

数量化 2 類

グループ（群）別に項目の選択データが与えられているとき、それらのグループを区別（判別）するのに最適な項目選択肢の数量化を行うのが数量化 2 類である。項目選択のデータは、各項目（アイテム） j に対する選択肢（カテゴリ） k から 1 つを選ぶ形式のアイテム・カテゴリ型であるが、岩坪（1987）では、次の表 1 の例があげられている。1 はその選択肢が選ばれたことを、0 は選ばれなかったことを表している。

表 1 グループ別に与えられたアイテム（項目）・カテゴリ（選択肢）型データ

群	個体	項目 1		項目 2		
		選択肢 1	選択肢 2	選択肢 1	選択肢 2	選択肢 3
1	1	1	0	1	0	0
	2	1	0	1	0	0
	3	1	0	0	1	0
2	1	1	0	0	1	0
	2	0	1	0	1	0
3	1	0	1	0	1	0
	2	0	1	0	0	1
	3	0	1	0	0	1

表 1 の形式のデータを表す変数 $n_i(\mathbf{n}, j, k)$ を次のように定義する。

個体 i がグループ \mathbf{n} に属しており、かつアイテム j のカテゴリ k が選ばれているとき

$$n_i(\mathbf{n}, j, k) = 1$$

それ以外るとき

$$n_i(\mathbf{n}, j, k) = 0$$

このとき、アイテム j のカテゴリ k に数値 $x(j, k)$ が与えられると、グループ \mathbf{n} の i 番目の個体の得点 $y_i(\mathbf{n})$ を次式で与えることができる

$$y_i(\mathbf{n}) = \sum_j \sum_k x(j, k) \cdot n_i(\mathbf{n}, j, k) \quad (1)$$

$y_i(\mathbf{n})$ の値のグループ内での変動を小さく、グループ間の変動が大きくなるように $x(j, k)$

の値を定める。すなわち、相関比の 2 乗

$$h^2 = \frac{V_B}{V_T}$$

が最大になるように $x(j, k)$ の値を定める。ここで、 V_B は $y_i(\mathbf{n})$ の群間分散を、 V_T は全体の分散を表す。

いま、 $n_i(\mathbf{n}, j, k)$ の項目選択肢間の分散共分散行列を A 、群間分散共分散行列を B とおくと、 h^2 を極大にする $x(j, k)$ は次式を満たす。

$$B\mathbf{x} = h^2 A\mathbf{x} \quad (2)$$

ここで、 \mathbf{x} は $x(j, k)$ を要素とするベクトルである。

h^2 は (2) 式の固有値となっているので、相関比の 2 乗を最大にするためには (2) 式の最大固有値に対応する固有ベクトルを $x(j, k)$ の値とする。複数個の解を求めるときは、(2) 式の固有値のうち大きさの順に必要な個数を取り、対応する固有ベクトルを用いる。

(2) 式より、固有値 (相関係数の 2 乗) に対応する項目選択肢 (アイテム・カテゴリ) の値 $x(j, k)$ が決まると、(1) 式より個体の得点 $y_i(\mathbf{n})$ も定まる。

プログラム PQ2.dpr は、 $x(j, k)$ 、 $y_i(\mathbf{n})$ 、および $y_i(\mathbf{n})$ のグループごとの平均値 $\overline{y(\mathbf{n})}$ を求めるものである。このプログラムを実行すると、図 1 の入力データファイルの名前を設定するダイアログボックスが表示される。

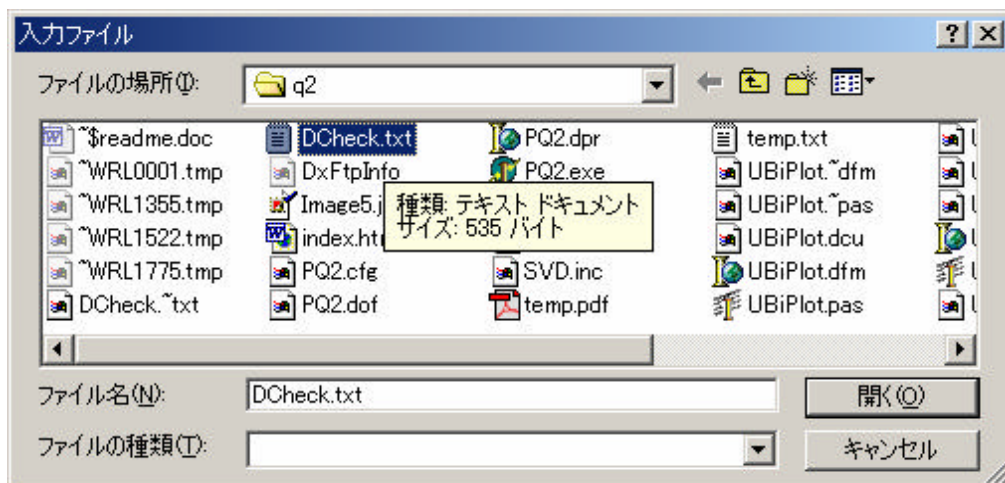


図 1 起動時に表示されるダイアログボックス

入力データファイルはテキストファイルであり、表 1 のデータの場合はリスト 1 のように用意する。

リスト1 入力データファイルの例

3	グループ数		
A			
B			
C			
2	アイテム数		
2	3	各アイテムのカテゴリー数	
-1	グループの区切り		
	1 0	1 0 0	
	1 0	1 0 0	
	1 0	0 1 0	
-2	グループの区切り（1 番目の区切り数より小さい値）		
	1 0	0 1 0	
	0 1	0 1 0	
-3	グループの区切り（1 番目の区切り数より小さい値）		
	0 1	0 1 0	
	0 1	0 0 1	
	0 1	0 0 1	
-9	データの終わり（1 番目の区切り数より小さい値）		

岩坪（1987）表 3.1

岩坪 (1987) 表 3.1

リスト1の入力データファイルは、次の形式で作成されている。

まず、第1行目にグループの数3を書く。第1行目に書かれた数値の右側は読み込まれないのでコメントなどを書くことができる。入力データファイルの各行は、必要なものが書かれた後の右側は読み込まれないので、自由にコメントを書くことができる。

第2行目以降に、グループのラベルとして英字1字を1行に1つずつ書く。この文字は各行の先頭左詰に書く。これらのラベルを書く行以外は、左側に適当に空白を置くことができるが、ラベルを書く行は空白を置くとその空白文字がラベルとして読み込まれる。

グループのラベル文字をすべて書いた後の行にアイテム(項目)の数を書く。続く行に、各アイテムに含まれるカテゴリ(選択肢)の数を順番に書く。リスト1では、1番目のアイテムのカテゴリ数2と2番目のアイテムのカテゴリ数3が書かれている。

カテゴリ数を書いた行の次の行に、負の整数を書く。この負数より小さい値がグループごとのデータの区切りとして使われる。

第1グループのデータを、この負数を書いた行の次行から1行に1人分ずつ書いていく。選択されたアイテムのカテゴリは1、選択されなかったカテゴリは0で選択・非選択を表す。第1グループのデータをすべて書き終えたら、グループの区切りとして書いた負数より小さい整数を書いて第1グループのデータの終わりの印とする。

第2グループ以降も同様に個人データとグループの区切りを示す整数値を書いていく。すべてのデータ(表1のデータの場合は第3グループまで)を書いた後の行は読み込まれないので、自由にコメントなどを書くことができる。

図1のダイアログボックスにおいて入力データファイル名を設定した後、「開く」ボタン

をクリックすると、計算結果を書き出す出力ファイルの名前の設定を求めるダイアログボックスが表示される（図2）。

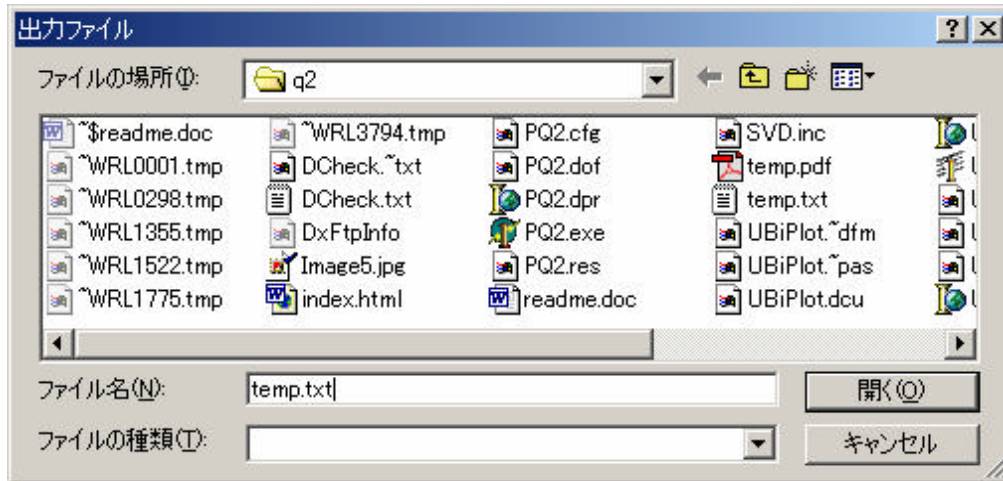


図2 出力用ファイル名の設定

計算結果は、設定した名前のファイルにテキストファイルとして書き出されるので、プログラムの実行終了後エディタなどで開いて見ることができる。

出力ファイル名の設定後、ダイアログボックスの「開く」ボタンをクリックすると、図3のようなフォームが表示される。

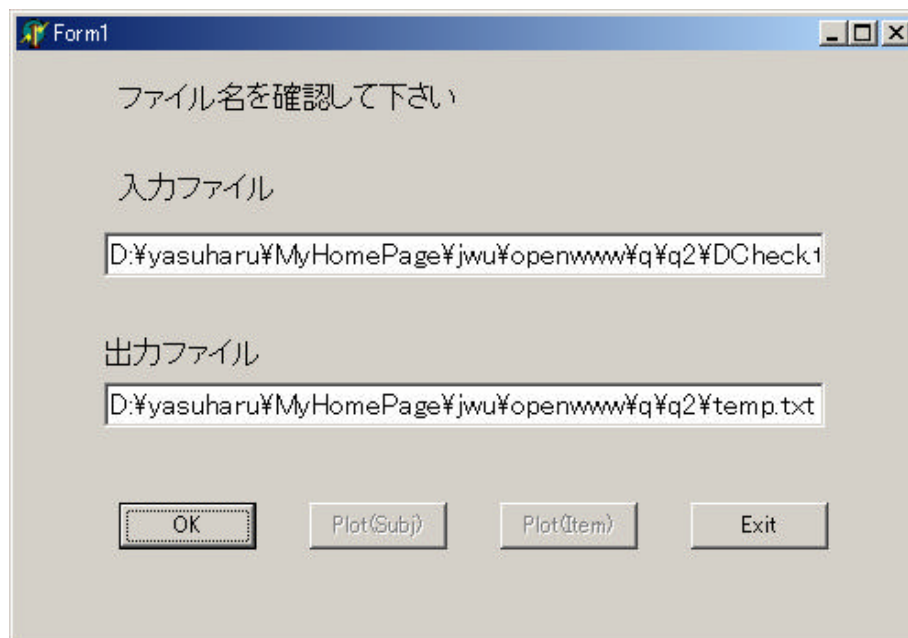


図3 ファイル名設定後に表示されるフォーム

フォーム上には、設定した入力データファイルと出力ファイルの名前が表示されている。設定したファイル名は、このフォーム上で変更することができる。「OK」ボタンをクリックすると、表示されているファイル名のファイルを入力ファイルおよび出力ファイルとして計算が始る。計算が終了すると、個人の値 $y_i(n)$ とグループの平均値 $\overline{y(n)}$ を図示するステップに進むためのボタン「Plot(Subj)」、およびアイテム・カテゴリ（項目・選択肢）の値 $x(j,k)$ とグループの平均値 $\overline{y(n)}$ を図示するステップに進むためのボタン「Plot(Item)」がイネーブル（クリック可能）になる。

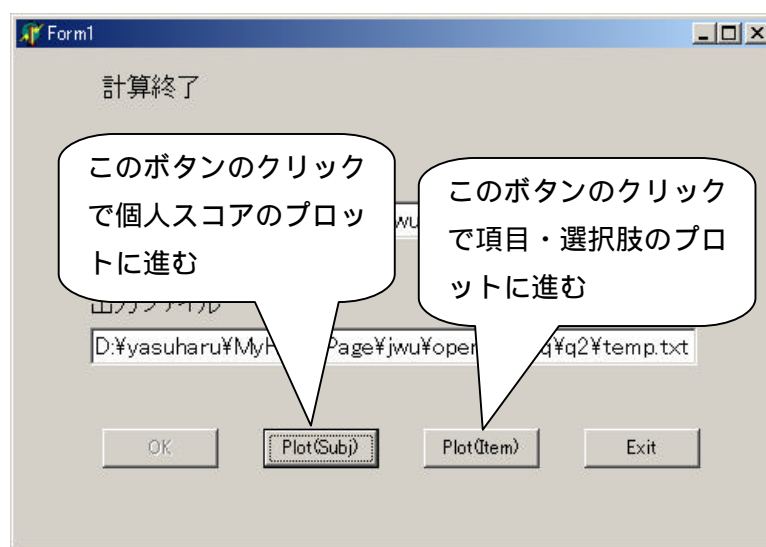


図 4 計算終了時のフォーム

「Plot(Subj)」ボタンをクリックすると図 5 のフォームが表示される。

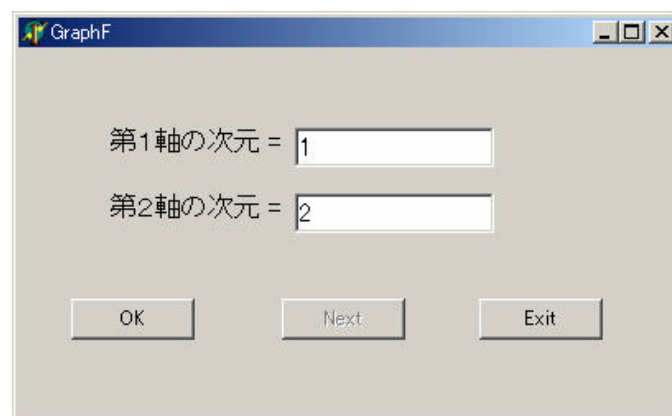


図 5 描画次元の選択

個人の値 $y_i(n)$ とグループの平均値 $\overline{y(n)}$ のプロットにおいて用いる解（相関比の2乗の大きさの順に対応する解を1軸（第1次元）、2軸（第2次元）・・・と数える）の軸（次元）を選ぶ。軸の選択後、「OK」ボタンをクリックすると、選んだ軸の解がプロットされる（図6）

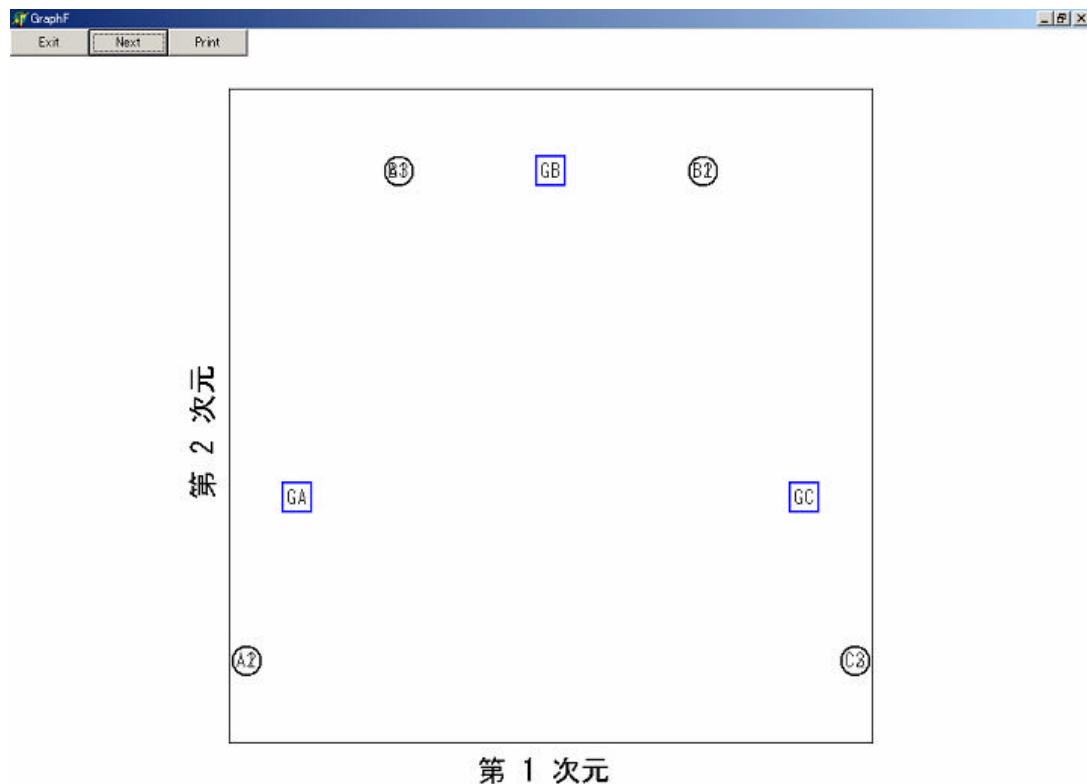


図6 個人の値とグループ平均値の描画

個人の値は小円によって表されている。小円内の先頭の文字は、グループのラベルとして入力されたものが表示されている。2番目の数値はグループ内の個人の番号である。入力データファイルに書かれている順番を表す。グループ内の個人の数が9人を超えるときは、10人目以降は?で表される。図6の表示では、グループAに属する個人は左側、グループBに属する個人は上方、グループCに属する個人は右側に位置している。グループの平均値は、小正方形で表されている。小正方形内の2文字目がそのグループを表すラベルである。それぞれのグループ内の個人を表す位置の平均位置にグループを表す小正方形が位置していることがわかる。ディスプレイに描画されている図は、図6の画面の左上のボタン「Print」をクリックするとプリンタに出力される。「Next」ボタンをクリックすると図5の描画次元の選択に戻る。描画したい次元を設定した後、「OK」ボタンをクリックすると設定した次元での描画が行われる。「Exit」ボタンのクリックで図4の計算終了時のフォ

ームに戻る。

アイテム・カテゴリの位置を表示するときは、図4のフォームにおいて「Plot(Item)」をクリックする。「Plot(Item)」のクリックで図7のフォームが表示される。

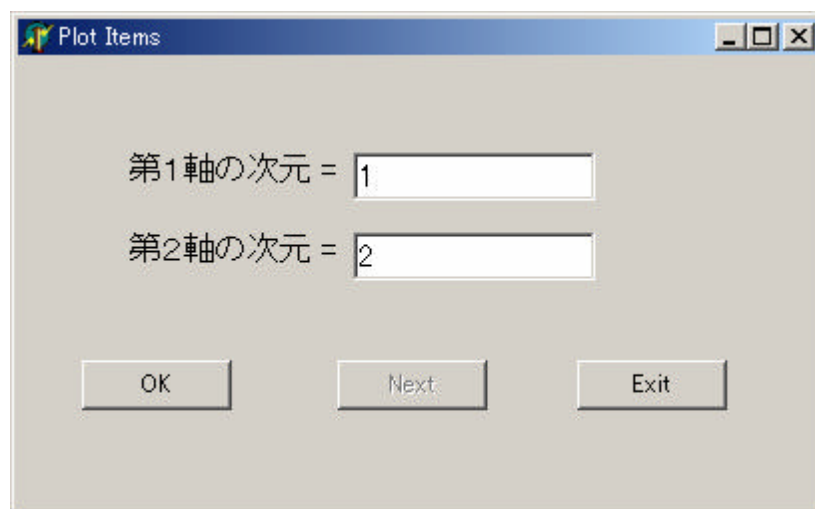


図7 描画次元の選択

図5の場合と同様に描画する軸（次元）を選び、「OK」ボタンをクリックすると選択した軸のアイテム・カテゴリ値 $x(j,k)$ とグループの平均値 $\overline{y(n)}$ がプロットされる（図8）。

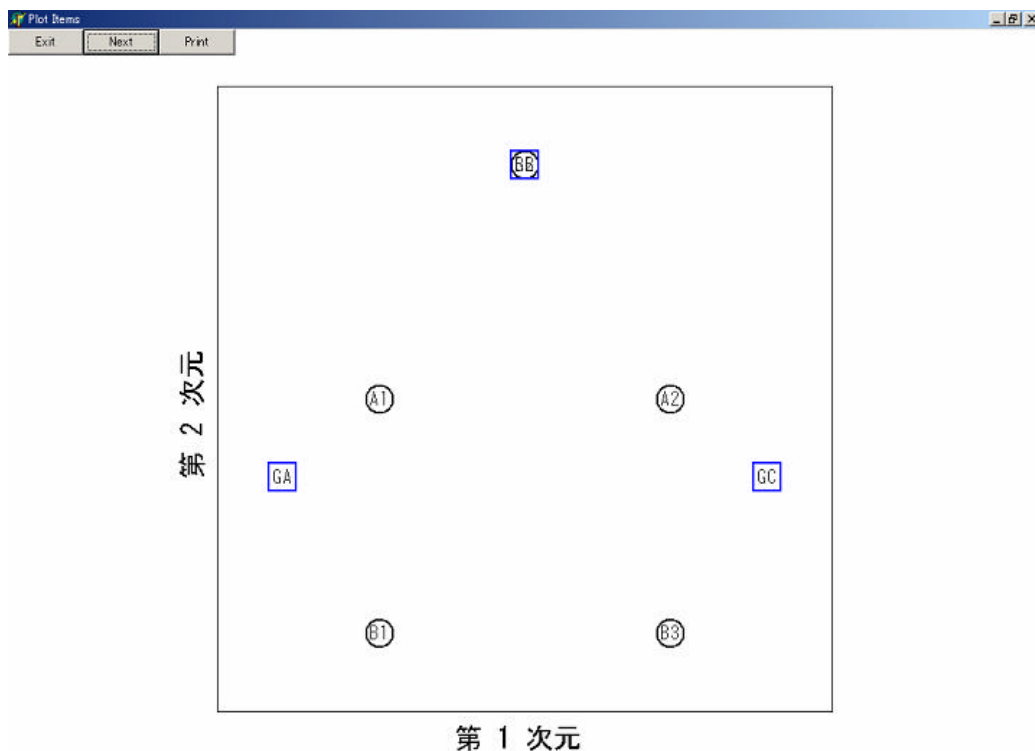


図 8 アイテム・カテゴリ値とグループ平均値のプロット

グループ平均値は図 6 の個人の値のプロットのとおり描かれている。アイテム・カテゴリの値は小円によって描かれ、小円内の最初の文字がそのアイテムを表す。アイテムは、入力ファイルにおける位置において左から A , B , . . . と表されている。アイテムを表す文字の右側の数値はカテゴリを表す。カテゴリはアイテム内における左からの位置を 1、2、. . . と数値で表している。図 8 から、アイテム A (1 番目のアイテム) のカテゴリ 1 およびアイテム B のカテゴリ 1 を選ぶ者はグループ A に、アイテム A のカテゴリ 2 およびアイテム B のカテゴリ 3 を選ぶ者はグループ C に属する傾向があることがわかる。アイテム B のカテゴリ 2 を選ぶ者はグループ B に属する傾向にある。これらの傾向は表 1 のデータにおいて確認できる。

「Exit」, 「Print」, 「Next」ボタンの使い方は、図 6 の場合と同じである。

図 4 のフォームにおいて「Exit」ボタンをクリックすると、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 で設定した名前の計算結果の出力ファイルを開いて見ることができる。表 1 のデータの場合、リスト 2 のようになっている。

リスト2 計算結果の出力例

入力データファイル...D:\yasuharu\MyHomePage\jwu\openwww\q2\DCheck.txt

グループ: A

1	1	0	1	0	0
2	1	0	1	0	0
3	1	0	0	1	0

グループ: B

1	1	0	0	1	0
2	0	1	0	1	0

グループ: C

1	0	1	0	1	0
2	0	1	0	0	1
3	0	1	0	0	1

個体総数 = 8

A =

0.250	-0.250	0.125	0.000	-0.125
-0.250	0.250	-0.125	0.000	0.125
0.125	-0.125	0.188	-0.125	-0.063
0.000	0.000	-0.125	0.250	-0.125
-0.125	0.125	-0.063	-0.125	0.188

B =

0.188	-0.188	0.125	0.000	-0.125
-0.188	0.188	-0.125	0.000	0.125
0.125	-0.125	0.104	-0.042	-0.063
0.000	0.000	-0.042	0.083	-0.042
-0.125	0.125	-0.063	-0.042	0.104

固有値 (相関比の 2 乗) =

0.8333333333333333
0.3333333333333333

基準化された最適スコア

A1 ->	-0.63246	0.00000
A2 ->	0.63246	-0.00000
B1 ->	-0.63246	-1.00000
B2 ->	0.00000	1.00000
B3 ->	0.63246	-1.00000

個体得点

グループ A

1	-1.265	-1.000
2	-1.265	-1.000
3	-0.632	1.000

グループ B

1	-0.632	1.000
2	0.632	1.000

グループ C

1	0.632	1.000
2	1.265	-1.000
3	1.265	-1.000

グループ平均得点

グループ A

-1.054	-0.333
--------	--------

グループ B

0.000	1.000
-------	-------

グループ C

1.054	-0.333
-------	--------

リスト2には、まず、入力データファイル名の書き出しの後、読み込んだデータが書き出されている。その後、個体の総数の出力の後、分散共分散行列 A 、群間分散共分散行列 B が書き出され、相関比の2乗が書き出されている。続いて、図8で表示されているアイテム・カテゴリの値が最適スコアとして、図6で表示されている個人の値が個体得点として書き出されている。最後にグループごとの平均得点がグループ平均得点として書き出されている。

参考文献

岩坪秀一「数量化法の基礎」朝倉書店、1987