

判別分析

1. 目的

例えば、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 の3種類で、
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。

前立腺癌			
ID	γ -Sm 値	尿酸値	PAP 値
佐藤	4.12	6.00	4.52
石井	3.82	5.58	2.13
深井	2.67	4.30	2.64
佐山	3.55	3.55	2.29
尾上	2.49	2.49	3.00
鬼頭	4.81	4.81	4.88

前立腺肥大			
ID	γ -Sm 値	尿酸値	PAP 値
土居	3.21	3.21	2.83
新井	0.95	7.5	2.25
新川	3.47	3.47	4.15
小室	2.16	9.3	1.76
杉山	2.18	3.75	2.59
田中	1.43	6.15	2.21
瀬尾	1.85	4.8	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かという問題に答えてくれるのが 判別分析です。

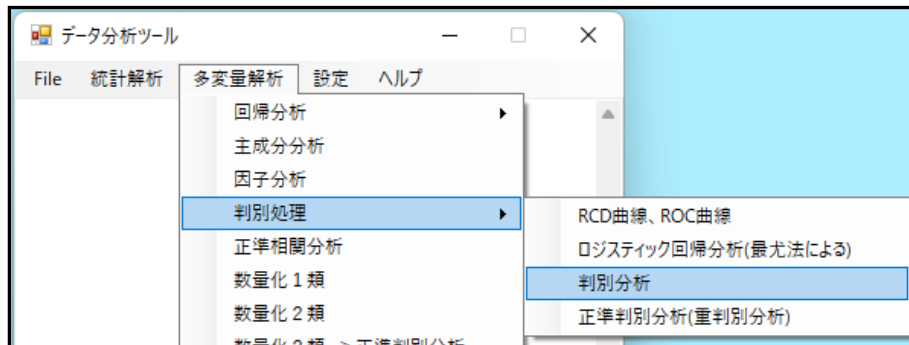
判別分析は、2つのグループに分類されたデータを元に、どちらに所属するか という問題に答えることができますが、所属グループの数が3つ以上の場合は、対応できません。
その場合は、別項で説明する 正準判別分析 を利用してください。

なお、多変量の場合、データの個数については、以下の注意が必要です。
多変量の数をPとした場合、2つのグループはともにP個の変量が定義されねばなりません。
また、2つのグループのデータ個数は それぞれ P 個より多く 定義してください。
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→判別分析」を選択します。



(2) パネルが表示されます。

有意水準 α (%) 5

2群のデータによる判別分析

群の平均分散(1変量の場合)

第1群個数: 平均値1: (不偏)分散値1:

第2群個数: 平均値2: (不偏)分散値2:

群の平均分散(多変量の場合)

第1群個数: 平均値1: 不偏分散共分散行列1: 表示

第2群個数: 平均値2: 不偏分散共分散行列2: 表示

2群の分散の評価結果

等分散の評価: ☐ 等分散とみなせる($\Sigma(1) = \Sigma(2)$) ☐ 等分散とみなせない($\Sigma(1) \neq \Sigma(2)$)

等分散の時: 判別式による評価

P1(第1群に属するのに、第2群と誤判別される確率 %):

P2(第2群に属するのに、第1群と誤判別される確率 %):

☐ 先頭行をラベルとして使用

第1群のデータ 表データを貼り付け クリア

第2群のデータ 表データを貼り付け クリア

NO ID X1

NO ID X1

以降 入力されるデータの 변수数にわけて、説明をします。

변수数が1つの場合(1次元の場合)、 (3) 項で。

변수数が2つの場合(2次元の場合)、 (4) 項で。

변수数が3つの場合(3次元の場合)、 (5) 項で。

(3) 変量数が 1 つの場合

以下を例に説明します。

あるクラスの生徒の学校での英語の模擬試験の得点を、2 グループに分けて表示しています。
左側は最近実施された英検で合格したグループ、右側は不合格のグループです。

	名前	得点
1	A	60
2	B	52
3	C	68
4	D	55
5	E	65
6	F	47
7	G	45
8	H	62
9	I	53

	名前	得点
1	J	49
2	K	40
3	L	52
4	M	37
5	N	55
6	O	38
7	P	45

これを見ると、学校の模試は低得点だが検定に合格した人（F、G）がいれば、
逆に学校の模試は高得点だが検定に不合格（N）というケースがあります。

学校の模試が、そのまま検定の結果につながるものではないとしても、
学校の模試で何点以上なら検定の合格圏内？ が判れば目安の一つになります。
そんな 目的に 利用できると思われます。

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。

	名前	得点
1	A	60
2	B	52
3	C	68
4	D	55
5	E	65
6	F	47
7	G	45
8	H	62
9	I	53

	名前	得点
1	J	49
2	K	40
3	L	52
4	M	37
5	N	55
6	O	38
7	P	45

1 列目に I D（名前）、2 列目に得点が 記述されています。

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

つまり、

- ・ 上図の左側表の部分をヘッダも含めてドラッグ&コピーして、第 1 群のデータ側の「表データを貼り付け」ボタンをクリックします。
- ・ 上図の右側表の部分をヘッダも含めてドラッグ&コピーして、第 2 群のデータ側の「表データを貼り付け」ボタンをクリックします。

なお、データ入力の決まりとして、各データには I D を必ずつけることとします。

先頭行をラベルとして使用をチェック

☒ 先頭行をラベルとして使用

第1群のデータ 表データを貼り付け クリア

No.	ID	X1
▶*		

データ分布1 データ分布2

第2群のデータ 表データを貼り付け クリア

No.	ID	X1
*		

「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

☒ 先頭行をラベルとして使用

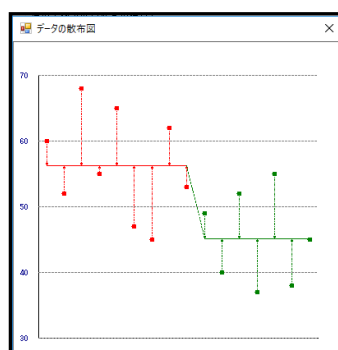
第1群のデータ 表データを貼り付け クリア

No.	名前	得点
▶	1 A	60
	2 B	52
	3 C	68
	4 D	55

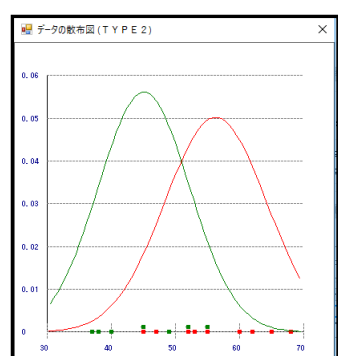
データ分布1 データ分布2

第2群のデータ 表データを貼り付け クリア

No.	名前	得点
▶	1 J	49
	2 K	40
	3 L	52
	4 M	37



「データ分布1」ボタンを押すと、データの散布状況を表示します。赤い点は第1グループのデータで、赤い水平な線は第1グループの平均値を示します。赤い点から赤い線への垂直な線は、平均値からのばらつきを表します。この線が長いと、グループ内のデータの分散（バラツキ）は大きいと直感的にわかります。緑の点は第2グループのデータで、緑の水平な線は第2グループの平均値を示します。緑の点から緑の線への垂直な線は、平均値からのばらつきを表します。



「データ分布2」ボタンを押すと、別のグラフを表示します。点データは下部に一直線状に表示されています。データが重なる部分は、わずかに位置をずらして表示しています。赤い曲線、緑の曲線は、2つのグループのデータが正規分布に従うと仮定した場合の確率分布を表現しています。ここでも、赤い点と緑の点が真ん中あたりでオーバーラップしているのが分かります。

データが グリッドにコピーされると、パネル上部の

2群のサンプルの判別分析

ボタンが押せる状態になります。

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。

2群のデータによる判別分析

有意水準 α (%): 5

2群のサンプルの判別分析

対象を指定して判別

群の平均分散(1変量の場合)

第1群個数: 9 平均値1: 56.33333 (不偏)分散値1: 63

第2群個数: 7 平均値2: 45.14286 (不偏)分散値2: 50.47619

群の平均分散(多変量の場合)

第1群個数: 平均値1: 不偏分散共分散行列1: 表示

第2群個数: 平均値2: 不偏分散共分散行列2: 表示

2群の分散の評価結果

等分散の評価: ☒ 等分散とみなせる ($\Sigma(1) = \Sigma(2)$) ☐ 等分散とみなせない ($\Sigma(1) \neq \Sigma(2)$)

等分散の時: 判別式による評価

P1 (第1群に属するのに、第2群と誤判別される確率 (%)): 23.05526

P2 (第2群に属するのに、第1群と誤判別される確率 (%)): 23.05526

この場合は1変量データの為、上図の赤の部分に、両群の平均、分散が表示されます。

2群の分散が等しいとみなせる(等分散)場合、「等分散とみなせる ($\Sigma(1) = \Sigma(2)$)」

にチェックが入っており、「等分散評価の詳細」ボタンを押すと次のパネルが表示されます。

2群のデータの等分散の検定 (Fisher検定)

有意水準 α (%): 5

両側検定

評価結果

自由度1: 8 平均値1: 56.33333 (不偏)分散値1: 63

自由度2: 6 平均値2: 45.14286 (不偏)分散値2: 50.47619

F値: 1.248113 帰無仮説(等分散)の採択域: (0.2149754 , 5.599623)

P値(%): 80.93661

分布関数

結果: 第1群と第2群とは等分散とみなせる

2グループの等分散検定は、上図のようにF分布を利用します。

F値が帰無仮説の採択域に入り、帰無仮説を棄却できない、つまり、

・第1グループと第2グループは 等分散とみなせる

という結論です。

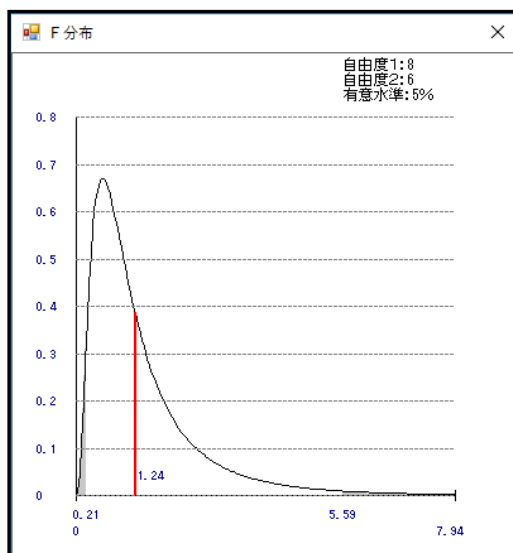
その様子は、「分布関数」ボタンを押すことで、確認できます。

次図のように F 値が 1.24 で、帰無仮説の棄却域は

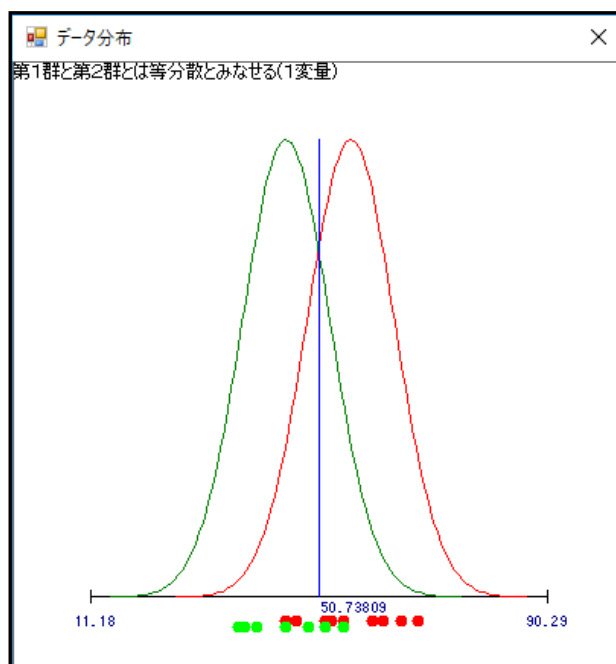
$[0, 0.21]$ 又は $[5.59, \infty]$

です。逆に 帰無仮説の採択域は $[0.21, 5.59]$ ですので、

F 値は 帰無仮説を 棄却できないことを 示します。



等分散の時には 「判別式による評価」ボタンを押すことができます。
次の図が表示されます。



この図から赤い曲線と青い曲線との交点の位置 50.7 を合否判定の基準とできます。

それでは 実際に 点数を入力して、可否を予想します。

対象を指定して判別 ボタンを 押します。

判別計算のパネルが表示されます。

NO	説明変数名	値
1	得点	0

第1群との汎距離 第2群との汎距離

判別結果

☒ 第1群に所属と判別 ☐ 第2群に所属と判別

判別図表示

得点欄に数字を、例えば 5 6 を入力し、**判別** ボタンを押します。

下に結果が表示されます。

第1群との汎距離 0.00192791 第2群との汎距離 2.045326

判別結果

☒ 第1群に所属と判別 ☐ 第2群に所属と判別

判別図表示

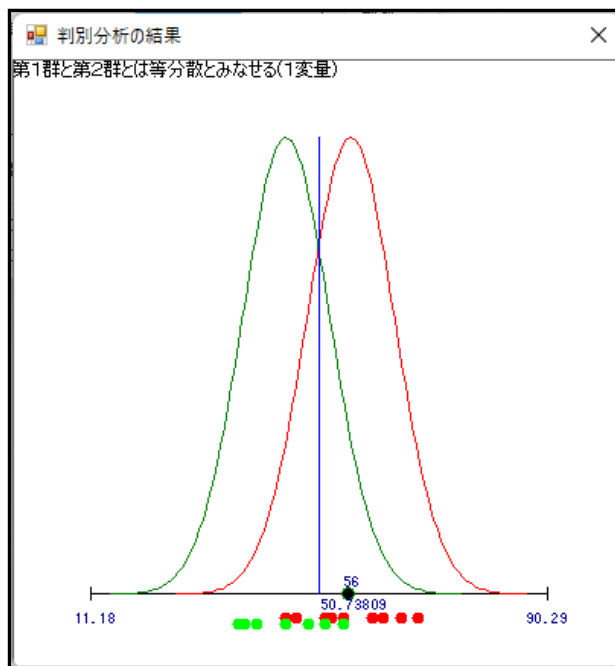
第1群の正規分布の平均位置から指定点までの距離は 0.0019 であり、

第2群の正規分布の平均位置から指定点までの距離は 2.0453 で、

第1群との 距離が近いということになります。

(なおここでの距離とはマハラノビス距離です)

また、**判別図表示** ボタンを押すと、



先ほどの図に 入力点数の位置を 黒丸で 示した図が 表示されました。

確かに 黒丸の位置は 判定基準の 50.7 より 大きく、合格圏内にあることが判ります。

ただし、この基準は 確実 というわけではなく、以下が示すように、

P 1 (第 1 群に属するのに、第 2 群と誤判別される確率 %):	23.05526
P 2 (第 2 群に属するのに、第 1 群と誤判別される確率 %):	23.05526

- ・ 第 2 群 (不合格) と判定したけど、実は第 1 群 (合格) だった
- ・ 第 1 群 (合格) と判断したけど、実は第 2 群 (不合格) だった

となりうる確率が 23% です と言ってます。

- (4) 変量数が 2つの場合
以下を例に説明します。
ある病院の 血液検査の結果です。
T C (総コレステロール)、T G (中性脂肪) が記述されています。
2 グループに分けていますが、左は 動脈硬化症患者 で 右は 正常者です。

動脈硬化症			正常		
ID.	TC	TG	ID.	TC	TG
動脈硬化-1	220	110	正常-1	180	130
動脈硬化-2	230	150	正常-2	180	150
動脈硬化-3	240	150	正常-3	190	160
動脈硬化-4	240	250	正常-4	190	180
動脈硬化-5	250	200	正常-5	200	160
動脈硬化-6	260	150	正常-6	200	170
動脈硬化-7	260	250	正常-7	200	240
動脈硬化-8	260	290	正常-8	210	160
動脈硬化-9	270	250	正常-9	210	180
動脈硬化-10	280	290	正常-10	210	250
			正常-11	220	180
			正常-12	220	260
			正常-13	220	300
			正常-14	230	250
			正常-15	240	320

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。

動脈硬化症			正常		
ID.	TC	TG	ID.	TC	TG
動脈硬化-1	220	110	正常-1	180	130
動脈硬化-2	230	150	正常-2	180	150
動脈硬化-3	240	150	正常-3	190	160
動脈硬化-4	240	250	正常-4	190	180
動脈硬化-5	250	200	正常-5	200	160
動脈硬化-6	260	150	正常-6	200	170
動脈硬化-7	260	250	正常-7	200	240
動脈硬化-8	260	290	正常-8	210	160
動脈硬化-9	270	250	正常-9	210	180
動脈硬化-10	280	290	正常-10	210	250
			正常-11	220	180
			正常-12	220	260
			正常-13	220	300
			正常-14	230	250
			正常-15	240	320

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

つまり、

- ・ 上図の左側表の部分をヘッダも含めてドラッグ&コピーして、第1群のデータ側の「表データを貼り付け」ボタンをクリックします。
- ・ 上図の右側表の部分をヘッダも含めてドラッグ&コピーして、第2群のデータ側の「表データを貼り付け」ボタンをクリックします。

なお、データ入力の決まりとして、各データには ID を必ずつけることとします。

☒ 先頭行をラベルとして使用

データ分布1

データ分布2

第1群のデータ

表データを貼り付け

クリア

	No.	ID.	TC	TG
▶	1	動脈硬化-1	220	110
	2	動脈硬化-2	230	150
	3	動脈硬化-3	240	150
	4	動脈硬化-4	240	250

第2群のデータ

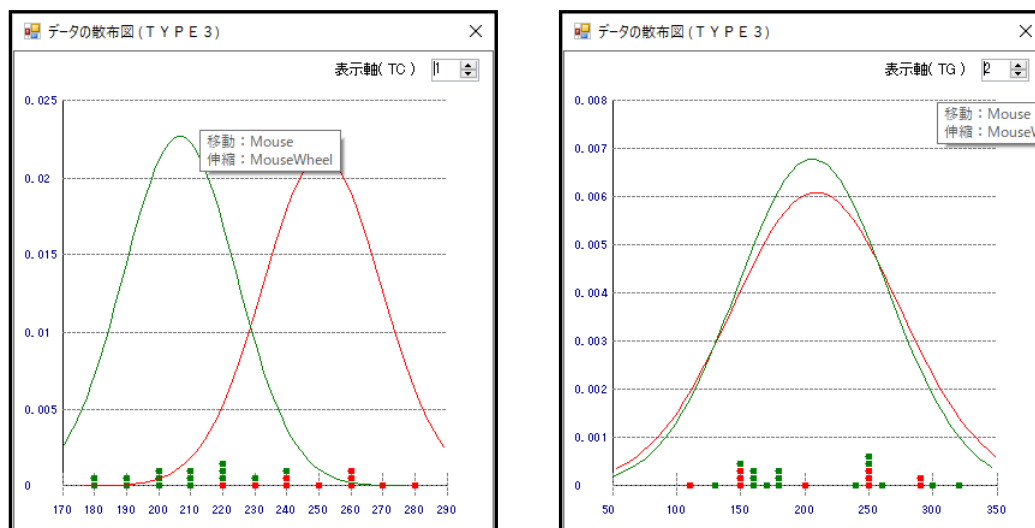
表データを貼り付け

クリア

	No.	ID.	TC	TG
▶	1	正常-1	180	130
	2	正常-2	180	150
	3	正常-3	190	160
	4	正常-4	190	180

2つのグリッドにデータが入力されました。

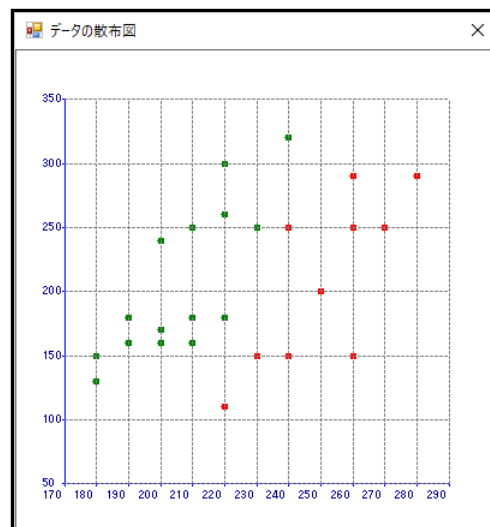
ここで、「データ分布2」ボタンを押すと、次のようなグラフが表示されます。



左図は変量をTCにして分布を表示したもの、右図は変量をTGにして分布を表示したものです。

どちらの変量を見ても、赤い点 と 緑の点との重なる部分が多く、特に右の分布曲線はほぼ重なっているため、赤い点（患者）と緑の点（正常）との分離は難しそうです。

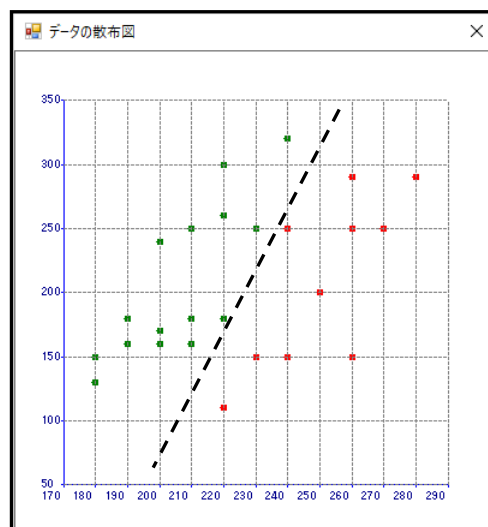
次に 「データ分布 1」 ボタンを押すと、次のようなグラフが表示されます。



図の 横軸はTCの値、縦軸はTGの値 を取ります。

1つの変量で見たときには、分からなかったのですが、2次元でデータを見ると、様子が違って見えます。

つまり、



のように 斜めに線を引くと、 赤の点、緑の点が 分離できそうです。

2変量の 判別分析は このようなアイデアで 分離を図るものです。

データが グリッドにコピーされると、パネル上部の

2群のサンプルの判別分析

ボタンが押せる状態になります。

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。

2群のデータによる判別分析

有意水準 α (%) : 5 2群のサンプルの判別分析 対象を指定して判別

群の平均分散(1変量の場合)

第1群個数 : 平均値1 : (不偏)分散値1 :

第2群個数 : 平均値2 : (不偏)分散値2 :

群の平均分散(多変量の場合)

第1群個数 : 10 平均値1 : TC : 251 不偏分散共分散行列1 : 表示

第2群個数 : 15 平均値2 : TC : 206.6667 不偏分散共分散行列2 : 表示

2群の分散の評価結果

等分散の評価 : ☒ 等分散とみなせる ($\Sigma(1) = \Sigma(2)$) 等分散評価の詳細

☐ 等分散とみなせない ($\Sigma(1) \neq \Sigma(2)$)

等分散の時 : 判別式による評価

P1 (第1群に属するのに、第2群と誤判別される確率 %) : 2.788576

P2 (第2群に属するのに、第1群と誤判別される確率 %) : 2.788576

この場合は2変量データの為、上図の赤の部分に、両群の平均、分散が表示されます。

平均値1 : TC : 251 TC : 251
TG : 209

平均値2 : TC : 206.6667 TC : 206.6667
TG : 206

2群の分散が等しいとみなせる(等分散)場合、「等分散とみなせる ($\Sigma(1) = \Sigma(2)$)」

にチェックが入っており、「等分散評価の詳細」ボタンを押すと次のパネルが表示されます。

2群のデータの等分散の検定

2群のデータの等分散の検定 (χ 自乗検定)

有意水準 α (%) : 5

評価結果

自由度 : 3

χ 自乗値 : 0.1060652 帰無仮説(等分散)の採択域 : (0 , 7.814716)

P値(%) : 99.10995 分布関数

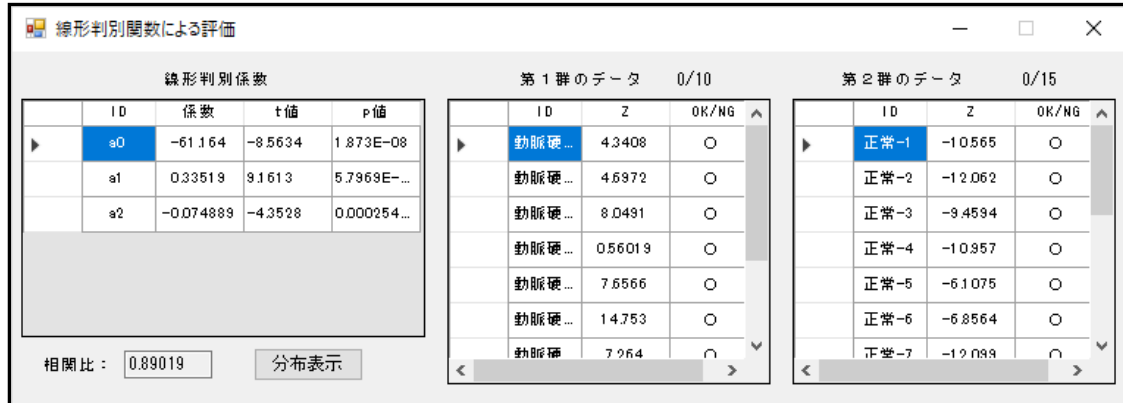
結 果 : 第1群と第2群とは等分散とみなせる

この場合の等分散検定は χ 自乗分布を用います。

χ 自乗値が 0.10 で、帰無仮説の採択域は [0.0, 7.8] なので、帰無仮説を棄却できない、つまり 等分散とみなす ということになります。

続いて **判別式による評価** ボタン を押します。

次の画面が表示されます。



左側の 線形判別係数は 2 グループを分離する直線の式を 表します。
直線の式は、

$$a_0 + a_1 * X + a_2 * Y = 0$$

であり、 上図により、 各係数は

$$a_0 = -61.164$$

$$a_1 = 0.335$$

$$a_2 = -0.074$$

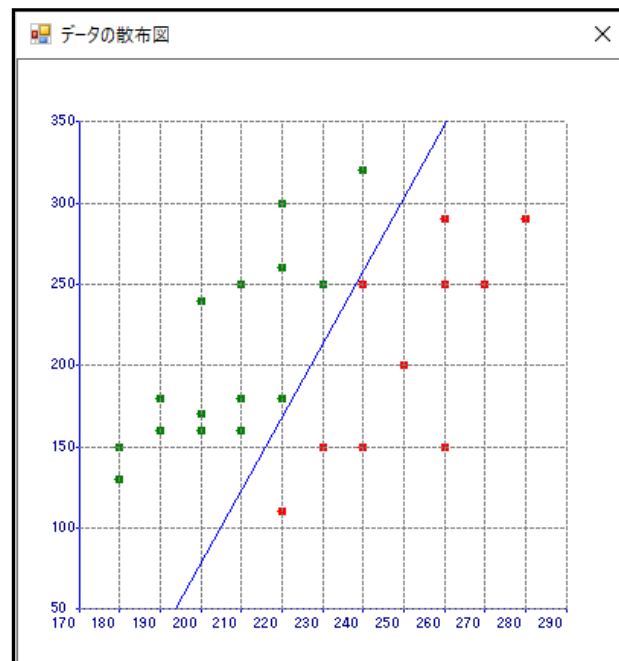
となります。

相関比は 0.89 で 分離の程度はいい とみなせるでしょう。

それぞれの t 値は 絶対値が 2.0 を越えてるので、説明変数としての効果はあることを示しています。

また それぞれの p 値は いずれも ほぼ 0 でするので、これも 説明変数としての効果はあることを示しています。

実際、この直線による 分離の様子は **分布表示** をクリックすると、



と表示され、青い斜めの線が 分離を示す線（線形判別式）となります。

青い線より左側が 緑の点、 右側が 赤の点 と きれいに 分離されています。

第 1 群のデータ 0/10				第 2 群のデータ 0/15			
	ID	Z	OK/NG		ID	Z	OK/NG
▶	動脈硬...	4.3408	○	▶	正常-1	-10.565	○
	動脈硬...	4.6972	○		正常-2	-12.062	○
	動脈硬...	8.0491	○		正常-3	-9.4594	○
	動脈硬...	0.56019	○		正常-4	-10.957	○
	動脈硬...	7.6566	○		正常-5	-6.1075	○
	動脈硬...	14.753	○		正常-6	-6.8564	○
	動脈硬...	7.264	○		正常-7	-12.099	○

上図の 左表の Z 値 は 第 1 群のデータの各点から分離直線への隔たりを計算したもの、
右表の Z 値は 第 2 群のデータの各点から分離直線への隔たりを計算したものです。

第 1 群の点は Z 値が 正の値を取り、第 2 群の点は Z 値が 負の値を取ります。
つまり 分離直線 を境にして、分離することになります。

第 1 群のデータの 0/10 は、データ数が 10 で、そのうち Z 値が 負のものの数を
表示しています。 この場合は 負のものの数は 0 ですから、すべて 正しく
分離されたことになります。

第 2 群のデータの 0/15 は、データ数が 15 で、そのうち Z 値が 正のものの数を、
表示しています。 この場合も 正のものの数は 0 ですから、すべて 正しく
分離されたことになります。

このように すべての点が いつも 正しく分離されるわけではなく、 場合によっては、
誤って分別される 場合もあります。

それでは 次に 新しい検査データ（TC, TG）を入力して、どちらになるかを
計算します。

対象を指定して判別 ボタンを 押します。

判別計算のパネルが表示されます。

判別分析計算に基づく判定

判別分析計算に基づく判定

NO	説明変数名	値
1	TC	0
2	TG	0

判別

第 1 群との汎距離 第 2 群との汎距離

判別結果
☐ 第 1 群に所属と判別
☐ 第 2 群に所属と判別

判別図表示

判別分析計算に基づく判定

NO	説明変数名	値
1	TC	225
2	TG	200

第1群との汎距離 4.432165 第2群との汎距離 2.985696

判別結果

☐ 第1群に所属と判別

☒ 第2群に所属と判別

判別図表示

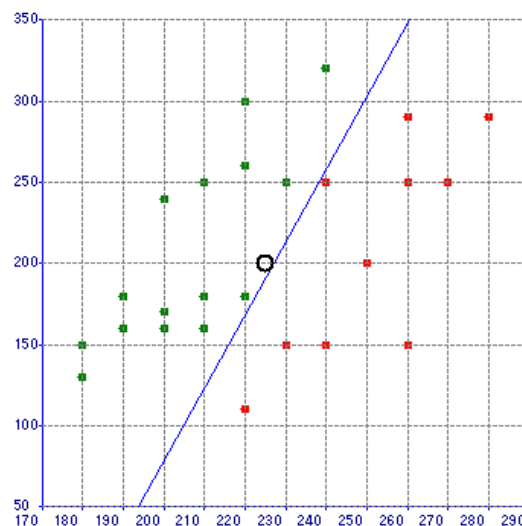
(TC, TG) = (225, 200) のデータを入力し、**判別** ボタンを押すと、

第1群の平均位置から指定点までの距離は 4.43 であり、

第2群の平均位置から指定点までの距離は 2.98 で、

第2群との 距離が近いということになります。
(なおここでの距離とはマハラノビス距離です)

また、**判別図表示** ボタンを押すと、



先ほどの図に 入力点数の位置を 黒丸で 示した図が 表示されました。

黒丸の位置は 分離直線ぎりぎりのところにあり、わずかに 第2群に近いことがわかります。

(5) 変量数が 3 つ以上の場合

変量数が 3 以上についての説明です。ここでは 変量数を3で説明します。
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。

前立腺癌			
ID	γ -Sm 値	尿酸値	PAP 値
佐藤	4.12	6.00	4.52
石井	3.82	5.58	2.13
深井	2.67	4.30	2.64
佐山	3.55	3.55	2.29
尾上	2.49	2.49	3.00
鬼頭	4.81	4.81	4.88

前立腺肥大			
ID	γ -Sm 値	尿酸値	PAP 値
土居	3.21	3.21	2.83
新井	0.95	7.5	2.25
新川	3.47	3.47	4.15
小室	2.16	9.3	1.76
杉山	2.18	3.75	2.59
田中	1.43	6.15	2.21
瀬尾	1.85	4.8	3.01

1 列目は ID, 2～4 列目にデータが記述されています。

2 列目以降は データによって、その列数は変わります。

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

- ・ 上表の左側「前立腺癌」の部分をヘッダも含めてドラッグ&コピーして、第1群のデータ側の「表データを貼り付け」ボタンをクリックします。
- ・ 上表の右側「前立腺肥大」の部分をヘッダも含めてドラッグ&コピーして、第2群のデータ側の「表データを貼り付け」ボタンをクリックします。

先頭行をラベルとして使用をチェック

☒ 先頭行をラベルとして使用

第1群のデータ **表データを貼り付け** クリア

No.	ID	X1
▶		

データ分布1 データ分布2

第2群のデータ **表データを貼り付け** クリア

No.	ID	X1
*		

「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

☒ 先頭行をラベルとして使用

第1群のデータ **表データを貼り付け** クリア

No.	ID	γ -Sm 値	尿酸値	PAP
▶	1 佐藤	4.12	6.00	
	2 石井	3.82	5.58	
	3 深井	2.67	4.30	
	4 佐山	3.55	3.55	

データ分布1 データ分布2

第2群のデータ **表データを貼り付け** クリア

No.	ID	γ -Sm 値	尿酸値	PAP
▶	1 土居	3.21	3.21	
	2 新井	0.95	7.5	
	3 新川	3.47	3.47	
	4 小室	2.16	9.3	

入力データを確認するためのボタンが「データ分布1」と「データ分布2」です。

☒ 先頭行をラベルとして使用

データ分布1

データ分布2

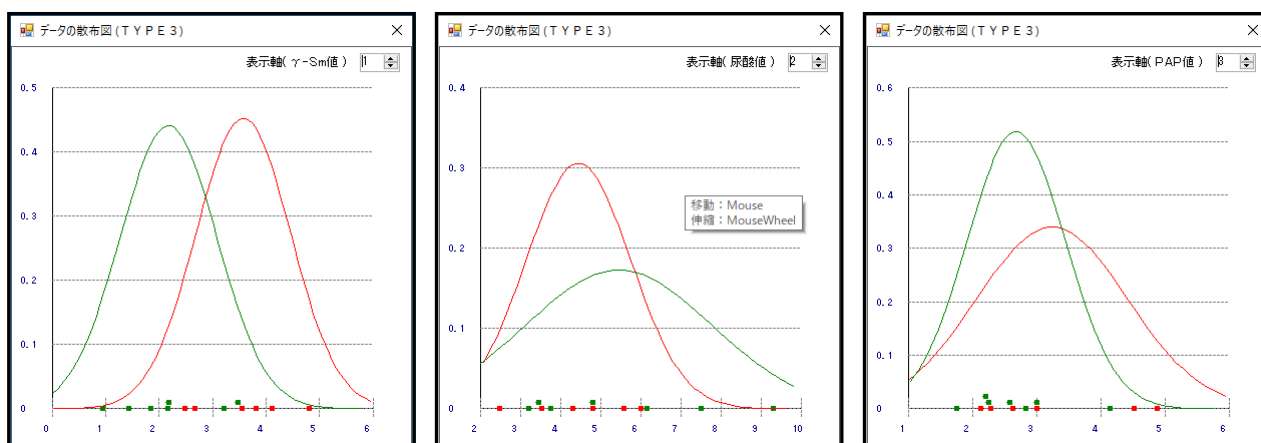
第1群のデータ 表データを貼り付け クリア

No.	ID	γ -Sm値	尿酸値	PAP
1	佐藤	4.12	6.00	
2	石井	3.82	5.58	
3	深井	2.67	4.30	
4	佐山	2.55	2.55	

第2群のデータ 表データを貼り付け クリア

No.	ID	γ -Sm値	尿酸値	PAP
1	土居	3.21	3.21	
2	新井	0.95	7.5	
3	新川	3.47	3.47	
4	小室	0.15	0.2	

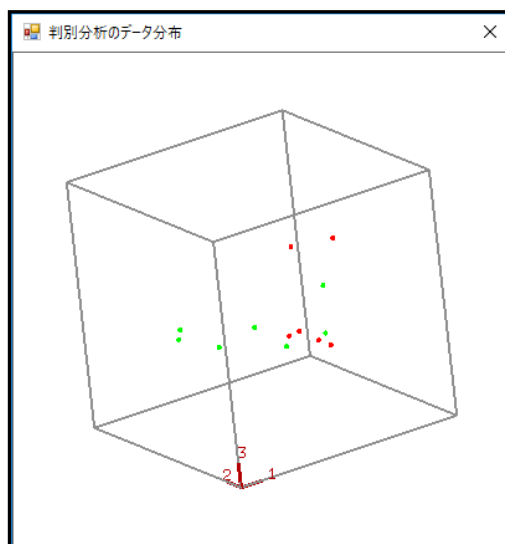
「データ分布2」ボタンを押すと、変量ごとのデータの散布状況を表示します。
 2変量の場合と同様に、1つの変量だけに注目すると、2つを分離するのは難しそうです。



「データ分布1」ボタンを押すと、データの散布状況を表示します。

入力データが3次元の場合、下のように表示して、各点が3次元空間の中でどのように位置されているかが把握できます。赤い点が第1群のデータで、緑の点が第2群のデータです。

マウスを用いて、図を回転させたり、拡大縮小したり、移動ができます。



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。

2群のサンプルの判別分析

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。

2群のデータによる判別分析

有意水準 α (%): 5 2群のサンプルの判別分析 対象を指定して判別

群の平均分散(1変量の場合)

第1群個数: 平均値1: (不偏)分散値1:
第2群個数: 平均値2: (不偏)分散値2:

群の平均分散(多変量の場合)

第1群個数: 6 平均値1: γ -S m値 : 3.576 不偏分散共分散行列1: 表示
第2群個数: 7 平均値2: γ -S m値 : 2.178 不偏分散共分散行列2: 表示

2群の分散の評価結果

等分散の評価: ☒ 等分散とみなせる ($\Sigma(1) = \Sigma(2)$) 等分散評価の詳細
☐ 等分散とみなせない ($\Sigma(1) \neq \Sigma(2)$)

等分散の時: 判別式による評価

P1 (第1群に属するのに、第2群と誤判別される確率 %): 19.95559
P2 (第2群に属するのに、第1群と誤判別される確率 %): 19.95559

この場合は3次元の多変量データの為、上図の赤の部分に、両群の平均、分散が表示されます。

上をクリックすると

下をクリックすると

平均値1: γ -S m値 : 3.576 平均値2: γ -S m値 : 3.576667
尿酸値 : 4.455
P A P 値 : 3.243333

平均値1: γ -S m値 : 3.576 平均値2: γ -S m値 : 2.178
尿酸値 : 5.454286
P A P 値 : 2.685714

分散とみなせる

と表示されます。

2群の分散が等しいとみなせる(等分散)場合、「等分散とみなせる ($\Sigma(1) = \Sigma(2)$)」

にチェックが入っており、「等分散評価の詳細」ボタンを押すと次のパネルが表示されます。

2群のデータの等分散の検定

2群のデータの等分散の検定 (x自乗検定)

有意水準 α (%): 5

評価結果

自由度: 6

x自乗値: 11.793 帰無仮説(等分散)の採択域: (0 , 12.59153)

P値(%): 6.674926 分布関数

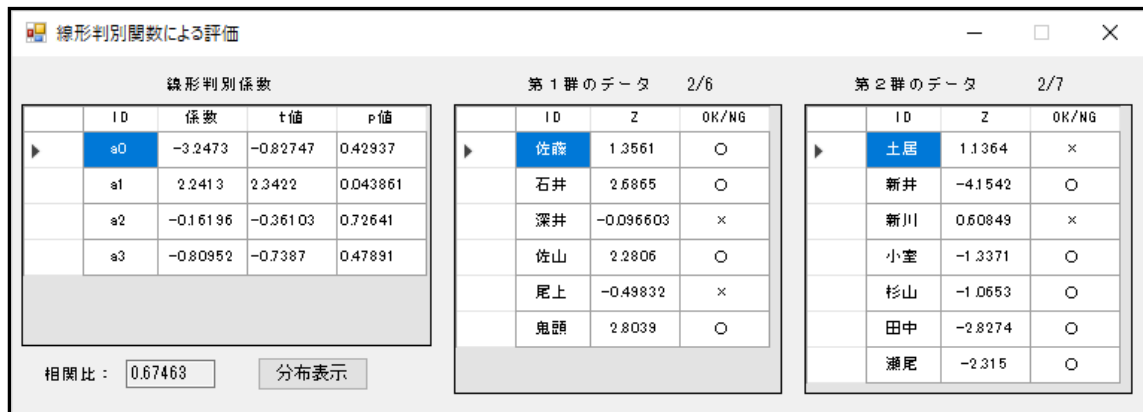
結果: 第1群と第2群とは等分散とみなせる

この場合の等分散検定も χ 自乗分布を用います。

χ 自乗値が 11.79 で、帰無仮説の採択域は [0.0, 12.59] なので、帰無仮説を棄却できない、つまり 等分散とみなす ということになります。

続いて **判別式による評価** ボタン を押します。

次の画面が表示されます。



左側の 線形判別係数は 2 グループを分離する平面の式を 表します。
平面の式は、

$$a_0 + a_1 * X + a_2 * Y + a_3 * Z = 0$$
であり、 上図により、 各係数は

$$\begin{aligned} a_0 &= -3.2 \\ a_1 &= 2.24 \\ a_2 &= -0.16 \\ a_3 &= -0.80 \end{aligned}$$

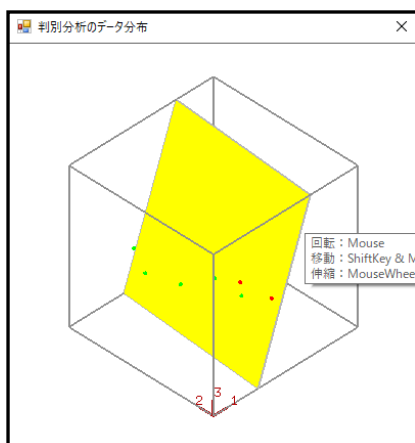
となります。

相関比は 0.67 で、そこそこに 分離できているとはいえ、ばっちりと分離できている とまでは 言えないようです。

t 値をみても、 a_1 は 2.24 で 2.0 を越えてますが、他はそうではないので、説明変数としての効果については 微妙です。

p 値についても、同様のことが言えます。

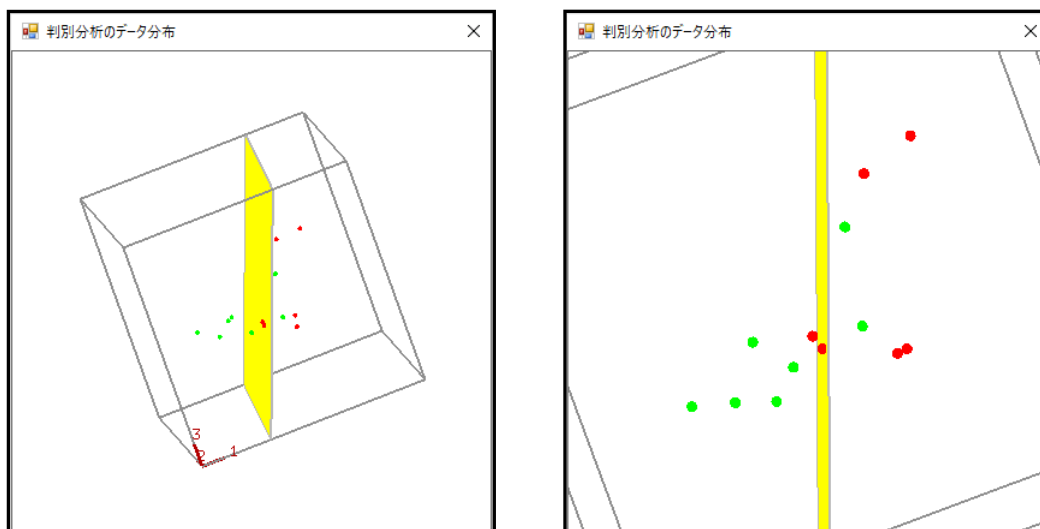
この平面による 分離の様子は **分布表示** をクリックすると、分かります。



3Dで サンプル点 (赤 と 緑) と 分離平面 (黄色) とを重畳表示しています。

マウス操作で、回転、移動、拡大縮小ができますので、見えやすいように、視点を 変更できます。

視点を変えたり、拡大したりして、黄色の分離平面が 赤の点、緑の点 を
どのように分離しているかを 確認することができます。



これを見ると、おおむね 分離できているようには見えますが、
赤の2点 が 緑側に、緑の2点が 赤側に 配置されているのが分かります。

第1群のデータ 2/6				第2群のデータ 2/7			
	ID	Z	OK/NG		ID	Z	OK/NG
▶	佐藤	1.3561	○	▶	土居	1.1364	×
	石井	2.6865	○		新井	-4.1542	○
	深井	-0.096603	×		新川	0.60849	×
	佐山	2.2806	○		小室	-1.3371	○
	尾上	-0.49832	×		杉山	-1.0653	○
	鬼頭	2.8039	○		田中	-2.8274	○
					瀬尾	-2.315	○

その様子は、上図の表からもわかります。

第1群のうちの2点が、分離平面より第2群側に 入っており、 逆に
第2群のうちの2点が、分離平面より第1群側に 入っています。

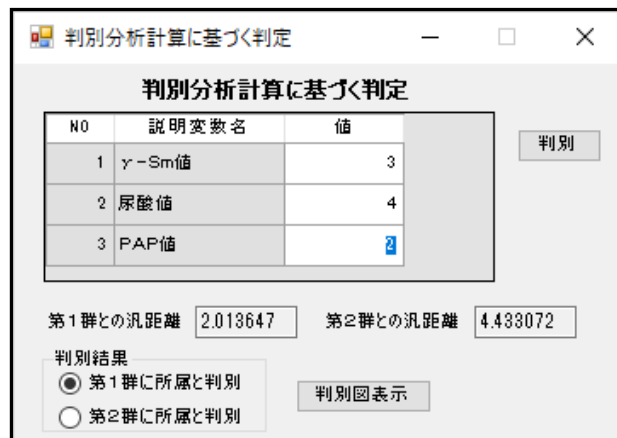
それでは 次に 新しい検査データ (γ -Sm 値、尿酸値、PAP値)を入力して、
どちらになるかを計算します。

対象を指定して判別 ボタンを 押します。

判別計算のパネルが表示されます。

γ -Sm 値 = 3、尿酸値 = 4、PAP 値 = 2

としてデータを入力し、判別ボタンを押すと、



N0	説明変数名	値
1	γ -Sm値	3
2	尿酸値	4
3	PAP値	2

第1群との汎距離 2.013647 第2群との汎距離 4.433072

判別結果
☒ 第1群に所属と判別
☐ 第2群に所属と判別

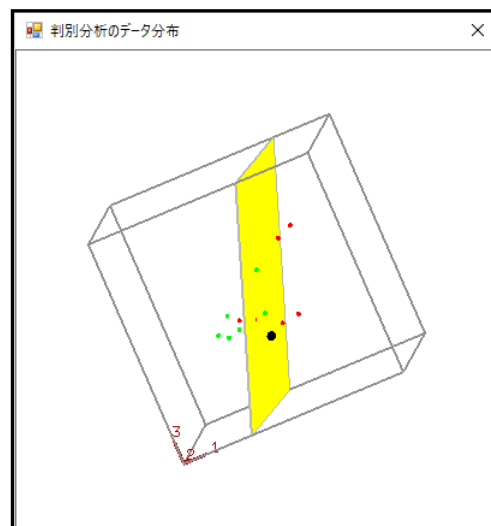
判別図表示

第1群の平均位置から指定点までの距離は 2.01 であり、

第2群の平均位置から指定点までの距離は 4.43 で、

第1群との 距離が近いということになります。
(なおここでの距離とはマハラノビス距離です)

また、**判別図表示** ボタンを押すと、 次の図が表示されます。



入力された 黒い点が、赤い点の側 第1群の側に 配置されているのが分かります。