JMP による保健・看護研究のための
医学統計学入門

第 1.1 版

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

1 章 量的データにおける統計解析 - 1 -
1.1 データの種類 - 1 -
1.2 仮説検定とは何か - 1 -
1.3 JMP による量的データの統計解析 - 2 -
1.3.1 データの概要と入力 - 2 -
1.3.2 量的データの要約とグラフ表示 - 3 -
1.3.3 量的データの比較(1)：2 標本 t 検定 - 5 -
1.3.4 量的データの比較(2)：Welch 検定 - 7 -
1.3.5 量的データの比較(3)：Wilcoxon 検定 - 8 -
1.3.6 対応のあるデータの解析 - 10 -
2 章 質的データにおける統計解析 - 13 -
2.1 クロス集計表に基づく統計解析 - 13 -
2.1.1 クロス集計表に基づく統計解析の基本 - 13 -
2.1.2 リスク比とオッズ比 - 16 -
2.1.3 クロス集計表による集計結果を直接入力する場合 - 18 -
2.2 対応のあるクロス集計表に基づく統計解析 - 20 -
2.3 カテゴリ数が多い場合の工夫：対応分析 - 23 -
3 章 一元配置の分散分析と多重比較 - 27 -
3.1 一元配置の分散分析及び多重比較の概要 - 27 -
3.1.1 一元配置の分散分析 - 27 -
3.1.2 多重比較 - 27 -
3.2 JMP による分散分析と多重比較の実行 - 28 -
3.2.1 分散分析の実行 - 28 -
3.2.2 分散分析の仮定の確認：等分散性の検定 - 30 -
3.2.3 多重比較の方法(パラメトリックの場合) - 31 -
3.3 3 群以上のデータに対するノンパラメトリック検定 - 33 -
3.3.1 Kruskal-Wallis 検定 - 33 -

3.3.2 多重比較の方法(ノンパラメトリックの場合)....- 34 -
4 章 相関分析・回帰分析....- 37 -
4.1 相関分析....- 37 -
4.1.1 相関分析の略説....- 37 -
4.1.2 JMP による相関分析....- 38 -
4.2 単回帰分析....- 40 -
4.2.1 単回帰分析の略説....- 40 -
4.2.2 JMP による単回帰分析....- 41 -
4.3 重回帰分析....- 43 -
4.3.1 重回帰分析の概要と JMP による実行....- 43 -
4.3.2 変数選択の方法....- 46 -
5 章 因子分析....- 51 -
参考文献....- 56 -
補遺 1：統計的検定の取捨選択....- 57 -

1 章 量的データにおける統計解析

1.1 データの種類

データの種類は，量的データと質的データの 2 種類に大別される．量的データとは，個々の観測値が数量で表されるデータであり，平均値あるいは中央値．数量には，計量データと計数データの 2 種類がある．計量データとは，血圧，腫瘍径，出血量などのように，数値に単位があるようなデータである．計量データは，小数点以下の値をとり，連続的に切れ目がないため，連続データと呼ぶこともある．一方で，計数値とは，ポリープの個数やリンパ節転移個数のように，個数あるいは回数として計測されたデータである．

質的データは，2 値データと多値データに分けられ，多値データは，更に名義カテゴリカル・データと順序カテゴリカル・データに分けられる．2 値データとは，奏効の有無，疾患の有無，治療の改善・非改善のように，アウトカムが 2 カテゴリで表されるデータである．これに対して，多値データは，3 個以上のカテゴリで表される．名義カテゴリカル・データとは，カテゴリが被験者の状態を表すラベルとして扱われるデータであり，疾患の種類や血液型がこれに該当する．一方で，疾患の進行程度を軽度，中程度，重度のカテゴリで測る場合，疾患の進行には，軽度＜中程度＜重度の順序関係が成り立つ．このように，カテゴリに順序関係が存在する場合を順序カテゴリカル・データという．

本章では，量的データの取り扱いについて述べる．

1.2 仮説検定とは何か

いま，抗がん剤治療中の胃癌患者に対して，術後補助化学療法開始時から栄養介入を実施した 53 名(栄養介入群)と実施しなかった 47 例(栄養非介入群)での治療後 6 カ月間での体重減少率を比較する研究を実施した．その結果，栄養介入群での体重減少率の平均値は 4.86%(標準偏差[1]：3.72)であり，栄養非介入群での体重減少率の平均値は 6.60%(標準偏差：4.90)であった．このとき，「栄養介入が術後補助化学療法を抑制したと判断してよいだろうか」．このことを統計学的に判断する方法が，仮説検定(検定)である．

仮説検定では，2 種類の仮説(帰無仮説 H_0，対立仮説 H_1)を設定する．帰無仮説 H_0 とは，言いたいことと反対の仮説(栄養介入の有無によって体重の平均減少率に違いがない)であり，対立仮説 H_1 とは，本来言いたい仮説(栄養介入の有無によって体重の平均減少率に違いがある)である．そして，帰無仮説 H_0 の「確からしさ」が小さいときに，帰無仮説 H_0 が誤っている(棄却される)と判断し，その逆仮説である対立仮説 H_1 が正しい(有意である)と判断する．

帰無仮説 H_0 が正しいとしたもとで，今回の研究結果が「どれぐらいの確率で生じるのか」を計算するとき，この確率は，p 値(有意確率)と呼ばれ「帰無仮説 H_0 の確からしさを表す確率」として解釈される．事例での p 値は 0.047 であることから，帰無仮説 H_0 の確からしさは 4.7%であることがわかる．

[1] 標準偏差とはデータのバラツキ具合を表す測度である．厳密には，平均値周りのバラツキを表している．今回の場合には，体重減少率の個人差を示す．

このとき，「帰無仮説 H_0 が誤っている(統計用語では，「棄却される」，「有意である」と呼ばれる)」と判断するには，p 値に対する閾値(通常は 0.05)を予め規定しなければならない．この閾値が有意水準 α である．有意水準 α =0.05 とするとき，この研究での p 値は，有意水準 α よりも小さいことから，帰無仮説 H_0 が棄却される．したがって，「栄養介入の有無によって体重の平均減少率に違いがある」と解釈できる．

1.3 JMP による量的データの統計解析

1.3.1 データの概要と入力

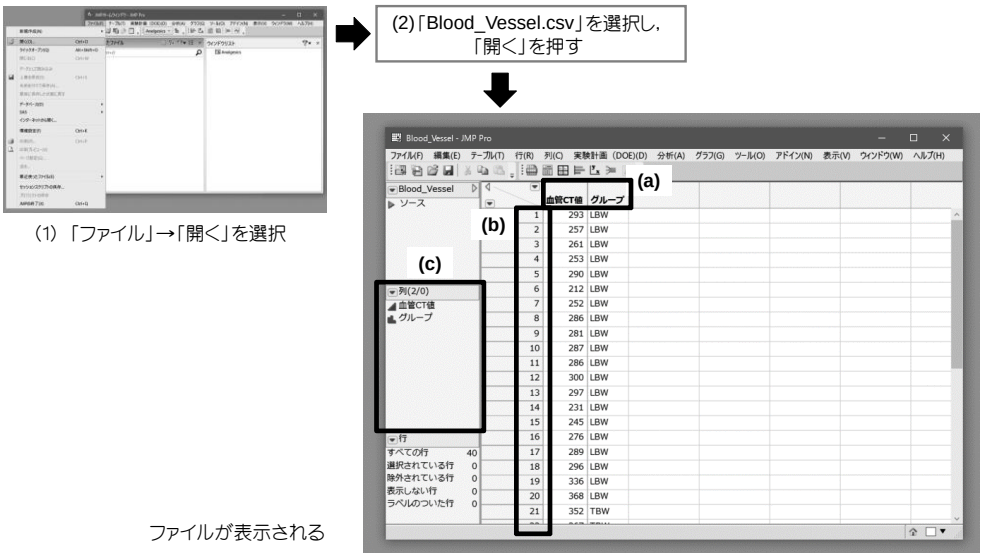
高度な画像診断を行うために，血管内に被験者の体格に応じた量の造影剤を注入して X 線画像がとられる．被験者の体脂肪を考慮した体重(LBW; Lean Body Weight)と通常の体重(TBW；Total Body Weight)から得られた血管 CT 値の平均に違いがあるだろうか(柳川他，2011).

このデータは，「Blood_Vessel.csv」に保存されている．このファイルを Excel で開いたものが以下である．

	A	B	C
1	血管CT値	グループ	
2	293	LBW	
3	257	LBW	
4	261	LBW	
5	253	LBW	
6	290	LBW	

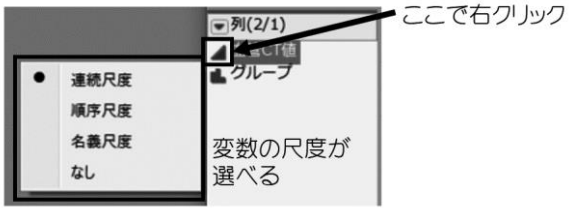
最初の列(1 列目)には，「血管 CT 値」，「グループ」が記載されている．これらが JMP で読み込んだときの変数(項目)名として認識される．「血管 CT 値」は各被験者の血管 CT 値であり，グループは測定された体重の種類(LBW，TBW)である．上記では，5 例が表示されているが，実際には 40 例(LBW:20 例，TBW：20 例)が存在する．

このファイルをに読み込む方法を以下に図示する．



ここで，(a)は変数名(項目名)であり，先ほどの Excel ワークシートの最初の列に対応している．また，(b)は，被験者の番号を表している．また，(c)は変数名とその変数のタイプを表している．変数のタイプとは，1.1 節で述べた変数の形式を意味している．「血管 CT 値」の(画面上では青色)は，量的データ(JMP では連続尺度)を表している．「グループ」の(画面上では赤色)は，名義カテゴリカル・データ(JMP では名義尺度)を表している．JMP には，更に，順序カテゴリカル・データ(JMP では順序尺度)を表す(画面上では緑色)がある．JMP では，これらの変数の形式によって統計手法の取捨選択が自動的に行われるので，注意が必要である．

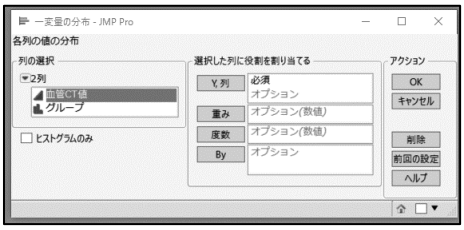
これらの変数の形式は，もとのファイルのデータが数字の場合には量的データ，文字の場合には名義カテゴリカル・データとして自動認識される．もし，意図しないデータ形式の場合には，アイコン上で右クリックすれば，



のように表示されるため，本来の形式を選択すればよい．今回の場合には必要ないのでそのまま続ける．

1.3.2 量的データの要約とグラフ表示

ここでは，グループ毎(LBW，TBW)でのデータの要約(平均値など)を見ていく．データの要約は，

量的データの要約
1.　「分析」→「一変量の分布」を選択する(下図のような画面が表示される).

2.　「Y, 列」に「血管 CT 値」,「By」[2]に「グループ」を割り当てる. 割り当て方は, ・「列の選択」にある「血管 CT 値」を左クリックし,「Y, 列」横の白い箱上にドラック&ドロップする[3]. ・「列の選択」にある「グループ」を左クリックし,「By」横の白い箱上にドラック&ドロップする.
3.　「OK」ボタンを押す.

で実行可能である．

すると，図 1.1 のような結果が表示される．これが，JMP の統計解析の初期画面といえる．なお，今回の場合には，「Y, 列」が量的データ(JMP では連続尺度)なので，この表示形式になっているが，質的データ(JMP では順序尺度，名義尺度)では，別の表示形式になる(詳しくは次章で述べる).

---

[2] 「By」とはカテゴリ毎に計算することを意味する．今回の場合には LBW，TBW 毎に計算する．もし，全体での要約を計算したい場合には By の部分を空白(何もしない)で良い.

[3] ドラック＆ドロップの代わりに「Y,列」と書かれたボタン(少し濃い灰色の部分)を左クリックしても良い.

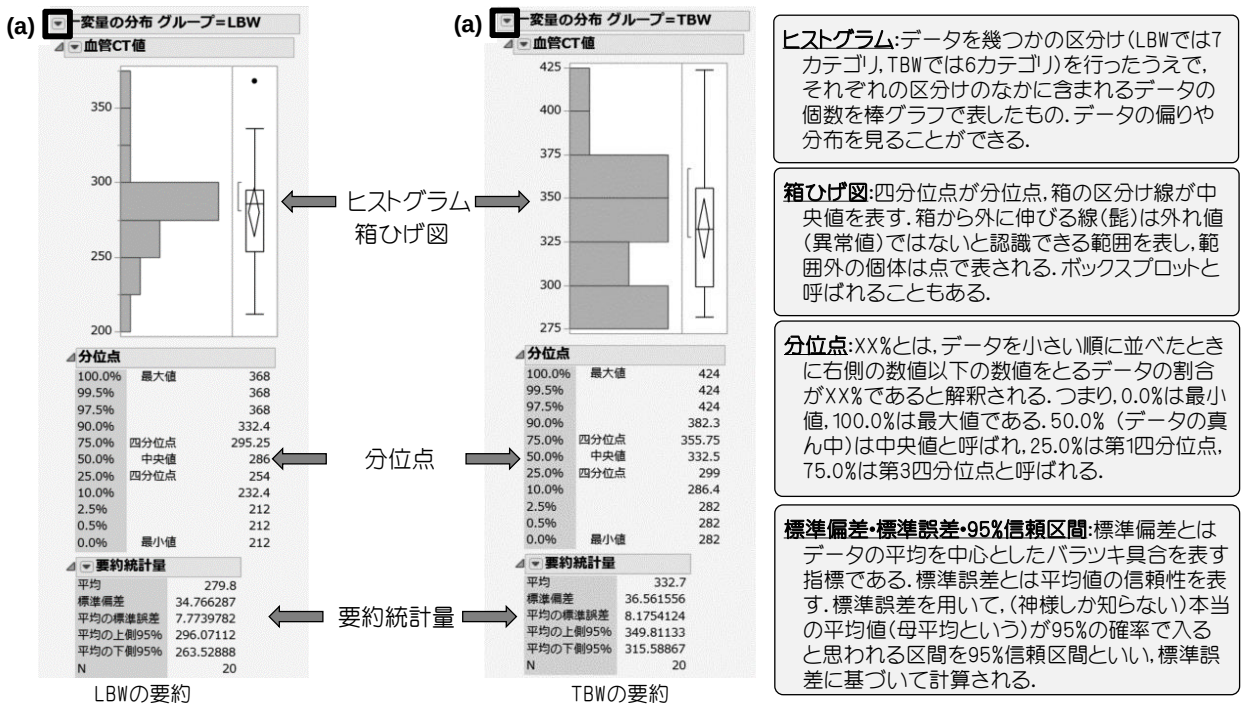


図 1.1：量的データの要約結果

図 1.1 より，平均値および中央値(データの真ん中の値)ともに，LBW よりも TBW のほうが大きいことがわかる．一方で，標準偏差にはあまり違いがないことから，個人差には大きな違いがないように思われる．また，ヒストグラムを見ると，LBW は 275～300 の範囲での度数(範囲内の個体数)が最大となっており，それを中心に左右対称に分布しているのに対して，TBW では，血管 CT 値が低い範囲の度数も高くなっている．因みにの右側には，JMP 結果における統計用語の簡単な説明を記載しているので参考にしてほしい．

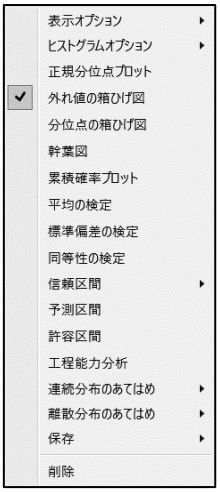
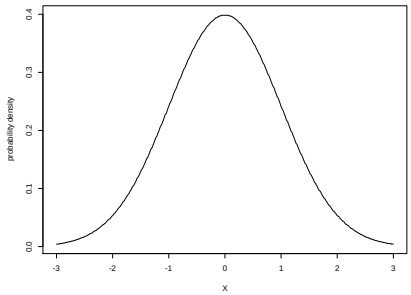


図 1.1 を「JMP の統計解析の初期画面」と前述したのは，ここから様々な解析が実行できるためである．図 1.1 の(a)で示した▼を左クリックすると，右図のメニューが表示される．

例えば，「正規分位点プロット」を用いることで，データが正規分布に従っているか否かを見ることができる(散布図に直線が引かれた図になっており，データ点に直線が当てはまっていれば正規分布に従っていると解釈できる)．また「平均の検定」を用いることで，平均値が任意の値であるかどうかを検定できる．

メモ 1-1：正規分布
正規分布とは，統計学の最も基本的な確率分布(統計学では個々のデータは，ある確率によって得られると考えている．このとき，得られたデータとその確率の対応関係のことを確率分布という)であり，自然現象や社会現象の多くの事象は正規分布に従っていると考えている．



また，統計学の多くの方法は，正規分布に基づいている．因みに，正規分布は，平均と標準偏差によって成り立っている．我々がデータを要約して評価する場合には，平均値を用いることが多いが，このことは，暗黙裡に正規分布を想定しているといえる(例えば，期末テストの成績を平均点で評価するなど)．図 1.1 の平均に対する 95%信頼区間(平均の上側 95%，平均の下側 95%)についても，データが正規分布に従っているという仮定のもとで計算されたものである．

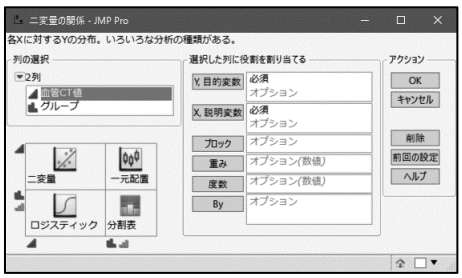
正規分布は，上図のような釣り鐘型の左右対称な形状を示している．TBW の分布形状に非常に類似していることから，TBW は高度に正規分布に従っていることが伺える．

1.3.3 量的データの比較(1)：2 標本 t 検定

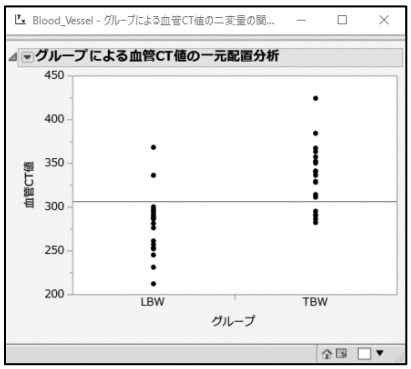
ここでは，検定を用いて LBW と TBW を比較する．このとき，正規分布を仮定したもとで平均値を比較する検定方法を 2 標本 t 検定という．

2 標本 t 検定の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される)．



2. 「Y，目的変数」に「血管 CT 値」，「X，説明変数」に「グループ」を割り当てる．
3. 「OK」ボタンを押す(下図のような画面が表示される)．

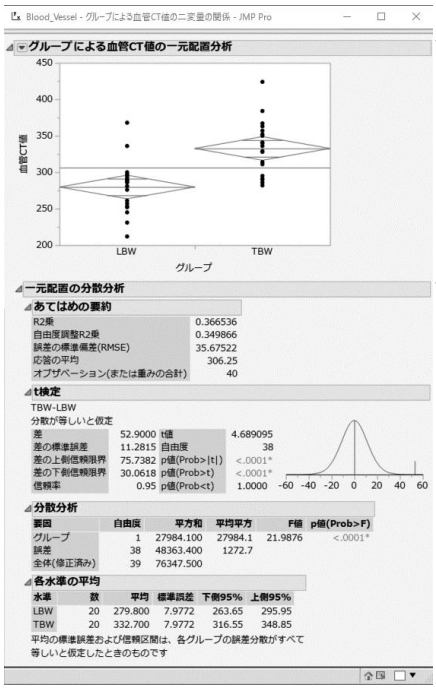


4　「▼」ボタンを左クリックし，「平均/ANOVA/プーリングした t 検定」を選択する．

上記の手順を実行すると図 1.2 の結果が表示される．ここで，(a)は各群のデータ点と平均値を表している．LBW の血管 CT 値は，TBW よりも低いことが伺える．このことは，各群での平均値と 95%信頼区間の出力(e)でも明らかである．

「平均/ANOVA/プーリングした t 検定」は，3 群以上の場合には 1 元配置の分散分析(詳しくは 3 章を参考)，2 群の場合には 2 標本 t 検定を自動的に選択する．(b)(d)のアウトプットは，1 元配置の分散分析

を解釈するのに用いることはあるものの，2 標本 t 検定では不必要である．したがって，これらの結果



(a) 各群のデータ点と平均値
- データ点は，観測値を表している．
- 水平線は，全体平均を表している．
- 菱形の横の長さは被験者数(標本サイズ)に対応している．
- 菱形の中央線は各群の平均値を表している．
- 菱形の縦の長さは平均の95%信頼区間を表している．

(b) 分散分析でのあてはめの要約
- 分散分析(3章)の適切性を表す要約
- 2標本t検定ではあまり意味がないので，無視してかまわない．

(c) 2標本t検定の結果
- このなかのp値が最も重要である．
- 差の信頼限界とは，TBWとLBWの平均値の差(TBW-LBW)に対する95%信頼区間を表している．

(d) 分散分析の結果
- p値は2標本t検定の「p値(Prob>|t|)」と一致する(無視)．

(e) 各群での平均値と95%信頼区間
- 各群での被験者数(数)，平均値，95%信頼区間(下側95%，上側95%)を表している．

図 1.2：2 標本 t 検定の結果

は，無視してもかまわない．

この出力結果において，最も重要なのが(c)である．「差」とは TBW の平均値(332.7)と LBW の平均値(279.8)の差を表している((c)の出力のタイトル「t 検定」の下側の表記「TBW-LBW」を参照)．また，「差の上側信頼限界」，「差の下側信頼限界」は，平均値の差に対する 95%信頼区間を表しており，この区間内に 0 を含まなければ，2 標本 t 検定においても有意差は認められない．この事例では，二つの信頼限界の範囲に 0 を含んでいないことから，有意であることが伺えるとともに，正値であることから，TBW の平均値のほうが LBW よりも統計学的にも大きいことが伺える．

「t 値」および「自由度」は p 値を計算するのに必要なものであり，p 値による解釈を行ううえでは必要ない(t 値の右の図は，t 値の位置と確率分布を表しているが，これを解釈する必要はない)．

図 1.2 では，帰無仮説「TBW と LBW の平均値には違いがない」に対する 2 標本 t 検定の結果として，3 種類の p 値が表示されている．それぞれの p 値の意味と解釈を示す．

p 値(Prob>|t|)	両側対立仮説(両側検定)という．2 群間に差がある(違いがある)ことを示す場合に用いる．p 値が 0.05 を下回る(有意である)とき，「TBW と LBW の平均値には違いがある」と解釈することができる．
p 値(Prob>t)	片側対立仮説(片側検定)という．A 群のほうが B 群よりも平均値が大きいことを示す場合に用いる．p 値が 0.05 を下回る(有意である)とき，「TBW の平均値のほうが LBW の平均値よりも大きい」と解釈することができる．

p 値(Prob<t)	片側対立仮説(片側検定)という．B 群のほうが A 群よりも平均値が大きいことを示す場合に用いる．p 値が 0.05 を下回る(有意である)とき，「LBW の平均値のほうが TBW の平均値よりも大きい」と解釈することができる．

看護研究を含む観察研究において，片側対立仮説(片側検定)を用いることは殆どない．そのため，実際の利用では，「p 値(Prob>|t|)」のみを見ればよい．因みに，「p 値(Prob>|t|)」を 2 倍すると，「p 値(Prob>t)」あるいは「p 値(Prob<t)」と同じになる(t が正値ならば前者，負値ならば後者になる)．そのため，「p 値(Prob>|t|)」(両側対立仮説)では有意でないのに，「p 値(Prob>t)」(片側対立仮説)では有意であるという結果になる(例えば p 値(Prob>|t|)が 0.03 であれば，p 値(Prob>t)は 0.06 になる)．このような場合でも，「p 値(Prob>|t|)」(両側対立仮説)を利用しなければならない．

この事例では，p 値(Prob>|t|)が<0.0001(0.0001 を下回る)ことから，有意水準 α=0.05 のもとで有意である．すなわち，TBW と LBW の平均値には違いがあることが統計学的にも認められた．

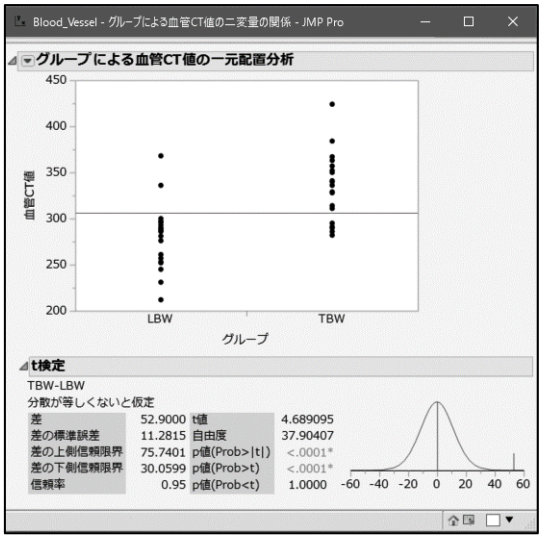
もし，p 値が 0.05 を上回った場合(有意でない)，帰無仮説が正しいと解釈してはいけない．この場合には，　対立仮説が正しいという根拠は得られなかっただけであり，2 群間の平均値が等しいとは解釈できない．

1.3.4 量的データの比較(2)：Welch 検定

2 標本 t 検定では，2 群間の標準偏差(バラツキ具合)が同じであることが仮定されている．これに対して，著しく標準偏差が異なる場合に用いることができるのが Welch 検定である．

Welch 検定の実行
1.　「2標本 t 検定の実行」の 1 から 3 までを行う.
2.　「▼」ボタンを左クリックし,「個々の分散を用いた t 検定」を選択する.

このときの実行結果以下に示す．



出力結果の形式は，図 1.2 の(c)と同じである．ただし，差の 95%信頼区間(差の上側信頼限界，差の下側信頼限界)が異なることに注意されたい．Welch 検定においても，p 値(Prob>|t|)が<0.0001(0.0001 を下回る)ことから，有意水準 α=0.05 のもとで有意である．

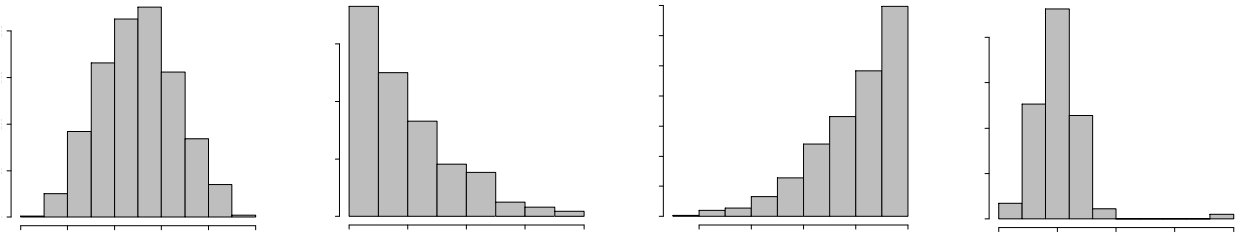
メモ 1-2：2 標本 t 検定と Welch 検定
2 標本 t 検定では，2 群間の標準偏差が同じであることが仮定されているのに対して，Welch 検定では，仮定する必要がない．そのため，一見すると Welch 検定を優先して使いたくなるかもしれない．

しかしながら，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．2 群間の比較の関心は，(1)2 群間に「何らかの」の違いに関心があるのか(同じでないことを示したいのか)，(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，標準偏差が異なることを示すことができれば(等分散性の検定[4])，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(後述)を用いることが推奨される．

さらに，「等分散性の検定」の結果で，有意であれば「Welch 検定」，有意でなければ「2 標本 t 検定」という取捨選択を推奨する文献があるが，この取捨選択は誤りである．

1.3.5 量的データの比較(3)：Wilcoxon 検定

下図は，ヒストグラムに対する幾つかのパターンである．



(a)　(b)　(c)　(d)

(a)は左右対称な分布形状を示しており，正規分布に従っていることが示唆される．これに対して，(b)及び(c)は著しく歪んだ分布形状を示している．また，(d)は左右対称な分布形状を示しているものの，外れ値(異常値)が示唆される．つまり，(b)～(c)では正規分布に従っていない可能性が高く，(d)では平均値が外れ値の影響を受ける可能性がある(つまり，平均値がデータ全体を代表する値とは言えない)．このような場合には，正規分布に従わない場合に用いることができる検定，すなわち，ノンパラメトリック検定[5]を利用する．

2 群を比較する場合に用いられるノンパラメトリック検定のなかで，最も代表的なものが Wilcoxon 検定(Mann-Whitney U 検定，Mann-Whitney-Wilcoxon 検定ともいう)である．Wilcoxon 検定とは，観測値を小さい順に順位付けを行い，その順位を用いて検定する．

Wilcoxon 検定の JMP による実行手順を以下に示す．

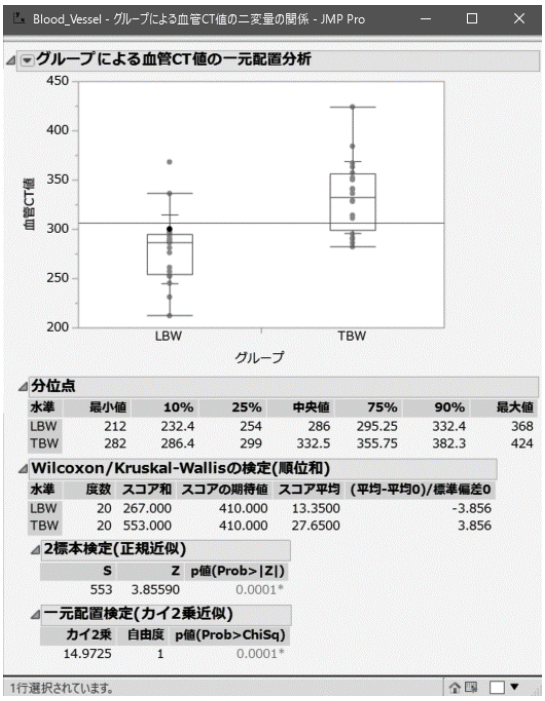
Wilcoxon 検定と分位点表示の実行
1.　「2標本 t 検定の実行」の 1 から 3 までを行う.
2.　「▼」ボタンを左クリックし,「分位点」を選択する[6].
3.　「▼」ボタンを左クリックし,「ノンパラメトリック」→「Wilcoxon 検定」を選択する.

[4] JMP にも実装されており，2 標本 t 検定，Welch 検定と同様に「▼」のメニューから，「等分散性の検定」を選べばよい．ただし，JMP では様々な検定結果が出てくるが，「Levene」あるいは「F 検定」を用いることが多い(内田他，2012).

[5] 正規分布に基づく検定(厳密には何らかの確率分布に基づく検定)をパラメトリック検定，そうでない場合をノンパラメトリック検定という．

[6] Wilcoxon 検定のみであれば，必ずしも必要ないが，その場合には，図 1.3 の(a)に箱ひげ図が表示されず，平均±標準偏差が表示される．

このときの結果を，図 1.3 に示す．(a)は，各群の箱ひげ図(中央値，第 1 四分位点(25%)，第 3 四分位点(75%)および外れ値を示す髭)を表している．LBW では 1 例の外れ値が示唆されているが，極端に大きいとはいえない．また，相対的な位置関係についても，LBW よりも TBW のほうが血管 CT 値が高いこ



(a) 各群の箱ひげ図

(b) 各群の分位点

(c) Wilcoxon検定の結果

・スコア平均とは，データ全体での順位の平均値を表しているが，スコアを解釈に用いることはないので無視

・2標本検定(正規近似)，一元配置検定(カイ2乗近似)の二つのp値がある．前者は，近似結果を補正しており，後者は補正なしでの結果である．

図 1.3：Wilcoxon 検定の結果

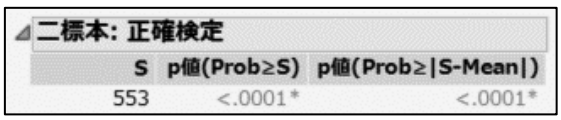
とが伺える．

(c)は，帰無仮説「LBW と TBW の血管 CT 値に相対的な違いがない」に対して，両側対立仮説(JMP では片側対立仮説が表示されない) 「LBW と TBW の血管 CT 値に相対的な違いがある」に対する，Wilcoxon 検定の結果である．Wilcoxon 検定では，p 値を計算する計算式は存在せず，近似を用いることがある．JMP では，「2 標本検定(正規近似)」及び「一元配置(カイ 2 乗近似)」の 2 種類が存在する．前者では，近似に対して補正を行うのに対して，後者では行っていない．したがって，「2 標本検定(正規近似)」での p 値を用いることを推奨する．

Wilcoxon 検定における p 値を計算するための近似の精度は，標本サイズが小さいほど悪くなる．そのため，被験者数が数 10 名の場合には，正確検定(正確な p 値の計算)を実施するほうが良い[7].

正確な Wilcoxon 検定の実行
1.　「2標本 t 検定の実行」の 1 から 3 までを行う.
2　「▼」ボタンを左クリックし,「ノンパラメトリック」→「正確検定」→「正確 Wilcoxon 検定」を選択する.

そうすると，下図が追記される．



[7] 正確検定は JMP Pro (本学でライセンス契約しているのは JMP Pro である)のみの機能である．また，50 例程度以上を超えるとハングアップ(ソフトがフリーズする)可能性があるので，注意が必要である．被験者数が多くなると，近似での結果も同じになるので，そちらを利用されたい．

ここで，「p 値(Prob≧S)」は，片側対立仮説(今回の場合は，TBW の血管 CT 値の相対的な位置は LBW よりも大きい)での p 値を表しており，「p 値(Prob≧|S-Mean|)」は両側対立仮説での p 値を表している．いずれも，有意水準 α=0.05 よりも小さいので有意である．

1.3.6 対応のあるデータの解析

医学系研究において「比較」を考えるとき，2 種類のデータの取得方法がある．一つは，被験者をランダムに 2 群に分け，それぞれの群に対して異なる介入を行ったり，背景因子が異なる 2 群を比較する場合などである．この場合には，それぞれの群を構成する被験者が異なる．このようなデータを独立 2 標本といい，前節までに述べた統計手法を用いて比較を行う．

もう一つは，介入前後でのアウトカム(検査値やアンケート調査)の変化を比較する場合や，2 種類の検査を同一被験者に実施して，検査結果の違いを比較する場合である．アウトカムが同一被験者からとられることを，対応のある場合，あるいはマッチドペアという．ここでは，対応のある場合の評価方法について述べる．

助産師が 5 年間の経験で分娩介助についてどのような意識の変革を起こすかを調べるため，資格取得直後と 5 年後に，分娩介助に関する 20 項目を自己評価してもらう研究が行われた(柳川他，2011)．資格取得後と 5 年間の経験後で，助産師の意識の差に違いがあるだろうか．

このデータは，「midwife.csv」に保存されている．このファイルを Excel で開いたものが以下である．

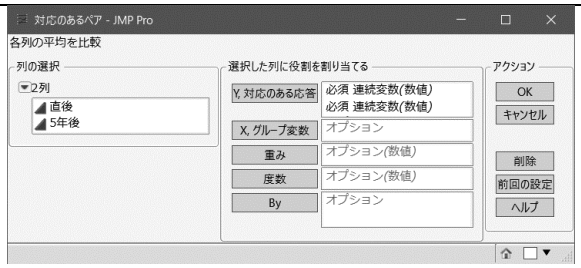
	A	B
1	直後	5年後
2	84	88
3	78	70
4	76	80
5	82	94

最初の列(1 列目)には，「直後」，「5 年後」が記載されている．これまでの血管造影の事例(独立 2 標本)では，アウトカム(血管 CT 値)とグループ(体重の種類)で構成されているのに対して，対応のあるデータでは，各列が個々の被験者に対応している．なお，ファイルの入力の方法は，1.3.1 節と同様なので，割愛する．

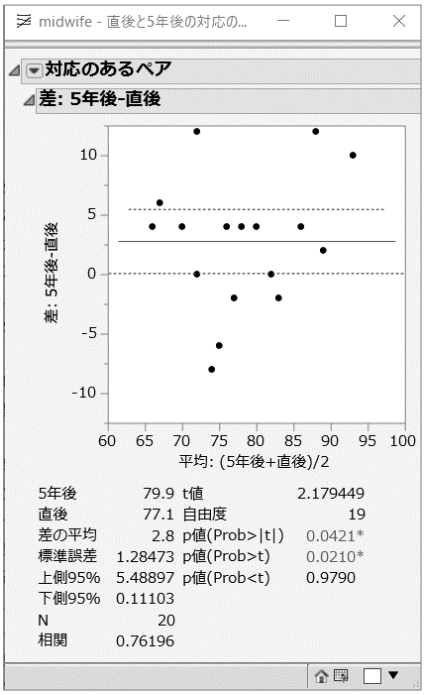
(a) 対応のある t 検定

データが正規分布に従うとき，対応のあるデータでは，対応のある t 検定を用いる．このデータでの関心は，「資格取得直後と 5 年後の分娩介助アンケートのスコア(以下，スコア)に変化があるか否か」にある．したがって，個々の被験者に対して，資格取得直後と 5 年後のスコアの差を計算し，その平均値が 0 に近くなければ変化したと考えることができる．従って，対応のある t 検定における帰無仮説は「資格取得前後と 5 年後のスコアの差の平均値は 0 である(資格取得前後と 5 年後のスコアに変化がない)」になる．対応のある t 検定の手順を以下に示す．

対応のある t 検定の実行
1.　「分析」→「発展的なモデル」→「対応のあるペア」を選択する(下図のような画面が表示される).



2. 「Y, 対応のある応答」に「直後」,「5 年後」の両方を割り当てる.
3. 「OK」ボタンを押す.



(a) 和と差のプロット

・X軸は被験者毎の平均，Y軸は5年前と5年後の差を表している.

・実線は差の平均値を表しており，点線は差に対する95%信頼区間を表している.

(b) 対応のあるt検定の結果･要約

・上側95%と下側95%は，差の平均に対する95%信頼区間

・Nは被験者数(標本サイズ)

・相関とは相関係数を表している．相関係数は，-1から1までの範囲をとり，0に近いほど相関関係が弱く，-1に近づくほど負の相関関係が強くなり，1に近づくほど正の相関関係が強くなる．詳しくは，5章を参照.

図 1.4：対応のある t 検定の結果

上記の手順を実行すると図 1.4 の結果が表示される．ここで，(a)は和と差のプロットを表している．ここで，個々の被験者の資格取得直後と 5 年後のスコアの変化を「スコア差」とすると，スコア差は

スコア差=(5 年後のスコア)－(直後のスコア)

である．Y 軸(縦軸)のデータ点の多くが 0 を上回っていることから，多くの助産師が 5 年後にスコアが上昇していることが伺える.

対応のある t 検定の結果は，(b)に表される.「5 年後」の 79.4 と「直後」の 77.1 は，資格取得 5 年後のスコアと資格取得直後の平均値である．また,「差の平均」は，スコア差の平均値であり,「上側 95%」と「下側 95%」は，スコア差の平均に対する 95%信頼区間である．この 95%信頼区間内に 0 を含まなければ，対応のある t 検定においても有意差は認められない．この事例では，二つの信頼限界の範囲に 0 を含んでいないことから，有意であることが伺える．また，正値であることから，5 年後のほうが取得直後に比べて，スコアが上昇している.

「t 値」および「自由度」は p 値を計算するのに必要なものであり，p 値による解釈を行ううえでは必要ない.

図 1.4 では，帰無仮説「資格取得前後と 5 年後のスコアの差の平均値は 0 である(資格取得前後と 5 年後のスコアに変化がない)」に対する対応のある t 検定の結果として，3 種類の p 値が表示されている．それぞれの p 値の意味と解釈を示す．

p 値(Prob>|t|)	両側対立仮説(両側検定)という．資格取得前後と 5 年後のスコアの差の平均値は 0 でないことを示す場合に用いる．p 値が 0.05 を下回る(有意である)とき，「資格取得前後と 5 年後のスコアに変化がある」と解釈することができる．
p 値(Prob>t)	片側対立仮説(片側検定)という．　スコア差(5 年後 - 直後)の平均値は 0 よりも大きい場合に用いる．p 値が 0.05 を下回る(有意である)とき，「5 年後のスコアのほうが資格取得前後に比べて上昇する」と解釈することができる．
p 値(Prob<t)	片側対立仮説(片側検定)という．　スコア差(5 年後 - 直後)の平均値は 0 よりも小さい場合に用いる．p 値が 0.05 を下回る(有意である)とき，「5 年後のスコアのほうが資格取得前後に比べて減少する」と解釈することができる．

看護研究を含む観察研究において，片側対立仮説(片側検定)を用いることは殆どない．そのため，実際の利用では，「p 値(Prob>|t|)」のみを見ればよい．

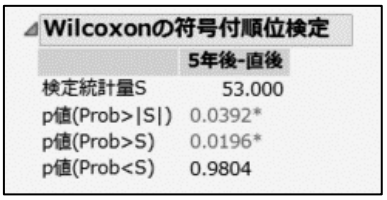
この事例では，p 値(Prob>|t|)が 0.0421 ことから，有意水準 α=0.05 のもとで有意である．すなわち，資格取得前後と 5 年後のスコアに変化があることが統計学的にも認められた．

(b) Wilcoxon 符号付順位検定

対応のある t 検定では，被験者毎のアウトカムの差(事例の場合には，資格取得 5 年後と取得直後でのスコアの変化の大きさ)を計算したうえで，その平均値が 0 であるか否かを評価している．すなわち，被験者毎のアウトカムの差が正規分布に従っていることが仮定される．一方で，1.3.5 節と同様の理由から，積分布に従っていない場合には，ノンパラメトリック検定の一つである，Wilcoxon 符号付順位検定(Wilcoxon 符号付き順位和検定)を利用しなければならない．

Wilcoxon 符号付順位検定の実行
1.　「対応のある t 検定」と同様の手順を用いて，図 1.4 の画面を表示させる．
2.　「対応のあるペア」の左側にある「▼」を左クリックし，「Wilcoxon 符号付順位検定」を選択する．

このとき，図 1.4 の結果の下部に以下が追加される．



Wilcoxonの符号付順位検定

	5年後-直後
検定統計量S	53.000
p値(Prob>|S|)	0.0392*
p値(Prob>S)	0.0196*
p値(Prob<S)	0.9804

ここでも，対応のある t 検定と同様に 3 種類の p 値が表示される．P 値(Prob>|S|)が両側検定での結果であり，この結果を解釈すればよい．この事例では，p 値(Prob>|S|)が 0.0392 ことから，有意水準 α=0.05 のもとで有意である．すなわち，資格取得前後と 5 年後のスコアに変化があることが統計学的にも認められた．

2 章 質的データにおける統計解析

2.1 クロス集計表に基づく統計解析

2.1.1 クロス集計表に基づく統計解析の基本

ここで用いるデータの略説を以下に示す.

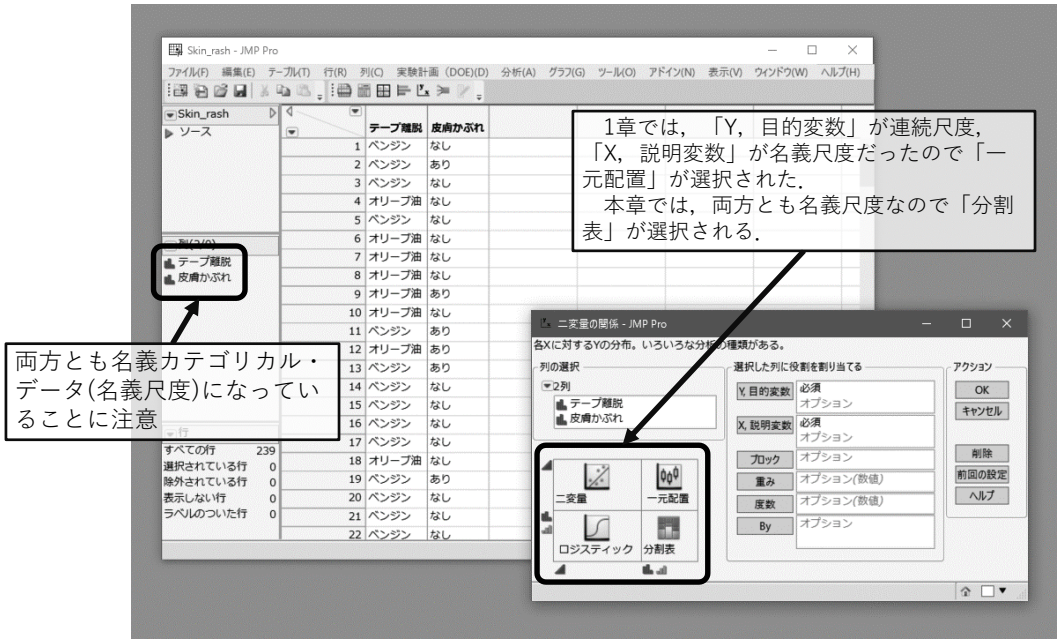
テープ剥離に用いるベンジンが皮膚かぶれの原因と考え，剥離材にオリーブ油を利用した．このとき，剥離剤(ベンジン・オリーブ油)と皮膚かぶれの発現の有無には違いがあるだろうか.

このデータは,「Skin_rash.csv」に保存されている．このファイルを Excel で開いたものが以下である.

	A	B	C
1	テープ離脱	皮膚かぶれ	
2	ベンジン	なし	
3	ベンジン	あり	
4	ベンジン	なし	
5	オリーブ油	なし	

最初の列(1 列目)には,「テープ離脱」として，剥離剤の種類(ベンジン，オリーブ油)が記載されており，2 列目には「皮膚かぶれ」の有無(あり，なし)が記載されている．なお，このデータの JMP での読み込み方法は，1 章と同じなので割愛する.

このデータを読み込んだときの画面とその説明を以下に示す.



このデータで読み込まれた 2 個の変数(項目)のデータは，いずれも文字(数値でない)である．JMP では，文字をデータとして読み込んだ場合に，自動的に名義カテゴリカル・データ(名義尺度)として認識される.

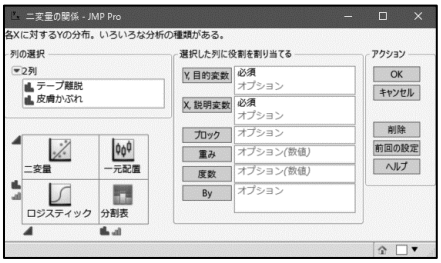
質的データの解析は，クロス集計表(分割表)に基づく解析になる．JMP では，1 章と同様に「2 変量の関係」で解析できる．上図の右下の「2 変量の関係」のメニュー画面の括弧で囲まれた部分は，変数の形式と，そのときに解析される統計手法が記載されている．これらを要約すると，次のようになる．

Y，目的変数	X，説明変数	選ばれる手法
連続尺度	連続尺度	二変量：単回帰分析，相関分析(4 章)
連続尺度	名義・順序尺度	一元配置：2 標本検定等(1 章)，分散分析(3 章),
名義・順序尺度	連続尺度	ロジスティック回帰(本セミナーでは扱わない)
名義・順序尺度	名義・順序尺度	分割表：クロス集計表およびその検定(本章)

JMP によるクロス集計表及び検定の方法を以下のとおりである．

クロス集計表及び検定の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される).



2. 「Y，目的変数」に「皮膚かぶれ」，「X，説明変数」に「テープ離脱」を割り当てる.
3. 「OK」ボタンを押す.

このときの結果を図 2.1 に示す．(a)はモザイクプロットの結果である．オリーブ油を剥離剤に用いた群よりも，ベンジンを剥離剤に用いた群のほうが，皮膚かぶれになっている割合が高いことが示唆される．

(b)はクロス集計表の結果である．全体パーセント，列パーセント，行パーセントの解釈は，次のとおりである．

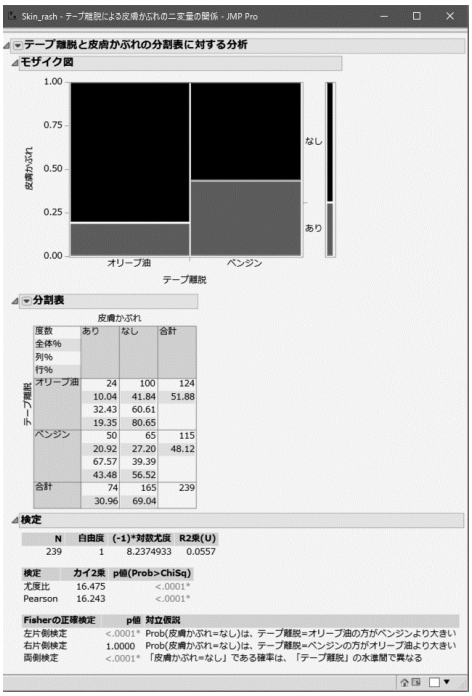
・全体パーセント：被験者全体のなかで，何パーセントの被験者が各セルに属しているかを表す．例えば，左上のセルの場合には，「オリーブ油を剥離剤に用いて，かつ皮膚かぶれになった割合は，10.04%である」と解釈される．

・列パーセント：皮膚かぶれの有無で分けたときの，それぞれの剥離剤の割合を表している(縦方向に 100%になるように計算している)．例えば，左上のセルの場合には，「皮膚かぶれになった被験者のうち，32.43%がオリーブ油を用いた」と解釈される．

・行パーセント：剥離剤の種類で分けたときの，皮膚かぶれの有無の割合を表している(横方向に 100%になるように計算している)．例えば，左上のセルの場合には，「オリーブ油を剥離剤に利用した被験者のうち，19.35%が皮膚かぶれになった」と解釈される．

行パーセントより，オリーブ油を剥離剤に用いたときの皮膚かぶれの割合は，19.35%であるのに対して，ベンジンの割合は，43.48%であることから，ベンジンのほうが皮膚かぶれの割合が多かった．

(c)は，2 種類の検定結果である．「尤度比」とは尤度比検定の結果であり，「Pearson」とはカイ 2 乗検定の結果である．いずれの場合にも，帰無仮説は「剥離剤の種類と皮膚かぶれの有無には関連性がない



(a) モザイクプロット

・モザイクプロットの横幅は，説明変数(剥離剤の種類)の被験者数(標本サイズ)に対応している（このデータでは，被験者数がほぼ同じであるため，違いを把握しにくい）．
・縦方向の割合は，各群の応答変数(皮膚かぶれの有無)の構成割合を表している．

(b) クロス集計表

・JMPでのクロス集計表は，各セルが度数(被験者数)，全体パーセント，列パーセント，行パーセントの順に構成されている．

(c) クロス集計表に対する検定の結果

・カイ2乗検定(Pearson)，尤度比検定では，p値の計算に近似式を用いている．その近似の方法によって2種類が存在する．通常，カイ2乗検定と言われるもののp値は，「Pearson」である．

(d) Fisherの正確検定の結果

・近似を用いずに検定する方法がFisherの正確検定であり，被験者数が少ない場合には，カイ2乗検定よりも正確である．

図 2.1：クロス集計表及び検定の結果

(剥離剤の種類によって皮膚かぶれの有無に違いがない)」に対して，対立仮説「剥離剤の種類と皮膚かぶれの有無には関連性がある(剥離剤の種類によって皮膚かぶれの有無に違いがある)」を検定する．p 値が若干異なるが，一般的に用いられているのは，カイ 2 乗検定であることから，「Pearson」での p 値を用いることが推奨される．

その結果，p 値は<0.0001(0.0001 未満)であることから，剥離剤によって皮膚かぶれの有無に違いが認められる．また，行パーセント点より，オリーブ油を用いたほうが，ベンジンを用いるよりも皮膚かぶれが起きにくいことがわかる．

(d)は，Fisher の正確検定の結果である．カイ 2 乗検定では，p 値の計算に近似式を使うため，被験者数が少ない場合には，精度の悪い p 値を計算する恐れがある．カイ 2 乗検定を適用可能な症例数のカットオフ値が明確に定義されているわけではないが，概ね 100 以上が必要だと考えられる．

また，カイ 2 乗検定では，片側対立仮説が存在しないのに対して，Fisher の正確検定では，片側対立仮説のもとでの p 値を計算することができる．それぞれの対立仮説の意味は，画面上に提示されている通りである．他方，通常は両側検定での p 値のみを解釈する．その結果，p 値は<0.0001(0.0001 未満)であることから，カイ 2 乗検定と同様の結果が与えられている．

メモ 2-1：カイ 2 乗検定
カイ 2 乗検定は，2 個のカテゴリカル変数の関連性を評価するための検定方法である．クロス集計表の目標には，2 種類が存在する． 連関関係の評価：お酒の嗜好(飲む，飲まない)と喫煙の嗜好(喫煙している，喫煙していない)に関する調査の場合，お酒を飲むことで喫煙者になるとも考えられるし，喫煙者だからお酒を飲むようになったとも考えらえる．このような評価の場合には，酒と煙草の嗜好の関連性(連関関係)を評価することになる．

因果関係の評価：2 種類の薬(薬 A，薬 B)と治療効果(効果あり，効果なし)に関する研究の場合，薬の違い(原因)によって治療効果に違いがある(結果)ことを評価する．このような，原因と結果の関係性が成り立つことを因果関係という．

カイ 2 乗検定は，いずれの場合にも適用できるが，2 つの目標のどちらにも共通することは，「2 個のカテゴリカル変数が関係ないものであるか(独立であるか)」ということを否定できればよいことになる．つまり，カイ 2 乗検定の帰無仮説は「2 変数は独立である」に対して，対立仮説「2 変数は独立でない」を検定することと同じ意味である．カイ 2 乗検定のことを独立性の検定と呼ぶこともある．

2.1.2 リスク比とオッズ比

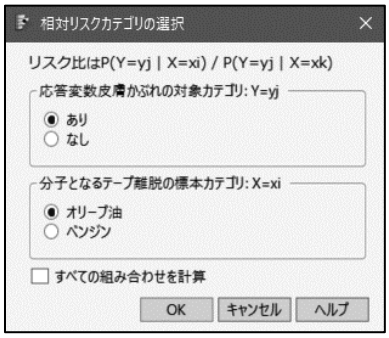
皮膚かぶれのデータにおいて，ベンジンを用いることはオリーブ油を用いるよりも，何倍ぐらい皮膚かぶれのリスクがあるのだろうか．このような場合に用いる測度がリスク比とオッズ比である．

(1) リスク比の計算

ここでは，リスク比の計算方法について述べる．

リスク比(相対リスク)の計算

1. 「クロス集計表及び検定の実行」と同じ手順を行い．図 2.1 を表示させる．
2. 「▼」を左クリックし，「相対リスク」を選択する(下記のような画面が表示される)．



3. 「目的変数の対象カテゴリ」を「あり」，「分子となる標本カテゴリ」を「オリーブ油」とする．
4. 「OK」ボタンを押す．

この手順を実施すると，図 2.1 のアウトプットの下部に以下が追記される．

相対リスク

説明	相対リスク	下側95%	上側95%
P(あり|オリーブ油)/P(あり|ベンジン)	0.445161	0.293858	0.674368

したがって，オリーブ油を利用することで，ベンジンに比べて，皮膚かぶれは，約 0.45 倍(おおよそ半分未満)になることがわかる．なお，「下側 95%」と「上側 95%」は，それぞれリスク比に対する 95%信頼区間の信頼限界値である．また，95%信頼区間が 1.00(オリーブ油とベンジンで皮膚かぶれの罹患が同じ)を含んでいないことから，オリーブ油の皮膚かぶれに対する罹患リスクはベンジンに比べて有意に小さいと言える．

(2) オッズ比の計算

次に，オッズ比を計算する．

オッズ比の計算

1. 「クロス集計表及び検定の実行」と同じ手順を行い．図 2.1 を表示させる．
2. 「▼」を左クリックし，「オッズ比」を選択する．

この手順を実施すると，図 2.1 のアウトプットの下部に以下が追記される．

オッズ比		
オッズ比	下側95%	上側95%
0.312	0.17499	0.556285

したがって，オリーブ油を利用することで，ベンジンに比べて，皮膚かぶれは，約 0.312 倍(おおよそ半分未満)になることがわかる．なお，「下側 95%」と「上側 95%」は，それぞれオッズ比に対する 95%信頼区間の信頼限界値である．また，95%信頼区間が 1.00(オリーブ油とベンジンで皮膚かぶれの罹患が同じ)を含んでいないことから，オリーブ油の皮膚かぶれに対する罹患リスクはベンジンに比べて有意に小さいと言える．因みに，オッズ比のほうが，リスク比に比べて 1.00(オリーブ油とベンジンで皮膚かぶれの罹患が同じ)よりも離れた値を示す．

メモ 2-2：リスク比とオッズ比

リスク比(相対リスクと呼ぶこともある)とオッズ比は，それぞれリスク，オッズの 2 群間の比によって計算される．いま，関心のあるイベント(事例では皮膚かぶれ)に対するリスクおよびオッズの定義は，

リスク ＝(関心のイベントが起きた被験者数) ÷ (被験者数)

オッズ比 ＝(関心のあるイベントが起きた割合) ÷ (関心のあるイベントが起きなかった割合)

で定義される．皮膚かぶれの事例におけるリスク比，オッズ比は

リスク比 ＝(オリーブ油でのリスク) ÷ (ベンジンでのリスク)

オッズ比 ＝(オリーブ油でのオッズ) ÷ (ベンジンでのオッズ)

である．いずれの測度でも，「オリーブ油はベンジンに比べて○○倍ほど皮膚かぶれになる」と解釈される．

リスク比のほうが解釈しやすそうだが，その利用はコホート研究に限定される．疫学研究の縦断研究には，コホート研究とケース・コントロール研究がある．いま，肺癌と喫煙習慣の関係を調査したいと考える．コホート研究の場合には，喫煙習慣のある被験者と喫煙習慣のない被験者に分けて，その後の経過を追跡し，肺癌に罹患したかどうかを調査する(原因で群分けを行い，その後の経過(結果の有無)を調査する)．ケースコントロール研究では，肺癌に罹患した被験者とそうでない被験者のデータを集め，喫煙習慣がなかったかを調査する(結果で群分けを行い，原因の有無を調査する)．リスクの定義をみればわかるように，リスクの計算には被験者数が必要になる．コホート研究では，原因をもとに被験者を集めるが，ケース・コントロール研究では，結果をもとに被験者を集める．そのため，被験者数の適切な集計を行うことができない．

また，リスク比には，数学的な問題もある．一つは，リスク比の場合には，0 倍～有限倍(分母のリスクで上限が決まる)までしか定義域がないのに対して，オッズ比の場合には，0 倍～∞倍まで定義可能である．もう一つは，ラベル付けの問題である．皮膚かぶれの事例では，関心のあるイベントを「皮膚かぶれの罹患」であったが，これを逆にして「皮膚かぶれの非罹患」とすることもできる．そうなると，「○○倍ほど皮膚かぶれになる」といっていたものが「1/○○倍ほど皮膚かぶれになる」と逆数での解釈になるはずだが，リスク比ではこのような数値にはならない．

さらに，今回は解説しないがロジスティック回帰分析では，オッズ比による解釈をおこなうため，最近では，研究の形式に依らず，オッズ比を用いることが多くなってきている．

JMP では，関心がある事象を自動的に判断するが，本来知りたい事象が逆になっている場合には，オッズ比の逆数をとればよい．例えば，喫煙者は非喫煙者に比べて肺癌になるリスクが高いことを知りたいのに，JMP では「肺癌でない」のほうを関心のある事象として計算しているとする．その結果が，0.5 だったとすると，1/0.5=2.0 倍が肺癌になるリスクでのオッズ比になる．

なお，JMP では，数値の場合には 0 に近い値，文字の場合には A に近い文字のほうを関心のある事象と自動認識する(グループ(説明変数)についても同様である)．

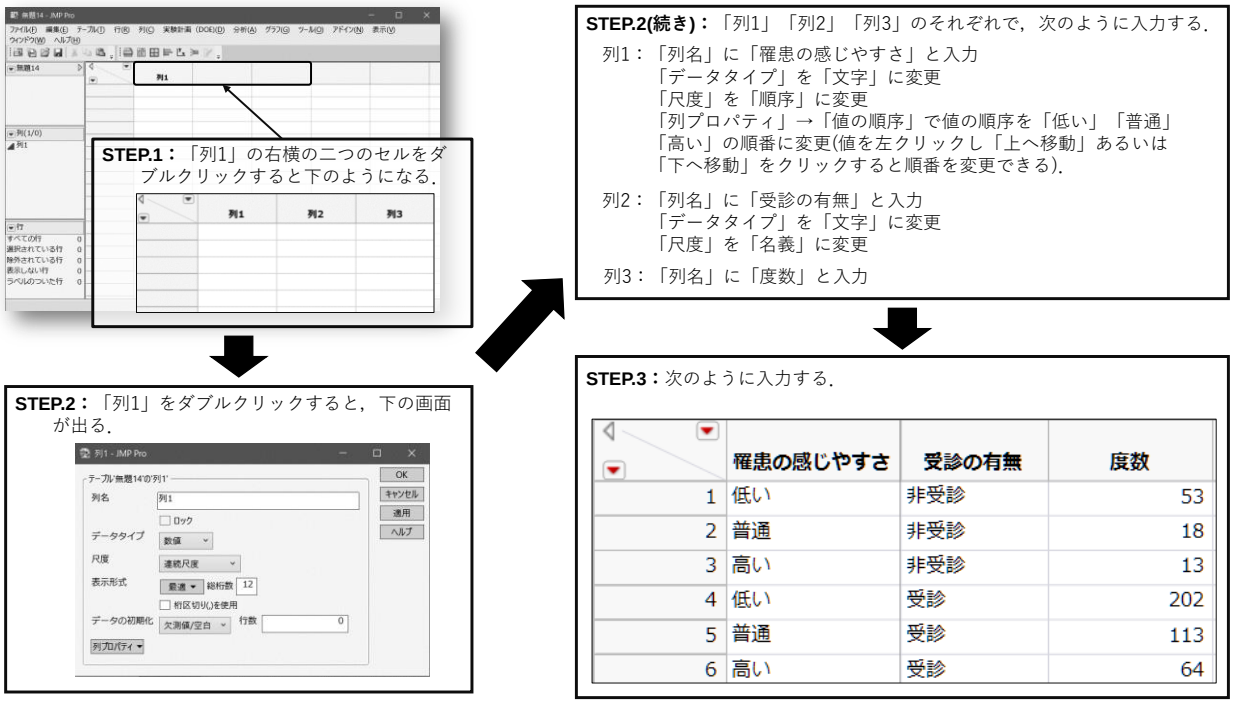
2.1.3 クロス集計表による集計結果を直接入力する場合

2.1.1 節では，生データからクロス集計表を作成する場合について述べた．ここでは，既にクロス集計表による集計が終了した場合について述べる．

アメリカのある地域では，遺伝性の特異的な疾患がある．この研究では，当該地域の住民 500 名を対象に，健康集団検査の受診の有無と特異的疾患に罹患しやすいと感じているか否か(低い，普通，高い)を調査している．その目的は，特異的疾患に罹患しやすいと感じている人ほど，健診に参加するという知見を得ることにある．　(Brown & Hollander, 1977).

	罹患しやすいと感じている度合		
受診の有無	低い	普通	高い
非受診	53	18	13
受診	202	113	64

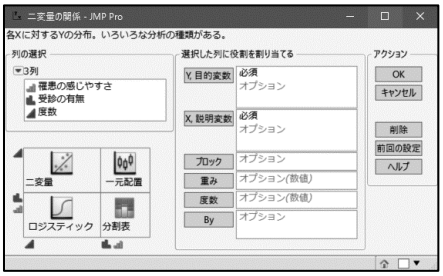
ここでは JMP に直接入力する方法について説明する．説明は以下のとおりである．



次に，クロス集計表及び検定を実行する．

集計データに基づくクロス集計表及び検定の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される).



2. 「Y, 目的変数」に「受診の有無」,「X, 説明変数」に「罹患の感じやすさ」,「度数」に「度数」を割り当てる.
3. 「OK」ボタンを押す.
4. 2×2 クロス集計表[8]以外では, Fisher の正確検定の結果が自動的に表示されない. そのため, クロス集計表の画面のなかの, タイトル「罹患の感じやすさと受診の有無の分割表に対する分析」左側の「▼」を左クリックし,「Fisher の正確検定」を選択する.

このときの結果を図 2.2 に示す．(a)のモザイクプロットより，罹患の感じやすさに対して，受診の割合が変化している傾向は認められない．このことは，(b)のクロス集計表における行パーセント(高い：83.12%，普通：86.26%，低い：79.22%)において，大きな差異がみとめられないことからもわかる．

ここでの検定では，帰無仮説「罹患の感じやすさと健康集団検診の受診には関連性がない」に対して対立仮説「罹患の感じやすさと健康集団検診の受診には関連性がある」を評価している．その結果，(c)

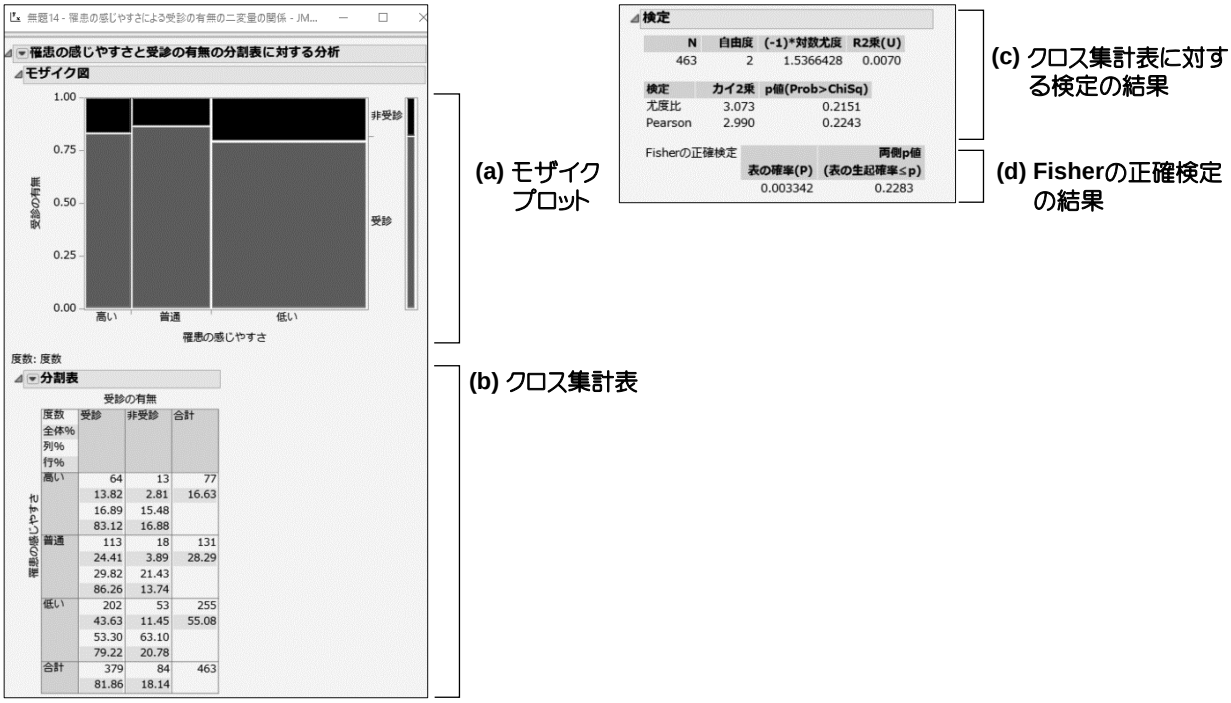


図 2.2：集計データに基づくクロス集計表及び検定の結果

[8] クロス集計表では，「列の数」×「行の数」クロス集計表(分割表)と呼ぶことがある．2.1.1 節の場合には，2 列×2 行なので 2×2 クロス集計表であり，2.1.3 節の場合には 3×2 クロス集計表である．

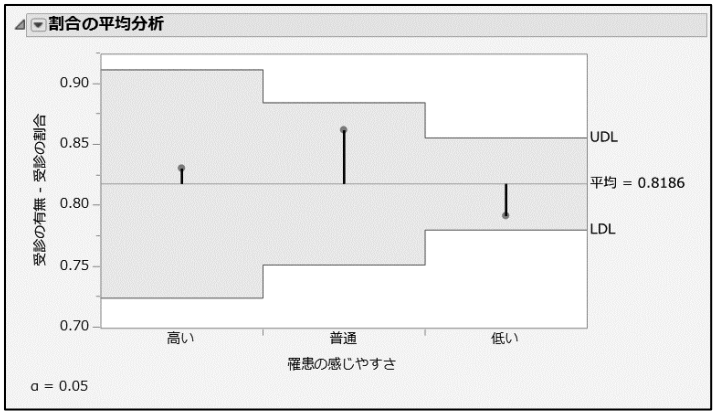
の「Pearson」(カイ 2 乗検定)の p 値が 0.2243 であり，有意水準 0.05 のもとで有意でない．また，(d)の Fisher の正確検定においても p 値が 0.2283 である．したがって，罹患の感じやすさと健康集団検診には，関連性が認められなかった．

なお，2×2 クロス集計表以外の Fisher の正確検定には片側対立仮説は存在しない(説明変数に対する目的変数の発現割合に大小関係が定義できないため[9])．また，「表の確率(P)」という項目が存在するが，この確率を統計学的に解釈することは困難なため，無視してよい．

さらに，説明変数(罹患の感じやすさ)に対する応答(受診の有無)の割合をグラフで表すことができる．

割合の平均値分析の計算
1.　「集計データに基づくクロス集計表及び検定の実行」と同じ手順を行い．図 2.2 を表示させる．
2.　「▼」を左クリックし，「割合の平均分析」を選択する．

上記を実行すると，次のようなグラフが追記される．



このグラフの縦軸は受診率であり，横線は全体での受診率(平均受診率)を表している．また，UDL および LDL で囲まれた範囲(色が塗られている部分)は，罹患の感じやすさの各カテゴリ(高い，普通，低い)における受診率が平均受診率と同じであるとしたときの範囲である．いずれのカテゴリでも，これらの範囲から外れておらず，このことからもカイ 2 乗検定および Fisher の正確検定において有意でないことが伺える．

2.2 対応のあるクロス集計表に基づく統計解析

質的データの場合においても，量的データと同様に対応がある場合が存在する．その例を以下に示す．

65 歳以上の高齢者を対象に，転倒予防訓練と運動機能の低下の有無に関する研究が実施された．この研究では，200 人の被験者に対して，転倒予防訓練前に運動機能検査を行い，3 カ月の転倒予防訓練後に同様の検査を実施している．ここでの目標は，転倒予防訓練後に運動機能の低下が改善していることを確認することにある．

このデータは，「fall_risk.csv」に保存されている．このファイルを Excel で開いたものが以下である．

[9] 罹患の感じやすさが高くなるほど受診率が増加する(あるいは減少する)といった説明変数(順序カテゴリカルデータ)の変化に対する目的変数の割合の変化を検定する方法に Cochran-Armitage 検定というものがあり，「▼」→「Cochran-Armitage の傾向性検定」から実行できる．

	A	B
1	訓練前	訓練後
2	低下なし	低下あり
3	低下あり	低下なし
4	低下あり	低下なし
5	低下なし	低下あり
6	低下あり	低下なし

最初の列(1 列目)には，「訓練前」として，転倒予防訓練前の運動機能検査での低下の有無が記載されており，2 番目の列には「訓練後」として，転倒予防訓練 3 カ月後の運動機能検査での低下の有無が記載されている．なお，このデータの JMP での読み込み方法は割愛する．

対応のあるクロス集計表の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される).

2. 「Y，目的変数」に「訓練後」，「X，説明変数」に「訓練前」を割り当てる.
3. 「OK」ボタンを押す.
4. タイトル「訓練前と訓練後の分割表に対する分析」左側の「▼」を左クリックし，「一致性の統計量」を選択する.

このときの結果を図 2.3 に示す．(a)および(c)は対応がない場合(2.1 節のようなデータの場合)に用いる結果であり，今回は必要ない情報である(無視しても構わない).

(b)は対応のあるクロス集計表である．対応のあるクロス集計表では，全体パーセントを利用して解釈を行う．すなわち，

(1) 訓練前，訓練後ともに運動機能が低下している割合は 20%

(2) 訓練前は運動機能が低下していたが，訓練後に低下が認められなかった割合は 30%

(3) 訓練前は運動機能の低下が認められなかったが，訓練後に低下していた割合は 15%

(4) 訓練前，訓練後ともに運動機能の低下が認められなかった割合は 35%

である．つまり，(1)(4)が転倒予防訓練で変化が認められなかった割合，(2)が転倒予防訓練の効果が認められた割合，(3)が転倒予防訓練後に運動機能が悪化した割合(転倒予防訓練が悪影響を及ぼしたのか，時間経過が影響したのかは不明)と解釈できる．

(d)は対応があるクロス集計表の結果である．カッパ係数は，訓練前後で運動機能の低下に変化が認められなかったことを表す測度(言いかえれば，訓練前後の運動機能低下の有無が一致しているか否かを表す測度)であるが，今回の目標は，訓練後に運動機能の低下が抑制できたか否かに関心があるため，訓練前後での運動機能の低下の有無の一致性には関心がない．

したがって，今回の場合には，McNemar 検定の結果を採用する．JMP では McNemar 検定を一般化した Bowker 検定が採用されているが，その結果は，McNemar 検定と一致する．その結果，p 値が 0.0016 なので，訓練前後で運動機能に変化が認められたと解釈することができる．

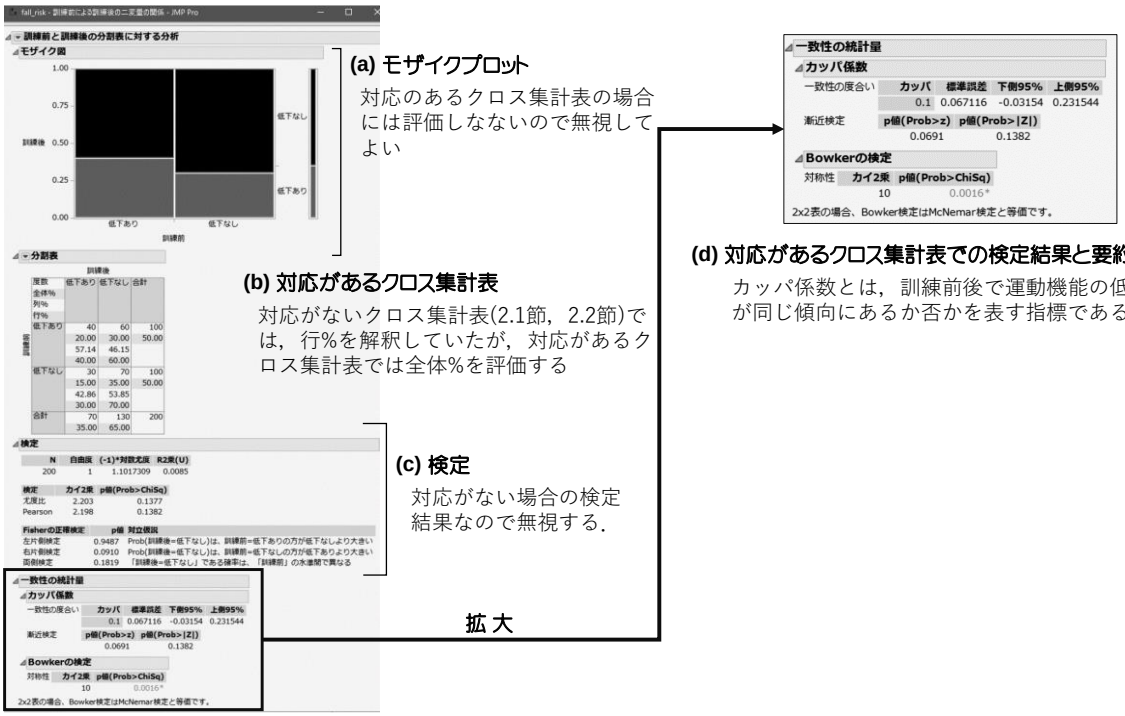


図 2.3：対応のあるクロス集計表及び検定の結果

メモ 2-3：カッパ係数と McNemar 検定

対応があるデータには，2 種類の考え方がある．

一致性を見る場合：例えば，検査 A と検査 B があったとする．検査 A は精度の高い検査方法であるが費用がかかり，検査 B は新しい検査方法で簡便に行えるとする．この 2 つの検査を同じ被験者に行った場合，次のような対応のあるクロス集計表が構成される．

		検査 B	
		陽性	陰性
検査 A	陽性	(a)	(b)
	陰性	(c)	(d)

この例の場合には，(a) 検査 A および検査 B ともに陽性，(d) 検査 A および検査 B ともに陰性の度数(被験者数)が診断結果が検査 A，検査 B で診断結果が異なる(b)(c)よりも大きくなることが期待される．

変化(違い)を見る場合：例えば，手術による不安感に関する調査を実施したとする．この調査では，医師からの手術に関する説明前に，手術に対する不安感(あり，なし)を質問したうえで，説明後に同じ質問を行い，医師の説明の適切性を検討している．この場合には，次のような対応のあるクロス集計表が作成される．

		説明後	
		あり	なし
説明前	あり	(a)	(b)
	なし	(c)	(d)

この例の場合には，説明前に不安ありだった患者が説明後に不安なしに変化することが期待される．

前者の場合には，カッパ係数を用いて評価を行い，後者の場合には McNemar 検定を用いて評価する．

カッパ係数

カッパ係数とは，対角(a)(d)の度数が高いか否かを評価する指標であり，-1 から 1 までの範囲をとる．カッパ係数が 1 に近づくほど(a)(d)の度数が高く(検査 A と検査 B の一致性が高い．つまり，同じ診断結果を示す)，カッパ係数が-1 に近づくほど(c)(b)の度数が高い(検査 A と検査 B で逆の診断結果が出る)．そして，0 に近づくほど偶然性(検査 A と検査 B は別物である)と解釈される．JMP における漸近検定では，帰無仮説「カッパ係数は 0 である」に対して両側対立仮説(Prob>|z|)「カッパ係数は 0 でない」，片側対立仮説(Prob>z) 「カッパ係数は 0 よりも大きい」を検定している．したがって，上記の検査の例では，1 に近づくほどポジティブな結論になる．

McNemar 検定

McNemar 検定では，説明前に不安ありの割合と説明後に不安ありの割合を比較する．つまり，帰無仮説「説明前に不安ありの割合と説明後に不安ありの割合」に対して，対立仮説「説明前に不安ありの割合」と「説明後に不安ありの割合は異なる」を検定している．したがって，上記の手術前説明の例の場合には，有意であれば説明前後で被験者の手術に対する不安意識に変化がみられると解釈できる．

2.3 カテゴリ数が多い場合の工夫：対応分析

クロス集計表では，2 変数のカテゴリが増加するほど解釈が困難になる傾向がある．このような場合には，対応分析(コレスポンデンス分析)を用いてグラフで解釈することが推奨される．

下表は，ある病院の 3 病棟(A，B，C)に入院している患者の食事形態を調査した結果である(内田他，2012)．ここでの関心は，病棟によって食事形態に違いがあるか否かを検討することにある．

病棟	食事形態				
	普通食	カロリー制限食	塩分制限食	軟菜食	嚥下食
A 病棟	18	9	6	14	5
B 病棟	11	7	5	21	7
C 病棟	6	11	11	12	5

データの入力は，2.1.3 節と同様に直接データを入力する場合を考える．ここでは入力の流れを略説する(2.1.3 節の図示を参考)．

STEP.1：「列 1」の右横の二つのセルをダブルクリックして，「列 2」「列 3」を作成する．

STEP.2：「列 1」「列 2」「列 3」をダブルクリックしながら，次のように入力する．

列 1：「列名」に「病棟」と入力，「データタイプ」を「文字」に変更，「尺度」を「名義」に変更．

列 2 :「列名」に「食事形態」と入力,「データタイプ」を「文字」に変更,「尺度」を「名義」に変更.

列 3 :「列名」を「度数」に変更

STEP.3：次のように入力する.

	病棟	食事形態	度数
1	A	普通食	18
2	A	カロリー制限食	9
3	A	塩分制限食	6
4	A	軟菜食	14
5	A	嚥下食	5
6	B	普通食	11
7	B	カロリー制限食	7
8	B	塩分制限食	5
9	B	軟菜食	21
10	B	嚥下食	7
11	C	普通食	6
12	C	カロリー制限食	17
13	C	塩分制限食	11
14	C	軟菜食	12
15	C	嚥下食	5

上記のデータに対するクロス集計表及び検定の作成方法は，1.1.3 節と同様である(「分析」→「二変量の関係」を選択して二変量の関係画面を表示する．そして,「Y，目的変数」に「病棟」,「X，説明変数」に「食事形態」,「度数」に「度数」を入力して OK ボタンを押す).

分析結果のなかから，分割表および検定結果を抜粋して以下に示す.

分割表

病棟	度数 全体% 列% 行%	食事形態 カロリー制限食	塩分制限食	軟菜食	普通食	嚥下食	合計
A		9	6	14	18	5	52
		5.84	3.90	9.09	11.69	3.25	33.77
		27.27	27.27	29.79	51.43	29.41	
		17.31	11.54	26.92	34.62	9.62	
B		7	5	21	11	7	51
		4.55	3.25	13.64	7.14	4.55	33.12
		21.21	22.73	44.68	31.43	41.18	
		13.73	9.80	41.18	21.57	13.73	
C		17	11	12	6	5	51
		11.04	7.14	7.79	3.90	3.25	33.12
		51.52	50.00	25.53	17.14	29.41	
		33.33	21.57	23.53	11.76	9.80	
合計		33	22	47	35	17	154
		21.43	14.29	30.52	22.73	11.04	

検定

N	自由度	(-1)*対数尤度	R2乗(U)
154	8	8.5394968	0.0358

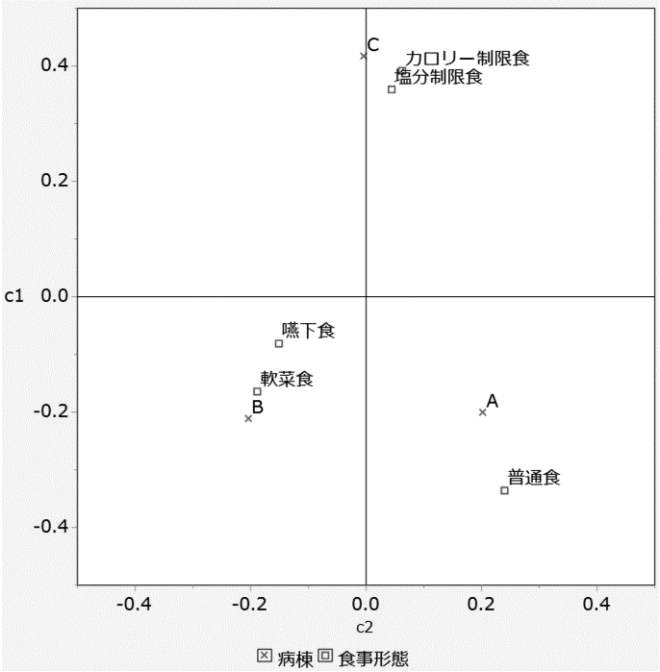
検定	カイ2乗	p値(Prob>ChiSq)
尤度比	17.079	0.0293*
Pearson	17.477	0.0255*

クロス集計表の結果より，病棟 A は普通食が多く(行パーセント：34.62%)，病棟 B は軟菜食(行パーセント：41.18%)が多く，病棟 C はカロリー制限食が多い(行パーセント：33.3%)ことが伺える．また，カイ 2 乗検定(Pearson)が 0.0255 であることから，有意水準 α=0.05 よりも小さい．したがって，病棟と食事形態には関連性があることが統計学的に示された．

一方で，食事形態が 5 カテゴリ，病棟が 3 カテゴリあることから，その全体の傾向を解釈することは難しい．クロス集計表のカテゴリ間の関係の近さをグラフで表す方法が対応分析である．

対応分析の計算
1.　クロス集計表を作成する(上述した方法).
2.　「▼」を左クリックし,「対応分析」を選択する.

これにより，下図が追記される．



対応分析では，病棟(A，B，C)及び食事形態(普通食，カロリー制限食，塩分制限食，軟菜食，嚥下食)の両方についてプロットされる．このプロットの見方は，距離が近いほど関係性が強く，一方で，遠いほど関係性が薄いと解釈できる．また，原点(0,0)に近づくほど平均的な傾向(特徴がない)ことを表す．

3 箇所の病棟が大きく散らばっていることから，病棟間で食事形態には違いが示唆される．また，病棟 A と普通食が近いことから，病棟 A は普通食の傾向が強いことが分かる．また，病棟 B は軟菜食，嚥下食の傾向が強く，病棟 C はカロリー制限食，塩分制限食の傾向が強いことがわかる．

3 章 一元配置の分散分析と多重比較

3.1 一元配置の分散分析及び多重比較の概要

3.1.1 一元配置の分散分析

3 種類の薬剤に対する研究の例を挙げる．この研究では，被験者を 3 群に分け，それぞれに対して 3 種類の薬剤(薬剤 A，薬剤 B，薬剤 C)のいずれかを投与されている．このとき，3 剤の効果に違いがあるか否かを評価する方法が一元配置の分散分析(一元配置 ANOVA[ANalysis Of VAriance])である．一元配置の分散分析では，帰無仮説「3 剤の薬剤に対する効果の平均値がすべて等しい」に対して，対立仮説「帰無仮説ではない」が評価される．

図 3.1 は，一元配置の分散分析に対する概念図である．有意でないとき，3 剤の平均効果は同じぐらいである．したがって，平均効果のバラツキを考えると，そのバラツキは小さい．有意であるとき，3 剤のいずれかの平均効果に違いが生じる．その結果として，3 剤の平均効果のバラツキは大きくなる．分散分析では，平均効果のバラツキが個々の群のバラツキに比べて大きい場合に有意差が認められる(p 値が小さくなる).

3.1.2 多重比較

図 3.2 は，128 症例をランダムに 2 群に分け，同じ薬剤を投与する臨床試験をシミュレーションによって 200 回実施したときの試験番号(Trial Number)と 2 標本 t 検定の p 値(p-value)を表している(X 軸：試

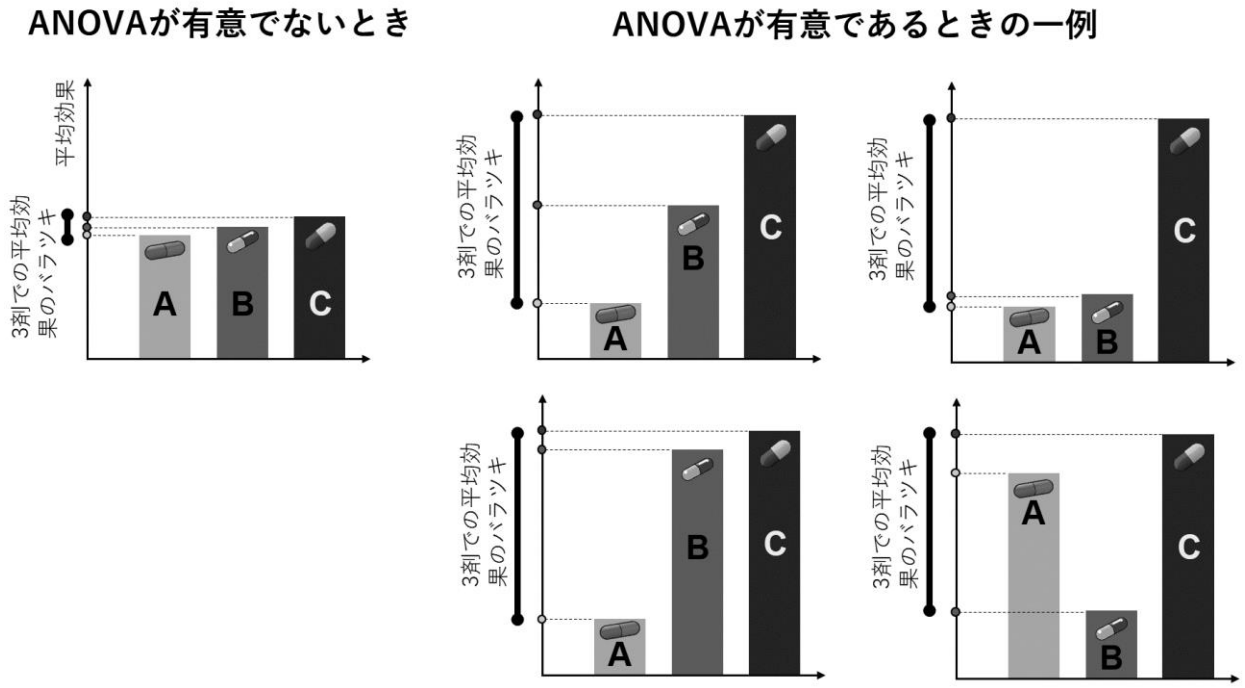


図 3.1：一元配置の分散分析における概念図

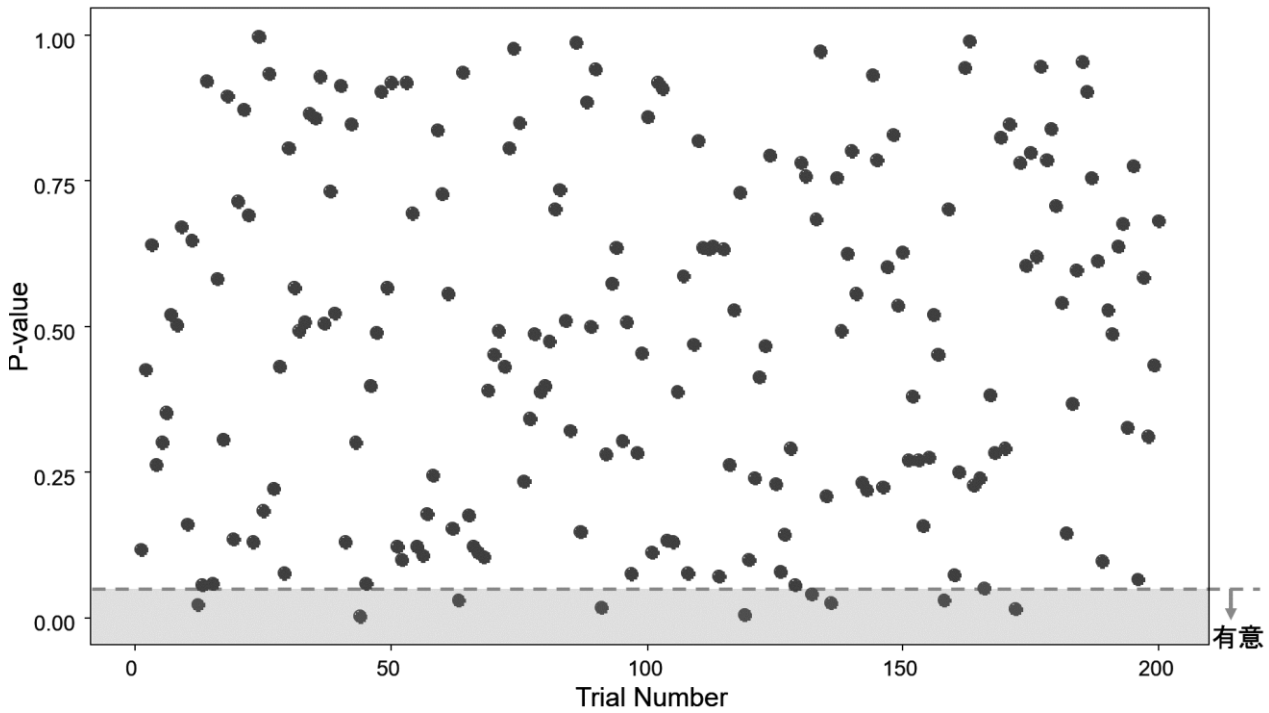


図 3.2：臨床試験のシミュレーションにおける結果

験番号，Y 軸：p 値)．ここで，横方向の点線は有意水準 0.05 を表している．2 群には同じ薬剤が投与されているので，本来は効果に違いがない．それにも関わらず，10 個の試験で有意差が認められている．

有意水準 0.05 は，帰無仮説が真実であったとしても，5%の確率で有意であると誤ってしまうことを意味する(そのため，有意水準 α は第 1 種の過誤あるいは α エラーと呼ばれる)．図 3.2 において，同じ薬剤を投与した臨床試験であるにも関わらず，5%の確率(10/200)で有意差が認められたのはそのためである．

先ほどの 3 剤の効果を比較するとき，全てのパターンで対比較するには，3 回の検定(A と B，A と C，B と C)が必要になる．この比較を有意水準 0.05 で検定した場合，3 剤における(真実の)平均効果が同じであったとしても，14.3%の確率でいずれかの検定が有意になる．つまり，もともとは有意水準 0.05 で比較していたとしても，「下手な鉄砲も数打てば当たる」効果で誤りの確率が増加している．このような状況に対処するための方法が多重比較である．

多重比較には，p 値を調整する方法と分散分析の結果を数理的に展開する方法がある．前者は，2 章までに述べた検定で得られた p 値を調整するだけなので，ざまざまな検定に適用することができる．後者の方法では，分散分析の結果と多重比較が対応付けられている方法や，あるいは特定のシチュエーションを想定した方法などが提案されているが量的データに対する方法に限定される．

3.2 JMP による分散分析と多重比較の実行

3.2.1 分散分析の実行

ここでは，疼痛患者に対する 3 種類の除痛薬のデータを用いる．

いま，14 名の疼痛患者をランダムに 3 群に分け，それぞれの群に除痛薬(A, B, C)を投与したときの投与後の痛みの程度を測定した．除痛薬によって痛みの程度に違いがあるだろうか．

このデータは，「Analgesics.csv」に保存されている．このファイルを Excel で開いたものが以下である．

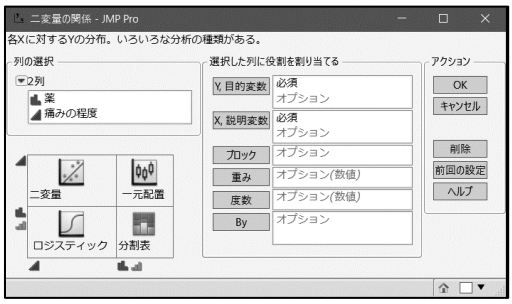
	A	B	C
1	薬	痛みの程度	
2	C	12.4	
3	A	7.69	
4	C	14	
5	A	9.69	
6	C	11.6	

最初の列(1 列目)には，「薬」として，除痛薬の種類(A,B,C)が記載されており，2 番目の列には「痛みの程度」として，投与後の痛みの程度が記載されている．なお，このデータの JMP での読み込み方法は割愛する．

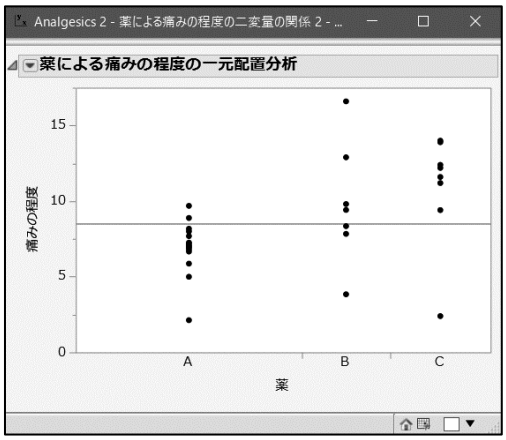
このとき，分散分析の実行方法を以下に示す．

一元配置の分散分析の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される)．



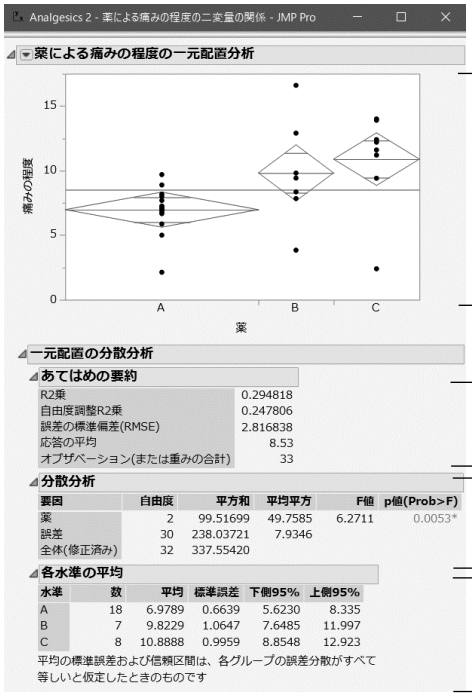
2. 「Y，目的変数」に「痛みの程度」，「X，説明変数」に「薬」を割り当てる．
3. 「OK」ボタンを押す(下図のような画面が表示される)．



4　「▼」ボタンを左クリックし，「平均/ANOVA」を選択する．

上記の手順を実行すると図 3.3 の結果が表示される．ここで，(a)は各群のデータ点と平均値を表している．菱形の横の長さは被験者数に対応している．薬 A の横幅が最も大きいのは，薬 A を投与された被験者数が最も多い(18 例)ためである．菱形の中央線は各群の平均値を表している．薬 A の痛みの程度が最も小さく，薬 C の痛みの程度が最も大きいことが示唆される．このことは，各群での平均値と 95%信頼区間の出力(e)でも明らかである．

(b)は分散分析での当てはめの要約である．分散分析は回帰分析(4 章)の一つと考えることができる．(b)は，分散分析の結果を回帰分析として考えたもとで，その当てはまりの良さを表している．ただし，その結果を解釈することは殆どないため，無視しても問題ない．



(a) 各群のデータ点と平均値
- データ点は，観測値を表している．
- 水平線は，全体平均を表している．
- 菱形の横の長さは被験者数(標本サイズ)に対応している．
- 菱形の中央線は各群の平均値を表している．
- 菱形の縦の長さは平均の95%信頼区間を表している．

(b) 分散分析でのあてはめの要約
分散分析は回帰分析(4章)の一つと考えることができる．当てはめの要約は，回帰分析における当てはまりの良さを意味するが，分散分析において解釈することは殆どない．

(c) 分散分析の結果
- このなかのp値が最も重要である．
- その他はp値を計算するのに用いるだけであり，無視

(d) 各群での平均値と95%信頼区間
- 各群での被験者数(数)，平均値，95%信頼区間(下側95%，上側95%)を表している．

図 3.3：一元配置の分散分析の結果

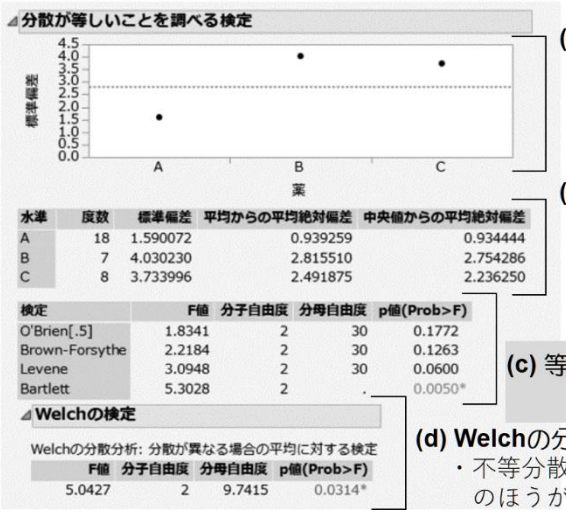
(c)は分散分析の結果であり，この表を分散分析表という．「自由度」，「平方和」，「平均平方」及び「F値」は p 値を計算する過程で必要になるが，結果の解釈のみを考えるのであれば，p 値のみを解釈すればよい．その結果，p 値は 0.0053 であり，有意水準 0.05 のもとで有意である．したがって，薬剤によって痛みの程度に違いがあることが統計学的に示された．

3.2.2 分散分析の仮定の確認：等分散性の検定

分散分析では，(2 標本 t 検定と同様に)すべての群の分散が等しいことが仮定される．これを等分散性という．等分散性の仮定が満たされない場合には，3.3 節の Kruskal-Wallis 検定を用いる必要がある．

等分散性の検定の実行
1.　「一元配置の分散分析の実行」の 1 から 3 までを行う．
2.　「▼」ボタンを左クリックし，「等分散性の検定」を選択する．

上記を実行すると次のような結果が得られる．



(a) 各群の標準偏差のグラフ
- 横線は全体での標準偏差を表している．
- 点は各群の標準偏差を表している．

(b) 標準偏差などのバラツキ
- 「度数」は被験者数を表している．
- 「平均からの平均絶対偏差」，「中央値からの平均絶対偏差」も「標準偏差」と同様にバラツキを表すが，あまり使わないので無視

(c) 等分散性を評価するための4種類の検定
- p値のみを見ればよい(検定の説明は本文記載)．

(d) Welchの分散分析の結果
- 不等分散の場合には，ノンパラメトリック検定(Kruskal-Wallis検定)のほうが良いため無視

(c)は等分散性(標準偏差を 2 乗した値が分散である)に対する 4 種類の検定結果(O'Brein 検定，Brown-Forstythe 検定，Levene 検定，Bartlett 検定) である．これらは，帰無仮説「すべての分散(バラツキ)が等しい」に対して，対立仮説「すべての分散が等しいわけではない」が検定される．

最も p 値が小さいのが Bartlett 検定である(p 値：0.0050)．Bartlett 検定はデータの正規分布からのずれが影響を及ぼすため，等分散性を適切に判断できない検定である．そのため，幾つかの文献では，基本的な等分散性の検定ということで Bartlett 検定の利用が説明されているものの，本来は推奨されない(Finney, 1995)．　Levene 検定も Bartlett 検定と同様に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場合には O'Brein 検定あるいは Brown-Forstythe 検定が推奨される．

そのため，今回は Levene 検定の結果を採用する．p 値は 0.060 であり，有意水準 0.05 を上回っており，有意でない．したがって，全ての分散が等しいわけではないという対立仮説を積極的には採用できないことから，分散分析の仮定(等分散性)を逸脱していないと考えられる．

3.2.3 多重比較の方法(パラメトリックの場合)

3.2.1 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．図 3.1 で示したように，分散分析で有意差が認められても，そのパターンは

A＝B≠C，A≠B＝C，A＝C≠B，A≠B≠C

が考えれれる．3 種類の薬剤のどれに違いがあるかを判断するために多重比較を用いることが多い．
JMP では，多重比較の方法として「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．

(1) **2 標本 t 検定**：「各ペア，Student の t 検定」は多重比較を行わずに全てのペアにおいて 2 標本 t 検定を行う場合である．

(2) **Tukey の HSD 検定**：「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．

(3) **Hsu の MCB 検定**：「最適値との比較(Hsu の MCB)」は任意の基準値(最適値)よりも各水準の平均値が高い(あるいは低い)か否かを明らかにする方法である．

(4) **Dunett の検定**：「コントロール群との比較(Dunnett)」は，群の一つが基準となっており (コントロール群)，コントロール群とその他の群の多重比較を行う場合に用いる．

今回は，すべてのペアでの多重比較なので，p 値を調整する方法として，「各ペア，Student の t 検定」の結果を用いる Bonferroni の多重比較法，及び，「Tukey の HSD 検定」について説明する．

<u>2 標本 t 検定に対する Bonferroni の多重比較調整</u>

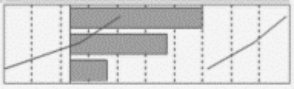
JMP では，Bonferroni の方法を自動的に実施することはできない．しかしながら，Bonferroni の方法では，2 標本 t 検定で得られた p 値を比較回数で掛けるだけなので非常に簡単である．先ず，「各ペア，Student の t 検定」を用いて，すべてのペアでの 2 標本 t 検定を実施する．

各ペアでの 2 標本 t 検定の実行
1.　「一元配置の分散分析の実行」の 1 から 3 までを行う.
2.　「▼」ボタンを左クリックし,「平均の比較」→「各ペア, Student の t 検定」を選択する.

上記を実施すると，図 3.3 の下部に新たに「平均の比較」の結果が得られる．ここでは，必要な部分のみを以下に示す.

差の順位レポート

水準	- 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
C	A	3.909861	1.196926	1.46541	6.354310	0.0027*
B	A	2.843968	1.254719	0.28149	5.406447	0.0308*
C	B	1.065893	1.457852	-1.91144	4.043224	0.4704



上図において「差」は(前の水準(薬) の平均値) - (後ろの水準(薬)の平均値)を表しており，「下側信頼限界」及び「上側信頼限界」は平均値の差の 95%信頼区間を表している.

次いで，得られた p 値を Bonferroni の方法に基づいて調整する(これを調整 p 値と呼ぶことがある).

- 薬 C 対 薬 A　0.0027 × 3 = 0.0081
- 薬 B 対 薬 A　0.0308 × 3 = 0.0924
- 薬 C 対 薬 B　0.4707 × 3 = 1.4121 → 1.0000 (p 値は 1 を超えないので調整時に 1 を上回った場合には 1 とする).

その結果，薬 A と薬 C の p 値のみが有意水準 0.05 を下回った．したがって，薬 A と薬 C の除痛効果には違いが認められた.

<u>Tukey の HSD による多重比較</u>

Bonferroni の方法は検定結果を比較回数で掛けるだけの単純な方法であることから，あらゆる検定(例えば，カイ 2 乗検定など)に応用可能である．一方で，有意になりにくい多重比較であることも広く知られている．その結果,「一元配置の分散分析で有意だった(3 剤の除痛効果は同じではない)だったのも関わらず，Bonferroni の多重比較を行うと，有意な組が見つからなかった(どの薬剤の対比較も有意でなかった)」ということが起こり得る.

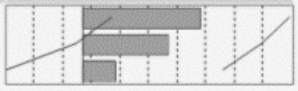
そのため，一元配置の分散分析で有意だった場合には，Tukey の HSD 検定(Tukey-Kramer の HSD 検定)を用いるのが一般的である．Tukey の HSD 検定は，一元配置の分散分析の結果を用いて多重比較を行うため，一元配置の分散分析で有意であれば「有意な組が必ず存在する」.

各ペアでの 2 標本 t 検定の実行
1.　「一元配置の分散分析の実行」の 1 から 3 までを行う.
2.　「▼」ボタンを左クリックし,「平均の比較」→「すべてのペア，Tukey の HSD 検定」を選択する.

上記を実施すると，図 3.3 の下部に新たに「平均の比較」の結果が得られる．ここでは，必要な部分のみを以下に示す.

差の順位レポート

水準	- 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
C	A	3.909861	1.196926	0.95904	6.860686	0.0075*
B	A	2.843968	1.254719	-0.24934	5.937274	0.0763
C	B	1.065893	1.457852	-2.52820	4.659989	0.7471



各ペアでの 2 標本 t 検定と同様の出力結果が与えられている．「差」および差の 95%信頼区間(「下側信頼限界」,「上側信頼限界」)は,「各ペア，Student の t 検定」と同じである．「p 値」は Tukey の HSD 検定での p 値になっている(Bonferroni の多重比較のように手計算を行う必要はない).

その結果，薬剤 A と薬剤 C において有意差が認められ，「差」が正値であることから薬剤 C のほうが薬剤 A に比べて痛みの程度が高かった．

3.3 3 群以上のデータに対するノンパラメトリック検定

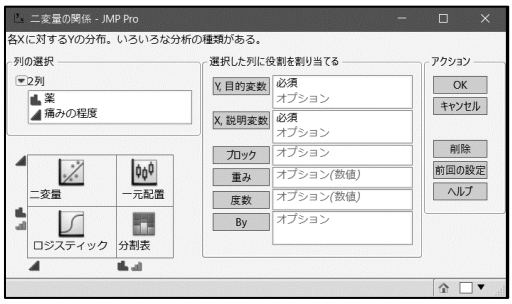
分散分析は，データが正規分布に従っており，かつバラツキ(標準偏差)が全ての群(グループ)で同じであると仮定できる場合に用いることができる．この過程を満たさない場合には，Kruskal-Wallis 検定(ノンパラメトリック検定の一つ)を用いることになる．つまり，パラメトリック検定において，2 群比較を実施する場合には 2 標本 t 検定，3 群以上の比較を実施する場合には一元配置の分散分析を用いるのに対して，ノンパラメトリック検定では，2 群比較を実施する場合には Wilcoxon 検定，3 群以上の比較を実施する場合には Kruskal-Wallis 検定を用いる．

3.3.1 Kruskal-Wallis 検定

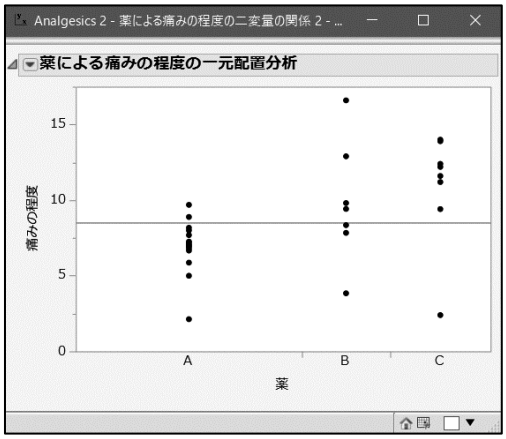
ここでは，先ほどの疼痛患者に対する除痛薬の例(Analgesics.csv)を用いて Kruskal-Wallis 検定を実施する．

Kruskal-Wallis 検定の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される)．



2. 「Y，目的変数」に「痛みの程度」，「X，説明変数」に「薬」を割り当てる．
3. 「OK」ボタンを押す(下図のような画面が表示される)．



4　「▼」ボタンを左クリックし，「ノンパラメトリック」→「Wilcoxon 検定」を選択する．

JMP では，2 群の場合(「X，説明変数」のカテゴリが 2 個の場合)には Wilcoxon 検定が実施され，3 群以上の場合(「X，説明変数」のカテゴリが 3 個以上の場合)には Kruskal-Wallis 検定が実施される．これは，Wilcoxon 検定を 3 群以上に拡張したものが Kruskal-Wallis 検定であることに由来する．

実行結果のなかで，Kruskal-Wallis 検定を実施することで追記された部分を以下に示す．

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
A	18	213.000	306.000	11.8333	-3.345
B	7	151.000	119.000	21.5714	1.387
C	8	197.000	136.000	24.6250	2.542

一元配置検定(カイ2乗近似)

カイ2乗	自由度	p値(Prob>ChiSq)
11.6801	2	0.0029*

上側の表(「スコア和」,「スコアの期待値」等が記載されている表)は，p 値を計算する過程で必要になるものの，それを解釈することは困難なので，無視してかまわない.

「一元配置検定(カイ２乗近似)」の「p 値(Prob>Chisq)」に記載されている p 値のみを解釈する．その結果，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.

3.3.2 多重比較の方法(ノンパラメトリックの場合)

JMP では，ノンパラメトリックの多重比較の方法として「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」,「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」がある.

(1) **Wilcoxon 検定**：「ペアごと Wilcoxon 検定」は多重比較を行わずに全てのペアにおいて Wilcoxon 検定を行う場合である.

(2) **Steel-Dwass 検定**：「すべてのペア Steel-Dwass 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．Tukey の HSD のノンパラメトリック版である.

(3) **Steel 検定**：「コントロール群との比較 Steel 検定」は群の一つが基準となっており (コントロール群)，コントロール群とその他の群の多重比較を行う場合に用いる．Dunett の多重比較のノンパラメトリック版である.

(4) **Dunn 検定**：Dunn 検定は，すべてのデータを併合したときの順位を用いて検定したうえで，Bonerroni 検定と同様に p 値を調整する方法である．JMP ではペアごとの比較をするすべてのペア 併合順位の Dunn 検定」及び，コントロール群と比較を行う「コントロール群との比較 併合順位の Dunn 検定」がある.

今回は，すべてのペアでの多重比較なので，p 値を調整する方法として，「ペアごと Wilcoxon 検定」の結果を用いる Bonferroni の多重比較法， Steel-Dwass 検定について説明する.

<u>(1) Wilcoxon 検定に対する Bonferroni の多重比較調整</u>

Wilcoxon 検定に対する Bonferroni の多重比較調整は，2 標本 t 検定の場合と同様に Wilcoxon 検定で得られた p 値を比較回数で掛けることで計算する．先ず，「ペアごと Wilcoxon 検定」を用いて，すべてのペアでの Wilccoxon 検定を実施する.

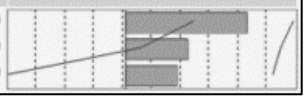
各ペアでの Wilcoxon 検定の実行
1. 「Kruskal-Wallis 検定の実行」の 1 から 3 までを行う.
2. 「▼」ボタンを左クリックし,「ノンパラメトリック」→「ノンパラメトリックな多重比較」→「ペアごと Wilcoxon 検定」を選択する.

上記を実施すると，「Wilcoxon 検定 ペアごとのノンパラメトリックな比較」の結果が得られる．ここでは，必要な部分のみを以下に示す．

Wilcoxon検定 ペアごとのノンパラメトリックな比較

q*	Alpha
1.95996	0.05

水準	- 水準	スコア平均の差	差の標準誤差	Z	p値	Hodges-Lehmann	下側信頼限界	上側信頼限界
C	A	9.659722	3.249444	2.972730	0.0030*	4.610000	2.60000	6.330000
B	A	7.638889	3.277685	2.330574	0.0198*	2.385000	0.58000	5.900000
C	B	1.741071	2.314550	0.752229	0.4519	1.985000	-4.40000	5.570000



上図において「Hodges-Lehmann」は群 1(左側の水準)と群 2（右側の水準）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ものであり，群間差の中央値(Hodges-Lehmann 推定量)である．「上側信頼限界」と「下側信頼限界」は，群間差の中央値に対する 95%信頼区間である．

次いで，得られた Wilcoxon 検定の p 値を Bonferroni の方法に基づいて調整する．

- 薬 C 対 薬 A　0.0030 × 3 = 0.0090
- 薬 B 対 薬 A　0.0198 × 3 = 0.0594
- 薬 C 対 薬 B　0.4519 × 3 = 1.3557 → 1.0000 (p 値は 1 を超えないので調整時に 1 を上回った場合には 1 とする).

その結果，薬 A と薬 C の p 値のみが有意水準 0.05 を下回った．したがって，薬 A と薬 C の除痛効果には違いが認められた．

(2) Stell-Dwass 検定による多重比較

Kruskal-Wallis で有意だった場合には，Steel-Dwass を用いるのが一般的である．Steel-Dwass 検定は，Kruskal-Wallis 検定で有意であれば「有意な組が必ず存在する」．すなわち，Steel-Dwass 検定とは，Tukey の HSD 検定のノンパラメトリック版と捉えることができる

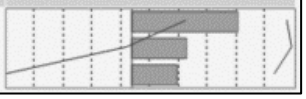
Stell-Dwass 検定の実行
1. 「一元配置の分散分析の実行」の 1 から 3 までを行う.
2. 「▼」ボタンを左クリックし,「ノンパラメトリック」→「ノンパラメトリックな多重比較」→「すべてのペア, Steel-Dwass 検定」を選択する.

上記を実施すると，「Steel-Dwass 検定 ペアごとのノンパラメトリックな比較」が得られる．ここでは，必要な部分のみを以下に示す．

Steel-Dwass検定 すべてのペアのノンパラメトリックな比較

q*	Alpha
2.34370	0.05

水準	- 水準	スコア平均の差	差の標準誤差	Z	p値	Hodges-Lehmann	下側信頼限界	上側信頼限界
C	A	9.659722	3.249444	2.972730	0.0083*	4.610000	2.31000	6.760000
B	A	7.638889	3.277685	2.330574	0.0517	2.385000	-0.27000	6.910000
C	B	1.741071	2.314550	0.752229	0.7323	1.985000	-5.44000	6.160000



各ペアでの Wilcoxon 検定と同様の出力結果が与えられている．　Steel-Dwass 検定での p 値において，薬 C と薬 A の p 値が 0.0083 であり有意だった．薬 C - 薬 A の差の中央値が 4.61 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．

4 章 相関分析・回帰分析

4.1 相関分析

4.1.1 相関分析の略説

相関関係とは，2 変数間の関連性を表す用語である．図 4.1 は相関関係を表す 3 種類の散布図である．図 4.1(a)は X(横軸)が増加するほど Y(縦軸)が増加している．このような状態を正の相関関係があるという．そして，図 4.1(b)は X(横軸)が増加するほど Y(縦軸)が減少している．このような状態を負の相関関係あるという．さらに，図 4.1(c)は X(横軸)が増加しても Y(縦軸)に変化はない．このような状態を無相関関係という．これらの相関関係を数値化したものが相関係数である．

相関係数には，次のような特徴がある．

・相関係数は，-1 から 1 までの範囲をとる．

・相関係数が正値の場合に正の相関関係があり，1 に近づくほど散布図のデータ点が比例(右肩上がり)の直線状に布置する(正の相関関係が強いと判断される).

・相関係数が負値の場合に負の相関関係があり，-1 に近づくほど散布図のデータ点が反比例(右肩下がり)の直線状に布置する(負の相関関係が強いと判断される).

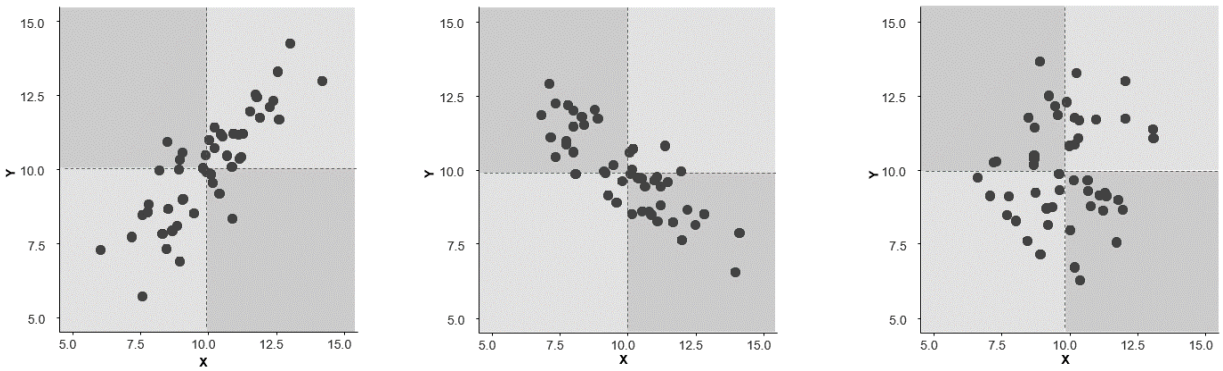
・相関係数が 0 に近づくほど無相関関係であることが示され，散布図のデータ点が一様に散らばる．

JMP では，3 種類の相関係数(相関係数(Pearson の相関係数)，Spearman の順位相関係数，Kendall の順位相関係数)が存在する．相関係数(Pearson の相関係数)は，最も一般的に用いられている相関係数であり，2 変数が正規分布に従っていることが仮定される．一方で，次のような場合には，順位相関係数を用いる．

(1) データが正規分布に従ってないとき．

(2) データに外れ値が存在するとき．

(3) データが順位尺度で得られていると．



(a) 正の相関関係　(b) 負の相関関係　(c) 無相関関係

図 4.1：相関関係の図示

Spearman の順位相関係数と Kendall の順位相関係数は，計算式が異なるので，得られる数値に若干の違いが認められるが，Spearman の順位相関係数を用いることが多いようである[10].

また，相関係数の大きさの目安であるが，諸説存在するものの，内田他(2013)では，次のように指摘されている.

相関係数の絶対値が 0.7 以上	強い相関関係がある.
相関係数の絶対値が 0.5 以上，0.7 未満	相関関係がある.
相関係数の絶対値が 0.3 以上，0.5 未満	弱い相関関係がある.
相関係数の絶対値が 0.3 未満	相関関係が殆どない.

4.1.2 JMP による相関分析

ここでは，コレステロール値と血圧(収縮期血圧)のデータを用いる.

健康診断を受診した 20 名の被験者のコレステロール値と収縮期血圧に相関関係があるか否かを分析する.

このデータは，「Col.csv」に保存されている．このファイルを Excel で開いたものが以下である.

	A	B
1	コレステロール	血圧
2	216	139
3	261	153
4	169	124
5	287	168
6	181	116

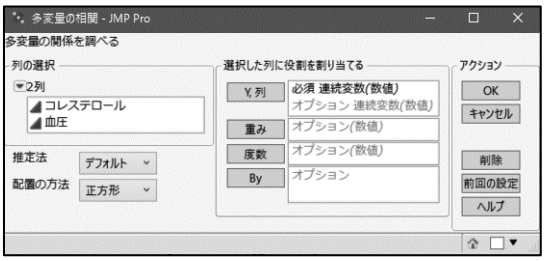
最初の列(1 列目)には，「コレステロール値」，2 番目の列には「血圧」として，収縮期血圧が記載されている．なお，このデータの JMP での読み込み方法は割愛する.

(1) 相関分析の実行

相関係数の計算方法を以下に示す.

相関分析の実行

1. 「分析」→「多変量」→「多変量の相関」を選択する(下図のような画面が表示される).



2. 「Y, 列」に「コレステロール」,「血圧」を割り当てる.
3. 「OK」ボタンを押す(下図のような画面が表示される).

上記を実行すると，相関係数(Pearson)の相関係数が図 4.2 のように表示される.

JMP の「多変量の相関」では，2 個以上の量的データに対する相関係数が全てのペアで計算される．そのため，表示形式が行列のようになっている．行列の「列 2 列目：血圧，行 1 列目：コレステロー

[10] Spearman の順位相関係数は，変数に順位付けしたうえで，相関係数を求めたものであり，意味が理解しやすいのが一つの理由である.

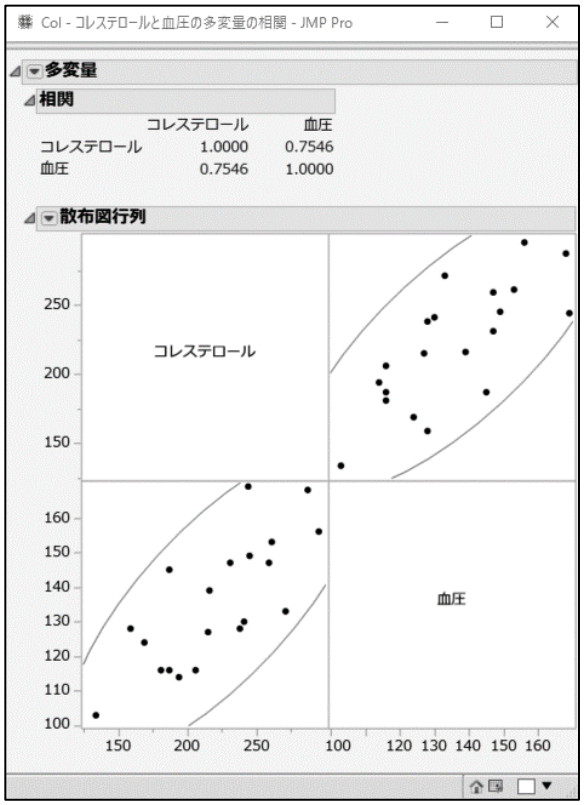


図 4.2：相関分析の結果

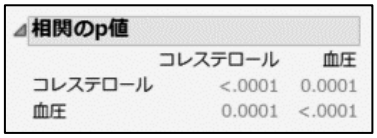
ル」の 0.7546 と「列 1 列目：コレステロール，行 2 列目：血圧」の 0.7546 は，いずれも血圧とコレステロールの相関係数を計算しており，数値は必ず同じになる．相関係数は 0.7546 なので，強い相関関係があることが示唆される．つまり，コレステロール値が増加するほど収縮期血圧も上昇する傾向にある．このことは，下側の散布図行列でも示唆される．

(2) 無相関性の検定

無相関性(2 変数に相関関係が無い)ことを検定により明らかにすることができる．これを無相関性の検定という．無相関性の検定では，帰無仮説「2 変数の相関係数は 0 である」に対して，両側対立仮説「2 変数の相関係数は 0 でない」が評価される．

無相関性の検定の実行
1.　「相関分析の実行」を行う.
2.　「▼」ボタンを左クリックし,「相関の p 値」を選択する.

このときの結果を以下に示す．



相関のp値

	コレステロール	血圧
コレステロール	<.0001	0.0001
血圧	0.0001	<.0001

p 値が<0.0001(0.0001 未満)であることから，有意水準 0.05 のもとで有意である．したがって，コレステロール値と収縮期血圧の相関係数は 0 でない(無相関関係でない)ことがわかった[11]．

[11] 「▼」ボタンを左クリックし，「相関の p 値」の代わりに「相関の信頼区間」をクリックすると，相関係数に対する 95%信頼区間を計算することができる．

(3) ノンパラメトリックな相関係数：Spearman の順位相関係数

2 変数のいずれか(あるいは両方)が正規分布に従わない場合，外れ値が存在する場合，あるいは順序尺度のデータの場合には，ノンパラメトリックな相関係数を用いたほうが良い．JMP では，Spearman の順位相関係数，Kendall の順位相関係数，Hoefdding の D 統計量の 3 種類が存在するが，ここでは，最も一般的に用いられる Spearman の順位相関係数について説明する．

無相関性の検定の実行
1. 「相関分析の実行」を行う.
2. 「▼」ボタンを左クリックし,「ノンパラメトリック相関係数」→「Spearman の順位相関係数」を選択する.

このときの結果を以下に示す．

ノンパラメトリック: Spearmanの順位相関係数(ρ)

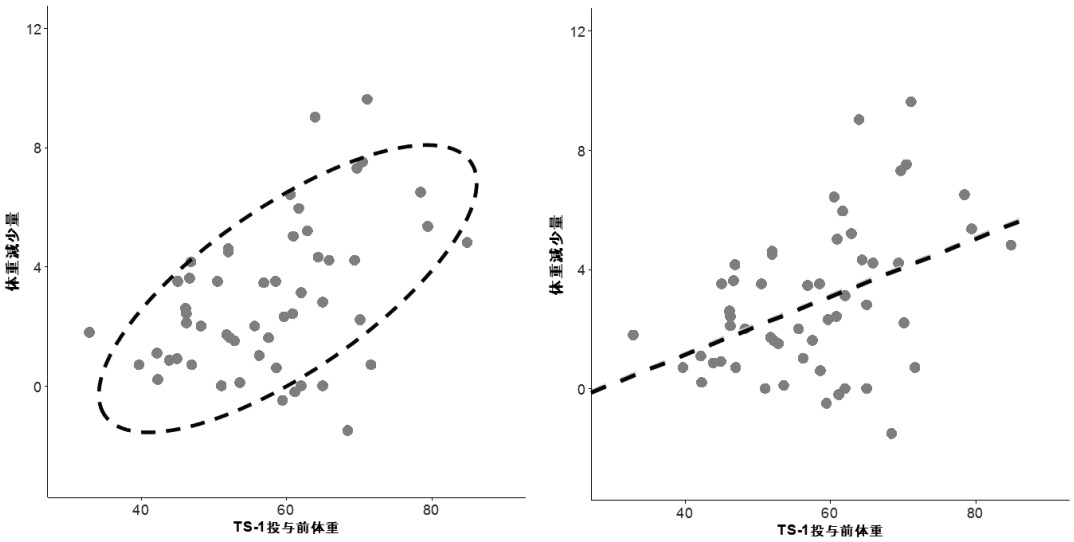
変数	vs. 変数	Spearmanの順位相関係数(ρ)	p値(Prob>|ρ|)	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
血圧	コレステロール	0.7855	<.0001*	

その結果，Spearman の順位相関係数は 0.7855 なので強い正の相関関係が認められた．また，無相関性の検定(「P 値(Prob>|ρ|)」の部分)の p 値が有意水準 0.05 を下回ることから，コレステロール値と収縮期血圧の相関関係が認められた(無相関関係ではなかった).

4.2 単回帰分析

4.2.1 単回帰分析の略説

図 4.3 は，TS-1 による補助化学療法が施行された 100 名の胃癌患者に対する投与前と投与後 6 カ月での体重減少量を表している．相関係数が 0.427 であり，正の相関が認められることから，TS-1 投与前の体重が重いほど体重減少量が大きいと解釈される．そのため，投与前の体重から，投与後 6 カ月での体重減少量を予測することも可能かもしれない．このように，一方の変数(複数の場合もある)からもう一方の変数を予測する統計的方法を回帰分析という．とくに，予測する側の変数は 1 個の場合は単回帰分



(a) 相関分析の図示　　(b) 回帰分析の図示

図 4.3：相関分析と回帰分析の違い

析と呼ばれ，複数の場合は重回帰分析と呼ばれる．このとき，予測する側の変数のことを説明変数，独立変数，入力変数と呼び，予測される側の変数を目的変数，従属変数，目的変数，出力変数という．ここでは，JMP の出力にあわせて，説明変数及び目的変数の名称を用いる．

相関分析と回帰分析は，一見すると同じような分析方法に見えるかもしれない．しかしながら，相関分析と回帰分析には明確な違いがある．相関分析とは，2 変数の関連性(相関関係)を分析する方法であり，正の相関が高いとは，片方の変数の値が上がれば，もう一方の変数の値が上がる(負の相関関係の場合には下がる)ことを表す(図 4.3(a))．一方で，回帰分析は，説明変数から目的変数を予測するための統計モデル(回帰直線)を推定する方法である(図 4.3(b))．

単回帰分析では，1 個の説明変数から目的変数を予測するための統計モデルを推定する．単回帰分析における統計モデルを単回帰直線あるいは単回帰モデルという．単回帰直線は，

$$(\text{目的変数}) = \beta_0 + \beta_1 \times (\text{説明変数}) + (\text{誤差})$$

で与えられる．ここで，単回帰直線の切片 β_0 および傾き β_1 は回帰係数(回帰パラメータ)と呼ばれる．また，誤差は単回帰直線で説明できなかった目的変数の予測値に対する乖離(誤差)である．

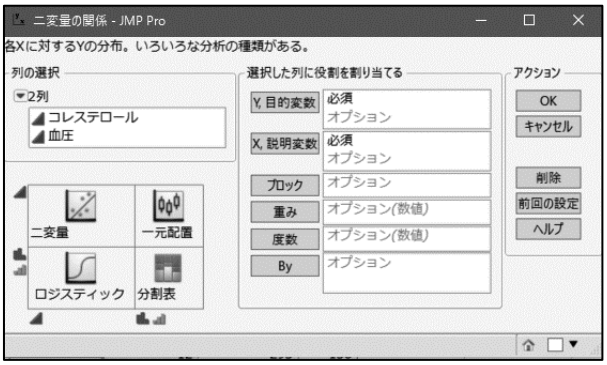
本節では，回帰係数の推定の方法，得られた回帰モデルに意味があるかなどを評価する統計的方法について述べる．

4.2.2 JMP による単回帰分析

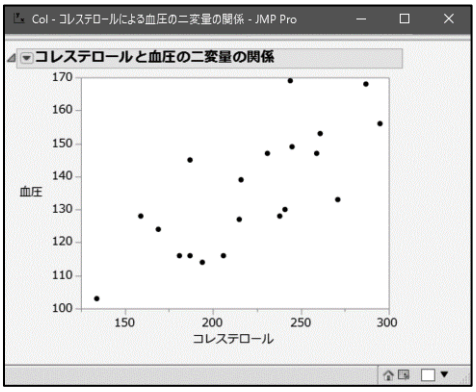
ここでは，4.1.2 節で用いたコレステロール値と血圧(収縮期血圧)のデータを用いる．

単回帰分析の実行

1. 「分析」→「2 変量の関係」を選択する(下図のような画面が表示される)．

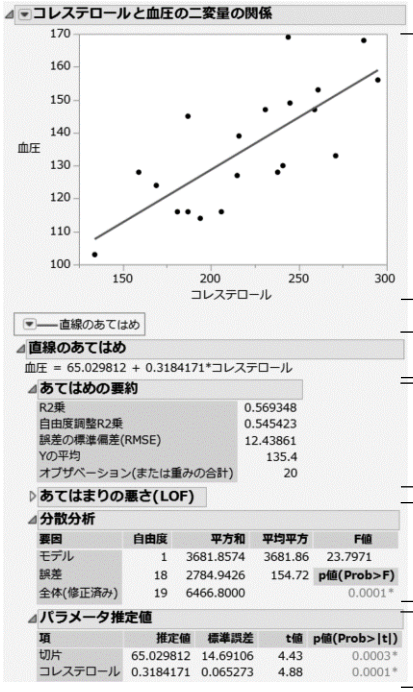


2. 「Y, 目的変数」に「血圧」，「X, 説明変数」に「コレステロール」を割り当てる．
3. 「OK」ボタンを押す(下図のような画面が表示される)．



4　「▼」ボタンを左クリックし，「直線のあてはめ」を選択する．

上記の実行結果を図 4.4 に示す．



(a) あてはめた回帰直線の図示

単回帰分析では，

(血圧) = (切片 β_0) + (傾き β_1)×(コレステロール)

の直線(単回帰直線)で与えられる．この図では，その計算(推定)された直線を表している．

データ点は観測値を表していおり，直線がデータ点に上手く当てはまっていれば，コレステロール値から収縮期血圧を予測できることを意味する．

(b) あてはめた回帰直線の数式

(c) あてはめの要約

ここでは，単回帰直線が当てはまっているか否かを評価する．
R2乗は当てはまっている割合を表しており，0～1の範囲をとる(1になるほど当てはまりが良い)．

(d) 回帰の分散分析

回帰の分散分析では，帰無仮説「回帰直線に意味がない」に対して，対立仮説「回帰直線に意味がある」を検定している．

(e) 回帰係数と検定結果

回帰係数の切片と傾き(コレステロール)の計算値(推定値)と帰無仮説「回帰係数は0である」に対して対立仮説「回帰係数は0でない」を検定している．

図 4.4：単回帰分析の結果

(a)は，あてはめた回帰直線の図示である．この直線は(b)の回帰直線の数式を表している．したがって，コレステロール値から血圧を予測する直線(回帰直線)は

(収縮期血圧) = 65.029812 + 0.3184171 × (コレステロール値)

である(研究の場合には小数点以下 3 桁程度で四捨五入する)．(a)のグラフより，データ点に対して直線がおおよそ当てはまっていることが示唆される．

(c)は，回帰直線の当てはまりを表している．ここで見なければならないのは，「R2 乗」であり寄与率と呼ばれる．寄与率とは，回帰直線がデータに何割当てはまっているかを表している指標であり，0 から 1 までの範囲をとる．この事例では，寄与率が 0.569348 であることから，全体の約 56.9%が当てはまっていることがわかる．因みに，相関係数の 2 乗値が寄与率である．「自由度調整 R2 乗」は重回帰分析で用いる寄与率であり，その意味は 4.3.1 節で述べる．また，「誤差の標準偏差」は，(データ) - (回帰直線での予測値)の標準偏差であるが，解釈には用いることは少ないので無視してよい．

(d)は，回帰の分散分析である．回帰の分散分析では，帰無仮説「回帰直線に意味がない」に対して，対立仮説「回帰直線に意味がある」を検定している．p 値(p 値(Prob>F))が 0.0001 であることから，有意水準 0.05 を下回る．したがって，回帰直線に意味があることが認められた．なお，その他の項目は，p 値を計算する過程で必要なものであるが，解釈することは殆どないため，無視してよい．

(e)は，回帰係数を表している．「推定値」は，(b)の数式の切片および傾き(コレステロール)を表しており，p 値(p 値(Prob>|t|))は，帰無仮説「回帰係数は 0 である(回帰係数に意味がある)」に対する対立仮説「回帰係数は 0 でない」の検定結果である．切片に対する p 値は 0.0003 であり，傾き(コレステロール)に対する p 値は 0.0001 なので，いずれの場合にも有意であった．したがって，切片及び傾きのいずれも有意であるため，

4.3 重回帰分析

4.3.1 重回帰分析の概要と JMP による実行

4.2 節では，説明変数が 1 個の場合の回帰分析法として，単回帰分析について説明した．一方で，医学系研究では，目的変数に対する複数の要因(説明変数)を評価するために，重回帰分析(いわゆる多変量解析)を用いる場面も少なくない．説明変数が 3 個以上のときの重回帰分析の予測式を重回帰式(重回帰モデル)という，

あるフィットネスクラブでは，有酸素運動の効果を把握するための調査を実施している．そこでは，被験者の「年齢」，「体重」といった背景情報，「走行時間」，「走行時の脈拍」，「休息時の脈拍」，「最大脈拍」が「酸素摂取量」とともにとられている．ここでは，「酸素摂取量」を予測するための重回帰式を計算することにある．

このデータは，「Fitness.csv」に保存されている．このファイルを Excel で開いたものが以下である．

	A	B	C	D	E	F	G
1	年齢	体重	酸素摂取量	走行時間	走行時の脈拍	休息時の脈拍	最大脈拍
2	42	68.15	59.57	8.17	166	40	172
3	38	81.87	60.06	8.63	170	48	186
4	43	85.84	54.3	8.65	156	45	168
5	50	70.87	54.63	8.92	146	48	155
6	49	81.42	49.16	8.95	180	44	185

なお，このデータの JMP での読み込み方法は割愛する．

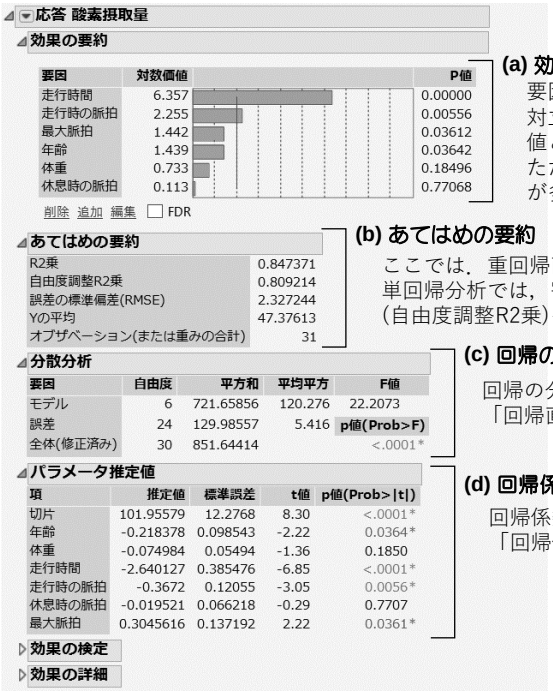
(1) 重回帰分析の実行

ここでは，「酸素摂取量」を目的変数，「年齢」「体重」「走行時間」「走行時の脈拍」「休息時の脈拍」「最大脈拍」を説明変数とした場合の重回帰分析を実行する．

重回帰分析の実行

1. 「分析」→「モデルのあてはめ」を選択する(下図のような画面が表示される).

2. 「Y，目的変数」に「酸素摂取量」，「モデル効果の構成」に「年齢」「体重」「走行時間」「走行時の脈拍」「休息時の脈拍」「最大脈拍」クリックし「追加」ボタンを押す．
3. 「強調点」を「最小レポート」に変更する.
4 「実行」ボタンを押す．



(a) 効果の要約
要因(説明変数)の回帰係数に対する検定(帰無仮説「回帰係数は0である」，対立仮説「回帰係数は0でない」)でのp値が記載されている．また，対数価値とは，p値に基づく影響の大きさを表す指標(高いほど影響が強い)である．ただし，影響の大きさの評価には，標準回帰係数(後述)を用いることのほうが多い．

(b) あてはめの要約
ここでは，重回帰直線(モデル)が当てはまっているか否かを評価する．
単回帰分析では，寄与率(R2乗)を用いていたが，重回帰分析では，自由度調整済み寄与率(自由度調整R2乗)を用いる．

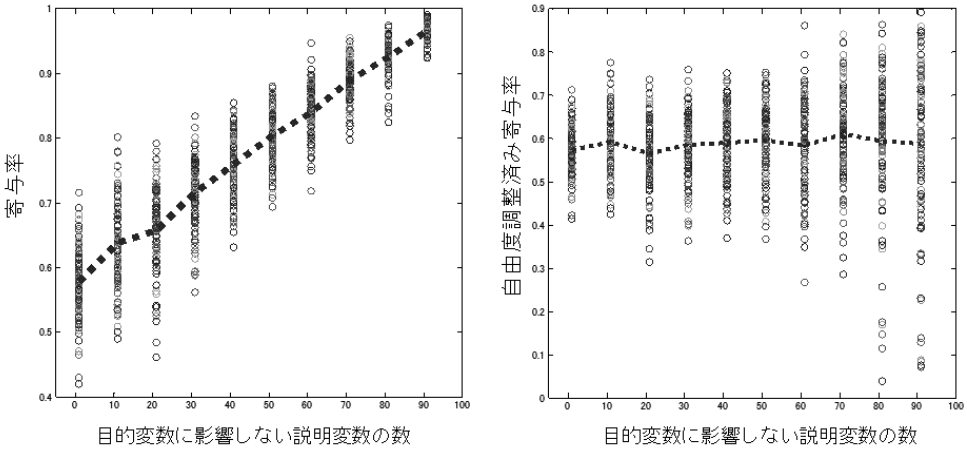
(c) 回帰の分散分析
回帰の分散分析では，帰無仮説「回帰直線に意味がない」に対して，対立仮説「回帰直線に意味がある」を検定している．

(d) 回帰係数と検定結果
回帰係数の計算値(推定値)と帰無仮説「回帰係数は0である」に対して対立仮説「回帰係数は0でない」を検定している．

図 4.5：重回帰分析の結果

このときの結果を図 4.5 に示す．(a)は，帰無仮説「回帰係数は 0 である」，に対して，対立仮説「回帰係数は 0 でない」を評価するための回帰係数に対する検定の結果である．「走行時間」，「走行時の脈拍」，「最大脈拍」，「年齢」の p 値は有意水準 0.05 のもとで有意であるのに対して，「体重」，「休息時の脈拍」は，有意でなかった．また，対数価値は，p 値に基づいて計算されるものであり，要因(説明変数)の影響の大きさを表している．ただし，一般的に説明変数の影響の大きさは，標準回帰係数で評価することが多い．標準回帰係数の計算方法は，後述する．

(b)は，回帰直線の当てはまりを表している．単回帰分析では，寄与率(R2 乗)を評価するが，重回帰分析では自由度調整済み寄与率(自由度調整 R2 乗)を評価する．図 4.6(a)は，説明変数の増加に伴う寄与



(a) 寄与率　　(b) 自由度調整済み寄与率

図 4.6：アウトカム(応答変数)に影響しない説明変数の数を増加したときの寄与率及び自由度調整済み寄与率の推移(各説明変数の数に対して 100 回のシミュレートを実施している．点線は平均値)

率の変化をシミュレーションを用いて表した散布図である(シミュレート回数=100)．ここで，X 軸は目的変数に影響しない説明変数(無意味な説明変数)の数，Y 軸は寄与率を表している．データ点は，個々のシミュレーションの結果であり，青色の点線は寄与率の平均値の推移を表している．寄与率はアウトカムに影響しない変数の増加に伴い上昇していることがわかる．図 4.6(b)は説明変数の増加に伴う自由度調整済み寄与率の変化を表している．説明変数が増加しても自由度調整済み寄与率が変化しないことがわかる．

この事例では，自由度調整済み寄与率が 0.8092 であることから，全体の約 80.9%が当てはまっていることがわかる．また，「誤差の標準偏差」は，(データ) - (回帰直線での予測値)の標準偏差であるが，解釈には用いることは少ないので無視してよい．

(c)は，回帰の分散分析である．回帰の分散分析では，帰無仮説「回帰直線に意味がない」に対して，対立仮説「回帰直線に意味がある」を検定している．p 値(p 値(Prob>F))が<0.0001 であることから，有意水準 0.05 を下回る．したがって，回帰直線に意味があることが認められた．なお，その他の項目は，p 値を計算する過程で必要なものであるが，解釈することは殆どないため，無視してよい．

(d)は，回帰係数及び p 値を表している．「推定値」は，回帰係数を表している．推定された直線(四捨五入して記載する)は，

(酸素摂取量)＝102.0 - 0.218×(年齢) - 0.075×(体重) - 2.640×(走行時間)
- 0.367×(走行時の脈拍) - 0.020×(休息時の脈拍)＋0.305×(最大脈拍)

である．p 値(p 値(Prob>|t|))は，帰無仮説「回帰係数は 0 である(回帰係数に意味がある)」に対する対立仮説「回帰係数は 0 でない」の検定結果である．(a)で述べたように，「走行時間」，「走行時の脈拍」，「最大脈拍」，「年齢」の p 値は有意水準 0.05 のもとで有意であった．

(2) 標準回帰係数

回帰分析の一つの目標は、目的変数を適切に予測するための回帰直線を推定することであるが(精度の高い予測式を作る)，もう一つの目標は，個々の説明変数が目的変数にどの程度の影響を及ぼすかを評価することである．一方で，説明変数は測度が異なるため(例えば，身長は cm で測定されており，体重は kg で測定されている)，推定された重回帰分析の回帰係数が目的変数に対する影響の強さを評価できない．各説明変数が目的変数に影響を及ぼす大きさを評価する係数は，標準回帰係数と呼ばれる．

標準回帰係数の計算
1.　「重回帰分析の実行」を行う.
2.　「▼」ボタンを左クリックし,「要因のスクリーニング」→「尺度化された推定値」を選択する.

このときの結果を以下に示す．

項	尺度化した推定値	標準誤差	t値	p値(Prob>|t|)
切片	47.376129	0.417985	113.34	<.0001*
年齢	-2.074592	0.936158	-2.22	0.0364*
体重	-1.220366	0.894149	-1.36	0.1850
走行時間	-7.735573	1.129445	-6.85	<.0001*
走行時の脈拍	-7.343994	2.411	-3.05	0.0056*
休息時の脈拍	-0.292812	0.993266	-0.29	0.7707
最大脈拍	5.6343895	2.53805	2.22	0.0361*

「尺度化した推定値」とは，標準回帰係数であり，絶対値の数値が大きいほど目的変数に対する影響が強いことを意味する．標準回帰係数の絶対値を大きい順に並べ替えたものを以下に示す(小数点第 3 位で四捨五入).

説明変数	標準回帰係数 （尺度化した推定値）	標準回帰係数の絶対値
走行時間	-7.736	7.736
走行時の脈拍	-7.344	7.344
最大脈拍	5.634	5.634
年齢	-2.075	2.075
体重	-1.220	1.220
休息時の脈拍	-0.293	0.293

その結果,「走行時間」の標準回帰係数の絶対値が最も高く(目的変数に影響を与えている)，次いで「走行時の脈拍」の標準回帰係数の絶対値が高かった．したがって，走行に関する要因が酸素摂取量に影響を及ぼしており，これらの回帰係数が負値であることから,「走行時間」,「走行時の脈拍」が高くなるほど，酸素摂取量が減少する傾向が示された.

4.3.2 変数選択の方法

(1) 変数選択(ステップワイズ法)の概要

重回帰分析では，複数の説明変数を評価することができる．一方で，少しでも多くの説明変数を重回帰直線(回帰モデル)に含めたほうが良い結果を導くかというと，そうではなく，不要な説明変数は「ノイズ」として重回帰直線(回帰モデル)に含まれるため,「不要な説明変数は含めるべきではない」.

フィットネスクラブのデータでは，6 個の説明変数がある．一方で,「体重」,「休息値の脈拍」の脈拍の回帰係数に対する検定は有意でなく，標準回帰係数も小さい．したがって，これらを削除しても重回帰直線の予測結果に影響がないかもしれない(むしろ，良くなるかもしれない).

不要な説明変数を削除することは，目的変数を予測するうえでの「ノイズ」を除去することにも繋がり，より安定的な重回帰直線の計算に繋がる．そのため，重回帰分析を実施する場合には，変数選択を併せて実施することが多い．このとき，目的変数を適切に予測するための説明変数を選択することは変数選択と呼ばれる.

目的変数(酸素摂取量)を予測するための最適な説明変数を選択するには，全ての説明変数のパターンを計算しなければならない．フィットネスクラブのデータの場合には，6 個の説明変数があることから，63 パターンの重回帰直線を計算し，最適な説明変数の組み合わせを選択することになる．63 パターンであれば，現在のコンピュータの能力であれば実行可能かもしれない．しかしながら，20 個の説明変数がある場合には，220-1=1,048,575 パターンでの評価を行わなければならず，計算が困難になる.

そのため，変数選択では，説明変数の組み合わせの全パターンを評価するのではなく，ステップワイズ法というアルゴリズムを用いることが多い．ステップワイズ法には次の 3 種類がある：

(a) 変数増加法：切片のみのモデルから出発し，1 個ずつ説明変数をモデルに加える方法.

(b) 変数減少法：全ての説明変数を含むモデルから出発し，1 個ずつ説明変数をモデルから除外する方法.

(c) 変数増減法：全ての説明変数を含むモデルから出発し，1 個づつ説明変数を加えるのか除外するのかを評価・実施する方法.

ステップワイズ法のアルゴリズムに対するゴールド・スタンダードは存在しない．一方で，医学系研究の領域では，削除される説明変数の数が比較的少ないという理由から，変数減少法が好まれる傾向にある．因みに，変数増加法は前進ステップワイズ法，変数減少法は後退ステップワイズ法と呼ばれることもある．

また，変数を増加(減少)させるか否かを評価する指標には，検定を用いる方法と情報量規準を用いる方法がある．情報量規準とは，推定された回帰直線(回帰モデル)の適切性を評価する測度(ものさし)であり，赤池の情報量規準(AIC；Akaike’s Information Criteria)，Bayes 流情報量規準(BIC)などの方法が提案されている．検定を用いる方法では有意水準 α を事前に設定したもとで評価しなければならず，恣意的に説明変数の数が制御されてしまう恐れがある．そのため，情報量規準を用いることが多くなっている．

(1) 変数選択(ステップワイズ法)の概要

ここでは，先ほどの重回帰分析と同様に，「酸素摂取量」を目的変数，「年齢」「体重」「走行時間」「走行時の脈拍」「休息時の脈拍」「最大脈拍」を説明変数とした場合のステップワイズ法を伴う重回帰分析を実行する．

ここでは，3 種類のステップワイズ法(変数増減法，変数減少法，変数増加法)の実行方法及び評価基準に BIC を用いた場合について述べる．JMP では，検定を用いる場合，修正 AIC(AICc)を用いる場合，BIC を用いる場合の 3 種類が存在する．前述したように，検定を用いる方法は推奨されない．修正 AIC と BIC の選択であるが，修正 AIC のほうが説明変数が多く選択される傾向にあるが，取捨選択に関する明確な基準は存在しない．

ステップワイズ法を伴う重回帰分析の実行

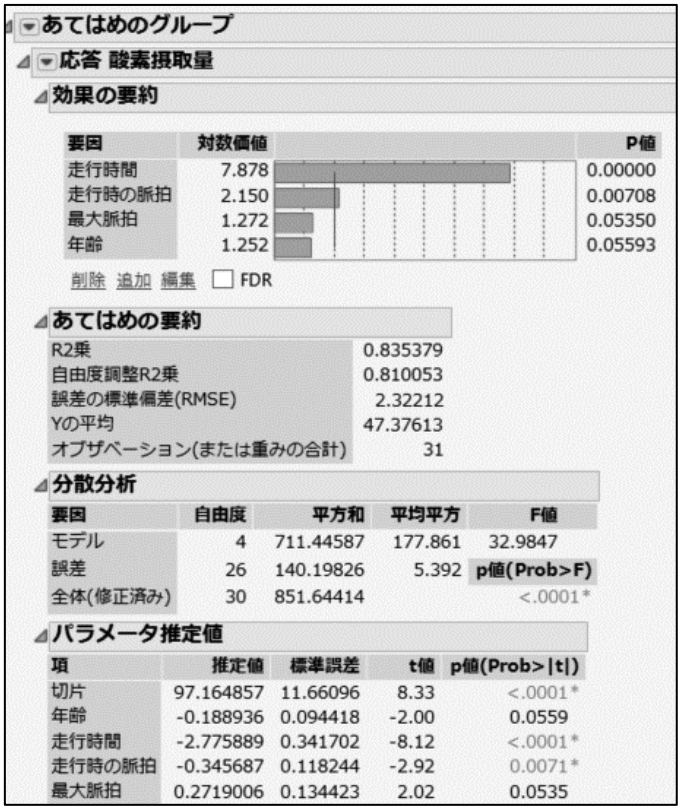
1. 「分析」→「モデルのあてはめ」を選択する(下図のような画面が表示される).

2. 「Y，目的変数」に「血圧」，「モデル効果の構成」に「年齢」「体重」「走行時間」「走行時の脈拍」「休息時の脈拍」「最大脈拍」クリックし「追加」ボタンを押す.

3.　「手法」を「ステップワイズ法」に変更する.
4　「実行」ボタンを押す(以下が表示される).

5a　変数増加法の場合:「停止ルール」に「最小 BIC」を選択,「方法」に「変数増加」を選択,「すべて削除」をクリックし,「実行」ボタンを押す.
5b　変数減少法の場合:「停止ルール」に「最小 BIC」を選択,「方法」に「変数減少」を選択,「すべて追加」をクリックし,「実行」ボタンを押す.
5c　変数増減法の場合:「停止ルール」に「最小 BIC」を選択,「方法」に「変数増減」を選択,「すべて追加」をクリックし,「実行」ボタンを押す.
6　「現在の推定値」の「追加」にチェックが入っているものが選択された変数になる. 選択されたモデルのもとで「モデルの実行」をクリックする.

変数減少法を用いて変数選択を実施後の重回帰分析の結果を以下に示す(重回帰分析の場合と出力は同じなので説明は割愛する).



あてはめのグループ

応答 酸素摂取量

効果の要約

要因	対数価値	P値
走行時間	7.878	0.00000
走行時の脈拍	2.150	0.00708
最大脈拍	1.272	0.05350
年齢	1.252	0.05593

削除 追加 編集 FDR

あてはめの要約

R2乗	0.835379
自由度調整R2乗	0.810053
誤差の標準偏差(RMSE)	2.32212
Yの平均	47.37613
オブザベーション(または重みの合計)	31

分散分析

要因	自由度	平方和	平均平方	F値
モデル	4	711.44587	177.861	32.9847
誤差	26	140.19826	5.392	p値(Prob>F)
全体(修正済み)	30	851.64414		<.0001*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	97.164857	11.66096	8.33	<.0001*
年齢	-0.188936	0.094418	-2.00	0.0559
走行時間	-2.775889	0.341702	-8.12	<.0001*
走行時の脈拍	-0.345687	0.118244	-2.92	0.0071*
最大脈拍	0.2719006	0.134423	2.02	0.0535

先ず,「体重」および「休息時の脈拍」が削除されていることを確認されたい．また,「年齢」,「最大脈拍」について，重回帰分析(変数選択無)の場合には有意だったにも関わらず，変数選択後の重回帰分析では，有意水準 0.05 を僅かに上回っている．

重回帰分析(変数選択無)の場合の自由度調整済み寄与率は 0.809 であったのに対して，変数選択後の重回帰分析では 0.810 であり，変数を 2 個削除したにも関わらず，自由度調整済み寄与率は僅かに上昇している．

その結果，変数選択によって新たに得られた重回帰式は

(酸素摂取量)＝97.2 - 0.189×(年齢) - 2.776×(走行時間) - 0.346×(走行時の脈拍)＋0.272×(最大脈拍)

であった．

5 章 因子分析

人間の感情は非常に複雑である．例えば，「ある人を好き」という感情があったとする．このとき，どこが好きなのかを考えると，「外見」，「優しい」，「気遣い」など，様々な要素がある．このとき，計量心理学などの分野では，「好き」という感情を直接的には測るのではなく，「外見」，「優しさ」，「気遣い」など具体的な内容について調査し，それらを用いて「好き」という感情を測ることを考える．つまり，直接的に測れるものを総合的に組み合わさせて「好き」という感情になると思われる．因子分析は，具体的に調査できる内容に基づいて，新しい潜在的な「ものさし」を作り出し，そのもとで様々な解釈を行う分析方法である．

このとき，直接的に測定できるものは観測変数(データ)と呼ばれ，通常は複数の項目(変数)構成される．そして，これらの変数を合成することで得られる潜在的な「ものさし」は，因子(潜在変数)と呼ばれる．因子分析(正式には，探索的因子分析)では，潜在的な「ものさし」を観測値(データ)から導き出す方法である．

因子分析は， (1) 因子数の決定，(2) 因子の回転，のステップに分けられる．本節では，ケーススタディに基づいて，因子分析を略説する．

(1) データの概要と入力

ここでは，病院における満足度調査アンケートに関するデータを用いる．

これは，ある病院で医療サービス向上のための満足度調査の結果である．ここでは，外来患者を対象に 5 項目で構成される 7 段階評価(1.非常に不満，2.不満，3.やや不満，4.どちらともいえない，5.やや満足，6.満足，7.非常に満足)での調査を実施している(内田他,2013 を一部改編).

問 1：受付の対応は丁寧でしたか？

問 2：診断・処置には満足しましたか？

問 3：診察時，医師から病気・怪我についての十分な説明はありましたか？

問 4：掲示板は見やすかったですか？

問 5：診察までの待ち時間の長さはどのように感じましたか？

このデータは，「questionarie.csv」に保存されている．このファイルを Excel で開いたものが以下である．

	A	B	C	D	E
1	受付の対応	診察・処置	病気・怪我の説明	掲示物	待ち時間（診察）
2	4	4	4	5	3
3	3	3	3	5	3
4	3	6	5	3	4
5	3	4	6	5	3

A 列から E 列は，それぞれ問 1 から問 5 に対応している．なお，このデータの JMP での読み込み方法は割愛する．

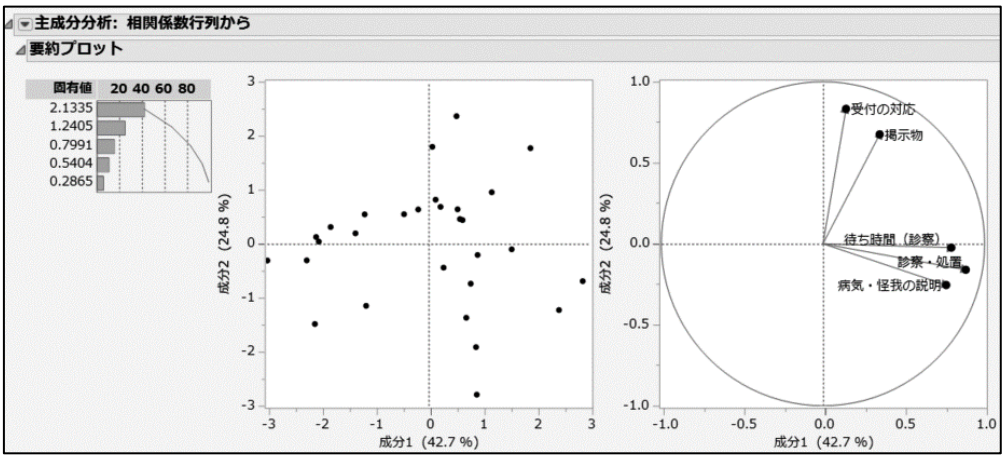
(2) 因子数の決定

因子数の決定では，観測変数の個数(変量数)の因子を計算(推定)から始める．因子の計算には様々な方法が提案されている．かつては，主成分分析を利用することも多かったが，最近では最尤法を用いることが多くなっているようである．一方で，最尤法では観測変数が正規分布に従うことが仮定される．

因子数の決定

1. 「分析」→「多変量」→「主成分分析」を選択すると以下の画面が表示される．

2. 「Y,列」に全ての変数を移動させる．
3. オプションのなかの「推定法」を「最尤法」に変更する．
4. 「OK」ボタンを押す(下図が得られる)[12]．



5. 「主成分分析：相関係数行列から」の左側にある「▼」から「スクリープロット」を選択する．
6. 「主成分分析：相関係数行列から」の左側にある「▼」から「固有値」を選択する．

「推定法」では，最初に「デフォルト」が設定されている．これは，

- 欠測がない場合には「リストワイズ」が選択される．
- 大規模なデータ[13]であり，かつ欠測がある場合は「ペアワイズ」が選択される．
- それ以外では，「REML」が選択される．

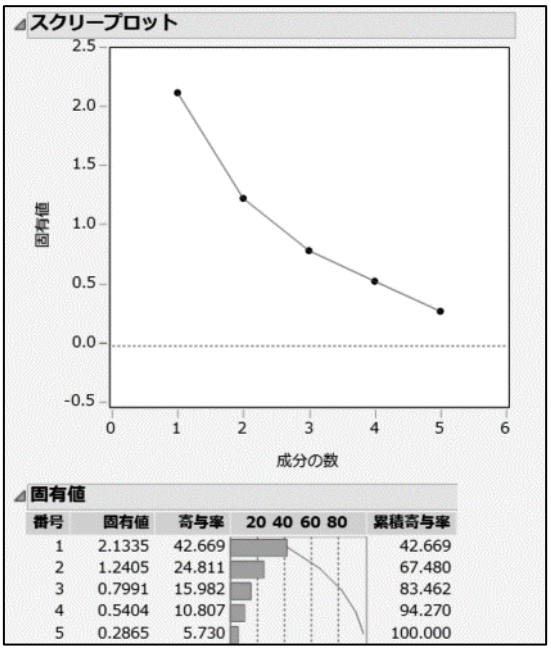
のいずれかが選択される．「REML」とは，制約付き最尤法のことで，今回用いた最尤法の性能を向上させたものである．「リストワイズ」とは，一般的な主成分分析(因子分析では主因子法という)の計算方法

[12] このときに得られる結果は，主成分分析と呼ばれる．主成分分析とは，多変数をより少ない変数で表現するために用いる方法である．例えば，プロ野球選手の成績(打率，打点，ホームラン数，安打数)は 4 変数である(つまり 4 次元)．この 4 次元のデータを 2 次元以下で表現するのが主成分分析の目的である．したがって，因子分析と若干意味が異なる．

[13] 11 変数以上，または 5001 例以上

を用いる場合である．「ペアワイズ」とは，欠測値を考慮して計算する方法である．JMP には，その他に外れ値を考慮した「ロバストの推定法」，非常に多くの観測変数を取り扱う場合に用いる「横長」が存在する．しかしながら，これらの方法は主成分分析に適用できる方法であり，因子分析のモデルの設定には，主成分分析(主因子法)と最尤法しか存在しないため，「デフォルト」(欠測がない場合には主因子法に対応)，あるいは「最尤法」を選択する．

その結果，下図が得られる．



因子分析における因子数の決定方法として，主なものには以下がある．

1. ガットマン基準：固有値が 1 以上の因子を採用する．
2. スクリー基準：スクリープロットを用いて，その推移がなだらかになる前までを抽出する，
3. 寄与率が 50～60%以上になる因子数を採用する．
4. 解釈が可能な因子構造を採用する．

このなかで，比較的多く用いられるのが 1 あるいは 2 である．1 の場合には，JMP アルゴリズム「因子数の決定」(上記記載の実行手順) のなかの「要約プロット」の左下の固有値を見て 1.0 以上の因子数を用いればよい．2 の場合には，JMP アルゴリズム「因子数の決定」のステップ 5 で得られるスクリープロットを省察し，なだらかになる直前までの因子数を用いればよい．3 の場合には，JMP アルゴリズム「因子数の決定」のステップ 6 で得られる累積寄与率が 50～60 以上となる因子数を用いればよい．4 の場合には，後述する因子数を変えながら探索的に実施する．

ここでは，ガットマン基準に用いて因子数を決定する．固有値が 1.00 を上回るのは 2 個なので，因子数は 2 個であると考える．なお，スクリー基準の場合には，プロットがなだらかになる部分は把握できず，寄与率の場合にはガットマン基準と同様に因子数に 2 個を選択することになる．

(3) 因子の回転

先ほど，因子数を 2 個に設定した．因子の回転とは，得られた因子に対して最も意味付けをしやすいように因子を回転させる．具体的には，回転によって得られる因子負荷量(因子の意味を解釈する指標であ

り，数値が 0 よりも離れているほど観測変数が因子の影響を受けている解釈することができる)が各因子で解釈できる状況をいう．解釈できる状況とは，例えば，4 個の観測変数から 2 因子を抽出した場合を考える．

(a) 解釈できない状況

	因子 1	因子 2
変数 A	0.50	0.50
変数 B	0.50	0.50
変数 C	0.50	0.50
変数 D	0.50	0.50

(b) 解釈できる状況

	因子 1	因子 2
変数 A	0.90	0.10
変数 B	0.90	0.10
変数 C	0.10	0.90
変数 D	0.10	0.90

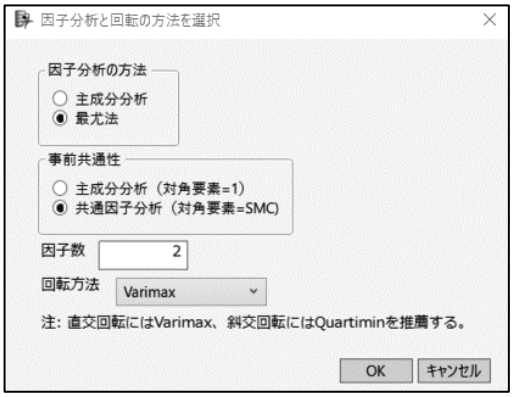
(a)の場合は，因子 1 と因子 2 が同じような特色を示しており，それぞれの因子の特徴を把握しにくい．一方で，(b)の場合には，因子 1 が変数 A と変数 B に影響を与えており，因子 2 が変数 C と変数 D に影響を与えていることがわかる．因子の回転は，可能な限り(b)の状況に近づけることが目的である．

因子回転は，直交解と斜交解の 2 種類に大別される．直交解として有名なのが Varimax 回転であり，斜交解では Promax 回転あるいは Quartimin 回転などがある．直交解は，結果を散布図等で図示できるものの，斜交解に比べて(b)の状況が得られにくい傾向にある．一方で斜交解は，(b)の状況を得られやすい傾向にあるものの，因子毎に解釈する必要がある(散布図によるグラフ表示ができない)．

ここでは，直交解の Varimax 回転を用いる(JMP では直交解に Varimax 回転，斜交解に Quartimin 回転が推奨されているが，斜交解に Promax 回転を用いる研究も多い)．

因子数の決定

1. 「因子数の決定」の 1 から 4 までを実行する．
2. 「主成分分析：相関係数行列から」の左側にある「▼」から「因子分析」を選択する(下記の画面が表示される)．



3. 「因子分析の方法」を「最尤法」にチェックする．
4. 「事前共通性」を「共通因子分析」にする[14]．
5. 「因子数」を「2」にする．
5. 「回転方法」を「Varimax」にする．
6. 「OK」ボタンを押す．

因子分析では，非常に多くのアウトプットが出力されるため，解釈をしながら見ていくこととする．

[14] 事前共通性の「共通性」について略説する．因子分析では，観測変数のデータは，因子による共通性と観測変数の独自性の和で与えられると考えられている．事前独自性とは，その重みの方法を表しているが，因子分析では「共通因子分析」以外を用いることは殆どない．

[A] 回転後の因子負荷量

回転後の因子負荷量では，次が表示される．

回転後の因子負荷量

	因子1	因子2
受付の対応	0.0347581	0.4957235
診察・処置	0.9987591	0.0498013
病気・怪我の説明	0.5925281	0.0335941
掲示物	0.0882174	0.5543414
待ち時間（診察）	0.6221213	0.1677376

因子負荷量とは，因子と観測変数の相関を表しており，数値が大きくなるほどその因子の影響が強いことを意味する．因子分析では，因子負荷量を見ながら因子に「意味付け」を行う．

・因子 1 :「診察・処置」「病気・怪我の説明」「待ち時間」が高いことから，診断・処置といった医療行為に関連する満足度に関する因子と解釈できる．

・因子 2 :「受付の対応」「掲示板」の因子負荷量が高いことから，医療行為以外の満足度に関する因子と解釈できる．

さらに，2 因子で十分か否かを検定する「有意性検定」を表示させる．

有意性検定

検定	自由度	カイ2乗	p値(Prob>ChiSq)
H0: 共通因子が1つもない。 HA: 少なくとも1つの共通因子がある。	10.000	29.585	0.0010*

検定	自由度	カイ2乗	p値(Prob>ChiSq)
H0: 2因子で十分である。 HA: もっと多くの因子が必要である。	1.000	1.738	0.1874

その結果，帰無仮説「共通因子(因子)が一つもない」に対する有意性検定が有意であることから，共通因子は少なくても一つは存在することが認められた．次いで，帰無仮説「2 因子で十分である」が棄却されなかったことから，2 因子の適切性が認められる．

(4) 因子分析のもう一つの実行方法

今回は，「分析」→「多変量解析」→「主成分分析」から実行しているが，もう一つの方法は，「分析」→「消費者調査」→「因子分析」で実行することである．これは，今回の主成分分析からの実行方法のメニューを作り直したものであり，若干簡便に利用できるようになっている．

参考文献

[12] 安武 繁：事例問題から学ぶ 看護疫学・保健統計学 重要事項をねこそぎ check!，医歯薬出版，2016.
[2] 石村貞夫・今福恵子・劉 晨・桃井保子：よくわかる医療・看護のための統計入門，東京図書, 2009.
[3] 片平洌彦：やさしい統計学―保健・医薬・看護・福祉関係者のために(第 6 版)，桐書房,2017.
[4] 新谷 歩：今日から使える医療統計，医学書院，2015.
[5] 比江島 欣慎：ぜんぶ絵で見る医療統計～身につく！研究手法と分析力，羊土社, 2017.
[6] 柳川 堯・西 晃央・椛 勇三郎・堤 千代：看護・リハビリ・福祉のための統計学，近代科学社, 2011.

補遺 1：統計的検定の取捨選択

		単群	2 群	3 群以上	回帰分析
計量データ	正規分布	母平均の検定	2 標本 t 検定	1 元配置分散分析	重回帰分析
	非正規分布	1 標本 Wilcoxon 検定	Wilcoxon 検定	Kruskal-Wallis 検定	
計量データ(対応)	正規分布	—	対応のある t 検定	反復測定分散分析	—
	非正規分布	—	Wilcoxon 符号付順位和検定	Friedman 検定	
2 値データ	少数例	母比率の検定(正確 p 値)	Fisher の正確検定	Fisher の正確検定	ロジスティック 回帰分析
	非少数例	母比率の検定	カイ 2 乗検定	カイ 2 乗検定 Cochran-Armitage 検定(傾向性)	
2 値データ(対応)	—	—	McNemar 検定	Cochran の Q 検定	—
カテゴリカル	通常	適合度検定	カイ 2 乗検定	—	ロジスティック回帰分析 比例オッズモデル(順序)
カテゴリカル(対応)	—	—	McNemar-Bowker 検定	—	
生存期間	—	1 標本ログランク検定	ログランク検定	一般化ログランク検定	Cox 比例ハザードモデル