

《多元统计与矩阵分析》样题

一、单项选择 (25 题, 每题 2 分, 共 50 分)

1. Q 型聚类是指对_____进行聚类

(A) 样品 (B) 变量 (C) 总体 (D) 元素

2. R 型聚类是指对_____进行聚类

(A) 样品 (B) 变量 (C) 总体 (D) 元素

3. 一元正态总体中, 用于参数检验的 F 分布推广到多元正态总体, 对应_____

(A) Wishart 分布 (B) Hotelling T^2 分布 (C) Wilks 分布 (D) Gauss 分布

4. 有因子分析模型,

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \boldsymbol{\varepsilon}$$

$$\mathbf{X} = (x_{ij})_{p \times n} \quad \mathbf{A} = (a_{ij})_{p \times m} \quad \mathbf{F} = (f_{ij})_{m \times n} \quad \boldsymbol{\varepsilon} = (\varepsilon_{ij})_{p \times n}$$

则下面肯定正确的是_____

(A) $\sum_{j=1}^m a_{ij} \leq 1$ (B) $\sum_{j=1}^m a_{ij} \geq 1$ (C) $\sum_{j=1}^m a_{ij}^2 \leq 1$ (D) $\sum_{j=1}^m a_{ij}^2 \geq 1$

5. 对二元(X, Y)数据进行主成分分析, 计算得到协方差矩阵如下,

$$\begin{matrix} & \mathbf{X} & \mathbf{Y} \\ \boldsymbol{\Sigma} = & \begin{bmatrix} 9 & -3 \\ -3 & 7 \end{bmatrix} \end{matrix}$$

则该数据的分布与图 1_____最类似。

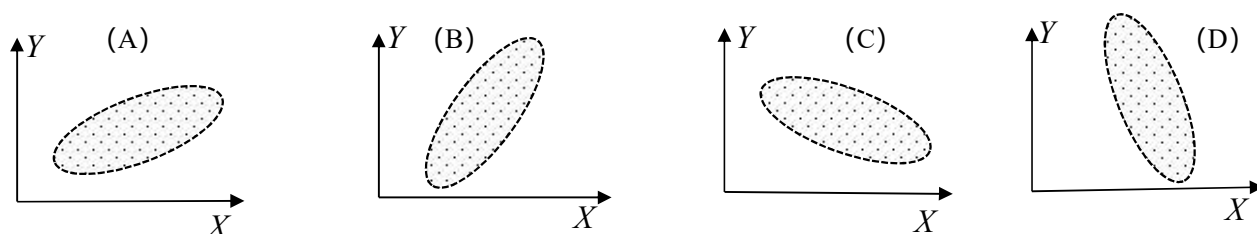
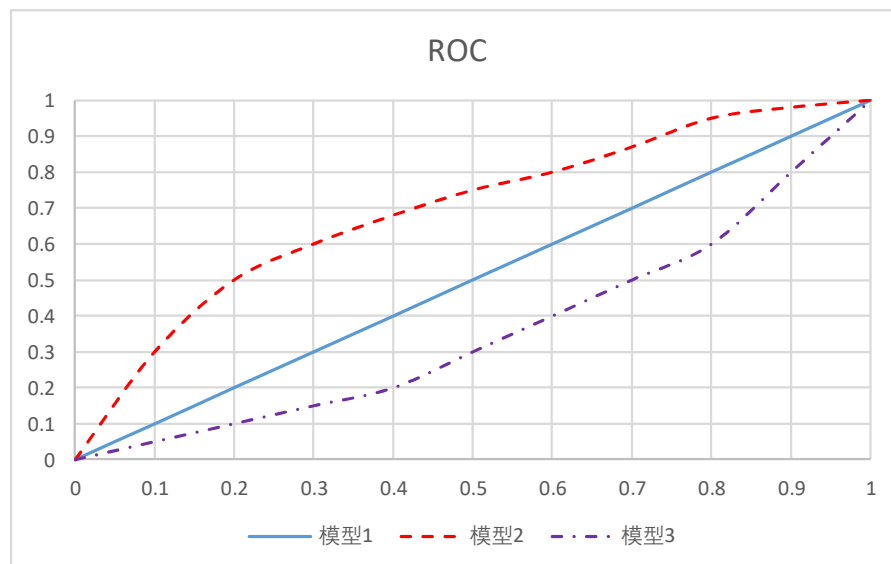


图 1 数据可视化

6. 下图展示了三个逻辑回归模型的 ROC 曲线。根据图中的结果，下面说法正确的是？

- (A) 模型 1 的分类性能是最差的
- (B) 模型 2 的分类性能是最好的
- (C) 模型 3 分类性能比模型 1 好
- (D) 模型 3 分类性能比模型 2 好



二、简答（4 题，每题 5 分，共 20 分）

1. 在因子分析中，共同因子的方差贡献怎么定义？有什么统计学意义？
2. 因子分析与对应分析异同？

三、证明题（2 题，共 15 分）

1. 设有因子分析模型：

$$X_i = w_{i1}F_1 + w_{i2}F_2 + \cdots + w_{im}F_m + r_i$$

$$(m \leq p, i = 1, 2, \dots, p)$$

方差 $\text{var}(r_i) = s_i^2$

试证明： $\sum_{j=1}^m w_{ij}^2 + s_i^2 = 1$

2. 有原始数据

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}' = (X_1 \ X_2 \ \dots \ X_p)'$$

试写出其第 k 主成分的数学表达式，并证明。

3. [因子分析]

P141 关于因子分析的因子旋转，有类似下面的关于因子共同度性质的一个证明。这一段证明中， $\sum_{j=1}^m \sum_{l=1}^m \sum_{t \neq l}^m a_{il} a_{it} \gamma_{lj} \gamma_{tj}$ 这段是等于 0 的，试给出详细证明（注意说明原理）。

$$\begin{aligned} \mathbf{B} &= (b_{ij})_{p \times m} = \left(\sum_{l=1}^m a_{il} \gamma_{lj} \right)_{p \times m} \\ h_i^2(\mathbf{B}) &= \sum_{j=1}^m b_{ij}^2 = \sum_{j=1}^m \left(\sum_{l=1}^m a_{il} \gamma_{lj} \right)^2 \\ &= \sum_{j=1}^m \sum_{l=1}^m a_{il}^2 \gamma_{lj}^2 + \sum_{j=1}^m \sum_{l=1}^m \sum_{t \neq l}^m a_{il} a_{it} \gamma_{lj} \gamma_{tj} \\ &= \sum_{l=1}^m a_{il}^2 \sum_{j=1}^m \gamma_{lj}^2 = \sum_{l=1}^m a_{il}^2 = h_i^2(\mathbf{A}) \end{aligned}$$

四、计算题（2 题，共 15 分）

1.

$$\text{设 } X = (X_1, X_2, X_3, X_4)^T \sim N_4(0, \Sigma), \text{ 协方差阵 } \Sigma = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix},$$

$$0 < \rho \leq 1$$

(1) 试从 Σ 出发求 X 的第一总体主成分；

(2) 试问当 ρ 取多大时才能使第一主成分的贡献率达 95% 以上。

2. 设三元总体 X 的协方差阵为 $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{pmatrix}$ ，从 Σ 出发，求总体主成分 F_1, F_2, F_3 ，并

求前两个主成分的累积贡献率。

3. 设三维随机向量 $X \sim N_3(\mu, \Sigma)$ ，其中 $\Sigma = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ ，问 X_1 与 X_2 是否独立？

$(X_1, X_2)'$ 和 X_3 是否独立？为什么？

4. 使用 Fisher 判别进行三维数据进行二分类，类别标记分别为 G_1, G_2 。假设

两类数据同分布，判别系数 $\alpha = (0.45 \ 0.23 \ 0.86)'$ ，两类数据的中心点分别为

$$\bar{x}^{(1)} = (2.1 \ 4.5 \ 9.7)', \quad \bar{x}^{(2)} = (10.4 \ 5.5 \ 1.1)'。$$

试写出其判别准则，并判断下面两个数据属于哪一类？

$$x_1 = (9.1 \ 5.5 \ 7.8)' \quad x_2 = (6.9 \ 5.6 \ 4.9)'$$