

正準判別分析（重判別分析）

1. 目的

13人の患者の検診データがあります。（名前等は仮）
各自の 検診データは γ -Sm 値、尿酸値、PAP 値 であり、
IDの佐藤～鬼頭の6人は前立腺癌を発症しており、土居～瀬尾の7人は前立腺肥大を発症しています。

ID	症状	γ -Sm 値	尿酸値	PAP 値
佐藤	前立腺癌	4.12	6.00	4.52
石井	前立腺癌	3.82	5.58	2.13
深井	前立腺癌	2.67	4.30	2.64
佐山	前立腺癌	3.55	3.55	2.29
尾上	前立腺癌	2.49	2.49	3.00
鬼頭	前立腺癌	4.81	4.81	4.88
土居	前立腺肥大	3.21	3.21	2.83
新井	前立腺肥大	0.95	7.50	2.25
新川	前立腺肥大	3.47	3.47	4.15
小室	前立腺肥大	2.16	9.30	1.76
杉山	前立腺肥大	2.18	3.75	2.59
田中	前立腺肥大	1.43	6.15	2.21
瀬尾	前立腺肥大	1.85	4.80	3.01

ここで新たな患者の検診データが知らされた場合、前立腺癌、前立腺肥大のどちらの症状に近いのかという問題を考慮します。

なお、ここでは、判別対象が 癌？ 肥大？ の二択ですので、「判別分析」の機能を用いても、判別できますが、判別対象が3つ以上になると「判別分析」ではできません。

正準判別分析（重判別分析とも言います）は3つ以上の判別を可能とするものです。

なお、判別対象（以降では 群と呼びます）ごとのデータの個数は、以下の注意が必要です。

上記の例ですと、群の数は2つで、“前立腺癌”と“前立腺肥大”とになります。
そして、“前立腺癌”のデータの個数は 6，“前立腺肥大”のデータの個数は 7 です。

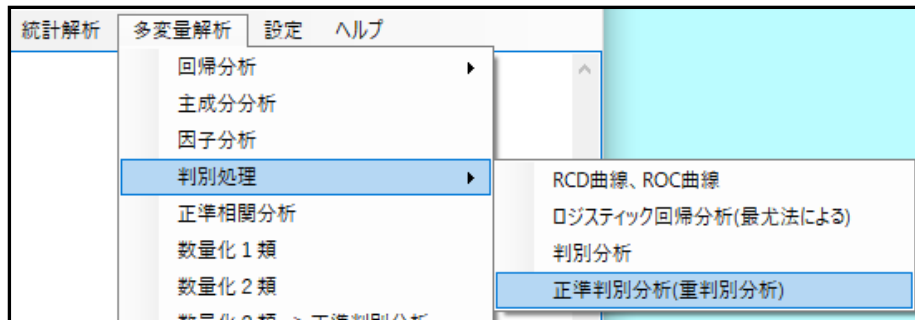
変量数をPとした場合、各群のデータ個数は Pより大きいことが必要です。

P個より少ないデータ数では、分散共分散行列が特異（行列のランクがPより小さい）となるため、計算が不可能となるからです。
また、データ数がP個に等しい場合、数値的に不安定になる場合があるので注意してください。

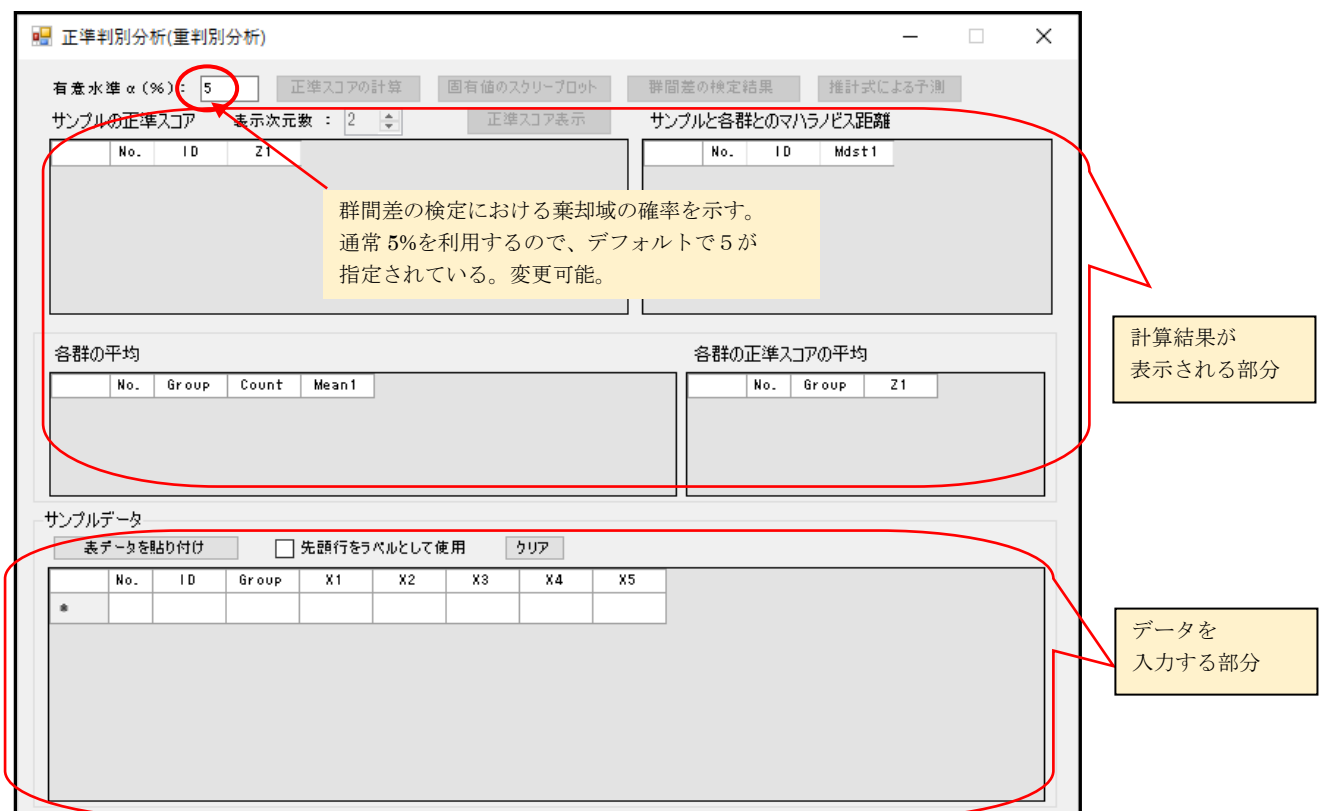
2. 使用法

(1) メニューの選択

メニューの「多変量解析→判別処理→正準判別分析(重判別分析)」を選択します。



(2) パネルが表示されます。



(3) データの入力

表計算に以下のデータが準備されているとして、そのデータをコピーして利用します。

ID	症状	γ -Sm 値	尿酸値	PAP 値
佐藤	前立腺癌	4.12	6.00	4.52
石井	前立腺癌	3.82	5.58	2.13
深井	前立腺癌	2.67	4.30	2.64
佐山	前立腺癌	3.55	3.55	2.29
尾上	前立腺癌	2.49	2.49	3.00
鬼頭	前立腺癌	4.81	4.81	4.88
土居	前立腺肥大	3.21	3.21	2.83
新井	前立腺肥大	0.95	7.50	2.25
新川	前立腺肥大	3.47	3.47	4.15
小室	前立腺肥大	2.16	9.30	1.76
杉山	前立腺肥大	2.18	3.75	2.59
田中	前立腺肥大	1.43	6.15	2.21
瀬尾	前立腺肥大	1.85	4.80	3.01

1 列目は ID、2 列目は サンプルが所属する群、3～5 列目にデータが記述されています。

3 列目以降は データによって、列数は変わります。

表計算のコピー対象部分をドラッグ&コピーし、「表データの貼り付け」ボタンを押します。

先頭行をラベルとして使用をチェック

サンプルデータ

表データを貼り付け ☒ 先頭行をラベルとして使用 クリア

No.	ID	Group	X1	X2	X3	X4	X5
*							

「表データを貼り付け」ボタンを押す

表のイメージが下のグリッドにコピーされる



サンプルデータ

表データを貼り付け ☒ 先頭行をラベルとして使用 クリア

No.	ID	症状	γ -Sm 値	尿酸値	PAP 値
1	佐藤	前立腺癌	4.12	6.00	4.52
2	石井	前立腺癌	3.82	5.58	2.13
3	深井	前立腺癌	2.67	4.30	2.64
4	佐山	前立腺癌	3.55	3.55	2.29
5	尾上	前立腺癌	2.49	2.49	3.00
6	鬼頭	前立腺癌	4.81	4.81	4.88

データ件数は 13 件、
変数数は 3、
グループ数は 2

OK

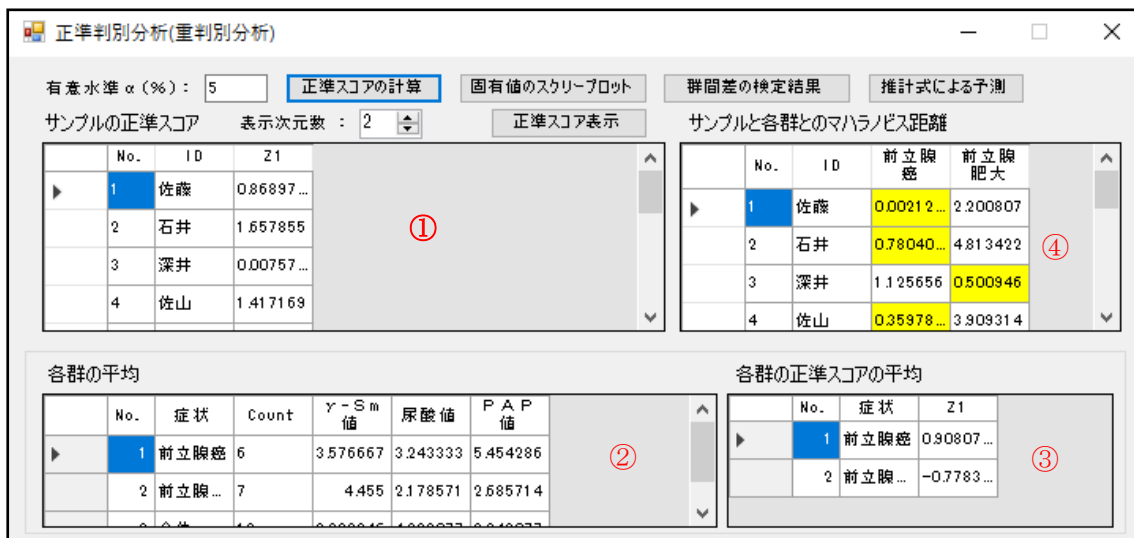
入力後に
このような
ポップアップが
表示されます

(4) 計算の実行

上記のデータが グリッドにコピーされると、パネル上部の
ボタンが押せる状態になります。

正準スコアの計算

そのボタンを押すと、計算が実行され、計算結果が表示されます。



各グリッドに計算結果が表示されます。

① サンプルの正準スコア

各サンプル（佐藤、石井、... 瀬尾）の3つのデータ（ γ -Sm 値、尿酸値、PAP 値）から、**正準スコア** という新たな 数量 を算出します。
（正準スコアについては 「考え方」 を 参照してください）

ここでは、佐藤 : 0.86897、石井 : 1.65785 ... と計算されます。
この場合は、1 次元のつまり1つの指標が 正準スコアとして 計算されます。

この正準スコアの指標の数（次元）は、元のデータの次元数 と 群の数とにより、決まります。
正準スコアの指標の次元（数） = \min （元のデータの次元数、群の数 - 1）

② 各群の平均

群ごとの 入力データの平均が表示されています。

③ 各群の正準スコアの平均

群ごとの正準スコアの平均が表示されています。

④ サンプルと各群とのマハラノビス距離

サンプル毎に 各群の正準スコアの平均との間の距離を表示しています。
この場合の距離はマハラノビス距離です（ユークリッド距離を分散で割った距離）。
この距離が一番小さい群に サンプルは 属していると考えるのが妥当です。
表では、属している群を 黄色で表現しています。

例えば

- ・ 佐藤は “前立腺癌” に近い
- ・ 石井も “前立腺癌” に近い

しかし、

- ・ 深井は “前立腺肥大” の方に近い

ということを示しています。

各グリッドは 表示サイズが小さく見にくいです。

ですが、グリッドの内容を、表計算シートにコピーすることができます。

グリッド①の サンプルの正準スコアを 表計算に コピーするには、以下のようになります。

① 左上隅のセルをクリックして、全体が青くなるのを見てから、Ctrl-Cを押す

No.	ID	Z1
1	佐藤	0.86897...
2	石井	1.657855
3	深井	0.00757...
4	佐山	1.417169

→

② 表計算ソフトを起動し、シートを開いてCtrl-Vを押す

No.	ID	Z1
1	佐藤	0.86897
2	石井	1.65786
3	深井	0.00758
4	佐山	1.41717
5	尾上	-0.23063
6	鬼頭	1.72748
7	土居	0.73874
8	新井	-2.39848
9	新川	0.42568
10	小室	-0.72800
11	杉山	-0.56683
12	田中	-1.61169
13	瀬尾	-1.30786

グリッド④の 「サンプルと各群との... 距離」を 表計算に コピーするには、同様にします。

③ 左上隅のセルをクリックして、全体が青くなるのを見てから、Ctrl-Cを押す

No.	ID	前立腺癌	前立腺肥大
1	佐藤	0.00212...	2.200807
2	石井	0.78040...	4.813422
3	深井	1.125656	0.500946
4	佐山	0.35978...	3.909314

→

④ 表計算ソフトを起動し、シートを開いてCtrl-Vを押す

No.	ID	前立腺癌	前立腺肥大
1	佐藤	0.00212	2.20081
2	石井	0.78040	4.81342
3	深井	1.12566	0.50095
4	佐山	0.35979	3.90931
5	尾上	1.79996	0.24330
6	鬼頭	0.93207	5.09249
7	土居	0.03981	1.86858
8	新井	15.17733	2.12876
9	新川	0.32303	1.17572
10	小室	3.71578	0.00208
11	杉山	3.01974	0.03629
12	田中	8.81380	0.56321
13	瀬尾	6.81645	0.22739

(*) データのコピーは、テキストのみが対象なので、右図のような、罫線や、背景の塗りつぶしなどはコピーされません。罫線などの修飾は 適宜行ってください。

これで、次の推定に必要な 準備ができました。

(5) 推定

ここで新たな患者の検診データが知らされたとします。その患者は 前立腺癌、前立腺肥大どちらに近いか を 計算します。

ボタン **推計式による予測** をクリックします。 下の パネルが 表示されます。

正準判別分析計算に基づく判定

判定対象のデータ入力:

NO	変 量 名	変 量 値
1	γ-Sm値	0
2	尿酸値	0
3	PAP値	0

判 別

入力データの正準スコア:

NO	Z	スコア
----	---	-----

分布表示

判定対象と各群の平均との汎距離:

NO	群	汎 距 離
----	---	-------

新たな患者の 検診データを、
γ-Sm 値：4、 尿酸値：3、 PAP 値：2
とします。「判定対象のデータ入力」 に データを入力し、

判別 ボタンを

正準判別分析計算に基づく判定

判定対象のデータ入力:

NO	変 量 名	変 量 値
1	γ-Sm値	4
2	尿酸値	3
3	PAP値	2

判 別

入力データの正準スコア:

NO	Z	スコア
1	1	2.207263

分布表示

判定対象と各群の平均との汎距離:

NO	群	汎 距 離
1	前立腺癌	2.343109
2	前立腺肥大	7.229258

こちらが小さい

正準スコアは
2.2 となりました

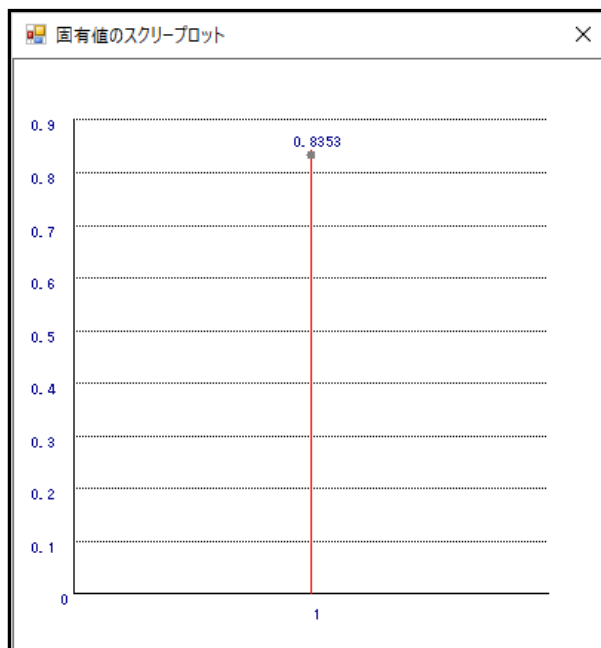
前立腺癌の症
状に近いと判
断された

(6) 固有値のスクリープロット

「正準スコアの計算」ボタンを押すと、「固有値のスクリープロット」ボタンが押せるようになります。

固有値のスクリープロット

このボタンを押すと、



指標の計算の参考になる 固有値を グラフ表示します。

ここでは、1つの固有値のみ計算されたわけですが、元のデータの変量数 や 群の数によって、計算される固有値の数は、複数になる場合があります。

その場合には、左側から 値の大きいもの順に グラフ表示されます。

(7) 群間差の検定結果

群間差の検定結果

ボタンを押すと、以下のパネルが表示されます。

正準判別分析

群間差の検定 (Fisher検定)

有意水準 α (%): 5 自由度 1: 3 自由度 2: 9

F 値: 2.505908217 帰無仮説の採択域: (0 , 3.86254835)

結 果: 群間差は有意とみなさない 分布関数

各群の正準スコアの平均が十分離れている場合には、つまり、群間差がある場合には、

- ・群間差は有意とみなせる

という結果となりますが、十分には離れていない場合には、

- ・群間差は有意とみなせない

と表示されます。