中国科学技术大学2014–2015学年第二学期考试参考答案

考试科目:多元统计分析

考试时间:2015年7月1日下午14:30–16:30

一、

(1)

$Cov(X_1, X_2) = -2 \qquad \therefore$不独立

$Cov((X_1, X_2), X_3) = 0 \qquad \therefore$独立

$\because (X_1, X_2)$与X_3独立 $\qquad \therefore \dfrac{X_1 + X_2}{2}$与$X_3$独立

$Cov(X_2, X_2 - \dfrac{5}{2}X_1 - X_3) = 10 \qquad \therefore$不独立

(2)

$$\begin{pmatrix} X_1 - 2X_2 \\ X_2 + 2X_3 \end{pmatrix} = AX \ \sim \ N(A\mu, A\Sigma A^T) = N\left(\begin{pmatrix} -5 \\ 9 \end{pmatrix}, \begin{pmatrix} 29 & -12 \\ -12 & 13 \end{pmatrix} \right)$$

其中

$$A = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

$$Corr(X_1 - 2X_2, X_2 + 2X_3) = -\frac{12}{\sqrt{377}}$$

(3a)

$$\because X_3\text{与}(X_1, X_2)\text{独立}$$
$$\therefore X_3 \Big| (X_1 = x_1, X_2 = x_2) \ \sim \ N(4, 2)$$

(3b)

$$E(X_3 \Big| X_1, X_2) = 4$$
$$\therefore Corr(E(X_3 \Big| X_1, X_2), X_1) = 0$$

(4a)

正确,S是以μ为中心的椭球

(4b)

正确

$$(X-\mu)^T\Sigma^{-1}(X-\mu) \ \sim \ \chi_3^2$$
$$\therefore P((X-\mu)^T\Sigma^{-1}(X-\mu)\le \chi_{0.05}^2(3)) = 0.95$$

(4c)
错误,与$|\Sigma|^{-\frac{1}{2}}$成正比

二、
(1)

$$n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) = 1.703$$
$$\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha) = 6.626$$
$$n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) < \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

$\therefore$接受H_0
(2)
μ的95%置信域为

$$n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) < \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha) = 6.626$$

(3)
同时T^2区间

$$\bar{x_i} - \sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{S_{ii}}{n}} \le \mu_i \le \bar{x_i} + \sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{S_{ii}}{n}}$$

代入得

$$0.516 \le \mu_1 \le 0.612$$
$$0.555 \le \mu_2 \le 0.651$$

同时庞弗罗尼区间

$$\bar{x_i} - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{ii}}{n}} \le \mu_i \le \bar{x_i} + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{ii}}{n}}$$

代入得

$$0.521 \le \mu_1 \le 0.607$$
$$0.560 \le \mu_2 \le 0.646$$

三、
(1)

$$\frac{\lambda_2}{7} = 0.213$$

$$Corr(Y_1, Y_2) = 0$$

$$Y_1 = e_1^T X = 0.285X_1 + 0.211X_2 + 0.294X_3 + 0.435X_4 + 0.453X_5 + 0.453X_6 + 0.434X_7$$

$$Corr(Y_1, Z_7) = 0.434$$

(2)

$$\Sigma = LL^T + \Psi$$

求得相应于X_7的共性方差为0.687,特殊方差为0.313

$$Corr(X_5, X_6) = 0.808$$

因子得分的回归方法是指:

我们先把L和Ψ当作已知的处理假设F和ϵ为联合正态分布的,于是利用多元正态分布的性质有

$$E(F|x) = L^T\Sigma^{-1}(x-\mu)$$

我们取极大似然估计$\hat{L}$和$\hat{\Psi}$为真值,我们得到

$$\hat{f}_j = \hat{L}_T S^{-1}(x_j - \bar{x}), \qquad j = 1, 2, \ldots, n$$

四、

解:

$$\begin{pmatrix} 0 & 33 & 37 & 24 & 31 & 36 & 39 \\ & 0 & 42 & 22 & 39 & 42 & 35 \\ & & 0 & 41 & 45 & 30 & 42 \\ & & & 0 & 41 & 32 & 40 \\ & & & & 0 & 46 & 48 \\ & & & & & 0 & 34 \\ & & & & & & 0 \end{pmatrix} \begin{array}{c} 7 \\ 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{array}$$

将4,6聚为一类46

$$\begin{aligned} d_{(46)1} &= max\{d_{41}, d_{61}\} = 40 \\ d_{(46)2} &= max\{d_{42}, d_{62}\} = 42 \\ d_{(46)3} &= max\{d_{43}, d_{63}\} = 41 \\ d_{(46)5} &= max\{d_{45}, d_{65}\} = 42 \\ d_{(46)7} &= max\{d_{47}, d_{67}\} = 33 \end{aligned}$$

新的距离矩阵为:

$$
\begin{array}{cc}
\begin{pmatrix}
0 & 37 & 31 & 36 & 39 & 33 \\
 & 0 & 45 & 30 & 42 & 42 \\
 & & 0 & 46 & 48 & 41 \\
 & & & 0 & 34 & 42 \\
 & & & & 0 & 40 \\
 & & & & & 0
\end{pmatrix} &
\begin{matrix} 7 \\ 5 \\ 3 \\ 2 \\ 1 \\ 46 \end{matrix}
\end{array}
$$

将2,5聚为一类25

$$
\begin{aligned}
d_{(25)(46)} &= max\{d_{2(46)}, d_{5(46)}\} = 42 \\
d_{(25)1} &= max\{d_{21}, d_{51}\} = 42 \\
d_{(25)3} &= max\{d_{23}, d_{53}\} = 46 \\
d_{(25)7} &= max\{d_{27}, d_{57}\} = 37
\end{aligned}
$$

新的距离矩阵为:

$$
\begin{array}{cc}
\begin{pmatrix}
0 & 31 & 39 & 33 & 37 \\
 & 0 & 48 & 41 & 46 \\
 & & 0 & 40 & 42 \\
 & & & 0 & 42 \\
 & & & & 0
\end{pmatrix} &
\begin{matrix} 7 \\ 3 \\ 1 \\ 46 \\ 25 \end{matrix}
\end{array}
$$

将3,7聚为一类37

$$
\begin{aligned}
d_{(37)(25)} &= max\{d_{3(25)}, d_{7(25)}\} = 46 \\
d_{(37)46} &= max\{d_{3(46)}, d_{7(46)}\} = 41 \\
d_{(37)1} &= max\{d_{31}, d_{71}\} = 48
\end{aligned}
$$

新的距离矩阵为:

$$
\begin{array}{cc}
\begin{pmatrix}
0 & 40 & 42 & 48 \\
 & 0 & 42 & 41 \\
 & & 0 & 46 \\
 & & & 0
\end{pmatrix} &
\begin{matrix} 1 \\ 46 \\ 25 \\ 37 \end{matrix}
\end{array}
$$

将1和46聚类为146

$$
\begin{aligned}
d_{(146)(25)} &= max\{d_{1(25)}, d_{(46)(25)}\} = 42 \\
d_{(146)37} &= max\{d_{1(37)}, d_{(46)(37)}\} = 48
\end{aligned}
$$

新的距离矩阵为:

$$\begin{array}{c}\begin{pmatrix} 0 & 46 & 48 \\ & 0 & 42 \\ & & 0 \end{pmatrix} \end{array} \begin{array}{c} 37 \\ 25 \\ 146 \end{array}$$

将25和146 聚类为(25146)

$$d_{(25146)(37)} = max\{d_{(25)(37)}, d_{(146)(37)}\} = 48$$

新的距离矩阵为:

$$\begin{pmatrix} 0 & 48 \\ & 0 \end{pmatrix} \begin{array}{c} 37 \\ (25146) \end{array}$$

最后37和(25146) 聚类为(3725146)

五、

解:

记有病为1,无病为0,则

$$ECM = c(1|0)P(1|0)\pi_0 + c(0|1)P(0|1)\pi_1$$

由题意$\pi_0 = 0.9, \quad \pi_1 = 0.1, \quad \frac{c(0|1)}{c(1|0)} = 20$ ，不妨设$c(0|1) = 20, \quad c(1|0) = 1$,则

$$ECM = 0.9P(1|0) + 2P(0|1)$$

要使得ECM最小,则A、B均为阳性应判为1,A、B均为阴性应判为0,
所以共有四种情况列举如下:

A、B均为阳性	A阴，B阳	A阳，B阴	A、B均为阴性	ECM
1	1	1	0	0.242
1	1	0	0	0.270
1	0	1	0	0.208
1	0	0	0	0.236

ECM最小对应的决策为

$$\delta(x) = \begin{cases} 1, & \text{A、B均为阳性或A阴B阳} \\ 0, & \text{A、B均为阴性或A阳B阴} \end{cases}$$

六、

(2)解:

似然比$=\Lambda=\dfrac{\max\limits_{\Sigma} L(\mu_0,\Sigma)}{\max\limits_{\mu,\Sigma} L(\mu,\Sigma)}=\left(\dfrac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{\frac{n}{2}}$

$$\Lambda^{\frac{2}{n}}=\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}=\frac{\left|\sum\limits_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|}{\left|\sum\limits_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right|}$$

令$A=\begin{bmatrix}\sum\limits_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' & \sqrt{n}(\bar{x}-\mu_0)\\ \sqrt{n}(\bar{x}-\mu_0)' & -1\end{bmatrix}=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}$

$\because |A|=|A_{22}||A_{11}-A_{12}A_{22}^{-1}A_{21}|=|A_{11}||A_{22}-A_{21}A_{11}^{-1}A_{12}|$

$\therefore (-1)\left|\sum\limits_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'\right|$

$=\left|\sum\limits_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|\left|-1-n(\bar{x}-\mu_0)'(\sum\limits_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})')^{-1}(\bar{x}-\mu_0)\right|$

又因为

$$\begin{aligned}\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'&=\sum_{j=1}^{n}(x_j-\bar{x}+\bar{x}-\mu_0)(x_j-\bar{x}+\bar{x}-\mu_0)'\\&=\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'\end{aligned}$$

$\therefore |n\hat{\Sigma}_0|=|n\hat{\Sigma}|\left(1+\dfrac{T^2}{n-1}\right)$

$\therefore \Lambda^{\frac{2}{n}}=\dfrac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}=\left(1+\dfrac{T^2}{n-1}\right)^{-1}$

$\therefore$似然比检验与基于T^2统计量的检验等价