

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月にリプレイ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ソフトウェアは、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、統計解析ソ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月にリプレイ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。

- (1) 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。

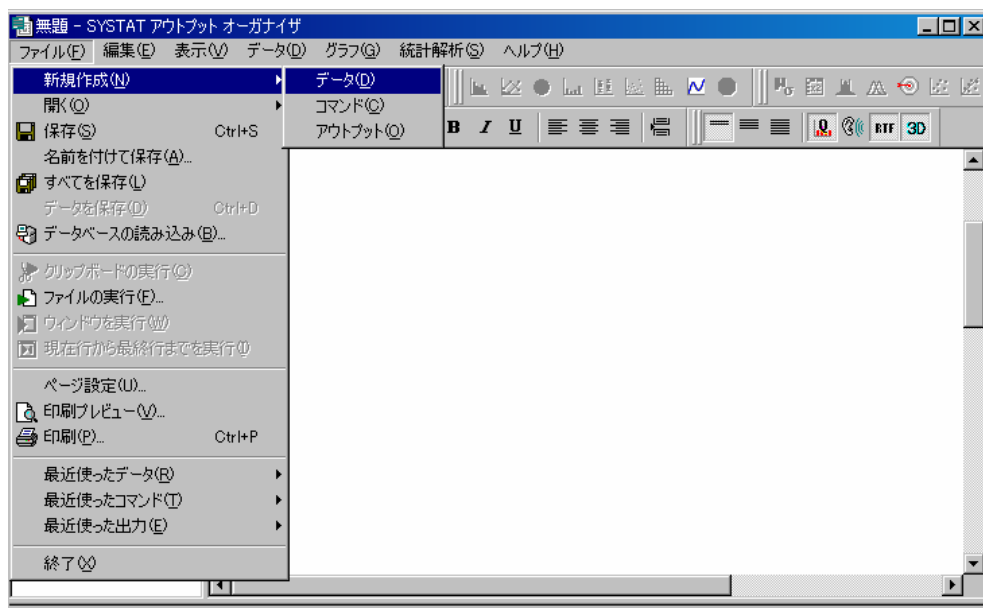


図 1. データの新規作成

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

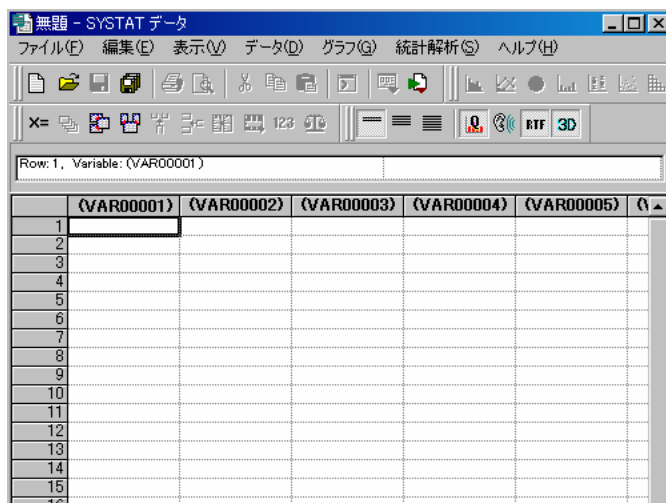


図 2 . データウィンドウ

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。

- (1) 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面が表示されるので、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）。

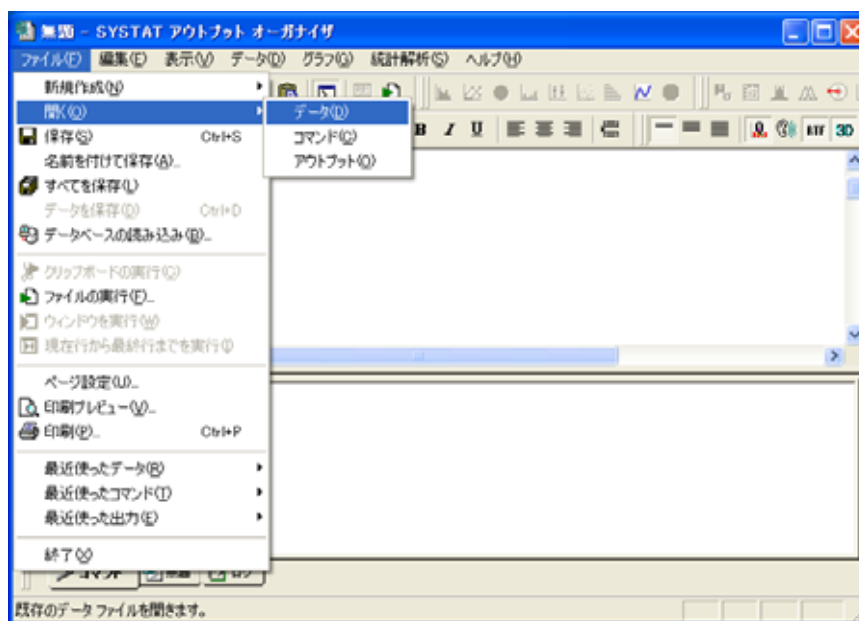


図 3 . 外部ファイルからデータを取り込む

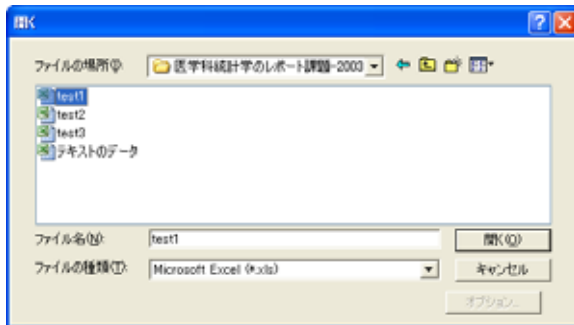


図4 . Excel ファイルを指定

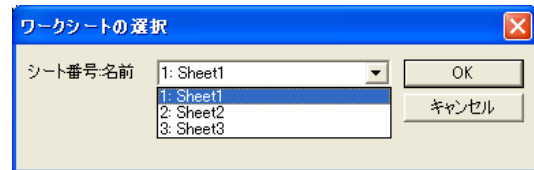


図5 . ワークシートを指定

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 群データの間に関連があるかどうかを「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

分散が等しくないを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)

(1) 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。

	A	B	(VAR00003)	(VAR00004)
1	3800	3910		
2	4000	3710		
3	3200	4310		
4	3800	4510		
5	5000	5100		
6	4100	4900		
7	2900	4310		
8	3400	3910		
9				
10				

図 6 . 解析データの入力

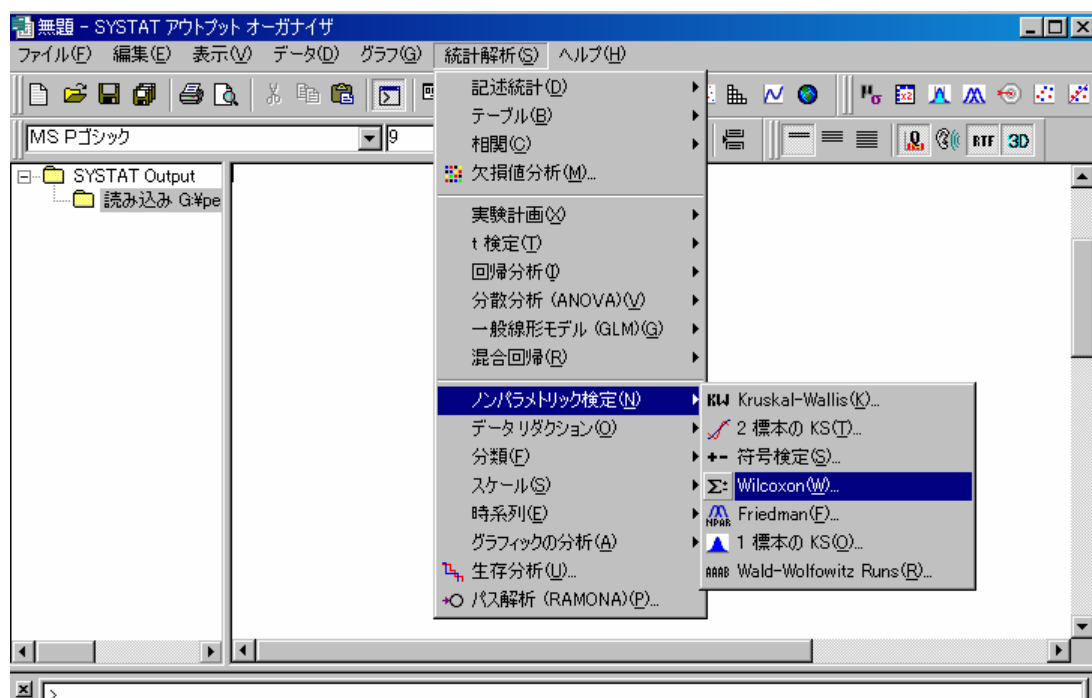


図 7 . ノンパラメトリック検定の「Wilcoxon」を選択



図 8 . 解析データをセットするためのウィンドウ

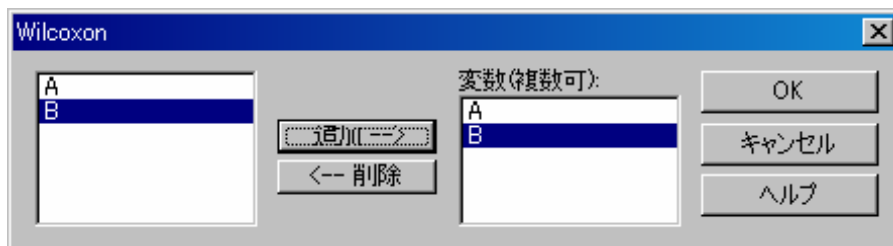


図 9 . 解析するデータを右側の変数ボックスに追加

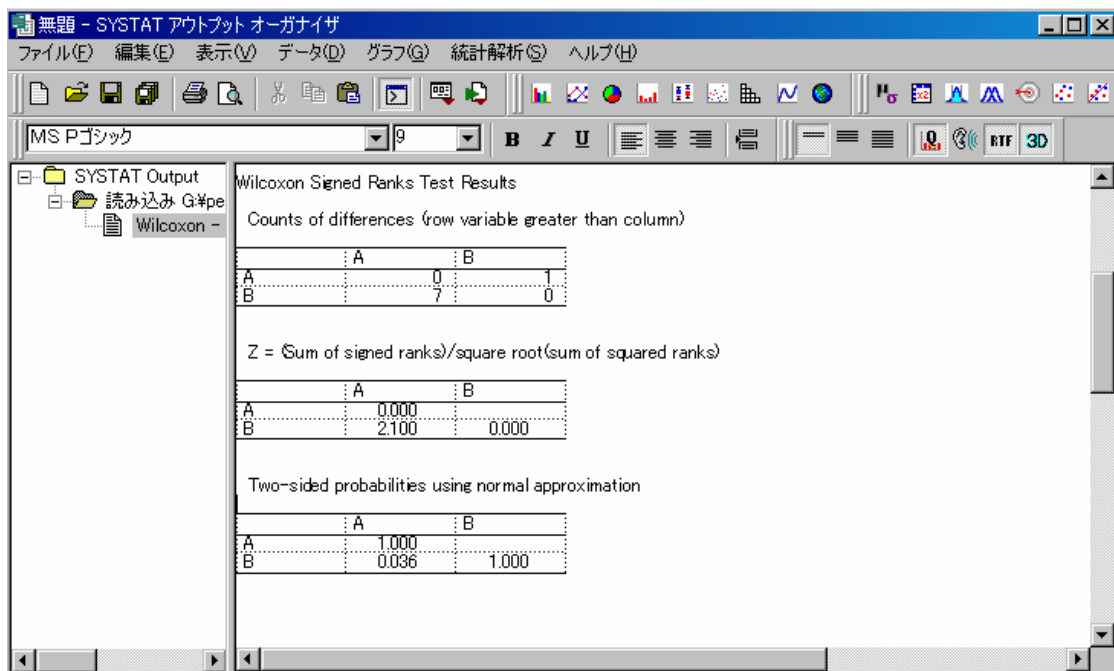


図 10 . 解析結果

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群データの間に関連があるかどうかを「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)

(1) SYSTATを起動し、表 2 のデータを図 1 1 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 1 1)。

(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 1 2)。

(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 1 3)。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 1 4)。

(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。

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

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。

	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 1 1 . 解析データの入力

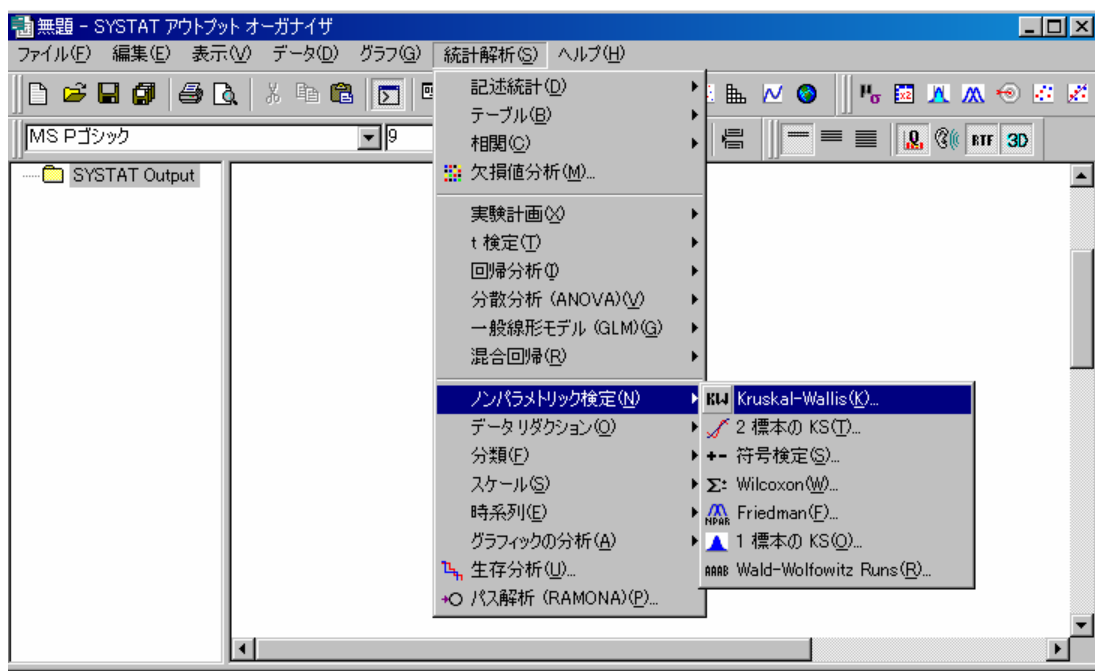


図 1 2 . ノンパラメトリック検定の「Kruskal-Wallis」を選択

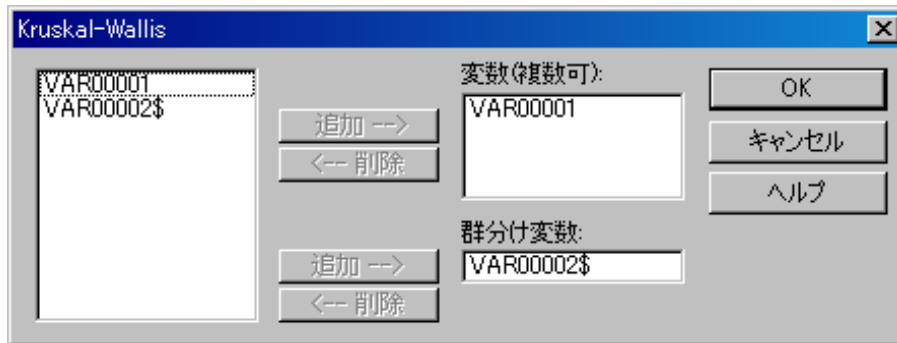


図 1 3 . データを変数ボックスと群分け変数ボックスに追加

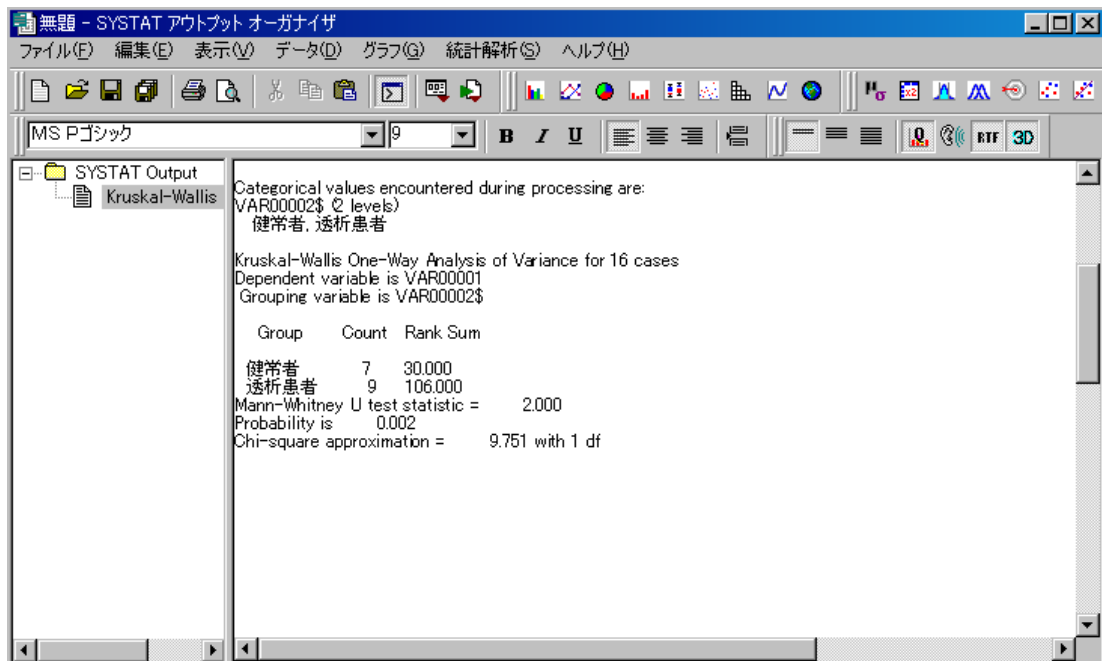


図 1 4 . 解析結果

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「SYSTAT」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 符号付順位和検定」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。