン曲線によるあてはめ」レポートの統計的詳細

この節では、「二変量」プラットフォームの「平滑化スプライン曲線によるあてはめ」レポートについて詳しく解説します。

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

## 「二変量正規楕円」レポートの統計的詳細

この節では、「二変量」プラットフォームの「二変量正規楕円」レポートについて詳しく解説します。

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

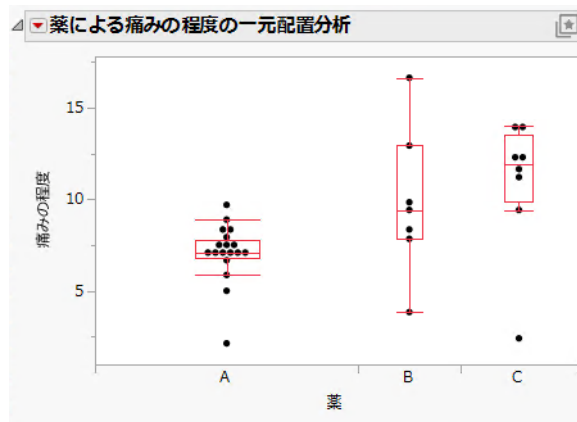
$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \quad \text{ここで、} s_{xy} = \frac{\sum w_i (x_i - \bar{x})(y_i - \bar{y})}{df}$$

この式の $w_i$ は、「重み」列を指定してあるときは $i$ 番目の観測（オブザベーション）における重みであり、「重み」列がないときは1です。

## 一元配置分析

「一元配置」プラットフォームでは、X変数のグループごとに、連続尺度のY変数の分布がどのように異なるかを分析します。グループ間で平均または分散の差を調べることができます。ここで用意されている統計分析には、分散分析（ANOVA）・平均分析（ANOM）・t検定・同等性検定・ノンパラメトリックな検定・正確検定・F検定・平均の多重比較検定があります。確率密度関数プロットや累積確率プロットを使えば、X変数のグループごとの応答の分布を視覚的に確認できます。たとえば、薬剤の種類（X）によって、連続尺度で測定した痛みの程度（Y）がどのように異なるかを調べることができます。

图 6.1 一元配置分析



## 目次

「一元配置」プラットフォームの概要	154
一元配置分析の例	154
「一元配置」プラットフォームの起動	157
データ形式	158
「一元配置」レポート	158
「一元配置」プラットフォームのオプション	159
「一元配置分析」レポート	168
「分位点」レポート	169
「平均/ANOVA/プーリングしたt検定」レポート	169
「平均と標準偏差」レポート	171
「t検定」レポート	172
平均分析レポート	173
平均の比較レポート	174
ノンパラメトリックな検定のレポート	177
ノンパラメトリックな多重比較のレポート	181
等分散性の検定レポート	183
同等性検定のレポート	185
ロバストなあてはめのレポート	188
検出力のレポート	189
「対応のあるあてはめ」レポート	190
一元配置プロットの要素	191
平均のひし形と標本サイズに比例したX軸	191
平均線・平均誤差バー・標準偏差線	192
比較円	193
「一元配置」プラットフォームの別例	195
平均分析の例	195
分散の平均分析の例	196
[各ペア, Studentのt検定]の例	197
[すべてのペア, TukeyのHSD検定]の例	199
[最適値との比較, HsuのMCB検定]の例	201
[コントロール群との比較, Dunnett検定]の例	202
[各ペアのステップワイズ, Newman-Keuls検定]の例	204
「平均の比較」で用意されている4種類の検定	204
Wilcoxon検定の例	205
等分散性検定の例	207
同等性の検定の例	208
[ロバストなあてはめ] オプションの例	210
[検出力] オプションの例	212



正規分位点プロットの例 .....	213
累積確率プロットの例 .....	214
[密度] の各オプションの例 .....	215
[対応のある列を設定] オプションの例 .....	216
データを積み重ねて一元配置分析を実行する例 .....	218
「一元配置」プラットフォームの統計的詳細 .....	224
比較円の統計的詳細 .....	224
検出力の統計的詳細 .....	226
平均分析の統計的詳細 .....	226
「あてはめの要約」レポートの統計的詳細 .....	227
等分散性検定の統計的詳細 .....	227
ノンパラメトリックな検定の統計量の統計的詳細 .....	228
[ロバストなあてはめ] の統計的詳細 .....	231

## 「一元配置」プラットフォームの概要

「一元配置」プラットフォームでは、連続尺度のY変数がカテゴリカルなX変数の複数の水準間でどのように分布しているかをすばやく視覚化できます。グループ間で平均または分散の差を調べることができます。手法には、分散分析（ANOVA）、平均分析（ANOM）、 $t$ 検定、同等性検定、ノンパラメトリックな検定、正確検定、 $F$ 検定、平均の多重比較検定があります。ほとんどの検定のレポートで、「一元配置」プロットに視覚化ツールが用意されています。また、確率密度関数プロットや累積確率プロットを使ってX変数の水準間での応答の分布を視覚化するオプションもあります。

検定は、「一元配置」の赤い三角ボタンのメニューに種類別にまとめられています。JMPでは、さまざまな種類の統計検定を試すことができます。どの検定が最も適しているかは、研究のデザインやデータの収集方法、データの分布、仮説によって異なります。

検定のレポートについては、次の節を参照してください。

- 「「分位点」レポート」
- 「「平均/ANOVA/プーリングした $t$ 検定」レポート」
- 「「平均と標準偏差」レポート」
- 「「 $t$ 検定」レポート」
- 「平均分析レポート」
- 「平均の比較レポート」
- 「ノンパラメトリックな検定のレポート」
- 「ノンパラメトリックな多重比較のレポート」
- 「等分散性の検定レポート」
- 「同等性検定のレポート」
- 「ロバストなあてはめのレポート」
- 「検出力のレポート」
- 「「対応のあるあてはめ」レポート」

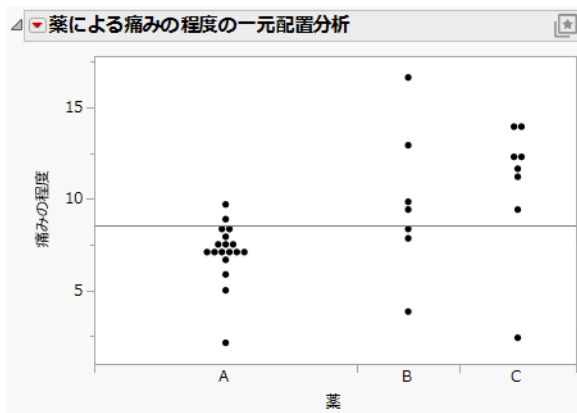
## 一元配置分析の例

この例では、「一元配置」プラットフォームを使って一元配置分散分析を行います。一元配置分散分析では、グループ平均の間に統計的に有意な差があるかどうかを検定することができます。33人の被験者が3種類の鎮痛剤（A、B、C）を服用したデータがあり、痛みの程度を数値で評価しています。ここでは、A、B、Cの鎮痛剤において、痛みの程度の平均に有意な差があるかどうかを調べます。

1. [ヘルプ] > [サンプルデータフォルダ] 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。
5. **[OK]** をクリックします。

図 6.2 一元配置分析の例



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

データの分散分析を行います。

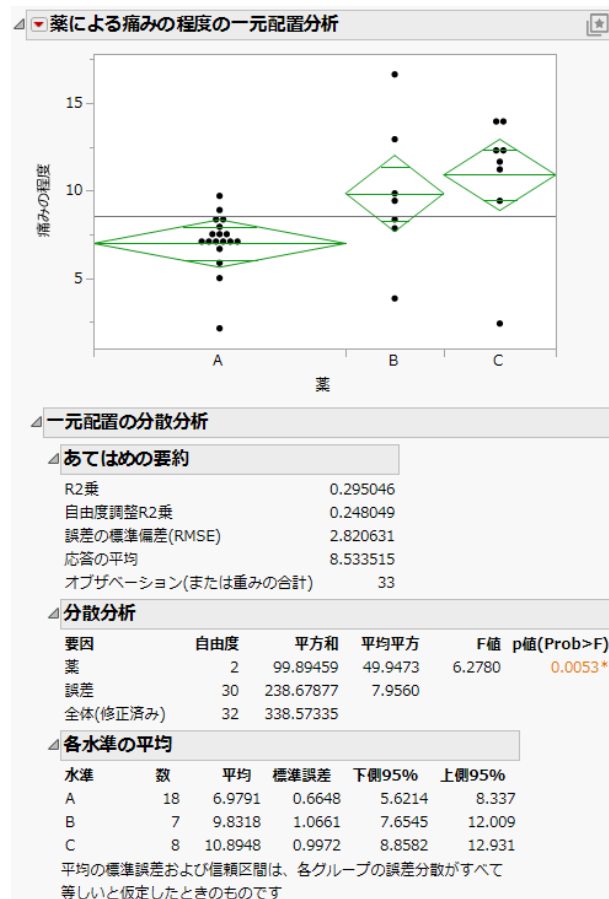
6. 赤い三角ボタンをクリックし、**[平均/ANOVA]** を選択します。

---

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

---

図 6.3 [平均/ANOVA] オプションの例



次の点を確認してください。

- プロットに信頼区間を示す平均のひし形が表示されます。
  - 各ひし形の中央に引かれている水平線はグループ平均を示します。これを目安にすれば、薬Aの平均が薬BおよびCの平均より低いことが一目でわかります。
  - ひし形の縦の長さがグループ平均の95%信頼区間を表します。

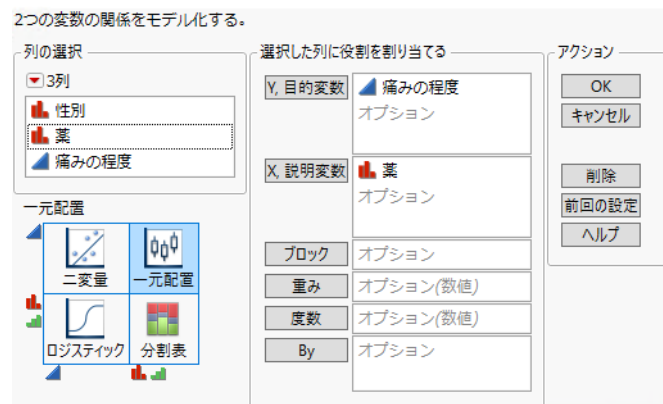
「平均のひし形と標本サイズに比例したX軸」を参照してください。

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

## 「一元配置」プラットフォームの起動

「一元配置」プラットフォームを起動するには、[分析] > [二変量の関係] を選択します。「二変量の関係」起動ウィンドウでは、4種類の分析が可能です。連続尺度のY変数と名義尺度または順序尺度のX変数を入力すると、自動的に「一元配置」プラットフォームが開きます。X変数とY変数の組み合わせごとにレポートが作成されます。

図6.4 一元配置の起動ウィンドウ



「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。起動ウィンドウには、次のようなオプションがあります。

**Y, 目的変数** 分析したい目的変数（応答変数）。変数のデータタイプは「数値」でなければなりません。

**X, 説明変数** 分析したい説明変数（予測変数）。変数のデータタイプは名義尺度または順序尺度でなければなりません。

**ブロック** ブロック変数とする列。指定すると、Y変数の値がブロック変数ごとに中心化されます。ブロックごとの各グループのセル度数が等しい場合、「一元配置」プラットフォームのいくつかのオプションは、二元配置分散分析として動作します。ブロックごとのグループのセル度数が異なる場合、ブロックがある一元配置分散分析モデルが当てはめられ、「一元配置」プラットフォームのオプションは使用できません。

**重み** データテーブルの各行に対する重みを含んだ列。値が0より大きい行だけが分析に用いられます。

**度数** 分析において各データ行に度数を割り当てます。分析対象のデータが要約されたものであるときに役立ちます。

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

## データ形式

「一元配置」プラットフォームでは、要約されていないデータ、および要約されているデータのいずれも分析できます。

**要約されていないデータ** 観測値ごとに1行ずつあり、Xの値が1列、Yの値が1列に入力されています。

**要約されたデータ** XとYの値の組み合わせが同じである複数の観測値が1行にまとめられているデータです。データテーブルには、各行の観測値の数を示す度数列が必要です。起動ウィンドウで、この列を「度数」に指定してください。

分散分析のデータでは、JMP以外の形式では、1行に複数の観測値が含まれていることもあります。JMPでそのような形式のデータを分析するには、データを読み込んでから、各行が1つの観測値を含む形式に構成し直す必要があります。「[データを積み重ねて一元配置分析を実行する例](#)」を参照してください。

---

**メモ:** 「二変量の関係」起動ウィンドウでは、連続尺度・順序尺度・名義尺度の列が使用できます。連続尺度の目的変数(Y)と名義尺度または順序尺度の説明変数(X)を指定すると、「一元配置」プラットフォームが開きます。それ以外の組み合わせでは、「二変量の関係」・「分割表」・「ロジスティック」のいずれかのプラットフォームが起動します。

---

---

## 「一元配置」レポート

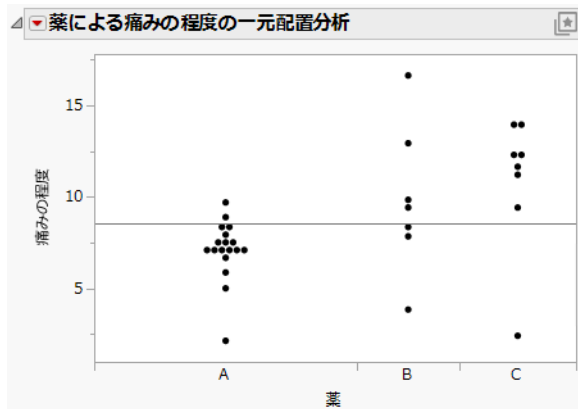
「一元配置」レポートには、まず、名義尺度または順序尺度のX変数と連続尺度のY変数の組み合わせごとに散布図が描かれます。赤い三角ボタンのメニューには、モデルのあてはめ、仮説の検定、同等性の検定、分布の比較、統計レポートの表示を行うためのオプションがあります。「[「一元配置」プラットフォームのオプション](#)」を参照してください。

---

**メモ:** 起動ウィンドウでブロック変数を指定した場合、一元配置プロットのY変数の値はブロック変数ごとに中心化されます。ブロック変数を使った分析では、使用できるオプションが異なります。

---

図6.5 一元配置プロット



### 変数を対話的に置換する

データテーブルの「列」パネルで該当する変数をクリックし、プロットの軸上までドラッグして変数を置換することもできます。

## 「一元配置」プラットフォームのオプション

「一元配置分析」の赤い三角ボタンのメニューには、検定・あてはめ・表示に関するオプションが用意されています。オプションには、分散分析（ANOVA）、ノンパラメトリックな検定、同等性の検定、グループ間の差を評価する多重比較の手法などがあります。確率密度関数プロットや累積確率プロットを使えば、X変数のグループごとの応答の分布を視覚的に確認できます。オプションにより、散布図プロットに視覚的な要素を追加したり、統計的なレポートを作成したり、追加の分析を実行したりできます。

ブロック変数を指定した場合、ブロックがある一元配置分散分析モデルがあてはめられます。なお、ブロック変数を指定した場合には、追加で使用できる分析オプションは限られています。

**メモ：** Y変数またはX変数を複数指定した場合は、「あてはめのグループ」メニューが表示されます。「あてはめのグループ」メニューのオプションを使うと、レポートの配置を変更したり、適合度の順に並べたりできます。『基本的な回帰モデル』を参照してください。

「一元配置分析」の赤い三角ボタンのメニューには、以下のオプションがあります。

**分位点** 「一元配置」レポートの先頭にあるグラフにおいて、箱ひげ図の表示／非表示を切り替えます。また、分位点レポートの表示／非表示を切り替えます。「[「一元配置分析」レポート](#)」を参照してください。

**平均/ANOVA** (X変数の水準数が3水準以上の場合のみ使用できます。)「一元配置」レポートの先頭にあるグラフにおいて、平均のひし形の表示／非表示を切り替えます。また、「分散分析」レポートの表示／非表示を切り替えます。「[「平均/ANOVA/プーリングしたt検定」レポート](#)」を参照してください。

**平均/ANOVA/プーリングしたt検定** (X変数の水準数が2水準の場合のみ使用できます。)  
「一元配置」レポートの先頭にあるグラフにおいて、平均のひし形の表示／非表示を切り替えます。また、「分散分析」レポートの表示／非表示を切り替えます。「一元配置の分散分析」レポートには、「プーリングしたt検定」レポートが含まれます。この検定では、2つのグループの分散が等しいと仮定されます。

**平均と標準偏差** 「一元配置」レポートの先頭にあるグラフにおいて、平均・誤差バー・標準偏差を示す線の表示／非表示を切り替えます。また、要約統計量表の表示／非表示を切り替えます。このレポートでの平均に対する標準誤差は、個々のグループごとの標準偏差を使用します。さらに詳しい説明については、「[平均線・平均誤差バー・標準偏差線](#)」を参照してください。

**個々の分散を用いたt検定** (X変数の水準数が2水準の場合のみ使用できます。) 異分散を仮定したt検定 (Welch検定) の表示／非表示を切り替えます。「[t検定レポート](#)」を参照してください。

**平均分析法** 平均分析 (ANOM) 手法を使って複数のグループを比較するオプションが用意されています。「[平均分析レポート](#)」を参照してください。平均分析手法について詳しくは、Nelson et al. (2005) を参照してください。

---

**メモ:** 起動ウィンドウでブロック変数を指定し、ブロック変数とX変数の水準の組み合わせごとの度数が同じである場合、平均分析グラフのみを使用できます。度数が等しくない場合、平均分析のオプションは使用できません。

---

**平均分析** グループ平均を全体平均と比較する平均分析チャートの表示／非表示を切り替えます。この分析法では、データがほぼ正規分布に従っているものと仮定します。「[平均分析の例](#)」を参照してください。

**順位変換平均分析** グループ平均をノンパラメトリックな手法で比較した順位変換平均分析チャートの表示／非表示を切り替えます。各グループで計算した順位平均を、全体の順位平均と比較します。[順位変換平均分析] では、順位を正規スコアに変換したデータに対して、通常のアノマの順序と棄却値が適用されます。

---

**ヒント:** この方法は、データが明らかに正規分布に従っておらず、また、変数変換によって正規分布に変換できない場合に使用することが考えられます。

---

**分散の平均分析** (X変数の水準数が3水準以上で、どのグループにも4つ以上の観測 (オブザベーション) がある場合のみ使用できます。) グループの標準偏差 (または分散) を誤差の標準偏差 (または誤差の平均平方) と比較する分散の平均分析 (ANOMV) チャートの表示／非表示を切り替えます。この分析法では、データがほぼ正規分布に従っているものと仮定します。分散の平均分析の詳細については、Wludyka and Nelson (1997) および Nelson et al. (2005) を参照してください。例として、「[分散の平均分析の例](#)」を参照してください。

**分散の平均分析-LeveneのADM** 非正規性に対してロバスト (頑健) なANOMV-LEVチャートの表示／非表示を切り替えます。この手法は、**中央値からの偏差の絶対値** (ADM; absolute deviations from the median) のグループ平均と、ADMの全体平均を比較します。詳細については、Levene (1960) または Brown and Forsythe (1974) を参照してください。



---

**ヒント:** この手法は非正規性に対してよりロバスト（頑健）であるため、データが正規分布に従わず、また、変数変換によって正規分布に変換できないと思われる場合に使用することが考えられます。Levene の ADM を用いた方法は、外れ値と非正規性に対してロバスト（頑健）です。

---

**範囲の平均分析**（バランス（釣り合い）のとれたデータ、および特定の標本サイズにのみ使用できます。）グループの範囲をその平均と比較する、範囲の平均分析チャートの表示／非表示を切り替えます。この検定は、尺度の指標として範囲を用いて、各グループのばらつきに差があるかどうかを検定します。Wheeler（2003）を参照してください。「[範囲の平均分析](#)」の制限」を参照してください。

**平均の比較** [平均の比較] メニューには、平均に対する多重比較を行うオプションがいくつか用意されています。「[平均の比較レポート](#)」を参照してください。

**各ペア, Student の t 検定** Fisher の LSD 法の結果の表示／非表示を切り替えます。Fisher の LSD 法は、すべてのペアの平均を比較する方法です。このオプションにより、一元配置プロットにおける比較円の表示／非表示も切り替えられます。

**すべてのペア, Tukey の HSD 検定** Tukey-Kramer 法の結果の表示／非表示を切り替えます。Tukey-Kramer 法は、すべてのペアの平均を比較し、かつ、その多重性を調整する方法です。このオプションにより、一元配置プロットにおける比較円の表示／非表示も切り替えられます。

**最適値との比較, Hsu の MCB 検定** Hsu 法の結果の表示／非表示を切り替えます。Hsu 法は、最大や最小となっている平均と他の平均を比較し、かつ、その多重性を調整する方法です。このオプションにより、一元配置プロットにおける比較円の表示／非表示も切り替えられます。

**コントロール群との比較, Dunnett 検定** Dunnett 法の結果の表示／非表示を切り替えます。Dunnett 法は、コントロール群をそれ以外のすべての群と比較し、かつ、その多重性を調整する方法です。このコマンドにより、一元配置プロットにおける比較円の表示／非表示も切り替えられます。

---

**ヒント:** 因子の列に「コントロール群」列プロパティを設定しておくと、[コントロール群との比較] を選択するたびにコントロール群を指定する必要がなくなります。『JMP の使用法』を参照してください。

---

**各ペアのステップワイズ, Newman-Keuls 検定** 「Newman-Keuls ステップワイズ法による各ペアの比較」レポートの表示／非表示を切り替えます。Newman-Keuls 検定では、検定全体における過誤率（family wise error rate）が守られていません。

**ノンパラメトリック** グループの位置をノンパラメトリックに比較するためのオプションがあります。「[ノンパラメトリックな検定のレポート](#)」を参照してください。

**Wilcoxon / Kruskal-Wallis 検定** Wilcoxon の順位スコアに基づく検定の表示／非表示を切り替えます。Wilcoxon の順位スコアは、データの順位そのものです。Wilcoxon の検定は、誤差がロジスティック分布に従っている場合に、最も検出力の高い順位検定となります。

X 変数の水準数が2水準であるとき、Wilcoxon 検定はMann-Whitney 検定と等価です。また、Wilcoxon 検定を3群以上に拡張した検定が、Kruskal-Wallis 検定です。X 変数の水準数が2水準であるとき、0.5の連続修正を行った結果が「Wilcoxon 2標本検定 (正規近似)」というタイトルで、連続修正を行わない結果が「Kruskal-Wallis 検定 (カイ2乗近似)」というタイトルで表示されます。これら2つの結果は、連続修正を行わなかったか、行わなかったかだけの違いです。

X 変数に3つ以上の水準があるときは、「Kruskal-Wallis 検定 (カイ2乗近似)」だけが表示されます。この結果は、通常の順位を使ったKruskal-Wallis 検定の枠組みでカイ2乗近似でp値を求めており、連続修正も行われていません。

このレポートについては、「[Wilcoxon / Kruskal-Wallis 検定、メディアン検定、Friedman 順位検定、Van der Waerden 検定のレポート](#)」を参照してください。例として、「[Wilcoxon 検定の例](#)」を参照してください。

**メディアン検定** 中央値検定（メディアン検定）の表示／非表示を切り替えます。メディアン順位スコアは、順位がメディアン（中央値）の順位の上か下かによって1または0の値を取ります。メディアン検定は、誤差が二重指数分布に従っている場合に、最も検出力の高い順位検定となります。このレポートについては、「[Wilcoxon / Kruskal-Wallis 検定、メディアン検定、Friedman 順位検定、Van der Waerden 検定のレポート](#)」を参照してください。

**Van der Waerden の検定** Van der Waerden 検定の表示／非表示を切り替えます。Van der Waerden 検定のスコアは、データの順位を $n+1$ で割ったものを標準正規分布の分位点関数で変換したものです。Van der Waerden の検定は、誤差が正規分布に従っている場合に、最も検出力の高い順位検定となります。このレポートについては、「[Wilcoxon / Kruskal-Wallis 検定、メディアン検定、Friedman 順位検定、Van der Waerden 検定のレポート](#)」を参照してください。

**Kolmogorov Smirnov 検定** (X 変数の水準数が2の場合のみ使用できます。) Kolmogorov-Smirnov 検定の表示／非表示を切り替えます。Kolmogorov Smirnov 検定は、各群の分布がすべて同じかどうかを経験的分布関数に基づき検定する。このレポートについては、「[Kolmogorov-Smirnov 二標本検定のレポート](#)」を参照してください。

**Friedman 順位検定** (起動ウィンドウで、ブロック変数を指定した場合で、かつ、各ブロック内の観測数（オブザベーション数）が等しい場合のみ使用できます。) Friedman 検定の表示／非表示を切り替えます。Friedman の順位スコアは、ブロック変数の各水準内でのデータの順位です。この検定は、ノンパラメトリックな反復測定分散分析と言えます。このレポートについては、「[Wilcoxon / Kruskal-Wallis 検定、メディアン検定、Friedman 順位検定、Van der Waerden 検定のレポート](#)」を参照してください。

**メモ:** 「Friedman 順位検定」の赤い三角ボタンのメニューには、Nemenyi 検定のオプションが含まれています。Nemenyi 検定は、Friedman 順位検定と同じように順位変換に基づく検定ですが、Tukey-Kramer 検定と同じようにすべてのペアに対する比較を行います。Friedman 順位検定は統計的に有意であってもどのペアが有意であるかは分かりません。一方、Nemenyi 検定ではペアに対する検定が行われます。

**メモ:**

- Wilcoxon 検定、メディアン検定、Van der Waerden 検定、Friedman の順位検定では、X 変数に水準が3つ以上ある場合、一元配置の枠組みでのカイ2乗近似が計算されます。
- 起動ウィンドウでブロック変数を指定し、ブロック変数と X 変数の水準の組み合わせごとの度数が同じである場合は、[ノンパラメトリック] のオプションのうち [Friedman 順位検定] だけは使用できます。度数が等しくない場合、[ノンパラメトリック] のオプションのいずれも使用できません。

**Jonckheere-Terpstra 検定** Jonckheere-Terpstra 検定の表示／非表示を切り替えます。この検定は、差の大きさに対して予め順序が想定できる状況での群間比較の傾向検定です。薬剤の投与量など、比較したい群にカテゴリな順序が想定できる場合に適した検定です。対立仮説は、カテゴリの順序です。

**二項分布** (X 変数の水準数が2水準の場合のみ使用できます。) 正確検定を行うオプションがあります。[「正確検定のレポート」](#)を参照してください。

**正確 Wilcoxon 検定** (X 変数の水準数が2水準の場合のみ使用できます。) Wilcoxon の順位スコアを使った正確検定の表示／非表示を切り替えます。[「Wilcoxon 検定の例」](#)を参照してください。

**正確メディアン検定** (X 変数の水準数が2水準の場合のみ使用できます。) 正確な中央値検定(メディアン検定) の表示／非表示を切り替えます。

**正確 Van Der Waerden 検定** (X 変数の水準数が2水準の場合のみ使用できます。) 正確な Van der Waerden 検定の表示／非表示を切り替えます。Van der Waerden 検定は、正規スコアに基づきます。

**正確 Kolmogorov Smirnov 検定** (X 変数の水準数が2水準の場合のみ使用できます。) Kolmogorov-Smirnov 検定の表示／非表示を切り替えます。Kolmogorov Smirnov 検定は、各群の分布がすべて同じかどうかを経験的分布関数に基づき検定します。

**注意:** 標本サイズが大きいと、正確検定の計算に時間がかかる場合があります。

**ノンパラメトリックな多重比較** グループの位置をノンパラメトリックに多重比較するためのオプションがあります。

**ペアごと Wilcoxon 検定** 各ペアに対する Wilcoxon 検定の表示／非表示を切り替えます。この方法では、多重比較の調整を行いません。これは、[平均の比較] メニューの [各ペア Student の t 検定] オプションのノンパラメトリック版です。[「\[ペアごと Wilcoxon 検定\]、\[すべてのペア Steel-Dwass 検定\]、\[コントロール群との比較 Steel 検定\] のレポート」](#)を参照してください。

**すべてのペア Steel-Dwass 検定** 各ペアに対する Steel-Dwass 検定の表示／非表示を切り替えます。これは、[平均の比較] メニューの [すべてのペア, Tukey の HSD 検定] オプションのノンパラメトリック版です。「[ペアごと Wilcoxon 検定]、[すべてのペア Steel-Dwass 検定]、[コントロール群との比較 Steel 検定] のレポート」を参照してください。

**コントロール群との比較 Steel 検定** グループ変数の各水準をコントロール群と比較する Steel 検定の表示／非表示を切り替えます。これは、[平均の比較] メニューの [コントロール群との比較, Dunnett 検定] オプションのノンパラメトリック版です。「[ペアごと Wilcoxon 検定]、[すべてのペア Steel-Dwass 検定]、[コントロール群との比較 Steel 検定] のレポート」を参照してください。

**すべてのペア 併合順位の Dunn 検定** 各ペアを比較する Dunn 検定の表示／非表示を切り替えます。この Dunn 検定では、比較対象のペアにおける順位でなく、すべてのデータを通じた順位が計算に使われます。表内の  $p$  値には、Bonferroni の調整が施されています。つまり、未調整の  $p$  値に比較の数を掛けたものです。調整済み  $p$  値が 1 を超えた場合は、1 と表示されます。「[すべてのペア 併合順位の Dunn 検定]、[コントロール群との比較 併合順位の Dunn 検定] のレポート」を参照してください。

**コントロール群との比較 併合順位の Dunn 検定** グループ変数の各水準をコントロール群と比較する Dunn 検定の表示／非表示を切り替えます。この Dunn 検定では、比較対象のペアにおける順位でなく、すべてのデータを通じた順位が計算に使われます。表内の  $p$  値には、Bonferroni の調整が施されています。つまり、未調整の  $p$  値に比較の数を掛けたものです。調整済み  $p$  値が 1 を超えた場合は、1 と表示されます。因子の列に「コントロール群」列プロパティを設定しておく、[コントロール群との比較 Steel 検定] または [コントロール群との比較 併合順位の Dunn 検定] を選択するたびにコントロール群を指定する必要がなくなります。『JMP の使用法』を参照してください。「[すべてのペア 併合順位の Dunn 検定]、[コントロール群との比較 併合順位の Dunn 検定] のレポート」を参照してください。

---

**メモ:** ノンパラメトリックな検定を選択するかどうかは、データによって決まります。なお、すべてのペアに対する仮説検定での検定の選び方は、Boos and Duan (2021) で解説されています。

---

**異分散** 4 種類の等分散性の検定と、標準偏差が等しくないときに平均を比較する Welch 検定の表示／非表示を切り替えます。等分散性の検定は、O'Brien, Brown-Forsythe, Levene, Bartlett の 4 種類です。「等分散性の検定レポート」を参照してください。

**同等性検定** 同等性検定・優越性検定・非劣性検定という 3 種類の検定を行えます。「同等性検定のレポート」を参照してください。同等性検定は、実質的に意味のある同等性（つまり、臨床的に意味がある同等性）を検出するための検定です。t 検定と Wilcoxon 検定が用意されています。

**平均** 平均の差に対する同等性検定・優越性検定・非劣性検定のいずれかを指定するウィンドウを起動します。呼び出されたウィンドウでは、分散に対する仮定やマージンなども指定します。

**標準偏差** 標準偏差の比に対する同等性検定・優越性検定・非劣性検定のいずれかを指定するウィンドウを起動します。呼び出されたウィンドウでは、マージンなども指定します。

**ロバスト** 平均に対してロバスト（頑健）な推定値を求めるためのオプションがあります。ロバストな手法では、外れ値が分析に及ぼす影響が小さくなります。裾の重い分布の場合、ロバストな推定の方が最小2乗法を使った推定より効率的です。「[ロバストなあてはめのレポート](#)」を参照してください。

**ロバストなあてはめ** 「一元配置」プロット上でロバスト平均の線の表示／非表示を切り替えます。また、Huber法によるグループ平均の推定値を含む「ロバストなあてはめ」レポートの表示／非表示も切り替えます。

**Cauchyのあてはめ** 「一元配置」プロット上でロバスト平均の線の表示／非表示を切り替えます。また、誤差がCauchy分布に従うと仮定してグループ平均を求めた「Cauchyのあてはめ」レポートの表示／非表示も切り替えます。このロバストな方法は、極端な外れ値に適しています。

**検出力** 分散分析における検出力を計算します。「[検出力のレポート](#)」を参照してください。検出力の計算の詳細については、『基本的な回帰モデル』を参照してください。

**$\alpha$ 水準の設定**  $\alpha$ 水準を指定することができます。一般的な値のリストから選択するか、[その他...]を選択して水準を指定します。

---

**メモ:**  $\alpha$ 水準は分析とレポートに使われます。これには、信頼限界、平均のひし形、比較円、多重比較分析などが含まれます。[同等性検定] オプション（同等性検定・非劣性検定・優越性検定）に使う  $\alpha$ 水準は別途設定します。

---

**正規分位点プロット** 各グループの分位点プロットに関する次のオプションがあります。「[正規分位点プロットの例](#)」を参照してください。

**分位点-実測値プロット** 「一元配置」レポートの先頭にあるグラフの右側に、分位点プロットを表示／非表示します。分位点プロットは、一元配置分析のプロットと縦軸（応答変数）を共有します。横軸にはグループごとの累積確率が示されます。カテゴリカルなX因子の水準ごとの分位点プロットが表示されます。

**実測値-分位点プロット** 横軸をY変数、縦軸を累積確率とする分位点プロットの表示／非表示を切り替えます。カテゴリカルなX変数の水準ごとの分位点プロットが表示されます。

**あてはめ線**（分位点プロットが開いている場合のみ使用できます。）開いている各分位点プロットにおいて、X変数の水準ごとの参照線を表示／非表示します。

**正規分位点のラベル**（分位点プロットが開いている場合のみ使用できます。）開いている各分位点プロットにおいて、正規分位点スケールを表示／非表示します。

**累積確率プロット** 累積分布関数（CDF）の表示／非表示を切り替えます。「[累積確率プロットの例](#)」を参照してください。

**密度** 各グループの密度を視覚化するオプションがあります。「[\[密度\]の各オプションの例](#)」を参照してください。

**密度の比較** 各グループの確率密度関数を重ね合わせたプロットの表示／非表示を切り替えます。

**密度の構成** 積み重ねて描かれた密度曲線の表示／非表示を切り替えます。各グループの密度曲線は、そのグループの度数で重み付けられます。この密度構成プロットは、X変数の範囲全体において、各グループが総密度にどれくらい寄与しているかを示します。

**密度の割合** X変数の各水準が密度全体にどれくらい寄与しているかを表すプロットの表示／非表示を切り替えます。この寄与度は、X変数の各値において、各グループの密度の総密度に占める割合で示されます。

**対応のある列を設定** 指定された対応する変数に基づいて、対応があるモデルをあてはめます。また、「一元配置」レポートの先頭にあるグラフにおいて、あてはめ線の表示／非表示を切り替えます。このオプションは、X変数における異なるグループの観測値が同じ個体から取られた場合など、分析で使うデータが対応のあるデータである場合に使用します。「[「対応のあるあてはめ」レポート](#)」を参照してください。

**保存** 現在のデータテーブルの新しい列に、次の数値を保存します。

**残差の保存** Y変数のデータ値から、グループ平均を引いた値を保存する。

**標準化データの保存** X変数の水準ごとに標準化した値を保存する。これは、Xの水準ごとに、Y変数を中心化して標準偏差で割ったものです。

**正規分位点の保存** X変数の各水準内で計算された正規分位点が保存されます。

**予測値の保存** X変数の水準ごとに計算した平均値を保存する。

**表示オプション** プロットに対して要素を追加または削除します。オプションのうち、関係のないものは表示されません。

**すべてのグラフ** 「一元配置」レポートの先頭にあるグラフの表示／非表示を切り替えます。

**点** 「一元配置」レポートの先頭にあるグラフにおいて、データ点の表示／非表示を切り替えます。

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

**平均のひし形** 「一元配置」レポートの先頭にあるグラフにおいて、平均のひし形の表示／非表示を切り替えます。各ひし形は平均の95%信頼区間を示し、ひし形の中心になる水平線は平均を示します。95%信頼区間は、プールした標準偏差を用いて計算されます。「[平均のひし形と標本サイズに比例したX軸](#)」を参照してください。

**平均線** 「一元配置」レポートの先頭にあるグラフにおいて、各グループの平均値の水平線の表示／非表示を切り替えます。「[平均線・平均誤差バー・標準偏差線](#)」を参照してください。

**平均の信頼区間** 各グループ平均の両側95%信頼区間の上限と下限を示す線の表示／非表示を切り替えます。95%信頼区間は、プールした標準偏差を用いて計算されます。

**平均誤差バー** 「一元配置」レポートの先頭にあるグラフにおいて、各グループの平均と、平均から上下へ1標準偏差だけ離れたところの誤差バーの表示／非表示を切り替えます。「[平均線・平均誤差バー・標準偏差線](#)」を参照してください。



**全体平均** 「一元配置」レポートの先頭にあるグラフにおいて、Y変数の全体平均の表示／非表示を切り替えます。

**標準偏差線** 「一元配置」レポートの先頭にあるグラフにおいて、各グループの平均から上下へ1標準偏差だけ離れたところの線の表示／非表示を切り替えます。「[平均線・平均誤差バー・標準偏差線](#)」を参照してください。

**比較円** （多重比較レポートが開いている場合のみ使用できます。）比較円の表示／非表示を切り替えます。「[比較円の統計的詳細](#)」を参照してください。例として、「[一元配置分析の実行](#)」を参照してください。

**平均をつなぐ** 「一元配置」レポートの先頭にあるグラフにおいて、グループ平均を結ぶ折れ線の表示／非表示を切り替えます。

**平均の平均** 「一元配置」レポートの先頭にあるグラフにおいて、グループ平均の平均値の表示／非表示を切り替えます。

**標本サイズに比例したX軸** （〔対応のある列を設定〕オプションが選択されている場合は使用できません。）横軸の間隔を指定します。選択した場合、間隔は各水準の観測数に比例します。「[平均のひし形と標本サイズに比例したX軸](#)」を参照してください。

**点の拡散** データ点の広がり指定します。選択した場合、データ点は区間の幅に広がります。

**点をずらす** データ点の広がりに加えるランダムな誤差の種類を指定します。このオプションを使うと、異なる方法でマーカーをずらし、点と点が重なるのを防ぐことができます。それぞれの方法の詳細については、『[グラフ機能](#)』を参照してください。

**対応を表す線** （〔対応のある列を設定〕オプションが選択されている場合のみ使用できます。）対応のある列が設定されていた場合、「一元配置」レポートの先頭にあるグラフにおいて、折れ線の表示／非表示を切り替えます。この折れ線は、対応する変数の各水準の平均を結ぶものです。

**対応を表す点線** （〔対応のある列を設定〕オプションが選択されていて、X変数のある水準に対応する列の値がすべて欠測値である場合のみ使用できます。）対応のある列が設定されていた場合、「一元配置」レポートの先頭にあるグラフにおいて、対応する変数の平均を結ぶ折れ線で、欠測セルに対する点線の表示／非表示を切り替えます。この時、欠測セルでの値は、二元配置分散分析モデルを使って求められます。

**ヒストグラム** 「一元配置」レポートの先頭にあるグラフの右側に、横方向のヒストグラムを表示／非表示します。

**ロバスト平均の線** （〔ロバスト〕のオプションが選択されている場合のみ使用できます。）「一元配置」レポートの先頭にあるグラフにおいて、各グループのロバスト平均の水平線の表示／非表示を切り替えます。

**凡例** 正規分位点プロット・累積分布関数（CDF）プロット・確率密度関数（PDF）プロットにおいて、凡例の表示／非表示を切り替えます。

以下のオプションの詳細については、『[JMPの使用法](#)』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMPレポートの設定に合わせて環境設定を変更したりできます。

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

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

---

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの [スクリプトの索引] を開いてください。また、[スクリプトの索引] には、この節で紹介されているオプションのスクリプト例もあります。

---

---

## 「一元配置分析」レポート

「一元配置」プラットフォームでは、グラフを作成したり、仮説検定を行ったり、モデルをあてはめたりすることで、カテゴリカル尺度のX変数と連続尺度のY変数の関係を調べることができます。この節では、分析を実行する各オプションで作成されるレポートの内容を詳しく説明します。

- 「「分位点」レポート」
- 「「平均/ANOVA/プーリングしたt検定」レポート」
- 「「平均と標準偏差」レポート」
- 「「t検定」レポート」
- 「平均分析レポート」
- 「平均の比較レポート」
- 「ノンパラメトリックな検定のレポート」
- 「ノンパラメトリックな多重比較のレポート」
- 「等分散性の検定レポート」
- 「同等性検定のレポート」
- 「ロバストなあてはめのレポート」
- 「検出力のレポート」
- 「「対応のあるあてはめ」レポート」



## 「分位点」レポート

「一元配置」プラットフォームの「分位点」レポートには、各グループの最小値と最大値が表示されます。また、10%点、25%点、50%点（中央値）、75%点、90%点もリストされます。50%点は、「中央値」や「メディアン」とも呼ばれます。また、25%点と75%点は、「四分位点」とも呼ばれます。

## 「平均/ANOVA/プーリングしたt検定」レポート

「一元配置」プラットフォームの「平均/ANOVA」オプションは、分散分析を行います。X変数の水準数が2水準の場合、このオプションは「平均/ANOVA/プーリングしたt検定」という名前で表示されます。レポートは、「あてはめの要約」、「分散分析」(ANOVA)、各グループの要約統計量の表で構成されています。X変数の水準数が2水準の場合、プーリングしたt検定の表が含まれます。起動ウィンドウでブロック変数を指定し、ブロックとX変数の水準の各組み合わせに含まれる度数が等しい場合は、「ブロック平均」表が表示されます。ブロック変数がその他の構成である場合は、「ブロックがある一元配置分散分析」レポートが作成されます。

### あてはめの要約

「一元配置」プラットフォームの「あてはめの要約」表には、次のような統計量が表示されます。

**R2乗** モデルによって説明できる変動の割合を表します。残りの変動は、モデルで説明できないため、ランダムな誤差とみなされます。モデルが完全にあてはまっていると、R2乗は1になります。「[「あてはめの要約」レポートの統計的詳細](#)」を参照してください。R<sup>2</sup>は「**決定係数**」とも言います。

---

**メモ:** R2乗が小さい場合は、必要な説明変数がモデルに含まれていない可能性があります。ただし、扱っているデータでの誤差が大きい場合は、意味のある分散分析モデルであっても、R2乗値が小さくなる場合があります。研究分野により求められるR2乗値が異なる場合がありますので、その分野に関する文献を参考にしてください。

---

**自由度調整 R2乗** 「自由度調整済みR2乗」、「自由度調整済み決定係数」などと呼ばれている指標であり、R2乗をモデルのパラメータ数で調整した指標です。自由度調整済みR2乗は、パラメータ数の異なるモデルを比較するのに役立ちます。「[「あてはめの要約」レポートの統計的詳細](#)」を参照してください。

**誤差の標準偏差 (RMSE)** 誤差の標準偏差の推定値。これは、「分散分析」レポートにおける「誤差」行の「平均平方」列にある値の平方根です。

**応答の平均** Y変数の全体平均（算術平均）。

**オブザベーション（または重みの合計）** 分散分析モデルを推定するために使われた標本サイズ（観測値の個数）。ただし、重み変数が使われた場合は、重みの合計です。「[「あてはめの要約」レポートの統計的詳細](#)」を参照してください。

## プーリングしたt検定

「一元配置」プラットフォームの「プーリングしたt検定」表は、「分散が等しい」という仮定のもとで2群の平均を比較したt検定の結果を示します。この表は、X変数の水準数が2水準の場合のみ作成されます。「[「t検定」レポート](#)」を参照してください。

## 分散分析

「一元配置」プラットフォームの「分散分析」表は、分散分析の結果をまとめたものです。分散分析は、データ全体の変動をいくつかの要因に分割します。

---

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

---

**要因** 変動の要因。(「モデル」・「誤差」・「全体(修正済み)」)。

**自由度** 変動の各要因の自由度 (DF)。

- 「全体(修正済み)」の自由度は、 $N - 1$ で求められます。 $N$ は分析に使用された標本サイズ(観測値の個数)です。
- モデルの自由度は $k - 1$ で、 $k$ はX変数の水準数です。

「誤差」の自由度は、「全体(修正済み)」とモデルの自由度の差( $N - k$ )です。

**平方和** 変動をもたらす各要因の平方和 (SS)。

- 「全体(修正済み)」の平方和は、各応答から全体平均までの距離の平方和です。「全体(修正済み)」の平方和は、基本モデルの平方和として、その他のモデルとの比較に使用されます。
- 「誤差」(残差)の平方和は、各点からそれぞれのグループ平均までの距離の平方和です。この平方和は、分散分析モデルをあてはめた後に残った、説明できない部分の平方和です。

全体平方和から誤差平方和を引いたものがモデルの平方和です。全変動のうちモデルによって説明されている部分を表します。

**平均平方** 平方和に関連する自由度で割った値。

- モデルの平均平方は、グループ平均が等しいという仮説のもとでの誤差分散推定値にもなっています。
- 「誤差」の平均平方は、モデルの平均平方とは無関係に誤差分散を推定した値であり、モデルの仮説による制約を受けません。

**F値** モデルの平均平方を誤差の平均平方(MSE)で割ったもの。「すべてのグループの母平均は等しい(真のモデルにおいて母平均には差がない)」という仮説が真のとき、誤差とモデルの平均平方が両方とも誤差分散の推定値となり、その比はF分布に従います。分散分析モデルをあてはめた結果、誤差の変動が全体の変動よりも有意に減少した場合、「F値」は期待値よりも大きくなります。

**p値(Prob>F)** すべてのグループの母平均の間に差がないという仮定のもとで、現在のF値よりも大きなF値を得る確率。p値が0.05以下だと、グループ平均に有意差がある証拠だとみなされます。

## 各水準の平均

「一元配置」プラットフォームの「各水準の平均」表は、名義／順序尺度のXの水準ごとに、Y変数の情報をまとめたものです。

**水準** X変数の水準。

**数** 各グループ内の標本サイズ（観測値の個数）。

**平均** 各グループの平均。

**標準誤差** グループ平均の標準偏差の推定値。このレポートの標準誤差は、「Yの分散がどの水準でも等しい」という仮定のもとに計算されます。この標準誤差は、「あてはめの要約」レポートにある誤差の標準偏差(RMSE)を、グループごとの標本サイズの平方根で割ったものです。

**下側95%、上側95%** このレポートでのグループ平均の両側95%信頼区間は、プールして求めた誤差分散推定値と自由度を使って計算したものです。

## ブロック平均

「一元配置」プラットフォームの「ブロック平均」レポートは、起動ウィンドウでブロック変数を指定し、ブロック変数とX変数の水準の組み合わせごとの度数が等しい場合にのみ表示されます。

## 「平均と標準偏差」レポート

「一元配置」プラットフォームの「平均と標準偏差」レポートは、X変数の水準ごとに要約統計量をまとめたものです。X変数の水準がグループ（処置群）を形成します。

**水準** X変数の水準。

**数** 各グループ内の標本サイズ（観測値の個数）。

**平均** 各グループの平均。

**標準偏差** 各グループの標準偏差。

**平均の標準誤差** グループ平均の標準誤差の推定値。この値は、標準偏差を、グループ平均の計算に使った値の個数の平方根で割ったものです。

**下側95%、上側95%** この両側95%信頼区間は、グループごとの標準偏差から計算されます。

---

**メモ:** この信頼区間は、特にグループごとの標準偏差が異なる場合、分散分析に基づく信頼区間（誤差分散をプールして求めた場合の信頼区間）とは異なります。

---

**標準偏差 下側95%、標準偏差 上側95%** この両側95%信頼区間は、グループごとの標準偏差から計算されます。デフォルトでは列は非表示です。列を表示するには、表を右クリックして **[列] > [標準偏差 下側95%]** と **[列] > [標準偏差 上側95%]** を選択します。

---

**メモ:** 標準偏差の信頼区間のプロットを表示するには、表を右クリックして【列】>【~標準偏差】を選択します。プロットに表示されている縦の点線は、標準偏差の平均を表します。グループの標準偏差の信頼区間に標準偏差の平均が含まれる場合、その信頼区間は青色で表示され、含まれない場合は赤色で表示されます。

---

## 「t検定」レポート

「一元配置」プラットフォームの「t検定」レポートは、2つのグループ平均を、等分散または異分散を仮定して比較したt検定の結果を示します。(X変数の水準数が2水準の場合のみ使用できます。)

「t検定」および「プーリングしたt検定」表には、どちらの群からどちらの群を引いた平均差であるのかという情報と、分散に関する仮定が表示されます。また、このレポートには、以下の統計量が含まれています。

**差** 2つのX水準における平均差の推定値。

---

**ヒント:** 差の順序を変更するには、「値の表示順序」列プロパティを使用します。

---

**差の標準誤差** 平均差の標準誤差を示します。

**差の上側信頼限界** 平均差の両側信頼区間における上限。

**差の下側信頼限界** 平均差の両側信頼区間における下限。

**信頼度** 信頼水準(1 -  $\alpha$ )を示します。信頼水準を変更するには、プラットフォームの赤い三角ボタンをクリックし、【 $\alpha$ 水準の設定】オプションから新しい $\alpha$ 水準を選択します。

**t値** t統計量。

**自由度** t検定で使用された自由度。

**p値(Prob>|t|)** 両側検定のp値。

**p値(Prob>t)** 上片側検定のp値。

**p値(Prob<t)** 下片側検定のp値。

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

**Cohenのd** (プーリングしたt検定でのみ使用可能。) 2群における平均の差を、誤差の標準偏差(RMSE)で割ったものです。この指標は、効果の大きさ(効果量; effect size)を表します。

## 平均分析レポート

「一元配置」プラットフォームの平均分析（ANOM; analysis of means）は、平均または分散の多重比較を行う方法です。平均分析のレポートに含まれるグラフには、次の要素が表示されます。

- 上側決定限界（UDL; Upper Decision Limit）
- 下側決定限界（LDL; Lower Decision Limit）
- 水平線（中心線）。分析の種類ごとに次の位置を中心線は示します。
  - 平均分析: 全体平均
  - 順位変換平均分析: 順位変換データの全体平均
  - 分散の平均分析: 誤差の標準偏差 (RMSE) (Y 軸のスケールが分散の場合は MSE)
  - 分散の平均分析 -Levene の ADM: 平均からの偏差の絶対値の全体平均
  - 範囲の平均分析: グループ範囲の全体平均

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

## 平均分析のグラフオプション

平均分析レポートのタイトルバーにある赤い三角ボタンをクリックすると、次のようなオプションが表示されます。

**有意水準の設定** リストの中から任意の有意水準を選択するか、**[その他 ...]** を選択して有意水準を指定します。有意水準を変更すると、自動的に上側決定限界と下側決定限界が更新されます。

---

**メモ:** [範囲の平均分析] の場合、選択できるのは 0.10、0.05、および 0.01 のみです。

---

**要約レポートの表示** 平均分析の手法に応じて作成される要約レポートの表示／非表示を切り替えます。

- [平均分析] の場合は、グループ平均と決定限界を示すレポートが作成されます。
- [順位変換平均分析] の場合は、順位のグループ平均と決定限界を示すレポートが作成されます。
- [分散の平均分析] の場合は、グループごとの標準偏差（もしくは、分散）と決定限界を示すレポートが作成されます。
- [分散の平均分析 -Levene の ADM] の場合は、ADM のグループ平均と決定限界を示すレポートが作成されます。
- [範囲の平均分析] の場合は、グループ範囲と決定限界を示すレポートが作成されます。

**Y 軸スケールを分散に**（[分散の平均分析] オプションでのみ使用できます。）グラフの縦軸のスケールを、標準偏差の単位または分散に切り替えます。

**表示オプション** 表示を変更する以下のオプションがあります。

**決定限界の表示** 決定限界線の表示／非表示を切り替えます。

**決定境界の陰影の表示** 決定境界を描く陰影の表示／非表示を切り替えます。

**中心線の表示** 中心線の表示／非表示を切り替えます。

**点のオプション: 垂線の表示** 垂線を表示します。これは、デフォルトのオプションです。[接続線の表示] オプションは、各グループの平均をつなぐ線を表示します。[点のみを表示] オプションは、各グループの平均を示す点のみを表示します。

### 【範囲の平均分析】の制限

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

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

## 平均の比較レポート

「一元配置」プラットフォームには、平均を比較するためのオプションが複数用意されています。ここでは、平均の比較レポートを作成するオプションと比較手法の種類について解説します。

### 平均の比較レポートのオプション

「一元配置」プラットフォームの「平均の比較」の赤い三角ボタンをクリックすると、メニューが開き、次のオプションが表示されます。

**差の行列** グループのすべてのペアの平均差で構成された行列の表示／非表示を切り替えます。

**棄却限界値** 平均の比較で使われる棄却値や有意水準 ( $\alpha$ ) の表示／非表示を切り替えます。

---

**ヒント:** 有意水準は、「一元配置分析」の赤い三角ボタンのメニューにある [  $\alpha$  水準の設定 ] オプションを使って変更できます。

---

**LSD 閾値行列** ([各ペアのステップワイズ] オプションの場合は使用できません。) 平均のペアごとの差から、その最小有意差を引いた行列の表示／非表示を切り替えます。正の値は、その平均のペアには有意差があることを示します。Hsu の MCB 検定の場合は、LSD 行列が2つ作成されます。1つは最小値との比較、もう1つは最大値との比較を示します。Dunnett 検定の場合は、上片側閾値行列と下片側閾値行列が作成できます。

**文字の接続レポート** ([各ペア]、[すべてのペア]、[各ペアのステップワイズ] の各オプションの場合のみ使用できます。) 文字によって有意差があるかどうかを示したレポートの表示／非表示を切り替えます。文字で繋がっていない平均のペアには有意差があります。

**差の順位レポート**（[各ペア] および [すべてのペア] オプションの場合のみ使用できます。）すべてのペアごとの平均差（正の値のほうの差）、標準誤差、信頼区間、 $p$  値、および、信頼区間も描いた差の棒グラフの表示／非表示を切り替えます。 $p$  値は、平均が等しいという仮説に対応しています。

**詳細な比較レポート**（[各ペア] オプションの場合のみ使用できます。）各比較の詳細レポートの表示／非表示を切り替えます。比較ごとに、水準間の差、差の標準誤差と信頼区間、 $t$  値、 $p$  値、および自由度が表示されます。検定結果を図示したグラフが、各レポートの右側に表示されます。

---

**メモ：** 差の標準誤差は、MSE と標本サイズに基づいてペアごとに求めた、プーリングした標準誤差です。

---

## 各ペア, Student の $t$ 検定

「一元配置」プラットフォームの [各ペア, Student の  $t$  検定] オプションは、Fisher の最小有意差 (LSD) 検定を行います。これは、Student の  $t$  検定に基づいてグループの水準の各ペアの差を検定するものです。この検定の例については、「[\[各ペア, Student の  \$t\$  検定\] の例](#)」を参照してください。

## すべてのペア, Tukey の HSD 検定

「一元配置」プラットフォームの [すべてのペア, Tukey の HSD 検定] オプションは、差のすべてのペアを検定します。この手法では、全体における有意水準を保ちながらすべてのペアが検定されます。HSD 検定の信頼区間は、Student の  $t$  検定の信頼区間より広くなります。Tukey の HSD 検定は、[各ペア] の検定（多重性調整を行っていない  $t$  検定）に比べて、その比較円は大きくなり、また、平均差に対する  $p$  値は大きくなります。

「棄却限界値」表にある「 $q^*$ 」の統計量は、 $q^* = (1/\sqrt{2}) \times q$  の式で求められます。この式で、 $q$  はスチューデント化された範囲分布の  $\alpha$  パーセント点です。この検定の例については、「[\[すべてのペア, Tukey の HSD 検定\] の例](#)」を参照してください。

[すべてのペア, Tukey の HSD 検定] オプションは、**Tukey** の HSD (Honestly Significant Difference) 検定、または **Tukey-Kramer** の HSD 検定 (Tukey 1953, Kramer 1956) を行います。この検定は、グループごとの標本サイズが同じときには、指定された有意水準の正確な検定です。グループごとの標本サイズが異なるときは、保守的な結果になります (Hayter 1984)。

## 最適値との比較, Hsu の MCB 検定

「一元配置」プラットフォームの [最適値との比較, Hsu の MCB 検定] は、特定の水準の平均が、それ以外の水準の平均の最大値または最小値とどう異なるかを調べます。Hsu (1996) を参照してください。この検定の例については、「[\[最適値との比較, Hsu の MCB 検定\] の例](#)」を参照してください。

Hsu の MCB 検定の分位点は、群によって異なります。各群の標本サイズが等しくない場合は、比較円のグラフは正確とはいえません。比較円の半径は、その群の平均に対する標準誤差と、最大分位点を掛け合わせたものになります。正確に有意差を評価したい場合には、グラフではなく、検定の  $p$  値を見てください。[「最小/最大との比較」](#) を参照してください。

[最適値との比較, Hsu の MCB 検定] オプションは、Hsu の MCB 検定 (Hsu 1996, Hsu 1981) を行います。

---

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

---

### 最小/最大との比較

「HsuのMCB検定」レポートには、平均の比較の全手法に共通する表に加え、「最小/最大との比較」表が含まれます。この表には次の値が表示されます。

**水準 群** (カテゴリカル変数のグループ)。

**p値 (vs. 最大)** ある水準の平均がその他の水準の最大平均より大きいという仮説検定の  $p$  値。(未知の) 真の平均の最大値以下に有意になっている平均を選ぶために使用できます。

**p値 (vs. 最小)** ある水準の平均がその他の水準の最小平均より小さいという仮説検定の  $p$  値。(未知の) 真の平均の最小値以上に有意になっている平均を選ぶために使用できます。

### コントロール群との比較, Dunnett 検定

「一元配置」プラットフォームの [コントロール群との比較, Dunnett 検定] オプションは、グループの平均をコントロール群と比較します。片側と両側の比較が使用できます。Student の  $t$  検定では、多重性調整をまったく行いません。Tukey-Kramer 検定では、すべてのペアの比較に関して多重性調整を行います。Dunnett 検定は、コントロール群との比較に関してだけで多重性の調整を行うので、これら2つの検定の中間にあたると言えます。この検定の例については、「[\[コントロール群との比較, Dunnett 検定\] の例](#)」を参照してください。Dunnett の検定の詳細については、Dunnett (1955) を参照してください。

### 各ペアのステップワイズ, Newman-Keuls 検定

「一元配置」プラットフォームの [各ペアのステップワイズ, Newman-Keuls 検定] オプションは、逐次的なステップワイズ手順により標本平均を比較します。各ステップでは、2群の平均の差に対して Tukey の HSD 検定を行います。この検定の例については、「[\[各ペアのステップワイズ, Newman-Keuls 検定\] の例](#)」を参照してください。

[各ペアのステップワイズ, Newman-Keuls 検定] オプションは、スチューデント化された範囲をもとにステップワイズ手順により検定を行い、平均に差があるかどうかを調べます。この検定は、「Newman-Keuls 検定」または「Student-Newman-Keuls 検定」と呼ばれています (Keuls, 1952)。この検定は実験全体の水準は保てないので妥当な検定ではありませんが、Tukey の HSD 検定ほど保守的ではありません。

---

**注意:** Newman-Keuls 検定では、検定全体における過誤率 (family wise error rate) が守られていません。結果の解釈には注意が必要です。

---

$J$  個の群における平均を以下のように検定します。

次のように定義します。

$J$  = 群数 (群の平均で昇順に並べたもの)



$N$  = 標本サイズ

$d$  = 自由度 ( $N - J$ )

$i$  = 比較対象の群平均のうち、最小平均の群番号

$j$  = 比較対象の群平均のうち、最大平均の群番号

$k$  = 各比較における  $j$  の最小値

処理の開始時に、 $i = 1, j = J, k = 2$  と設定します。

1. 群  $i$  と  $j$  に対して、Tukey の HSD 検定を行います。ここで、適切な分位点を見つける群数は、 $j - i + 1$  です。
  - この検定の結果が有意である場合は、群  $i$  と  $j$  に有意な差があると判断します。 $j$  の値を 1 減らします。そして、 $j$  が  $k$  より小さい場合は、 $i$  の値を 1 増やし、 $k = \max(i, j) + 1, j = J$  と設定して、[手順2](#)に進みます。また、 $j$  が  $k$  以上である場合は、[手順2](#)に進みます。
  - この検定の結果が有意でない場合は、 $i$  と  $j$  に有意な差があるとは判断しません。 $i$  の値を 1 増やし、 $k = i + 1, j = J$  と設定し、[手順2](#)に進みます。
2.  $k$  の値に基づいて、処理を続行するか、または終了するかを判定します。
  - $k$  が  $i$  より大きく、 $J$  以下である場合は、[手順1](#)を繰り返します。
  - $k$  が  $i$  以下か、または  $J$  より大きい場合は、処理を終了します。まだ検定していない範囲はすべて、有意な差がないとみなされます。

各ステップにおいて実行される Tukey の HSD 検定の分位点は、そのステップでの並び替え済みの群平均の個数に基づいており、ステップごとに異なります。Newman-Keuls 検定のレポートの「最小  $q^*$ 」(検定に使われた最小の分位点)は、上記の手順で使用した、スチューデント化された範囲の最小分位点を、2 の平方根で割った値です。

検定の結果は、「文字の接続レポート」に示されます。

Newman-Keuls 検定の詳細については、Howell (2013) を参照してください。

---

**メモ：** [各ペアのステップワイズ, Newman-Keuls 検定] オプションでは、比較円は表示されません。この手法では、ステップごとに比較円の大きさが異なるためです。

---

## ノンパラメトリックな検定のレポート

「一元配置」プラットフォームには、平均を比較するためのノンパラメトリックなオプションが用意されています。ここでは、ノンパラメトリックなオプションのレポートを取り上げます。

ノンパラメトリックな検定では、(分布の形状が同じであるという仮定が正しければ)「グループ間で平均または中央値が等しい」などの帰無仮説に対する検定が行われます。ノンパラメトリックな検定では、「順位スコア」と呼ばれる、 $Y$  変数の順位から変換された値が使用されます。Hajek (1969) および SAS Institute Inc. (2023a) を参照してください。ノンパラメトリックな多重比較では、ペアの比較を繰り返すことによる全体的な誤りの確率も調整されます。「[ノンパラメトリックな多重比較のレポート](#)」を参照してください。

**Wilcoxon / Kruskal-Wallis 検定、メディアン検定、Friedman 順位検定、Van der Waerden 検定のレポート**

「一元配置」プラットフォームの [Wilcoxon / Kruskal-Wallis 検定]・[メディアン検定]・[Friedman 順位検定]・[Van der Waerden 検定] では、要約統計量と検定結果を示す表が2〜3つ作成されます。要約統計量の表には以下の列があります。

**水準** Xの水準。

**度数** 各群の標本サイズ（各水準の度数）。

**スコア和** 各水準の順位スコアの和。

**スコアの期待値** 水準間に差はないという帰無仮説のもとでのスコアの期待値。

**スコア平均** 各水準の平均順位スコア。

**(平均 - 平均 0)/ 標準偏差 0** 標準化したスコア。「平均 0」は帰無仮説のもとでの平均スコアの期待値で、「標準偏差 0」は帰無仮説のもとでの平均スコアの標準偏差です。帰無仮説は、「(すべての群で分布の形状が同じという仮定が正しいもとで) どの群の平均やメディアンもまったく同じである」というものです。

**「2 標本検定 (正規近似)」表**

X変数の水準数が2の場合、「2 標本検定 (正規近似)」表には次の列が含まれます。

**S** 標本サイズが小さいほうの水準の順位スコアの和。

**Z** 正規近似の検定統計量。ここでの結果は、0.5の連続修正を行っています。[「2 標本の場合における正規近似」](#)を参照してください。

**p 値(Prob>Z)** 正規近似の片側検定の  $p$  値。ここでの結果は、0.5の連続修正を行っています。 $p$  値は、0.5連続修正を行った検定統計量に基づき、標準正規分布により求められます。

**p 値(Prob>|Z|)** 正規近似の  $p$  値。ここでの結果は、0.5の連続修正を行っています。 $p$  値は、0.5連続修正を行った検定統計量に基づき、標準正規分布により求められます。

---

**メモ:** Wilcoxon/Kruskal-Wallis の検定（順位和）レポートでは、この表は「Wilcoxon 2 標本検定 (正規近似)」表という名前です。

---

**「一元配置検定 (カイ 2 乗近似)」表**

「一元配置（または Kruskal-Wallis）検定 (カイ 2 乗近似)」表は、カイ 2 乗近似による検定結果を示しています。Conover (1999) を参照してください。グループ数が2群の場合、Kruskal-Wallis 検定は Wilcoxon 検定と等価です。ただし、グループ数が2群の場合、JMP の「Wilcoxon 検定」では 0.5 連続修正を行っており、「Kruskal-Wallis 検定」では 0.5 連続修正を行っていません。

**カイ 2 乗** カイ 2 乗検定統計量の値。[「一元配置のカイ 2 乗近似」](#)を参照してください。

**自由度** 検定の自由度。

**p値 (Prob>ChiSq)** 検定の $p$ 値。この $p$ 値は $X$ の水準数から1を引いた値を自由度にしたカイ2乗分布に基づいています。グループ数が2群の場合、この $p$ 値は、Wilcoxon検定の $p$ 値です。ただし、ここでの結果は、0.5連続修正を行わないWilcoxon検定の結果となっています。

---

**メモ:** Wilcoxon/Kruskal-Wallisの検定(順位和) レポートでは、この表は「Kruskal-Wallis検定 (カイ2乗近似)」表という名前です。

---

## Friedman-Nemenyi 検定

「Friedman 順位検定」の赤い三角ボタンのメニューには、Nemenyi 検定のオプションが含まれています。この検定は、Tukey-Kramerの多重比較検定のノンパラメトリック版です。Friedman 順位検定が統計的に有意である場合は、Nemenyi 検定を使ってどのペアが異なるかを評価します。要約表には以下の列があります。

**水準** 比較対象のペアのうち、1つ目の水準。

**- 水準** 比較対象のペアのうち、2つ目の水準。

**スコア平均の差** 1つ目の水準(「水準」)の順位スコアの平均から、2つ目の水準(「-水準」)の順位スコアの平均を引いた値。

**Z** 検定統計量。この検定統計量は、「(「すべての群の分布の形状が同じ」という仮定が正しい場合において) 分布の位置に差がない」という帰無仮説のもとで、漸近的に標準正規分布に従います。

**p値** Zに基づく漸近検定の $p$ 値。

レポートに、検定の要約統計量を表示した2つのアウトラインが追加されます。

**ペアごとのp値** ペアごとの $p$ 値の行列。要約表にある値と同じです。

**文字の接続レポート** 文字によって有意差があるかどうかを示したレポート。文字で繋がっていない平均のペアには有意差があります。

## Kolmogorov-Smirnov 二標本検定のレポート

「一元配置」プラットフォームの「Kolmogrov Smirnov 二標本検定」レポートでは、要約表と「漸近 Kolmogrov-Smirnov 検定」表の2つが作成されます。要約表には以下の列があります。

**水準**  $X$ 変数の2つの水準。

**度数** 各水準の度数。

**最大時の経験分布関数** 2つの経験累積分布関数(EDF)の差が最大になるときの、各水準のEDFの値。「合計」という行には、2つのEDFの差が最大になるときの、プーリングしたEDF(データセット全体のEDF)の値が示されます。

**最大時の平均からの乖離** 各水準について、以下の手順に従って求めた値です。

- 該当の水準における経験分布関数と、データセット全体（合計）のプーリングした経験分布関数との差を、群間差が最大となっている個所で求める。
- この差に、その水準における観測値の度数の平方根を掛ける。

「漸近 Kolmogorov-Smirnov 検定」表には次の列があります。

**KS** Kolmogorov-Smirnov 統計量は次式で計算されます。

$$KS = \max_j \sqrt{\frac{1}{n} \sum_i n_i (F_i(x_j) - F(x_j))^2}$$

この式では、次のような表記を使用しています。

- $x_j (j = 1, \dots, n)$  は、標本サイズ
- $n_i$  は、X の  $i$  番目の水準の標本サイズ
- $F$  は、プーリングした経験累積分布関数
- $F_i$  は、X の  $i$  番目の水準の経験累積分布関数

---

**メモ:** この Kolmogorov-Smirnov 統計量は X 変数に 3 つ以上の水準がある場合でも当てはまりますが、JMP では、X 変数の水準数が 2 である場合にのみ Kolmogorov-Smirnov 検定が使えます。

---

**Ksa** 漸近的な Kolmogorov-Smirnov 統計量。  $KS\sqrt{n}$  で求められます。ここで、 $n$  は、全体の標本サイズ（観測値数）です。

**D=max|F1-F2|** 2 群の EDF の差で、その絶対値が最大のもの。2 つの標本を比較するために使用される Kolmogorov-Smirnov 統計量です。

**p 値 (Prob > D)** 両側検定の  $p$  値。「2 群の分布はまったく同じである」という帰無仮説のもとで、D が求められた値よりも大きくなる確率です。

**D+ = max(F1-F2)** 有意であれば、「1 つ目の水準が 2 つ目の水準を上回る」という対立仮説が採択される片側検定の検定統計量。

**p 値 (Prob > D+)** D+ に基づく検定の  $p$  値。

**D- = max(F2-F1)** 有意であれば、「2 つ目の水準が 1 つ目の水準を上回る」という対立仮説が採択される片側検定の検定統計量。

**p 値 (Prob > D-)** D- に基づく検定の  $p$  値。

## 正確検定のレポート

「一元配置」プラットフォームでは、X変数が2水準の場合に、各種ノンパラメトリック検定の正確検定を実行することができます。Wilcoxon / Kruskal-Wallis 検定、メディアン検定、および Van der Waerden 検定では、次の列を含む「二標本: 正確検定」表が作成されます。

**S** 標本サイズが小さいほうの水準の順位スコアの和。両水準における観測値の個数が等しい場合は、値の順序で並べて後になる水準を使用します。

**p 値 (Prob ≤ S)** 片側検定の  $p$  値。この検定の方向は、X 変数の2つの水準間にある差の符号によって決まります。

**p 値 (Prob ≥ |S-Mean|)** 両側検定の  $p$  値。

---

**メモ:** 正確 Kolmogorov-Smirnov 検定の表には、漸近検定と同じ統計量が表示されます。ただし、 $p$  値は正確な確率です。「[Kolmogorov-Smirnov 二標本検定のレポート](#)」を参照してください。

---

## ノンパラメトリックな多重比較のレポート

「一元配置」プラットフォームの [ノンパラメトリックな多重比較] には、ノンパラメトリックな多重比較を行うためのいくつかの検定手法が用意されています。これらは順位を用いた検定で、[ペアごと Wilcoxon 検定] 以外では多重比較の調整が行われます。これらの検定の詳細については、Dunn (1964) および Hsu (1996) を参照してください。ここでは、ノンパラメトリックなオプションのレポートを取り上げます。

### [ペアごと Wilcoxon 検定]、[すべてのペア Steel-Dwass 検定]、[コントロール群との比較 Steel 検定] のレポート

これらの多重比較のレポートには、検定結果、信頼区間、文字の接続レポートが表示されます。これらの検定では、データ値は比較対象となる2つの水準内での順位に変換されます。

**q\*** 信頼区間の計算に使用される分位点。

**Alpha** 信頼区間の計算に使用される  $\alpha$  水準。信頼水準は、一元配置分析のメニューから [  $\alpha$  水準の設定 ] を選択することによって変更できます。

**水準** 比較対象のペアのうち、1つ目の水準。

**- 水準** 比較対象のペアのうち、2つ目の水準。

**スコア平均の差** 1つ目の水準 (「水準」) の順位スコアの平均から、2つ目の水準 (「-水準」) の順位スコアの平均を引いた値 (連続修正が適用されます)。

1つ目の水準の標本サイズを  $n_1$ 、2つ目の水準の標本サイズを  $n_2$  とします。まずデータが、これら2水準内で順位付けされます。同順位の場合は、平均順位が使用されます。1つ目の水準の順位の合計を  $\text{ScoreSum}_1$ 、2つ目の水準の順位の合計を  $\text{ScoreSum}_2$  とします。

平均スコアの差が正の数である場合、「スコア平均の差」は次の式で計算されます。

$$\text{スコア平均の差} = (\text{ScoreSum}_1 - 0.5)/n_1 - (\text{ScoreSum}_2 + 0.5)/n_2$$

平均スコアの差が負の数である場合、「スコア平均の差」は次の式で計算されます。

$$\text{スコア平均の差} = (\text{ScoreSum}_1 + 0.5)/n_1 - (\text{ScoreSum}_2 - 0.5)/n_2$$

**差の標準誤差** スコア平均の差の標準誤差。

**Z** 検定統計量。この検定統計量は、「(「すべての群の分布の形状が同じ」という仮定が正しい場合において) 分布の位置に差がない」という帰無仮説のもとで、漸近的に標準正規分布に従います。

**p値** Zに基づく漸近検定のp値。

**Hodges-Lehmann** 位置の差に対するHodges-Lehmann推定値。1つ目の水準のデータ値と、2つ目の水準のデータ値のすべてのペアから計算された差の中央値です。「差のプロット」の棒グラフは、Hodges-Lehmann推定値の大きさを表します。

**下側信頼限界** Hodges-Lehmann推定値の両側信頼区間の下限。

---

**メモ:** なお、標本サイズが大きすぎてメモリ不足となるだろう場合には、計算されません。

---

**上側信頼限界** Hodges-Lehmann推定値の両側信頼区間の上限。

---

**メモ:** なお、標本サイズが大きすぎてメモリ不足となるだろう場合には、計算されません。

---

**相違点のプロット** スコア平均の差の棒グラフ。

## 文字の接続レポート

文字によって有意差があるかどうかを示したレポート。文字で繋がっていない平均のペアには有意差がありません。

## [すべてのペア 併合順位のDunn検定]、[コントロール群との比較 併合順位のDunn検定] のレポート

Dunn検定は、すべての群のデータから求められた順位に基づいています。[コントロール群との比較 併合順位のDunn検定] では、コントロール群となる水準を選択する必要があります。

**水準** 比較対象のペアのうち、1つ目の水準。

**- 水準** 比較対象のペアのうち、2つ目の水準。

**スコア平均の差** 1つ目の水準(「水準」)の順位スコアの平均から、2つ目の水準(「-水準」)の順位スコアの平均を引いた値(連続修正が適用されます)。ここでの順位は、データセット全体における順位です。同順位の場合は、平均順位が使用されます。連続修正については、「スコア平均の差」を参照してください。

**差の標準誤差** スコア平均の差の標準誤差。

**Z** 検定統計量。この検定統計量は、「(「すべての群の分布の形状が同じ」という仮定が正しい場合において) 分布の位置に差がない」という帰無仮説のもとで、漸近的に標準正規分布に従います。

**p 値** Zに基づく漸近検定のp値。

## 文字の接続レポート

文字によって有意差があるかどうかを示したレポート。文字で繋がっていない平均のペアには有意差があります。

## 等分散性の検定レポート

「一元配置」プラットフォームには、4種類の等分散性の検定があります。等分散性の検定のうちの3つは、各グループ内でのばらつきを測定するための新しいY変数を作成し、それに対して分散分析を行います。もう1つはBartlettの検定です。Bartlettの検定は、正規分布のもとでの尤度比検定に類似しています。起動ウィンドウでブロック変数を指定した場合、[等分散性の検定] オプションは使用できません。

---

**メモ:** 等分散性の検定としては、分散の平均分析 (ANOMV) 法も使用できます。「[平均分析レポート](#)」を参照してください。

---

等分散性の検定として、以下のものが用意されています。

**O'Brien** グループ平均が元のY変数のグループごとの不偏分散と等しくなるような従属変数が新しく作成されます。O'Brien検定は、グループごとの分散に対する分散分析と言えます (O'Brien 1979; Olejnik and Algina 1987)。

**Brown-Forsythe** 各グループにおける中央値と各観測値との差の絶対値を応答として分散分析を実行し、F検定を行います (Brown and Forsythe 1974)。

**Levene** 各グループにおける平均と各観測値との差の絶対値を応答として分散分析を実行し、F検定を行います (Levene 1960)。ばらつきは  $z_{ij} = |y_{ij} - \bar{y}_i|$  として計算されます。

**Bartlett** 標本分散の重み付きの算術平均を、標本分散の重み付きの幾何平均と比べます。幾何平均は必ず算術平均以下になり、算術平均と幾何平均が等しくなるのは、すべての値が等しい場合です。すべての標本分散が等しいと、それらの算術平均と幾何平均は等しくなります。グループごとの分散の間に見られる変動が大きいほど、2つの平均の差が大きくなります。2つの平均から計算された値が、 $\chi^2$ 分布によって近似されます (カイ2乗分布は、F分布の分母自由度を無限大にし、分子自由度を掛けたものです)。カイ2乗検定統計量の値が大きいということは、算術平均/幾何平均の比が大きい、つまりグループごとの分散の違いが大きいことを意味します。Bartlett検定のカイ2乗検定統計量を自由度で割ったものが、F値として出力されます。Bartlettの検定は、正規性の仮定が満たされないケースに対してロバスト (頑健) ではありません (Bartlett and Kendall 1946)。

**両側F検定** (X変数の水準数が2水準の場合にのみ使用可能) 検定の対象が2つのグループだけである場合、等分散性の検定として標準的なF検定も行われます。F検定は、小さい方の分散値に対する大きい方の分散推定値の比です。両側検定にする目的で、F分布のp値に2が掛けられます。

---

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

---

「[等分散性検定の例](#)」を参照してください。

---

**ヒント:** 等分散性の検定からグループの分散が有意に異なることがわかったときは、通常のANOVA検定ではなく、Welchの検定を使用した方がよいかもしれません。Welchの統計量は、通常の一元配置分散分析におけるF検定の枠組みを用いていますが、グループごとの不偏分散の逆数で加重された平均が計算に使われています (Welch 1951; Brown and Forsythe 1974; Asiribo and Gurland 1990)。水準が2つしかないとき、Welchの一元配置分散分析は分散が等しくない場合のt検定と等価です。

---

## 「分散が等しいことを調べる検定」レポート

「一元配置」プラットフォームの「分散が等しいことを調べる検定」レポートには、標準偏差のプロットと要約表が表示されます。最初の表には以下の列があります。

**水準** 因子の水準。

**度数** 各水準の度数。

**標準偏差** X変数の水準ごとの応答変数の標準偏差。O'Brien検定における変換された変数のグループ平均は、標準偏差の2乗（分散）です。データが1つしかない水準については、標準偏差は計算されません。

**平均からの平均絶対偏差** グループ平均からの偏差の絶対値を平均した値。Levene 検定における変換された変数のグループ平均は、それらの絶対偏差を平均したものです。

**中央値からの平均絶対偏差** グループ中央値からのYの偏差の絶対値を平均した値。Brown-Forsythe 検定における変換された変数のグループ平均は、それらの絶対偏差を平均したものです。

2つ目の表は、等分散性の検定をまとめたもので、次の列があります。

**検定** 各検定の名前。

**F 値** 計算されたF統計量。「[等分散性検定の統計的詳細](#)」を参照してください。

**分子自由度** 分子で使用される自由度。因子の水準数を $k$ とすると、分子には $k-1$ 個の自由度があります。O'Brien、Brown-Forsythe、およびLeveneの統計量の計算では、データが1つしかない水準は無視されます。その場合の分子自由度は、計算に使用された水準数からさらに1を引いたものになります。

**分母自由度** 分母で使用される自由度。O'Brien、Brown-Forsythe、およびLeveneでは、検定統計量の計算に使用された因子水準の数だけ自由度が引かれます。因子の水準数を $k$ とすると、分母自由度は $n - k$ となります。

**p 値** 「分散がすべての水準にわたって等しい」という仮定のもとで、現在のF値より大きなF値を得る確率。

---

**メモ:** X変数のいずれかの水準で標本サイズが5に満たない場合は、警告が表示されます。標本サイズが小さいときの上記の検定の性能については、Brown and Forsythe (1974) と Miller (1972) を参照してください。

---



## 「Welchの検定」レポート

**F 値** 平均が等しいかどうかを調べる検定の  $F$  統計量。

**分子自由度** 検定の分子で使用される自由度。因子の水準数を  $k$  とすると、分子には  $k-1$  個の自由度があります。標本サイズが1しかない水準は、Welchの一元配置分散分析の計算には使用されません。その場合の分子自由度は、計算に使用された水準数からさらに1を引いたものになります。

**分母自由度** 検定の分母で使用される自由度。[「等分散性検定の統計的詳細」](#)を参照してください。

**p 値 (Prob>F)** 「母平均がすべての水準において等しい」という仮定のもとで、現在の  $F$  値より大きな  $F$  値を得る確率。

**t 検定** ( $X$  変数の水準数が2水準の場合のみ使用できます。)  $F$  値と  $t$  値の関係を示すために出力されます。 $t$  値は、 $F$  値の平方根です。

## 同等性検定のレポート

「一元配置」プラットフォームの「同等性検定」サブメニューにあるオプションは、平均や標準偏差に対して検定を行います。設定ウィンドウで検定を定義し、同等性検定・優越性検定・非劣性検定を選択して含めることができます。検定方法として、 $t$  検定と Wilcoxon 検定のいずれかを選べます。同等性検定のレポートは、図と要約表で構成されます。

同等性検定では、「2つの片側検定」(TOST; two one-sided tests) という手法を使い、2つの平均の差、または、2つの標準偏差の比が実質的に同等かどうかを検定します。平均に関する同等性検定については、Schuirmann (1987) を参照してください。平均に関する同等性検定では、「母平均の差は、閾値を超えている」という帰無仮説を検定するために、片側検定を2回、行います。2回の検定が両方ともこの帰無仮説を棄却すれば、「母平均の差は、上限値と下限値のどちらの閾値も超えていない」とみなされます。したがって、その場合、それら2群の母平均は「実質的に同等」とみなされます。[「同等性の検定の例」](#)を参照してください。

---

**メモ:** 同等性・優越性・非劣性の検定で3群以上ある場合に複数の群間比較を行う場合、有意水準の多重性調整は行われません。

---

## 同等性検定の設定ウィンドウ

「同等性検定」サブメニューでオプションを選択すると、検定を定義するためのウィンドウが開きます。

**検定の種類**  $t$  検定か Wilcoxon 検定かを指定します。

**t 検定** 検定として  $t$  検定を用います。

**Wilcoxon 検定** 検定として Wilcoxon 検定を用います。

**対立仮説** 検定の種類を選択します。どの対立仮説の検定を行うかを選択します。

**同等性（上下限）** 同等性検定を行います。このオプションは、グループ間の違いが同等性のマージンより大きくないことを示したいときに使います。

**優越性（片側）** 優越性検定を行います。このオプションは、一方のグループが他方より優れている（良い）ことを示したいときに使います。

**非劣性（片側）** 非劣性検定を行います。このオプションは、一方のグループが他方より劣っていないことを示したいときに使います。

**対立仮説の方向** （優越性検定と非劣性検定の場合にのみ指定します。）対立仮説の方向を指定します。

**仮説のプロット** 帰無仮説および対立仮説を図で表現したもの。

**比較方法** 比較の方法を指定します。

**すべてのペア** グループの水準のすべてのペア間で比較を行います。

**コントロール群との比較** 各水準を特定のコントロール群と比較します。

**分散に対する仮定** ( $t$  検定でのみ使用可能。) 検定の計算に使用する分散の仮定を指定します。

**等しい分散** プールした分散に基づいた検定が行われます。グループ間で分散が等しいと仮定したい場合は、このオプションを使用します。

**異分散** グループごとに異なる分散に基づいて検定が行われます。

**マージンと有意水準** 検定の有意水準。

**差** （平均に対する検定での差に対するマージンを指定します。）同等性検定・優越性検定・非劣性検定のマージン。このマージン ( $\Delta$ ) には、実質的に意味がある差を指定します。同等性検定の場合は、ここに指定する「差」は0より大きくなければなりません。

**比** （標準偏差に対する検定での比に対するマージンを指定します。）同等性検定・優越性検定・非劣性検定のマージン。標準偏差の比で指定します。このマージンは、実質的に意味を持つ標準偏差の比を示します。指定された「比」に対して、同等性の範囲は (比,  $1/\text{比}$ ) と定義されます。同等性検定の場合、ここに指定する「比」は1とは異なる値でなければなりません。

**アルファ** 検定の有意水準。

## 「同等性検定」レポート

検定のレポートでは、冒頭に対立仮説が表示されます。比較ごとに、以下の列を含む「検定」レポートが作成されます。

**差** （平均に対する検定で表示されます。） $t$  検定の場合は平均の差の推定値、Wilcoxon 検定の場合はスコア平均の差。

**差の標準誤差**（平均に対する  $t$  検定で表示されます。）平均の差に対する標準誤差。

**比**（標準偏差に対する検定で表示されます。）標準偏差の比に対する点推定値。

**下限値  $t$  値、上限値  $t$  値**（平均に対する  $t$  検定で表示されます。）非劣性検定・優越性検定の場合は、片方のみ計算されます。）片側検定の下限または上限の  $t$  比。

**下限値  $z$  値、上限値  $z$  値**（平均に対する Wilcoxon 検定で表示されます。非劣性検定・優越性検定の場合は、片方のみ計算されます。）片側検定の下限または上限の  $z$  比。

**下限値  $F$  値、上限値  $F$  値**（標準偏差に対する検定で表示されます。非劣性検定・優越性検定の場合は、片方のみ計算されます。）片側検定の下限または上限の  $F$  比。

**下限値  $p$  値、上限値  $p$  値**（非劣性検定・優越性検定の場合は、どちらか一方の  $p$  値のみ計算されます。）下限値または上限値の  $t$  比に対応する有意確率（ $p$  値）。

**最大  $p$  値**（同等性検定で表示されます。）2つの片側  $t$  検定の  $p$  値のうち大きいほうの値。

**Hodges-Lehmann**（平均に対する Wilcoxon 検定で表示されます。）グループの位置の差に対する Hodges-Lehmann 推定値。各グループから1つずつ取った2つのデータ値をペアとし、それらすべてのペアの差の中央値です。この推定量はフォレストプロットにもプロットされます。

**両側 90% 下限、両側 90% 上限** 平均の差または標準偏差の比に対する、信頼係数  $1-2\alpha$  の両側信頼区間の下限および上限。

**判定** 指定の有意水準で行った仮説検定の判定。

## 同等性検定のオプション

「同等性検定」・「優越性検定」・「非劣性検定」の赤い三角ボタンのメニューには、以下のオプションが含まれます。

**検定のレポート** 平均や標準偏差に対する同等性検定・優越性検定・非劣性検定の表示／非表示を切り替えます。「[「同等性検定」レポート](#)」を参照してください。

**散布図**（平均に対する  $t$  検定で表示されます。）散布図の表示／非表示を切り替えます。ペアごとの比較の信頼区間を示す散布図が作成されます。散布図は、受容域ごとに色分けされています。平均の差を示す軸上に、実質的に同等な区間、優越しているとみなす区間、非劣性であるとみなす区間がそれぞれの色で表示されます。このプロットは、「[ディフォグラム](#)」(diffogram) や、「平均－平均散布図」(mean-mean scatterplot) などとも呼ばれています。

---

**ヒント:** 点の上にマウスポインタを置くと、比較されているグループと差の推定値が表示されます。

---

散布図には次のオプションがあります。

**参照線の表示** 散布図上で点の参照線の表示／非表示を切り替えます。散布図の点が多い場合、この表示は推奨しません。マウスポインタを点の上に置くとラベルが表示されるので、点が多い場合にはその方法でラベルを確認してください。

**フォレストプロット** フォレストプロットの表示／非表示を切り替えます。標準偏差の平均または比の差に対する信頼区間がプロットされます。これらの信頼区間は、平均の差、または標準偏差の比を示す軸上に表示されます。色の付いた領域は、同等性・優越性・非劣性を示す区間です。

---

**ヒント:** 点の上にマウスポインタを置くと、比較されているグループと差または比の推定値が表示されます。

---

**ペアの比較** (平均に対する  $t$  検定で表示されます。) すべてのペアの比較についての「実質的な同等性」レポートの表示／非表示を切り替えます。このレポートには、「[同等性検定](#) レポート」にある値に加え、各ペアの検定のグラフが表示されます。

**削除** レポートウィンドウから、該当の検定に関するレポートを削除します。

## ロバストなあてはめのレポート

「一元配置」プラットフォームの [ロバスト] オプションには、外れ値 (極値) の影響を受けにくい2つの手法、[ロバストなあてはめ] と [Cauchy のあてはめ] があります。どちらのオプションも、一元配置プロット上にロバストな平均を示す直線を追加します。

- [ロバストなあてはめ] オプションは、Huber の  $M$  推定を行います。Huber の  $M$  推定は、ロバスト (頑健) な推定方法であり、応答変数における外れ値が推定結果に与える影響が小さいです。Huber の  $M$  推定では、Huber の損失関数を最小にするパラメータ推定値を求めます。Huber の損失関数は、誤差が小さい場合は2次曲線として増加し、誤差が大きい場合は線形に増加します。ロバスト推定 (頑健推定) の詳細については、Huber (1973) および Huber and Ronchetti (2009) を参照してください。「[\[ロバストなあてはめ\] オプションの例](#)」を参照してください。Huber の損失関数については、「[\[ロバストなあてはめ\] の統計的詳細](#)」を参照してください。
- このオプションでは、誤差が Cauchy 分布に従うと仮定されます。Cauchy 分布は正規分布よりも裾が広く、その結果、外れ値が推定に与える影響が小さくなります。このオプションは、データにある外れ値の割合が大きい場合に有用です。しかし、データが正規分布に近く、外れ値が少ない場合は、このオプションの推定結果は間違っただけになる可能性があります。[Cauchy のあてはめ] オプションは、最尤推定によってパラメータ推定値を算出します。

---

**メモ:** 起動ウィンドウでブロック変数を指定した場合、[ロバスト] オプションは使用できません。

---

[ロバスト] のオプションで作成されるレポートには、2つの表があります。表には、以下の列があります。

**$\sigma$**  通常の最小2乗推定における「誤差の標準偏差」(RMSE) に相当する  $\sigma$ 。

**カイ2乗** モデルが応答変数の単なる平均より予測モデルとして優れているという仮説の検定統計量。

**$p$  値** 「傾きが0である」に対するカイ2乗検定の  $p$  値。

**対数価値** 対数価値は、 $p$  値を  $-\log_{10}(p \text{ 値})$  と変換した値です。

**水準**  $X$  変数の水準。

**ロバスト 平均** Huber または Cauchy の手法を用いて推定したロバストな平均。

**標準誤差** パラメータ推定値の標準誤差を推定した値。

## 検出力のレポート

「一元配置」プラットフォームの「検出力」オプションを選択すると、特定の仮説検定に対して、検出力に関連した結果が表示されます。「[\[検出力\] オプションの例](#)」を参照してください。詳細については、「[検出力の統計の詳細](#)」を参照してください。

- **最小有意値 (LSV: Least Significant Value)** は、現在得られているデータにおいて、 $p$  値が特定の値（たとえば 0.05）になるようなパラメータ（もしくはパラメータ関数）の値です。つまり、現在のデータにおいて、ある値（たとえば 0.05）以下に  $p$  値がなって有意と判断される最小の効果の大きさを表します。最小有意値 (LSV) は、確率のスケールではなく、パラメータのスケール上で有意度を測定していると言えます。その値によって、計画とデータがどのくらい敏感かがわかります。
- **最小有意数 (LSN: Least Significant Number)** は、現在得られているデータにおいて、 $p$  値が特定の値（たとえば 0.05）になるような標本サイズです。つまり、 $\alpha$ （アルファ; 有意水準）、 $\sigma$ （シグマ; 誤差の標準偏差）、 $\delta$ （デルタ; 効果の大きさ、効果量）が与えられているときに、結果が有意となるほど推定値の分散が減少するために必要な標本サイズが、最小有意数 (LSN) です。最小有意数 (LSN) は、有意性を得るためにはどれだけの多くのデータが必要かを考える参考になります。最小有意数 (LSN) の値だけの標本サイズがあるとき、検出力は約 50% になります。
- **検出力** とは、グループ間に実際に差がある場合に、検定が有意となる確率（ $p$  値  $< \alpha$  となる確率）を指します。検出力は、標本サイズ、効果の大きさ（効果量）、誤差の標準偏差、および有意水準の関数です。検出力からは、その実験で、与えられた有意水準において差（効果の大きさ）をどれだけ的確に検出できるかがわかります。

---

**メモ:** 一元配置でグループが 2 群しかないときは、「検出力の詳細」で計算された最小有意値 (LSV) は Student の  $t$  検定における最小有意差 (LSD) と等しくなります。

---

## 検出力の詳細

「一元配置」プラットフォームの「検出力の詳細」表は、「モデルのあてはめ」プラットフォームのものと同じです。検出力の計算については、『基本的な回帰モデル』を参照してください。

「 $\alpha$ 」（アルファ; 有意水準）、「 $\sigma$ 」（シグマ; 誤差の標準偏差）、「 $\delta$ 」（デルタ; 効果の大きさ、効果量）、「数」（標本サイズ）の各列には値を 1～2 つ入力するか、連続した値の範囲として開始値、終了値、および間隔を入力します（[図 6.29](#)）。入力した値のすべての組み合わせに対して、検出力が計算されます。

**アルファ ( $\alpha$ )** 0～1 の値を取る有意水準（通常は 0.05、0.01、0.10 のどれかを使用します）。デフォルトでは 0.05 に設定されています。

**$\sigma$**  誤差の標準偏差。デフォルトとして、データから求められた誤差平均平方の平方根である RMSE が設定されています。

$\delta$  標準化していない効果の大きさ (effect size; 効果量)。効果の大きさを計算する方法については、『基本的な回帰モデル』を参照してください。最初のフィールドには仮説の平方和を  $n$  で割ったものの平方根、つまり  $\delta = \sqrt{SS/n}$  が設定されています。

**数 ( $n$ )** すべてのグループの標本サイズの合計。デフォルトでは、最初のフィールドに現在の標本サイズが設定されています。

**検出力を求める** 「 $\alpha$ 」、「 $\sigma$ 」、「 $\delta$ 」、および「数」の関数として、検出力 (有意な結果が出る確率) が計算されます。

**最小有意数を求める** 指定された「 $\alpha$ 」、「 $\sigma$ 」、「 $\delta$ 」で、およそ 50% の検出力を得るために必要な標本サイズ。

**最小有意値を求める**  $p$  値が  $\alpha$  以下となるパラメータまたは線形検定の値。これは「 $\alpha$ 」、「 $\sigma$ 」、標本サイズ (「数」)、および推定値の標準誤差の関数です。この機能は、X 変数の水準数が 2 つの場合にだけ使用でき、通常は個々のパラメータに適用します。

**調整済み検出力と信頼区間** 事後的に検出力を調べるときは、誤差標準偏差とパラメータに対して推定値を使用します

- 誤差標準偏差とパラメータに推定値を単純に代入して F 分布の非心度パラメータを計算するとバイアス (偏り) が生じます。それを調整するために、非心度パラメータに対してよりバイアスが小さい推定値を使って計算したものが、調整済み検出力です。
- また、調整済み検出力の信頼区間は、非心度推定値の信頼区間から算出されています。

調整済み検出力と信頼限界が計算されるのは、元の「 $\delta$ 」の推定値に対してだけで、 $\delta$  にその他の値を指定して計算することはできません (調整済み検出力は、 $\delta$  の推定値における確率的なばらつきを調整したもののからです)。

## 「対応のあるあてはめ」レポート

「一元配置」プラットフォームの [対応のある列を設定] オプションでは、対応がある場合の分析を行うために、対応付けのための変数 (ID 変数) を指定します。[対応のある列を設定] オプションは、異なるグループのデータが同じ個体から取られた場合など、データに対応がある場合に使用します。起動ウィンドウでブロック変数を指定した場合、[対応のある列を設定] オプションは使用できません。「[\[対応のある列を設定\] オプションの例](#)」を参照してください。

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

[対応のある列を設定] オプションでは、主に 2 つの処理が行われます。

- グループ変数 (「二変量の関係」における X 変数) と対応付けのための変数の両方を含んだ加法モデルが (反復比例あてはめアルゴリズムを使用して) あてはめられます。反復比例あてはめアルゴリズムは個体数が何百もあるときに非常に有効です。個体数が多い場合に、線形モデルを普通にあてはめると、計算に時間がかかり、膨大なメモリが必要になります。

- 一元配置プロットに、グループ間で対応する点をつなぐ線が表示されます。同じ ID を持つ観測値が複数ある場合は、それらの観測値の平均をつなぐ線が描かれます。一元配置プロットから線を消すには、**表示オプション** > **対応を表す線** を選択します。

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

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

---

## 一元配置プロットの要素

「一元配置」プラットフォームでは、連続尺度の  $Y$  変数とカテゴリカル尺度の  $X$  変数の関係を示すプロットが作成されます。このプロットを見ると、 $X$  変数の各水準における  $Y$  変数の分布がわかります。「一元配置」プラットフォームのオプションの多くは、このプロットに視覚的な要素を追加します。「一元配置」の赤い三角ボタンのメニューにある **表示オプション** サブメニューでは、表示要素を追加したり、削除したりできます。ここでは、以下の表示要素について詳しく説明します。

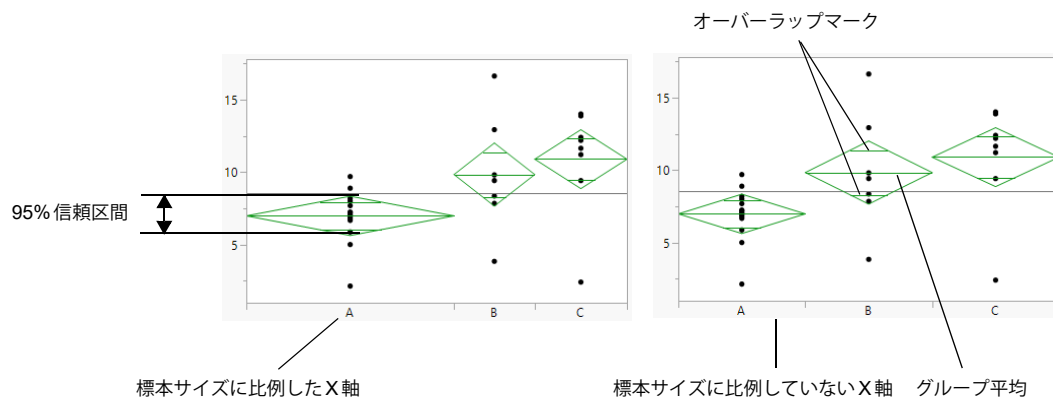
- 「平均のひし形と標本サイズに比例した  $X$  軸」
- 「平均線・平均誤差バー・標準偏差線」
- 「比較円」

### 平均のひし形と標本サイズに比例した $X$ 軸

「一元配置」プラットフォームでは、横軸が、デフォルトで標本サイズに比例した軸として表示されます。つまり、グループ水準間の間隔は、各グループに属するデータ点の数に比例します。**表示オプション** の **平均のひし形** オプションを選択するか、分散分析 (ANOVA) の検定を実行すると、プロット上に平均のひし形が表示されます。平均のひし形は、 $X$  変数の各水準における標本平均と信頼区間を示します。



図 6.6 [平均のひし形] と [標本サイズに比例したX軸] のオプション



### 平均のひし形の詳細

- 各ひし形の上下の頂点は、各グループの平均に対する  $(1 - \alpha) \times 100\%$  の両側信頼区間を表します。このひし形の両側信頼区間は、誤差の分散がすべての観測値において等しいという前提のもとに計算されています。よって、ひし形の高さは、グループ内にある標本サイズの平方根の逆数に比例します。
- [標本サイズに比例した X 軸] オプションが選択されているときは、横軸に沿った各グループの幅（ひし形の横幅）が、そのグループ（X 変数の各水準）の標本サイズに比例します。そのため、幅の狭いひし形は縦長になる傾向があります。それは、標本サイズが小さいグループでは、平均の信頼区間が広がるためです。
- 各ひし形の中央を横切る線はグループ平均を表します。
- ひし形内部の上下に引かれた線は、オーバーラップマークです。グループの標本サイズが等しいとき、ひし形がオーバーラップマークの内側で重なっている場合は、2つのグループ平均が所定の信頼水準において有意に異なることを示します。オーバーラップマークは、次の式で計算されます。

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \frac{\sqrt{2}}{2}$$

あるグループの上側のオーバーラップマークを標本サイズが等しい別のグループの下側のオーバーラップマークと比較した場合、分散が等しいと仮定すると、2標本  $t$  検定と等価になります。

- 平均のひし形は、プラットフォームのメニューから [平均/ANOVA/プーリングした  $t$  検定] または [平均/ANOVA] オプションを選択すると表示されますが、赤い三角ボタンをクリックし、メニューから [表示オプション] > [平均のひし形] を選択すれば、いつでも表示／非表示を切り替えることができます。

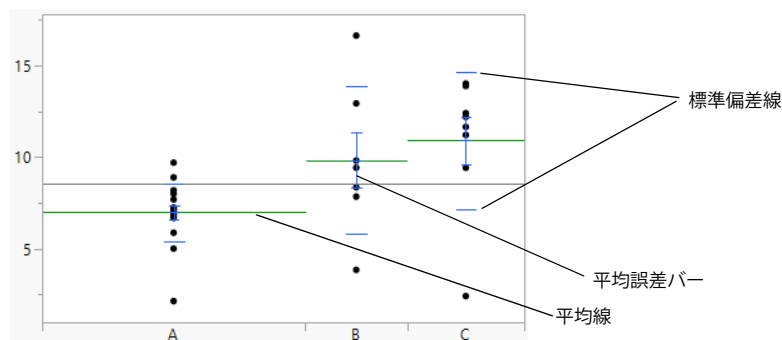
### 平均線・平均誤差バー・標準偏差線

「一元配置」プラットフォームで [表示オプション] > [平均線] オプションを選択すると、一元配置プロットに平均線が追加されます。平均線は、X 変数の水準ごとにおける応答変数の平均を示します。



平均誤差バーと標準偏差線は、赤い三角ボタンをクリックし、メニューから**【平均と標準偏差】**オプションを選択すると表示されます。平均誤差バーと標準偏差線の表示／非表示を個別に切り替える場合は、**【表示オプション】** > **【平均誤差バー】** または **【標準偏差線】** をそれぞれ選択します。

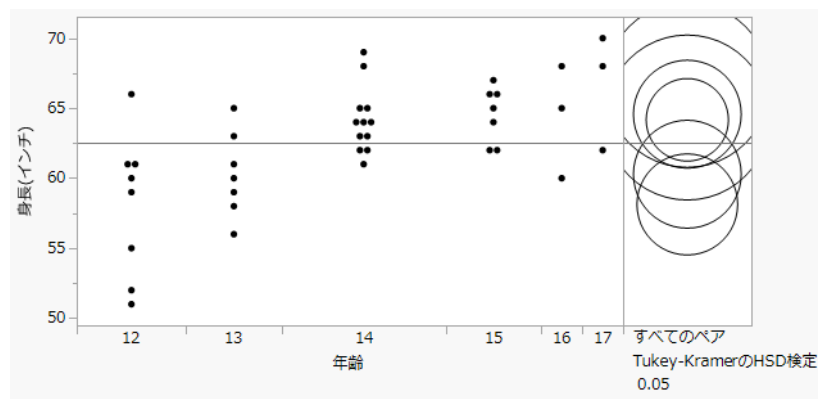
図6.7 平均線・平均誤差バー・標準偏差線



## 比較円

「一元配置」プラットフォームで多重比較検定（[各ペアのステップワイズ, Newman-Keuls 検定]を除く）を行うと、一元配置プロットに**比較円**が表示されます。比較円は、グループ平均を視覚的に比較したものです。図6.8は、[すべてのペア, TukeyのHSD 検定]の比較円です。円の半径は、検定と有意水準によって決まります。[「比較円の統計的詳細」](#)を参照してください。

図6.8 グラフによるグループ平均の比較



グループ平均の各ペアを、比較円の交わりを見て比較することができます。外側の交角によってグループ平均が有意に異なるかどうかわかります（図6.9）。

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



## 「一元配置」プラットフォームの別例

ここでは、「一元配置」プラットフォームを使った例をご紹介します。

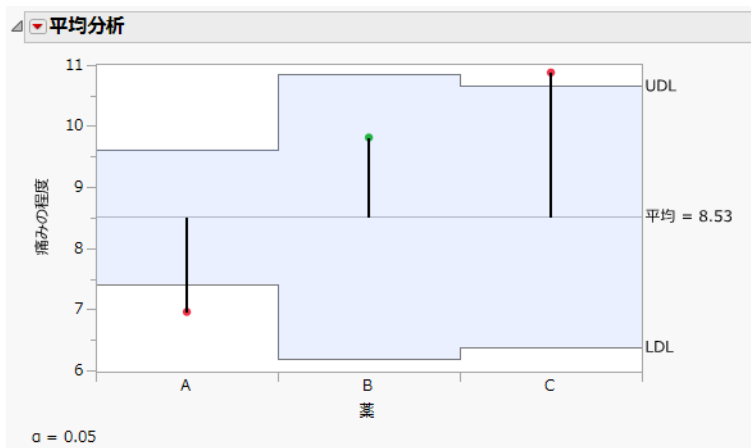
- 「平均分析の例」
- 「分散の平均分析の例」
- 「[各ペア,Studentのt検定] の例」
- 「[すべてのペア,TukeyのHSD検定] の例」
- 「[最適値との比較,HsuのMCB検定] の例」
- 「[コントロール群との比較,Dunnett検定] の例」
- 「[各ペアのステップワイズ,Newman-Keuls検定] の例」
- 「[平均の比較] で用意されている4種類の検定」
- 「Wilcoxon検定の例」
- 「等分散性検定の例」
- 「同等性の検定の例」
- 「[ロバストなあてはめ] オプションの例」
- 「[検出力] オプションの例」
- 「正規分位点プロットの例」
- 「累積確率プロットの例」
- 「[密度] の各オプションの例」
- 「[対応のある列を設定] オプションの例」
- 「データを積み重ねて一元配置分析を実行する例」

### 平均分析の例

この例では、「一元配置」プラットフォームを使い、平均分析（ANOM; analysis of means）によって平均の差を検定します。平均分析は、多重性を調整している検定です。

1. [ヘルプ] > [サンプルデータフォルダ] 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。
5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[平均分析法] > [平均分析] を選択します。

図 6.11 「平均分析」のグラフ



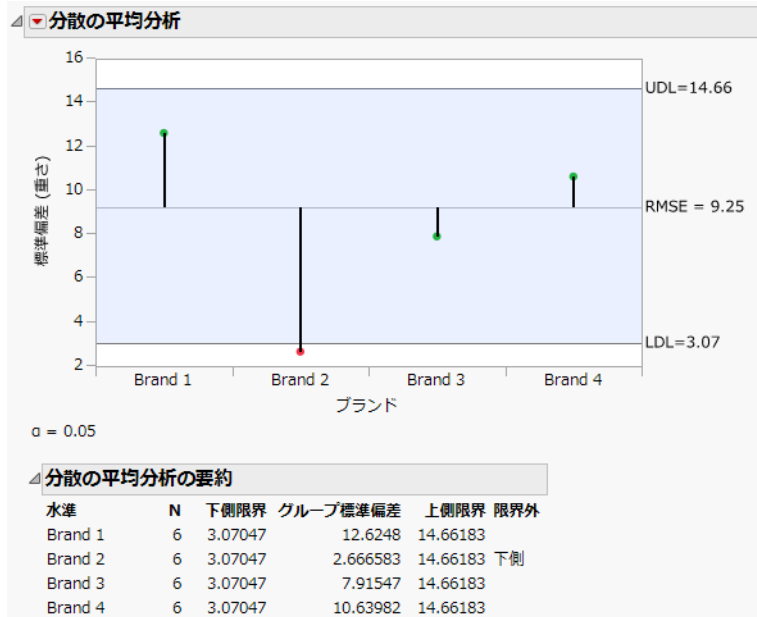
グラフを見ると、薬AとCの平均は、全体平均と有意に異なります。薬Aの平均は低く、薬Cの平均は高くなっています。各薬で標本サイズが異なっているため、薬ごとに決定限界が違ってきます。

## 分散の平均分析の例

この例では、「一元配置」プラットフォームを使い、分散の平均分析（ANOMV; analysis of means for variances）で分散の違いを検定します。4つのブランドのバネ製品について、0.10インチ伸ばすのに必要な重量を調べました。ブランドごとに6製品ずつ調べました。分散の平均分析（ANOMV）の統計的検定は、非正規分布に対して頑健ではありませんが、データは正規分布に従っていることが確認されています。ANOMVによって、ブランド間で分散が有意に異なるかどうかを調べます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Spring Data.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「重さ」を選択し、[Y, 目的変数] をクリックします。
4. 「ブランド」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[平均分析法] > [分散の平均分析] を選択します。
7. 「分散の平均分析」の赤い三角ボタンをクリックし、[要約レポートの表示] を選択します。

図6.12 「分散の平均分析」チャート



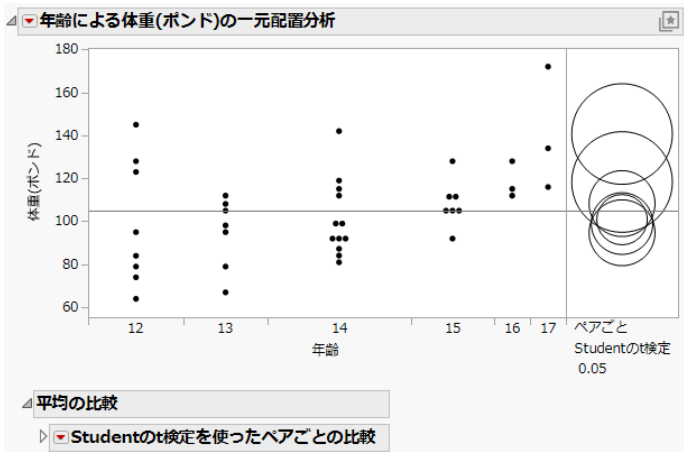
「Brand 2」の標準偏差が下側決定限界を下回っていることがわかります。つまり、「Brand 2」は、他のブランドと比較して、分散が有意に小さいことを示しています。

## 「各ペア, Studentのt検定」の例

この例では、「一元配置」プラットフォームにおけるt検定で、全てのペアの平均差を検定します。

1. 「ヘルプ」>「サンプルデータフォルダ」を選択し、「Big Class.jmp」を開きます。
2. 「分析」>「二変量の関係」を選択します。
3. 「体重(ポンド)」を選択し、「Y, 目的変数」をクリックします。
4. 「年齢」を選択し、「X, 説明変数」をクリックします。
5. 「OK」をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、「平均の比較」>「各ペア, Studentのt検定」を選択します。

図 6.13 「各ペア, Student の t 検定」の比較円の例



この平均の比較方法は、Fisher の最小有意差 (LSD) 法を用いて個々の平均ペア間で差を評価します。この方法では過誤率が守られず、多重比較の調整が行われないため、グループの数が多い場合は他の比較方法を使用してください。

図 6.14 「Student の t 検定を使ったペアごとの比較」レポートの例



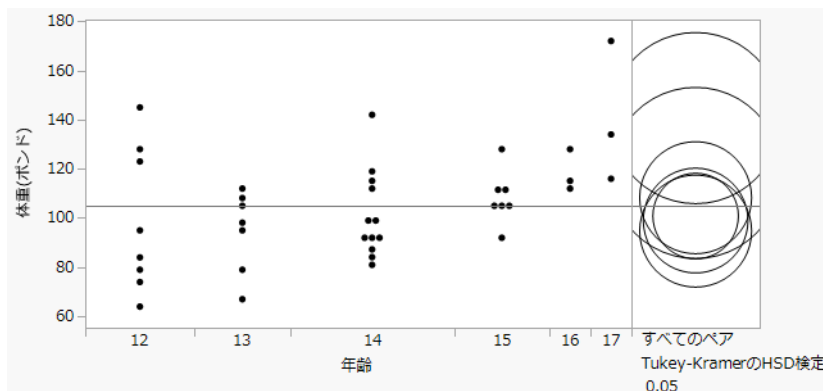
図6.14の先頭に表示されている行列は、観測された平均の差の絶対値から、LSD（最小有意差；検定が有意となる差）を引いたものです。値が正の場合、観測された平均の差がLSDよりも大きいことを意味していますので、2つの平均には統計的に有意な差があると結論できます。

## 「すべてのペア ,Tukey のHSD 検定」の例

この例では、「一元配置」プラットフォームにおける Tukey の HSD 検定を用います。

1. 「ヘルプ」 > 「サンプルデータフォルダ」を選択し、「Big Class.jmp」を開きます。
2. 「分析」 > 「二変量の関係」を選択します。
3. 「体重(ポンド)」を選択し、「Y, 目的変数」をクリックします。
4. 「年齢」を選択し、「X, 説明変数」をクリックします。
5. 「OK」をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、「平均の比較」 > 「すべてのペア ,Tukey のHSD 検定」を選択します。

図6.15 「すべてのペア ,Tukey のHSD 検定」の比較円





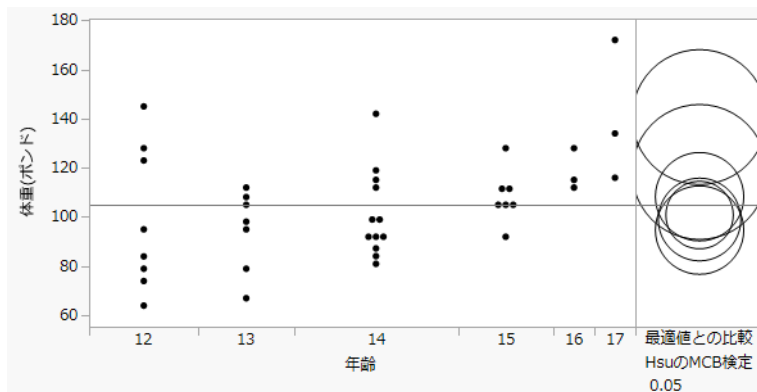


## 〔最適値との比較,HsuのMCB検定〕の例

この例では、「一元配置」プラットフォームにおけるHsuのMCB検定を用います。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「体重(ポンド)」を選択し、[Y, 目的変数] をクリックします。
4. 「年齢」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[平均の比較] > [最適値との比較 ,Hsu の MCB 検定] を選択します。

図6.17 〔最適値との比較,HsuのMCB検定〕の比較円の例



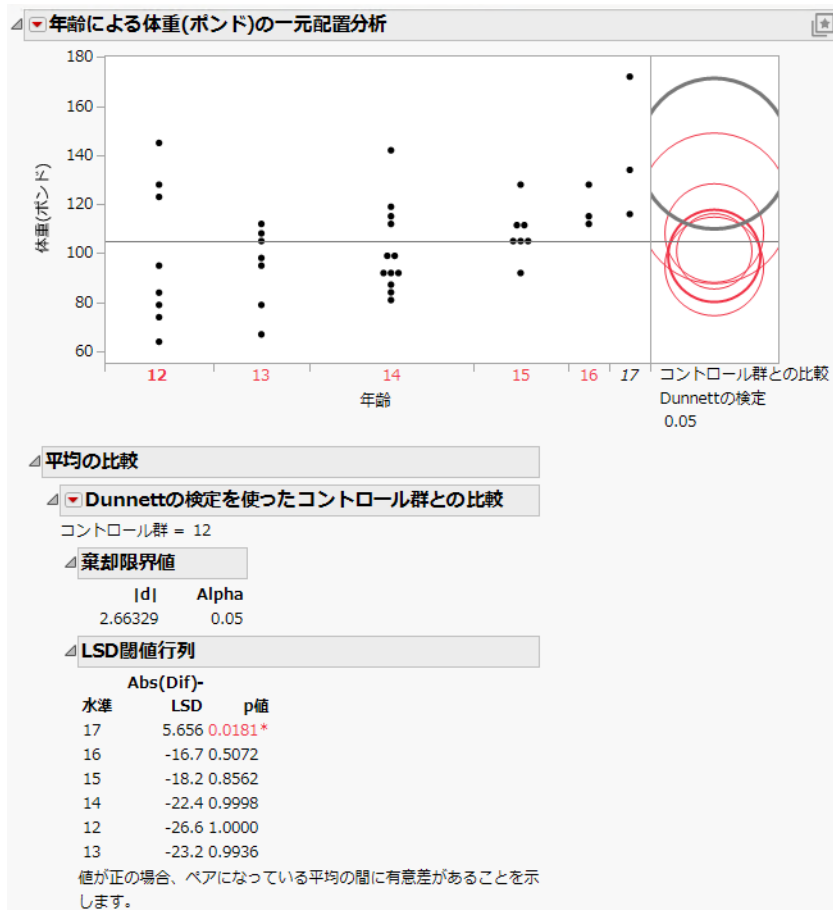


5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[平均の比較] > [コントロール群との比較,Dunnett検定]を選択します。
7. コントロール群とする年齢を「12」に設定します。

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

8. [OK] をクリックします。

図6.19 [コントロール群との比較(Dunnett)] の比較円の例



比較円や「LSD 閾値行列」を見ると、水準17だけが、コントロール群である水準12と有意に異なることがわかります。

**ヒント:** 片側検定を行うには、「Dunnettの検定を使ったコントロール群との比較」の赤い三角ボタンのメニューから「下片側 Dunnett」または「上片側 Dunnett」を選択します。

## 「各ペアのステップワイズ, Newman-Keuls 検定」の例

この例では、「一元配置」プラットフォームにおける Newman-Keuls 検定を用います。

1. 「ヘルプ」>「サンプルデータフォルダ」を選択し、「Big Class.jmp」を開きます。
2. 「分析」>「二変量の関係」を選択します。
3. 「体重(ポンド)」を選択し、「Y, 目的変数」をクリックします。
4. 「年齢」を選択し、「X, 説明変数」をクリックします。
5. 「OK」をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、「平均の比較」>「各ペアのステップワイズ, Newman-Keuls 検定」を選択します。

図 6.20 「Newman-Keuls ステップワイズ法による各ペアの比較」レポートの例

平均の比較

Newman-Keulsステップワイズ法による各ペアの比較

警告: Newman-Keuls検定では、検定全体における過誤率(family-wise error rate)が守られていません。結果の解釈には注意が必要です。

最小q\*

最小q\*

Alpha

2.45044

0.05

文字の接続レポート

水準		平均	標準誤差
17	A	140.67	11.531
16	A B	118.33	11.531
15	A B	108.29	7.549
14	B	100.83	5.765
12	B	99.00	7.061
13	B	94.71	7.549

同じ文字で接続されていない水準には有意差があります。

「文字の接続レポート」によると、水準17は、16と15を除くすべての水準と有意に異なっています。

## 「平均の比較」で用意されている4種類の検定

この例では、「一元配置」プラットフォームにおける「平均の比較」に関する4種類の検定を比べてみます。

1. 「ヘルプ」>「サンプルデータフォルダ」を選択し、「Big Class.jmp」を開きます。
2. 「分析」>「二変量の関係」を選択します。
3. 「体重(ポンド)」を選択し、「Y, 目的変数」をクリックします。

4. 「年齢」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[平均の比較] のオプションをすべて選択します。この例での「コントロール群との比較,Dunnett」オプションでは、コントロール群として17歳を選択してください。

4種類の検定はどれもグループ平均の差を検定するものです。しかし、同じ仮説に対する検定で、結果が異なることがあります。

図6.21 4つの多重比較検定の比較円

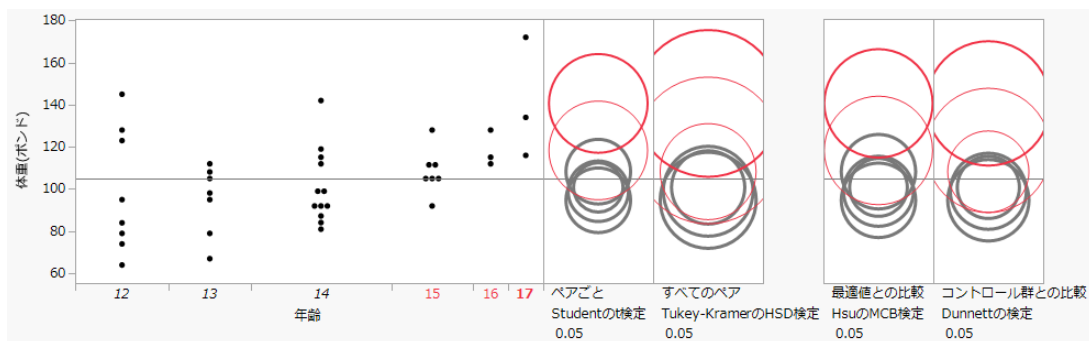


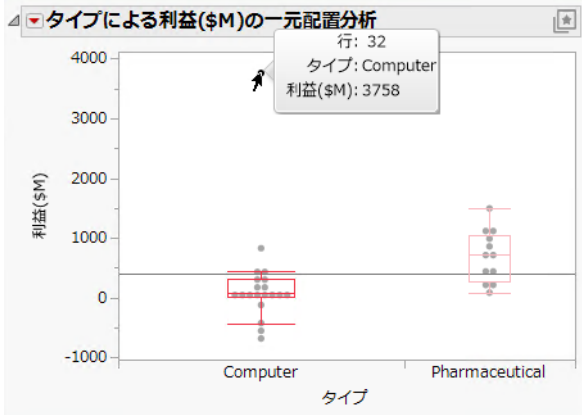
図6.21では、「年齢」が17歳のグループが強調表示されています。その他の円には、17歳のグループとの関係に基づいて色が付けられています。Studentのt検定とHsu検定では、「年齢」が15歳のグループ（上から3番目の円）が、グレーで表示されています。これは、このグループが17歳のグループと有意に異なることを示しています。一方、Tukey検定とDunnett検定では、15歳と17歳は赤で表示されています。これは、これらのグループ間に有意差が認められないことを示しています。

## Wilcoxon 検定の例

「一元配置」プラットフォームでWilcoxon検定を使い、会社の業種によって利益の平均に差があるかどうかを調べてみましょう。2つの異なる業種の会社（製薬業界12社とコンピュータ業界20社）のデータを使用します。

1. [ヘルプ] > [サンプルデータフォルダ] 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. 「利益(\$M)」を選択し、[Y, 目的変数] をクリックします。
4. 「タイプ」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[表示オプション] > [箱ひげ図] を選択します。

図 6.22 「Computer」の利益の分布



箱ひげ図を見ると、グループ内の利益は正規分布に従わず、対称でもないようです。また、32行目に非常に大きな値があり、通常のt検定ではこの1つのデータ値が検定結果に大きく影響する恐れがあります。

7. 「一元配置分析」の赤い三角ボタンをクリックし、[ノンパラメトリック] > [Wilcoxon / Kruskal-Wallis 検定] を選択します。

図 6.23 Wilcoxon 検定の結果

Wilcoxon/Kruskal-Wallisの検定(順位和)					
水準	度数	スコア和	スコアの期待値	スコア平均	(平均-平均0)/標準偏差0
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289

Wilcoxon 2標本検定 (正規近似)			
S	Z	p値(Prob>Z)	p値(Prob> Z )
283	3.28916	0.0005*	0.0010*

Kruskal-Wallis検定 (カイ2乗近似)		
カイ2乗	自由度	p値(Prob>ChiSq)
10.9470	1	0.0009*

0.5の連続修正を行った場合（「正規近似」の結果）と0.5の連続修正を行わなかった場合（「カイ2乗近似」の結果）の両方で、 $p$ 値が0.0010以下であり、有意となっています。これにより、これら2つの分布には有意な違いがあり、会社のタイプによって利益の分布は異なっていると結論付けることができます。

正規近似およびカイ2乗近似の検定は、いずれも検定統計量の漸近分布に基づいたものです。正確検定を行うこともできます。

8. 「一元配置分析」の赤い三角ボタンをクリックし、[ノンパラメトリック] > [正確検定] > [正確 Wilcoxon 検定] を選択します。

図 6.24 正確 Wilcoxon 検定の結果

Wilcoxon/Kruskal-Wallisの検定(順位和)					
水準	度数	スコア和	スコアの期待値	スコア平均	(平均-平均0)/標準偏差0
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289
Wilcoxon 2標本検定 (正規近似)					
S	Z	p値(Prob>Z)	p値(Prob> Z )		
283	3.28916	0.0005*	0.0010*		
Kruskal-Wallis検定(カイ2乗近似)					
カイ2乗	自由度	p値(Prob>ChiSq)			
10.9470	1	0.0009*			
二標本: 正確検定					
S	p値(Prob≥S)	p値(Prob≥ S-Mean )			
283	0.0003*	0.0005*			

検定の結果では、 $S = 283$  となっています。この値は、「タイプ」のうち標本サイズが小さいほうの水準（「Pharmaceutical」）のオブザベーションの順位の和です。「2群での母集団分布が同じである」という仮定のもとで、 $S$  とその期待値との差の絶対値が、観測された値（観測された  $S$  とその期待値との差の絶対値）以上となる確率は、0.0005 です。これは、Wilcoxon 検定の両側検定に対する結果であり、「利益の分布が「タイプ」によって異なる」という帰無仮説を棄却する根拠となります。

この例では、外れ値があるので、Wilcoxon 検定のほうが分散分析や分散が異なる  $t$  検定よりも適切でしょう。Wilcoxon 検定は、32 行目の大きな値の影響を受けにくく、正規分布を前提としていないからです。

## 等分散性検定の例

「一元配置」プラットフォームを使って2つのグループの分散が等しいことを検定します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「身長(インチ)」を選択し、[Y, 目的変数] をクリックします。
4. 「性別」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. [一元配置分析] の赤い三角ボタンをクリックし、[等分散性の検定] を選択します。

図 6.25 「分散が等しいことを調べる検定」レポートの例



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

## 同等性の検定の例

「一元配置」プラットフォームを使い、犬の治療に使用するモルヒネの代替薬剤がモルヒネと比べて劣っていないかどうかを調べます。最初の2つの時点の間で差を比較します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Dogs.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。



3. 「差」を選択し、**[Y, 目的変数]** をクリックします。
4. 「薬」を選択し、**[X, 説明変数]** をクリックします。
5. **[OK]** をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、**[同等性検定] > [平均]** を選択します。

図6.26 「同等性/優越性/非劣性の検定」ウィンドウ

検定の種類

☒ t検定 ☐ Wilcoxon検定

対立仮説

☒ 同等性（上下限）  
☐ 優越性（片側）  
☐ 非劣性（片側）

分散に対する仮定

☒ 等しい分散  
☐ 異分散

同等性の仮説

差   
Alpha

比較方法

☒ すべてのペア  
☐ コントロール群との比較

7. **[非劣性（片側）]** を選択します。
8. 「分散に対する仮定」は **[等しい分散]** のままにします。
9. 「差」を1に設定します。これは、非劣性のマージンです。
10. **[OK]** をクリックします。

図 6.27 非劣性の検定の例



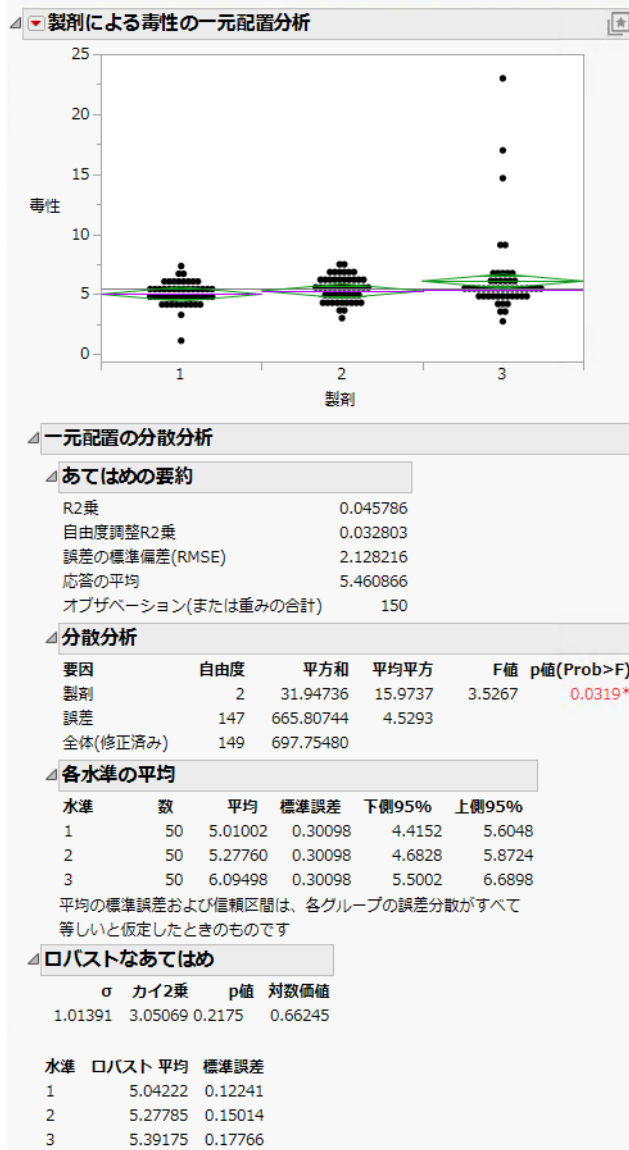
2つの薬剤の標本平均の差は、-0.639です。平均差に対する下側信頼限界は-1.93で、下側非劣性限界の-1.0を下回っています。代替薬剤が標準薬剤より劣っていないと結論することはできません（つまり、標準薬剤に対する代替薬剤の非劣性を示すことはできませんでした）。

## 【ロバストなあてはめ】オプションの例

「一元配置」プラットフォームを使い、3つのグループのうちの1つに外れ値が含まれるデータに対して「ロバストなあてはめ」を実行してみましょう。3種類の製剤で毒性を測定したデータを使います。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Drug Toxicit.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「毒性」を選択し、[Y, 目的変数] をクリックします。
4. 「製剤」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. [一元配置分析] の赤い三角ボタンをクリックし、[平均/ANOVA] を選択します。
7. [一元配置分析] の赤い三角ボタンをクリックし、[ロバスト] > [ロバストなあてはめ] を選択します。

図 6.28 ロバストなあてはめの例



標準の「分散分析」レポートを見ると、 $p$  値が0.0319であるため、「3つの製剤は有意に異なる」という結論に導かれてしまう可能性があります。しかし、「ロバストなあてはめ」レポートを確認すれば、 $p$  値が0.2175になっているため、「3つの製剤は有意に異なる」とは結論できません。いくつかのデータで毒性が不自然に高く記録され、通常の分散分析に対して思わしくない影響を与えているのかもしれない。

## 〔検出力〕 オプションの例

この例では、「一元配置」プラットフォームの〔検出力〕 オプションを用います。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Typing Data.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「タイプ速度」を選択し、[Y, 目的変数] をクリックします。
4. 「ブランド」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. [一元配置分析] の赤い三角ボタンをクリックし、[検出力] を選択します。
7. 「開始値」行の「 $\delta$ 」(左から3列目の1番上のボックス) に「2」、「数」に「11」と入力します。
8. 「終了値」行の「 $\delta$ 」に「6」、「数」に「17」と入力します。
9. 「間隔」行の「 $\delta$ 」と「数」に「2」と入力します。
10. [検出力を求める] チェックボックスをオンにします。

図 6.29 「検出力の詳細」 ウィンドウの例

**検出力の詳細**

ブランド  
クリックして1つか2つ、または連続する値をそれぞれに入力する。

	$\alpha$	$\sigma$	$\delta$	数
開始値:	0.050	4.27033	2	11
終了値:	.	.	6	17
間隔	.	.	2	2

☒ 検出力を求める  
☐ 最小有意数を求める  
☐ 最小有意値を求める  
☐ 調整済み検出力と信頼区間

完了 キャンセル ヘルプ

シーケンスのすべての組み合わせに対して計算。

11. [完了] をクリックします。

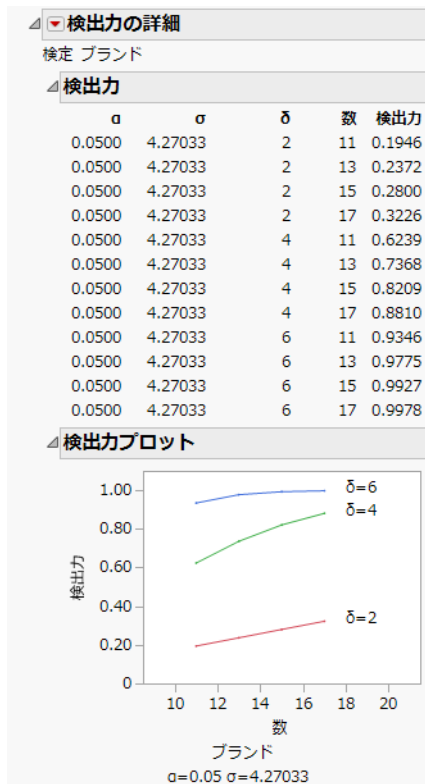
**メモ:** [完了] ボタンは、必要なオプションがすべて適用されるまでグレー表示のままです。

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

検出力の値をプロットにするには、次の手順に従います。

12. 表の下赤い三角ボタンをクリックし、[検出力プロット] を選択します。

図6.30 「検出力」レポートの例



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

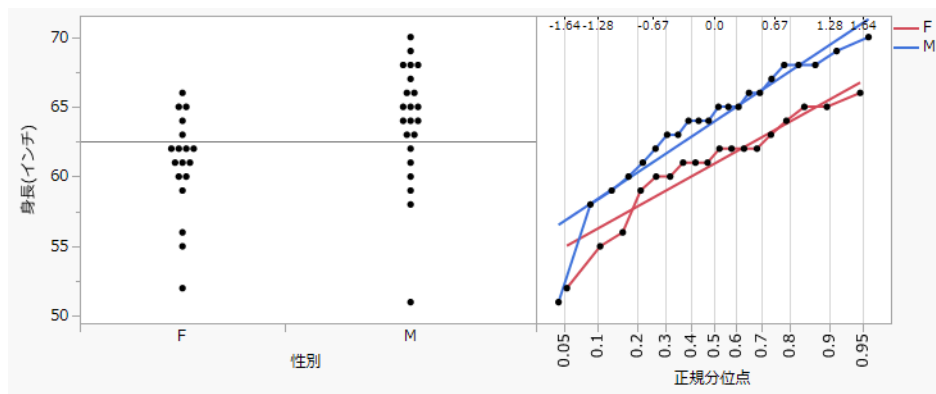
検出力は「 $\delta$ 」と「数」の組み合わせごとにプロットされています。一般的な性質として、「数」（標本サイズ）と「 $\delta$ 」（平均の差）が大きくなるほど検出力は上がりますが、この例でもその性質を確認することができます。

## 正規分位点プロットの例

「一元配置」プラットフォームを使って正規分位点プロットを作成します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「身長(インチ)」を選択し、[Y, 目的変数] をクリックします。
4. 「性別」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、[正規分位点プロット] > [分位点-実測値プロット] を選択します。

図 6.31 正規分位点プロットの例



次のことを注意してください。

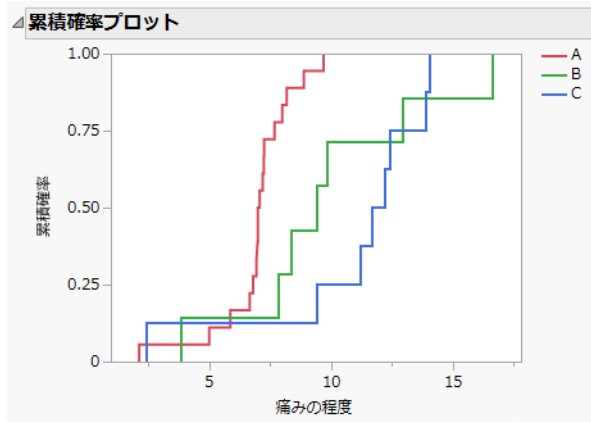
- あてはめ線はデフォルトで表示されます。
- データ点はあてはめ線のごく近くに分布しています。つまり、データは正規分布に従っていることを示唆しています。

## 累積確率プロットの例

「一元配置」プラットフォームを使って累積確率プロットを作成します。

1. [ヘルプ] > [サンプルデータフォルダ] 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。
5. [OK] をクリックします。
6. [一元配置分析] の赤い三角ボタンをクリックし、[累積確率プロット] を選択します。

図 6.32 累積確率プロットの例



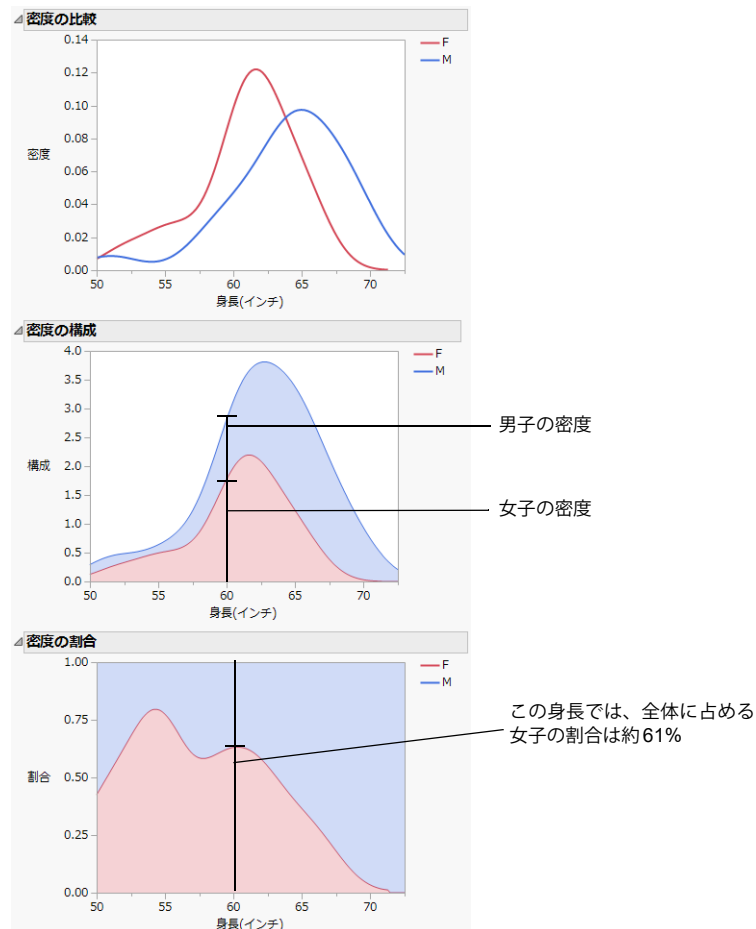
累積確率プロットには、X変数の水準ごとに曲線が描かれます。

## 【密度】の各オプションの例

この例では、「一元配置」プラットフォームの【密度】オプションをご紹介します。

1. 【ヘルプ】 > 【サンプルデータフォルダ】を選択し、「Big Class.jmp」を開きます。
2. 【分析】 > 【二変量の関係】を選択します。
3. 「身長(インチ)」を選択し、【Y, 目的変数】をクリックします。
4. 「性別」を選択し、【X, 説明変数】をクリックします。
5. 【OK】をクリックします。
6. 「一元配置分析」の赤い三角ボタンをクリックし、【密度】 > 【密度の比較】、【密度】 > 【密度の構成】、および【密度】 > 【密度の割合】の3つのオプションをすべて選択します。

図 6.33 「密度」の各オプションの例



## 「対応のある列を設定」オプションの例

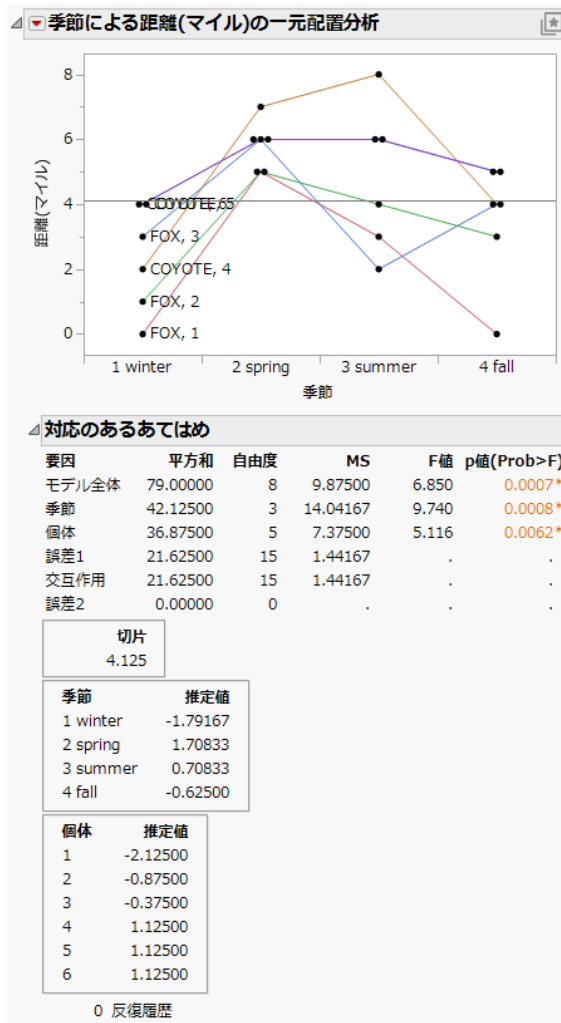
「一元配置」プラットフォームを使い、6種類の動物がそれぞれの季節に移動した距離のデータを分析します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Matching.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「距離(マイル)」を選択し、[Y, 目的変数] をクリックします。
4. 「季節」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. [一元配置分析] の赤い三角ボタンをクリックし、[対応のある列を設定] を選択します。
7. 対応のある列として「個体」を選択します。



8. **[OK]** をクリックします。

図 6.34 「対応のあるあてはめ」レポートの例



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

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

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

## データを積み重ねて一元配置分析を実行する例

分散分析のデータは、JMP 以外の場面においては、1 行に複数の観測値が含まれている形式で保存されているかもしれません。JMP でそのような形式のデータを分析するには、データを読み込んでから、各行が 1 つの観測値を含む形式に構成し直す必要があります。たとえば、スプレッドシート形式のデータがあるとしましょう。このスプレッド形式のデータには、3 つの異なる製造ラインで作られた商品のデータが、3 列に表示されています。これを JMP のデータテーブルに変更するには、各行が 1 個の商品のデータに対応するように、積み重ねた形式にする必要があります。

### 状況と目的

この例では、ファイル「Fill Weights.xlsx」を使用します。このファイルには、3 つの異なる製造ラインから無作為抽出された箱入りシリアルの重さが含まれています。データの形式は、[図 6.35](#) のとおりです。

- 「ID」の列には、箱入りシリアルの ID が含まれています。
- 「Line」の列には、各製造ラインから抽出された箱入りシリアルの重さ（単位はオンス）が含まれています。

図 6.35 データの形式

Weights					
ID	Line A	ID	Line B	ID	Line C
215	12.42	705	13.63	254	11.73
287	12.49	670	12.56	282	11.40
381	12.80	715	12.87	938	12.78
		683	13.09	597	12.19
		514	13.31	179	12.25
		517	12.64		
		946	12.75		

箱入りシリアル目標の重さは、12.5 オンスです。調べたいのは、この 3 つの製造ラインがすべて目標を満たしているかどうかということですが、まずは、各製造ラインの重さの平均値を比較してみることになります。各ラインの重さの平均を比較するには、一元配置分析を使用できます。

「一元配置」プラットフォームを使用するには、次の手順に従います。

- データを JMP に読み込みます。「[データの読み込み](#)」を参照してください。
- 各行が観測値を 1 個だけ含む形式になるように、データを構成し直します。このためには、箱入りシリアルの ID、製造ラインの名前、重さの列を作成する必要があります。「[データの積み重ね](#)」を参照してください。

## データの読み込み

この例では、Microsoft Excel のデータを JMP に読み込む2つの方法を紹介します。どちらか1つの方法で読み込むか、または両方とも試してみてください。

- **【ファイル】 > 【開く】** から Microsoft Excel ファイルを選択し、Excel 読み込みウィザードを使ってデータを読み込みます。[「Excel 読み込みウィザードを使ったデータの読み込み」](#)を参照してください。この方法は任意の Excel ファイルに使用でき、便利です。
- Microsoft Excel のデータを、JMP のデータテーブルにコピー＆ペーストします。[「Excel データのコピー＆ペースト」](#)を参照してください。小さなデータファイルの場合は、この方法を使用できます。

Microsoft Excel からデータを読み込む方法については、『JMP の使用法』を参照してください。

### Excel 読み込みウィザードを使ったデータの読み込み

1. **【ヘルプ】>【サンプルデータフォルダ】**を選択し、「Samples/Import Data」フォルダにある「Fill Weights.xlsx」を開きます。

Excel 読み込みウィザードが表示されます。

2. **「列見出しの開始行」**に「3」と入力します。

Excel ファイルの1行目には表の説明が入力されており、2行目は空白です。列の見出しは3行目から始まります。

3. **「列見出しの行数」**に「2」と入力します。

Excel ファイルでは、3行目と4行目に列の見出しが含まれています。

4. **【読み込み】**をクリックします。

図 6.36 Excel 読み込みウィザードで作成した JMP のデータテーブル

	Weights-ID	Weights-Line A	Weights-ID 2	Weights-Line B	Weights-ID 3	Weights-Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.40
3	381	12.80	715	12.87	938	12.78
4	.	.	683	13.09	597	12.19
5	.	.	514	13.31	179	12.25
6	.	.	517	12.64	.	.
7	.	.	946	12.75	.	.

7行のデータが作成され、各行に3つのラインのデータが表示されています。また、3つの製造ラインそれぞれのIDと重みの列があり、全部で6列になっています。

IDの列名の「Weights」の部分は、不要であり、誤解を招く恐れがあります。ここで列名を変更することもできますが、データの積み重ねを行った後で列名を変更したほうが効率的です。

5. **「データの積み重ね」**に進みます。

## Excel データのコピー＆ペースト

1. Microsoft Excel でファイル「Fill Weights.xlsx」を開きます。
2. 表内の、「Weights」という見出し行以外の部分を選択します。
3. 右クリックして【コピー】を選択します。
4. JMP で【ファイル】>【新規作成】>【データテーブル】を選択します。
5. 【編集】>【列名とともに貼り付け】を選択します。

【列名とともに貼り付け】は、データ値だけでなく、列名も一緒にコピーしたい場合に使用するオプションです。

図6.37 【列名とともに貼り付け】で作成した JMP のデータテーブル

	ID	Line A	ID 2	Line B	ID 3	Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.40
3	381	12.80	715	12.87	938	12.78
4	.	.	683	13.09	597	12.19
5	.	.	514	13.31	179	12.25
6	.	.	517	12.64	.	.
7	.	.	946	12.75	.	.

6. 「データの積み重ね」に進みます。

## データの積み重ね

1つのオブザベーションを1行に配置するには、[列の積み重ね] オプションを使用します。[列の積み重ね] オプションの詳細については、『JMPの使用法』を参照してください。

1. JMP のデータテーブルで、【テーブル】>【列の積み重ね】を選択します。
2. 6列すべてを選択し、【積み重ねる列】をクリックします。
3. 【複数系列の積み重ね】チェックボックスをオンにします。

ここで積み重ねるのは、「ID」と「Line」の2つの系列です。「系列の数」はデフォルトで「2」に設定されているので、変更する必要はありません。また、各系列は隣接しておらず、ID、Line A、ID、Line B、ID、Line C と、交互になっています。このため、【隣接】はオフのままにしておきます。

4. 【行による積み重ね】をオフにします。
5. 【欠測値の行を除外】をオンにします。
6. 「出力テーブル名」に「積み重ねたテーブル」と入力します。
7. 【OK】をクリックします。

この新しいデータテーブルで、「データ」および「データ2」には、それぞれIDと重さが含まれています。

8. 「ラベル」列の見出しを右クリックし、【列の削除】を選択します。

「ラベル」に表示されている項目は、元のデータテーブルで箱のIDの見出しとして表示されていたものであり、ここでは不要です。

9. 列の見出しをダブルクリックし、次のように列名を変更します。

- 「データ」→「ID」
- 「ラベル 2」→「ライン」
- 「データ 2」→「重さ」

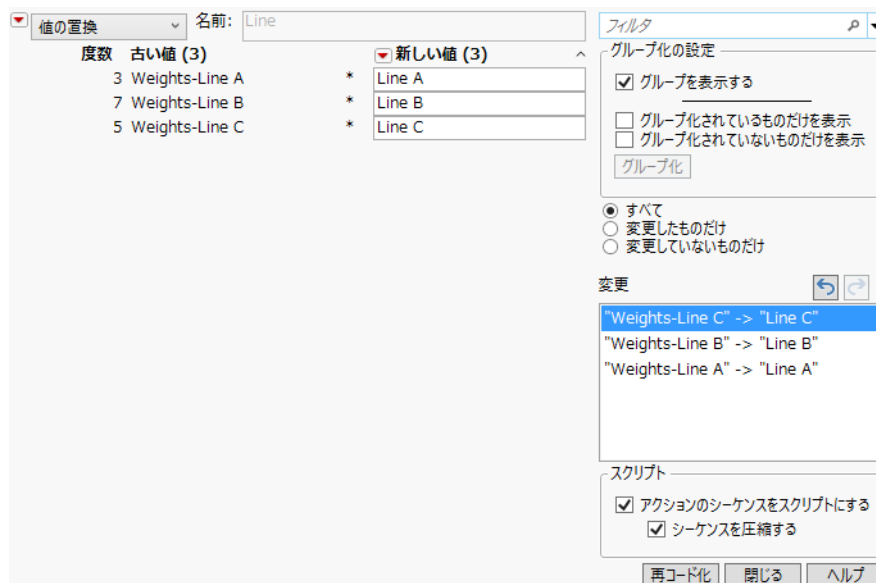
10. 「列」パネルで、「ID」の左側のアイコンをクリックし、**[名義尺度]** を選択します。

「ID」列の値は数値ですが、箱を識別する番号であるため名義尺度として扱う必要があります。この例では重大な支障はありませんが、列には常に適切な尺度を割り当てるようにしましょう。

11. (**[ファイル]** > **[開く]** を使って Excel のデータを読み込んだ場合のみ) 次の手順を行います。

1. 「ライン」列の見出しをクリックして列を選択し、**[列]** > **[再コード化]** を選択します。
2. **[列の新規作成]** をクリックして、**[値の置換]** を選択します。
3. 「新しい値」の値を、[図 6.38](#) のように変更します。

図 6.38 列の値の再コード化



4. **[再コード化]** をクリックします。

これで、JMPの分析に適した構造のデータテーブルが作成されました。1つの行は、1つの箱入りシリアルに対応しています。1列目には箱のID、2列目には製造ライン、3列目には重さが含まれています ([図 6.39](#))。

図 6.39 再コード化したデータテーブル

	ID	ライン	重さ
1	215	ライン A	12.42
2	287	ライン A	12.49
3	381	ライン A	12.8
4	705	ライン B	13.63
5	670	ライン B	12.56
6	715	ライン B	12.87
7	683	ライン B	13.09
8	514	ライン B	13.31
9	517	ライン B	12.64
10	946	ライン B	12.75
11	254	ライン C	11.73
12	282	ライン C	11.4
13	938	ライン C	12.78
14	597	ライン C	12.19
15	179	ライン C	12.25

## 一元配置分析の実行

ここでは、以下のことを行います。

- 一元配置の分散分析を実行し、3つの製造ラインの重さの平均値に差があるかどうかを検定する。
- 比較円を作成し、どのラインに差があるかを調べる。
- 点に ID のラベルを付ける。これは、箱の重さを測り直したり、詳しく調査したりしたいときに便利です。

まず、「積み重ねたテーブル」がアクティブになっていることを確認します。

1. **[分析] > [二変量の関係]** を選択します。
2. 「重さ」を選択し、**[Y, 目的変数]** をクリックします。
3. 「ライン」を選択し、**[X, 説明変数]** をクリックします。
4. **[OK]** をクリックします。
5. **[一元配置分析]** の赤い三角ボタンをクリックし、**[平均/ANOVA]** を選択します。

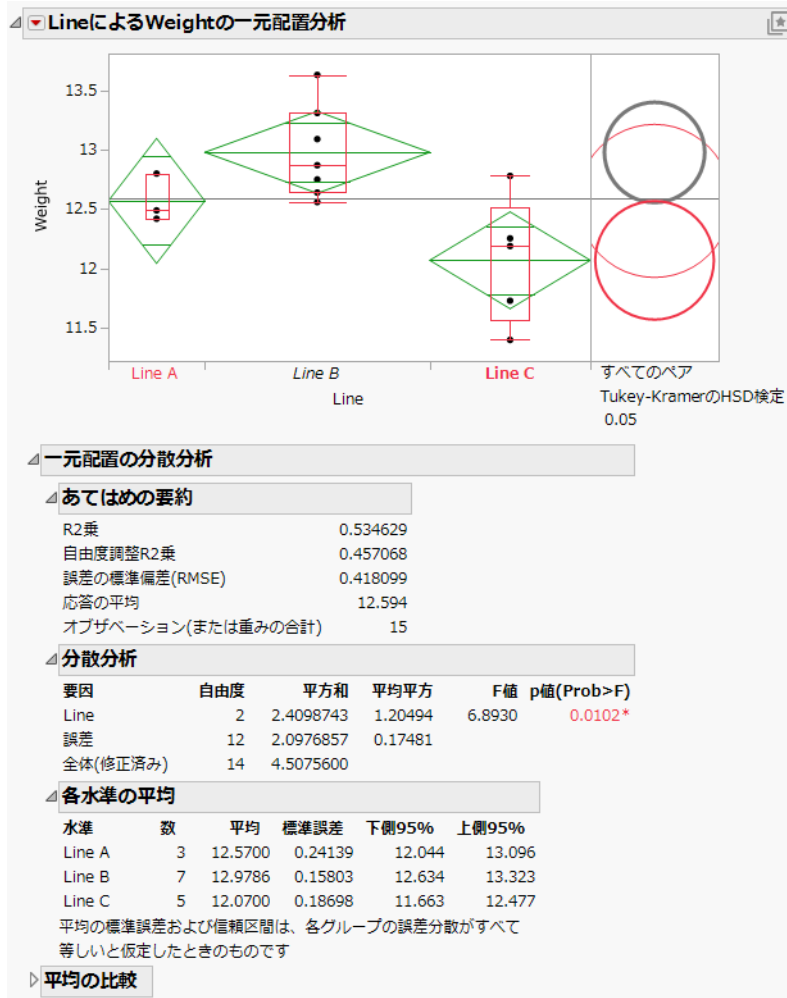
一元配置分析のプロットに表示されるひし形は、各製造ラインの平均値の95%信頼区間を示します。平均のひし形の外にある点は、外れ値のように見えますが、そうではありません。このことを確認するために、箱ひげ図を追加してみましょう。

6. 「一元配置分析」の赤い三角ボタンのメニューをクリックし、**[表示オプション] > [箱ひげ図]** を選択します。  
すべての点は、箱ひげ図の境界線内に収まっており、外れ値ではないことがわかります。
7. データテーブルの「列」パネルで「ID」を右クリックし、**[ラベルあり/ラベルなし]** を選択します。
8. プロットの点の上にカーソルを置くと、その点の「ID」、「ライン」、「重さ」の値が表示されます (図 6.40)。
9. 「一元配置分析」の赤い三角ボタンをクリックし、**[平均の比較] > [すべてのペア, Tukey の HSD 検定]** を選択します。

プロットの右側に比較円が表示されます。

10. 下の比較円をクリックします。

図 6.40 ラインごとの重さの一元配置分析



「分散分析」レポートによると、 $p$  値は0.0102となっており、この結果は「いずれかの平均が他の平均と異なっている」ことを示す証拠になっています。プロットにおいて、選択している「ラインC」の比較円は、赤で表示されています。「ラインB」の円は太いグレーで示されており、「ラインC」と「ラインB」の平均は0.05の有意水準で有意差があると言えます。「ラインA」と「ラインB」の平均には、統計的に有意な差は見られません。

一元配置分析プロットに表示される平均のひし形は、各製造ラインの平均値に対する両側95%信頼区間を示しています。「一元配置の分散分析」レポートの「各水準の平均」に、両側95%信頼区間の下限と上限が数値で示されています。どちらの結果からも、「ラインB」と「ラインC」の信頼区間に、目標値の12.5が含まれていないことがわかります（「ラインB」は目標値を上回っており、「ラインC」は下回っています）。この2つの製造ラインについては、重さが目標値と異なっている原因を調べ、対処する必要があります。

## 「一元配置」プラットフォームの統計的詳細

この節では、「一元配置」プラットフォームの統計的詳細を説明します。

- [「比較円の統計的詳細」](#)
- [「検出力の統計的詳細」](#)
- [「平均分析の統計的詳細」](#)
- [「「あてはめの要約」レポートの統計的詳細」](#)
- [「等分散性検定の統計的詳細」](#)
- [「ノンパラメトリックな検定の統計量の統計的詳細」](#)
- [「\[ロバストなあてはめ\] の統計的詳細」](#)

### 比較円の統計的詳細

「一元配置」プラットフォームの比較円は、多重比較検定の**最小有意差**（LSD）をグラフィカルに表したものです。最小有意差は、確率の分位点に2つの平均の差の標準誤差を掛けたものです。[ペアごと] オプションでは、FisherのLSDが使用され、確率の分位点はStudentの $t$ 統計量です。ここでは、[ペアごと] オプションを使ったときの比較円の計算方法をご紹介します。LSDは次のように定義されます。

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

2つの独立した平均の差の標準誤差は、次の関係から計算されます。

$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{Std}(\hat{\mu}_1)]^2 + [\text{Std}(\hat{\mu}_2)]^2$$

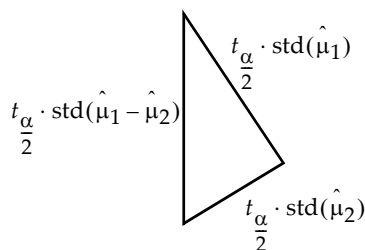
平均の間に相関がない場合、統計量の関係は次のようになります。

$$\text{LSD}^2 = [t_{\alpha/2} \text{Std}((\hat{\mu}_1 - \hat{\mu}_2))]^2 = [t_{\alpha/2} \text{Std}(\hat{\mu}_1)]^2 + [t_{\alpha/2} \text{Std}(\hat{\mu}_2)]^2$$

これらの平方値を図にすると、ピタゴラスの定理に基づく直角三角形になります（[図6.41](#)）。



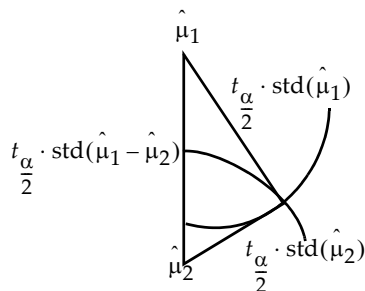
図6.41 2平均の差の関係



三角形の斜辺が、平均を比較する際の目安になります。観測された平均の差が、斜辺の長さ（最小有意差）より大きい場合に限り、平均は有意に異なります。

では、2つの平均がちょうど境界線上にあり、観測された平均の差が最小有意差に等しい場合にはどうなるでしょうか。縦軸に2つの平均を置き、それを頂点として三角形を描きます。また、それぞれの平均を中心として、それぞれの信頼区間の長さを直径とした円を描きます。

図6.42  $t$ 検定統計量の幾何学的関係



各円の半径は対応する三角形の辺の長さと等しく、 $t_{\alpha/2} \text{std}(\hat{\mu}_i)$  で表されます。

円は三角形の2辺と同じ角度で交わるため、次のような関係が成り立ちます。

- 平均の差が最小有意差と等しい場合、平均を中心とした信頼区間の円は直角に交わります。つまり、接線の角度が直角になります。

では、平均の差が最小有意差より大きい場合や、逆に小さい場合に、円がどのように交わるかを考えてみましょう。

- 円の外側の交角が直角より大きいとき、平均は有意には**異なりません**。円の外側の交角が直角より小さいとき、平均は有意に**異なります**。外側の交角が90度より小さいということは、平均の差が最小有意差より大きいことを意味します。
- 円が交わらない場合、平均は有意に異なり、円が入れ子になる場合、平均は有意には**異なりません** (図6.9)。

Studentの $t$ 分布の分位点ではなく、他の分位点を用いれば、いろいろな多重比較検定に対しても同様のグラフを描くことができます。

## 検出力の統計的詳細

「一元配置」プラットフォームでは、検出力の計算に非心  $F$  分布が使用されます。計算式は、次のとおりです (O'Brien and Lohr 1984)。

$$\text{Power} = \text{Prob}(F > F_{crit}, v_1, v_2, nc)$$

ここで

この計算式で得られる検出力は、非心度パラメータ  $nc$ 、分子自由度  $v_1$ 、分母自由度  $v_2$  の  $F$  分布に従う確率変数が、 $F_{crit} = F_{(1-\alpha, v_1, v_2)}$  よりも大きくなる確率を指します。この棄却値は、 $v_1$  および  $v_2$  の自由度を持つ  $F$  分布の累積確率  $1 - \alpha$  における分位点です。

$v_1 = r - 1$  は、分子自由度。

$v_2 = r(n - 1)$  は、分母自由度。

$n$  は、各グループの標本サイズ。

$r$  は、グループの個数。

$nc = n(CSS)/\sigma^2$  は、非心度パラメータ。

$$CSS = \sum_{g=1}^r (\mu_g - \mu)^2 \text{ は、仮説の平方和。}$$

$\mu_g$  は、 $g$  番目のグループの平均。

$\mu$  は、全体平均。

$\sigma^2$  は、誤差の分散。

## 平均分析の統計的詳細

### 順位変換

$n$  個の観測値があるとします。順位変換平均分析で使われるスコアは、次のように計算されます。

- まず、すべての観測値に小さい値から順番に順位を与えます。この時、同順位も考慮します。同順位のものには、平均順位を割り当てます。
- 順位を  $R_1, R_2, \dots, R_n$  とします。
- $i$  番目の観測値に対しての変換は、次のように求められます。

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

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

## 「あてはめの要約」レポートの統計的詳細

この節では、「一元配置」プラットフォームの「あてはめの要約」レポートについて詳しく解説します。

### R2 乗

連続尺度の応答のあてはめでは、 $R^2$  は、モデルの「分散分析」レポートの統計量を使用して、次のように計算されます。

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

### 自由度調整 R2 乗

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

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

ここで、「誤差」の平均平方は「分散分析」レポートにある値です。分母の平均平方は、「全体 (修正済み)」の平方和を、対応する自由度で割ったものです。[「分散分析」](#)を参照してください。

## 等分散性検定の統計的詳細

この節では、「一元配置」プラットフォームの「等分散性の検定」レポートについて詳しく解説します。

### F 値

O'Brien の検定では、元の応答変数を変換して、新しい応答変数を作成し、それに対して一元配置分散分析を行います。変換後の応答変数は、そのグループ平均が、元の応答変数のグループごとの不偏分散と等しくなるように作成されます。この変換後の応答変数は、次式で計算されます。

$$r_{ijk} = \frac{(n_{ij} - 1.5)n_{ij}(y_{ijk} - \bar{y}_{ij})^2 - 0.5s_{ij}^2(n_{ij} - 1)}{(n_{ij} - 1)(n_{ij} - 2)}$$

$n$  は、観測値  $y_{ijk}$  の個数を表します。

Brown-Forsythe 検定は、 $z_{ij} = |y_{ij} - \tilde{y}_i|$  に対する一元配置分散分析の  $F$  検定です。ここで、 $\tilde{y}_i$  は  $i$  番目の水準における  $Y$  の中央値です。

Leven 検定は、 $z_{ij} = |y_{ij} - \bar{y}_i|$  に対する一元配置分散分析の  $F$  検定です。ここで、 $\bar{y}_i$  は  $i$  番目の水準における  $Y$  の平均です。

Bartlett 検定は、次式により計算されます。

$$T = \frac{v \log \left( \sum_i \frac{v_i}{v} s_i^2 \right) - \sum_i v_i \log(s_i^2)}{1 + \left( \frac{\sum_i \frac{1}{v_i} - \frac{1}{v}}{3(k-1)} \right)} \quad \text{この式で、} v_i = n_i - 1 \text{ および } v = \sum_i v_i$$

$n_i$  は  $i$  番目の水準における度数、 $s_i^2$  は  $i$  番目の水準における  $Y$  の不偏分散です。この Bartlett 検定の検定統計量は、カイ 2 乗分布に従います。表中の Bartlett 検定の  $F$  値に出力されている値は、この Bartlett のカイ 2 乗検定統計量を自由度で割ったものです。

### Welch の検定の $F$ 値

Welch の検定の  $F$  値は、次の式で計算されます。

$$F = \frac{\left[ \frac{\sum_i w_i (\bar{y}_i - \bar{y}_{..})^2}{k-1} \right]}{\left\{ 1 + \frac{2(k-2)}{k^2-1} \left[ \sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i-1} \right] \right\}} \quad \text{この式で } w_i = \frac{n_i}{2}, u = \sum_i w_i, \bar{y}_{..} = \sum_i \frac{w_i \bar{y}_i}{u}$$

$n_i$  は  $i$  番目の水準における度数、 $\bar{y}_i$  は  $i$  番目の水準における  $Y$  の平均、 $s_i^2$  は  $i$  番目の水準における  $Y$  の不偏分散です。

### Welch 検定の分母自由度

Welch 検定の分母自由度は、次の近似式によって計算されます。

$$df = \frac{1}{\left( \frac{3}{k^2-1} \right) \left[ \sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i-1} \right]}$$

ここで、 $w_i$ 、 $n_i$ 、 $u$  は  $F$  値の式と同じ定義です。

## ノンパラメトリックな検定の統計量の統計的詳細

この節では、「一元配置」プラットフォームの Wilcoxon 検定、メディアン検定、Van der Waerden 検定、Friedman の順位検定の統計量を求める式を紹介します。

## 表記

これらの順位検定はスコアに基づいています。以下の説明では、次の表記を使用するものとします。

$j = 1, \dots, n$  標本全体の観測値に対する通し番号。

$i = 1, \dots, k$   $X$ の水準。 $k$ は水準の総数。

$n_1, n_2, \dots, n_k$   $X$ の第 $k$ 水準における標本サイズ。

$R_j$   $j$ 番目の観測値の中間順位。**中間順位**は、同順位がない場合には、通常の順位です。同順位があった場合は、それらの同順位のもの順位の平均（中間値）です。

$\alpha$  中間順位を変換する関数。各検定でスコアを定義するために使用します。

起動ウィンドウでブロック変数を指定した場合、次の表記が使用されます。

$b = 1, \dots, B$  ブロック変数の水準。 $B$ はブロックの総数。

$R_{bi}$  ブロック $b$ 内における、第 $i$ 観測値の中間順位。

関数 $\alpha$ は、以下のようにスコアを定義します。

### Wilcoxonのスコア

$$\alpha(R_j) = R_j$$

### メディアンのスコア

$$\alpha(R_j) = \begin{cases} 1 & R_j > \text{中央値} \text{ の場合} \\ 0 & R_j < \text{中央値} \text{ の場合} \\ t & R_j = \text{中央値} \text{ の場合} \end{cases}$$

メディアン（中央値）で同順位となる観測値の個数を $n_t$ とし、 $n_u$ をメディアンより大きい観測値の個数とすると、 $t$ は次のように求められます。

$$t = \frac{\text{floor}(n/2) - n_u}{n_t}$$

### Van der Waerdenのスコア

$$\alpha(R_j) = (R_j / (n + 1)) \text{ における標準正規分布の分位点}$$

### Friedmanの順位スコア

$$\alpha(R_{bi}) = R_{bi}$$

## 2標本の場合における正規近似

Xの水準が2水準の場合には、カイ2乗近似の統計量だけではなく、正規近似の統計量も計算されます。この節で使用される表記については、「表記」を参照してください。「2標本検定(正規近似)」レポートに表示される統計量の定義は、以下のとおりです。

**S** 統計量  $S$  は、観測値の個数が少ないほうの水準の  $\alpha(R_j)$  を合計した値です。両水準における観測値の個数が等しい場合は、値の順序で並べて後になる水準を使用します。

**Z**  $Z$  は次のように定義されます。

$$Z = (S - E(S)) / \sqrt{\text{Var}(S)}$$

---

**メモ:** X変数が2水準である場合、Wilcoxon検定だけ、0.5の連続修正が「正規近似」のほうでは行われます(「カイ2乗近似」のほうは0.5の連続修正を行っていません)。(S - E(S))がゼロより大きい場合は、分子から0.5を引きます。(S - E(S))がゼロより小さい場合は、分子に0.5を足します。

---

$E(S)$  帰無仮説のもとでの  $S$  の期待値。観測値の個数が少ないほうの水準、または、両水準における観測値の個数が等しい場合は値の順序で並べて後になる水準の観測値の個数を、 $n_l$  で表します。

$$E(S) = \frac{n_l}{n} \sum_{j=1}^n \alpha(R_j)$$

$\text{Var}(S)$  全観測値の平均スコアを  $ave$  とすると、 $S$  の分散は次のように定義されます。

$$\text{Var}(S) = \frac{n_1 n_2}{n(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

### Friedmanの順位検定における、2標本の場合の正規近似

Friedmanの順位検定では、2標本の場合の正規近似の計算方法は、 $S$  の分散を除き上記と同じです。 $S$  の分散は、次のような式で計算されます。

$$\text{Var}(S) = \frac{1}{k(k-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

## 一元配置のカイ2乗近似

---

**メモ:** 通常の平均順位スコアに基づく順位検定は、2群の時は「Wilcoxon検定」や「Mann-Whitney検定」と呼ばれ、3群以上の場合には「Kruskal-Wallis検定」と呼ばれています。

---

この節で使用される表記については、「表記」を参照してください。カイ2乗検定の統計量の計算には、次の値が使用されます。

$T_i$   $X$ の $i$ 番目の水準のスコアの合計。

$E(T_i)$   $i$ 番目の水準の合計スコアの期待値。水準間に差はないという帰無仮説のもとで、次のように定義されます。

$$E(T_i) = \frac{n_i}{n} \sum_{j=1}^n \alpha(R_j)$$

$Var(T)$  全観測値の平均スコアを $ave$ とすると、 $T$ の分散は次のように定義されます。

$$Var(T) = \frac{1}{(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

検定統計量は次の式で求められます。この検定統計量は、自由度が $k-1$ の漸近的なカイ2乗分布に従います。

$$C = \left( \sum_{i=1}^k (T_i - E(T_i))^2 / n_i \right) / Var(T)$$

### Friedmanの順位検定における、一元配置のカイ2乗近似

Friedmanの順位検定におけるカイ2乗検定の統計量は、次のような式で計算されます。

$$C = \frac{\sum_{i=1}^k (T_i - E(T_i))^2 / n_i}{\frac{1}{(k-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2 / n_i}$$

## 「ロバストなあてはめ」の統計的詳細

「一元配置」プラットフォームの「ロバストなあてはめ」オプションは、HuberのM推定を行います。HuberのM推定は、外れ値に対してロバスト（頑健）な推定方法であり、応答変数における外れ値が推定結果に与える影響が小さいです。HuberのM推定では、次式によって定義されているHuberの損失関数を最小にするパラメータ推定値を求めます。

$$l(e) = \sum_i \rho(e_i)$$

この式で、

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & |e| < k \text{ の場合} \\ k|e| - \frac{1}{2}k^2 & |e| \geq k \text{ の場合} \end{cases}$$

$e_i$ は残差を表します

Huberの損失関数は、誤差が小さい場合は2次曲線として増加し、誤差が大きい場合は線形に増加します。ロバスト推定（頑健推定）の詳細については、Huber（1973）およびHuber and Ronchetti（2009）を参照してください。「[\[ロバストなあてはめ\] オプションの例](#)」を参照してください。



# 第7章

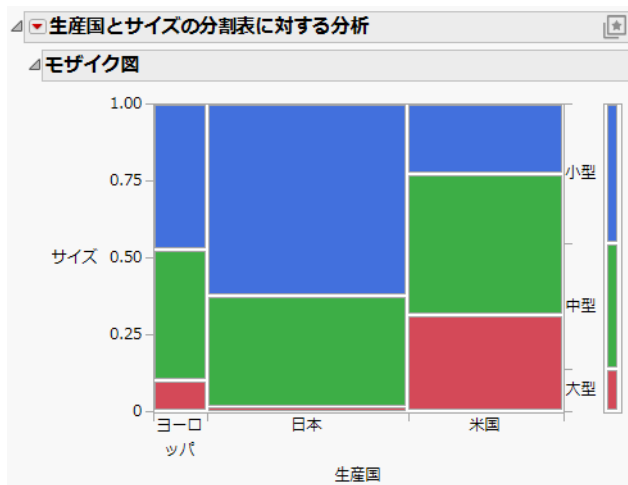
## 分割表分析

### 2つのカテゴリカルな変数の関係を調べる

2つのカテゴリカルな変数の関係を調べるには、「分割表」プラットフォームを用いてください。カテゴリカル変数は、順序尺度と名義尺度のどちらでも使えます。分割表分析の結果として、モザイク図、度数および割合の分割表、有意性のカイ2乗検定などが表示されます。また、割合の平均分析、対応分析、関連の指標の計算なども行えます。

分割表分析は、「二変量の関係」プラットフォームにおいて、XおよびYが、**両方ともカテゴリカル変数**のときに行われる分析です。

図7.1 分割表分析の例



## 目次

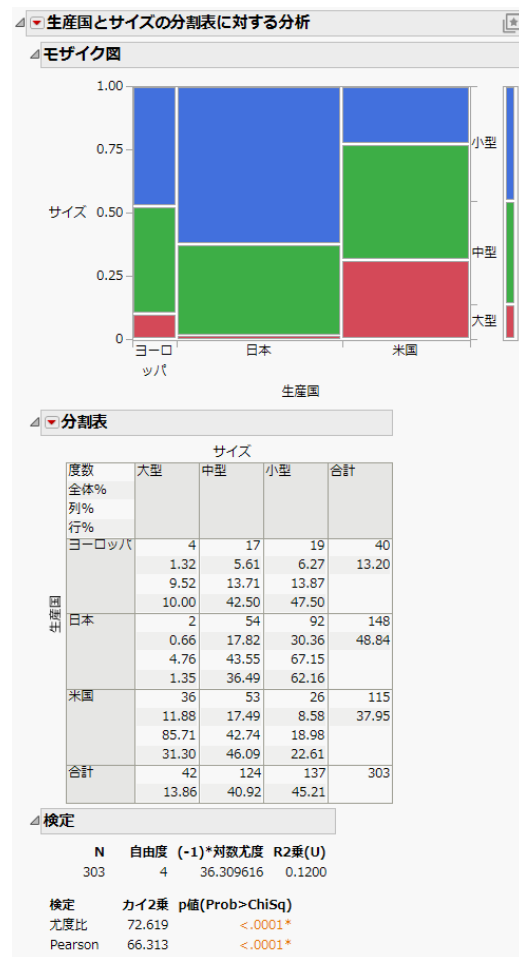
分割表分析の例	235
「分割表」プラットフォームの起動	236
データ形式	237
「分割表」レポート	237
モザイク図	239
分割表	241
「検定」レポート	242
「分割表」プラットフォームのオプション	243
「分割表分析」レポート	246
「割合の平均分析」レポート	246
「対応分析」レポート	247
Cochran-Mantel-Haenszel 検定	248
[一致性の統計量] オプション	249
「相対リスク」レポート	249
割合の2標本検定	250
「関連の指標」レポート	250
Cochran-Armitage の傾向検定	252
「Fisher の正確検定」レポート	252
同等性検定のレポート	252
「分割表」プラットフォームの別例	254
割合の平均分析の例	255
対応分析の例	256
Cochran-Mantel-Haenszel 検定の例	257
[一致性の統計量] オプションの例	259
[相対リスク] オプションの例	260
割合の2標本検定の例	261
[関連の指標] オプションの例	262
Cochran-Armitage の傾向検定の例	263
相対リスクの同等性検定の例	264
「分割表」プラットフォームの統計的詳細	265
一致性の統計量の統計的詳細	265
オッズ比の統計的詳細	266
「検定」レポートの統計的詳細	266
対応分析の統計的詳細	267

## 分割表分析の例

2つのカテゴリカルな変数の関係性を調べるには、「分割表」プラットフォームを用いてください。この例では、所有する車についてのアンケート結果を使用します。このデータには、調査対象者の属性（「性別」、「既婚/未婚」、「年齢」と、調査対象者が所有する車の属性（「生産国」・「サイズ」・「タイプ」）の情報が含まれています。ここでは、車のサイズ（小型・中型・大型）と生産国の関係性について調べます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「サイズ」を選択し、[Y, 目的変数] をクリックします。
4. 「生産国」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。

図 7.2 分割表分析の例



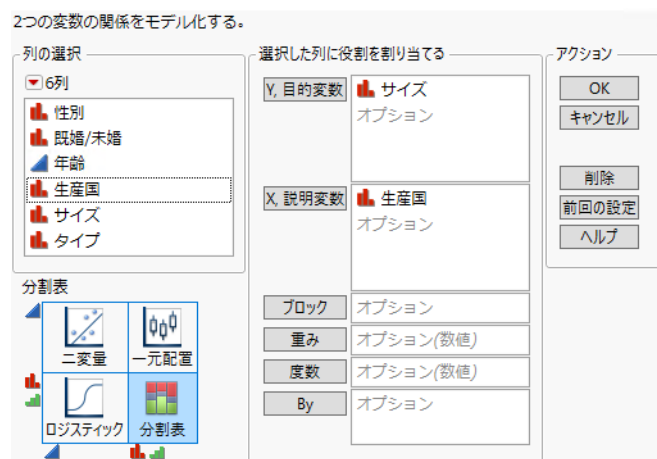
モザイク図と凡例から、次のことがわかります。

- 日本車において、大型車に分類されるものはごくわずかです。
- ヨーロッパ車の大半は、小型車か中型車に分類されます。
- 米国車の大半は、大型車か中型車に分類されます。

## 「分割表」プラットフォームの起動

「分割表」プラットフォームを起動するには、[分析] > [二変量の関係] を選択します。「二変量の関係」起動ウィンドウでは、4種類の分析が可能です。順序尺度または名義尺度のY変数と、順序尺度または名義尺度のX変数を入力すると、「分割表」プラットフォームが開きます。

図7.3 「分割表」起動ウィンドウ



「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。「分割表」起動ウィンドウには、次のようなオプションがあります。

**Y, 目的変数** 分析したい目的変数（応答変数）。順序尺度か名義尺度でなければなりません。

**X, 説明変数** 分析したい説明変数（予測変数）。順序尺度か名義尺度でなければなりません。

**ブロック** ブロック変数とする列。指定されたブロック変数を層別因子として、Cochran-Mantel-Haenszel 検定が行われます。分類変数を指定した分析は、層別解析とも呼ばれます。

**重み** データテーブルの各行に対する重みを含んだ列。値が0より大きい行だけが分析に用いられます。

**度数** 分析において各データ行に度数を割り当てます。分析対象のデータが要約されたものであるときに役立ちます。

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

## データ形式

「分割表」プラットフォームでは、要約されていないデータ、および要約されているデータのいずれも分析できます。

**要約されていないデータ** 観測値ごとに1行ずつあり、Xの値が1列、Yの値が1列に入力されています。

**要約されたデータ** XとYの値の組み合わせが同じである複数の観測値が1行にまとめられているデータです。データテーブルには、各行の観測値の数を示す度数列が必要です。起動ウィンドウで、この列を「度数」に指定してください。

集計されたカテゴリカルデータの例については、「[割合の平均分析の例](#)」を参照してください。

---

**メモ：**「二変量の関係」起動ウィンドウでは、連続尺度、順序尺度、名義尺度の列が使用できます。[Y, 目的変数]の列と[X, 説明変数]の列がどちらも順序尺度か名義尺度である場合、「分割表」プラットフォームが起動します。それ以外の組み合わせでは、「二変量の関係」・「一元配置」・「ロジスティック」のいずれかのプラットフォームが起動します。

---

---

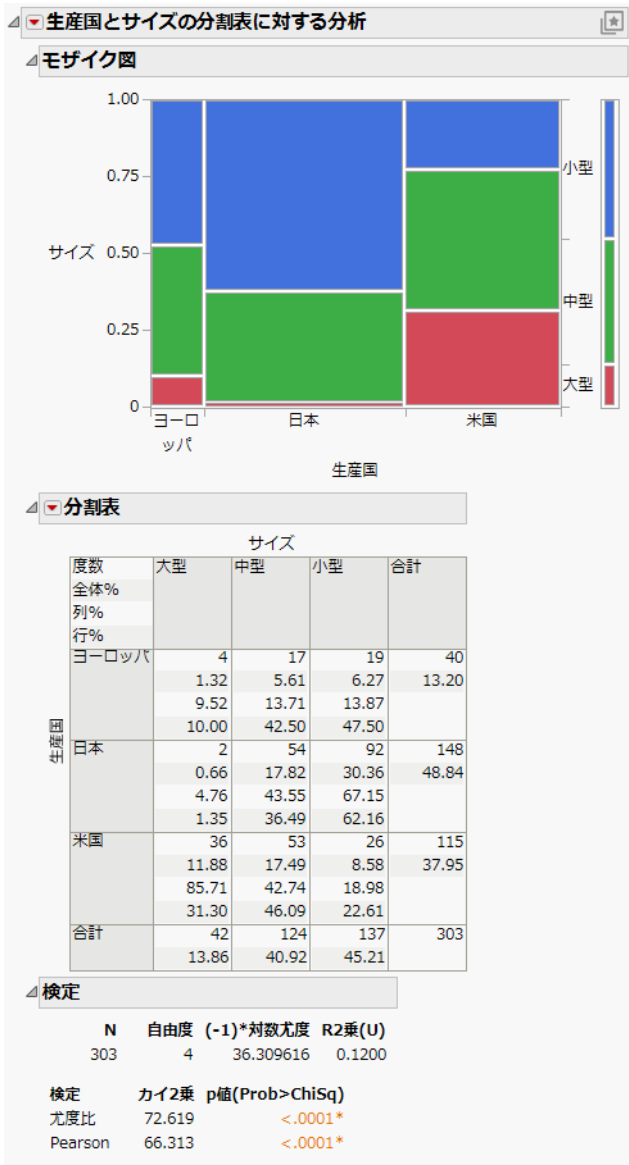
## 「分割表」レポート

開いたばかりの「分割表」レポートには、モザイク図、分割表、そして応答変数（Y）の水準が因子変数（X）の水準間で異なるかどうかを調べる検定が表示されます。赤い三角ボタンのオプションを選択することにより、いろいろな分析や検定を追加していただけます。「[「分割表」プラットフォームのオプション](#)」を参照してください。

この節では、「分割表」レポートの以下のセクションをご紹介します。

- [「モザイク図」](#)
- [「分割表」](#)
- [「検定」レポート](#)

図 7.4 「分割表」レポートの例



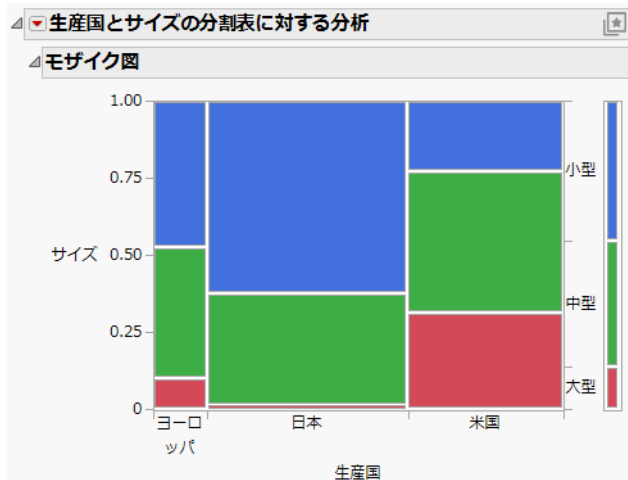
変数を対話的に置換する

プロット内で変数を対話的に置き換えるには、一方の軸の変数をドラッグしてもう一方の軸にドロップします。データテーブルの「列」パネルで該当する変数をクリックし、プロットの軸上までドラッグして変数を置換することもできます。

## モザイク図

「分割表」プラットフォームの「モザイク図」は、2元度数表（分割表）をグラフにしたものです。モザイク図は大きさの異なる長方形に分割したプロットで、それぞれの長方形における縦の長さが、X変数の各水準内における、Y変数の各水準の割合を表します。モザイク図は1981年にHartiganとKleinerによって考案され、Friendly（1994）が改良を加えました。

図7.5 モザイク図の例



このモザイク図で、次の点に注目してください。

- 横軸における長方形の幅は、X変数の各水準の観測数（オブザベーション数）を示します。
- プロットの右側に表示されている縦長のプロットは、X変数の水準をすべて合わせた場合にY変数の各水準が占める割合を示します。これらの割合は、X変数とY変数の間に関連がないとする帰無仮説を表します。
- 左側の縦軸のスケールには応答確率が表示され、軸全体は1の確率、つまり標本全体を表します。

**ヒント:** モザイク図の長方形をクリックすると、その領域が強調表示され、関連するデータテーブルで対応する行が選択されます。プロットにラベルを付けるには、長方形を右クリックします。

## モザイク図のポップアップメニュー

「分割表」レポートのモザイク図上を右クリックして、色を変更したり、セルのラベルを指定したりすることができます。図の各長方形に個別にメニューが用意されています。

**色の設定** 水準に割り当てられている色を確認し、変更することができます。「色の設定」を参照してください。

**セルのラベル** モザイク図の各セルに表示するラベルの種類を指定します。

**ラベルなし** ラベルを表示しません。ラベルが付いていた場合は、それが表示されなくなります。

**度数ラベル** 標本サイズ（観測値の度数）を各セルに表示します。

**パーセントラベル** 応答水準の割合（確率）を各セルに表示します。

**値ラベル** Y変数の水準を各セルに表示します。

**行ラベル** セルが表しているすべての行の行番号を表示します。

**線の色** 各セルを囲む線の色を指定します。

**線種** 各セルを囲む線の種類を指定します。

**線の幅** 各セルを囲む線の幅を指定します。

**透明度** セルの色の透明度を指定します。

---

**メモ:** ポップアップメニューのその他のオプションについては、『JMPの使用法』を参照してください。

---

## 色の設定

モザイク図のポップアップメニューで「色の設定...」オプションを選択すると、「値の色を選択」ウィンドウが開き、水準に割り当てられている色が表示されます。

デフォルトのモザイク図の色は、Y応答列が順序尺度か名義尺度か、「値の色」列プロパティが設定されているかどうかによって異なります。水準の色を変更するには、2列目の色の楕円形をクリックし、新しい色を選択します。

「値の色を選択」ウィンドウには、以下のオプションがあります。

**マクロ** 次のいずれかの方法で色を変更します。

**両端の間でグラデーション** 変数のすべての水準に色のグラデーションが適用されます。

**選択した点の間でグラデーション** 選択した水準の範囲にだけ色のグラデーションが適用されます。水準の範囲を選択するには、水準の範囲をドラッグするか、またはShiftキーを押しながら最初と最後の水準をクリックします。

**色の逆転** 色の順序を逆転させます。

**元の色に戻す** 「値の色を選択」ウィンドウで行った変更をすべて破棄します。

**カラーテーマ** カラーテーマに従って各値の色を変更します。

**列の色を保存** 変更後の色をデータテーブルに保存します。カラーテーマを変更し、このチェックボックスをオンにすると、関連するデータテーブルの列に「値の色」列プロパティが追加されます。データテーブルからこのプロパティを編集する場合は、**[列] > [列情報]** を選択します。



## 分割表

「分割表」は、 $r$ 行 $\times$  $c$ 列の2元度数表です。 $X$ 変数の水準と同数( $r$ )の行、 $Y$ 変数の水準と同数( $c$ )の列ができます。

図7.6 分割表の例

分割表				
	度数	サイズ		
		大型	中型	小型
全体%				
列%				
行%				
国産車	ヨーロッパ	4	17	19
		1.32	5.61	6.27
		9.52	13.71	13.87
		10.00	42.50	47.50
日本		2	54	92
		0.66	17.82	30.36
		4.76	43.55	67.15
		1.35	36.49	62.16
米国		36	53	26
		11.88	17.49	8.58
		85.71	42.74	18.98
		31.30	46.09	22.61
合計		42	124	137
		13.86	40.92	45.21

この分割表について、次の点を確認してください。

- この分割表において、「度数」・「全体%」・「列%」・「行%」は、行と列が変わる各セル内の数値を指しています。
- 最後の列には、各行の合計度数と割合が表示されています。
- 最下行には、各列の合計度数と割合が表示されています。

図7.6で、「米国」製の「大型」車に着目してください。次の表で、車の分割表から導き出せる結論について説明します。

表7.1 例の分割表による結果

数	説明	表のラベル
36	米国製大型車の台数	度数
11.88%	全車両303台のうち米国製大型車が占める割合 (36/303)	全体%
85.71%	大型車42台のうち米国車が占める割合 (36/42)	列%
31.30%	米国車115台のうち大型車が占める割合 (36/115)	行%
37.95%	全車両303台のうち米国車が占める割合 (115/303)	全体%
13.86%	全車両303台のうち大型車が占める割合 (42/303)	全体%

**ヒント:**

- 「分割表」で統計量の表示／非表示を切り替えるには、「分割表」の赤い三角ボタンをクリックし、該当する統計量を選択します。
- 「分割表」の統計量の表示形式を変更するには、表内で該当する値をクリックし、[表示形式...]を選択します。

**分割表に表示される統計量**

「分割表」プラットフォームの「分割表」には、以下の統計量が含まれます。

**度数** セル度数・周辺度数（行和や列和）・総度数。総度数は、全体の標本サイズとも呼ばれます。

**全体%** セル度数や周辺度数の、総度数に占める割合。

**行%** 各セル度数の、行和に占める割合。

**列%** 各セル度数の、列和に占める割合。

**期待値** 行と列が独立していると仮定して計算された各セルの期待度数 ( $E$ )。対応する行合計と列合計の積を全体合計で割った値です。

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

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

**累積列度数** 表の上から順に計算した、列和の累積。

**累積列%** 表の上から順に計算した、列和の累積割合。

**累積行度数** 表の左から順に計算した、行和の累積。

**累積行%** 表の左から順に計算した、行和の累積割合。

**データテーブルに出力** 分割表に含まれている統計量を新しいデータテーブルに表示します。

**表示形式** 分割表において、統計量の書式を変更するためのウィンドウを開く。

**「検定」レポート**

「分割表」プラットフォームの「検定」レポートには、Y変数の水準の割合がX変数の水準に依存せず独立しているかどうかを評価する2つの検定が表示されます。

「検定」レポートには、以下の統計量が含まれます。

**N** 観測（オブザベーション）の合計個数。

**自由度** 検定に関係する自由度。自由度は  $(c - 1)(r - 1)$  に等しくなります。 $c$  は列数、 $r$  は行数です。

**(-1)\* 対数尤度** 対数尤度の符号を逆にしたもの。適合度や不確定性を表す指標です（連続尺度の応答変数における平方和と同じような指標です）。『基本的な回帰モデル』を参照してください。

**R2 乗 (U)** 全体の不確定性のうち、モデルによって説明される割合。

- $R^2$  が 1 ということは、X 変数の水準から Y 変数の水準を完璧に予測できることを意味しています。
- $R^2$  が 0 ということは、「標本全体で応答の割合が同じである」という状態に比べて、モデルを使っても予測が向上しないことを意味しています。

**検定** 「すべての標本グループにおいて、応答の割合（確率）は等しい」という帰無仮説を検定する 2 種類のカイ 2 乗検定。

**カイ 2 乗** カイ 2 乗検定の検定統計量。

**p 値 (Prob>ChiSq)** 「X 変数と Y 変数の間に関係が存在しない」という仮定のもとで、現在のカイ 2 乗値より大きなカイ 2 乗値を得る確率。両方の変数が 2 水準しかない場合は、Fisher の正確検定も実行されます（両側検定と片側検定の  $p$  値が表示されます）。

このレポートの統計量について詳しくは「[「検定」レポートの統計的詳細](#)」を参照してください。

### Fisher の正確検定

このレポートには、 $2 \times 2$  分割表に対する Fisher の正確検定の結果が表示されます。 $2 \times 2$  分割表の場合は自動的に検定結果が表示されます。 $2 \times 2$  および  $r \times c$  分割表に対する Fisher の正確検定の詳細については、「[「Fisher の正確検定」レポート](#)」を参照してください。

---

## 「分割表」プラットフォームのオプション

「分割表分析」の赤い三角ボタンのメニューには、その他の分析を行うオプションが含まれています。

---

**メモ:** Y 変数または X 変数を複数指定した場合は、「あてはめのグループ」メニューが表示されます。「あてはめのグループ」メニューのオプションを使うと、レポートの配置を変更したり、適合度の順に並べたりできます。『基本的な回帰モデル』を参照してください。

---

「分割表分析」の赤い三角ボタンのメニューには、以下のオプションがあります。

**モザイク図** モザイク図の表示／非表示を切り替えます。モザイク図は、分割表をグラフに表したものです。「[モザイク図](#)」を参照してください。

**分割表**  $r$  行  $c$  列の 2 元度数表の表示／非表示を切り替えます。X 変数の水準と同数 ( $r$ ) の行、X 変数の水準と同数 ( $c$ ) の列ができます。「[分割表](#)」を参照してください。

**検定** X 変数の水準間で応答割合が同じかどうかの検定の表示／非表示を切り替えます。これらの検定は、連続尺度の応答に対する分散分析を、カテゴリーカルな応答にしたものと解釈できます。「[「検定」レポート](#)」を参照してください。

**$\alpha$ 水準の設定** 信頼区間の $\alpha$ 水準を変更します。サブメニューから、よく使われている有意水準（[0.10]・[0.05]・[0.01]）を選択するか、**【その他】**を選択して任意の有意水準を入力してください。

**割合の平均分析**（応答変数の水準数が2つの場合にのみ使用できます。）グループごとの割合を比較するための、割合の平均分析（ANOMP）の表示／非表示を切り替えます。ANOMPは、X変数の水準ごとの応答の割合を、割合の全体平均と比較します。**「割合の平均分析」レポート**を参照してください。

**対応分析** 対応分析の表示／非表示を切り替えます。対応分析は、度数表の行で行割合のパターンが類似しているもの、および、度数表の列で列割合のパターンが類似しているものが近くになるようにプロットします。対応分析のグラフには、分割表の各行と各列につき1つの点がプロットされます。**「対応分析」レポート**を参照してください。

**Cochran-Mantel-Haenszel検定** Cochran-Mantel-Haenszel検定の表示／非表示を切り替えます。

Cochran-Mantel-Haenszel検定は、3つ目の分類変数をブロック変数として、2つの分類変数の間に関係があるかどうかを判断するための検定です。このオプションを選択すると、ブロック変数を指定するウィンドウが開きます。**「Cochran-Mantel-Haenszel検定」**を参照してください。

**一致性の統計量**（XとYにおける水準がまったく同じ場合にのみ使用できます。）水準間の一致性に関するレポートの表示／非表示を切り替えます。レポートには、 $\kappa$ 統計量（Agresti 1990）、標準誤差、信頼区間、一致性に対する仮説検定が表示されます。また、Bowkerの対称性検定も表示されます。なお、2×2表に対するBowker検定は、「McNemar検定」と呼ばれています**「[一致性の統計量] オプション」**を参照してください。

**相対リスク**（X変数とY変数の水準数がいずれも2水準である場合にのみ使用できます。）相対リスク（リスク比）の表示／非表示を切り替えます。**「相対リスク」レポート**を参照してください。

「相対リスク」レポートには、相対リスク（リスク比）に対する信頼区間も表示されます。**【 $\alpha$ 水準の設定】オプション**を使用して $\alpha$ 水準を変更できます。

**リスク差**（X変数とY変数の水準数がいずれも2水準である場合にのみ使用できます。）リスク差（割合の差）の表示／非表示を切り替えます。

「リスク差」レポートには、リスク差（割合の差）に対する信頼区間も表示されます。**【 $\alpha$ 水準の設定】オプション**を使用して $\alpha$ 水準を変更できます。

**オッズ比**（X変数とY変数の水準数がいずれも2水準である場合にのみ使用できます。）オッズ比の表示／非表示を切り替えます。**「オッズ比の統計的詳細」**を参照してください。

「オッズ比」レポートには、オッズ比に対する信頼区間も表示されます。**【 $\alpha$ 水準の設定】オプション**を使用して $\alpha$ 水準を変更できます。

**割合の2標本検定**（X変数とY変数の水準数がいずれも2水準である場合にのみ使用できます。）割合に対する2標本検定の表示／非表示を切り替えます。X変数の2水準間で、Y変数の割合が異なるかどうかを比較します。**「割合の2標本検定」**を参照してください。

**関連の指標** 分割表における変数間の関連性指標の表示／非表示を切り替えます。**「関連の指標」レポート**を参照してください。

**Jonckheere-Terpstra 検定** Jonckheere-Terpstra 検定の表示／非表示を切り替えます。この検定は、比較したい群に対して予め順序が想定できる状況での群間比較の傾向検定です。薬剤の投与量など、比較したい群にカテゴリカルな順序が想定できる場合に適した検定です。対立仮説は、「比較する群に順序性がある」です。

**Cochran-Armitage の傾向検定** (片方の変数が2水準の変数で、もう一方が順序変数の場合にのみ使用できます。) 二項割合に対する Cochran-Armitage 傾向検定の表示／非表示を切り替えます。Cochran-Armitage 傾向検定は、群に順序性がある場合に使われる検定です。[「Cochran-Armitage の傾向検定」](#)を参照してください。

**二項分布** 次の正確検定を実行するためのオプションが表示されます。

**Fisher の正確検定** Fisher の正確検定の表示／非表示を切り替えます。Fisher の正確検定は、2つのカテゴリカル変数間の関連を調べる検定です。この検定は、大標本近似に基づくものではなく、周辺和を固定したもとの並び替えに基づく、条件付き正確検定です。[「Fisher の正確検定」レポート](#)を参照してください。

**Cochran-Armitage の傾向検定** (いずれかの変数が2水準である場合にのみ使用できます。) 正確な Cochran-Armitage 傾向検定の表示／非表示を切り替えます。[「Cochran-Armitage の傾向検定」](#)を参照してください。

**一致性の統計量** (いずれかの変数が2水準である場合にのみ使用できます。)  $\kappa$  統計量に対する正確検定の表示／非表示を切り替えます。[「一致性の統計量」オプション](#)を参照してください。

---

**メモ:** 「度数」変数に整数でない値がある場合、正確検定は計算できません。また、全体の標本サイズが 32767 より大きく、かつ、分割表が  $2 \times 2$  より大きい場合も、正確検定のオプションは使用できません。

---

**実質的な同等性の検定** (X 変数と Y 変数の水準数がいずれも2水準である場合にのみ使用できます。) リスク差もしくは相対リスク (リスク比) に対する同等性検定・優越性検定・非劣性検定を実行するための以下のオプションを含みます。[「同等性検定のレポート」](#)を参照してください。

**リスク差** リスク差に対する検定の設定ウィンドウが開きます。この設定ウィンドウにて、同等性検定、優越性検定、非劣性検定といった検定の種類を選択したり、マージンを指定したりできます。

**相対リスク** 相対リスク (割合の比) に対する検定の設定ウィンドウが開きます。この設定ウィンドウにて、同等性検定、優越性検定、非劣性検定といった検定の種類を選択したり、マージンを指定したりできます。

**表示オプション** モザイク図に変更を加えるための以下のオプションを含みます。

**横向きモザイク図** モザイク図の向きを水平方向または垂直方向に変更します。

**データテーブルに出力** レポートの表から、JMP データテーブルを作成します。

以下のオプションの詳細については、『JMP の使用法』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMPレポートの設定に合わせて環境設定を変更したりできます。

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

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

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの [スクリプトの索引] を開いてください。また、[スクリプトの索引] には、この節で紹介されているオプションのスクリプト例もあります。

---

## 「分割表分析」レポート

「分割表」プラットフォームでは、2つのカテゴリカル変数間の関係をグラフで調べることができます。この節では、分析を実行する各オプションで作成されるレポートの内容を詳しく説明します。

- 「[割合の平均分析](#)」レポート
- 「[対応分析](#)」レポート
- 「[Cochran-Mantel-Haenszel 検定](#)」
- 「[\[一貫性の統計量\] オプション](#)」
- 「[相対リスク](#)」レポート
- 「[割合の2標本検定](#)」
- 「[関連の指標](#)」レポート
- 「[Cochran-Armitage の傾向検定](#)」
- 「[Fisher の正確検定](#)」レポート
- 「[同等性検定のレポート](#)」

### 「割合の平均分析」レポート

「分割表」プラットフォームの割合の平均分析 (ANOMP; analysis of means for proportions) は、グループごとの割合が、割合の全体平均と異なるかどうかを調べる多重比較方法です。平均分析の手法について詳しくは、Nelson et al. (2005) を参照してください。「[割合の平均分析の例](#)」も参照してください。

Y変数が2水準である場合、このオプションによって、X変数の水準ごとの応答の割合を、それらの割合の全体平均と比較できます。この手法では、二項分布の正規近似が使用されます。標本サイズが小さすぎる場合は、警告が表示されます。

「割合の平均分析」の赤い三角ボタンのメニューには、以下のオプションがあります。

**有意水準の設定** 「割合の平均分析」チャートの決定限界の計算に使われる有意水準を指定します。

**要約レポートの表示** X変数の各水準における応答割合と決定限界を含むレポートの表示／非表示を切り替えます。レポートには、決定限界を超えているかどうか也表示されます。

**応答水準の切り替え** 応答変数のカテゴリで、分析において注目するカテゴリを変更します。

**表示オプション** 以下のオプションを使って「割合の平均分析」チャートに変更を加えることができます。

**決定限界の表示** 割合に対する平均分析のグラフにおいて、決定限界を示す線の表示／非表示を切り替えます。

**決定限界の陰影の表示** 割合に対する平均分析のグラフにおいて、決定限界に対する陰影の表示／非表示を切り替えます。

**中心線の表示** 割合に対する平均分析のグラフにおいて、中心線の表示／非表示を切り替えます。

**点のオプション** 「割合の平均分析」チャートにおける点のスタイルを指定します。垂線、接続線、点のみから選択できます。デフォルトでは、全体平均の水平線からの垂線でグラフが描かれます。

## 「対応分析」レポート

「分割表」プラットフォームの対応分析は、度数表の中から度数のパターンが似ている行や列を見つけるための手法です。対応分析のグラフでは、各行と各列につき1つの点がプロットされます。水準数が多く、モザイク図から有用な情報を引き出すのが難しい場合は、対応分析が役立ちます。[「対応分析の例」](#)を参照してください。

「対応分析」のグラフには、行と列のプロファイルが含まれます。行ごとの割合、つまり、行内にある度数をその行の合計度数で割ったものを、**行プロファイル**と言います。2つの行の行プロファイルが似ているとき、対応分析のグラフにおいて、それらの行を表す点は互いに近い位置にあります。プロットにおける2点間の平方距離は、「カイ2乗距離」という2行の等質性を示す検定統計量を近似したものです。

列プロファイルも、行プロファイルと同じように定義され、同じように解釈できます。行の点と列の点の距離には意味がありませんが、原点からの方向には意味があり、その関係を見ることでプロットを解釈できます。

「対応分析」の赤い三角ボタンのオプションを使用して、「分割表」レポートに三次元散布図を追加したり、データテーブルに列プロパティを追加したりすることができます。

**三次元対応分析** 対応分析の3次元プロットの表示／非表示を切り替えます。

**値の表示順序を保存** データテーブル内のX変数とY変数の列に、対応分析の第1次元目の順序を保存します。対応分析における第1次元目の座標の大きさで水準を並べたときの順序を、「値の順序」列プロパティに保存します。



## 「詳細」レポート

「分割表」プラットフォームの「詳細」レポートは、対応分析の統計情報をまとめたもので、プロットに使われている値をリストしています。

**特異値** 標準化した分割表を特異値分解したときの特異値。「[対応分析の統計的詳細](#)」を参照してください。

**慣性** 特異値を2乗した値。各次元でデータの変動がどれぐらい説明されるかを示します。

**割合** 全次元の慣性に対する、各次元の慣性が占める割合。

**累積** 慣性の累積割合。

---

**ヒント:** 最初の2次元によって全体の慣性のほとんどが捉えられているときは、2次元だけの対応分析プロットによって、分割表のばらつきが十分に表されていることを意味しています。

---

**X変数 c1、c2、c3** 対応分析プロットで使われている座標。

**Y変数 c1、c2、c3** 対応分析プロットで使われている座標。

## Cochran-Mantel-Haenszel 検定

「分割表」プラットフォームのCochran-Mantel-Haenszel 検定では、3つ目の分類変数をブロック変数（層別変数）としたときに、2つの分類変数の間に関係が見られるかどうかを調べます。3つ目の分類変数でブロックに分けることを、層別とも言います。

---

**メモ:** 「[Cochran-Mantel-Haenszel 検定の例](#)」を参照してください。

---

「Cochran-Mantel-Haenszel 検定」レポートには、各種検定の結果をまとめた表があります。検定ごとに、カイ2乗統計量、検定の自由度、有意確率（p 値 (Prob>ChiSq)）が表示されます。レポートには以下の検定が含まれます。

**スコアの相関**（YとXの両方が順序尺度か間隔尺度である場合のみに結果を解釈できます。YとXのいずれかが名義尺度の時には解釈が難しいです。）層別変数の少なくとも1つの層において、XとYの間に線形関係があるという対立仮説を検定します。

**X間でのスコア比較**（Yが順序尺度か間隔尺度である場合のみに結果を解釈できます。Yが名義尺度の時には解釈が難しいです。）層別変数の少なくとも1つの層において、ある列の行スコアの平均が、他の列の行スコアの平均と異なるという対立仮説を検定します。

**Y間でのスコア比較**（Xが順序尺度か間隔尺度である場合のみに結果を解釈できます。Xが名義尺度の時には解釈が難しいです。）層別変数の少なくとも1つの層において、ある列の行スコアの平均が、他の列の行スコアの平均と異なるという対立仮説を検定します。

**カテゴリーの一般連関** 層別変数の少なくとも1つの水準で、XとYの間に何らかの関連があることを検定します。



---

**ヒント:** 1つの値しか含まない列をデータテーブルに用意して、その列を層別変数に指定し、Cochran-Mantel-Haenszel 検定を行うことで、(層別変数がない場合における) 行スコアの平均に対する検定を行います。

---

## 【一貫性の統計量】 オプション

2つのカテゴリカル変数が同じ水準を持つ場合は、「分割表」プラットフォームの「一貫性の統計量」を使ってカッパ統計量を計算できます (Agresti 1990)。カッパ統計量の点推定値だけでなく、標準誤差・信頼区間・仮説検定も計算されます。「[【一貫性の統計量】 オプションの例](#)」を参照してください。

レポートに表示されるカッパ統計量とその  $p$  値は近似値です。「[一貫性の統計量の統計的詳細](#)」を参照してください。検定の  $p$  値として、正確な値も計算できます。「[Fisher の正確検定 レポート](#)」を参照してください。

「カッパ係数」レポートには、次の統計量が含まれます。

**カッパ** カッパ統計量。

**標準誤差** カッパ統計量の標準誤差。

**下側 95%** カッパ統計量に対する両側信頼区間の下限。

**上側 95%** カッパ統計量に対する両側信頼区間の上限。

**$p$  値 (Prob>Z)** カッパに対する片側漸近検定の有意確率 ( $p$  値)。「カッパ統計量が0である」という帰無仮説を検定します。

**$p$  値 (Prob>|Z|)** カッパに対する両側漸近検定の有意確率 ( $p$  値)。

Bowker 検定の「検定」レポートには、以下の統計量が含まれます。

**カイ2乗** 分割表の対称性に対する Bowker 検定の検定統計量。Bowker 検定の帰無仮説は、「分割表における割合が対称になっている (すべてのセルにおいて  $p_{ij}=p_{ji}$  である)」というものです。Y 変数と X 変数の水準数がどちらも 2 水準である場合、Bowker 検定は McNemar 検定と同じです。

**$p$  値 (Prob>ChiSq)** Bowker の検定統計量の有意確率 ( $p$  値)。Bowker 検定の帰無仮説は、「正方分割表において、確率が対称である」というものです。

## 「相対リスク」 レポート

「分割表」プラットフォームの「相対リスク」オプションは、 $2 \times 2$  の分割表のリスク比を計算します。レポートには信頼区間も表示されます。この手法について詳しくは、Agresti (1990, 3.4.2 節) を参照してください。「[【相対リスク】 オプションの例](#)」を参照してください。

「相対リスク」オプションを選択すると、「相対リスクカテゴリの選択」ウィンドウが開きます。1 対の応答と因子について、または、応答と因子のすべての組み合わせについて相対リスクを計算することができます。

---

**メモ:** 両方の変数に「コントロール群」と「イベントを示す水準」の列プロパティが指定されている場合、「相対リスクカテゴリの選択」ウィンドウは表示されません。

---

## 割合の2標本検定

「分割表」プラットフォームでは、信頼区間を計算し、2つの割合の差に対して仮説検定を行うことができます。この分析は、X変数とY変数の水準数がどちらも2であるときに可能です。[「割合の2標本検定の例」](#)を参照してください。

**説明** 検定の説明。

**割合の差** X変数の2水準間における割合の差。

**下側95%** 割合の差に対する両側信頼区間の下限。この信頼区間は、調整を伴ったWald法によって計算されています。

**上側95%** 割合の差に対する両側信頼区間の上限。この信頼区間は、調整を伴ったWald法によって計算されています。

**調整済みWald検定 (帰無仮説)** 片側検定と両側検定の帰無仮説の説明。

**確率** 検定の有意確率 ( $p$  値)。

---

**ヒント:** 表の下ラジオボタンを使って検定の対象となる応答水準を変更します。

---

## 「関連の指標」レポート

「分割表」プラットフォームで「関連の指標」オプションを選択すると、関連に関する統計量が計算されます。[「\[関連の指標\] オプションの例」](#)を参照してください。

「関連の指標」レポートには、次の統計量の点推定値・標準誤差・信頼区間が表示されます。

**ガンマ** 順序的な関係の指標。同順位のペアを無視したときの、一致するペアと一致しないペアの確率の差として計算されます。-1から1の値を取ります。

**Kendallのタウ-b** ガンマに似ていますが、同順位のペアの数に対して修正が行われます。-1から1の値を取ります。

**Stuartのタウ-c** ガンマに似ていますが、表のサイズによる調整、および、同順位のペアに対する修正が行われます。-1から1の値を取ります。

**SomersのD** Kendallのタウ-bに、非対称性の修正を加えた指標です。SomersのDは、同順位のペアに対する修正を、独立変数の同順位に対してだけ行います。-1から1の値を取ります。

- 「C|R」は、行変数Xを独立変数、列変数Yを従属変数とすることを示します。
- 同様に、「R|C」は、列変数Yを独立変数、行変数Xを従属変数とすることを示します。

---

**メモ:** ガンマ、Kendallのタウ-b、Stuartのタウ-c、SomersのDは、X変数の増加に伴い、Y変数も増加する傾向にあるかどうかを示す順序的な関係を示す指標です。これらの統計量は、順序関係の一致するペアと一致しないペアの個数から計算されます。データのペアにおいて、X値がより大きい方が、Y値もより大きくなっているペアを、「順序関係が一致するペア」と言います。逆に、X値が大きい方が、Y値が小さくなっているペアを「順序関係が一致しないペア」と言います。これらの指標は、両変数が順序尺度の場合だけに適しています。

---

**非対称ラムダ** C|RとR|Cで異なります。0から1の値を取ります。

- 「非対称ラムダ (C|R)」は、行変数Xで列変数Yを予測したときに、予測的中率がどれくらい向上するかを示します。
- 「非対称ラムダ (R|C)」は、列変数Yから行変数Xを予測したときに、予測的中率がどれくらい向上するかを示します。

**対称ラムダ** 2つの非対称ラムダ係数の平均のような指標です。0から1の値を取ります。

**不確実性係数** C|RとR|Cで異なります。0から1の値を取ります。

- 「不確実性係数 (C|R)」は、行変数Xによって説明される列変数Yの不確実性の割合を示します。
- 「不確実性係数 (R|C)」は、列変数Yによって説明される行変数Xの不確実性の割合を示します。

**不確実性係数 (対称)** 上記2つの不確実性係数を、対称性をもつように変えた指標です。0から1の値を取ります。

---

**メモ:** ラムダおよび不確実性係数は、順序尺度と名義尺度の変数に適しています。

---

関連の指標の計算方法については、SAS Institute Inc. (2023b) の「FREQ Procedure」章を参照してください。次の文献にも詳細情報が記載されています。

- Brown and Benedetti (1977)
- Goodman and Kruskal (1979)
- Kendall and Stuart (1979)
- Snedecor and Cochran (1980)
- Somers (1962)

## Cochran-Armitageの傾向検定

「分割表」プラットフォームにある Cochran-Armitage の傾向検定は、順序変数の水準間における二項割合の傾向を調べます。この検定は、片方の変数が2水準の変数で、もう一方が順序変数の場合にだけ適切な検定です。2水準の変数が応答変数、順序尺度の変数が説明変数として扱われます。帰無仮説は「割合の傾向がない」、つまり、「説明変数の全水準において二項割合が同じ」というものです。「[Cochran-Armitageの傾向検定の例](#)」を参照してください。

---

**メモ:** この検定の検定統計量と有意確率 ( $p$  値) は近似値です。正確な傾向検定も実行できます。

---

## 「Fisherの正確検定」レポート

$r \times c$  表に対する Fisher の正確検定は、2変数間の関連を検定します。 $r \times c$  表とは、 $r$  行、 $c$  列の表を指します。Fisher の正確検定は、行和および列和を固定して、超幾何分布を用いて  $p$  値を求めます。

この検定は、大標本近似ではなく、行和と列和を固定したもとで求められる条件付き正確検定です。標本サイズが小さい場合や、度数が0のセルがある場合など、尤度比検定や Pearson 検定では妥当な結果が得られないときに適しています。

「Fisherの正確検定」レポートには、次のような情報が表示されます。

**表の確率 (P)** 現在、観測されている表が生じる確率。検定の  $p$  値ではありません。

**両側  $p$  値 (表の生起確率  $\leq p$ )** 両側検定の有意確率 ( $p$  値)。

$2 \times 2$  分割表に対しては、1つの行または1つの列のすべてが0ではない限り（この場合、検定を実行できません）、Fisher の正確検定が自動的に実行されます。「[検定](#)」[レポート](#)」を参照してください。

## 同等性検定のレポート

「分割表」プラットフォームの [同等性検定] サブメニューにあるオプションは、リスク差または相対リスクに対して同等性検定・優越性検定・非劣性検定を行います。

「同等性検定」レポートは、図と要約表で構成されます。[同等性検定] サブメニューでオプションを選択すると、検定の詳細を設定するウィンドウが開きます。

### 同等性の検定に対する設定ウィンドウ

[同等性検定] サブメニューでオプションを選択すると、検定を定義するためのウィンドウが開きます。

**対立仮説** 検定の種類を選択します。どの対立仮説の検定を行うかを選択します。

**同等性（上下限）** 同等性検定を行います。このオプションは、グループ間の違いが同等性のマージンより大きくないことを示したいときに使います。

**優越性（片側）** 優越性検定を行います。このオプションは、一方のグループが他方より優れている（良い）ことを示したいときに使います。

**非劣性（片側）** 非劣性検定を行います。このオプションは、一方のグループが他方より劣っていないことを示したいときに使います。

**対立仮説の方向** （優越性検定と非劣性検定の場合にのみ指定します。）対立仮説の方向を指定します。

**仮説のプロット** 帰無仮説および対立仮説を図で表現したもの。

**Yでの対象となるカテゴリ** 割合を計算する際の分子となるY変数の水準。

**Xでの対象となるカテゴリ** コントロール群となるX変数の水準。（リスク差では引かれる方の群、相対リスクでは分子となる群）

**マージンと有意水準** 検定の有意水準。

**差** （リスク差に対する検定で表示されます。）同等性検定・優越性検定・非劣性検定のマージン。このマージン（ $\Delta$ ）には、実質的に意味がある差を指定します。同等性検定の場合は、ここに指定する「差」は0より大きくなければなりません。

**比** （相対リスクの検定での比に対するマージンを指定します。）同等性検定・優越性検定・非劣性検定のマージン。リスク比に対するマージンです。このマージン（ $\Delta$ ）には、実質的に意味があるリスク比を指定します。値の範囲は（比, 1/比）と定義されます。同等性検定の場合、ここに指定する「比」は1とは異なる値でなければなりません。

**アルファ** 検定の有意水準。

## 「同等性検定」レポート

検定のレポートでは、冒頭に対立仮説が表示されます。比較ごとに、以下の列を含む「検定」レポートが作成されます。

**差** （リスク差に対する検定で表示されます。）リスク差の点推定値。

**比** （相対リスクに対する検定で表示されます。）リスク比の点推定値。

**差の標準誤差** （リスク差に対する検定で表示されます。）リスク差の標準誤差。

**下限値 標準誤差、上限値 標準誤差** （優越性検定・非劣性検定の場合は、どちらか一方のみ計算されます。）仮説の下限値または上限値の下で計算された標準誤差。

**下限値 z 値、上限値 z 値** （優越性検定・非劣性検定の場合は、どちらか一方のみ計算されます。）片側検定の下限または上限のz比。

**下限値 p 値、上限値 p 値** （非劣性検定・優越性検定の場合は、どちらか一方のp値のみ計算されます。）下限値または上限値のz比に対応する有意確率（p値）。

**最大p値** （同等性検定の場合にのみ計算されます。）2つの片側t検定のp値のうち大きいほうの値。

**両側 90% 下限、両側 90% 上限** リスク差またはリスク比に対する、信頼係数  $(1-2\alpha)$  の両側信頼区間の限界。

**判定** 指定の有意水準で行った仮説検定の判定。

## 同等性検定のオプション

「同等性検定」・「優越性検定」・「非劣性検定」の赤い三角ボタンのメニューには、以下のオプションが含まれます。

**検定のレポート** リスク差やリスク比の同等性検定・優越性検定・非劣性検定を要約したレポートの表示／非表示を切り替えます。「[同等性検定のレポート](#)」を参照してください。

**フォレストプロット** フォレストプロットの表示／非表示を切り替えます。「フォレストプロット」は、リスク差または相対リスクに対する信頼区間をプロットしたグラフです。信頼区間は、リスク差または相対リスクを示すスケール上にプロットされます。色の付いた領域は、同等性・優越性・非劣性を示す区間です。

---

**ヒント:** 点の上にポイントを置くと、比較されているグループとリスク差またはリスク比の推定値が表示されます。

---

**削除** 「分割表に対する分析」レポートウィンドウから検定のレポートが削除されます。

---

## 「分割表」プラットフォームの別例

この節では、「分割表」プラットフォームを使った例をご紹介します。

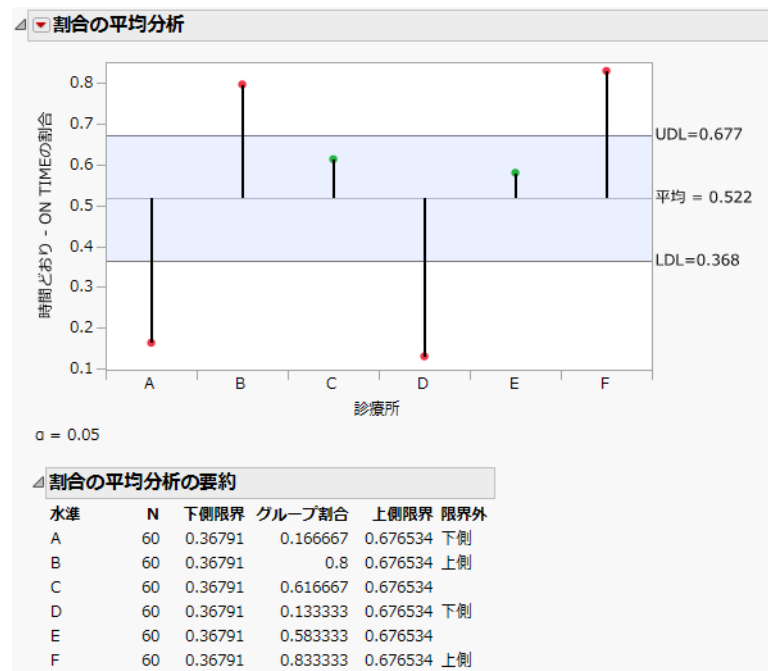
- 「割合の平均分析の例」
- 「対応分析の例」
- 「Cochran-Mantel-Haenszel 検定の例」
- 「[一貫性の統計量] オプションの例」
- 「[相対リスク] オプションの例」
- 「割合の2標本検定の例」
- 「[関連の指標] オプションの例」
- 「Cochran-Armitage の傾向検定の例」
- 「相対リスクの同等性検定の例」

## 割合の平均分析の例

「分割表」プラットフォームを使い、ある地域にある6軒の診療所において予約時間通りに診察された患者の割合を調べます。6つの診療所ごとに、1週間分のデータから、60の診察予約データが無作為抽出しました。予約時間から5分以内に診察室に通された場合には、定刻どおりの診察とみなします。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Office Visits.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「時間どおり」を選択し、[Y, 目的変数] をクリックします。
4. 「診療所」を選択し、[X, 説明変数] をクリックします。
5. 「度数」を選択し、[度数] をクリックします。
6. [OK] をクリックします。
7. 「分割表に対する分析」の赤い三角ボタンをクリックし、[割合の平均分析] を選択します。
8. 「割合の平均分析」の赤い三角ボタンをクリックし、[要約レポートの表示] および [応答水準の切り替え] を選択します。

図7.7 割合の平均分析の例



「割合の平均分析」のプロットは、予約時間どおりに診察を受けた患者の割合を、診療所ごとに示しています。次の点に注意してください。

- 予約時間どおりに受診できた割合が一番高いのは、診療所Fです。診療所Bがそれに次いでいます。

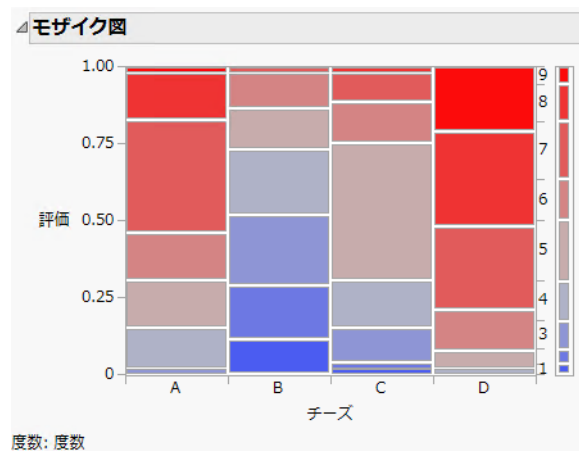
- 診療所Dは、予約時間どおりに受診できた割合が一番低く、診療所Aがそれに次いでいます。
- 診療所CとEは平均に近く、決定限界を超えていません。

## 対応分析の例

「分析表」プラットフォームを使い、チーズの試食試験の結果に対して対応分析を行います。試食試験では、4種類のチーズ添加物を9段階で評価し、その度数を記録しました。対応分析は、変数間に関連があることが確認されているとき、その関連がどのようなものであるかをグラフで表します。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Cheese.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「評価」を選択し、[Y, 目的変数] をクリックします。  
「評価」は、1（最下位）から9（第1位）までを範囲としています。
4. 「チーズ」を選択し、[X, 説明変数] をクリックします。  
「A」・「B」・「C」・「D」はチーズの4つの添加物を表します。
5. 「度数」を選択し、[度数] をクリックします。
6. [OK] をクリックします。

図7.8 チーズデータのモザイク図



モザイク図を見ると、チーズの種類によって評価順位が異なることがわかります。特にチーズBは、一貫して他のチーズより評価が低くなっています。このような状況での次の分析としては、対応分析が考えられます。

7. 対応分析のプロットを表示するには、「分割表に対する分析」の赤い三角ボタンをクリックし、[対応分析] を選択します。



図7.9 対応分析プロットの例

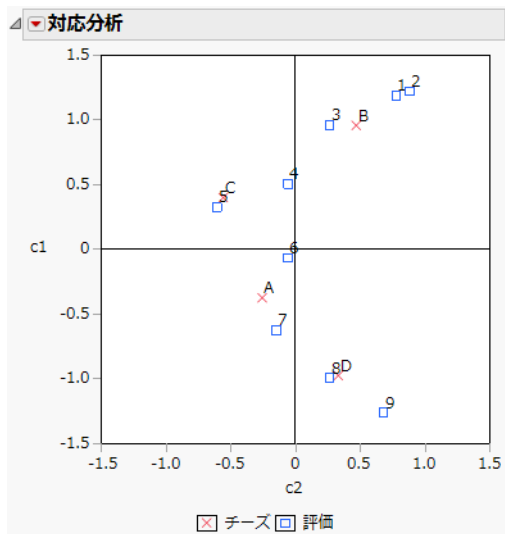


図7.9は、対応分析をグラフ化したもので、第1軸には「c1」、第2軸には「c2」という名前が付いています。次の点に注意してください。

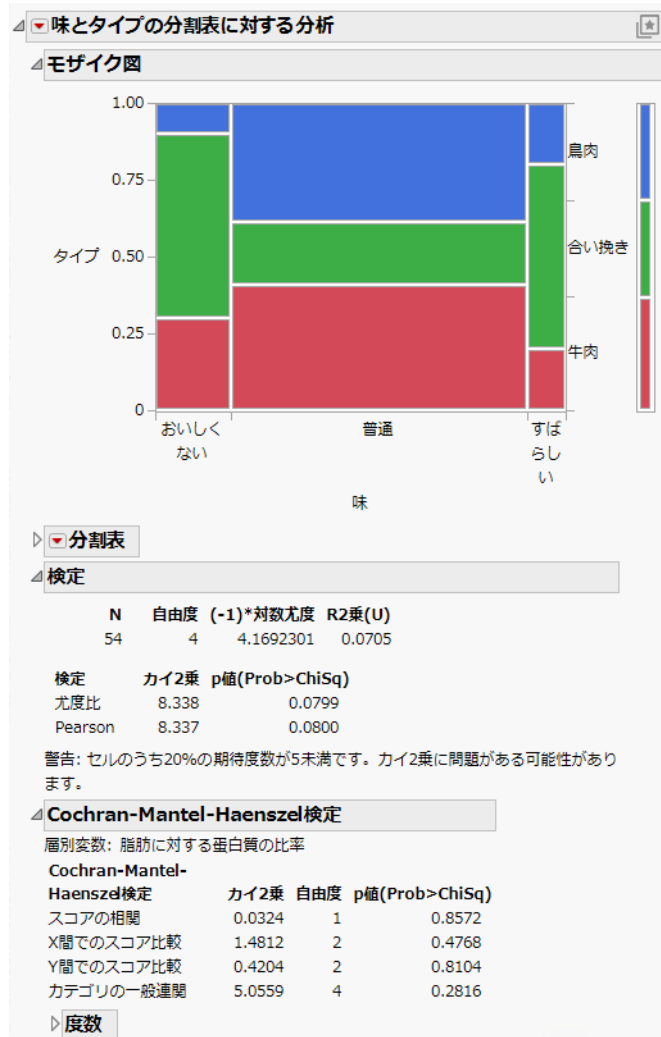
- c1軸は、全体的な満足度を表しているようです。c1軸の上部に1（最低の評価）、下部に9（最高の評価）という評価が分布しています。
- チーズDは、評価が8と9である割合が、他のチーズよりも高くなっています。
- チーズBは、評価が1～3である割合が、他のチーズよりも高くなっています。
- チーズCとAは、4～7という中間的な評価の割合が、他のチーズよりも高くなっています。

## Cochran-Mantel-Haenszel 検定の例

「分割表」プラットフォームを使ってホットドックのタイプと味の関係を調べます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Hot Dogs.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「タイプ」を選択し、[Y, 目的変数] をクリックします。
4. 「味」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の赤い三角ボタンをクリックして [Cochran-Mantel-Haenszel 検定] を選択します。
7. 層別変数として「脂肪に対する蛋白質の比率」を選択し、[OK] をクリックします。

図7.10 Cochran-Mantel-Haenszel 検定の例



次の点に注意してください。

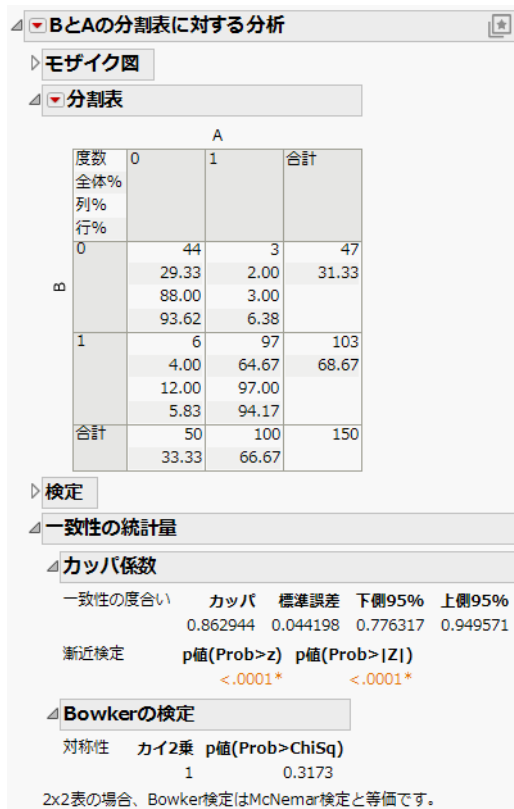
- 「検定」レポートを見ると、Pearson のカイ 2 乗検定に対する  $p$  値は 0.0799 で、5% よりは少し大きいです。有意水準を 10% とすると、ホットドッグの「味」と「タイプ」の間に何らかの関係があるかもしれないことを示唆しています。
- 「Cochran-Mantel-Haenszel 検定」のレポートを見ると、カテゴリの一般連関の  $p$  値が 0.2816 となっています。「脂肪に対する蛋白質の比率」で層別化した場合は、ホットドッグの「味」と「タイプ」の間に関係があるとは言えません。

## 【一致性の統計量】 オプションの例

「分割表」プラットフォームを使って2人の判定者の関係を調べます。3人の判定者が50個の部品をそれぞれ3回ずつ評価し、結果をデータテーブルにまとめました。判定者AとBの関係を調べてみましょう。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Attribute Gauge.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「A」を選択し、[Y, 目的変数] をクリックします。
4. 「B」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の赤い三角ボタンをクリックして【一致性の統計量】を選択します。

図7.11 「一致性の統計量」レポートの例



一致性の統計量は0.86と、1に近い大きな値となっています。また、その $p$ 値は<.0001と小さくなっています。分割表を見ても、対角線上における度数が大きく、一致性が高いことを裏付けています。2名の評価者が同じ評価をしている場合（つまり、ともに0と評価するか、または、ともに1と評価している場合）、一致性の統計量は大きくなります。

## 「相対リスク」オプションの例

「分割表」プラットフォームを使って調査対象者における性別ごとの既婚の割合を調べてみましょう。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「既婚/未婚」を選択し、[Y, 目的変数] をクリックします。
4. 「性別」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の赤い三角ボタンをクリックして [相対リスク] を選択します。  
「相対リスクカテゴリの選択」ウィンドウが表示されます。

図7.12 「相対リスクカテゴリの選択」ウィンドウ

「相対リスクカテゴリの選択」ウィンドウについて、次の点を確認してください。

- 1 対の応答と因子についてのみ相対リスクを計算するには、このダイアログボックスで対象の変数を指定します。たとえば、図7.12のウィンドウで [OK] をクリックすると、次のように計算されます。

$$\frac{P(Y = \text{未婚} | X = \text{女性})}{P(Y = \text{未婚} | X = \text{男性})}$$

- 応答と因子のすべての組み合わせ ( $2 \times 2 = 4$ ) について相対リスクを計算するには、[すべての組み合わせを計算] チェックボックスをオンにします (図7.13)。
7. [すべての組み合わせを計算] チェックボックスをオンにして、すべての組み合わせについて計算を行います。他のすべての設定はデフォルト値のままにしておきます。

図7.13 「相対リスク」レポートの例

性別と既婚/未婚の分割表に対する分析

モザイク図

分割表

既婚/未婚

度数

既婚

未婚

合計

全体%

列%

行%

女性

95

43

138

31.35

14.19

45.54

48.47

40.19

68.84

31.16

男性

101

64

165

33.33

21.12

54.46

51.53

59.81

61.21

38.79

合計

196

107

303

64.69

35.31

性別

検定

相対リスク

説明

相対リスク

下側95%

上側95%

P(既婚|女性)/P(既婚|男性)

1.124623

0.953193

1.326885

P(既婚|男性)/P(既婚|女性)

0.889187

0.753645

1.049105

P(未婚|女性)/P(未婚|男性)

0.803329

0.587179

1.099046

P(未婚|男性)/P(未婚|女性)

1.24482

0.90988

1.703057

相対リスクは次のように計算します。

1. 「相対リスク」レポートの最初の項目 (P(既婚|女性)/P(既婚|男性)) を例にして説明します。
2. P(既婚|女性) および P(既婚|男性) といった確率は、分割表に表示されています。これらの確率は、分割表では、女性および男性の各行における、既婚の「行%」に相当します。

$$P(\text{既婚} | \text{女性}) = 0.6884$$

$$P(\text{既婚} | \text{男性}) = 0.6121$$

これらの値をもとに、相対リスクは次式で計算されます。

$$P(\text{既婚} | \text{女性}) / P(\text{既婚} | \text{男性}) = \frac{0.6884}{0.6121} = 1.1247$$

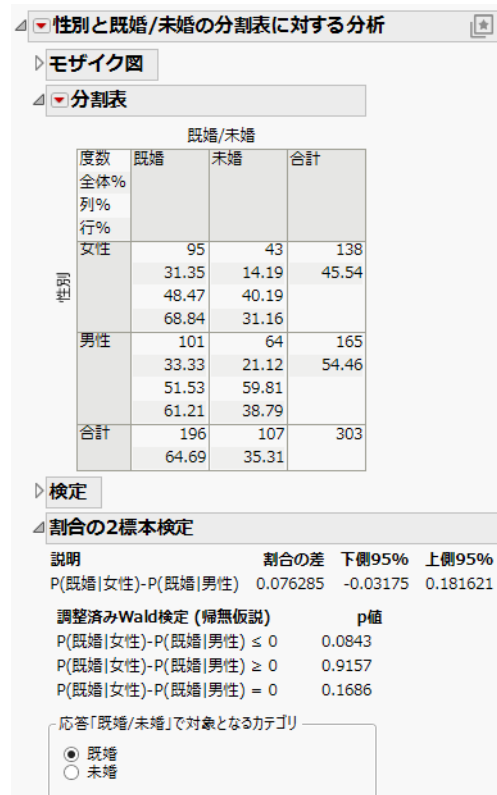
## 割合の2標本検定の例

「分割表」プラットフォームを使って、女性と男性について既婚者の割合を調べてみましょう。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「既婚/未婚」を選択し、[Y, 目的変数] をクリックします。
4. 「性別」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。

6. 「分割表に対する分析」の赤い三角ボタンをクリックし、[割合の2標本検定] を選択します。

図7.14 「割合の2標本検定」レポートの例



この例では、女性と男性で既婚者の割合を比較します。「分割表」の「行%」で次の値を確認します。

$$P(\text{既婚} | \text{女性}) = 0.6884$$

$$P(\text{既婚} | \text{男性}) = 0.6121$$

2つの割合の差は0.0763で、これがレポートの「割合の差」に表示されています。割合の差に対する両側信頼区間は、[-0.03175, 0.181621] となっています。この信頼区間に対応した修正Wald法による両側検定の $p$ 値は、0.1686となっています。これは、Pearsonのカイ2乗検定による $p$ 値(0.1665)に近い値です。一般に、割合の差の検定には、修正Wald検定よりも、Pearsonのカイ2乗検定の方がよく使用されます。

## 【関連の指標】 オプションの例

「分割表」プラットフォームを使って性別と配偶者の有無に関連があるかどうかを調べます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。

- 「既婚/未婚」を選択し、[Y, 目的変数] をクリックします。
- 「性別」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。
- 「分割表に対する分析」の赤い三角ボタンをクリックして [関連の指標] を選択します。

図7.15 「関連の指標」レポートの例

関連の指標				
指標	値	標準誤差	下側95%	上側95%
ガンマ	0.1667	0.1184	-0.0654	0.3987
Kendallのタウ-b	0.0795	0.0570	-0.0321	0.1911
Stuartのタウ-c	0.0757	0.0543	-0.0307	0.1821
SomersのD (C R)	0.0763	0.0547	-0.0309	0.1835
SomersのD (R C)	0.0828	0.0593	-0.0335	0.1991
非対称ラムダ(C R)	0.0000	0.0000	0.0000	0.0000
非対称ラムダ(R C)	0.0000	0.0000	0.0000	0.0000
対称ラムダ	0.0000	0.0000	0.0000	0.0000
不確実性係数(C R)	0.0049	0.0070	0.0000	0.0186
不確実性係数(R C)	0.0046	0.0066	0.0000	0.0176
不確実性係数(対称)	0.0047	0.0068	0.0000	0.0181

分析の対象である変数（性別と既婚/未婚）が名義尺度なので、ラムダと不確実性係数を使用します。いずれの値も小さくなっているため、関連性は弱いようです。

## Cochran-Armitageの傾向検定の例

「分割表」プラットフォームを使って、購入する車両のサイズと男女の比率に関連があるかどうかを調べてみましょう。

- [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。  
この検定の目的に合わせて、「サイズ」を順序尺度に変更します。
- 「列」パネルで、「サイズ」の横にあるアイコンを右クリックし、[順序尺度] を選択します。
- [分析] > [二変量の関係] を選択します。
- 「性別」を選択し、[Y, 目的変数] をクリックします。
- 「サイズ」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。
- 「分割表に対する分析」の赤い三角ボタンをクリックして [Cochran-Armitageの傾向検定] を選択します。

図7.16 「Cochran-Armitageの傾向検定」レポートの例

Cochran-Armitageの傾向検定			
漸近検定	Z	p値(Prob<Z)	p値(Prob> Z )
	-0.37271	0.3547	0.7094

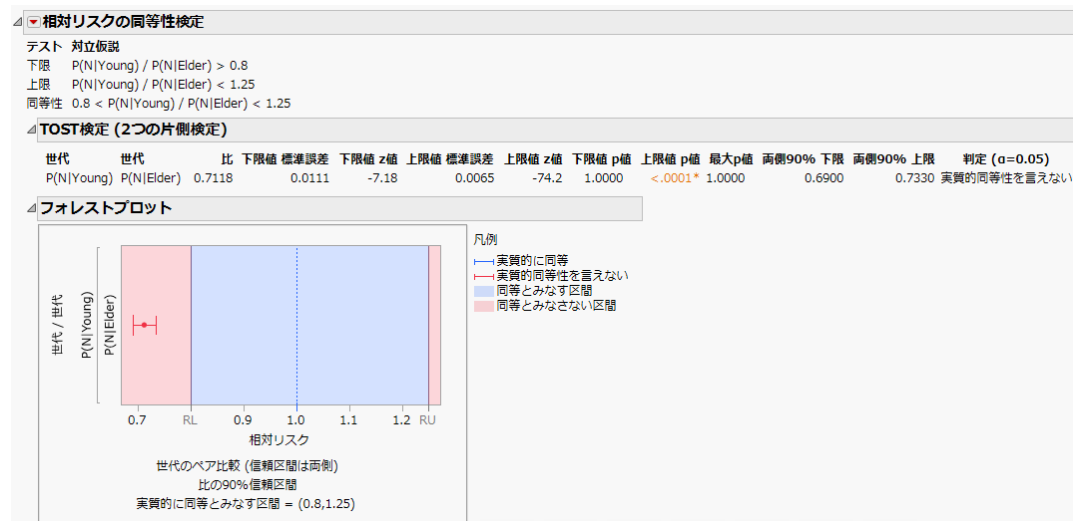
両側検定の  $p$  値 (0.7094) は大きい値になっています。このことから、各サイズの車両を購入する男性と女性の割合に関係があるとは結論できません。

## 相対リスクの同等性検定の例

「分割表」プラットフォームを使って、保険金を請求しない割合がドライバーの世代間で同等かどうかを検定します。ここでは、「ドライバーの世代間の母相対リスクが0.8～1.25である場合、割合は実質的に同等である」とみなすことにします。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Auto Raw Data.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「請求(あり/なし)」を選択し、[Y, 目的変数] をクリックします。
4. 「世代」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「分割表に対する分析」の赤い三角ボタンをクリックして [同等性検定] > [相対リスク] を選択します。
7. 「マージンと有意水準」パネルで、「比」に「0.8」と入力します。
8. 「Xでの対象となるカテゴリ」パネルで「Young」を選択します。
9. [OK] をクリックします。

図7.17 相対リスクの同等性検定



年配の (elder) ドライバーと比較したときの、若い (young) ドライバーが保険金を請求しない標本相対リスクは、0.7118です。2つの片側検定 (TOST 検定) による同等性検定は有意ではありません。よって、「2つのグループが保険金を請求する割合は実質的に同等である」とは結論できません。



## 「分割表」プラットフォームの統計的詳細

この節では、「分割表」プラットフォームで使われている統計手法の詳細を説明します。

- [「一致性の統計量の統計的詳細」](#)
- [「オッズ比の統計的詳細」](#)
- [「「検定」レポートの統計的詳細」](#)
- [「対応分析の統計的詳細」](#)

### 一致性の統計量の統計的詳細

この節では、「分割表」プラットフォームにおける一致性の統計量について詳しく解説します。2名の評価者が、 $n$ 個の対象を判定したとします。2人の判定者の判定が完全に一致している場合は、カッパ係数が1になります。観測された一致が、偶然によって一致する期待値を上回ると、カッパ係数は正の値になります。カッパ係数が大きいほど、一致性の高いことを示します。実際にはほとんどあり得ないケースですが、観測された一致性が偶然による一致性を下回ると、カッパが負になります。カッパ係数の最小値は、周辺度数に応じて-1から0までの値を取ります。

カッパ係数は、次式により計算されます。

$$\hat{\kappa} = \frac{P_0 - P_c}{1 - P_c} \quad \text{この式で、} P_0 = \sum_i p_{ii} \quad \text{および} \quad P_c = \sum_i p_{i+} p_{+i}$$

ここで、 $p_{ij}$  は  $(i, j)$  番目のセルの個体の比率であり、 $\sum_i \sum_j p_{ij} = 1$  が成り立つことに注意してください。

単純なカッパ係数の漸近分散は、次のような推定式で計算されます。

$$\text{var} = \frac{A + B - C}{(1 - P_c)^2 n} \quad \text{この式で} \quad A = \sum_i p_{ii} [1 - (p_{i+} + p_{+i})(1 - \hat{\kappa})]^2, \quad B = (1 - \hat{\kappa})^2 \sum_{i \neq j} p_{ij} (p_{+i} + p_{j+})^2,$$

$$C = [\hat{\kappa} - P_c(1 - \hat{\kappa})]^2$$

Cohen (1960) と Fleiss et al. (1969) を参照してください。

対称性に対する Bowker 検定は、「分割表における確率が対称である ( $p_{ij}=p_{ji}$  である)」という帰無仮説を検定します。

## オッズ比の統計的詳細

「分割表」プラットフォームでは、オッズ比は次の式で計算されます。

$$\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}$$

上の式で、 $p_{ij}$  は  $2 \times 2$  分割表の第  $i$  行第  $j$  列の度数です。

## 「検定」レポートの統計的詳細

この節では、「分割表」プラットフォームの「検定」レポートについて詳しく解説します。

### R2 乗 (U)

R2 乗 (U) は次の式で計算されます。

$$\frac{\text{モデルの「(-1)* 対数尤度」}}{\text{「全体の修正済み」の「(-1)* 対数尤度」}}$$

分母の「(-1)\* 対数尤度」は、標本全体において応答確率が一定であるモデルのものです。

### 検定

この節では、2つのカイ2乗検定の計算について説明します。

「尤度比」カイ2乗検定の検定統計量は、「検定」レポートに表示されている「(-1)\* 対数尤度」を2倍した値です。この「(-1)\* 対数尤度」は、グループごとに応答確率をあてはめたモデルの対数尤度から、標本全体において応答確率が一定であるモデルの対数尤度を引いたものです。尤度比カイ2乗値は、行と列の主効果だけのモデルのデビアンズ（逸脱度）になっています。尤度比カイ2乗値は、 $G^2$  という記号で示されることがあり、次式で求めます。

$$G^2 = 2 \left[ \sum_{ij} (-n_{ij}) \ln(p_j) - \sum_{ij} -n_{ij} \ln(p_{ij}) \right] \quad \text{この式で、} p_{ij} = \frac{n_{ij}}{N} \text{ および } p_j = \frac{N_j}{N}$$

この式は簡略に次式のようにも表せます。

$$G^2 = 2 \sum_i \sum_j n_{ij} \ln \left( \frac{n_{ij}}{e_{ij}} \right)$$

「Pearson」カイ2乗検定の検定統計量は、各セルにおける観測度数と期待度数の差の平方和から計算されます。Pearsonのカイ2乗検定は、大標本では度数が正規分布に従う傾向にあるという性質を用いています。Pearsonのカイ2乗統計量は、デビアンズ（逸脱度）と同じように、行と列の主効果だけのモデルの適合度を示します。Pearsonのカイ2乗統計量は、次のように定義されます。

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

上式において、**O**はセルの観測度数、**E**はセルの期待度数です。和記号は、すべてのセルの合計を表しています。2×2分割表に対してときどき採用されている0.5連続修正は、JMPでは行っていません。

## 対応分析の統計的詳細

この節では、「分割表」プラットフォームの対応分析について詳しく解説します。

対応分析では、次のような特異値分解が行われます。

$$\mathbf{D}_r^{-0.5}(\mathbf{P} - r\mathbf{c}')\mathbf{D}_c^{-0.5} = \mathbf{U}\mathbf{D}\mathbf{diag}(\Lambda)\mathbf{V}'$$

ここで

**P**は、合計度数（総和）で度数を割った値を要素とする行列

**r**は**P**の行和。**c**は**P**の列和

**D<sub>r</sub>**と**D<sub>c</sub>**は、それぞれ**r**および**c**の値から成る対角行列です。

**Λ**は、特異値の列ベクトル（この特異値は、「詳細」レポートに表示されます）

特異値分解の詳細については、『多変量分析』を参照してください。

「詳細」レポートの行の座標（**rc**）と列の座標（**cc**）は、次のように計算されます。

$$rc = \mathbf{D}_r^{-0.5}\mathbf{U}\mathbf{D}\mathbf{diag}(\Lambda)$$

$$cc = \mathbf{D}_c^{-0.5}\mathbf{V}\mathbf{D}\mathbf{diag}(\Lambda)$$



# 第8章

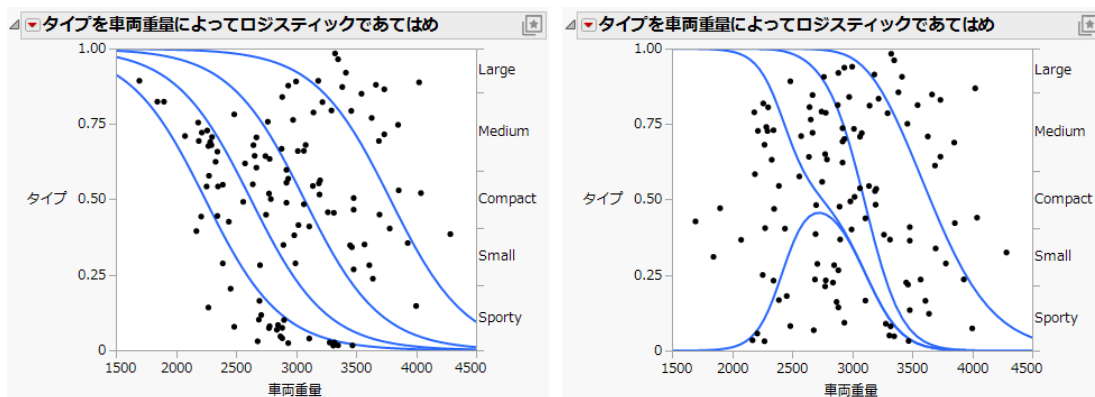
## ロジスティック分析 カテゴリカルなY変数と連続尺度のX変数の関係を調べる

「ロジスティック」プラットフォームでは、連続尺度の説明変数（x）とカテゴリカルな応答変数に対してロジスティック回帰モデルをあてはめます。ROC 曲線（受診者動作特性曲線）、リフトチャート、オッズ比が表示できます。あてはめたモデルにより、X 変数の各値における応答の確率を推定できます。逆推定を行って、Y 変数の特定の確率値に対する X の値を推定することもできます。

「ロジスティック」プラットフォームは、「二変数の関係」プラットフォームのうち、応答変数が**名義尺度もしくは順序尺度で、説明変数が連続尺度**の場合に実行される分析です。応答変数が名義尺度のときには名義ロジスティック回帰、順序尺度のときには順序ロジスティック回帰が行われます。

- 名義ロジスティック回帰モデルは、名義尺度の応答変数の各水準の確率を表す一連の曲線を求めます。[図8.1](#)の右の図が名義ロジスティック回帰モデルの一例です。
- 順序ロジスティック回帰モデルは、順序尺度の応答変数が「イベントを示す水準」以下となる確率を推定します。このモデルは、ロジスティック曲線を1つだけ計算し、それを水平にずらすことですべての順序カテゴリの確率を推定します。このモデルは、名義ロジスティックよりも複雑でなく、順序尺度の応答変数に適しています。[図8.1](#)の左の図が順序ロジスティック回帰モデルの一例です。

図8.1 順序・名義ロジスティック回帰の例



## 目次

「ロジスティック」プラットフォームの概要 .....	271
名義ロジスティック回帰の例 .....	271
「ロジスティック」プラットフォームの起動 .....	273
データ形式 .....	274
「ロジスティック」レポート .....	274
ロジスティックプロット .....	275
「反復履歴」レポート .....	276
「モデル全体の検定」レポート .....	276
「あてはめの詳細」レポート .....	277
「パラメータ推定値」レポート .....	278
「ロジスティック」プラットフォームのオプション .....	278
ロジスティック分析レポート .....	280
ROC 曲線 .....	280
逆推定 .....	281
ロジスティック回帰の別例 .....	281
順序ロジスティック回帰の例 .....	281
ロジスティックプロットの例 .....	283
ROC 曲線の例 .....	285
逆推定の例 .....	286
「ロジスティック」プラットフォームの統計的詳細 .....	289

## 「ロジスティック」プラットフォームの概要

ロジスティック回帰は、連続尺度のX変数の値に対するカテゴリカルなY変数の水準の確率を求めます。ロジスティック回帰は古くから普及している手法で、医薬品開発における用量反応データや、マーケティングの購買選択など、いろいろな分野で使用されています。ロジスティック回帰の文献 (Hosmer and Lemeshow 1989) のほかに、多くのカテゴリカル分析のテキスト (Agresti 1990) で取り上げられています。

同じ状況でも、連続変数をY、カテゴリカル変数をXと考えて判別分析を行うこともできます。判別分析では、連続量のデータが固定された説明変数ではなく正規分布に従ったランダムな応答であると仮定されます。『多変量分析』を参照してください。

「二変量の関係」プラットフォームにおけるロジスティック単回帰分析は、説明変数が1つしかないモデルの結果をグラフで表現したものです。一方、「モデルのあてはめ」プラットフォームで行われるロジスティック回帰モデルでは複数の説明変数を扱えます。説明変数が2つ以上ある場合のロジスティック回帰モデルについては、『基本的な回帰モデル』を参照してください。正規分布関数を使ったロジスティック回帰（**プロビット分析**）については、『予測モデルおよび発展的なモデル』を参照してください。

### 名義ロジスティック回帰

名義ロジスティック回帰モデルは、応答変数がある水準になる確率を、説明変数の滑らかな関数によって表します。予測確率は0～1の値を取り、X変数のどの値においてもY変数の水準の予測確率の和は1になります。

ロジスティック確率プロットでは縦軸が確率を表します。Y変数の水準数を $k$ とすると、 $k-1$ 本の滑らかな曲線が作成され、それによって確率の合計 (=1) がY変数の水準ごとに分けられます。ロジスティック回帰のあてはめでは、「観測された応答データが生じる確率の自然対数」を合計した値の符号を逆にしたものが最小化されます。この推定方法を「最尤法」と呼びます。

### 順序ロジスティック回帰

Y変数が順序尺度の場合、順序性を考慮したロジスティック回帰が行われます。Y変数が特定的水準以下になる累積確率がロジスティック曲線で表されます。曲線の形状はどの水準でも同じで、水平方向の位置が異なるだけです。

順序ロジスティックモデルでは、Y変数の水準数を $r$ としたとき、累積ロジスティック曲線の $r-1$ 個の異なる切片を推定しますが、傾きは1つです。順序モデルは計算されるパラメータが名義モデルより少ないため、適切であれば順序モデルを使用した方が効率的です。

## 名義ロジスティック回帰の例

「ロジスティック」プラットフォームを使って、連続尺度の因子と名義尺度の応答変数の関係を調べます。このデータは、ウサギを12匹ずつ5群に分け、連鎖球菌を注入した実験から得たものです。体内に連鎖球菌が認められたウサギに対して、用量を変えてペニシリンを投与します。ウサギの治癒率に対して、用量の自然対数が効果を持つかどうかを調べます。

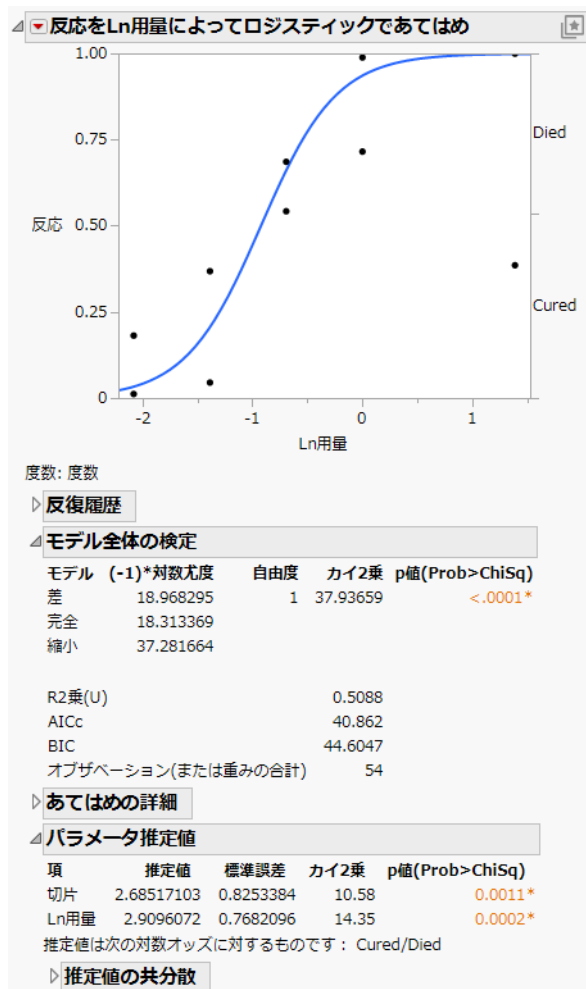
1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Penicillin.jmp」を開きます。

2. [分析] > [二変量の関係] を選択します。
3. 「反応」を選択し、[Y, 目的変数] をクリックします。
4. 「Ln用量」を選択し、[X, 説明変数] をクリックします。

「度数」列が自動的に [度数] に設定されています。これは、「度数」列にあらかじめ [度数] の役割が割り当てられているためです。

5. 「イベントを示す水準」から [Cured] を選択します。
6. [OK] をクリックします。

図8.2 名義ロジスティックのレポートの例





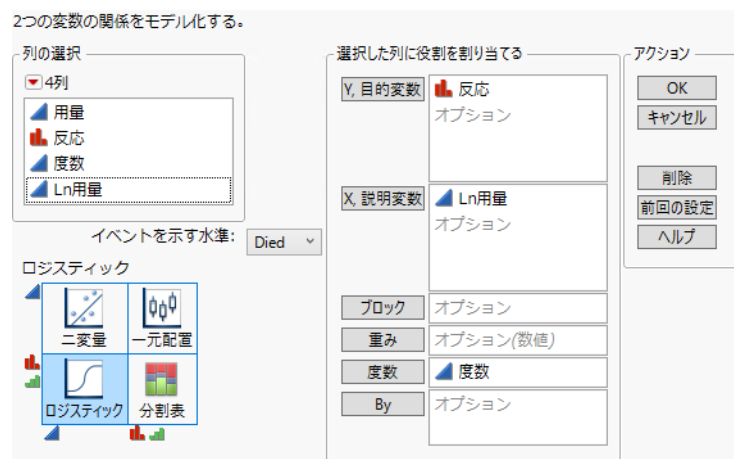
治癒の予測確率が「Ln 用量」の関数としてあてはめられたモデルを示すグラフが作成されます。 $p$  値は統計的に有意で、「薬の用量は、ウサギが治癒する確率に影響している」と結論できます。右の縦軸に表示されている周辺分布は、X 変数と Y 変数の間に関連がない場合の、Y 変数の各水準の確率を示します。

**ヒント:** 2 水準のうちのどちらかを「イベント」とするかは、起動ウィンドウで「イベントを示す水準」を指定するか、またはデータテーブルにおいて「値の表示順序」列プロパティを使用することで変更できます。

## 「ロジスティック」プラットフォームの起動

「ロジスティック」プラットフォームを起動するには、[分析] > [二変数の関係] を選択します。「二変数の関係」起動ウィンドウでは、4 種類の分析が可能です。順序尺度または名義尺度の Y 変数と連続尺度の X 変数を指定すると、「ロジスティック」プラットフォームが起動されます。

図 8.3 ロジスティックの起動ウィンドウ



「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。起動ウィンドウには、次のようなオプションがあります。

**Y, 応答変数** 分析したい目的変数（応答変数）。目的変数は従属変数とも呼ばれます。順序尺度か名義尺度でなければなりません。

**X, 説明変数** 分析したい説明変数（予測変数）。説明変数は独立変数とも呼ばれます。連続尺度でなければなりません。

**ブロック** （ロジスティック分析では使用できません。）ブロック変数とする列。

**重み** データテーブルの各行に対する重みを含んだ列。値が0より大きい行だけが分析に用いられます。

**度数** 分析において各データ行に度を割り当てます。分析対象のデータが要約されたものであるときに役立ちます。

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

**イベントを示す水準**（名義尺度の応答変数を1つ指定し、その変数が2値である場合に使用できます。）2水準のうちにどちらの水準に興味があるか（オッズの分子にどちらの水準をもってくるか）を指定できます。

---

**ヒント:** ロジスティックモデルでイベントとして使用するデフォルトの水準は、水準の順序によって決まります。デフォルトの水準を変更するには、「値の表示順序」列プロパティを使います。

---

## データ形式

「ロジスティック」プラットフォームでは、要約されていないデータ、および要約されているデータのいずれも分析できます。

**要約されていないデータ** 観測値ごとに1行ずつあり、Xの値が1列、Yの値が1列に入力されています。

**要約されたデータ** XとYの値の組み合わせが同じである複数の観測値が1行にまとめられているデータです。データテーブルには、各行の観測値の数を示す度数列が必要です。起動ウィンドウで、この列を「度数」に指定してください。

「ロジスティック」プラットフォームの要約されたデータの例は、「[名義ロジスティック回帰の例](#)」にあります。

---

**メモ:** 「二変量の関係」起動ウィンドウでは、連続尺度、順序尺度、名義尺度の列が使用できます。順序尺度または名義尺度の目的変数（Y）と連続尺度の説明変数（X）という組み合わせに対して、「ロジスティック」プラットフォームが起動されます。それ以外の組み合わせでは、「二変量」、「一元配置」、「分割表」のいずれかのプラットフォームが起動します。

---

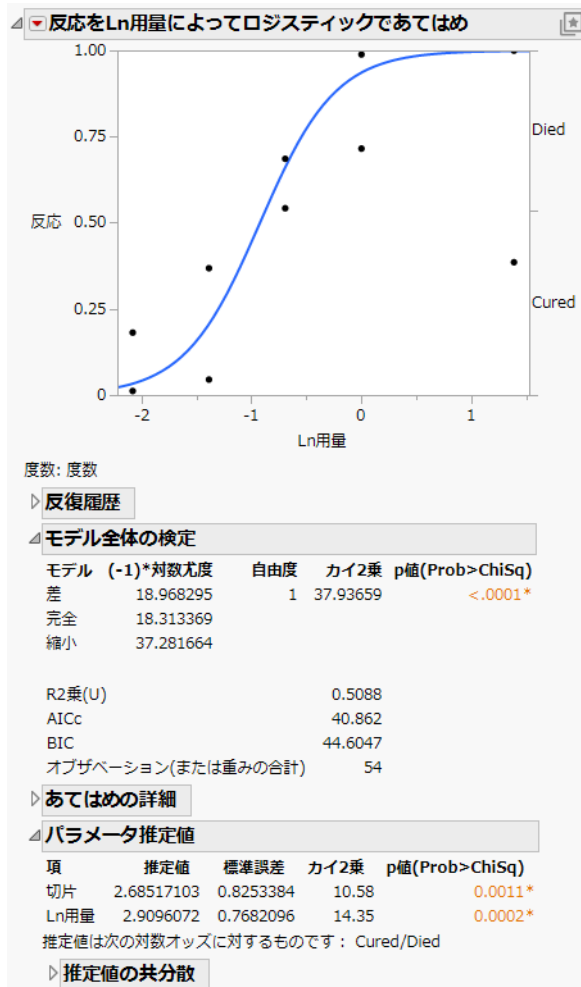
## 「ロジスティック」レポート

「ロジスティック」レポートには、最初、ロジスティックプロット、「反復履歴」、「モデル全体の検定」、「あてはめの詳細」、「パラメータ推定値」が表示されます。赤い三角ボタンのメニューを使って、ロジスティックプロットの表示を変更したり、他のプロットを追加したりできます。「[「ロジスティック」プラットフォームのオプション](#)」を参照してください。

この節では、「ロジスティック」レポートの以下のセクションをご紹介します。

- 「[ロジスティックプロット](#)」
- 「[反復履歴](#)」レポート
- 「[モデル全体の検定](#)」レポート
- 「[あてはめの詳細](#)」レポート
- 「[パラメータ推定値](#)」レポート

図8.4 「ロジスティック」レポートの例



## 変数対話的に置換する

データテーブルの「列」パネルで該当する変数をクリックし、プロットの軸上までドラッグして変数を置換することもできます。

## ロジスティックプロット

「ロジスティック」プラットフォームのロジスティックプロットは、あてはめられたモデルを図にしたものです。右側の縦軸は、Y変数の各水準のオブザベーション数の割合で区切られています。プロットに描かれている曲線は、Y変数の水準のうち1つを除くすべての水準をモデル化したものです。曲線は、Y変数の該当する水準以下となる予測累積確率を示します。左側の縦軸は、確率のスケールを示します。特定の水準における予測確率は、その水準の曲線とその下の水準の曲線の縦の距離として測定できます。

ロジスティックプロット上の点は、データテーブルの各行に対応しています。各点の横方向の位置は、連続尺度のX変数の値によって決まります。各点の縦方向の位置は、カテゴリカル尺度のY変数の値によって決まります。縦方向の位置で見たとき、各点は、その水準の曲線とその下の曲線の間にランダムに散らばります。縦方向の位置はランダムに決められますが、点をずらして表示することで、密集している場所がわかりやすくなります。乱数シード値は固定されているので、同じモデルを複数回あてはめても、各点のY座標は変わりません。

## 「反復履歴」レポート

「ロジスティック」プラットフォームの「反復履歴」レポートには、各反復計算とロジスティックモデルの収束判定基準が表示されます。「反復履歴」は名義ロジスティック回帰の場合にのみ表示されます。

## 「モデル全体の検定」レポート

「ロジスティック」プラットフォームの「モデル全体の検定」レポートは、モデルが、「データ全体で応答確率が一定である」と仮定したときよりも良くあてはまっているかどうかを示します。このレポートは、連続尺度の応答に対する分散分析表に相当します。このレポートに表示される検定は、ロジスティック回帰モデルがデータにどの程度あてはまっているかを評価する尤度比カイ2乗検定です。

あてはめられたモデルから計算される確率の自然対数を取り、その合計の符号を逆にしたものが「(-1)\*対数尤度」です。カテゴリカルな応答における「(-1)\*対数尤度」は、連続尺度の応答における平方和に相当します。現在のモデルの「(-1)\*対数尤度」から、全体で確率が一定であると仮定したモデルの「(-1)\*対数尤度」を引いた差を2倍したものが、尤度比検定のカイ2乗統計量になります。この検定統計量は、X変数の値とY変数の水準の間に関連がないとする仮説を検定します。

「R2乗(U)」（ $R^2$ ）は0～1の値を取ります。 $R^2$ の値が大きいほど、モデルのあてはまりが良いと考えられます。ただし、カテゴリカルな応答のモデルでは、この $R^2$ が高い値となることはほとんどありません。

「モデル全体の検定」レポートには、次の情報が表示されます。

**モデル** 変動の要因。

**差** 完全（full）モデルと縮小（reduced）モデルの差。このモデルは、X変数があてはまりにどの程度貢献しているかを表します。

**完全** 切片とX変数を含んだモデル。

**縮小** 切片しか含まないモデル。

**(-1)\*対数尤度** 各モデルのばらつきの指標。『基本的な回帰モデル』を参照してください。

**自由度** 指定したモデルと切片だけのモデルとの自由度の差。言い換えると、切片を除いたモデルの自由度です。

**カイ2乗** 尤度比カイ2乗。「応答確率が標本全体で一定であるとしたときよりも、モデルの適合度は良いわけではない」という帰無仮説を検定する値です。この検定統計量は、「完全」と「縮小」の負の対数尤度における差を2倍した値です。「[「ロジスティック」プラットフォームの統計的詳細](#)」を参照してください。

**p 値 (Prob>ChiSq)** 「真のモデルが切片だけのモデルである」という仮定のもとで、現在の値より大きなカイ2乗値を得る確率。通常、 $p$  値が0.05より小さいとモデルが有意だと考えられます。

**R2 乗 (U)** 不確定性のうち、あてはめたモデルに起因する部分。負の対数尤度の「差」を「縮小」で割った値で定義される。「R2 乗 (U)」の値が1の場合、イベントが生じる予測確率は1に等しくなり、あてはめられたモデルは完全に応答値を予測しています。ロジスティックモデルの場合、予測確率が確定的であることはほとんどないため、「R2 乗 (U)」の値は小さくなるのが普通です。[「ロジスティック」プラットフォームの統計的詳細](#)を参照してください。

「R2 乗 (U)」は、「不確定性係数 (U)」や「McFadden の疑似 R2 乗」と呼ばれることもあります。

**AICc** 修正済み赤池情報量規準。『基本的な回帰モデル』を参照してください。

**BIC** ベイズ情報量規準。『基本的な回帰モデル』を参照してください。

**オブザベーション (または重みの合計)** データの標本サイズ。「モデルのあてはめ」ウィンドウで [度数] または [重み] の列を指定した場合、この値はその列の値の合計となります。

## 「あてはめの詳細」レポート

「ロジスティック」プラットフォームの「あてはめの詳細」レポートには、以下の適合度指標とその定義の式が表示されます。

**エントロピー R2 乗** 現在のモデルと、切片だけのモデルの対数尤度を比較します。「R2 乗 (U)」と同じです。[「ロジスティック」プラットフォームの統計的詳細](#)を参照してください。

**一般化 R2 乗** この指標は、一般的な回帰モデルに適用できるものです。尤度関数  $L$  に基づいており、最大値が1になるように尺度化されています。一般化 R2 乗は、通常の R2 乗（正規分布に従う連続尺度の応答変数に対する標準最小2乗法の R2 乗）を一般化したものです。この一般化 R2 乗は、「Nagelkerke の  $R^2$ 」、または「Craig and Uhler の  $R^2$ 」とも呼ばれており、Cox and Snell の疑似  $R^2$  を最大が1になるように尺度化したものです。Nagelkerke (1991) を参照してください。

**平均 -Log p**  $-\log(p)$  の平均。 $p$  は、実際に観測された応答水準に対する予測確率です。

**RASE** 誤差の標準偏差。誤差は、応答と  $p$ （実際に観測された応答水準に対する予測確率）の差です。

**平均 絶対偏差** 応答と  $p$ （実際に発生したイベントの確率の予測値）の差の絶対値の平均。

**誤分類率** 予測確率が最も大きい応答の水準が、観測された水準と一致しない割合。

**N** 測定値の個数。

「エントロピー R2 乗」と「一般化 R2 乗」は、1に近いほど適合度が良いことを示します。 $[\text{平均 } -\text{Log } p] \cdot [\text{RASE}] \cdot [\text{平均 絶対偏差}] \cdot [\text{誤分類率}]$  は、値が小さいほど、適合度が良いことを示します。

## 「パラメータ推定値」レポート

名義ロジスティックモデルは、 $k-1$  個ずつの切片パラメータと傾きパラメータを持ちます。ここで、 $k$  は、応答変数の水準数です。一方、順序ロジスティックモデルでは、 $k-1$  個の切片パラメータと1個だけの傾きパラメータを持ちます。「ロジスティック」プラットフォームの「パラメータ推定値」レポートには、以下の推定値が表示されます。実際に必要になることはまずありませんが、各パラメータ推定値を個別に検定することも可能です。

**項** ロジスティックモデル内のパラメータ。応答変数の各水準（最後の水準を除く）における因子ごとに、切片と傾きが表示されます。順序ロジスティックモデルには、傾きが1つしかありません。

**推定値** ロジスティックモデルで計算されたパラメータ推定値。

**標準誤差** 各パラメータ推定値の標準誤差。各項を0と比較する統計量の計算に使用されます。

**カイ2乗** 各パラメータが0であるという仮説を検定するWald検定。Waldのカイ2乗は $(\text{推定値} \div \text{標準誤差})^2$ という式で計算されます。

**p値(Prob>ChiSq)** カイ2乗検定の観測有意確率 ( $p$  値)。

### 推定値の共分散

パラメータ推定値の分散推定値および共分散推定値が表示されます。分散推定値の平方根は、「パラメータ推定値」表の「標準誤差」列に表示される数値と一致します。

**メモ:** 「推定値の共分散」レポートは、応答変数が名義尺度の場合のみ表示され、順序尺度の場合は表示されません。

---

## 「ロジスティック」プラットフォームのオプション

「ロジスティックであてはめ」の赤い三角ボタンのメニューには、ロジスティックプロットとあてはめのレポートに関するオプションがあります。

**メモ:** Y変数またはX変数を複数指定した場合は、[あてはめのグループ] メニューが表示されます。[あてはめのグループ] メニューのオプションを使うと、レポートの配置を変更したり、適合度の順に並べたりできます。『基本的な回帰モデル』を参照してください。

「ロジスティックであてはめ」の赤い三角ボタンをクリックすると、次のようなオプションが表示されます。

**オッズ比** （応答が2水準の場合のみ使用できます。）「パラメータ推定値」レポートにおいて、オッズ比の列を追加または削除します。『基本的な回帰モデル』を参照してください。

**逆推定** （応答が2水準の名義尺度の場合のみ使用できます。）与えられた応答変数の値（複数指定することも可）から、説明変数の値を推定します。「[逆推定](#)」を参照してください。

**ロジスティックプロット** ロジスティックプロットの表示／非表示を切り替えます。「[ロジスティックプロット](#)」を参照してください。

**プロットのオプション** ロジスティックプロットに関する以下のオプションがあります。

**点の表示** ロジスティックプロットの点の表示／非表示を切り替えます。

**応答率折れ線の表示** ロジスティックプロットにおける、応答率折れ線の表示／非表示を切り替えます。応答率折れ線は、X変数の各値に対して複数の点がある場合に役立ちます。各X値において複数の点がある場合には応答率を求めることができ、あてはめられたロジスティック曲線とそれらの応答率を比較することができます。度数が少なくて応答率が0か1になってしまうグループが多数生じるのを防ぐため、点が3つ以上あるX値のグループに対してだけ、応答率は表示されます。

**線の色** プロットの曲線の色を指定できます。

**ROC 曲線** モデルの受診者動作特性（ROC）曲線の表示／非表示を切り替えます。ROC 曲線は、Xの値ごとに、「1-特異度」に対して感度をプロットしたものです。「[ROC 曲線](#)」を参照してください。

**PR 曲線** 応答変数の各水準に対するPR 曲線（適合率-再現率曲線; Precision-Recall curve）の表示／非表示を切り替えます。PR 曲線は、異なる閾値における適合率と再現率をプロットしたものです。『予測モデルおよび発展的なモデル』を参照してください。

**リフトチャート** モデルのリフトチャートを表示します。リフトチャートは、モデルの予測能力を示すもので、データの部分ごとのリフト値をプロットしています。リフト値は一意な予測確率値ごとに求められます。予測確率が特定の閾値以上になっているデータ行の、全体に対する割合が、リフトチャートの横軸です。縦軸にプロットされるリフト値は、応答変数の興味があるほうの水準に関して、横軸の予測確率以上となっているグループで正しく分類されている割合を、全体の割合で割ったものです。リフトチャートの詳細については、『予測モデルおよび発展的なモデル』を参照してください。

**確率の計算式の保存** データテーブルに新たな列を作成し、モデルによって予測される確率の計算式を保存します。新しい列に保存される計算式は、次のとおりです。

- X 因子の線形結合式（ロジット）
- 各水準が生じる確率の予測式
- 最も生じる確率が高い応答変数の値を計算する予測式

以下のオプションの詳細については、『JMPの使用法』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMP レポートの設定に合わせて環境設定を変更したりできます。



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

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

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの [スクリプトの索引] を開いてください。また、[スクリプトの索引] には、この節で紹介されているオプションのスクリプト例もあります。

## ロジスティック分析レポート

「ロジスティック」プラットフォームでは、グラフを表示し、モデルをあてはめて連続量のX変数とカテゴリカルなY変数の関係を調べることができます。この節では、分析を実行する各オプションで作成されるレポートの内容を詳しく説明します。

- 「[ROC 曲線](#)」
- 「[逆推定](#)」

### ROC 曲線

「ロジスティック」プラットフォームには、ロジスティック回帰モデルの受診者動作特性（ROC）曲線をあてはめるオプションがあります。[ROC 曲線] オプションは、起動ウィンドウで指定された「イベントを示す水準」を**陽性**の応答水準としてROC曲線を作成します。「[ROC 曲線の例](#)」を参照してください。

ある疾患にかかっているかどうかを、次のようにX変数の値によって診断するとします。

- Xの値が特定の閾値より大きいときは、その疾患にかかっていると診断します。
- Xの値が閾値より小さいときは、その疾患にかかっていないと診断します。

たとえば、特定の種類のガンを診断する検査では、ある血液成分の濃度を測定します。診断における閾値をいろいろな値に変更し、それに伴って変化した偽陽性率と偽陰性率をプロットしてみれば、診断の良し悪しが分かります。分析の目的は、真陰性と真陽性を最も高める、X変数の閾値を求めることです。受診者動作特性（ROC）曲線を見ると、これらの値がどれだけ急激に変化するかがわかります。理想的な診断では、ROC曲線における曲線下の面積（AUC; Area Under Curve）が大きくなります。

次の2つは、医学の分野で使用されている概念です。

- **感度**は、ある状態が存在しているときに、その状態が存在していることを、与えられた X 変数の閾値によって正確に予測する確率です。Xの閾値が与えられているときに、状態が本当は存在しているのに存在していないと予測される確率は、「1-感度」です。
- **特異度**は、ある状態が存在していないときに、その状態が存在しないことを、診断によって正確に予測する確率です。



受診者動作特性（ROC）曲線は、X変数の値ごとに、「1-特異度」に対して感度をプロットしたものです。ROC曲線より下の面積（AUC）は、曲線によって示唆された情報を要約するための指標として使用されます。

検査による予測が完全であれば、1つの閾値を境にして、それより大きな確率予測値となっているすべての対象者で実際にある状態になっており、それより小さな確率予測値となっているすべての検査対象者で実際にある状態になっていない、というデータになっています。その場合、曲線は完璧な感度を持ち、グリッド上の(0,1)の点を通ります。ROC曲線がこの理想の点に近ければ近いほど、モデルによる判別能力は高くなります。逆に、予測することがまったくできない場合は、ROC曲線はグリッドの対角線上を通ります（DeLong et al. 1988）。

ROC曲線は、偽陽性率と偽陰性率の関係をグラフ化したものです。この関係は通常、曲線下の面積（AUC）に基づいて評価されます。この曲線下の面積（AUC）は、レポートにおいてプロットの下に表示されています。また、プロット内には、ROC曲線に接する45度の線が黄色で表示されます。この線と曲線の交点は、感度と特異度の和を最大にするカットオフ値を示しています。

## 逆推定

「ロジスティック」プラットフォームを使って逆推定を行えば、通常とは逆の方向での予測が可能です。逆推定では、X因子の値から応答（Y）の値を予測するのではなく、応答変数の値からX因子の値を予測します。ただし、ロジスティック回帰の場合、応答変数の値そのものではなく、応答変数の各水準が生じる確率を使用します。この機能は、応答が2水準の名義尺度の場合のみ使用できます。

「モデルのあてはめ」プラットフォームにも、逆推定およびその信頼区間を計算するオプションがあります。逆推定の詳細については、『基本的な回帰モデル』を参照してください。

---

メモ：「[逆推定の例](#)」も参照してください。

---

## ロジスティック回帰の別例

ここでは、「ロジスティック」プラットフォームを使った例をご紹介します。

- 「[順序ロジスティック回帰の例](#)」
- 「[ロジスティックプロットの例](#)」
- 「[ROC曲線の例](#)」
- 「[逆推定の例](#)」

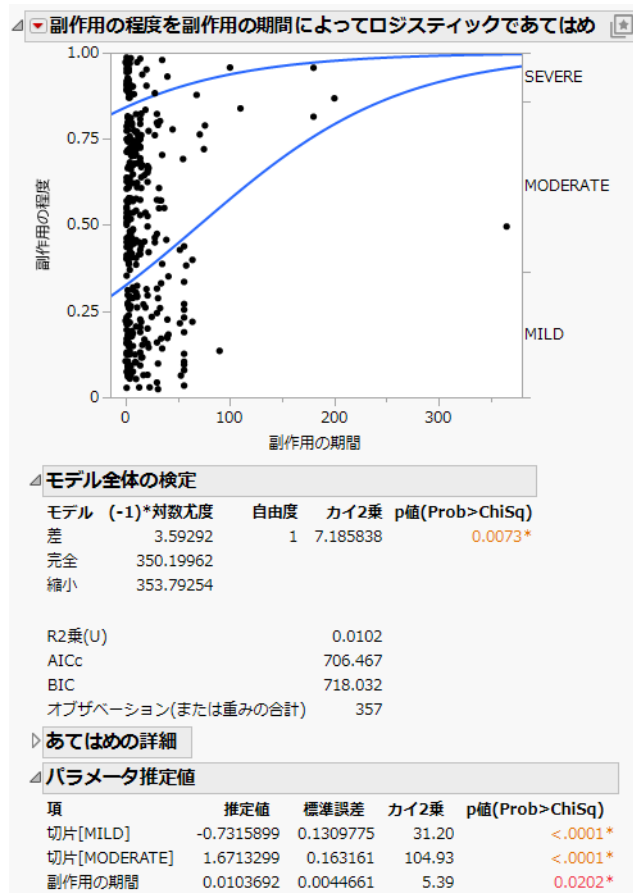
## 順序ロジスティック回帰の例

順序尺度の目的変数と、連続尺度の説明変数の関係を調べる方法をご紹介します。この例では、副作用の程度を、副作用の期間の関数として、順序ロジスティック回帰によりモデル化してみます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「AdverseR.jmp」を開きます。

2. 「副作用の程度」の左側のアイコンを右クリックし、順序尺度に変更します。
3. [分析] > [二変量の関係] を選択します。
4. 「副作用の程度」を選択し、[Y, 目的変数] をクリックします。
5. 「副作用の期間」を選択し、[X, 説明変数] をクリックします。
6. [OK] をクリックします。

図8.5 順序ロジスティックのレポートの例



このプロットのデータ点は、横軸ではX座標の値の位置に、縦軸ではY値の曲線の下にランダムな位置に描画されます。

「モデル全体の検定」レポートと「パラメータ推定値」レポートの詳細については、「[「ロジスティック」レポート](#)」を参照してください。「パラメータ推定値」レポートを見ると、切片パラメータは最後の応答水準を除いたすべての水準について計算されていますが、傾きパラメータは1つしかありません。応答水準間の間隔は切片パラメータによっても変わってきます。この例では、副作用の程度が大きくなるにつれて、切片パラメータも大きくなっています。

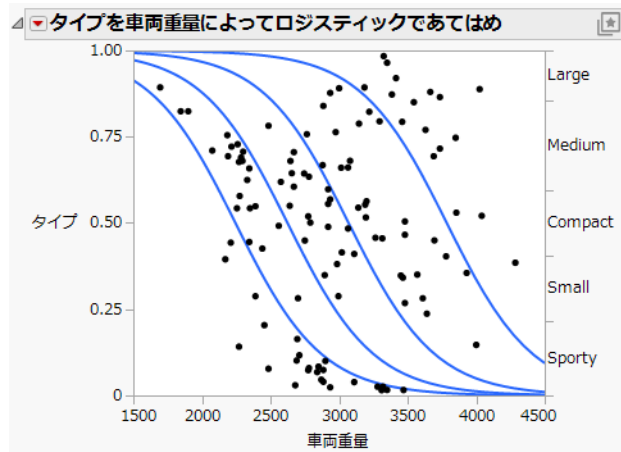
## ロジスティックプロットの例

この例では、車両重量によって、116台の車両のタイプを予測するとしましょう。車両のタイプは小さいものから順に、「Sporty」・「Small」・「Compact」・「Medium」・「Large」とします。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Physical Data.jmp」を開きます。
2. 「列」パネルで、「タイプ」の左側のアイコンを右クリックし、[順序尺度] を選択します。
3. 「タイプ」を右クリックし、[列情報] を選択します。
4. [列プロパティ] メニューから [値の表示順序] を選択します。
5. 上から順に「Sporty」・「Small」・「Compact」・「Medium」・「Large」となっていることを確認します。
6. [OK] をクリックします。
7. [分析] > [二変量の関係] を選択します。
8. 「タイプ」を選択し、[Y, 目的変数] をクリックします。
9. 「車両重量」を選択し、[X, 説明変数] をクリックします。
10. [OK] をクリックします。

レポートウィンドウが表示されます。

図8.6 「タイプ」を「車両重量」によってロジスティックであてはめたプロットの例



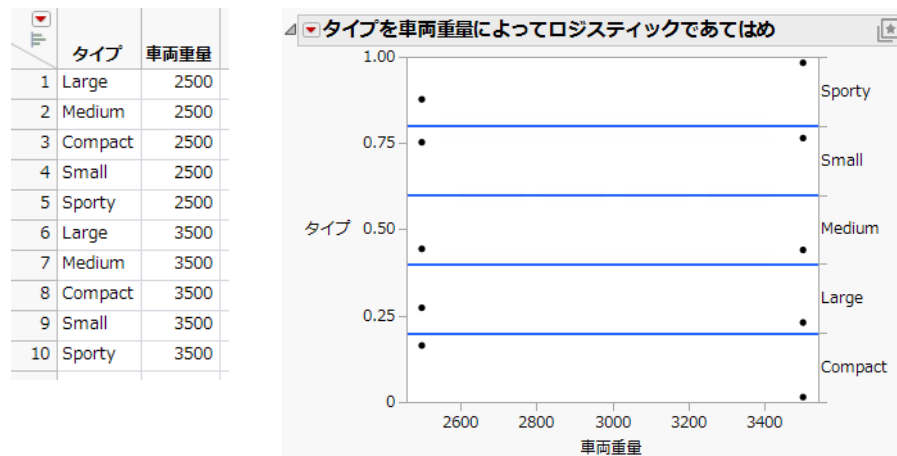
次の点を確認してください。

- 最初（一番下）の曲線は、所定の「車両重量」の車両が「Sporty」である確率を表しています。
- 2 番目の曲線は、車両が「Small」または「Sporty」である確率を表しています。1 番目と 2 番目の曲線の垂直距離が、車両が「Small」である確率です。
- プロットを見ると、予想どおり、重量が重い車両ほど「Large」である確率が高くなります。

- プロット上の点の X 座標はデータの値です。Y 座標は応答カテゴリの内側においてランダムにプロットされます。

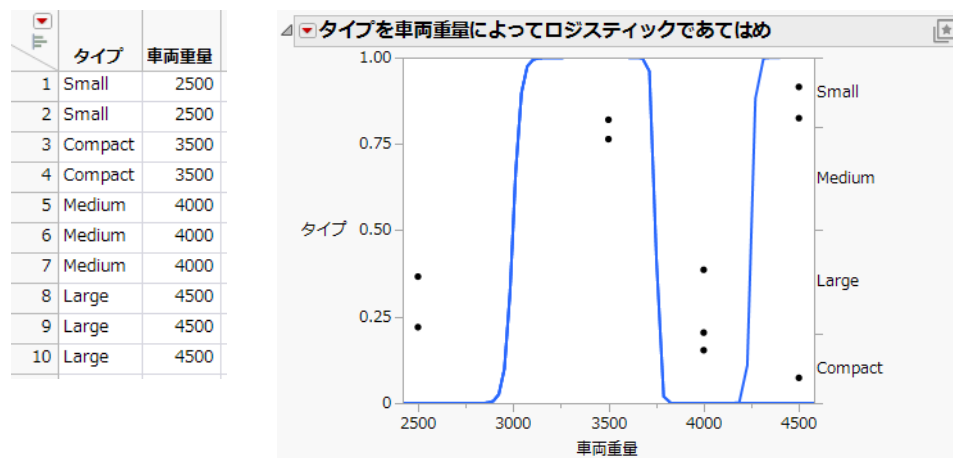
X 変数が応答変数に対して効果を持たない場合、ロジスティック曲線は水平になります。つまり、応答変数の各水準の確率が、説明変数の範囲全体にわたって一定になります。図 8.7 のロジスティックプロットが、ロジスティック曲線が水平となる例です。このような場合では、「タイプ」の確率を予測するのに、「車両重量」は役立ちません。

図 8.7 予測が無意味なデータとロジスティックプロットの例



説明変数の値から応答変数の値を完全に予測できる場合、ロジスティック曲線は事実上垂直になります。応答の予測値はどの説明変数の値においてもほぼ確か（確率がほとんど 1）です。図 8.8 のロジスティックプロットでは、「車両重量」が「タイプ」をほぼ完全に予測しています。

図 8.8 予測が完全となるデータとロジスティックプロット



この場合、パラメータ推定値が非常に大きくなり、回帰レポート上に「**不安定**」と表示されます。そのような場合には、[一般化線形モデル]において[Firthバイアス調整推定値]を使用することを検討してください。『基本的な回帰モデル』を参照してください。

---

**メモ:** 図8.7および図8.8のプロットを再現するには、図中のデータテーブルを作成してから、この節の冒頭で紹介した手順7～10を実行します。

---

## ROC 曲線の例

「ロジスティック」プラットフォームを使ってROC曲線をあてはめます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Penicillin.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「反応」を選択し、[Y, 目的変数] をクリックします。
4. 「Ln用量」を選択し、[X, 説明変数] をクリックします。  
「度数」が自動的に「度数」に設定されます。これは、「度数」の役割が、あらかじめ「度数」列に割り当てられているためです。
5. 「イベントを示す水準」から [Cured] を選択します。
6. [OK] をクリックします。
7. 「ロジスティックであてはめ」の赤い三角ボタンをクリックし、[ROC 曲線] を選択します。

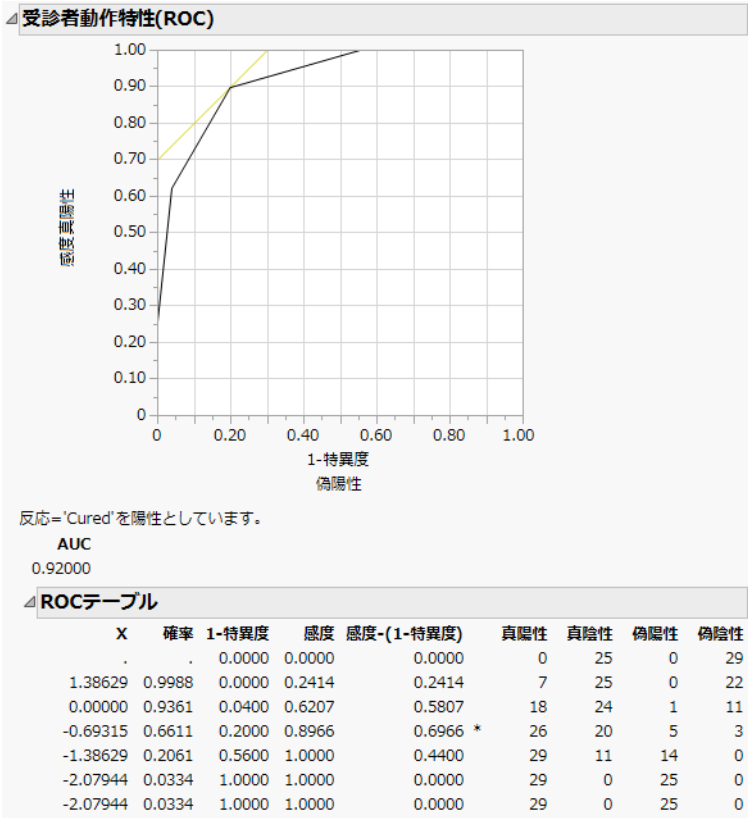
---

**メモ:** この例では、名義尺度の応答のROC曲線を示します。順序尺度のROC曲線の詳細については、『予測モデルおよび発展的なモデル』を参照してください。

---

この例では、「反応」を応答として、「Ln用量」を説明変数としています。応答を予測する確率が、前述した説明によって計算され、それらの結果に基づきROC曲線が描かれます。なお、「ROCテーブル」では、「感度 - (1 - 特異度)」が最も大きい行にアスタリスクが付けられます。この点に対応するX値が、感度と特異度の和を最大化する閾値となります。ROC曲線のプロットの黄色い線は、この点における接線です。

図 8.9 ROC 曲線と「ROC テーブル」の例



ROC 曲線がプロットの左上にあり、AUC が 0.5 を上回っているため、モデルの予測能力は高いと言えます。

逆推定の例

「ロジスティック」プラットフォームでは、十字ツールを使ってグラフ上で逆推定の大まかな値を求めることができます。応答の水準が 2 つのときは、[逆推定] オプションを選択すれば正確な逆推定が計算されます。逆推定では、Y 変数のイベントを示す水準の与えられた確率から、それに対応する X 変数の推定値と、その信頼区間を求めます。

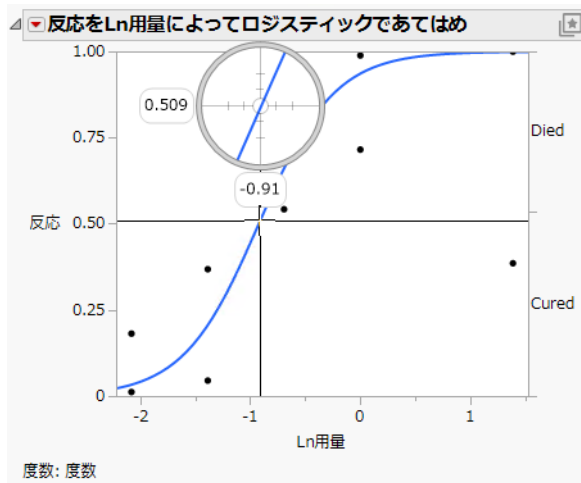
1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Penicillin.jmp」を開きます。
  2. [分析] > [二変量の関係] を選択します。
  3. 「反応」を選択し、[Y, 目的変数] をクリックします。
  4. 「Ln 用量」を選択し、[X, 説明変数] をクリックします。
- 「度数」が自動的に [度数] に設定されます。これは、[度数] の役割が、あらかじめ「度数」列に割り当てられているためです。

5. 「イベントを示す水準」から **[Cured]** を選択します。
6. **[OK]** をクリックします。

## 十字ツール

7. 十字ツールをクリックします。
8. ロジスティックプロットの左側の縦軸上で、(応答) 確率が0.5の位置に十字ツールを置きます。
9. 十字ツールのカーソルをロジスティック曲線のほうへ動かし、曲線と交わった点で、十字ツールの下に表示される「Ln用量」の値を読みます。

図8.10 ロジスティックプロット上の十字ツールの例



この例では、「Ln用量」の値が約-0.9のときに、治癒率と致死率が等しくなります。逆推定の正確な値を求めるには、「逆推定」オプションを使用します。

## 【逆推定】オプション

10. 「ロジスティックであてはめ」の赤い三角ボタンをクリックし、**【逆推定】** (図8.11) を選択します。
11. 「信頼水準」の入力ボックスに「0.95」と入力します。
12. 「信頼水準」のドロップダウンリストから**「両側」**を選択します。
13. 対象となる応答確率を指定します。この例では、「0.5」と「0.9」を入力します。つまり、治癒率が0.5と0.9であるときの「Ln用量」の値を求めます。

図 8.11 「逆推定」ウィンドウ

逆推定したい確率値を1つ以上、指定してください。

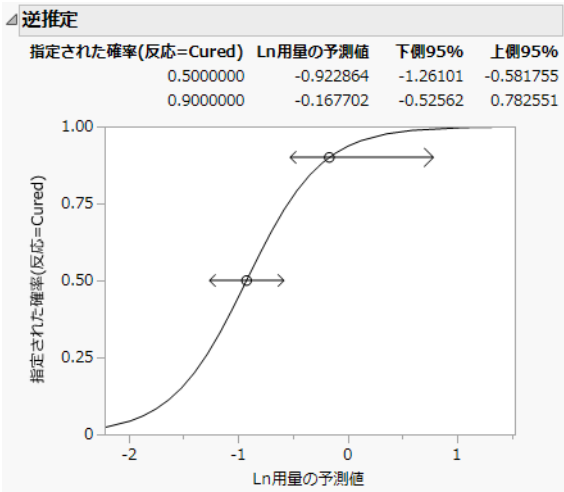
Ln用量 (予測対象)	信頼水準	確率(反応=Cured)
	0.95	0.5
		0.9
		.
		.
		.
		.
		.
		.

両側

OK キャンセル ヘルプ

14. [OK] をクリックします。

図 8.12 逆推定プロットの例



X変数の値の推定値とその信頼区間は、「逆推定」レポートに表とグラフの両方の形式で表示されます。たとえば、治癒率が90%になるときの「Ln用量」の両側信頼区間は、-0.526から0.783までとなっています。



---

## 「ロジスティック」プラットフォームの統計的詳細

ここでは、「モデル全体の検定」レポートの統計的詳細について説明します。

### カイ2乗

モデル全体に対する尤度比検定のカイ2乗統計量は、次式により計算されます。この統計量は、教科書では、ときどき、 $G^2$ という記号で表されています。

$$G^2 = 2(\sum -\ln p(\text{縮小モデル}) - \sum -\ln p(\text{モデル}))$$

ここでの総和は、全データに対する総和を表しています。

### R2乗(U)

R2乗(U)は次の式で計算されます。

$$\frac{\text{差の「(-1)*対数尤度」}}{\text{縮小モデルの「(-1)*対数尤度」}}$$

これらの数値は「モデル全体の検定」レポートに表示されます。

---

**メモ:** 上式によって計算された「R2乗(U)」は、「**McFaddenの疑似R2乗**」と呼ばれることもあります。

---



# 第9章

## 表の作成 対話的に集計表を作成する

「表の作成」プラットフォームでは、記述統計量の集計表（ピボットテーブル）を対話的に作成できます。簡単かつ柔軟な方法で、データを要約した表を作成できます。具体的には、データテーブルの列を集計表の行または列に割り当ててから、興味のある要約統計量を割り当てます。

図9.1 表の作成例

	性別											
	F						M					
	年齢						年齢					
	12	13	14	15	16	17	12	13	14	15	16	17
	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)
最小値	64	67	81	92	112	116	79	79	92	104	128	134
平均	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
最大値	145	112	142	112	115	116	128	105	119	128	128	172

		性別							
		女性				男性			
		既婚/未婚				既婚/未婚			
		既婚		未婚		既婚		未婚	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	サイズ	年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
ヨーロッパ	大型	34.0	7.071	28.0	.	.	26.0	.	.
	中型	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	小型	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
日本	大型	25.0	.	.	.	32.0	.	.	.
	中型	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	小型	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
米国	大型	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	中型	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	小型	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
生産国									
ヨーロッパ		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
日本		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781
米国		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179

		平均		
タイプ	会社規模	利益(\$M)	売上(\$M)	従業員一人あたりの利益
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
Pharmaceutical	すべて	240.87	5652.02	6159.015
	big	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
	small	156.95	1083.75	38337.19
	すべて	690.08	5070.25	23546.12
すべて		409.32	5433.86	12679.18
タイプ				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
すべて		409.32	5433.86	12679.18

## 目次

「表の作成」プラットフォームの使用例 .....	293
「表の作成」プラットフォームの起動 .....	298
「表の作成」プラットフォームのダイアログウィンドウ .....	300
統計量の追加 .....	301
「表の作成」プラットフォームのレポート .....	304
分析列 .....	305
グループ列 .....	305
列テーブルと行テーブル .....	306
表の編集 .....	307
「表の作成」プラットフォームのオプション .....	307
テスト集計パネルの表示 .....	308
列のポップアップメニュー .....	309
「表の作成」プラットフォームの別例 .....	310
さまざまな表を作成して配置を変更する例 .....	310
複数のデータ列に対する表の例 .....	315
ページ列の使用例 .....	317
グループ列を積み重ねる例 .....	319
一意のID列を使用した例 .....	320
グループ列をパッキングする例 .....	322

## 「表の作成」プラットフォームの使用例

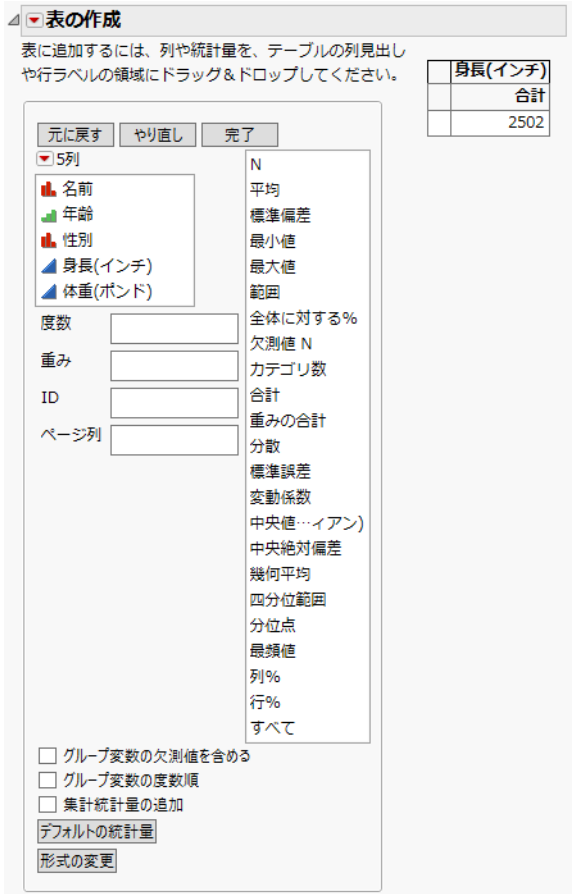
ここでは、データの要約を表形式にまとめる方法を紹介します。男子生徒と女子生徒の身長を測定したデータがあります。このデータを基に、男子生徒と女子生徒それぞれの、および、全体の平均身長を示す表を作成します。

図9.2 平均身長を示す表

	身長(インチ)
性別	平均
F	60.9
M	63.9
すべて	62.6

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [表の作成] を選択します。  
「身長(インチ)」が分析対象の変数であるため、これを表の一番上に表示します。
3. 「身長(インチ)」をクリックし、「列のドロップゾーン」にドラッグします。

図9.3 「身長」変数の追加後



性別ごとに要約したいので、性別を表の左側に配置します。

4. 「性別」をクリックしてドラッグし、数値「2502」の横にある空のセルにドロップします。

図9.4 「性別」変数の追加後

**表の作成**

表に追加するには、列や統計量を、テーブルの列見出し  
や行ラベルの領域にドラッグ&ドロップしてください。

性別	身長(インチ)
F	1096
M	1406

元に戻す

やり直し

完了

▼ 5列

名前

年齢

性別

身長(インチ)

体重(ポンド)

度数

重み

ID

ページ列

☐ グループ変数の欠測値を含める  
☐ グループ変数の度数順  
☐ 集計統計量の追加  

デフォルトの統計量

形式の変更

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値…イアン)

中央絶対偏差

幾何平均

四分位範囲

分位点

最頻値

列%

行%

すべて

合計の代わりに平均を表示します。

5. [平均] をクリックしてドラッグし、「合計」の上にドロップします。

図9.5 [平均] の追加後

表の作成

表に追加するには、列や統計量を、テーブルの列見出し  
や行ラベルの領域にドラッグ&ドロップしてください。

元に戻す やり直し 完了

▼ 5列

名前

年齢

性別

身長(インチ)

体重(ポンド)

度数

重み

ID

ページ列

☐ グループ変数の欠測値を含める

☐ グループ変数の度数順

☐ 集計統計量の追加

デフォルトの統計量

形式の変更

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値…イアン)

中央絶対偏差

幾何平均

四分位範囲

分位点

最頻値

列%

行%

すべて

	身長(インチ)
性別	平均
F	60.9
M	63.9

さらに、男子と女子を合わせた全体の平均も表示します。

6. [すべて] をクリックしてドラッグし、「性別」の上にドロップします。または、[集計統計量の追加] チェックボックスをオンにします。



図9.6 「すべて」の追加後

**表の作成**

表に追加するには、列や統計量を、テーブルの列見出しや行ラベルの領域にドラッグ&ドロップしてください。

元に戻す

やり直し

完了

▼ 5列

名前

年齢

性別

身長(インチ)

体重(ポンド)

度数

重み

ID

ページ列

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値…イアン)

中央絶対偏差

幾何平均

四分位範囲

分位点

最頻値

列%

行%

すべて

☐ グループ変数の欠測値を含める  
☐ グループ変数の度数順  
☐ 集計統計量の追加  

デフォルトの統計量

形式の変更

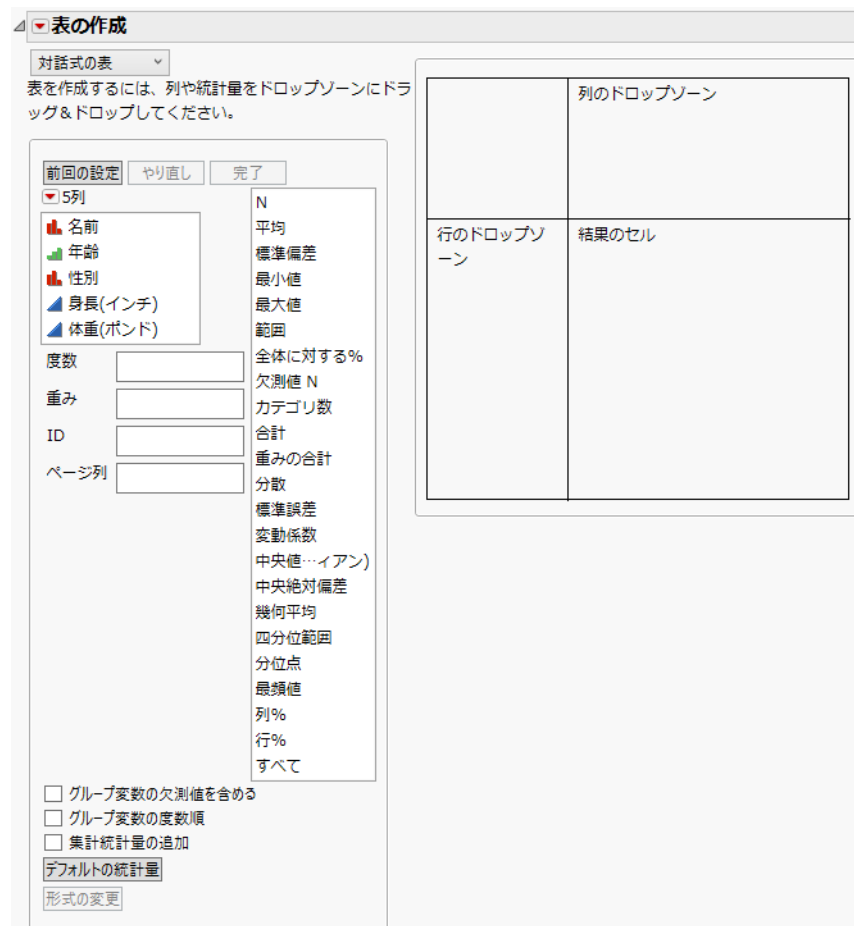
7. (オプション) **完了** をクリックします。

完成した表に、女子と男子の平均身長と全体の平均身長が表示されます。

## 「表の作成」プラットフォームの起動

「表の作成」プラットフォームを起動するには、[分析] > [表の作成] を選択します。

図9.7 「表の作成」の対話的な表



赤い三角ボタンのメニューのオプションについては、「[「表の作成」プラットフォームのオプション](#)」を参照してください。「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。

「表の作成」ウィンドウには、次のオプションがあります。

**対話式の表／ダイアログボックス** 2つのモード間で切り替えます。項目をドラッグ&ドロップしながら独自の表を作成していく場合は、[対話式の表] を、決まった形式の簡単な表を作成する場合は [ダイアログボックス] を選択してください。「[「表の作成」プラットフォームのダイアログウィンドウ](#)」を参照してください。

**〔統計量〕オプション** 標準的な統計量のリストが表示されます。統計量をリストから表にドラッグすれば、その統計量が表に組み込まれます。[「統計量の追加」](#)を参照してください。

**列のドロップゾーン** 列または統計量をドラッグしてこのゾーンにドロップし、列を作成します。

---

**メモ:** 〔統計量〕オプションに表示されているものと同じ名前の列がデータテーブルにある場合は、必ずリストから列名をドラッグ&ドロップしてください。そうしないと、表で同名の統計量が代用される場合があります。

---

**行のドロップゾーン** 列または統計量をドラッグしてこのゾーンにドロップし、行を作成します。

---

**ヒント:** 列のリストから1つまたは複数の列を選択するか、統計量を1つまたは複数選択し、Altキー（macOSの場合はOptionキー）を押しながらドロップゾーンをクリックすることによっても、行または列を作成できます。

---

**結果のセル** ドラッグ&ドロップした列や統計量に基づいて、結果が各セルに表示されます。

**度数** この役割を割り当てた列の値は、各行の度数（繰り返し数）を表します。このオプションは、分析対象のデータが要約されたものである場合に、度数を割り当てのに便利です。

**重み** データに重み（重要度、影響度など）をつけるのに使うデータテーブル列を指定します。

**ID** 一意な値の識別子となるデータテーブル列を指定します。[「一意のID列を使用した例」](#)を参照してください。

**ページ列** 名義尺度または順序尺度の列のカテゴリごとに個別の表を作成します。[「ページ列の使用例」](#)を参照してください。

**グループ変数の欠測値を含める** グループ変数にある欠測値を1つのグループとして集計した結果を、表に追加します。これをチェックしない場合、欠測値は表に含まれません。列プロパティとして定義した欠測値のコードも考慮されます。

**グループ変数の度数順** 度数の大きい順に並べ直してグループを表示します。

**集計統計量の追加** すべての行と列の集計統計量を追加します。

**デフォルトの統計量** 分析列または非分析列（グループ変数など）をドラッグ&ドロップしたときにデフォルトで表示される統計量を変更できます。

**形式の変更** 特定の統計量の表示形式を変更できます。[「数値の表示形式の変更」](#)を参照してください。

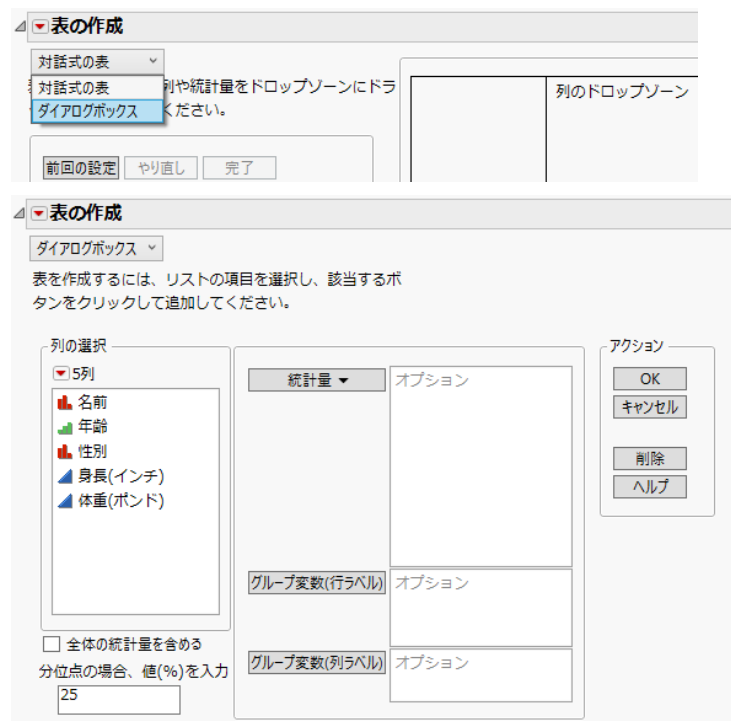
**プロットスケールの変更** （赤い三角ボタンのメニューで **〔チャートの表示〕** が選択されている場合にのみ表示されます。）統一した独自のスケール（カスタムスケール）を指定できます。

**プロットスケールの統一** （赤い三角ボタンのメニューで **〔チャートの表示〕** が選択されている場合にのみ表示されます。）このチェックボックスをオフにすると、各列のデータに基づき、それぞれの棒グラフのスケールが決められます。

## 「表の作成」プラットフォームのダイアログウィンドウ

「表の作成」プラットフォームでは、ドラッグ&ドロップ操作で対話式の表を作成する代わりに、ダイアログボックスを使って簡単な表を作成することもできます。この場合、[分析] > [表の作成] の選択後に、メニューから[ダイアログボックス]を選択してください(図9.8を参照)。表の作成後、必要であれば赤い三角ボタンのメニューから[設定パネルの表示]を選択し、新しい項目を表にドラッグ&ドロップして、より複雑な表にすることもできます。

図9.8 ダイアログボックスの使用



ダイアログボックスを使用したときの「表の作成」ウィンドウには、次のオプションがあります。

**全体の統計量を含める** グループ列のカテゴリに対して、それらを集計した要約統計量も表示します。

**分位点の場合、値(%)を入力** 求めたい分位点の累積確率をパーセントで入力します。たとえば、75%の分位点を求めた場合、データの75%がその分位点より小さくなっています。ここで指定した設定は、すべてのグループ列に適用されます。

**統計量** 列を選択したら、その列に関して求めたい統計量をこのリストから選択します。「統計量の追加」を参照してください。

**グループ変数(行ラベル)** 行ラベルとして使用する列を選択します。

**グループ変数(列ラベル)** 列ラベルとして使用する列を選択します。

## 統計量の追加

「表の作成」プラットフォームでは、設定パネルに表示される標準的な統計量を、表に追加していくことができます。列の操作と同じく、任意の統計量をリストから表にドラッグできます。次のことを注意してください。

---

**ヒント:** 列と統計量の両方を同時に選択して、表にドラッグできます。

---

- グループ列によって定義されるカテゴリに属する全データから、分析列の値に基づいて統計量が計算されます。
- 統計量も、表の行と列のどちらにでも指定できますが、どちらかに統一しておく必要があります。
- 連続尺度の列をデータ領域にドラッグすると、分析列として扱われます。

---

**メモ:** 分析列は、統計量の計算に使用される列であり、連続尺度の数値列でなければいけません。「[分析列](#)」を参照してください。

---

「表の作成」では、次の統計量を使用できます。

**N** その列にある欠測値でないデータの数です。分析列がない場合のデフォルトの統計量です。

**平均** 欠測値以外のデータ値から計算される算術平均です。つまり、欠測値以外の値の合計を【**重みの合計**】で割った値です（なお、重みが定義されている場合は、合計を求める際に、その【**重み**】変数の値で重み付けます）。

**標準偏差** 欠測値以外のデータ値から計算される標準偏差です。つまり、分散の平方根です。

**最小値** 欠測値以外のデータ値における最小値です。

**最大値** 欠測値以外のデータ値における最大値です。

**範囲** 【**最大値**】と【**最小値**】の差です。

**全体に対する %** データ全体の合計に対する割合を計算します。分母はすべての観測値の合計で、分子はカテゴリごとの合計です。分析列がない場合、[全体に対する %] は度数の合計に対する割合になります。分析列がある場合、[全体に対する %] は分析列の全合計に対する割合になります。つまり、分母はすべてのデータ値の合計で、分子はカテゴリに属するデータ値の合計になります。表に統計量のキーワードをドラッグすると、異なる割合を求めることができます。

- [全体に対する %] の見出しに、表の中から1つまたは複数のグループ列をドロップすると、分母の定義が変わります。このときは、グループごとの合計が分母に使われます。
- 表の行と列にグループ列を指定して2元表を作成したとき、列合計に対する割合を取得したい場合は、表の行に配置されたすべてのグループ列をドラッグして、表の【**全体に対する %**】の見出しにドロップします（この操作による結果は、[列 %] を用いた場合と同じです）。同様に、行合計に対する割合を取得したい場合は、表の列のすべてのグループ列をドラッグして、表の【**全体に対する %**】の見出しにドロップします（この操作による結果は、[行 %] を用いた場合と同じです）。

**欠測値 N** 欠測値の個数です。

**カテゴリ数** 分析列にあるカテゴリの個数です。

**合計** 欠測値以外のデータ値の合計です。表に他の統計量が表示されていない場合の、分析列のデフォルトの統計量です。

**重みの合計** 列内のすべての「重み」の合計です。「重み」に割り当てられた列がない場合、「**重みの合計**」は欠測値以外の値の個数になります。

**分散** 欠測値以外のデータ値から計算される分散（不偏分散）です。つまり、平均からの偏差を2乗したものの合計を、非欠測値の個数から1を引いた数で割った値です。

**標準誤差** 平均の標準誤差です。つまり、標準偏差をNの平方根で割った値です。列に「重み」役割が割り当てられている場合、分母には重みの和の平方根が使われます。

**変動係数** ばらつきの度合いを示す指標で、標準偏差を平均で割り、100を掛けて求められます。

**中央値(メディアン)** 50%のパーセント点です。下半分のデータが中央値より小さく、上半分のデータが中央値より大きくなっています。

**中央絶対偏差** 「中央値からの偏差の絶対値」の中央値。

**幾何平均** データの積の $n$ 乗根。利子の計算などによく使われます。データに特に大きな値が含まれていて分布が歪んでいる場合などに、便利です。

---

**メモ:** データに負の値がある場合、幾何平均は欠測値になります。また、データに0の値がある場合（かつ負の値がない場合）は0になります。

---

**四分位範囲** 第3四分位点と第1四分位点の差です。

**分位点** 指定したパーセントに対する分位点を求めます。たとえば、75%の分位点を求めた場合、データの75%がその分位点より小さくなっています。異なる分位点を指定するには、「**分位点**」をクリックして表にドラッグし、表示されるボックスに分位点のパーセントを入力します。

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

**列%** 分析列がない場合は、セルの度数が列の合計度数に占める割合が表示されます。分析列がある場合、「**列%**」は、分析列の値の、列合計に対する割合になります。統計量を上部に配置した表の場合、複数の行がある表（統計量が縦に並べて表示されている表）に「**列%**」を追加できます。

**行%** 分析列がない場合は、セルの度数が行の合計度数に占める割合が表示されます。分析列がある場合、「**行%**」は、分析列の値の、行合計に対する割合になります。統計量を左側に配置した表の場合、複数の列がある表（統計量が横に並べて表示されている表）に「**行%**」を追加できます。

**すべて** グループ列のカテゴリに対して、それらを集計した要約統計量も表示します。

## 数値の表示形式の変更

各セルの表示形式は、分析列と統計量によって決まります。度数の場合、デフォルトでは小数点以下は表示されません。統計量の場合、分析列および指定した統計量の種類に基づき、適切な表示形式に設定されます。デフォルトの表示形式を変更するには、次の手順を行います。

1. 「表の作成」ウィンドウの下部にある **【形式の変更】** ボタンをクリックします。
2. 表示されるパネルで、フィールドの幅、カンマ、小数点以下の桁数を入力します（図9.9）。
3. セル値をパーセント形式で表示するには、小数点以下の桁数の後にカンマを追加し、「**パーセント**」と入力します。
4. （オプション）最適な表示形式を自動的に設定させたい場合には、テキストボックスに「**最適**」と入力します。

すると、各セル値の精度を考慮した上で、その値を表示するのに最適な表示形式が選ばれます。

5. 変更内容を確定し、「形式」セクションを閉じるには **【OK】** をクリックします。または、[形式の設定] をクリックすると、「**形式**」セクションが開いたまま、変更内容が反映されます。

図9.9 数値の表示形式の変更

形式

☐ 同じ小数形式を適用

特定の統計量の表示形式を変更できます。各形式は、フィールド幅と小数桁数の2つの整数で構成されます。「最適」形式を適用するには、2番目の整数の代わりに'Best'というキーワードを入力します。「パーセント」形式を指定するには、2番目の整数の後に'Percent'というキーワードを入力します。

N

9, 最適

形式の設定 キャンセル OK

## 「表の作成」プラットフォームのレポート

「表の作成」プラットフォームでは、複数の表が上下左右に連結された表を表示できます。列テーブルまたは行テーブルが1つだけの表も作成可能です。

図9.10 「表の作成」の出力

	性別											
	F						M					
	年齢						年齢					
	12	13	14	15	16	17	12	13	14	15	16	17
	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)	体重(ポンド)
最小値	64	67	81	92	112	116	79	79	92	104	128	134
平均	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
最大値	145	112	142	112	115	116	128	105	119	128	128	172

		性別							
		女性				男性			
		既婚/未婚				既婚/未婚			
		既婚		未婚		既婚		未婚	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	サイズ	年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
ヨーロッパ	大型	34.0	7.071	28.0	.	.	26.0	.	.
	中型	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	小型	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
日本	大型	25.0	.	.	.	32.0	.	.	.
	中型	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	小型	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
米国	大型	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	中型	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	小型	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
生産国		年齢	年齢	年齢	年齢	年齢	年齢	年齢	年齢
ヨーロッパ		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
日本		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781
米国		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179

		平均			
タイプ	会社規模	利益(\$M)	売上(\$M)	従業員一人あたりの利益	
Computer	big	1089.9	20597.48	4530.478	
	medium	-85.75	3018.85	-3462.51	
	small	44.94	1758.06	7998.815	
	すべて	240.87	5652.02	6159.015	
Pharmaceutical	big	894.42	7474.04	17140.70	
	medium	698.98	4261.06	24035.11	
	small	156.95	1083.75	38337.19	
	すべて	690.08	5070.25	23546.12	
すべて	すべて	409.32	5433.86	12679.18	
タイプ		利益(\$M)	売上(\$M)	従業員一人あたりの利益	
Computer		240.87	5652.02	6159.015	
Pharmaceutical		690.08	5070.25	23546.12	
すべて		409.32	5433.86	12679.18	

表を対話的に作成する際は、次のような作業を繰り返します。

- 項目（列または統計量）を適切なリスト内でクリックし、（行または列の）ドロップゾーンまでドラッグします。「[表の編集](#)」および「[列テーブルと行テーブル](#)」を参照してください。
- ドラッグ&ドロップ操作を繰り返して表に適宜項目を追加します。表は、項目を追加するたびに更新されます。すでに列や行の見出しがある場合は、既存の項目を基準にして別の項目を追加していくことができます。

クリック操作やドラッグ操作の際には、次の点に注意してください。

- JMP では、データ列の尺度を基に列の役割を判断します。連続尺度のデータ列は分析列と想定されます。「[分析列](#)」を参照してください。順序尺度または名義尺度のデータ列はグループ列と想定されます。「[グループ列](#)」を参照してください。
- 複数の列を表にドラッグ&ドロップした場合は、次のように処理されます。
  - ドラッグ&ドロップされた複数のデータ列が共通のデータ値を持っている場合は、各列の列名とそれらの列が持つカテゴリから構成される1つの集計表が生成されます。この時、各セルは、指定された各列と、それらの列の各値から構成されます。



- ドラッグ&ドロップされた複数の列が共通のデータ値をもっていない場合は、個別の表に分かれます。
- このデフォルトの結果は、列を右クリックして **【テーブルの結合】** または **【テーブルを分離】** を選択すれば変更できます。**「列のポップアップメニュー」** を参照してください。
- 複数のデータ列を表において入れ子状にする場合は、空の表に第1列をドラッグ&ドロップしてから、追加の列をその第1列にドラッグ&ドロップします。
- 適切に作成された表では、グループ列、分析列、統計量がそれぞれまとまって配置されます。分析列のリストの中に統計量が配置されたり、グループ列のリストの中に分析列が挿入されたりすることはありません。
- 「表の作成」上の設定パネルだけではなく、データテーブルのテーブルパネルからも、「表の作成」の表にデータ列をドラッグすることができます。

---

**メモ:** 開いているデータテーブルで、データを追加したり、行を削除したり、データを再コード化したりした場合は、「表の作成」の表も更新されます。

---

## 分析列

「表の作成」プラットフォームの分析列は、統計量の計算に使用する任意の数値列であり、連続尺度である必要があります。グループ列によって区別されるカテゴリごとに、分析列の統計量が計算されます。

---

**メモ:** 分析列は、表の行と列のどちらにでも指定できますが、どちらかに統一しておく必要があります。

---

## グループ列

「表の作成」プラットフォームにおいてデータをカテゴリ別に分類するために使う列（名義尺度または順序尺度）を、グループ列と呼びます。文字、整数値、小数値などを使用できますが、カテゴリ数は制限する必要があります。

次のことを注意してください。

- 複数のグループ列が入れ子状になっている場合は、それらのグループ列の組み合わせによって各組み合わせが作成されます。たとえば、「性別」と「既婚/未婚」の2つのグループ列がある場合、「女性で既婚」、「女性で未婚」、「男性で既婚」、「男性で未婚」の4つの組み合わせを持つ表が作成されます。
- グループ列が入れ子状になっている場合は、入れ子になっている値を積み重ねて表の横方向のスペースを節約することができます。グループ列を積み重ねるには、グループ列を右クリックし、**「グループ列の積み重ね」** オプションを選択します。入れ子になっているグループ列の組み合わせによって積み重ねが行われます。
- グループ列は、表の列と行のどちらでも使用できます。これらのグループ列を組み合わせ、表のセルが決まります。
- グループ列に欠測値がある場合、そのデータはデフォルトでは表に含まれません。**「グループ変数の欠測値を含める」** オプションをクリックすると、それらのデータを含めることができます。

- なお、欠測値として扱うコードやデータ値を指定するには、「欠測値のコード」列プロパティを使用します。[グループ変数の欠測値を含める] オプションをクリックすると、「欠測値のコード」に指定したデータも含まれます。欠測値のコードの詳細については、『JMPの使用法』を参照してください。

列テーブルと行テーブル

「表の作成」プラットフォームの1つの表は、列と列の見出しを基準として作成されます。ここでは、これらの見出しを**行テーブル**および**列テーブル**と呼ぶことにします (図9.11)。

「表の作成」プラットフォームの行テーブルと列テーブルの例

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
3. 「サイズ」をドラッグし、「行のドロップゾーン」にドロップします。
4. 「生産国」をドラッグし、「サイズ」見出しの左側にドロップします。
5. [平均] をドラッグし、「N」見出しの上にドロップします。
6. [標準偏差] をドラッグし、「平均」見出しの下にドロップします。
7. 「年齢」をドラッグし、「平均」見出しの上にドロップします。
8. 「タイプ」をドラッグし、表の右端にドロップします。
9. 「性別」をドラッグし、表の下にドロップします。

図9.11 行テーブルと列テーブル

2つの列テーブル

生産国	サイズ	年齢		タイプ		
		平均	標準偏差	スポーツ	ファミリー	ワーク
ヨーロッパ	大型	30.5	5.802	0	1	3
	中型	30.9	6.194	8	9	0
	小型	30.8	5.388	13	5	1
日本	大型	28.5	4.95	0	1	1
	中型	30.2	4.668	15	32	7
	小型	29.7	5.928	41	33	18
米国	大型	33.8	6.167	1	28	7
	中型	31.1	6.976	14	35	4
	小型	30.4	5.846	8	11	7
性別						
女性		30.6	6.386	41	76	21
男性		30.8	5.643	59	79	27

2つの行テーブル

複数の列テーブルがある場合は、表の左側にある見出しを複数の列テーブルが共有します。この例では、「生産国」と「性別」が共有されています。同様に、複数の行テーブルがある場合は、表の上部にある見出しを複数の行テーブルが共有します。この例では、「年齢」と「タイプ」の両方が2つの行テーブル間で共有されています。

## 表の編集

「表の作成」プラットフォームでは、表に追加した項目を編集することができます。具体的には、項目の削除、列ラベルの削除、統計量とラベルの編集を行います。

### 項目の削除

表に項目を追加した後で、次のいずれかの方法で項目を削除できます。

- 項目をドラッグし、表の外側にドロップします。
- 最後に追加した項目を削除するには、**[元に戻す]** をクリックします。
- 項目を右クリックし、**[削除]** を選びます。

### 列ラベルの削除

グループ列を表の上部に配置した場合、グループ列の各カテゴリの上に、そのグループ列の列名が表示されますが、この列名を削除することもできます。列テーブルから列名を削除するには、列名を右クリックし、**[列ラベルの削除]** を選びます。列ラベルを再度挿入するには、カテゴリのいずれかを右クリックし、**[列ラベルの復元]** を選びます。

### 統計量の名前・ラベルの編集

統計量の名前やラベルを編集できます。たとえば、「平均」の代わりに「平均値」とすることが可能です。編集する項目を右クリックし、**[項目ラベルの変更]** を選びます。表示されるボックスに、新しいラベルを入力します。または、名前やラベルをダブルクリックして編集ボックスへ直接入力することもできます。

ある統計量の名前を、別の既存の統計量名に変更した場合、実際に計算される統計量もその別の統計量になるので注意してください。別の既存の統計量の名前にすることは、統計量をいったん表から削除し、その別の統計量を追加するのとまったく同じです。

---

## 「表の作成」プラットフォームのオプション

「表の作成」の赤い三角ボタンをクリックすると、メニューに次のオプションが表示されます。

**設定パネルの表示** 対話的に操作するための設定パネルを表示します。

**表の表示** 要約データを表形式で表示します。

**チャートの表示** 表で示されている要約統計量を棒グラフで表示します。この棒グラフによって、要約統計量の相対的な大きさを視覚的に比較できます。デフォルトでは、すべての列の棒グラフに同じスケールが適用されます。**[プロットスケールの統一]** チェックボックスをオフにすると、各列のデータに基づき、それぞれの棒グラフのスケールが決まります。**[プロットスケールの変更]** ボタンを使用して、統一したカスタムスケールを指定できます。チャートの0の線は、左端か中央に位置します。データがすべて負に

なっている場合、もしくはデータがすべて正となっている場合には、0を基準線としてチャートが作成され、それ以外の場合は、チャートの中央が0になります。

**濃淡の表示** 表内に複数の行がある場合に、グレーの濃淡を適用します。

**ツールヒントの表示** 表上にカーソルを合わせると、関連項目のヒントが表示されます。

**テスト集計パネルの表示** 元のデータから無作為抽出した値を使ってテスト集計を行うための、テスト集計パネルを表示します。データ量が膨大である場合に特に便利です。「[テスト集計パネルの表示](#)」を参照してください。

**データテーブルに出力** 表からデータテーブルを作成します。行テーブルが複数ある場合は、それぞれの行テーブルで構成が異なるかもしれませんが、行テーブルごとに1つのデータテーブルが作成されます。

**フルパスの列名** 作成されるデータテーブルの列名に、グループ列のフルパス（列名=水準名の形式）を使用します。

以下のオプションの詳細については、『JMPの使用法』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMPレポートの設定に合わせて環境設定を変更したりできます。

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

---

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの[スクリプトの索引]を開いてください。また、[スクリプトの索引]には、この節で紹介されているオプションのスクリプト例もあります。

---

「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。

## テスト集計パネルの表示

「表の作成」プラットフォームで大量のデータを使用する場合、そこから抽出した一部のデータを使ってさまざまな表のレイアウトを試し、最適なものを見つけたいことがあるでしょう。そのような場合において、JMPでは、指定された標本サイズだけのデータを無作為抽出した後、表を作成することもできます。テスト集計機能を使うには、次の手順を行います。

1. 「表の作成」の赤い三角ボタンをクリックし、[\[テスト集計パネルの表示\]](#)を選択します。

2. 図9.12に示すように、「**標本サイズ(>1)または標本抽出率(<1)**」の下のボックスに、標本サイズを入力します。標本サイズは、アクティブなデータテーブルに対する割合、または無作為抽出するデータの行数のどちらかとして認識されます。

図9.12 テスト集計パネル

3. [リサンプリング] をクリックします。
4. JMPデータテーブルでサンプリングされたデータを確認するには、[テストデータの表示] ボタンをクリックします。テスト集計パネルを非表示にすると、データテーブルのすべてのデータを使って表が再生成されます。

## 列のポップアップメニュー

「表の作成」レポートの表で列を右クリックすると、次のオプションが表示されます。

**削除** 選択した列を削除します。

**列のパッキング** (統計量の列を右クリックすると表示されます。) オプションを使って表の統計量の列を結合し、コンパクトな表示にすることができます。「[グループ列をパッキングする例](#)」を参照してください。

**バック** 選択されている統計量の列を1列にまとめます。

**アンパック** 選択されている統計量を個別の列に分割します。

**テンプレート** 1列にまとめられた列の、統計量の表示形式を定義することができます。[バック] をクリックした時点で選択されている列のうち、一番左の列を ^FIRST、それ以外の列を ^OTHERS として指定します。また、括弧内で統計量を区切る区切り文字も指定できます。

**グループ変数として使用** 分析列をグループ列に変更します。

**分析列として使用** グループ列を分析列に変更します。

**項目ラベルの変更** (個別の列または入れ子状の列の場合にのみ表示されます。) 新しいラベルを入力できます。

**テーブルの結合 (カテゴリごとの列)** (個別の列または入れ子状の列の場合にのみ表示されます。) 個々の列または入れ子状の列を結合します。「[複数のデータ列に対する表の例](#)」を参照してください。

**グループ変数を入れ子にする** グループ変数を縦または横に入れ子にします。

**テーブルを分離** (結合したテーブルの場合にのみ表示されます。) 列ごとに個別の表を作成します。

**列ラベルの削除** 列テーブルから列名を削除します。

**列ラベルの復元** 非表示になっていた列名を列テーブルに表示します。

**度数順** グループ列の各水準を、度数が大きい順に並べます。このオプションは、特定のグループ列について、[グループ変数の度数順] の設定を無効にするためにも使用できます。

**グループ列の積み重ね** (複数のグループ列がある場合のみ使用できます。) 入れ子状になっているグループ列の値が結合した形で表示され、表の横幅が狭くなります。値のインデントは、グループ列の階層を示します。「[グループ列を積み重ねる例](#)」を参照してください。

---

## 「表の作成」プラットフォームの別例

ここでは、「表の作成」プラットフォームを使った例をご紹介します。

- 「[さまざまな表を作成して配置を変更する例](#)」
- 「[複数のデータ列に対する表の例](#)」
- 「[ページ列の使用例](#)」
- 「[グループ列を積み重ねる例](#)」
- 「[一意のID列を使用した例](#)」
- 「[グループ列をパッキングする例](#)」

### さまざまな表を作成して配置を変更する例

この例では、「表の作成」プラットフォームを使い、次の手順で表の作成と編集を行います。

1. 「[度数表を作成する](#)」
2. 「[統計量の表を作成する](#)」
3. 「[表の配置を変更する](#)」

#### 度数表を作成する

日本車・ヨーロッパ車・米国車の所有者数を表示する表を作成してみましょう。これらの所有者数を車のサイズ別に集計してみましょう。

図9.13 車の所有者数を示す表

生産国	サイズ	N
ヨーロッパ	大型	4
	中型	17
	小型	19
日本	大型	2
	中型	54
	小型	92
米国	大型	36
	中型	53
	小型	26

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Poll.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
3. 「生産国」をクリックし、「行のドロップゾーン」にドラッグします。
4. 「サイズ」をクリックしてドラッグし、「生産国」見出しの右側にドロップします。

図9.14 「生産国」と「サイズ」を表に追加

表の作成

表に追加するには、列や統計量を、テーブルの列見出しや行レベルの領域にドラッグ&ドロップしてください。

元に戻す やり直し 完了

▼6列

性別

既婚/未婚

年齢

生産国

サイズ

タイプ

度数

重み

ID

ページ列

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値…イアン)

中央絶対偏差

幾何平均

四分位範囲

分位点

最頻値

列%

行%

すべて

☐ グループ変数の欠測値を含める

☐ グループ変数の度数順

☐ 集計統計量の追加

デフォルトの統計量

形式の変更

生産国	サイズ	N
ヨーロッパ	大型	4
	中型	17
	小型	19
日本	大型	2
	中型	54
	小型	92
米国	大型	36
	中型	53
	小型	26

統計量の表を作成する

次に、各サイズの車の所有者について、その年齢の平均と標準偏差を表示してみましょう。



図9.15 年齢と標準偏差を年齢別に示した表

生産国	サイズ			
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
			標準偏差	6.194
日本	小型	年齢	平均	30.8
			標準偏差	5.388
	大型	年齢	平均	28.5
			標準偏差	4.95
米国	中型	年齢	平均	30.2
			標準偏差	4.668
	小型	年齢	平均	29.7
			標準偏差	5.928
	大型	年齢	平均	33.8
			標準偏差	6.167
	中型	年齢	平均	31.1
			標準偏差	6.976
	小型	年齢	平均	30.4
			標準偏差	5.846

1. 図9.14の状態から開始します。「年齢」をクリックしてドラッグし、「サイズ」見出しの右側にドロップします。
2. [平均] をクリックしてドラッグし、「合計」の上にドロップします。
3. [標準偏差] をクリックしてドラッグし、「平均」の下側にドロップします。

表の「平均」の下に「標準偏差」が表示されます。なお、[標準偏差] を「平均」の上側にドロップした場合は、「平均」の上に「標準偏差」が表示されます。

図9.16 「年齢」、[平均]、[標準偏差] を表に追加

表の作成

表に追加するには、列や統計量を、テーブルの列見出しや行ラベルの領域にドラッグ&ドロップしてください。

元に戻す

やり直し

完了

▼6列

性別

既婚/未婚

年齢

生産国

サイズ

タイプ

度数

重み

ID

ページ列

N

平均

標準偏差

最小値

最大値

範囲

全体に対する%

欠測値 N

カテゴリ数

合計

重みの合計

分散

標準誤差

変動係数

中央値…イアン)

中央絶対偏差

幾何平均

四分位範囲

分位点

最頻値

列%

行%

すべて

☐ グループ変数の欠測値を含める

☐ グループ変数の度数順

☐ 集計統計量の追加

デフォルトの統計量

形式の変更

生産国	サイズ			
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
日本	大型	年齢	平均	30.8
			標準偏差	5.388
	中型	年齢	平均	28.5
米国	大型	年齢	平均	4.95
			標準偏差	30.2
	中型	年齢	平均	4.668
	大型	年齢	平均	29.7
			標準偏差	5.928
	中型	年齢	平均	33.8
	大型	年齢	平均	6.167
			標準偏差	31.1
	中型	年齢	平均	6.976
	大型	年齢	平均	30.4
			標準偏差	5.846

表の配置を変更する

「サイズ」を表の上部に移動し、クロス表形式にします。

図9.17 「サイズ」を上部に移動

	サイズ					
	大型		中型		小型	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
生産国	年齢	年齢	年齢	年齢	年齢	年齢
ヨーロッパ	30.5	5.802	30.9	6.194	30.8	5.388
日本	28.5	4.95	30.2	4.668	29.7	5.928
米国	33.8	6.167	31.1	6.976	30.4	5.846

表の配置を変更します。

1. 図9.16の状態から開始します。「サイズ」見出しをクリックしてドラッグし、表見出しの一番右側にドロップします。

図9.18 「サイズ」の移動

生産国	サイズ			Tabular
ヨーロッパ	大型	年齢	平均	30.5
			標準偏差	5.802
	中型	年齢	平均	30.9
			標準偏差	6.194
日本	小型	年齢	平均	30.8
			標準偏差	5.388
	大型	年齢	平均	28.5
			標準偏差	4.95
米国	中型	年齢	平均	30.2
			標準偏差	4.668
	小型	年齢	平均	29.7
			標準偏差	5.928
	大型	年齢	平均	33.8
			標準偏差	6.167
	中型	年齢	平均	31.1
			標準偏差	6.976
	小型	年齢	平均	30.4
			標準偏差	5.846

			サイズ		
生産国			大型	中型	小型
ヨーロッパ	年齢	平均	30.5	30.9	30.8
		標準偏差	5.802	6.194	5.388
日本	年齢	平均	28.5	30.2	29.7
		標準偏差	4.95	4.668	5.928
米国	年齢	平均	33.8	31.1	30.4
		標準偏差	6.167	6.976	5.846

2. 「年齢」をクリックしてドラッグし、「大型」・「中型」・「小型」見出しの下にドロップします。
3. 「平均」と「標準偏差」の両方を選択してドラッグし、「大型」見出しの下にドロップします。

これで、見やすい表になりました。車のサイズと生産国別に、所有者の年齢に関する平均と標準偏差がすぐわかります。

## 複数のデータ列に対する表の例

人気者になるのに重要だと思う要素（成績・スポーツ・容貌・お金持ち）を、生徒が評価した結果のデータがあります。「表の作成」プラットフォームを使ってこれらの要素をすべて1つの表にまとめ、また、性別と都会/田舎の組み合わせで集計してみましょう。

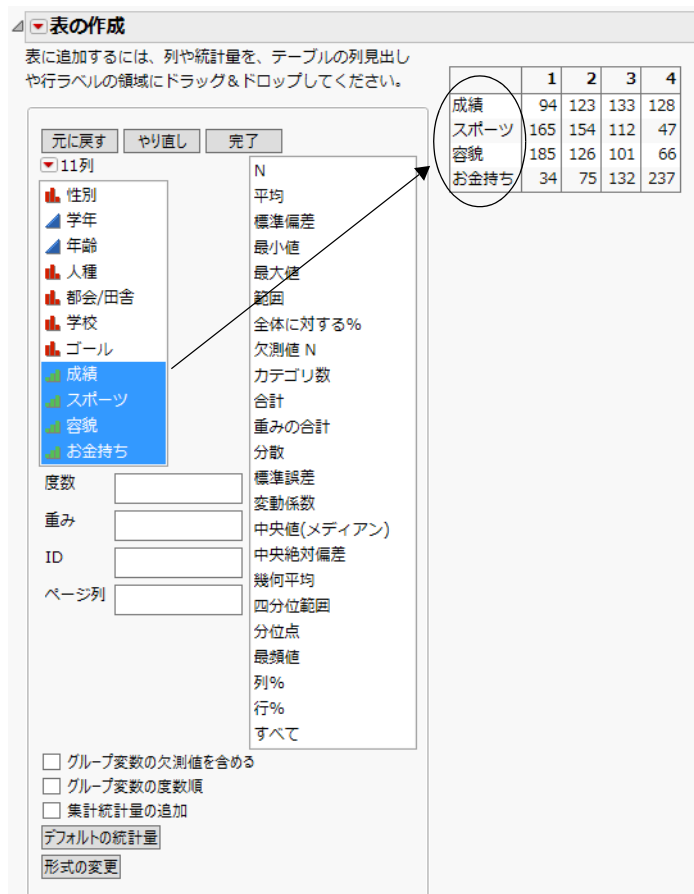
図9.19 人口統計データの追加

		全体に対する%														
		都会/田舎														
		Rural					Suburban					Urban				
性別		1	2	3	4	すべて	1	2	3	4	すべて	1	2	3	4	すべて
boy	成績	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	スポーツ	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	容貌	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
girl	お金持ち	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
	成績	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	スポーツ	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	容貌	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	お金持ち	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

1. 「ヘルプ」>「サンプルデータフォルダ」を選択し、「Children's Popularity.jmp」を開きます。
2. 「分析」>「表の作成」を選択します。

3. 「成績」・「スポーツ」・「容貌」・「お金持ち」を選択してドラッグし、「行のドロップゾーン」にドロップします。

図9.20 カテゴリの列



4つの列は同じ水準値を持つため、これら4つの集計結果をまとめた1つの表が表示されます。

要素ごとの4段階評価の割合を表にしましょう。

4. 「性別」をドラッグし、左側の空の見出しにドロップします。
5. 「全体に対する %」をドラッグし、数字の見出しの上にドロップします。
6. 「すべて」をドラッグし、数字の見出し「4」の右側にドロップします。

図9.21 「性別」・「全体に対する%」・「すべて」を表に追加

		全体に対する%				
性別		1	2	3	4	すべて
boy	成績	8.16%	12.76%	14.02%	12.55%	47.49%
	スポーツ	26.57%	12.55%	6.49%	1.88%	47.49%
	容貌	9.21%	15.48%	12.34%	10.46%	47.49%
girl	お金持ち	3.56%	6.69%	14.64%	22.59%	47.49%
	成績	11.51%	12.97%	13.81%	14.23%	52.51%
	スポーツ	7.95%	19.67%	16.95%	7.95%	52.51%
	容貌	29.50%	10.88%	8.79%	3.35%	52.51%
	お金持ち	3.56%	9.00%	12.97%	26.99%	52.51%

人の属性データを追加して、さらに細かく集計します。

7. 「都会/田舎」をドラッグし、「全体に対する%」見出しの下にドロップします。

図9.22 「都会/田舎」を表に追加

		全体に対する%														
		都会/田舎														
		Rural					Suburban					Urban				
性別		1	2	3	4	すべて	1	2	3	4	すべて	1	2	3	4	すべて
boy	成績	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	スポーツ	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	容貌	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
girl	お金持ち	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
	成績	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	スポーツ	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	容貌	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	お金持ち	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

これで、田舎（「Rural」）・郊外（「Suburban」）・都会（「Urban」）に住む男子（「boy」）にとって、人気者になるために一番重要な要素は「スポーツ」だということがわかります。田舎・郊外・都会に住む女子（「girl」）は、「容貌」を一番大事な要素だと考えているようです。

## ページ列の使用例

男子生徒と女子生徒の身長を測定したデータがあります。「表の作成」プラットフォームを使って、身長の平均を生徒の年齢別にまとめた表を作成してみましょう。さらにデータを性別で層別し、性別ごとの表を作成します。このとき、層別の列をページ列として追加します。ページ列を指定すると、グループごとにページが作成されます。

図9.23 生徒の性別ごとの平均身長

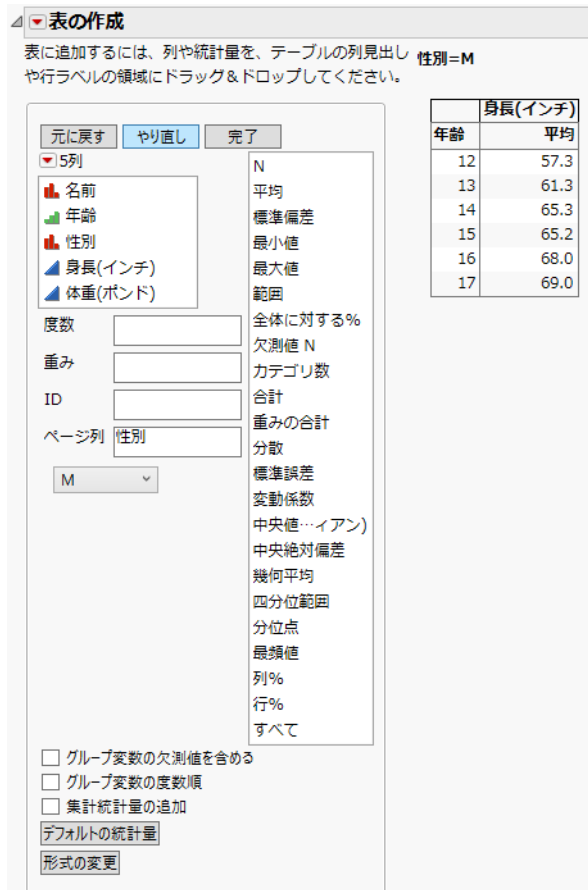
性別=F			性別=M		
		身長(インチ)			身長(インチ)
年齢	平均		年齢	平均	
12	58.6		12	57.3	
13	59.0		13	61.3	
14	62.6		14	65.3	
15	63.0		15	65.2	
16	62.5		16	68.0	
17	62.0		17	69.0	

女子

男子

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Big Class.jmp」を開きます。
2. [分析] > [表の作成] を選択します。  
「身長(インチ)」が分析対象の変数であるため、これを表の一番上に表示します。
3. 「身長(インチ)」をクリックし、「列のドロップゾーン」にドラッグします。  
年齢別の統計量が必要なため、年齢を左側に表示します。
4. 「年齢」をクリックしてドラッグし、数値「2502」の横にある空のセルにドロップします。
5. 「平均」をクリックしてドラッグし、「合計」のセルにドロップします。
6. 「性別」をクリックしてドラッグし、「ページ列」のフィールドにドロップします。
7. 「ページ列」の下にあるドロップダウンリストから「F」を選択すると、女子の平均身長だけが表示されます。
8. 同様に、「ページ列」の下にあるドロップダウンリストから「M」を選択すると、男子の平均身長だけが表示されます。[選択行なし] をクリックして、性別を指定せず、全体の値を表示させることもできます。

図9.24 ページ列の使用例



## グループ列を積み重ねる例

この例では、一連のグループ変数を使ってアンケート回答者の雇用に関する情報を分析します。変数名や値の中には文字数の多いものがあり、表の横幅が極端に広がってしまうため、グループ列を積み重ねて表示することになります。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Consumer Preferences.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
3. [現雇用先での勤続年数] と [給与] を選択してドラッグし、「列のドロップゾーン」にドロップします。
4. [N] と [平均] を選択してドラッグし、「現雇用先での勤続年数」の下にある [合計] というセルにドロップします。
5. 「結婚状態」・「学齢期の子供」・「仕事に対する満足度」を選択してドラッグし、表の左側の空白のセルにドロップします。

- 6. 「結婚状態」の見出しを右クリックし、[グループ列の積み重ね] を選択します。
- 7. [集計統計量の追加] チェックボックスをオンにします。

図9.25 積み重ねたグループ列

結婚状態/学齢期の子供/仕事に対する満足度	現雇用先での勤続年数		給与	
	N	平均	N	平均
Single	177	8	177	\$54,983
Yes	49	7.918367	49	\$60,594
Not at all satisfied	3	5.666667	3	\$45,667
Somewhat satisfied	26	7.769231	26	\$49,465
Extremely satisfied	20	8.45	20	\$77,300
No	128	8.03125	128	\$52,835
Not at all satisfied	13	7	13	\$46,346
Somewhat satisfied	71	8.352113	71	\$48,241
Extremely satisfied	44	7.818182	44	\$62,166
Not Single	271	9.608856	271	\$59,944
Yes	136	9.676471	136	\$63,935
Not at all satisfied	9	7.333333	9	\$59,167
Somewhat satisfied	70	8.8	70	\$59,725
Extremely satisfied	57	11.12281	57	\$69,859
No	135	9.540741	135	\$55,923
Not at all satisfied	7	11.57143	7	\$48,714
Somewhat satisfied	75	8.706667	75	\$51,134
Extremely satisfied	53	10.45283	53	\$63,651
すべて	448	8.973214	448	\$57,984

表には、グループ変数の組み合わせごとの給与と現雇用先での勤続年数の平均、および回答者数が表示されます。たとえば、独身（「single」）で学齢期の子供がいて（「Yes」）、仕事に非常に満足している（「Extremely satisfied」）と答えた人は20人います。この20人の回答者について見ると、現雇用先での平均勤続年数は8.45年、平均給与は\$77,300です。

表には、「結婚状態」と「学齢期の子供」の集計値も表示されます。たとえば、仕事に対する満足度を考慮しない場合、独身で学齢期の子供がいない回答者は128人です。この128人の回答者について見ると、現雇用先での平均勤続年数は8.03年、平均給与は\$52,835です。

一意のID列を使用した例

この例では、米国におけるニカルジピンの治験データを加工したものを使用して、体の各器官に副作用が現れる回数を薬剤とプラシボで比較します。なお、ここでは、1人の患者で同じ器官で複数回、副作用が発現しても、「1人」と数えることにします。

- 1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Nicardipine jmp」を開きます。
- 2. [分析] > [表の作成] を選択します。
- 3. 症例数が含まれた列である「症例」をクリックし、「列のドロップゾーン」にドラッグします。
- 4. 「器官分類または器官別大分類」を選択してドラッグし、表の左側の空白のセルにドロップします。
- 5. 「計画された治療群」を選択してドラッグし、「器官分類または器官別大分類」列の左側にドロップします。



この時点で、表には、各器官分類または器官別大分類における副作用の度数が、薬剤とプラシボに分けて表示されています。しかし、調べたいのは、各器官分類または器官別大分類において副作用が発現した一  
意な患者の人数です。

6. 「被験者 ID」を選択してドラッグし、「ID」にドロップします。
7. 【完了】をクリックします。

図9.26 一意の ID を使用して作成した表

計画された治療群	器官分類または器官別大分類	症例 合計
NIC .15	BLOOD AND LYMPHATIC SYSTEM DISORDERS	696
	CARDIAC DISORDERS	520
	ENDOCRINE DISORDERS	84
	EYE DISORDERS	40
	GASTROINTESTINAL DISORDERS	380
	GENERAL DISORDERS AND ADMINISTRATION SITE CONDITIONS	824
	HEPATOBIILIARY DISORDERS	504
	IMMUNE SYSTEM DISORDERS	20
	INFECTIONS AND INFESTATIONS	500
	INJURY, POISONING AND PROCEDURAL COMPLICATIONS	36
	INVESTIGATIONS	400
	METABOLISM AND NUTRITION DISORDERS	728
	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS	12
	NEOPLASMS BENIGN, MALIGNANT AND UNSPECIFIED (INCL CYSTS AND POLYPS)	20
	NERVOUS SYSTEM DISORDERS	1064
	PSYCHIATRIC DISORDERS	104
	RENAL AND URINARY DISORDERS	300
	REPRODUCTIVE SYSTEM AND BREAST DISORDERS	4
	RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS	868
	SKIN AND SUBCUTANEOUS TISSUE DISORDERS	28
	VASCULAR DISORDERS	1472
Placebo	BLOOD AND LYMPHATIC SYSTEM DISORDERS	728
	CARDIAC DISORDERS	588
	EAR AND LABYRINTH DISORDERS	4
	ENDOCRINE DISORDERS	120
	EYE DISORDERS	48
	GASTROINTESTINAL DISORDERS	352
	GENERAL DISORDERS AND ADMINISTRATION SITE CONDITIONS	808
	HEPATOBIILIARY DISORDERS	444
	IMMUNE SYSTEM DISORDERS	12
	INFECTIONS AND INFESTATIONS	476
	INJURY, POISONING AND PROCEDURAL COMPLICATIONS	12
	INVESTIGATIONS	364
	METABOLISM AND NUTRITION DISORDERS	720
	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS	12
	NEOPLASMS BENIGN, MALIGNANT AND UNSPECIFIED (INCL CYSTS AND POLYPS)	20
	NERVOUS SYSTEM DISORDERS	1124
	PSYCHIATRIC DISORDERS	20
	RENAL AND URINARY DISORDERS	160
	REPRODUCTIVE SYSTEM AND BREAST DISORDERS	12
	RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS	776
	SKIN AND SUBCUTANEOUS TISSUE DISORDERS	12
	VASCULAR DISORDERS	1520

この表は、薬剤またはプラシボを投与された患者のうち、各器官分類または器官別大分類において副作用が発現した人数を示します。たとえば、薬剤が投与された患者のうち、blood and lymphatic system（血管とリンパ管）に少なくとも1つの副作用が出た患者は696人です。

## グループ列をパッキングする例

この例では、さまざまなメーカーのハイブリッド車とガソリン車で燃費の中央値を比較します。[列のパッキング] オプションを使うと、一般道・高速道・総合の燃費を簡単に比較できます。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Hybrid Fuel Economy.jmp」を開きます。
2. [分析] > [表の作成] を選択します。
3. 「一般道走行の燃費(マイル/ガロン)」・「高速道走行の燃費(マイル/ガロン)」・「総合的な燃費(マイル/ガロン)」を選択してドラッグし、「列のドロップゾーン」にドロップします。
4. 「中央値(メディアン)」を選択してドラッグし、「一般道走行の燃費(マイル/ガロン)」の下の「合計」セルにドロップします。
5. 「中央値(メディアン)」というラベルが付いたいずれかのセルを選択してドラッグし、「一般道走行の燃費(マイル/ガロン)」のセルのすぐ上にドロップします。
6. 「総合的な燃費(マイル/ガロン)」のセルを選択してドラッグし、「一般道走行の燃費(マイル/ガロン)」のセルの左側にドロップします。
7. 「車種クラス」と「エンジン」を選択してドラッグし、表の左側の空白のセルにドロップします。
8. 表内で燃費の3列すべてを選択し、いずれかを右クリックします。
9. ポップアップメニューから [列のパッキング] > [バック] を選択します。

---

**ヒント:** パッキングした列の統計量の表示形式を調整するには、パッキングした列の見出しを右クリックして [列のパッキング] > [テンプレート] を選択します。

---

10. (オプション) いずれかのグループ列の見出しを右クリックし、[グループ列の積み重ね] を選択します。
11. [完了] をクリックします。

図9.27 列をパッキングした表

	中央値(メディアン)
車種クラス/エンジン	総合的な燃費(マイル/ガロン) (一般道走行の燃費(マイル/ガロン), 高速道走行の燃費(マイル/ガロン))
SUV 4WD	. ( , .)
Gas	18 (16 , 22)
Hybrid	21 (20 , 24)
SUV 2WD	. ( , .)
Gas	21 (18 , 25)
Hybrid	30 (32 , 28)
Midsized Cars	. ( , .)
Gas	26 (22.5 , 32)
Hybrid	38 (38 , 36)
Large Cars	. ( , .)
Gas	17 (15 , 22)
Hybrid	20 (17 , 24)
Trucks 2WD	. ( , .)
Gas	14 (13 , 18)
Hybrid	21 (20 , 23)
Trucks 4WD	. ( , .)
Gas	14 (12 , 18)
Hybrid	21 (20 , 23)
Compact Cars	. ( , .)
Hybrid	35 (35 , 34)
Subcompact Cars	. ( , .)
Gas	25 (22 , 31.5)

ハイブリッド車の燃費の中央値は、どの車両クラスでも高いことがわかります。たとえば、Large Cars（大型車）を見ると、ハイブリッド車の総合的な燃費は20、ガソリン車の総合的な燃費は17です。



# 第10章

## シミュレーション パラメトリックなモデルの乱数シミュレーション

シミュレーション機能を用いると、パラメトリックおよびノンパラメトリックな乱数シミュレーションを柔軟に行えます。シミュレーションは、次のような場合に使用します。

- ・ パラメトリックなブートストラップを行う。
- ・ 複雑なモデルの検出力計算を行う。
- ・ 複雑なモデルにおいて、予測値や信頼区間などの統計量の分布を近似する。
- ・ 並び替え検定を行う。
- ・ 説明変数に対する仮定が、モデルにおよぼす効果を調べる。
- ・ 様々な仮定に基づいて、モデルを分析する。
- ・ 新しい統計手法や、既存の統計手法を評価する。

このシミュレーション機能は、ブートストラップを行うときと同じ操作で実行できます。レポートの任意の場所で右クリックすると、[シミュレーション] オプションが表示されます。

図 10.1 検出力のシミュレーション



## 目次

シミュレーション機能の概要 .....	327
シミュレーション機能の例 .....	327
シミュレーション機能の起動 .....	333
「シミュレーション」 ウィンドウ .....	333
シミュレーション結果のテーブル .....	334
「シミュレーション 結果」 レポート .....	335
「検出力のシミュレーション」 レポート .....	335
シミュレーション機能の別例 .....	336
並び替え検定の例 .....	336
一般化回帰モデルに含める説明変数の検討例 .....	338
検出力の事前計算の例 .....	343

## シミュレーション機能の概要

シミュレーション機能は、レポート内の統計量の列をシミュレーションした結果を表示します。この機能を用いるには、統計量の列を右クリックして、[シミュレーション] を選択し、「シミュレーション」ウィンドウで、乱数データに置き換えたい列を指定します。乱数データに置き換えたい列を、「切り替え元の列」と呼びます。この列は、どんなものでもよく、応答変数であっても説明変数であっても構いません。その後、生成する乱数を示す計算式を含む列を指定します。乱数の計算式を含む列を、「切り替え先の列」と呼びます。「切り替え元の列」に対して、「切り替え先の列」の乱数が生成されます。

**メモ:** このシミュレーションの機能を用いるには、データテーブルに、乱数を生成する計算式を含んだ列を準備しておく必要があります。

処理の流れは次のとおりです。まず、「切り替え先の列」の計算式に基づいて、乱数データが生成されます。次に、そこで生成された「切り替え先の列」の乱数を、「切り替え元の列」のデータと置き換えて、元のレポート全体が再計算されます。この処理を、 $N$  回繰り返します。ここで、 $N$  はシミュレーション回数 (標本数) です。

シミュレーションが終了すると、次のようなデータテーブルが出力されます。

- データテーブルの各行は、各シミュレーションの結果です。
- 各列は、各統計量に対応しています。
- シミュレーション結果を見るためのスクリプトも作成されます。

**ヒント:** シミュレーションを実行すると、元のレポートに表示されているすべての分析が、再実行されます。これにより、特定の列を選択してシミュレーションを行っても、すべての分析が再実行されるために、時間がかかる場合があります。処理に時間がかかる場合は、プラットフォームのレポートから、無関係なオプションを除いてください。

シミュレーション機能は、次のものを除くすべてのプラットフォームで利用できます: アソシエーション分析、特性要因図、多次元尺度構成、多重因子分析、信頼性ブロック図、信頼性予測、修理可能システムのシミュレーション、応答のスクリーニング、テキストエクスプローラ。

## シミュレーション機能の例

変更が困難な変数の分散成分について、セミパラメトリックな信頼区間を計算してみましょう。この例では、温度、時間、触媒の量が、反応におよぼす効果を調べたいとします。温度は、変更が非常に困難な変数 (一次単位因子) です。時間は、変更が困難な変数 (二次単位因子) です。触媒の量は、変更が容易な変数です。一次単位因子と二次単位因子については、『実験計画 (DOE)』を参照してください。

過去の研究から、一次単位と二次単位の標準偏差は、それぞれ、誤差の標準偏差の約 2 倍と約 1.5 倍であることが分かっています。JMP がデフォルトで求めている信頼区間は、Wald 法による近似によるもので、分散成分の推定量が漸的に正規分布に従うものとして算出されています。この Wald 法による近似は精度が悪く、信頼区間の幅は狭くなる傾向があります。ここでは、分散成分の標本分布をシミュレートで求めて、分位点に基づいて信頼区間を計算してみます。

## 計画の作成

ここでは、カスタム計画を使用して、2段分割計画を作成します。

---

**ヒント:** この節の手順を省略するには、[ヘルプ] > [サンプルデータフォルダ] を選択し、「Design Experiment」フォルダの「Catalyst Design.jmp」を開きます。そして、「Catalyst Design.jmp」データテーブルの「DOE シミュレート」スクリプトの横にある緑の三角ボタンをクリックします。この操作が済んだら、「モデルのあてはめ」に進んでください。

---

1. [実験計画(DOE)] > [カスタム計画] を選択します。
2. 「因子」アウトラインで、「N個の因子を追加」の横のボックスに「3」と入力します。
3. [因子の追加] > [連続変数] を選択します。
4. 因子の名前をダブルクリックして、「温度」、「時間」、「触媒」に変更します。  
値は、デフォルトの「-1」と「1」をそのまま使用します。
5. 「温度」の [容易] をクリックして、[非常に困難] に変更します。  
これで、「温度」が一次単位因子になります。
6. 「時間」の [容易] をクリックして、[困難] に変更します。  
これで、「時間」が二次単位因子になります。
7. [続行] をクリックします。
8. 「モデル」パネルで [交互作用] > [2次] を選択します。  
すると、モデルにすべての2次交互作用が追加されます。
9. 「カスタム計画」の赤い三角ボタンをクリックし、メニューから [応答のシミュレート] を選択します。  
この操作を行っておくと、[テーブルの作成] をクリックして計画のデータテーブルを作成したときに、「応答をシミュレート」ウィンドウが開きます。

---

**メモ:** 乱数シード値 (手順10) と開始点の数 (手順11) を設定すると、以下の数値例と同じ実験設定が得られます。同じ結果でなくても良い場合は、これらの手順は不要です。

---

10. (オプション) 「カスタム計画」の赤い三角ボタンをクリックして、[乱数シード値の設定] を選択します。  
「12345」と入力して [OK] をクリックします。
11. (オプション) 「カスタム計画」の赤い三角ボタンをクリックして、[開始点の数] を選択します。「1000」と入力して [OK] をクリックします。
12. [計画の作成] をクリックします。
13. [テーブルの作成] をクリックします。

---

**メモ:** 乱数を用いているため、「Y」および「Yのシミュレーション」の値は、図10.2とは異なったものになります。

---



**図 10.2** 計画テーブル

Catalyst Design		一次単位	二次単位	温度	時間	触媒	Y	Yのシミュレーション
ロックされたファイル C:		1	1	1	1	-1	-0.046768023	0.08266896
計画	カスタム計画	2	1	1	1	1	6.539690219	7.99771819
基準	D-最適計画	3	1	1	1	1	6.5397108257	7.57655845
▶ モデル		4	1	2	-1	1	0.494801276	-0.3226987
▶ 混合モデルのあてはめ		5	1	2	-1	-1	-2.65841644	-2.6853901
▶ 計画の評価		6	1	2	1	-1	-0.200740508	-3.3961459
▶ DOEシミュレート		7	2	3	-1	1	3.1335699565	-0.4130434
▶ DOEダイアログ		8	2	3	-1	1	2.2739679213	1.03061945
▶ Yのシミュレーションのモデル		9	2	3	-1	1	1.1704395547	-1.3268726
		10	2	4	-1	-1	-0.693941035	2.7031519
		11	2	4	-1	-1	2.3525055616	2.48510044
		12	2	4	-1	-1	-1.545149968	-0.9948507
■ 一次単位 *		13	3	5	-1	1	-1.4838427	-0.8433382
■ 二次単位 *		14	3	5	-1	1	-1.375362439	-0.3655945
▲ 温度 *		15	3	5	-1	1	0.7434199911	1.0999211
▲ 時間 *		16	3	6	-1	-1	-1.11989643	0.61987196
▲ 触媒 *		17	3	6	-1	-1	3.4346315881	3.88157347
▲ Y *		18	3	6	-1	1	0.2987179478	-0.6335796
▲ Yのシミュレーションモデル		19	4	7	1	-1	-1.245681444	2.88701227
		20	4	7	1	-1	0.4951842993	-0.369875
		21	4	7	1	-1	-0.6602484	-1.5421839
		22	4	8	1	1	0.7779088461	1.6495913
		23	4	8	1	1	7.9432230978	7.67379998
		24	4	8	1	-1	1.0668195454	0.53690221

**図 10.3 「応答をシミュレート」ウィンドウ**

応答をシミュレート

効果	Y
切片	1
温度	1
時間	1
触媒	1
温度*時間	1
温度*触媒	1
時間*触媒	1

係数のリセット

分布

☒ 正規      誤差  $\sigma$ :

一次単位  $\sigma$ :

二次単位  $\sigma$ :

☐ 二項

☐ Poisson

適用

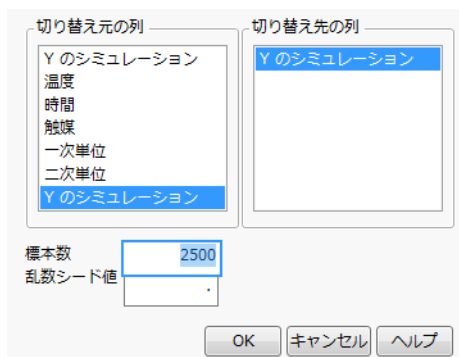


## 信頼区間の生成

ここでは、分散成分の推定値をシミュレートして、信頼区間を求めます。

1. 「REML法による分散成分推定値」アウトラインで、「分散成分」列を右クリックして、[シミュレーション]を選択します。

図10.5 「シミュレーション」ウィンドウ



このシミュレーションでは、現モデルの推定に用いた「Yのシミュレーション」列のデータ値が、シミュレーションごとに「Yのシミュレーション」列の計算式で生成される乱数に置き換えられます。右クリックして選択された「分散成分」列がシミュレーションされます。右クリックした「分散成分」列が強調表示され、この列が「パラメータ推定値」テーブルに表示されている各効果に対してシミュレーションされることを示します。

2. 「標本数」に「200」と入力します。
3. (オプション)「乱数シード値」に「456」と入力します。

この操作を行うと、1行目を除き、図10.6と同じ値が生成されます。

4. [OK] をクリックします。

1行目の値は、図10.6とは異なる可能性があります。

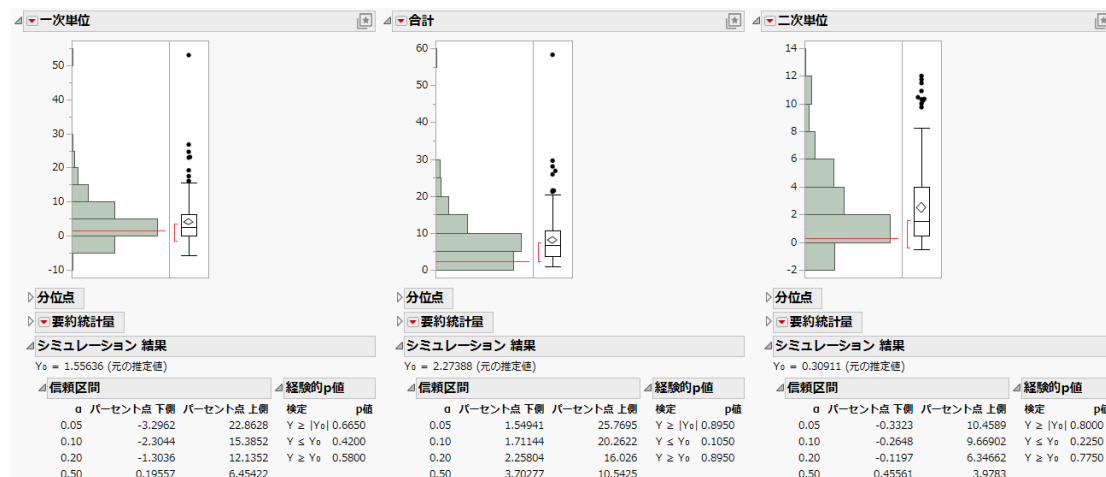
図 10.6 分散成分のシミュレーション結果（一部）

最小2乗法によるあてはめ シミ...		Y	SimID	一次単位	合計	残差	二次単位
乱数シード値	456	1 Y Simulated	0	1.5563639995	2.2738839302	0.4084102799	0.3091096508
▶ 連結したデータテーブルの作成		2 Y Simulated	1	0.1160090855	11.825684937	1.7118437836	9.9978320675
▶ 一変量の分布		3 Y Simulated	2	3.3652848741	8.6529959081	0.7259348275	4.5617762065
		4 Y Simulated	3	22.95631251	26.836360491	0.8416572537	3.0383907272
		5 Y Simulated	4	12.156574946	13.143778421	0.9872034755	-0.176636748
		6 Y Simulated	5	-0.060658009	1.3514412755	1.3514412755	-0.153702674
		7 Y Simulated	6	11.804365807	13.894478258	0.9837455205	1.1063669307
		8 Y Simulated	7	2.6824797817	3.7015003812	0.8559296173	0.1630909822
		9 Y Simulated	8	26.765100099	28.022776455	1.0422128078	0.2154635485
		10 Y Simulated	9	12.75877132	17.850598404	1.0302645404	4.0615625437
		11 Y Simulated	10	-0.433818317	3.0738531832	1.2569403028	1.8169128804
		12 Y Simulated	11	2.4634884406	11.105096973	1.6719518802	6.9696566519
		13 Y Simulated	12	4.2166345958	7.2251377019	1.0326091926	1.9758939136
		14 Y Simulated	13	2.8027318323	4.6661579836	0.7050109246	1.1584152266
		15 Y Simulated	14	0.3194564512	6.53640214	0.9281404099	5.2888052789
		16 Y Simulated	15	-0.270350632	3.1414624009	1.6586718874	1.4827905134
		17 Y Simulated	16	-0.192580268	1.7108138759	0.9471174937	0.7636963822
		18 Y Simulated	17	0.756815441	2.927052512	0.9782751744	1.1919618966
		19 Y Simulated	18	4.8135329401	7.4052160139	1.5524955876	1.0391874862
		20 Y Simulated	19	7.795916077	13.992557456	1.2884483808	4.9081929984
		21 Y Simulated	20	1.7125911677	2.8121549503	0.9736955177	0.1258682649

「最小2乗法によるあてはめ シミュレーション結果（分散成分）」データテーブルの最初の行は、元データから計算された「分散成分」の推定値であり、除外の状態になっています。その他の行は、シミュレーションで生成された乱数です。

## 5. 「一変量の分布」スクリプトを実行します。

図 10.7 分散成分に対して「一変量の分布」を実行した結果（一部）



各分散成分の信頼区間は、この「シミュレーション結果」レポートに示されているとおりです。 $\alpha=0.05$ の行の95%信頼区間を、REMLレポート（図10.4）の信頼区間と比較してみましょう。

- 一次単位の分散成分の95%信頼区間は、シミュレーションでは-3.296～22.863です。REMLレポートのWald法では、-1.973～5.086です。
- 二次単位の分散成分の95%信頼区間は、シミュレーションでは-0.332～10.459です。REMLレポートのWald法では、-0.605～1.223です。

シミュレーションによって求めた信頼区間は、1組の値を使ってREML法で計算した信頼区間に比べ、はるかに広がっています。信頼区間の精度をさらに高めるには、シミュレーションの回数を増やしてみてください。

---

## シミュレーション機能の起動

シミュレーション機能を起動するには、レポートウィンドウで計算値の列を右クリックし、[シミュレーション]を選択します。なお、このシミュレーション機能を用いるには、予め、乱数を生成する計算式の列をデータテーブルに準備しておく必要があります。[シミュレーション]オプションは、ブートストラップに対応しているすべてのレポートを含め、多くのレポートで使用できます。

---

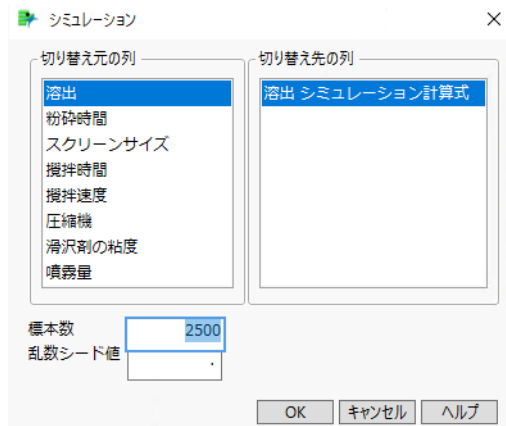
**メモ:** シミュレーション機能は、By変数を使用したレポートでは使用できません。

---

## 「シミュレーション」ウィンドウ

「シミュレーション」ウィンドウで列と標本数を指定し、[OK]をクリックしてシミュレーションを開始します。進行状況バーと[ここで停止]オプションが表示されます。進行状況を示すバーの上には、その時点で計算が終了した標本数が表示されます。[ここで停止]をクリックすると、その時点までの計算結果が「シミュレーション結果」データテーブルに表示されます。進行状況を示すウィンドウには、実行中の分析名も表示されます。

図 10.8 「シミュレーション」 ウィンドウ (Tablet Production.jmp)



「シミュレーション」ウィンドウには、以下のパネルとオプションがあります。

**切り替え元の列** 切り替え先の列によって置き換えられる列。

**切り替え先の列** 切り替え元の列を置き換える列。この列の計算式に従ってシミュレーションした値を使って、分析が繰り返されます。このリストボックスには、計算式を含んだ列だけが表示されます。

**標本数** データをシミュレーションし、レポートを再実行する回数。デフォルトの値は2500です。

**乱数シード値** 乱数を決めるためのシード値。この乱数シード値は、シミュレーション結果を再現したいときに使用します。

## シミュレーション結果のテーブル

シミュレーションを行うと、結果をまとめたデータテーブルが作成されます。次のことを注意してください。

- データテーブルにおける最初の行は、元のレポートの結果となっています。この行には、除外の属性が与えられます。
- その他の行には、シミュレーションの結果が示され、その行数は、「シミュレーション」ウィンドウで「標本数」に指定した数と同じです。
- 元のレポートにおける表の1列目が、結果のデータテーブルにおける列になります。結果のデータテーブルでは、これらの見出しが列名になります。
- 結果のデータテーブルには、「一変量の分布」レポートを作成する「**一変量の分布**」スクリプトが含まれます。「一変量の分布」スクリプトを実行すると、結果のデータテーブルの各列について、ヒストグラム、分位点、要約統計量、シミュレーションの結果が表示されます。この「一変量の分布」レポートには、基本的な情報のほか、次の情報が表示されます。
  - ヒストグラムに表示される元の推定値を示す赤い線。

- 元の推定値、信頼区間、経験 $p$ 値を含む「シミュレーション 結果」レポート。「[「シミュレーション 結果」レポート](#)」を参照してください。
- シミュレーションの対象にした統計量が $p$ 値である場合は、「検出力のシミュレーション」レポートも表示されます。「[「検出力のシミュレーション」レポート](#)」を参照してください。
- $p$ 値の列に対してシミュレーションを行った場合にのみ、「[検出力の分析](#)」スクリプトが表示されます。このスクリプトは「一変量の分布」レポートとともに、 $p$ 値のヒストグラムと、「検出力のシミュレーション」レポートを作成します。「[「検出力のシミュレーション」レポート](#)」を参照してください。

## 「シミュレーション 結果」レポート

「シミュレーション 結果」レポートには、次の情報が含まれています。

**元の推定値** 「シミュレーション結果」データテーブルの最初の行にも表示されている元の推定値。この値は $Y_0$ と記されています。

**信頼区間** 0.05、0.10、0.20、および0.50の有意水準における、分位点に基づく信頼区間の下限および上限を示します。

**経験的 $p$ 値** 両側検定と片側検定の経験的 $p$ 値は、元の推定値を基準にしてシミュレートした結果から求められます。これらの経験 $p$ 値は、レポートの「検定」列に示されている範囲内にシミュレートされた値が含まれている割合です。

## 「検出力のシミュレーション」レポート

「検出力のシミュレーション」レポートには、次の情報が含まれています。

**アルファ ( $\alpha$ )** 有意水準。0.01、0.05、0.10、および0.20が有意水準として使われています。

**棄却された度数** 検定によって棄却された標本数を、有意水準別に表示します。

**棄却された割合** 検定によって棄却された標本の割合を、有意水準別に表示します。

**下側95%と上側95%** 棄却された標本の割合について、両側95%信頼区間の下限および上限を示します。この信頼区間の計算には、Wilsonのスコア法を使用しています。Wilson (1927) を参照してください。

---

**ヒント:** 信頼区間を狭くするには、シミュレーションの標本数を増やしてください。

---

## シミュレーション機能の別例

この節では、シミュレーション機能を使用したその他の例をご紹介します。

- 「並び替え検定の例」
- 「一般化回帰モデルに含める説明変数の検討例」
- 「検出力の事前計算の例」

なお、『品質と工程』には、正規分布に従わない変数について、PpKや不適合率の信頼区間をシミュレートする方法が紹介されています。

### 並び替え検定の例

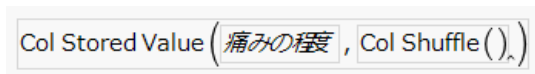
この例では、乱数シミュレーションに基づいて、並び替え検定を行ってみましょう。3種類の鎮痛剤の効果に違いがあるかどうかに興味があります。標本サイズが非常に小さく、通常の分散分析では妥当な結果とならない恐れがあります。そこで、シミュレーションに基づく並び替え検定を行ってみましょう。

まず、痛みの程度の測定値を、3種類の薬に対してランダムにシャッフルする計算式を作成します。「鎮痛剤の効果がない」という帰無仮説のもとでは、観測された測定値に対して各処置が同様の確率で割り付けられると考えられます。測定値をランダムにシャッフルして得られる $F$ 値は、標本サイズが大きければ近似的に $F$ 分布に従います。ここでは、通常の分散分析のように $F$ 分布から $p$ 値を計算するのではなく、乱数による並び替えによって $p$ 値を求めてみましょう。

#### シミュレーションの計算式の定義

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Analgesics.jmp」を開きます。
2. [列] > [列の新規作成] を選択します。
3. 「列名」に「痛みの程度（シャッフル）」と入力します。
4. 「列プロパティ」リストから [計算式] を選択します。
5. 関数のリストから [行] > [Col Stored Value] を選択します。
6. 「列」のリストから、[痛みの程度] をダブルクリックします。
7. エディタパネルの上部の記号リストから、挿入キー (^) をクリックします。
8. 関数のリストから、[乱数] > [Col Shuffle] を選択します。

図10.9 完成した式



この計算式は、「痛みの程度」列の値をランダムにシャッフルします。

9. 計算式エディタウィンドウで [OK] をクリックします。



10. 列情報ウィンドウで **[OK]** をクリックします。

### 並び替え検定の実行

1. **[分析] > [二変量の関係]** を選択します。
2. 「痛みの程度」を選択し、**[Y, 目的変数]** をクリックします。
3. 「薬」を選択し、**[X, 説明変数]** をクリックします。
4. **[OK]** をクリックします。
5. 赤い三角ボタンをクリックし、**[平均/ANOVA]** を選択します。

図 10.10 「分散分析」レポート

分散分析					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
薬	2	99.89459	49.9473	6.2780	0.0053*
誤差	30	238.67877	7.9560		
全体(修正済み)	32	338.57335			

ここで、 $F$  値が6.2780であることに注目してください。

6. 「分散分析」アウトラインの「 $F$  値」列を右クリックし、**[シミュレーション]** を選択します。
7. 「切り替え元の列」のリストから、「痛みの程度」をクリックします。
8. 「切り替え先の列」のリストから、「痛みの程度 (シャッフル)」をクリックします。
9. 「標本数」に「1000」と入力します。
10. (オプション)「乱数シード値」に「456」と入力します。

こうすると、後述する結果と同じ結果が再現されます。

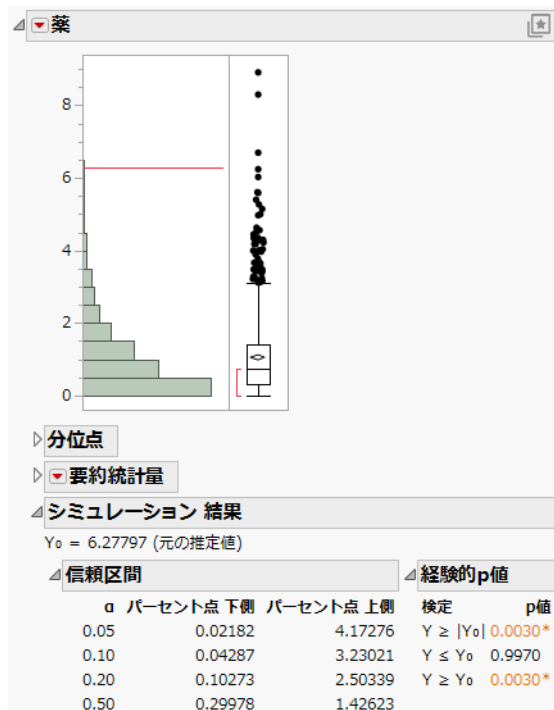
図 10.11 設定後の「シミュレーション」ウィンドウ

切り替え元の列		切り替え先の列	
痛みの程度		痛みの程度 (シャッフル)	
薬			
標本数	1000		
乱数シード値	456		
OK		キャンセル ヘルプ	

11. **[OK]** をクリックします。

シミュレーションの結果を示すテーブルでは、「全体(修正済み)」と「誤差」は空白になっています。 $F$  値だけが、「薬」の列に出力されています。

12. このテーブルで、「一変量の分布」スクリプトを実行します。

図 10.12 シミュレーションによって得た  $F$  値の帰無分布

観測された  $F$  値の 6.2780 は、ヒストグラムにおいて赤い線で表示されます。この値は、シミュレーションによって得た  $F$  値の帰無分布における上位 0.5% 点よりも大きくなっています。このことより、今回の実験結果は、「3 種類の薬が痛みの程度に与える効果のいずれかは異なる」ことを示す強い証拠であると言えます。

## JMP PRO 一般化回帰モデルに含める説明変数の検討例

一般化回帰モデルを作成し、ゼロでない項を使って縮小モデルをあてはめてみましょう。縮小モデルでシミュレーション機能を使用し、特定の説明変数をモデルに含めるべきかどうかを検討します。

製薬会社が持つ過去のデータに基づき、錠剤の溶出速度とその速度に影響をおよぼす可能性のある要因について分析しましょう。溶出速度が 70 に満たない錠剤は、不適合と見なされます。どの要因が溶出速度に影響するのかを調べてみましょう。

### モデルのあてはめ

この節では、一般化回帰モデルのあてはめを行います。

**ヒント:** この節の手順を省略するには、「Tablet Production.jmp」データテーブルの「一般化回帰」スクリプトの横の緑の三角ボタンをクリックし、モデルを作成してください。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Tablet Production.jmp」を開きます。
2. [分析] > [モデルのあてはめ] を選択します。
3. 「溶出」を選択し、[Y] をクリックします。
4. 「粉碎時間」から「噴霧器圧力」までを選択し、[追加] をクリックします。
5. 「手法」のリストから [一般化回帰] を選択します。
6. [実行] をクリックします。
7. 「モデルの設定」パネルで [適応型] チェックボックスを選択します。
8. 「モデルの設定」パネルで [実行] をクリックします。

図 10.13 適応型 Lasso に基づくモデル

元の説明変数に対する推定値						
項	推定値	標準誤差	Waldカイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
切片	108.64064	24.332978	19.933984	<.0001*	60.94888	156.3324
粉碎時間	0.1302779	0.028185	21.365104	<.0001*	0.0750363	0.1855195
スクリーンサイズ[3-5]	4.1877689	0.5414927	59.810874	<.0001*	3.1264626	5.2490751
スクリーンサイズ[4-5]	2.3907767	0.5672167	17.765609	<.0001*	1.2790525	3.5025009
ステアリン酸マグネシウム供給業者[Jones Inc-Smith Ind]	0	0	0	1.0000	0	0
乳糖供給業者[Bond Inc-James Ind]	0	0	0	1.0000	0	0
砂糖供給業者[Sour-Sweet]	0	0	0	1.0000	0	0
タルク供給業者[Rough-Smooth]	0	0	0	1.0000	0	0
攪拌時間	0.7223431	0.1381847	27.325418	<.0001*	0.4515059	0.9931802
攪拌速度	0.2130883	0.2389206	0.7954486	0.3725	-0.255187	0.681364
圧縮機[Compress1-Compress2]	-0.538111	0.3993672	1.8155137	0.1778	-1.320857	0.244634
圧力	0	0	0	1.0000	0	0
滑沢剤供給業者[Coat-Mac]	0	0	0	1.0000	0	0
滑沢剤供給業者[Down-Mac]	0	0	0	1.0000	0	0
滑沢剤の粘度	0.1834341	0.0500838	13.414234	0.0002*	0.0852717	0.2815964
注入温度	0	0	0	1.0000	0	0
排出温度	0	0	0	1.0000	0	0
噴霧量	-0.204356	0.0420142	23.658261	<.0001*	-0.286703	-0.12201
噴霧器圧力	0	0	0	1.0000	0	0
正規分布 パラメータ	推定値	標準誤差	Waldカイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
尺度	2.0224258	0.1641142	151.86341	<.0001*	1.700768	2.3440837

「適応型 Lasso（検証法: AICc, 分布: 正規）」レポートのパラメータ推定値を見てみましょう。パラメータ推定値がゼロでないものは、「粉碎時間」、「スクリーンサイズ」、「攪拌時間」、「攪拌速度」、「圧縮機」、「滑沢剤の粘度」、および「噴霧量」で、これらが「溶出」に関係している可能性があります。

## 効果のさらなる削除

モデルの縮小に進む前に、「Tablet Production.jmp」データテーブルで、列が選択されていないことを確認します。列が選択されたままだと、後述の手順1でも選択された状態になります。列の選択を解除しておく、誤ってゼロの項の列を含めてしまうことが避けられます。

この節の手順を省略するには、「Tablet Production.jmp」データテーブルの「一般化回帰 縮小モデル」スク립トの横の緑の三角ボタンをクリックし、縮小モデルを作成してください。

1. 「適応型 Lasso（検証法: AICc, 分布: 正規）」レポートの赤い三角ボタンをクリックし、[非ゼロの効果を使って再起動] > [非ゼロの効果を使って再起動] を選択します。

「モデルのあてはめ」ウィンドウが開き、「モデル効果の構成」リストに、パラメータ推定値がゼロでない項が表示されます。応答は「Y」にすでに指定されています。「手法」は「一般化回帰」が選択されています。

2. 「実行」をクリックします。
3. 「モデルの設定」パネルで「**適応型**」チェックボックスを選択します。
4. 「モデルの設定」パネルで「**実行**」をクリックします。

図10.14 適応型Lassoによる縮小モデル

元の説明変数に対する推定値						
項	推定値	標準誤差	Waldカイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
切片	95.142391	23.430178	16.489099	<.0001*	49.220085	141.0647
粉碎時間	0.1394103	0.0275609	25.585971	<.0001*	0.0853919	0.1934288
スクリーンサイズ[3-5]	4.3323833	0.534237	65.763638	<.0001*	3.285298	5.3794685
スクリーンサイズ[4-5]	2.6331283	0.5457852	23.275584	<.0001*	1.563409	3.7028476
攪拌時間	0.7583048	0.1385246	29.966364	<.0001*	0.4868017	1.029808
攪拌速度	0.4575802	0.2289168	3.9955744	0.0456*	0.0089116	0.9062488
圧縮機[Compress1-Compress2]	-0.877986	0.4127286	4.5252866	0.0334*	-1.686919	-0.069053
滑沢剤の粘度	0.198673	0.0486798	16.656386	<.0001*	0.1032624	0.2940836
噴霧量	-0.212753	0.0410929	26.805238	<.0001*	-0.293294	-0.132213
正規分布 パラメータ	推定値	標準誤差	Waldカイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
尺度	1.9982636	0.1574951	160.97992	<.0001*	1.689579	2.3069483

「攪拌速度」の信頼区間（「下側95%」）が、ゼロに非常に近くなっていることに注目してください。次に、シミュレーションによって、現在の推定値が真値であった場合、適応型Lassoを行うと「攪拌速度」の推定値がどれぐらいの割合でゼロになるかを調べてみましょう。

### 攪拌速度の推定値がどれぐらいゼロになるかの考察

以下の手順では、縮小モデル（図10.14）のレポートを使用します。

1. 「適応型Lasso（検証法: AICc, 分布: 正規）」レポートの赤い三角ボタンをクリックし、「**列の保存**」>「**シミュレーション計算式の保存**」を選択します。

「Tablet Production.jmp」データテーブルに、「溶出 シミュレーション計算式」という新しい列が追加されます。

2. （オプション）データテーブルの「列」パネルで、「溶出 シミュレーション計算式」の右側にある＋記号をクリックします。

図 10.15 シミュレーション計算式

Random Normal

95.142391047

+ 0.1394103417 • 粉碎時間

+ Match(スクリーンサイズ)  $\left( \begin{array}{l} \text{"3"} \Rightarrow 4.3323832797 \\ \text{"4"} \Rightarrow 2.6331282831 \\ \text{"5"} \Rightarrow 0 \\ \text{else} \Rightarrow . \end{array} \right)$

+ 0.7583048404 • 攪拌時間

+ 0.4575801859 • 攪拌速度

+ Match(圧縮機)  $\left( \begin{array}{l} \text{"Compress1"} \Rightarrow -0.877985955 \\ \text{"Compress2"} \Rightarrow 0 \\ \text{else} \Rightarrow . \end{array} \right)$

+ 0.1986730026 • 滑沢剤の粒度

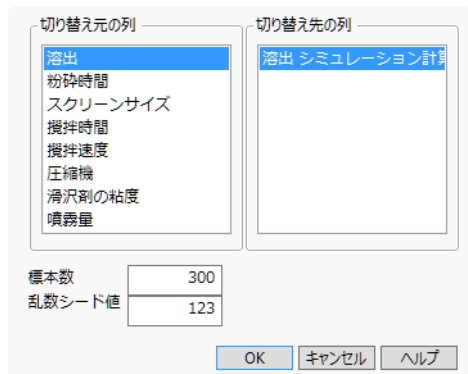
+ -0.212753206 • 噴霧量

, 1.9982636301

この計算式は、あてはめたモデルと「溶出」の分布に従って、各行の値をシミュレートします。「溶出」の分布は、標準偏差が約1.998の正規分布に従うと推定されています。

3. **【キャンセル】** をクリックします。
4. 縮小モデルのレポートウィンドウに戻ります。「元の説明変数に対する推定値」レポートで、「推定値」列を右クリックし、**【シミュレーション】** を選択します。  
「切り替え元の列」リストで**【溶出】** が選択されていることを確認します。
5. **【標本数】** に「300」と入力します。  
「溶出」列を「溶出 シミュレーション計算式」列のシミュレーション値に置き換えて、分析を300回実行します。
6. (オプション) **【乱数シード値】** に「123」と入力します。  
こうすると、後述する結果と同じ結果が再現されます。

図 10.16 設定後の「シミュレーション」ウィンドウ



7. **[OK]** をクリックします。

データテーブルの最初の行は、分析対象のデータから計算された推定値であり、除外されています。その他の行は、シミュレーションで生成された乱数です。

8. 「**一変量の分布**」スクリプトを実行します。

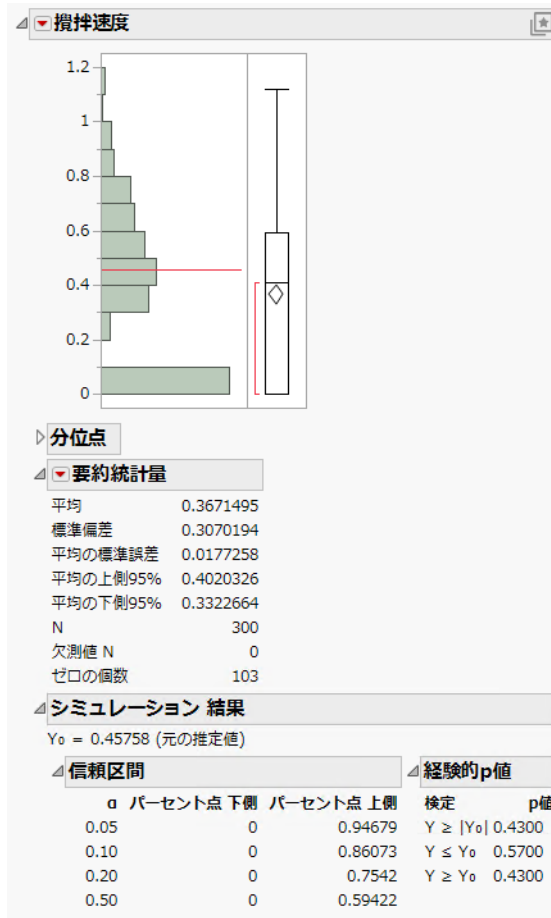
9. Ctrl キーを押しながら、「スクリーンサイズ[3-5]」の赤い三角ボタンをクリックし、**[表示オプション] > [要約統計量のカスタマイズ]** を選択します。

10. **[ゼロの個数]** を選択します。

11. **[OK]** をクリックします。

12. 「**攪拌速度**」のレポートにスクロールします。

図 10.17 シミュレーションによる攪拌速度の推定値のヒストグラム



「要約統計量」レポートによると、シミュレーションのうち  $103/300 = 34.3\%$  で、「攪拌速度」の推定値がゼロになっていることがわかります。

## 検出力の事前計算の例

この例では、シミュレーション機能を使って非線形モデルの検出力の事前計算を行います。6つの連続尺度の因子が、部品検査の可否におよぼす主効果を調べたいとしましょう。応答は二項分布に従い、実験は60回行えるものとします。

一般化線形モデルで、リンク関数としてロジット関数を使用して、成功確率をモデル化します。リンク関数としてロジット関数を使用すると、次のロジスティックモデルが当てはめられます。

$$\pi(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_6 X_6)}}$$

ここで、 $\pi(\mathbf{X})$  は、 $\mathbf{X} = (X_1, X_2, \dots, X_6)$  という因子設定のもと、部品が合格する確率を表します。

線形予測子  $L(\mathbf{X})$  は、次の式で表されます。

$$L(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \dots + \beta_6 X_6$$

線形予測子の係数が次のような場合だったときの検出力を求めてみましょう。

係数	値
$\beta_0$	0
$\beta_1$	1
$\beta_2$	0.9
$\beta_3$	0.8
$\beta_4$	0.7
$\beta_5$	0.6
$\beta_6$	0.5

線形予測子の切片は0なので、全因子が0に設定されている場合、部品が合格する確率は50%に等しくなります。第  $i$  因子の各水準における合格率は、その他の因子がすべて0である場合、次のようになります。

因子	$X_i = 1$ の場合の合格率	$X_i = -1$ の場合の合格率	差
$X_1$	73.11%	26.89%	46.2%
$X_2$	71.09%	28.91%	42.2%
$X_3$	69.00%	31.00%	38.0%
$X_4$	66.82%	33.18%	33.6%
$X_5$	64.56%	35.43%	29.1%
$X_6$	62.25%	37.75%	24.5%

たとえば、 $X_1$  以外の全因子が0に設定されている場、検出したい差は46.2%になります。検出したい差が最小になるのは、 $X_6$  以外の全因子が0に設定された場合で、その差は24.5%です。



## 計画の作成

ここでは、実験に対するカスタム計画を作成します。

---

**メモ:** この節の手順を省略するには、[ヘルプ] > [サンプルデータフォルダ] を選択し、「Design Experiment」フォルダの「Experiment.jmp」を開きます。「DOE シミュレート」スクリプトの横にある緑色の三角形をクリックし、「[応答をシミュレート] の設定」に進んでください。

---

1. [実験計画 (DOE)] > [カスタム計画] を選択します。

---

**メモ:** カスタム計画は非線形な状況に最適とはいえませんが、この例では単純化のため、「非線形計画」プラットフォームではなく「カスタム計画」プラットフォームを使用します。「非線形計画」プラットフォームで作成した計画のほうが直交計画より優れていることを示す例については、『実験計画 (DOE)』を参照してください。

---

2. 「因子」アウトラインで、「N 個の因子を追加」の横のボックスに「6」と入力します。
3. [因子の追加] > [連続変数] を選択します。
4. [続行] をクリックします。  
主効果だけの計画を作成したいため、「モデル」アウトラインには変更を加えません。
5. 「実験の回数」の「ユーザ定義」テキストボックスに「60」と入力します。
6. 「カスタム計画」の赤い三角ボタンをクリックし、メニューから [応答のシミュレート] を選択します。  
この操作を行っておくと、[テーブルの作成] をクリックして計画のデータテーブルを作成したときに、「応答をシミュレート」ウィンドウが開きます。

---

**メモ:** 乱数シード値 (手順7) と開始点の数 (手順8) を設定すると、この例と同じ実験が得られます。同じ結果でなくても良い場合は、これらの手順は不要です。

---

7. (オプション) 「カスタム計画」の赤い三角ボタンをクリックして、[乱数シード値の設定] を選択します。「12345」と入力して [OK] をクリックします。
8. (オプション) 「カスタム計画」の赤い三角ボタンをクリックして、[開始点の数] を選択します。「1」と入力して [OK] をクリックします。
9. [計画の作成] をクリックします。
10. [テーブルの作成] をクリックします。

---

**メモ:** 乱数を用いているため、「Y」および「Yのシミュレーション」の値は、図10.18とは異なったものになります。

---

図 10.18 計画テーブル (一部)

Custom Design			X1	X2	X3	X4	X5	X6	Y	Y Simulated	
Design	Custom Design	Σ	1	-1	1	-1	1	1	-1	-0.796505268	-0.796505268
Criterion	D Optimal		2	1	1	1	1	-1	1	4.6366447636	4.6366447636
Model			3	-1	-1	-1	-1	-1	-1	-5.765575838	-5.765575838
Evaluate Design			4	-1	-1	1	-1	1	-1	-0.701336672	-0.701336672
Generalized Regression			5	1	-1	-1	-1	-1	1	-1.670504226	-1.670504226
DOE Simulate			6	1	-1	-1	1	-1	-1	-0.132721143	-0.132721143
DOE Dialog			7	-1	1	1	1	1	-1	3.245090941	3.245090941
			8	-1	1	-1	-1	1	-1	0.2920735414	0.2920735414
			9	-1	1	-1	-1	-1	-1	-2.720270723	-2.720270723
			10	1	1	-1	1	1	1	4.0238266527	4.0238266527
			11	-1	-1	1	1	-1	-1	-2.229218491	-2.229218491
			12	-1	1	-1	1	-1	1	1.4625517974	1.4625517974
			13	1	1	-1	1	-1	-1	-1.073024419	-1.073024419
			14	-1	-1	-1	-1	1	1	-1.709272769	-1.709272769
			15	1	1	-1	1	1	-1	2.355634722	2.355634722
			16	1	-1	-1	-1	-1	-1	-3.14978945	-3.14978945

図 10.19 「応答をシミュレート」ウィンドウ

**応答をシミュレート**

効果 Y

切片 1

X1 1

X2 1

X3 1

X4 1

X5 1

X6 1

係数のリセット

**分布**

☒ 正規 エラー σ: 1

☐ 二項

☐ Poisson

適用

計画のデータテーブルと「応答をシミュレート」ウィンドウが表示されます。計画のデータテーブルには、次の2つの列があります。

- 「Y」列: 「応答をシミュレート」ウィンドウの設定に従ってシミュレートされたデータ値。
- 「Yのシミュレーション」列: 計算式が含まれており、その計算結果が表示されている。「応答をシミュレート」ウィンドウでモデルを設定すると、そのモデルに従った乱数が生成されます。計算式を表示するには、「列」パネルで列名の右側にある+記号をクリックしてください。

次の節では、二項分布に従う応答をシミュレートして、一般化線形モデルをあてはめます。

## 「応答をシミュレート」の設定

ここでは、二項分布に従う応答データをシミュレートします。成功率はロジスティックモデルに従うものとします。「応答をシミュレート」の詳細については、『実験計画(DOE)』を参照してください。

**メモ:** この節の手順を省略するには、「二項分布のシミュレート」スクリプトの横にある緑の三角ボタンをクリックします。この操作が済んだら、「一般化線形モデルのあてはめ」に進んでください。

1. 「応答をシミュレート」ウィンドウ (図10.19) で、「Y」の下に以下の値を入力します。
  - 「切片」に「0」と入力します。
  - 「X1」は、デフォルト値の「1」をそのまま使います。
  - 「X2」に「0.9」と入力します。
  - 「X3」に「0.8」と入力します。
  - 「X4」に「0.7」と入力します。
  - 「X5」に「0.6」と入力します。
  - 「X6」に「0.5」と入力します。
2. 「分布」アウトラインから「二項」を選択します。  
 「N」の値は「1」のままにしておきます。これは、それぞれの試行回数が1回だけであることを示します。

図10.20 設定後の「応答をシミュレート」ウィンドウ

効果	Y
切片	0
X1	1
X2	0.9
X3	0.8
X4	0.7
X5	0.6
X6	0.5

係数のリセット

**分布**

☐ 正規  
☒ 二項 N   
☐ Poisson

適用

3. 「適用」をクリックします。  
 データテーブルの「Yのシミュレーション」列が、二項分布に従う乱数を生成する計算式に置き換えられます。「Y 試行回数」の列には、各実験における試行回数が表示されます。
4. (オプション) 「列」パネルで、「Yのシミュレーション」の右側にある+記号をクリックします。

図 10.21 「Yのシミュレーション」に挿入される、二項分布に従う乱数の計算式

$$\text{Random Binomial} \left( 1, \frac{1}{1 + \text{Exp} \left( -1 \cdot \left( 0 + 1 \cdot X_1 + 0.9 \cdot X_2 + 0.8 \cdot X_3 + 0.7 \cdot X_4 + 0.6 \cdot X_5 + 0.5 \cdot X_6 \right) \right)} \right)$$

5. [キャンセル] をクリックします。

### 一般化線形モデルのあてはめ

ここでは、一般化線形モデルを使って、ロジスティックモデルをあてはめます。

1. データテーブルの「モデル」スクリプトの横の緑の三角ボタンをクリックします。
2. [Y] ボタンの横の変数「Y」をクリックし、[削除] ボタンをクリックします。
3. 「Yのシミュレーション」をクリックし、[Y] ボタンをクリックします。

応答変数Yとして、二項分布に従う乱数の列が設定されます。

4. 「手法」リストから[一般化線形モデル]を選択します。
5. 「分布」リストから[二項]を選択します。

[リンク関数] メニューに[ロジット]が表示されます。

6. [実行] をクリックします。

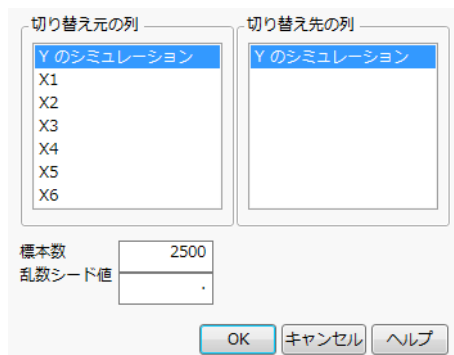
二項分布のデータ1組に対して、モデルがあてはめられます。

### 検出力の考察

この例の冒頭で、線形予測子および係数と、係数によって決められる確率を示しました。ここでは、尤度比検定のp値をシミュレートして、それらの係数に対する検出力を調べます。

1. 「効果の検定」アウトラインで、「p 値(Prob>ChiSq)」列を右クリックして[シミュレーション]を選択します。

図 10.22 「シミュレーション」ウィンドウ



「切り替え元の列」リストで「Yのシミュレーション」列が選択されていることを確認します。この列は、モデルのあてはめに使用されたデータ値を含んでいます。「切り替え先の列」でも「Yのシミュレーション」を選択すると、切り替え元の「Yのシミュレーション」データ値を、「Yのシミュレーション」の計算式で指定されている乱数に置き換えて、シミュレーションが実行されます。

レポートで選択した「**p値 (Prob>ChiSq)**」列は、「係数がゼロである」という帰無仮説に対する尤度比検定の $p$ 値です。「効果の検定」表にある、各効果に対するこの「**p値 (Prob>ChiSq)**」をシミュレートしましょう。

2. 「**標本数**」に「**500**」と入力します。
3. 「**[OK]**」をクリックします。

「一般化線形モデル シミュレーション結果」データテーブルが表示されます。

**メモ:** シミュレーションは乱数によって行われるため、結果は図10.23とは異なる可能性があります。

図 10.23 「シミュレーション結果」テーブル（一部）

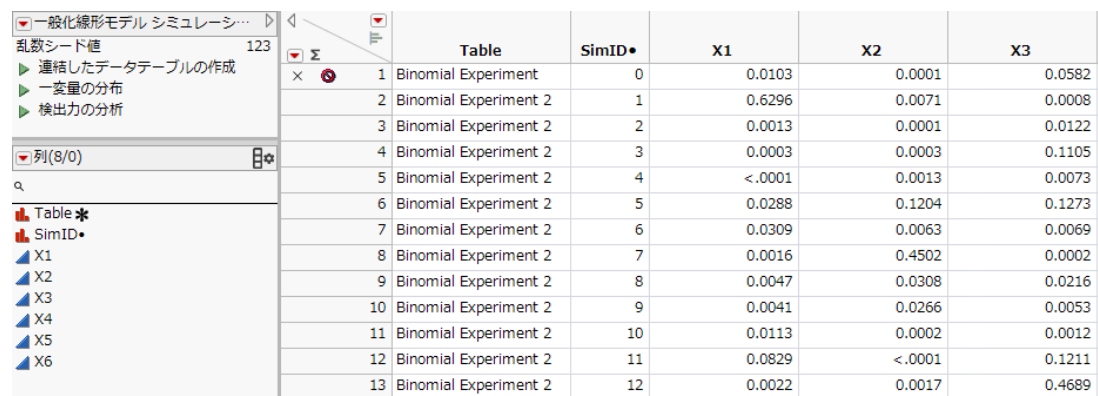


Table	SimID	X1	X2	X3
1 Binomial Experiment	0	0.0103	0.0001	0.0582
2 Binomial Experiment 2	1	0.6296	0.0071	0.0008
3 Binomial Experiment 2	2	0.0013	0.0001	0.0122
4 Binomial Experiment 2	3	0.0003	0.0003	0.1105
5 Binomial Experiment 2	4	<.0001	0.0013	0.0073
6 Binomial Experiment 2	5	0.0288	0.1204	0.1273
7 Binomial Experiment 2	6	0.0309	0.0063	0.0069
8 Binomial Experiment 2	7	0.0016	0.4502	0.0002
9 Binomial Experiment 2	8	0.0047	0.0308	0.0216
10 Binomial Experiment 2	9	0.0041	0.0266	0.0053
11 Binomial Experiment 2	10	0.0113	0.0002	0.0012
12 Binomial Experiment 2	11	0.0829	<.0001	0.1211
13 Binomial Experiment 2	12	0.0022	0.0017	0.4689

テーブルの最初の行は、観測されたデータから得られた「**p値 (Prob>ChiSq)**」であり、除外の状態になっています。2行目以降の500行は、シミュレーションしたデータから計算された結果です。

4. 「**検出力の分析**」スクリプトを実行します。

**メモ:** 応答値はシミュレートされた乱数であるため、検出力のシミュレーションの結果は、図10.24とは異なる可能性があります。

図 10.24 一変量の分布（最初の3つの効果）



各主効果のヒストグラムには、シミュレーションによって得られた500個の「p値(Prob>ChiSq)」が描かれています。また、「検出力のシミュレーション」アウトラインには、500回のシミュレーションのうち有意性検定で棄却されたものの割合が表示されます。

次に、見やすくするために、レポートを縦に積み重ねて、プロットを非表示にしてみましょう。

- 「一変量の分布」の赤い三角ボタンをクリックし、**[積み重ねて表示]** を選択します。
- Ctrlキーを押しながら、「X1」の赤い三角ボタンをクリックし、**[外れ値の箱ひげ図]** の選択を解除します。
- Ctrlキーを押しながら「X1」の赤い三角ボタンをクリックし、**[ヒストグラムオプション]** を選択し、**[ヒストグラム]** の選択を解除します。

**メモ:** 応答値はシミュレートされた乱数であるため、検出力のシミュレーションの結果は、[図 10.25](#) とは異なる可能性があります。

図 10.25 検出力（最初の3つの効果）

X1									
シミュレーション 結果					検出力のシミュレーション				
Y <sub>0</sub> = 0.01029 (元の推定値)									
信頼区間				経験的p値					
α	パーセント点 下側	パーセント点 上側	検定	p値		α	棄却された度数	棄却された割合	下側95% 上側95%
0.05	1.51e-7	0.36019	Y ≥  Y <sub>0</sub>	0.2740		0.01	362	0.724	0.68322 0.76136
0.10	7.23e-7	0.19148	Y ≤ Y <sub>0</sub>	0.7260		0.05	434	0.868	0.83551 0.89488
0.20	6.57e-6	0.08284	Y ≥ Y <sub>0</sub>	0.2740		0.10	460	0.92	0.89289 0.9407
0.50	0.00013	0.01229				0.20	477	0.954	0.93192 0.96915
X2									
シミュレーション 結果					検出力のシミュレーション				
Y <sub>0</sub> = 0.00014 (元の推定値)									
信頼区間				経験的p値					
α	パーセント点 下側	パーセント点 上側	検定	p値		α	棄却された度数	棄却された割合	下側95% 上側95%
0.05	1.08e-6	0.44811	Y ≥  Y <sub>0</sub>	0.8280		0.01	316	0.632	0.58887 0.67312
0.10	4.39e-6	0.24295	Y ≤ Y <sub>0</sub>	0.1720		0.05	407	0.814	0.77755 0.84567
0.20	2.82e-5	0.12448	Y ≥ Y <sub>0</sub>	0.8280		0.10	440	0.88	0.84858 0.90563
0.50	0.00031	0.02695				0.20	467	0.934	0.90876 0.95262
X3									
シミュレーション 結果					検出力のシミュレーション				
Y <sub>0</sub> = 0.05823 (元の推定値)									
信頼区間				経験的p値					
α	パーセント点 下側	パーセント点 上側	検定	p値		α	棄却された度数	棄却された割合	下側95% 上側95%
0.05	0.00001	0.67078	Y ≥  Y <sub>0</sub>	0.3400		0.01	209	0.418	0.37555 0.4617
0.10	0.00004	0.53067	Y ≤ Y <sub>0</sub>	0.6600		0.05	320	0.64	0.59701 0.68086
0.20	0.0002	0.36561	Y ≥ Y <sub>0</sub>	0.3400		0.10	364	0.728	0.68737 0.76516
0.50	0.00203	0.11806				0.20	417	0.834	0.79886 0.86404

「検出力のシミュレーション」アウトラインの「棄却された割合」は、シミュレーションのうち、 $p$  値が  $\alpha$  より小さくなったものの割合を示します。たとえば、係数が0.8、確率の差が38%である「X3」で、棄却された割合を見てみると、有意水準0.05の検定で  $379/500 = 0.758$  となっています。表10.1は、シミュレーションで計算された、有意水準0.05の検定に対する検出力を、効果ごとに示したものです。「検出したい差」が小さくなるにつれ、検出力も小さくなっています。たとえば、「検出したい差」が24.5%（X6）における検出力は、約0.37にとどまっています。

**メモ:** 応答値はシミュレートされた乱数であるため、検出力のシミュレーションの結果は、表10.1とは異なる可能性があります。

表 10.1 検出力のシミュレーション（有意水準 0.05）

因子	$X_i = 1$ の場合の 合格率	$X_i = -1$ の場合の 合格率	検出する差	シミュレーションで得られた検出力 ( $\alpha=0.05$ で棄却された割合)
$X_1$	73.11%	26.89%	46.2%	0.852
$X_2$	71.09%	28.91%	42.2%	0.828
$X_3$	69.00%	31.00%	38.0%	0.758
$X_4$	66.82%	33.18%	33.6%	0.654
$X_5$	64.56%	35.43%	29.1%	0.488
$X_6$	62.25%	37.75%	24.5%	0.372



# 第11章

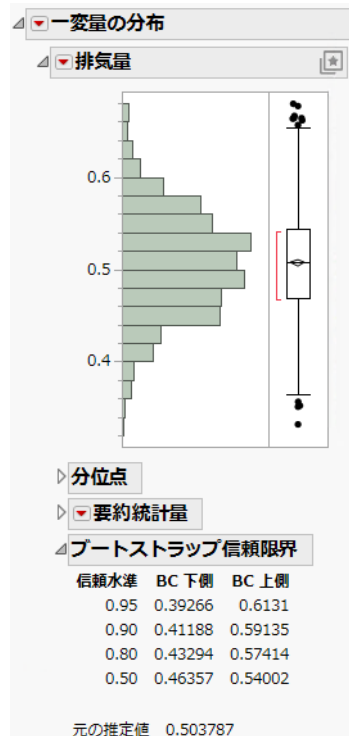
## ブートストラップ 標本再抽出によって統計量の分布を近似する

ブートストラップは、データから標本を何度も無作為抽出することにより、統計量の標本分布を近似する手法です。ブートストラップを使えば、統計量の標本分布における特徴（平均・バイアス・標準誤差・信頼区間など）を推測することができます。ブートストラップは、特に以下のような状況で役立ちます。

- 統計量の標本分布を理論的に導出するのが複雑であるか、不明である
- 推測を行うための仮定が満たされないため、パラメトリックな推測が行えない

**メモ：**「ブートストラップ」オプションは、通常の赤い三角ボタンのコマンドではなく、レポートの表を右クリック（コンテキストクリック）すると呼び出されるメニューに用意されています。

図 11.1 傾きパラメータに対して、ブートストラップを実行した結果



## 目次

ブートストラップの概要.....	355
ブートストラップを利用できる JMP プラットフォーム .....	356
ブートストラップの例 .....	357
「ブートストラップ」ウィンドウのオプション.....	359
ブートストラップ結果テーブル（積み重ね） .....	360
結果を積み重ねていないデータテーブル .....	361
ブートストラップで生成された統計量の分析.....	362
ブートストラップの別例.....	363
ブートストラップの統計的詳細 .....	368
「小数の重み」の統計的詳細 .....	368
バイアス修正済みの信頼限界の統計的詳細.....	368

## ブートストラップの概要

ブートストラップでは、得られたデータから何度も復元抽出することにより、統計量の分布を求めます。なお、JMPのブートストラップでは、「データの各行は互いに独立である」と仮定します。

単純なブートストラップでは、標本サイズが $n$ である元データから、標本サイズが $n$ であるブートストラップ標本が復元抽出されます。このブートストラップ標本には、同じデータ行が重複して抽出されたり、逆に、抽出されないデータ行があったりすることに注意してください。ブートストラップ標本において、同じデータ行が重複している回数を、「**ブートストラップの重み**」と呼びます。ブートストラップにおける反復ごとに、以下のような情報が求められ、分析全体が再実行され、統計量が再計算されます。

- ブートストラップの標本 ( $n$  個のオブザベーションから成るデータセット)
- ブートストラップの重み (分析プラットフォームの度数変数)

上記の計算を反復して、目的の統計量の分布を求めます。

ただし、単純なブートストラップが適切でない状況もあります。たとえば、データセットが小さかったり、ロジスティック回帰で分離の問題があったりする場合などです。このような場合に対して、JMPでは、ベイズ流のブートストラップ (Bayesian bootstrap) も用意されています。ベイズ流のブートストラップでは、重みが小数点以下を含むものになっています。この小数の重みは、合計すると $n$ になります。この小数の重みを、各分析用のプラットフォームで度数変数として用いることにより、目的の統計量の分布を求めることができます。小数の重みについての詳細は、「[小数の重み](#)」および「[「小数の重み」の統計的詳細](#)」を参照してください。

目的の統計量に対してブートストラップを適用するには、結果の表から統計量の列を選択し、[ブートストラップ] を選択します。

**メモ:** [ブートストラップ] オプションは、通常の赤い三角ボタンのコマンドではなく、レポートの表を右クリック (コンテキストクリック) すると呼び出されるメニューに用意されています。

JMPでは、多くの統計プラットフォームでブートストラップを使用できます。ブートストラップを使用できるプラットフォームのリストについては、「[ブートストラップを利用できる JMP プラットフォーム](#)」を参照してください。目的の統計量を計算するのに使われたデータから、標本抽出が行われます。なお、分析するとき「度数」の列が用いられている場合には、各オブザベーションはその度数だけ繰り返し発生しているものとして処理されます。また、分析するとき「重み」の列が用いられている場合には、ブートストラップにも、その重みが反映されます。

**ヒント:** ブートストラップを実行すると、元のレポートに表示されているすべての分析が、再実行されます。そのため、特定の列を選択してブートストラップを行っても、すべての分析が再実行されるために、時間がかかる場合があります。処理に時間がかかる場合は、プラットフォームのレポートから、無関係なオプションを除いてください。

## ブートストラップを利用できる JMP プラットフォーム

ブートストラップは、以下の統計プラットフォームで使用できます。

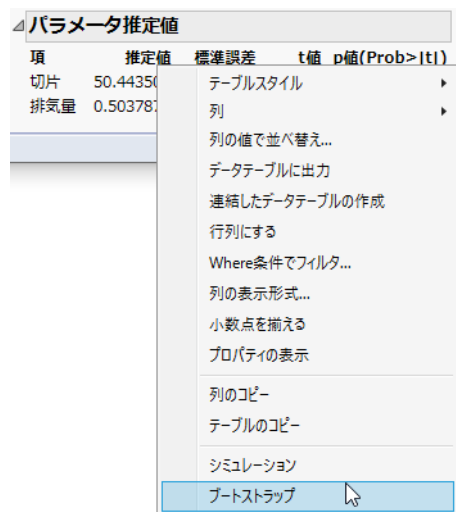
- ブースティングツリー
- ブートストラップ森
- カテゴリカル
- 破壊劣化
- 判別分析
- 一変量の分布
- 疲労モデル
- 曲線のあてはめ
- 寿命の二変量
- 生存時間(パラメトリック)のあてはめ
- 比例ハザードのあてはめ
- 二変量の関係
- 一般化線形モデル
- 一般化回帰
- 寿命の一変量
- ロジスティック
- 対数線形-分散
- 多重対応分析
- 多変量の相関
- ニューラル
- 非線形回帰
- 生存時間(パラメトリック)
- PLS 回帰
- パーティション
- 主成分分析
- 比例ハザード
- 標準最小2乗
- 構造方程式モデル
- 生存時間分析
- アップリフト

## ブートストラップの例

この例では、回帰分析で信頼限界を求めるときの前提である等分散性が満たされていません。回帰モデルの傾きに対する信頼限界が不適切な可能性があるため、ブートストラップ法を用いることにしました。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Car Physical Data.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「馬力」を選択し、[Y, 目的変数] をクリックします。
4. 「排気量」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「排気量と馬力の二変量の関係」の赤い三角ボタンをクリックし、[直線のあてはめ] を選択します。  
傾きの推定値は0.503787（約0.504）です。
7. (オプション) 「パラメータ推定値」レポートを右クリックし、[列] > [下側95%] を選択します。
8. (オプション) 「パラメータ推定値」レポートを右クリックし、[列] > [上側95%] を選択します。  
回帰分析における傾きの信頼限界は、0.4249038および0.5826711となっています。
9. 「パラメータ推定値」レポートの「推定値」列を右クリックし、[ブートストラップ] を選択します。

図11.2 [ブートストラップ] オプション

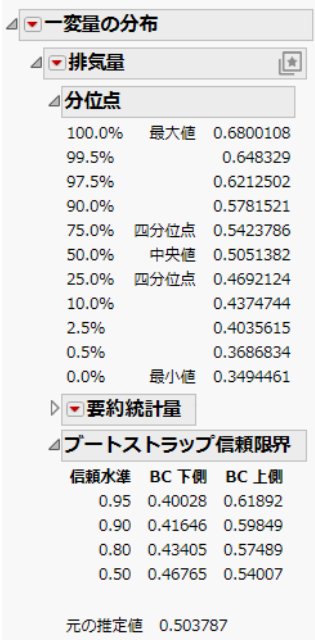


【選択された列の分割】 オプションを選択する場合は、どの列で右クリックしたかによって、分析対象となる列が変わる点に注意してください。「[「ブートストラップ」ウィンドウのオプション](#)」を参照してください。

10. 「ブートストラップ標本数」に「1000」と入力します。
11. (オプション) 図11.3の結果と一致させるには、「乱数シード値」に「12345」と入力します。

12. [OK] をクリックします。
- ブートストラップ処理が行われ、傾きと切片の結果を積み重ねていない「ブートストラップ結果」データテーブルが作成されます。
- 次に、ブートストラップによって生成された傾きの分布を分析します。
13. 「ブートストラップ結果」テーブルで、「一変量の分布」スクリプトを実行します。
- 「一変量の分布」レポートには、「ブートストラップ信頼限界」レポートが含まれます。

図 11.3 ブートストラップのレポート



傾きの推定値（手順6）は、0.504です。ブートストラップで求めた95%の信頼水準に基づくと、傾きは0.40028～0.61892であると推定できます。つまり、排気量が1単位増加すると、95%の信頼水準で、馬力は0.40028～0.61892だけ増加すると言えます。ブートストラップによる傾きの信頼区間（0.400～0.619）は、通常の回帰分析（手順7および手順8）による信頼区間（0.425～0.583）に比べ、わずかに広がっています。

**メモ：**「ブートストラップ信頼限界」レポートの「BC 下側」列および「BC 上側」列は、**バイアス修正済み（bias-corrected）の信頼限界**です。[「バイアス修正済みの信頼限界の統計的詳細」](#)を参照してください。

## 「ブートストラップ」ウィンドウのオプション

ブートストラップを実行するには、レポートウィンドウ内の表において、統計量を含む数値列を右クリックし、メニューから**「ブートストラップ」**を選択します。選択した列が強調表示され、「ブートストラップ」ウィンドウが表示されます。このウィンドウでオプションを選択し、**[OK]** をクリックすると、選択した列のすべての統計値に対してブートストラップが実行され、結果を含むデフォルトのデータテーブルが作成されます。

---

**メモ:** 「ブートストラップ」オプションは、By 変数を使用したレポートでは使用できません。

---

「ブートストラップ」ウィンドウには、次のオプションがあります。

**ブートストラップ標本数** データからの無作為抽出と、そこからの統計量の計算といった一連の処理を反復する回数を設定します。回数を増やせば、それだけ統計量の状態をより正確に推定することができます。デフォルトでは、2500 に設定されています。

**乱数シード値** 後でブートストラップ分析を再実行するときに、同じ結果を再現したい場合は、同じ乱数シード値を入力してください。デフォルトでは、シード値は設定されていません。

**小数の重み** バイズ流のブートストラップ分析を実行します。ブートストラップを実行するたびに、各データ行に対する重みが計算されます（重みの計算方法については、「[「小数の重み」の統計的詳細](#)」を参照してください）。この重みを使って、目的の統計量が計算されます。デフォルトでは、このオプションは選択されておらず、通常の単純なブートストラップが実行されます。

---

**ヒント:** 標本サイズが小さかったり、ロジスティック回帰で分離の可能性があったりする場合は、「[小数の重み](#)」オプションを使用してください。

---

「[小数の重み](#)」オプションを選択すると、「[小数の重み](#)」では、ブートストラップの各抽出において、各データ行に対して小数の重みが求められます。なお、これらの重みの合計は、目的の統計量を計算したデータの標本サイズ  $n$  になっています。重みの計算方法の詳細については、「[「小数の重み」の統計的詳細](#)」を参照してください。

**選択された列の分割** ブートストラップで得られた統計量を、結果のデータテーブルにおいて列ごとに**分割**します。このデータテーブルでは、各行が、ブートストラップ標本1組に対応しています（ただし、最初の1行目は、観測されたデータに対するものです）。

このオプションをオフにすると、積み重ねた形式のデータテーブルだけが作成されます。この形式では、ブートストラップにおける抽出ごとに、元の表全体が縦に積み重ねられていきます。この形式では、元の表における各行が、結果のデータテーブルにおける複数の各行に対応しています。そして、元の表における各列が、結果のデータテーブルにおける各列と対応しています。例として、「[ブートストラップ結果テーブル（積み重ね）](#)」を参照してください。

**分割が成功したら積み重ねたテーブルを破棄**（[選択された列の分割] オプションを選択した場合のみ選択可能）このオプションによって、作成されるデータテーブルの数が決まります。

このチェックボックスを**オフ**にした場合は、以下の2つのデータテーブルが作成されます。

- 積み重ねた形式のデータテーブル。元の表を積み重ねた形式で、表のすべての列に対する結果を含んでいます。
- 分割した形式のデータテーブル。積み重ねた形式のデータテーブルを分割し、元の表で選択した列に対する結果だけを含んでいます。

このチェックボックスを**オン**にすると、[選択された列の分割] の処理が成功した場合は、積み重ねた形式のデータテーブルは破棄されます。

## ブートストラップ結果テーブル（積み重ね）

積み重ねた形式のブートストラップ結果テーブルには、統計量とブートストラップの抽出の組み合わせの数だけ行があり、1つの列にブートストラップの結果がまとめられた状態になっています。デフォルトでは、ブートストラップを行うと、結果が縦に積み重なったデータテーブルが作成されます（[図 11.4](#)）。ただし、[分割が成功したら積み重ねたテーブルを破棄] オプションを選択すると、多くの場合このデータテーブルは破棄されるため、表示されません。

[図 11.4](#)は、「Car Physical Data.jmp」を「二変量の関係」プラットフォームで分析し、「パラメータ推定値」レポートに対してブートストラップを実行した結果です。「[ブートストラップの例](#)」を参照してください。

図 11.4 ブートストラップ結果テーブル（積み重ね）

	Σ	項	バイアス	推定値	標準誤差	t値	p値(Prob> t )	BootID●
×	⊗	1 切片		50.443509471	6.7436209999	7.48	<.0001	0
×	⊗	2 排気量		0.5037874592	0.0398202611	12.65	<.0001	0
		3 切片		45.385604758	5.4724273046	8.29	<.0001	1
		4 排気量		0.5451674092	0.0306151772	17.81	<.0001	1
		5 切片		40.843862813	7.0508187326	5.79	<.0001	2
		6 排気量		0.5854988173	0.0457041828	12.81	<.0001	2
		7 切片		47.765642104	5.4677610087	8.74	<.0001	3
		8 排気量		0.4943459579	0.0305765677	16.17	<.0001	3
		9 切片		59.758060908	7.0436720036	8.48	<.0001	4
		10 排気量		0.4385540785	0.042053201	10.43	<.0001	4
		11 切片		66.341062413	6.4663383596	10.26	<.0001	5
		12 排気量		0.3969640071	0.0393100058	10.10	<.0001	5
		13 切片		41.234989734	5.8543901826	7.04	<.0001	6
		14 排気量		0.5723548142	0.034172131	16.75	<.0001	6
		15 切片		43.876867815	7.8085946052	5.62	<.0001	7
		16 排気量		0.5731377467	0.0439823115	13.03	<.0001	7



縦に積み重なったデータテーブルには、次のような特徴があります。

- 各ブートストラップ標本に対して、元の表における1列目で示されている統計量の種類ごとに、別々の行が生成されます。表におけるこの1列目の列（統計量の種類を示す列）は、結果のデータテーブルでも保存されます。この例では、各ブートストラップ標本に対し、元の表における「項」列における「切片」と「排気量」の結果が、別々の行に保存されています。
- 分析で使ったデータテーブル列が含まれます。この例では、Xの「排気量」と、Yの「馬力」です。
- 元の表にあるすべての列が、結果のデータテーブルにも保存されます。この例では、「～バイアス」・「推定値」・「標準誤差」・「t値」・「p値（Prob>|t|）」といった列があります。ここで、「～バイアス」は表にある列なのですが、今回の分析結果にはバイアスが生じているパラメータ推定値はないため、いずれも空白になっている点に注意してください。
- 「BootID・」列は、ブートストラップ標本の識別IDです。「BootID・」が0の行には、元の推定値が含まれています。この行には、×印のマーカと、「除外」の行属性が与えられます。この例では、各ブートストラップ標本につき、結果が2行ずつ（「切片」と「排気量」）生成されています。
- このデータテーブル名の末尾は、「ブートストラップ結果（積み重ね）」となっています。

【選択された列の分割】 オプションを選択した場合、結果を積み重ねていないデータテーブルも表示されます。  
「結果を積み重ねていないデータテーブル」を参照してください。

## 結果を積み重ねていないデータテーブル

ブートストラップ結果を積み重ねていないデータテーブルには、抽出の数だけ行があり、ブートストラップの結果ごとに個別の列があります。ブートストラップを実行するときに、【選択された列の分割】を選択すると、結果を積み重ねていないデータテーブルが作成されます。

この例では、図11.4（積み重ね）にある項の推定値が、図11.5（積み重ねていない）では2つの列（「排気量」、「切片」）に分割されています。

図11.5 結果を積み重ねていないデータテーブル

		X	Y	Table	BootID・	切片	排気量
<div> <div>排気量による馬力の二変量 ブー...</div> <div>乱数シード値 12345</div> <div>連結したデータテーブルの作成</div> <div>一変量の分布</div> </div> <div>列(6/0)</div> <div> <div>X *</div> <div>Y</div> <div>Table</div> <div>BootID・</div> <div>切片</div> <div>排気量</div> </div>	×	1 排気量	馬力	Car Physical Data	0	50.443509471	0.5037874592
		2 排気量	馬力	Car Physical Data	1	45.385604758	0.5451674092
		3 排気量	馬力	Car Physical Data	2	40.843862813	0.5854988173
		4 排気量	馬力	Car Physical Data	3	47.765642104	0.4943459579
		5 排気量	馬力	Car Physical Data	4	59.758060908	0.4385540785
		6 排気量	馬力	Car Physical Data	5	66.341062413	0.3969640071
		7 排気量	馬力	Car Physical Data	6	41.234989734	0.5723548142
		8 排気量	馬力	Car Physical Data	7	43.876867815	0.5731377467
		9 排気量	馬力	Car Physical Data	8	33.125519594	0.637861847
		10 排気量	馬力	Car Physical Data	9	45.293111747	0.5441967613
		11 排気量	馬力	Car Physical Data	10	45.87058516	0.5423927826
		12 排気量	馬力	Car Physical Data	11	50.686966854	0.5137395469

積み重ねていない形式の結果のデータテーブルには、次のような特徴があります。

- ブートストラップ標本ごとに、1つの行が作成されます。
- 分析で使ったデータテーブル列が含まれます。この例では、Xの「排気量」と、Yの「馬力」です。
- 元の表における各行が、結果のデータテーブルにおいて、別々の列になります。
- 「ブートストラップ」ウィンドウで乱数シード値を指定すると、結果のデータテーブルに、「乱数シード値」というテーブル変数とその値が表示されます。
- 結果を積み重ねていないテーブルには、「ソース」および「一変量の分布」というテーブルスクリプトがあります。「一変量の分布」というテーブルスクリプトを実行すると、このブートストラップ標本に基づく統計量（ブートストラップ信頼限界など）をすばやく得ることができます。
- 「BootID•」列は、ブートストラップ標本の識別 ID です。「BootID•」が 0 の行には、元の推定値が含まれています。この行には、×印のマーカースと、「除外」の行属性が与えられます。積み重ねていない形式のデータテーブルでは、各ブートストラップ標本につき、1つの行が作成されます。
- データテーブル名の末尾は、「ブートストラップ結果 (<列名>）」となっています（ここで、<列名>は、ブートストラップを実行した列名です）。

---

## ブートストラップで生成された統計量の分析

ブートストラップの結果は、「一変量の分布」プラットフォームを使って分析します。

- 積み重ねていない形式のデータテーブルにおいては、生成されたデータテーブルに「一変量の分布」スクリプトが保存されるので、それを実行できます。
- 積み重ねた形式のデータテーブルにおいては、[分析] > [一変量の分布] を選択し、列を選択して役割を割り当てます。ほとんどのケースでは、元の表で1列目になっているものを [BY] に割り当てればうまくいきます。

「一変量の分布」レポートにおいては、2種類のブートストラップ信頼限界が計算されます。「BootID•」列があるデータテーブルに対して「一変量の分布」プラットフォームを実行すると、「ブートストラップ信頼限界」レポートが自動的に生成されます（[図 11.6](#)）。

「一変量の分布」レポートには、2種類のブートストラップ信頼限界が計算されています。

- 「分位点」レポートにおいて、通常の**分位点によるブートストラップ信頼区間**を知ることができます。通常の分位点に基づくたとえば両側 95% 信頼区間の下限と上限は、2.5% と 97.5% の分位点です。
- また、「ブートストラップ信頼限界」レポートには、**バイアス修正済みのブートストラップ信頼限界**が表示されます。信頼水準が 95%、90%、80%、50% の両側信頼区間が示され、「BC 下側」列および「BC 上側」列に、それぞれ下限と上限が表示されます。バイアス修正済みのブートストラップ信頼限界の計算については、「[バイアス修正済みの信頼限界の統計的詳細](#)」を参照してください。

図 11.6 「ブートストラップ信頼限界」レポート

「ブートストラップ信頼限界」レポートの下にある「**元の推定値**」は、観測された元データから算出された推定値です。

「ブートストラップ信頼限界」レポートの解釈の詳細については、「[ブートストラップの概要](#)」を参照してください。分位点に基づいて信頼区間を求める方法（バイアス修正をしないもの、およびバイアス修正を行うもの）については、Efron（1981）が説明しています。

## ブートストラップの別例

この例では、標本サイズの小さいデータに対して、「小数の重み」オプション（ベイズ流のブートストラップ法）を使う利点を説明します。使用するデータは、7種類の異なる土壌ごとに3つのサンプルを採取し、応答Yを測定したものです。ここで、土壌「wabash」における平均に対する信頼区間を求めるとします。

各土壌について、3つしか測定値がないので、単純なブートストラップ法では、抽出されたブートストラップ標本に、土壌が「wabash」であるデータが1つも含まれない場合が出てきます。「小数の重み」オプションを使えば、各ブートストラップ標本において、常に「wabash」のデータが含まれます。

「wabash」の標本について、平均値の分布を2つのブートストラップ法で調べてみましょう。

- 「[単純なブートストラップ分析](#)」
- 「[ベイズ流のブートストラップ法](#)」

## 単純なブートストラップ分析

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Snapdragon.jmp」を開きます。
2. [分析] > [二変量の関係] を選択します。
3. 「Y」を選択し、[Y, 目的変数] をクリックします。
4. 「土壌」を選択し、[X, 説明変数] をクリックします。
5. [OK] をクリックします。
6. 「土壌によるYの一元配置分析」の赤い三角ボタンをクリックし、[平均/ANOVA] を選択します。
7. 「各水準の平均」レポートで、「平均」列を右クリックし、[ブートストラップ] を選択します。
8. 「ブートストラップ標本数」に「1000」と入力します。
9. (オプション) 図11.7の結果と一致させるには、「乱数シード値」に「12345」と入力します。
10. [OK] をクリックします。

図11.7 単純なブートストラップの結果

	Table	X	Y	BootID*	clarion	clinton	compost	knox	o'neill	wabash	webster
1	Snapdragon	土壌	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	Snapdragon	土壌	Y	1	32.2333	30.9000	28.0000	35.7000	34.3500	31.9000	31.1000
3	Snapdragon	土壌	Y	2	32.4333	30.3000	30.2000	33.9667	31.2000	38.0000	•
4	Snapdragon	土壌	Y	3	32.5000	30.7500	29.2000	34.0333	32.8000	38.0000	31.1000
5	Snapdragon	土壌	Y	4	31.5000	29.4000	29.9000	35.7500	35.0400	34.9500	•
6	Snapdragon	土壌	Y	5	32.5000	30.6000	31.8000	35.7000	34.3000	35.0500	31.5667
7	Snapdragon	土壌	Y	6	32.4000	30.1000	29.8500	34.4500	36.0000	38.2000	30.4000
8	Snapdragon	土壌	Y	7	31.9000	29.3400	•	34.4000	33.6000	35.9667	31.4500
9	Snapdragon	土壌	Y	8	32.1400	31.3500	29.6667	33.1000	35.1000	37.9333	30.6333
10	Snapdragon	土壌	Y	9	32.7000	30.8000	30.2000	35.2600	31.2000	34.8500	32.5000
11	Snapdragon	土壌	Y	10	32.1000	32.1000	28.0000	33.1000	34.1143	34.0000	31.8000
12	Snapdragon	土壌	Y	11	31.5000	30.3000	•	33.1000	34.4000	37.2125	30.6333
13	Snapdragon	土壌	Y	12	32.7000	30.4200	31.8000	34.0333	35.1000	35.0500	•
14	Snapdragon	土壌	Y	13	32.1667	•	30.1000	34.9000	33.9000	38.2000	31.8000
15	Snapdragon	土壌	Y	14	32.3000	•	29.2000	33.9667	34.4000	35.6000	31.9400
16	Snapdragon	土壌	Y	15	32.2500	29.1000	29.6667	33.1000	35.1000	31.9000	•
17	Snapdragon	土壌	Y	16	31.5000	30.4200	30.0000	•	35.2800	37.8000	31.1000
18	Snapdragon	土壌	Y	17	32.4333	29.7000	29.2000	33.1000	34.4000	34.0000	31.1000
19	Snapdragon	土壌	Y	18	32.3800	32.1000	29.9000	35.8000	34.6800	35.0500	31.8000
20	Snapdragon	土壌	Y	19	32.0600	30.1000	30.2800	•	33.1500	31.9000	31.1000
21	Snapdragon	土壌	Y	20	31.9000	29.1000	28.4000	35.9000	34.8000	34.8500	30.8200

図11.7において欠測値となっている個所は、そのときのブートストラップ標本の抽出において、該当する土壌タイプのデータ値が1つも選択されなかったことを意味します。

11. [分析] > [一変量の分布] を選択します。
12. 「wabash」を選択し、[Y, 列] をクリックします。
13. [OK] をクリックします。

図 11.8 「wabash」の平均値の分布（ベイズ流ブートストラップ）

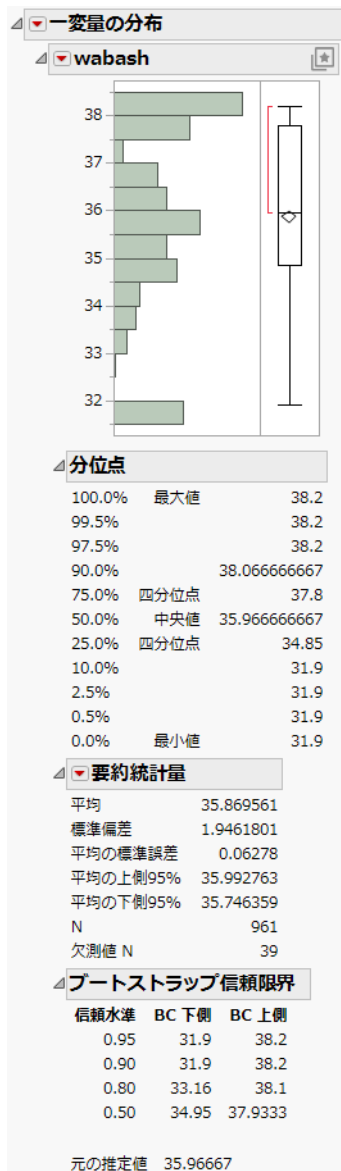


図 11.8 は、「wabash」の平均値の分布を、単純なブートストラップ分析で求めた結果です。次の点に注意してください。

- 「要約統計量」レポートによると、ブートストラップを実行した結果、「wabash」の平均値を含む行数は、 $N = 961$  でした。これは、ブートストラップを 1,000 回実行し、そのうち 39 回では、「wabash」の 3 つの測定値がいずれもブートストラップ標本に選ばなかったということを意味しています。

- 標本の平均値のヒストグラムは滑らかではなく、両端にピークの値があります。「wabash」の3つの値は、38.2, 37.8, 31.9です。分布の下端のピークは、31.9のみがブートストラップ標本に含まれた場合の結果です。分布の上端のピークは、38.2と37.8だけがブートストラップ標本に含まれた場合の結果です。

次に、ブートストラップ標本が無い状態になるのを回避するために、[小数の重み] オプション（ベイズ流のブートストラップ法）を使用してみましょう。

### ベイズ流のブートストラップ法

- 「一元配置分析」レポートの「各水準の平均」レポートで、「平均」を右クリックし、[ブートストラップ]を選択します。
- 「ブートストラップ標本数」に「1000」と入力します。
- （オプション）図11.9の結果と一致させるには、「乱数シード値」に「12345」と入力します。
- [小数の重み] チェックボックスをオンにします。
- [OK] をクリックします。

図11.9 ベイズ流ブートストラップによる結果

	Table	X	Y	BootID*	clarion	clinton	compost	knox	o'neill	wabash	webster
1	Snapdragon	土壌	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	Snapdragon	土壌	Y	1	31.9365	31.3493	29.3234	35.4497	34.7105	33.7270	31.4071
3	Snapdragon	土壌	Y	2	32.4189	30.3474	29.5143	35.6281	32.7006	34.1027	32.1212
4	Snapdragon	土壌	Y	3	32.2339	30.2001	31.4102	34.8758	33.6674	38.0389	31.5428
5	Snapdragon	土壌	Y	4	32.4054	30.3242	30.6227	33.8495	32.9495	36.8344	31.9607
6	Snapdragon	土壌	Y	5	32.2262	30.8672	29.7999	33.6759	32.2022	35.5792	31.9058
7	Snapdragon	土壌	Y	6	32.3222	31.9732	28.8823	35.6307	34.8863	35.3014	30.1325
8	Snapdragon	土壌	Y	7	31.9948	30.8828	29.2516	35.3424	33.2094	36.3367	30.8386
9	Snapdragon	土壌	Y	8	31.6254	29.7677	28.6390	34.4697	33.0662	36.6183	31.4667
10	Snapdragon	土壌	Y	9	32.3499	29.9416	29.5732	35.2564	31.9583	35.8246	29.9302
11	Snapdragon	土壌	Y	10	32.5228	30.4506	28.4859	34.9088	35.8126	34.6317	31.6100
12	Snapdragon	土壌	Y	11	31.9057	30.0711	29.0693	35.4018	34.4654	33.2086	31.0309
13	Snapdragon	土壌	Y	12	31.7275	29.6189	29.1609	34.3984	33.6840	35.0815	31.5446
14	Snapdragon	土壌	Y	13	32.5893	30.5210	28.6054	33.4594	34.0958	33.7692	31.7129
15	Snapdragon	土壌	Y	14	32.2473	30.5199	31.6080	35.5617	34.1706	35.7146	31.8023
16	Snapdragon	土壌	Y	15	32.2329	29.7275	30.1770	35.3286	32.7820	37.4866	30.9706
17	Snapdragon	土壌	Y	16	31.5831	29.7508	28.7122	33.5304	34.5348	37.1092	31.2752
18	Snapdragon	土壌	Y	17	32.3545	31.3237	29.1542	35.5890	32.2606	37.2005	30.8430
19	Snapdragon	土壌	Y	18	32.3811	29.6241	30.6138	35.4308	33.2024	33.0787	31.2926
20	Snapdragon	土壌	Y	19	31.7488	29.7763	28.7327	34.7007	33.8910	34.0573	29.9064

ベイズ流ブートストラップでは、結果のデータテーブルに欠測値はありません。各標本抽出において、「Snapdragon.jmp」データテーブルの21行がすべて含まれるように、ブートストラップの重みが設定されています。

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

図 11.10 「wabash」の平均値の分布（ベイズのブートストラップ）



ベイズ流のブートストラップ法による結果では、「wabash」の平均が、より滑らかに分布しています。1,000個のブートストラップ標本のすべてに、「wabash」の3つの測定値が含まれています。ベイズ流のブートストラップ法では、各標本抽出において、異なる小数の重みを適用して、「wabash」の標本平均を求めています。

「ブートストラップ信頼限界」レポートでは、平均値に対する両側95%信頼区間は、32.6396～37.8168となっています。



## ブートストラップの統計的詳細

この節では、ブートストラップの統計的詳細について説明します。

- 「[小数の重み](#)」の統計的詳細
- 「[バイアス修正済みの信頼限界の統計的詳細](#)」

### 「小数の重み」の統計的詳細

この節では、ブートストラップ分析における小数の重みの計算方法を解説します。[小数の重み] オプションをオンにした場合は、ベイズ流のブートストラップ (Bayesian bootstrap; Rubin 1981) が実行されます。ある 1 つのブートストラップ標本において、同じデータ行が重複して抽出される回数を、「**ブートストラップの重み**」と呼びます。通常のブートストラップ法では、各ブートストラップの重みは、単純な復元抽出法によって決められます。

ベイズ流のブートストラップ法では、ブートストラップ抽出における抽出確率を確率変数とみなし、無情報の事前分布を仮定して、その事後分布を求めます。そして、抽出確率の乱数を、その事後分布から生成します。その抽出確率から、ブートストラップの重みを求めます。具体的には、以下の手順に従います。

- 形状パラメータが  $(n - 1)/n$ 、尺度パラメータが 1 である、独立なガンマ分布に従う  $n$  個の乱数を生成する。

---

**メモ:** Rubin (1981) は、ガンマ分布の形状パラメータとして 1 を使用しています。JMP で採用している形状パラメータは、重みの平均と分散が、単純なブートストラップの重みの平均と分散に等しくなるように設定されています。

---

- $S$  ( $n$  個の乱数の和) を計算する。
- $n$  個の乱数のそれぞれに  $N/S$  を掛けたものを、小数の重みとする ( $N$  は行数、または、度数変数が指定されている場合は度数の合計)。

---

**メモ:** 分析において度数変数を使用した場合には、行ごとに、ガンマ分布の形状パラメータに、度数を掛け合わせます。ただし、度数の合計は、1 より大きくなっているものとします。すると、形状パラメータは  $f_i(N - 1)/N$  に等しくなります (ここで、 $f_i$  は  $i$  番目の行の度数の値、 $N$  は度数の値の合計です)。

---

この手順で計算された各行の小数の重みは、その平均と分散が、単純なブートストラップの重みの平均と分散と等しくなっています。また、この手順で計算された各行の小数の重みは、正の値であり、すべてを合計すると  $N$ 、その平均は 1 となっています。

### バイアス修正済みの信頼限界の統計的詳細

この節では、バイアス修正済み (bias-corrected) のブートストラップ信頼限界 (以下、「BC 信頼限界」と述べます) について説明します。BC 信頼限界は、結果のデータテーブルで「一変量の分布」スクリプトを実行すると、「ブートストラップ信頼限界」レポートに示されます。BC 信頼限界は、ブートストラップ分布の非対称性に対処して、分位点の信頼限界を補正しています。Efron (1981) を参照してください。



## 表記

- $p^*$ は、ブートストラップ標本の推定値が、元の推定値以下となっている割合。
- $z_0$ は、標準正規分布の $p^*$ 分位点。
- $z_\alpha$ は、標準正規分布の $\alpha$ 分位点。

## BC 信頼限界の両端

信頼水準が $(1 - \alpha)$ である BC 信頼限界は、ブートストラップ分布における以下の分位点です。

- 下限は、次の累積確率に対する分位点。

$$\Phi\left(2z_0 + z_{\frac{\alpha}{2}}\right)$$

- 上限は、次の累積確率に対する分位点。

$$\Phi\left(2z_0 + z_{1 - \frac{\alpha}{2}}\right)$$



# 第12章

## テキストエクスプローラ 非構造化テキストを探索する

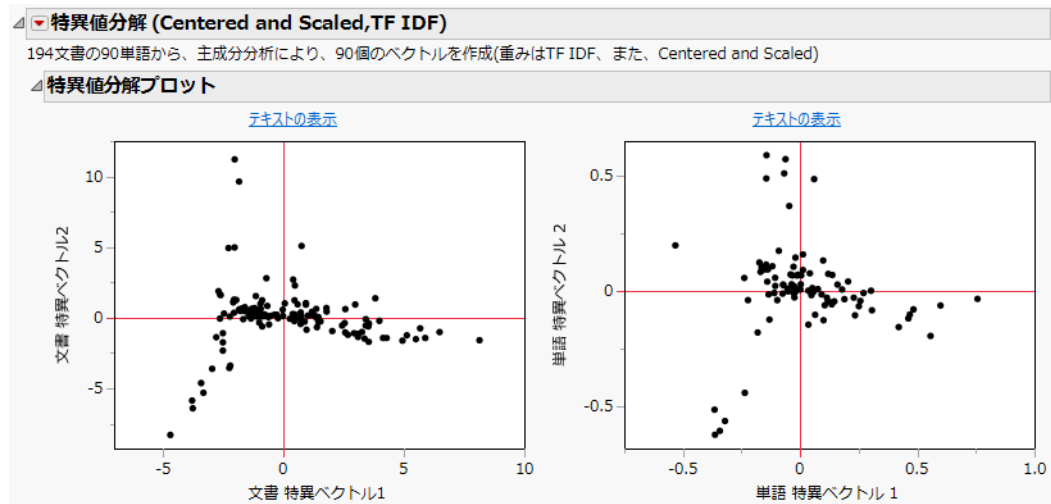
**JMP PRO** このプラットフォームの機能の多くは、JMP Pro だけの機能です。JMP Pro だけで動作する機能の説明には、このマークが付いています。

「テキストエクスプローラ」は、アンケート調査の自由回答や、医療現場のインシデントレポートなど、構造化されてないテキストデータを分析するためのプラットフォームです。このツールを使って、似ている単語を1つにまとめたり、誤っている語句を再コード化したりできます。また、テキストデータにおける単語の出現頻度を把握できます。

**JMP PRO** JMP Pro では、特異値分解に基づいて、似ている文書をトピックごとに分類することができます。特異値分解の結果をもとに、同じ単語が使われている文書や、1つの文書内で一緒に使われていることが多い単語を、クラスターにまとめることができます。また、潜在クラス分析によっても文書をクラスターに分けることができます。

**JMP PRO** JMP Pro では、文書に含まれる重要な単語や感情を特定することもできます。単語選択分析は、ある応答変数を最もよく説明する単語を特定します。感情分析は、辞書に基づいて文書内の感情語を特定し、それらの語に対して、肯定的・否定的・全体的なスコアを付けます。感情分析は、基本的な自然言語処理（NLP）にも対応しています。ただし、感情分析は英語文書しかサポートしていません。

図12.1 テキストエクスプローラの特異値分解プロット



## 目次

「テキストエクスプローラ」プラットフォームの概要 .....	373
テキスト処理の手順 .....	375
「テキストエクスプローラ」プラットフォームの例 .....	376
「テキストエクスプローラ」プラットフォームの起動 .....	380
正規表現エディタでのRegexのカスタマイズ .....	382
「テキストエクスプローラ」レポート .....	387
要約レポート .....	388
単語と句のリスト .....	388
「テキストエクスプローラ」プラットフォームのオプション .....	392
テキストを前処理するオプション .....	392
テキストの統計分析に関するオプション .....	398
保存に関するオプション .....	400
テキストエクスプローラのレポートオプション .....	401
潜在クラス分析 .....	402
潜在意味分析（特異値分解） .....	403
特異値分解レポート .....	404
特異値分解レポートのオプション .....	405
トピックの分析 .....	407
「トピック分析」レポート .....	408
「トピック分析」レポートのオプション .....	408
線形判別分析 .....	409
「判別分析」レポート .....	410
「判別分析」レポートのオプション .....	410
単語選択分析 .....	411
単語選択分析の設定 .....	411
「単語選択分析」レポート .....	412
「単語選択分析」レポートのオプション .....	414
感情分析 .....	415
「感情分析」レポート .....	416
「感情分析」レポートのオプション .....	418
「テキストエクスプローラ」プラットフォームの別例 .....	420

## 「テキストエクスプローラ」プラットフォームの概要

「テキストエクスプローラ」プラットフォームは、構造化されていないテキストを分析し、その意味を把握する助けになります。構造化されていないテキストデータは珍しくありません。たとえば、アンケートの自由回答欄や、製品のレビューコメント、インシデントレポートといった例があります。

テキスト分析は、通常、単語の整理（キュレーティング）と、統計分析とを交互に行いながら進めていきます。

### 単語の整理

テキスト分析では、独自の用語が使用されます。**単語（term, token）**はテキストを構成する最小単位で、文章におけるワード（word）のようなものです。単語を定義する方法は、正規表現の使用をはじめ、多数の方法があります。テキストを単語に分解することを、**トークン化（tokenization）**と呼びます。

- **句**は少数の単語から構成されます。このプラットフォームには、句を単語として扱うオプションも用意されています。
- **文書**は単語の集合です。JMP のデータテーブルでは、テキスト列に含まれる各行の非構造化テキストが、1つの文書に対応します。
- **コーパス**は複数の文書の集合を指します。

通常は、いくつかの単語を除外して分析したほうがよいでしょう。これらの除外する単語は、「**ストップワード**」と呼ばれます。「テキストエクスプローラ」プラットフォームには、いくつかのストップワードが予め定義されていますが、分析者が自分でストップワードを追加することもできます。ストップワードは、単語としては除外されますが、句の中には含まれていても構いません。ただし、ストップワードで始まる句やストップワードで終わる句は除外されます。

単語を再コード化することもできます。この機能は、同義語を1つにまとめるのに便利です。

**語幹抽出**とは、単語の語尾の変化を取り除き、**語幹**が共通する単語をまとめる処理のことを指します。たとえば、「jump」、「jumped」、「jumping」は、すべて「jump」として扱われます。JMPで行われる語幹抽出は、文字列処理言語のSnowballで使われている方法と似ています。句に対しては、単語単体に対する語幹抽出と同じように、句に含まれる各単語を語幹抽出します。

---

**メモ：**「テキストエクスプローラ」起動ウィンドウで東アジア言語を指定した場合、単語整理の段階は上記の説明とは異なります。東アジア言語に対しては語幹抽出が使用できないため、違う方法でトークン化が行われます。「[東アジア言語のトークン化](#)」を参照してください。

---

## 単語リストの分析

「テキストエクスプローラ」プラットフォームでは、テキスト分析に **bag of words** (単語の袋) というアプローチを採用しています。このアプローチでは、句における構成を除き、語順は考慮されません。分析は、単語の出現頻度だけに基づいて行われます。

正規表現 (Regex)、ストップワード、再コード化、語幹抽出などの整理 (キュレーティング) を行って単語リストを作成したら、その整理した単語リストに対して統計分析を行ってください。このプラットフォームでは、**文書単語行列** (DTM; Document-Term Matrix) に対する統計分析を行えます。文書単語行列の各行は、1つの文書に対応しています (JMP データテーブルにおけるテキスト列の1セルが、1文書に対応しています)。文書単語行列の各列は、整理した後の各単語に対応しています。文書単語行列に対する統計分析では、語順を無視する **bag of words** アプローチを採用しています。もっとも単純な場合では、文書単語行列は、単語の出現頻度となっています。それ以外の重みを用いる方法もいくつかあり、これについては「[保存に関するオプション](#)」で説明します。

**JMP PRO** このプラットフォームの統計分析では、最初に文書単語行列に対して特異値分解 (SVD; Singular Value Decomposition) を実行します。これにより、単語情報に関するデータの次元を減らすことができます。特異値分解の詳細については、『多変量分析』を参照してください。また、階層型クラスター分析によって、単語や文書をクラスターリングできます。クラスター分析によって、似たような単語や文書をグループにまとめることができます。

## 「テキストエクスプローラ」プラットフォームの分析手順

「テキストエクスプローラ」プラットフォームの通常の手順は以下のとおりです。

1. どのようにトークン化を行うかを指定します。予め用意されている正規表現や、自分でカスタマイズした正規表現などを指定します。
2. レポート上で、ストップワードの追加や、単語リストへの句の追加、単語の再コード化、語幹抽出の例外指定などを行います。
3. 語幹抽出の方法を指定します。

---

**メモ:** 東アジア言語では語幹抽出を使用できません。

---

4. 単語と句の出現頻度、特異値分解、クラスター分析などの手法を使って、重要な単語や句を見極めます。

---

**メモ:** **JMP PRO** 特異値分解とクラスター分析は、JMP Proでのみ使用できます。

---

5. 必要に応じて、単語テーブル、文書単語行列、特異ベクトルなどの結果を保存し、それらをもとに追加の分析を行います。

---

**メモ:** **JMP PRO** 特異ベクトルを保存するオプションは、JMP Proでのみ使用可能です。

---

6. 場合によっては、句、再コード化、ストップワードの設定を、再分析するときに利用できるよう保存します。

## テキスト処理の手順

「テキストエクスプローラ」プラットフォームのテキスト処理は、トークン化、句の抽出、単語の抽出という3つの段階に分けられます。

### トークン化

この段階では次の処理が実行されます。

1. 英語などの欧米言語のアルファベットに対しては、大文字を小文字に変換します。
2. 「基本ワード」によるトークン化、または正規表現（[Regex]）に基づくトークン化を実行し、文字をトークンに分けます。
3. 指定された再コード化の定義に従って、トークンを再コード化します。再コード化は、語幹抽出の前に行われることに注意してください。

---

**メモ:** トークンの再コード化は、レポートウィンドウで指定した順序に関わらず、内部で一括処理されます。

---

### 東アジア言語のトークン化

「言語」オプションを「日本語」・「中国語（簡体字）」・「中国語（繁体字）」・「韓国語」のいずれかに設定した場合、その言語専用の辞書を使ってテキストの解析が行われます。上記の言語のいずれかを初めて指定したときに、パブリックソースから辞書がダウンロードされ、JMPデータテーブルに保存されます。このJMPデータテーブルは、TextExplorer ディレクトリにあるその言語専用のサブディレクトリに保存されます。TextExplorerディレクトリの場所は、次のとおりです。

- Windows の場合: C:\¥ユーザー ¥<ユーザ名>\AppData\Roaming\JMP\JMP¥TextExplorer¥
- macOS の場合: /Users/<ユーザ名>/Library/Application Support/JMP/TextExplorer/

言語専用の辞書に単語を追加・削除するには、TextExplorerディレクトリにあるその言語のサブディレクトリでdictionary-User.jmp データテーブルを開き、編集します。dictionary-User.jmp データテーブルには、「Data」と「action」の2つの列があります。辞書に単語を追加するには、dictionary-User.jmp データテーブルに行を追加し、その行の1番目の列に該当する単語、2番目の列に「add」という文字列を入力します。辞書から単語を削除するには、dictionary-User.jmp データテーブルに行を追加し、その行の1番目の列に該当する単語、2番目の列に「delete」という文字列を入力します。

### 句の抽出

この段階では、コーパス（文書の集合）に出現する句を抽出します。抽出された句を1つの単語として単語リストに追加することもできます。句の先頭または末尾にストップワードがくることはありませんが、句の途中に含めることはできません。

## 単語の抽出

この段階では、前の段階で抽出されたトークンと句の情報に基づいて、単語リストを作成します。

各トークンについて、次の処理を行います。

1. 起動ウィンドウで指定された、単語あたりの最小文字数および最大文字数の条件を満たしているかどうかをチェックします。数字のみのトークンは、この処理から除外されます。
2. トークンを単語として見なせるかどうかをチェックします。トークン化の方法として「基本ワード」を選択した場合は、少なくとも1文字のアルファベットまたはユニコード文字が含まれていることが条件となります。数字のみのトークンは、この処理から除外されます。「Regex」を選択した場合は、正規表現 (regular expression) を使用して、トークンを構成する文字を識別します。
3. トークンがストップワードでないかどうかをチェックします。
4. 語幹抽出およびその例外を適用します。

ある句を1つの単語として単語リストに追加すると、次の処理が行われます。

1. 句を単語リストに追加します。このとき、句に含まれる各単語に対して、単語に対して行われた語幹抽出と同じ語幹抽出を行います。その語幹抽出で同じになった句がある場合は、単語リストにて1つにまとめられます。
2. 追加された句に含まれる単語に関して、その単語単体での出現頻度を、該当の句に含まれている分だけ減らします。

---

## 「テキストエクスプローラ」プラットフォームの例

JMPで記述形式による回答を分析する方法を紹介します。この例では、ペットに関するアンケートへの回答を分析してみましょう。

1. 「ヘルプ」>「サンプルデータフォルダ」を選択し、「Pet Survey.jmp」を開きます。
2. 「分析」>「テキストエクスプローラ」を選択します。
3. 「質問への回答」を選択し、「テキスト列」ボタンをクリックします。
4. 「言語」から「英語」を選択します。
5. 「OK」をクリックします。



図 12.2 「テキストエクスプローラ」レポートの最初の画面の例



このデータには、194文書、異なる単語が372単語あることが先頭の情報からわかります。また、トークンの総数は2075個です。もっとも頻繁に使用されている単語は「cat」で、55回出現しています。

- 「「質問への回答」のテキストエクスプローラ」の赤い三角ボタンをクリックし、[単語のオプション] > [語幹抽出] > [すべての単語] を選択します。
- 「句」のリストから「cat food」と「dog food」を選択して右クリックし、[句の追加] を選択します。「cat food」と「dog food」が「単語」リストに追加されます。
- 「単語」リストをスクロールダウンして、「cat food」と「dog food」があることを確認します。それぞれ4回ずつ出現していることがわかります。

図 12.3 「単語」 リストを編集後にスクロール

単語	度数
barn.	4
cat. food.	4
couch.	4
dog. food.	4
door.	4
duck.	4
get.	4
great.	4
guard.	4
job.	4
know.	4
made.	4
make.	4
now.	4
watch.	4
anymor.	3
ate.	3
back.	3
bath.	3
best.	3
box.	3
eat.	3

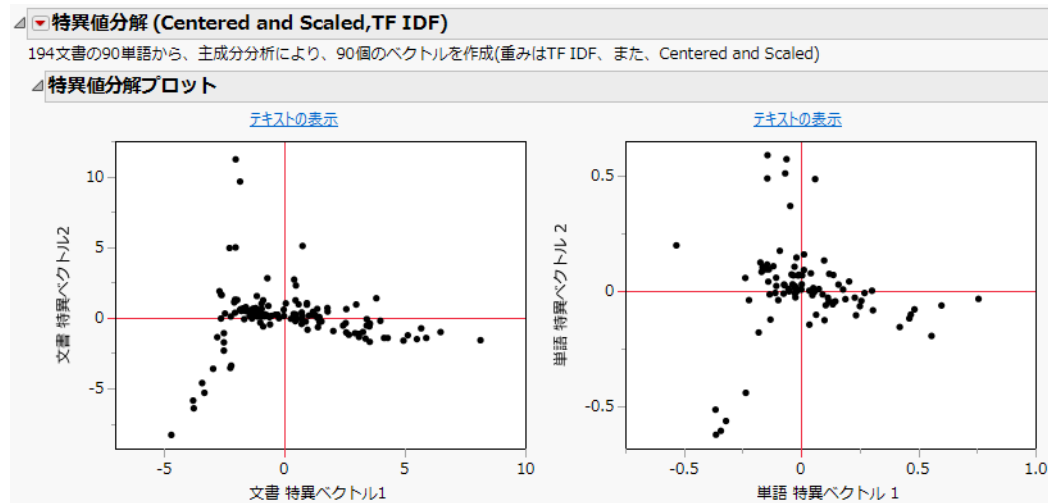
「句」リストの「cat food」と「dog food」は、このレポートでは単語として扱うことにしたため、グレーで表示されます。

**JMP PRO** この後の手順は、JMP Proでのみ実行できます。

9. **JMP PRO** 「「質問への回答」のテキストエクスプローラ」の赤い三角ボタンをクリックし、**[潜在意味分析, 特異値分解]** を選択します。
10. **JMP PRO** デフォルト値をそのままにして、**[OK]** をクリックします。

2つの特異値分解プロットが表示されます。左のプロットには、文書空間の最初の2つの特異ベクトルが表示されています。右のプロットには、単語空間の最初の2つの特異ベクトルが表示されています。

図12.4 特異値分解プロット



11. **JMP PRO** 左の特異値分解プロットの、一番右側にある7つの点を選択します。  
この7つの回答は、他の回答から離れてクラスターを形成しています。このクラスターを詳しく調べるために、質問への回答文を読んでみましょう。
12. **JMP PRO** 左の特異値分解プロットの上にある **「テキストの表示」** をクリックします。

図12.5 選択した文書のテキスト

```
There was this funny video of a cat trying to jump into someones lap, but fell into the pool instead. [56]
My kids are always trying to make videos of the cats doing funny things. [78]
The cat jumps out of my lap when he hears someone shake that box of cat food. [79]
The cats are always trying to sit in my lap while I am sitting down to dinner. [93]
The funny cat video where the cats jumped through the window right into the bathtub was hilarious. [142]
We made this funny video of the cat trying to climb the wall to chase a laser pointer. [153]
My cat always enjoys sitting in my lap and watching cat videos on my tablet. [185]
```

ウィンドウが開き、選択した7つの文書のテキストが表示されます。これらの回答は、「funny」、「cat」、「video」という単語が組み合わせて使われているという点で、共通しています。また、3つの文書は、第1次元の特異ベクトルが他の文書よりも大きい正の値を取っており、第1次元において他の文書と異なることを示唆しています。

特異ベクトルについてさらに詳しく分析すると、その次元が何を表しているのかがわかる可能性があります。たとえば、プロットの右端のほうにある回答の多くが、猫に関する内容であるのに対し、左端のほうにある回答の多くは、犬に関する内容です。したがって、第1次元の特異ベクトルは、回答における猫と犬の違いを抽出していると言えます。

## 「テキストエクスプローラ」プラットフォームの起動

「テキストエクスプローラ」プラットフォームを起動するには、[分析] > [テキストエクスプローラ] を選択します。

図 12.6 「テキストエクスプローラ」起動ウィンドウ

自由形式のテキストを分析する。

列の選択

▼ 1列

質問への回答

言語

表示言語 ▼

句あたりの最大単語数 4

句の最大数 5000

単語あたりの最小文字数 1

単語あたりの最大文字数 50

語幹抽出 しない ▼

トークン化 Regex ▼

☐ Regexのカスタマイズ

☐ 数字を単語として扱う

選択した列に役割を割り当てる

テキスト列	必須(文字)
検証	オプション(数値)
ID	オプション
By	オプション

アクション

OK

キャンセル

削除

前回の設定

ヘルプ

「列の選択」の赤い三角ボタンのメニューのオプションについては、『JMPの使用法』を参照してください。「テキストエクスプローラ」起動ウィンドウには、以下のオプションがあります。

**テキスト列** テキストデータを含む列を割り当てます。複数の列を指定した場合には、列ごとに別々の分析が実行されます。

**JMP PRO 検証** JMP Proでは、検証列を選択することができます。「列の選択」リストで列を選択せず、[検証] ボタンをクリックすると、データテーブルに検証列を新規作成することができます。検証列を作成する方法の詳細については、『予測モデルおよび発展的なモデル』を参照してください。

検証列を指定すると、文書単語行列の計算には影響しませんが、潜在クラス分析、潜在意味分析、トピック分析、判別分析で学習セットのみが使用されるようになります。検証列は、[単語選択分析] オプションで実行される一般化回帰における検証法として使われます。

**ID** [積み重ねた文書単語行列の保存] オプションによって保存されるデータテーブルにおいてIDとなる列を指定します。このデータテーブルは、アソシエーション分析を行うのに適しています。また、「潜在クラス分析」レポートでIDとなる列を指定するためにも使えます。

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

**メモ:** By変数を指定した場合、[Regexのカスタマイズ] オプションの設定はBy変数の全水準に適用されます。

**言語** テキスト処理に使用する言語を指定します。これは、語幹抽出、および、ストップワード・再コード化・句のビルトインリストに影響します。JMPを実行している言語と同じ言語でなくても構いません。特に設定を変更しない限り、JMPの「表示言語」と同じ言語に設定されます。

---

**メモ:** 「言語」オプションを「日本語」・「中国語（簡体字）」・「中国語（繁体字）」・「韓国語」のいずれかに設定した場合、その言語専用の辞書を使ってテキストの解析が行われます。上記の言語のいずれかを初めて指定したときに、パブリックソースから辞書がダウンロードされ、JMPデータテーブルに保存されます。この辞書では、単語の追加や削除もできます。「[東アジア言語のトークン化](#)」を参照してください。

---

**句あたりの最大単語数** 句リストの句を構成するときの、1句あたりに含める単語数の最大値を指定します。

**句の最大数** 句のリストにおける上限を指定します。ここに指定された上限以下の出現回数をもつ句のみが、句のリストに表示されます。

**単語あたりの最小文字数** 1単語あたりに含める文字数の最小値を指定します。

**単語あたりの最大文字数** 1単語あたりに含める文字数の最大値（最大2000まで）を指定します。

**語幹抽出** （英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）先頭が同じで末尾が異なる語を同じとみなすかどうかを指定します。次のオプションがあります。

**しない** 語幹抽出を行いません。

**同じ語幹の単語** 1つの語幹から2つ以上の単語に派生しているもののみ、同じ語幹にまとめます。

**すべての単語** 派生している単語がないものを含め、すべての単語を語幹にまとめます。

---

**メモ:** 語幹抽出のオプションは、単語リストに追加された句にも適用されます。この処理では、句に含まれる各単語に語幹抽出を行い、その語幹抽出によって同じになる句は1つにまとめられます。たとえば、「dogs bark」、「dog barks」という句は、両方とも「dog·bark」という1つの句にまとめられます。語幹抽出のオプションが選択されている場合、単語リストから句を削除することはできません。

---

**トークン化** （英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）句を単語（トークン）に分解する方法を指定します。トークン化には、次のオプションがあります。

**Regex** 予め用意されているデフォルトの正規表現（ビルトインの正規表現）によって、テキストを構文解析します。ビルトインの正規表現を追加、削除、または編集したい場合は、**[Regexのカスタマイズ]** オプションを選択してください。「[正規表現エディタでのRegexのカスタマイズ](#)」を参照してください。

**基本ワード** 一般的によく使用される区切り文字に基づいて、単語を抽出します。この区切り文字には、スペース・タブ・改行文字のほか、英語での句読点の多くが含まれています。単語として数字も含めたい場合は、**[数字を単語として扱う]** オプションを選択してください。このオプションを選択しなかった場合、区切り文字に基づいて抽出された単語のうち、数字のみのものは除外されます。

---

**ヒント:** デフォルトの区切り文字を表示するには、「トークン化」で「基本ワード」を指定してレポートを作成し、「テキストエクスプローラ」レポートの赤い三角ボタンから **[表示オプション] > [区切り文字の表示]** を選択します。

---

**Regex のカスタマイズ**（「トークン化」で「Regex」を指定したときのみ有効）「テキストエクスプローラ 正規表現エディタ」ウィンドウで、正規表現の設定を編集できます。特殊な単語に対応するには、このオプションを使用します。たとえば、電話番号や、文字と数字から構成される単語を抽出するように設定できます。「Regex のカスタマイズ」オプションを使用するのは、デフォルトの正規表現の結果が望ましくない場合のみにしてください。たとえば、デフォルトの正規表現で識別できない構造がテキストに含まれている場合などに使用してください。「[正規表現エディタでの Regex のカスタマイズ](#)」を参照してください。

**数字を単語として扱う**（「トークン化」で「基本ワード」を指定したときのみ有効）数字を単語として抽出します。このオプションを選択すると、数字の単語については、「単語あたりの最小文字数」の設定は無視されます。

起動ウィンドウで **[OK]** をクリックすると、「Regex のカスタマイズ」を選択している場合は、「テキストエクスプローラ 正規表現エディタ」ウィンドウが表示されます。そうでない場合は、「テキストエクスプローラ」レポートが表示されます。

---

**メモ:** テキスト処理では、アルファベットの大文字と小文字は区別されません。テキスト全体が内部で小文字に変換されてから、トークン化や分析が行われます。正規表現処理や、「テキストエクスプローラ」レポートに出力される単語の集計処理にも、小文字に変換されたテキストが使用されます。

---

## 正規表現エディタでの Regex のカスタマイズ

「Regex のカスタマイズ」オプションを選択すると、「テキストエクスプローラ 正規表現エディタ」ウィンドウが表示されます。このウィンドウから、電話番号・時間・金額など、予め提供されている、さまざまなビルトインの正規表現 (regular expression) を選択できます。また、独自の正規表現を定義することもできます。

---

**メモ:** 「Regex のカスタマイズ」オプションを使用するのは、デフォルトの正規表現の結果が望ましくない場合のみにしてください。たとえば、デフォルトの正規表現で識別できない構造がテキストに含まれている場合などに使用してください。

---

---

**ヒント:** 起動ウィンドウで「言語」オプションを「日本語」・「中国語（簡体字）」・「中国語（繁体字）」・「韓国語」のいずれかに設定した場合、正規表現のリストには、選ばれた言語に対応した正規表現が1つ以上含まれます。独自の正規表現を追加したい場合は、その言語に対応した正規表現の後に追加していくことを推奨します。東アジア言語は区切り文字がないので、英語などの言語での正規表現を適用すると長い文字列が1つの単語として認識されてしまいます。そのため、各言語に対応した正規表現の前に、独自の正規表現を使用するのは避けてください。

---

図 12.7 テキストエクスプローラ 正規表現エディタ



## スクリプトエディタボックスでの構文解析

ウィンドウの上部には、スクリプトエディタボックスが表示され、サンプルテキストの構文解析結果が表示されます。構文解析結果での色は、「正規表現エディタ」に示されている色に対応しています。

- 「行」の**最初・前・次・最後**に移動するボタンをクリックすると、分析対象のテキストがスクリプトエディタボックスに表示されます。ここで、各行のテキストがどのように構文解析されるのかを確認できます。スクリプトエディタボックスに行番号を直接入力することもできます。
- **[列に保存]** ボタンをクリックすると、この正規表現によるトークン化の結果が、データテーブルに新しい列として追加されます。正規表現を変更する方法については、「**正規表現の編集**」を参照してください。**[列] > [ユーティリティ] > [テキストマッチによる列の作成]** から正規表現エディタを表示したときは、**[列に保存]** ボタンは表示されません。

**メモ:** **[列に保存]** ボタンをクリックしたときの結果には、正規表現のみがテキストのマッチに使用され、ストップワード・再コード化・語幹抽出、および、句・単語あたりの最小／最大文字数の設定は使用されません。

## 正規表現の追加

トークン化に使用する正規表現を追加するには、正規表現のリストの下にある＋記号をクリックします。すると、「正規表現ライブラリからの選択」ウィンドウが表示されます。このウィンドウには、予め用意されている正規表現（ビルトインの正規表現）がすべて表示されます。また、直前に「正規表現エディタ」で編集した正規表現も表示されます。予め用意されている正規表現は「ビルトイン」と記されています。分析者が独自に定義した正規表現（カスタムの正規表現）は、保存した名前で表示されます。同じ名前のカスタムの正規表現は、最新のもののだけがライブラリに保存されます。

**【前回の設定】** ボタンをクリックすると、正規表現エディタを前回使用したときの正規表現のリスト（**【列に保存】** ボタンまたは **【OK】** ボタンをクリックしたときに表示されていたもの）が自動的に追加されます。

このリストから1つまたは複数の正規表現を選択して **【OK】** をクリックすると、その正規表現がトークン化に使用されるようになります。ライブラリからカスタムの正規表現を削除するには、**【選択アイテムを削除】** ボタンを使用します。正規表現のライブラリは、ユーザごとに別の JSL ファイルとして、「TextExplorer」というディレクトリに保存されます。このディレクトリの場所は、次のとおりです。

- Windows の場合: C:\ユーザー \<ユーザ名>\AppData\Roaming\JMP\JMP\TextExplorer\
- macOS の場合: /Users/<ユーザ名>/Library/Application Support/JMP/TextExplorer/

このファイルは、他のユーザと共有しても構いませんが、直接編集しないでください。編集する場合は、「正規表現エディタ」を使用してください。

## 正規表現の編集

トークン化では、「正規表現エディタ」パネルで指定された正規表現が順番に適用されます。適用される順序を変更するには、リストから正規表現を選択して、リストの下にある上矢印または下矢印をクリックしてください。または、リスト内で正規表現をドラッグ&ドロップしても、適用される順序を変更できます。現在選択している正規表現には、青い三角が付きます。正規表現をリストから削除するには、正規表現を選択して、リストの下にあるマイナス記号をクリックしてください。なお、「レフトオーバー」という正規表現は、リストから削除することはできず、常にこのリストの最後に表示されます。

「正規表現エディタ」パネルのリストから正規表現を選択すると、編集可能なフィールドにその正規表現の属性が表示されます。これらのフィールドは、すべてクリックして編集できます。

正規表現には次のような属性があります。

**タイトル** 現在のウィンドウで各正規表現を識別するための名前を指定します（正規表現ライブラリにも、この名前で表示されます）。

**Regex** 正規表現（regular expression）の定義を指定します。正規表現にマッチする文字列を取得するため、最低1組の括弧が必要です。

**結果** 正規表現にマッチしたテキストを、どのように置き換えるかを指定します。置き換えるものとして、静的なテキスト、空白、正規表現にマッチしたテキストなどを指定できます。以下のように定義してください。

- － 正規表現にマッチしたテキストを、特定の静的なテキストに置き換えるには、その静的なテキストをそのまま指定します。



- マッチした文字列を無視するには、このフィールドを空白にします。
- 正規表現にマッチした一番外側の括弧の結果を使用するには、このフィールドに「¥1」と指定します。
- 正規表現にマッチした結果全体を使用するには、このフィールドに「¥0」と指定します。

**例**（オプション）正規表現の処理の例を、テキストと色を使って記述します。

**コメント**（オプション）正規表現の説明や処理を記述します。

**色** 正規表現にマッチしたテキストに与える色を指定します。スクリプトエディタボックスおよび「例」フィールドにおいて、指定した色が使われます。色を変更するには、矢印ボタンを使用します。

---

**メモ:**「Regex」フィールドの定義が無効の場合は、正規表現リストのその正規表現名の横に赤い×印が表示されます。

---

## 独自の正規表現を作成する方法

自分自身で独自の正規表現を定義したい場合（カスタムの正規表現を定義したい場合）には、以下の手順に従います。

1. 正規表現のリストの下にある＋記号をクリックします。
2. 「正規表現ライブラリからの選択」ウィンドウで、[未定義] が選択されていることを確認します。
3. [OK] をクリックします。
4. 「正規表現エディタ」パネルで、正規表現の定義を編集します。
5. 「タイトル」フィールドに、定義した正規表現に対する固有の名前を指定します。

---

**ヒント:** 正規表現の定義フィールドを編集するときは、ログウィンドウを開いて見える状態にしておくと便利です。一部のエラーメッセージは、ログウィンドウにのみ表示されます。ログウィンドウを開くには、[表示] > [ログ] を選択します。正規表現のトラブルシューティングについては、<https://regexr.com/> をはじめ、インターネット上の多くの資料で解説されています。

---

## 単語の区切り文字のリスト

[単語の区切り文字のリスト] をクリックすると、トークン化処理において単語の区切りと見なす文字のリストを指定できます。**区切り文字**は、ある1つの単語の先頭文字には使うことができません。ただし、正規表現で許可されていれば、単語の途中にある文字としては含めることはできます。[単語の区切り文字のリスト] をクリックするとウィンドウが表示され、このウィンドウで区切り文字の追加や削除を行えます。デフォルトでは、空白（ホワイトスペース）のみが区切り文字として指定されています。「区切り文字」ウィンドウで、区切り文字のリストに加えた変更を取り消すには、[リセット] ボタンをクリックします。区切り文字のリストに加えた変更は、そのときのトークン化処理にのみ適用されます。

指定した正規表現と、必須の「レフトオーバー」正規表現によって、以下のような処理が行われます。

1. テキストストリームの現在の文字を、区切り文字のリストと比べます。
  - － 現在の文字が、区切り文字のリストのいずれかと一致したら、その文字を除外し、そこまでの時点で蓄積された一時的な「レフトオーバー」文字列を処理し、次の文字に移動して、[手順1](#)を繰り返します。
  - － 現在の文字が、区切り文字のリストと一致しない場合は、[手順2](#)に進みます。
2. 現在の文字から始まる文字列を、指定された正規表現とひとつずつ比較します（「レフトオーバー」の正規表現は除きます）。
  - － 現在の文字から始まる文字列が、いずれかの正規表現とマッチしたら、以下で述べる一連の処理を実行します。ここまでの一時的な「レフトオーバー」文字列に蓄積された文字を処理します。そして、「結果」フィールドに定義されている文字列を、単語として保存します。最後に、正規表現にマッチした文字列のすぐ後の1文字を、現在の文字に設定します。これらの処理が済んだら、[手順1](#)に戻ります。
  - － 現在の文字から始まる文字列が、「レフトオーバー」以外のどの正規表現にも一致しなかった場合は、[手順3](#)に進みます。
3. 現在の文字を一時的な「レフトオーバー」文字列に追加して、テキストストリーム内の次の文字を現在の文字に設定します。[手順1](#)に戻ります。
  - － 一時的な「レフトオーバー」文字列には、他の正規表現に一致するまで、1文字ずつ追加されます。
  - － 「レフトオーバー」正規表現の「結果」はデフォルトでは空白で、一時的に蓄積された「レフトオーバー」文字列を破棄する設定になっています。

#### ヒント：

- 「レフトオーバー」正規表現の「結果」を「¥1」に設定した場合は、句読点などを区切り文字に追加することを検討してください。指定した句読点は、結果から除かれます。
- 目的の単語を取得するために、「レフトオーバー」正規表現の結果を「¥1」に変更する代わりに、以下を行うことを検討してください。
  - － 正規表現ライブラリからより多くの正規表現を追加する。
  - － カスタムな正規表現（自分自身の独自の正規表現）を定義する。

上記のような処理は、データテーブルの各行において、テキストの最後まで実行されます。

### 結果をデータテーブルの列に保存する

**[列に保存]** ボタンをクリックすると、正規表現によるトークン化の結果が、データテーブルに新しい列として追加されます。新しい列は、データタイプが「文字」で、「テキストエクスプローラ」起動ウィンドウで「テキスト列」に指定した名前に、番号が追加されたものになります。[列] > [ユーティリティ] > [テキストマッチによる列の作成] を指定すると、正規表現ユーティリティを直接起動することができます。それにより、データテーブルの列に正規表現を適用することもできます。『JMPの使用法』を参照してください。

**メモ：** 正規表現を独自に定義した場合、適用されるトークン化は、元のそのままのテキストに対して、その正規表現を適用したものです。小文字に変換したテキストに対する結果ではない点に注意してください。

## 「テキストエクスプローラ 正規表現エディタ」を閉じる

「テキストエクスプローラ 正規表現エディタ」ウィンドウで **[OK]** をクリックすると、次の一連の処理が行われます。

1. 「テキストエクスプローラ 正規表現エディタ」ウィンドウで独自に定義した、カスタムの正規表現が、正規表現ライブラリに保存されます。

**注意：**ユーザーが独自に定義した、カスタムの正規表現がライブラリに保存されるのは、正規表現を定義した後に **[OK]** をクリックしたときのみです。次にライブラリを使用できるのは、最後に保存した正規表現だけです。正規表現ライブラリに定義した正規表現を保存したい場合は、固有の名前を付けてください。以前の正規表現を後で使用したい場合は、「テキストエクスプローラ」レポートウィンドウで、スク립トを保存することもできます。

2. 「テキストエクスプローラ」レポートが生成されます。このレポートに、指定した正規表現の設定を使ってテキストをトークン化した結果が表示されます。

## 「テキストエクスプローラ」レポート

「テキストエクスプローラ」レポートには、要約レポートと「単語と句のリスト」レポートが表示されます。

図12.8 「テキストエクスプローラ」レポートの例



要約レポート

「テキストエクスプローラ」レポートの上部には、以下の要約統計量が表示されます。

**単語の数** 単語リストに含まれる単語の数。

**ケース数** コーパスに含まれる文書の数。

**トークンの総数** コーパスに含まれる単語の総数。

**ケースあたりのトークン** トークンの総数を、ケースの数で割った値。

**無記入でないケースの数** 少なくとも1つの単語を含む文書の数。

**無記入でないケースの割合** コーパスの全文書数に対し、少なくとも1つの単語を含む文書数の割合。

単語と句のリスト

「テキストエクスプローラ」レポートの「単語と句のリスト」レポートには、トークン化が行われた後の単語と句の一覧が表示されます。このレポートの例については、[図 12.8](#)を参照してください。「単語」リストの「度数」は、コーパス中のその単語の出現頻度です。「句」リストの「度数」は、コーパス中のその句の出現頻度です。「N」は、句を構成する単語数です。

デフォルトでは、「単語」リストは出現頻度の降順に並べられます。出現頻度が同じ単語は、アルファベット順に並べられます。「句」リストも出現頻度の降順に並べられます。出現頻度が同じ句は、「N」の降順に並べられます。出現頻度とNの両方が同じ句は、アルファベット順に並べられます。両リストとも、アルファベット順に並べ替えるオプションが用意されています。

「句」リストには、起動ウィンドウで設定した【**句あたりの最大単語数**】と【**句の最大数**】が反映されています。なお、データテーブルに1度しか登場しない句は、このリストに表示されません。

指定した句を1つの単語として扱うこともできます。単語として扱われる句は、それが指定された保存場所（スコープ）ごとに、「句」リストで色分けされます（[表 12.1](#)）。句を単語として登録するときの保存場所については、[「\[単語のオプション\] から開く管理ウィンドウ」](#)を参照してください。

表 12.1 単語に指定された句の色

保存場所	色
ビルトイン	赤
ユーザライブラリ	緑
プロジェクト	青
列プロパティ	オレンジ
ローカル	グレー

## 単語と句のオプションメニュー

「単語と句のリスト」の各表で、一番左の列を右クリックすると、オプションが表示されます。各表をデータテーブルとして保存するには、「度数」列を右クリックして、「データテーブルに出力」を選択します。

### 「単語」リストのポップアップメニュー

「単語」リストを右クリックすると、以下のポップアップメニューが表示されます。

**行の選択** データテーブルで、選択された単語を含む行を選択します。

**テキストの表示** 選択された単語を含む文書を表示します。

---

**メモ:** デフォルトでは、最初の10,000個の文書が表示されます。文書数が10,000個を超える場合は、文書数を増やすためのウィンドウが表示されます。

---

**アルファベット順** 単語リストの並び順を指定します。このオプションが選択されている場合、単語はアルファベット順に並べられます。このオプションが選択されていない場合、単語は度数の降順で並べられます。

**数値の順序** ([アルファベット順]を選択した場合にのみ使用可能) 単語リストの並び順を指定します。このオプションが選択されている場合、並び替え対象項目がテキスト部分と数値部分に分割されてから、数値部分の数字の順番で並べられます。[数値の順序]オプションの並び替え規則については、『JMPの使用法』を参照してください。

**コピー** 選択された単語をクリップボードにコピーします。

**色** 選択された単語に色を付けます。

**ラベル** 単語の特異値分解プロットで、選択された単語に対応する点にラベルを付けます。

**この単語を含む句を選択** 選択された単語を含む句を、「句」リストから選択します。

**指示変数の保存** 選択された単語ごとに指示変数の列を作成し、データテーブルに保存します。指示変数の値は、その単語を含む文書は1、含まない文書は0になります。

**計算式の保存** 選択された単語ごとに計算式列を作成し、データテーブルに保存します。計算式の値は、その単語を含む文書は1、含まない文書は0になります。新しい文書に対してこの計算式を使うと便利です。

**再コード化...** 1つまたは複数の単語の値を変更します。単語を選択してから、このオプションを選択してください。このオプションを選択すると、「再コード化」ウィンドウが表示されます。『JMPの使用法』を参照してください。

**ストップワードの追加** 選択された単語をストップワードに追加し、単語リストから削除します。この操作を行うと、句のリストも更新されます。

---

**メモ:** 語幹をストップワードとして追加すると、その語幹に対応するトークンすべてがストップワードとして追加されます。

---

**語幹の例外の追加**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）選択された単語を、語幹抽出の例外リストに追加します。

**句の削除**（語幹抽出の方法を「しない」に設定し、「単語」リストから句を選択したときのみ使用できます。）選択した句を「単語」リストから削除します。また、単語の「度数」列も更新します。

---

**メモ:** 句が感情句として追加されている場合、現在の「感情分析」レポートの感情語のリストからもその句が削除されます。

---

**JMP PRO 感情の追加**（現在のレポートウィンドウで「感情分析」レポートが開かれている場合にのみ使用可能）選択した単語を、現在の「感情分析」レポートの感情語のリストに追加します。

---

**メモ:** 語幹を感情語として追加すると、その語幹に対応するトークンすべてが感情語として追加されます。

---

**フィルタの表示** 「単語」リストの上の検索フィルタの表示／非表示を切り替えます。「[検索フィルタのオプション](#)」を参照してください。

**データテーブルに出力** レポートの表から、JMP データテーブルを作成します。

**連結したデータテーブルの作成** 現在の表と同じ種類である複数の表を連結して、1つのJMP データテーブルを作成します。

## 「句」リストのポップアップメニュー

「句」リストを右クリックすると、以下のポップアップメニューが表示されます。

**行の選択** データテーブルで、選択された句を含む行を選択します。

**テキストの表示** 選択された句を含む文書を表示します。

**指示変数の保存** 選択された句ごとに指示変数の列を作成し、データテーブルに保存します。指示変数の値は、その句を含む文書は1、含まない文書は0になります。

**アルファベット順** 句リストの並び順を指定します。このオプションが選択されている場合、句はアルファベット順に並べられます。このオプションが選択されていない場合、句は度数の降順で並べられます。

**数値の順序**（「アルファベット順」を選択した場合にのみ使用可能）句リストの並び順を指定します。このオプションが選択されている場合、並び替え対象項目がテキスト部分と数値部分に分割されてから、数値部分の数字の順番で並べられます。「数値の順序」オプションの並び替え規則については、『JMPの使用法』を参照してください。

**コピー** 選択された句をクリップボードにコピーします。

**この句を含むものを選択** 選択された句を含む、それより長い句を、「句」リストから選択します。

**この句に含まれたものを選択** 選択された句に含まれる、それより短い句と単語を、「句」リストと「単語」リストから選択します。

**句の追加** 選択された句を「単語」リストに追加し、「度数」列とともに更新します。

**ストップワードの追加** 選択された句をストップワードに追加します。この操作を行うと、単語リストも更新されます。

**JMP PRO 感情句の追加** (現在のレポートウィンドウで「感情分析」レポートが開かれている場合にのみ使用可能) 選択された句を、「単語」リストと、現在の「感情分析」レポートの感情語リストに追加します。

**フィルタの表示** 「句」リストの上の検索フィルタの表示／非表示を切り替えます。「[検索フィルタのオプション](#)」を参照してください。

**データテーブルに出力** レポートの表から、JMP データテーブルを作成します。

**連結したデータテーブルの作成** 現在の表と同じ種類である複数の表を連結して、1つのJMP データテーブルを作成します。

### 検索フィルタのオプション

検索ボックス横の下向き矢印ボタンをクリックして、検索を絞り込みます。

**指定の単語を含む** 指定した単語を一部に含む項目を返します。「ease oom」を検索すると、「Release Zoom」のような文字列を返します。

**指定の句を含む** 指定した文字列を完全に含む項目を返します。「text box」を検索すると、「text」のすぐ後に「box」（「Context Box」と「Text Box」など）を含む文字列を返します。

**指定の句で始まる** 指定した文字列で始まる項目を返します。

**指定の句で終わる** 指定した文字列で終わる項目を返します。

**句全体が一致する** 指定した文字列全体で構成されている項目を返します。「text box」を検索すると、「text box」となっている文字列だけを返します。

**正規表現** 検索ボックスでは、ワイルドカード (\*) とピリオド (.) を使用できます。「get.\*name」を検索すると、「get」の後に1つ以上の単語を含む文字列を検索します。「Get Color Theme Names」、「Get Name Info」、「Get Effect Names」などが返されます。

**結果を逆にする** 指定した検索条件に一致しない項目を返します。

**すべての単語が一致する** 指定したすべての単語を含む項目を返します。「t test」を検索すると、検索文字列のどちらか一方、または両方を含む、「Pat Test」、「Shortest Edit Script」、「Paired t test」などの文字列を返します。

**大文字/小文字を無視** 検索基準で大文字と小文字を区別しません。

**語全体が一致する** 「すべての単語が一致する」の設定に基づいて、指定した単語全体を含む項目を返します。「data filter」を検索する場合、「すべての単語が一致する」も選択されていると、「data」と「filter」の両方を含む文字列が返されます。

## 「テキストエクスペローラ」プラットフォームのオプション

ここでは、「テキストエクスペローラ」プラットフォームに用意されているオプションについて説明します。

- 「[テキストを前処理するオプション](#)」
- 「[テキストの統計分析に関するオプション](#)」
- 「[保存に関するオプション](#)」
- 「[テキストエクスペローラのレポートオプション](#)」

### テキストを前処理するオプション

「テキストエクスペローラ」レポートの赤い三角ボタンで用意されているオプションのうち、テキストを前処理するものには以下のものがあります。

**表示オプション** レポートの表示内容を制御するサブメニューが表示されます。

**ワードクラウドの表示** 「ワードクラウド」レポートの表示／非表示を切り替えます。「ワードクラウド」レポートの赤い三角ボタンをクリックすると、ワードクラウドのレイアウトやフォントを変更するメニューが表示されます。「[ワードクラウド](#)」の[オプション](#)」を参照してください。

ワードクラウドの幅は、対話的に変更できます。幅を変更すると、自動的に高さが調整されます。「単語」リストの単語と、ワードクラウドの単語はリンクしています。

**単語リストの表示** 「単語と句のリスト」レポートにおいて、単語リストの表示／非表示を切り替えます。

**句リストの表示** 「単語と句のリスト」レポートにおいて、句リストの表示／非表示を切り替えます。

**単語と句のオプションの表示** 「単語と句のリスト」レポートのボタンの表示／非表示を切り替えます。これらのボタンは、各リストのポップアップメニューで利用可能なオプションをボタンで表示したものです。「[単語と句のリスト](#)」を参照してください。

**要約の表示** 要約表の表示／非表示を切り替えます。「[要約レポート](#)」を参照してください。

**ストップワードの表示** 分析に使用したストップワードリストの表示／非表示を切り替えます。プラットフォームを起動した最初には、予め用意されているストップワード（ビルトインのストップワード）が使用されます。ストップワードを追加するには、「単語」リストから単語を右クリックし、ポップアップメニューから「[ストップワードの追加](#)」を選択します。「[単語のオプション](#)」から開く[管理ウィンドウ](#)」を参照してください。

**再コード化の表示** 再コード化された単語リストの表示／非表示を切り替えます。「[単語のオプション](#)」から開く[管理ウィンドウ](#)」を参照してください。

**指定した句の表示** ユーザによって単語に指定された句のリストの表示／非表示を切り替えます。「[単語のオプション](#)」から開く[管理ウィンドウ](#)」を参照してください。



**語幹の例外の表示**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）語幹抽出の例外リストの表示／非表示を切り替えます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」を参照してください。

**区切り文字の表示**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定し、「トークン化」オプションで「基本ワード」を選択した場合にのみ使用可能）トークン化の方法として「基本ワード」を使用したときの区切り文字の表示／非表示を切り替えます。区切り文字のセットを変更するには、JSLでAdd Delimiters() または Set Delimiters() メッセージを使用してください。

**語幹レポートの表示**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定し、「語幹抽出」オプションを「しない」以外に設定した場合にのみ使用可能）「語幹抽出」レポートの表示／非表示を切り替えます。このレポートには、語幹抽出の結果が2つの形式で表示されます。左の表には、語幹に対して、その語幹に該当する単語が表示されます。右の表には、単語に対して、その単語の語幹が表示されます。

**選択されている行を表示** ウィンドウが開き、現在選択している行の文書のテキストが表示されます。

**表に対してフィルタを設定** 検索フィルタの表示／非表示をまとめて切り替えます。検索フィルタが表示されるのは、「ストップワード」表、「指定した句」表、「語幹の例外」表、「単語」リスト、「句」リスト、「語幹抽出」レポート内の表です。フィルタツールの詳細については、「[検索フィルタのオプション](#)」を参照してください。

**単語のオプション** 「単語」リストに関連するサブメニューが表示されます。

**語幹抽出**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）語幹抽出のオプションについては、「[「テキストエクスプローラ」プラットフォームの起動](#)」を参照してください。

**ビルトインのストップワードを含める** トークン化において、JMPで予め用意されているストップワードを用いるかどうかを指定します。

**ビルトインの句を含める** JMPで予め用意されている句を単語リストに含めるかどうかを指定します。

**ストップワードの管理** ストップワードの追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定されたストップワードをローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」を参照してください。

**再コード化の管理** 再コード化する単語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された再コード化をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」を参照してください。

**句の管理** 単語として扱う句の追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された句をローカルでは除外することができます。「[\[単語のオプション\]](#) から開く管理ウィンドウ」を参照してください。

**語幹の例外の管理**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）語幹抽出の例外とする単語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）として、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された語幹抽出の例外をローカルでは除外することができます。「[\[単語のオプション\]](#) から開く管理ウィンドウ」を参照してください。

**解析のオプション** 構文解析の方法やトークン化に関連するサブメニューが表示されます。

**トークン化**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定した場合にのみ使用可能）トークン化のオプションについては、「[「テキストエクスプローラ」プラットフォームの起動](#)」を参照してください。

**Regexのカスタマイズ**（「トークン化」オプションで「Regex」を選択した場合にのみ使用可能）正規表現（regular expression）をカスタマイズするウィンドウを表示します。現在の「テキストエクスプローラ」レポートに適用する正規表現の設定を編集できます。

---

**メモ:** プラットフォームの起動ウィンドウでBy変数を指定したときは、「Regexのカスタマイズ」オプションの設定は、By変数の全水準に自動的に適用されます。

---

**数字を単語として扱う**（英語・ドイツ語・スペイン語・フランス語・イタリア語のいずれかに「言語」オプションを設定し、「トークン化」オプションで「基本ワード」を選択した場合にのみ使用可能）数字を単語としてトークン化します。このオプションを指定すると、数字については「単語あたりの最小文字数」の設定は無視されます。

## 「ワードクラウド」のオプション

「ワードクラウド」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**レイアウト** ワードクラウドにおける単語の配置を指定します。デフォルトは「度数順」です。

**度数順** 単語の出現頻度順に、左から右へと表示します。

**アルファベット順** 単語のアルファベット順に、左から右へと表示します。

**中心から配置** 単語を雲状に表示します。表示されている単語の大きさは出現頻度を表しています。

**色分け** ワードクラウドにおける単語の色分けを指定します。デフォルトは「なし」です。

**なし** 「単語」リストと同じ色を付けます。

**単色** すべての単語に同じ色を付けます。この色は、凡例から変更できます。

**グレー濃淡** 単語にグレーの濃淡を付けます。

**任意の色** 単語にさまざまな色を付けます。この色は、凡例から変更できます。

**列の値による色分け...** 単語の色として、グラデーションがある色を用います。このグラデーションは、[列による単語のスコア] で計算されるスコアに基づいています。色と濃淡は、凡例から変更できます。

**フォント** ワードクラウドにおける単語のフォント、スタイル、サイズを指定します。

**凡例の表示** ワードクラウドにおいて、凡例の表示／非表示を切り替えます。

## [単語のオプション] から開く管理ウィンドウ

ストップワード・再コード化・句、および、語幹の例外に関する情報は、保存場所（スコープ）を指定することができます。これらの情報はいくつかの場所に保存することができ、「ユーザ」はテキストエクスプローラのユーザライブラリに、「プロジェクト」は現在のプロジェクトに、「列」は分析対象列の列プロパティに、「ローカル」はプラットフォームのスクリプトに、それぞれ保存されます。特定の「テキストエクスプローラ」レポートの定義や例外をローカルとして保存するには、そのレポートを作成するスクリプトにそれらの設定を保存してください。

[単語のオプション] から、ストップワード・再コード化・句、および、語幹の例外を管理するウィンドウを開くことができます。これら4つのウィンドウの構成はほぼ同じです。図12.9に、「ストップワードの管理」ウィンドウを示します。「句の管理」と「語幹の例外の管理」ウィンドウは、このウィンドウと同じです。「再コード化の管理」ウィンドウは、少し異なります。「再コード化の管理」を参照してください。

図12.9 「ストップワードの管理」ウィンドウ



## ストップワードの管理

「ストップワードの管理」ウィンドウには、ストップワードのリストが保存場所（スコープ）ごとに表示されます。各リストの下には、テキストボックスと追加ボタンがあり、これらによって独自のストップワードを追加できます。また、ストップワードを別の保存場所にドラッグして移動したり、コピー＆ペーストしたりできます。ウィンドウの最下部にある2つのボタンを使えば、選択した項目を左隣りまたは右隣りの保存場所に移すことができます。選択した項目を現在の保存場所から削除するには、[x] ボタンをクリックします。既存の項目を編集するには、その項目をダブルクリックして、テキストを変更します。

**言語** JMPで予め準備されているストップワード（ビルトインのストップワード）と、どのユーザライブラリに保存するかについて、言語を指定します。なお、[項目の適用]を選択した場合は、変更内容がメインのユーザライブラリに保存されます。言語の設定は、「ビルトイン」・「ユーザ」・「プロジェクト」のストップワードにのみ適用されます。

**ビルトイン(ロック)** 指定された言語について、予め準備されているストップワードが一覧表示されています。ここに表示されているストップワードをローカルでは除外したい場合は、「ローカルの例外」に移します。

**ユーザ** 指定された言語のユーザライブラリに含まれるストップワードが一覧表示されています。

**プロジェクト** （プロジェクトからテキストエクスプローラを起動した場合にのみ、表示されます）現在のプロジェクトに含まれる、指定された言語のストップワードが表示されます。

**列** テキスト列の列プロパティ「ストップワード」に含まれるストップワードが一覧表示されています。

**ローカル** ローカルのストップワードが一覧表示されています。テキストエクスプローラを起動するときに、JSLにおいてローカルのストップワードを指定することもできます。ローカルのストップワードは、現在の「テキストエクスプローラ」レポートでのみ適用されます。

**ローカルの例外** 現在の「テキストエクスプローラ」プラットフォームで、例外的に除外するストップワードが一覧表示されています。テキストエクスプローラを起動するときに、JSLにおいてローカルの例外を指定することもできます。「ローカルの例外」にリストされた単語は、他のいずれかの保存場所で指定されたストップワードを無効にします。

**読み込み** ストップワードをテキストファイルから読み込みます。読み込まれたストップワードは、クリップボードにコピーされます。クリップボードから、「ビルトイン」以外のいずれかのリストボックスに貼り付けてください。

**書き出し** ストップワードを、クリップボードやテキストファイルに書き出します。「書き出し」ウィンドウが表示されるので、書き出したいストップワードの保存場所を選択し、書き出し先を指定します。

ユーザライブラリファイルは、TextExplorerディレクトリに保存されます。このディレクトリの場所は、次のとおりです。

- Windowsの場合: C:\ユーザー \<ユーザ名>\AppData\Roaming\JMP\JMP\TextExplorer\<言語>
- macOSの場合: /Users/<ユーザ名>/Library/Application Support/JMP/TextExplorer/<言語>/

メインのユーザライブラリファイルは、**TextExplorer**ディレクトリのすぐ下に保存されます。このファイルは、言語に関わらず共通です。

[OK] をクリックすると、「ユーザ」に加えた変更はユーザライブラリに、「プロジェクト」に加えた変更はプロジェクトに、「列」に加えた変更は列プロパティに、それぞれ保存されます。「ローカル」と「ローカルの例外」に指定した項目は、「テキストエクスプローラ」レポートにおいて保存されたスクリプトにだけ保存されます。

ストップワードをユーザライブラリに保存すると、「**stopwords.txt**」というファイルが作成されます。また、列プロパティに保存すると、「ストップワード」という列プロパティが作成されます。

## 再コード化の管理

「再コード化の管理」ウィンドウの構成は、「ストップワードの管理」ウィンドウと少し異なり、各リストの下に、2つのテキストボックスが表示されます。上のボックスに指定した古い値が、下のボックスに指定した新しい値に再コード化されます。

再コード化情報をユーザライブラリに保存すると、「**recodes.txt**」というファイルが作成されます。また、列プロパティに保存すると、「再コード化」という列プロパティが作成されます。

## 句の管理

句をユーザライブラリに保存すると、「**phrases.txt**」というファイルが作成されます。また、列プロパティに保存すると、「句」という列プロパティが作成されます。

## 語幹の例外の管理

語幹の例外をユーザライブラリに保存すると、「**stemExceptions.txt**」というファイルが作成されます。また、列プロパティに保存すると、「語幹の例外」という列プロパティが作成されます。

---

**メモ:** 「語幹の例外の管理」ウィンドウで「ローカルの例外」に指定された単語は、語幹の例外から除外され語幹抽出処理の対象となります。

---

## 否定語の管理

「否定語の管理」ウィンドウは、「感情分析」レポートから起動できます。[「感情分析」](#)を参照してください。

否定語をユーザライブラリに保存すると、「**negations.txt**」というファイルが作成されます。また、列プロパティに保存すると、「否定語」という列プロパティが作成されます。

---

**メモ:** 「ローカル」または「ローカルの例外」に表示されている単語は、現在開いている「感情分析」レポートにのみ適用されます。

---

### JMP PRO 強意語の管理

「強意語の管理」ウィンドウは、「感情分析」レポートから起動できます。このウィンドウの構成は、「ストップワードの管理」ウィンドウと似ていますが、単語リストの下に、テキストボックスに加えて乗数の設定パネルがある点が異なります。この設定パネルでは、強意語を新しく追加するときに、強意の度合いを乗数で指定できます。「感情分析」を参照してください。

強意語をユーザライブラリに保存すると、「intensifiers.txt」というファイルが作成されます。また、列プロパティに保存すると、「強意語」という列プロパティが作成されます。

---

**メモ:**「ローカル」または「ローカルの例外」に表示されている単語は、現在開いている「感情分析」レポートにのみ適用されます。

---

### JMP PRO 感情語の管理

「感情語の管理」ウィンドウは、「感情分析」レポートから起動できます。このウィンドウの構成は、「ストップワードの管理」ウィンドウと似ていますが、単語リストの下に、テキストボックスに加えてスコアの設定パネルがある点が異なります。この設定パネルでは、感情語を新しく追加するときに、感情スコアを指定できます。「感情分析」を参照してください。

感情語をユーザライブラリに保存すると、「sentiments.txt」というファイルが作成されます。また、列プロパティに保存すると、「感情語」という列プロパティが作成されます。

---

**メモ:**「ローカル」または「ローカルの例外」に表示されている単語は、現在開いている「感情分析」レポートにのみ適用されます。

---

## JMP PRO テキストの統計分析に関するオプション

「テキストエクスペローラ」レポートの赤い三角ボタンで用意されているオプションのうち、統計分析に関するものには以下のものがあります。

**潜在クラス分析** 2値の文書単語行列に対して潜在クラス分析を行います。なお、分析の計算には、疎な行列に対する数値計算ルーチンが使用されています。「潜在クラス分析」を参照してください。

このオプションを選択すると、「設定」ウィンドウが開き、以下のオプションが表示されます。

**単語の最大数** 潜在クラス分析に含める単語の最大数。

**単語の最小度数** 潜在クラス分析に含める単語の最小出現頻度。

**クラスターの数** 潜在クラス分析におけるクラスターの個数。

**潜在意味分析, 特異値分解** 文書単語行列に対して、決められた次元までの特異値分解を実行します。「潜在意味分析 (特異値分解)」を参照してください。

**判別分析** 文書単語行列に基づき、各文書がどのグループまたはカテゴリに属するのかを予測します。「[線形判別分析](#)」を参照してください。

**単語選択分析** 特定の応答変数をどの単語が最もよく説明するかを分析します。単語選択分析は、応答変数が何かしらの評価である場合の感情分析にも役立ちます。「[単語選択分析](#)」を参照してください。

**感情分析** （「言語」オプションを英語に設定した場合にのみ使用可能）辞書に登録されている感情語を文書から抜き出し、各文書における肯定的感情、否定的感情、全体的感情をスコアリングします。「[感情分析](#)」を参照してください。

## JMP PRO 特異値分解の設定ウィンドウ

「テキストエクスプローラ」プラットフォームの統計分析は、文書単語行列 (DTM; Document-Term Matrix) に対して行われます。文書単語行列の各列は、単語リストに含まれている各単語に対応しています（ただし、列数の最大値は、「単語の最大数」で指定したものです）。文書単語行列の各行は、1つの文書に対応しています（JMP データテーブルにおけるテキスト列の1セルが、1文書に対応しています）。文書単語行列の各要素をどのように算出するかは、「設定」ウィンドウにて指定します。

[図 12.10](#) に、特異値分解の「設定」ウィンドウを示します。「テキストエクスプローラ」の赤い三角ボタンから、文書単語行列の特異値分解を実行するオプションを選択すると、「設定」ウィンドウが開き、以下のオプションが表示されます。

**単語の最大数** 特異値分解に含める単語の最大数。

**単語の最小度数** 特異値分解に含める単語の最小出現頻度。

**重み** 文書単語行列の要素を算出する方法。この方法については、「[文書単語行列の設定ウィンドウ](#)」を参照してください。

**特異ベクトルの数** 特異値分解で求める特異ベクトルの個数。デフォルト値は100です。ただし、文書数または単語数が100を下回るときは、その値がデフォルト値になります。

**中心化と尺度化** 文書単語行列を中心化や尺度化するか否かを指定します。[中心化と尺度化]、[中心化]、または[中心化しない]のいずれかを選択します。デフォルトでは、中心化と尺度化の両方が適用されます。

図 12.10 特異値分解の「設定」ウィンドウ



## 保存に関するオプション

「テキストエクスプローラ」レポートの赤い三角ボタンには、新たなデータテーブルを保存したり、既存のデータテーブルに新たな列を保存したり、または、列プロパティに情報を保存したりするための以下のオプションが含まれています。

**文書単語行列の保存** 文書単語行列の各列を、データテーブルの列に保存します（列数の最大値は、「単語の最大数」に指定した数によって決まります）。



**積み重ねた文書単語行列の保存** 文書単語行列を、積み重ねた形式で JMP データテーブルに保存します。この積み重ねた形式のデータテーブルは、「アソシエーション分析」プラットフォームで分析を行いたい場合に適しています。『予測モデルおよび発展的なモデル』を参照してください。「テキストエクスプローラ」の起動ウィンドウで ID 変数を指定した場合は、この ID によって、単語が元のテキストデータテーブルのどの行からきたのかを識別できます。積み重ねたテーブルには、アソシエーション分析を起動するスクリプトも含まれています。

**文書単語行列の計算式の保存** 尺度の形式がベクトルである計算式を、データテーブルの列に保存します。ベクトルのサイズは、[単語の最大数]、[単語の最小度数]、および [重み] にユーザが指定した値によって決まります。列の計算式には、JSL 関数の `Text Score()` が使用されます。この関数の詳細については、[ヘルプ] > [スクリプトの索引] を参照してください。

**単語テーブルの保存** 「単語と句のリスト」における単語リストのデータテーブルを作成します。このデータテーブルには、各単語、各単語の出現回数、そして、各単語を含む文書数が含まれます。このオプションを選択してデータテーブルを作成した後に、[列による単語のスコア] を選択すると、そのテーブルにスコアの列が追加されます。

**列による単語のスコア** 指定した列の値に基づいて単語のスコアを計算し、データテーブルに保存します。各単語のスコアは、指定した列の各値に、単語の出現頻度に基づく重み付けをしたうえで、平均値をとったものです。このオプションを選択する前に、[単語テーブルの保存] を選択してデータテーブルがすでに作成されている場合は、そのデータテーブルにスコアの列が追加されます。それ以外の場合は、単語テーブルが自動的に新規作成されます。指定した列が連続変数でない場合は、その列の水準ごとに、スコアの列が作成されます。

## 文書単語行列の設定ウィンドウ

「テキストエクスプローラ」レポートの赤い三角ボタンから [文書単語行列の保存] または [文書単語行列の計算式の保存] を選択すると、設定ウィンドウが開き、以下のオプションが表示されます。

**単語の最大数** 文書単語行列に含める単語の最大数。

**単語の最小度数** 文書単語行列に含める単語の最小出現頻度。

**重み** 文書単語行列の要素を算出する方法。



以下のオプションがあります。

**2値** 各文書において、該当の単語が1つでもある場合は1を、1つもない場合は0を割り当てます。これがデフォルトの重みとなります（特異値分解をそれ以前に実行した場合は、そのときに指定した値が表示されます）。

**3値** 各文書において、該当の単語が2つ以上ある場合には2、1つしかない場合は1、1つもない場合は0を割り当てます。

**度数** 各文書内の単語の出現頻度を割り当てます。

**度数の対数**  $\log_{10}(1+x)$  を割り当てます（ここで、 $x$  は各文書内の単語の出現頻度です）。

**TF IDF**  $TF \times \log_{10}(nDoc / nDocTerm)$  を割り当てます。“TF IDF” は、**単語度数 (term frequency)** と **文書度数の逆数 (inverse document frequency)** の略です。これは、特異値分解におけるデフォルトの重みです。計算式の各項の定義は以下のとおりです。

$TF$  = 文書内における単語の出現頻度

$nDoc$  = コーパスに含まれる文書の数

$nDocTerm$  = その単語を含む文書の数

---

**メモ:** 特異値分解を実行した後に「文書単語行列の保存」または「文書単語行列の計算式の保存」を選択した場合、これらの「設定」ウィンドウには、直前に行った特異値分解で指定した値が表示されます。

---

## テキストエクスプローラのレポートオプション

以下のオプションの詳細については、『JMPの使用法』を参照してください。

**ローカルデータフィルタ** データをフィルタリングするためのローカルデータフィルタの表示／非表示を切り替えます。

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

**プラットフォーム環境設定** 現在のプラットフォームの環境設定を表示したり、現在のJMPレポートの設定に合わせて環境設定を変更したりできます。

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

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

**メモ:** このプラットフォームには、他にもスクリプトだけで使用できるオプションがあります。[ヘルプ] メニューの [スクリプトの索引] を開いてください。また、[スクリプトの索引] には、この節で紹介されているオプションのスクリプト例もあります。

## JMP PRO 潜在クラス分析

「テキストエクスプローラ」プラットフォームで潜在クラス分析を実行すると、内容の似たような文書が同じクラスターに属するように分類されます。「潜在クラス分析」レポートには、モデルの指定内容、バイズ情報量規準値 (BIC)、[テキストの表示] ボタンが表示されます。「クラスターの混合確率」表から1つまたは複数のクラスターを選択し、[テキストの表示] ボタンをクリックすると、そのクラスターに分類される確率の高い文書のテキストが表示されます。

「潜在クラス分析」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**表示オプション** 「潜在クラス分析」レポートに表示する内容を指定します。デフォルトでは、「クラスターごとのワードクラウド」以外のレポートがすべて表示されます。

**クラスター混合確率** 該当のクラスターに文書が含まれる事前確率を示した表の表示／非表示を切り替えます。

**ヒント:** 「クラスターの混合確率」表で1つまたは複数の行を選択すると、そのクラスターに事後確率により割り当てられるデータ行も選択されます。

**クラスターごとの単語確率** ある文書が特定のクラスターに属しているという条件のもとで、その文書に各単語が登場する条件付き確率の推定値を示します。単語は、デフォルトで、コーパス全体での出現頻度の降順で並べられます。

「最も特徴的なクラスター」には、その単語が最も高い割合で出現するクラスターが示されます。

「最も確率の高いクラスター」には、その単語を含む文書をランダムに選んだときに、その文書が入る確率が最も高いクラスターが示されます。

**クラスター別上位単語** 各クラスターで最もスコアの高い上位10個の単語の表示／非表示を切り替えます。クラスター  $c$  の単語  $t$  のスコア  $S_{t,c}$  は、次のように計算されます。

$$S_{t,c} = 100 \cdot \text{mean}(p_t) \cdot \log_{10} \left( \frac{p_{t,c}}{\text{mean}(p_t)} \right)$$

ここで、 $\text{mean}(p_t)$  は、単語  $t$  のクラスターごとの単語確率の平均値、 $p_{t,c}$  は単語  $t$  のクラスター  $c$  における単語確率です。

**多次元尺度構成プロット** 多次元尺度構成プロット（MDSプロット）の表示／非表示を切り替えます。このプロットは、クラスターの類似度を2次元で表します。多次元尺度構成プロットの詳細については、『多変量分析』を参照してください。文書を選択して「テキストの表示」ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

**行ごとのクラスター確率** 「混合確率」表の表示／非表示を切り替えます。この表には、各クラスターに属する事後確率が行ごとに示されます。「最尤クラスター」には、この事後確率が最も高いクラスターが示されます。

**クラスター別ワードクラウド** クラスターごとのワードクラウドの表示／非表示を切り替えます。

**クラスター名の変更** クラスターにわかりやすい名前を付けることができます。

**確率の保存** 「混合確率」表の値を、データテーブルの対応する行に保存します。

**確率の計算式を保存** 各クラスターに属する事後確率と、その事後確率が最大となるクラスターに分類する計算式の列を、データテーブルに保存します。

スコアを求める計算式は、JSL関数のText Score()が使用されます。なお、この際、計算方法を示す引数として、“LCA”が指定されます。

**クラスターによる色分け** データテーブルの行に、最尤クラスターごとに色を付けます。

**削除** 「テキストエクスプローラ」レポートから、「潜在クラス分析」レポートを削除します。

潜在クラス分析の詳細については、『多変量分析』を参照してください。

---

**メモ:** 「テキストエクスプローラ」プラットフォームでも潜在クラス分析が実行できますが、そこで使われているアルゴリズムは文書単語行列の希薄性を利用しています。そのため、「テキストエクスプローラ」プラットフォームでの潜在クラス分析の結果は、「潜在クラス分析」プラットフォームでの結果と正確には一致しません。

---

---

## 潜在意味分析（特異値分解）

「テキストエクスプローラ」プラットフォームの潜在意味分析では、文書単語行列（DTM; Document-Term Matrix）を、指定された次元まで特異値分解（SVD; Singular Value Decomposition）します。この手法では、文書単語行列を特異値分解して、少数の次元まで縮小します。潜在意味分析は、文書単語行列に対する主成分分析（PCA; principal components analysis）です。

文書単語行列に対する特異値分解では、 $\mathbf{U}$ ,  $\mathbf{S}$ ,  $\mathbf{V}'$  の3つの行列を使って文書単語行列を近似します。これらの行列の関係は、次のとおりです。

$$DTM \approx \mathbf{U}\mathbf{S}\mathbf{V}'$$

$nDoc$ をDTMの文書数（行数）、 $nTerm$ をDTMの単語数（列数）、 $nVec$ を指定された特異値ベクトルの数とします。なお、 $nVec$ は、 $nDoc$ と $nTerm$ のうちの小さい方の値以下になります。 $\mathbf{U}$ は $nDoc \times nVec$ の行列で、各列が文書単語行列の左特異ベクトルになっています。 $\mathbf{S}$ は、次元数が $nVec$ の対角行列で、 $\mathbf{S}$ の対角成分が、文書単語行列の特異値となります。 $\mathbf{V}'$ は、 $nVec \times nTerm$ の行列です。 $\mathbf{V}'$ の行（または $\mathbf{V}$ の列）は右特異ベクトルです。

右特異ベクトル（に特異値を掛けたもの）は、似たような意味やトピックをもつ単語間の関連を捉えます。たとえば、3つの単語が同じ文書でよく一緒に使用される場合は、 $\mathbf{SV}'$ において、その3つの単語に対応する数値がどれも大きくなります。一方、左特異ベクトル $\mathbf{US}$ は、特異値分解で得られた右特異ベクトルに文書を射影したものです。

潜在意味分析では、間接的な関連も捉えることができます。たとえば、2つの単語が同じ文書で使用されていなくても、それぞれがまた別の同じ単語とよく一緒に使用されている場合、そのような間接的な関連を捉えることができます。なお、2つの文書に共通する単語がまったくなくても、それぞれの文書に含まれる単語が縮減された次元では近くに位置すれば、それらの文書も似たような座標に射影されます。

テキストデータを特異値分解すると、数値のベクトルに変換されますので、その変換された数値データに基づいて、クラスタリング・分類・回帰分析といった統計分析を行えるようになります。他のJMPプラットフォームの統計分析を行いたい場合には、保存に関するオプションを使って、特異値分解の結果を保存してください。

デフォルトでは、文書単語行列を中心化、尺度化して、 $nDoc-1$ で割ってから、特異値分解が実行されます。この分析は、文書単語行列の相関行列に対して主成分分析を実行することと同じです。

特異値分解の「設定」ウィンドウで、[中心化] または [中心化しない] を選択することもできます。

- [中心化] を選択すると、文書単語行列を中心化して、 $nDoc-1$  で割ってから、特異値分解が実行されます。この分析は、文書単語行列の共分散行列に対して主成分分析を実行することと同じです。
- [中心化しない] を選択すると、文書単語行列を  $nDoc$  で割ってから、特異値分解が実行されます。この分析は、尺度化しない文書単語行列に対して主成分分析を実行することと同じです。

なお、文書単語行列の特異値分解で使われるアルゴリズムでは、中心化されている場合を含め、この行列が疎であることを活用しています。

## JMP PRO 特異値分解レポート

「テキストエクスプローラ」プラットフォームで [潜在意味分析, 特異値分解] オプションを選択すると、2つの特異値分解プロットと特異値の表が生成されます。

## JMP PRO 特異値分解プロット

1つ目のプロットは、各文書を点で示します。各文書の点の座標は、最初の2つの左特異ベクトル（ $\mathbf{U}$ 行列の最初の2列）に、特異値の対角行列（ $\mathbf{S}$ ）を掛け合わせたものです。このプロットは、「主成分分析」プラットフォームのスコアプロットと同じです。プロットの各点は文書（データテーブルの行）を表し、点を選択すると、対応する行がデータテーブルで選択されます。

2つ目のプロットは、各単語を点で示します。各単語の点の座標は、最初の2つの右特異ベクトル（ $\mathbf{V}'$  行列の最初の2行）に、特異値の対角行列（ $\mathbf{S}$ ）を掛け合わせたものです。このプロットは、「主成分分析」プラットフォームの負荷量プロットと同じです。プロットの各点は、「単語」リストの各行に対応しています。

特異値分解プロットの上には、[テキストの表示] ボタンがあります。プロット上の点を選択してこのボタンをクリックすると、ウィンドウが開き、該当するテキストが表示されます。

## **JMP PRO** 特異値

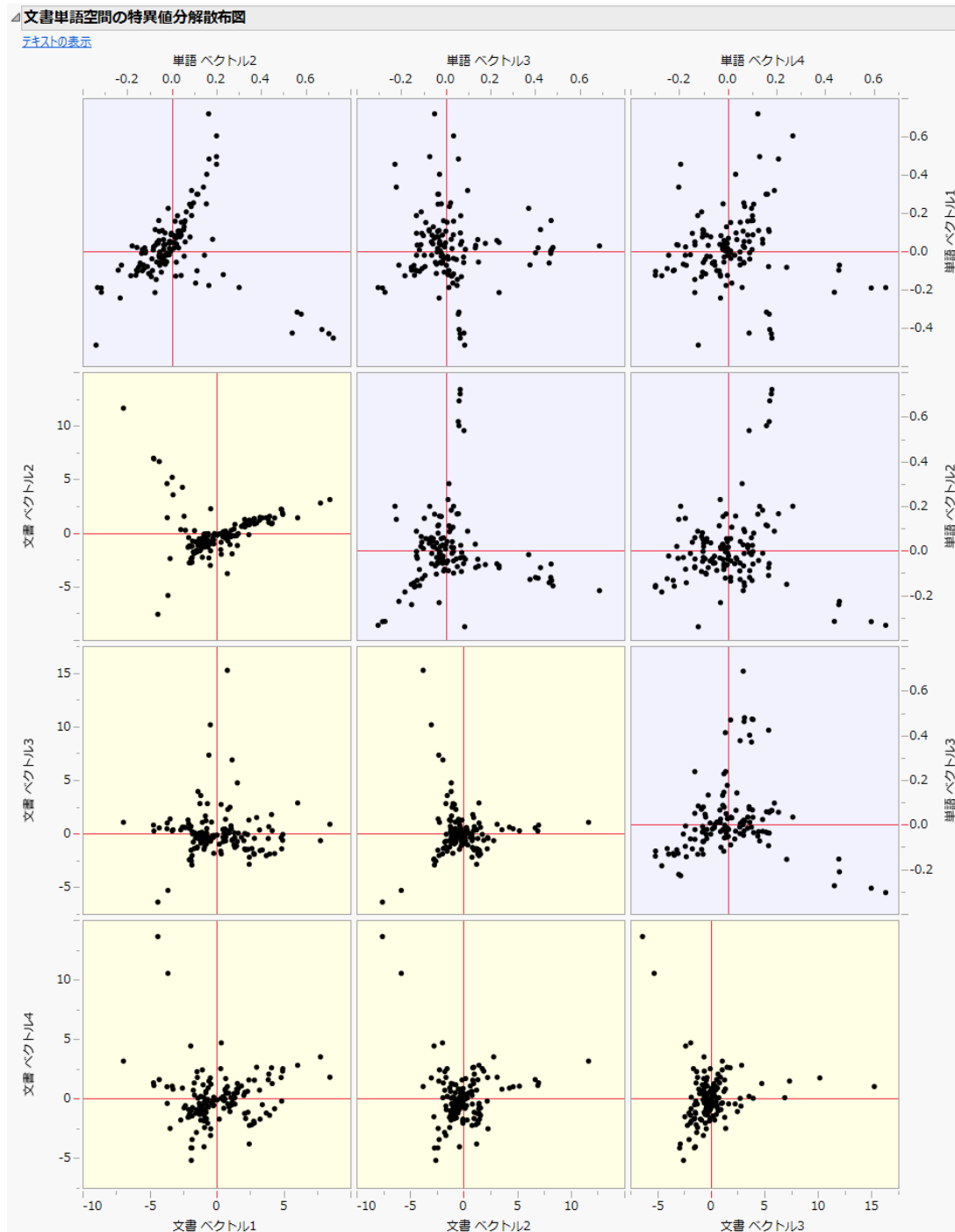
2つの特異値分解プロットの下には、特異値の表が表示されます。この特異値は、文書単語行列の特異値分解における  $\mathbf{S}$  行列の対角要素です。この表には「固有値」の列もあり、対応する主成分分析の固有値が示されます。「主成分分析」プラットフォームと同様に、変動の割合（寄与率）と、その累積（累積寄与率）の列もあります。これらの値は、文書単語行列におけるばらつきが各次元でどれくらい説明されているかを示しています。「累積寄与率」の値に基づいて、文書単語行列のばらつきのどの程度が説明されているかを見て、使用する特異ベクトルの数を決めることができます。

## **JMP PRO** 特異値分解レポートのオプション

「テキストエクスプローラ」プラットフォームの「特異値分解」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**特異値分解 散布図行列** 単語と文書の特異ベクトルを描いた散布図行列の表示／非表示を切り替えます。このオプションを選択すると、次元を指定するウィンドウが呼び出されます。最初の2次元だけでなく、指定した次元までの特異値分解の散布図が描かれます。文書を選択して [テキストの表示] ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

図 12.11 文書単語空間の特異値分散布図



**トピック分析, 特異値分解を回転** 文書単語行列の特異値分解した結果を Varimax 回転します。そして、単語を「トピック」と呼ばれるグループにまとめます。このオプションを複数回使用して、異なる次元でのトピックを見ることができます。「[トピックの分析](#)」を参照してください。

**クラスター 単語** 単語に対する階層型クラスター分析の表示／非表示を切り替えます。樹形図の右側に、クラスター数を設定するオプションと、クラスターをデータテーブルに保存するオプションが表示されます。このデータテーブルには、単語ごとに、度数、単語を含む文書数、割り当てられたクラスターが保存されます。階層型クラスター分析と樹形図の詳細については、『多変量分析』を参照してください。

**クラスター 文書** 文書に対する階層型クラスター分析の表示／非表示を切り替えます。樹形図の右側に、クラスター数を設定するオプション、クラスターをデータテーブルの列に保存するオプション、樹形図で選択した枝の文書を表示するオプションが表示されます。

**近傍点の選択** 文書の特異値分解プロットで選択されている点の中から、近傍点を見つけて選択します。近傍点を見つけるアルゴリズムでは、特異値分解の **U** 行列を使用します。このオプションを選択した場合は、近傍点の個数を指定する必要があります。デフォルトでは、10 個の近傍点を選択されます。

**文書 特異ベクトルの保存** ユーザが指定した数の主成分スコア（左特異ベクトルに特異値をかけたもの）を、データテーブルの列に保存します。最初の2列が、文書の特異値分解プロットで使用されているものです。「[潜在意味分析（特異値分解）](#)」を参照してください。

**特異ベクトルの計算式の保存** 文書の主成分スコアを求める計算式を、データテーブルの「ベクトル」尺度の列に保存します。列の計算式には、JSL 関数の **Text Score()** が使用されます。この関数の詳細については、[ヘルプ] > [スクリプトの索引] を参照してください。

**単語 特異ベクトルの保存** 新しいデータテーブルを作成し、単語が行に、ユーザが指定したまでの次元が列になっている主成分負荷量行列を保存します。単語テーブルがすでに開かれている場合は、そのテーブルに列を追加します。最初の2列が、単語の特異値分解プロットで使用されているものです。「[潜在意味分析（特異値分解）](#)」を参照してください。

**削除** 「テキストエクスプローラ」レポートウィンドウから「特異値分解」レポートを削除します。

---

## **JMP PRO** トピックの分析

「テキストエクスプローラ」プラットフォームで「トピック分析, 特異値分解を回転」オプションを選択すると、文書単語行列の特異値分解して得られた主成分負荷量が Varimax 回転されます。この際、回転する主成分負荷量の次元として、文書単語行列から維持したいトピック数を指定します。トピック数を指定すると、「トピック分析」レポートが表示されます。

トピックの分析は、主成分分析の結果を回転させることと同じです。Varimax 回転は、主成分負荷量（特異ベクトルに特異値をかけたもの）のいくつかの座標がぐっきりと大きくなるように（少数の単語における数値が大きくなるように）、回転します。回転後のベクトルは、少数の単語だけが大きな値となるため、テキストの内容を解釈しやすくなります。負の値は、該当のトピックだけにはその単語が含まれていない傾向を表します。各トピックにおいて、負の値となっている単語は、正の値の単語よりも少なくなるようにしています。

## JMP PRO 「トピック分析」レポート

「テキストエクスプローラ」プラットフォームの「トピック分析」レポートには、回転後の負荷量が多い単語がトピックごとに表示されます。「トピック分析」レポートには、特異値分解を回転した結果のレポートがあります。

「トピック別上位負荷量」レポートには、単語の表がトピックごとに表示されます。この表には、回転した後の負荷量の絶対値が大きい単語が表示され、負荷量の絶対値の降順で並べられます。これらの表は、各トピックの概念を判断するために役立ちます。

「トピック分析」レポートには、以下のレポートも表示されます。

**トピック負荷量** 単語の負荷量をトピックごとに示す行列が表示されます。この行列は、主成分負荷量を回転させた後の負荷量行列です。

**トピック別ワードクラウド** ワードクラウドがトピックごとに表示されます。

**トピックスコア** トピックと文書スコアの行列が表示されます。トピックに対する文書のスコアが高い場合は、そのトピックとの関連が強いことを示します。

**トピックスコアのプロット** トピックのスコアのプロットが文書ごとに表示されます。また、「テキストの表示」ボタンもあります。文書を選択して[テキストの表示] ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

「トピックスコアのプロット」は、「トピックスコア」レポートの数値をグラフに描いたものです。各パネルは、トピック（「トピックスコア」レポートの行列の1列）に対応しています。パネル中の各点は、コーパス中の文書（「トピックスコア」レポートの行列の1行）に対応しています。

**各トピックによって説明される分散** 各トピックによって説明される分散の表が表示されます。この表には、各トピックによって説明される変動の割合（寄与率）とその累積（累積寄与率）も表示されます。

**回転行列** Varimax 回転の回転行列が表示されます。

## JMP PRO 「トピック分析」レポートのオプション

「テキストエクスプローラ」プラットフォームの「トピック分析」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**トピックの散布図行列** 回転した後の負荷量ベクトルの散布図行列の表示／非表示を切り替えます。文書を選択して[テキストの表示] ボタンをクリックすると、ウィンドウが開き、その文書のテキストが表示されます。

**表示オプション** 「トピック分析」レポートの各要素の表示／非表示を切り替えるオプション。「[「トピック分析」レポート](#)」を参照してください。

**トピック名の変更** トピックに内容のわかりやすい名前を付けることができます。



**文書 トピックベクトルの保存** ユーザが指定した数の、主成分スコアを回転したものを、データテーブルの列に保存します。

**トピックベクトルの計算式の保存** 主成分スコアを回転したものを求める計算式を、データテーブルの「ベクトル」尺度の列に保存します。列の計算式には、JSL関数の`Text Score()`が使用されます。この関数の詳細については、[ヘルプ] > [スクリプトの索引] を参照してください。

**単語 トピックベクトルの保存** [単語テーブルの保存] オプションで作成したデータテーブルに、トピックベクトルの列を保存します。

**削除** 「特異値分解」レポートから「トピック分析」レポートを削除します。

---

## 線形判別分析

「テキストエクスプローラ」プラットフォームの判別分析は、文書単語行列の列に基づき、各文書がどのグループまたはカテゴリに属するのかを予測します。具体的には、各文書が1つの応答列のどのカテゴリに分類されるのかを予測します。[判別分析] オプションを使用する場合、カテゴリやグループを含む応答列を選択する必要があります。各文書がどのグループに属するのかは、文書単語行列の列に基づいて予測されます。判別分析の詳細については、『多変量分析』を参照してください。

「テキストエクスプローラ」プラットフォームの判別分析は、中心化した文書単語行列の特異値分解を基にしています。応答列の各グループのグループ平均を使って文書単語行列が中心化されます。文書単語行列が疎であることを活用しているため、「判別分析」プラットフォームより高速です。

### 判別分析の「設定」ウィンドウ

「テキストエクスプローラ」プラットフォームの[判別分析] オプションは、文書単語行列に対して行われます。文書単語行列の各列は、単語リストに含まれている各単語に対応しています（ただし、列数の最大値は、「単語の最大数」で指定したものです）。文書単語行列の各行は、1つの文書に対応しています（JMPデータテーブルにおけるテキスト列の1セルが、1文書に対応しています）。文書単語行列の要素をどのように算出するかは、「設定」ウィンドウにて指定します。

「テキストエクスプローラ」の赤い三角ボタンのメニューから[判別分析] オプションを選択すると、「設定」ウィンドウが開き、以下のオプションが表示されます。

**単語の最大数** 判別分析に含める単語の最大数。

**単語の最小度数** 判別分析に含める単語の最小度数。

**重み** 文書単語行列の要素を算出する方法。この方法については、「[文書単語行列の設定ウィンドウ](#)」を参照してください。

**特異ベクトルの数** 判別分析に含める特異ベクトルの数。デフォルト値は100です。ただし、文書数または単語数が100を下回るときは、その値がデフォルト値になります。

## JMP PRO 「判別分析」レポート

「テキストエクスプローラ」プラットフォームの「判別分析」レポートでは、デフォルトで「分類の要約」と「判別スコア」という2つのレポートが開かれます。その他のレポートは、デフォルトでは閉じています。

「判別分析」レポートには、以下のレポートが含まれます。

**単語の平均** 判別分析に使用された単語の表が表示されます。単語は、文書単語行列の列に対応しています。この表には、各単語の各グループにおける平均、全体平均、および群内標準偏差が含まれます。

**各グループ平均への距離の2乗** 各文書の、各グループへのMahalanobis距離の2乗を示します。Mahalanobis距離の詳細については、『多変量分析』を参照してください。

**各グループに属する確率** 文書がそれぞれのグループに属する確率を示します。

**分類の要約** 判別スコアの要約レポート。「判別分析」プラットフォームの「スコアの要約」レポートに対応します。

**判別スコア** 各文書に対して、予測された分類などの結果が表示されます。「判別分析」プラットフォームの「判別スコア」レポートと同じ内容です。

## JMP PRO 「判別分析」レポートのオプション

「テキストエクスプローラ」プラットフォームの「判別分析」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**正準プロット** 正準スコアをプロットしたグラフの表示／非表示を切り替えます。正準スコアは、分析変数の線形結合のなかで、グループを最も分離するものです。応答変数に3つ以上の水準があるときは、正準座標の数を指定しなければなりません。3以上に指定すると、正準プロットの行列が生成されます。

**確率の保存** データテーブルの新しい列に、各グループに属する確率を保存します。事後確率が最も高いグループを示す列も作成されます。「最尤<応答変数の名前>」の列には、最も事後確率の高い水準が含まれます。

確率の各列には、該当の文書がその応答水準に属する事後確率が含まれます。また、確率の各列には、「応答確率」列プロパティが保存されます。「応答確率」列プロパティの詳細については、『JMPの使用法』を参照してください。

**確率の計算式を保存** データテーブルの新しい列に、各グループに属する事後確率を求める計算式を保存します。最初の列には、各グループに属する事後確率を求める計算式を保存します。この計算式では、Text Score() 関数が使われます。各グループに属する事後確率の列と、応答の予測値の列も作成されます。

**正準スコアの保存** データテーブルの新しい列に正準スコアを保存します。正準スコアは、分析変数の線形結合のなかで、グループを最も分離するものです。k番目の正準スコアの列名は、「正準<k>」となります。

**削除** 「テキストエクスプローラ」レポートウィンドウから、「判別分析」レポートを削除します。

## JMP PRO 単語選択分析

「テキストエクスプローラ」プラットフォームの単語選択分析では、特定の応答変数を最もよく説明する単語を特定します。具体的には、「一般化回帰」プラットフォームで、文書単語行列に対して変数選択を実行することによって、特定の応答変数に最も影響を与えている単語を見つけます。感情分析と同様に、2値などの応答変数に対して分析を実行できます。あてはめられたモデルは、指定した応答変数の列に対して、適切な応答変数の分布を使用します。

**ヒント:** 単語選択分析の例については、[ヘルプ] > [サンプルデータフォルダ] を選択し、「Chips.jmp」を開いて、「テキストエクスプローラ: 単語の選択」スクリプトを実行してください。

## JMP PRO 単語選択分析の設定

「単語選択分析」レポートの「設定」では、応答列の選択や、応答のイベントを示す水準の指定、モデルの設定の調整を行えます。モデルの設定を指定したら、[実行] ボタンをクリックしてモデルを実行します。すると、あてはめられたモデルが「要約」レポートに表示されます。「単語選択分析の「要約」レポート」を参照してください。

### イベントを示す水準

応答列を選択すると、「イベントを示す水準」アウトラインが表示されます。

- 応答列が名義尺度である場合は、ロジスティック回帰モデルにおける「イベントを示す水準」を選択してください。「イベントを示す水準」のグループと、他のすべての水準を併合したグループとのオッズ比に対するロジスティック回帰モデルがあてはめられます。
- 応答列が順序尺度である場合、その応答列の全水準がデフォルトでモデルに含められます。モデルから特定の水準を除外するには、ローカルデータフィルタを使ってその除外する水準を選択してください。モデルに含めた水準は数値として扱われ、正規分布でモデル化されます。

**メモ:** 応答列が順序尺度である場合、応答列のデータタイプが数値である場合のみ、単語選択分析を行えます。

- 応答列が連続尺度である場合、モデルから除外したい応答の値があるときは、ローカルデータフィルタのヒストグラムを使って、それらの除外したい応答の値を選択してください。応答列の値は、正規分布でモデル化されます。
- 応答列の尺度が「多重応答」である場合、応答の水準のうちの1つまたは複数を、2値のロジスティック回帰モデルのイベントを示す水準として選択してください。複数の水準を選択した場合、それらの水準のいずれかが応答列に含まれていれば、その文書は「イベントを示す水準」に属すると見なされます。なお、**[ANDで組み合わせる]** オプションを選択すると、選択した水準すべてが応答列に含まれることが、「イベントを示す水準」にその文書が属するための条件となります。

## モデルの設定

デフォルトでは、一般化回帰モデルの推定法として[弾性ネット]、検証法として[AICc]が使用され、早期停止が許可されています。この設定は、「モデルの設定」アウトラインで変更できます。『基本的な回帰モデル』を参照してください。

---

**メモ:**「テキストエクスプローラ」の起動ウィンドウで検証列を指定した場合、「単語選択分析」レポートの一般化回帰プラットフォームでは、検証法として[検証列]が使用されます。

---

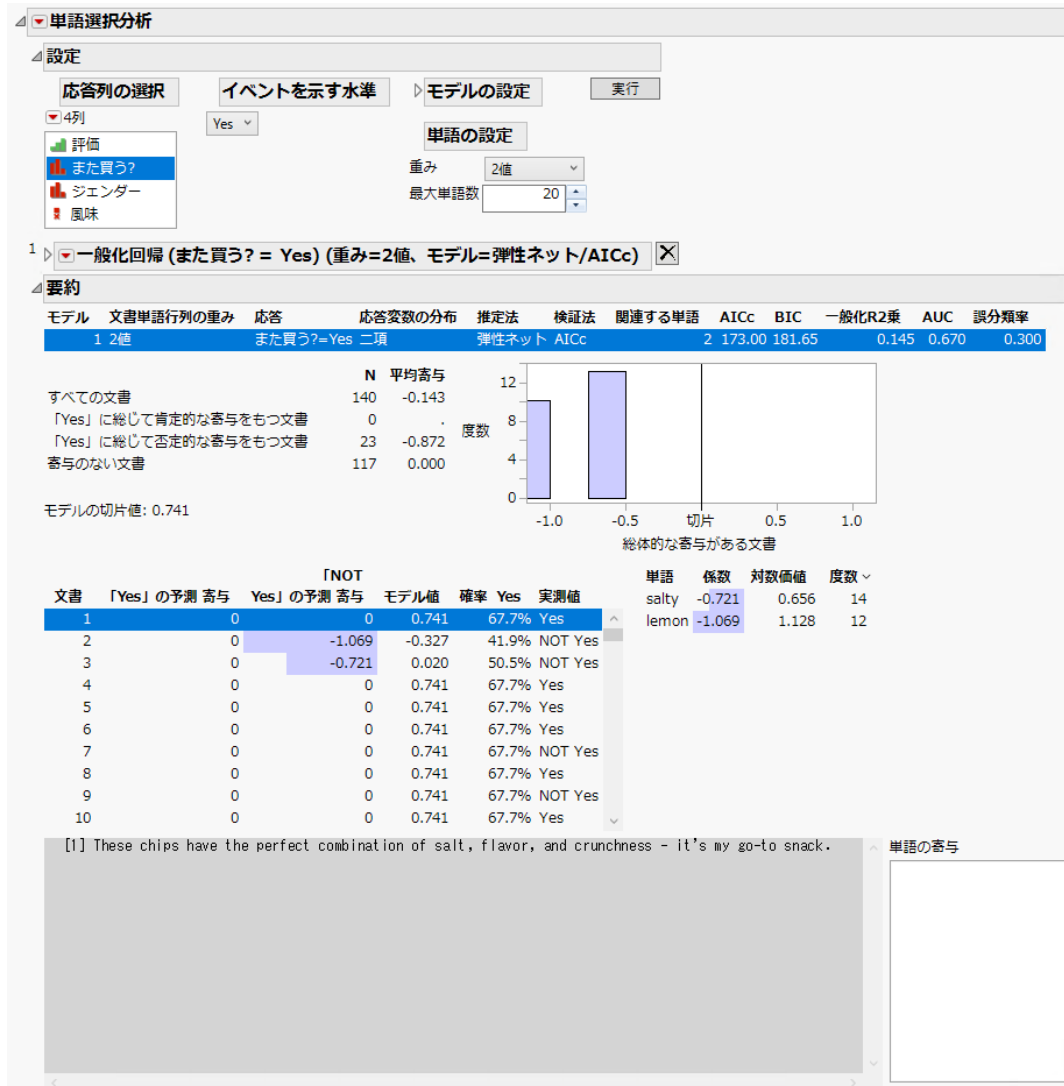
## 単語の設定

「単語の設定」では、回帰モデルで使用する文書単語行列を定義します。重みを用いる方法や、文書単語行列の列として含める単語の最大数を変更できます。なお、コーパスでの出現回数が10回未満の単語は、モデル用の文書単語行列には含められません。文書単語行列のオプションの詳細については、「[文書単語行列の設定ウィンドウ](#)」を参照してください。

## **JMP** PRO 「単語選択分析」レポート

「テキストエクスプローラ」プラットフォームで単語選択分析を実行すると、「単語選択分析」レポートの下に3つのレポートが表示されます。「設定」レポートには、一般化回帰の設定項目が表示されます。「[単語選択分析の設定](#)」を参照してください。単語選択分析を実行するたびに、「設定」レポートの下に「一般化回帰」レポートが作成されます（デフォルトで閉じられています）。『基本的な回帰モデル』を参照してください。一番下に、「要約」レポートがあります。

図 12.12 「単語選択分析」レポート



## 単語選択分析の「要約」レポート

「要約」レポートには、「モデルの比較」表、要約表およびヒストグラム、文書スコアの表、単語スコアの表、テキストボックスが表示されます。

「モデルの比較」表の各行には、あてはめられたモデルが表示されます。その下には、現在選択されているモデルの結果が表示されます。

要約表には、文書の度数と平均スコアが、すべての文書およびモデルから得た応答の予測値ごとに表示されます。「平均寄与」は、文書スコアの表に示されている寄与値の平均です。「要約」レポートのヒストグラムには、文書の全体的な寄与値の分布が表示されます。このヒストグラムは対話的に操作できます。棒をクリックすると、文書スコアの表でその文書が強調表示されます。

文書スコアの表には、各文書の肯定的な寄与、否定的な寄与と、予測値、実測値が表示されます。この予測値は、応答変数の分布が二項分布の場合は、文書が「イベントを示す水準」である確率になります。応答変数の分布が正規分布の場合は、各文書にあてはめられたモデルから求められる予測値になります。この表で行を選択すると、表の下テキストボックスにその文書のテキストが表示されます。

単語スコアの表には、あてはめたモデルによって選択された単語と、そのモデル係数、対数値、コーパス内での出現回数が表示されます。この表で行を選択すると、表の下テキストボックスにその文書のテキストが表示されます。

このテキストボックスには、文書スコアの表で文書を選択したときはその文書のテキストが、単語スコアの表で単語を選択したときはその単語を含む文書が表示されます。

## 「単語選択分析」レポートのオプション

「テキストエクスプローラ」プラットフォームの「単語選択分析」の赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**文書のスコアを保存** （「要約」レポートで一般化回帰モデルがあてはめられた後にのみ使用可能）文書スコアの列を、データテーブルに追加します。各文書の肯定的な寄与、否定的な寄与、および予測値が、新しい列として追加されます。

**単語スコアの文書単語行列を保存** （「要約」レポートで一般化回帰モデルがあてはめられた後にのみ使用可能）現在選択されている分析で使われている単語の列を、データテーブルに追加します。「単語選択分析」レポートの「単語の設定」で指定した重みの方法で計算された単語のスコアが、新しい列として追加されます。

**各予測を計算式として保存** （「要約」レポートで一般化回帰モデルがあてはめられた後にのみ使用可能）現在選択されている分析の予測式を、データテーブルに新しい列として追加します。

**単語クラウドの表示** 「要約」レポートにおいて、単語クラウドの表示／非表示を切り替えます。単語クラウドは、現在選択されている分析のモデル係数に基づいて表示されます。係数の絶対値が単語の大きさに、係数の符号が単語の色に反映されます。

**削除** 「テキストエクスプローラ」レポートウィンドウから「単語選択分析」レポートを削除します。

## JMP PRO 感情分析

「テキストエクスプローラ」プラットフォームの感情分析は、辞書に基づいて文書内の感情語を特定し、それらの語に対して、肯定的・否定的・全体的なスコアを付けます。JMPの感情分析は、英語のみしか扱えません。JMPの感情分析では、「各文書において、1つのトピックに対して2項対立な感情が自由記述形式で表されている」と仮定しています。感情分析では、基本的な自然言語処理（NLP）を使用しています。自然言語処理の詳細については、<https://opennlp.apache.org/>を参照してください。自然言語処理を使用したくない場合は、「文書の解析」オプションの選択を解除してください。

**ヒント:** 感情分析の例については、[ヘルプ] > [サンプルデータフォルダ] を選択し、「Chips.jmp」を開いて、「テキストエクスプローラ:感情分析」スクリプトを実行してください。

### メモ:

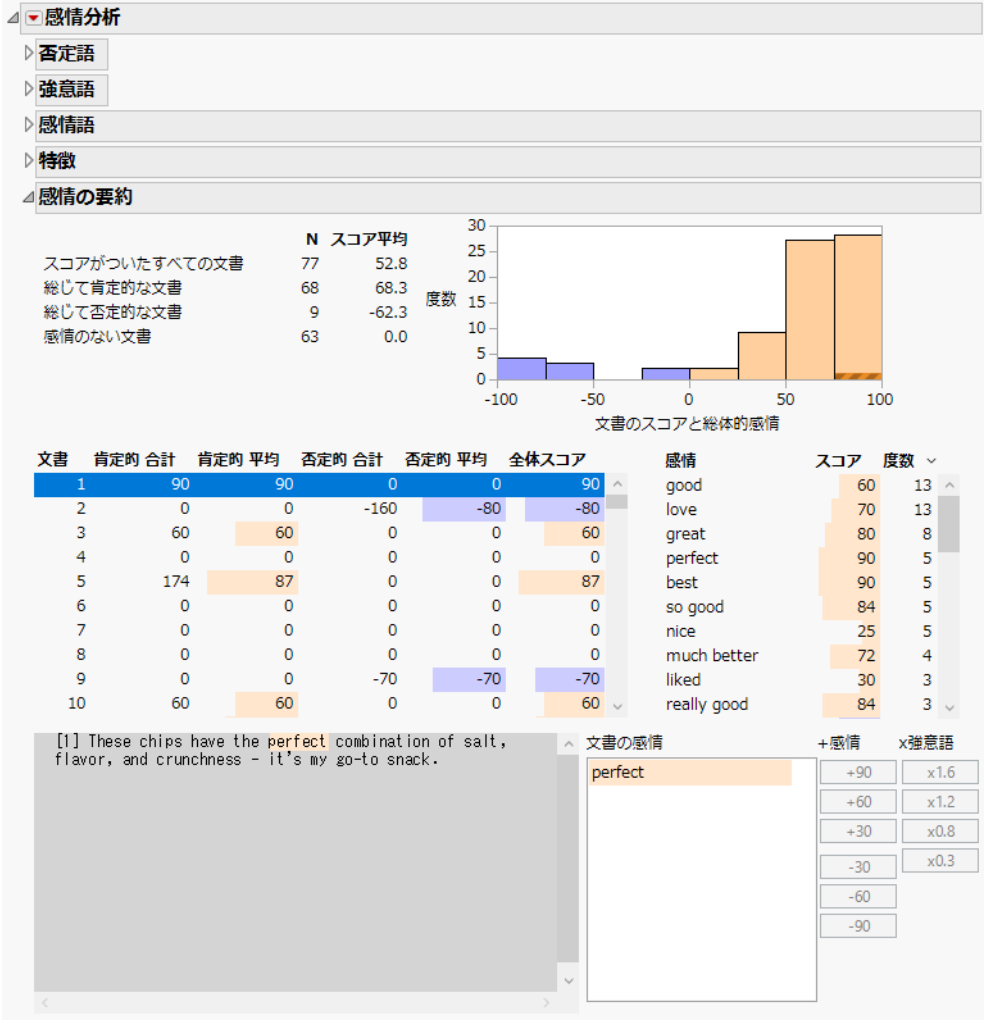
- JMPの感情分析では、1つ1つの単語が、否定・強意・感情のいずれか1つだけに分類されます。
  - 否定語は、直後にある感情語が意味する感情を否定します。「**否定語**」を参照してください。
  - 強意語は、直後にある感情語が意味する感情の度合いを表します。「**強意語**」を参照してください。
  - 感情語は、肯定的もしくは否定的な感情を表す単語です。「**感情語**」を参照してください。
- 感情分析では、ひとまとまりの文字列を、何らかの感情として認識します。「感情語」レポートまたは「感情語の管理」ウィンドウで、JMPで予め提供している感情語（ビルトインの感情語）とそのデフォルトのスコアを確認できます。
- ストップワードに指定されている単語を、否定語・強意語・感情語のいずれかに指定した場合、「感情分析」レポートを開いている間だけ、その単語はストップワードから除外されます。この一時的な除外は、「テキストエクスプローラ」レポート全体に適用されます。「感情分析」レポートを閉じると、その単語は再びストップワードに追加されます。

JMP  
PRO

「感情分析」レポート

「テキストエクスプローラ」プラットフォームの「感情分析」レポートは、デフォルトでは、「感情の要約」レポートのみが開かれた状態となっています。その他のレポートは、デフォルトでは閉じています。

図 12.13 「感情分析」レポート





「感情分析」レポートには、次のレポートが含まれます。

### 否定語

現在の感情分析に含まれる否定語の一覧です。否定語は、直後にある感情語に割り当てられている感情レベルを否定します。このリストを右クリックすると、メニューが開き、選択した単語を削除するオプションなどが表示されます。JMPには、否定語の一覧が予め用意されています。テキストにある単語を否定語として追加したり否定語から削除したりするには、「感情の要約」レポートにあるテキストボックスにて、追加・削除したい単語の上にマウスポインタを置きます。[否定語の管理] オプションも使えます。

### 強意語

強意語およびその乗数の一覧が表示されます。強意語は、直後にある感情語に割り当てられている感情レベルを調整します。このリストを右クリックすると、メニューが開き、選択した単語を削除するオプションなどが表示されます。JMPには、強意語とそれに対する乗数の一覧が予め用意されています。「強意語」アウトラインにおいて、強意語に対する乗数の値を変更できます。テキストにある単語を強意語として追加したり強意語から削除したりするには、「感情の要約」レポートにあるテキストボックスにて、追加・削除したい単語の上にマウスポインタを置きます。オプション [強意語の管理] オプションも使えます。

### 感情語

感情語およびそのスコアの一覧が表示されます。感情語は、肯定的または否定的な感情のスコアが割り当てられている単語です。このリストを右クリックすると、メニューが開き、選択した単語を削除するオプションなどが表示されます。JMPには、感情語とそれに対する感情スコアの一覧が予め用意されています。その一覧における単語の感情スコアの値は変更できます。

「感情語」アウトラインにおいて、感情語を新しく追加することもできます。「考えられる感情」表に、感情語の候補とその度数が表示されているので、その表から単語を選択し、「+感情」の下にあるボタンのいずれかをクリックすると、その単語を感情語として追加できます。ボタンに示されている値以外のスコアを指定するには、感情語の一覧に追加してから、スコアの数値を編集してください。

「考えられる感情」表の単語を選択すると、その単語を含む文書が右側に表示されるので、コーパス内でその単語が使用された文脈を確認できます。

### 特徴

「特徴」アウトラインにおいて、コーパス内の特徴語にスコアを与えることができます。特徴とは、感情語によって形容や説明されている単語です。[検索] ボタンをクリックすると、考えられる特徴の一覧が表示されます。「考えられる特徴」表から語句を1つまたは複数選択すると、右側のテキストボックスにその語句を含む文書の該当部分が表示されます。[選択されている特徴にスコアをつける] ボタンをクリックすると、選択した特徴にスコアをつけた結果が「感情の要約」レポートに表示されます。

---

**メモ：** [文書の解析] オプションを選択すると、感情語と同じ節内で使用されている特徴にスコアがつけられます。

---

## 感情の要約

「感情の要約」レポートには、現在の設定に基づいて感情分析を行った結果が表示されます。このレポートには、要約の表およびヒストグラム、文書スコアの表、感情語の表、テキストボックス、および、感情語と強意語を追加するための設定パネルが表示されます。

要約の表には、スコアの基準別に文書数とスコア平均が表示されます。スコア平均は、[スコアリング] オプションの設定に従います。「**感情分析**」レポートのオプション」を参照してください。要約のヒストグラムには、文書の全体的な感情スコアの分布が表示されます。このヒストグラムは対話的に操作できます。棒をクリックすると、文書スコアの表で該当文書が強調表示されます。

文書スコアの表には、各文書の肯定的感情および否定的感情のスコアの合計および平均と、全体的な感情スコアが表示されます。この表で行を選択すると、表の下のテキストボックスにその文書のテキストが表示されます。[スコア列] を指定した場合は、そのスコア列の値も表に表示されます。

---

**ヒント:** 文書スコアの表で、カーソルをセルに合わせると、スコアの計算が表示されます。

---

感情語の表には、感情語ごとにスコアと度数（コーパス内での出現数）が表示されます。

---

**ヒント:** 感情語が複数の単語から構成される場合は、その感情語の「スコア」列のセルにカーソルを合わせると、どのようにスコアが計算されているかを示す計算式が表示されます。

---

テキストボックスには、文書スコアの表で選択した文書のテキストか、感情語の表で選択した感情語の文が表示されます。文書スコアの表で文書を選択すると、その文書に含まれる感情の一覧がテキストボックスの右側に表示されます。

---

**ヒント:** テキストボックスで、否定語・強意語・感情語のいずれかに分類される語句にカーソルを合わせると、その分類と[削除] ボタンが表示されます。[削除] ボタンをクリックすると、その語句を分類から除外できます。

---

テキストボックスで単語を選択すると、設定パネルが有効になります。その選択した単語を感情語として追加するには、テキストボックス内で単語を選択し、「+感情」の下いずれかのボタンをクリックします。単語を強意語として追加するには、テキストボックス内で単語を選択し、「×強意語」の下ボタンをクリックします。

## 「感情分析」レポートのオプション

「テキストエクスプローラ」プラットフォームの「感情分析」レポートの赤い三角ボタンをクリックすると、以下のオプションが表示されます。

**スコアリング** 文書の全体スコアの計算方法を指定します。次のオプションがあります。

**尺度化** 肯定的な語句と否定的な語句のスコアを合計し、その合計値を文書に含まれる語句の数で割った値を、全体スコアとします。

**最小値/最大値** 肯定的スコアの最大値と、否定的スコアの最小値の合計を、全体スコアとします。

**スコア列** 算出された感情との比較対象となる既知の情報が含まれているデータテーブルの列を指定します。指定したスコア列は、文書スコアの表に追加されます。

---

**ヒント:** スコア列と全体スコアの列をグラフで比較すると、感情スコアの検討に役立ちます。

---

**文書の解析** 文書の解析に自然言語処理（NLP）を使用するかどうかを指定します。自然言語処理の詳細については、<https://opennlp.apache.org/>を参照してください。

**文書のスコアを保存** 文書スコアの列を、データテーブルに追加します。肯定的感情の合計、否定的感情の合計、肯定的感情の平均、否定的感情の平均、全体的感情の各スコアが、データテーブルに新しい列として追加されます。

**文書別に感情スコアの度数を保存** 感情語の列をデータテーブルに追加します。追加された各列には、文書内でのその感情語の出現回数が含まれます。

**否定語の表示** 「否定語」レポートの表示／非表示を切り替えます。

**強意語の表示** 「強意語」レポートの表示／非表示を切り替えます。

**感情語の表示** 「感情語」レポートの表示／非表示を切り替えます。

**特徴検索の表示** 「特徴」レポートの表示／非表示を切り替えます。

**感情クラウドの表示** 「感情の要約」レポートにおいて、感情語の単語クラウドの表示／非表示を切り替えます。

**ビルトインの否定語を含める** 感情分析で使用する否定語に、ビルトインの否定語を含めるかどうかを指定します。

**ビルトインの強意語を含める** 感情分析で使用する強意語に、ビルトインの強意語を含めるかどうかを指定します。

**ビルトインの感情語を含める** 感情分析で使用する感情語に、ビルトインの感情語を含めるかどうかを指定します。

**否定語の管理** 否定語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）としては、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された否定語をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」を参照してください。

**強意語の管理** 強意語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）としては、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された強意語をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」を参照してください。

**感情語の管理** 感情語の追加や削除を行うウィンドウを表示します。保存場所（スコープ）としては、「ユーザ」・「列」・「ローカル」があります。「ローカルの例外」では、いずれかの保存場所で指定された感情語をローカルでは除外することができます。「[\[単語のオプション\] から開く管理ウィンドウ](#)」を参照してください。

## JMP PRO 「テキストエクスプローラ」プラットフォームの別例

この例では、国家運輸安全委員会による、2001年に米国で発生した航空機のインシデントレポートを使用します。まず、事故原因の調査結果に関する記述を分析します。次に、トピックごとに分類してみて、各トピックの内容を探ります。

1. [ヘルプ] > [サンプルデータフォルダ] を選択し、「Aircraft Incidents.jmp」を開きます。
2. [行] > [列の値による色/マーカー分け] を選択します。
3. 列のリストで「死亡者」を選択し、[OK] をクリックします。  
致命的な事故の行が、赤で示されます。
4. [分析] > [テキストエクスプローラ] を選択します。
5. 「列の選択」リストで「事故原因」を選択し、[テキスト列] をクリックします。
6. 「言語」から [英語] を選択します。
7. 「言語」として [英語] を選択した後、「語幹抽出」リストから [すべての単語] を選択します。
8. 「トークン化」リストから [基本ワード] を選択します。
9. [OK] をクリックします。

図 12.14 「事故原因」のテキストエクスプローラ



レポートによると、約51,000個のトークンと、約1,900個の異なる単語が含まれていることがわかります。

10. 「単語」リストの「pilot」を右クリックし、**[行の選択]**を選択します。

データテーブルで選択されている行数を見ると、インシデントレポートに「pilot」を含む単語が1,300回以上出現していることがわかります。

11. 「pilot」を右クリックして、**[ストップワードの追加]**を選択します。

「pilot」を含む単語は、他の単語より頻繁に出現しているため、文書間の違いを読み取るためには役立ちません。「pilot」を語幹とするすべての単語が、ストップワードに追加されます。

**JMP PRO** この後の手順は、JMP Proでのみ実行できます。

12. **JMP PRO** 「「事故原因」のテキストエクスプローラ」の赤い三角ボタンをクリックし、**[潜在意味分析, 特異値分解]**を選択します。

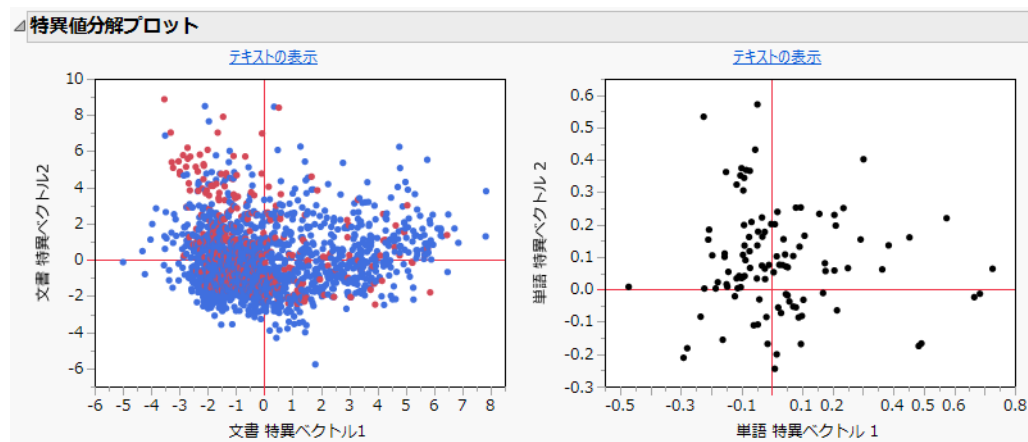
これは、特異値分解の結果を回転するトピック分析を行う前の、最初のステップです。

13. **JMP PRO** 「設定」ウィンドウで、「単語の最小度数」に「50」と入力します。

約51,000個の全トークン中、少なくとも0.1%の割合を占める単語を抽出します。

14. **JMP PRO** **[OK]** をクリックします。

図 12.15 事故原因の特異値分解プロット



文書の特異値分解プロットでは、「fatal」と「nonfatal」の違いはあまり見られません。

15. **JMP PRO** 「特異値分解 (中心化・尺度化, TF IDF)」の赤い三角ボタンをクリックし、**[トピック分析, 特異値分解を回転]**を選択します。

単語のグループから形成されるトピックを表示します。

16. **JMP PRO** 「トピックの個数」に「5」と入力します。

17. **JMP PRO** **[OK]** をクリックします。

図 12.16 「事故原因」のトピック別上位負荷量

トピック別上位負荷量									
トピック 1		トピック 2		トピック 3		トピック 4		トピック 5	
単語	負荷量	単語	負荷量	単語	負荷量	単語	負荷量	単語	負荷量
power-	0.67567	altitud-	0.48052	factor-	0.5093	control-	0.5125	fuel-	0.4864
loss-	0.66539	low-	0.45137	condit-	0.4677	direct-	0.4957	personnel-	0.4630
forc-	0.62046	dark-	0.42289	unsuit-	0.3984	experi-	0.4382	mainten-	0.4546
engin-	0.61866	night-	0.40408	accid-	0.3909	student-	0.4273	result-	0.4153
suitabl-	0.58926	maintain-	0.39283	select-	0.3842	lack-	0.3625	preflight-	0.3930
lack-	0.53292	instrument-	0.39239	associ-	0.3806	maintain-	0.3616	exhaust-	0.3663
reason-	0.47828	clearanc-	0.37881	area-	0.3628	instructor-	0.3559	inspect-	0.3640
undetermin-	0.46599	airspe-	0.33612	compens-	0.3352	supervis-	0.3282	plan-	0.3474
terrain-	0.37013	condit-	0.33473	failur-	-0.3207	power-	-0.3269	reason-	-0.3468
total-	0.31932	stall-	0.33010	wind-	0.3202	failur-	0.3171	undetermin-	-0.3438
land-	0.29402	flight-	0.32838	inadequ-	0.3080	reason-	-0.3142	improp-	0.3389
		contin-	0.31729	result-	-0.2845	undetermin-	-0.3101	inadequ-	0.3353
		maneu-	0.30931	terrain-	0.2663	crosswind-	0.3085	subsequ-	0.3246
		weather-	0.29484	airspe-	-0.2542	aircraft-	0.3007	maintain-	-0.3117
		adequ-	0.28082			factor-	0.2787	due-	0.2882

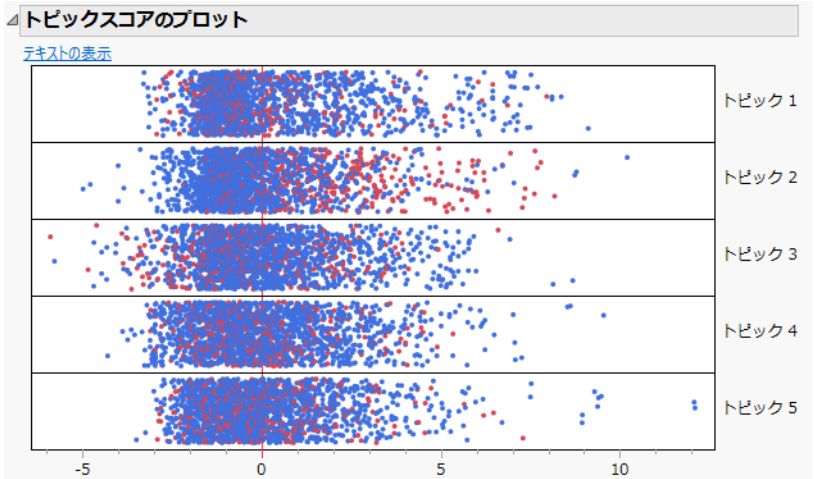
各トピック中の負荷量が高い単語は、インシデントレポートの主題を見極める手がかりとなります。

たとえばトピック 1 では、「power」、「loss」、「engine」という単語が高い負荷量を記録しており、エンジンの出力低下に起因する事故を示唆しています。これは、「loss of engine power」という句が、インシデントレポート全体に 273 回出現しているのに対応しています。

トピック 2 では、「low」、「dark」といった単語が高い負荷量を記録しており、暗さや低空と関連する事故を示唆しています。


18. **JMP PRO** 「トピックスコアのプロット」のグレーの開閉ボタンをクリックします。

図 12.17 事故原因のトピックスコアのプロット



トピックスコアのプロットでは、コーパスに含まれる文書がそれぞれ 1 つの点で表されます。プロットで点を選択して、特定の文書のテキストを詳しく調べることができます。

トピック2について調べたいとしましょう。

19.  トピック2のプロット上で一番右に位置する3つの点を選択し、グラフの左上にある**【テキストの表示】**ボタンをクリックします。

新しいウィンドウが開き、トピック2の中で最もスコアが高い3つの文書のテキストが表示されます。それを見ると、やはりトピック2と低空に関連があることがわかります。

この後にも、様々な分析が考えられます。テキスト分析は、反復的なプロセスです。トピック情報を参考に、ストップワードを追加したり、特定の句を単語に指定したりするなどして、単語リストを整理してもよいでしょう。また、文書単語行列・主成分スコア（左特異ベクトル）・回転した主成分スコアをデータテーブルに保存して、JMPの他のプラットフォームで分析を進めてもよいでしょう。そのとき、テキスト分析で得られた情報と、データにある他の情報との関係を調べてみるのもよいでしょう。





---

『基本的な統計分析』では、以下の文献を参考にしています。

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Agresti, A., and Coull, B. A. (1998). "Approximate is Better Than 'Exact' for Interval Estimation of Binomial Proportions." *American Statistician* 52:119–126.
- Ament, S., Gregoire, J., and Gomes, C. (2019). "Exponentially-Modified Gaussian Mixture Model: Applications in Spectroscopy." *arXiv preprint arXiv:1902.05601*.
- Asiribo, O., and Gurland, J. (1990). "Coping with Variance Heterogeneity." *Communication in Statistics: Theory and Methods* 19:4029–4048.
- Bablok, W., Passing, H., Bender, R., and Schneider, B. (1988). "A General Regression Procedure for Method Transformation." *Journal of Clinical Chemistry Clinical Biochemistry* 26:783–90.
- Bartlett, M. S., and Kendall, D. G. (1946). "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation." *Supplement to the Journal of the Royal Statistical Society* 8:128–138.
- Bissell, A. F. (1990). "How Reliable is Your Capability Index?" *Applied Statistics* 30:331–340.
- Boos, D. D., and Duan, S. (2021). "Pairwise Comparisons Using Ranks in the One-Way Model." *American Statistician* 75:414–423.
- Brown, M. B., and Benedetti, J. K. (1977). "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables." *Journal of the American Statistical Association* 72:305–315.
- Brown, M. B., and Forsythe, A. B. (1974). "Robust Tests for Equality of Variances." *Journal of the American Statistical Association* 69:364–367.
- Chen, S.-X., and Hall, P. (1993). "Empirical Likelihood Confidence Intervals for Quantiles." *The Annals of Statistics* 21:1166–1181.
- Chou, Y.-M., Owen, D. B., and Borrego, S. A. (1990). "Lower Confidence Limits on Process Capability Indices." *Journal of Quality Technology* 22:223–229.
- Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74:829–836.
- Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Education Psychological Measurement* 20:37–46.
- Conover, W. J. (1972). "A Kolmogorov Goodness-of-fit Test for Discontinuous Distributions." *Journal of the American Statistical Association* 67:591–596.
- Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York: John Wiley & Sons.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. New York: John Wiley & Sons.

- Cureton, E. E. (1967). "The Normal Approximation to the Signed-Rank Sampling Distribution when Zero Differences are Present." *Journal of the American Statistical Association* 62:1068–1069.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics* 44:837–845.
- Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences*. Pacific Grove, CA: Duxbury Press.
- Dunn, O. J. (1964). "Multiple Comparisons Using Rank Sums." *Technometrics* 6:241–252.
- Dunnett, C. W. (1955). "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50:1096–1121.
- Efron, B. (1981). "Nonparametric Standard Errors and Confidence Intervals." *The Canadian Journal of Statistics* 9:139–158.
- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. 2nd ed. Boca Raton, Florida: CRC.
- Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969). "Large-Sample Standard Errors of Kappa and Weighted Kappa." *Psychological Bulletin* 72:323–327.
- Friendly, M. (1994). "Mosaic Displays for Multi-Way Contingency Tables." *Journal of the American Statistical Association* 89:190–200.
- Goodman, L. A., and Kruskal, W. H. (1979). *Measures of Association for Cross Classification*. New York: Springer-Verlag.
- Gupta, S. S. (1965). "On Some Multiple Decision (Selection and Ranking) Rules." *Technometrics* 7:225–245.
- Hajek, J. (1969). *A Course in Nonparametric Statistics*. San Francisco: Holden-Day.
- Hartigan, J. A., and Kleiner, B. (1981). "Mosaics for Contingency Tables." In *Computer Science and Statistics: Proceedings of the Thirteenth Symposium on the Interface*, edited by W. F. Eddy, 268–273. New York: Springer-Verlag.
- Hayter, A. J. (1984). "A Proof of the Conjecture That the Tukey-Kramer Method Is Conservative." *Annals of Mathematical Statistics* 12: 61–75.
- Hosmer, D. W., and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.
- Howell, D. C. (2013). *Statistical Methods for Psychology*. 8th ed. Belmont, California: Wadsworth.
- Hsu, J. (1981). "Simultaneous Confidence Intervals for All Distances from the 'Best'." *Annals of Statistics* 9:1026–1034.
- Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall.
- Huber, P. J. (1973). "Robust Regression: Asymptotics, Conjecture, and Monte Carlo." *Annals of Statistics* 1:799–821.
- Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*. 2nd ed. New York: John Wiley & Sons.

- Iman, R. L. (1974). "Use of a t-statistic as an Approximation to the Exact Distribution of Wilcoxon Signed Ranks Test Statistic." *Communications in Statistics—Simulation and Computation* 3:795–806.
- Jones, M. C., and Pewsey, A. (2009). "Sinh-Arcsinh Distributions." *Biometrika* 96:761–780.
- Kendall, M., and Stuart, A. (1979). *The Advanced Theory of Statistics*. 4th ed. Vol. 2. New York: Macmillan.
- Keuls, M. (1952). "The Use of the 'Studentized Range' in Connection with an Analysis of Variance." *Euphytica* 1.2:112–122.
- Kramer, C. Y. (1956). "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." *Biometrics* 12:307–310.
- Lehmann, E. L., and D'Abrera, H. J. M. (2006). *Nonparametrics: Statistical Methods Based on Ranks*. Rev. ed. San Francisco: Holden-Day.
- Levene, H. (1960). "Robust Tests for the Equality of Variance." In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann. Palo Alto, CA: Stanford University Press.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.
- Meeker, W. Q., Hahn, G. J., and Escobar, L. A. (2017). *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. New York: John Wiley & Sons.
- Miller, A. J. (1972). "Letter to the Editor." *Technometrics* 14:507.
- Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78:691–692.
- Nelson, P. R., Wludyka, P. S., and Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia: Society for Industrial and Applied Mathematics.
- Neter, J., Wasserman, W., and Kutner, M. H. (1996). *Applied Linear Statistical Models*. 4th ed. Boston: Irwin.
- O'Brien, R. G. (1979). "A General ANOVA Method for Robust Tests of Additive Models for Variances." *Journal of the American Statistical Association* 74:877–880.
- O'Brien, R., and Lohr, V. (1984). "Power Analysis For Linear Models: The Time Has Come." *Proceedings of the Ninth Annual SAS User's Group International Conference* 840–846. Cary, NC: SAS Institute Inc.
- Olejnik, S. F., and Algina, J. (1987). "Type I Error Rates and Power Estimates of Selected Parametric and Nonparametric Tests of Scale." *Journal of Educational Statistics* 12:45–61.
- Palmer, E. M., Horowitz, T. S., Torralba, A., and Wolfe, J. M. (2011). "What are the Shapes of Response Time Distributions in Visual Search?" *Journal of Experimental Psychology: Human Perception and Performance* 37:58–71.

- Passing, H., Bablok, W. (1983). "A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods". *Journal of Clinical Chemistry Clinical Biochemistry* 21:709–20.
- Passing, H., Bablok, W. (1984). "Comparison of Several Regression Procedures for Method Comparison Studies and Determination of Sample Sizes". *Journal of Clinical Chemistry Clinical Biochemistry* 22:431–45.
- Pratt, J. W. (1959). "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures." *Journal of the American Statistical Association* 54:655–667.
- Reinsch, C. H. (1967). "Smoothing by Spline Functions." *Numerische Mathematik* 10:177–183.
- Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
- Rubin, D. (1981). "The Bayesian Bootstrap." *The Annals of Statistics* 9:130–134.
- SAS Institute Inc. (2023a). "Introduction to Nonparametric Analysis." In *SAS/STAT® User's Guide*. Cary, NC: SAS Institute Inc.  
[https://go.documentation.sas.com/api/collections/pgmsascdc/9.4\\_3.5/docsets/statug/content/intronpar.pdf](https://go.documentation.sas.com/api/collections/pgmsascdc/9.4_3.5/docsets/statug/content/intronpar.pdf).
- SAS Institute Inc. (2023b). "The FREQ Procedure." In *SAS/STAT® User's Guide*. Cary, NC: SAS Institute Inc.  
[https://go.documentation.sas.com/api/collections/pgmsascdc/9.4\\_3.5/docsets/statug/content/freq.pdf](https://go.documentation.sas.com/api/collections/pgmsascdc/9.4_3.5/docsets/statug/content/freq.pdf).
- Schuurmann, D. J. (1987). "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15:657–680.
- Slifker, J. F., and Shapiro, S. S. (1980). "The Johnson System: Selection and Parameter Estimation." *Technometrics* 22:239–246.
- Snedecor, G. W., and Cochran, W. G. (1980). *Statistical Methods*. 7th ed. Ames, Iowa: Iowa State University Press.
- Somers, R. H. (1962). "A New Asymmetric Measure of Association for Ordinal Variables." *American Sociological Review* 27:799–811.
- Stephens, M. A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association* 69:730–737.
- Tan, C. Y., and Iglewicz, B. (1999). "Measurement-Methods Comparisons and Linear Statistical Relationship." *Technometrics* 41:192–201.
- Tamhane, A. C., and Dunlop, D. D. (2000). *Statistics and Data Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Tukey, J. W. (1953). "The Problem of Multiple Comparisons." In *Multiple Comparisons, 1948–1983*, edited by H. I. Braun, vol. 8 of *The Collected Works of John W. Tukey* (published 1994), 1–300. London: Chapman & Hall. Unpublished manuscript.
- Welch, B. L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach." *Biometrika* 38:330–336.
- Wheeler, D. J. (2003). *Range Based Analysis of Means*. Knoxville, TN: SPC Press.

- Wilson, E. B. (1927). "Probable Inference, the Law of Succession, and Statistical Inference." *Journal of the American Statistical Association* 22:209–212.
- Wludyka, P. S., and Nelson, P. R. (1997). "An Analysis-of-Means-Type Test for Variances From Normal Populations." *Technometrics* 39:274–285.

