中国科学技术大学2015–2016学年第一学期考试参考答案

考试科目:多元统计分析

考试时间:2016年1月19日上午8:30–10:30

一、(24分)

(1)解:

$X_2 \sim N(1, 13)$

(2)解:

$$\because \begin{pmatrix} Y_1 \\ Y_2 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & -1 & 3 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

$$\therefore E\begin{pmatrix} Y_1 \\ Y_2 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & -1 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 10 \\ 17 \end{pmatrix}$$

$$\therefore Cov\begin{pmatrix} Y_1 \\ Y_2 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & -1 & 3 \end{pmatrix} \begin{pmatrix} 6 & 1 & -2 \\ 1 & 13 & 4 \\ -2 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & -1 & 3 \end{pmatrix}^T = \begin{pmatrix} 29 & -1 & 10 \\ -1 & 9 & 12 \\ 10 & 12 & 21 \end{pmatrix}$$

$\therefore Y \sim N(17, 21)$

(3)解:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 8 \\ 10 \end{pmatrix}, \begin{pmatrix} 29 & -1 \\ -1 & 9 \end{pmatrix}\right)$$

(4)解:

$$\because \begin{pmatrix} X_1 \\ X_3 \\ \frac{1}{2}(X_1+X_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

$$\therefore E\begin{pmatrix} X_1 \\ X_3 \\ \frac{1}{2}(X_1+X_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}$$

$$\therefore Cov\begin{pmatrix} X_1 \\ X_3 \\ \frac{1}{2}(X_1+X_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 6 & 1 & -2 \\ 1 & 13 & 4 \\ -2 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}^T = \begin{pmatrix} 6 & -2 & \frac{7}{2} \\ -2 & 4 & 1 \\ \frac{7}{2} & 1 & \frac{21}{4} \end{pmatrix}$$

$$\therefore \begin{pmatrix} X_1 \\ X_3 \\ \frac{1}{2}(X_1+X_2) \end{pmatrix} \sim N_3\left(\begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 6 & -2 & \frac{7}{2} \\ -2 & 4 & 1 \\ \frac{7}{2} & 1 & \frac{21}{4} \end{pmatrix}\right)$$

(5)解:

$\because E(X^{(2)}|X^{(1)}) = \mu^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(X^{(1)} - \mu^{(1)})$

$\because Cov(X^{(2)}|X^{(1)}) = \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

$$\therefore E(X_3|X_1 = x_1, X_2 = x_2) = 4+(-2,4)\begin{pmatrix} 6 & 1 \\ 1 & 13 \end{pmatrix}^{-1}\begin{pmatrix} x_1 - 3 \\ x_2 - 1 \end{pmatrix} = \frac{1}{77}(-30x_1+26x_2+372)$$

$$\therefore Cov(X_3|X_1 = x_1, X_2 = x_2) = 4-(-2,4)\begin{pmatrix} 6 & 1 \\ 1 & 13 \end{pmatrix}^{-1}\begin{pmatrix} -2 \\ 4 \end{pmatrix} = \frac{144}{77}$$

$$\therefore (X_3|X_1 = x_1, X_2 = x_2) \sim N(\frac{1}{77}(-30x_1+26x_2+372), \frac{144}{77})$$

(6)解:

$$\because \begin{pmatrix} Y_1 \\ Y_2 \\ Y \end{pmatrix} \sim N_3\left(\begin{pmatrix} 8 \\ 10 \\ 17 \end{pmatrix}, \begin{pmatrix} 29 & -1 & 10 \\ -1 & 9 & 12 \\ 10 & 12 & 21 \end{pmatrix}\right)$$

$$\therefore E(Y|Y_1 = y_1, Y_2 = y_2) = 17+(10,12)\begin{pmatrix} 29 & -1 \\ -1 & 9 \end{pmatrix}^{-1}\begin{pmatrix} y_1 - 8 \\ y_2 - 10 \end{pmatrix}$$
$$= \frac{1}{130}(51y_1+179y_2+12)$$

$$\therefore Cov(Y|Y_1 = y_1, Y_2 = y_2) = 21-(10,12)\begin{pmatrix} 29 & -1 \\ -1 & 9 \end{pmatrix}^{-1}\begin{pmatrix} 10 \\ 12 \end{pmatrix} = \frac{36}{65}$$

$$\therefore (Y|Y_1 = y_1, Y_2 = y_2) \sim N(\frac{1}{130}(51y_1+179y_2+12), \frac{36}{65})$$

二、(14分)

解:

经计算可得:

$$E[Y-(\beta_0+\beta^TX)]^2 = \beta_0^2+\beta^T(\Sigma_{xx}+\mu_x\mu_x^T)\beta+2\beta_0\beta^T\mu_x-2(\beta_0\mu_y+\beta^T\sigma_{xy}+\beta^T\mu_x\mu_y)$$
$$+\sigma_{yy}+\mu_y^2$$

上式右边记为$f(\beta_0,\beta)$,对β_0和β求导并令其为0可得:

$$\begin{cases} \dfrac{\partial f}{\partial \beta_0} = 2\beta_0 + 2\beta^T - 2\mu_y = 0 \\ \\ \dfrac{\partial f}{\partial \beta} = 2(\Sigma_{xx} + \mu_x\mu_x^T)\beta + 2\beta_0\mu_x - 2\sigma_{xy} - 2\mu_x\mu_y = 0 \end{cases}$$

解此方程组可得:

$$\begin{cases} \beta_0 = \mu_y - \sigma_{yx}\Sigma_{xx}^{-1}\mu_x \\ \\ \beta = \Sigma_{xx}^{-1}\sigma_{xy} \end{cases}$$

将β_0和β带入$f(\beta_0,\beta)$得:

$$\min_{\beta_0,\beta} E[Y - (\beta_0 + \beta^T X)]^2 = \sigma_{yy} - \sigma_{yx}\Sigma_{xx}^{-1}\sigma_{xy}$$

三、(14分)
解:
$T^2 = n(\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$

其中$n = 20, p = 3$,带入计算得

$T^2 = 0.4715$

$\dfrac{(n-1)p}{n-p} F_{p,n-p}(\alpha) = \dfrac{19 \times 3}{17} F_{3,17}(0.05) = 10.7194$

$\therefore T^2 < \dfrac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$,接受原假设.

四、(16分)
(1)解:

因子载荷矩阵$L = (\sqrt{\lambda_1}\phi_1, \sqrt{\lambda_2}\phi_2) = \begin{pmatrix} 0.882 & -0.181 \\ 0.826 & -0.404 \\ 0.714 & 0.693 \end{pmatrix}$

(2)解:
共性方差:

$$h_1^2 = 0.882^2 + 0.181^2 = 0.811$$
$$h_2^2 = 0.826^2 + 0.404^2 = 0.845$$
$$h_3^2 = 0.714^2 + 0.693^2 = 0.969$$

第1个因子的贡献:$\frac{\lambda_1}{p} = \frac{1.96}{3} = 0.653$

第2个因子的贡献:$\frac{\lambda_2}{p} = \frac{0.68}{3} = 0.227$

其统计意义是用少数几个公共因子解释协方差的结构,当后$p-m$ 个特征值较小时可以略去它们对方差的贡献.

五、(20分)
(1)
两因素方差分析模型:

$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1 \ldots a, \quad j = 1 \ldots b, \quad k = 1 \ldots c$

(2)
可以把因子分析看作是主成分分析的一个扩充,两者都力图逼近协方差矩阵.然而基于因子分析模型的近似是更为精细的,因子分析的主要问题是数据是否与一个规定的结构一致.

(3)
将多元观测值x变换成一元观测值y,使得由总体π_1和π_2 导出的y
尽可能的分开.

六、(12分)
解:

$$D = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ 2.5 & 1.5 & 0 & & \\ 6 & 5 & 3.5 & 0 & \\ 8 & 7 & 5.5 & 2.5 & 0 \end{bmatrix}$$

由于$\min_{i,k}(d_{ik}) = d_{12} = 1$,所以对象1和2合并形成聚类(12).

$d_{(12)3} = min\{d_{13}, d_{23}\} = min\{2.5, 1.5\} = 1.5$

$d_{(12)4} = min\{d_{14}, d_{24}\} = min\{6, 5\} = 5$

$d_{(12)3} = min\{d_{15}, d_{25}\} = min\{8, 7\} = 7$

新的距离矩阵为:

$$\begin{array}{c} (12) \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & & \\ 1.5 & 0 & & \\ 5 & 3.5 & 0 & \\ 7 & 5.5 & 2 & 0 \end{bmatrix}$$

将(12)和3聚类为(123)

$d_{(123)4} = min\{d_{(12)4}, d_{34}\} = min\{5, 3.5\} = 3.5$

$d_{(123)5} = min\{d_{(12)5}, d_{35}\} = min\{7, 5.5\} = 5.5$

新的距离矩阵为:

$$\begin{array}{c} (123) \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & \\ 3.5 & 0 & \\ 5.5 & 2 & 0 \end{bmatrix}$$

将4和5聚类为(45)

$d_{(123)(45)} = min\{d_{(123)4}, d_{(123)5}\} = min\{3.5, 5.5\} = 3.5$

最终的距离矩阵变为:

$$\begin{array}{c} (123) \\ (45) \end{array} \begin{bmatrix} 0 & \\ 3.5 & 0 \end{bmatrix}$$

最后聚为一类(12345)