統計解析ソフトウェア「SYSTAT」の利用法

医学部・地域医療科学教育研究センター 竹生政資

takefu@cc.saga-u.ac.jp

1.はじめに

統計解析ソフトウェアは,医学および看護学の研究において欠かすことのできないツー ルの一つであり,2004年2月にリプレースされた医学サブセンターのすべての教育研究用 クライアント PC端末には「SYSTAT」という統計解析ソフトウェアがインストールされて いる.

最も広く利用されている統計解析フトウェアは, おそらく Microsoft Excel であろう. もちろんこれは,統計解析「専用」のソフトウェアではないが,基本的な統計解析機能が ほぼすべて提供されている上に,データの入力や整理を行うための表計算機能が充実して おり,さまざまなデータ形式のファイルを読み込んだり書き出したりすることもでき,少 なくとも「一次的な」データの加工ツールとして使うには非常に便利である.

Microsoft Excel の「分析ツール」には以下の統計解析機能が提供されている.

基本統計量分析,ヒストグラム分析,順位と百分位数分析

分散分析,相関分析,共分散分析,回帰分析

t 検定, F 検定, z 検定

「分析ツール」では,データの範囲指定や解析パラメータなどを GUI を使って「対話的」 に設定することができ,初心者でも簡単に利用することができる.また,「分析ツール」に は含まれていないが,ワークシート上で直接関数を入力して利用できる統計関数もある. 例えば, χ^2 検定(CHITEST 関数)やその他いろいろな統計分布関数の計算などはワーク シート関数として提供されている.

このように Microsoft Excel は,統計解析フトウェアの代用として使うことも十分可能で あるが,ただこれは統計解析専用のソフトウェアではないので,基本的な統計解析機能し か提供されていない.例えば,ノンパラメトリック分析,多変量解析,生存分析などの医 学および看護学の研究でしばしば必要とされる重要な機能は含まれておらず,これらの解 析を行なうためには,SYSTAT,SPSS,JMP,SAS などの統計解析専用のソフトウェアを 利用する必要がある.

1999 年 2 月に導入された医学サブセンターの旧システムでは,StatView という統計解 析ソフトウェアを導入した.StatView は,価格が手ごろで,医学および看護学の研究にと って必要十分な統計解析機能を備えており,Macintosh 版と Windows 版の両方があり,ま たすべての統計解析が統一された GUI で「対話的」に「やさしく」利用することができる ので,当時,本学の医学部ユーザの間ではもっとも人気が高かった.2004 年 2 月にリプレ ースされた新システムでも「StatView」を導入したいと思ったが,残念ながら,新システ ムの入札仕様策定の時点ですでに販売中止となっており,代わって「SYSTAT」を導入する こととなった.

以下では, Microsoft Excel では解析できないノンパラメトリック分析の例として, 「Wilcoxon 符合付順位和検定」と「Wilcoxon 順位和検定」の二つを紹介する.その前にまず, Microsoft Excel で作成されたデータをSYSTAT に取り込む方法について説明しよう.

2 . Microsoft Excel のデータを SYSTAT に取り込む方法

データの入力や編集を行う機能は SYSTAT にも用意されているが,おそらく多くのユー ザは,データの入力や編集などの「一次的な」データの加工ツールとしては Microsoft Excel を利用していると思われるので,まずはじめに Microsoft Excel で入力されたデータを SYSTAT に取り込む方法について説明する.以下の二つの方法がある.

まず一つの方法は,次のように Microsoft Excel のデータを「コピー」してから,SYSTAT に直接「貼り付け」を行う.これが最も簡単で手っ取り早い方法である.

- SYSTAT を起動して「ファイル」メニューから「新規作成」をクリックし,続いて「デ ータ」をクリックする(図1).
- (2) データウィンドウが開く(図2). このウィンドウは解析データを入力するためのもの である.
- (3) Microsoft Excelを起動し、データの入力されているワークシートを表示させ、取り込みたいデータの領域(複数の行と列にまたがってもよい)を範囲指定してから、 Microsoft Excelの編集メニューから「コピー」を実行する.
- (4) 次に SYSTAT のデータウィンドウ(図2)の中のコピー場所をクリックしてから、このデータウィンドウの「編集」メニューから「貼り付け」を実行する.

🔠 無題 - SYSTAT アウトプ	ット オーガナイ	ቻ					
ファイル(<u>F</u>) 編集(E) 表示	R∭ データ	<u>.D) グラフ(G</u>)	統計解析(S)	ヘルプ(日)			
新規作成(<u>N</u>)	•	データ(<u>D</u>)		🖄 🌒 ha	瞳 感 軸	м 🛛 🗌 н.	- 園 山 瓜 😔 図 🖉
開<(_)	+	コマンド(<u>C</u>)					
☐ 保存(S)	Ctrl+S	アウトプット	() B 2	′⊻∣≣	き き 信		RIE 3D
名前を付けて保存(A)]						
🗊 すべてを保存(L)							
データを保存(D)	Otrl+D						
受 データベースの読み込み((<u>B</u>)						
♪ クリップボードの実行(C)							
🛃 ファイルの実行(E)							
🗐 ウ心ドウを実行(型)							
🗊 現在行から最終行までを	実行の						
ページ設定(U)							
[∂_ 印刷プレビュー(⊻)							
	Ctrl+P						
	•						
最近使ったコマンド(工)	+						
最近使った出力(E)	+						
							-
	11						

図1.データの新規作成

もう一つの方法は, Microsoft Excel のファイル名とその中の ワークシート名を指定して取り 込む方法である. Microsoft Excel ではデータの入力はワー クシートとよばれる格子状の表 の中で行うが,ひとつのファイル (Microsoft Excel の用語では 「ブック(Book)という」)の中 にはいくつかのワークシートが 含まれているので(デフォルトで は Sheet1, Sheet2, Sheet3 の 三つ), SYSTAT との間でデータ



図2.データウィンドウ

のやり取りをするためには,あらかじめ「どのファイルのどのワークシート」にデータが 入力されているかを確認しておく必要がある.

データを SYSTAT に取り込む具体的な手順は以下のとおりである.

- SYSTAT を起動して「ファイル」メニューから「開く」を選択するとサブメニューが 表示されるので「データ」を選択する(図3).
- (2) ファイル選択画面が表示されるので,画面下の「ファイルの種類」で「Microsoft Excel(*.xls)」を選択してから,画面上の「ファイルの場所」で SYSTAT に取り込むデータが保存されている場所を指定し,目的のファイルを指定して「開く」ボタンをクリックする(図4).
- (3) Excel のどのシ

ートかを選択 する画が表 示されるので, 解析データの 入っているワ ークシート名 を選択して 「OK」ボタン をクリックす る(図5).

新規作成N		NO BRAN
間(©) ■ 保存(S) Chi+S	データ(1) コマンド(2) B Z 型 単語 副 信	0 () en
 26前を行けて(14存(b) すべてを(条存(l) データを(条存(l)) データを(条存(l)) Chi+D データベースの読み込み(b) 	QP+7+7-10	
 かりッフボードの実行(2) ・		
ページ設定(U) (2)、印刷ブルビュー(V) 〇 印刷(V)		
最近使ったデータ(B) 最近使ったコマンド(I) 最近使った出力(E)		
終了⊗		

図3.外部ファイルからデータを取り込む

RK .					?
ファイルの場所の	🗀 医学科統計学のレポート課題-2003 💌	e (<u>ت</u>		
Sitest2					
B test3 R テキストのデータ					
a change e					
ファイル名他	test1		- [獣の	
ファイルの種類(①)	Microsoft Excel (#xis)		•	キャンセル	
				わらい。	

ワークシートの選	択	
シート番号:名前	1: Sheet1 2: Sheet1 3: Sheet3	OK キャンセル

図5.ワークシートを指定

図4. Excel ファイルを指定

3.関連2群データの差のノンパラメトリック検定

これは「お互いに関連のある」2群データの間に有意差があるかどうかを「ノンパラメトリック」な方法により検定するものである.通常「Wilcoxon 符合付順位和検定」が使われる.この検定は、「パラメトリック」な方法による「t 検定」に相当する.Microsoft Excelの「分析ツール」には次の三つの「t 検定」

一対の標本による平均の t 検定

等分散を仮定した二標本による t 検定

分散が等しくないと仮定した二標本による t 検定

が提供されているが、「Wilcoxon 符合付順位和検定」は上の「一対の標本による平均の t 検 定」の「ノンパラメトリック」版に対応するものである.

以下に,具体的なデータ例と「Wilcoxon 符合付順位和検定」の手順を示す.

例題1.ある病院の眼科患者8名の眼部水晶体の厚さ(mm)を二つの方法,A法とB
法で測定した(表1).差が認められるか,両側検定を行え.
(「新版 医学への統計学」,古川俊之監修・丹後俊郎著,朝倉書店,p.88,例題5.8)

- SYSTAT を起動し,表1のデータを2群データとして第1列と第2列に入力する.あるいは,もしデータが Microsoft Excel のワークシート上に入力されている場合は,すでに述べた第2節の方法によって SYSTAT に取り込む(図6).
- (2) 「統計解析」メニューを開き、「ノンパラメトリック検定」を選択し、さらに「Wilcoxon」を選択する(図7).
- (3) 解析データをセットするためのウィンドウ(図8)が表示されるので,左側のボックスに表示されている二つのデータ(ここではAとB)をそれぞれをクリックして「追加」ボタンを押すと,右側の「変数」ボックス内に追加される(図9).終わったら「OK」ボタンをクリックする.

方法A	方法B
3.80	3.91
4.00	3.71
3.20	4.31
3.80	4.51
5.00	5.10
4.10	4.90
2.90	4.31
3.40	3.91

表1. 例題1のデータ

- (4) 結果が表示され,両側検定の p-value として「0.036」(最後の表の A 列 B 行の数値を 読む)が得られる(図10).
- (5) したがって,方法 A と方法 B は有意水準 0.05 で「有意差あり」という結論が得られる.

「「「「」」、「「」」、「「」」、「「」」、「「」」、「」、「」、「」」、「「」」、「」、「	ー SYSTAT データ	2		
ファイル(E) 編集(E) 表	€示(⊻) データ([<u>)</u>) グラフ(<u>G</u>) 希	苑計解析(<u>S</u>) へ,
₿ ¤	je 🖪 🎒 🎒	🗟 🕹 🗈	C 🖂 🖷	😫 🛛 🖬 🗠
∭×= Ę	5 67 17 17 1	- 🔢 🧮 123	• ∰ 📰 ≡	= 📰 <u>.</u> 🕅
Row: 1,	Variable: A		3.8	
	Α	В	(VAR00003)	(VAR00004)
1	3.800	3.910		
2	4.000	3.710		¢
3	3.200	4.310		•
4	3.800	4.510		
5	5.000	5.100		
6	4.100	4.900		
7	2.900	4.310		
8	3.400	3.910		
9				
10				

図6.解析データの入力



図7.ノンパラメトリック検定の「Wilcoxon」を選択

Wilcoxon		×
A B (追加 (肖耶	変数(複数可):	OK キャンセル ヘルプ

図8.解析データをセットするためのウィンドウ



図9.解析するデータを右側の変数ボックスに追加

→ 無題 - SYSTAT アウトプッ コライル(5) 得集(5) ませ	トオーガナイザ 「□□×」
	▼ 9 ▼ B Z U ≡ ≡ ≡ ∥
□- ← SYSTAT Output □- 伊 読み込み G¥pe ● Wilcoxon -	Wilcoxon Signed Ranks Test Results Counts of differences (row variable greater than column) IA IB A 0 B 7 Z = Sum of signed ranks)/square root(sum of squared ranks) IA IB
• • • • • • • • • • • • • • • • • • •	B 2100 0.000 Two-sided probabilities using normal approximation A B A 1.000 B B 0.036 1.000

図10.解析結果

4.独立2群データの差のノンパラメトリック検定

これはお互いに関連のない「独立な」2群データの間に有意差があるかどうかを「ノン パラメトリック」な方法により検定するものである.通常「Wilcoxon 順位和検定」が使わ れる(「Mann-Whitney の U 検定」もこの方法と等価な検定法である).この検定は、「パ ラメトリック」な方法による「t 検定」に相当する.Microsoft Excel の「分析ツール」に は次の三つの「t 検定」

一対の標本による平均の t 検定

等分散を仮定した二標本による t 検定

分散が等しくないと仮定した二標本による t 検定

が提供されているが、「Wilcoxon 順位和検定」は上の「等分散を仮定した二標本による t 検 定」または「分散が等しくないと仮定した二標本による t 検定」の「ノンパラメトリック」 版に対応するものである.

以下に,具体的なデータ例と「Wilcoxon 順位和検定」の手順を示す.

例題2.透析患者の免疫グロブリンの一つ IgG値(mg/100ml)が健常者に比べて高い かどうかを調べるために,40歳代男性の透析患者9名,同年代の病院職員の健常者7 名の IgGを測定し,表2のデータが得られた.これから,透析患者の IgGの方が健常 者に比べて高いと認められるか.

(「新版 医学への統計学」,古川俊之監修・丹後俊郎著,朝倉書店,p.80,例題 5.5)

- SYSTATを起動し,表2のデータを図11の形式に変換して 入力する.あるいは,もしデータがMicrosoft Excelのワーク シート上に入力されている場合は,すでに述べた第2節の方 法によってSYSTATに取り込む(図11).
- (2) 「統計解析」メニューを開き、「ノンパラメトリック検定」
 を選択し、「Kruskal-Wallis」を選択する(図12).

透析患者	健常者
1326	1220
1418	1080
1820	980
1516	1420
1635	1170
1720	1290
1580	1116
1452	
1600	
表2.例題	[2のデータ

 (3) 解析データをセットするためのウィンドウが表示されるので, 左側ボックスに表示されている二つのデータ(ここでは VAR00001 と VAR00002\$)をそれぞれをクリックし, 右側

の「変数」ボックスと「群分け変数」ボックスに追加する(図13).終わったら「OK」 ボタンをクリックする.

- (4) 結果が表示され, p-value として「0.002」が得られる(図14).
- (5) したがって, 有意水準 0.01 で, 透析患者の IgG の方が健常者に比べて高いという結果 が得られる.

ここで注意すべき点は, Kruskal-Wallis は本来ノンパラメトリック分散分析の手法であるが,特別な場合として「独立2群データの差の検定」を含むので, SYSTAT では対応の

ない独立2群データの「Wilcoxon 順位和検定」を「Kruskal-Wallis」で代用するように設計されていることである「Wilcoxon 順位和検定」は頻繁に用いられる検定手法であるので,できれば「Wilcoxon 順位和検定」として独立した解析メニューを提供して欲しいところで

あるが,分散分析において水準数が2の場合の「Kruskal-Wallis」による検定は, 「Wilcoxon順位和検定」による「独立2群 データの差」の検定と数学的には完全に等 価であるため,SYSTATでは統計解析メニ ューの中に「Wilcoxon順位和検定」に相当 する項目は存在しない.

また,もう一つの注意点は,用意すべき データは表2のように2列にそれぞれの群 のデータを入力するのではなく,図11の ように数値データはすべて第1列にまとめ て入力し,第2列には「透析患者」や「健 常者」のような「群分け」データを入力し なければならないことである.

1 またい ひんしん ひんしん ひんしん しんしん しんしん しんしん しんしん しんし	- SYSTAT デー	\$	
ファイル(E) 編集(E) 录	表示(⊻) データ([D) グラフ(<u>G</u>) 紡
0	ê 🖬 🕼 🗧) 🖪 X 🖻	6 🗩 🖷
x=	b 😰 💾 😤	3= 🕅 🛄 123	• 📭 📗 💻
Bow 1	Variable: 0/AR000	103.)	
1		/	
	VAR00001	VAR00002\$	(VAR00003)
1	1326.000	透析患者	
2	1418.000	透析患者	
3	1820.000	透析患者	
4	1516.000	透析患者	
5	1635.000	透析患者	
6	1720.000	透析患者	
7	1580.000	透析患者	
8	1452.000	透析患者	
9	1600.000	透析患者	
10	1220.000	健常者	
11	1080.000	健常者	
12	980.000	健常者	
13	1420.000	健常者	
14	1170.000	健常者	
15	1290.000	健常者	
16	1116.000	健常者	
17			

図11.解析データの入力



図12. ノンパラメトリック検定の「Kruskal-Wallis」を選択





🔠 無題 - SYSTAT アウトプット	ヽオーガナイザ×
_ファイル(E) 編集(E) 表示(⊻)データ(D) グラフ(G) 統計解析(S) ヘルプ(H)
D 🛎 🖬 🕼 🖨 D.	. X 🗈 🛍 🔽 🖳 🎝 📗 🛚 🖉 🎍 🖬 🖾 🛳 🗠 🚳 🗍 🥵 🐼
MS Pゴシック	<u>▼</u> ⁹ ▼ B <i>I</i> <u>U</u> ≡ ≡ ≡ ⊡ ③ ③ № № № 3D
⊡ C SYSTAT Output	Categorical values encountered during processing are: VAR00002\$ © levels) 健常者, 透析患者 Kruskal-Wallis One-Way Analysis of Variance for 16 cases
	Group Count Rank Sum
F	健常者 7 30.000 透析患者 9 106.000 Mann-Whitney U test statistic = 2.000 Probability is 0.002 Chi-square approximation = 9.751 with 1 df

図14.解析結果

5.おわりに

本稿では、医学サブセンターの新システムに導入された統計解析ソフトウェア「STSTAT」 について、Microsoft Excel では解析できないノンパラメトリック検定の例として「Wilcoxon 符合付順位和検定」と「Wilcoxon 順位和検定」(独立2群の場合として「Kruskal-Wallis」 で代用)の二つの例を紹介した.

SYSTAT は,この他にも生存分析や回帰分析などさまざまな統計解を行うことができ, また豊富なグラフィックス機能も持っているので,オンラインマニュアルを参考にして各 自で活用していただきたい.