

統計解析ソフトウェア JMP Pro 入門 セミナー

ver.2.0

下川敏雄

和歌山県立医科大学 医学部 医療データサイエンス学
和歌山県立医科大学附属病院 臨床研究センター

目次

第 0 章	JMP の基本操作	1
0.1	Excel からのデータ入力	1
0.1.1	データのインポート	1
0.1.2	変数のラベル付け	2
0.1.3	個体の除外	3
第 1 章	仮説検定の基礎	4
1.1	定量データの比較：2 標本検定・共分散分析 (ANCOVA)	4
1.1.1	データの選択	4
1.1.2	2 標本データの比較	5
1.1.3	対応のあるデータの解析	11
1.1.4	ベースラインがある場合・調整因子がある場合の比較：共分散分析	12
1.2	カテゴリカルデータの評価	14
1.2.1	カイ 2 乗検定・Fisher の正確検定	14
1.2.2	傾向性の検定 (Cochran-Armitage 検定)	15
1.2.3	覚書：度数を入力した場合のクロス集計表の構成	16
1.2.4	対応のあるクロス集計表の解析	18
1.2.5	調整因子がある場合の検定 (Cochran-Mantel-Haenszel 検定)	20
第 2 章	分散分析	22
2.1	多標本の比較	22
2.1.1	一元配置の分散分析	23
2.1.2	パラメトリックな多重比較法	24
2.1.3	Kruskal-Wallis 検定	24
2.1.4	ノンパラメトリックな多重比較法	25
2.2	多元配置の分散分析	26
第 3 章	一般化線形モデル	29
3.1	重回帰分析	29
3.2	ロジスティック回帰分析	32
3.3	Poisson 回帰分析	35
第 4 章	生存時間解析	38
4.1	生存時間曲線の推定及び解析	38
4.2	Cox 比例ハザード・モデル	39

第0章 JMP の基本操作

0.1 Excel からのデータ入力

0.1.1 データのインポート

ここでは、Excel で作成されたデータ Example.xlsx を JMP により読み込む。このデータは、2 種類の治療法のアウトカムを比較したときの結果である。Excel でのデータの内容は

	A	B
1	治療群	効果
2	1	35.4
3	1	37.8
4	1	45.0
5	1	45.3
6	1	30.1
7	1	32.3
8	1	40.5
9	1	29.7
10	1	32.5
11	1	31.3
12	1	36.3
13	2	32.5
14	2	27.1
15	2	33.2
16	2	28.6
17	2	32.2
18	2	30.2
19	2	30.9
20	2	30.0
21	2	50.5

のようになっている。ここでの注意点は、20 例目の症例は打ち間違いであることにある。

このデータを JMP に入力する方法は、

「ファイル」→「開く」

を選択し、「Example.xlsx」を開く。そうすると、図 1 のような入力ウィザードが表示される。




このウィザードでは、Excel のワークシートの選択、列見出し (変数名) の開始行・行数、データの開始行・開始列を選択することができる。Example.xlsx では、最初の列 (1 列) に変数名が入力されており、A 行からデータが入っているため、デフォルトの状態ですべての項目が「読み込み」に入力すればよい。Excel 読み込みウィザードでは、その他の設定を行うこともできるが、実際には、インポート後にも変更可能である。




図 1: JMP のインポート・ウィザード画面

	治療群	結果
1	1	35.4
2	1	37.8
3	1	45.0
4	1	45.3
5	1	30.1
6	1	32.3
7	1	40.5
8	1	29.7
9	1	32.5
10	1	31.3
11	1	36.3
12	2	32.5
13	2	27.1
14	2	33.2
15	2	28.6
16	2	32.2
17	2	30.2

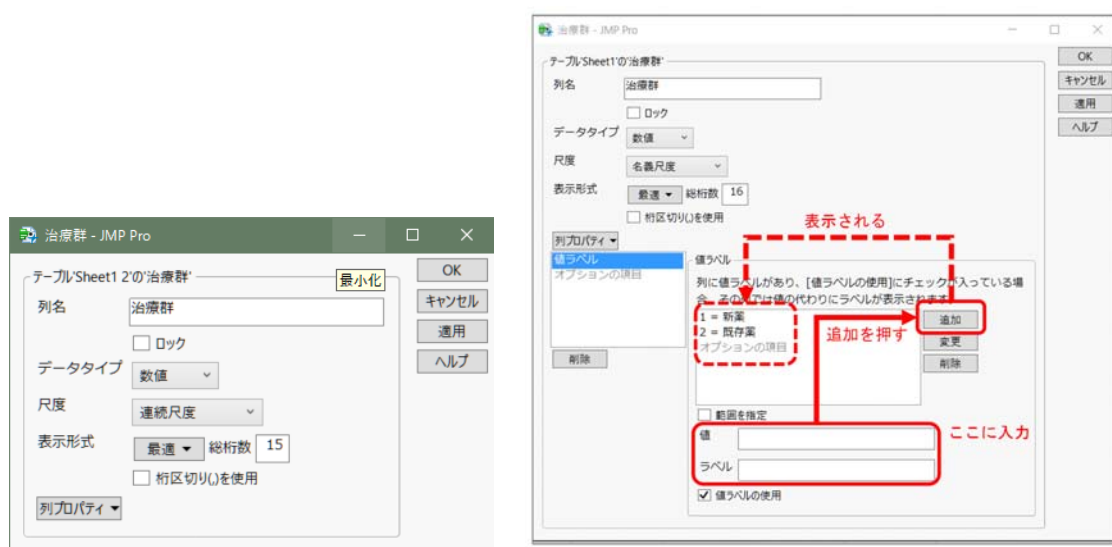
図 2: JMP にインポートされた画面

図 2 は、JMP へのインポートの結果である。JMP の変数の形式には、連続尺度 , 順序尺度 , 及び名義尺度  が存在する。

Example.xlsx では、新薬に「1」、既存薬に「2」のラベルを付けている。そのため、変数「治療群」は連続尺度として認識されている。そのため、名義尺度に変換しなければならない。この変換は、図 2 の左側中央にある「変数」のなかの治療群の右側にある  を右クリックし、そのなかから「名義尺度」を選択することで変数の尺度を変更することができる。

0.1.2 変数のラベル付け

JMP Pro では変数にラベルを付けることができる。ラベル付けの方法を以下に示す。



(a) 変数のプロパティ画面

(b) 変数のラベル入力画面

図 3: 変数のラベル入力

行	治療群	効果
1	新薬	35.4
2	新薬	37.8
3	新薬	45.0
4	新薬	45.3
5	新薬	30.1
6	新薬	32.3
7	新薬	40.5
8	新薬	29.7
9	新薬	32.5
10	新薬	31.3
11	新薬	36.3
12	既存薬	32.5
13	既存薬	27.1
14	既存薬	33.2
15	既存薬	28.6
16	既存薬	32.2
17	既存薬	30.2
18	既存薬	30.9
19	既存薬	30.0
20	既存薬	50.5


図 4: JMP にインポートされた画面

変数のラベル付け

- [1] 図 2 の上側の変数「治療群」と書かれた部分をダブルクリックする (変数のプロパティ画面が表示される)。
- [2] 「列プロパティ」→「値ラベル」を選択する (図 1.1(b) が表示される)。
- [3] 「値」に「1」と入力し、「ラベル」に「新薬」と入力し、「追加」ボタンを押す。
- [4] 「値」に「2」と入力し、「ラベル」に「既存薬」と入力し、「追加」ボタンを押す。
- [5] 「OK」ボタンを押す。

これらの手順を実施することで、図 4 のように変換できる (左側の「列」のなかのラベルを付けた変数「治療群」にはアスタリスク「*」が表示されていることに注意されたい)。

0.1.3 個体の除外

前述したように個体 21 は、打ち間違いのデータであり、除外する必要がある。このような場合には、個体番号 1, 2, ..., 20 の列 (灰色の列) の「20」のセルを右クリックし「除外する / 除外しない」をクリックする。すると、「20」の右側に  が表示され、解析から除外することができる (図 2 のなかの左下の「行」の「除外されている行」が「1」になっていることに注目されたい)。

第1章 仮説検定の基礎

1.1 定量データの比較：2 標本検定・共分散分析 (ANCOVA)

1.1.1 データの選択



ここでは、JMP のサンプルデータ (「Big Class.jmp」) を利用する。サンプルデータの場所はユーザーが選択した場所 (デフォルトでは C:\Program Files\SAS\JMP\13) 以下にあるフォルダ Samples\Data に保存されている。

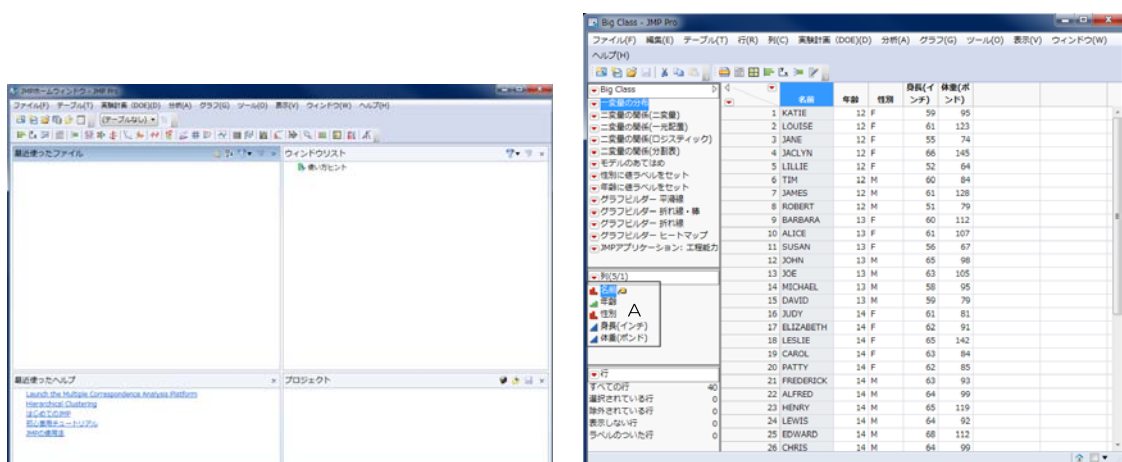
JMP を起動すると図 1.1(a) が表示される。ここで、

「ファイル」→「開く」

を選択し、「Big Class.jmp」を選択する¹。図 1.1(b) は、データを読み込んだときに開くデータテーブル画面である。このデータは、12 歳から 17 歳までの子供の性別、身長 (インチ)、および体重 (ポンド) からなるシンプルな事例データである。

ここで、左上の「Big Class」、「一変量の分布」、…、「JMP アプリケーション：工程能力」は、サンプルデータに含まれるマクロである。マクロとは、自動的に統計計算するためのプログラミング言語であり、本演習とは関係がない (実行するとグラフや統計計算が自動的に行われ、アウトプットが表示される)。

左中の A で表されている領域は、右側のデータの各変数の形式に対応している。「名前」「性別」の  は名義尺度を表しており、「年齢」の  は順序尺度を表しており、そして、「身長 (インチ)」、「体重 (ポンド)」

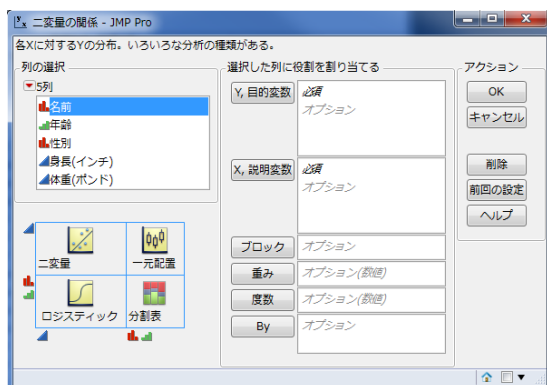


(a) ホーム画面

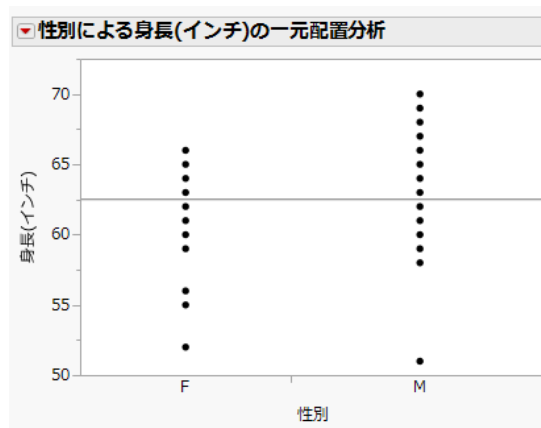
(b) データテーブル画面

図 1.1: JMP の画面構成

¹ サンプルサイズを開く場合には、「ヘルプ」→「サンプルデータライブラリ」を選択すると直接サンプルデータを選択できる。



(a) 二変量の関係でのウィンドー画面



(b) 二変量の関係での解析画面

図 1.2: 二変量の関係での画面

の▲は定量 (量的) データを表している。JMP では、変数の形式に対応して統計手法を自動判別する。そのため、変数の形式を確認することは重要である。なお、変数の形式の変更は A の部分の当該の変数を右クリック (Windows の場合) し「列情報」を選択してからウィンドー内の「尺度」のプルダウンメニューから選択すればよい。また、一括して変換したい場合には、当該の複数の変数を選択し (CTRL キーを押しながら左クリックで選択し)、右クリックで「列属性の一括設定」を選択してからウィンドー内の真ん中の「属性」プルダウンメニューから選択すればよい。

1.1.2 2 標本データの比較

2 標本データの比較 (2 標本 t 検定, Welch 検定, Wilcoxon 検定等) はすべて同じ手続きで行われる。ここでは、身長 (インチ) に性差があるか否かを検定する。

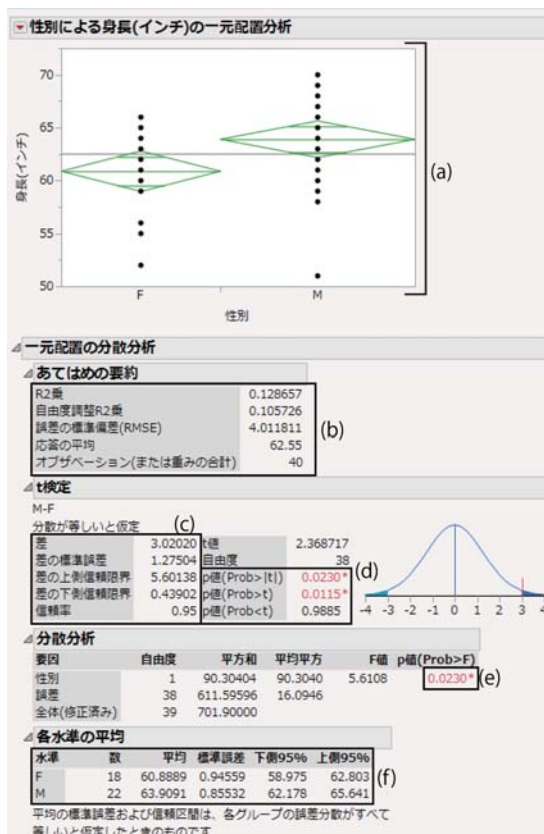
2 標本データの比較を実行するための流れ

- [1] 「分析」→「二変量の関係」を選択する (図 1.2(a) が表示される)。
- [2] 「列の選択」から「身長 (インチ)」を「Y, 目的変数」, 「性別」を「X, 説明変数」にドラッグする (「身長 (インチ)」を選択した後で「Y, 目的変数」を左クリック, 「性別」を選択した後で「X, 説明変数」を左クリックでも同様である)。
- [3] 「OK」ボタンを押す。

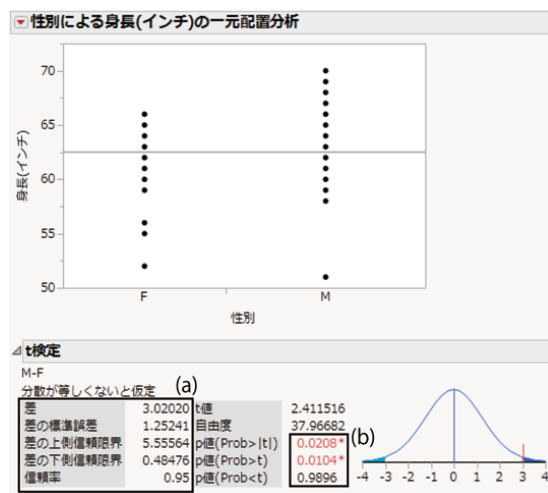
上記の作業を行うと、図 1.2(b) が表示される。これは、性別 (説明変数) で分けられた身長 (インチ) (応答変数) の 1 次元散布図である。女子 (F) に比べて男子 (M) のほうが相対的に身長が高いことが示唆される。仮説検定は、▲を左クリックすることで手法を選択しながら進める。

2 標本 t 検定の場合

2 標本 t 検定 (母分散が等しいと仮定できる場合の t 検定) の計算の流れを以下に示す。




(I) 2 標本 t 検定の結果



(II) Welch 検定の結果

図 1.3: 2 標本データの比較に対する検定の諸型

2 標本 t 検定の方法

- [0] 上記「2 標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「平均/ANOVA/プーリングした t 検定」を選択する。

上記の作業を行うと、図 1.3(I) が表示される。ここで、(a) は先ほどの 1 次元散布図に緑色のひし形が追加された図である。ここで、中央の水平線はデータ全体での標本平均を表しており、ひし形の縦方向の頂点は母平均に対する 95%信頼区間を表しており、横方向の大きさは標本サイズに対応している。すなわち、ひし形の横幅が大きいほど標本サイズが大きいということを意味する。

(b) は一元配置の分散分析におけるモデル適合の結果であるが、2 標本 t 検定での比較の場合には不必要な情報である。

(c) は母平均の差 $\mu_2 - \mu_1$ (ここに μ_1 は女子の身長に対する母平均、 μ_2 は男子の身長に対する母平均) の点推定値 (標本平均の差) $\bar{x}_2 - \bar{x}_1$ と母平均の差に対する 95%信頼区間である。

(d) は 2 標本 t 検定の結果である。JMP では片側対立仮説と両側対立仮説の 3 種類の仮説に対して p 値を同時に表示する。表示の意味を以下に示す：

p 値 (Prob>|t|) 両側検定 (対立仮説 $H_1: \mu_1 \neq \mu_2$) での p 値 (男子と女子で身長之母平均に差がある)
p 値 (Prob>t) 片側検定 1 (対立仮説 $H_1: \mu_1 < \mu_2$) での p 値 (男子の身長之母平均は女子よりも高い)
p 値 (Prob<t) 片側検定 2 (対立仮説 $H_1: \mu_1 > \mu_2$) での p 値 (女子の身長之母平均は男子よりも高い)

その結果、両側検定、片側検定 1 で有意水準 0.05 のもとで有意差が認められた (両側検定: 0.0230, 片側検定 1: 0.0115)。研究計画前の仮説によって、どの結果を採用するかは異なるが、いずれにしても、身長には性差が認められる。


(e) は一元配置の分散分析での結果である。2 水準での一元配置の分散分析と 2 標本 t 検定の結果は一致するため、一元配置の分散分析の p 値 (d) は両側検定での 2 標本 t 検定での p 値と一致している。

(f) は群毎での標本平均及び 95% 信頼区間である。その結果、女子よりも男子のほうが身長が高いことが伺える。このことは、片側検定 1 が有意だったことから明らかである。

Welch 検定の場合

Welch 検定 (母分散が等しいと仮定できない場合の t 検定) の計算の流れを以下に示す。

Welch 検定の方法

- [0] 上記「2 標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「個々の分散を用いた t 検定」を選択する。

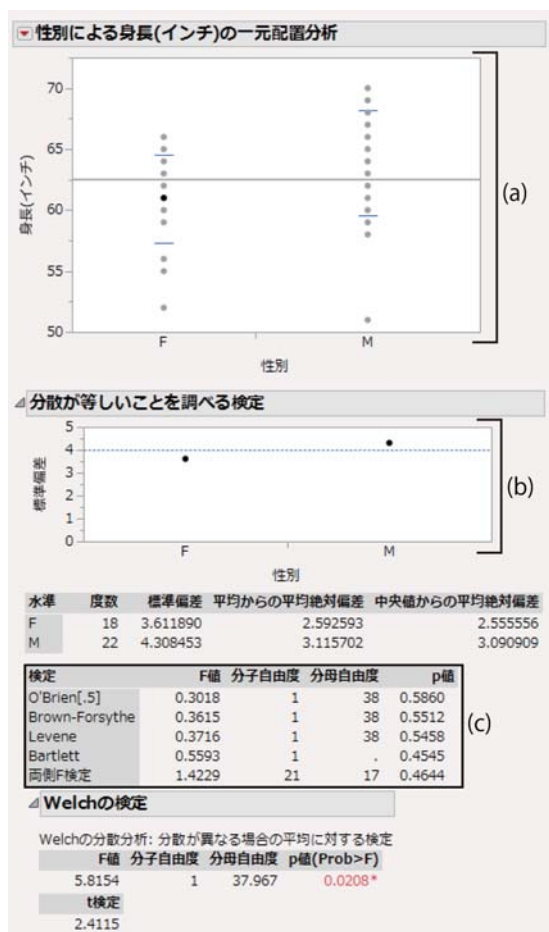
上記の作業を行うと、図 1.3(II) が表示される。Welch 検定の場合には、1 次元散布図に標本平均及び 95% 信頼区間を表示することがなく、また、各群に対する標本平均及び 95% 信頼区間を計算しないことに注意されたい。

(a) は母平均の差 $\mu_2 - \mu_1$ に対する点推定値 (標本平均の差) $\bar{x}_2 - \bar{x}_1$ と母平均の差に対する 95% 信頼区間である。母平均の差に対する 95% 信頼区間の結果は、等分散を仮定した場合の 95% 信頼区間と若干数値が異なる。今回の場合には、等分散性を仮定しないほうが信頼区間の幅が狭くなっているが、いかなる場合においても同様であるとは限らないので注意が必要である。

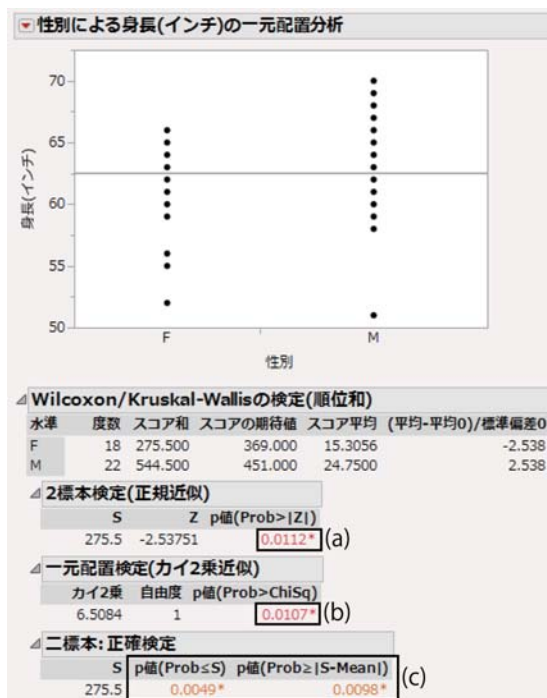
(b) は Welch 検定での結果である。2 標本 t 検定の場合と同様に、JMP では片側対立仮説と両側対立仮説の 3 種類の仮説に対して p 値を同時に表示する。その結果、両側検定、片側検定 1 で有意水準 0.05 のもとで有意差が認められた (両側検定: 0.0208, 片側検定 1: 0.0104)。したがって、2 標本 t 検定と同様の結論が与えられた。

覚書: 2 標本 t 検定と Welch 検定の使い分け

正規分布は、平均と分散 (標準偏差) の 2 つの測度 (パラメータ) によって構成される。独立 2 標本の場合には、母集団における分散が同じであるか否かによって検定方法が異なる。母集団における分散が同じであると仮定できる場合には、2 標本 t 検定を用いることができ、異なる場合には、Welch の検定を用いることが考えられる。



(III) 等分散性の検定



(IV) Wilcoxon 検定の結果


1.3(続き) : 2 標本データの比較に対する検定の諸型

ただし, Welch の検定の利用については, 批判的な意見が報告されている. 独立 2 標本での比較の関心は, (1) 母集団の違いに差があるのか, (2) 平均値の差にあるのか, に大別される. 関心の対象が (1) である場合には, 不等分散であることを示すことができれば (等分散性の検定), Welch 検定を用いる必要は必ずしも存在しない. 関心の対象が (2) である場合においても, 試験結果の分散 (標準偏差) に明らかな違いがなければ, 2 標本 t 検定で十分であることがいくつかの文献で指摘されている. また, 母集団の分散が明らかに異なる場合には, 母集団が正規分布に従っていないことが想定されるため, ノンパラメトリック検定を用いることが推奨される

等分散性の検定

医療機器の精度などを比較する場合には, 平均よりもむしろ分散が問題になる. このような場合には等分散性の検定が利用される. 等分散性の検定の計算の流れを以下に示す.

等分散性の検定

- [0] 上記「2 標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「等分散性の検定」を選択する。

上記の作業を行うと、図 1.3(III) が表示される。


(a) は 1 次元散布図に標本平均 \pm 標準偏差の区間幅を横線で表したものである。また、水平線はデータ全体での標本平均を表している。また、(b) は標準偏差を表している。ここで、水平線は併合した場合の標準偏差であり、等分散性が示唆される場合には、水平線付近にプロットされる

(c) は等分散性に対する 5 種類の検定結果 (O'Brein 検定, Brown-Forstythe 検定, Levene 検定, Bartlett 検定, 両側 F 検定) である。最も p 値が小さいのが Bartlett 検定である (p 値: 0.4545)。Bartlett 検定は非正規性 (正規分布からのずれ) に敏感 (感度が高い) であり、等分散性を適切に判断できない検定であるため推奨されない (Finney, 1995)。両側 F 検定は 2 個の分散の比を利用した検定であり、最も知られている検定法の一つである。Levene 検定も同様に広く知られている方法である。O'Brein 検定及び Brown-Forstythe 検定は Levene 検定の修正版 (頑健版) であり、外れ値がある場合などには Levene 検定よりも良いとされている。また、O'Brein 検定と Brown-Forstythe 検定では若干の p 値の違いがあるものの、その検定の動機はほぼ同じである。したがって、正規分布のもとでバラツキを評価する (分散・標準偏差に意味を持たせる) のであれば Levene 検定が推奨され、非正規分布のもとでバラツキを評価する (ノンパラメトリックな状況) 場合には O'Brein 検定あるいは Brown-Forstythe 検定が推奨される。本データではいずれの場合においても等分散性が仮定できそうである。

Wilcoxon 検定 (Mann-Whitney U test)

Wilcoxon 検定 (ノンパラメトリック検定) の計算の流れを以下に示す。

Wilcoxon 検定の方法

- [0] 上記「2 標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「ノンパラメトリック」→「Wilcoxon 検定」を選択する。

上記の作業を行うと、図 1.3(IV) が表示される。もし、3 標本以上の場合には Kruskal-Wallis 検定になる。Wilcoxon 検定では検定統計量の帰無分布は近似により与えられる。(a) は正規近似の場合であり、(b) はカイ 2 乗近似の場合である。本来であれば、p 値が一致するのに若干異なるのは、正規近似の場合には連続性の補正を行っているためである。標本サイズが少ない場合には連続性の補正がある (a) のほうが精度が良いとされている。因みに、標本サイズが増えれば 2 個の p 値はほぼ同じになる。さらに、(c) は

「ノンパラメトリック検定」→「正確検定」→「正確 Wilcoxon 検定」

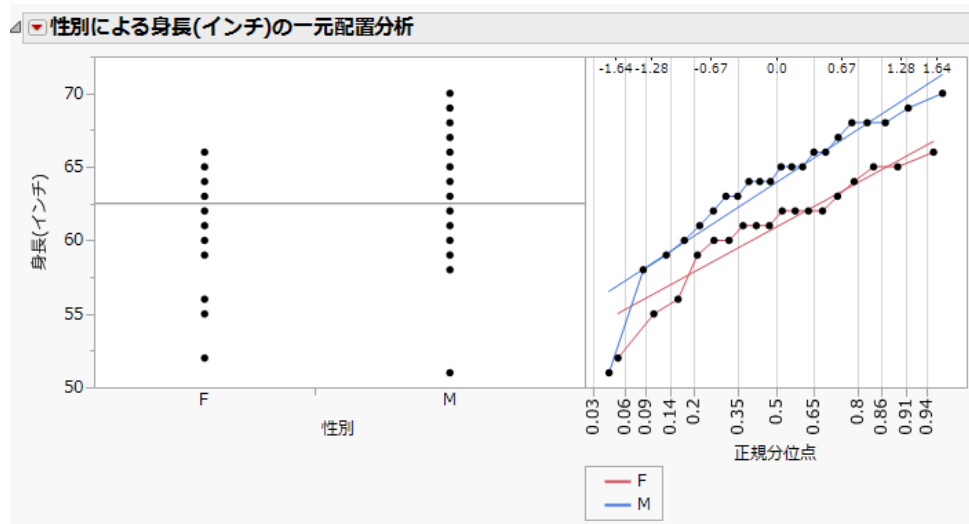


図 1.4: 正規性の点検：正規確率プロット

の結果である．ここでは， p 値 ($\text{Prob} \leq S$) と p 値 ($\text{Prob} \geq |S - \text{Mean}|$) の 2 個の p 値が記載される．前者は片側検定での p 値を表しており，後者は両側検定での p 値である．症例数が少ない医学系研究では正確な p 値 (exact p value) が好まれる傾向にあるため，Wilcoxon 検定の結果として (c) を採用することが推奨される．

JMP ではその他のノンパラメトリック法として，メディアン検定 (中央値の差の検定)，Van der Waerden の検定 (Wilcoxon 検定と同様に順位を用いた検定)，Kormogorov-Smirnov 検定 (経験累積分布関数の最大差を用いた検定) が実装されている．

覚書：正規性の評価

JMP では正規性の検定として Shapiro-Wilk 検定が実装されている²．一方で，グラフを見て正規性を点検することもできる．正規性を点検するためのグラフを正規確率プロット (normal probability plot) という．正規確率プロットの描き方を以下に示す．

正規確率プロットの描写方法




- [0] 上記「2 標本データの比較を実行するための流れ」を実行する．
- [1]  を左クリックする．
- [2] 「正規分位点プロット」→「実測値 - 分位点プロット」を選択する．

図 1.4 の右側は正規確率プロットの結果である．データ点が直線付近に布置していれば，正規性が満たされていると判断できる．他方，直線形状からずれている (歪んだ分布形状をもつ) 場合には非線形な形状を示す．本データではいずれの群もおおよそ直線状に布置しており，正規性が示唆される．

²Shapiro-Wilk 検定は次の手順で行える：「分析」→「一変量」で変数を選択 (今回の場合には「身長 (インチ)」を「Y, 目的変数」，「性別」を「By」に入れる)．解析結果のウィンドーの部分の「身長 (インチ)」での  を左クリックする．「連続分布のあてはめ」→「正規分布」を選択する．「正規のあてはめ」での  を左クリックし「適合度」を選択する．

1.1.3 対応のあるデータの解析

治療前後でのアウトカムの比較，あるいは同一被験者に異なる介入を行ったときの比較を行う場合（例えば，クロスオーバー試験），対応のあるデータとして取り扱うことになる．このとき用いられる方法が，パラメトリックの場合には対応のある t 検定，ノンパラメトリックの場合には Wilcoxon 符号付き順位検定が用いられる．

ここでは，サンプル・データのなかから，温度計のデータ Therm.jmp を用いる．このデータは，20 人の被験者に関して，口腔体温計と鼓膜体温計（耳式体温計）で計測した体温が記録されている．したがって，同一被験者から 2 種類の体温計での計測結果が得られている．「ヘルプ」→「サンプルデータライブラリ」で検索すると簡単に見つけることができる．ここで，「名前」は被験者の名前，「口部」は口腔体温計での計測結果，「鼓膜」は鼓膜体温計での計測結果，「差」は鼓膜体温計 口腔体温計の差を表している．

なお，独立 2 標本の場合にはアウトカムを表す変数とグループを表す変数を作成するのに対して，対応のあるデータの場合には，それぞれの変数が異なる介入でのアウトカムを表していることに注意されたい．

対応のある t 検定

対応のある t 検定の計算の流れを以下に示す．

対応のある t 検定

- [1] 「分析」→「発展的なモデル」→「対応のあるペア」を選択する（対応のあるペアのウィンドーが表示される）．
- [2] 「口部」及び「鼓膜」を「Y，対応のある応答」にドラックする．
- [2] 「OK」ボタンを押す．

このときの結果を図 1.5(A) に示す．ここに (a) は，X 軸が「(鼓膜 + 口部)/2」（各個体の体温の平均）を表しており，Y 軸が「鼓膜-口部」を表している．また，赤色の実線は「鼓膜-口部」の平均値であり，点線は 95%信頼区間である．すべてのデータ点の Y 軸の値が 0 を上回っていることから，鼓膜体温計の測定値のほうが口腔体温計の測定値に比べて高いことが分かる．(b) は「鼓膜-口部」の 95%信頼区間 ((a) のグラフの赤色の点線) の値であり，(c) は対応のある t 検定での p 値である． μ_D を体温計の測定差（鼓膜-口部）の平均値とすると， p 値の意味は次のとおりである：

- | | |
|-----------------|--|
| p 値 (Prob> t) | 両側検定 (対立仮説 $H_1: \mu_D = 0$) での p 値
(鼓膜体温計の測定値と口腔体温計の測定値には違いがある) |
| p 値 (Prob>t) | 片側検定 1 (対立仮説 $H_1: \mu_D < 0$) での p 値
(鼓膜体温計の測定値のほうが口腔体温計の測定値に比べて高い) |
| p 値 (Prob<t) | 片側検定 2 (対立仮説 $H_1: \mu_D > 0$) での p 値
(鼓膜体温計の測定値のほうが口腔体温計の測定値に比べて低い) |

両側対立仮説 (p 値 (Prob> |t|)) における解釈では， p 値が 0.0001 未満で有意差が認められることから，鼓膜体温計の測定値と口腔体温計の測定値には違いがあり，(a) のプロットより鼓膜体温計の測定値のほうが口腔体温計の測定値に比べて有意に高いことがわかる³．

³対応のある t 検定は，閾値平均値を 0 としたときの「差」に対する 1 標本 t 検定と同じである．実際に「分析」→「一変量の

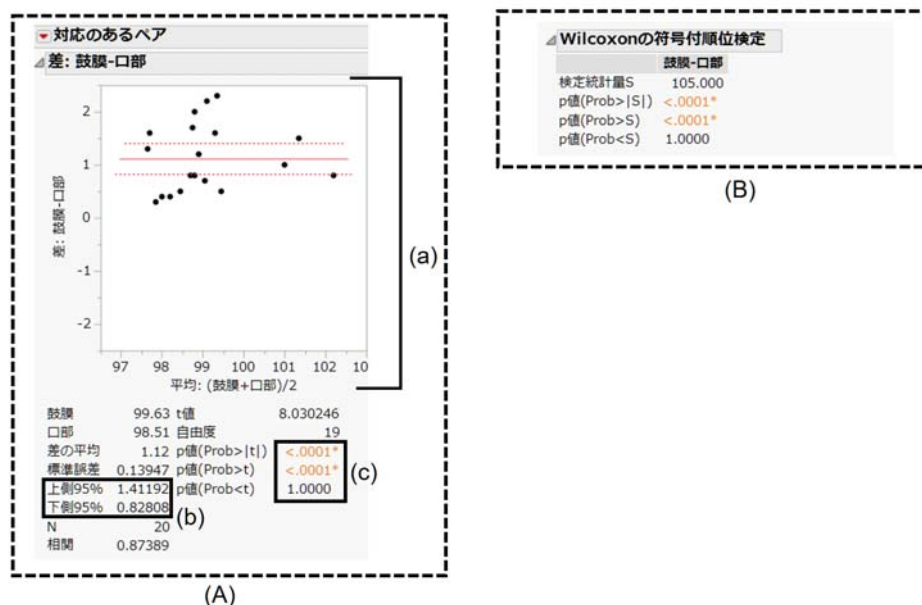



図 1.5: 対応のあるデータの解析

Wilcoxon 符号付き順位検定

Wilcoxon 符号付き順位検定 (ノンパラメトリック検定) の計算の流れを以下に示す。


Wilcoxon 符号付き順位検定の方法


- [0] 上記「対応のある t 検定」を実行する。
- [1]  を左クリックする。
- [2] 「Wilcoxon の符号付順位検定」を選択する。

上記の作業を行うと、図 1.5(B) が追加表示される。ここでも両側対立仮説値 (Prob> |S|) と片側対立仮説 (p 値 (Prob>t), p 値 (Prob<t)) の両方が表示される。いずれにしても、対応のある t 検定と同様の結果が得られている。

1.1.4 ベースラインがある場合・調整因子がある場合の比較：共分散分析

ベースラインがある場合・調整因子がある場合の比較として共分散分析が頻用される。共分散分析は 2 個の回帰直線 (処理群, 対照群) の傾きが等しいと仮定したもとで切片を比較する検定方法である。したがって、切片の比較に先立って、2 個の回帰直線の傾きが等しい (平行性) の仮定を評価しなければならない。このとき、JMP では共分散分析単独のメニューがない。

今回のデータでは 12 歳から 17 歳までの被験者が集計されている。成長するほど身長が高くなることが想定されるため、年齢で調整したうえで評価する。このときデータセットの「年齢」は順序尺度  になっ

分布」を選択し、「Y, 列」に「差」をドラッグして「OK」ボタンを押して解析結果を表示させる。次いで、「差」の  ボタンを左クリックして、「平均の検定」を選択したもとで「OK」ボタンを押すと対応のある t 検定と同じ値が得られる。すなわち、対応のある t 検定とは個体間の差をとったもとで閾値を 0 としたときの 1 標本 t 検定と同じになる

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	39.457057	4.812681	8.20	<.0001*
(年齢-13.975)*性別[F]	-0.600896	0.343014	-1.75	0.0883
年齢	1.6360307	0.343014	4.77	<.0001*
性別[F]	-1.227546	0.502444	-2.44	0.0196*

(I) 平行性の仮定に対する評価

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	39.416521	4.945343	7.97	<.0001*
年齢	1.6466542	0.352418	4.67	<.0001*
性別[F]	-1.214868	0.516246	-2.35	0.0240*

(II) 処理効果の評価

図 1.6: 共分散分析の結果

ているので、定量データ⁴に変換する必要がある⁴。

まず、平行性の仮定を検討する

共分散分析における平行性の仮定の検討

- [1] 「分析」→「モデルのあてはめ」を選択する（モデルのあてはめのウィンドーが表示される）。
- [2] 「身長（インチ）」を選択して「Y」をクリックする。
- [3] 「性別」及び「年齢」を選択して「モデル効果の構成」のなかの「追加」をクリック（「性別」を選択してから「CTRL」キーを押しながら「年齢」）する。
- [4] 「性別」及び「年齢」を選択して「モデル効果の構成」のなかの「交差」をクリックする。
- [5] 「強調点」を「最小レポート」にする。
- [6] 「実行」ボタンを押す。

このときの結果の一部を図 1.6(I) に示す。ここに A は平行性に関する検定の結果を表している。見方は、p 値が有意でなければ仮定が崩れていないことを示す（有意だと平行性の仮定が満たされないので ANCOVA を用いて評価できない）。今回の場合は、p 値が 0.0883 で有意でないため、ANCOVA による評価を行うことができる。

次いで、性別による差を評価する。

共分散分析における切片の差の検討

- [1] 「分析」→「モデルのあてはめ」を選択する（モデルのあてはめのウィンドーが表示される）。
- [2] 「身長（インチ）」を選択して「Y」をクリックする。
- [3] 「性別」及び「年齢」を選択して「モデル効果の構成」のなかの「追加」をクリック（「性別」を選択してから「CTRL」キーを押しながら「年齢」）する。
- [4] 「強調点」を「最小レポート」にする。
- [5] 「実行」ボタンを押す。

このときの結果の一部を図 1.6(II) に示す。ここに A は切片の差に関する検定の結果を表している。p 値が 0.0240 なので有意水準 0.05 のもとで有意である。したがって、年齢によって調整した場合においても身長に性差が認められる。

⁴データの型変換はデータウィンドー左中央付近の「年齢」を右クリックして「列情報...」から変換できる

1.2 カテゴリカルデータの評価

1.2.1 カイ 2 乗検定・Fisher の正確検定

ここでは、車に関する調査をまとめたサンプルデータ「Car Poll.jmp」を利用する。本ファイルは、「ヘルプ」→「サンプルデータライブラリ」で検索すると簡単に見つけることができる。これは、所有する自動車について、調査対象者の性別、既婚/未婚、年齢、及び自動車の（購入した自動車の）生産国、サイズ（大型、中型、小型）、タイプ（ファミリー、スポーツ、ワーク）を調査したデータである。

ここでは、自動車のサイズと生産国の関係について解析する（生産国によって自動車のサイズに違いがあるか否か）。カイ 2 乗検定の実行方法を以下に示す。

カイ 2 乗検定の実行

- [1] 「分析」→「二変量の関係」を選択する（二変量の関係のウィンドーが表示される）。
- [2] 「サイズ」を選択して「Y, 目的変数」をクリックする。
- [3] 「生産国」を選択して「X, 説明変数」をクリックする。
- [4] 「OK」ボタンを押す。

図 1.7(I)~(III) はこのときの出力結果である。モザイクプロット (図 1.7(I)) とは、クロス集計表の列方向の割合 (説明変数 X) の比率を長方形の横幅 (今回の場合は生産国)、行方向の割合 (従属変数 Y) の比率を長方形の縦幅 (今回の場合はサイズ) で表現したグラフィカル手法である。その結果、ヨーロッパ車が最も少なく、日本車が最も多いことが示唆された。また、日本車には大型車が非常に少ない一方で小型車が最も多かった。さらに、米国車は大型車が多いものの、小型車が少ないことが示唆された。


図 1.7(II) はクロス集計表である。それぞれのパーセント点は、

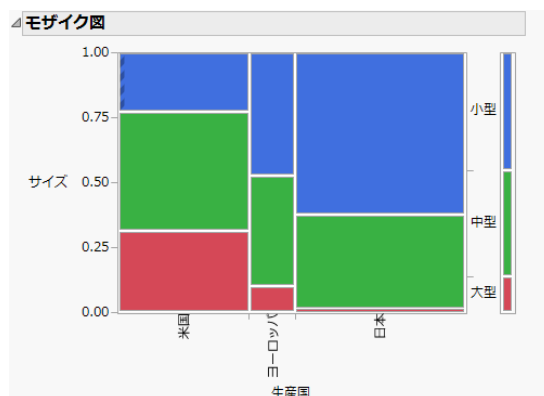
パーセント	確率の意味	例示
全体%	$\Pr(\text{生産国 (X)} \cap \text{サイズ (Y)})$	米国製でかつ大型車は 11.88%
列%	$\Pr(\text{生産国 (X)} \mid \text{サイズ (Y)})$	大型車のうち米国製は 85.71%
行%	$\Pr(\text{サイズ (Y)} \mid \text{生産国 (X)})$	米国製のうち大型車は 31.30%

と解釈できる。

図 1.7(III) はクロス集計表での仮説検定の結果である⁵。JMP では尤度比と Pearson の 2 種類の結果が表示される。尤度比では帰無仮説からの乖離を尤度 (likelihood) という概念を用いて検定しており、より複雑なクロス集計表を解析する対数線形モデル (log-linear model) の特別な場合として計算している。他方、Pearson は通常のカイ 2 乗検定の結果である。いずれの場合においても帰無分布は同じであるものの、検定統計量は若干異なる。本データではいずれの検定においても p 値が < 0.0001 であり高度に有意である。

2×2 分割表では Fisher の正確検定が自動で表示されるが、 $m \times k$ 分割表では表示されない。これは、 $m \times k$ 分割表での正確検定 (拡張 Fisher の正確検定) は JMP Pro のみに実装されているためである。計算方法を以下に示す。

⁵ 2×2 分割表では Fisher の正確検定が自動で表示される。また、相対リスク、オッズ比、割合の 2 標本検定 (母比率の差の検定) が  のメニューから選択できる



(I) モザイクプロット (モザイク図)

分割表				
国別	サイズ			
	大型	中型	小型	合計
	度数			
	全体%			
	列%			
	行%			
米国	36	53	26	115
	11.88	17.49	8.58	37.95
	85.71	42.74	18.98	
	31.30	46.09	22.61	
ヨーロッパ	4	17	19	40
	1.32	5.61	6.27	13.20
	9.52	13.71	13.87	
	10.00	42.50	47.50	
日本	2	54	92	148
	0.66	17.82	30.36	48.84
	4.76	43.55	67.15	
	1.35	36.49	62.16	
合計	42	124	137	303
	13.86	40.92	45.21	

(II) クロス集計表

検定				
	N	自由度	(-1)*対数尤度	R2乗(U)
	303	4	36.309616	0.1200
検定	カイ2乗	p値(Prob>ChiSq)		
尤度比	72.619	<.0001*		
Pearson	66.313	<.0001*		


(III) カイ 2 乗検定

Fisherの正確検定	両側p値
表の確率(P) (表の生起確率 $\leq p$)	
2.01e-19	<.0001*

(IV) Fisher の正確検定

図 1.7: クロス集計表に対する検定の結果


Fisher の正確検定の実行

- [0] 上述の「カイ 2 乗検定の実行」を実行する
- [1]  を左クリックする。
- [2] 「正確検定」→「Fisher の正確検定」を選択する。

このときの結果を図 1.7(IV) に示す。この場合においても、p 値が < 0.0001 であり高度に有意である。

1.2.2 傾向性の検定 (Cochran-Armitage 検定)

先ほどと同様に、車に関する調査をまとめたサンプルデータ「Car Poll.jmp」を利用する。ここでは、自動車のサイズと性別の関係について解析する (女性はより小さな自動車を所有している・男性はより大きな自動車を所有している)。Cochran-Armitage 検定の実行方法を以下に示す。

- [1] 「分析」→「二変量の関係」を選択する（二変量の関係のウィンドーが表示される）。
- [2] 「サイズ」を選択して「Y, 目的変数」をクリックする。
- [3] 「性別」を選択して「X, 説明変数」をクリックする。
- [4] 「OK」ボタンを押す。
- [5]  を左クリックする。
- [6] 「Cochran-Armitage の傾向検定」を選択する。

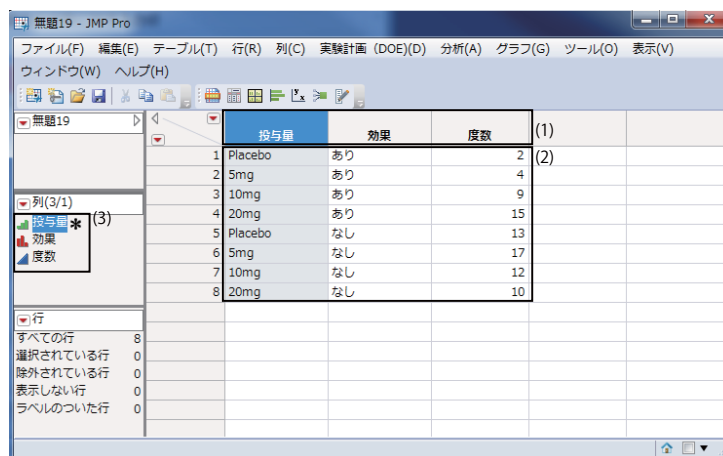
このときの結果を図 1.8(I) に示す。p 値が 0.7094 で有意でなかった。したがって、女性はより小さな自動車所有している（男性はより大きな自動車所有している）という傾向変化は認められなかった。

1.2.3 覚書：度数を入力した場合のクロス集計表の構成

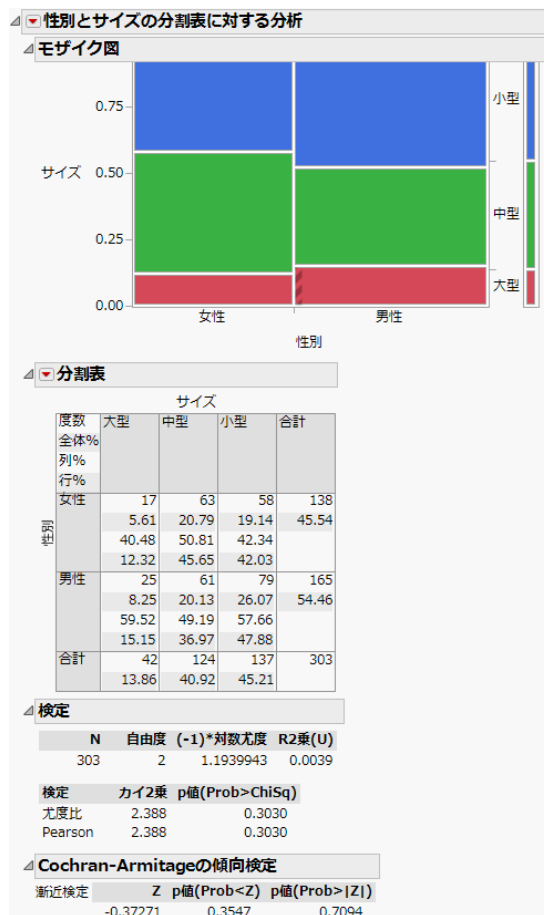
JMP では度数を手入力したデータに基づいてクロス集計表を作ることができる。下表はある薬剤を 82 人に投与したときの用量-反応関係を評価した試験の結果である。

		投与量				
		Placebo	5mg	10mg	20mg	合計
効果	あり	2	4	9	15	30
	なし	13	17	12	10	52

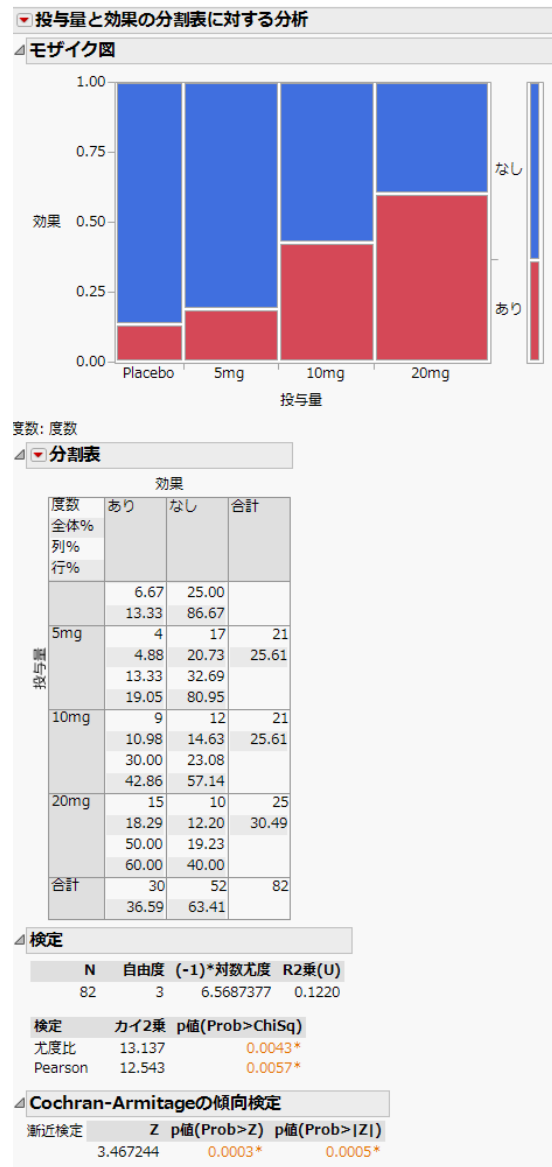
この度数分布表をデータとして手入力したうえで Cochran-Armitage 検定を行う。「ファイル」→「新規作成」→「データテーブル (D)」を選択する。表示されたデータテーブルのウィンドー下図の順番で入力する。



- (1) 変数の作成及び変数名を入力するために灰色の部分ダブルクリックすると変数が定義できる。そこに左から「投与量」、「効果」、「度数」とそれぞれ入力する。
- (2) それぞれの変数に対して上図のようにデータを入力する。
- (3) 各変数を「右クリック」→「列情報」を選択し、ウィンドーを表示させる。それぞれの変数に対しては、以下のように入力する。なお、変数の尺度のみを変換する場合には変数名左のアイコンを左クリックすればよい。



(I) 自動車購入データの結果




(II) 用量-反応データの結果

図 1.8: Cochran-Armitage 検定の結果

- ・「投与量」:「尺度」を「順序尺度」に変換する．そして、値の順序を「Placebo」「5mg」「10mg」、
「20mg」にする．なお順番はカテゴリを選択して「上へ移動」あるいは「下へ移動」で実行する．
- ・「効果」:「尺度」を「名義尺度」に変換する．
- ・「度数」:「尺度」を「連続尺度」に変換する．

上記の手順で作成したもとでクロス集計表のもとの Cochran-Armitage 検定を実行する．手順は以下のとおりである．

- [1] 「分析」→「二変量の関係」を選択する(二変量の関係のウィンドーが表示される)。
- [2] 「効果」を選択して「Y, 目的変数」をクリックする。
- [3] 「投与量」を選択して「X, 説明変数」をクリックする。
- [4] 「度数」を選択して「度数」をクリックする。
- [5] 「OK」ボタンを押す。
- [6]  を左クリックする。
- [7] 「Cochran-Armitage の傾向検定」を選択する。

このときの結果を図 1.8(II) に示す。p 値が 0.0005 で有意だった。また、モザイクプロットより、用量が増加するほど効果の上昇が認められた。

1.2.4 対応のあるクロス集計表の解析

治療前後でのアウトカムの変化、あるいは 2 種類の検査での所見の有無など、同一被験者からカテゴリカル変数としてアウトカムが得られる場合を「対応のあるデータ」という。ここでは、ある治療法が症状を改善するか否かを調べるために、治療前の症状の有無、治療後の症状の有無を集計した結果を用いる。このときのクロス集計表は

		治療後		
		症状あり	症状なし	合計
治療前	症状あり	35	12	47
	症状なし	4	16	20
合計		39	28	67

のように与えられる。通常のクロス集計表では、列方向(縦方向)に介入(要因)をおき、行方向(横方向)にアウトカム(結果)をおいて構成される。これに対して、対応のあるクロス集計表では、列方向(治療前)、行方向(治療後)ともにアウトカムを表している。例えば、左上のセルの 35 は「(治療前は症状あり)かつ(治療後は症状あり)」の被験者数を表している。このクロス集計表では、「(治療前は症状あり)かつ(治療後は症状なし)」の被験者数と「(治療前は症状なし)かつ(治療後は症状あり)」の被験者数を比較することになり、このような検定を McNemar 検定という。

ここでは、前述の度数を入力した場合のクロス集計表の作り方を参考に行う。「ファイル」→「新規作成」→「データテーブル(D)」を選択する。表示されたデータテーブルのウィンドー下図の順番で入力する。

- (1) 変数の作成及び変数名を入力するために灰色の部分ダブルクリックすると変数が定義できる。そこに左から「治療前」、「治療後」、「人数」とそれぞれ入力する。
- (2) 各変数を「右クリック」→「列情報」を選択し、ウィンドーを表示させる。それぞれの変数に対しては、以下のように入力する。なお、変数の尺度のみを変換する場合には変数名左のアイコンを左クリックすればよい。

	治療前	治療後	人数
1 あり	あり	あり	35
2 あり	なし	あり	12
3 なし	あり	なし	4
4 なし	なし	なし	16

図 1.9: 対応のあるクロス集計表のデータテーブル


- ・「治療前」:「尺度」を「名義尺度」に変換する。
- ・「治療後」:「尺度」を「名義尺度」に変換する。
- ・「人数」:「尺度」を「連続尺度」に変換する。

(2) それぞれの変数に対してデータを入力する。

これにより、1.2.4 のようなデータテーブルが作成される。

上記の手順で作成したもとでクロス集計表のもとの McNemar 検定を実行する。手順は以下のとおりである。

— 度数データに基づく McNemar 検定の実行 —

- [1] 「分析」→「二変量の関係」を選択する(二変量の関係のウィンドーが表示される)。
- [2] 「治療前」を選択して「Y, 目的変数」をクリックする。
- [3] 「治療後」を選択して「X, 説明変数」をクリックする。
- [4] 「人数」を選択して「度数」をクリックする。
- [5] 「OK」ボタンを押す。
- [6]  を左クリックする。
- [7] 「一致性の統計量」を選択する。

このときの結果を図 1.2.4 に示す。(a) は対応のあるクロス集計表でのモザイクプロットである。モザイクプロットでの四角形の面積は被験者の割合に対応している。「(治療前は症状あり) かつ (治療後は症状あり)」の被験者数が多いことが伺える。「(治療前は症状あり) かつ (治療後は症状なし)」の四角形の面積のほうが、「(治療前は症状なし) かつ (治療後は症状あり)」よりも小さいことから、治療後に症状が改善した割合が高いように思われる。(b) はこのときの対応のあるクロス集計表である。ここで、各セルの 3 列目と 4 列目は、それぞれ列パーセント点(縦方向に割合を計算したもの)と行パーセント点(行方向に割合を計算したもの)になるが、対応のあるクロス集計表では、全体パーセント(2 列目)のみに意味があるので注意されたい。

検定(×で表した部分)については、独立 2 標本(すなわち対応のないクロス集計表)で用いられる検定の結果なので、この結果を解釈してはならず、Bowker の検定の結果(c)のみを解釈しなければならない。

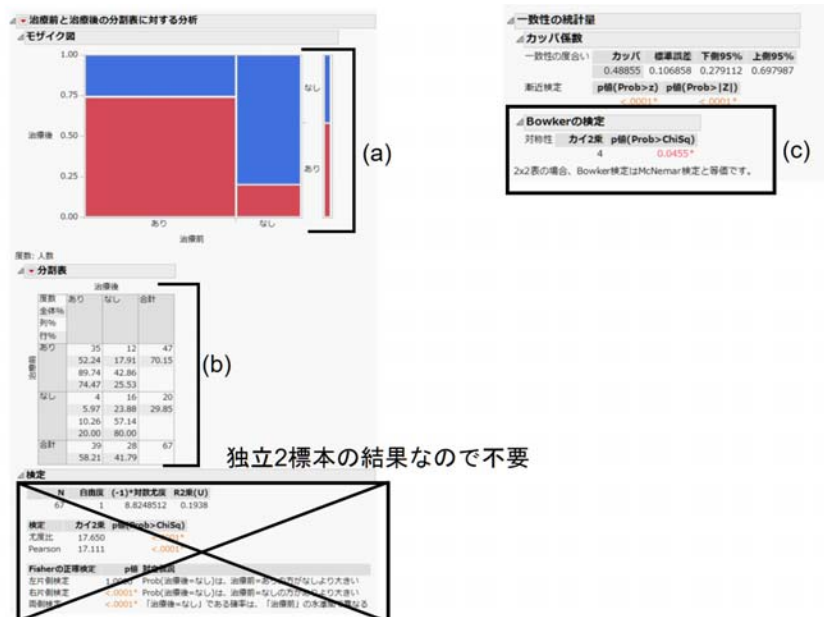



図 1.10: McNemar 検定の結果

Bowker の検定の結果は、 2×2 クロス集計表の場合には、McNemar 検定と一致する (McNemar 検定は 2×2 分割表の対称性を検定する方法である)。その結果、p 値が 0.0455 であることから、治療前後で症状の有無に違いが認められた。

1.2.5 調整因子がある場合の検定 (Cochran-Mantel-Haenszel 検定)

カイ 2 乗検定では、自動車のサイズと生産国の関係について解析したが、男性ほど大型の自動車を好み、女性ほど小型の自動車を好むかもしれない。そのため、性別を交絡因子 (調整因子) としたときの Cochran-Mantel-Haenszel 検定 (Mantel-Haenszel 検定) を行う。

Cochran-Mantel-Haenszel 検定の実行

- [1] 「分析」→「二変量の関係」を選択する (二変量の関係のウィンドーが表示される)。
- [2] 「サイズ」を選択して「Y, 目的変数」をクリックする。
- [3] 「生産国」を選択して「X, 説明変数」をクリックする。
- [4] 「OK」ボタンを押す。
- [5]  を左クリックする。
- [6] 「Cochran-Mantel-Haenszel 検定」を選択する (交絡因子を選択するウィンドーが表示される)。
- [7] 「性別」を選択する。
- [8] 「OK」ボタンを押す。

このときの結果を図 1.11 に示す。Cochran-Mantel-Haenszel 検定の結果は A の部分を見ればよい。p 値

Cochran-Mantel-Haenszel検定			
層別変数: 性別			
Cochran-Mantel-Haenszel検定	カイ2乗	自由度	p値(Prob>ChiSq)
スコアの相関	62.4205	1	<.0001*
X間でのスコア比較	63.3521	2	<.0001*
Y間でのスコア比較	64.6912	2	A <.0001*
カテゴリの一般連関	65.8979	4	<.0001*

図 1.11: Cochran-Mantel-Haenszel 検定の結果

が < 0.0001 で高度に有意だった。したがって、性別で調整した場合においても自動車のサイズと生産国には関連性が認められた。

第2章 分散分析

2.1 多標本の比較

ここでは、疼痛に対する3種類の除痛薬(A,B,C)を比較するサンプルデータ「Analgesics.jmp」を利用する。本ファイルは、「ヘルプ」→「サンプルデータライブラリ」で検索すると簡単に見つけることができる。これは、3種類の薬(A,B,C)の効果に対して、痛みの程度および被験者の性別がとられている。多標本比較(一元配置の分散分析、パラメトリックな多重比較法、Kruskal-Wallis検定、ノンパラメトリックな多重比較法)においても、2標本比較と同様の手順で実行できる。

多標本データの比較を実行するための流れ

- [1] 「分析」→「二変量の関係」を選択する(二変量の関係ウィンドーが表示される)。
- [2] 「列の選択」から「痛みの程度」を「Y, 目的変数」「薬」を「X, 説明変数」にドラックする(「痛みの程度」を選択した後で「Y, 目的変数」を左クリック「薬」を選択した後で「X, 説明変数」を左クリックでも同様である)。
- [3] 「OK」ボタンを押す。

このときの結果を2.1に示す。ここに水平線は全体平均を表している。薬剤Aは全体平均よりも下側に布置しており、薬剤Cは全体平均よりも上側に布置していることが示唆される。

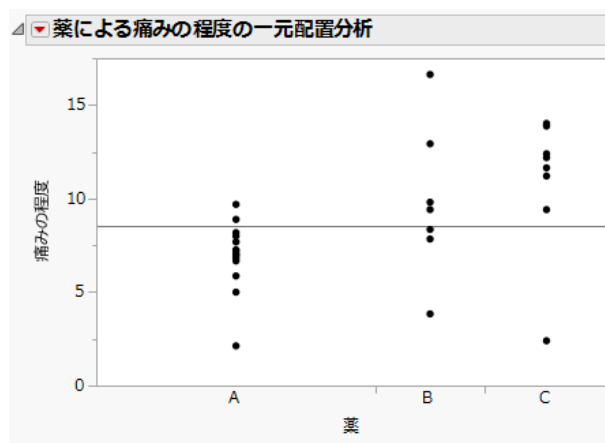
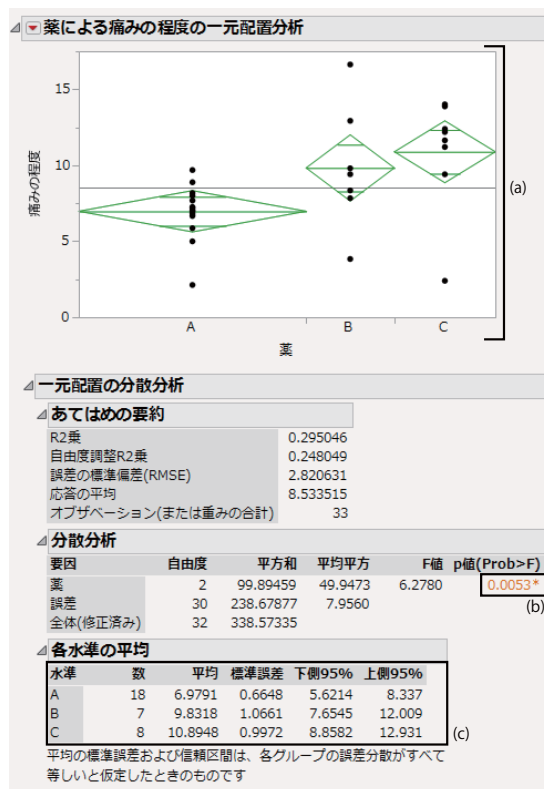
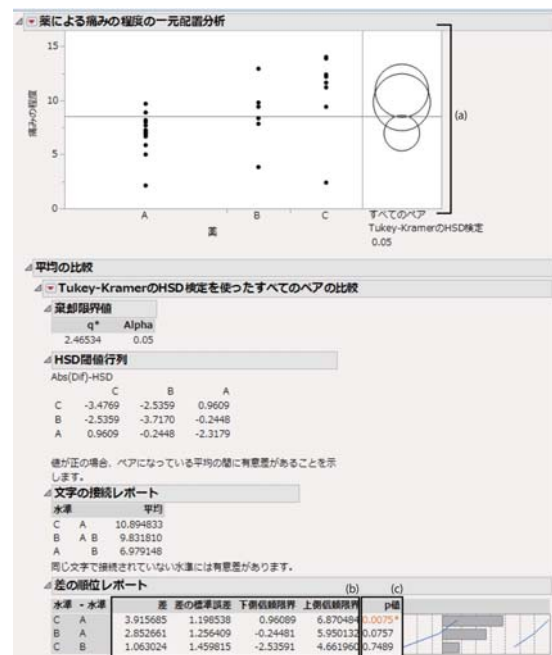


図 2.1: 多標本比較での一次元散布図



(I) 一元配置の分散分析の結果



(II) Tukey の HSD 検定 (多重比較) の結果


図 2.2: パラメトリック (正規分布を仮定したもとでの) 多標本比較

2.1.1 一元配置の分散分析

一元配置の分散分析では、3種類の薬 (A,B,C) の平均効果に違いがあるか否かを評価する。一元配置の分散分析の流れを以下に示す (2 標本 t 検定と同じ流れで実行できる)。

一元配置の分散分析の方法

[0] 上記「多標本データの比較を実行するための流れ」を実行する。

[1]  を左クリックする。

[2] 「平均/ANOVA」を選択する。

このときの結果を 2.2(I) に示す。(a) は 2 標本 t 検定と同様のグラフである。緑色の中央の横線は標本平均を表しており、ひし形の縦方向の頂点は母平均に対する 95% 信頼区間を表しており、横方向の大きさは標本サイズに対応している。すなわち、ひし形の横幅が大きいほど標本サイズが大きいということを意味する。

(b) は一元配置の分散分析の結果である。p 値が 0.0053 であることから、有意水準 $\alpha = 0.05$ のもとで有意である。したがって、薬によって除痛効果に違いがあることが認められた。

(c) は要因 (薬剤) の各水準 (A,B,C) における標本平均および 95% 信頼区間を表している。すなわち、これらは (a) のひし形の縦方向の頂点に対応している。


2.1.2 パラメトリックな多重比較法

前節の一元配置の分散分析では薬によって除痛効果に違いがあることが認められた。ただし、そのパターンは

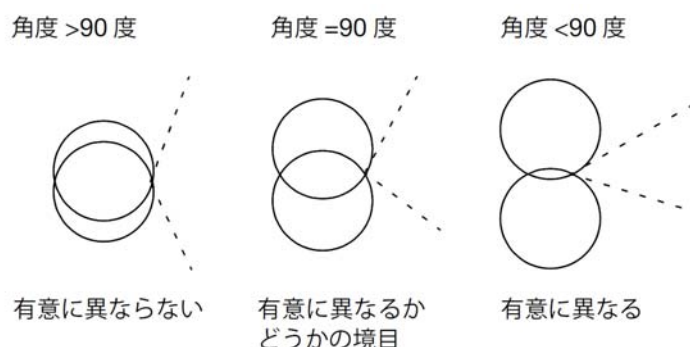
$$A = B \neq C, A \neq B = C, A = C \neq B, A \neq B \neq C$$

のいずれかである。どの状況であるかを判断するために多重比較を行うことが多い。JMP では多標本比較の方法として「各ペア, Student の t 検定」、「すべてのペア, Tukey の HSD 検定」、「最適値との比較 (Hsu の MCB)」、「コントロール群との比較 (Dunnett)」がある。「各ペア, Student の t 検定」は多重比較を行わずにペアワイズに 2 標本 t 検定を行う場合である。「すべてのペア, Tukey の HSD 検定」はペアワイズに多重比較のもとで母平均を比較する場合である。「最適値との比較 (Hsu の MCB)」は任意の基準値 (最適値) よりも各水準の母平均が高い (あるいは低い) か否かを 1 標本検定のもとで明らかにする方法である。「コントロール群との比較 (Dunnett)」は、水準の一つがコントロール群であり、そのコントロール群との多重比較を行う場合に用いる。今回は、コントロール群がない状況でのペアワイズ比較なので「すべてのペア, Tukey の HSD 検定」を利用する。

Tukey の HSD 検定の方法

- [0] 上記「多標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「平均の比較」→「すべてのペア, Tukey の HSD 検定」を選択する。

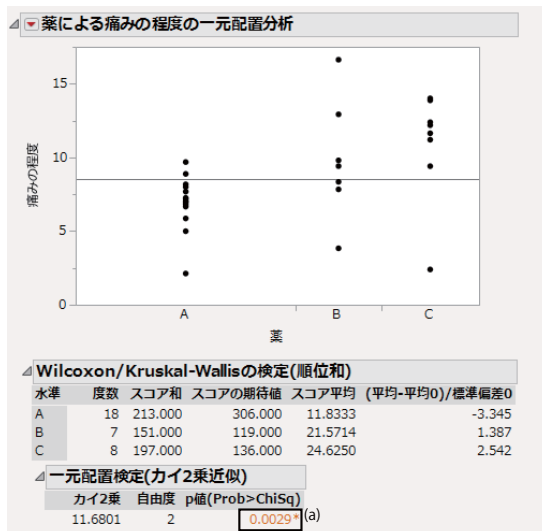
このときの結果を 2.2(I) に示す。(a) は比較円と呼ばれるものであり、Tukey の多重比較検定をグラフィカルに実行できる。その解釈は 2 個の円の接点からの角度で次のように解釈される。



(b) は母平均の差に対する 95%信頼区間を表しており、(c) がこのときの p 値である。その結果、薬剤 A と C において有意差が認められ、薬剤 C のほうが薬剤 A に比べて痛みの程度が高かった。

2.1.3 Kruskal-Wallis 検定

ここでは、一元配置の分散分析のノンパラメトリック版である Kruskal-Wallis 検定を述べる。Kruskal-Wallis 検定の流れを以下に示す (Wilcoxon 検定と同じ流れで実行できる)。




(I) Kruskal-Wallis 検定の結果



(II) Steel-Dwass 検定 (多重比較) の結果

図 2.3: ノンパラメトリック (正規分布を仮定しないもとでの) 多標本比較

Kruskal-Wallis 検定の方法

- [0] 上記「多標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「ノンパラメトリック」→「Wilcoxon 検定」を選択する。

このときの結果を 2.3(I) に示す。(a) は Kruskal-Wallis 検定における p 値である。p 値は 0.0029 なので有意水準 0.05 のもとで有意である。すなわち、3 種類の薬剤で痛みの程度に違いがあることが認められた。

2.1.4 ノンパラメトリックな多重比較法

JMP ではパラメトリックな多重比較法に対応する形式でノンパラメトリックな多重比較法が実装されている。以下はパラメトリックな多重比較法とノンパラメトリックな多重比較法の対応表である。

パラメトリック	ノンパラメトリック	略説
Student の t 検定	Wilcoxon 検定	多重比較調整無しでの比較
Tukey の HSD 検定	Steel-Dwass 検定	ペアワイズな比較
Dunnett の多重比較検定	Steel 検定	コントロール群との比較
-	Dunn 検定 (全てのペア)	Bonfferoni 型の多重比較
-	Dunn 検定 (コントロール群との比較)	Bonfferoni 型の多重比較
Hsu の MCB	-	任意の基準値 (最適値) との比較

ここで「-」は対応する検定が実装されていないことを表す。また、Dunn 検定はすべての順位を用いた方法であり (Kruskal-Wallis 検定と同様の順位和をとったうえで検定統計量を構成する。したがって、Wilcoxon 検定の p 値とは異なる)、多重比較調整は Bonfferoni 型で実施される。

ここでは、すべてのペアで比較するために Steel-Dwass 検定を利用する。

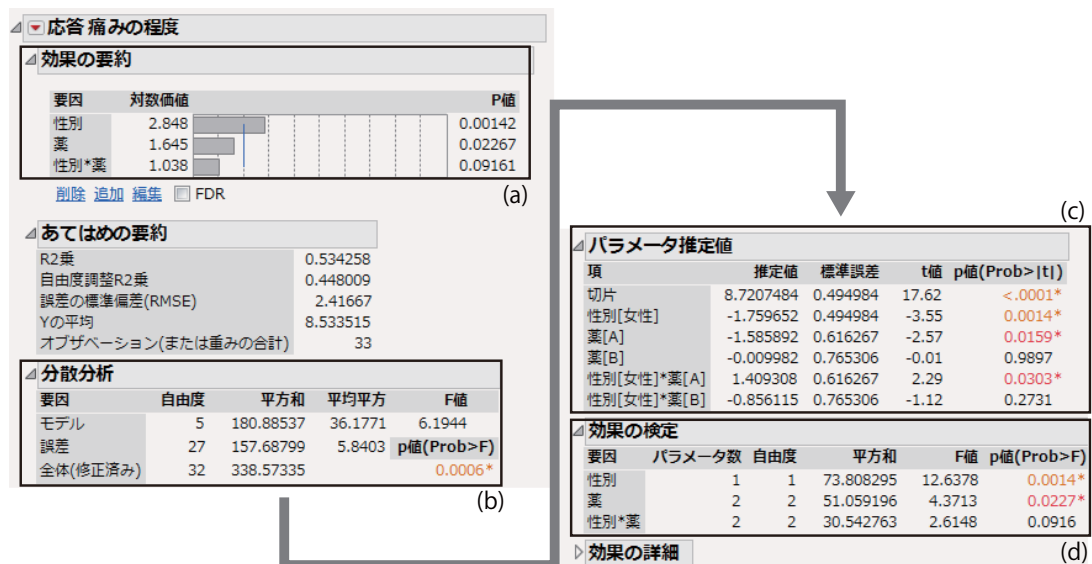


図 2.4: 二元配置の分散分析の結果

Steel-Dwass 検定の方法


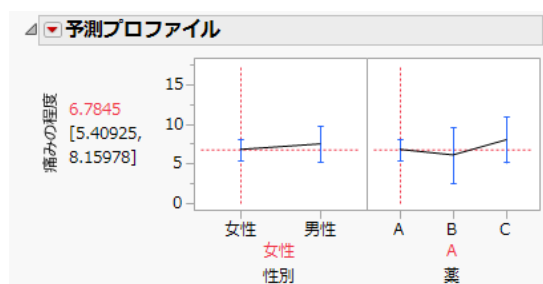
- [0] 上記「多標本データの比較を実行するための流れ」を実行する。
- [1]  を左クリックする。
- [2] 「ノンパラメトリック」→「ノンパラメトリックな多重比較」→「すべてのペア Steel-Dwass 検定」を選択する。

図 2.3(II) は、Steel-Dwass 検定の結果である。(a) はそれぞれの多重比較での p 値であり、薬剤 A 対 薬剤 C において有意だった。(b) は群 1 と群 2 に含まれる個体のすべての組み合わせで差をとり、その中央値をとった値 (Hodges-Lehmann 推定量) であり、群間の相対的な位置の差を表している。その結果、薬剤 C - 薬剤 A の位置が 4.654333 なので、薬剤 C のほうが痛みの程度の値が相対的に高いことがわかった。

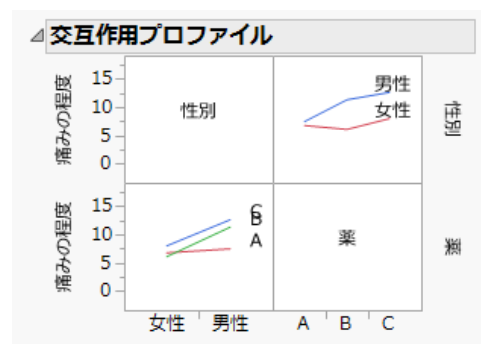
2.2 多元配置の分散分析

ここでは、前節「多標本の比較」と同様に、疼痛に対する 3 種類の除痛薬 (A,B,C) を比較するサンプルデータ「Analgesics.jmp」を利用する。本ファイルは、「ヘルプ」→「サンプルデータライブラリ」で検索すると簡単に見つけることができる。これは、3 種類の薬 (A,B,C) の効果に対して、痛みの程度および被験者の性別がとられている。前節では、3 種類の薬 (A,B,C) の効果のみを因子としていたが、今回は性別も因子に加えた、二元配置の分散分析を考える。二元配置の分散分析では、交互作用を検討することが考えられる。交互作用とは、例えば「薬剤 A を男性に投与すると女性に投与するよりも有効である」というような相乗効果が認められる状況などに生じる効果である。今回は、

$$y_{jki} = \mu + \eta_j + \gamma_k + (\eta\gamma)_{jk} + \epsilon_{jki}$$



(I) プロファイル・プロット



(II) 交互作用のグラフ

図 2.5: 主効果及び交互作用のグラフ

というモデルを考える．ここに μ は全平均， $\eta_j (j = 1, 2)$ は性別による効果， $\gamma_k (k = 1, 2, 3)$ は薬剤による効果， ϵ_{jki} は誤差である．

二元配置の分散分析の方法を以下に示す．

二元配置の分散分析

- [1] 「分析」→「モデルのあてはめ」を選択する（モデルのあてはめのウィンドーが表示される）．
- [2] 「痛みの程度」を選択して「Y」をクリックする．
- [3] 「性別」及び「薬」を選択して「モデル効果の構成」のなかの「追加」をクリック（「性別」を選択してから「CTRL」キーを押しながら「薬」）する．
- [4] 「性別」及び「薬」を選択して「モデル効果の構成」のなかの「交差」をクリックする．
- [5] 「強調点」を「最小レポート」にする．
- [6] 「実行」ボタンを押す．

二元配置の分散分析の結果を図 2.4 に示す．JMP では分散分析単独での解析は出来ず，回帰分析と出力が同じになる（分散分析は回帰分析の一部として捉えることが出来るため）．(a) は分散分析の p 値および対数値を表している．対数値とは $-\log_{10}(\text{p 値})$ で表される指標であり，大きいほど影響（価値）が高いと解釈される（医学統計学ではあまり用いないので注意されたい）．他方，(d) は分散分析表で結果を表している．(a) の p 値と (d) の p 値は同じ結果である．その結果，「性別」及び「薬」の主効果において有意だった¹．

分散分析において，「性別」及び「薬」の主効果において有意な結果が得られた．次いで，性別毎に薬剤が疼痛に対してどのような影響を及ぼしているかをグラフィカルに評価する方法を述べる．プロファイル・プロットを表示する手順は次のとおりである．

¹(b) の分散分析表は回帰モデルの有意性（回帰分析に意味があるか評価）を検定するための分散分析表である．ここでの回帰モデルは「性別」「薬」に対してダミー変数を設定（例えば性別であれば 1 を「女性」，0 を「男性」）としたもとで重回帰分析を実施した場合である．また，(c) は重回帰分析のパラメーター推定値及び p 値を表している．ここでは分散分析を実施しているので，割愛する．

—— プロファイル・プロットの表示 ——


- [0] 上記「二元配置の分散分析」を実行する．
- [1]  を左クリックする．
- [2] 「因子プロファイル」→「プロファイル」を選択する．

図 2.5(I) は主効果に対するプロファイル・プロットである．ここで，横軸は各因子の水準，縦軸が応答（疼痛）平均を表している．また，X 軸の軸名の上の赤は基準（赤色の点線）を表している．すなわち，「性別」では「女性」を基準に赤色の点線が描写されており，「薬」では「薬 A」を基準に赤色の点線が描写されている．その結果，「女性」よりも「男性」のほうが痛みの程度が高く，「薬 C」の痛みの程度が最も高いことが示唆された．また，一元配置の分散分析では「薬 A」の痛みの程度が最も低かったが，性別で調整された二元配置の分散分析では「薬 B」のほうが若干低かった．

次いで，交互作用のプロファイルを省察する方法について述べる．交互作用のグラフを表示する手順は次のとおりである．

—— 交互作用のグラフ表示 ——


- [0] 上記「二元配置の分散分析」を実行する．
- [1]  を左クリックする．
- [2] 「因子プロファイル」→「交互作用プロット」を選択する．

図 2.5(II) は交互作用プロットである．その結果，「薬剤 A」では痛みの程度に性差が示唆されなかったものの，「薬剤 B」及び「薬剤 C」では「女性」に比べて「男性」のほうが痛みの程度が高かった．二元配置の分散分析では有意差は認められなかったものの，「薬剤 B」及び「薬剤 C」では有効性に性差が示唆された．

第3章 一般化線形モデル

医学研究において用いられる重回帰分析，ロジスティック回帰分析といった分析方法は一般化線形モデルの枠組みで捉えることができる．本節では，一般化線形モデルの諸型の利用方法について述べる．

3.1 重回帰分析

ここでは，運動効果についての調査によって測定されたサンプルデータ「Fitness.jmp」を利用する．本ファイルは，「ヘルプ」→「サンプルデータライブラリ」で検索すると簡単に見つけることができる．これは，「名前」「性別」「年齢」「体重」といった被験者属性，「走行時間」「走行時の脈拍」「休息時の脈拍」「最大脈拍」といった運動負荷のもとでとられた測度，および酸素摂取量がとられている．ここでの目標は，「体重」「走行時間」「走行時の脈拍」「休息時の脈拍」及び「最大脈拍」から「酸素摂取量」を予測するための回帰モデルを推定することにある．

今回の事例では，応答（従属変数）に対して5個の説明変数（独立変数）を用いて回帰モデルを推定することになるが，説明変数が複数個の場合の回帰分析を，重回帰分析という．重回帰分析では多重共線性などの観点から不必要な説明変数を削除する，すなわち，ステップワイズ法を利用することが少なくない．そのため，本解析においてもステップワイズ法を利用する．ステップワイズ法には3種類（変数増加法，変数減少法，変数増減法）があるが，本解析では変数減少法について解説する．

後退ステップワイズ法による重回帰分析の手順を以下に示す．



図 3.1: 重回帰分析におけるステップワイズ法のメニュー

SSE	DfE	RMSE	R2乗	自由度調整R2乗	Cp	p	AICc	BIC
161.77233	27	2.4477679	0.8100	0.7889	2.8284105	4	151.5925	156.3624

現在の推定値

ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	"p値(Prob>F)"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	80.9007896	1	0	0.000	1
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	体重	0	1	4.989591	0.827	0.37137
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	走行時間	-2.9701867	1	443.2028	73.971	3.25e-9
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	走行時の脈拍	-0.3751142	1	55.14175	9.203	0.00529
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	休息時の脈拍	0	1	0.350744	0.056	0.81999
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	最大脈拍	0.35421891	1	41.34703	6.901	0.01403

ステップ履歴

ステップ	パラメータ	アクション	"p値"	逐次平方和	R2乗	Cp	p	AICc	BIC
1	体重	追加	0.3817	22.55181	0.0265	105.33	2	196.73	200.143
2	走行時間	追加	0.0000	611.6719	0.7449	9.6712	3	157.858	162.059
3	走行時の脈拍	追加	0.1777	14.38346	0.7618	9.3748	4	158.595	163.365
4	休息時の脈拍	追加	0.8313	0.360424	0.7623	11.317	5	161.64	166.744
5	最大脈拍	追加	0.0121	45.83023	0.8161	6	157.051	162.22	
6	休息時の脈拍	削除	0.8600	0.199033	0.8158	4.0318	5	153.721	158.825
7	体重	削除	0.3714	4.989591	0.8100	2.8284	4	151.592	156.362
8	最大脈拍	削除	0.0140	41.34703	0.7614	7.4298	3	155.787	159.984
9	走行時の脈拍	削除	0.1567	15.36208	0.7434	7.8825	2	155.397	158.81
10	走行時間	削除	0.0000	632.9001	0.0000	106.93	1	195.102	197.541
11	最適	特定	.	.	0.8100	2.8284	4	151.592	156.362

図 3.2: 重回帰分析におけるステップワイズ法の実行結果

後退ステップワイズ (変数減少法) による重回帰分析の実行 (停止ルールに BIC を用いた場合)

- [1] 「分析」→「モデルのあてはめ」を選択する (モデルのあてはめのウィンドーが表示される) .
- [2] 「酸素摂取量」を選択して「Y」をクリックする .
- [3] 「体重」、「走行時間」、「走行時の脈拍」、「休息時の脈拍」、「最大脈拍」を選択し、「モデル効果の構成」のなかの「追加」をクリックする .
- [4] 「手法」→「ステップワイズ法」を選択する .
- [5] 「実行」ボタンを押す (図 3.1 が表示される) .
- [6] 「停止ルール (図 3.1 のなかの (a))」→「最小 BIC」を選択する .
- [7] 「方向 (図 3.1 のなかの (b))」→「変数減少」を選択する .
- [8] 「停止ルール」横の「全て追加 (図 3.1 のなかの (c))」をクリックする^a .
- [9] 「実行」ボタンを押す .

^a 「方向」を「変数増加」(前進ステップワイズ法) の場合には「全て追加」にしてはならない (切片のみから変数選択が出発するため) . また「変数増減」の場合には「停止ルール」を「閾値 p 値」のみしか実行できない

ここで「停止ルール (図 3.1 のなかの (a))」には、最小 BIC (Bayesian Information Criteria) 以外に、最小 AICc (修正 Akaike's Information Criteria) , 閾値 p 値がある . 最小 BIC 及び最小 AICc は回帰モデルの適合度を情報量規準という概念で数値化することで、最良のモデルを選択する . 他方、閾値 p 値とは変数の増加、あるいは減少の可否を検定により判断しながら実行する . そのため、閾値 p 値では有意水準と同様に閾値が必要になる . いずれの停止ルールもコンセンサスが得られているが、閾値 p 値では統計担当者恣意的に閾値 p 値を設定するため、より客観的な BIC あるいは AICc が推奨される .

このときの結果 (図 3.1 の履歴) を図 3.2 に示す . (a) は変数選択された変数をチェックボックスのチェックで表している . 本解析では「走行時間」、「走行時の脈拍」、「最大脈拍」が選択されている . (b) は選択された 3 個の変数を用いたときの重回帰分析における回帰パラメータに対する有意性検定での p 値である . いずれの回帰パラメータも帰無仮説「真の回帰パラメータは 0 である」に対して、p 値が有意水準 0.05 のも

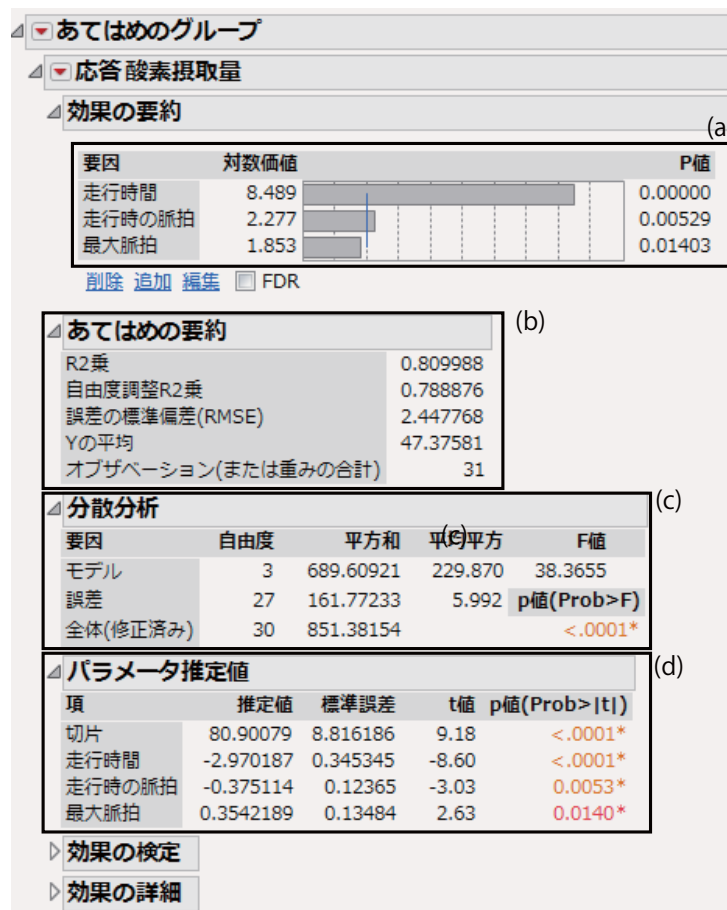


図 3.3: 選択された変数を用いた重回帰分析の結果

とで有意であり、回帰パラメータの有意性が認められた。(c) はステップワイズ法における変数の増加・減少の過程を表している。ステップ 1 からステップ 5 までは、手順「後退ステップワイズ (変数減少法) による重回帰分析の実行」の [8] を実行したときに全変数がモデルに追加されたことを表しており、ステップ 6 以降が後退ステップワイズ法でのプロセスを表している。ステップ 6 からステップ 7 までは BIC が減少している。すなわち、変数を削除することでモデル適合度が向上している。他方、ステップ 8-10 では BIC が上昇している。これは、ステップ 8-10 で削除された変数「最大脈拍」「走行時の脈拍」「走行時間」は意味のある変数であり、削除することでモデル適合度が大幅に悪化することを意味する。そのため、ステップ 11 において、ステップ 7 のモデルが最良モデルとして選択されている。

手順「後退ステップワイズ (変数減少法) による重回帰分析の実行」の実行後に「モデルの実行 (図 3.1 のなかの (d))」をクリックすると図 3.3 が表示される。これは、後退ステップワイズ法で選択された変数を用いて重回帰分析を行った結果を表している。(a) は重回帰分析の p 値および対数価値を表している。対数価値とは $-\log_{10}(p \text{ 値})$ で表される指標であり、大きいほど影響 (価値) が高いと解釈される。(b) は回帰モデルのあてはまりのよさを表している。ここでは、「自由度調整 R2 乗」(自由度調整済み寄与率) を診る。その結果、0.788876 であり、推定された回帰モデルが全体の約 78.9% を説明できる。(c) は回帰の分散分析表 (帰無仮説「回帰モデルに意味はない」) である。その結果、p 値は < 0.0001 であり、回帰モデルに意味が

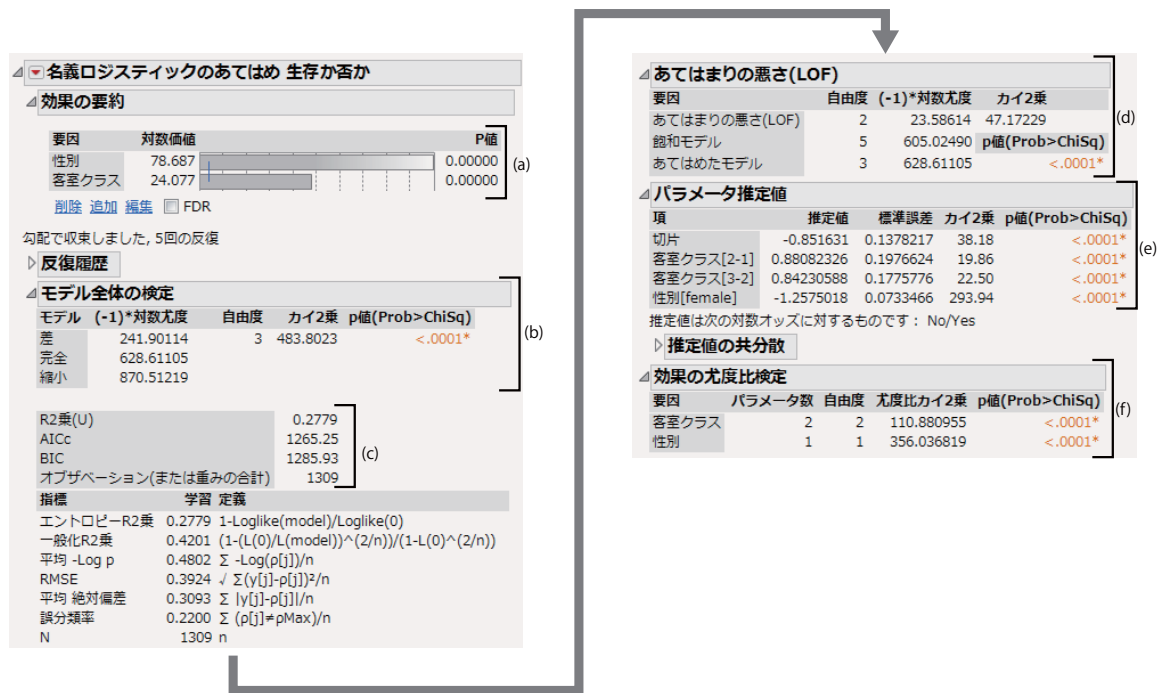


図 3.4: ロジスティック回帰分析の結果

あることが示唆された。(d) はパラメータ推定値である。選択されたすべての説明変数の回帰パラメータが有意(帰無仮説「回帰パラメータは0である」)だった。

3.2 ロジスティック回帰分析

ここでは、タイタニック号の生存の有無に関するデータを利用する。このサンプルデータは、「Titanic Passengers.jmp」に含まれている。これは、タイタニックの乗客と生存に対する影響要因を探索することを意図したデータである。本解析では、タイタニック号の水難事故による「生存か否か」に「性別」(名義尺度)、「乗客クラス」(順序尺度)が寄与しているか否かを評価する。ここで「生存か否か」では、「Yes」に1、「No」に0の識別番号が与えられている。JMPの注意点は、関心のある事象を0、それ以外を1と認識する¹。したがって、ここでは、タイタニック号の水難事故によって死亡した(=生存しなかった)要因を探ることになる。

ロジスティック回帰分析の実行

- [1] 「分析」→「モデルのあてはめ」を選択する(モデルのあてはめのウィンドーが表示される)。
- [2] 「生存か否か」を選択して「Y」をクリックする(「手法」が「名義ロジスティック」に自動的に変更される。これは「生存か否か」が名義尺度であることに依る)。
- [3] 「乗客クラス」、「性別」を選択し、「モデル効果の構成」のなかの「追加」をクリックする。
- [4] 「実行」ボタンを押す。

¹JMPでは、より若い数値、アルファベットに対して関心がある事象を割り振る。

このときの結果を図 3.4 に示す．因みにここでは，(a) はロジスティック回帰分析の共変量に対する尤度比検定での p 値および対数値を表している．対数値とは $-\log_{10}(p \text{ 値})$ で表される指標であり，大きいほど影響 (値) が高いと解釈される (医学統計学ではあまり用いないので注意されたい) ．

(b) はモデルの適合度を表しており， p 値は null モデル (切片のみのモデル) と full モデル (共変量を含めたモデル) の尤度比 (モデルの当てはまりの比) で検定を行った結果である． p 値が < 0.0001 なので推定された統計モデルに意味があることが示された ．

(c) はロジスティック回帰モデルの適合度の強さを表している．AICc 及び BIC は複数のモデル (例えば，「性別」だけを共変量とした場合や「乗客クラス」のみを共変量とした場合) を比較するための統計量であり，小さいほど適合度が良いと判断される．また， R^2 乗はロジスティック回帰モデルにおける寄与率を表している．ただし，重回帰モデルでの寄与率と異なり，ロジスティック回帰モデルでは寄与率が小さくなる傾向にあるため，評価が困難であるため，あまり利用されない ．

(d) はロジスティック回帰の不適合度 (Lack Of Fit) を表している．「飽和モデル」とは全ての共変量及び交互作用を含んだモデル (これ以上細分化できないほどの組み合わせとした場合のモデル) の負の対数尤度を表しており，「あてはめたモデル」は今回の主効果のみのモデルでの負の対数尤度である．あてはまりの悪さ (LOF) とは「飽和モデル-あてはめたモデル」である．ここでの p 値が有意であるということは交互作用を含んだ複雑なモデルのほうがより当てはまりがよくなることを意味する．本解析では， p 値が < 0.0001 なので，「乗客クラス \times 性別」の交互作用を含むことが推奨されている² ．

(e) は回帰パラメータの推定結果である．下側に「No/Yes」であることから，死亡 (No) がイベントで計算されている．「乗客クラス」は順序尺度であるため，ダミー変数で構成されている「乗客クラス [2-1]」とは 2~3 等席が 1，1 等席が 0 のダミー変数である．回帰パラメータが 0.8808... である．オッズ比は， $\exp(t \text{ 回帰パラメータ})$ であることから，回帰パラメータが正值ということは，オッズ比が 1 を上回ることを意味する．すなわち，2~3 等席の死亡リスクは，1 等席の死亡リスクよりも高いことが伺える．「乗客クラス [3-2]」とは 3 等席が 1，1~2 等席が 0 のダミー変数である．回帰パラメータが 0.8423... であることから，3 等席の死亡リスクは，更に高いことが示唆される (オッズ比の計算は後ほど述べる) ．また，「性別 [female]」は，男性が 1，女性が 0 のダミー変数である．回帰パラメータが $-1.25...$ (負値) なので，男性のほうが女性に比べて死亡リスクが高いことが分かる ．

(f) は要因を含むモデルと null モデル (切片のみのモデル) での尤度比検定である．「乗客クラス」は「乗客クラス [2-1]」及び「乗客クラス [3-2]」の 2 個のダミー変数によるモデルでの尤度比検定であり，「性別」は「性別 (female)」のみの尤度比検定である．いずれも p 値が < 0.0001 なのでいずれの要因も有意である ．

² 「ロジスティック回帰分析の実行」の [3] において「乗客クラス」「性別」を選択し「モデル効果の構成」のなかの「マクロ」→「完全実施要因」を選択すると飽和モデルとなる．因みに交互作用を含めることで性差は 3 等席において小さくなることがわかる

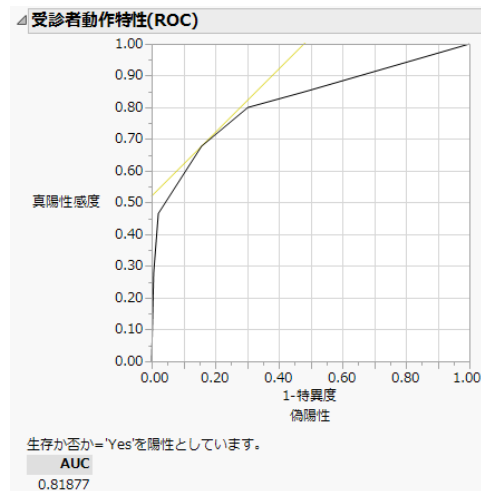




図 3.5: ROC 曲線の表示

ROC 曲線の表示

- [0] 上記「ロジスティック回帰分析の実行」を実行する。
- [1]  を左クリックする。
- [2] 「ROC 曲線」を選択する (どちらを陽性とするかを選択するウィンドーが表示される)。
- [3] 「Yes」を陽性とする。

次に推定されたロジスティックモデルを用いて ROC 曲線を描写する。このときの結果を図 3.5 に示す。最後にオッズ比の計算を行う。

オッズ比の計算

- [0] 上記「ロジスティック回帰分析の実行」を実行する。
- [1]  を左クリックする。
- [2] 「オッズ比」を選択する。

このときの結果を図 3.6 に示す。いま、「乗客クラス [2-1]」の回帰パラメータの推定値を $\hat{\beta}_{21}$ 、「乗客クラス [3-2]」の回帰パラメータの推定値を $\hat{\beta}_{32}$ とするとき、(a) のオッズ比の計算は、

水準 1	水準 2	計算方法
2	1	$\exp(\hat{\beta}_{21})$
3	1	$\exp(\hat{\beta}_{21} + \hat{\beta}_{32})$
3	2	$\exp(\hat{\beta}_{32})$
1	2	$1 / \exp(\hat{\beta}_{21})$
1	3	$1 / \{ \exp(\hat{\beta}_{21} + \hat{\beta}_{32}) \}$
2	3	$1 / \exp(\hat{\beta}_{32})$

で計算されている。その結果、2 等席の乗客は 1 等席の乗客に比べて約 2.41 倍の死亡リスクがあり、3 等席

オッズ比					
生存か否か: No対Yesのオッズ比に対して オッズ比の検定と信頼区間は、尤度比に基づいて計算されています。					
客室クラスのオッズ比					
水準1	/水準2	オッズ比	p値(Prob>ChiSq)	下側95%	上側95%
2	1	2.4128853	<.0001*	1.6416174	3.5647057
3	1	5.6020306	<.0001*	4.016299	7.8707409
3	2	2.3217144	<.0001*	1.6402042	3.2921658
1	2	0.4144416	<.0001*	0.2805281	0.6091553
1	3	0.1785067	<.0001*	0.1270528	0.2489854
2	3	0.4307162	<.0001*	0.3037514	0.6096802
性別のオッズ比					
水準1	/水準2	オッズ比	p値(Prob>ChiSq)	下側95%	上側95%
male	female	12.366652	<.0001*	9.3148147	16.560614
female	male	0.0808626	<.0001*	0.0603842	0.1073559

図 3.6: オッズ比の表示

の乗客は1等席の乗客に比べて5.60倍、2等席の乗客に比べて2.32倍の死亡リスクがあることが示唆された。さらに、男性は女性に比べて12.37倍の死亡リスクだった。因にすべてのオッズ比が有意だった。

3.3 Poisson 回帰分析

ここでは、カブトガニの巣穴に関するデータを用いる。メスのカブトガニの巣穴には、オスのカブトガニが1匹ずつ棲んでいる。この調査では近くにサテライトと呼ばれる他のオスの巣穴の数が何個存在するかを応答とし、それに影響を与える要因としてメスの特徴を用いて、Poisson 回帰分析により解析する。このデータは、サンプルデータ CrabSatellites に存在する。実行を以下に示す。

Poisson 回帰分析の実行

- [1] 「分析」→「モデルのあてはめ」を選択する(モデルのあてはめのウィンドーが表示される)。
- [2] 「サテライト数」を選択して「Y」をクリックする
- [3] 「手法」のなかから「一般化線形モデル」を選択する。
- [4] 「分布」のなかから「Poisson」を選択する。
- [5] 「リンク関数」のなかから「対数」を選択する。
- [6] 「色」「緑棘」「幅」「体重」を選択し、「モデル効果の構成」のなかの「追加」をクリックする。
- [7] 「実行」ボタンを押す。

このときの結果を図 3.7 に示す。(a) はモデルの適合度を表しており、上側の p 値は null モデル(切片のみのモデル)と full モデル(共変量を含めたモデル)の尤度比(モデルの当てはまりの比)で検定を行った結果である。p 値が < 0.0001 なので推定された統計モデルに意味があることが示された。また、下側の適合度検定の「Pearson」及び「デビアンズ」は残差に基づいて評価している。

(b) は Poisson 回帰分析の共変量に対する尤度比検定での p 値および対数値を表している。対数値と

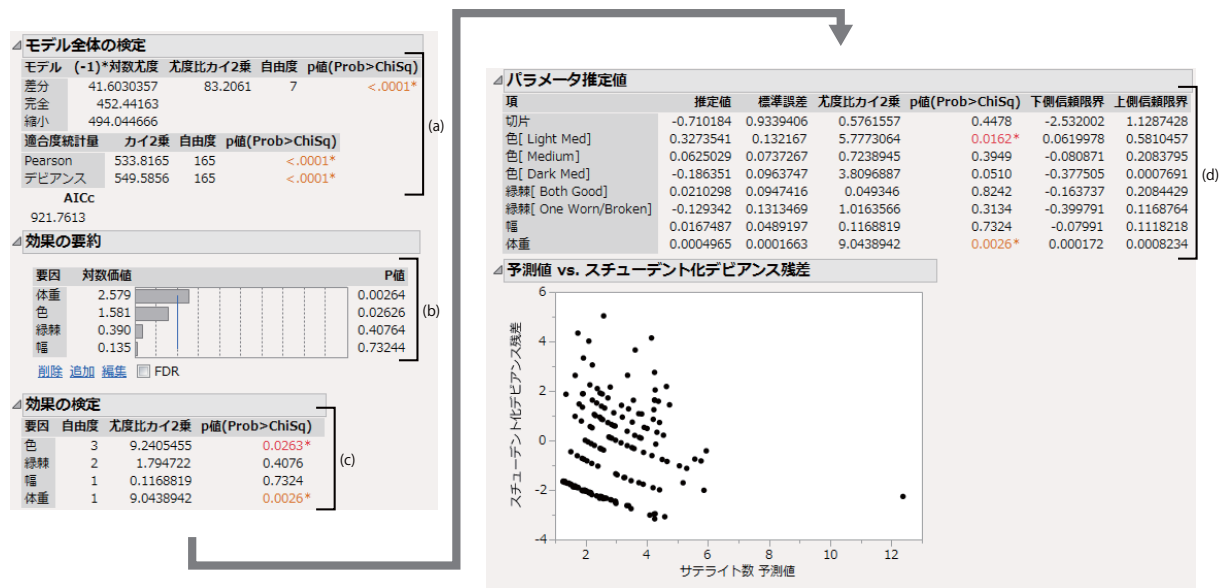


図 3.7: Poisson 回帰分析の結果

は $-\log_{10}(p \text{ 値})$ で表される指標であり、大きいほど影響 (価値) が高いと解釈される (医学統計学ではあまり用いないので注意されたい)。

(c) は要因を含むモデルと null モデル (切片のみのモデル) での尤度比検定である。「色」及び「体重」において有意である。

(d) は回帰パラメータの推定結果である。共変量「色」及びは「緑棘」が名義尺度であることからダミー変数が用いられている。「色 [color]」は color が真ならば 1, 偽ならば 0 を意味している。「緑棘」についても同様である。その結果、「色 [Light Med]」及び「体重」のみで有意だった。(c) において「色」及び「体重」が有意だったが、「色」は「Light Med」のみ他の色のメスのカブトガニとサテライト数が異なるようである。

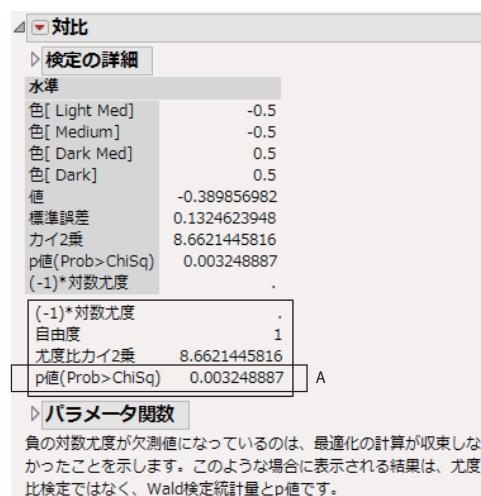



図 3.8: Poisson 回帰分析の対比の計算

次に甲羅の色が濃い (Dark, Dark Med) メスと薄いメス (Light Med, Medium) のあいだに違いがあるか否かを評価する．このような評価を「対比 (contrast)」という．

—— 対比の表示 ——

- [0] 上記「Poisson 回帰分析の実行」を実行する．
- [1]  を左クリックする．
- [2] 「対比」を選択する (どの共変量の対比を計算するかを選択するウィンドーが表示される) ．
- [3] 「色」を選択して「実行」ボタンを押す (対比を計算するするウィンドーが表示される)
- [4] 「Dark」, 「Dark Med」では「+」ボタン, 「Medium」, 「Light Med」では「-」ボタンを押す (「Dark」, 「Dark Med」は 0.5, 「Medium」, 「Light Med」は-0.5 になる) ．
- [5] 「完了」ボタンを押す ．

このときの結果を図 3.8 に示す．A は対比に対する尤度比検定の結果である．p 値が 0.0032... であることから，有意水準 $\alpha = 0.05$ のもとで有意である．すなわち，甲羅の色が濃い (Dark, Dark Med) メスと薄いメス (Light Med, Medium) のあいだにサテライト数の違いが認められた．

第4章 生存時間解析

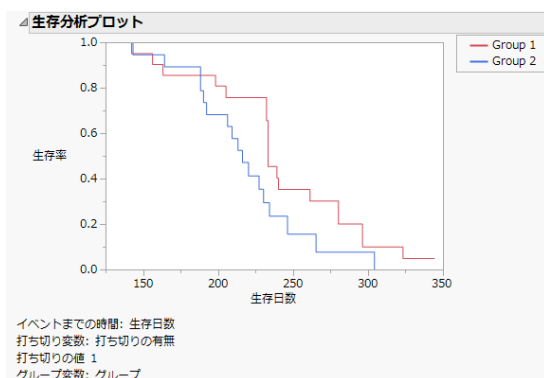
4.1 生存時間曲線の推定及び解析

ここでは、サンプルデータ Rats.jmp を利用する。このデータは、2 群に分けられたラットを発がん性物質にさらし、生存時間を測定したデータである。JMP では生存時間分布の推定と生存時間分布の比較が同時に行われる。その手順を以下に示す。

生存時間分布の推定と比較の実行

- [1] 「分析」→「信頼性/生存時間分析」→「生存時間分析」を選択する。
- [2] 「生存日数」を選択して「Y, イベントまでの時間」を押す。
- [3] 「グループ」を選択して「グループ変数」を押す。
- [4] 「打ち切りの有無」を選択して「打ち切り」を押す。
- [5] 「OK」ボタンを押す。

このときの結果を 4.1 に示す。4.1(I) は Kaplan-Meier プロットである。生存日数が約 180 日までは Group2 のほうが生存率が高かったが、以降は Group1 のほうが高かった。4.1(II) の (a) は要約指標を表している。「中央値時間」は中央生存時間 (MST; Median Survival Time) を表しており、下側 95%、上側 95% はそれぞれ MST の 95% 信頼区間の下側、上側信頼限界を表している。また、「組み合わせ」は全体での生存曲線の推定値である。その結果、Group1 のほうが Group2 に比べて MST が 17 日長いことが示唆された。(b) は生存時間分布を比較した結果である。ログランク検定 (ログランク)、一般化 Wilcoxon 検定 (Wilcoxon) のいずれも有意差は認められなかった。一般化 Wilcoxon 検定がログランク検定よりも p 値が大きいのは、



(I) Kaplan-Meier プロット

要約					
グループ	故障数	打ち切り数	平均	標準誤差	
Group 1	19	2	240.795	バイアスあり 11.206	
Group 2	17	2	218.757	9.40318	
組み合わせ	36	4	230.729	バイアスあり 7.57346	

分位点					
グループ	中央値時間	下側95%	上側95%	25%寿命	75%寿命
Group 1	233	232	280	232	280
Group 2	216	190	234	190	234
組み合わせ	232	213	239	201.5	261

グループ間での検定			
検定	カイ2乗	自由度	p値(Prob>ChiSq)
ログランク	3.1227	1	0.0772
Wilcoxon	2.6510	1	0.1035

(II) 統計量及び検定結果



図 4.1: 生存時間曲線の推定と比較の結果

Group 1							
生存日数	生存率	故障率	生存標準誤差	故障数	打ち切り数	リスク集合の大きさ	
0.000	1.0000	0.0000	0.0000	0	0	21	
142.000	0.9524	0.0476	0.0465	1	0	21	
156.000	0.9048	0.0952	0.0641	1	0	20	
163.000	0.8571	0.1429	0.0764	1	0	19	
198.000	0.8095	0.1905	0.0857	1	0	18	
204.000	0.8095	0.1905	0.0857	0	1	17	
205.000	0.7589	0.2411	0.0941	1	0	16	
232.000	0.6577	0.3423	0.1053	2	0	15	
233.000	0.4554	0.5446	0.1114	4	0	13	
239.000	0.4048	0.5952	0.1099	1	0	9	
240.000	0.3542	0.6458	0.1072	1	0	8	
261.000	0.3036	0.6964	0.1031	1	0	7	
280.000	0.2024	0.7976	0.0902	2	0	6	
296.000	0.1012	0.8988	0.0678	2	0	4	
323.000	0.0506	0.9494	0.0493	1	0	2	
344.000	0.0506	0.9494	0.0493	0	1	1	

Group 2							
生存日数	生存率	故障率	生存標準誤差	故障数	打ち切り数	リスク集合の大きさ	
0.000	1.0000	0.0000	0.0000	0	0	19	
143.000	0.9474	0.0526	0.0512	1	0	19	
164.000	0.8947	0.1053	0.0704	1	0	18	
188.000	0.7895	0.2105	0.0935	2	0	17	
190.000	0.7368	0.2632	0.1010	1	0	15	
192.000	0.6842	0.3158	0.1066	1	0	14	
206.000	0.6316	0.3684	0.1107	1	0	13	
209.000	0.5789	0.4211	0.1133	1	0	12	
213.000	0.5263	0.4737	0.1145	1	0	11	
216.000	0.4737	0.5263	0.1145	1	1	10	
220.000	0.4145	0.5855	0.1145	1	0	8	
227.000	0.3553	0.6447	0.1124	1	0	7	
230.000	0.2961	0.7039	0.1082	1	0	6	
234.000	0.2368	0.7632	0.1015	1	0	5	
244.000	0.2368	0.7632	0.1015	0	1	4	
246.000	0.1579	0.8421	0.0934	1	0	3	
265.000	0.0789	0.9211	0.0728	1	0	2	
304.000	0.0000	1.0000	0.0000	1	0	1	

図 4.2: 年次生存率の省察

一般化 Wilcoxon 検定が初期の生存時間の差に敏感である一方で、後期の生存時間の差には鈍感なためである（初期の生存率の差に関心がある場合には一般化 Wilcoxon 検定の結果を用いるが、一般的にはログランク検定の結果を用いる）。

年次生存率を省察する場合には Group1(Group1 を省察する場合) の左側のボタンをクリックする。図 4.2 は Group1, Group2 のそれぞれのボタンをクリックした場合である。例えば、6ヵ月生存率 (6ヵ月を $30.5 \times 6 = 183$) の場合には、「生存日数」が 183 日以下で最も生存日数が高い場所を見ればよい ((A) の部分)。すなわち、Group1 での 6ヵ月生存率は 0.8571 であり、Group2 での 6ヵ月生存率は 0.8947 だった。したがって、6ヵ月生存率では Group2 のほうが高かった (交差ハザードであったため)。

4.2 Cox 比例ハザード・モデル

ここでは、サンプルデータ VA Lung Cancer.jmp を用いる。このデータは、肺がんに対する無作為化比較試験の予後因子の評価を実施するデータである。共変量「セル」は「Adeno」「Large」「Small」「Squamous」の 4 水準の名義尺度、「処置」は標準治療 (Standard)、試験治療 (Test) の 2 水準の名義尺度、「前治療」は「Yes」「No」の 2 水準の名義尺度、「年齢」、「月数」(肺がん診断から試験開始までの期間)、「KPS」(一般全身状態) は連続尺度である。

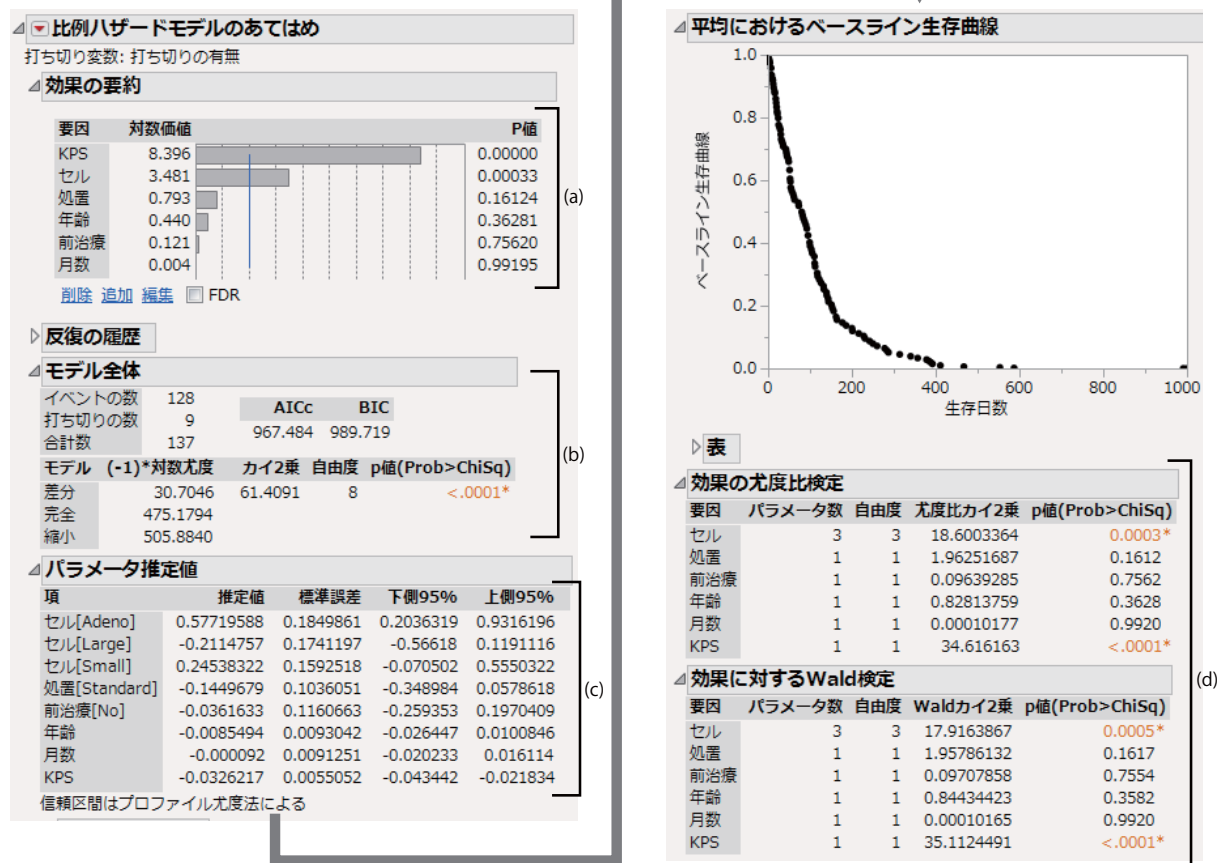


図 4.3: 比例ハザード・モデルの結果

比例ハザード・モデルの実行方法を以下に示す。

比例ハザード・モデルの実行

- [1] 「分析」→「信頼性/生存時間分析」→「比例ハザードのあてはめ」を選択する。
- [2] 「生存日数」を選択して「イベントまでの時間」を押す。
- [3] 「セル」「処置」「前治療」「年齢」「月数」「KPS」を選択して「モデル効果の構成」のなかの「追加」を押す。
- [4] 「打ち切りの有無」を選択して「打ち切り」を押す (打ち切りの値が Censored になっていることを確認する)。
- [5] 「実行」ボタンを押す。

このときの結果を図 4.3 に示す。(a) は比例ハザード・モデルの共変量に対する尤度比検定での p 値および対数値を表している。対数値とは $-\log_{10}(p \text{ 値})$ で表される指標であり、大きいほど影響 (価値) が高いと解釈される (医学統計学ではあまり用いないので注意されたい)。その結果、「KPS」および「セル」の影響が強く、有意だった。

(b) がモデル適合度を表しており、p 値は null モデル (基線ハザードのみの場合のモデル) での尤度との尤

▲ リスク比

▲ 単位リスク比

連続変数が1単位だけ変化した場合

項	リスク比	下側95%	上側95%	逆数
年齢	0.991487	0.9739	1.010136	1.0085861
月数	0.999908	0.979971	1.016244	1.000092
KPS	0.967905	0.957488	0.978403	1.0331596

▲ 範囲リスク比

連続変数が範囲全体で変化した場合

項	リスク比	下側95%	上側95%	逆数
年齢	0.669099	0.288517	1.606371	1.4945466
月数	0.992119	0.175518	3.998028	1.0079435
KPS	0.05484	0.020935	0.143244	18.23482

(a)

▲ セルのリスク比

水準1	/水準2	リスク比	p値(Prob>ChiSq)	下側95%	上側95%
Large	Adeno	0.4544481	0.0096*	0.2502748	0.824169
Small	Adeno	0.7176217	0.2325	0.4199109	1.2418767
Small	Large	1.579106	0.0825	0.9433383	2.6898478
Squamous	Adeno	0.3047391	<.0001*	0.1683479	0.5501962
Squamous	Large	0.6705696	0.1591	0.3845788	1.1714599
Squamous	Small	0.4246514	0.0014*	0.2444471	0.7213156
Adeno	Large	2.2004712	0.0096*	1.2133433	3.9956087
Adeno	Small	1.3934918	0.2325	0.8052329	2.3814575
Large	Small	0.6332697	0.0825	0.3717682	1.060065
Adeno	Squamous	3.2814957	<.0001*	1.8175335	5.9400784
Large	Squamous	1.4912695	0.1591	0.8536357	2.600247
Small	Squamous	2.3548727	0.0014*	1.3863557	4.090864

(b)

▲ 処置のリスク比

水準1	/水準2	リスク比	p値(Prob>ChiSq)	下側95%	上側95%
Test	Standard	1.3363418	0.1612	0.8907214	2.0096664
Standard	Test	0.7483115	0.1612	0.497595	1.1226855

▲ 前治療のリスク比

水準1	/水準2	リスク比	p値(Prob>ChiSq)	下側95%	上側95%
Yes	No	1.0750063	0.7562	0.6742989	1.6798514
No	Yes	0.9302271	0.7562	0.5952908	1.4830218

図 4.4: ハザード比の計算結果


度比検定を表している。p 値が < 0.0001 であることからモデルに意味があることが示された。

(c) はパラメータ推定の結果であり，(d) は予後因子に対する尤度比検定及び Wald 検定の結果である。因みに (a) の p 値は尤度比検定の結果を表している。

それぞれの予後因子によるハザード比は

ハザード比の表示

[0] 上記「比例ハザード・モデルの実行」を実行する。

[1]  を左クリックする。

[2] 「リスク比」を選択する。

このときの結果を図 4.4 に示す。(a) は連続尺度に対するハザード比である「単位リスク比」とは共変量が 1 変化した場合のハザード比である。例えば「年齢」であれば年齢が 1 歳増加するときに何倍死亡リスクが 0.991487 倍増加することを表している「範囲リスク」は共変量の最小値に対して最大値において何倍

死亡リスクが増加するかを表している。

(b) は名義尺度に対するハザード比である。「セル」は 4 水準の名義尺度であるため、その組み合わせを用いてハザード比が計算されている。