

[利用の手引き]

統計解析ソフトウェア SPSS の操作入門

犬塚裕樹[†]

Hiroki Inutsuka[†]

[†] 久留米大学 医学部看護学科

[†] Kurume University School of Nursing

1. はじめに

本学医学部には、コンピュータが利用できるコンピュータ実習室がある。そこには、学内 LAN に接続されたパソコンが 140 台ほどあり、学部、院生に対する毎日の講義の他、教員と学生、さらには病院の医療従事者などの職員の研究にも利用されている。コンピュータ操作一般について困っている場合には、現場に常駐の情報・コンピュータに詳しい同室の管理者が対応してくれる。このような利用サポートも、多くの利用者の存在の一因となっている。

このコンピュータ実習室で最も多く利用されるソフトウェアの種類の 1 つに統計ソフトウェアがある。特に、JMP と SPSS が利用されている。

本稿では、SPSS の操作の基礎について、次の項目を説明する。

- (1) EXCEL からのデータの読み込み
- (2) 変数へのラベル付け
- (3) 度数分布のグラフ表示
- (4) 頻度分布のグラフ表示
- (5) データの選択
- (6) 2 つの変数間の関連
- (7) 2 グループ間での順序変数データ頻度分布の違いの検出

2. EXCEL からのデータの読み込み

SPSS に入力する方法として、データが入力された EXCEL ファイルを SPSS に読み込ませることが可能である。図 1 のように、EXCEL ファイルに、あるアンケート調査の回答の結果として、10 件のデータを入力した場合をアつかう。これは、「ID」、「性」など 8 つの項目をもっている。実際の多くのデータ解析でつかわれる数に比べて、極端に少ない件数で説明しよう。

	A	B	C	D	E	F	G	H
1	ID	性	年齢	体重	身長	Q1	Q2	Q3
2	1	1	20	70	160	1	4	3
3	2	1	22	7	155	2	4	3
4	3	1	18	75	158	2	4	4
5	4	2	18	80	148	1	3	4
6	5	2	25	83	160	1	4	3
7	6	1	30	70	164	1	2	4
8	7	1	28	68	155	2	3	3
9	8	2	24	65	157	1	2	3
10	9	2	23	68	159	2	4	4
11	10	2	23	70	165	2	4	4
12								

図 1 データが入力された EXCEL ファイル

SPSS を起動し、メニューの「ファイル」をクリックし、「開く」をクリックして、ファイルウィンドウを開く。この EXCEL ファイ

ルを指定し、OK のボタンをクリックする。

すると、図 2 のような画面になる。これは「データ ビュー」表示である。EXCEL 画面と同じような表示となっている。ウィンドウの左下部にある「データ ビュー」と「変数 ビュー」のスイッチボタンをクリックすることで、これらの 2 つの画面を切り替えることができる (図 3)。

	ID	性	年齢	体重	身長	Q1	Q2	Q3
1	1.0	1.0	20.0	70.0	160.0	1.0	4.0	3.0
2	2.0	1.0	22.0	7.0	155.0	2.0	4.0	3.0
3	3.0	1.0	18.0	75.0	158.0	2.0	4.0	4.0
4	4.0	2.0	18.0	80.0	148.0	1.0	3.0	4.0
5	5.0	2.0	25.0	83.0	160.0	1.0	4.0	3.0
6	6.0	1.0	30.0	70.0	164.0	1.0	2.0	4.0
7	7.0	1.0	28.0	68.0	155.0	2.0	3.0	3.0
8	8.0	2.0	24.0	65.0	157.0	1.0	2.0	3.0
9	9.0	2.0	23.0	68.0	159.0	2.0	4.0	4.0
10	10.0	2.0	23.0	70.0	165.0	2.0	4.0	4.0
11								
12								

図 2 SPSS のデータビュー画面

3. 変数へのラベル付け

「変数 ビュー」をクリックすると、図4の画面になる。この画面では、変数ごとに、「変数の型」、「数値の桁数」、「小数桁数」、「ラベル」、「値のラベル」、「欠損値の指定」、さらに尺度の型などが設定できる。

変数名の「性」には性別を入力している。女性が1、男性を2とコード化する場合を説明する。2行目の変数「性」において、「ラベル」と「値」を入力しよう。



図3 データビューと変数ビューのボタン

名前	型	幅	小数桁数	ラベル	値	欠損値	列	配置	尺度	役割
1 ID	数値	12	1		なし	なし	12	右	名義(N)	入力
2 性	数値	12	1		なし	なし	12	右	名義(N)	入力
3 年齢	数値	12	1		なし	なし	12	右	スケール(S)	入力
4 体重	数値	12	1		なし	なし	12	右	名義(N)	入力
5 身長	数値	12	1		なし	なし	12	右	スケール(S)	入力
6 Q1	数値	12	1		なし	なし	12	右	名義(N)	入力
7 Q2	数値	12	1		なし	なし	12	右	名義(N)	入力
8 Q3	数値	12	1		なし	なし	12	右	名義(N)	入力
9										

図4 変数ビュー画面

図4の「ラベル」の列の2行目に「性」と入力し、同じ行の「値」の列のセルをクリックする。すると、図5のような「値ラベル」のウィンドウが開く。

「値(U)」に「1」、「ラベル(L)」に「女性」と入力する。その後、「追加(A)」をクリックする。同様に、男性についても「値」と「ラベル」を入力する。すると、図5の表示になる。そこで再度「追加(A)」をクリックする。

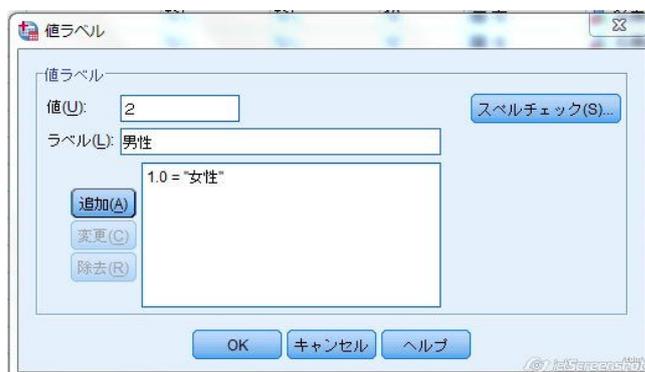


図5 値ラベル画面

そこで、「OK」ボタンをクリックすると、図6の変数ビュー画面が表示される。

	名前	型	幅	小数桁数	ラベル	値	欠損値	列
1	ID	数値	12	1		なし	なし	12
2	性	数値	12	1	性	{1.0, 女性}...	なし	12
3	年齢	数値	12	1		なし	なし	12
4	体重	数値	12	1		なし	なし	12
5	身長	数値	12	1		なし	なし	12
6	Q1	数値	12	1		なし	なし	12
7	Q2	数値	12	1		なし	なし	12
8	Q3	数値	12	1		なし	なし	12
9								
10								
11								

図 6 変数ビュー画面

変数「ID」、「性」、「Q1」、「Q2」、「Q3」の入力値は整数値であるために、「小数桁数」を1から0に変更する。その結果が図7である。

	名前	型	幅	小数桁数	ラベル	値	欠損値	列
1	ID	数値	12	0		なし	なし	12
2	性	数値	12	0	性	{1, 女性}...	なし	12
3	年齢	数値	12	0		なし	なし	12
4	体重	数値	12	1		なし	なし	12
5	身長	数値	12	1		なし	なし	12
6	Q1	数値	12	0		なし	なし	12
7	Q2	数値	12	0		なし	なし	12
8	Q3	数値	12	0		なし	なし	12
9								

図 7 変数ビュー画面

ここで、「データ ビュー」ボタンをクリックして切り替えると図8のようになる。

	ID	性	年齢	体重	身長	Q1	Q2	Q3
1	1	1	20	70.0	160.0	1	4	3
2	2	1	22	72.0	155.0	2	4	3
3	3	1	18	75.0	158.0	2	4	4
4	4	2	18	80.0	148.0	1	3	4
5	5	2	25	83.0	160.0	1	4	3
6	6	1	30	70.0	164.0	1	2	4
7	7	1	28	68.0	155.0	2	3	3
8	8	2	24	65.0	157.0	1	2	3
9	9	2	23	68.0	159.0	2	4	4
10	10	2	23	70.0	165.0	2	4	4

図 8 データビュー画面

4. 度数分布のグラフ表示

変数「性」に関する回答数の度数をグラフ作画しよう。棒グラフで表示することにする。

まず、メニューの「グラフ(G)」をクリックする。表示されるリストの中の「レガシーダイアログ(L)」をクリックする。

つぎに、サブリストの中の「棒」をクリックする。すると、図9のウィンドウが表示される。

図9の「単純」をクリックし、下部にある「定義」ボタンをクリックする。すると、図10が現れる。



図9 棒グラフ・ウィンドウ

図10の「カテゴリ軸」窓に変数「性」を入力する。すなわち、左端の窓のリストの「性」をクリックし、「カテゴリ軸」窓の左のエンターボタンをクリックする。すると図11のようになる。

そこで、図11の「棒の表現内容」欄において、「ケースの数」にチェックがはいつていることを確認した上で、「OK」ボタンをクリックする。

すると、図12のように「性」の度数が棒グラフで表示される。



図10 単純棒グラフ・ウィンドウ



図11 カテゴリ軸への入力

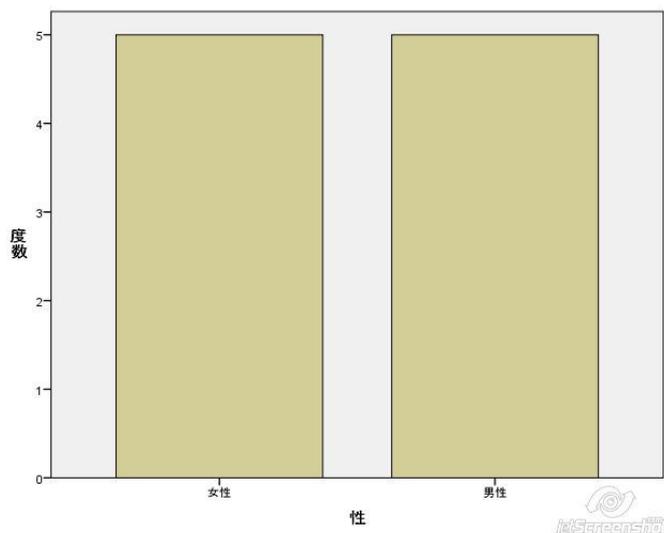


図 12 性ごとの度数分布

なお、「性」に関して、データビューでは、「1」と「2」の数値を入力している。しかし、変数ビューにおいて、ラベルの「値」には、1は「女性」、2は「男性」と設定しているため、このグラフ中では、数字のかわりに「女性」と「男性」で表示されている。この機能によりわかりやすい表示となっている。

5. 頻度分布のグラフ表示

つぎに、変数「性」の度数を、全データ数に対する割合で表したグラフを作成しよう。割合での表示をおこなうためには、図 11 において、「ケースの%」をクリックし、「ケースの数」から図 13 のように変更する。その後、「OK」ボタンをクリックする。

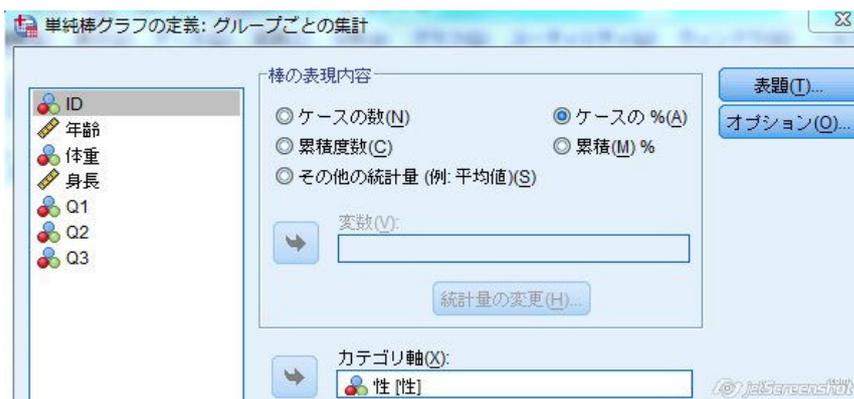


図 13 単純グラフウィンドウ

すると、図 14 のように、棒グラフの縦軸が、パーセントに変更されたグラフが作成される。

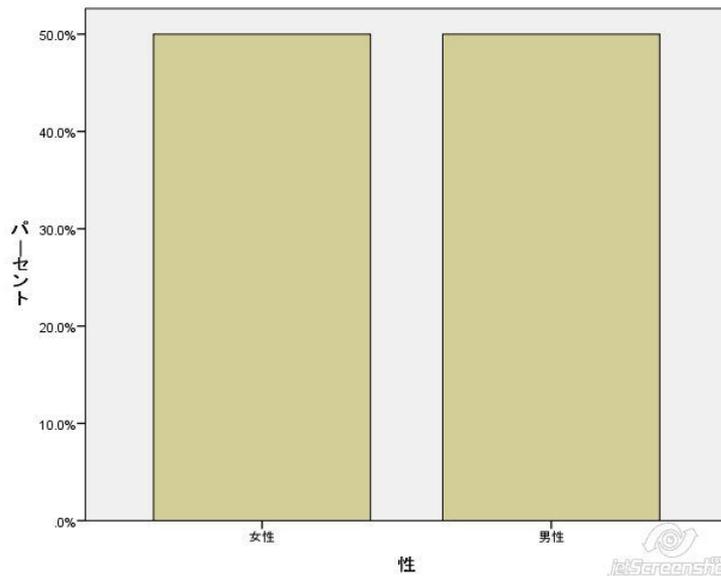


図 14 性ごとの割合の分布

6. データの選択

変数「Q1」は、アンケートの質問項目に対して「はい」と「いいえ」で答えている回答結果である。「はい」に「1」、「いいえ」に「2」の数値をわりあてている。なお、同様の操作で「Q2」の回答結果の頻度分布も作成することができる。

ここでは、「女性」にしぼって Q1 の回答結果を表示しよう。全データのうち、女性だけのデータを選択することができる。他の「男性」のデータは除外して分析する。

メニューの「データ」をクリックし、「ケースの選択 (S)」をクリックする。すると図 15 が表示される。

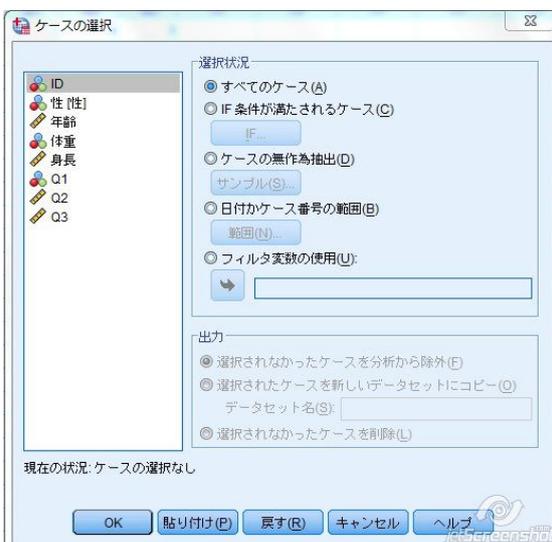


図 15 ケースの選択ウィンドウ

この画面において、「IF 条件が満たされるケース」をクリックし、「IF」ボタンをクリックする。すると図 16 の画面が表示される。左窓のなかの「性」をクリックし、中央の窓に入力する。その後、「=」ボタンをクリックし、女性のコードである「1」を入力する。そして、「続行」ボタンをクリックする。

データビューの表示では、図 17 のようになる。斜線が表示され、分析から除外されるデータが示されている。この段階で、「Q1」の頻度分布を表示させる操作をおこなおう。前述のグラフ表示の操作をおこなうと、図 18 が表示される。



	ID	性	年齢	体重	身長	Q1	Q2	Q3	filter_\$
	1	1	20	70.0	160.0	1	4	3	1
	2	1	22	72.0	155.0	2	4	3	1
	3	1	18	75.0	158.0	2	4	4	1
	4	2	18	80.0	148.0	1	3	4	0
	5	2	25	83.0	160.0	1	4	3	0
	6	1	30	70.0	164.0	1	2	4	1
	7	1	28	68.0	155.0	2	3	3	1
	8	2	24	65.0	157.0	1	2	3	0
	9	2	23	68.0	159.0	2	4	4	0
	10	2	23	70.0	165.0	2	4	4	0
	11								

図 16 データビュー表示でのデータ選択の結果

図 17 IF 条件の定義ウィンドウ

この図から、女性のうち、「Q1」で「1」と回答した割合は 40%であったことがわかる。

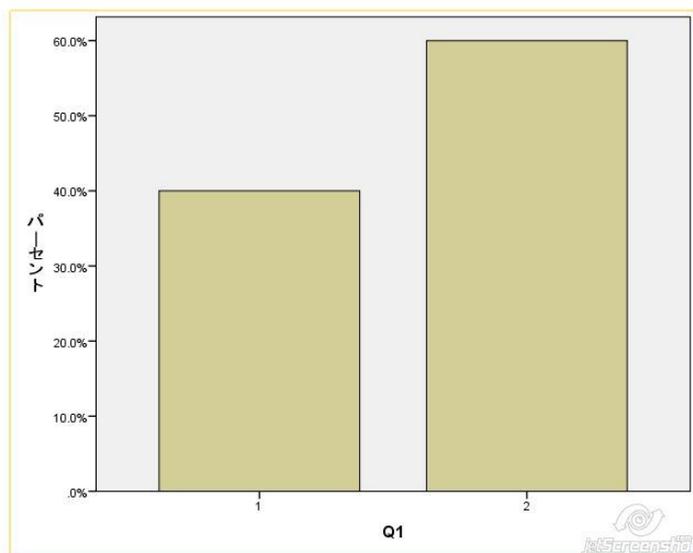


図 18 「Q1」で「1」と回答した割合

7.2 つの変数間の関連

変数「性」と「Q1」の間に統計的関連があるかどうかを調べよう。そのためには、2×2 のクロス表を利用できる。統計的関連があるかどうか、ということは、表現を変えれば、性別間で、「はい」と「いいえ」の回答の頻度に違いがあるかどうかということである。

メニューの「分析」をクリックし、サブリストの「記述統計 (E)」／「クロス集計表 (C)」をクリックする。すると、図 19 が表示される。

そこで、図 19 の「行」に「性」をクリックで指定し、「列」には「Q1」をクリックで指定する。最後に、「OK」ボタンをクリックする。

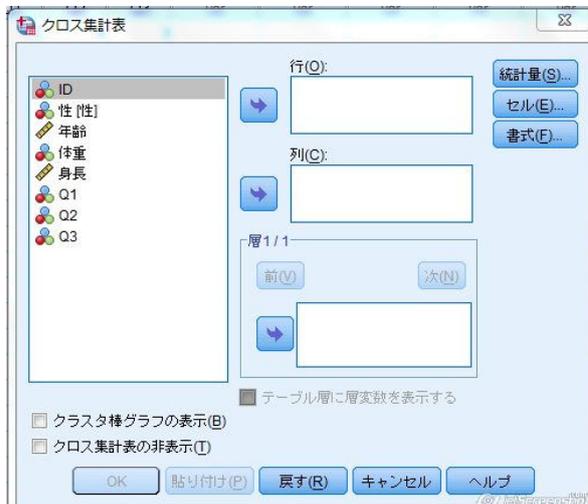


図 19 クロス集計表ウィンドウ

すると、図 20 が表示される。ここには、「性」と「Q1」のそれぞれの周辺度数分布が示されている。女性と男性は、どちらも 5 人である。そして、「Q1」において「1」と回答しているのは 5 人である。さらに、「性」と「Q1」の結合度数が示されている。たとえば、「女性」が「Q1」の質問に対して「1」と回答した人数は 2 人で、「2」と回答した人数が 3 人であることがわかる。

➡ クロス集計表

[データセット1] C:\Users\hinutsu\Desktop\SPSS原稿作成\SPSS#無題2.sav

処理したケースの要約						
	ケース					
	有効数		欠損		合計	
	N	パーセント	N	パーセント	N	パーセント
性 * Q1	10	100.0%	0	0.0%	10	100.0%

性と Q1 のクロス表				
度数		Q1		合計
		1	2	
性	女性	2	3	5
	男性	3	2	5
合計		5	5	10

図 20 クロス集計表

このクロス表から、「性」と「Q1」の間の、統計的な関連性について、統計検定をおこなうことができる。統計検定は、カイ 2 乗検定がつかえる。

図 19 において、「統計量 (S)」 ボタンをクリックする。すると、図 21 のようなウィンドウが表示される。



図 21 統計量の指定

図 21 のように、「カイ 2 乗(H)」にチェックを入れる。その後、「続行」をクリックする。

すると、図 19 に戻る。このウィンドウにおいて、「セル(E)」 をクリックすると図 22 が現れる。このウィンドウで、「パーセンテージ」のグループの「行」にチェックを入れておく。これによって行の数値データに割合の%が表示される。

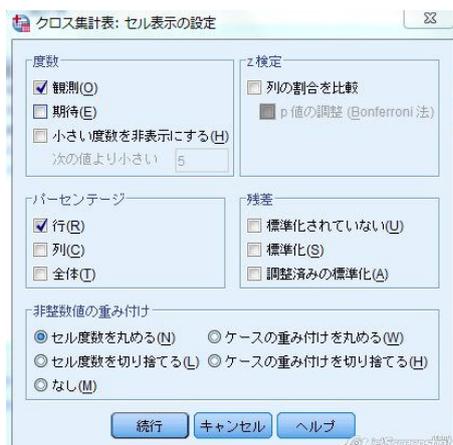


図 22 セル表示の設定

その後、「続行」をクリックして、ウィンドウを閉じる。図 19 の画面にもどったら「OK」 ボタンをクリックする。

その結果が図 23 の表である。カイ 2 乗検定の表には、漸近有意確率（両側）が 0.527 と示されている。有意水準が 0.05 において有意とはなっていない。

すなわち、有意水準 5%で、「性」と「Q1」の間には関連があるとはいえない、つまり、「性」によって「Q1」の回答結果の割合に違いはないという結論となる。ただし、この結果は、「性」と「Q1」はどんな関連

もないということを証明したことにはならない、ということに注意しなければならない。

本来、これらの 2 つの変数間には関連があるとしても、データ数が少ない場合、有意に関連があるとの結論にならない。この分析では、データ数が少ないために有意とはなっていないものと思われる。もともと、「性」と「Q1」に関連があるかどうかとは、どうということなのかを理解しておくことが重要である。

「Q1」の回答結果に影響を及ぼす因子が「性」であるかどうかを調べる場合に、この分析は有用である。

性と Q1 のクロス表

		Q1		合計
		1	2	
性	女性	度数 2	3	5
		性の % 40.0%	60.0%	100.0%
	男性	度数 3	2	5
		性の % 60.0%	40.0%	100.0%
合計		度数 5	5	10
		性の % 50.0%	50.0%	100.0%

カイ 2 乗検定

	値	自由度	漸近有意確率 (両側)	正確有意確率 (両側)	正確有意確率 (片側)
Pearson のカイ 2 乗	.400 ^a	1	.527	1.000	.500
連続修正 ^b	.000	1	1.000		
尤度比	.403	1	.526		
Fisher の直接法					
線型と線型による連関	.360	1	.549		
有効なケースの数	10				

a. 4 セル (100.0%) は期待度数が 5 未満です。最小期待度数は 2.50 です。

b. 2x2 表に対してのみ計算



図 23 カイ 2 乗検定結果

8.2 グループ間での順序変数データ頻度分布の違いの検出

Q2 の回答結果は順序変数とする。1、2、3、4 というように反応の強さの順でコード化されている。1 が「強く同感している」、2 は「同感している」、「3 は「同感していない」、4 は「全く同感していない」といった具合である。

これらの反応が性別で異なるのかどうかを調べよう。2 グループ間での順序変数の分布の違いを調べるので、Mann-Whitney (マンホイットニー) 検定がつかえる。帰無仮説は「女性、男性の 2 つのグループ間で、Q2 に対する回答の頻度分布は等しい」である。

メニューの「分析」をクリックし、表示されるサブリストから「ノンパラメトリック検定 (N)」をクリックする。さらに表示されるリストから「独立サンプルの検定 (I)」をクリックする。

すると、図 24 の画面が表示される。「分析のカスタマイズ」をクリックする。その後、「フィールド」のタブをクリックする。表示されるつぎの画面で、これまでの操作と同様に、グループ窓に「性」をマウスで移動する。さらに検定フィールドに「Q2」をマウスで移動する。

つぎに「設定」のタブをクリックする。すると、図 25 の画面が表示される。

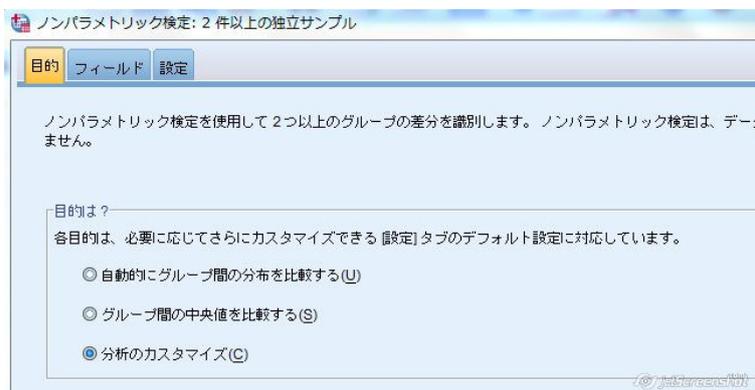


図 24 2 件以上の独立サンプル：目的



図 25 2 個以上の独立したサンプル：設定

ここで、図 25 の「検定のカスタマイズ」をクリックし、Mann-Whitney の U (2 サンプル) にチェックをいれる。その後、ウィンドウの下部にある「実行」ボタンをクリックする。

すると、図 26 の検定結果が示される。ここでは、帰無仮説が有意水準 5% で棄却されなかったことが示された。つまり、「Q2」の分布は「性」のカテゴリについて差があうとはいえない結果である。

ノンパラメトリック検定

[データセット1] C:\Users\hinutusu\Desktop\SPSS原稿作成\SPSS\無題2.sav

仮説検定の要約			
帰無仮説	テスト	有意確率	決定
1 Q2 の分布は 性 のカテゴリで同じです。	独立サンプルによる Mann-Whitney の U の検定	1.000 ¹	帰無仮説を採択します。

漸近的な有意確率が表示されます。有意水準は .05 です。

¹ この検定の正確な有意確率が表示されます。

図 26 検定結果

9. おわりに

最近では、マウス操作だけで簡単に分析できる統計解析ソフトウェアが広く普及し、統計解析はとても便利になった。しかし、医療分野の学術論文を読んで気になることがしばしばある。

基礎的な統計の部分でおかしな使い方がされていることがある。たとえば、明らかな順序変数に対して「ノンパラメトリック分析をおこなう」と論文の初めに述べておきながら、結論的な重要なところでは、平均値や標準偏差を求めて、それらの比較をして結論を導いている。用語の問題では、「有意水準は $p < 0.05$ とする」という決まった同一の記載がしばしばみられる。この記載には少なからず違和感がある。たいしたことではないものの、おやっと思ってしまう。

統計解析の便利なソフトウェアが出てきても、数式をつかった統計学の基本の理解はぜひ必要ではないかと考えている。

参考文献

[1] SPSS でやさしく学ぶアンケート処理 [第4版] 石村貞夫、加藤千恵子、劉チェン 東京図書 (2015)