岡本安晴　2002.8

判別分析

Fisher’s Linear Discriminant Function

1. モデル

各個体から p 変量のデータが得られているとする。これらの個体が g 個のグループに分けられているとき、第 k グループの i 番目の個体のデータを

$$\mathbf{X}_{ki} = \begin{bmatrix} X_{ki1} \\ \vdots \\ X_{kip} \end{bmatrix}$$

と表す。 X_{kij} は第 k グループの i 番目の個体の j 番目の変量の値を表す。

データ全体の平均を

$$\mathbf{m} = \frac{1}{N}\sum_{k=1}^{g}\sum_{i=1}^{N_k}\mathbf{X}_{ki}$$

とおく。ここで、 $N = \sum_{k=1}^{g} N_k$ であり、 N_k は第 k グループのデータ（個体）の総数を表す。

第 k グループの平均を以下のようにおく。

$$\mathbf{m}_k = \frac{1}{N_k}\sum_{i}\mathbf{X}_{ki}$$

データを中心化した（平均を０とした）ものを

$$\mathbf{x}_{ki} = \mathbf{X}_{ki} - \mathbf{m}$$

とおく。このとき、$\mathbf{x}_{ki}$ のデータ全体にわたる平均は

$$\bar{\mathbf{x}}_{\bullet\bullet} = \frac{1}{N}\sum_{k,i}\mathbf{x}_{ki} = \mathbf{m} - \mathbf{m} = \mathbf{0}$$

であり、グループ k における平均は

$$\bar{\mathbf{x}}_{k\bullet} = \frac{1}{N_k}\sum_{i}\mathbf{x}_{ki} = \mathbf{m}_k - \mathbf{m}$$

となる。

判別関数を $\mathbf{x}_{ki}$ の変量の線形結合

$$y_{ki} = \mathbf{v}'\mathbf{x}_{ki}$$

として構成し、 y_{ki} のグループ間の変動とグループ内の変動との比が最大になるように $\mathbf{v}$ を決める。グループ間の変動は以下のように算出される。

$$SS_{between} = \sum_{k}\sum_{i}\left(\mathbf{v}'\bar{\mathbf{x}}_{k\bullet} - \mathbf{v}'\bar{\mathbf{x}}_{\bullet\bullet}\right)^2$$

$$= \sum_k \sum_i (\mathbf{v}'\overline{\mathbf{x}}_{k\bullet})^2$$

$$= \sum_k \sum_i \{\mathbf{v}'(\mathbf{m}_k - \mathbf{m})\}\{\mathbf{v}'(\mathbf{m}_k - \mathbf{m})\}'$$

$$= \mathbf{v}'\left\{\sum_k \sum_i (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})'\right\}\mathbf{v}$$

$$= \mathbf{v}'\mathbf{A}\mathbf{v}$$

ここで、

$$\mathbf{A} = \sum_{k=1}^{g} \sum_{i=1}^{N_k} (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})'$$

$$= \sum_k N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})'$$

である。

グループ内変動は以下のように算出される。

$$SS_{within} = \sum_{k=1}^{g} \left\{ \sum_{i=1}^{N_k} (\mathbf{v}'\mathbf{x}_{ki} - \mathbf{v}'\overline{\mathbf{x}}_{k\bullet})^2 \right\}$$

$$= \sum_{k=1}^{g} \left[\sum_{k=1}^{N_k} \{\mathbf{v}'(\mathbf{X}_{ki} - \mathbf{m}) - \mathbf{v}'(\mathbf{m}_k - \mathbf{m})\}^2 \right]$$

$$= \sum_{k=1}^{g} \left[\sum_{k=1}^{N_k} \{\mathbf{v}'(\mathbf{X}_{ki} - \mathbf{m}_k)\}^2 \right]$$

$$= \sum_{k=1}^{g} \left[\sum_{k=1}^{N_k} \{\mathbf{v}'(\mathbf{X}_{ki} - \mathbf{m}_k)\}\{\mathbf{v}'(\mathbf{X}_{ki} - \mathbf{m}_k)\}' \right]$$

$$= \mathbf{v}'\left\{\sum_{k,i} (\mathbf{X}_{ki} - \mathbf{m}_k)(\mathbf{X}_{ki} - \mathbf{m}_k)'\right\}\mathbf{v}$$

$$= \mathbf{v}'\mathbf{W}\mathbf{v}$$

ここで、

$$\mathbf{W} = \sum_{k=1}^{g} \sum_{i=1}^{N_k} (\mathbf{X}_{ki} - \mathbf{m}_k)(\mathbf{X}_{ki} - \mathbf{m}_k)'$$

である。

級間変動と級内変動の比を

$$Q_0 = \frac{SS_{between}}{SS_{within}} = \frac{\mathbf{v}'\mathbf{A}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}}$$

とおく。Q_0 の最大化を

$$\mathbf{v}'\mathbf{W}\mathbf{v} = 1$$

の条件の下で行う。このために次の目的関数をおく。

$$Q = \mathbf{v}'\mathbf{A}\mathbf{v} - \lambda(\mathbf{v}'\mathbf{W}\mathbf{v} - 1)$$

　Q の偏導関数は次式で与えられる。

$$\frac{\partial Q}{\partial \mathbf{v}} = 2\mathbf{A}\mathbf{v} - 2\lambda\mathbf{W}\mathbf{v}$$

　上式の値を０とおいて次式を得る。

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{W}\mathbf{v} \tag{1}$$

　（１）式を満たす $\mathbf{v}$ を求める。まず、$\mathbf{W}$ の固有値分解を考える。

$$\mathbf{W} = \mathbf{U}\mathbf{L}\mathbf{U}'$$

$$= \mathbf{U}\mathbf{L}^{1/2}\left(\mathbf{U}\mathbf{L}^{1/2}\right)' \tag{2}$$

　（２）式を（１）式に代入すると以下のようになる。

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{U}\mathbf{L}^{1/2}\left(\mathbf{U}\mathbf{L}^{1/2}\right)'\mathbf{v}$$

$$\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\mathbf{A}\left\{\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\right\}' \cdot \left(\mathbf{U}\mathbf{L}^{1/2}\right)'\mathbf{v} = \lambda\left(\mathbf{U}\mathbf{L}^{1/2}\right)'\mathbf{v} \tag{3}$$

　ここで、

$$\left[\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\mathbf{A}\left\{\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\right\}'\right]' = \left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\mathbf{A}\left\{\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\right\}'$$

が成り立っている。すなわち、λ と $\left(\mathbf{U}\mathbf{L}^{1/2}\right)'\mathbf{v}$ は実対称行列 $\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\mathbf{A}\left\{\left(\mathbf{U}\mathbf{L}^{1/2}\right)^{-1}\right\}'$ の固有値と固有ベクトルになっている。

　（３）式を満たす λ と $\mathbf{v}$ は（１）式より次式を満たす。

$$\mathbf{v}'\mathbf{A}\mathbf{v} = \lambda\mathbf{v}'\mathbf{W}\mathbf{v}$$

　したがって、

$$Q_0 = \frac{SS_{between}}{SS_{within}} = \frac{\mathbf{v}'\mathbf{A}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}} = \lambda$$

が成り立つ。すなわち、固有値 λ は群間変動量と群内変動量の比の値になっている。

　0 でない固有値の数を n とおき、n 個の固有値は

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$$

と並べられているとする。

　j 番目の固有値に対する固有ベクトル $\mathbf{v}_j$ によって構成される判別関数の分散 θ_j は以下のように計算される。

$$y_{ki}^{(j)} = \mathbf{v}'_j \mathbf{x}_{ki} = \mathbf{v}'_{ki} \left(\mathbf{X}_{ki} - \mathbf{m} \right)$$

とおけば、

$$\begin{aligned}
\theta_j &= \frac{1}{N} \sum_{k=1}^{g} \sum_{i=1}^{N_k} \left(y_{ki}^{(j)} \right)^2 \\
&= \frac{1}{N} \sum_{k,i} \left\{ \mathbf{v}'_j \left(\mathbf{X}_{ki} - \mathbf{m} \right) \right\} \left\{ \mathbf{v}'_j \left(\mathbf{X}_{ki} - \mathbf{m} \right) \right\}' \\
&= \mathbf{v}'_j \left\{ \frac{1}{N} \sum_{k,i} \left(\mathbf{X}_{ki} - \mathbf{m} \right) \left(\mathbf{X}_{ki} - \mathbf{m} \right)' \right\} \mathbf{v}_j \\
&= \mathbf{v}'_j \left(\frac{1}{N} \mathbf{T} \right) \mathbf{v}_j
\end{aligned}$$

となる。ここで、

$$\mathbf{T} = \sum_{k=1}^{g} \sum_{i=1}^{N_k} \left(\mathbf{X}_{ki} - \mathbf{m} \right) \left(\mathbf{X}_{ki} - \mathbf{m} \right)'$$

はデータの全変動量を表す。

　したがって、j 番目の判別関数を平均が０、分散を１になるように標準化したもの（因子、factor）は以下のように与えられる。

$$\begin{aligned}
f_{ki}^{(j)} &= \frac{1}{\sqrt{\theta_j}} y_{ki}^{(j)} \\
&= \frac{1}{\sqrt{\theta_j}} \mathbf{v}'_j \mathbf{x}_{ki} \\
&= \frac{1}{\sqrt{\theta_j}} \mathbf{v}'_j \left(\mathbf{X}_{ki} - \mathbf{m} \right) \\
&= \mathbf{b}'_j \left(\mathbf{X}_{ki} - \mathbf{m} \right)
\end{aligned}$$

ここで、

$$\mathbf{b}_j = \theta_j^{-1/2} \mathbf{v}_j$$

とおいた。

　$\mathbf{X}_{ki}$ あるいは $\mathbf{x}_{ki}$ を標準化した値は次式で与えられる。

$$\mathbf{z}_{ki} = \mathbf{D}_{diag}^{-1/2}\mathbf{x}_{ki} = \mathbf{D}_{diag}^{-1/2}(\mathbf{X}_{ki} - \mathbf{m})$$

ここで、

$$\mathbf{D}_{diag} = diag\left(\frac{1}{N}\mathbf{T}\right)$$

である。

　以上より、$\mathbf{z}_{ki}$ と $f_{ki}^{(j)}$ との相関係数ベクトル $\mathbf{s}^{(j)}$ は以下のように算出される。

$$\begin{aligned}
\mathbf{s}^{(j)} &= \frac{1}{N}\sum_{k,i}\mathbf{z}_{ki} f_{ki}^{(j)} \\
&= \frac{1}{N}\sum_{k,i}\mathbf{z}_{ki}(\mathbf{b}'_j\mathbf{x}_{ki})' \\
&= \frac{1}{N}\sum_{k,i}\mathbf{z}_{ki}(\mathbf{D}_{diag}^{1/2}\mathbf{z}_{ki})'\mathbf{b}_j \\
&= \left(\frac{1}{N}\sum_{k,i}\mathbf{z}_{ki}\mathbf{z}'_{ki}\right)\mathbf{D}_{diag}^{1/2}\mathbf{b}_j \\
&= \mathbf{R}\mathbf{c}_j
\end{aligned}$$

ここで、

$$\mathbf{R} = \frac{1}{N}\sum_{k,i}\mathbf{z}_{ki}\mathbf{z}'_{ki}\text{、}$$

$$\mathbf{c}_j = \mathbf{D}_{diag}^{1/2}\mathbf{b}_j$$

である。

　したがって、因子構造行列は次式で与えられる。

$$\begin{aligned}
\mathbf{S} &= \begin{bmatrix}\mathbf{s}^{(1)} & \cdots & \mathbf{s}^{(n)}\end{bmatrix} \\
&= \mathbf{R}\begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_n\end{bmatrix} \\
&= \mathbf{RC}
\end{aligned}$$

ここで、

$$\mathbf{C} = \begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_n\end{bmatrix}$$

とおいた。

k 番目までの判別関数を採用したとき、残りの $n-k$ 個の判別関数による判別が有意であることの χ^2 検定は次式により行うことができる。

$$\chi^2 = -\left(N - \frac{p+g}{2} - 1\right)\log\Lambda'、\quad df = (p-k)(g-k-1) \qquad （4）$$

ここで、

$$\Lambda' = \prod_{j=k+1}^{n} \frac{1}{1+\lambda_j}$$

であり、p は $\mathbf{X}_{ki}$ における変数（変量）の数、g はグループの数、N はデータの総数 $\sum_k N_k$ である。

$k=0$ のときの（4）式は、グループ間の差を 1 元配置の多変量分散分析で検定するものとなっている。

2. プログラム

プログラム PDiscrim.dpr は、判別分析を行うものである。このプログラムを実行すると図 2.1 のフォームが表示される。

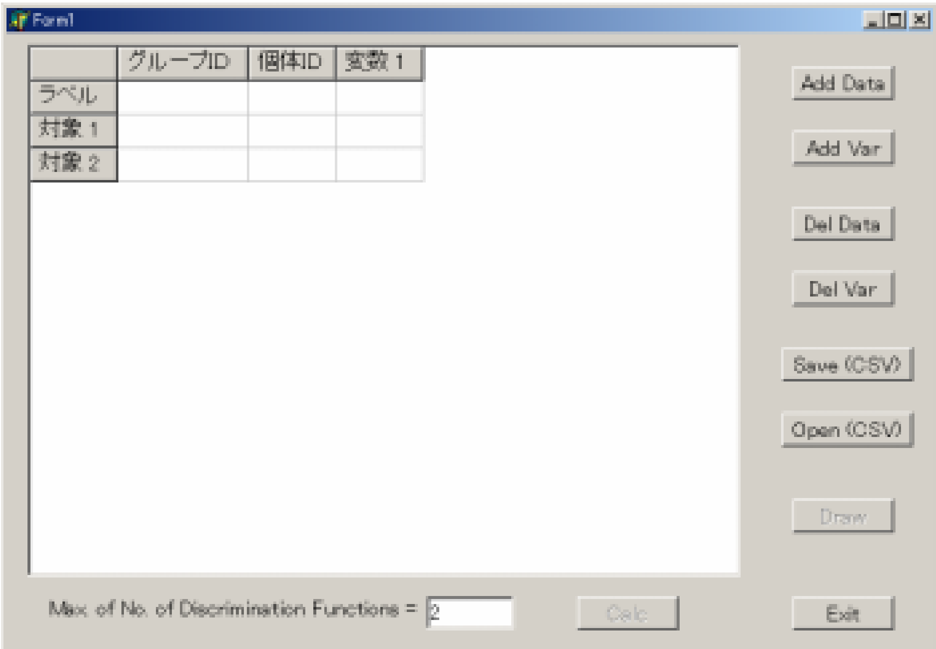

図 2.1　実行開始時のフォーム

「Add Data」ボタンのクリックでグリッドのアクティブなセルの下に行が追加挿入され、「Add Var」ボタンのクリックでアクティブなセルの右側に列が追加挿入される。グリッド内のセルはそのクリックによりアクティブになる。行数と列数をデータに合わせて必要なだけ用意してデータ値を図 2.2 のように設定する。行あるいは列の追加挿入はデータ設定作

業の途中においてもできる。「Del Data」ボタンあるは「Del Var」ボタンをクリックすると、アクティブなセルを含む行あるいは列が削除される。

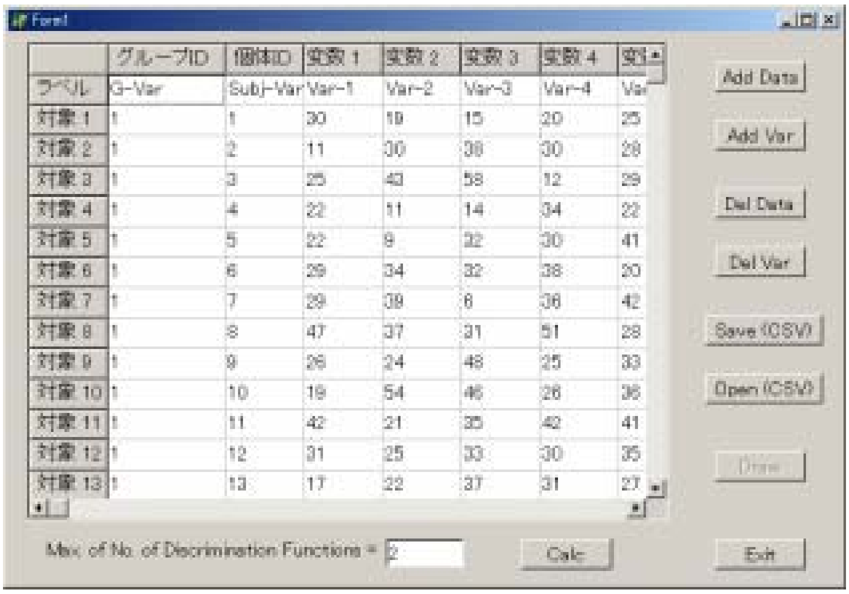

図 2.2　データの設定

グリッド内のラベル行は、各列の変数に対するラベルとして適当な文字列を設定する。ラベル行の次行からデータ値を設定していく。グループＩＤの欄には、データの属するグループを示す文字を設定する。文字は半角文字 1 字である。個体ＩＤの欄は個体の識別文字列を設定する。図 2.2 のデータの場合は図 2.3 に示されるように、グループごとに通し番号が付けられている。

図 2.3　通し番号の例

変数１の欄から各変数（変量）の値を設定していく。設定されたデータは「Save(CVS)」ボ

タンのクリックでファイルに保存できる。保存したデータは「Open(CSV)」ボタンのクリックで読み込むことができる。データはＣＳＶ形式で保存されるので Excel などで開いて見ることもできる。図 2.2 のデータを CSV 形式で保存してものを Excel で開くと図 2.4 のようになる。

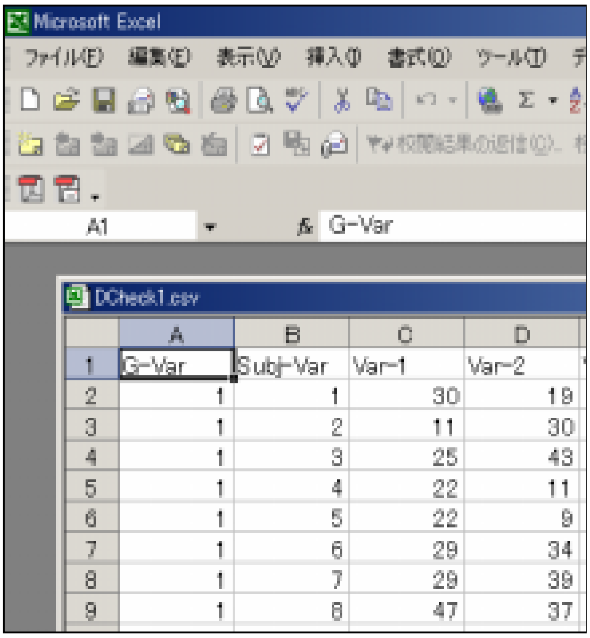

図 2.4　Excel で開いたデータ

逆に、Excel で図 2.4 の形式で用意したデータは、拡張子を.csv として CSV 形式で保存すると「Open(CSV)」ボタンのクリックでグリッド内に読み込むことができる。

データの設定後、グリッドの下のエディットコンポーネント内に求める判別関数の最大数を設定する。判別関数は 0 でない固有値の数だけ求められるが、変数の数が多いときは大量の出力になる。求められる判別関数の最大個数を適当な数に設定しておくことにより大量の出力を避けることができる。

判別関数の最大個数の設定後、「Calc」ボタンをクリックすると計算が始まる。「Calc」ボタンのクリックでまず図 2.5 のようなダイアログボックスが表示される。

図 2.5　「Calc」ボタンのクリックで表示されるダイアログボックス

図 2.5 のダイアログボックスで設定した名前のテキストファイルに計算結果が書き出される。テキストファイルなのでプログラムの実行終了後にエディタなどで開いて見ることができる。ファイル名の設定後、「開く」ボタンをクリックすると計算が始まり、終了すると図 2.6 のメッセージボックスが表示される。

図 2.6　計算終了時に表示されるメッセージボックス

このメッセージボックスには図 2.5 で設定したファイル名が表示される。「OK」ボタンのクリックで図 2.7 のようなフォームに戻る。

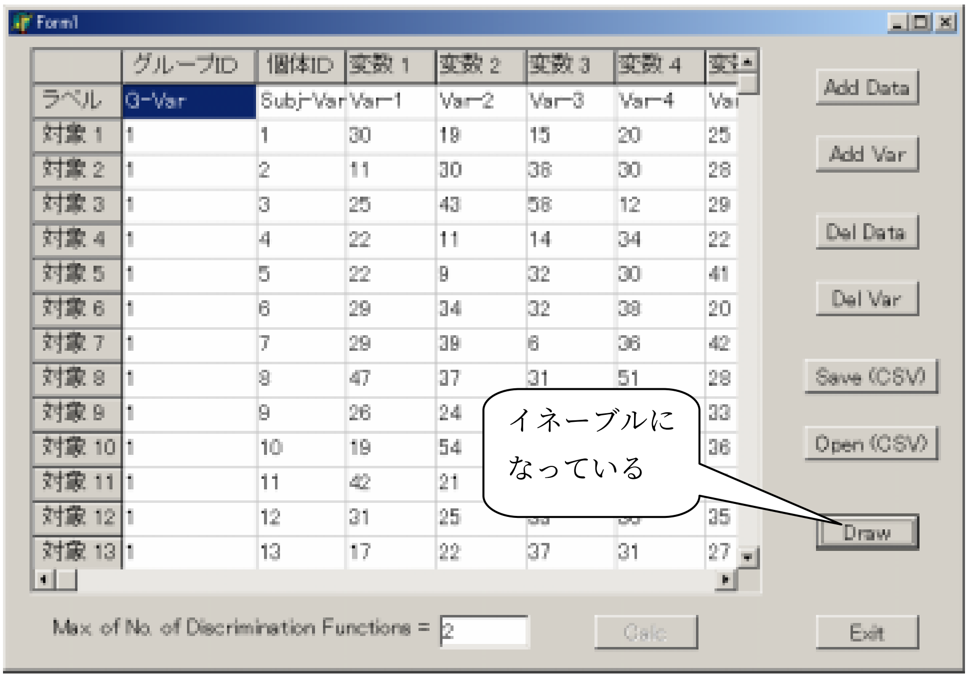

図 2.7　「Draw」ボタンがイネーブルになっている

計算終了後の図 2.7 のフォームでは「Draw」ボタンがイネーブルになっている。「Draw」ボタンは判別関数の値の分布を図示するためのものである。「Draw」ボタンのクリックで図 2.7 のフォームが表示される。

図 2.7　判別関数の選択

横軸にとる判別関数を第 1 軸の次元として設定し、縦軸にとる判別関数を第 2 軸の次元として設定した後、「OK」ボタンをクリックする。「OK」ボタンのクリックで図 2.8 のように判別関数の値の分布図が描画される。

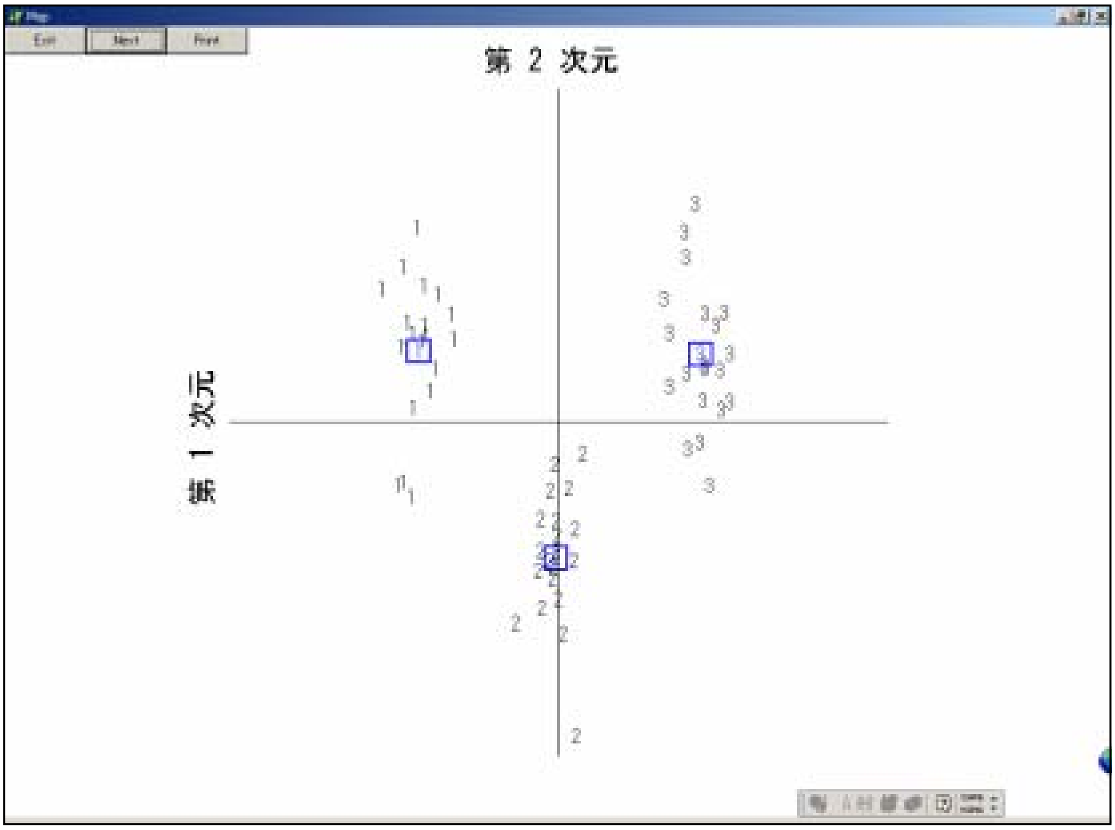

図 2.8　判別関数の図示

個々のデータに対する判別関数の値に対応する位置にグループ ID の文字が書かれ、グループごとの平均値を表す点が内部にグループ ID の文字が書かれた青の小正方形で表示されている。

図 2.8 の画面の左上の「Next」ボタンのクリックで図 2.7 のフォームが再表示されるので、判別関数を選びなおして描画を行うことができる。「Print」ボタンのクリックで描画されているものがプリンタに出力される。「Exit」ボタンのクリックで図 2.9 のフォームに戻る。

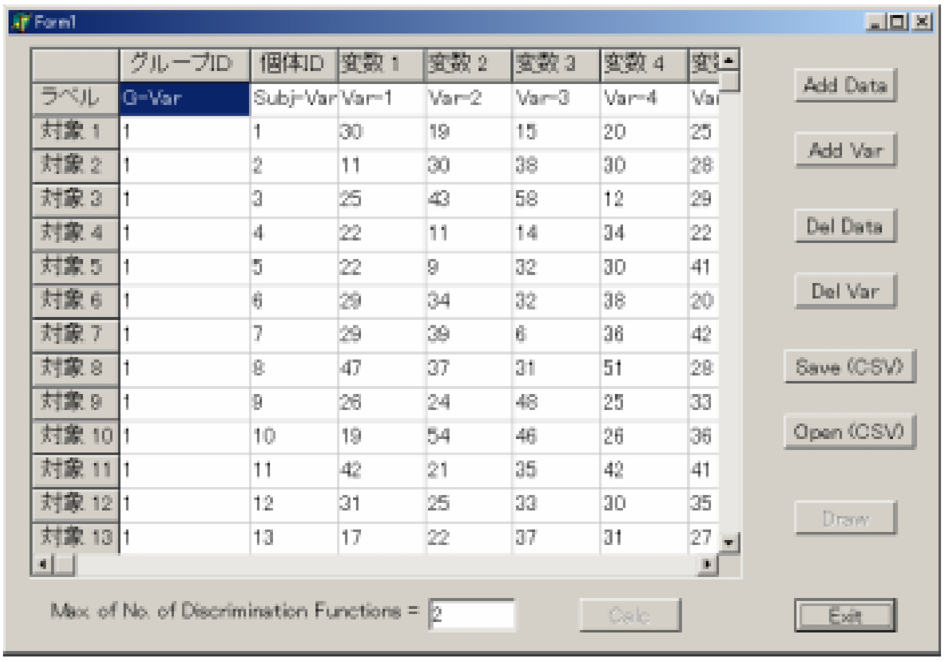

図 2.9　描画後のフォーム

図 2.9 のフォームの「Exit」ボタンのクリックでプログラムの実行終了となる。

プログラムの実行終了後、図 2.5 で設定した名前のテキストファイルを開くとリスト 2.1 のような内容になっている。

リスト 2.1　出力ファイルの内容

```
Variables =
Group ID   ==> G-Var
Subject ID ==> Subj-Var
Var  1 ==> Var-1
Var  2 ==> Var-2
Var  3 ==> Var-3
Var  4 ==> Var-4
Var  5 ==> Var-5
Var  6 ==> Var-6
Var  7 ==> Var-7
Var  8 ==> Var-8
Var  9 ==> Var-9
Var 10 ==> Var-10
Var 11 ==> Var-11
Var 12 ==> Var-12
Var 13 ==> Var-13
Var 14 ==> Var-14
Var 15 ==> Var-15

Data =
    1    1   30   19   15   20   25   43   40   23   29   38    9    0   14  -17    7
    1    2   11   30   38   30   28   44   18   28   33   40    0   -2    5   11  -16
    1    3   25   43   58   12   29   41   29   20   39   23   12    6   18    7    7
                                  ・
                                  ・
                                  ・
    3   18   53   47   68   57   68   79   68   56   67   82    1  -18  -21  -16    6
```

```
    3   19   61   80   72   61   78   58   65   77   76   57   -2  -14    5   -1    1
    3   20   67   79   76   77   62   72   66   77   83   81    8   -3    5   14    6

No. of Groups = 3
Group-1    ID Char. = 1
Group-2    ID Char. = 2
Group-3    ID Char. = 3

Means and SDs (Global) =
Var-1 <Var-1>           mean = 53.2           sd = 19.39227
Var-2 <Var-2>           mean = 52.4           sd = 21.50364
Var-3 <Var-3>           mean = 54.53333       sd = 18.79403
Var-4 <Var-4>           mean = 53.16667       sd = 19.3099
Var-5 <Var-5>           mean = 54.13333       sd = 18.96002
Var-6 <Var-6>           mean = 46.95          sd = 20.48367
Var-7 <Var-7>           mean = 47.03333       sd = 18.42821
Var-8 <Var-8>           mean = 50.15          sd = 19.11616
Var-9 <Var-9>           mean = 45.2           sd = 20.70411
Var-10 <Var-10>         mean = 45.73333       sd = 18.5552
Var-11 <Var-11>         mean = 0.8333333      sd = 10.50265
Var-12 <Var-12>         mean = 0.55           sd = 12.25755
Var-13 <Var-13>         mean = 2.316667       sd = 10.67785
Var-14 <Var-14>         mean = -1.366667      sd = 8.631274
Var-15 <Var-15>         mean = -0.95          sd = 8.444772

Means for Group 1 <1>    Nk = 20
Var-1 <Var-1>           mean = 29.85          sd = 9.76358
Var-1 <Var-1>           mean = 28.45          sd = 12.35101
Var-1 <Var-1>           mean = 32.65          sd = 11.81218
Var-1 <Var-1>           mean = 30.35          sd = 8.368244
Var-1 <Var-1>           mean = 32.3           sd = 9.80867
Var-1 <Var-1>           mean = 29.1           sd = 8.938121
Var-1 <Var-1>           mean = 30.25          sd = 9.648186
Var-1 <Var-1>           mean = 33.7           sd = 12.38588
Var-1 <Var-1>           mean = 28.8           sd = 8.219489
Var-1 <Var-1>           mean = 29.4           sd = 8.475848
Var-1 <Var-1>           mean = 0.85           sd = 8.386149
Var-1 <Var-1>           mean = 0.05           sd = 10.04229
Var-1 <Var-1>           mean = 2.8            sd = 8.902809
Var-1 <Var-1>           mean = -2.7           sd = 9.066973
Var-1 <Var-1>           mean = -1.05          sd = 9.046961

Means for Group 2 <2>    Nk = 20
Var-2 <Var-2>           mean = 59.65          sd = 7.805607
Var-2 <Var-2>           mean = 57.65          sd = 10.62674
Var-2 <Var-2>           mean = 60.5           sd = 8.303614
Var-2 <Var-2>           mean = 58.55          sd = 9.54712
Var-2 <Var-2>           mean = 61.1           sd = 12.52158
Var-2 <Var-2>           mean = 39.1           sd = 8.3
Var-2 <Var-2>           mean = 42.35          sd = 8.53976
Var-2 <Var-2>           mean = 45.1           sd = 11.99958
Var-2 <Var-2>           mean = 35.6           sd = 8.996666
Var-2 <Var-2>           mean = 40.75          sd = 9.218867
Var-2 <Var-2>           mean = 1.85           sd = 10.52746
Var-2 <Var-2>           mean = 1.85           sd = 14.45087
Var-2 <Var-2>           mean = 1.85           sd = 10.37919
Var-2 <Var-2>           mean = -1.55          sd = 7.459725
Var-2 <Var-2>           mean = -1.65          sd = 8.284172

Means for Group 3 <3>    Nk = 20
Var-3 <Var-3>           mean = 70.1           sd = 9.974467
Var-3 <Var-3>           mean = 71.1           sd = 13.07249
Var-3 <Var-3>           mean = 70.45          sd = 9.129485
Var-3 <Var-3>           mean = 70.6           sd = 10.1951
Var-3 <Var-3>           mean = 69             sd = 8.899438
```

```
Var-3 <Var-3>          mean = 72.65          sd = 8.320306
Var-3 <Var-3>          mean = 68.5           sd = 9.399468
Var-3 <Var-3>          mean = 71.65          sd = 6.366121
Var-3 <Var-3>          mean = 71.2           sd = 10.01798
Var-3 <Var-3>          mean = 67.05          sd = 11.40384
Var-3 <Var-3>          mean = -0.2           sd = 12.15154
Var-3 <Var-3>          mean = -0.25          sd = 11.76807
Var-3 <Var-3>          mean = 2.3            sd = 12.43423
Var-3 <Var-3>          mean = 0.15           sd = 9.029258
Var-3 <Var-3>          mean = -0.15          sd = 7.894777

No. of Nonzero eigen values = 2

Lambda =
           42.762
           2.0259
           0
           0
           0
           0
           0
           0
           0
           0
           0
           0
           0
           0
           0

After the acceptance of the first 0 discriminant functions
          Chi-square = 244.2994  with df = 30   p = 0.00000
After the acceptance of the first 1 discriminant functions
          Chi-square = 55.36074  with df = 14   p = 0.00000

No. of Discrimination Functions adopted = 2

Eigen Vectors =
  0.19124 -0.30891
  0.16468 -0.06019
  0.23630 -0.28104
  0.41742 -0.36492
  0.33213 -0.34717
  0.56902  0.21945
  0.15389  0.19243
  0.23986  0.11091
  0.15510  0.61570
  0.29949  0.14080
  0.20948 -0.19882
  0.05096 -0.14235
 -0.11026 -0.02664
 -0.09162 -0.10291
  0.10071  0.02793

Means of Discrimination Functions =

Discrimination Function - 1
     Group-1 <1> = 84.08726
     Group-2 <2> = 137.4928
     Group-3 <3> = 194.0015

Discrimination Function - 2
     Group-1 <1> = -4.584198
     Group-2 <2> = -32.65635
     Group-3 <3> = -5.10763
```

```
Discrimination Factors =
    1 Group-1  subject <1>  =     -1.208    1.634
    2 Group-1  subject <2>  =     -1.153    1.143
    3 Group-1  subject <3>  =     -1.143    0.803
                            .
                            .
                            .
   58 Group-3  subject <18> =      1.085    1.595
   59 Group-3  subject <19> =      0.959    0.740
   60 Group-3  subject <20> =      1.432    0.915

Structure Matrix =
                    Discrim. Func. 1    Discrim. Func. 2
Var-1 <Var-1>              0.85321            -0.30427
Var-2 <Var-2>              0.81617            -0.22708
Var-3 <Var-3>              0.82683            -0.29067
Var-4 <Var-4>              0.85749            -0.25784
Var-5 <Var-5>              0.79502            -0.33323
Var-6 <Var-6>              0.88241             0.31385
Var-7 <Var-7>              0.86007             0.20271
Var-8 <Var-8>              0.82286             0.21212
Var-9 <Var-9>              0.85107             0.38399
Var-10 <Var-10>            0.84102             0.21556
Var-11 <Var-11>           -0.04241            -0.08283
Var-12 <Var-12>           -0.01134            -0.09144
Var-13 <Var-13>           -0.01883             0.03814
Var-14 <Var-14>            0.13660             0.01567
Var-15 <Var-15>            0.04498             0.07076
```

リスト 2.1 では、まず各変数のラベルが出力され、読み込んだデータ値が書き出されている。データ値の書き出しに続いて、読み込まれたデータにおけるグループ数とグループ ID の文字が出力され、さらにデータ全体およびグループごとの平均値と標準偏差が出力されている。次に、求まった０でない固有値の数とすべての固有値が書き出されている。

χ^2 の値は式（４）によって算出されたものである。「p =」に続く値は p 値である。その後、以下の計算の対象となる変別関数の個数と対応する固有ベクトルが書き出されている。

続いて、判別関数ごとにグループの平均値が書き出されている。これらの値の分布の様子は、標準化された判別関数（因子、factor）に対するものとして図 2.8 のように視覚的に表示することができる。標準化された判別関数の個体別の値は「Discrimination Factors =」の行に続いて書き出されている。

最後に、データ変数の値と判別関数との相関係数が Structure Matrix（構造行列）として出力されている。構造行列によって、各変数が判別関数にどのように関係しているのかが判る。

参考文献

Cooley,W.W. and Lohnes,P.R. (1971) *Multivariate data analysis.* Wiley.

Mardia,K.V., Kent,J.T. and Bibby,J.M. (1979) *Multivariate analysis*. Academic Press.

Overall,J.E. and Klett,C.J. (1972) *Applied multivariate analysis*. McGraw-Hill Book Company.