中国科学技术大学 2014 — 2015 学年第二学期考试试卷

考试科目 ______多元统计分析______ 得分 ____________

学生所在系 ________ 姓名 __________ 学号 ____________

（考试时间： 2015 年 7 月 1 日下午 14:30 — 16:30）

一、(计 26 分) 设 $X\sim N_3(\mu,\Sigma)$, 其中 $\mu=(-3,1,4)^T$,

$$\Sigma=\begin{bmatrix}1&-2&0\\-2&5&0\\0&0&2\end{bmatrix},$$

1. (2分 × 4) 下列各对随机变量独立吗？为什么？

   X_1 和 X_2； (X_1,X_2) 和 X_3； $\frac{X_1+X_2}{2}$ 和 X_3； X_2 和 $X_2-\frac{5}{2}X_1-X_3$ 。

2. (3分 × 2) 求 X_1-2X_2 与 X_2+2X_3 的联合分布及其相关系数。
3. (3分 × 2) 求:

   (a). 在给定 $(X_1=x_1,X_2=x_2)$ 条件下， X_3 的条件分布;

   (b). $E(X_3|X_1,X_2)$ 与 X_1 的相关系数。

4. (3分 × 2) 设

$$S=(X-\mu)^T\Sigma^{-1}(X-\mu)\le\chi^2_{0.05}(3),$$

   则下面结论对吗？为什么？

   (a). S 的中心为 μ; (b). 来自总体 X 的一个随机观测，其落入 S 的概率为 0.95;

   (c). S 的体积与 Σ 的行列式的值成比例.

二、(计 19 分) 从二元正态总体 X 中抽取容量为 42 的样本, 样本均值向量为 $\bar{x}=(0.564,0.603)^T$, 样本协方差阵为

$$S=\begin{bmatrix}0.0144&0.0117\\0.0117&0.0146\end{bmatrix},$$

S 的特征值为 0.026 和 0.002, 相应的特征向量为 $(0.704,0.710)^T$ 和 $(-0.710,0.704)^T$.

1. (5 分) 在显著性水平 $\alpha=0.05$ 下检验假设

$$H_0:\ \mu=\begin{bmatrix}0.562\\0.589\end{bmatrix},\quad H_1:\ \mu\ne\begin{bmatrix}0.562\\0.589\end{bmatrix}.$$

2. (6 分) 求 μ 的 95% 的置信域，并画出草图。
3. (8 分) 对 μ 的分量分别求同时 T^2 区间和同时庞弗罗尼区间。

三、(计 21 分) 来自总体 $X=(X_1,\cdots,X_7)^T$ 的容量为 3000 的样本所得的样本相关系数矩阵为

$$R=\begin{pmatrix}1&0.402&0.395&0.301&0.305&0.399&0.340\\0.402&1&0.618&0.150&0.135&0.206&0.183\\0.395&0.618&1&0.321&0.289&0.363&0.345\\0.301&0.150&0.321&1&0.846&0.759&0.661\\0.305&0.135&0.289&0.846&1&0.797&0.800\\0.399&0.206&0.363&0.759&0.797&1&0.736\\0.340&0.183&0.345&0.661&0.800&0.736&1\end{pmatrix}$$

1. (10=2+2+3+3 分) R 的特征值为 3.82,1.49,0.65,0.36,0.34,0.23,0.11, 相应的特征向量为 (0.285,0.211,0.294,0.435,0.453,0.453,0.434), (-0.351,-0.643,-0.515,0.240,0.282,0.167,0.182), (0.877,-0.246,-0.387,-0.113,-0.079,0.028,-0.027), (-0.088,0.686,-0.693,0.126,0.127,0.023,-0.090), (-0.076,-0.098,-0.112,-0.604,-0.024,-0.065,0.776), (0.112,-0.010,0.029,0.330,0.270,-0.873,0.208), (-0.023,0.020,-0.074,0.500,-0.787,0.024,0.352).

   求第二主成分的贡献率；估计第一样本主成分和第二样本主成分的相关系数；给出第一样本主成分的得分公式；求第一样本主成分与标准化后的 X_7 样本观测之间的相关系数.

2. (11=3+3+5 分) 对数据进行因子分析， 2 个公因子模型下因子载荷的极大似然估计如下

$$\begin{bmatrix} 0.270 & 0.031 & 0.209 & 0.856 & 0.957 & 0.812 & 0.799 \\ 0.483 & 0.792 & 0.764 & 0.165 & 0.122 & 0.256 & 0.221 \end{bmatrix}^T$$

   求相应于 X_7 的公共性方差和特殊方差；求 X_5 和 X_6 相关系数的估计；简述因子得分的回归方法是如何实现的，必要的步骤要写清楚。

四、(计 10 分) 利用完全连接法进行聚类分析，距离矩阵如下：

$$\begin{pmatrix} 0 & 33 & 37 & 24 & 31 & 36 & 39 \\ & 0 & 42 & 22 & 39 & 42 & 35 \\ & & 0 & 41 & 45 & 30 & 42 \\ & & & 0 & 41 & 32 & 40 \\ & & & & 0 & 46 & 48 \\ & & & & & 0 & 34 \\ & & & & & & 0 \end{pmatrix}.$$

五、(计 10 分) 设某种疾病的发病率为 10% , 针对有病和无病的两组对象分别各自进行了 A 和 B 的两种检测，检测精度见下表：

检测结果	A和B均为阳性	A阳性B阴性	A阴性B阳性	A和B均为阴性
患病组概率	0.90	0.01	0.05	0.04
无病组概率	0.04	0.06	0.08	0.82

若将有病误判为无病的损失是将无病误判为有病的损失的 20 倍，试构造分类准则使得期望的错分代价最小。

六、(计 14 分)

1. (7 分) 设 X 和 Y 是 p 维和 q 维随机变量，且存在二阶矩，设 $p \le q$ 。它们的第 j 对典型变量分别为 $a_j^T X$ 和 $b_j^T Y$ ，典型相关系数为 λ_j $(j = 1, \cdots, p)$. 令 $X^* = CX + \alpha$, $Y^* = DY + \beta$, 其中 C, D 分别为 $p \times p, q \times q$ 阶非奇异阵， α, β 分别为 p 维， q 维随机向量。试证明 X^* 、 Y^* 的第 j 对典型变量为 $C^{-1} a_j X^*$ 、 $D^{-1} b_j Y^*$ ，并求出相应的典型相关系数。

2. (7 分) 设 $X_1, \cdots, X_n$ 是来自正态总体 $N_p(\mu, \Sigma)$ 的一个随机样本，对于如下的假设检验：

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$$

   证明基于 Hotelling T^2 统计量的检验与基于似然比检验统计量的检验是等价的。

附： F 和 t 分布的部分上 α 分位点：

$F_{2,40}(0.0125) = 4.899$, $F_{2,40}(0.025) = 4.051$, $F_{2,40}(0.05) = 3.232$, $F_{2,40}(0.1) = 2.440$
$F_{2,41}(0.0125) = 4.886$, $F_{2,41}(0.025) = 4.042$, $F_{2,41}(0.05) = 3.226$, $F_{2,41}(0.1) = 2.437$
$F_{2,42}(0.0125) = 4.873$, $F_{2,42}(0.025) = 4.033$, $F_{2,42}(0.05) = 3.220$, $F_{2,42}(0.1) = 2.434$
$t_{40}(0.0125) = 2.329$, $t_{40}(0.025) = 2.021$, $t_{40}(0.05) = 1.684$, $t_{40}(0.1) = 1.303$
$t_{41}(0.0125) = 2.327$, $t_{41}(0.025) = 2.020$, $t_{41}(0.05) = 1.683$, $t_{41}(0.1) = 1.303$
$t_{42}(0.0125) = 2.325$, $t_{42}(0.025) = 2.018$, $t_{42}(0.05) = 1.682$, $t_{42}(0.1) = 1.302$