# 浙江大学20<u>18</u>—20<u>19</u> 学年 <u>秋冬</u> 学期

# 《多元统计分析》课程期末试卷

课程号：<u>06120340</u>， 开课学院：<u>数学学院</u>

考试试卷：A卷、B卷√（请在选定项上打√）

考试形式：闭√、开卷（请在选定项上打√），允许带<u>计算器</u>入场

考试日期：<u>2019</u> 年<u>1</u>月<u>25</u>日, 考试时间：<u>120</u> 分钟

**诚信考试，沉着应考，杜绝违纪。**

**考生姓名：____________ 学号：____________ 所属院系：____________**

| 题序 | 一 | 二 | 三 | 四 | 五 | 六 | 七 | 八 | 总分 |
|---|---|---|---|---|---|---|---|---|---|
| 得分 | | | | | | | | | |
| 评卷人 | | | | | | | | | |

一、填空题（33分）

1. 设随机向量$X=(X_1,X_2,X_3)'\sim N_3(\mu,\Sigma)$, 且$\mu=(1,-1,2)'$,

$$\Sigma=\begin{pmatrix}4&-1&-1\\-1&2&0\\-1&0&1\end{pmatrix},\quad A=\begin{pmatrix}3&-1&1\\-1&0&2\end{pmatrix},\quad d=\begin{pmatrix}1\\-1\end{pmatrix},$$

则$Y=AX+d$的分布为____________; $X_1$关于$X_2,X_3$的最优线性预测的表达式：____________; 最优线性预测函数的方差为____________。

2. 设$X_1,\cdots,X_{n_1}$ 和$Y_1,\cdots,Y_{n_2}$ 是分别来自总体分布为$N_p(\mu_i,\Sigma)$, $i=1,2$ 的两组相互独立的样本。记

$$\overline{X}=\frac{1}{n_1}\sum_{t=1}^{n_1}X_t,\quad \overline{Y}=\frac{1}{n_2}\sum_{t=1}^{n_2}Y_t,\quad A_x=\sum_{t=1}^{n_1}(X_t-\overline{X})(X_t-\overline{X})',\quad A_y=\sum_{t=1}^{n_2}(Y_t-\overline{Y})(Y_t-\overline{Y})',$$

$A = A_x + A_y$。则（1）当$n \to \infty$ 时，$\dfrac{A}{n_1+n_2-2}$ 依概率收敛于________________；(2) $\dfrac{n_1 n_2}{(n_1+n_2)}(\overline{X}-\overline{Y})'\Sigma^{-1}(\overline{X}-\overline{Y})$ 的分布为_____________; (3) $\dfrac{n_1 n_2(n_1+n_2-2)}{(n_1+n_2)}(\overline{X}-\overline{Y})'A^{-1}(\overline{X}-\overline{Y})$ 的分布为________________。

3. 已知$X=(X_1,X_2,X_3)'$的协方差矩阵的特征根分别为$\lambda_1=5.83$, $\lambda_2=2$, $\lambda_3=0.17$, 所对应的特征向量分别为$l_1=(0.383,-0.924,0)'$, $l_2=(0,0,1)'$, $l_3=(0.924,0.383,0)'$, 则$X$的第二主成分为________________；前两个主成分的累计贡献率为________________；第一个主成分对$X_3$的贡献率为________________。

4. 已知$G_1$, $G_2$ 两类数据服从多元正态分布$N_p(\mu_1,\Sigma)$, $N_p(\mu_2,\Sigma)$，从历史数据知这两类数据所占的比例相同, 错判所带来的损失也相同，则贝叶斯判别的解为：__________________; Fisher线性判别方程为：___________。

二、(17分) 已知总体$X=(X(1),X(2))' \sim N_p(\mu,\Sigma)$, 其中$\mu,\Sigma$未知, $p$为给定的整数。设$X_1,\cdots,X_n$ 是来自总体$X$ 的样本。

(1) 试给出$\mu,\Sigma$ 的极大似然估计量$\widehat{\mu}$, $\widehat{\Sigma}$;

(2) 判断$\widehat{\mu}$, $\widehat{\Sigma}$ 是否为无偏估计量?

(3) 若$p=2$, 从观测数据测得样本均值为$\overline{x}=\begin{pmatrix}1\\-1\end{pmatrix}$, 样本离差阵为$\Sigma=\begin{pmatrix}4&3\\3&9\end{pmatrix}$. 令$\theta$为$\mathrm{E}(X(1)|X(2))$ 的方差，试求$\theta$ 的极大似然估计量。

三、(18分) 设两类总体$X,Y$ 服从协方差相同的二元正态分布。现分别从总体$X,Y$ 中抽取3个个体，其相应的样本观察值如下：

$$\boldsymbol{X}=\begin{pmatrix}2&12\\4&10\\3&8\end{pmatrix},\quad \boldsymbol{Y}=\begin{pmatrix}5&7\\3&9\\4&5\end{pmatrix}.$$

(1) 试问在显著水平$\alpha=0.05$下, $X$与$Y$的均值是否有显著的差异? ($F_{2,3}(0.05)=9.55$.)

(2) 根据上面的数据信息，给出Fisher线性判别函数；

(3) 判断观测值$x_0=(2,12)$ 应属于哪一类?

四、（10分）下面是5个样品两两间的距离矩阵如下：

$$D=\left(\begin{array}{c|ccccc} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \hline x_1 & 0 & 8 & 2 & 4 & 9 \\ x_2 & 8 & 0 & 6 & 5 & 7 \\ x_3 & 2 & 6 & 0 & 10 & 2 \\ x_4 & 4 & 5 & 10 & 0 & 3 \\ x_5 & 9 & 7 & 2 & 3 & 0 \end{array}\right),$$

试用最大距离法对样品作系统聚类，并画出谱系聚类图。

五、（10分）已知$X=(X_1,X_2,X_3,X_4)'$为标准化随机向量，其协方差矩阵为

$$\Sigma=\begin{pmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{pmatrix}.$$

(1)试利用$\Sigma=AA'+D$求解因子载荷矩阵$A$ 和特殊因子的协方差阵$D$;

(2) 计算$X_3$的共同度$h_3^2$及对应于观测值$x_0=(1,2,3,4)'$的因子得分。

六、（12分）在100个学生中进行阅读速度$X_1$, 阅读能力$X_2$, 运算速度$Y_1$ 和运算能力测验，由所测得成绩算出相关系数矩阵如下：

$$R=\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix},$$

其中

$$\Sigma_{11}=\begin{pmatrix} 1 & 0.63 \\ 0.63 & 1 \end{pmatrix},\quad \Sigma_{12}=\Sigma_{21}'=\begin{pmatrix} 0.24 & 0.59 \\ -0.06 & 0.07 \end{pmatrix},\quad \Sigma_{22}=\begin{pmatrix} 1 & 0.42 \\ 0.42 & 1 \end{pmatrix}.$$

(1) 试求第一对典型相关变量和典型相关系数;

(2) 假设$X = (X_1, X_2)'$和$Y = (Y_1, Y_2)'$都服从正态分布，试给出检验阅读能力$X_2$ 与运算能力$Y_2$ 独立性的方法。