

多変量解析 ~ 数量化3類 ~

発表日 6月2日

発表者 田島 勇樹



10.1 数量化3類とは

- 各サンプルがそれぞれの変数に反応しているかどうかを示すデータに基づいて、主成分分析と同じ目的で適用する手法

10.1 適用例と解析の目的

- 表10.1は児童10人の得意科目のデータである。

表10.1児童の得意科目のデータ(印が得意科目(質的変数))

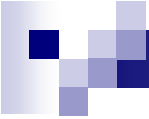
児童No.	国語	社会	算数	理科	音楽	図工	体育
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

- ・低い次元でデータのばらつきを解釈できないか
- ・どのような数量を与えたらよいか
- ・数量化によって説明力はどれくらいあるか
- ・科目や児童をどのように分類できるか

などを検討する。

10.2 解析の流れ(1)

- (1) 第 i 変数に数量 x_i 、第 j サンプルに数量 j_i をわりあて、反応がある(印のある)ところに数量 $[x_i, j_i]$ を与えて相関係数を考える。
- (2.1) 相関係数が最大になるような数量 x_i と j_i を求める過程で得られる行列の固有値を求める。
- (2.2) 最大固有値は常に1なので除外する。第2固有値とそれに対応する固有ベクトルを求めて成分1の変数スコア $x^{(1)}$ とサンプルスコア $j^{(1)}$ を求める。
- (2.3) 第3固有値とそれに対応する固有ベクトルを求めて成分2の変数スコア $x^{(2)}$ とサンプルスコア $j^{(2)}$ を求める。以下同様に成分3、成分4などを求める。



10.2 解析の流れ(2)

- (3.1) それぞれの成分の寄与率および累積寄与率を求める。
数量化3類では固有値は1以下である。したがって、主成分分析のときに用いた「累積寄与率が80%以上」という目安だけが意味を持つ。
- (3.2) 数量化3類では成分1や成分2の寄与率は主成分分析の場合のように大きくなることが多い。そのような場合、主に成分1と成分2に重点をおいて考察する。
- (4) 変数スコアとサンプルスコアの散布図(横軸が成分1、縦軸が成分2)を描いて、変数とサンプルの特徴付けと分類を行う。

10.2 基本的な考え方

- 表10.1に基づいて説明する。

同じ得意科目など反応が似ているグループに分類する。

表10.2は

- ・反応の似た児童(サンプル)が近く
- ・反応のされ方が似た科目(変数)が近くに配置されたもの

これよりサンプルや変数のグルーピングが可能

表10.2 表10.1の並び替えたもの

児童No.	音楽	図工	算数	理科	国語	社会	体育
2							
6							
7							
1							
9							
4							
10							
8							
3							
5							

10.2成分の導出(1)

- 各変数、サンプルにどのような数量を与えるか？
 - ・変数 $[a_1, a_2, a_3]$ に数量 $[x_1, x_2, x_3]$ を与える
 - ・変数 $[1, 2, 3, 4]$ に数量 $[y_1, y_2, y_3, y_4]$ を与える

「行と列を並び替え、 印を対角線に集める」

「 x と y の相関係数を最大にする x と y を与える」

表10.3

サンプルNo.	変数		
	a1	a2	a3
1			
2			
3			
4			



サンプルNo		変数		
		a1	a2	a3
		x1	x2	x3
1	y1		[x2,y1]	[x3,y1]
2	y2		[x2,y2]	
3	y3			[x3,y3]
4	y4	[x1,y4]		[x3,y4]

10.2成分の導出(2)

- 相関係数の値はxとyの平行移動に関しては不変なので下の式のように仮定できる

$$\bar{x} = \frac{x_1 + 2x_2 + 3x_3}{6} = 0 \quad , \quad \bar{y} = \frac{2y_1 + y_2 + y_3 + 2y_4}{6} = 0 \quad (10.1)$$

相関係数 $r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (10.2)$

(10.1)の制約条件($\sum x = \sum y = 0$)より

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = x_1y_4 + x_2y_1 + x_2y_2 + x_3y_1 + x_3y_3 + x_3y_4 \quad (10.3)$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = x_1^2 + 2x_2^2 + 3x_3^2 \quad (10.4)$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 2y_1^2 + y_2^2 + y_3^2 + y_4^2 \quad (10.5)$$

10.2成分の導出(3)

- $[x,y]$ から $[u,w]$ への変換

$$u_1 = \sqrt{1}x_1, \quad u_2 = \sqrt{2}x_2, \quad u_3 = \sqrt{3}x_3 \quad (10.6)$$

$$w_1 = \sqrt{2}y_1, \quad w_2 = \sqrt{1}y_2, \quad w_3 = \sqrt{1}y_3, \quad w_4 = \sqrt{2}y_4 \quad (10.7)$$

根号の中は変数の数を表す。

S_{xy} と S_{xx} と S_{yy} を u と w で置き換える。

$$S_{xy} = \frac{u_1}{\sqrt{1}} \cdot \frac{w_4}{\sqrt{2}} + \frac{u_2}{\sqrt{2}} \cdot \frac{w_1}{\sqrt{2}} + \frac{u_2}{\sqrt{2}} \cdot \frac{w_2}{\sqrt{1}} + \frac{u_3}{\sqrt{3}} \cdot \frac{w_1}{\sqrt{2}} + \frac{u_3}{\sqrt{3}} \cdot \frac{w_3}{\sqrt{1}} + \frac{u_3}{\sqrt{3}} \cdot \frac{w_4}{\sqrt{2}} \quad (10.8)$$

$$S_{xx} = u_1^2 + u_2^2 + u_3^2 \quad (10.9)$$

$$S_{yy} = w_1^2 + w_2^2 + w_3^2 + w_4^2 \quad (10.10)$$

10.2 S_{xy} の最大化(1)

相関係数の最大化にあたって

$$S_{xx} = u_1^2 + u_2^2 + u_3^2 = 1 \quad (10.11)$$

$$S_{yy} = w_1^2 + w_2^2 + w_3^2 + w_4^2 = 1 \quad (10.12)$$

上の制約条件の下で S_{xy} の最大化を考える

制約なしでは S_{xy} はいくらでも大きくなる 制約が必要

制約があり最大化する

ラグランジュの未定乗数法

未定乗数: λ, η

$$f(u_1, u_2, u_3, w_1, w_2, w_3, w_4, \lambda, \eta) = S_{xy} - \frac{\lambda}{2}(S_{xx} - 1) - \frac{\eta}{2}(S_{yy} - 1) \quad (10.13)$$

λ と η を2で割る 微分した後の結果を見やすくするため

10.2 S_{xy} の最大化(2)

- (10.13)の式を $u_1, u_2, u_3, w_1, w_2, w_3, w_4$ で偏微分して0とおく。

$$\frac{\partial f}{\partial u_1} = \frac{1}{\sqrt{1}} \cdot \frac{w_4}{\sqrt{2}} - \lambda u_1 = 0 \quad (10.14)$$

$$\frac{\partial f}{\partial u_2} = \frac{1}{\sqrt{2}} \cdot \frac{w_1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{w_2}{\sqrt{1}} - \lambda u_2 = 0 \quad (10.15)$$

$$\frac{\partial f}{\partial u_3} = \frac{1}{\sqrt{3}} \cdot \frac{w_1}{\sqrt{2}} + \frac{1}{\sqrt{3}} \cdot \frac{w_3}{\sqrt{1}} + \frac{1}{\sqrt{3}} \cdot \frac{w_4}{\sqrt{2}} - \lambda u_3 = 0 \quad (10.16)$$

$$\frac{\partial f}{\partial w_1} = \frac{1}{\sqrt{2}} \cdot \frac{u_2}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{u_3}{\sqrt{3}} - \eta w_1 = 0 \quad (10.17)$$

$$\frac{\partial f}{\partial w_2} = \frac{1}{\sqrt{1}} \cdot \frac{u_2}{\sqrt{2}} - \eta w_2 = 0 \quad (10.18)$$

$$\frac{\partial f}{\partial w_3} = \frac{1}{\sqrt{1}} \cdot \frac{u_3}{\sqrt{3}} - \eta w_3 = 0 \quad (10.19)$$

$$\frac{\partial f}{\partial w_4} = \frac{1}{\sqrt{2}} \cdot \frac{u_1}{\sqrt{1}} + \frac{1}{\sqrt{2}} \cdot \frac{u_3}{\sqrt{3}} - \eta w_4 = 0 \quad (10.20)$$

10.2 S_{xy} の最大化(3)

(10.14)式 $\times u_1$ + (10.15)式 $\times u_2$ + (10.16)式 $\times u_3$ より

$$S_{xy} = \lambda \quad (10.21)$$

(10.17)式 $\times w_1$ + (10.18)式 $\times w_2$ + (10.19)式 $\times w_3$ + (10.20)式 $\times w_4$ より

$$S_{xy} = \eta \quad (10.22)$$

(10.21)式と(10.22)式より

$$S_{xy} = \lambda = \eta \quad (10.23)$$

(10.17)式 ~ (10.20)式より η に λ を代入

$$w_1 = \frac{1}{\lambda} \left(\frac{1}{\sqrt{2}} \cdot \frac{u_2}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{u_3}{\sqrt{3}} \right) \quad (10.24), \quad w_2 = \frac{1}{\lambda} \left(\frac{1}{\sqrt{1}} \cdot \frac{u_2}{\sqrt{2}} \right) \quad (10.25)$$

$$w_3 = \frac{1}{\lambda} \left(\frac{1}{\sqrt{1}} \cdot \frac{u_3}{\sqrt{3}} \right) \quad (10.26), \quad w_4 = \frac{1}{\lambda} \left(\frac{1}{\sqrt{2}} \cdot \frac{u_1}{\sqrt{1}} + \frac{1}{\sqrt{2}} \cdot \frac{u_3}{\sqrt{3}} \right) \quad (10.27)$$

10.2 行列とベクトル表現(1)

(10.24)式 ~ (10.27)式を(10.14)式 ~ (10.16)式に代入

$$\begin{aligned}\frac{1}{2}u_1 + \frac{\sqrt{3}}{6}u_3 &= \lambda^2 u_1 \\ \frac{3}{4}u_2 + \frac{\sqrt{6}}{12}u_3 &= \lambda^2 u_2 \\ \frac{\sqrt{3}}{6}u_1 + \frac{\sqrt{6}}{12}u_2 + \frac{2}{3}u_3 &= \lambda^2 u_3\end{aligned}\tag{10.28}$$

これを行列とベクトルで表すと

$$\begin{bmatrix} 1/2 & 0 & \sqrt{3}/6 \\ 0 & 3/4 & \sqrt{6}/12 \\ \sqrt{3}/6 & \sqrt{6}/12 & 2/3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \lambda^2 \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\tag{10.29}$$

10.2 行列とベクトル表現(2)

目的は $S_{xy} = \lambda$ の最大化

(10.29)式の最大固有値とそれに対応する(長さ1の)固有ベクトルを求める

(10.29)式の固有値と対応する(長さ1の)固有ベクトル

$$\begin{aligned} \text{第1固有値: } 1 & \quad [u_1, u_2, u_3] = \left[\frac{\sqrt{6}}{6}, \frac{2\sqrt{3}}{6}, \frac{3\sqrt{2}}{6} \right] \\ \text{第2固有値: } 2/3 & \quad [u_1, u_2, u_3] = \left[-\frac{\sqrt{30}}{10}, \frac{2\sqrt{15}}{10}, -\frac{\sqrt{10}}{10} \right] \\ \text{第3固有値: } 1/4 & \quad [u_1, u_2, u_3] = \left[\frac{2\sqrt{30}}{15}, \frac{\sqrt{15}}{15}, -\frac{3\sqrt{10}}{15} \right] \end{aligned} \quad (10.30)$$

$$\lambda = S_{xy} = S_{xy} / \sqrt{S_{xx} S_{yy}} = r_{xy} \text{ (制約条件より } S_{xx} = S_{yy} = 1 \text{) より } \lambda^2 \leq 1$$

10.2 成分の求め方(1)

・成分1を求める

まず第1固有値であり最大固有値である1の固有ベクトルを(10.6)式に代入

$$[x_1, x_2, x_3] = \left[\frac{u_1}{\sqrt{1}}, \frac{u_2}{\sqrt{2}}, \frac{u_3}{\sqrt{3}} \right] = \left[\frac{6}{\sqrt{6}}, \frac{6}{\sqrt{6}}, \frac{6}{\sqrt{6}} \right] \quad (10.31)$$

となり(10.1)式を満たさない。

よって S_{xy} の最大値は第2固有値 $\lambda^2 = 2/3$ に対応して $\lambda = \sqrt{2/3}$ となる
同様に第2固有値に対応する $[x_1, x_2, x_3]$ を求めると

$$[x_1, x_2, x_3] = \left[\frac{u_1}{\sqrt{1}}, \frac{u_2}{\sqrt{2}}, \frac{u_3}{\sqrt{3}} \right] = \left[-\frac{3\sqrt{30}}{30}, \frac{3\sqrt{30}}{30}, -\frac{\sqrt{30}}{30} \right] \quad (10.32)$$

この数量は(10.1)の制約条件を満たす。

10.2 成分の求め方(2)

(10.32)式の値を(10.24)~(10.27)式を用いて $[w_1, w_2, w_3, w_4]$ を求める。

$$[w_1, w_2, w_3, w_4] = \left[\frac{\sqrt{10}}{10}, \frac{3\sqrt{5}}{10}, -\frac{\sqrt{5}}{10}, -\frac{2\sqrt{10}}{10} \right] \quad (10.33)$$

(10.7)式を用いて、 $[y_1, y_2, y_3, y_4]$ を求める。

$$[y_1, y_2, y_3, y_4] = \left[\frac{w_1}{\sqrt{2}}, \frac{w_2}{\sqrt{1}}, \frac{w_3}{\sqrt{1}}, \frac{w_4}{\sqrt{2}} \right] = \left[\frac{\sqrt{5}}{10}, \frac{3\sqrt{5}}{10}, -\frac{\sqrt{5}}{10}, -\frac{2\sqrt{5}}{10} \right] \quad (10.34)$$

制約条件(10.1)式を満たす。

成分1だけで変数やサンプルの分類が十分に行えない

成分2や成分3、...と求める。

ここでは表10.3のデータを用いたので成分2までしか求められない。

一般に

(p と n の小さいほう-1)個の成分が求まる。

p : 変数の個数

n : サンプルの個数

10.2成分の求め方(3)

- 成分1は(10.32)、(10.34)で求めた。成分2と区別するために添え字(1)をつける。

$$\left[x_1^{[1]}, x_2^{[1]}, x_3^{[1]} \right] \quad \left[y_1^{[1]}, y_2^{[1]}, y_3^{[1]}, y_4^{[1]} \right] \quad \left[u_1^{[1]}, u_2^{[1]}, u_3^{[1]} \right] \quad \left[w_1^{[1]}, w_2^{[1]}, w_3^{[1]}, w_4^{[1]} \right]$$

同様に成分2も添え字(2)をつけて表現する。

$$\left[x_1^{[2]}, x_2^{[2]}, x_3^{[2]} \right] \quad \left[y_1^{[2]}, y_2^{[2]}, y_3^{[2]}, y_4^{[2]} \right] \quad \left[u_1^{[2]}, u_2^{[2]}, u_3^{[2]} \right] \quad \left[w_1^{[2]}, w_2^{[2]}, w_3^{[2]}, w_4^{[2]} \right]$$

新たに $u^{(1)}$ と $u^{(2)}$ は無相関、 $w^{(1)}$ と $w^{(2)}$ は無相関という制約条件を課す。

$$\sum u_i^{(1)} u_i^{(2)} = \sum w_i^{(1)} w_i^{(2)} = 0 \quad (10.35)$$

第2固有ベクトルと第3固有ベクトルは直行しており(10.35)式に代入すると成り立つ。

よって第3個有値に対応する $[x_1^{[2]}, x_2^{[2]}, x_3^{[2]}]$ を求めると

$$\left[x_1^{[2]}, x_2^{[2]}, x_3^{[2]} \right] = \left[\frac{u_1^{[2]}}{\sqrt{1}}, \frac{u_2^{[2]}}{\sqrt{2}}, \frac{u_3^{[2]}}{\sqrt{3}} \right] = \left[\frac{4\sqrt{30}}{30}, \frac{\sqrt{30}}{30}, -\frac{2\sqrt{30}}{30} \right] \quad (10.36)$$

10.2成分の求め方(4)

(10.36)式は制約条件(10.1)式を満たす。

(10.24) ~ (10.27)式を用いて $[w_1^{[2]}, w_2^{[2]}, w_3^{[2]}, w_4^{[2]}]$ 、 $[y_1^{[2]}, y_2^{[2]}, y_3^{[2]}, y_4^{[2]}]$ を求める。

$$[w_1^{[2]}, w_2^{[2]}, w_3^{[2]}, w_4^{[2]}] = \left[-\frac{\sqrt{15}}{15}, \frac{\sqrt{30}}{15}, -\frac{2\sqrt{30}}{15}, \frac{2\sqrt{15}}{15} \right] \quad (10.37)$$

$$[y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}] = \left[\frac{w_1^{(2)}}{\sqrt{2}}, \frac{w_2^{(2)}}{\sqrt{1}}, \frac{w_3^{(2)}}{\sqrt{1}}, \frac{w_4^{(2)}}{\sqrt{2}} \right] = \left[-\frac{\sqrt{30}}{30}, \frac{\sqrt{30}}{15}, -\frac{2\sqrt{30}}{15}, \frac{\sqrt{30}}{15} \right] \quad (10.38)$$

よって制約条件(10.1)式を満たす。

10.3 寄与率および累積寄与率

成分1、成分2の寄与率

$$\text{成分1の寄与率} = \frac{\lambda_2^2}{\lambda_2^2 + \lambda_3^2 + \dots} \quad (10.39)$$

$$\text{成分2の寄与率} = \frac{\lambda_3^2}{\lambda_2^2 + \lambda_3^2 + \dots}$$

累積寄与率

$$\text{成分1までの累積寄与率} = \frac{\lambda_2^2}{\lambda_2^2 + \lambda_3^2 + \dots} \quad (10.40)$$

$$\text{成分2までの累積寄与率} = \frac{\lambda_2^2 + \lambda_3^2}{\lambda_2^2 + \lambda_3^2 + \dots}$$

表10.3のデータより

成分1の寄与率:0.727 成分2の寄与率:0.273

成分1の累積寄与率:0.727 成分2の累積寄与率:1.000

10.4 変数スコアとサンプルスコアの散布図(1)

・表10.1のデータに基づいた数量化3類の分析

表10.7 第1固有値を除く固有値、寄与率、累積寄与率

No.	固有値	寄与率	累積寄与率
1	0.561	0.437	0.437
2	0.279	0.218	0.655
3	0.197	0.153	0.808
4	0.125	0.098	0.906
5	0.091	0.071	0.977
6	0.029	0.023	1

表10.7より成分2までで累積寄与率は0.655なので成分1、成分2において考察する。

変数スコアとサンプルスコアを主成分分析の手順に準じて、それぞれの成分に対して分散が固有値に一致するように $\sqrt{(n-1)\lambda^2}$ 倍する。

10.4 変数スコアとサンプルスコアの散布図(2)

■ 変数スコアとサンプルスコアの表

表10.8 変数スコア

科目	成分1 x 1	成分2 x 2
国語	-0.581	-0.336
社会	-0.84	-0.335
算数	0.394	-0.077
理科	0.152	-0.553
音楽	1.318	0.102
図工	0.805	0.605
体育	-0.949	1

表10.9 サンプルスコア

児童	成分1 y 1	成分2 y 2
1	0.167	-0.179
2	1.12	0.397
3	-1.021	0.628
4	-0.291	-0.616
5	-1.194	0.629
6	1.012	0.097
7	0.83	-0.333
8	-0.74	-0.106
9	0.111	0.964
10	-0.291	-0.616

10.4 変数スコアとサンプルスコアの散布図(2)

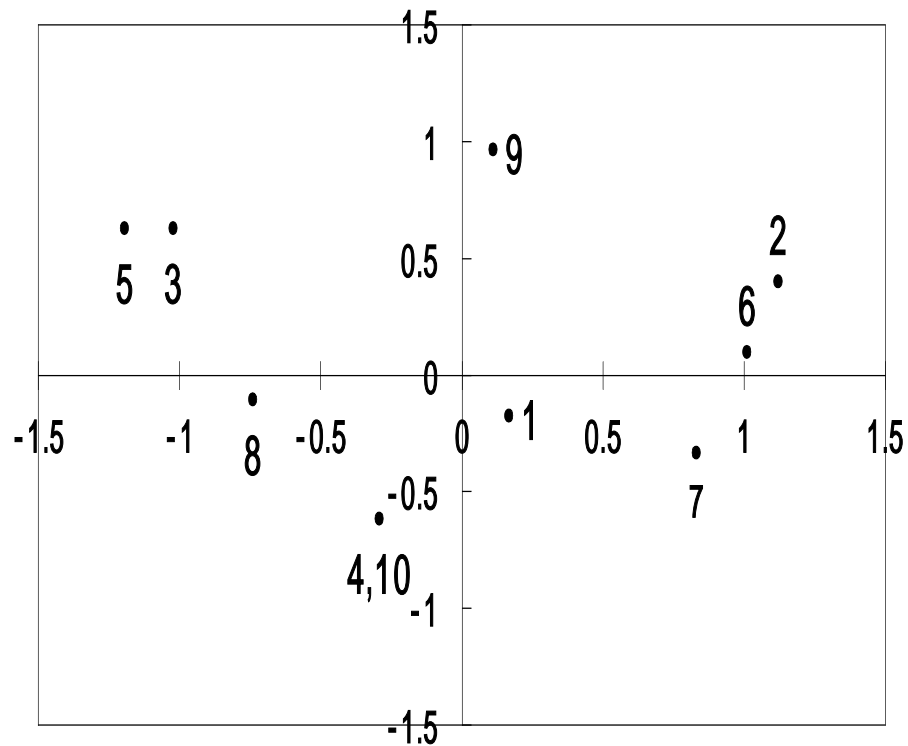


図10.1 変数スコア散布図

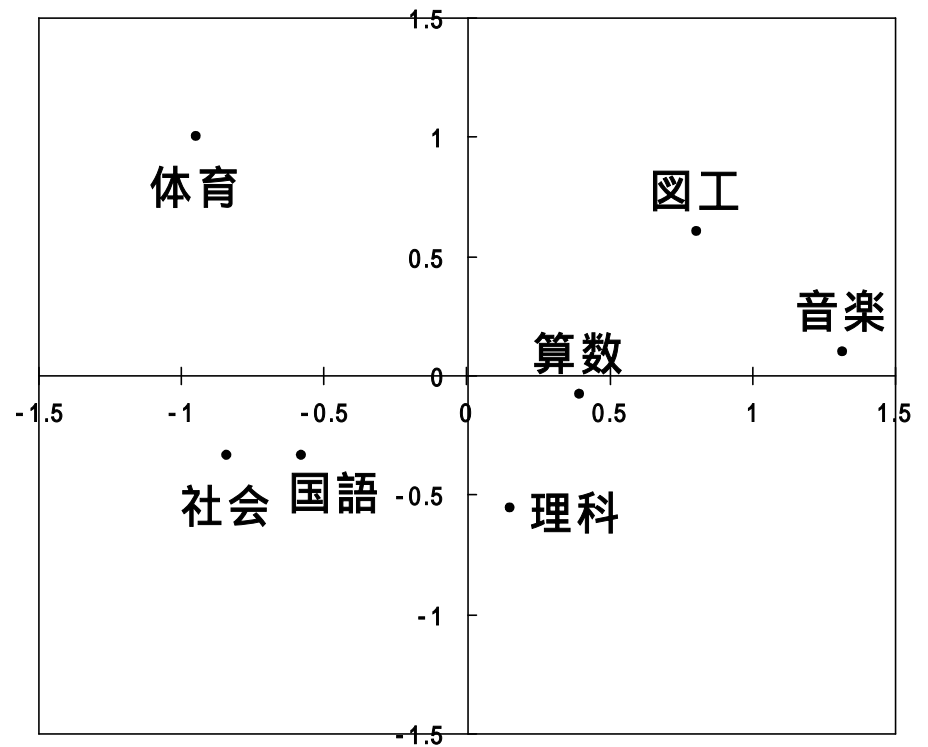


図10.2 サンプルスコア散布図