

[利用の手引]

## 統計解析ソフト JMP の利用 :

### 解析の基本

犬塚裕樹<sup>†</sup>

Hiroki Inutsuka<sup>†</sup>

<sup>†</sup>久留米大学 医学部看護学科

<sup>†</sup> Kurume University, School of Nursing

## 1. はじめに

統計解析ソフトウェア JMP は、扱いやすく、現在、教育や研究の場に広く普及している。本学では、御井キャンパスや旭町キャンパスの情報教育施設に設置されているほとんどのコンピュータで自由に扱える環境にある。医学部看護学科では、この JMP は学生教育において統計関連の多くの授業のみならず、卒業研究で最も利用されている統計ソフトウェアの1つである。

本稿では、学部学生の卒業研究レベルでの利用で、一般に使われると予想される基礎的な操作手順について説明をする。

内容は、

- [1] 度数分布：名義尺度
- [2] 度数分布：連続尺度
- [3] 変数間の相関：名義尺度
- [4] 変数間の相関：連続尺度
- [5] 計算式により新しい変数を生成する
- [6] 分割表の入力
- [7] ノンパラメトリック変数の2群間での比較

JMP にはウィンドウズパソコン用や Macintosh 用のものがある。いずれの操作もほとんど同じであるが、表示が一部異なる。本稿では、Macintosh JMP を利用した場合で説明する。JMP Pro 11 を利用した。



## 2. データの入力

解析にもちいるデータは、図1のように、データベース形式で、エクセルファイルに入力されているとしよう。各1行にある1人のすべての情報が入力されているという形式で

	A	B	C	D	E	F	G	H
1	ID	氏名	性	年齢	身長	体重	Q1	Q2
2	1	a	1	25	160	53	1	1
3	2	b	2	26	158	58	1	1
4	3	c	1	30	162	50	1	2
5	4	d	1	27	165	60	2	2
6	5	e	2	32	160	52	2	2
7	6	f	2	21	162	53	1	3
8	7	g	2	22	165	55	2	3
9	8	h	1	35	169	65	2	2
10	9	i	1	36	159	50	2	2
11	10	j	1	32	158	48	1	2
12								
13								

図1 エクセルファイルのデータ

ある。

この形式のエクセルデータは JMP に取り込み、解析をおこなうことができる。

### 3. JMP の起動

JMP を起動しよう。すると図 2 の画面が表示される。

「データテーブルの新規作成」をクリックする。

すると、図 3 の新規シートが表示される。このシートには、エクセルシートから「コピー／ペースト」の機能をつかうことで、データを取り込むことができる。

エクセルシートのファイルを開き、データの部分をコピーする。そのあとで、JMP のこの新規シートの画面を表示している状態で、メニューの「編集」／「列名とともにペースト」をクリックすると、JMP のデータシートに入力される。項目名は JMP の項目名の部分に入力されている。

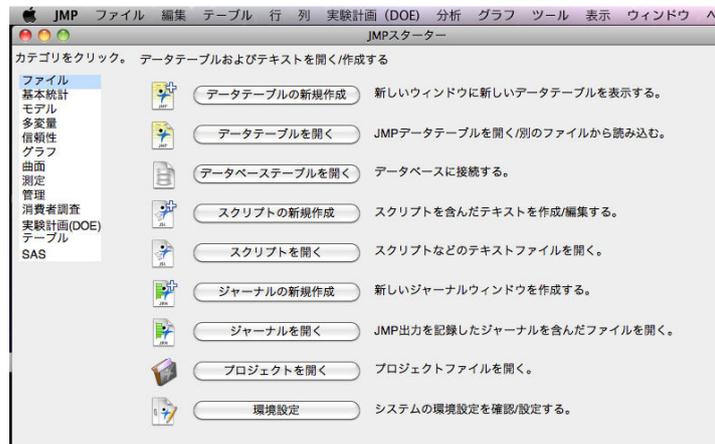


図 2 JMP 初期画面

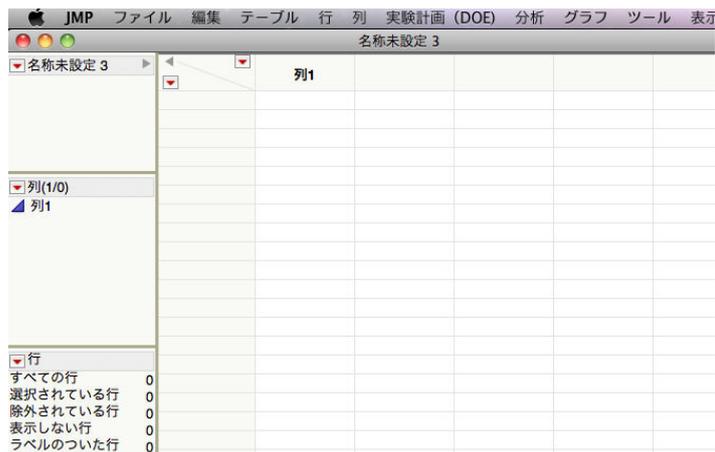


図 3 JMP 新規シート

### 4. JMP における変数

項目名としてあらわされる変数は、「連続尺度」、「順序尺度」、「名義尺度」の 3 つに区分される。JMP では、これらの区分指定が特に重要である。この情報をもとに JMP は解析時、解析するうえで必要なメニューを表示してくれる。たとえば、「性別」は、男性、女性の水準をもつ「名義尺度」である。

尺度の指定はシートの項目名のところでおこなうことができる。図 4 のように、「連続尺度」、「順序尺度」、「名義尺度」の区分は、変数名の左隣にマークとして表示される。



図 4 変数名リスト

## 5. 度数分布：名義尺度

データの分析を始めるにあたって、最初に必要なことは、データの全体的な基本特性を知ることである。まず、項目の度数分布をみよう。

ここでは、「性別」の度数分布をみよう。「性別」の変数が「名義尺度」となっていることを確認する。「名義尺度」は、赤色でヒストグラムのデザインとなっている。

メニューの「分析」をクリックし、表示されるポップアップメニューの中の「一変量の分布」をクリックする。

すると、図5のような表示があらわれる。このデータでは、性別は「女性」が1、「男性」が2とされている。

まず、「8列」ウィンドウ内の「性」をクリックし、さらに、中央に位置する「Y, 列」をクリックする。すると、「Y, 列」ウィンドウの中に「性」という変数名が入力表示される。そこで、「OK」をクリックする。

すると、図6のようなヒストグラムが表示される。この図をふつうに見慣れている横表示にしよう。

「一変量の分布」の左隣の▽表示をクリックし、「積み重ね」をクリックする。

すると、図7のような横の表示にかわる。ヒストグラムと同時に、右の部分には、水準「1」の度数が6件で、60%の割合であることがわかる。これが、名義尺度の場合の表示となっている。



図5 一変量分布の画面

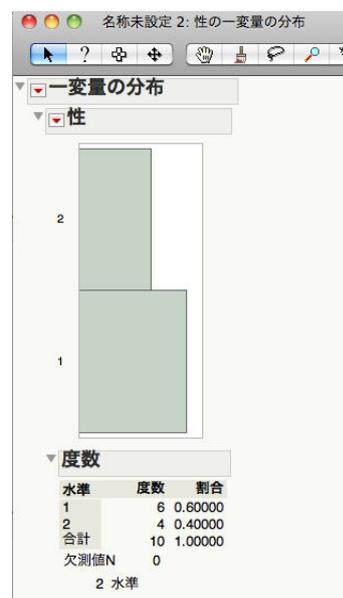


図6 ヒストグラム



図7ヒストグラム (積み重ね)

## 6. 度数分布：連続尺度

ここでは、連続尺度である「年齢」の度数分布をみてみよう。

メニューの「分析」をクリックしたあと、「一変量」をクリックする。

「8列」ウィンドウ内の「年齢」をクリックし、「Y,列」をクリックする。その後、「OK」をクリックする。

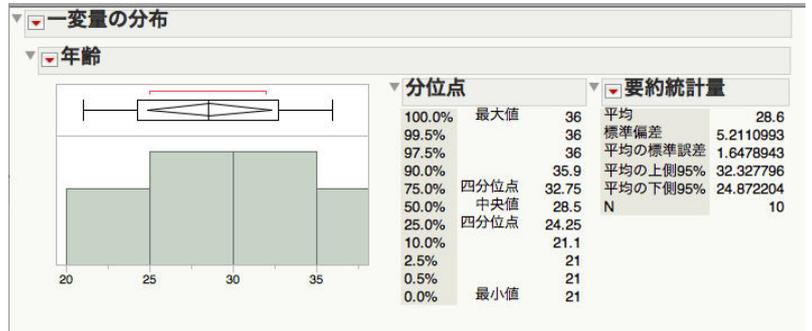


図8 年齢のヒストグラム

すると、図8のようなウィンドウが表示される。ここでは、ヒストグラムと分位点、および平均、標準偏差なども表示される。

このウィンドウに表示されているヒストグラムの表示を少し変更しよう。ヒストグラムの年齢の1表示区画の幅枠を小さくする。そのためには「手のひらツール」をつかう。

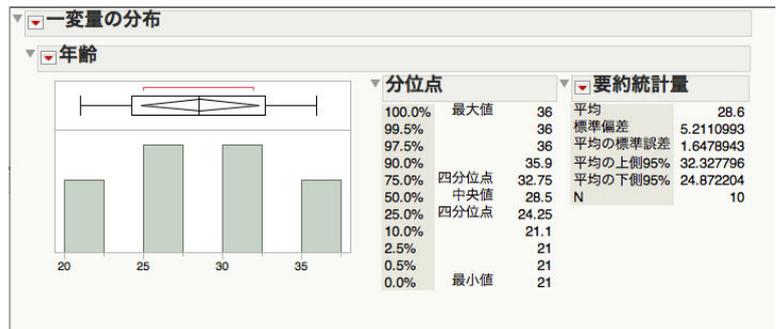


図9 年齢のヒストグラム

「手のひらツール」をクリックし指定したあと、ヒストグラム上でマウスを下にドラッグすると、図9のような区画幅が小さい表示にかわった。

多くの変数の度数分布を描きたいとき、図5での変数の指定で、1個だけでなく複数個の指定ができ、同時に複数の度数分布をもとめることができるのが便利である。

体重と身長を同時に描こう。

メニューの「分析」をクリックし、表示されるポップアップメニューの中の「一変量の分布」をクリックする。

変数の「体重」と「身長」をそれぞれクリックし、「Y,列」をクリックする。すると図10のようにボックスの中に「体重」と「身長」が登録される。



図10 一変量の分布の画面

そこで、「OK」をクリックする。すると、図 11 のように、2 つの変数の度数分布が同時に表示される。

体重と身長 の度数分布が、男性、女性でどのように異なっているかを知りたい場合に、簡単にその処理ができる。つまり、体重と身長 の度数分布を「性」の層別に描くことが簡単にできるのである。

図 10 の画面で「By」を利用する。「性」をクリックし、さらに「By」をクリックして、「OK」をクリックすると、図 12 のような表示になる。「性」が女性 (1) と男性 (2) ごとに「体重」と「身長」の度数分布が示されている。

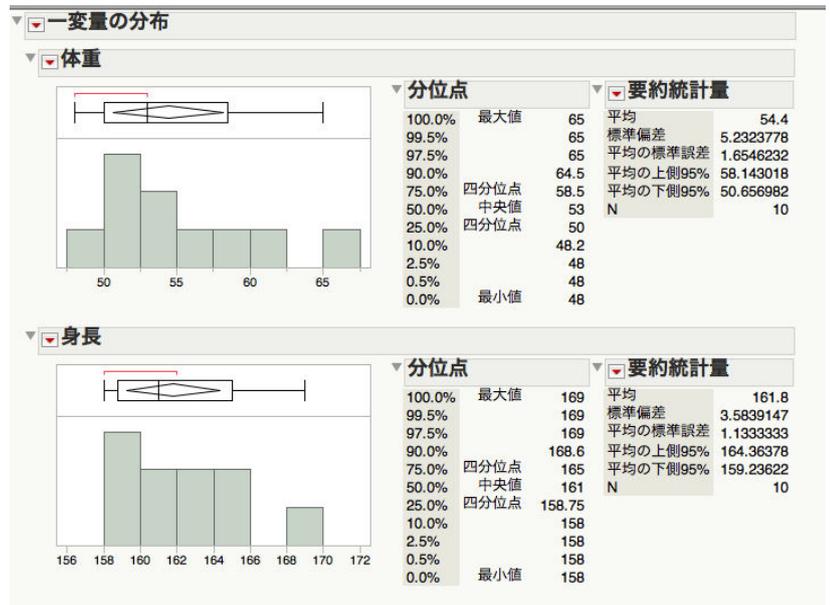


図 11 複数のヒストグラム

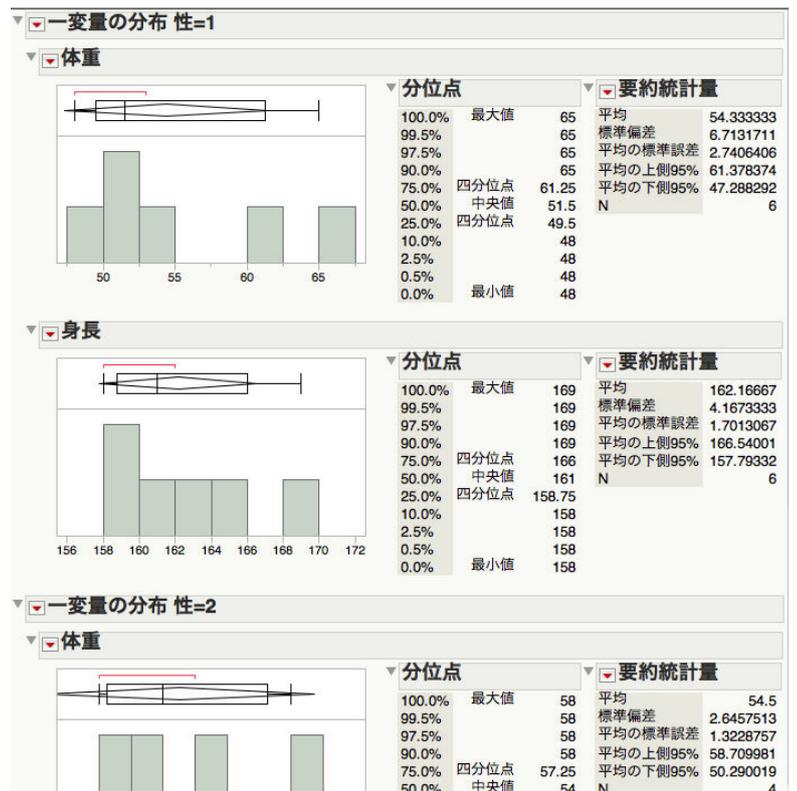


図 12 複数のヒストグラム (層別)

## 7. 変数間の相関：名義尺度

名義尺度の場合で変数間の相関を調べよう。「Q1」と「Q2」は、ともに名義尺度としよう。「Q1」は1と2の categorie をもつ。「Q2」は1、2、3の categorie をもっている。これらの2つの変数(項目)間に相関関係があるかどうかを調べる。

メニューの「分析/2変数の関係」をクリックする。すると図13の画面になる。

「9列」ウィンドウから「Q1」を選択し、「X, 説明変数」をクリックし、「説明変数」のウィンドウ内に入れる。「Q2」をクリックし、「Y, 目的変数」をクリックし、「目的変数」のウィンドウ内に入れる。

そして、「OK」をクリックする。

すると、図14のモザイク

図と図15の分割表が表示される。モザイク図は、Q1とQ2の周辺分布分とQ1とQ2のクロス分布を読み取ることができる。赤い部分がQ1の1の categorie を示している。青はQ1の2の categorie を示している。2変数の全6つの categorie の内、Q1で2で、Q2では2のクロス部分の割合がもっとも多いことがわかる。

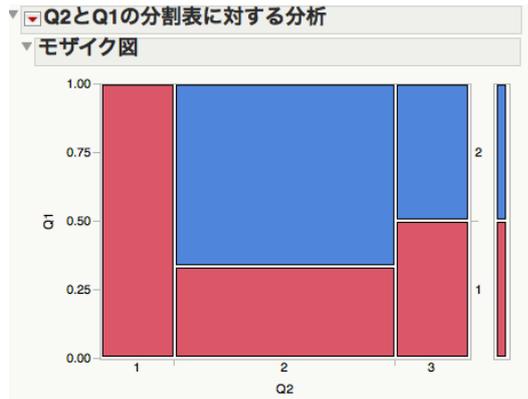


図14 モザイク図

		Q1		
		1	2	
Q2	1	20.00	0.00	20.00
	2	40.00	0.00	40.00
	3	10.00	10.00	20.00
		33.33	66.67	100.00

図15 分割表

この割合は面積で示されているため、最も面積の大きい部分である。

分割表には、カイ自乗検定による2変数間の独立性の検定結果も表示されている。

この分割表には2変数の水準ごとの度数に加え、全体、行、および列における各度数の割合が%で示されている。これから、Q2の水準が変化したときに、Q1の水準を選んでいる件数の割合が変化しているかどうかを知ることができる。すなわち、Q2の水準値が1のとき、Q1の水準値が1と2で、それぞれ100%と0%である。それに対して、Q2の水準値が2で、Q2の1と2の割合は、それぞれ33%と67%になっており割合が変化している。つまり、Q1の取りうる値はQ2の値に影響を受けていることがわかるのである。

このことを、統計検定をつかって、事前に設定した

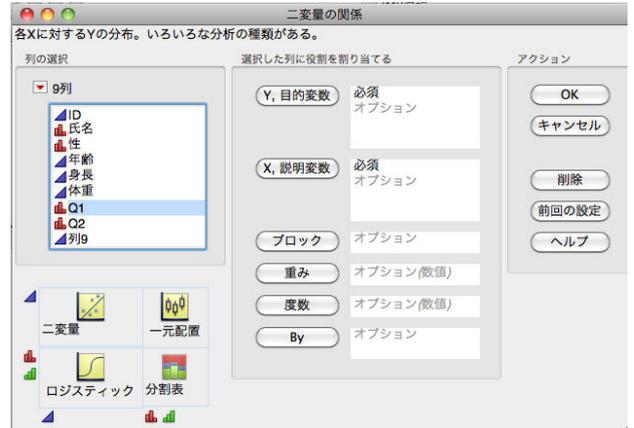


図13 二変数の関係の画面

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	3.452	0.1780
Pearson	2.667	0.2636

警告: セルのうち20%の期待度数が5未満です。カイ2乗に問題がある可能性があります。  
警告: 平均セル度数が5未満です。尤度比カイ2乗に問題がある可能性があります。

図16 検定結果

有意水準のもとで有意かどうかをみることができる。図 16 の検定結果の表がそれを示している。尤度比に関して p 値が示されている。p 値は 0.178 となっていることから、「Q1 と Q2 は互いに無関係である」という帰無仮説が、有意水準 5% で棄却されないことがわかる。有意水準 5% で、Q1 と Q2 はお互いに独立しているとの証明にはならないが、お互いに影響を及ぼし合っているとは言い切れないという結論になる。

## 8. 変数間の相関：連続尺度

つぎに、連続尺度間の相関を調べよう。身長と体重の間の相関をみる。

図 13 の画面において、「体重」をクリックして「目的変数」のウィンドウへ、「身長」をクリックして「説明変数」のウィンドウに登録する。

すると、連続尺度の場合は、散布図が描かれる。つぎの解析として、性別で身長と体重との関係が異なるかどうか気になる。そこで、性別ごとに、身長と体重の関係をみよう。ここでも「By」が利用できる。層別にわけることで、見えてくるものがある場合がある。図 17 のように「性」別に身長と体重の散布図が表示された。これで層別に比較することができる。

もし、ここで、「身長」と「体重」の関係のしかたに「性」別で違いができれば、「身長」と「体重」の関係に「性」の要因が影響を与えていることが示されたことになる。

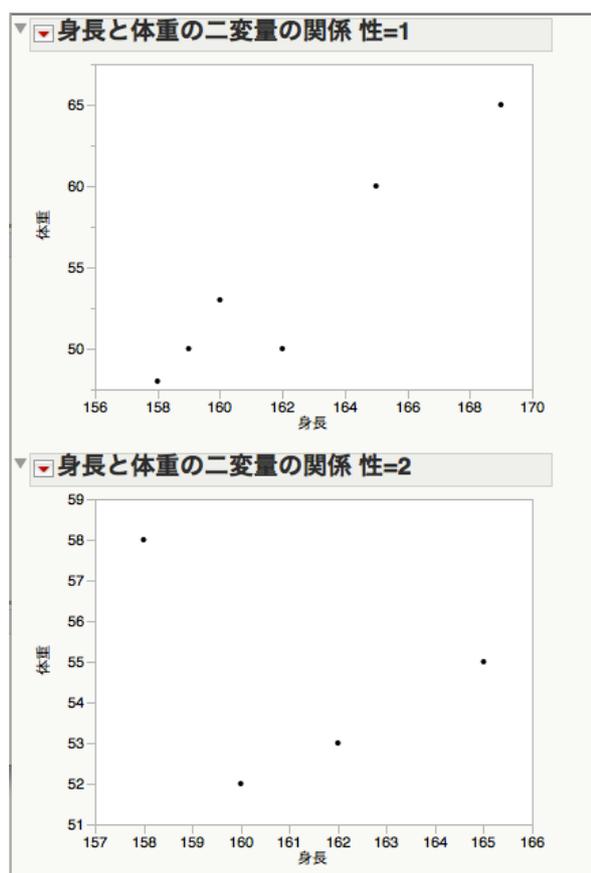


図 17 二変量の関係の結果

## 9. 「計算式」により新しい変数を生成する

「計算式」機能をつかい、行数を入力する新しい項目をつくろう。これは、ID など、通し番号をつける際に利用できる。データシートにおいて、新しい列でダブルクリックする。すると、図 18 のように、自動的に「列9」という名前の変数が追加され、新しい列が生成される。

その列をクリックして選択した上で、メニューの「列」をクリックするとリストが表示される。その中から「計算式...」を選ぶ。すると図 19 の表示になる。

右のウィンドウ内の「関数 (グルー

長	体重	Q1	Q2	列9
160	53	1	1	•
158	58	1	1	•
162	50	1	2	•
165	60	2	2	•
160	52	2	2	•
162	53	1	3	•
165	55	2	3	•
169	65	2	2	•
159	50	2	2	•
158	48	1	2	•

図 18 列の作成

ブ別)」のウィンドウ内の「行」をクリックする。すると表示されるリストの中から、「Row」をクリックする。最後に「OK」をクリックする。

すると、データシートの「列 9」の項目の列に、1 から始まる通し番号が入力された。

もう一つの例を説明しよう。「計算式」をつかって、体重と身長から、BMI の値をもつ項目を作成する。

BMI の定義式は

$$\text{BMI} = \text{体重} / \text{身長}^2$$

である。単位は体重が【Kg】、身長が【m】である。手元にもっているデータでは身長が【cm】であることに注意する。

新しい列をつくり、BMI という項目名としよう。そのためには新しい列でダブルクリックし、項目名のところでさらにダブルクリックし、ここに「BMI」と文字を入力する。

この列でクリックし、メニューの「列」をクリックする。表示されたリストから「計算式...」をクリックし、図 19 の画面を表示させる。

左の「テーブル列」ウィンドウの中にある「体重」をクリックする。すると中央下部にある大きな枠の中に「体重」が入力される。さらに、中央の演算子ボタンと「身長」ボタンを適切に選択指定し、図 20 のように BMI の定義式を作成する。2 乗は「x<sup>y</sup>」のボタンを利用する。このボタンは一般に n 乗など累乗を計算する際につかう。「身長」をクリックした後、「x<sup>y</sup>」をクリックし、その後、「2」をクリックすると、身長の 2 乗の標記ができる。その後、「OK」をクリックする。

すると、図 21 のように、BMI の項目に BMI 値が計算され入力された。

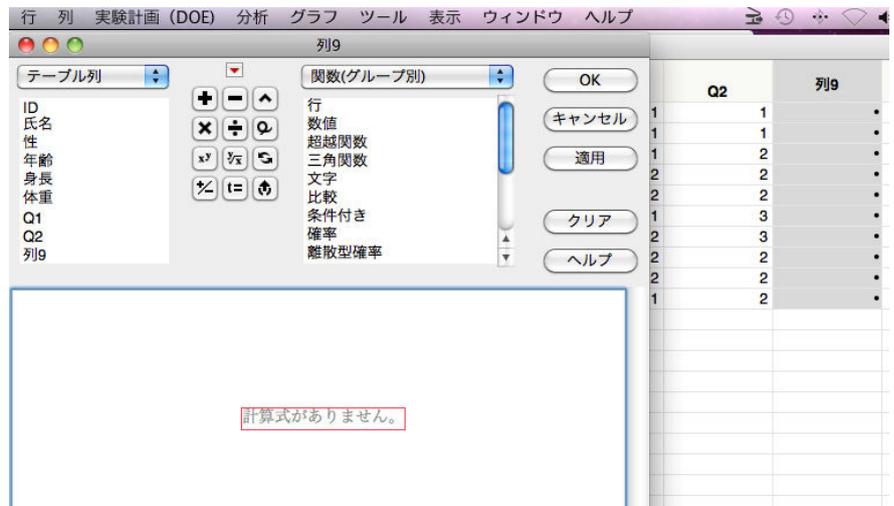


図 19 計算式の画面

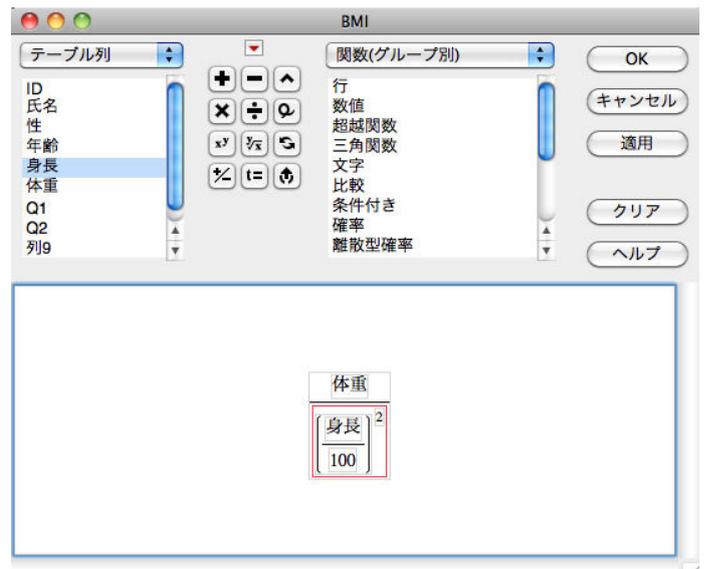


図 20 計算式の画面での式の入力

	身長	体重	Q1	Q2	列9	BMI
25	160	53	1	1	1	20.703125
26	158	58	1	1	2	23.233456177
30	162	50	1	2	3	19.051973784
27	165	60	2	2	4	22.038567493
32	160	52	2	2	5	20.3125
21	162	53	1	3	6	20.195092212
22	165	55	2	3	7	20.202020202
35	169	65	2	2	8	22.758306782
36	159	50	2	2	9	19.777698667
32	158	48	1	2	10	19.227687871

図 21 BMI 値の入力結果

## 10. 分割表による入力

通常での分割表の表示では、データベース形式で入力されているデータに対して、分割表による解析などがおこなわれる。しかし、場合によっては、すでに分割表のデータがあって、このデータをつかって、2つの因子間の独立性をもとめたい場合が少なくない。じつは、JMPには直接分割表が作成できる入力方法が用意されている。

図 22 のような「熱」と「風邪」の変数をもつ分割表のデータを例に説明をしよう。「A1」は熱がある、「A2」は熱がない。「B1」は風邪を引いている、「B2」は風邪を引いていない、としている。

		風邪		
		B1	B2	
熱	A1	10	5	15
	A2	4	8	12
		14	13	

図 22 分割表

まず、新規データシートを作成する。項目として名義尺度の「熱」と名義尺度の「風邪」を作成する。「熱」の水準の A1 と A2 として、それぞれ「1」と「2」をつかおう。「風邪」の水準の B1 と B2 として、それぞれ「1」と「2」をつかおう。また、このクロス表の度数の値は「度数」として、図 23 のように入力する。

名称未設定	熱	風邪	度数
1	1	1	10
2	1	2	5
3	2	1	4
4	2	2	8

図 23 データシートへの入力

メニューの「分析」／「二変数の関係」をクリックすると、図 24 の画面が現れる。この画面で「熱」を説明変数、「風邪」を目的変数の枠に入力する。さらに、度数を「度数」の枠に入力する。その後、「OK」をクリックする。

すると、図 25 の表示になる。ここには、モザイク図と、入力された分割表が表示される。さらに、ここには、図 26 に示されている表で、



図 24 二変数の関係の画面

2 変数間の独立の検定結果のカイ自乗値や p 値が表示される。

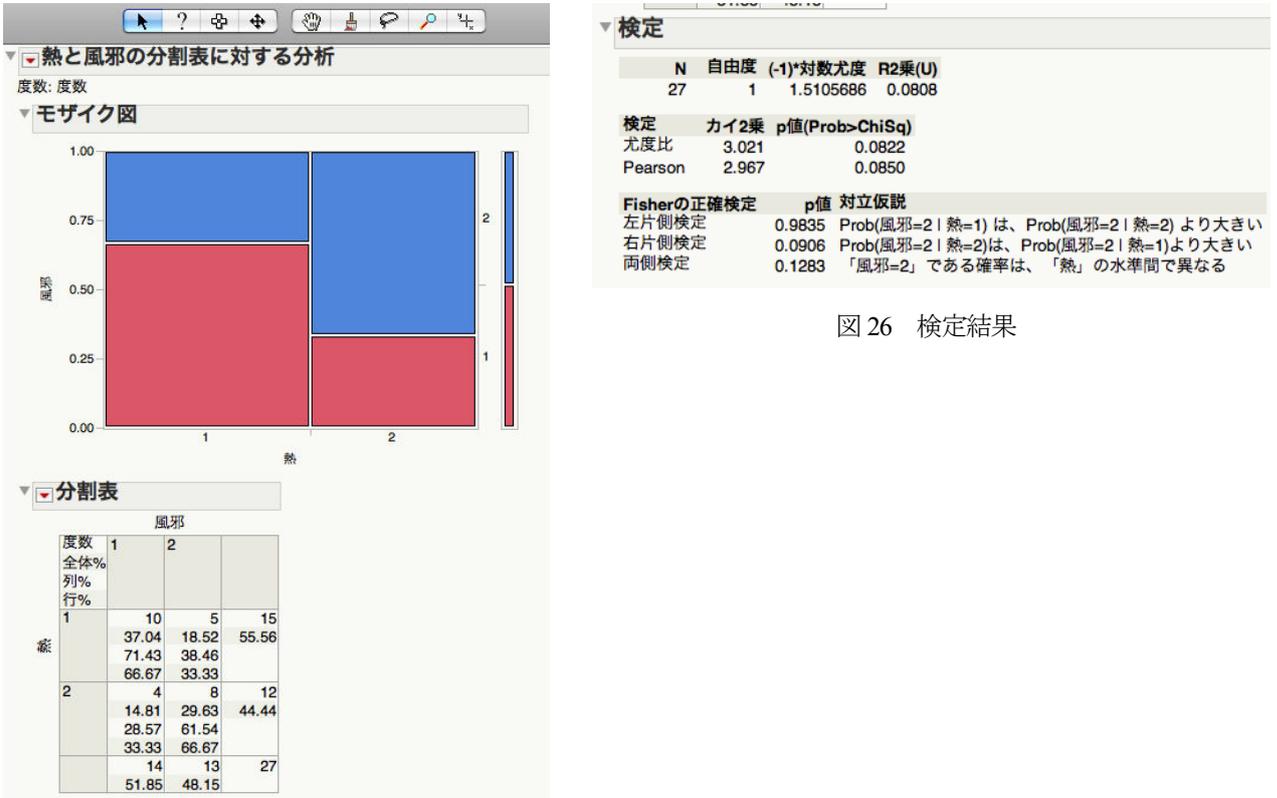


図 25 モザイク図と分割表

検定	N	自由度	(-1)*対数尤度	R2乗(U)
	27	1	1.5105686	0.0808
尤度比			3.021	0.0822
Pearson			2.967	0.0850
<b>Fisherの正確検定</b>			<b>p値</b>	<b>対立仮説</b>
左片側検定			0.9835	Prob(風邪=2   熱=1) は、Prob(風邪=2   熱=2) より大きい
右片側検定			0.0906	Prob(風邪=2   熱=2) は、Prob(風邪=2   熱=1) より大きい
両側検定			0.1283	「風邪=2」である確率は、「熱」の水準間で異なる

図 26 検定結果

## 11. ノンパラメトリック変数の 2 群間での比較

「Q2」が、ここでは、順序尺度として、男女間で「Q2」の回答に差があるかどうかを調べよう。すなわち、たとえば、「1」が「よかった」、「2」が「ふつう」、「3」が「悪かった」というように、回答に順序をもつ場合である。

まず、Q2 を連続尺度に切り替える必要がある。「Q2」を図 27 の表示に切り替える。



図 27 変数リスト

メニュー「分析」をクリックし、「二変数の関係」をクリックして指定する。

すると、図 13 の画面が表示される。ここで、「Q2」を「Y,目的変数」に、「性」を「X,説明変数」に指定し、それぞれの枠に登録する。そこで、「OK」をクリックすると、図 28 のような表示が現れる。これは、2 群間で連続変数の平均値の差を検定する場合の表示と同じである。この段階から、さらに操作を進める。

サブメニューを表示するために、「性による Q2 の一元配置分析」と書かれている項目の左にある赤

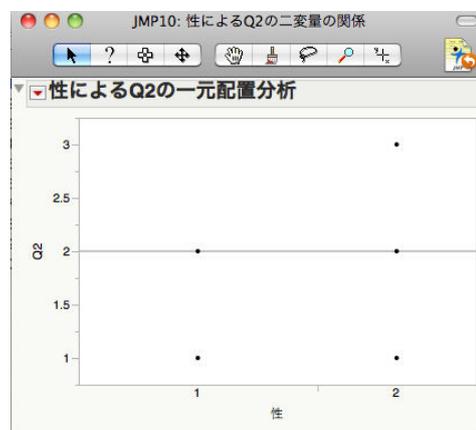
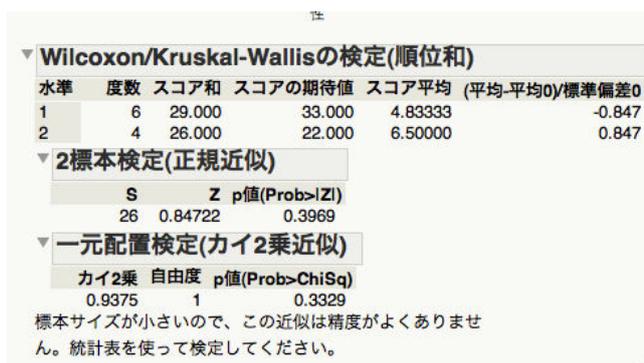


図 28 一元配置分析の画面

い▽マークをクリックする。

表示されるリストの中から、「ノンパラメトリック」／「Wilcoxon 検定」を選ぶ。

すると、図 29 の解析結果の表がグラフの下に追記される。



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
1	6	29.000	33.000	4.83333	-0.847
2	4	26.000	22.000	6.50000	0.847

S	Z	p値(Prob> Z )
26	0.84722	0.3969

カイ2乗	自由度	p値(Prob>ChiSq)
0.9375	1	0.3329

標本サイズが小さいので、この近似は精度がよくありません。統計表を使って検定してください。

図 29 Wilcoxon の検定結果

## 12. おわりに

JMP などの統計ソフトウェアを使えば、ほとんどクリック操作だけで、簡単に統計解析ができる。そのため、研究の解析部分をすばやく進めることができる。一方、それだけではなく、簡単に解析結果をだすことができるために、解析ソフトウェアをつかうことで統計学の勉強をすることが可能となっている。

統計学の勉強には、データ数を極端に少なくすることで、解析においてどのような操作がなされているかをかなり身近に感じることができるのではないかと思う。そういう意味もあって、本稿では、少ないデータ数で解析方法を説明してみた。解析を進める際には、データのグルーピングが非常に重要になる。どのような方針でグループ分けをするのかなどをみる際にも少数のデータでの操作での練習は役にたつように思う。

卒業研究の終盤において、学生が統計ソフトウェアをつかう操作の速さには目をみはるものがある。身近に、自由に高機能の統計ソフトウェアがつかえる環境があるのは、とてもありがたく感じる。

## 参考文献

- [1] 田久浩志、林俊克、小島隆矢、JMP による統計解析入門第 2 版。Ohmsha、2010.