データ解析 2005 年度期末試験問題

静岡大学工学部システム工学科

安藤 和敏

2006 年 2 月 15 日

注意事項

- (関数) 電卓のみ持ち込み可. 携帯電話の電卓機能の使用も不可.
- 試験の時間は 10:20-11:40 である.
- 問題用紙は持ち帰ってよい.
- 解答及び採点の結果は, Web ページ (`http://coconut.sys.eng.shizuoka.ac.jp/data/`) で公開する.

問題 1. (配点 24)

以下の設問に答えよ.

設問 (1) X, Y, Z, W を確率変数として, a, b, c, d を定数とする. $U = aX + bY$ と $V = cZ + dW$ の共分散 s_{UV} が,

$$s_{UV} = acs_{XZ} + ads_{XW} + bcs_{YZ} + bds_{YW}$$

で与えられることを示せ. $\bar{U} = a\bar{X} + b\bar{Y}, \bar{V} = c\bar{Z} + d\bar{W}$ という事実を使ってもよい. ◇

設問 (2) 表 1 に示すような 3 変数 X, Y, Z, W に関するデータが得られているとする. 表 1 のデータの

表 1: X, Y, Z に関するデータ

No.	X	Y	Z	W
1	83	-111	-23	18
2	78	-116	27	6
3	60	-73	78	6
4	84	-130	40	3
5	79	-115	54	18
6	54	-87	43	23
7	68	-103	18	-18
8	59	-96	74	2
9	82	-114	-22	18
10	84	-130	26	12

分散共分散行列 S が

$$S = \begin{bmatrix} 137.21 & -188.61 & -247.94 & 20.36 \\ -188.61 & 328.72 & 289.83 & -9.22 \\ -247.94 & 289.83 & 1193.83 & -81.67 \\ 20.36 & -9.22 & -81.67 & 142.18 \end{bmatrix}$$

で与えられるとき, $U = 2X + 3Y$ の分散 s_U^2, 及び, U と $V = 4Z - W$ の共分散 s_{UV} を求めよ. (小数点以下 1 桁まで)◇

表 2: x, y, z, w に関するデータ

No.	x	y	z	w
1	182.0	67.5	85.1	93.7
2	176.3	63.5	83.5	93.0
3	179.0	63.0	87.0	96.3
4	172.8	88.0	101.0	94.1
5	170.0	61.5	83.0	91.8
6	164.5	58.5	83.5	88.5
7	170.6	57.0	83.0	88.1
8	170.4	79.0	98.0	92.5
9	165.0	63.3	89.5	89.7
10	172.5	66.0	91.4	91.4

問題 2. (配点 25)

表 2 に示すような 4 変数 x, y, z, w に関するデータが得られているとする. このデータに対して, 主成分分析のモデル

$$u = ax + by + cz + dw$$

を考える. 表 2 のデータの分散共分散行列 S を計算すると,

$$S = \begin{bmatrix} 31.24 & 8.28 & -2.28 & 11.16 \\ 8.28 & 92.22 & 57.72 & 12.11 \\ -2.28 & 57.72 & 42.25 & 5.53 \\ 11.16 & 12.11 & 5.53 & 6.68 \end{bmatrix}$$

であった. さらに, S の固有値は, $1.48\ (\simeq 1.22^2), 2.96\ (\simeq 1.72^2), 35.97\ (\sim 6.00^2), 131.98\ (\simeq 11.49^2)$ の 4 つであり, これらの固有値に属する固有ベクトルは, それぞれ,

$$\begin{bmatrix} -0.366 \\ 0.036 \\ -0.195 \\ 0.909 \end{bmatrix}, \begin{bmatrix} -0.130 \\ 0.553 \\ -0.786 \\ -0.243 \end{bmatrix}, \begin{bmatrix} 0.919 \\ 0.031 \\ -0.229 \\ 0.320 \end{bmatrix}, \begin{bmatrix} 0.068 \\ 0.832 \\ 0.540 \\ 0.110 \end{bmatrix}$$

であった.

以下の設問に答えよ.

設問 (1) 第 1 主成分 $u_1 = a_1x+b_1y+c_1z+d_1w$ の係数 a_1, b_1, c_1, d_1 と第 2 主成分 $u_2 = a_2x+b_2y+c_2z+d_2w$ の係数 a_2, b_2, c_2, d_2 を求めよ. (小数点以下 3 桁まで) $\diamond$

設問 (2) 第 1 主成分と第 2 主成分の寄与率, 及び, 第 2 主成分までの累積寄与率を求めよ. (小数点以下 3 桁まで) $\diamond$

設問 (3) No. 10 のデータに対する主成分得点 u_1 と u_2 を求めよ. (小数点以下 2 桁まで) $\diamond$

問題 3. (配点 51)

表 3 に示すような 5 変数 x, y, z, u, v についての $n = 10$ 個のデータが得られているとする．このデータ

表 3: 5 変数 x, y, z, u, v に関するデータ

No	x	y	z	u	v
1	x_1	y_1	z_1	u_1	v_1
2	x_2	y_2	z_2	u_2	v_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	x_i	y_i	z_i	u_i	v_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	z_n	u_n	v_n

に対して, 2 因子の因子分析モデルは,

$$x = a_x F + b_x G + e_x, \tag{1}$$

$$y = a_y F + b_y G + e_y, \tag{2}$$

$$z = a_z G + b_z G + e_z, \tag{3}$$

$$u = a_u G + b_u G + e_u, \tag{4}$$

$$v = a_v G + b_v G + e_v \tag{5}$$

として与えられる．ここで, いくつかの仮定が置かれる．

仮定 1 共通因子と独自因子は互いに独立である．

$$s_{Fe_x} = 0, \quad s_{Fe_y} = 0, \quad s_{Fe_z} = 0, \quad s_{Fe_u} = 0, \quad s_{Fe_v} = 0,$$
$$s_{Ge_x} = 0, \quad s_{Ge_y} = 0, \quad s_{Ge_z} = 0, \quad s_{Ge_u} = 0, \quad s_{Ge_v} = 0$$

仮定 2 共通因子同士は互いに独立である．

$$s_{FG} = 0$$

仮定 3 x, y, z, u, v は標準化されている．

$$s_x^2 = 1, \quad s_y^2 = 1, \quad s_z^2 = 1, \quad s_u^2 = 1 \quad s_v^2 = 1$$

$$s_{xy} = r_{xy}, \quad s_{xz} = r_{xz}, \quad s_{xu} = r_{xu}, \quad s_{xv} = r_{xv},$$
$$s_{yz} = r_{yz}, \quad s_{yu} = r_{yu}, \quad s_{yv} = r_{yv}, \quad s_{zu} = r_{zu},$$
$$s_{zv} = r_{zv}, \quad s_{uv} = r_{uv},$$

設問 (1) x の分散 s_x^2 を, $s_F^2, s_G^2, s_{e_x}^2, s_{FG}, s_{Fe_x}, s_{Ge_x}$ を用いて表現せよ. (ここでは, 仮定 1〜3 は用いない.) ⋄

設問 (2) x と y の共分散 s_{xy} を, $s_F^2, s_{FG}, s_{Fe_y}, s_G^2, s_{Ge_y}, s_{e_xe_y}, s_{Fe_x}, s_{Ge_x}$ を用いて表現せよ. (ここでは, 仮定 1〜3 は用いない.) ⋄

設問 (1),(2) の結果に, 仮定 1〜3 を用いると,

$$1 = a_x^2 + b_x^2 + s_{e_x}^2,\ r_{xy} = a_xa_y + b_xb_y$$

が得られる. 他の変数についても同様にして計算すると以下の方程式が得られる.

$$\begin{bmatrix} 1-s_{e_x}^2 & r_{xy} & r_{xz} & r_{xu} & r_{xv} \\ r_{xy} & 1-s_{e_y}^2 & r_{yz} & r_{yu} & r_{yv} \\ r_{xz} & r_{yz} & 1-s_{e_z}^2 & r_{zu} & r_{zv} \\ r_{xu} & r_{yu} & r_{zu} & 1-s_{e_u}^2 & r_{uv} \\ r_{xv} & r_{yv} & r_{zv} & r_{uv} & 1-s_{e_v}^2 \end{bmatrix} = \begin{bmatrix} a_x & b_x \\ a_y & b_y \\ a_z & b_z \\ a_u & b_u \\ a_v & b_v \end{bmatrix} \begin{bmatrix} a_x & a_y & a_z & a_u & a_v \\ b_x & b_y & b_z & b_u & b_v \end{bmatrix} \tag{6}$$

設問 (3) 式 (6) の左辺の行列の $1 - s_{e_x}^2$ は, x の共通性と呼ばれ, h_x^2 と表される. h_x^2 どんな量を表現していると解釈されるかを述べよ. ⋄

同様に, y の共通性 h_y^2, $\cdots$, v の共通性 h_v^2 も定義される. SMC 法を用いて, h_x^2, h_y^2, h_z^2, h_u^2, h_v^2 の推定値を求めた結果, これらの推定値

$$h_x^2 = 0.300, h_y^2 = 0.297, h_z^2 = 0.206, h_u^2 = 0.420, h_v^2 = 0.418$$

が得られた.

設問 (4) 式 (6) の左辺の行列の対角成分にこれらの推定値を代入した行列 R_F は何と呼ばれるか? ⋄

設問 (5) R_F の固有値は, $0.396\ (\simeq 0.629^2), 0.552\ (\simeq 0.743^2), 0.614\ (\simeq 0.784^2), 1.051\ (\simeq 1.025^2), 2.387\ (\simeq 1.545^2)$ の 5 つであり, これらの固有値に属する固有ベクトルは, それぞれ,

$$\begin{bmatrix} -0.144 \\ 0.163 \\ -0.024 \\ -0.695 \\ 0.685 \end{bmatrix}, \begin{bmatrix} 0.801 \\ -0.424 \\ -0.380 \\ -0.162 \\ 0.092 \end{bmatrix}, \begin{bmatrix} -0.109 \\ -0.731 \\ 0.614 \\ 0.087 \\ 0.262 \end{bmatrix}, \begin{bmatrix} -0.334 \\ -0.202 \\ -0.582 \\ 0.521 \\ 0.487 \end{bmatrix}, \begin{bmatrix} 0.464 \\ 0.466 \\ 0.373 \\ 0.459 \\ 0.466 \end{bmatrix}$$

であった.

主因子法によって, ${}^t[a_x, a_y, a_z, a_u, a_v]$ と ${}^t[b_x, b_y, b_z, b_u, b_v]$ を求めよ. ⋄

設問 (6) 反復主因子法とはどのようなものかを簡潔に説明せよ. ⋄

設問 (7) 因子 F の寄与率 C_F と因子 G の寄与率 C_G, 及び, 因子全体の寄与率 C_{all} を求めよ. (小数点以下 2 桁まで)⋄

学籍番号		氏名	

問題 1(1) の解答欄 (配点 16)

$$
\begin{aligned}
s_{UV} &= \frac{1}{n}\sum_{i=1}^{n}(U_i-\bar{U})(V_i-\bar{V}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\{a(X_i-\bar{X})+b(Y_i-\bar{Y})\}\{c(Z_i-\bar{Z})+d(W_i-\bar{W})\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\{ac(X_i-\bar{X})(Z_i-\bar{Z})+ad(X_i-\bar{X})(W_i-\bar{W}) \\
&\quad +bc(Y_i-\bar{Y})(Z_i-\bar{Z})+bd(Y_i-\bar{Y})(W_i-\bar{W})\} \\
&= ac\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Z_i-\bar{Z})+ad\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(W_i-\bar{W}) \\
&\quad +bc\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})(Z_i-\bar{Z})+bd\frac{1}{n}\sum_{i=1}(Y_i-\bar{Y})(W_i-\bar{W})\} \\
&= acs_{XZ}+ads_{XW}+bcs_{YZ}+bds_{YW}
\end{aligned}
$$

問題 1(2) の解答欄 (配点 8)

$s_U^2 =$ 1244.0　　$s_{UV} =$ 1481.4

問題 2(1) の解答欄 (配点 8)

$a_1 =$ <u>0.068</u> $b_1 =$ <u>0.832</u> $c_1 =$ <u>0.540</u> $d_1 =$ <u>0.110</u>

$a_2 =$ <u>0.919</u> $b_2 =$ <u>0.031</u> $c_2 =$ <u>-0.229</u> $d_2 =$ <u>0.320</u>

問題 2(2) の解答欄 (配点 9)

$C_1 =$ <u>0.766</u> $C_2 =$ <u>0.209</u> $C_{\text{all}} =$ <u>0.975 (0.974)</u>

問題 2(3) の解答欄 (配点 8)

$u_1 =$ <u>126.05</u> $u_2 =$ <u>168.89</u>

問題 3(1) の解答欄 (配点 8)

$s_x^2 = \underline{a_x^2 s_F^2 + b_x^2 s_G^2 + s_{e_x}^2 + 2a_x b_x s_{FG} + 2a_x s_{Fe_x} + 2b_x s_{Ge_x}}$

問題 3(2) の解答欄 (配点 8)

$s_{xy} = \underline{a_x a_y s_F^2 + a_x b_y s_{FG} + a_x s_{Fe_y} + b_x a_y s_{FG} + b_x b_y s_G^2 + b_x s_{Ge_y} + a_y s_{Fe_x} + b_y s_{Ge_x} + s_{e_x e_y}}$

問題 3(3) の解答欄 (配点 8)

<u>x の持つ情報量のうち, 共通因子 F, G で説明できる割合い</u>

学籍番号		氏名	

問題 3(4) の解答欄 (配点 4)

因子決定行列

問題 3(5) の解答欄 (配点 10)

$a_x =$ 0.717 $b_x =$ -0.342

$a_y =$ 0.720 $b_y =$ -0.207

$a_z =$ 0.576 $b_z =$ -0.597

$a_u =$ 0.710 $b_u =$ 0.534

$a_v =$ 0.720 $b_v =$ 0.499

問題 3(6) の解答欄 (配点 4)

推定した共通性と計算された共通性が等しくなるまで主因子法を繰り返す.

問題 3(7) の解答欄 (配点 9)

$C_F =$ 0.48 $C_G =$ 0.21 $C_{\mathrm{all}} =$ 0.69