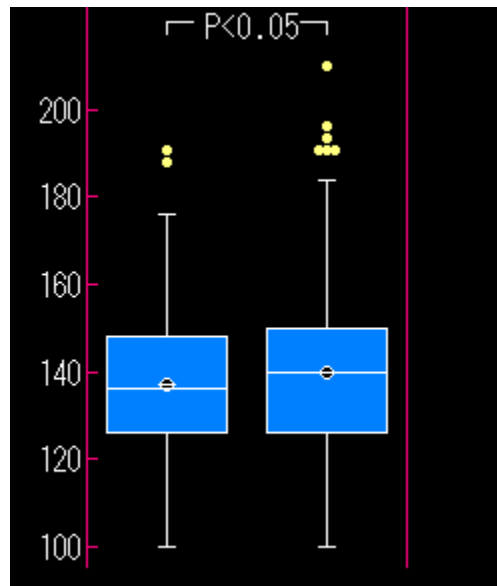


The Statistical Package for the Biosciences (SPBS) User's Guide

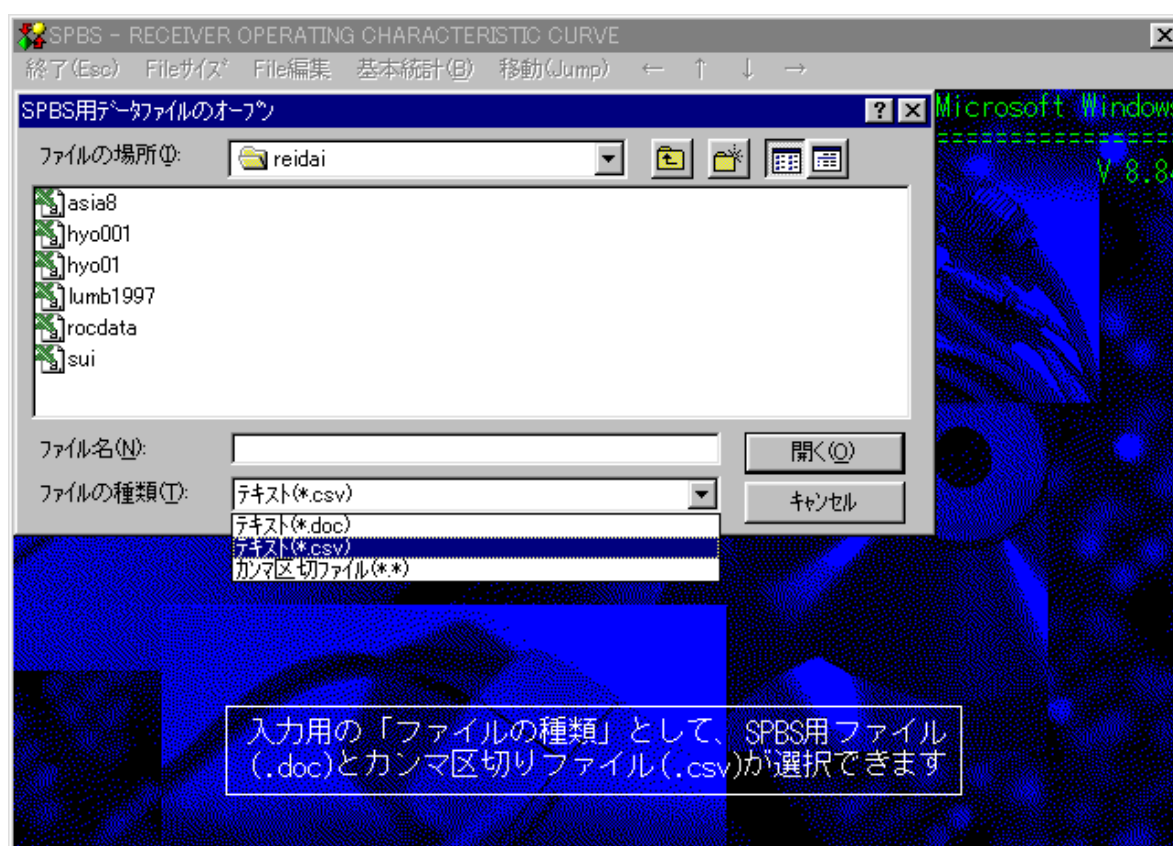


**Winesteen Institute of Community Medicine
Toda, Saitama, Japan**

The Statistical Package for the Biosciences

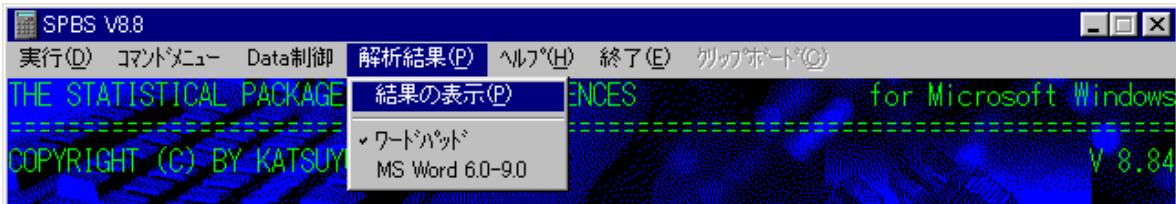
The Statistical Package for the Biosciences (以下、**SPBS 統計パッケージ**と略)は生物科学領域のデータを、探索的に解析し仮説を設定する、説明的に利用し生物学的現象の解明を推進する、検証的に扱い、論文の結果に統計的根拠を提示する、の3つの目的に沿って、1986年にMS-DOS版パーソナルコンピュータで可動する初心者向け対話型統計ソフトとして誕生した。

SPBS 統計パッケージは、当初のMS-DOS版SPBSをMicrosoftのWindows 95/98/Me/NT4/2000システム上で動くように変更し、かつ最新の統計学の流れに沿うように改良を加えたものである。以前のMS-DOS版(16ビットモード)ではひとつのデータファイルの配列サイズが64キロバイト(KB)に制限されていたために、変数とサンプル数の積が最大16,000個(単精度データ)であった。Windows版(32ビットモード)ではひとつのデータファイルの配列サイズ制限枠が2ギガバイト(GB)であるので、Windows版SPBS統計パッケージでは最大変数で999個、最大サンプルで32,000個を扱うことが可能となった。また「データファイルの作成と編集」および「メニュープログラム」の「コマンドメニュー」を除く全てのプログラムで、表計算ソフト等で作成される「カンマ区切りファイル」を直接入力して解析することができる。その上、表計算ソフトの数値データをクリップボードにコピーすると、その数値データをSPBS用データファイルに変換することも、また自動的に取り込んで解析することも可能である。

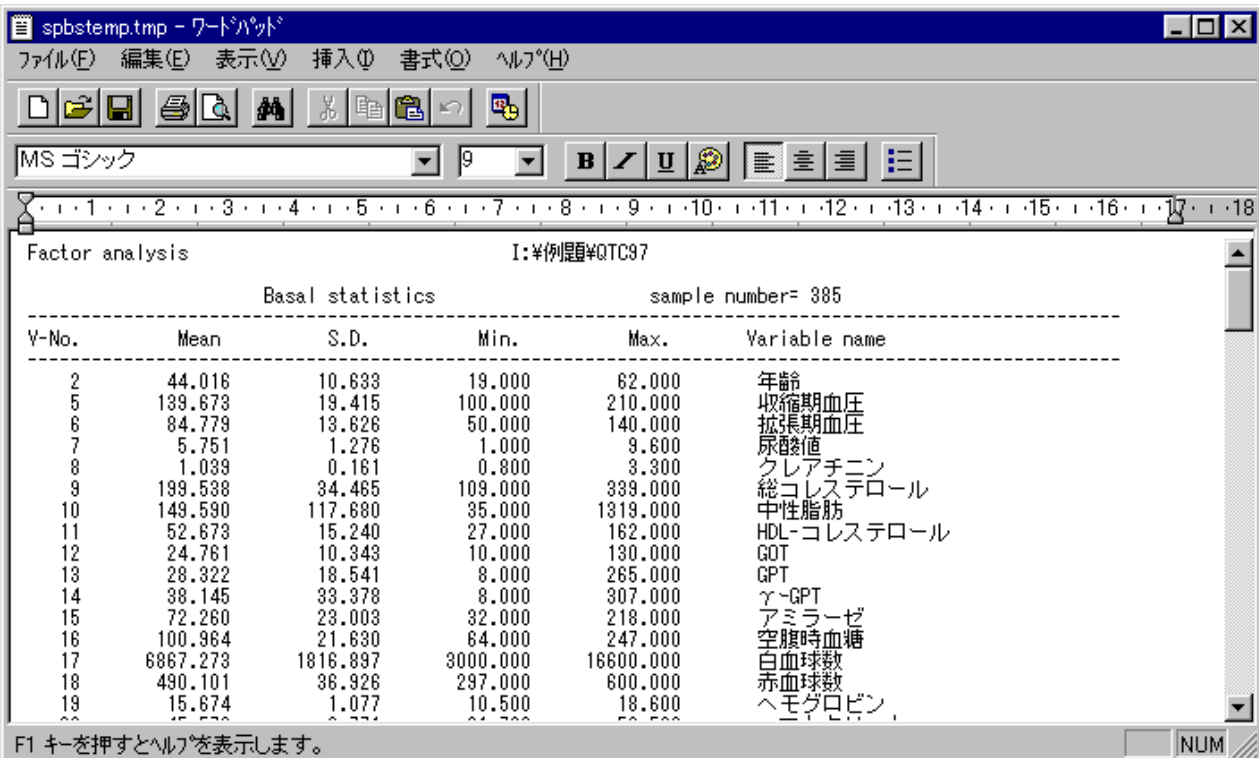


SPBS 統計パッケージでは、数値入力(およびその直後のエンターキー)と変数名入力を除き、全てマウスで操作する(勿論、キーボードにより全ての操作をおこなうことも可能)。画面で指定がない限り、<終了>ボタンは「Esc」キー、<続行>ボタンは「Esc」以外の任意のキー、<実行>ボタンは「Enter」キーと対応しており、また<Yes>および<No>ボタンは「Y」または「N」キーと対応している。さらにプログラムの説明として、SPBS 統計パッケージ「解析方法の選択」の「HELP」ボタンをクリックすると「トピックの検索: SPBS - HELP」が表示される。

注意点として、SPBS 統計パッケージを使用するときは必ずコントロールパネルの「画面」の「ディスプレイの詳細」のフォントサイズが「小さいフォント」に選択されていなければならない(画面サイズは、640×480、800×600、1024×768のいずれを使用することもできるが、カラーパレットは16ビット以上が好ましい)。また、Microsoft社製インターネットエクスプローラ Ver4.0以上がインストールされていないコンピュータでは作動しないことがある。この種のコンピュータでは、インターネットエクスプローラ Ver4.0以上を一度インストールし、不要ならその後アンインストールする必要がある(アンインストールしても%windows%system フォルダ内にSHLWAPI.DLLファイルは残り、SPBS 統計パッケージは作動する)。



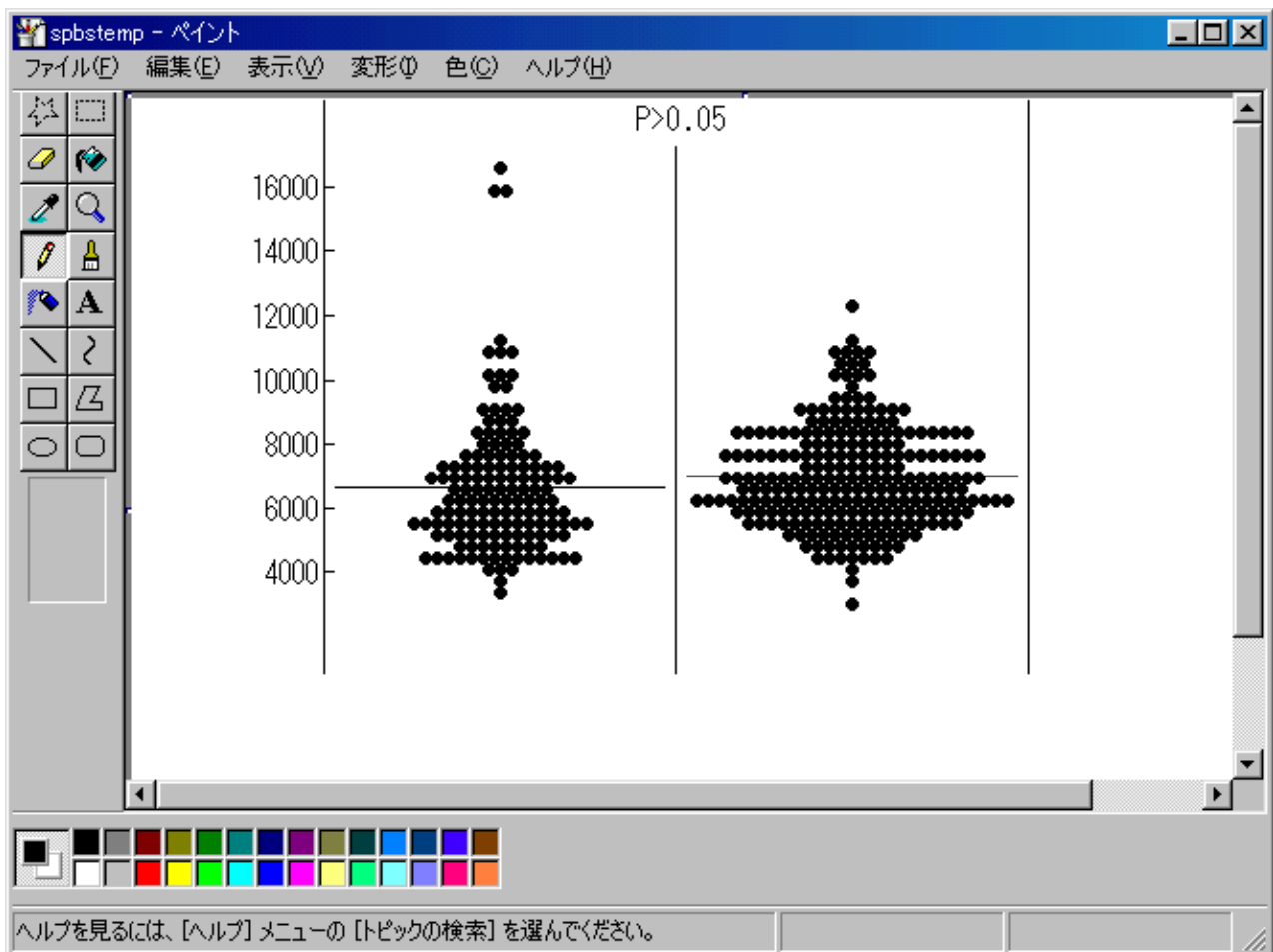
メインメニューの上に「結果の表示(P)」ボタンがある。解析プログラムが終了した後、このボタンをクリックすると解析結果が Microsoft 社製『ワードパッド』上に表示される。解析結果のプリンターによる印刷、ディスクへの保存、他のソフトへの貼り付けは、『ワードパッド』上で実行できる (Windows 95/98/Me/NT4/2000 システムに内蔵されている『ワードパッド』がインストールされていないと SPBS 統計パッケージは作動しない)。また解析を続行する場合、『ワードパッド』を終了させた後でないと次の解析結果が表示されない。『ワードパッド』のプリンター印刷は、A4紙で左右の余白が 20mm、フォントが「MSゴシック」、文字サイズが 9 ポイントを想定して出力されるように設計している (Windows 98/Me/NT4/2000 を使用している人は『ワードパッド』の文字サイズが 10 ポイントに設定されているので、印刷前に上図の「編集」の「すべて選択」をクリックして文字サイズを 9 ポイントに変更してから「印刷」ボタンをクリックする)。なお、解析結果がない時にこのボタンを押しても『ワードパッド』は現れない。



SPBS 統計パッケージの解析プログラムのうち「有意差検定」、「一次回帰式」、「分散分析」、「相関分析」、「順位相関係数」、「主成分分析」等々ではグラフィック解析結果が自動的にディスプレイに表示される。これは、使用したデータの中に解析に相応しくない数値が含まれているかどうか (解析データの分布状況) をチェックする目的で加えられている。このグラフィック解析結果を、Microsoft 社製『Word』、JUSTSYSTEM 社製『一太郎』等のワープロソフトの文書中に貼り付けることができる。この場合、下のように「上の図をプリントしますか?」と画面に表示された時に「<Paint>」ボタンをクリックすると、Microsoft 社製の『ペイント』が自動的に起動し、グラフィック解析結果を表示する。『ペイント』上では、グラフの範囲指定を行った後に「編集」の「コピー」をクリックし (この操作でグラフィック解析結果はクリップボード上に移動する)、ワープロソフト上で「編集」の「形式を選択して貼り付け」《ビットマップ<DIB>」を選択すると貼り付けが終了する (SPBS 統計パッケージで解析を続行する場合には『ペイント』を必ず終了しなければならない)。



グラフィック解析結果をプリンターで直接印刷することも可能である。すなわち、「上の図をプリントしますか?」で「Yes」ボタンをマウスでクリックすると、プリント印刷される (こちらのグラフの方が、『ペイント』経由で印刷されるものより高品質・高精度の画質が得られる。これは、プリンター印刷時の線質まで考えて設計されているためである)。なお、この場合 Windows 95/98/Me/NT4/2000 システムにプリンターが登録されていないと (プリンターの設置場所は「LPT1:」に初期設定されているので、異なっている場合には「コマンドメニュー」の中に「プリンターの設定場所の変更」プログラムを利用して設定場所を変更する必要がある。詳細は「SPBS User's Guide」の『0. コマンドメニュー』の『 . プリンター場所の設定』に記述されている)。



SPBS 統計パッケージでは、データ内の数値「-99」および「-999」を欠損値として扱うように設計されている。この数値を含むデータファイルでは、（計算途中で除外されるので）解析の度にサンプル数が異なることがありうるので注意を要す。また、解析結果の有意性は一部で「 $P < 0.01$ 」、「 $P < 0.05$ 」、「n.s.」等々の形で表示されることがあるが、多くは確率 P 値または 95% 信頼区間値が算出される（確率 P 値が必要な場合には「各種検定・分析」の「確率 P 値の計算」を利用）。なお、内部の演算は倍精度で行っているが解析データは単精度（大体 6 桁までの数字）のみの扱いなので、解析結果の精度は大体 6 桁位と考えるべきである。

SPBS 統計パッケージのインストール後のプログラムファイルの消失については、ハードディスク内の「SPBS」フォルダーの「spbshck.exe」をマイコンピュータ上でダブルクリックすると、プログラムチェックを実行し、消失プログラムを見つけ出す。この場合、消失プログラムのみを CD-ROM 上から「SPBS」フォルダーにコピーすれば SPBS 統計パッケージは修復する。

SPBS 統計パッケージは、SAS や SPSS 等が大容量（かつ高価）のソフトであるのに対し、誰でも容易にインストールして利用できる（ハードディスクの占有容量を 20MB 未満に抑えた）ソフトを目指して作成されている。このため、「共分散構造分析」、「対応のある多重ロジスティック分析」、「多元配置分散分析」などの煩雑な解析、サンプル数が 32,000 を越えるデータ解析、専門家でないといと解らない統計量の算出等々はサポートしていない。したがって、これらが必要な人は前述の SAS や SPSS を購入する必要がある。

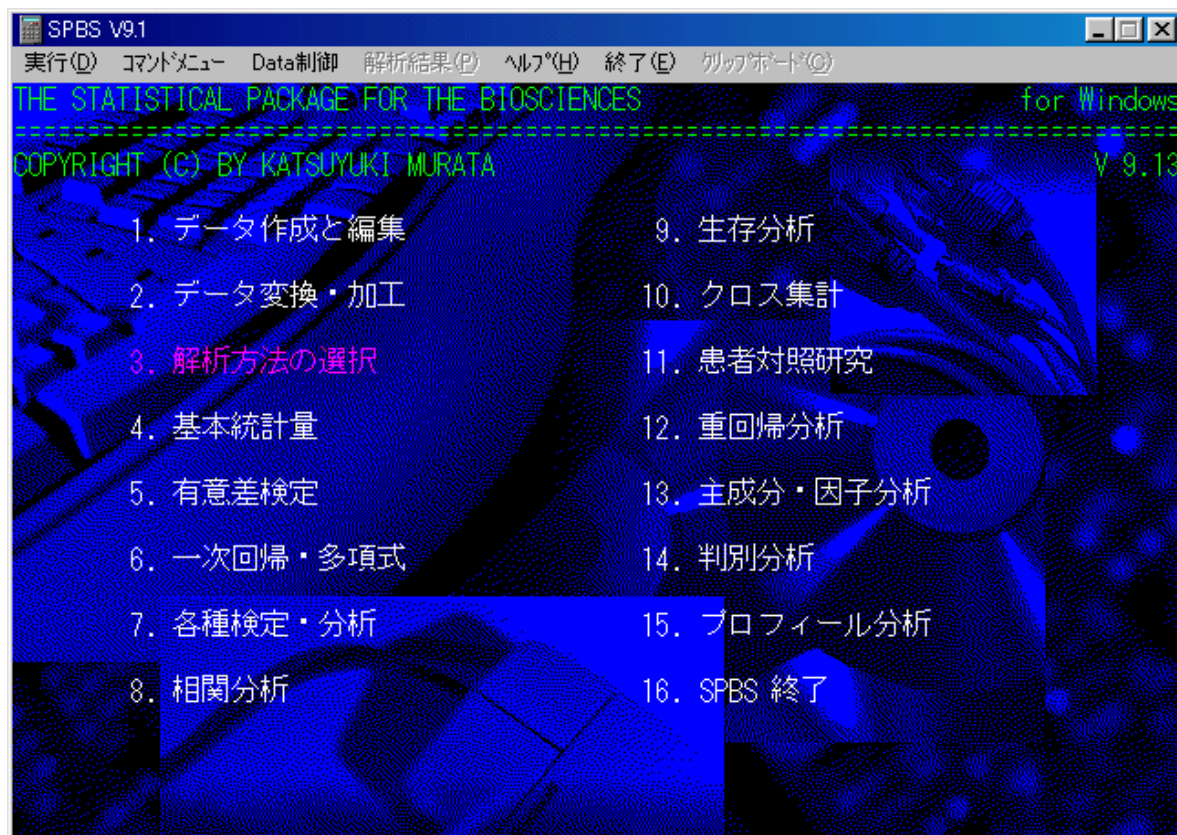
作成者はこのソフト使用により生じるいかなる不利益に対しても一切の責任を負わない。 利用者はこれを了承の上で使用することができる。なお、SPBS 統計パッケージは現在のところ Windows 版しかサポートしていない。これは Mac ユーザーへの当てつけではなく、著者の手元に適当な Mac 用の変換ソフト（プログラム開発ソフト）がないためである。したがって、Mac 上で使用したい人は Windows 98 エミュレータソフトを Mac にインストールする方法が考えられる。

July 27, 2001

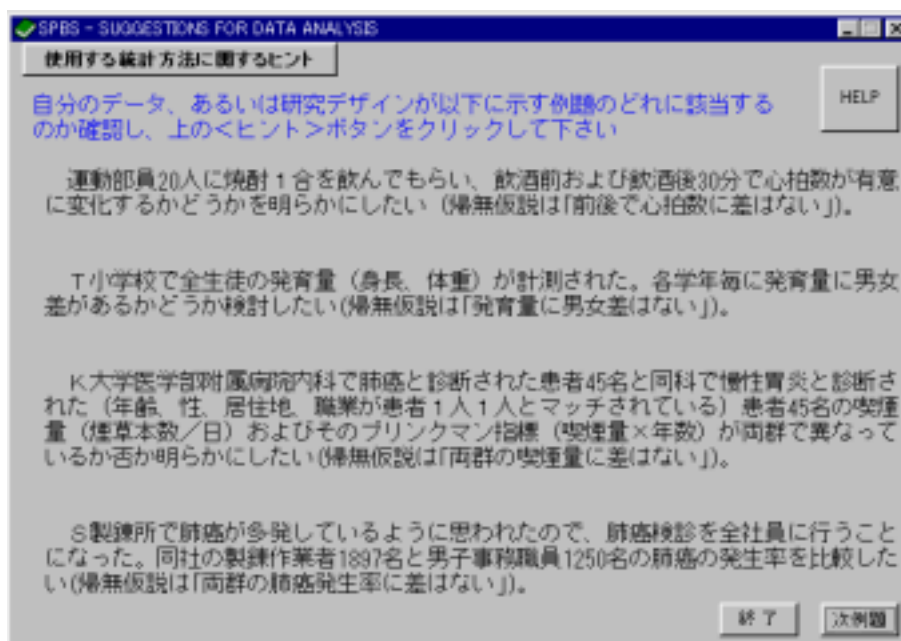
Winestem Institute of Community Medicine
Winestem Villagefield, MD
www.comworks.co.jp/spbs/

．SPBS統計パッケージのメインメニュー

下図はSPBS統計パッケージのメインメニュー画面である。



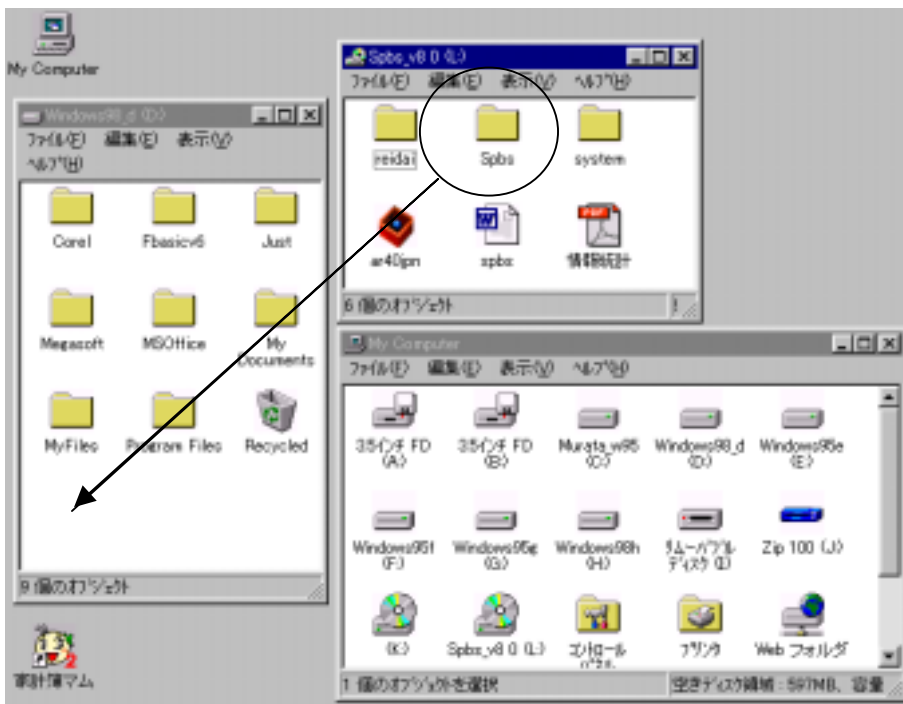
下の画面は、上の画面の「3．解析方法の選択」をダブルクリックした時に現れる。この画面の<使用する解析方法に関するヒント>ボタンをクリックすると、どのプログラムを使用すればよいのを表示し、さらにその解析プログラムに進むこともできる（<解析実行>ボタンが下図の<終了>ボタンの左側に現れるのでクリックする）。



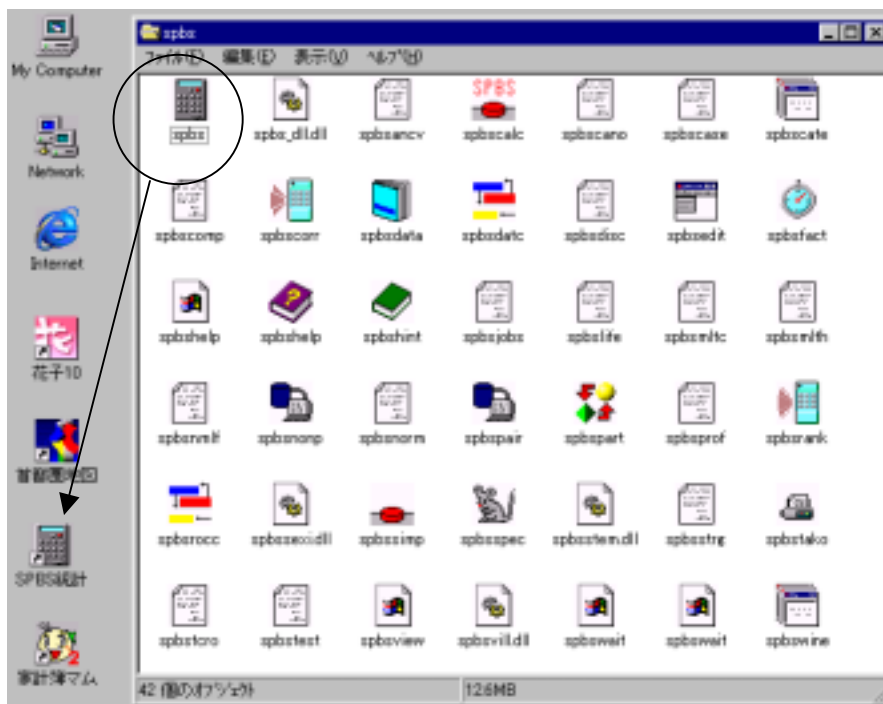
また、<HELP>ボタンをクリックすると、「トピックの検索：SPBS - HELP」に移動する。

. SPBS 統計パッケージのインストール方法

Windows 95/98/Me/NT4/2000 のエクスプローラかマイコンピュータを利用して、CD-ROM 内の「¥SPBS」というフォルダーをハードディスク内の任意の場所にまるごとコピーする。

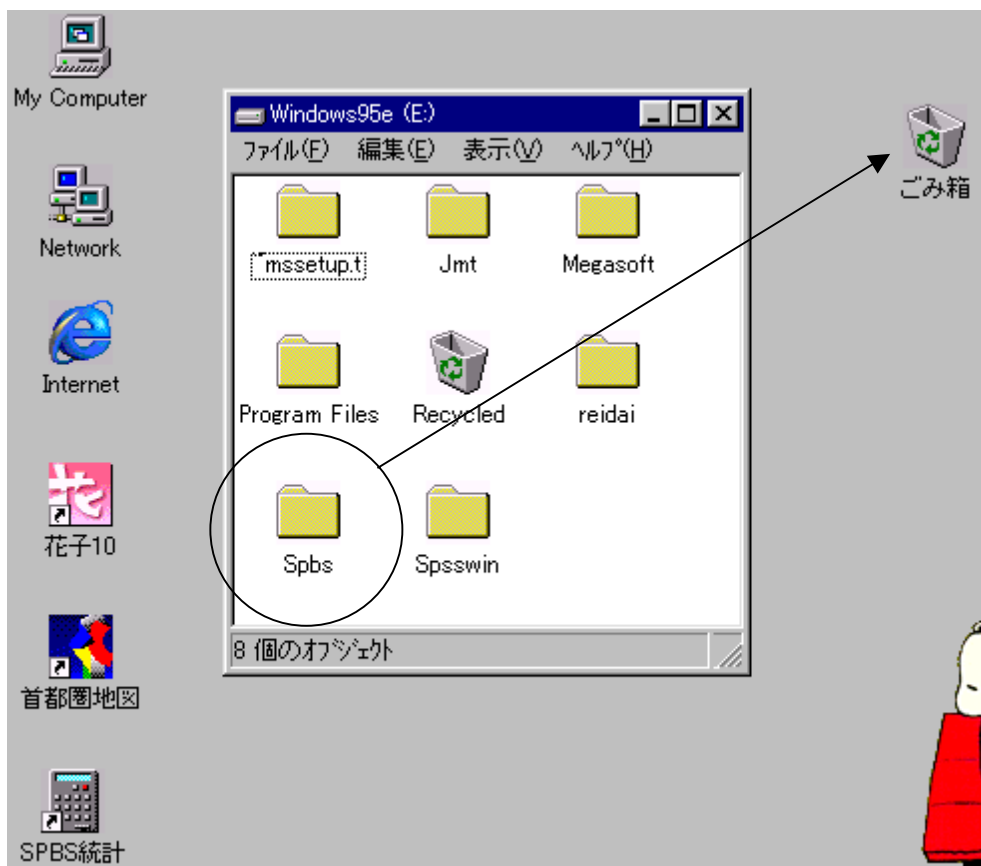


ハードディスク内にコピーされた「¥SPBS」フォルダーをマイコンピュータで開けて SPBS.EXE ファイル（これ以外のファイルではプログラムが上手く動作しない）上にマウスを置き、左クリックしたままデスクトップにドラッグする（外付けハードディスクの場合には右クリックして、「ショートカットを作成」のボタンをクリックし、出来上がったショートカットをデスクトップに移動する）。これでショートカットがデスクトップに作成される。以後、このショートカットをダブルクリックすると、プログラムが起動する。



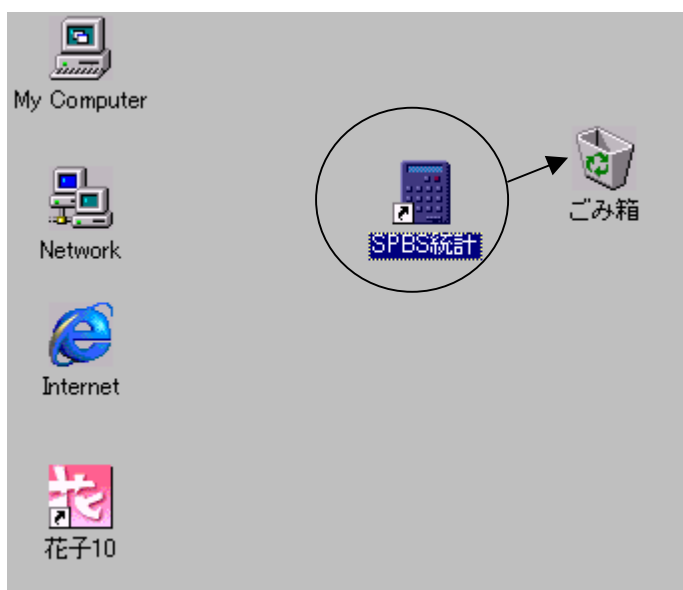
. SPBS 統計パッケージのアンインストール方法

ハードディスク内の「¥SPBS」フォルダーをゴミ箱にドラッグする。



次に、SPBS.EXE のショートカットをゴミ箱にドラッグする。

Windows 95/98/Me/NT4/2000 システムへの書き込みは一切ないので、以上の操作で全ての SPBS 統計パッケージのプログラムはハードディスクから抹消される。



. The Statistical Package for the Biosciences (SPBS)のプログラム内容

. コマンドメニュー

1. SPBS 用データファイルの確認
2. SPBS 用データファイルの複製
3. SPBS 用データファイルの消去
4. プリンター場所の設定 (LPT1 以外の場合)
5. データ制御プログラムの起動

. データの作成・編集 (データの変換・加工)

1. データファイルの作成と修正
2. データファイル Excel (CSV) SPBS file 変換
3. 固定長 file SPBS file の相互変換
4. データファイルの変数名の確認・変更
5. データファイルの縦横変換
6. データファイルの比較検討
7. データファイルの編集 (データの変換・加工)
8. 連続量データのカテゴリデータへの変換

. 基本統計量

1. 正規性の検討
2. 日本地図の作成

. 有意差検定

1. 2 群間の有意差検定
2. 一元配置分散分析 (3 群以上の対応のない有意差検定)
3. 二元配置分散分析 (3 群以上の対応のある有意差検定)
4. 2 群間の多重比較
5. 反復測定データの有意差検定 (多重比較)

. 一次回帰・多項式

1. 一次回帰式
2. 多項式

. 各種検定・分析

1. 比率の検定
2. 共分散分析
3. 内的整合性の検討
4. 確率 P 値の計算
5. ROC 曲線

. 相関分析

1. 相関分析 (2 相関係数の比較を含む)
2. 順位相関係数
3. 偏相関係数
4. 正準相関分析

. 生存分析

1. Kaplan-Meier 法による生存曲線
2. Cox の比例ハザードモデル
3. Weibull 生存分析

. クロス集計

1. クロス集計 (第3段階までの層別化)
2. 繰り返しのある二元配置分散分析

. 患者対照研究

1. χ^2 検定と Odds 比
2. 相乗作用指数
3. Mantel-Haenszel の χ^2 検定

. 重回帰分析

. 主成分・因子分析

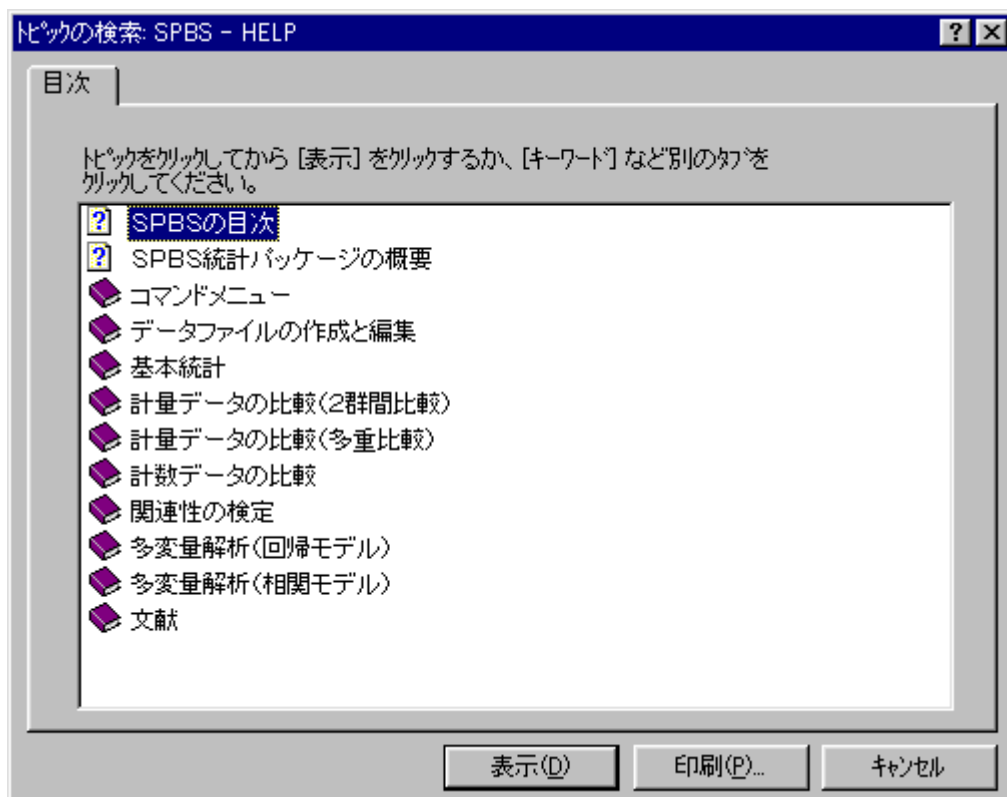
1. 主成分分析
2. 因子分析
3. 数量化理論 類

. 判別分析

1. 判別分析
2. 多重ロジスティック分析

. Profile 分析

. 参考文献



この「トピックの検索: SPBS - HELP」は『3. 解析方法の選択』の中の HELP ボタンあるいはメインメニューの上にある <ヘルプ> をクリックすると現れる。

プログラムの解説

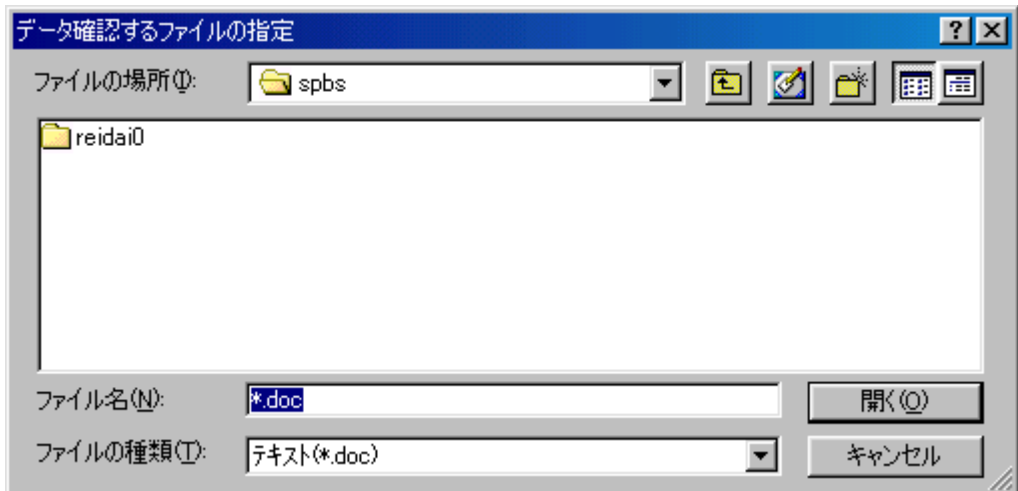


0. コマンドメニュー

(1) SPBS用データファイルの確認



SPBS 統計パッケージではファイル形式として、「XXX.doc」と「XXX.txt」を用いている(“doc”および“txt”を拡張子と呼ぶ)。前者は変数名ファイルとして、後者は解析データファイルとして使用している。「SPBS 用ファイルのオープン」時には前者のみが表示され、SPBS 統計パッケージを実行している間に“.txt”を見る



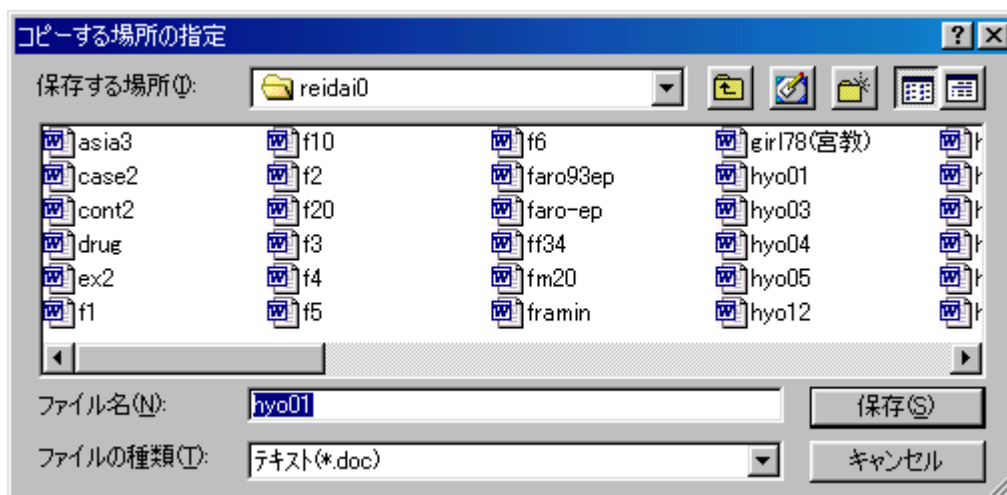
ことは殆どないが、この2つのファイルがないと実際の解析はできない。このプログラムでは、これら両者のファイルがあるかどうかと(エラーが発生する時は、後者の“txt”ファイルが存在しないか、または“doc”ファイルに書かれている変数およびサンプルの数と解析データファイルの数が一致しないことを意味する)、自分が使おうとし

変数	1	2	3	4	5	6	7	8	9	10	11
1	1	45	148	100	250	189	19	33	111	29.8	0
2	2	47	152	96	228	141	76	44	111	29.8	0
3	3	38	128	80	168	183	18	33	100	27.8	0
4	4	30	118	56	207	86	20	11	90	28.2	0
5	5	35	134	88	171	129	34	20	100	29.1	0
6	6	31	136	82	210	108	20	33	185	22.1	0
7	7	57	152	98	285	93	13	26	98	24.1	0
8	8	31	132	80	201	103	29	22	95	25.7	1
9	9	58	138	82	182	132	37	41	106	24.0	0
10	10	26	114	74	165	77	15	18	83	21.1	1
11	11	26	120	68	163	80	22	15	34	20.4	1
12	12	58	154	98	186	220	19	65	98	22.3	0
13	13	53	152	100	197	94	15	76	104	24.6	1
14	14	36	124	70	223	137	45	53	80	26.3	1
15	15	33	116	68	226	105	21	21	96	28.8	0
--- 以下省略 ---											
有効サンプル	60	60	60	60	60	60	60	60	60	60	60
欠損値数	0	0	0	0	0	0	0	0	0	0	0
平均値	30	43	137	86	198	138	30	40	99	23.7	0
最小値	1	23	102	56	133	45	12	8	67	19.8	0
最大値	60	58	172	112	260	538	78	282	143	28.2	1
変数名	1	ID番号							9	FBS	
	2	年齢							10	EMI	
	3	収縮期血圧							11	動物形態	
	4	拡張期血圧									
	5	総コレステロール									
	6	中性脂肪									
	7	LDL									
	8	γ-GTP									

ているデータファイルが どのような変数名を入れたファイルなのか、 実際の数値としてどのような値が入っているのか、および 有効サンプル数、欠損値データ数、平均値、最小値、最大値のチェックをおこなう。なお、サンプル数が15個を越える場合にはそれ以後のデータは省略され画面で見ることができないが、<印刷>ボタンをクリックすると全データがMicrosoft 社製「ワードパッド」上に打ち出される。

(2) SPBS用データファイルの複製

このプログラムは、あるフォルダーにある SPBS 用データファイルを別のフォルダーにコピーするとき使用する。この場合、フォルダーのあるドライブ名が同一のドライブであろうが異なっていようが問題としない。前述のように、SPBS 用ファイルは2種類（“txt”および“doc”）で1つのペアをなしているが、このプログラムでは確実に SPBS 用データファイルとして利用できる形でコピーされる（マイコンピュータとかエクスプローラを使用してコピーすることができるが、ペアであることを忘れると解析できなくなる）。



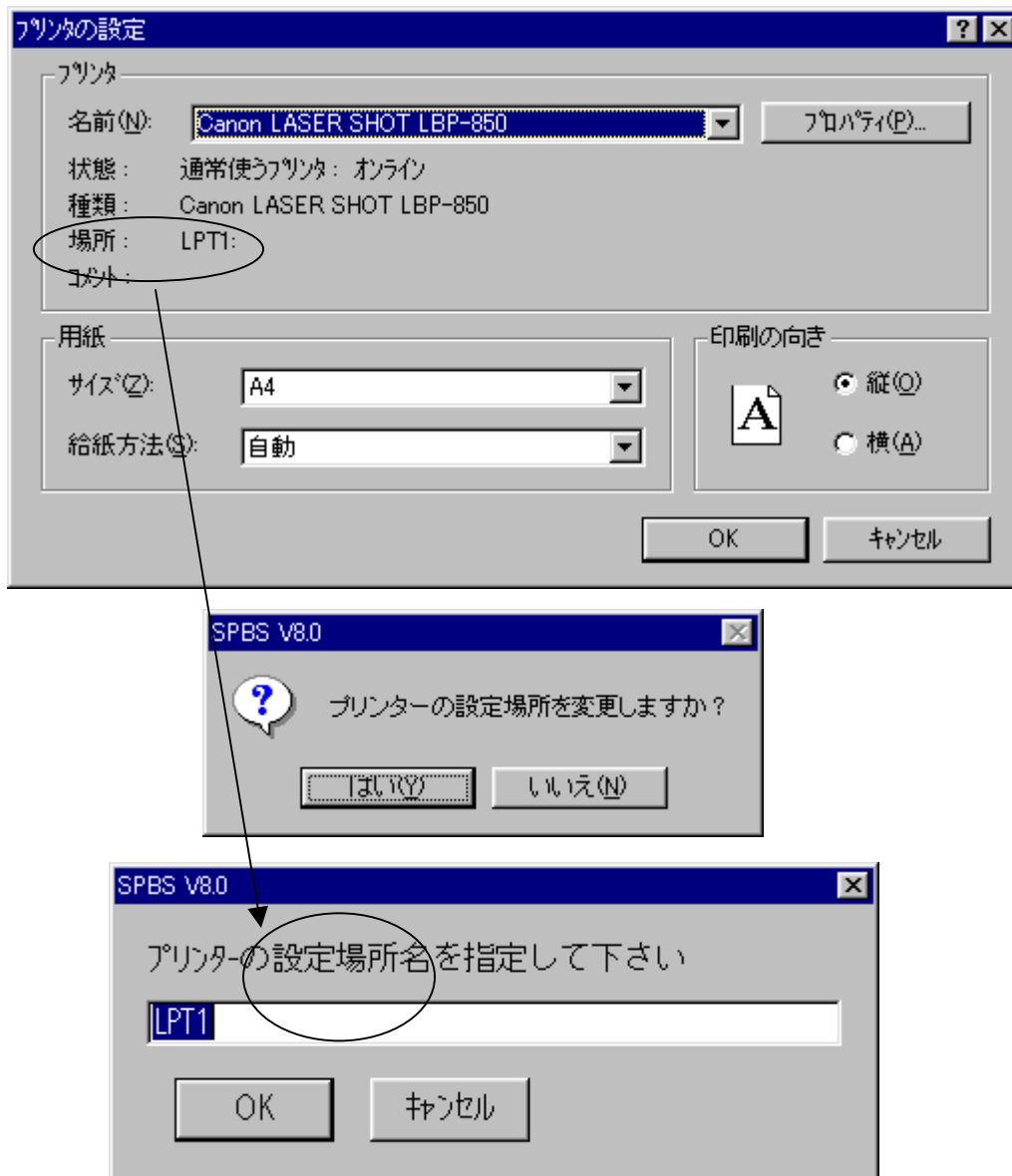
(3) SPBS用データファイルの消去

このプログラムは、SPBS 用のデータファイル（この場合には、「XXX.doc」と「XXX.txt」の両者）やその他のデータファイルをハードディスク（またはディスク）から消去する。しかしながら、ここで一旦消去するとデータの復帰は不可能なので、細心の注意が必要である。

(4) プリンター場所の設定

通常、Windows 用のプリンターは「LPT1:」に設置されているが、これ以外の設置場所に接続されている場合にはプリンターの設置場所を変更する必要がある。「プリンター場所の設定」を選択すると、最初に下図が現れるので、この図のプリンターの「場所」を確認する必要がある。「LPT1:」の場合、何ら変更する必要はない（デフォルト値は「LPT1:」）ので、この図の<キャンセル>ボタン、次の図の<No>ボタンをクリックする。自分の接続しようとするプリンターの「場所」が「LPT1:」以外であれば、それをメモしておき、次の図で<はい(Y)>ボタンをクリックして、プリンターの場所を入力する。

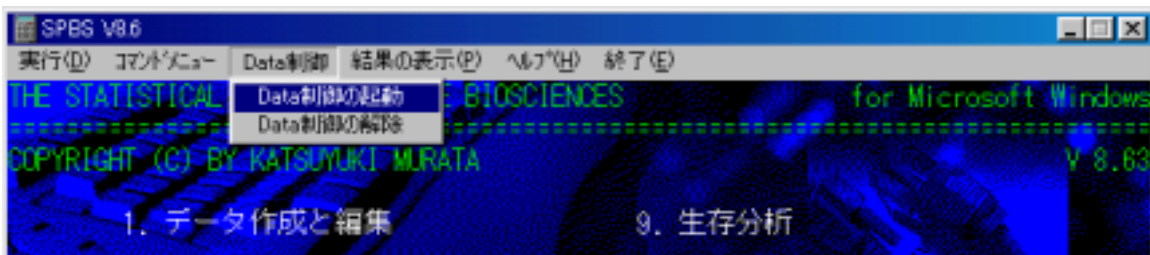
なお、この設定が必要なのは SPBS 統計パッケージで描かれた図を直接プリンター印刷する時のみであり、「ワードパッド」や「ペイント」を使用している時は一切関係がない。



「ワードパッド」や「ペイント」で解析結果をファイルに保存する時は、自分でおこなった解析の内容が後で分かるように新規に「名前を付けて保存」しなければならない。SPBS で付けるファイル名は一時的な名前（「spbstemp.tmp」または「spbstemp.bmp」）であり、次の解析プログラムに移動する際、あるいは終了時に自動的に消去される。

(5) データ制御プログラムの起動

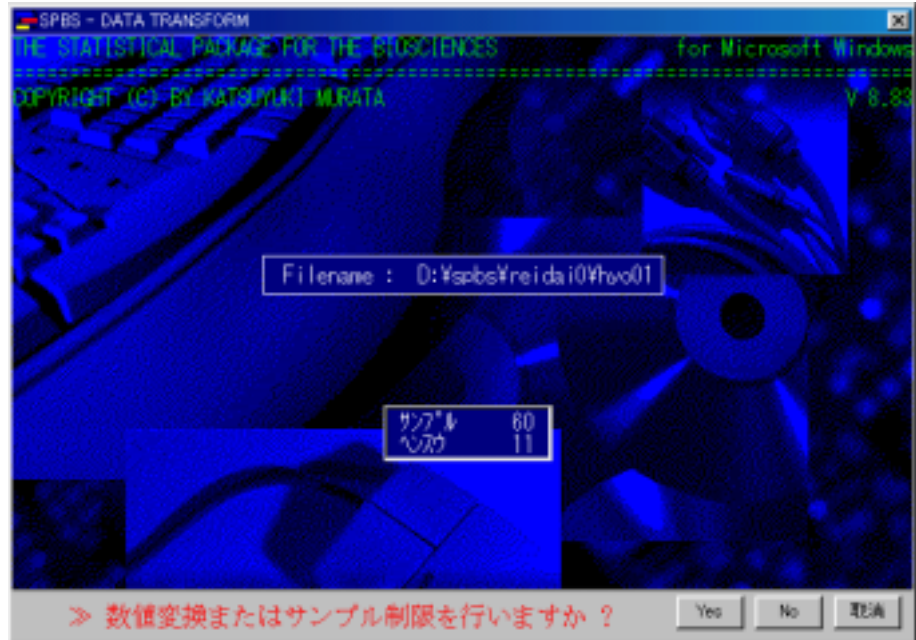
解析プログラムでデータファイルを入力した直後に、データファイルの編集や特定の変数を「対数変換」、「ルート変換」または「Data 制限」したい場合にこのプログラムを起動する。ここでは、次図の<Data 制御>をクリックし<データ制御の起動>を選択する。また、解除する時は<Data 制御の解除>をクリックする。



「データ制御プログラムの起動」を選択すると、各解析プログラムでデータファイルを入力する度に「数値変換またはサンプルを制限しますか？」と尋ねられるようになる（このような操作が不要な場合は、上の操作を再度おこない、＜解除＞ボタンをクリックする）。

ここでの「数値変換」ないし「サンプル制限」は元のデータファイルに一切の書き込みをせず、当該プログラムにおいてのみ有効となるので、同一データファイルを色々な解析プログラムで使用する時は、毎回変換・制限命令をおこなう必要がある。

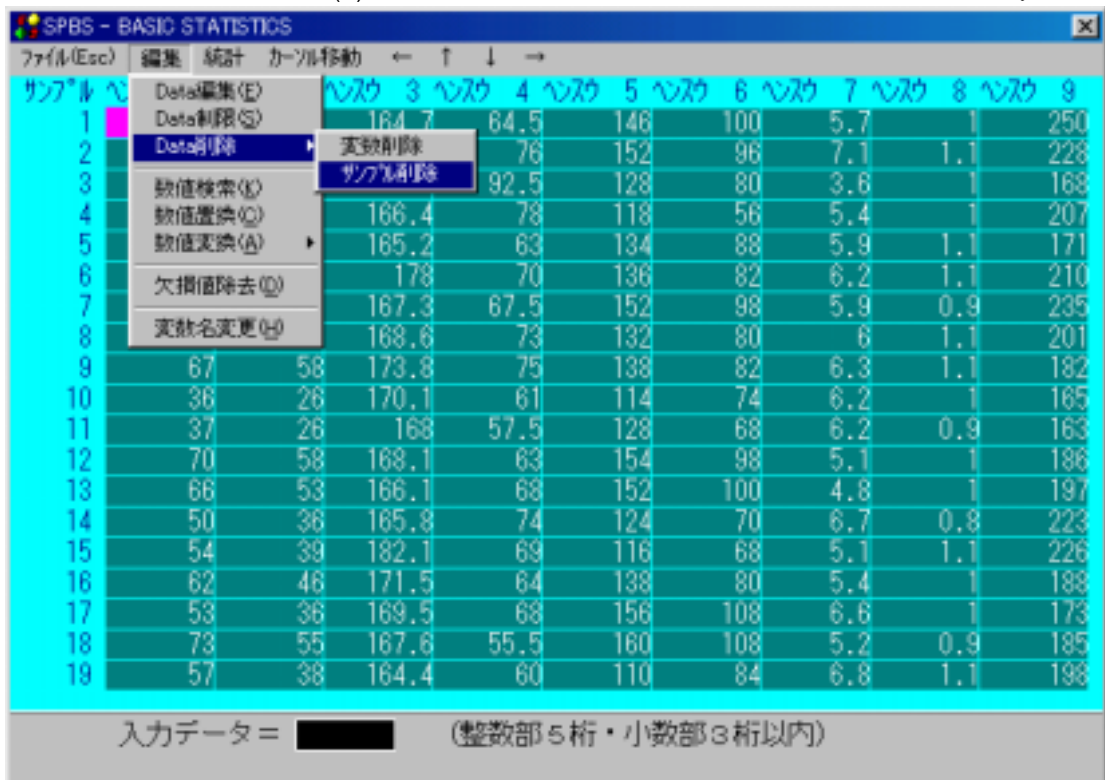
Data 制限は大きなファイルを呼び出し、ファイル内の或る集団（例えば、30～45歳の女性集団）を取り出して解析する場合に利用する（紫色のカーソルを絞り込み用の変数番号の位置に移動し、＜編集＞の中にある＜Data 制限＞ボタンをクリックし、使用するサンプル値の範囲を入力すると実行できる）。



この他、＜編集＞の中には＜数値検索＞、＜数値置換＞、＜欠損値除去＞が、また＜Data 削除＞の中に＜サンプル削除＞と＜変数削除＞がある。但し、変数の追加やサンプルの追加はここではできない（変数あるいはサンプルの追加は、次ページ以降の「ファイルの作成と編集」の「(7)データファイルの編集」か「データの变换・加工」を用いる）。さらに、このプログラムの中で

基本統計量をチェックすることができる。

また、＜終了＞の中に＜File 取消＞があるので、データファイルの選択が間違っていた時、これをクリックすると現在のデータファイルを取消し、データ入力の最初の画面に戻る。



1. ファイルの作成と編集



(1) データファイルの作成と修正

A. データファイルの作成



解析するデータをディスク内ファイルとして保存する。同一データで繰り返し計算する場合、データを長期間

保存したい場合、その他各種データを操作する場合に有用である（以後これをデータファイルと呼ぶ）。なお、データファイルがなければ「Fileの作成と編集」の以下の操作はできない。

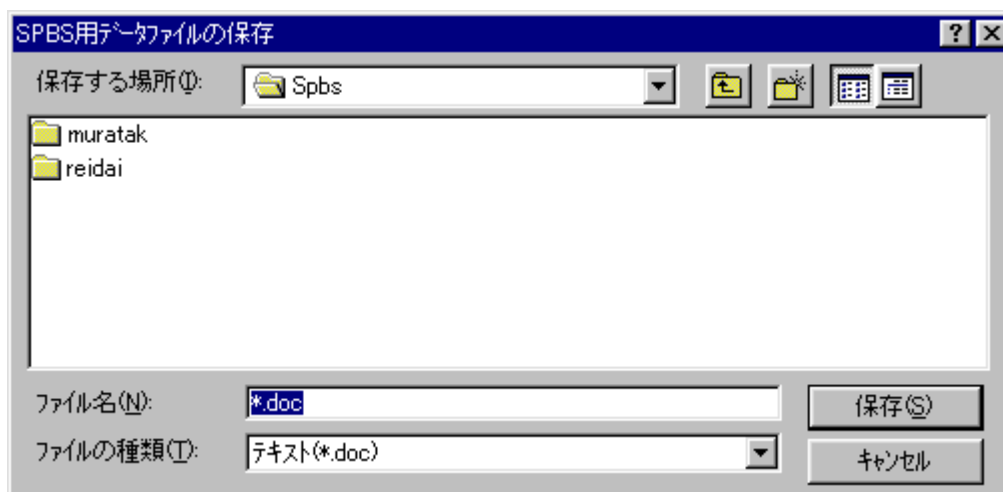
最初にデータの変数（項目数）およびサンプル数を入力する。指定できる変数は最大 999 個、サンプル数 32,000 個である。データの入力、矢印キー（またはマウス）で入力場所までカーソルを移動し、「入力データ」のセル内に数値を入れる。次に入力する場所へ向かう矢印キーを押す。この場合、「入力データ」のセル内には数字、「-」および「・」記号以外は入力できない。また、入力セルが空白（黒くなる）状態でない時に矢印キーを押すと、元（紫色のセル内）のデータが変わってしまうので注意する必要がある。なお、データファイル作成時に、「-99」と「-999」はSPBS 統計パッケージ内の欠損値となっているので、これらの数値は欠損値の時のみ使用し、そうでない時は使用しない（これらの数値があると、各種プログラムの中で自動的に欠損値扱いして処理する）。

データ入力が全て終了したら、<終了>ボタン（または Esc キー）を押す。また縦あるいは横方向にジャンプしたい時は、<JUMP>ボタンをクリックし、ジャンプする個数（最大 999 を越えることはできない）を入れ、矢印キーで移動したい方向を指定する（データの最大変数、サンプル数を超える個数を入力するとジャンプ命令は解除される）。

矢印（、、、）ボタンあるいはキーは各々の方向にセルを 1 つ移動する。また、<<、>>の各ボタンは、それぞれの方向に 1 ページ分（すなわち、横方向では 8 個、縦方向では 19 個）移動する。但し、サンプル数が 19 個未満、あるいは変数が 9 個未満の時にはこれらのボタンは表示されない。

データファイルを作成する時に、グループ番号（例えば、0=対照群、1=患者群）を表す変数を設けると「有意差検定」、「クロス集計」、「判別函数」、「多重ロジスティック分析」、「分散分析」、「生存分析」等を実行する際、1 つのデータファイルを入力し、このグループ番号でファイルを分割して解析することができるようになる。

<計算機>ボタンをクリックすると、Microsoft 社製の電卓が画面上に現れる。これを利用して計算結果の数値をデータ入力画面に入力する場合には、電卓のパネルにある「編集」の「コピー」ボタンをクリックすると、SPBS のデータ入力画面に計算値を移動させることができる（ただし、画面では小数点第 3 位までしか表示されない）。



B. データファイルの修正

データファイルの作成の後で、入力データを訂正したい場合に使用する。エディター画面は「データファイルの編集」と基本的に大体同じであるので、当該頁を参照されたい。

ファイル

<ファイル>ボタンを押す際に、修正後のデータを保存する場合<変換終了>を、また間違っ修正した場合<File 取消>をクリックする（<無変換保存>は<一時保存>ボタンを最後に押した時点までのデータファイルに書き換えられる）。また、入力作業等で長時間このプログラムを使用している時は、<一時保存>を頻回に使用することをお薦めする。これは Windows で時折起こるクラッシュの際に、入力あるいは修正中のデータが全て消失することに対する安全策である（但し、「データファイルの修正」を使用する前のファイルは、データファイルの読み書き中のクラッシュでない限り、無傷である）。長時間修正作業をし、時折<一時保存>ボタンを押している状態で万が一コンピュータがクラッシュまたはフリーズした時は、SPBS 統計パッケージが入っている（インストールされている）フォルダーに「sptempo.doc」および「sptempo.txt」として保存されているので、SPBS 統計パッケージをスタートさせる前にファイル名を変更（例えば「aaa.doc」と「aaa.txt」に変更してから）すれば、クラッシュ前に<一時保存>ボタンを押した時までのデータファイルは使用可能となる（“doc”および“txt”を変えると読み込み不可となる）。この操作をおこなう前に SPBS 統計パッケージをスタートさせ、かつ再び「データファイルの修正」をおこなうと、この一時ファイルは消去されてしまう。

Data 編集

データ編集では、カーソルが置いてある変数において、以下の操作をおこなうことができる。変数間の加減乗除や当該変数に定数を加減乗除することも可能である。なお、データの並べ替え（Sort）をおこなうと、全ての変数においてサンプルの並べ替えが実行されるので、注意する必要がある。

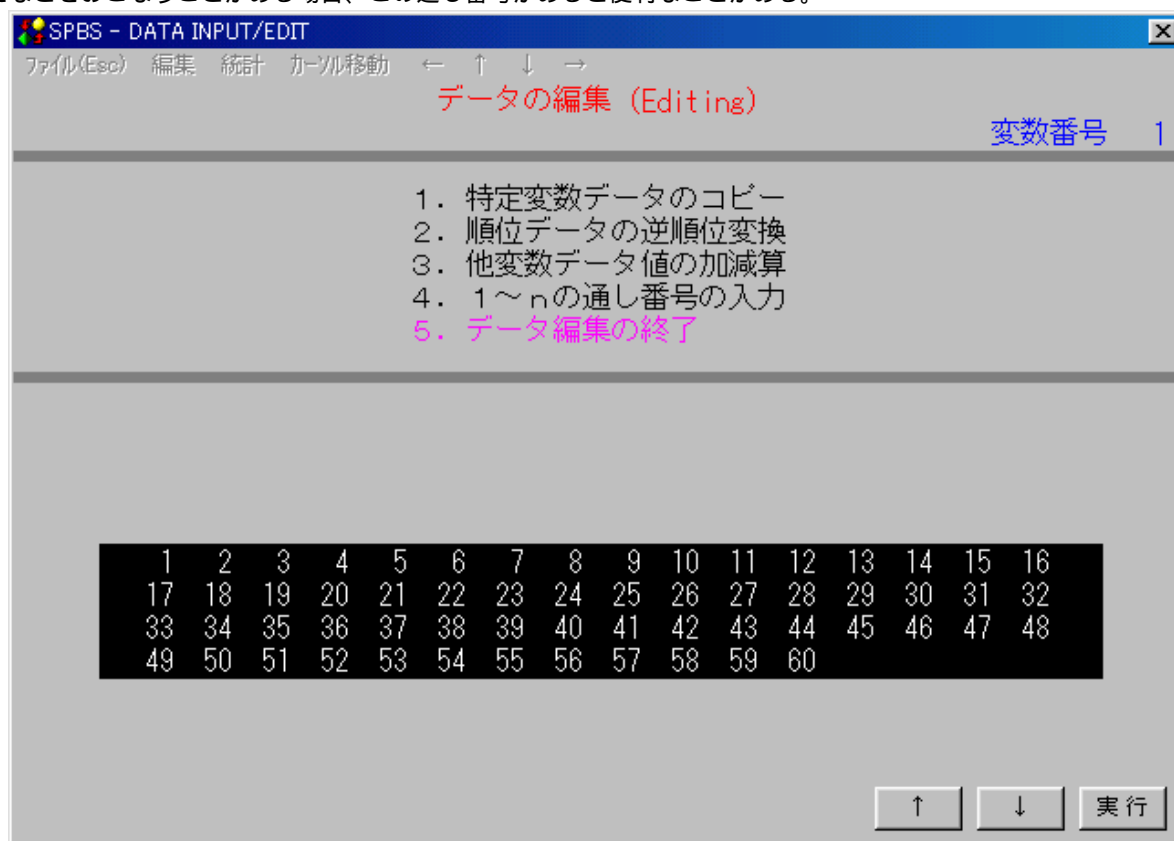
特定変数データのコピー：データ（変数）の追加をおこなった後で、そこに既存の変数番号のデータをコピーする

ことができる。例えば、新規に追加した変数に「体重 (Kg)」項目をコピーし、その値を「身長 (m)」項目で2度除算すると、BMIを算出したことになる。

順位データの逆順位変換：0から15までの整数値からなる変数の場合、この順序を最小値と最大値から、逆順にすることができる。これは、アンケート調査などで、故意に逆順に並べたりすることがあり、これを正順に戻す時に使用する。

他変数データ値の加減算：当初の値に、別の変数の値を加算したり、減算したりすることができる。例えば、食前データと食後データがあるような場合に、食前後の差を算出することが可能である。

1～nの通し番号の入力：1から総サンプル数までの通し番号を入力することができる。あとで、サンプルの並べ替えなどをおこなうことがある場合、この通し番号があると便利なおことがある。



Data 制限

カーソルのある当該変数の数値（最小値～最大値）内の特定範囲を指定すると、その区間の数値をもったサンプルをデータファイルから使用する、削除することができる。この操作をおこなった後は、データ数が当然のことながら減少する。例えば、「1 = 男、2 = 女、0 = 無記入」が入力されている変数で、男子のみ抽出したデータファイルを作成する場合には、の「使用する」を選択し、抽出する最小値および最大値とともに1を入力する。すると、女子 (= 2)、無記入者 (= 0) を除いたデータファイルが作成される。同様に、変数に年齢が入力されている場合、特定の年齢階級の値を入力すると、その年齢階級のデータのみ選択できる。この操作は反復繰り返すことができるので、例えば男子30～45歳のデータを抽出することも可能である。

また、データに「1 = 良い、2 = どちらとも言えない、3 = 悪い」が入力されている変数で、曖昧な「どちらとも言えない」を除いたデータファイルを作成することもできる。すなわち、「削除」を選択し、最小値および最大値に2を入力すると、2のみが削除されたデータファイルが作成される。

なお、この「Data 制限」は他の解析プログラムでデータファイルを選択した直後に「データの数値制限をしますか?」と出た際に（このためには、メインメニューにある「データ制御の起動」にチェックが入っていなければならない）、「はい(Y)」を選択し、実行することができる。この場合、当該解析においてのみ書き換えられる。

数値検索

数値検索では、指定した数値が当該変数内に何個あるのか、何番のサンプル番号にあるのかを調べることができる。

数値置換

数値置換では、指定した数値を検索し、カーソルのある当該変数内の数値を置換する。この数値置換プログラムを解除したい場合は「-9999」を入れる。

数値の種類（数種）が15個未満の時は「リコーディング (Recoding)」プログラムが自動的に起動する。すなわち、全ての数種が画面に現れ、それらに変換後の数値を入れることによって一括して変換をすることができる。例えば、部位番号、職種番号を再構成する時などに大いに利用できる。

数値変換

数値変換機能には、以下の7つの関数を保持している。計算は元データが8桁の単精度数字で処理しているため、変換後の値も単精度である。このうち、対数は底を10とした場合と、eとした場合の2種類ある。これらの「対数」および「ルート」変換値については、変換後に「元に戻す」をクリックすると、変換前のデータに戻すことができる。

The screenshot shows the SPBS - DATA INPUT/EDIT window with a menu open over the '数値変換(A)' option. The menu includes: 無変換, 対数(10), 対数(e), Root(√), 2値変換, and 元に戻す. The data table below shows 19 samples and 9 variables.

サンプル	1	2	3	4	5	6	7	8	9
1	146	100	250	183	19	33	111		
2	152	96	228	141	76	44	111		
3	128	80	168	183	18	33	100		
4	118	56	207	86	20	11	90		
5		88	171	129	34	20	100		
6		82	210	108	20	33	105		
7		98	235	93	19	26	98		
8		80	201	103	29	22	95		
9	9	5	82	182	132	37	41	106	
10	10	2	74	165	77	15	18	83	
11	11	26	128	68	163	80	22	15	94
12	12	58	154	98	186	220	19	65	98
13	13	53	152	100	197	94	15	76	104
14	14	36	124	70	223	137	45	53	80
15	15	39	116	68	226	105	21	21	96
16	16	46	138	80	188	65	19	19	95
17	17	36	156	108	173	405	34	66	86
18	18	55	160	108	185	70	19	28	104
19	19	38	110	84	198	45	17	9	91

入力データ = (整数部5桁・小数部3桁以内)

欠損値処理

SPBS 統計パッケージでは、欠損値に「-999」または「-99」を割り当てている。カーソルのある当該変数内の欠損値を含むサンプルデータを削除するプログラムである。このため、この操作を実行したあとでは、データファイルのサンプル数が入力時に比べ、減少する。

変数名変更

変数の内容が元の内容と変わった時に、ここで修正することができる。変数名は全角 17 文字、半角 34 文字まで入力可能であり、それ以上は自動的にカットされる。

統計

個々の統計量： カーソルの置いてある変数の平均値、標準偏差、最小値、最大値（欠損値がある場合には、その欠損値を除いたデータで算出）、サンプル数が表示される。「サンプル数」の前に「*」印がついている場合は、欠損値があることを示す。

全体の統計量： データファイルにある全ての変数の平均値、標準偏差、最小値、最大値が（欠損値がある場合には、その欠損値を除いたデータで）算出される。また、サンプル数が5000以下の時には、さらに中央値も算出される。このデータは「印刷」ボタンをクリックすると、プリンターに印刷される。なお、サンプル数が2未満の場合には算出されない。

カーソルの移動

カーソルの移動をおこなう。ここでは「↑」、「↓」、「←」、「→」のそれぞれの方向に1頁分移動する。また、「任意の移動」では、左右前後に最大 999 個移動することができる。最初に移動したい1~999の数値を入力し、その後矢印キーを押す（矢印キーは自分の移動したい方向を押す）。

このほか、画面の数値データの周辺にあるサンプル番号、変数番号が書かれているところをクリックすると、所定の数移動することができる。画面の「サンプル」と書かれているところをクリックすると、左上のセルが変数1、サンプル1のデータがくるように移動する。サンプル数の最下部（起動時の場合、サンプル19）をクリックすると、データファイルの最終サンプル番号のデータが一番下になるように移動する。また、変数の最右端（起動時の場合、変数9）をクリックすると、データファイルの最終変数番号のデータが右端になるように移動する。これらの移動の仕方は、メインメニューの「ヘルプ」をクリックした際の「データファイルの作成と修正」のところに詳述されている。

データ作表画面の右（下）端の変数（サンプル）番号をマウスでクリックすると、最後部の変数（サンプル）番号のデータが画面の右（下）端に移動する。同様に、左（上）端の変数（サンプル）番号をマウスでクリックすると、1番目の変数（サンプル）データが画面の左（上）端に移動する。

同様に、データ作表画面の上述の位置より1～2個中央寄りをマウスでクリックすると、矢印キー1回分または5（縦向きは10）回分の操作に相当する（画面の上端2、3番目の変数番号は、同6、7番目は、左端2、3番目のサンプル番号は、同17、18番目はに移動）。

(2) データファイル Excel (CSV) SPBS file の変換

Microsoft 社製の Excel の CSV ファイル（個々のサンプルのデータをカンマで区切った形式で、拡張子が CSV になっているファイル）を SPBS 用のファイルに変換する。また、SPBS 用ファイルから Excel の CSV ファイルに戻すことができる（Excel では最大変数が 256 列に制限されているので、SPBS 用ファイルの変数が 256 を越えると、Excel 用 CSV ファイルは作成されない）。Excel 内の横第一列に変数名が入っている場合には、その変数名を読み取り SPBS 統計パッケージの変数名データに自動的に取り込む（これ以外の文字列は全て欠損値扱いする）。SPBS 統計パッケージでは半角 34 文字（または全角 17 文字）までを変数名データとして使用できる。

変数名ファイルは、SPBS 用に変換した後作成するようにプログラムされている。Excel ファイルに変数名がある場合は上述のように同時に変換するが、無い場合には必ず半角 34 字（全角で 17 字）以内で入力した方がよい。なお、Excel ファイルの変数名に半角の「カンマ」、「セミコロン」等があると変換後変数名エラーが発生するので、これらは使用前に変換しなければならない。

Excel ファイルのデータ内に「M」（=male）とか「F」（=female）の英文字が入っていると、これらは欠損値扱いとして SPBS 用ファイルでは「-999」となる。Excel 上で事前に数値化してから、このプログラムを使用する。また、SAS プログラムなどで欠損値としてしばしば使用される「・」（ピリオド）もこのプログラムでは「-999」になる。これら欠損値は「データファイルの編集」の<欠損値処理>を用いて削除あるいは置換することができる（SPBS 統計パッケージでは「-999」を欠損値として除外して計算する）。

(3) 固定長データの変換（クリップボードデータの変換を含む）

このプログラムでは、SAS やフォートランのデータのように固定長のデータを SPBS 用データに読み代える操作をおこなう。入力データはテキストリーダーで読める形式（テキストファイル形式）、各サンプルデータが一行で書かれている、文字データは含まれていないことが条件である。したがって、一人分のデータが（エンターキーで区切られた形で）数行に渡って入力されているデータの場合には、一行に修正した後でなければ変換できない

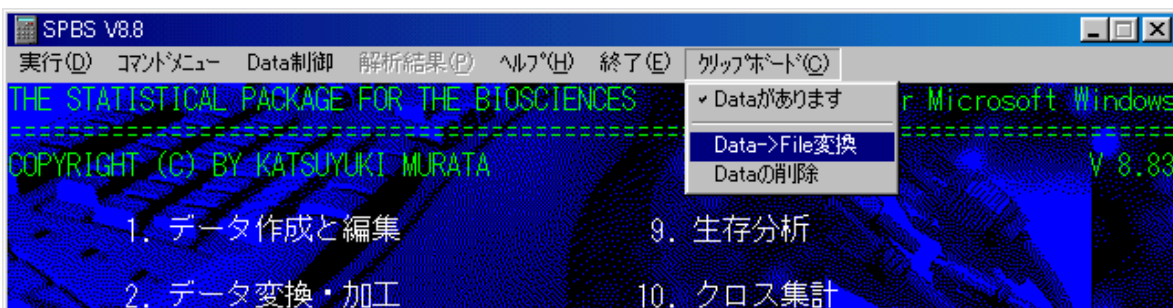
固定長データの変換（FORTRAN DATA ⇔ SPBS DATA 変換）

File name: D:\\$SPBS¥reidai¥hyo01.dat

1	10	20	30	40	50	60				
1	45	146	100	250	183	19	33	111	23.8	0
2	47	152	96	228	141	76	44	111	23.8	0
3	38	128	80	168	183	18	33	100	27.8	0
4	30	118	56	207	86	20	11	90	28.2	0
5	35	134	88	171	129	34	20	100	23.1	0

変数を何個設定しますか（一括変換の時は「*」入力）

（変換前の一人当たりのデータ長は 1000 カラムを超えてはならない。また、変換後のデータの変数は 100 個以下になるように設定しなければならない）。また、サンプル数は 32,000 個未満でなければならない。



「固定長 SPBS 用」への一括変換も可能である。この場合、データ間にスペースが最低 1 個以上入っている必要がある。また、「SPBS 用 固定長」のプログラムを使用すると、SPBS 用データから SAS 用のデータファイルを作成することができる。すなわち、「固定長でかつ変数データ間に（最低 1）スペースを挿入した」ファイルに置換される。SPBS 統計パッケージにない解析を SAS でおこないたい時に利用できる。

データを入力する際 Excel のカンマ区切り（CSV）ファイルを取扱うことができるので、結果的に「Excel CSV file

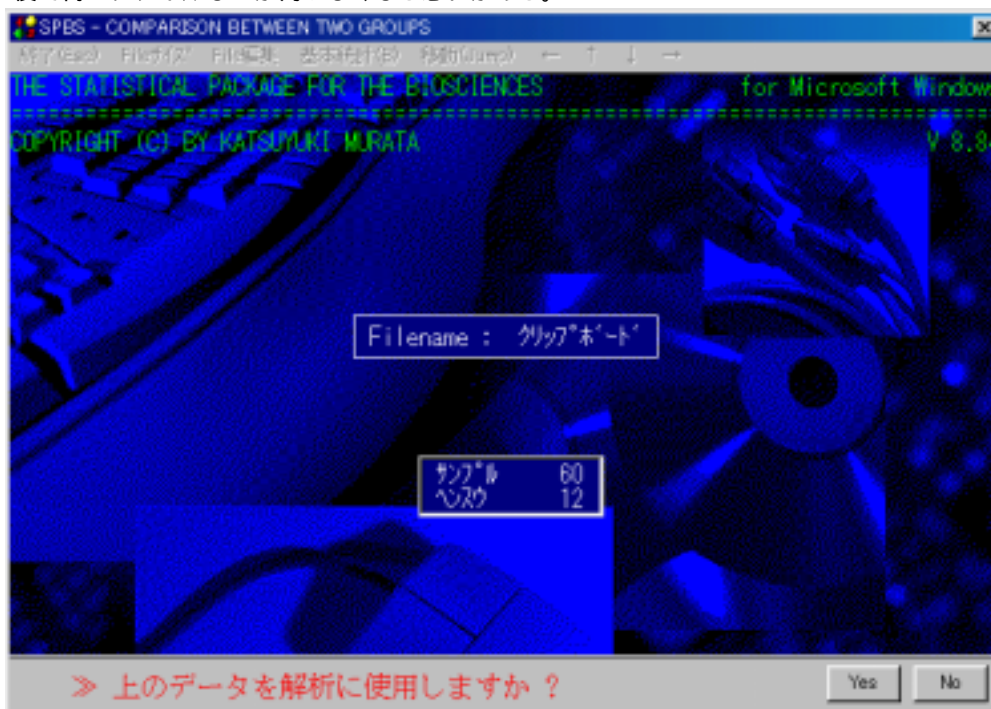
固定長ファイル」の変換を1度に行うことが可能である。

	A	B	C	D	E	F	G	H	I	J
1	ID番号	年齢	収縮期血圧	拡張期血圧	総コレステ	中性脂肪	GOT	γ-GTP	FBS	BMI
2	1	45	146	100	250	183	19	33	111	23
3	2	47	152	96	228	141	76	44	111	23
4	3	38	128	80	168	183	18	33	100	27
5	4	30	118	56	207	86	20	11	90	28
6	5	35	134	88	171	129	34	20	100	23
7	6	31	136	82	210	108	20	33	105	22
8	7	57	152	98	235	93	19	26	98	24
9	8	31	132	80	201	103	29	22	95	25
10	9	58	138	82	182	132	37	41	106	24
11	10	26	114	74	165	77	15	18	83	21
12	11	26	128	68	163	80	22	15	94	20
13	12	58	154	98	186	220	19	65	98	22
14	13	53	152	100	187	94	15	76	104	24
15	14	36	124	70	223	137	45	53	80	26
16	15	39	116	68	226	105	21	21	96	20
17	16	46	138	80	188	65	19	19	95	21
18	17	36	156	108	173	406	34	66	86	23
19	18	55	160	108	185	70	19	28	104	19

クリップボードデータの利用

表計算ソフトの数値データ（**但し、変数名を除く**）をクリップボードにコピーすると、その数値データを SPBS 用データファイルに変換できる。さらに、SPBS 統計パッケージの解析結果の数値部分（例えば、因子得点）をクリップボードにコピーすると、その数値データも SPBS 用データファイルに変換できる。但し、いずれの場合にも変数名ファイルをその直後に作成しないと、後で何のファイルなのか判らなくなる恐れがある。

表計算ソフト（または「ワードパッド」）上でコピーすると、SPBS 統計パッケージのメインメニュー画面上方の **<クリップボード>** という黒文字が表示される。これ以後は、この文字が消えるまでは当該データを各々の解析プログラムで使用できる。つまり、解析プログラムを選択して「データファイルによる入力」を選択すると、右図のように画面に「Filename: クリップボード」と表示されるので、「上のデータを解析に使用しますか?」に対し **<Yes>** ボタンをクリックする。データは自動変換され、解析で使用することができる。



(4) データファイルの変数名の変更

データファイルの確認では、変数名とともに各変数のサンプル 15 個、有効サンプル数、欠損値数、平均値、最小値、最大値を画面に表示することもできる。

データの変数名ファイルの変更では、間違っ て入力した変数名を修正することができる。変数名ファイルの拡張子は“.doc”で「Word」の保存ファイルと同じである。データファイルの変数、変数名、サンプル数の情報はこのファイル内に入っているの で、削除するとデータファイル（ファイルの拡張子は“.txt”）の認識が不可能になる。

変数名の文字数は半角で 34（全角では 17）文字まで使用できるが、一部の全角文字で

ID No.	1	2	3	4	5	6	7	8	9	10	11
1	45	146	109	259	183	19	83	111	29.8	8	
2	47	152	98	228	141	76	44	111	28.8	8	
3	38	123	99	168	169	16	83	188	27.8	8	
4	38	118	96	207	88	26	11	98	28.2	8	
5	55	134	99	171	129	34	26	188	29.1	8	
6	31	139	92	219	169	26	29	188	27.1	8	
7	57	152	98	229	83	19	26	98	24.1	8	
8	31	132	99	201	163	29	22	95	26.7	8	
9	58	138	92	182	182	37	41	188	24.8	8	
10	28	114	74	165	77	15	18	88	21.1	8	
11	28	128	98	163	80	22	15	94	28.4	8	
12	58	154	98	188	230	19	86	98	22.9	8	
13	58	152	109	197	84	15	76	184	24.8	8	
14	38	124	73	229	137	45	83	88	26.3	8	
15	38	118	98	229	166	21	21	98	28.8	8	
16	48	139	99	198	65	19	19	98	21.8	8	
17	38	138	108	173	406	34	86	86	23.7	8	
18	58	163	109	186	79	19	28	184	19.8	8	
19	28	118	84	188	85	15	8	91	28.8	8	

入力出来ない場合がある。これは半角文字中の禁止文字（例えば、<カンマ、セミコロン、ダブルクォーツ等>）が全角文字を構成する記号の中に含まれる場合である。このような時は別の日本語に置換する等の工夫が必要となる。

変数の数および変数名が全く同じデータファイル（サンプル数は異なっても構わない）の場合には、既に作成されている別の変数名ファイルを転用することができる。「このファイルを別のデータの変数名ファイルに変更しますか？」と尋ねてきた時に「Y」を入れる（サンプル数が違う時は、新たなサンプル数を入力する）。

(5) データファイルの縦横変換

通常、データファイルを作成する時、SAS や SPBS 統計パッケージでは縦方向にサンプル 1, 2, …、横方向に変数 1, 2, …と入力しますが、Excel でこの縦横（行と列）を逆に入力してしまったりすることがある。この種のデータの縦横を逆にする。

(6) 2つのファイルの比較検討

SPBS 統計パッケージでデータ解析をおこなう時、しばしば新たなデータファイルを作成することがある（SPBS 統計パッケージの「クロス集計」の中の「繰り返しのある二元配置分散分析」では、性・年齢階級別の比較ファイルを分割することなくおこなうことができる。しかし「サンプル制限」機能を使用しない時には、性・年齢を含む各種データが1つのファイル内に存在する場合ファイルを男女別に分割しなければならない）。このため、データファイルが幾つも作成され、似通ったファイルが沢山できてしまう恐れがある。このような時、このプログラムを使用して、幾つかの変数を指定して2つのファイルに同じサンプルのデータがどれほどあるのかを調べることができる。また、「データファイルの編集」の中の「ID 準拠によるデータファイルの合成」を利用して1つのファイルに再編集することも可能である。

(7) データファイルの編集

上書き保存

入力時のデータファイル名に上書き保存する。この上書き保存が終了すると、プログラムはメインメニューに戻る。入力時のファイル名と異なるファイル名にしたい場合には、次の「名付けて保存」を使用する。

名付けて保存

入力時のデータファイルをそのまま保存しておき、新規のファイル名を付けて保存したい場合にこれを使用する。この保存が終了すると、プログラムは自動的にメインメニューに戻る。ここでは、EXCEL のカンマ区切りファイル（.CSV）や固定長ファイル（.DAT）に変換して保存することも可能となっている。SPBS 用ファイルでない場合、ファイル名を指定する時に「.CSV」または「.DAT」を同時に書き込まないと、SPBS 用ファイルと認識され保存される。なお、固定長ファイルの場合には、変数名は保存されない。

一時保存

データ入力時に、コンピュータからしばらく離れていたりしていると、画面が凍結（フリーズ）状態になることがある。また、何かの折りに電源が切れてそれまで入力していたデータが全て吹っ飛ぶことがある。これを避けるために、頻回「一時保存」をしておくと、一時的ファイル「sptempo.doc」（変数名ファイル）と「sptempo.txt」（数値ファイル）という名前で保存される。これら2つのファイルは「データファイルの編集（変換と加工）」を起動（あるいは正常に終了）した時点で消去するので、SPBS 統計パッケージを起動する前に両者のファイル名を変更しておく必要がある（変更時、両者とも同じ名前を使用しないと、SPBS 統計パッケージではエラーが生じる）。

一時 File の読み込み

データ編集などで間違った操作をして、その直前のデータに戻りたい場合に、このボタンをクリックする。上の「一時保存」をクリックした時のデータが読み込まれる（一度も「一時保存」ボタンを押していないときは、入力時のデータが読み込まれる）。

ファイルの合成

最初に入力したファイルに更なるデータファイルを合成したい時に使用する。「変数の増加」、「サンプル数の増加」、「ID 番号で照合した2つのファイルの合成」ができるが、前2者はデータファイルのサンプル数が等しいか、変数の数が等しくないと合成できない。

変数の増加：サンプルの数が等しいデータファイルを合成する。同じサンプルのデータを測定1と測定2とで作成した場合などに使用する。

サンプル数の増加：変数の数が等しいデータファイルを合成する。同じ項目（変数）でフォーマットされた2つのデータファイルを合成したい場合に使用する。

ID 準拠の合成：2つのデータファイルがあり、共通する ID が存在する時にこのプログラムを使用することができる。この場合、**共通する ID 番号**のみのデータファイルを作成することも、**全ての ID 番号**を一つにすることも可能である。しかしながら、それぞれのデータファイル内に同じ ID 番号が存在すると、この合成を中止する。この場合、何番の ID 番号が重なっているのかを解析後に示すので、データ編集の途中でチェックすることができる（ID 準拠の合成に失敗した時は、最初に入力したデータファイルが画面に現れる）。同じ ID 番号のデータは削除するか、別の ID 番号に修正するまで合成することはできない。

保存せず終了

データファイルの編集の途中で保存する必要がない場合に、このボタンをクリックする。この後、メインメニューに戻る。

保存せず新規編集

データファイルの編集の途中で保存する必要がない場合で、別のデータファイルを編集したくなった時にこのボタンをクリックする。なお、この「データファイルの編集（変換と加工）」をデータファイル作成として使用したい場合には、入力ファイルの指定の際に、「キャンセル」ボタンをクリックすると、新規に作成するファイル名、サンプル数、変数の数を尋ねてくるので、その後でデータ入力をおこなうことができる。

Data 編集

データ編集では、カーソルが置いてある変数において、以下の操作をおこなうことができる。変数間の加減乗除や当該変数に定数を加減乗除することも可能である。なお、データの並べ替え（Sort）をおこなうと、全ての変数においてサンプルの並べ替えが実行されるので、注意する必要がある。

特定変数データのコピー：データ（変数）の追加をおこなった後で、そこに既存の変数番号のデータをコピーすることができる。例えば、新規に追加した変数に「体重（Kg）」項目をコピーし、その値を「身長（m）」項目で2度除算すると、BMI を算出したことになる。

順位データの逆順位変換：0 から 15 までの整数値からなる変数の場合、この順序を最小値と最大値から、逆順にすることができる。これは、アンケート調査などで、故意に逆順に並べたりすることがあり、これを正順に戻す時に使用する。

他変数データ値の加減算：当初の値に、別の変数の値を加算したり、減算したりすることができる。例えば、食前データと食後データがあるような場合に、食前後の値を算出することが可能である。

1 ~ n の通し番号の入力：1 から総サンプル数までの通し番号を入力することができる。あとで、サンプルの並べ替えなどをおこなうことがある場合、この通し番号があると便利なおことがある。

データの並べ替え (Sort)：カーソル（紫色の数値）のある変数を小 大へ、あるいは大 小に並べ替える。同順位のデータが多数ある時には、当該変数の他にもう1つ変数を選んで、並べ替えることも可能である。なお、この「データの並べ替え」をおこなうと、当該変数以外の**全ての変数のデータが連動して**並べ替わる。

小数点以下の値の桁調整：数値変換をおこなうと、7 ~ 8 桁の数値になってしまい、小数点第3位以下の数値は画面に現れないにもかかわらずデータに含まれている。このままデータ解析をおこなうと、画面で表示された数値とデータファイルに保存された数値が異なっていることになり、一見結果が違っているように見える。このプログラムでは、小数点以下、第1位（結果として整数になる）、第2位、第3位、第4位（SPBS 統計パッケージの画面では小数点第3位までしか表示できない）で四捨五入することができる。

他変数データ値の乗除算：当初（カーソルが置いてあった変数）の値に、別の変数の値を乗算したり、除算することができる。除算の場合、当該変数に0の値が1つでもあると除算が実行されないため、「数値検索」で事前に調べておく必要があるかもしれない。

ダミー変数(0, 1)の作成：多重ロジスティック分析をおこなう際に、カテゴリーデータを使用することが多々ある。このカテゴリーが(0, 1)の時に算出された Odds 比は或る病態（または推定発生因子）の有無が当該疾患（目的変数）の有無にどれくらい寄与するかを示す。しかし、1 ~ 4 等のカテゴリーになると、等間隔の影響を及ぼすのかどうかについて何ら情報を与えてくれない。このような時、1 ~ 4 のカテゴリーを 1=(0,0,0)、2=(1,0,0)、3=(0,1,0)、4=(0,0,1)のようなダミー変数に置換（Partial 法）して計算すると、1と2、1と3、1と4の関係が明瞭になる（この場合、1が対照群と見なされる）。

カテゴリーデータを多変量解析する時、順序データ（例えば、1 = 大変悪い、2 = 悪い、3 = どちらとも言えない、4 = 良い、5 = 大変良い）の場合はそのまま使用することが可能であるが、この順序が全く意味を持たないこともある

(例えば、1 = 頭部、2 = 頸部、3 = 胸部、...)。後者の場合、データをダミー変数に変換すると、解析結果の意味ある評価をおこなうことができる。ここでは、Partial 法に基づいてダミー変数を作成する。例えば、当該変数の元の数値が 1 ~ 5 であれば、データファイルの最後の変数に 4 つの変数が追加され、1 = (0,0,0,0)、2 = (1,0,0,0)、3 = (0,1,0,0)、4 = (0,0,1,0)、5 = (0,0,0,1) が割り当てられる。欠損サンプルに対しては、この 4 変数に「-999」が挿入される。データの数値の種類(数種)が 2 以上、6 以内であれば自動で割り振りされるが、連続量の場合にはカテゴリー数を尋ねてきて、その後で各カテゴリーの境界値を入力する(カテゴリーが 5 個であるならば、境界値は「カテゴリー」- 1、すなわち 4 個となる)。なお、入力データの変数の数が既に 994 以上の場合は、このプログラムが作動しない(SPBS 統計パッケージの使用最大変数は 999 に固定されているため)。

定数データ値の加算：カーソルのある当該変数の値に、定数を加算することができる。但し、定数が 0 の場合には無視される。

定数データ値の減算：カーソルのある当該変数の値から、指定した定数を減算することができる。但し、定数が 0 の場合には無視される。

定数データ値の乗算：カーソルのある当該変数の値に、定数を乗算することができる。但し、定数が 0 または 1 の場合には無視される。

定数データ値の除算：カーソルのある当該変数の値を、指定した定数で除算することができる。但し、指定定数が 0 または 1 の場合には、無視される。

Data 制限

カーソルのある当該変数の数値(最小値~最大値)内の特定範囲を指定すると、その区間の数値をもったサンプルをデータファイルから使用する、削除することができる。この操作をおこなった後は、データ数が当然のことながら減少する。例えば、「1 = 男、2 = 女、0 = 無記入」が入力されている変数で、男子のみ抽出したデータファイルを作成する場合には、この「使用する」を選択し、抽出する最小値および最大値とともに 1 を入力する。すると、女子(= 2)、無記入者(= 0)を除いたデータファイルが作成される。同様に、変数に年齢が入力されている場合、特定の年齢階級の値を入力すると、その年齢階級のデータのみ選択できる。この操作は反復繰り返すことができるので、例えば男子 30~45 歳のデータを抽出することも可能である。

また、データに「1 = 良い、2 = どちらとも言えない、3 = 悪い」が入力されている変数で、曖昧な「どちらとも言えない」を除いたデータファイルを作成することもできる。すなわち、「削除」を選択し、最小値および最大値に 2 を入力すると、2 のみが削除されたデータファイルが作成される。

なお、この「Data 制限」は他の解析プログラムでデータファイルを選択した直後に「データの数値制限をしますか?」と出た際に、「はい(Y)」を選択し、実行することができる(このためには、メインメニューにある「データ制御の起動」にチェックが入っていないなければならない)。この場合、元データは書き替えられない。

Data 削除

「データ削除」では、変数丸ごと削除、特定のサンプルデータの削除、変数の並べ替えをおこなうことができる。

変数削除：カーソルのある変数を丸ごと削除することができる。当該変数番号はデータ表示されている画面の下に表示される。

サンプル削除：カーソルのあるサンプルのみを削除する。当該サンプル番号はデータ表示されている画面の下に表示される。

変数並べ替え：入力されたデータファイルの変数の並べ替えが必要な時にこれを使用することができる。この場合、変数の数を指定するので変数の数を減らすことはできるが、変数を元のデータよりも増加させることはできない。新しく作成するデータファイルの変数番号(1, 2, ...)に元の変数番号を入力することで並べ替え操作は終了する。変数番号を入力しないで、「エンターキー」のみ押すと、新しく作成するデータの変数番号と同じ変数番号が自動的に入力されてしまう。また、数値キーで入力すると、元の変数名が自動的に表示されるので、その時点でチェックする。

Data 挿入

「データ挿入」では、データファイルに新たな変数やサンプルを追加することができる。

変数追加：このボタンを 1 回クリックすると、その都度最終変数の後に 1 変数が追加される。また、追加と同時に、新たな変数名の入力を要求してくる。後でデータ解析することを考えると、この時点で変数名を入力しておくことを勧める。さらに、複数の変数を追加することも可能である。

サンプル追加：このボタンを 1 回クリックする度に、1 サンプルデータが追加される。このデータは最終サンプルの後に追加される。また、複数のサンプルを追加することも可能である。

数値検索

数値検索では、指定した数値が当該変数内に何個あるのか、何番のサンプル番号にあるのかを調べることができる。

数値置換

数値置換では、指定した数値を検索し、カーソルのある当該変数内の数値を瞬時に置換する。この数値置換プログラムを解除したい場合は「-9999」を入れる。

数値の種類(数種)が 15 個未満のときには「リコーディング(Recoding)」プログラムが自動的に起動する。すなわち、全ての数種が画面に現れ、それらに変換後の数値を入れることによって一括して変換をすることができる。例えば、部位番号、職種番号を再構成する時などに大いに利用できる。

数値変換

数値変換機能には、以下の7つの関数を保持している。計算は元データが8桁の単精度数字で処理しているため、変換後の値も単精度である。このうち、対数は底を10とした場合と、eとした場合の2種類ある。これらの「対数」および「ルート」変換値については、変換後に「元に戻す」をクリックすると、変換前のデータに戻すことができる。特殊変換の中にある「2値変換」、「logit変換」、「基準化」、「逆変換」、「Fisher変換」、「順位値変換」は元に戻すことができない（前者においても、何度もデータ変換していると、桁落ちのため元の数値に戻らないことがあるので、繰り返し使用しない方がよい）。



「2値変換」は、元のデータを0と1に変換する。このため、境界値を指定する必要がある。「Logit変換」は、数値xを $1/(1 + \exp(-X))$ に当てはめる。得られる数値は0から1の値になるが、データに±15以上の値があると、実際問題計算しても0または1になるので、その数値より大きい値が含まれる場合には、この変換をおこなわない。「基準化」は、集団の平均値 Mean および標準偏差 SD を算出し、平均値0、標準偏差1になるように変換する。すなわち、データ値がXであるなら、 $(X - \text{Mean}) / \text{SD}$ で算出される。さらに、正規確率値を算出することも、その正規確率値を基準化データに戻すことも可能である。「順位値変換」は当該変数に（小さい方から大きい方に）順位を付ける（同順位の場合には、それらの中央値を入れる）。

分割変数

ここでは、2変数を用いてデータを4分割または3分割する。この分割した時の位置を示す番号（1, 2, 3, 4または、1, 2, 3）は新規の変数に書き込まれる（この変数は最後に追加される）。

2変数による4分割：2変数の各々の分割値（境界値）を入力すると、4分表が作成され、それぞれに含まれる数を表示するとともに、1, 2, 3, 4が割り当てられ、追加された変数に自動的に入力される。

2変数による3分割：高血圧の定義のように、収縮期血圧 140mmHg、160mmHg および拡張期血圧 90mmHg、95mmHg を入力すると、正常範囲（= 1）、境界範囲（= 2）、異常範囲（= 3）が判定され、追加された変数に自動的に入力される。

欠損値処理

SPBS 統計パッケージでは、欠損値に「-999」または「-99」を割り当てている。この数値のデータをどう処理するかで、2通りの方法が用意されている。計算に当たって、どちらの方法を採用するかは研究者が決めるべき事柄であるが、可能な限り前者の方法をお勧めする。制作者としては、SPBS 統計パッケージの解析プログラムで自動的に「-999」や「-99」を除外して解析しているので、このプログラムが必要かどうか疑問視している。

欠損サンプルの除去：カーソルのある当該変数内の欠損値を含むサンプルデータを削除するプログラムである。このため、この操作を実行したあとでは、データファイルのサンプル数が入力時に比べ、減少する。

集団平均値の挿入：欠損値以外のデータで計算した集団平均値を、欠損値があるサンプルに挿入するプログラムである。

変数名変更

変数の内容が元の内容と変わった時に、ここで修正することができる。変数名は全角 17 文字、半角 34 文字まで入力可能であり、それ以上は自動的にカットされる。

基本統計量

カーソルの置いてある変数の平均値、標準偏差、最小値、最大値（欠損値がある場合には、その欠損値を除いたデータで算出）、サンプル数が表示される。「サンプル数」の前に「*」印がついている場合は、欠損値があることを示す。

全体の統計量

データファイルにある全ての変数の平均値、標準偏差、最小値、最大値（欠損値がある場合には、その欠損値を除いたデータで算出）が算出される。また、サンプル数が5000以下の時には、さらに中央値も算出される。このデータは「印刷」ボタンをクリックすると、プリンターに印刷される。なお、サンプル数が2未満の場合には算出されない。

Basal Statistics		sample number= 5098				
V-No.	Mean	S.D.	Min.	Max.	Defect	Variable name
1	3.143	2.279	1.000	10.000	0	事業所
2	42.260	11.940	17.000	76.000	0	健診時年齢
3	1.137	0.343	1.000	2.000	0	性別
4	167.043	7.084	141.000	191.600	8	身長
5	64.039	10.296	34.300	122.100	4	体重
6	94.866	21.085	51.000	424.000	9	空腹時血糖
7	22.886	2.977	13.900	40.600	10	B M I

Defect は欠損値の数。Mean、SD、Min、Max は欠損値を除外した値

カーソルの移動

カーソルの移動をおこなう。ここでは「**<**」、「**>**」、「**↑**」、「**↓**」のそれぞれの方向に1頁分移動する。また、「任意の移動」では、左右前後に最大999個移動することができる。最初に移動したい1~999の数値を入力し、その後矢印キーを押す（矢印キーは自分の移動したい方向を押す）。

このほか、画面の数値データの周辺にあるサンプル番号、変数番号が書かれているところをクリックすると、所定の数移動することができる。画面の「サンプル」と書かれているところをクリックすると、左上のセルが変数1、サンプル1のデータがくるように移動する。サンプル数の最下部（起動時の場合、サンプル19）をクリックすると、データファイルの最終サンプル番号のデータが一番下になるように移動する。また、変数の最右端（起動時の場合、変数9）をクリックすると、データファイルの最終変数番号のデータが右端になるように移動する。これらの移動の仕方は、メインメニューの「ヘルプ」をクリックした際の「データファイルの作成と修正」のところに詳述されている。

データ作表画面の右（下）端の変数（サンプル）番号をマウスでクリックすると、最後部の変数（サンプル）番号のデータが画面の右（下）端に移動する。同様に、左（上）端の変数（サンプル）番号をマウスでクリックすると、1番目の変数（サンプル）データが画面の左（上）端に移動する。

同様に、データ作表画面の上述の位置より1~2個中央寄りをマウスでクリックすると、矢印キー1回分または5（縦向きは10）回分の操作に相当する（画面の上端2、3番目の変数番号は、同6、7番目は、左端2、3番目のサンプル番号は、同17、18番目はに移動）。

(8) 連続量データの離散量データへの変換

既にあるデータファイル（連続量データ）を離散量データ（0, 1, 2, …, 8）に変換し、「離散量データによる患者対照研究」の解析用データファイルを作成することができる。すなわち、連続量数値を幾つかの区間（カテゴリ）に分け、個々の数値に対し区間番号（カテゴリ番号）を変換する。この場合、カテゴリ数は最大9個で、必ず0から始まる。また、各カテゴリの境界値は各区間の上限値となる。

最初データファイルを入れ、次に変換する変数番号を指定する。カテゴリ数の指定の後、各カテゴリの境界値（上限値）を「必ず、小さい方から」入力する（境界値の数は[カテゴリ数]-1）。例えば、データファイルに1日当りの喫煙本数が入っている場合、非喫煙と喫煙に分類するためにはカテゴリ数を2、その上限値を0とすれば、非喫煙者に[0]が、喫煙者に[1]が割り当てられる。なお、データファイルの保存の際に、最初と同じファイル名を書くと元のファイルは消失する。

データファイルが1つの場合には、対照群と患者群に分ける数値は **0=「対照群」、1=「患者群」** のように、対照群が患者群より小さい数値なるように割り振る必要がある。

2. 基本統計量

データファイル内における特定の 변수番号の平均値、標準偏差、最小値、最大値、中央値等を算出し、かつ度数分布を表示する。

データの <input checked="" type="radio"/> 入力方法	<input checked="" type="radio"/> Datafile	<input type="radio"/> Keyboard (平均値、SDの算出)	実行	終了
区間の <input checked="" type="radio"/> 設定	<input type="radio"/> 自分で設定	<input checked="" type="radio"/> 平均、SDより自動設定		

区間設定は「自動の場合」と「自分で設定する場合」の2通りあり、設定区間は10段階となっている(固定)。「自動の場合」には、各境界値が以下のように自動的に算出される。

第1段階	平均 - 2.0 × 標準偏差	第2段階	平均 - 1.5 × 標準偏差
第3段階	平均 - 1.0 × 標準偏差	第4段階	平均 - 0.5 × 標準偏差
第5段階	平均値	第6段階	平均 + 0.5 × 標準偏差
第7段階	平均 + 1.0 × 標準偏差	第8段階	平均 + 1.5 × 標準偏差
第9段階	平均 + 2.0 × 標準偏差		

自分で境界値を定める場合、境界値(全部で9)を必ず順次小さい方から入力する必要がある。大きい数値から入れた時、あるいは同じ値を入れた時には計算できない。

境界値は各区分の上限値となる。例えば、1日当りの煙草本数の度数分布を表す場合に第1段階の境界値に0を入力すると、煙草を吸わない人の数が第1区分に示される。

サンプル数が20以上の時には、正規分布の適合度検定(Hosmer-Lemeshow 検定)をおこなう(正規分布確率から各期待度数を算出し、それと観測度数のずれを χ^2 検定する)。検定値が有意($P < 0.05$)な時は正規分布近似が否定される(この場合には、標準偏差、最小値、最大値の代わりに中央値、25パーセンタイル値、75パーセンタイル値がディスプレイに表示される)。サンプル数が25個以上の場合、歪度および尖度の検定(Fisherの補正值)も同時におこなう。

○変数番号 19 の度数分布表 [ヘモグロビン] A:¥QTC97

平均値	15.674	標準偏差	1.077	サンプル数 =	385
最小値	10.500	最大値	18.600		
5%タイル	14.100	95%タイル	17.300		
25%タイル	15.100	75%タイル	16.400	中央値 =	15.6
適合度検定	2値 =	13.391	P=0.099		
歪度(skewness)の検定	歪度 =	-0.6062	P<0.001		
尖度(kurtosis)の検定	尖度 =	2.3987	P<0.001		

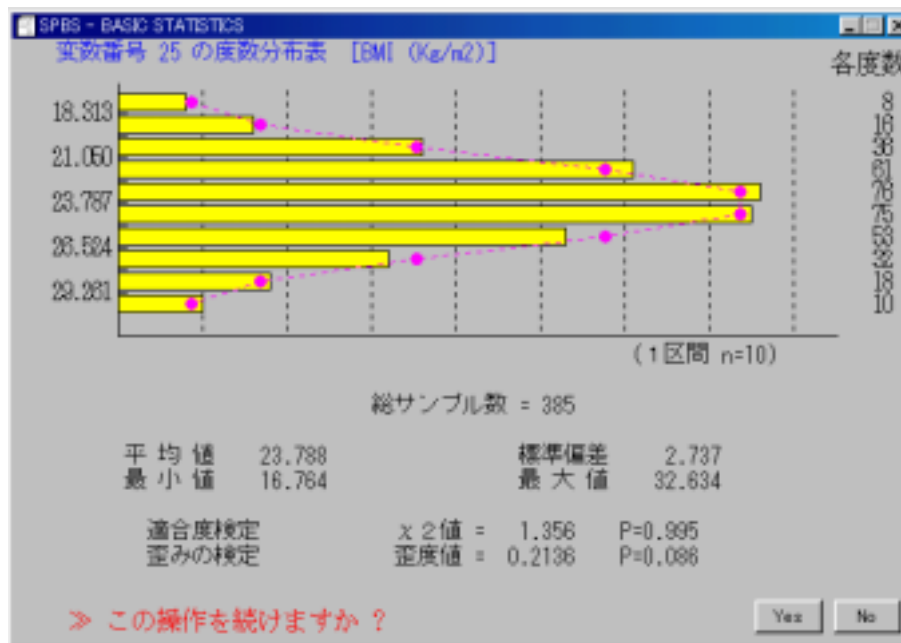
区間	1	2	3	4	5	6	7	8	9	10
度数	9	8	32	52	92	81	50	40	14	7
割合	(2.3)	(2.1)	(8.3)	(13.5)	(23.9)	(21.0)	(13.0)	(10.4)	(3.6)	(1.8)

画面に「一括して処理しますか?」と現れた時に<Yes>ボタンをクリックすると、次画面で選択した全ての変数の基本統計を算出する。但し、変数名に[カテ]と付いている時はその変数をカテゴリー変数と見なし、上述の平均値、標準偏差、最小値、最大値等々を算出しないで、各カテゴリー(特定の数値)に含まれるサンプルの数を算出する。従って、カテゴリー変数でなく連続量変数と考えている変数では一括処理をしないで、1つ1つ変数を選択しておこなわねばならない。なお、このカテゴリー変数と見なす基準はある変数内に含まれる数値の種類(数種)が12以下の場合に限られる。

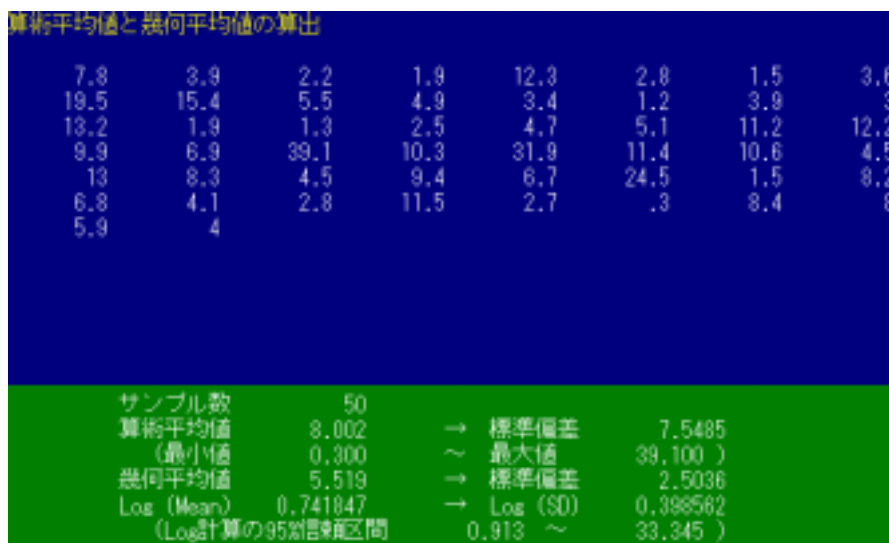
○変数番号 17 の度数表 [Whale-intake(P)]

数種	0	1	2	3	4	5	8	10
度数	34	33	41	15	24	1	1	1
(%)	22.67	22.00	27.33	10.00	16.00	0.67	0.67	0.67

総サンプル数 = 150



この他「キーボード入力」を選択すると、算術平均（標準偏差、最小値、最大値）や幾何平均（ \log 平均値の 10^{mean} 、 \log 標準偏差の 10^{SD} 、 $10^{(\text{mean} \pm 1.96 \times \text{SD})}$ ）を算出する。但し、「キーボード入力」はサンプル数が 2 ~ 88 個の時のみ使用可能であり、89 以上では使用できない。この解析結果は、下図のように画面に表示されるが、メインメニューに戻った時に「解析結果の表示」ボタンを押すと見る事ができる。

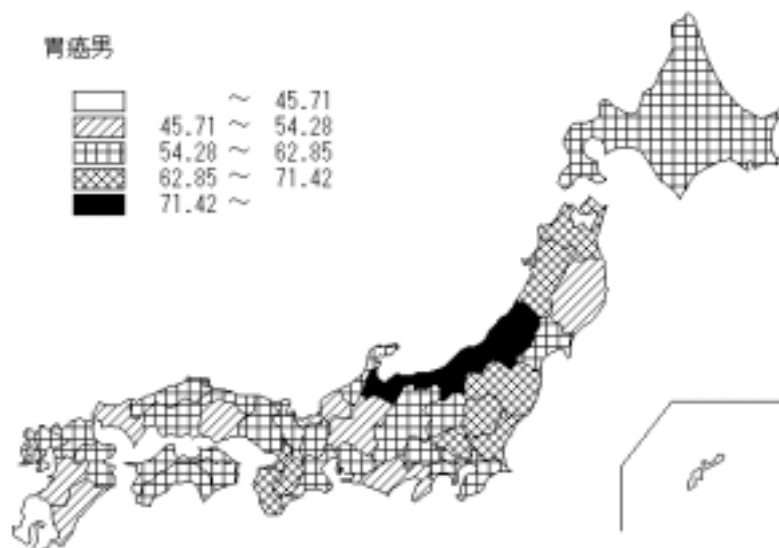


正規確率の Q - Q プロット (quantile-quantile plot) を表示したい場合には、区間の設定は「自動設定」であり、「一括処理」を選択しない。ここで描かれる図は、当該変数の値（順序統計量）を横軸とし、Weibull法による累積比率 $P_i (=i/(n+1))$ に対応する標準正規分布のパーセント点（正規スコア）を縦軸としている。同時に、この相関係数 r と「ズレ確率」（後者は、作成者が勝手につけた名前。「1 - 寄与率」で定義され、寄与率 (contribution ratio) は $r \times r$ ）を算出する。原則的に、この Q - Q プロットで直線性が高ければ正規分布近似と見なすことができるので、この「ズレ確率」が 0.05 以下であれば相当直線性が高いと考えて差し支えないと考える（ここで用いている適合度検定はサンプル数が 100 を超えると有意になりやすくなり、結果として「正規分布の適合度が悪い」となりやすい。同様に、歪度および尖度の検定もサンプル数の増加とともに有意となりやすい）。

データファイルからの入力の場合で、サンプル数が 46 または 47 の時に限りファイル入力後に「**日本地図を作図**しますか？」と尋ねてくる。もし、データが < 北海道、青森、岩手、宮城、秋田、山形、福島、栃木、群馬、埼玉、千葉、東京、神奈川、新潟、富山、石川、福井、山梨、長野、岐阜、静岡、愛知、三重、滋賀、京都、大阪、兵庫、奈良、和歌山、鳥取、島根、岡山、広島、山口、徳島、香川、愛媛、高知、福岡、佐賀、長崎、熊本、大分、宮崎、鹿児島、沖縄 > の順に並んでいれば、これらの県の地図に 5 色調（白色から黒色まで）の「塗りつぶし」をおこなうことが可能である。5 色調の塗りつぶし基準は、自分で 4 つの境界値（いずれも上限値）を入れる場合と自動で計算させる場合の 2

種の選択ができる。後者の基準（境界）値は白色から黒色向きに、平均値 - 1.5 標準偏差、平均値 - 0.5 標準偏差、平均値 + 0.5 標準偏差、平均値 + 1.5 標準偏差となる。また、最初に設定した基準値を別の変数に対しても使用することが可能である。このプログラムで描かれる図は自動的に「ペイント」に移動するので、図の加工は「ペイント」上でおこなうことになる。なお、46 または 47 都道府県データに欠損値（-99 または -999）がある場合には、都道府県データがズレてしまうので、日本地図を作図してはならない。解析結果では、平均値および標準偏差の他に、当該変数の値を小さい方から大きい方に並べ替え、県名と併せて表示する。

このプログラムでは「自分で4区間を入力しますか?」と表示される画面の右上隅に「M」の表示が現れる。この時（「Yes」または「No」を選択しないで）キーボードで「M」または「m」を入力すると、正規分布関数と当該変数の平均値および標準偏差を用いて、正規分布確率で P=0.2、P=0.4、P=0.6、P=0.8（P は Probability）に相当する値を自動的に算出する。そして、これらの境界値を用いて塗りつぶし作業をした日本地図を描く。前述の平均値 ± 1.5 標準偏差および平均値 ± 0.5 標準偏差の境界値に納得できない場合にはこれを試してみるのも一法であろう。



変数 23 胃癌男

平均値 58.562 標準偏差 8.570 (最小値 28.8 ~ 最大値 77.8)

順位替え後の県名とその値

沖 縄	28.8	神奈川	58.2
鹿児島	42.9	鳥 取	58.8
熊 本	45.9	島 根	58.8
岡 山	48.5	千 葉	58.9
岩 手	49.6	東 京	59.3
静 岡	49.8	宮 城	59.5
福 井	51.0	長 野	59.5
宮 崎	52.4	滋 賀	59.9
岐 阜	53.4	石 川	60.0
山 口	53.9	群 馬	62.0
大 分	54.9	長 崎	62.3
徳 島	55.1	埼 玉	63.7
広 島	55.6	茨 城	64.4
愛 知	55.6	大 阪	64.7
福 岡	55.7	青 森	66.0
三 重	56.2	栃 木	66.5
山 梨	56.5	奈 良	67.0
高 知	56.5	福 島	68.8
愛 媛	56.9	和歌山	68.9
京 都	57.1	秋 田	69.7
兵 庫	57.1	富 山	75.6
北 海 道	57.3	山 形	75.7
佐 賀	57.5	新 潟	77.8
香 川	58.2		

3. 有意差検定 (2 グル - プ間の検定)

2 グル - プ間の有意差検定をおこなう。この場合、正規分布するデータ、正規分布するかどうかわからないデータのいずれも取扱うことができる。また2つのグル - プが対応する場合、あるいは対応しない場合のいずれでも計算可能である。最初の画面で初期設定は「対応のない検定」、入力形式は「データファイル」、「データファイル数1」となっている。もし「対応のある検定」、入力形式が「キ - ボ - ド」あるいは「データファイル数2」の場合にはこれらを変更してからおこなう（なお、入力形式がデータファイルの場合データファイルのサンプル数、変数に対する制限はないが、「グル - プ間の有意差検定」では変数の数に制限がある）。

1つのデータファイルで有意差検定をおこなう場合は、データファイル内の異なる2変数を選択すれば計算を実行する（「対応のある検定」、「対応のない検定」のいずれも可能）。1つのファイルのある特定変数で2群に分けて解析することができるが、これは「対応のない検定」の場合に限られる。

2つのデータファイルで有意差検定をおこなう場合、両データファイルにおける各任意の変数間の有意差検定を実行する。「第1群の変数番号は」と最初尋ねてくるがこれは第1群のファイルを指し、「第2群の変数番号は」は第2群のファイルを指す。変数を選択する際、矢印キーか数字キーで変数番号を指定し、「終了」ボタンをクリックすると解析を実行する。なお、「対応のある検定」を指定しているにもかかわらず、2つのデータファイルのサンプル数が異なっている時には、自動的に「対応のない検定」に変更される。

(1) 正規分布する場合

各変数間のデータが正規分布する場合に使用する。「対応のない検定」では、最初に分散が等しいかどうかを両側F検定する。次に平均値の差の検定をおこなう。平均値の差の検定では、分散が等しい場合と異なる場合の2通りの結果が示される。従って、F検定の結果を踏まえて最終的な解釈をおこなう必要がある。なお、分散が等しい場合にはStudentのt検定を、異なる場合にはWelchの検定を用いる。

対応のあるt検定は、このプログラムの最初の画面で選択された場合に実行される。差の検定をおこなう両変数のサンプルが対応のあるデータの場合にはこの方法を用いる（逆に偶然サンプル数が一致しているだけの場合は「対応のあるt検定」を使用しない！ 例え、動物実験の対照群と薬物投与群の有意差検定の場合、1対1に最初からペアが組まれている場合を除き、各々同数であっても「対応のないt検定」になる）。

年齢のように、原因因子と結果因子の両者に影響を及ぼす可能性のある因子を交絡因子（共変量）と呼ぶ。交絡因子の影響を除外して当該変数の有意差を比較する方法を共分散分析と呼ぶ。このプログラムでは交絡変数を1個のみ用いて解析することができる。交絡変数が2以上の時は「共分散分析」を使用する。交絡因子を指定すると多変量解析扱いになるので、欠損値（例えば-99、-999）があれば自動的に削除され、当初のサンプル数と異なってくる。

欠損値の処理は、「対応のない検定」の場合には個々の変数内で削除されるが、「対応のある検定」の場合には2つの変数をペアとみなして、いずれかに欠損値があれば、ペア単位で削除する。

第1群の変数 (9): FBS G:%hyo01.txt
第2群の変数 (9): FBS G:%hyo01.txt

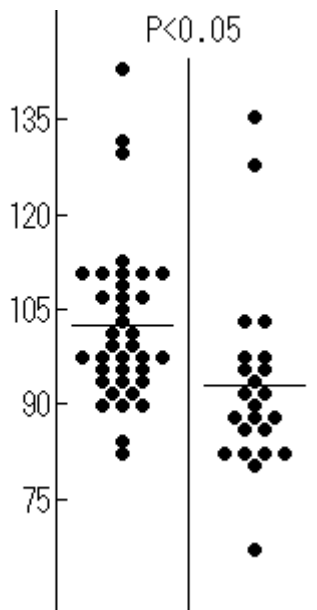
2群に分割した変数は 11 番です

	1 グル-プ	2 グル-プ
平均	102.3243	92.8261
標準偏差	12.6492	14.7298
サンプル	37	23

[対応のないt検定]

F(22 , 36)= 1.35603 P=0.408 (両側検定)--> 「等分散の場合」を見よ！

	等分散 (Student)	不等分散 (Welch)
t 値	2.654	2.561
自由度	58	41
確 率	P=0.010	P=0.014
* * 危険率5%で有意差があります * *		



左図の縦軸を変更する事ができる。「図をプリントしますか？」と表示された時に、キーボードの「m」または「M」を入力すると、画面上で最大値と最小値を尋ねてくる。表示される最大値より小さい値、あるいは表示される最小値より大きい値を入れると無視される。この操作をおこなう場合、プリンターには印刷することができず、「ペイント」画面に図が現れる。

(2) 正規分布しない場合

データが正規分布しない場合には、Wilcoxon 検定をおこなう。この解析の検定表はプログラム中には含まれていないので、参考文献の検定表を調べる必要がある (Glantz, 1987)。この結果は「正規分布する場合」の検定結果の後に、算出される。

対応のない検定を選択した場合には、Wilcoxon の順位和検定と Median 検定をおこなう。また、対応のある検定を選択した場合には、Wilcoxon の符号付順位検定をおこなう (後者の方法では、ペアのサンプル数が 6 以上、25 以下の場合に限り Wilcoxon 検定表の結果が出てくる。また、正規分布近似による確率 P 値を算出する)。

Wilcoxon の順位和検定では順位和を算出するが、検定表をプログラムが持っていないので統計の本で有意性を調べる必要がある。この順位和から有意性の検定をおこなう方法として「Mann-Whitney の

U 検定」があり、この解析結果が算出される。

Median (中央値) 検定は、2 群の全データの中央値を基準として各群でそれぞれ大きい数、小さい数の個数を算出し、 2×2 分割表を作成した後で χ^2 検定する (中央値と同じ値はこの分割表から削除される)。概して検出力は低い、分布を意識する必要もなく、直感的に差があるかどうかを知ることができる。

[Median Test]

中央値は 96.500 です

	1 グループ	2 グループ
Median <	24	6
Median >	13	17

χ^2 値 (Yates) = 7.05 P=0.008

[Wilcoxon の順位和検定]

[Group 1] の順位和 : 1355.0
 [Group 2] の順位和 : 475.0

Mann-Whitney U 検定 Z = 3.446 P<0.001

(3) データの対応の有無

使用するデータが 2 群間で対応のある場合とそうでない場合で解析法が異なってくる。対応のある比較をする場合には、このプログラムを起動した際に対応のある検定を選択する (初期設定は「対応のない検定」になっている)。

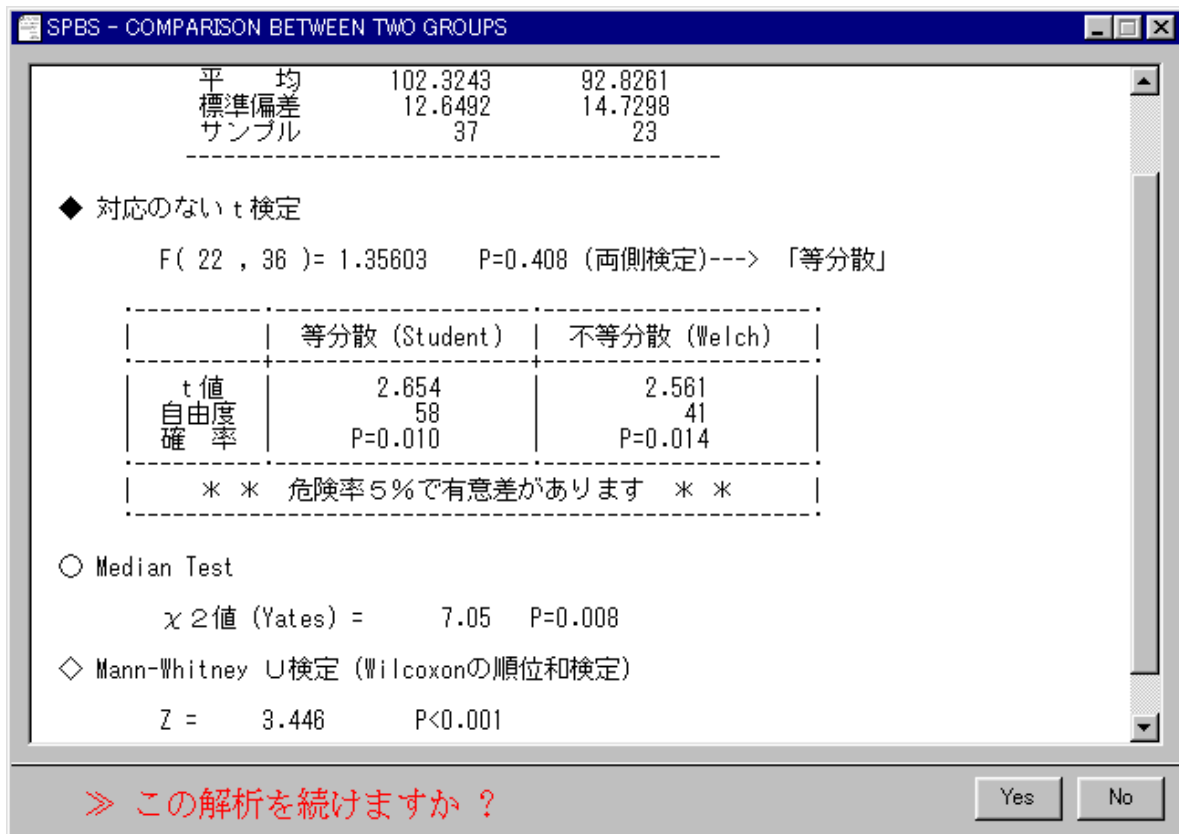
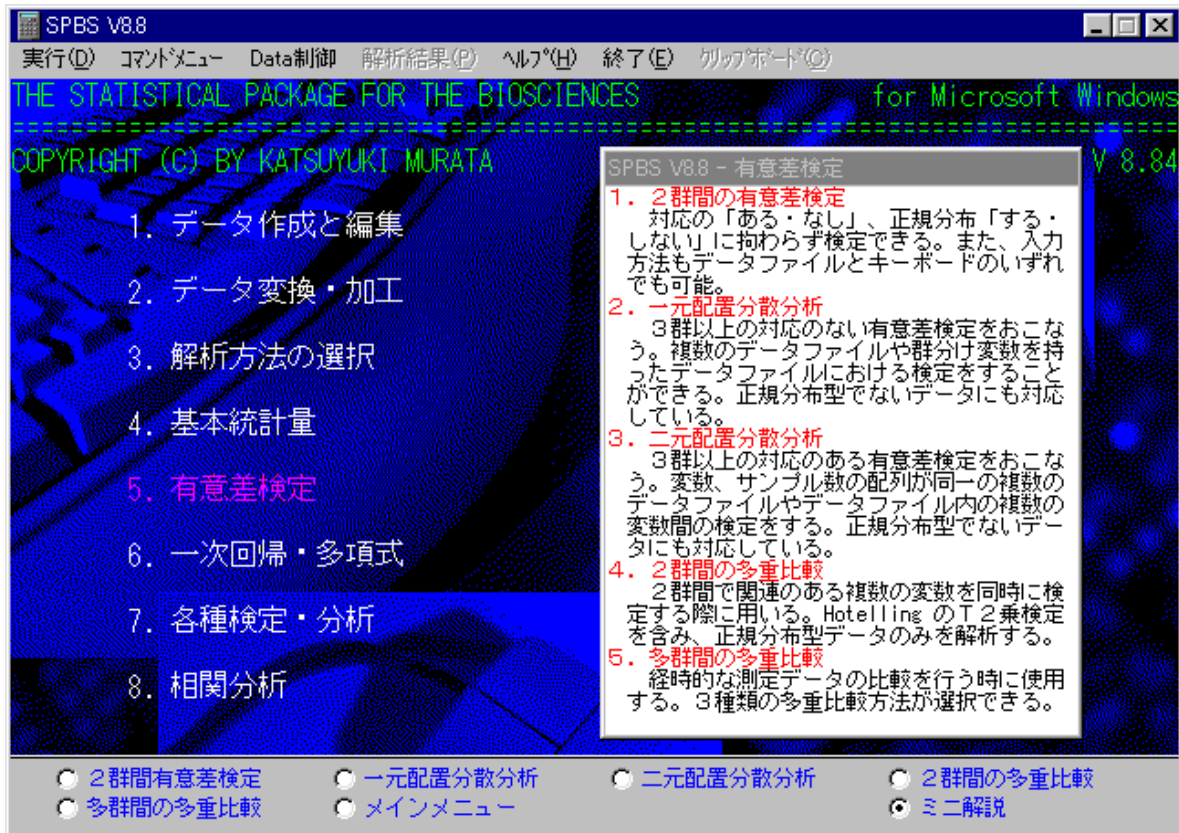
解析方法	<input checked="" type="radio"/> 対応のない検定	<input type="radio"/> 対応のある検定
データの入力法	<input type="radio"/> Keyboard	<input checked="" type="radio"/> Data file
データファイルの数	<input checked="" type="radio"/> One datafile	<input type="radio"/> Two datafiles
		実行 終了

(4) 平均値、標準偏差、サンプル数が既知の有意差検定

医学論文等には、図表の中に既にサンプル数、平均値、標準偏差が記述されているものがある。これらのデータの解析結果に疑問があるような場合、所謂「電卓」感覚でこのプログラムを使用することができる。但し、論文ではデータが桁落ちしている可能性があるため、あくまで検算のつもりで利用した方がよい。

なお、メインメニューの「5. 有意差検定」をダブルクリック (または右クリック) すると、下図のように 5 つの選

択肢が画面に現れる。この中のどれを選択すれば良いのか困ったときには「ミニ解説」のチェックボタンをクリックすると、簡単な説明が現れる。この「ミニ解説」は再度チェックボタンをクリックするか、解析プログラムまたはメインメニューを選択すると消える。この種の「ミニ解説」はこのほか「7. 各種検定・分析」、「8. 相関分析」で現れるようになっている（このミニ解説画面はメインメニューへの貼り付けでないので、移動させることができる）。



4. 一元配置分散分析

サンプル数は異なるが、変数の数および変数の内容が同じであるデータファイルを用いて、それらの各変数の有意差検定をおこなう時に使用する。この場合最大 12 個の変数を取扱うことができ、それら全ての組合せの有意差検定（等分散の場合には Student t-test を、また等分散でない場合には Welch t-test を使用）をおこなう。更に、等分散の仮定下で、一元配置分散分析の結果が有意でかつグループ数が 3 以上の時には Scheffe の多重比較も同時におこなう。

変数の配列が全く同じデータファイル（サンプル数は異なっても可）を最大 12 個まで読み込んで上の解析をおこなうことができる。

データファイルが 1 つの場合には、ある特定の変数の値を基準に 2 ~ 12 個のサブグループを作成し、その分類に使用した以外の変数のデータで有意差検定をおこなう、複数（12 個以下）の変数間の有意差検定をおこなうことができる（後者の場合、解析前に欠損値が何か尋ねる）。前者の場合、例えば年齢が変数番号 1 に入っているとすると、この変数で区切るべき上限値を入力し、層別化データを作成する。その後の操作は複数のデータファイルで解析をおこなう時と全く同様におこなう。また、キーボードからデータを直接入力して解析をおこなうこともできる。なお、これらのどの方法をおこなうかは、この解析プログラムをメインメニューで選択した直後に現れる画面（下図）で決定する。



一元配置分散分析の結果が有意でかつグループ数が 3 個以上（但し、最小サンプル数が 10 以上）の時には、箱ヒゲ図を SAS 準拠で描く。

グループ間の検定を全変数で解析する時は「全変数について順次計算する」を選択する（特定変数のみ解析したい時は「n」を入力する）。Bartlett の等分散性の検定で $P > 0.05$ の時は等分散と考え、一元配置分散分析の結果を参照する。また、Bartlett 検定で $P < 0.05$ の時は Kruskal-Wallis rank-sum test の結果を参照する。また、基本統計で示されたグループの平均値が増加傾向あるいは減少傾向を示す時には、傾向性検定（累積²法による Trend Test）の結果も出力する（広津千尋、1992）。

なお、*、**、***（または ¥）は各々 $p < 0.05$ 、 $p < 0.01$ 、 $p < 0.001$ を示す。「n.s.」は not significant（有意でない）の省略形である。

変数番号： 20 臍帯血中のセレン濃度

基本統計

グループ No	平均	標準偏差	(最小値 - 最大値)		サンプル数
1	112.550	18.816	75.200	199.600	148
2	112.738	16.194	75.400	175.400	99
3	117.525	15.966	87.600	159.800	55
4	115.650	16.552	87.100	149.600	32
5	122.793	21.900	83.700	168.700	41

Bartlett の等分散性の検定

検定統計量 B = 7.945 P=0.094 (自由度 4) 等分散

分散分析表 (一元配置分散分析)

	平方和	自由度	不偏分散	F 値
グループ	4218.11	4	1054.53	3.274 (P=0.012)
残差	119188.94	370	322.13	
全体	123407.05	374		

傾向性検定 (Trend Test)

累積²法

F = 7.392 (df1= 2.113 df2= 370) P<0.05

分散の検定 (: 検定法 : F-値)

	1	2	3	4	5
1		1.350	1.389	1.292	1.355
2	S		1.029	1.045	1.829*
3	S	S		1.075	1.881*
4	S	S	S		1.751
5	S	W	W	S	

*は両側F検定(5%)で有意であることを示す

Sは「等分散の場合(Student t-test)」、Wは「不等分散の場合(Welch test)」を示す

有意差検定 [: 危険率(P) : t-値]

	1	2	3	4	5
1		-0.081	-1.741	-0.862	-2.974
2	0.935		-1.767	-0.879	-2.654
3	0.083	0.079		0.521	-1.303
4	0.390	0.381	0.604		-1.534
5	0.003	0.010	0.197	0.130	

Scheffe の多重比較 [平均値の差 μ の信頼区間と有意性]

グループ間	95%信頼区間		99%信頼区間		有意性
1 - 2	-7.403	μ	7.026	-8.744	μ 8.368 n.s.
1 - 3	-13.750	μ	3.799	-15.382	μ 5.431 n.s.
1 - 4	-13.932	μ	7.732	-15.947	μ 9.747 n.s.
1 - 5	-20.049	μ	-0.436	-21.873	μ 1.387 P<0.05
2 - 3	-14.132	μ	4.557	-15.869	μ 6.295 n.s.
2 - 4	-14.211	μ	8.387	-16.312	μ 10.489 n.s.
2 - 5	-20.374	μ	0.265	-22.293	μ 2.184 n.s.
3 - 4	-10.478	μ	14.229	-12.776	μ 16.527 n.s.
3 - 5	-16.732	μ	6.197	-18.864	μ 8.329 n.s.
4 - 5	-20.249	μ	5.964	-22.687	μ 8.401 n.s.

Kruskal-Wallis rank-sum test

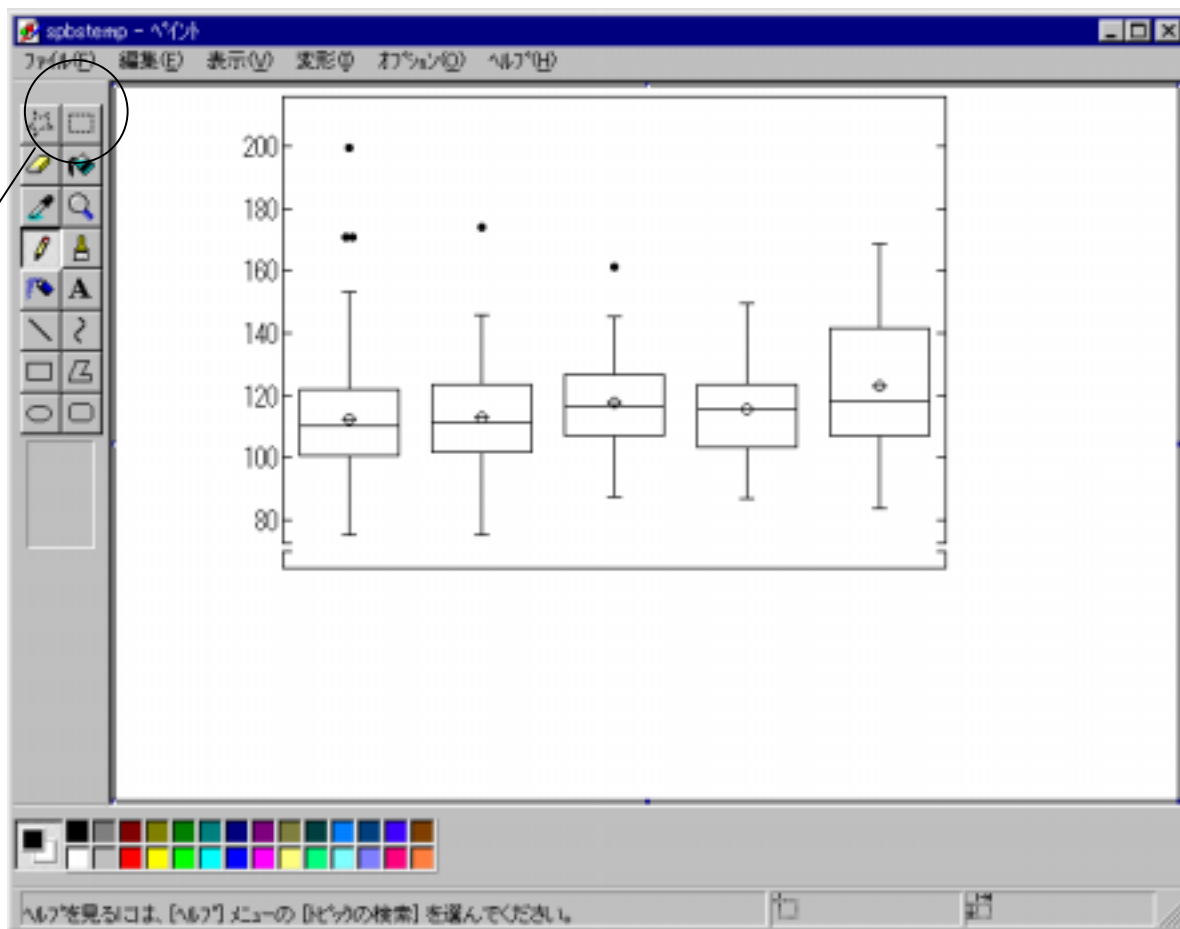
H値 = 10.736 P=0.030 (自由度 = 4)

Dunn の全順位法による多重比較

Group 間	差の統計量 H_{ij}	有意性
1 - 2	-0.500	n.s.
1 - 3	-2.140	n.s.
1 - 4	-1.050	n.s.
1 - 5	-2.770	n.s.
2 - 3	-1.624	n.s.
2 - 4	-0.687	n.s.
2 - 5	-2.283	n.s.
3 - 4	0.600	n.s.
3 - 5	-0.731	n.s.
4 - 5	-1.205	n.s.

注) Bonferroni の不等式による調整(正規分布近似)

下の図はペイントに解析結果を箱ヒゲ図で出力した例を示す。このペイント上で必要な図の部分を取り取って、ワープロソフトに貼り付けることが可能である。



「ペイント」のこのボタンをクリックすると描かれた図の範囲選択がおこなえる。ワープロソフトの画面に貼れる領域を考慮して、ここで範囲指定し、「編集」の「コピー」をクリックするとワープロソフト上で貼り付けが可能となる。なお、この図の縦軸はコンピュータが自動的に算出するので、2グループ間の比較（有意差検定）の図のように変更することはできない。

「一元配置分散分析」および「二元配置分散分析」（このほか「プロフィール分析」）では、各々のプログラムの起動画面で「H」または「h」キーを押すと、簡単な説明が画面に現れる（画面の「Help」ボタンでも同じ）。

5 . 二元配置分散分析

サンプル数、変数の数および変数の内容が同じで、かつ各サンプルが対応しているデ - タファイルを用いて、それらの各変数の有意差検定をおこなう時に使用する。この場合最大 12 個のデ - タファイルを取扱い、それら全部の組合せの検定をおこなう。

二元配置分散分析のグループ間変動が有意な時は Scheffe の多重比較の結果を参照する（サンプル間変動の有意性検定結果は算出されるが、どのサンプル間なのかの多重比較はおこなわない）。また、データが正規分布を示さないと考えられる場合は Friedman 検定をおこなう方がよい。

データ入力方法 Friedman 検定	<input type="radio"/> One datafile	<input type="radio"/> Some datafiles	<input checked="" type="radio"/> Keyboard	<input type="radio"/> Help	実行	終了
	<input checked="" type="checkbox"/> 実行					

同一デ - タファイル内における複数の変数（3 個以上、12 個まで）を用いて、それら全ての組み合わせによる有意差検定をおこなうこともできる（この場合には「グループ数」を 1 とする）。この場合、デ - タ入力後、同時に有意差検定をおこなう変数の数を入力する。

キーボードからデータ入力する場合（初期設定はキーボードによるデータ入力形式になっている）には、入れようとする変数（項目）の数がグループ数に相当する。この後、サンプル数を指定すればデータ入力ができる。なお、一元配置分散分析の時と同様に、サンプル数が 10 以上で比較するグループ数が 3 個以上の時には SAS 準拠の箱ヒゲ図を出力する。

Dunnett による多重比較の基準 t 値も同時に出力されることがある。この場合、複数のケース（あるいはコントロール）群と 1 つのコントロール（あるいはケース）群の組合せ数となる（Dunnett, 1955）。

- 新しい変数番号 1 飲酒前（収縮期血圧）
- 新しい変数番号 2 飲酒後 30 分
- 新しい変数番号 3 飲酒後 1 時間
- 新しい変数番号 4 飲酒後 2 時間

基本統計

グループ	平均	標準偏差	(最小値 - 最大値)	
1	124.000	11.408	102.000	142.000
2	128.125	12.252	102.000	146.000
3	127.875	12.447	100.000	140.000
4	124.875	12.241	104.000	140.000

サンプル数 = 16

分散分析表 (二元配置分散分析)

	平方和	自由度	不偏分散	F 値
変数間	209.6875	3	69.8958	3.747 (P=0.017)
サンプル間	7935.9375	15	529.0625	28.366 (P<0.001)
残差	839.3125	45	18.6514	
全体	8984.9375	63		

Scheffe の多重比較

変数間の比較 (処理数= 4 , サンプル数= 16)

5%水準差 = 4.434 1%水準差 = 5.452

グループ間	95%信頼区間	99%信頼区間	有意性
1 - 2	-8.559 μ 0.309	-9.577 μ 1.327	n.s.

1 - 3	-8.309	μ	0.559	-9.327	μ	1.577	n.s.
1 - 4	-5.309	μ	3.559	-6.327	μ	4.577	n.s.
2 - 3	-4.184	μ	4.684	-5.202	μ	5.702	n.s.
2 - 4	-1.184	μ	7.684	-2.202	μ	8.702	n.s.
3 - 4	-1.434	μ	7.434	-2.452	μ	8.452	n.s.

 μ : 2 変数間の平均値の差

* * * * t - 値 * * * * (Paired samples)				
	1	2	3	4
1	0.000			
2	-2.771	0.000		
3	-2.443	0.176	0.000	
4	-1.163	1.755	1.668	0.000

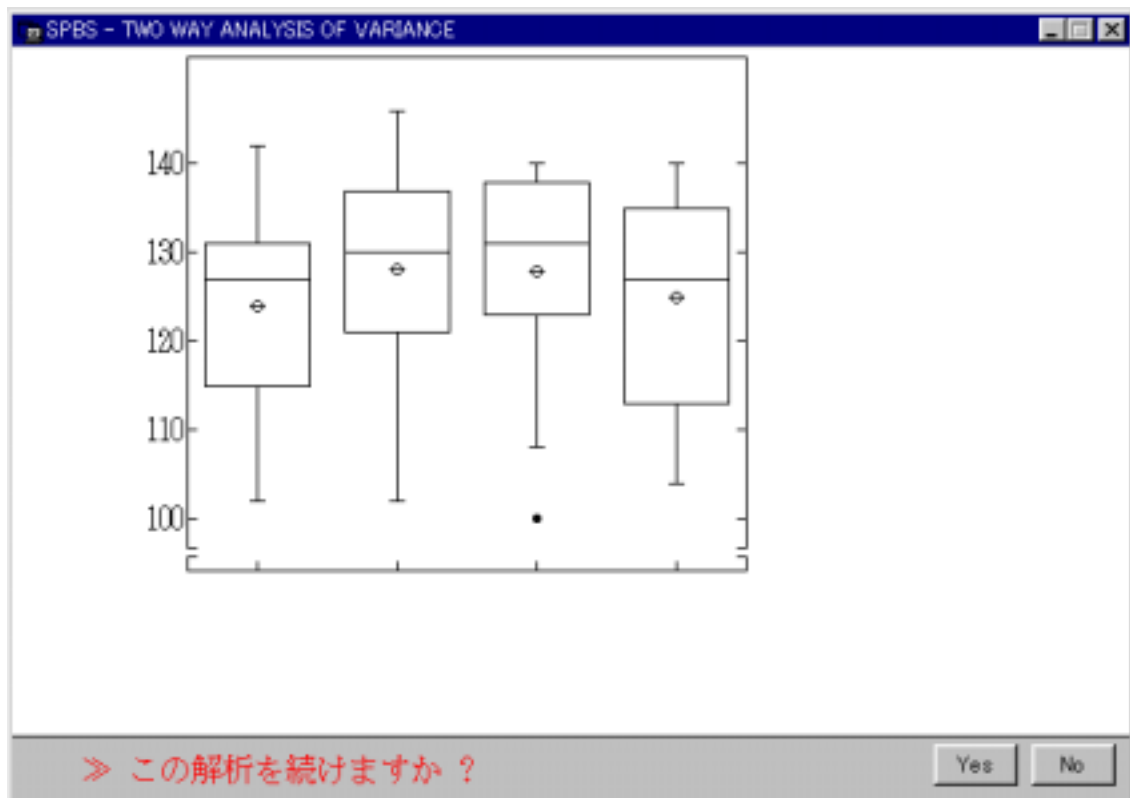
多重比較(Dunnett's multiple comparison)の場合の基準値

[組合せ数 3 個] $t(p<0.05)= 2.64$ $t(p<0.01)= 3.45$

* * * * 有意性 (確率 P) * * * * (単純比較の結果)				
	1	2	3	4
1	1.000			
2	0.014	1.000		
3	0.027	0.863	1.000	
4	0.263	0.100	0.116	1.000

グループ間変動の有意性検定 (Friedman 検定)

検定統計量 $Q = 4.97$ (自由度 3) $P=0.174$



6. 2群間の有意差検定（多重比較）

複数の変数からなる2群の平均値を比較する場合、それら変数が全く独立であれば個々の変数について「平均値の差の検定（t検定）」をおこなう。しかし、それら変数が相互に独立でない時は多重比較による統計学的検定を使用する。このプログラムでは多重比較による2群間の有意差検定をMorrisonの同時信頼区間（simultaneous confidence intervals）の概念に基づいて検定する。この理論はHotellingの T^2 統計量の算出から始まる。なお、Bonferroniの方法の時にはHotellingの T^2 統計量を算出しない。

Datafile の数	<input checked="" type="radio"/> One Datafile	<input type="radio"/> Two Datafiles	実行	終了
2群の分割方法	<input type="radio"/> 特定のサンプル番号	<input checked="" type="radio"/> 特定変数の数値		

最初に、2群間のデータが対応しているか否かを指定し、次に入力するデータファイル数が1個なのか2個なのか指定する。データファイルを2個入力する場合、2つのデータファイルの変数配列は同じでなければならないが、ファイルの変数の数は999個以下であれば制限がない。但し、計算に使用する変数の数はサンプル数によって制限される（対応のある場合には、最大使用変数の数は $< \text{サンプル数} - 1 >$ 。また、対応のない場合の最大使用変数は $< 2 \text{群の合計サンプル数} - 1 >$ となる。いずれの場合でも、サンプル数よりも多い変数を指定することはできない。

データの対応の有無には関係なくHotellingの T^2 統計量が最初に算出される（このHotellingの T^2 統計量の値は判別分析の際にも算出され、同じデータ・変数を使って解析すると、ここで算出された値と同じ値となる）。次に、解析に使用した変数の範囲で2群間に有意差があるかどうかを検定する。この後、2群間の個々の変数における平均値の差の同時信頼区間が算出される（この信頼区間に0を含む場合、有意差なし）。なお、Bonferroniの方がMorrisonの方法より緩い結果が得られる。また、「対応のあるデータの場合」には単純比較（対応のあるt検定）による有意差検定の結果も同時に算出される。これ以上の詳しい数学的解説（具体的な計算方法）はMorrison（1976）の原文を参照されたい。

変数番号	使用状況	変数名
16	未使用	空腹時血糖
17	未使用	白血球数
18	未使用	赤血球数
19	未使用	ヘモグロビン
20	未使用	ヘマトクリット
21	使用	心拍数
22	使用	RR Interval (sec)
23	未使用	QT interval
24	使用	QTc interval
25	未使用	BMI (Kg/m ²)
26	未使用	TG/HDLc ratio
27	未使用	2群に分ける変数のため使用禁止

Multiple Comparison for Two-Group Data

E:¥reidai¥QTC97

** 2群間の対応のない有意差検定（多重比較） **

使用する変数名
21 心拍数

22 RR Interval (sec)
24 QTc interval

2群の総和の分散共分散 (自由度 383)

	21	22	24
21	188.264		
22	-1.97097	0.0224459	
24	159.430	-1.72216	430.728

2群の分散共分散行列が等しいか否かの検定

λ 値 = 77.34 P<0.001

2群は等分散ではありません!

Bonferroni Method

同時信頼区間 (Simultaneous confidence intervals)

9 5 % 信頼区間			9 9 % 信頼区間			
21	-7.223	μ	-0.282	-8.015	μ	0.511
22	0.015	μ	0.091	0.006	μ	0.100
24	-13.382	μ	-2.884	-14.581	μ	-1.685

注) μ は両群の平均値の差を表す

多重比較の結果 (サンプル数: 第1群 145, 第2群 240)

	第1群 (平均値 ± 標準偏差)		第2群 (平均値 ± 標準偏差)		有意性
21	71.131	± 15.203	74.883	± 12.745	P<0.05
22	0.871	± 0.169	0.818	± 0.137	P<0.01
24	395.559	± 20.253	403.692	± 21.050	P<0.001

Morrison Method

Hotelling T² 統計量 22.466 F(3 , 381)= 7.449 P<0.001

両群には有意差がありません

同時信頼区間 (Simultaneous confidence intervals)

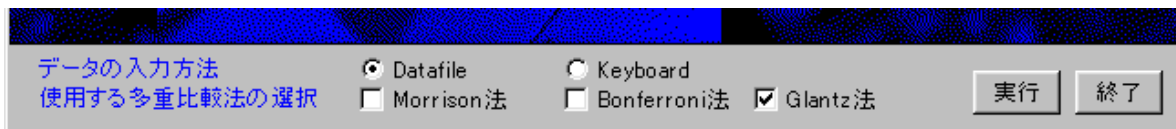
	9 5 % 信頼区間		9 9 % 信頼区間		有意性		
21	-7.815	μ	0.311	-8.659	μ	1.155	P=0.083
22	0.009	μ	0.097	-0.001	μ	0.107	P=0.011
24	-14.279	μ	-1.987	-15.555	μ	-0.711	P=0.004

注) μ は両群の平均値の差を表す

注) 有意性検定の自由度はHotellingの統計量の検定と同じ

7. 反復測定データの有意差検定 (多重比較)

同一集団に対し、ある項目の測定を3回以上反復実施した場合（例えば、複数回の投薬処置前後の測定）あるいはある物質濃度の日内変動を調べた場合、それら異なる条件下の平均値の変動が有意であったか否か、どの条件間で有意差があるのかどうかをこの解析プログラムで検討することができる。したがって、対応のある一元配置分散分析法と考えて差し支えない。また、変数間の有意差検定には多重比較（同時信頼区間、simultaneous confidence intervals）の概念を用いているので、普通の対応のある t 検定の結果より厳しい基準になる [同時推定 (simultaneous inference)] をおこなう場合に単一比較の基準（例えば、df=20, P=0.05 の t 値は t=2.086）を用いると、確率論的に実質的な有意水準が高くなる（すなわち、5%水準が実際にはそれ以上の水準で検定したことになる）。従って、この誤用を避けるために多重比較が使用される。



この解析のデータ入力形式はディスクまたはキーボードのいずれでも選択できるが、3変数以上でなければ多重比較ができない。また、サンプル数が多重比較する変数の数より3以上多くないと Morrisson の方法は使用できない（これ以外の方法はサンプル数による制限を受けない）。この repeated measurement の詳しい数学的な解説は Morrisson (1976) および Glantz (1987) 本に記述されている。

【 多重比較に使用する変数 】

基本統計

変数番号	平均	標準偏差	(最小値 - 最大値)	変数名
3	124.000	11.408	102.000 - 142.000	飲酒前 (収縮期血圧)
4	128.125	12.252	102.000 - 146.000	飲酒後30分
5	127.875	12.447	100.000 - 140.000	飲酒後1時間
6	124.875	12.241	104.000 - 140.000	飲酒後2時間

サンプル数 = 16

対照群との多重比較 (Glantz Method)

Analysis of Variance Table for Repeated Measurements

Source of variation	SS	df	MS
Treatments	209.68750	3	69.8958
Residual	839.31250	45	18.6514
F-value = 3.747 P=0.017			

各変数間の有意差検定 (多重比較)

変数間	平均値差 μ	t-value	有意性
3 - 4	-4.125	-2.702	n.s.
3 - 5	-3.875	-2.538	n.s.
3 - 6	-0.875	-0.573	n.s.
4 - 5	0.250	0.164	n.s.
4 - 6	3.250	2.128	n.s.
5 - 6	3.000	1.965	n.s.

基準 t 値: 2.760 (P<0.05) 3.345 (P<0.01) 4.108 (P<0.001)

変数間の有意差検定 (Bonferroni Method)

同時信頼区間 (Simultaneous confidence intervals)

変数間	9 5 %信頼区間		9 9 %信頼区間		有意性		
3 - 4	-8.644	μ	0.394	-9.814	μ	1.564	n.s.
3 - 5	-8.691	μ	0.941	-9.937	μ	2.187	n.s.
3 - 6	-3.159	μ	1.409	-3.749	μ	1.999	n.s.
4 - 5	-4.075	μ	4.575	-5.194	μ	5.694	n.s.
4 - 6	-2.373	μ	8.873	-3.828	μ	10.328	n.s.
5 - 6	-2.460	μ	8.460	-3.873	μ	9.873	n.s.

注) μ は 2 変数間の平均値の差を表す

変数間の有意性検定 (Morrison Method)

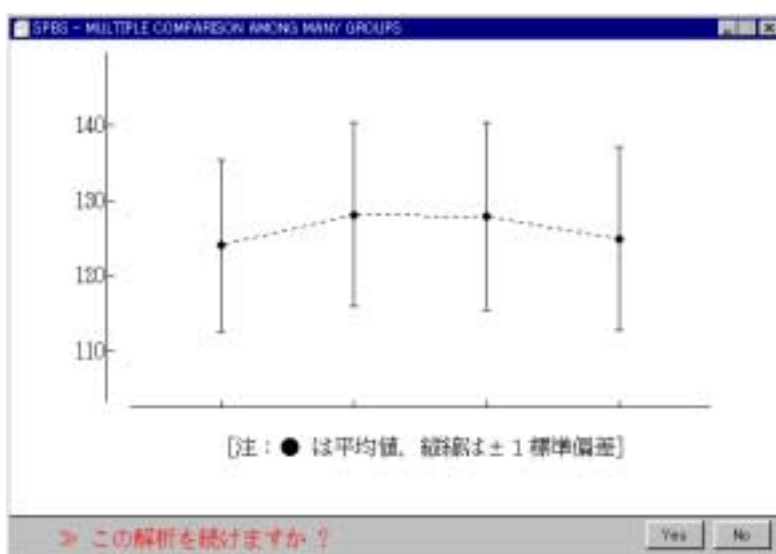
$$F(3, 13) = 3.615 \quad P=0.043$$

変数間には有意な変動があります

同時信頼区間 (Simultaneous confidence intervals)

変数間	9 5 %信頼区間		9 9 %信頼区間		有意性		
3 - 4	-9.239	μ	0.989	-10.760	μ	2.510	P=0.135
3 - 5	-9.325	μ	1.575	-10.945	μ	3.195	P=0.211
3 - 6	-3.459	μ	1.709	-4.227	μ	2.477	P=0.761
4 - 5	-4.644	μ	5.144	-6.099	μ	6.599	P=0.999
4 - 6	-3.113	μ	9.613	-5.004	μ	11.504	P=0.472
5 - 6	-3.178	μ	9.178	-5.015	μ	11.015	P=0.514

注) μ は 2 変数間の平均値の差を表す



反復測定データの有意差検定(多重比較)では、残念ながら一元配置分散分析、二元配置分散分析の時の箱ヒゲ図は描かれない。左の図のように各群の平均値±標準偏差で表示される。

	第 1 群		第 2 群		総 和 群	
	独立変数 X	従属変数 Y	独立変数 X	従属変数 Y	独立変数 X	従属変数 Y
平 均	5.936	1.326	6.588	2.026	6.262	1.676
標準偏差	0.516	0.198	0.636	0.275	0.663	0.425
(最小)	4.900	1.000	4.900	1.400	4.900	1.000
(最大)	7.000	1.800	7.900	2.500	7.900	2.500

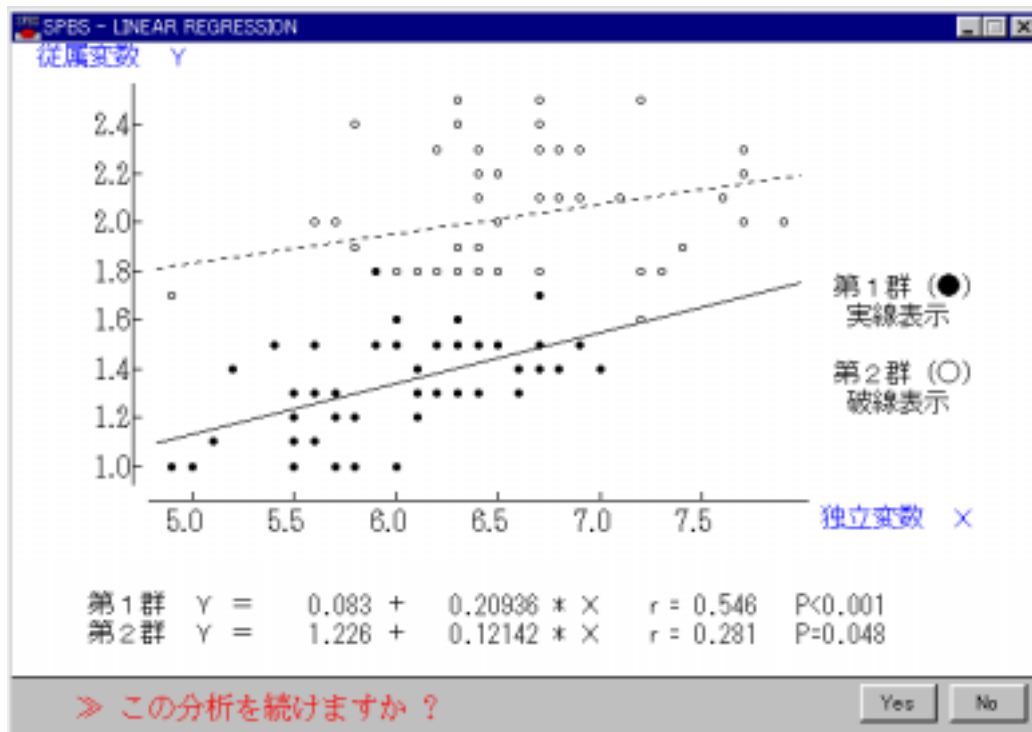
回 帰 残 差

	自由度	Xの総和	XYの総和	Yの総和	b値(傾き)	平方和(残差)
第1の群	49	13.055	2.733	1.916	0.209357	1.3440
第2の群	49	19.813	2.406	3.696	0.121416	3.4041
両群の和	98	32.868	5.139	5.612		4.7481
Pooled	98	32.868	5.139	5.612	0.156347	4.8090

分 散 分 析 表

	平方和	自由度	平均平方	F 値
b	0.0609	1	0.0609	1.231 P=0.270
群毎の残差の和	4.7481	96	0.0495	
全体	4.8090	97		

両直線の勾配は有意な相違がありません。



一次回帰式で独立変数の数値変換（対数またはルート変換）をおこなった場合、その横（X）軸を 変換後の値、または 元（変換前）の数値のいずれか選択した後に描く。では回帰直線となるが、では回帰曲線となる。

9. 多項式の適合度検定

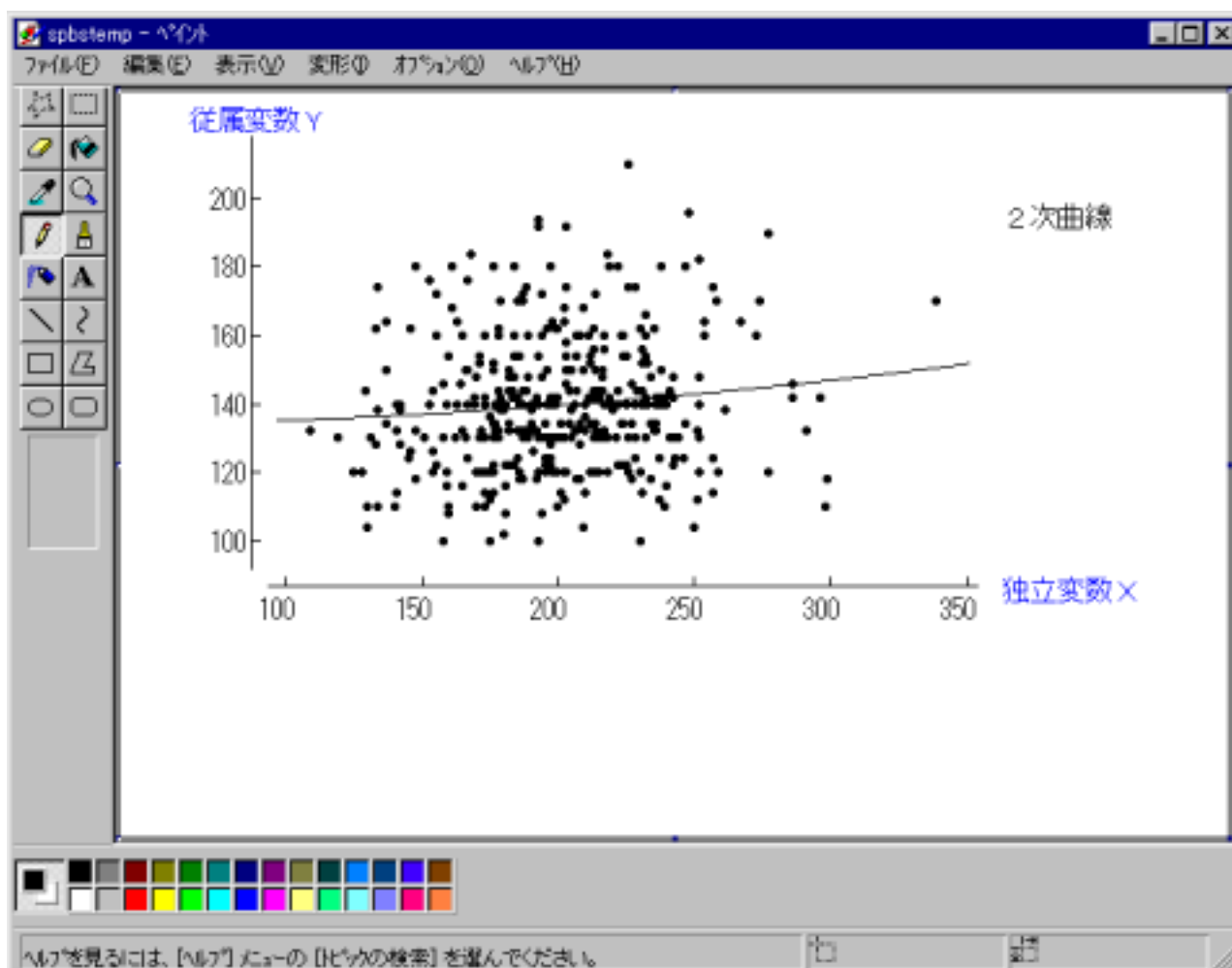
多項式（但し、5次まで）のあてはめが何次まで有効であるか重回帰分析を用いて解析する。このプログラムの場合、重回帰分析の増減法を用いておこなうので有意（ $P < 0.05$ ）な項（次数）のみが最終多項式に選択される。

多項式とは：
$$Y = B_0 + B_1 X + \dots + B_i X^i + \dots + B_n X^n \quad (B_i \text{は偏回帰係数})$$

最初に多項式のY（従属変数）とX（独立変数）を指定する。するとXの1乗値から5乗値までおよびYの総当たりの相関係数を算出し、その後重回帰分析を開始する。有効次数は最後ステップの最大項の値である。但し、数理統計学的にどの項とも有意でない、すなわち有意な多項式が見つからない場合もある。

次数の決定	<input checked="" type="radio"/> 逐次、次数変換	<input type="radio"/> 有意な次数固定	実行	終了
残差の算出	<input type="radio"/> 印刷可	<input checked="" type="radio"/> 印刷不可		

有意な多項式が算出された場合、X - Y平面にデータをプロットすることができる。また、観測値、予測値および残差を必要に応じて算出することができる。この解析結果の解釈にあたっては文献等を参考にして熟慮する必要がある。すなわち、「何故一回帰式で説明ができないで2次、・・・、5次式になったのか」に対する生物学的な理由付けが必要となるからである。数理統計学的なモデルを作るまでは容易であるが、合理的な説明のないモデルは全く意味を持たないので、誤用しないよう注意しなければならない。



10. 比率の検定

(1) 対応のない 2 × 2 分割表の検定 (Fisher の直接確率と χ^2 検定)

比率の検定には通常 χ^2 検定が用いられる (離散量の検定)。2 群のある 2 特性 (例えば、薬の効果の有無、煙草を吸う吸わない、など) に対してこの「出現頻度が両群で変わらない (帰無仮説)」かどうかを検定する。通常、Yates の修正値による結果を使用する。

Fisher の直接確率は正規近似の方法に基づく検定法の 1 つである。2 × 2 分割表のデータがあるとき、その出現度数の分布が得られる確率 (P) が求められる。4 分割表の中の数値がいずれも 5 以上の場合には普通 χ^2 検定を用いる。したがって、分割表の中に 4 以下の数値を含む時、Fisher の直接確率を用いる (通常、両側検定の結果を使用)。なお、数値が大き過ぎると計算できないことがある (全観察度数が 40 以下の時にはこの方法を用いた方が良いと考えられている)。

このほか、Odds 比と相対危険 (relative risk) も同時に算出する。相対危険は入力方向により算出結果が大いに異なるので、注意して入力する必要がある。

対応のない 2 × 2 分割表の検定 (Fisher 直接確率と χ^2 乗検定)

X1= 383 X2= 57
X3= 107 X4= 39

直接確率 P = 0.0001453 (片側検定)
直接確率 P = 0.0002538 (両側検定)

χ^2 = 15.1472 P<0.001
 χ^2 (Yates) = 14.1595 P<0.001

Case-control study の場合

Odds 比 = 2.45 [95%信頼区間 1.55 ~ 3.88]

Cohort study の場合

相対危険 (D+/-) = 1.86 [95%信頼区間 1.39 ~ 2.50]
相対危険 (E+/-) = 2.06 [95%信頼区間 1.44 ~ 2.96]

(2) 対応のある χ^2 検定 (McNemar 検定)

対応のある 2 × 2 分割表の検定には、McNemar 検定を用いる。これは、対応のある患者 - 対照研究のように、患者と対照のデータがペアになっている場合の離散 (比率) データの解析である。この時、ある要因がある場合を (+)、無い場合を (-) として表すと、各々のペアは (+,+), (-,+), (+,-), (-,-) のいずれかになる。したがって、このプログラムを使用する場合には、そのペア数が既に算出されていることが前提となる。このペア数を入力すると χ^2 値と有意水準を算出する。また、McNemar の直接確率値も同時に算出する。

▼ McNemar検定 (対応のある χ^2 乗検定)

		対 照 群	
		(-)	(+)
患 者 群	(-)	12 χ^2_{21}	
	(+)	23 χ^2_{22}	
		総ペア数	

McNemar 検定は、対応のある患者・対照群間において、要因 A が両群で有意に異なっているかどうかを検討 (χ^2 乗検定) する手法です

McNemar 検定 (対応のある χ^2 乗検定)

X11= 8 X21= 2
 X12= 6 X22= 34

McNemar 検定 : 2 乗値 = 1.1250 P=0.2888
 McNemar exact test : 直接確率 = 0.2890625

Odds 比 = 3.0000

95% 信頼区間 0.54 ~ 30.39

(3) M x N の χ^2 検定

2 以上の群に 2 以上の特性項目があるとき、各群の各特性の母出現率に差がない（帰無仮説）かどうかを検定する。この場合自由度が 2 以上であればその確率が同時に算出される。自由度は、M 群に N 項目あるとすれば、 $(M - 1) \times (N - 1)$ となる。

M x N の χ^2 乗検定

23	12	36
36	32	52
44	28	19
56	41	24

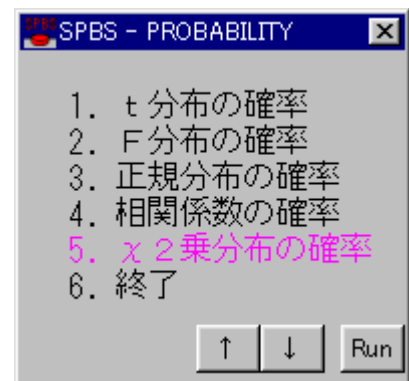
χ^2 値 = 33.1310
 自由度 = 6
 危険率 = P<0.001

(4) Mantel-Haenszel 検定

2 x 2 分割表の χ^2 検定の特殊な形として、Mantel-Haenszel 検定がある。これは同じ 2 x 2 分割表の χ^2 検定でも、層別化されたデータの場合に使用する。例えば、癌の発生は多くの場合加齢とともに発生頻度が高くなるので、30 歳から 79 歳までの集団で有害因子の曝露の有無と癌の発症の有無で 2 x 2 の分割表を作成して通常の χ^2 検定をおこなうと集団間の年齢差を無視した結果になる可能性がある。すなわち、対照群には若年者が多数を占め、患者群には高齢者が多数占めているような状況もありうる。このような場合、年齢による補正が必要になる。Mantel-Haenszel 検定は、上述の例では年齢階級別（例えば 10 歳階級毎）に 2 x 2 分割表を作成し、全体として有害因子の曝露影響が癌の発生に関連していたのかを判定する（自由度 1 の χ^2 検定）。結果として、層別化に使用した変数の影響を取り除いたことになる。このプログラムでは Mantel-Haenszel 法による Odds 比および Odds 比の 95% 信頼区間を Robins et al (1986) の方法により算出する。

(5) t 分布、F 分布、 χ^2 分布

SPBS 統計パッケージでは、プリントスペースを極力減らすために、全ての箇所確率値 P を算出してない。どうしても確率 P 値が必要な場合に、SPBS 統計パッケージで算出された t 値、F 値、 χ^2 値あるいは相関係数と自由度を入力すると P 値を求めることができる。ここでの確率値は通常の値よりも 1 桁多い小数点第 4 位まで出力する。



(6) 割合の区間推定 (95% 信頼区間)

選挙前の無作為抽出による標本調査で「C 政党の支持率は 40% と推定される」という結論が得られたとしよう。しかしながら、調べた標本数が提示されなければ、この 40% という数字がどの程度信頼できるかは明らかでない。

ある集団（人数 N）の中で、A 人が C 政党を支持していたとすると、その支持率は A/N で表される（N=A+B）。この時の標本割合 $P=A/N$ がどの程度信頼できるかは、この P の値がどのような範囲にある数値なのかを示すことによって推定可能となる。Q=1-P として、 $(P \cdot Q/N)$ は P の標準偏差であり、推定値 P の誤差を測る目安となる。実際には、標本割合の 95%（時に 99%）信頼区間を標準偏差を用いて算出することにより、当該数値の精度（信頼度）を推し量ることになる（Fleiss JL, 1981）。

このプログラムでは、上の A（割合算出の際の分子に相当）と N（割合算出の際の分母）をキーボードより入力すると、95% および 99% 信頼区間が算出される。下の例は、割合が 40% であっても、標本数が 5 の場合と 1000 の場合で信頼区間の値が如何に異なるかを如実に表している。

割合の区間推定 (95% 信頼区間)

割合算出のための分子の値	2		
割合算出のための分母の値	5		
母集団割合	= 0.40000	95% 信頼区間	0.072584 ~ 0.829576
標本偏差	= 0.219089	99% 信頼区間	0.049686 ~ 0.877577

11. 共分散分析 (2群の有意差検定)

2グループ (主効果) 間の正規分布するデータの有意差検定をおこなう。この場合、幾つかの交絡因子 (共変量) が介在していることが自明であるデータの比較をおこなう時にこの共分散分析を使用する。このプログラムでは、「交絡因子間の相関 (交互作用) は無い」との仮定下に解析をおこなっているため、交絡因子を無思慮に多数選択する事は避けるべきである。サンプル数が多い時には層別化して解析するのの一法である。

Datafile の数 One Datafile Two Datafiles
 指定変数による一括計算 実行 補正值データの一時保存

データファイルは1つまたは2つ入力することができるが、それぞれに制限がある。1つの場合には、2グループに分割するための基準変数をデータファイルに含んでいなければならない。このため、共分散分析用のデータファイルは変数が最低でも3以上必要となる。2つの場合は、データファイルの各変数の配列 (順番) が同一でかつ変数が2以上ないと解析できない。また、サンプル数が解析に使用する変数の数の (最低でも) 3 ~ 5 倍ないと結果の信頼性が十分でないと考えられている。

共分散分析は線形回帰式を用いて元データを補正 (調整) し、その上で有意であるか否かを検定する。線形関係のない変数を加えると誤った結論を導くことになるので細心の注意が必要である。

単に2グループ間の有意差検定 (Student の t 検定および Welch t 検定) をおこなうこともできるが、別途「2グループ間の有意差検定」というプログラムが SPBS 統計パッケージには用意されている。そのプログラムでは、正規分布形にこだわらない有意差検定ができるように、Wilcoxon の順位和検定等が含まれている。

2グループのデータを図として描くことができる。この場合、「有意差検定 (2グループ間の検定)」は粗データであるのに対し、共変量で補正された値で描かれる。

なお、本解析はGLM (一般線形モデル) の Type の平方和モデルに準拠しており、主効果と共変量 (交絡因子) との交互作用の検定も可能である。但し、交互作用が検討できるのは共変量に加えた変数のみであり、「主効果」とは < 2群に分類した変数 > を指す。

Analysis of Covariance

第1群の変数 (17): 白血球数 B:¥QXY96
 第2群の変数 (17): 白血球数 B:¥QXY96

2群に分類した変数 27 - 作業 (0=日勤 1=交勤)

	1グループ	2グループ
平均	6480.1653	7137.9464
標準偏差	1774.9469	2050.9294
最小値	3300.000	3400.000
最大値	13400.000	14300.000
サンプル	121	224

対応のない t 検定

F(223 , 120) = 1.33515 n.s. (両側検定) --> 「等分散の場合」を見よ!

	等分散 (Student)	不等分散 (Welch)
t-値	-2.976	-3.107
df	343	277
確率	P=0.003	P=0.002
* * 危険率 1% で有意差があります * *		

共分散分析 共変量 (Covariate) を調整した時の結果

調整後のデータ 1グループ 2グループ 2群の差

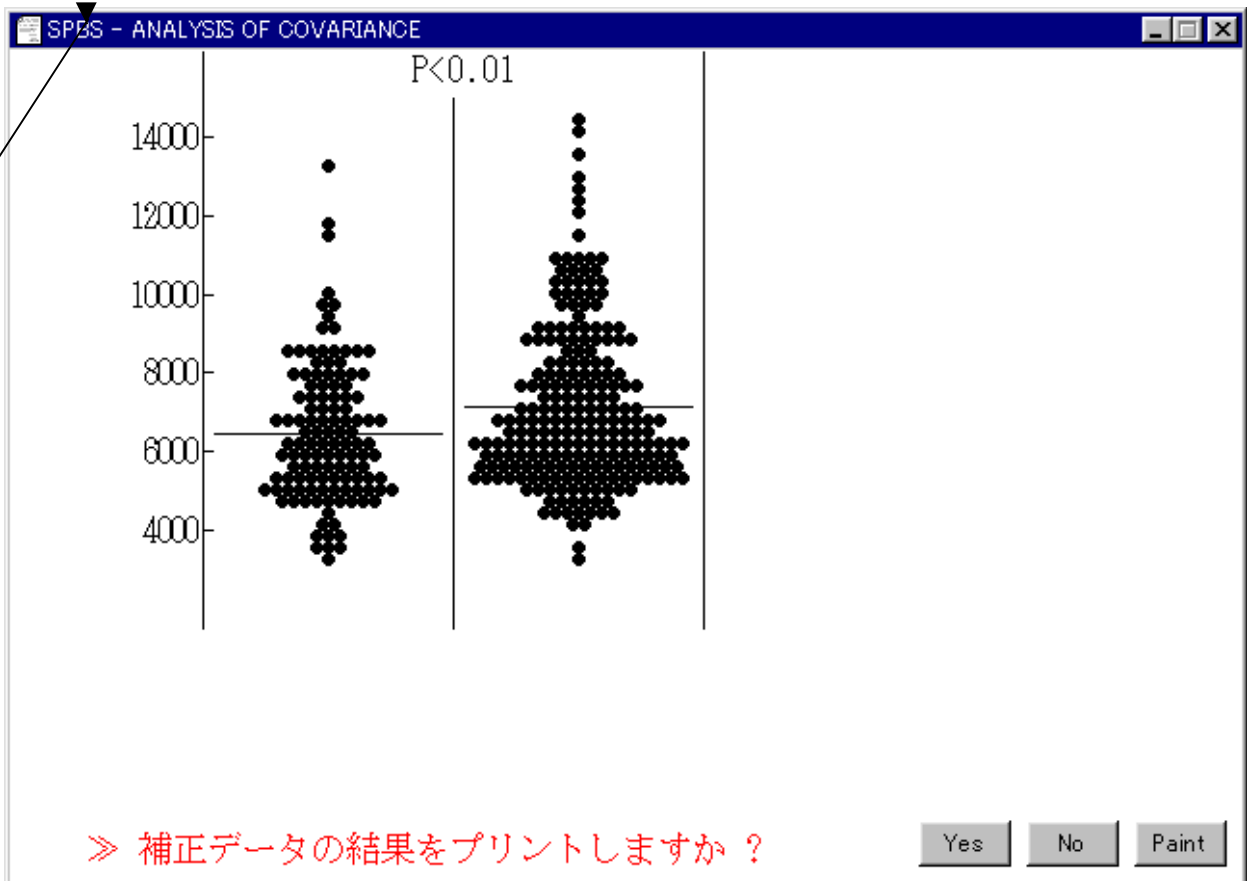
平均値 : 6478.479 7138.857 -660.378
 標準偏差 : 1757.730 2049.435 221.986 (推定の標準誤差)

調整後の2群の有意性 t value = -2.975 (df = 341) P=0.003

共変量(交絡変数)の推定回帰係数 () [Type の平方和モデル]

推定回帰モデルの有意性 F = 3.709 (df1 = 3 , df2 = 341) P=0.012

(2) = 10.860083 P=0.283 年齢
 (25) = 39.044230 P=0.329 喫煙本数



上の図は2群の年齢と肥満度 (BMI) で補正された白血球数のプロットである。この図の縦軸を自由に変更する事ができる。「図をプリントしますか?」と表示された時に、キーボードで「m」または「M」を入力すると、画面上で最大値と最小値を尋ねてくる。表示される最大値より小さい値、あるいは表示される最小値より大きい値を入力すると無視 (変更取り消し) される。但し、このオプションは「ペイント」画面に対してのみ有効で、プリンターへの印刷はできない。

◀ **ヒント** 画面のコピーはペイントを用いなくとも上のタイトルバー (「SPBS - ANALYSIS OF COVARIANCE」と書いてある青色部分) を右クリックして「コピー」の「テキスト/グラフィクス」を選択するとクリップボードに画面が移動する。この操作テクニックは SPBS 統計パッケージの「解析方法の選択」を除く全てのプログラムで利用できるようになっている (文字情報のみの場合は、上の操作の内「コピー」の「テキスト」を選択する)。



12. 内的整合性の検討

Cronbach の係数（信頼性係数）を算出し、項目群の内的整合性を検討する。この場合、ファイルの中からカテゴリーデータのみからなる関連する項目を複数個指定した後で、まずそれらの項目群の相関行列を算出する。その後、信頼性係数を算出する。これは、項目の折半を必要とせず、反復測定も必要なく、信頼性係数の唯一の推定値が求められる方法であり、「同一特性を測っている質問項目は内的整合性のある項目群から構成されていなければならない」との考えに基づいて信頼性を評価する。通常、最終的な項目数としては5項目以上が必要であると考えられている。信頼性係数の水準に関する確たる基準はないが、一般に0.8以上が必要だと考えられており、この水準以上であれば各項目と項目全体（項目の総和）との間に高い整合性があると判断される。

これらの係数を算出した時、計算に使用した合成変数を保存することもできる。このため、メッセージボックスの指示に注意を払いながら実行する必要がある。また、Cronbach の係数を算出する際に相関係数も算出されるが、その有意性は「*」、「**」および「¥¥」の記号で表現される（各々 $P<0.05$ 、 $P<0.01$ 、 $P<0.001$ を表す）。さらに、このプログラムでは各変数に重み付けすることもできるようになっている。使用する変数を決定した後で、ある変数の重み付けを2としたい場合には、重み付けを変更しようと思っている変数の「重み付け」部位を左クリックすると、重み付け値が入力できるようになっている（この重み付けを変更しても相関係数の絶対値は変わらないが、係数の値は変わる）。

このプログラムでは、項目群内の得点が逆方向になっている場合に得点を逆方向に変更することができる。例えば、質問紙の中で同じ内容を正反対の意味で尋ねる項目が挿入されることがあるが、このような質問は他の類似質問との間に有意な負の相関があることになり、その得点を単純に項目群に加算することはできない。この変更を終えた後で、Cronbach の係数を算出することも、またその加算したデータを保存することもできるように設計されている。但し、ここで変換できるのは得点が整数であり、かつ0以上で20以下の場合に限られる。

Cronbach's coefficient

例題nima

基本統計

変数番号	平均	標準偏差	(最小値 - 最大値)	変数名
8	2.869	0.891	1.000 - 4.000	家庭ストレス
9	2.049	1.418	1.000 - 5.000	仕事の過酷さ
10	1.986	0.939	1.000 - 3.000	仕事上の地位
11	2.332	0.968	1.000 - 5.000	職場ストレス
12	1.341	0.580	1.000 - 4.000	労働時間
13	1.091	0.347	1.000 - 4.000	残業時間
15	1.241	0.666	1.000 - 5.000	通勤時間
18	3.220	0.904	1.000 - 5.000	健康状態
	16.128	2.704	9.000 - 25.000	使用変数の総和

サンプル数 = 701

欠損値を除外しています

変数間の単相関係数 (* $P<0.05$ ** $P<0.01$ ¥¥ $P<0.001$)

	8	9	10	11	12	13	15	18
8	1.000¥¥							
9	0.030	1.000¥¥						
10	-0.052	0.105**	1.000¥¥					
11	0.026	0.030	0.026	1.000¥¥				
12	0.018	-0.018	-0.062	-0.144¥¥	1.000¥¥			
13	-0.077*	-0.006	0.022	-0.044	0.449¥¥	1.000¥¥		
15	0.034	0.038	0.028	-0.071	0.046	0.034	1.000¥¥	
18	0.054	0.038	-0.069	0.362¥¥	-0.078*	-0.000	-0.057	1.000¥¥
	0.357¥¥	0.599¥¥	0.368¥¥	0.458¥¥	0.181¥¥	0.196¥¥	0.257¥¥	0.447¥¥

内的整合性の検討

(欠損値を除外)

Cronbach's coefficient = 0.15153

(通常は 0.8 以上)

13. ROC 曲線

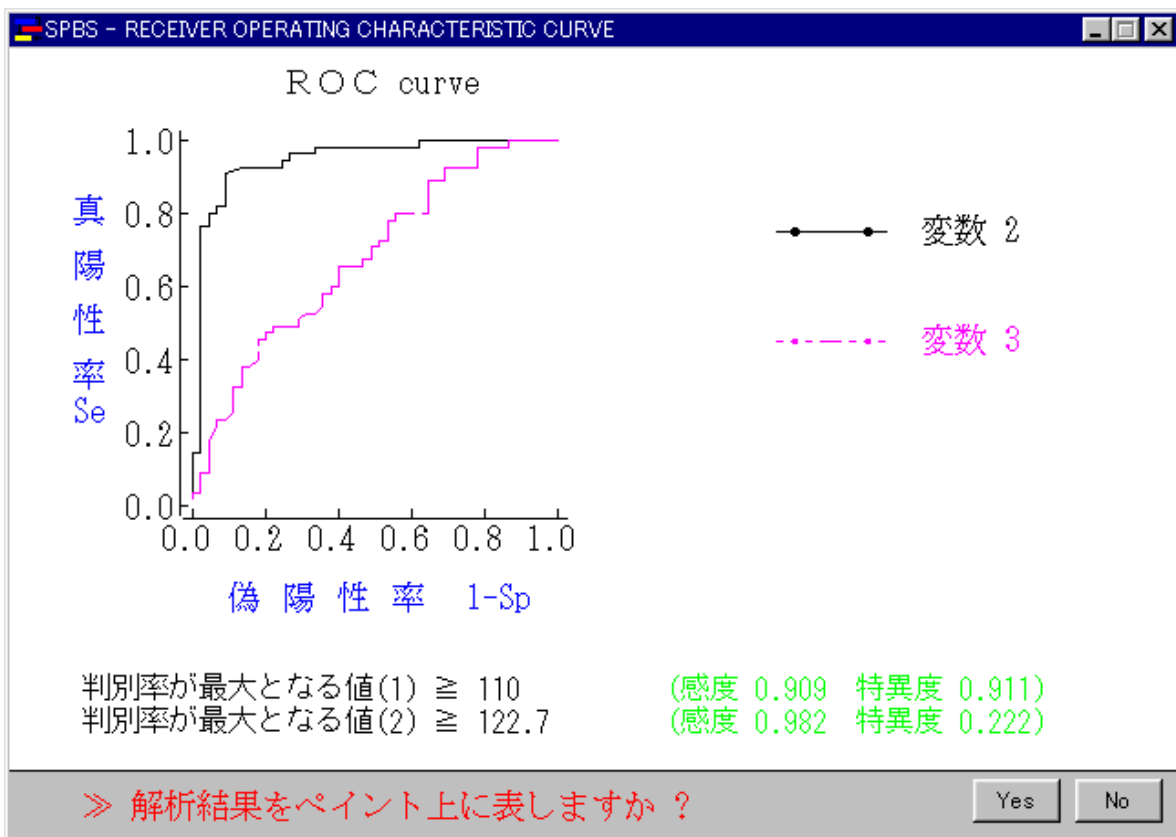
確定診断結果を表す変数とスクリーニング検査の変数より、真陽性率（感度）と偽陽性率（1 - 特異度）を算出し、ROC 曲線を描く。スクリーニング検査では、高い感度と特異度が要求されるが、どの辺りにカットオフ値（判定値）を定めるかを決定する前にこの ROC 曲線を描く必要がある。なお、スクリーニング検査のデータが小さい方から大きい方に並べ替えて各々の検査値の感度および特異度を逐次出力するが、感度を優先させているので、同一感度の場合特異度が最も大きくなった時の値しか出力されない。

最初に、確定診断結果（0 - 1 データ、または連続量でも可）を表す変数を入力する。次に、3つの方法で ROC 曲線を描く検査データを選択できる。

同じ確定診断結果に対し、2つの検査のどちらが判定率（感度および特異度）が高いかどうかを検討することができる。この比較は、2つの ROC 曲線の面積を算出し、その面積の大きさで判定率の比較をおこなう。Hanley & McNeil（1982、1983）の論文に大体準拠しているが、保守的な方法 $SE(\text{Area}_1 - \text{Area}_2) = \{ (SE_2(\text{Area}_1) + SE_2(\text{Area}_2)) \}$ で標準誤差を算出している。

検査値データを共変量（交絡因子）で補正した後で、ROC 曲線を描くこともできる。性・年齢などに影響されやすい検査の場合に、これらを補正した場合としない場合の ROC 曲線を比較することができる。

1つの検査データのみを選択し、その ROC 曲線を描くこともできる。



解析結果では、最初に判別率が最大となる値をカットオフ値として2群に分け、その差の比較をおこなう。次に判別率が最大時の病気の有無、検査陽性・陰性者の各人数（2 × 2 分割表）、その後 ROC 曲線下の面積、感度、特異度、95%信頼区間を算出する。なお、感度および特異度の 95%信頼区間の算出には直接確率計算法による推定を用いている（矢野栄二ほか、1999）。

14. 相関分析 (単相関, 偏相関, 重相関係数)

2つの変数に有意な関係があるかどうかを推定する時に使用する。ある変数Yがある変数Xに依存するかどうかは回帰分析で解析するが、ここでは単に関係を問題としている。この関係には正の関係、負の関係および全く関係が認められない場合の3通りがある。これらは単相関係数により判断される。また、この有意性の検定にはt検定が用いられている。本解析では相関係数とその有意性を計算する。なお、サンプル数が少ない場合には順位相関係数の方が良く使用される。



偏相関係数、重相関係数の算出は変数が12以下でなければおこなえない。この時、多重有意性検定の基準相関係数も算出する。変数の数が2の時には単相関係数のみであり、勿論偏相関係数や重相関係数は算出されないが、その代わりに画面で散布図を見ることができる(但し、X軸およびY軸を自分で決められない。X軸、Y軸を自分で決めて描かせたい場合には、「Spearmanの順位相関係数」のプログラムを利用することができる)。

2変数の場合にはグラフが画面上に現れるとともに、黒字で「グラフ作図」ボタンが現れる。プリンター経由で印刷したい場合、あるいは「ペイント」上に描きたい場合にこのボタンを操作する。

相関係数間の独立性検定は、検討している当該相関係数の間(下の例では3つの変数間)で依存関係があるかどうかを検討する。独立性が棄却されれば、「当該変数間に依存関係がある」ことを示唆する。この考え方はMorrison(1976)の本の中に詳述されている。

多重有意性検定(multiple significance test)用の基準相関係数(r)もMorrison(1976)の本の中で記述されている方法を用いて算出している(下式のZは有意水準の正規分布の値、mは組合せ数、Nは相関係数算出時のサンプル数)。組合せ数によって基準値が異なるので、何個の相関係数で多重有意性検定をおこなっているのかを確認してから利用する。

$$|r| > \tanh\left[\frac{Z}{(N-3)}\right]$$

使用変数名

2 年齢
11 HDL コレステロール [Log]
25 BMI (Kg/m²)

単相関係数行列	[:有意性(P)	:相関係数]
2 -- 11 -- 25 --		
2	-0.039	0.062
11	0.470	-0.307
25	0.251	<0.001

自由度 = 343

相関係数間の独立性検定(Hypothesis of Independence)

2値 = 35.359 P<0.001 (自由度 df = 3)

これら相関係数間の独立性が否定されました。

多重有意性検定(Multiple significance test)用の基準相関係数

[組合せ数 3の場合]	P<0.05 基準 : r = 0.129	P<0.01 基準 : r = 0.157
[組合せ数 2の場合]	P<0.05 基準 : r = 0.121	P<0.01 基準 : r = 0.151

偏相関係数行列	[:有意性(P)	:偏相関係数]
2 11 25		

2	-0.021	0.052
11	0.698	-0.306
25	0.336	<0.001

重相関係数 と F 値

番号	2	0.0653	F(2 , 342)=	0.7327	P=0.481
番号	11	0.3078	F(2 , 342)=	17.8920	P<0.001
番号	25	0.3111	F(2 , 342)=	18.3259	P<0.001

このプログラムでは、既に算出されている2つの相関係数が有意に異なっているかどうかを正規分布を用いて検定することができる。算出式は以下の通りである（Zの値が正規分布近似する）。この場合、2つの相関係数の値（R1、R2）の他にそれぞれの相関係数を算出した際の自由度（df1、df2）を順次キーボードで入力する。

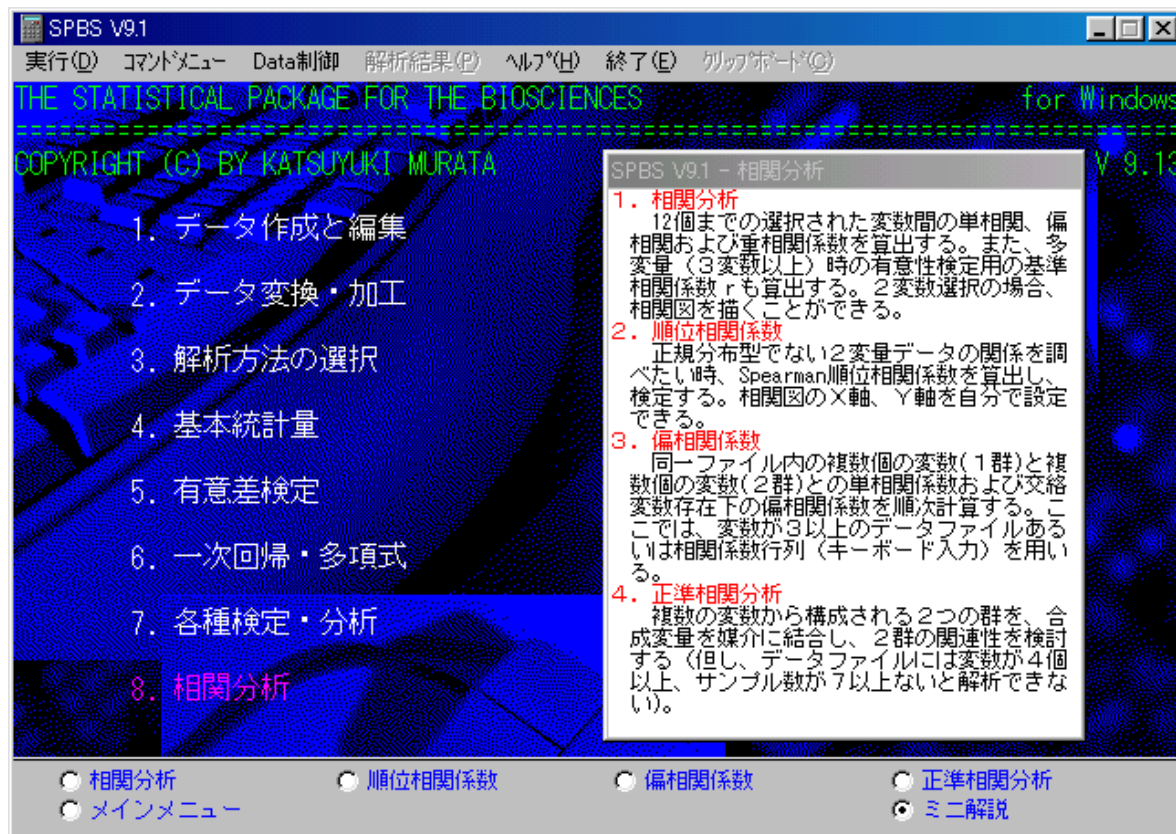
$$Z = \frac{\log\{(1+R1)/(1-R1)\} - \log\{(1+R2)/(1-R2)\}}{2 \cdot \left\{1/(df1-1) + 1/(df2-1)\right\}}$$

2群の相関係数の差の検定（正規分布近似）

1番目の相関係数 r= 0.432 (自由度 123)
 2番目の相関係数 r= 0.245 (自由度 235)

統計量 T= 1.901 P=0.057

メインメニューの「8. 相関分析」をダブルクリックすると、下図が現れる。ここでは、「相関分析」、「順位相関分析」、「偏相関係数」、「正準相関分析」を選択できる。また、「ミニ解説」をクリックすると、下図のように簡単な説明が見られる。



15. スピアマンの順位相関係数

2変量の関係を Spearman の順位相関係数で検討する。単相関係数の場合、2変量データが各々正規分布することが前提であるが、順位相関係数ではデータの分布形に依存しない。特に、サンプル数が少なくかつ平均値から著しくかけ離れたデータが存在する時、単相関係数では有意な相関があるように示されることがある。このような場合には、特に Spearman の順位相関係数による検定が推奨される。なお、サンプル数が 26 以上になると検定表がないので、正規分布近似による t 検定 (t 値は同時に算出される) をおこなう。



データ入力後、最初に順位相関係数を算出する 2 変数がどれなのか尋ねてくる。順位相関係数の他に単相関係数と X - Y 平面上のデータプロットを表示するので、特にかげ離れたデータがないかどうかを画面で見ながらチェックできる。また、画面のデータを「ペイント」上またはプリンターに描くことができる。

同順位のデータを含む場合、同順位データを補正した順位相関係数が算出される (Siegel, 1956) 。

Spearman's rank correlation coefficient

D:¥SPBS¥abc

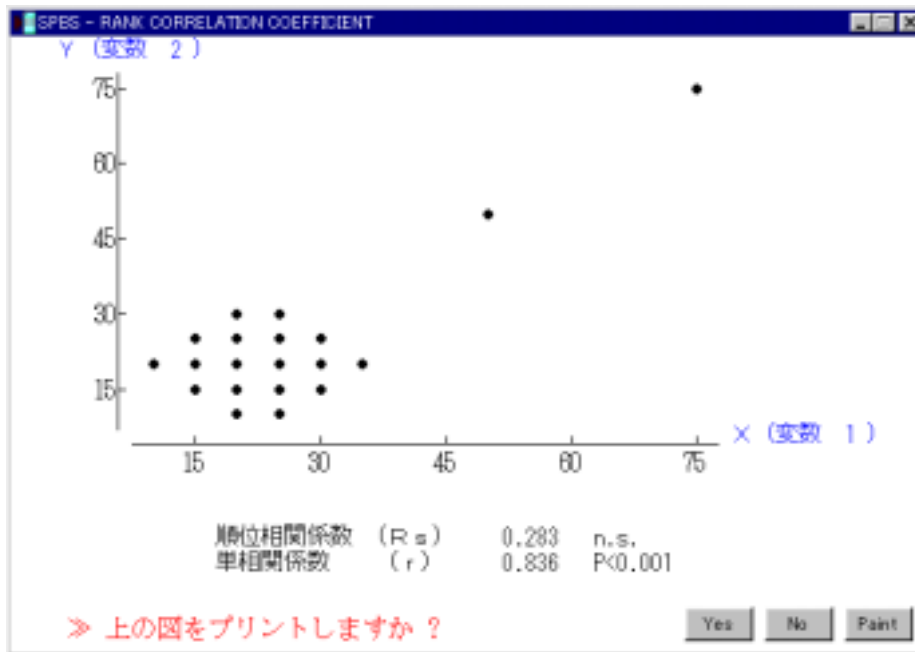
順位相関係数行列	: 確率 P 値		: 相関係数				
	1	2	3	4	5	6	7
1		-0.115	0.033	-0.050	-0.059	-0.024	-0.034
2	<0.001		-0.234	-0.149	0.079	0.422	0.213
3	0.018	<0.001		-0.499	-0.460	-0.233	-0.268
4	<0.001	<0.001	<0.001		0.577	0.018	0.091
5	<0.001	<0.001	<0.001	<0.001		0.223	0.840
6	0.087	<0.001	<0.001	0.192	<0.001		0.274
7	0.014	<0.001	<0.001	<0.001	<0.001	<0.001	

確率は近似 t 値を用いて算出しています (自由度 5078)

変数番号 4 身長
 変数番号 2 健診時年齢

順位相関係数 (R s) -0.149 P<0.001
 単相関係数 (r) -0.130 P<0.001

--> 単相関係数 r の 95% 信頼区間 : -0.157 r -0.103



注意点

左の図は、肝機能検査の GOT と GPT の 20 名の検査結果を图示したものである。これを見ると単相関係数は有意水準 $P < 0.001$ で有意であるが、順位相関係数は有意でない。単相関係数の場合、平均よりかけ離れたデータ (50, 50) と (75, 75) があるため、それにより全体の平均値が右上に引っ張られてしまい、一見 “ 有意な関係がある ” ような錯覚に陥ってしまう。このように、データを必ずプロットし、かけ離れたデータがないかどうか確認することが大切である。

16. 偏相関係数

2つの変数に有意な関係があるかどうかを推定する時に使用する。ある変数Yがある変数Xに依存するかどうかは回帰分析で解析するが、ここでは単に関係を問題にしている。この関係には正の関係、負の関係および全く関係が認められない場合の3通りがある。これらは単相関係数により判断される（この有意性の検定にはt検定を用いる）。本解析では相関係数、偏相関係数およびその有意性を計算する（相関係数を算出する変数の数は縦軸は制限ないが、横軸は12個である。縦横軸については下例を参照）。なお、サンプル数が少ない場合には順位相関係数の方が良く使用される。

入力方法	<input checked="" type="radio"/> Datafile	実行	終了
	<input type="radio"/> Keyboard (相関係数行列)		

偏相関係数は交絡因子（第3の変数）の影響を除外する目的で算出される。疫学的な考え方を適用すると、多くの生物学的データ（変数）が性・年齢の影響を受けているので、それらの変数間の関係を調べる時にはデータを層別化（性別、年齢別に分ける）した後に相関係数を算出すべきである。サンプル数が少ない場合には層別化が不可能であることが多く、算術的に交絡因子の影響を除外した偏相関係数を使用することが多い。相関係数の多重有意性検定（multiple significance test）用の基準値（r）の算出にはMorrison（1976）の方法を使用している。

使用する変数の一覧

横列の変数番号と変数名

V005 収縮期血圧
V006 拡張期血圧
V021 心拍数

縦列の変数番号と変数名

V009 総コレステロール
V010 中性脂肪
V011 HDL-コレステロール
V016 空腹時血糖

単相関係数 [* P<0.05 ** P<0.01 ¥¥ P<0.001 (自由度 = 383)]

	V005	V006	V021
V009	0.112*	0.173¥¥	0.048
V010	0.209¥¥	0.165**	0.042
V011	-0.019	-0.032	-0.024
V016	0.205¥¥	0.188¥¥	0.135**

多重比較(Multiple comparison)の場合の相関係数

[組合せ数 4の場合]	P<0.05基準 : r = 0.127	P<0.01基準 : r = 0.153
[組合せ数 3の場合]	P<0.05基準 : r = 0.122	P<0.01基準 : r = 0.149
[組合せ数 12の場合]	P<0.05基準 : r = 0.146	P<0.01基準 : r = 0.169

交絡変数 : 年齢

偏相関係数 (自由度 = 382)

	V005	V006	V021
V009	0.042	0.087	0.032
V010	0.177	0.122	0.033
V011	-0.023	-0.039	-0.025
V016	0.134	0.091	0.122

偏相関係数の確率値 (P)

	V005	V006	V021
V009	0.4069	0.0896	0.5260
V010	0.0005	0.0163	0.5159
V011	0.6598	0.4481	0.6320
V016	0.0084	0.0760	0.0170

17. 正準相関分析

正準相関分析はそれぞれ2変数以上から成る2群の関係を統計的に検討する。すなわち、2群の相関が最大になる各々の総合特性値（正準得点）を算出し（この総合特性値を算出する線形回帰式の係数が正準変量係数）、各々の群の間に有意な関係があるか否かを検定する（田中豊ほか、1984）。従って、この解析法は重回帰分析と主成分分析をミックスした如き方法といえる。

データ入力後、それぞれの変数が第一、第二群のいずれに属するかを尋ねてくる（両群とも最低2変数以上なければ解析を実行しない）。次に正準得点を印字（作図）するか否かを尋ねてくる。

解析結果には、全変数の単相関係数、分散共分散、正準相関係数とその有意性、正準変量係数、正準変量と元の変量との相関係数、および正準得点を算出する。また正準相関係数が有意な場合には第一二群の正準得点を二次平面上にプロットすることができる（但し、サンプル数が16以上の時のみ。この正準得点は標準化正準変量係数と基準化データの積より算出している）。

総合特性値の意味付けは、正準変量と元の変量との相関係数が高いもの（寄与率が高いもの）を中心に、解析者が考案することになっている。

第1群の使用変数番号

- X(1) 反復横跳び
- X(2) 垂直跳び
- X(3) 背筋力
- X(4) 握力
- X(5) 踏台昇降
- X(6) 立位体前屈

第2群の使用変数番号

- X(7) 伏臥上体そらし
- X(8) 50m走
- X(9) 走り幅跳び
- X(10) ボール投げ
- X(11) 懸垂
- X(12) 持久走

基本統計

変数番号	平均	標準偏差	(最小値 - 最大値)	
1	47.421	3.366	41.000	54.000
2	59.684	6.921	50.000	74.000
3	114.474	15.660	90.000	155.000
4	44.632	5.355	29.000	53.000
5	66.911	14.449	45.900	98.000
6	15.342	5.951	2.000	25.000
7	59.737	8.143	39.000	75.000
8	7.137	0.351	6.600	8.000
9	441.053	42.389	362.000	522.000
10	27.816	2.749	21.000	33.000
11	7.579	3.782	2.000	21.000
12	366.763	31.185	298.000	447.000

サンプル数 = 38

[相関係数行列 (自由度 36) * P<0.05, ** P<0.01]

	1	2	3	4	5	6	7	8	9	10	11	
1	1.000**											
2	0.301	1.000**										
3	0.164	0.387*	1.000**									
4	-0.029	0.009	0.319	1.000**								
5	0.246	-0.101	-0.243	-0.037	1.000**							
6	0.072	0.456**	0.193	0.052	0.052	1.000**						
7	-0.166	0.251	-0.018	0.204	0.323*	0.281	1.000**					
8	-0.401*	-0.300	-0.303	-0.283	-0.430**	-0.080	-0.257	1.000**				
9	0.361*	0.640**	0.559**	0.271	-0.184	0.260	0.150	-0.443**	1.000**			
10	0.412*	0.508**	0.554**	-0.041	-0.012	0.331*	0.039	-0.265	0.499**	1.000**		
11	0.280	0.560**	0.322*	0.247	0.142	0.236	0.084	-0.463**	0.607**	0.356*	1.000**	
12	-0.470**	-0.243	-0.480**	-0.093	-0.018	-0.289	0.192	0.074	-0.469**	-0.521**	-0.420**	1.000**

12
12 1.000**

[分散共分散行列 (自由度 37)]

	1	2	3	4	5	6	7
1	11.331						
2	7.001	47.898					
3	8.660	41.964	245.229				
4	-0.516	0.340	26.747	28.671			
5	11.979	-10.124	-54.924	-2.861	208.759		
6	1.447	18.787	17.996	1.670	4.442	35.420	
7	-4.562	14.131	-2.250	8.873	38.019	13.633	66.307
8	-0.473	-0.729	-1.661	-0.532	-2.176	-0.167	-0.733
9	51.491	187.747	371.083	61.533	-112.898	65.495	51.798
10	3.809	9.670	23.846	-0.610	-0.460	5.416	0.869
11	3.560	14.647	19.043	5.003	7.734	5.310	2.589
12	-49.357	-52.509	-234.452	-15.522	-7.979	-53.674	48.882

	8	9	10	11	12
8	0.123				
9	-6.583	1796.808			
10	-0.255	58.145	7.560		
11	-0.614	97.266	3.704	14.304	
12	0.814	-619.771	-44.639	-49.535	972.510

[正準相関係数]

	r^2	2 値	自由度	有意性
第 1 正準相関	0.852	0.726	95.248	36 P<0.001
第 2 正準相関	0.752	0.565	56.394	25 P<0.001
第 3 正準相関	0.716	0.512	31.446	16 P=0.012
第 4 正準相関	0.411	0.169	8.994	9 P=0.438
第 5 正準相関	0.284	0.081	3.421	4 P=0.490
第 6 正準相関	0.083	0.007	0.340	1 P=0.560

[標準化正準変数係数 (上段=第 1 群、下段=第 2 群)]

変数 \ 軸	1	2	3	4	5	6
1	0.3941	-0.4654	0.4928	-0.3914	-0.1110	-0.6819
2	0.3054	0.3289	-1.0301	-0.1998	0.5164	-0.1644
3	0.5657	-0.2795	0.3763	0.3004	0.1599	0.8829
4	0.0923	0.3685	-0.2346	-0.6559	-0.7160	0.0078
5	0.2915	0.8978	0.3137	0.1202	0.3838	0.2391
6	0.1011	-0.0098	0.1879	0.7868	-0.7148	-0.3555
7	0.0808	0.6896	-0.1958	0.5780	-0.4930	0.2580
8	-0.4369	-0.3989	-0.7432	0.6078	-0.2475	-0.3903
9	0.1815	-0.7298	-0.7776	-0.5207	-0.5337	0.5788
10	0.3532	-0.2604	-0.0961	0.7574	0.8790	0.1880
11	0.0468	0.5031	-0.5764	-0.0130	0.3738	-1.0485
12	-0.3680	-0.0693	-0.8429	-0.2071	0.8868	0.3242

[正準変数係数 (上段=第 1 群、下段=第 2 群)]

変数 \ 軸	1	2	3	4	5	6
1	0.1171	-0.1383	0.1464	-0.1163	-0.0330	-0.2026

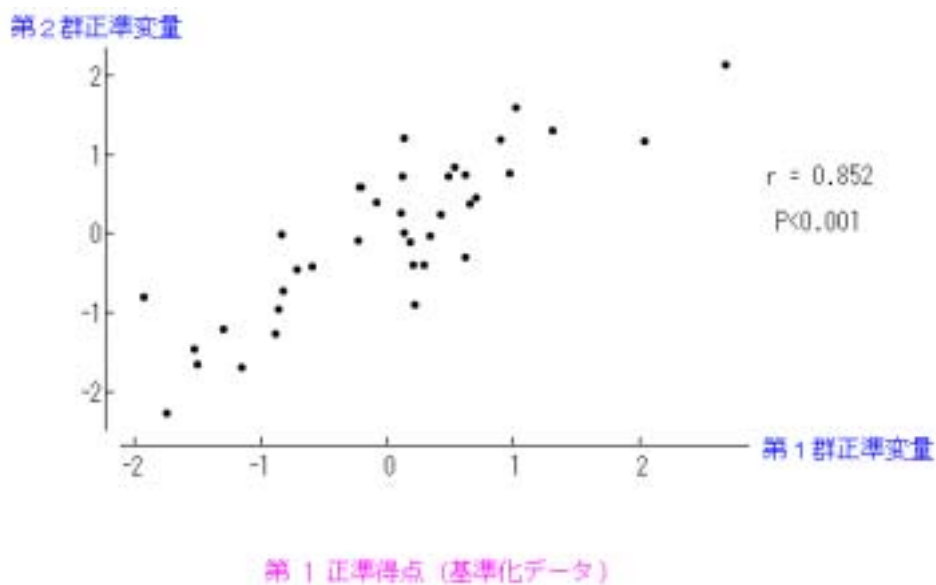
2	0.0441	0.0475	-0.1488	-0.0289	0.0746	-0.0238
3	0.0361	-0.0178	0.0240	0.0192	0.0102	0.0564
4	0.0172	0.0688	-0.0438	-0.1225	-0.1337	0.0015
5	0.0202	0.0621	0.0217	0.0083	0.0266	0.0165
6	0.0170	-0.0016	0.0316	0.1322	-0.1201	-0.0597

7	0.0099	0.0847	-0.0240	0.0710	-0.0605	0.0317
8	-1.2461	-1.1376	-2.1198	1.7334	-0.7060	-1.1133
9	0.0043	-0.0172	-0.0183	-0.0123	-0.0126	0.0137
10	0.1284	-0.0947	-0.0350	0.2755	0.3197	0.0684
11	0.0124	0.1330	-0.1524	-0.0034	0.0988	-0.2772
12	-0.0118	-0.0022	-0.0270	-0.0066	0.0284	0.0104

[正準変量と元の変量との相関 (上段=第 1 群、下段=第 2 群)]

変数 \ 軸	1	2	3	4	5	6
1	0.655**	-0.203	0.343*	-0.297	0.134	-0.553**
2	0.660**	-0.011	-0.685**	0.140	0.173	-0.214
3	0.727**	-0.331*	-0.056	0.072	-0.118	0.583**
4	0.259	0.262	-0.140	-0.514**	-0.709**	0.280
5	0.222	0.804**	0.466**	0.036	0.255	-0.145
6	0.398*	0.118	-0.170	0.697**	-0.474**	-0.297

7	0.167	0.701**	-0.336*	0.332*	-0.273	0.427**
8	-0.681**	-0.422**	-0.119	0.480**	-0.224	-0.253
9	0.764**	-0.242	-0.480**	-0.236	-0.248	0.096
10	0.771**	-0.277	-0.061	0.462**	0.331*	0.048
11	0.646**	0.239	-0.401*	-0.205	0.064	-0.564**
12	-0.674**	0.300	-0.279	-0.195	0.409*	0.416**



18. 生存分析 (Survival Analyses)

生存分析には2種類のプログラムが用意されている。1つはKaplan-Meier法による生存曲線の推定で、2つの生存率を時間軸上に表す。今1つはCoxの比例ハザードモデルである。2つのプログラムとも、最初に「死亡」か「観察打ち切り」かを示す変数を指定する必要がある(このプログラム内では「死亡」=1、「打ち切り」=0としています)。「打ち切り」がない場合には全員死亡したものとして解析をすることができる。次に「観察(生存)期間」を表す変数を指定する。勿論、この変数がないデータファイルでは生存分析をおこなうことができない。「観察期間」は日数、週数、月数いずれでも良いが、生命表の図を描く際になるべく全て整数で表わせる数値の方が推奨される。

解析方法は	<input checked="" type="radio"/> Kaplan-Meier法による生存曲線の推定 <input type="radio"/> 多変量生存分析(Cox法 or Weibull法)	<input type="button" value="実行"/> <input type="button" value="終了"/>
-------	---	---

Kaplan-Meier 法による生存曲線の推定をおこなう際には、2群に分類する変数が必要となる(3群以上の生存曲線を同時に作成することはできない)。生存期間が2群で有意に異なっているかどうかの検定にはGeneralized Wilcoxon および Log-Rank 検定を用いる。Coxの比例ハザードモデルで多変量による生存分析をおこなう時には、最初にこのKaplan-Meier法による生存曲線を描き、2群の生存曲線がクロスしていないことを確認する必要がある。2群の生存曲線がクロスしている場合には、比例ハザード性がないことを示しており、比例ハザードモデルを適用できない。

生存関数の算出に当たって、同じ時点に「死亡」データと「打ち切り」データが存在する場合、「打ち切り」データをどう処理するかが問題になる。1つの考え方は、「死亡」直後に「打ち切り」があったと考え、当該死亡率を算出する生存者数に「打ち切り」データを加える。もう1つの考え方は、同一時点に「打ち切り」が存在した時に「打ち切り」データを0.5人として生存者数に加える。最近は前者の考え方を採用している統計パッケージが多いので(大橋靖雄・浜田知久馬、1995)、SPBS統計パッケージでも前者の考え方を採用している。

Survival curve by Kaplan-Meier method

A: ¥白血病

2群に分けた変数 2 抗癌剤使用(1)

第1群のサンプル数 = 21 第2群のサンプル数 = 21

Kaplan-Meier の推定量 (第1群)

生存期間	死亡数/生存数	当該生存率	生存率(SR)	SRの標準誤差
1	2 / 21	0.9048	0.9048	0.0641
2	2 / 19	0.8947	0.8095	0.0857
3	1 / 17	0.9412	0.7619	0.0929
4	2 / 16	0.8750	0.6667	0.1029
5	2 / 14	0.8571	0.5714	0.1080
8	4 / 12	0.6667	0.3810	0.1060
11	2 / 8	0.7500	0.2857	0.0986
12	2 / 6	0.6667	0.1905	0.0857
15	1 / 4	0.7500	0.1429	0.0764
17	1 / 3	0.6667	0.0952	0.0641
22	1 / 2	0.5000	0.0476	0.0465
23	1 / 1	0.0000	0.0000	0.0000

Kaplan-Meier の推定量 (第2群)

生存期間	死亡数/生存数	当該生存率	生存率(SR)	SRの標準誤差
6	3 / 21	0.8571	0.8571	0.0764
7	1 / 17	0.9412	0.8067	0.0869
10	1 / 15	0.9333	0.7529	0.0963
13	1 / 12	0.9167	0.6902	0.1068
16	1 / 11	0.9091	0.6275	0.1141

22	1 / 7	0.8571	0.5378	0.1282
23	1 / 6	0.8333	0.4482	0.1346

打切り症例数 = 12

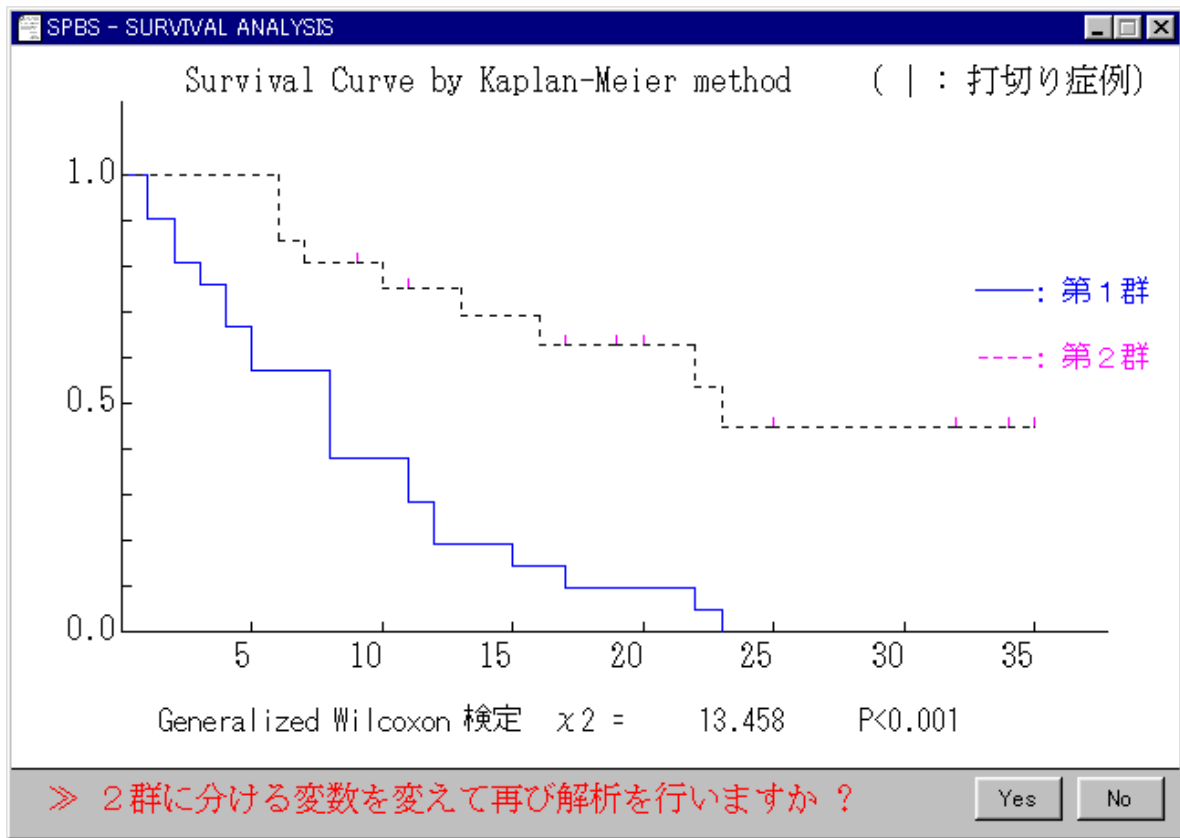
Generalized Wilcoxon 検定の結果

順位統計量 = -271 分散 $V_w = 5457.112$
 $z = 13.458$ $P < 0.001$

全期間を通じて 第2群の生存率 > 第1群の生存率

Log-Rank Test (対数順位検定) の結果

順位統計量 = 10.251 分散 $V_w = 6.257$
 $z = 16.793$ $P < 0.001$ (相対危険 = 4.179, 95%信頼区間 2.109 ~ 8.281)



Cox の比例ハザードモデル

では Newton-Raphson 法を用いているので、時折死亡期間を説明する変数の寄与度を計算している間に (解析結果が収束しないで) 終了してしまうことがある。また、説明変数は最低 1 個以上必要とする (説明変数は比例ハザード性がある変数を使用しなければならない。これについては、Kaplan-Meier 法による生存曲線の推定の項を参照)。まず、最初にモデルが有意であるかどうかを尤度比検定でおこなう。次に、その説明変数が有意かどうかについて Wald 検定 (χ^2 検定) する。個々の説明変数が有意であっても、モデルが有意でない時には割り引いて考える必要がある (多項目による多重性の問題)。下の例では、術中照射が膀胱手術の生命予後に有意な影響を及ぼしていると考えられるが、モデルが有意でないので、解析結果を強調できるかどうかやや疑問である。

[Cox's Proportional Hazard Model]

A: 膀胱癌

使用変数名 解析の症例数 = 83
 変数 1 手術時の年齢
 変数 2 性別 (0=男性)
 変数 3 術中照射の有無

変数 4 占拠部位 (0=頭部のみ)
 変数 5 膵内胆管への浸潤の程度

説明変数間における推定分散共分散値の行列

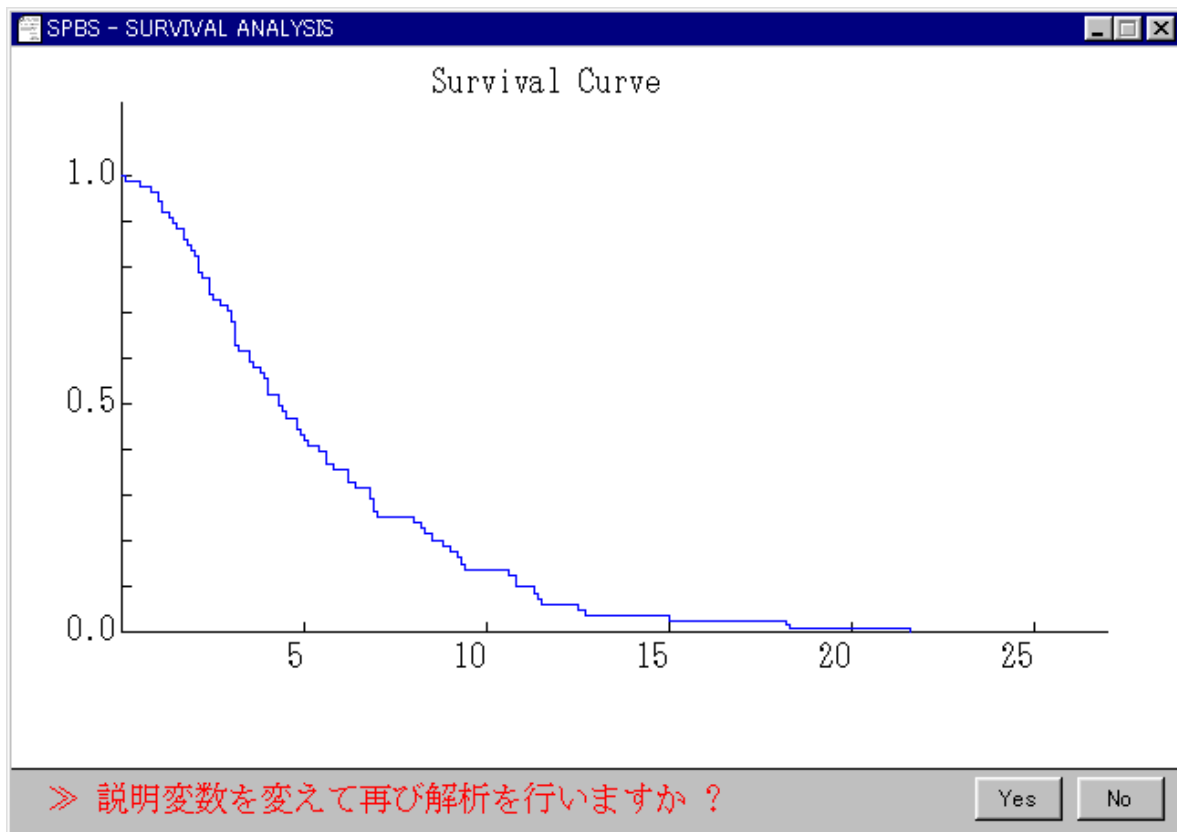
	1	2	3	4	5
1	0.000166849				
2	0.000038433	0.0529083			
3	-0.000733327	-0.000602497	0.0939419		
4	0.000129395	-0.0050793	-0.0027152	0.0839605	
5	0.000022833	-0.0042760	0.0145839	0.0079209	0.0103415

Proportional Hazard Model による推定 (Wald 検定)

変数	推定値()	標準誤差	推定値 / 標準誤差	2乗値	有意性	exp()	exp()の95% C.V.
1	0.0141872	0.0129170	1.098	1.206	P=0.272	1.014	(0.989 ~ 1.040)
2	-0.0290901	0.23002	-0.126	0.016	P=0.899	0.971	(0.619 ~ 1.525)
3	-0.83773	0.30650	-2.733	7.471	P=0.006	0.433	(0.237 ~ 0.789)
4	0.50106	0.28976	1.729	2.990	P=0.084	1.650	(0.935 ~ 2.912)
5	-0.0385508	0.10169	-0.379	0.144	P=0.704	0.962	(0.788 ~ 1.174)

全尤値 (共変量によらない対数尤度) 570.430
 偏尤値 (共変量からの最大対数尤度) 560.160 (AIC = 570.16)

モデルの尤度比検定 (-2 LOG L) 10.269 (自由度 = 5) P=0.068



19. クロス集計と繰返しのある二元配置分散分析

個人の各種データを含む大きなデータファイルから、特定の 변수により細分類した新しいデータファイルを作成する。例えば、広範囲の年齢層からなるデータファイルから年齢階級毎のデータファイルを作成することができる。これにより、「グル-ブ間の有意差検定」による解析をおこなうことができる。

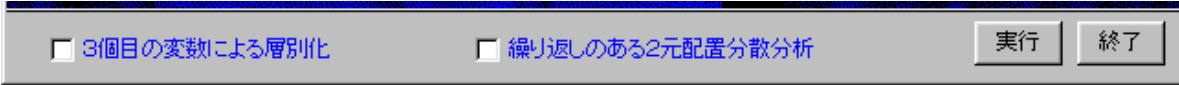
最初に、どの 변수を基準に分類するのか尋ねてくる。次に、何区分するのか尋ねる（最大 99 区分まで可能）。その後、各区分の境界値（各区分に含まれる上限値）を小さい方から順次入力する。この基準変数で分類したデータを別々のデータファイルとして保存できる。

上述の基準変数を設定した後で、2 番目の 변수を指定することにより 2 変数間のクロス集計表を作成する（但し、2 番目の 변수の分類は最大 14 区分まで）。このクロス集計表の作成は、データファイル内の（基準変数を除く）全変数で一括して自動的に実行することも可能である（但し、数値が 15 種類以上ある 변수の時はスキップする）。また、クロス集計では²検定をおこなう（表中の縦横一列が全て 0 の場合中断）。

*** クロス集計表 ***

変数番号 1	変数番号 2	1	2	3
第 1 区分	(29)	7	9	4
第 2 区分	(39)	7	5	6
第 3 区分	(49)	9	6	10
第 4 区分	(59)	8	6	9
縦列の総計		31	26	29

[注] ()内は各区分の上限値を示す
 [注] 変数番号 2 = 薬剤
 [注] 変数番号 1 = 年齢 2 値 = 3.596 (自由度 6) 危険率 P=0.731



クロス集計の後で、連続データを含む 3 番目の 변수で繰返しのある二元配置分散分析（model は「Type の平方和」を使用）をおこなうことができる。この場合、2 番目の分類が 8 区分以下で、セル内に含まれるサンプルが（標準偏差の算出のため）2 以上必要となる（欠損値 < -99, -999 > が含まれている場合には除外して計算するが、セル内の数が 2 未満になると計算できない）。繰返しのある二元配置分散分析で交互作用が有意となった場合には、層別化データ（各区分のデータ）毎に一元配置分散分析をおこなう必要がある（プログラムは「有意差検定」の中の「一元配置分散分析」を用いる）。なお、繰返しがない場合には「二元配置分散分析（3 群以上の対応のある有意差検定）」を用いる。

○ 2 元配置分散分析結果： 変数 3 - 血中濃度

	変数 2 -1	変数 2 -2	変数 2 -3
変数 1 - 1	7	9	4
平均	29.714	28.444	21.750
標準偏差	4.786	2.963	6.397
変数 1 - 2	7	5	6
平均	30.714	28.400	20.333
標準偏差	3.638	3.130	2.582
変数 1 - 3	9	6	10
平均	27.667	25.333	18.300
標準偏差	5.339	7.394	4.547
変数 1 - 4	8	6	9
平均	21.250	21.333	12.778
標準偏差	7.025	9.626	4.438

分散分析表 (二元配置分散分析)

	平方和	自由度	不偏分散	F 値
モデル	2596.89647	11	236.08150	8.178 P<0.001
変数 1 間	939.05746	3	313.01915	10.843 P<0.001
変数 2 間	1367.72068	2	683.86034	23.690 P<0.001
交互作用	20.03870	6	3.33978	0.116 P=0.994
残 差	2136.18492	74	28.86736	
全 体	4733.08140	85		

Model は「Type の平方和」を使用

なお、最大 3 個の変数で分類することができる。この他、データファイル入力時に何回でも「データ制限」(特定の変数内の最小値および最大値をキーボードで入力することにより当該範囲のデータのみを抜き出す操作)をおこなうことができる。

* * クロス集計表 * *

変数番号 2	変数番号 3	1	2	3	4
	(29)	(39)	(49)	(59)	
第 1 区分 (1)	87	83	263	106	
第 2 区分 (2)	97	56	99	90	
縦列の総計	184	139	362	196	

[注] () 内は各区間の上限値を示す

[注] 変数番号 3 = 年齢

[注] 変数番号 2 = 性別(1=M 2=F)

2 値 = 39.307 (自由度 3) 危険率 P<0.001

○ 3 個目の変数による層別化: 変数 4 - 喫煙状況(1=Smoker 2=Nonsmoker)

<区分値 1 >

	変数 3 -1	変数 3 -2	変数 3 -3	変数 3 -4	
変数 2 - 1	44	49	151	54	298
変数 2 - 2	10	7	12	8	37
	54	56	163	62	335
2 値 =	5.659 (自由度 3) 危険率 P=0.129				

<区分値 2 >

	変数 3 -1	変数 3 -2	変数 3 -3	変数 3 -4	
変数 2 - 1	43	34	112	52	241
変数 2 - 2	87	49	87	82	305
	130	83	199	134	546
2 値 =	20.236 (自由度 3) 危険率 P<0.001				

20. 患者対照研究〔離散量デ - タのみ〕

患者・対照研究をおこなう際には、まず患者群と対照群を定義し、次に推定原因因子を定義する。この後幾つかの交絡因子に配慮しながら、患者群と対象群の間で推定原因因子の曝露頻度を比較する。また、コホート研究で“曝露集団が非曝露集団と比べ当該疾患への発症率が何倍であるか（相対危険）”に相当する Odds 比を算出し、推定原因因子と疾患との関連の強さを推定する。このプログラムでは、比率検定と関連の強さに関する統計的処理のみをおこなう。実際の研究では、患者と対照群の設定方法の妥当性、信頼性、サイズの決定、交絡因子の除去等々、幾つかの問題点を解決しなければならない。そして、この対象および推定原因因子の設定が不十分だと、最終的な算出結果である Odds 比が意味を持たなくなる場合も出てくる。

(1) 対応のない場合 (M x 2 表、²検定、Odds 比) の計算

患者群と対照群とで、影響因子により出現頻度が異なるかどうかを検定する。使用する数値は 0 から 8 までの整数に限られる（これ以外の数値は解析不可）。

まずデ - タの入力では、デ - タファイルが 1 つなのか 2 つなのか尋ねてくる。解析する 2 つの変数番号（対照群と患者群）を入力すると、M x 2 表を作成し、² 値を算出する。さらに、0 と 1（または 1 以上）のカテゴリ - で ² 値、Odds 比を算出する。なお、データファイルの変数構造は第 1 群、第 2 群とも同一でなければ解析ができない。

データファイルが 1 つの場合には、2 通りの実行の仕方がある。1 つは、対応がなくともサンプル数が等しく、患者群データと対照群データが異なった変数番号の所に存在する場合である。この時は「データを分割」しないで次のステップに進めば、プログラムは各々の変数番号を尋ねてくる。もう 1 つは、同一ファイルに患者群と対照群を区別する変数と実際に解析に使用する変数（データ）が含まれている場合である。後者の場合には「データを分割する」変数番号を指定して、（同じ配列の 2 つのファイルを作成したかの如くに）以下の ² 検定をおこなうことができる。

なお、対照群、患者群の代わりに非曝露群、曝露群のデータで、発症有（+）・無（-）の情報がデータファイルに入力されていた（コホート研究）場合には、その相対危険（relative risk）および 95% 信頼区間も算出する。この場合、対照群 = 「非曝露群」、患者群 = 「曝露群」を対応させないと、相対危険度（RR）の値が間違ってしまう。

Case-control study

A: ¥対照群

A: ¥患者群

第 1 の変数番号（対照群）： 5 喫煙本数（1 日当り）

第 2 の変数番号（患者群）： 5 喫煙本数（1 日当り）

患者対照研究 [(-) = 0, (+) = 1 or over]

カテゴリ - 番号	対照群	患者群
(-)	15	4
(+)	12	23
2 値 =	9.826	P=0.002
2 値(Yates)=	8.120	P=0.004
自由度 =	1	
Odds 比 =	7.19	[95% 信頼区間 1.95 ~ 26.51]

< Cohort study の場合 >

相対危険(RR) = 1.92 [95%信頼区間 1.22 ~ 3.01]

(2) 対応のある場合 (2 x 2 表、²検定、Odds 比) の計算

患者群と対照群が 1 対 1 に対応する場合に、影響因子が両群の出現頻度に関係しているかどうか検定する（ここでは McNemar 検定を用いる）。対応のある検定では、2 x 2 分割表の各セルの総数は患者 - 対照群のペア - の数になる。操作法は「対応のない場合」と同様である。

データ入力の際に、曝露の有無を逆に入れてしまっている場合がある（例えば 0=曝露有、1=曝露無）が、この時は『最初の区分が「影響因子なし（-）」、あとの区分が「影響因子あり（+）」ですか？』と尋ねてきた時に、「n」を入力すると計算の際に修正する（通常は「y」を入力する）。

Case-control study

E: ¥reidai ¥対照群

E: ¥reidai ¥患者群

第1の変数番号(対照群) : 5 喫煙本数(1日当り)
 第2の変数番号(患者群) : 5 喫煙本数(1日当り)

対応のある患者対照研究

		対 照	
		(-)	(+)
患 者	(-)	1	3
	(+)	14	9

(-):影響因子なし...[カテゴリ-番号] 0
 (+):影響因子あり...[カテゴリ-番号] 1

2 値 : $\chi^2 = 5.882$ P=0.015
 McNemar exact test (2項分布)による直接確率 P = 0.0127

Odds比 = 4.67
 [95% 信頼区間 1.30 ~ 25.33]

(3) 相乗作用指数の算出

ある結果因子(例えば喉頭癌)に対し、2つの原因因子(例えば喫煙と飲酒)が考えられる場合がある。このような時、2つの原因因子がお互いに全く独立している場合には、相加作用(2つの相対危険の和)があると言えるが、相乗作用がある場合には単純な相対危険の和でなくなる。2つの原因因子の相対危険(Odds比)から相乗作用指数(および95%信頼区間)を算出し、有意な相乗作用があるのか否かをこのプログラムは検討する。統計的な考え方 Rothman(1986)の本に記載されている。

データが[0=(-)、1=(+)]からなる患者群用および対照群用データファイルがある場合で、それら2つのデータファイルの変数名(項目名)の配列(サンプル数と変数)が同一のときに限りディスクからの入力が可能である。また、4分表のデータをキーボードから直接入力することによっても計算できる。相乗作用指数とその95%信頼区間(後者の結果はプリントのみ)が算出される。

相乗作用指数の算出 (患者群) B: ¥患者群
 (対照群) B: ¥対照群

要因Aの変数名: 喫煙本数(1日当り)
 要因Bの変数名: 飲酒量(1週間当りの100ml)

		患者群 (Case)		対照群 (Control)	
		要 因 A		要 因 A	
		(-)	(+)	(-)	(+)
要因B	(-)	2	9	7	7
	(+)	2	14	8	5

各セルのOdds比

		要 因 A	
		(-)	(+)
要因B	(-)	1	4.5
	(+)	0.9	9.8

相乗作用指数 = 2.6 (90%信頼区間 0.5 ~ 13.3; 95%信頼区間 0.4 ~ 18.1)

(4) Mantel-Haenszel 法による対応のない χ^2 検定

2×2 分割表の χ^2 検定の特殊な形として、Mantel-Haenszel 検定がある。これは同じ 2×2 の χ^2 検定でも、層別化されたデータの場合に使用される。例えば、癌の発生は多くの場合加齢とともに発生頻度が高くなるので、30 歳から 79 歳までの集団で有害因子の曝露の有無と癌の発症の有無で 2×2 分割表を作成して通常の χ^2 検定をおこなうと、患者および対照集団間の年齢差を無視した結果になる可能性がある。すなわち、対照群には若年者が多数を占め、患者群には高齢者が多数占めているような状況もありうる。したがって、このような場合には年齢による補正が必要になる。

Mantel-Haenszel 検定は、上述の例では年齢階級別（例えば 10 歳階級毎）に 2×2 分割表を作成し、全体として有害因子の曝露影響が癌の発生に関連していたのかを判定する（自由度 1 の χ^2 検定）。結果として、層別化に使用した変数の影響を取り除いたことになる。このプログラムでは Mantel-Haenszel 法による Odds 比およびその 95% 信頼区間が Robins et al (1986) の方法により算出される。

データファイルが 1 つの場合は患者 - 対照の 2 群に分ける変数と層別化に使用する変数、および検定に使うカテゴリーデータ変数が必要となる。またデータファイルが 2 つの場合には、変数の配列が全く同一のデータファイルでなければ計算ができない。結果として、上記の他に層別化した時の 2×2 分表、総計の χ^2 検定および Odds 比（95% 信頼区間を含む）が出力される。

Case-control study

2 群に分割した変数は 19 番です

層別化変数は 3 年齢

Mantel-Haenszel法

発症因子の変数番号 11 職場ストレス

	対照群	患者群
発症 (-)	X1	X3
因子 (+)	X2	X4

この境界値(上限値) 3

第 1 層別化データ (層別化の上限値 29)

	72	57
	31	5

第 2 層別化データ (層別化の上限値 39)

	60	54
	14	1

第 3 層別化データ (層別化の上限値 49)

	131	122
	24	6

第 4 層別化データ

	62	65
	20	2

累計

	325	298
	89	14

Mantel-Haenszel法による χ^2 値 = 39.263 P<0.001 (df=1)

Mantel-Haenszel法によるOdds ratio = 0.176

Odds ratioの95%信頼区間 0.098 ~ 0.316

総計の χ^2 (Yates) = 40.900 P<0.001

総計の Odds比 = 0.17 [95%信頼区間 0.10 ~ 0.31]

2.1. 重回帰分析

例えば、血圧を規定する因子は多数（例えば、ホルモン、心拍出量、循環血液量あるいは腎排出量）ある。これに対し、これまでは1つの因子がある状態を規定するという仮定のもとに、関連性が調べられてきた。単純な現象は後者の仮定で説明できたかもしれないが、幾つかの因子がかかる規定因子として挙げられた時、この中のどれが最も影響するのかを決めるのは容易ではない。重回帰分析は、ある状態（これを従属変数Yとする）とこれを説明すると考えられる因子（これを独立変数 X_i とする）と以下の関係が成立すると仮定して、最小2乗法により最良不偏推定値（不偏で分散が最小な推定値）を求める手法である。このとき偏回帰係数を算出し、どの変数が有意な影響を及ぼすかをt検定により推定する。

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n \quad (B_i \text{ は偏回帰係数})$$

データ入力後、まず相関分析表がプリントされる。次に、変数増減法をおこなうかどうか尋ねてくる。それから、それぞれの変数が従属変数、独立変数あるいは使用しない変数のいずれであるか尋ねる。この過程の中で、自分の目的とする方法になるよう入力する。

重回帰分析に使用できる独立変数は基本的に連続量であるが、2値変数（0, 1）も使用可能である。この場合、実測値（従属変数）と予測値（回帰モデルで算出される値）の差（残差）が正規分布しなければならない。SPBS 統計パッケージでは、この残差が正規分布するかどうか適合度検定を用いてチェックするので、解析の最後（Durbin-Watson 比の後）に出てくる「残差の正規性検定」の結果を見る必要がある。

解析方法	<input checked="" type="radio"/> 変数固定法	<input type="radio"/> 変数増減法 (P=0.05)
従属変数の指定	<input checked="" type="radio"/> 選択された全変数	<input type="radio"/> 任意の変数
独立変数の指定	<input type="radio"/> 選択された全変数	<input checked="" type="radio"/> 任意の変数
		実行 終了

データが時系列データであるときには、Durbin-Watson 比を検討する必要がある（奥野忠一ほか、1983）。なお、SPBS 統計パッケージでは「観測値、予測値および残差を算出しますか？」と尋ねてきた時に<はい(Y)>を選択すると、残差のプロット（基準化したデータ）とともに「連(run)の数による標準正規分布近似値とそのP値」を解析結果として示す（下の例題では省略されている）。

重回帰分析をおこなう時、多重共線性の問題が出てくる。特に、独立変数間の相関係数が高い変数があるときに、この問題が生じやすくなる。解析結果の中で、独立変数と従属変数の相関係数が重回帰分析で算出される偏回帰係数と符号（正・負）が異なっている場合には、独立変数の中に強い相関をもった別の独立変数が含まれている可能性があるため、独立変数の選択にあたって十分検討する必要がある。

(1) 一般形の重回帰分析

従属変数を説明すると考えられる全変数（独立変数）を入れて解析する。この場合、各独立変数の偏回帰係数、標準化回帰係数、そのt値および全変数による重相関係数、寄与率、Durbin-Watson 比を算出する。また必要なら観察値、予測値および残差を算出する。これらの意味は参考文献を参照されたい。

(2) 増加法による重回帰分析

重回帰分析の独立変数を順次増やしていく方法である。最初に従属変数と最も相関の高い独立変数を1つ入れ（相関行列を必ず見ること）、以後は分散分析表が算出されるたびに、〔残りの独立変数の回帰への寄与〕を見ながら最もt値の大きい変数番号を1個ずつ入力する（偏回帰係数が危険率5%で有意な変数には*印が画面に表示される）。なお、独立変数を増やす基準は解析者が決める。

(3) 減少法による重回帰分析

重回帰分析の独立変数を順次減らしていく方法である。最初に独立変数を全部入れ、以後は〔標準回帰係数〕の横にある〔t値〕を見ながら回帰への寄与の小さい変数を減らす。但し、変数に対してサンプル数が少ない時は自由度が負になってしまい、計算が中断することがある。

(4) 増減法による重回帰分析

重回帰分析の独立変数を順次増やしたり減らしたりして、重回帰式の寄与率が高くなるように独立変数を取捨選択する方法である。本プログラムでは、取捨選択基準が危険率5%の場合に限り自動的に解析できるように設計されている（本プログラムが起動した時に、これをおこなうかどうか尋ねる）。また、そのほかの基準で解析する場合には相関行列および画面に表われるt値から順次変数の取捨選択を自分でおこなう。

Basal Statistics

sample number= 60

V-No.	Mean	S.D.	Min.	Max.	Names of variables
2	43.233	10.067	23.000	58.000	年齢
3	136.933	16.670	102.000	172.000	収縮期血圧
5	197.867	30.225	133.000	260.000	総コレステロール
6	137.800	87.598	45.000	538.000	中性脂肪
7	29.650	15.131	12.000	78.000	GOT
8	39.917	38.300	8.000	262.000	-GTP
9	98.683	14.150	67.000	143.000	FBS
10	23.672	2.151	19.800	28.200	BMI

Correlation Matrix [: P value : Correlation coefficient]

	2	3	5	6	7	8	9	10
2		0.401	0.224	0.203	0.162	0.124	0.293	0.129
3	0.002		0.108	0.154	0.339	0.128	0.244	0.043
5	0.085	0.413		0.066	0.213	-0.179	-0.065	0.258
6	0.120	0.241	0.619		0.231	0.195	-0.127	0.183
7	0.216	0.008	0.103	0.076		0.428	0.199	0.288
8	0.345	0.329	0.172	0.135	<0.001		0.223	0.198
9	0.023	0.060	0.620	0.335	0.128	0.087		0.119
10	0.325	0.744	0.046	0.161	0.026	0.130	0.364	

従属変数の番号 3 --- 収縮期血圧

分散分析表

	平方和	自由度	不偏分散	F 値
回帰	4265.07	7	609.2953	2.612
残差	12130.67	52	233.2820	
全体	16395.73	59	277.8938	

重相関係数 (R) 0.51003 P=0.022
 自由度調整済重相関係数 (R*) 0.40067

寄与率 0.26013

回帰式	標準回帰係数	t 値	[変数名]
+0.54042 * X(2)	0.32634	2.445 (P=0.018)	年齢
-0.00758 * X(5)	-0.01375	-0.100 (P=0.921)	総コレステロール
+0.01209 * X(6)	0.06352	0.486 (P=0.629)	中性脂肪
+0.33833 * X(7)	0.30709	2.135 (P=0.038)	GOT
-0.02890 * X(8)	-0.06641	-0.465 (P=0.644)	-GTP
+0.14247 * X(9)	0.12093	0.901 (P=0.372)	FBS
-0.75099 * X(10)	-0.09690	-0.742 (P=0.461)	BMI
+108.244			

Durbin-Watson 比 = 1.84291

残差の正規性の検定 2 値 = 2.476 P=0.871 (正規分布仮説は妥当)

2.2. 主成分分析

主成分分析は、相互に相関のある多種類の特性値のもつ情報を、相互に無相関な少数個の総合特性値に要約する手法である。実際のデータ解析をおこなうときに考慮すべき事項は成書（奥野忠一ほか、1983）で詳しく述べられているので参照されたい。

入力方法は、データファイルからの場合と相関係数をキーボードより入力する場合の2通りが選択できる。後者の場合は、主成分得点を算出することができない。前者の場合には、特定変数の中の或る範囲に属する数値データのみを用いて主成分分析することも可能である。

このプログラム起動時に、主成分負荷量の算出の後に Varimax 回転をするかどうかを選択する。この回転をおこなうことにより、因子の解釈が容易になる（一方、回転する因子数を変えられることより、任意性があることになる）。すなわち、各変数の共通性（主成分負荷量の2乗総和）を変えないようにして、各成分がもっとも分離されやすい（最大になる）ように直交回転させる。

2つの主成分（因子）負荷量および主成分（因子）得点をX-Y平面に描くことができる。また、主成分（因子）得点は、データファイルとしてディスク内に保存することもできる。

主成分分析と因子分析のどちらを使用すべきかという点に関しては、研究者の間で意見が分れている。前者は一義的に解が決定してしまうのに対して、後者は使用する手法（主因子法、最尤法、直交回転、斜交回転等々）により異なった解を得ることができる反面、恣意的になり易い点が指摘されている。



Principal component analysis

E:¥spbs¥身体計測

サンプル数 = 50

(Correlation Matrix) ----- * P<0.05 ** P<0.01 ¥¥ P<0.001

	1	2	3	4	5	6	7	8	9	10	11
1	1.000¥¥										
2	0.387**	1.000¥¥									
3	0.178	0.148	1.000¥¥								
4	0.046	0.255	0.306*	1.000¥¥							
5	0.070	0.084	0.424**	0.479¥¥	1.000¥¥						
6	0.139	0.274	0.446**	0.412**	0.397**	1.000¥¥					
7	0.072	0.282*	-0.010	0.233	-0.191	-0.054	1.000¥¥				
8	0.050	0.296*	0.302*	0.525¥¥	0.372**	0.323*	0.192	1.000¥¥			
9	0.010	0.223	0.283*	0.409**	0.276	0.612¥¥	0.075	0.233	1.000¥¥		
10	-0.173	0.078	0.207	0.579¥¥	0.413**	0.246	0.055	0.626¥¥	0.394**	1.000¥¥	
11	-0.147	0.173	0.047	0.411**	0.056	0.265	0.398**	0.405**	0.456¥¥	0.356*	1.000¥¥

主成分	固有値	寄与率%	累積寄与率%	2値	自由度	有意性
1	3.761	34.195	34.195	175.81	65	P<0.001
2	1.586	14.417	48.612	88.520	54	P=0.002
3	1.508	13.713	62.325	66.765	44	P=0.015
4	1.037	9.424	71.749	36.487	35	P=0.399
5	0.705	6.408	78.156	19.830	27	P=0.838
6	0.553	5.030	83.186	14.497	20	P=0.804
7	0.498	4.526	87.712	14.000	14	P=0.450
8	0.436	3.960	91.672	10.677	9	P=0.299
9	0.377	3.428	95.101	8.279	5	P=0.141
10	0.334	3.041	98.142	5.720	2	P=0.057
11	0.204	1.858	100.000			

[固有ベクトル]

変数	Vector										
	1	2	3	4	5	6	7	8	9	10	11
1	0.052	-0.194	0.653	-0.172	-0.137	0.301	-0.617	0.031	0.062	0.000	0.096
2	0.213	0.119	0.557	-0.079	-0.372	-0.316	0.555	0.030	0.218	-0.156	0.075
3	0.273	-0.367	0.124	-0.008	0.732	-0.312	-0.009	0.158	0.209	-0.270	0.030
4	0.409	0.069	-0.066	-0.182	-0.003	0.492	0.166	0.093	-0.391	-0.561	-0.213

5	0.309	-0.408	-0.144	-0.232	-0.007	0.400	0.284	-0.388	0.365	0.374	0.023
6	0.352	-0.251	0.110	0.447	-0.041	-0.124	0.035	-0.198	-0.605	0.222	0.358
7	0.113	0.585	0.272	-0.082	0.476	0.241	0.156	0.146	-0.096	0.470	0.067
8	0.376	0.114	-0.058	-0.420	-0.073	-0.467	-0.266	-0.217	-0.196	0.227	-0.484
9	0.349	-0.006	-0.020	0.590	-0.159	0.105	-0.093	0.390	0.280	0.179	-0.472
10	0.361	0.076	-0.352	-0.276	-0.217	-0.087	-0.161	0.516	0.116	0.078	0.545
11	0.292	0.468	-0.098	0.265	0.041	-0.003	-0.268	-0.537	0.338	-0.298	0.228

[因子(主成分)負荷量]

変数	主成分										
	1	2	3	4	5	6	7	8	9	10	11
1	0.102	-0.244	0.802	-0.175	-0.115	0.224	-0.435	0.020	0.038	0.000	0.043
2	0.412	0.150	0.684	-0.080	-0.312	-0.235	0.392	0.020	0.134	-0.090	0.034
3	0.529	-0.462	0.152	-0.008	0.614	-0.232	-0.006	0.104	0.128	-0.156	0.013
4	0.793	0.087	-0.081	-0.185	-0.003	0.366	0.117	0.061	-0.240	-0.324	-0.096
5	0.599	-0.513	-0.177	-0.236	-0.006	0.297	0.200	-0.256	0.224	0.216	0.010
6	0.683	-0.316	0.136	0.455	-0.034	-0.092	0.025	-0.131	-0.371	0.128	0.162
7	0.219	0.736	0.334	-0.084	0.400	0.180	0.110	0.096	-0.059	0.272	0.030
8	0.729	0.144	-0.072	-0.428	-0.062	-0.348	-0.188	-0.143	-0.120	0.132	-0.219
9	0.677	-0.007	-0.024	0.601	-0.133	0.078	-0.065	0.257	0.172	0.103	-0.213
10	0.700	0.096	-0.432	-0.281	-0.182	-0.065	-0.114	0.340	0.071	0.045	0.247
11	0.565	0.589	-0.121	0.269	0.035	-0.002	-0.189	-0.354	0.207	-0.172	0.103

【解析をおこなおうとするが、途中でプログラムが終了する時のチェックポイント】

本統計量を算出した時に、標準偏差 (SD) が 0 である変数を主成分分析の変数に選択することはできない。したがって、データファイル作成段階あるいは主成分分析の基本統計の中に SD が 0 の変数がないかどうかチェックする必要がある。

概して解析に使用する変数に比べて当該サンプル数が少ないと計算できない、あるいは計算しても精度があまりに悪い。したがって、SPBS では解析に使用する変数の数を減らし、(サンプル数) > (変数の数) となった時点で解析し始める。

変数 X と変数 Y が全く同じ数値 (あるいは単に数倍した数値) であるのに、「主成分分析」の使用変数にこれら両変数を選択している場合には解析不能となる。この場合、自分では同じ数値を入力している訳ではないにもかかわらず、たまたま得られたデータの変数 X と変数 Y で全て同じ数値になっていた例も含まれる。これを回避するためには、SPBS メインメニューの「8 . 相関分析」の「相関分析」で全変数の相関係数を算出し、単相関係数が ± 1 の変数がないかどうかチェックする。また、変数 X と変数 Y で単相関係数が ± 1 となることがわかったら、これらの変数 X か Y のいずれか 1 つのみ選択し、両者を同時に選択しないようにすると解析が可能となる。

なお、市川ら (1993) は主成分分析における解釈をおこなう際の注意点として、以下の 5 点を挙げている。

しばしば外れ値が主成分を決定することがある。すなわち、ある少数の観察データの得点のみが大きくなるような主成分が得られることがある。この場合、外れ値を除くか否かで解釈の結果が大きく変化するので、主成分に一般的な意味付けをすることはできなくなる。

可能な限り、観察データの選択に対する結果の安定性を調べるべきである。この最も簡単な方法としては、データを分割して解析を繰り返してみることである。

相関行列、分散共分散行列、平方和積行列のうちどれを用いるか、さらには後二者の場合には各変数の測定単位をどうするかということによって解析結果は変化する。どの方法を用いるかは研究者の考え方によるが、SPBS 統計パッケージでは相関行列を用いる方法しか選択できない。

同じ変数を次々に重複させ付け加えていくと、第 1 主成分得点はこの変数に収束していく。このように、主成分は用いる変数の選び方に依拠する。

主成分得点を散布図にプロットしてみると、個々の変数の視察では見出せなかった外れ値が明らかになることがある。最終的な結論を導くというより、データの予備的吟味として主成分分析を利用することは、無難でかつ有力な方法である。

23. 因子分析

因子分析ははじめ心理学者が考案し、人間に対する各種の心理テストの結果から、その人の潜在能力あるいは潜在知能指数の値を推定することが目的で作られた。この算出方法は幾通りかあるが、本解析では主因子法を用いている。この方法は共通性を重相関係数の2乗(SMC)で求めた後、反復推定により潜在因子を算出する。さらに、回転前の因子負荷量を Varimax 法で回転し、因子負荷量を算出する。因子得点は、必要ならば算出し、またデータファイルとして保存することもできる。なお、因子負荷量は最大17因子までしか算出しない。

最初の手順は主成分分析の場合とほとんど同様であるが、途中で抽出因子数を尋ねてくる。この時、相関行列の固有値とSMC法による固有値の値を見て、抽出因子数を入力する(固有値が1以上である個数を指定することが多い)。無入力状態でエンターキーを押すと、このプログラムで設定した個数が自動的に当てはめられる。

因子の反復推定あるいは Varimax 回転の際、解が収束しない場合が度々ある。また、因子分析の計算途中で共通性(Communality)が1以上になると、計算を打ち切り、抽出因子数の変更あるいは使用変数の変更を要求する(共通性が1以上の変数がある場合、下の因子負荷量のプロットを描いた際に半径1の円からはみ出てしまいます)。この時は、

抽出因子数を増減して再度計算をおこなう、解析に使用する変数を変えておこなう、「主成分分析」で Varimax 回転した結果を用いる、のいずれかを選択せざるを得ない。なお、使用変数が20個以上の場合、計算に多少時間がかかる。



因子負荷量および因子得点のプロットはいずれも可能である。前者は「ペイント」だけでなくプリンター印刷も可能であるが、後者は「ペイント」にしかプロットしない。

解析手順は部分的に田中豊ほか(1984)を参照している。更に詳しく知りたい方は Dixon & Masey (1983) の本を参照されたい。

(Correlation Matrix)

	1	2	3	4	5	6
1	1.000¥¥					
2	-0.042**	1.000¥¥				
3	0.032*	-0.238¥¥	1.000¥¥			
4	-0.031*	-0.130¥¥	-0.557¥¥	1.000¥¥		
5	-0.042**	0.071¥¥	-0.470¥¥	0.594¥¥	1.000¥¥	
6	0.001	0.297¥¥	-0.129¥¥	-0.016	0.124¥¥	1.000¥¥

* P<0.05 ** P<0.01 ¥¥ P<0.001 (自由度 = 5078)

対角要素にSMC(Squared Multiple Correlation)と相関係数行列を用いた時の固有値の算出

	対角 S M C (非反復解法)	相関行列
1	1.5577	2.1197
2	0.5612	1.3416
3	0.0058	0.9986
4	-0.0320	0.7400
5	-0.1622	0.4948
6	-0.2656	0.3054

抽出する因子数 = 2

主因子法による因子負荷量と共通性

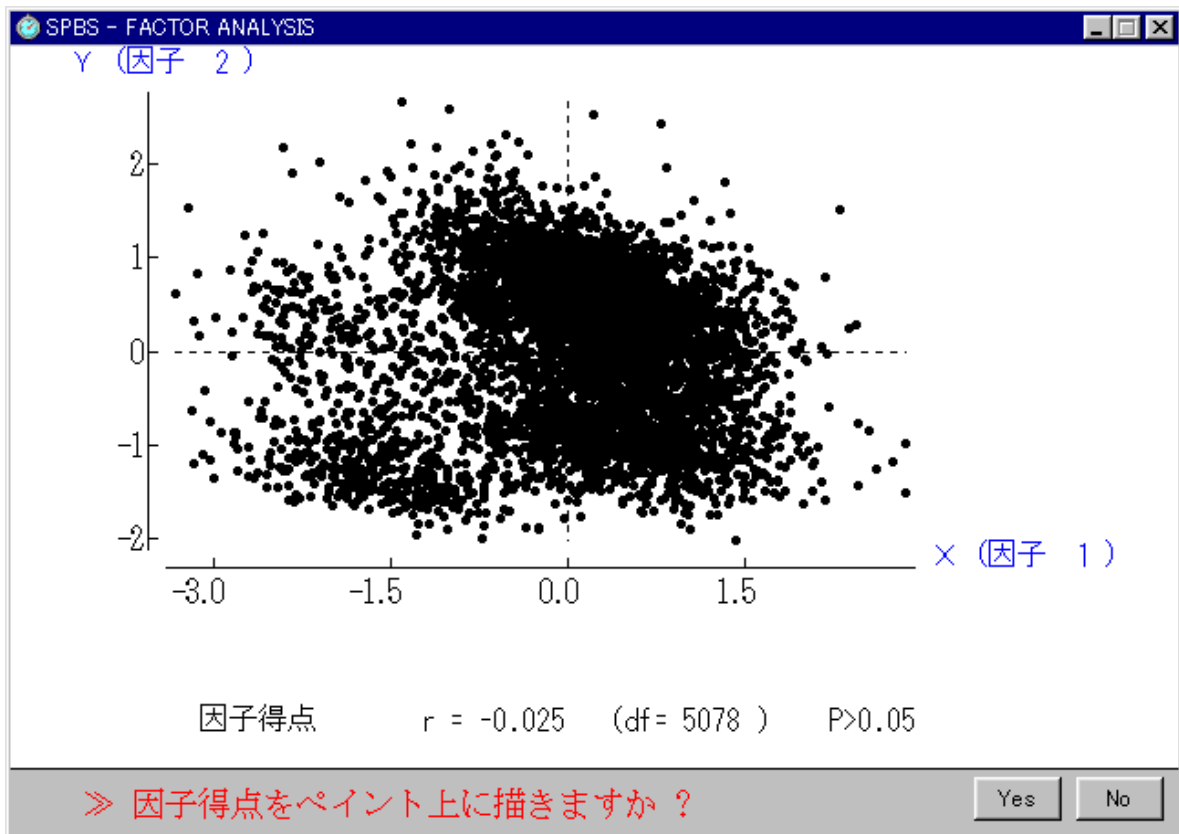
[Varimax回転前の因子負荷量]

変数	1	2	共通性
1	-0.048	-0.033	0.003
2	0.126	0.798	0.653
3	-0.693	-0.177	0.511
4	0.877	-0.308	0.865
5	0.680	0.004	0.463

6	0.123	0.354	0.140
男女	1.746	0.889	

[Varimax回転後の因子負荷量]

変数	1	2	共通性
1	-0.042	-0.040	0.003
2	-0.005	0.808	0.653
3	-0.655	-0.286	0.511
4	0.916	-0.162	0.865
5	0.671	0.114	0.463
6	0.064	0.369	0.140
男女	1.723	0.912	



データを層別化すべきかどうかを調べるために、因子得点をプロットすることがある。上のデータの場合には男 (=0) 女 (=1) データがあり、そのため右上と左下の2群に分離しているように見える。このような場合には男女別々に解析する方が良いと考えられる。

【解析をおこなおうとするが、途中でプログラムが終了する時のチェックポイント】

本統計量を算出した時に、標準偏差 (SD) が 0 である変数を因子分析の変数に選択することはできない。したがって、データファイル作成段階あるいは因子分析の基本統計の中に SD が 0 の変数がないかどうかチェックする必要がある。

概して解析に使用する変数に比べて当該サンプル数が少ないと計算できない、あるいは計算しても精度があまりに悪い。したがって、SPBS では解析に使用する変数の数を減らし、(サンプル数) > (変数の数) となった時点で解析し始める。

変数 X と変数 Y が全く同じ数値 (あるいは単に数倍した数値) であるのに、「因子分析」の使用変数にこれら両変数を選択している場合には解析不能となる。この場合には、自分では同じ数値を入力している訳ではないにもかかわらず、たまたま得られたデータの変数 X と変数 Y で全て同じ数値になっていた例も含まれる。これを回避するためには、SPBS メインメニューの「8. 相関分析」の「相関分析」で全変数の相関係数を算出し、単相関係数が ±1 の変数がないかどうかチェックする。また、変数 X と変数 Y で単相関係数が ±1 となることがわかったら、これらの変数 X か Y のいずれか 1 つのみ選択し、両者を同時に選択しないようにすると解析が可能となる。

24. 数量化理論 類

数量化理論 (theory of quantification) は現象解析のための方法として、実際に得られたデータから如何にして有効な情報を引き出すかという過程から導き出されてきた一連の手法の総称で、1950年頃に林知己夫によって提唱された(宮原英夫・丹後俊郎、1995)。たとえば、調査票の「はい」、「いいえ」という回答を統計的に取り扱う場合には、「はい」をX1(=1)、「いいえ」をX2(=0)で表し、これを確率変数として取り扱う。

数量化で取り扱うデータは多変量データといわれるものである。多変量データとは問題にしている事象、およびその事象を引き起こすと考えられる要因の多元的測定の結果得られるデータをいう。数量化 類はカテゴリーデータの主成分分析と考えることができ、缶詰のレツテルの好みの官能検査分析の過程で考え出された。缶詰のレツテルの好きなものに 印をつけた調査結果から、似たもの同士が近くに、似ていないものは離れるように各被験者とレツテルの両方に得点を与える。ちょうど被験者とレツテルについて縦横の表の形に並べたデータを、縦、横とも並べ替えて対角線上に反応が集まるようにする。これは両者の相関係数が最大になるように得点を与えることになる。数量化 類はこのように類似したパターンを分けるということで、パターン分類法と呼ぶこともある。

最初の画面で「全項目」か「全項目 - 1 (ダミー変数)」か、あるいは「解析データ」や解析後の「サンプル得点」を印刷するかどうかが尋ねてくる。元データが「0」と「1」のみからなるデータの場合には、「全項目 - 1」(ダミー変数)を選んで解析をおこなう。また、「全項目」を選ぶと、SASの対応分析 (correspondence analysis) と同じ解が得られる。「解析データ」は最終的に解析で使用される形(0と1からなるデータで、すなわち下の例題の「解析データ」)として表される。なお、サンプル得点の印刷をおこなわなくともサンプル得点をデータファイルとして別途保存するかどうかが尋ねてくる。

データ内の変数選択の際には、同一変数内に異なった数値の種類(数種)が2以上かつ10以下である場合のみ使用可能である。また、0と1の「解析データ」の形に変えた時に、同一サンプルの値が全て0(無反応データ)の場合には自動的に除外される。SPBS統計パッケージによる数量化 類の解析後の因子数は、(使用したカテゴリー数に依存するが)最大10個まで表示する(2~10因子)。さらに、「カテゴリーの数量」や「サンプル得点」を二次平面上(前者はプリンター印刷も可、後者は『ペイント』のみ)に描くことができる。

詳細な使用例および解釈例は宮原英夫・丹後俊郎(1995)の本にあるので参照されたい。また、解析手順は部分的に田中豊ほか(1984)を参照している。

[解析データ]	<全項目数>																			
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	1	0	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0
2	1	0	1	0	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0
3	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
4	0	1	0	1	0	1	1	0	1	0	1	0	0	1	1	0	1	0	1	0
5	0	1	0	1	0	1	0	1	1	0	1	0	0	1	0	1	0	1	1	0
6	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	0	1	1	0
7	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1	0	1
8	0	1	0	1	0	1	0	1	1	0	0	1	0	1	1	0	1	0	1	0
9	0	1	1	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0
10	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0
11	0	1	0	1	1	0	1	0	1	0	0	1	0	1	1	0	1	0	1	0
12	0	1	1	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0

数量化理論 III類									
	1	2	3	4	5	6	7	8	9
固有値	0.441	0.216	0.136	0.092	0.038	0.034	0.023	0.013	0.006
寄与率	44.10%	21.56%	13.57%	9.24%	3.84%	3.41%	2.33%	1.29%	0.55%
累積値	44.10%	65.66%	79.23%	88.47%	92.31%	95.72%	98.05%	99.33%	99.89%

[カテゴリーに付与された数量]

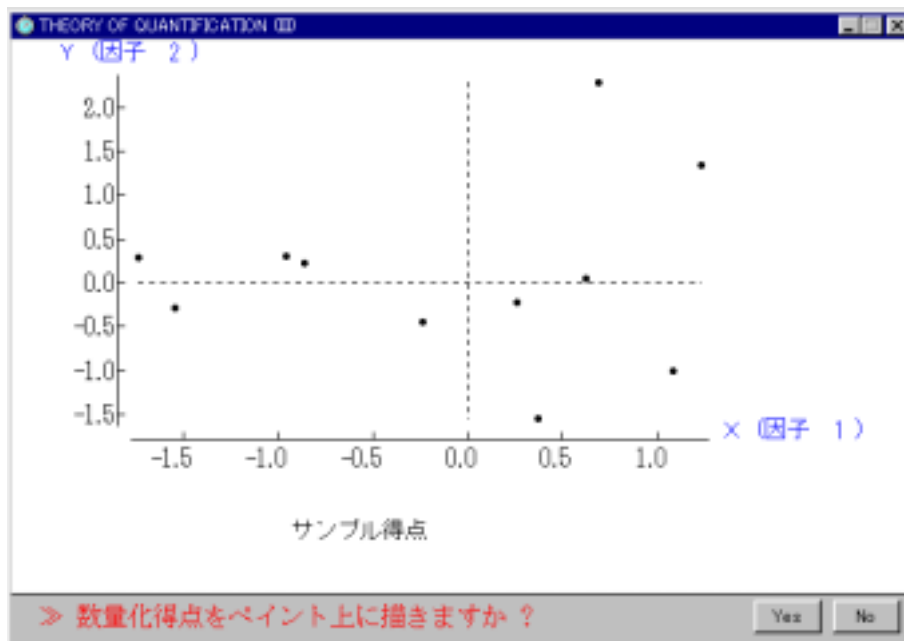
Categories / Factors										
No.	1	2	3	4	5	6	7	8	9	10
1	1.443	3.905	-0.925	-0.577	-0.976	2.936	3.205	1.421	3.052	-0.552

2	-0.289	-0.781	0.185	0.115	0.195	-0.587	-0.641	-0.284	-0.610	0.110
3	1.535	0.873	-2.087	-0.928	1.168	0.678	-0.090	0.584	-3.075	0.186
4	-0.768	-0.437	1.044	0.464	-0.584	-0.339	0.045	-0.292	1.537	-0.093
5	1.270	-2.552	-2.079	-0.174	1.587	-0.415	2.387	-1.822	1.579	1.822
6	-0.423	0.851	0.693	0.058	-0.529	0.138	-0.796	0.607	-0.526	-0.607
7	1.164	-0.856	0.431	-0.746	-1.048	-1.480	1.134	1.217	-0.317	-1.008
8	-1.164	0.856	-0.431	0.746	1.048	1.480	-1.134	-1.217	0.317	1.008
9	0.319	-0.671	0.721	-0.943	0.237	0.953	-0.367	0.390	0.167	0.306
10	-0.958	2.014	-2.164	2.829	-0.710	-2.858	1.100	-1.169	-0.501	-0.918
11	0.555	1.624	1.964	-0.888	0.622	-1.233	0.742	-1.122	-0.503	1.563
12	-0.397	-1.160	-1.403	0.634	-0.444	0.881	-0.530	0.801	0.359	-1.117
13	1.416	0.719	-0.954	-0.976	0.138	-1.287	-2.274	-0.157	1.647	-0.180
14	-1.011	-0.514	0.681	0.697	-0.099	0.919	1.624	0.112	-1.177	0.129
15	0.707	-0.077	0.106	0.999	-0.507	0.059	-0.273	0.600	-0.036	1.059
16	-2.122	0.230	-0.319	-2.997	1.521	-0.176	0.818	-1.800	0.107	-3.177
17	0.958	-0.147	0.244	0.212	-0.960	0.630	-0.148	-1.471	-0.329	-0.586
18	-1.916	0.294	-0.488	-0.424	1.921	-1.259	0.297	2.942	0.658	1.172
19	0.493	-0.002	0.425	0.623	0.907	0.031	0.013	0.110	0.106	-0.583
20	-2.466	0.009	-2.127	-3.116	-4.536	-0.157	-0.065	-0.551	-0.528	2.914

[サンプルに付与された得点]

Samples / Factors

No.	1	2	3	4	5	6	7	8	9	10
1	1.230	1.340	0.168	-1.041	-0.484	0.772	0.752	1.920	0.250	-1.199
2	0.687	2.287	-0.849	0.690	0.102	0.311	0.226	-1.597	0.204	1.162
3	-1.734	0.294	-1.175	-0.327	-1.131	-1.061	0.470	-0.749	-0.488	-1.730
4	0.257	-0.218	1.763	0.195	-0.901	-0.493	0.874	-0.117	-2.270	0.868
5	-0.953	0.312	1.215	-0.838	2.673	-0.040	0.394	-0.488	0.102	-0.514
6	-0.862	0.225	-0.367	2.217	0.611	-0.832	-0.193	1.948	0.172	0.476
7	-1.541	-0.285	-0.392	-1.568	-0.648	1.003	-0.490	0.624	0.409	1.926
8	-0.237	-0.448	0.615	1.186	-0.375	2.257	-1.446	-0.567	-0.259	-1.116
9	1.081	-1.002	-1.197	-0.389	0.649	-0.291	-0.516	-0.029	-0.684	0.031
10	0.622	0.048	1.319	-0.356	-0.780	-1.688	-1.680	-0.354	1.527	-0.055
11	0.369	-1.550	0.097	0.619	-0.365	0.353	2.126	-0.563	1.721	0.121
12	1.081	-1.002	-1.197	-0.389	0.649	-0.291	-0.516	-0.029	-0.684	0.031



25. 判別分析

似通った2つのある疾患の多数の患者について、既存の診断結果により、数種の検査データ（これらは全て連続量）がそれぞれの疾患毎に得られていると仮定する。このデータを利用して両疾患の疑いのある新しい患者に対し、このいずれかを判定したい場合にこの分析方法を用いる。従って、どちらの群に属するかを決めるため、ここでは判別関数を算出する。

Datafileの数	<input checked="" type="radio"/> One Datafile	<input type="radio"/> Two Datafiles	
変数の選択方法	<input type="radio"/> 変数増減法	<input checked="" type="radio"/> 変数減少法	<input type="radio"/> 変数固定法
判別得点	<input type="radio"/> 印刷可	<input checked="" type="radio"/> 印刷不可	

データの輸入はディスクからのみである。この場合データファイルを1個または2個作成しておかなければならない。データが1個のときは2群に分類する基準を、特定のサンプル番号まで、およびそれ以後とするか、特定の変数番号の数値で分類するか、のいずれかを選択することができる。

判別得点は解析に使用した変数（ただし、使用した変数が10以下）とともに解析結果に表示することができる。判別得点のヒストグラムは最終段階で表示する。また、サンプル数が70以上では、実際のデータを当てはめて算出した確率P値を低いものから高い順に10等分し、データと理論値が一致するか否かの検討（Hosmer-Lemeshow検定）をおこなう。

変数増減法（および減少法）の場合、線形判別関数の全変数が $P < 0.05$ の有意水準になるまで変数の増減を繰り返す。変数減少法でおこなう場合、（有意な関連をもつ変数がなければ）変数が2個で強制的に終了する。

下の例題は、月経の未潮者と既潮者の身長、体重、睡眠時間、就寝・起床時間の規則性、就寝後すぐに寝られるかどうかを調査した結果を判別分析した（元の論文では「多重ロジスティック回帰分析」を使用）。ここでは、「月経の未潮・既潮に係る要因として身長、体重、睡眠時間が影響を及ぼす」という結論になる。

Discriminat analysis

E:¥reidai¥Menarch

12 番の変数で2群に分類（境界値は .5、境界値は第1群）

各群の基本データ

	第1群	第2群	有意差検定
変数 3 ----- 身長 (cm)			
平均	140.299	156.233	t= 18.524
標準偏差	8.285	5.227	P<0.001
サンプル	134	120	df= 227
変数 4 ----- 体重 (Kg)			
平均	33.604	47.267	t= 17.871
標準偏差	5.948	6.229	P<0.001
サンプル	134	120	df= 252
変数 5 ----- 睡眠時間 (min)			
平均	484.254	421.750	t= -10.301
標準偏差	54.246	42.221	P<0.001
サンプル	134	120	df= 247
変数 8 ----- 睡眠時間の規則性 (0=regular;1=sometimes;1=irregula)			
平均	0.731	0.750	t= 0.337
標準偏差	0.445	0.435	P=0.736
サンプル	134	120	df= 252
変数 9 ----- 寝つき (0=10min;0=30min;1=30min over)			
平均	0.261	0.192	t= -1.317
標準偏差	0.441	0.395	P=0.189
サンプル	134	120	df= 252

[2群のプールした分散共分散 (自由度 252)]

	3	4	5	8	9
3	49.1330				
4	29.1125	36.9980			
5	-122.239	-66.5101	2394.87		

8	0.00296138	-0.0803127	-1.68399	0.193763	
9	-0.150057	0.109960	0.301069	0.0125118	0.176388

[分散共分散行列の検定]

2 値 = 49.25 P<0.001 (自由度 = 15)

[Discriminant Analysis]

		各係数の D^2	t 値	有意性	変数名
Z =	+0.163357 x X (3)	5.740	4.025	P<0.001	身長 (cm)
	+0.221455 x X (4)	5.421	5.018	P<0.001	体重 (Kg)
	-0.011485 x X (5)	6.101	-2.595	P=0.010	睡眠時間 (min)
	+0.110396 x X (8)	6.373	0.236	P=0.813	睡眠時間の規則性 (0=regular;1=irregula
	-0.381483 x X (9)	6.350	-0.777	P=0.438	寝つき (0=10min;0=30min;1=30min over)
	-27.9675				

Mahalanobis の平方距離 (D^2)

D^2 = 6.3751 F (5 , 248) = 79.436 P<0.001

[参考] Hotelling の T^2 : 403.59

誤分類確率 P (Misclassification probability, 1 方の群を他群と誤まる確率)

P = 10.34 % (正規分布による確率計算値)

[判別率 (観察データによる判別正当率)]

	(+)	(-)	判別率	P(z)による判別率
Group 1 :	19	115	0.858	0.822
Group 2 :	112	8	0.933	0.876
Total :			0.894	0.847

P (Z) は Logistic 関数を当てはめた時のものです

判別率の 95% 信頼区間 (近似値) : 0.856 ~ 0.932

[判別得点のヒストグラム]

```

- < X : Group 1    0 : Group 2 >
+ |0
:
8.13 |00
      |0000
      |0000000
      |00000000
      |000000000000000000000000
      |0000000000000000
      |X000000000000000000
      |XXXX000000000000000000
      |XXXXXX00000000000000
      |XXXXXXXX00000000
0 +   |XXXXXXXXXX000
      |XXXXXXXXXX00
      |XXXXXXXXXXXXXXXXXX0
      |XXXXXXXXXX00

```

```

XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXX
XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXX
-8.13 |XXXX
      :
      - |XXXXX
      -

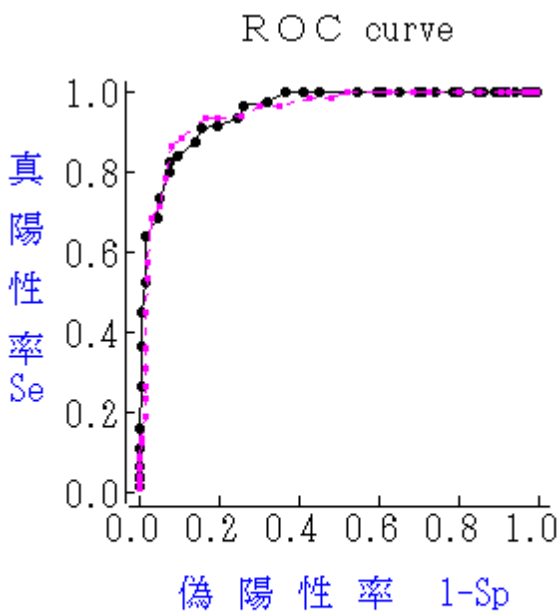
```

10分位法による検討（各分位のサンプル数は 25 個です）

分位	観察数	期待数	平均確率	0	0.2	0.4	0.6	0.8	1
1	0	0.0	0.0009	*					
2	0	0.2	0.0076	*					
3	1	0.8	0.0312	*+					
4	2	2.9	0.1172	+ *					
5	5	8.1	0.3252		+	*			
6	14	15.7	0.6284				+	*	
7	20	20.3	0.8111					*	
8	24	22.0	0.8813					*	+
9	25	22.8	0.9135					*	+
10	25	23.1	0.9236					*	+

2 値 = 2.150 P=0.828 * は期待度合、+ は観測度合を表示

参考 この集団の未潮者と既潮者を判別する身長と体重をROC曲線で調べてみると以下ようになる。
 なお、黒線(1)は身長、紫線(2)は体重を示す。



身長による判別率が最大となる値 152
 敏感度 0.825 特異度 0.925

	月経 (-)	月経 (+)
検査陰性者	124	21
検査陽性者	10	99
2値 =	142.46	P<0.001

ROC 曲線下の面積 0.951

体重による判別率が最大となる値 42
 敏感度 0.867 特異度 0.918

	月経 (-)	月経 (+)
検査陰性者	123	16
検査陽性者	11	104
2値 =	154.13	P<0.001

ROC 曲線下の面積 0.947

判別率が最大となる値(1) \geq 152
 判別率が最大となる値(2) \geq 42

26. 多重ロジスティック回帰分析

多重ロジスティック回帰分析は疫学における前向き研究によって得られたデータに基づく病因究明の方法論として開発され、危険因子 (Risk Factor) の解析に用いられる。すなわち、説明変数を慢性疾患の危険因子とした場合の、死亡あるいは発病などの事象 (外的基準) の生起に対する危険因子の寄与の程度を解析する。これには最尤法と判別分析法の2種類の解法があるが、ここでは最尤法を用いている。

最尤推定を実際にコンピュータで実行する場合その解法は大別して3種類あるが、SPBS 統計パッケージでは Newton-Raphson 法を採用している。この方法は他の解法に比べ短時間で解析を終了するが (MS-DOS 版では Walker-Duncan の論文に従って解析プログラムを作成したが、Newton-Raphson 法に比べ数十倍の時間を要するため Windows 版では断念した)、解が初期値 (プログラムの中で自動的に与えられる) に大きく影響され易い欠点があると言われている (柴田義貞ほか、1978)。また、解が収束しないで途中で解析を中断することもある。基本的な解析手順は柳井晴夫、高木広文 (1986) を参照した。

回帰モデルの有意性検定には尤度比検定を使用している。また、その説明変数が有意かどうかについては Wald 検定 (χ^2 検定) を使用している。個々の説明変数が有意であっても回帰モデルが有意でない時 (尤度比検定が $P > 0.05$) は、ロジスティック回帰モデルの適合性がよくないことを意味するので、説明変数の有意性を割り引いて考える必要がある (多項目による多重性の問題)。

説明変数が2値 (0, 1) データの場合 Odds 比とその95%信頼区間が算出される。なお、Odds 比の95%信頼区間は Wald 検定の結果に基づいて算出される (SAS では尤度比検定に基づいて95%信頼区間が算出されるが、この場合 χ^2 検定の結果と95%信頼区間の結果に整合性がなく、奇異な印象を受ける。一方の Wald 検定の場合には、原則的にデータの正規性を仮定しているが、解析結果の解釈に一貫性がある)。

2群の分割方法	<input type="radio"/> 特定のサンプル番号	<input checked="" type="radio"/> 特定変数の数値	実行	終了
解析手順	<input type="radio"/> 変数減少法	<input checked="" type="radio"/> 変数固定法		
解析データ	<input type="radio"/> 印刷可	<input checked="" type="radio"/> 印刷不可		

データを2群に分けるので、特定のサンプル番号までを第1群、残りを第2群としておくか、2群に分ける数値が入った変数を設けていなければならない。サンプル数が70以上の場合、実際のデータを当てはめて算出した確率P値を低いものから高い順に10等分し、データと理論値が一致するか否かの検討をおこなう (Hosmer-Lemeshow 検定)。

多重ロジスティック分析の理論・解説および計算の考え方は Anderson et al (1980) あるいは丹後俊郎ほか (1996) の本にあるが、更に実務的な勉強をしたい人には、Kleinbaum (1994) の本を推奨する。

変数 2 (低体重児 0=2500g 以上 1=2500g 未満) のデータで分割 (基準値は 0.5)

解析に使用する変数

- 3 --- 年齢
- 4 --- 最終月経時の体重
- 6 --- 妊娠時の喫煙 (1=Yes 0=No)

<< 第1群のサンプル数 = 59 第2群のサンプル数 = 130 >>

[Newton-Raphson 法による結果]

分散共分散行列

	3	4	6
3	0.001071		
4	-0.0000288837	0.0000376366	
6	-0.0000056819	0.0000554815	0.106196

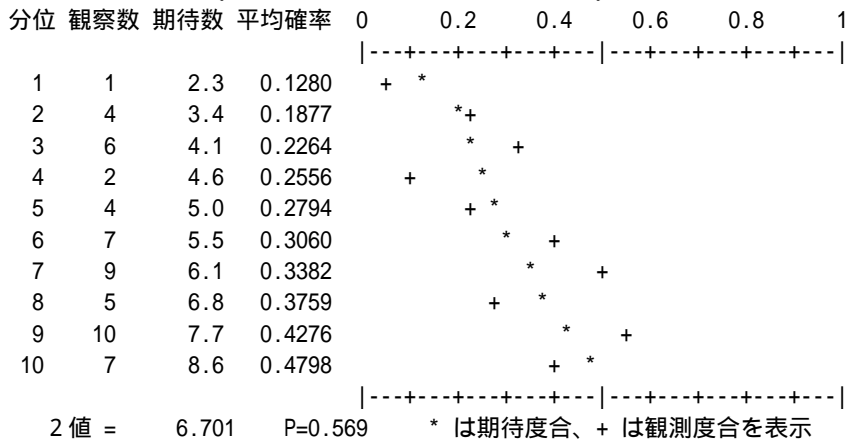
推定値、標準誤差およびその有意性 (Wald 検定)

Y =	推定値 ()	標準誤差	2 値	有意性	exp()	exp() の 95% 信頼区間
	-0.038995 * X(3)	0.032726	1.420	P=0.233	0.962	(0.902 ~ 1.025)
	-0.012139 * X(4)	0.006135	3.915	P=0.048	0.988	(0.976 ~ 1.000)
	+0.670764 * X(6)	0.325878	4.237	P=0.040	1.956	(1.033 ~ 3.704)
	+1.368225	1.014262				

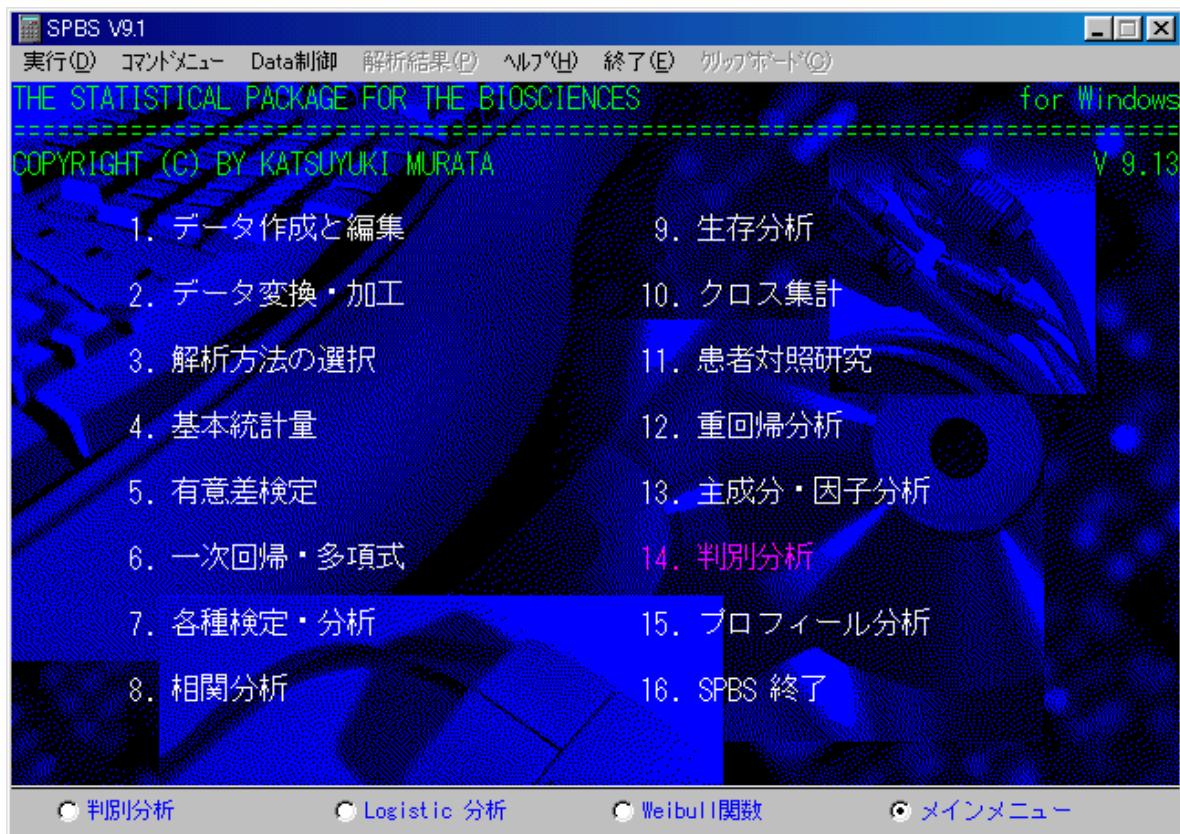
尤度 (定数項のみの対数尤度) 234.672
 尤度 (回帰モデルの対数尤度) 222.879 (AIC = 230.88)

モデルの尤度比検定 (-2 LOG L) 11.793 (自由度 = 3) P=0.008

10 分位法による検討 (各分位のサンプル数は 18 個です)



この多重ロジスティック回帰分析は『メインメニュー』の中の『判別分析』をダブルクリックすると下図が現れ、この画面下の「Logistic 分析」をクリックすると解析プログラムに移動する。



27. プロフィール分析

Morrison (1976) によって考案されたプロフィール分析は、2群のn個(3以上)の特性(変数)の各平均値のパターンが、並行であるか、もし並行ならば、どちらのパターンが有意に大きいのか、またその時各荷重平均値間に有意な差があるのか、を検討する方法である。これは変数間の差分を分散分析する方法と考えられる。

並行性の棄却は統計的に強調できる。しかしながら、並行性(Parallelism)が棄却された場合、この分析の以後の検討は全く意味を持たない。並行性が支持されたときは次の、の仮説の検討結果を参照する。

最初に検討する項目(変数)が同一に配列された2つのデータファイルを作成しておく必要がある。また、この解析にはF検定が使用されているので、2群の総サンプル数が使用する変数+3より多くなければ解析が実行されないようにプログラムされている。計算にあたっては、まず2群の各変数の有意差検定をおこない、その後に、前述の仮説をtまたはF検定により検討する。

プロフィール分析を使用するにあたっては、群内の使用変数の単位が等しいこと、群間の単位系も等しいことが条件である。但し2の場合に関しては、(群間の単位系が異なっても)各々の群で同一サンプル毎にデータを基準化すれば、「並行性」と「荷重平均値間の有意差」についてのみ検討することができると考えられる。

Profile analysis

第1群のファイル名 : D:\SPBS\reidai\自殺男

第2群のファイル名 : D:\SPBS\reidai\自殺女

各群の基本データ(対応なし)

	第1群	第2群	有意差検定
変数 1 ----- 1月の死亡率			
平均	13.038	11.650	t= 1.253
標準偏差	2.961	3.295	P=0.220
サンプル	16	16	df= 30
変数 2 ----- 2月の死亡率			
平均	17.063	15.725	t= 1.018
標準偏差	3.241	4.135	P=0.317
サンプル	16	16	df= 30
変数 3 ----- 3月の死亡率			
平均	22.788	20.025	t= 2.696
標準偏差	2.626	3.147	P=0.011
サンプル	16	16	df= 30
変数 4 ----- 4月の死亡率			
平均	23.275	22.175	t= 0.922
標準偏差	2.919	3.777	P=0.364
サンプル	16	16	df= 30
変数 5 ----- 5月の死亡率			
平均	21.069	18.244	t= 2.389
標準偏差	2.753	3.846	P=0.023
サンプル	16	16	df= 30
変数 6 ----- 6月の死亡率			
平均	19.300	18.250	t= 0.723
標準偏差	3.038	4.954	P=0.475
サンプル	16	16	df= 30
変数 7 ----- 7月の死亡率			
平均	16.438	15.506	t= 0.790
標準偏差	3.188	3.477	P=0.436
サンプル	16	16	df= 30
変数 8 ----- 8月の死亡率			
平均	14.975	13.944	t= 1.317
標準偏差	1.735	2.608	P=0.198
サンプル	16	16	df= 30
変数 9 ----- 9月の死亡率			
平均	12.900	11.475	t= 1.655
標準偏差	2.150	2.690	P=0.108
サンプル	16	16	df= 30

変数	10	-----	10月の死亡率		
平均	10.850		9.613	t=	1.787
標準偏差	1.911		2.006		P=0.084
サンプル	16		16	df=	30
変数	11	-----	11月の死亡率		
平均	8.400		7.194	t=	2.051
標準偏差	1.537		1.781		P=0.049
サンプル	16		16	df=	30
変数	12	-----	12月の死亡率		
平均	12.394		11.300	t=	1.193
標準偏差	2.241		2.904		P=0.242
サンプル	16		16	df=	30

1, 2 群の総和の分散共分散 (自由度 30)

	1	2	3	4	5	6	7	8
1	9.8119							
2	-0.41458	13.802						
3	2.3632	-5.1283	8.3989					
4	-1.8538	-2.2748	-1.0845	11.393				
5	-1.1045	-2.3875	1.8532	-2.9982	11.183			
6	4.6627	-4.5993	2.2433	4.4723	0.51117	16.883		
7	1.4187	-3.1283	1.4655	1.6412	3.3185	3.0912	11.126	
8	1.4193	1.7456	0.17692	0.35925	-1.7354	2.0732	-1.4990	4.9063
9	0.61100	1.2980	1.0383	0.94500	-1.6228	0.29400	-0.095917	-0.75975
10	-0.71400	2.9345	-1.2292	2.2115	-2.3118	-1.0687	-1.5107	0.32638
11	-1.0945	1.4038	-2.7152	0.56892	-1.7875	-3.4058	-1.5506	-1.3159
12	0.64812	0.92654	0.76629	-1.9258	-2.2421	-4.4867	-0.20454	1.4442
	9	10	11	12				
9	5.9297							
10	1.5215	3.8372						
11	-0.12475	0.56571	2.7676					
12	0.35400	0.32317	-0.020000	6.7276				

1. 両プロファイルの並行性の検討 (対応なし)

$$F(11, 20) = 0.282 \quad P=0.982$$

両群の平均値プロファイルの並行性(Parallelism)が支持されました。
 以後の結果を検討して下さい。

2. 両プロファイルの同一性 (有意差) の検討

$$t(df=30) = 5.128 \quad P<0.001$$

両プロファイルの同一性が否定されました: 第1群のProfile > 第2群のProfile

3. 荷重平均値を用いたプロファイルの変数間の変動の検討

$$F(11, 20) = 124.736 \quad P<0.001$$

プロファイルの変数間には有意な変動があります。

各変数間の平均値の差における95%信頼区間とその有意性

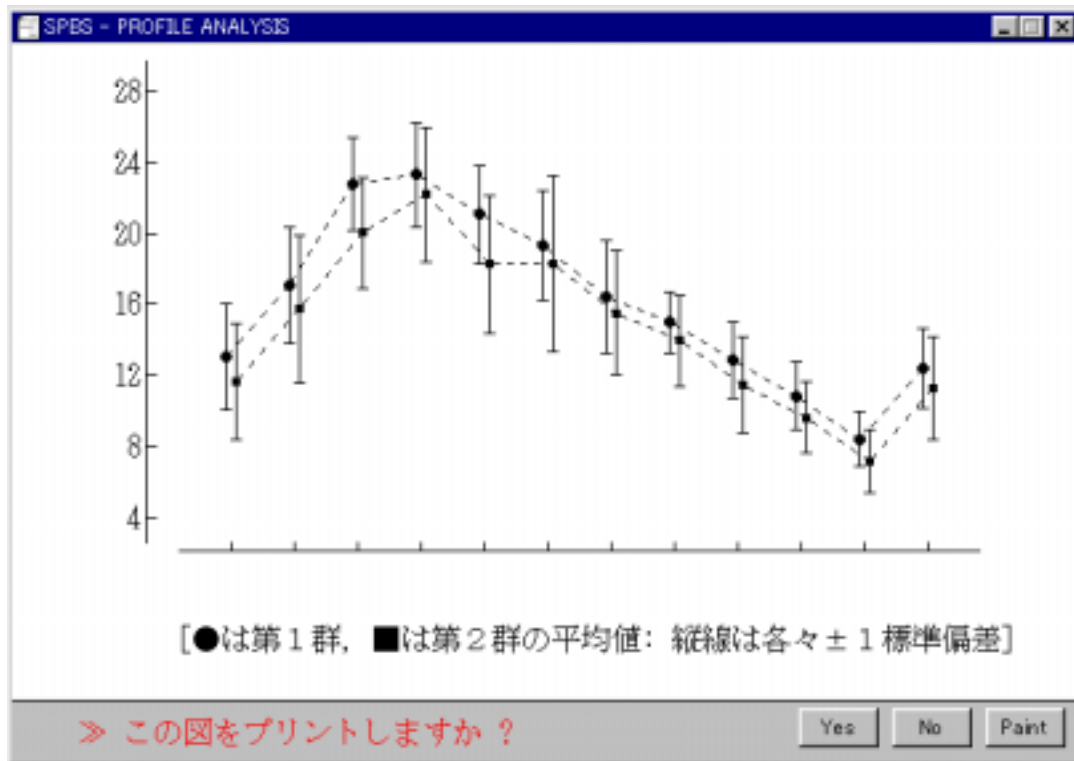
変数間	荷重平均の差	その95%信頼区間	有意性
1 - 2	-4.050	-8.456 μ	0.356 n.s.
1 - 3	-9.063	-12.335 μ	-5.790 P<0.01

1 - 4	-10.381	-14.829	μ	-5.934	P<0.01
1 - 5	-7.313	-11.605	μ	-3.020	P<0.01
1 - 6	-6.431	-10.145	μ	-2.717	P<0.01
1 - 7	-3.628	-7.419	μ	0.163	n.s.
1 - 8	-2.116	-5.187	μ	0.956	n.s.
1 - 9	0.156	-3.239	μ	3.552	n.s.
1 - 10	2.112	-1.348	μ	5.573	n.s.
1 - 11	4.547	1.122	μ	7.971	P<0.01

..... 途中省略

7 - 9	3.784	0.084	μ	7.485	P<0.05
7 - 10	5.741	1.962	μ	9.520	P<0.01
7 - 11	8.175	4.502	μ	11.848	P<0.01
7 - 12	4.125	0.317	μ	7.933	P<0.05
8 - 9	2.272	-0.860	μ	5.404	n.s.
8 - 10	4.228	1.693	μ	6.763	P<0.01
8 - 11	6.663	3.802	μ	9.523	P<0.01
8 - 12	2.613	-0.023	μ	5.248	n.s.
9 - 10	1.956	-0.354	μ	4.267	n.s.
9 - 11	4.391	1.725	μ	7.056	P<0.01
9 - 12	0.341	-2.740	μ	3.421	n.s.
10 - 11	2.434	0.350	μ	4.519	P<0.05
10 - 12	-1.616	-4.422	μ	1.191	n.s.
11 - 12	-4.050	-6.802	μ	-1.298	P<0.01

μ : 荷重平均の変数間の差 ($M_i - M_j$)



28 . Weibull 生存分析

多重ロジスティック回帰分析の場合、ある特定期間内、具体的には追跡調査期間内にある疾患が発症する確率が目的変数となっているが、多重Weibull分布の場合は、観察の開始から任意の時間 t までに発症や死亡などの事象が生起する確率分布を対象とする（古川・田中、1983）。

多重Weibull分布では、確率分布として生命表分析や機会の故障解析で用いられるWeibull分布を採用する。Weibull分布とは、時間 t までに事象を生起させる確率 $p(t)$ とすれば、 $p(t) = 1 - \exp(-t^m/t_0)$ で与えられる分布で、形状母数 m の値によって、 $0 < m < 1$ ならば“初期故障”、 $m = 1$ ならば“偶発故障”、 $m > 1$ ならば“摩耗故障”のように各故障モードを表すことができる。 t_0 はこの事象生起の時間的スケールを表し、尺度母数と呼ばれている。多重Weibull関数は尺度母数 t_0 を各個体の属性変量の関数としたもので、 $t_0 = \exp(-w)$ で、 $w = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$ で与えられる。 w が大きくなると尺度母数 t_0 が小さくなり、事象生起の時間的スケールがより小さくなるので、 w が死亡の発生などを説明する変数となる。なお、以前は生存分析にしばしばこのWeibull関数が使用されたが、近年はCoxの比例ハザードモデルの方が好んで使われる。

Weibull Function

D:¥SPBS¥reidai¥膵癌

基本データ

解析の症例数 = 83

変数	平均値	標準偏差	最小値	最大値	変数名
1	61.133	9.651	38.000	83.000	手術時の年齢
2	0.422	0.497	0.000	1.000	性別 (0=男性)
3	0.735	0.444	0.000	1.000	術中照射の有無
4	0.783	0.415	0.000	1.000	占拠部位 (0=頭部のみ)
5	2.422	1.298	1.000	4.000	膵内胆管への浸潤の程度

説明変数間の推定分散共分散行列

(Iteration= 8)

	1	2	3	4	5
1	0.000172370				
2	-0.000004736	0.0521144			
3	-0.000776344	-0.000602279	0.0923346		
4	0.000152854	-0.0035578	-0.000672795	0.0835348	
5	0.000027219	-0.0042850	0.0144573	0.0084652	0.0101814

Weibull Functionによる推定 (Wald検定)

(注: 変数 0 は定数項)

変数	推定値()	標準誤差	推定値 / 標準誤差	2乗値	有意性
0	-3.11307	0.96367			
1	0.0146093	0.0131290	1.113	1.238	P=0.266
2	-0.0293205	0.22829	-0.128	0.016	P=0.898
3	-0.89367	0.30387	-2.941	8.650	P=0.003
4	0.51626	0.28902	1.786	3.191	P=0.074
5	-0.0356282	0.10090	-0.353	0.125	P=0.724

形状母数(shape parameter) 1.43035 (標準誤差 0.12046)

全尤値 (共変量によらない対数尤度) 438.520

偏尤値 (共変量からの最大対数尤度) 426.832 (AIC = 440.83)

モデルの尤度比検定 (-2 LOG L) 11.688 (自由度 = 5) P=0.039

謝 辞

帝京大学法学部山岡和枝助教授、医学部衛生学公衆衛生学教室の野中浩一講師、矢野栄二教授より本書および添付ソフトに対し有益な助言を賜りました。この場を借りて感謝申し上げます。

参考図書および文献

本書および添付ソフトの作成にあたっては、以下の参考資料（本および論文）を参照した。

1. 縣俊彦. *EBM - 臨床医学研究の方法論*. 中外医学, 1998.
2. Anderson S, Auquier A, Hauck WW, Oakes D, Vandaele W, Weisberg HI: *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*. John Wiley & Sons (New York), 1980.
3. Chalmers I, Altman DG. *Systematic Reviews*. BMJ Publishing Group (London), 1995.
4. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A: A method for assessing the quality of a randomized control trial. *Controlled Clin Trials* 2: 31-49, 1981.
5. Colton T: *Statistics in Medicine*. Little, Brown & Co. (Boston), 1974.
6. Cox DR: Regression models and life-tables. *J. Roy. Statist. Soc. Br.* 34: 187-202, 1972.
7. Dixon RA, Munro JF, Silcocks PB: *The Evidence Based Medicine Workbook – Critical appraisal for clinical problem solving*. Reed Educational and Professional Publishing Ltd (Oxford), 1997 (野崎貞彦, 横山英世監訳: *EBMワークブック*. 医歯薬出版, 1999).
8. Dixon WJ, Massey FJ: *Introduction to Statistical Analysis, 4th edn*. McGraw-Hill Book Co. (New York), 1983.
9. Drummond MF, Stoddart GL, Torrance GA. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press (Oxford), 1987.
10. Dunnett CW: A multiple comparison procedure for comparing several treatments with a control. *Am. Statist. Assoc. J.* 50: 1096-1121, 1955.
11. 疫学研究におけるインフォームド・コンセントに関する研究と倫理ガイドライン策定研究班. 疫学研究におけるインフォームド・コンセントに関するガイドライン、Ver 1.0. 疫学研究におけるインフォームド・コンセントに関する研究と倫理ガイドライン策定研究班, 2000.
12. Fleiss JL: *Statistical Methods for Rates and Proportions, 2nd edn*. John Wiley & Sons (New York), 1981.
13. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials (3rd Ed)*. Williams & Wilkins (Baltimore), 1996 (福井次矢監訳: *臨床疫学 EBM 実践のための必須知識*. メディカル・サイエンス・インターナショナル, 1999).
14. Garner TI, Dardis R. Cost-effectiveness analysis of end-stage renal disease treatment. *Med Care* 25: 25-34, 1987.
15. Glantz SA: *Primer of Biostatistics, 2nd edn*. McGraw-Hill Book Co. (New York), 1987.
15. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29-36, 1982.
17. Hanley JA, McNeil BJ: A method of comparing the area under receiver operating characteristic curves derived from the same cases. *Radiology* 148: 839-843, 1983.
18. 橋本英樹: RCTは最高の evidence なのか. *Medicina* 35: 2076-2079, 1998.
19. 広津千尋: *臨床試験データの統計解析*. 廣川書店, 1992.
20. 市川伸一、大橋靖雄、岸本淳司、浜田知久馬: *SASによるデータ解析入門 [第2版]*. 東京大学出版会, 1993.
21. 今泉洋子、稲葉 裕: *保健統計学*. 培風館, 1997.
22. 開原成允: *マイコンによる医療統計処理*. 中山書店, 1983.
23. Kleinbaum DG: *Logistic Regression: A Self-Learning Text*. Springer-Verlag (New York), 1994.
24. Kleinbaum DG, Kupper LL, Muller KE, Zizam A: *Applied Regression Analysis and Other Multivariable Methods (3rd ed)*. Duxbury Press (Pacific Grove), 1998.
25. Lloyd E (ed): *Handbook of Applicable Mathematics (Part A & B)*. John Wiley & Sons (Chichester), 1984.
26. 前谷俊三: *臨床生存分析 - 生存データと予後因子の解析*. 南江堂, 1996.
27. Miettinen OS: Estimation of relative risk from individually matched series. *Biometrics* 26: 75-85, 1970.

28. 宮原英夫、丹後俊郎：医学統計ハンドブック．朝倉書店．1995．
29. 三宅浩次：t-検定の誤用を避けるために 多重比較の統計学的検定 ．医学のあゆみ 136: 600-603, 1986．
30. Morrison DF: *Multivariate Statistical Methods, 2nd edn.* McGraw-Hill Book Co. (New York), 1976．
31. 永田 靖、吉田道弘：統計的多重比較法の基礎．サイエンティスト社，1997．
32. Namboodiri K, Suchindran CM: *Life Table Techniques and Their Applications.* Academic Press (Orlando), 1987．
33. 中道 実：社会調査方法論．恒星社厚生閣，1997．
34. 大橋靖雄、浜田知久馬：生存時間解析 - SAS による生物統計．東京大学出版会，1995．
35. 奥野忠一(編)：応用統計ハンドブック．養賢堂，1984．
36. 奥野忠一、久米 均、芳賀敏郎、吉沢 正：多変量解析法．日科技連，1983．
37. Petitti DB. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis, 1st ed.* Oxford Univ Press (New York), 1994 (福井次矢, 青木則明監訳: *E B Mのためのデータ統合型研究 - メタ分析、決断分析、費用効果分析の理論と実際. メディカル・サイエンス・インターナショナル*, 1999)．
38. Ransohoff DF, Gracie WA, Wolfenson LB, Neuhauser D: Prophylactic cholecystectomy or expectant management for silent gallstones. *Ann Intern Med* 99: 199-204, 1983．
39. Robins J, Greenland S, Breslow NE: A general estimator for the variance of the Mantel-Haenszel Odds ratio. *Am. J. Epidemiol.* 124: 719-723, 1986．
40. Rothman KJ: The estimation of synergy or antagonism. *Am. J. Epidemiol.* 103: 506-511, 1976．
41. Rothman KJ, Greenland S: *Modern Epidemiology (2nd ed)*. Lippincott-Raven Publishers (Philadelphia), 1998．
42. Rushton L: Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. *Occup. Environ. Med.* 57: 1-9, 2000．
43. Sackett DL, Richardson WS, Rosenberg W, Haynes RB: *Evidence-based MEDICINE.* Churchill Livingstone (London), 1997 (久繁哲徳監訳: *根拠に基づく医療-EBMの実践と教育の方法.* オーシーシー, 1998)．
44. 佐久間昭：医学統計Q & A．金原出版，1990．
45. 柴田義貞、開原成允、小沢利男：多重 logistic モデルによる risk factor の分析 ．医学のあゆみ 106: 207-212, 1978．
46. 新村秀一：パソコンによるデータ解析 統計ソフトを使いこなす．講談社，1995．
47. Siegel S: *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill Kogakusha (Tokyo), 1956．
48. Silagy C, Haines A (Ed): *Evidence Based Practice in Primary Care.* BMJ Books (London), 1998．
49. Snedecor GW, Cochran WG: *Statistical Methods, 6th edn.* Iowa State University Press (Ames), 1967．
50. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 341: 418-422, 1993．
51. 田中 豊、垂水共之、脇本和昌：パソコン統計解析ハンドブック、 ．多変量解析編．共立出版，1984．
52. 丹後俊郎：医学への統計学．朝倉書店，1983．
53. 丹後俊郎、山岡和枝、高木晴良：ロジスティック回帰分析 - SAS を利用した統計解析の実際 - ．朝倉書店，1996．
54. Thompson SG, Pocock SJ: Can meta-analysis be trusted? *Lancet* 338: 1127-1130, 1991．
55. 鳥居敏雄、高橋暁正、土肥一郎：医学・生物学のための推計学．東京大学出版会，1981．
56. Wilson JMG, Jungner G: *Principles and practice of screening for disease.* WHO (Geneva), 1968．
57. Weinstein MC, Fineberg HV: *Clinical Decision Analysis.* WB Saunders Co. (Philadelphia), 1980．
58. 柳井晴夫、高木広文(編)：多変量解析ハンドブック．現代数学社，1986．
59. 矢野栄二、小林廉毅、山岡和枝(編)：Evidence Based Medicine による健康診断．医学書院，1999．
60. 矢野栄二(編)：医療と保健における評価 - Toward Evidence Based Medicine - ．南江堂，1999．

索引

ID 準拠によるデータファイルの合成	21	ダミー変数の作成 (Partial 法)	22
ROC 曲線	48	適合度検定	25, 42, 64, 72, 75
一元配置分散分析	31	データ制御プログラムの起動	12
一次回帰式	40	データの検索・置換 (Recode)	16, 22
因子分析	68	データの削除・追加	22
Weibull 生存分析	80	データの小数点以下の桁調整	21
Wilcoxon 順位和検定	29	データの対数・ルート変換	23
Wilcoxon 符号付検定	29	データの並べ替え	21
Welch の検定	28, 45	データの変数名ファイルの確認・変更	20
Excel (CSV) file SPBS file の相互変換	18	データファイルの合成	21
SPBS のヒント	4	データファイルのサイズの変更	22
SPBS 用データファイルの確認	10	データファイルの作成と修正	14
SPBS 用データファイルの消去	11	データファイルの縦横変換	20
SPBS 用データファイルの複製	11	特異度	48
SPBS 用データファイルの変換	11	内的整合性の検討	47
Odds 比	43, 61	2 群間の多重比較	36
χ^2 検定	43, 44, 59, 61	2 群間の有意差検定	28, 45
確率 P 値	44	二元配置分散分析 (繰り返しのある)	59
Kaplan-Meier 法による生存曲線	55	二元配置分散分析 (繰り返し無し)	34
感度	48	二次回帰式	38
幾何平均	26	2 変数によるデータの 3 または 4 分割	23
基本統計量	24, 25	日本地図の作成	26
共分散分析	45	箱ヒゲ図	32, 35
Kruskal-Wallis 順位検定	31	パーセントイル	25
クロス集計	59	Bartlett 検定	31
Cronback の 係数	47	反復測定データの有意差検定	38
Cox の比例ハザードモデル	57	判別分析	72
算術平均	25	比率の検定	43
Scheffe の多重比較	31, 34	ファイル内データの加減乗除	22
重回帰分析	64	ファイルの比較検討	20
主成分分析	66	Fisher 直接確率	43
数量化理論 類	70	Friedman 順位検定	34
Student の t 検定	28, 45	プリンター場所の設定 (LPT1 以外の場合)	12
Spearman 順位相関係数	51	Profile 分析 (並行性の検定)	77
正規分布	25	変数内の順位データの逆順位変換	16, 21
正準相関分析	53	偏相関係数	50, 52
生存分析	55	Hotelling の T^2 統計量	36, 72
相関 (または相関分析)	50, 52	母分散の検定 (F 検定)	28
相乗作用指数	62	Bonferroni の多重比較	6, 39
相対危険 (Relative risk)	62	McNemer 検定	3, 62
対応のある t 検定	28	Mantel-Haenszel 検定	4, 63
対応のない t 検定	28	Mann-Whitney 検定	29
多項式	42	Median test	29
多重共線性	64	Morrison の多重比較	6, 39
多重有意性検定 (または多重比較)	31, 34, 38	有意差検定	28, 45
多重ロジスティック回帰分析	75	連続量データのカテゴリデータへの変換	24

The Statistical Package for the Biosciences (SPBS) User's Guide

Winestem Villagefield, MD

Winestem Institute of Community Medicine, Toda, Saitama, Japan