山东大学 多元统计分析 课程试卷（A）卷　　课程号 28230460　　2022-2023 学年 1 学期

一	二	三	四	五	六	七	八	总分	复核人

得分	阅卷人

一、试述主成分分析的基本思想．（10 分）

综合

主成分分析是将多个指标化为少数几个综合指标的统计方法。在实际问题中，研究多指标的问题是经常遇到的，由于变量个数太多，并且彼此之间存在一定的相关性，使得增加了分析问题的复杂性。

主成分分析就是设法把原来的多个指标重新组合成较少的几个新的不相关变量来代替原来的变量，而这几个综合变量又能尽可能地多反映原来变量的信息。

得分	阅卷人

二.试述 Wishart、Hotelling T^2、Wilks 三个分布的定义．（10 分）

Wishart：

设 $X_{(\alpha)} \sim N_p(0,\Sigma)$ $(\alpha=1,\cdots,n)$ 相互独立，记 $X=(X_{(1)},\cdots,X_{(n)})'$ 为 $n\times p$ 矩阵，则称随机阵

$$W=\sum_{\alpha=1}^{n} X_{(\alpha)}X'_{(\alpha)} = X'X$$ 的分布为 Wishart 分布，

记为 $W \sim W_p(n,\Sigma)$．

Hotelling T^2：设 $X_{(\alpha)}$ 是来自 p 元总体 $N_p(\mu,\Sigma)$ 的随机样本，$\bar{X}$ 和 A 分别是正态总体 $N_p(\mu,\Sigma)$ 的样本均值向量和样本离差阵，则

$$T^2=(n-1)[\sqrt{n}(\bar{X}-\mu)]'A^{-1}[\sqrt{n}(\bar{X}-\mu)]$$
$$=(n-1)n(\bar{X}-\mu)'A^{-1}(\bar{X}-\mu)$$
$$\sim T^2(p,n-1)$$

Wilks：设 $A_1 \sim W_p(n_1,\Sigma)$，$A_2 \sim W_p(n_2,\Sigma)$，$(\Sigma>0, n_1\geq p)$ 且 A_1 与 A_2 独立，则称广义方差之比

$$\Lambda=\frac{|A_1|}{|A_1+A_2|}$$ 为 Wilks 统计量，

记为 $\Lambda \sim \Lambda(p,n_1,n_2)$

得分	阅卷人

三、设 $\begin{cases} y_1 = a+\varepsilon_1, \\ y_2 = 2a-b+\varepsilon_2, \\ y_3 = a+2b+\varepsilon_3, \end{cases}$ $\varepsilon=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix} \sim N_3(0,\sigma^2 I_3)$.

试求参数 a,b 的最小二乘估计．（10 分）

$$\hat{\beta}=\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}=(X'X)^{-1}X'Y=$$

得分	阅卷人

四、已知随机变量 X 的密度为 $f(x)=\begin{cases} ax+b, & 0<x<1 \\ 0, & 其他 \end{cases}$，且 $P(X>\frac{1}{2})=\frac{5}{8}$，

（1）求 a,b；（2）计算 $P(\frac{1}{4}<X\leq\frac{1}{2})$．（10 分）

阅卷人

五、试证：随机向量 X 的协方差阵 $D(X)=\Sigma$ 是对称非负定阵（15 分）

$cov(X_i, X_j) = cov(X_j, X_i)$

$\therefore \Sigma = \Sigma'$

任意 $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$，$\alpha'\Sigma\alpha = \alpha' E[(X-E(X))(X-E(X))']\alpha$

$= E[(\alpha'(X-E(X)))^2] \geq 0$

$\therefore \Sigma \geq 0$

阅卷人

六、 证明对称幂等矩阵的特征值只能取 0 或 1。（15 分）

Pf: $\because A\xi = \lambda\xi$

若 A 为对称幂等矩阵

$\therefore A^2\xi = \lambda^2\xi = A\xi = \lambda\xi$

$\therefore \lambda^2 = \lambda$

$\therefore \lambda$ 为 0 或 1.

得分	阅卷人

七 设随机变量 (X,Y) 具有概率密度

$$f(x,y) = \begin{cases} \frac{1}{8}(x+y), & 0 \le x \le 2, 0 \le y \le 2, \\ 0, & others \end{cases}$$ 求 ρ_{XY}（15 分）

得分	阅卷人

八、设总体 X 在区间 $[0,\theta]$ 上服从均匀分布，其中 θ 是未知参数。

$X_1, X_2, \cdots, X_n$ 是总体 X 的一组样本，$x_1, x_2, \dots x_n$ 是 X 的一组样本观测值，求 θ 的矩估计，试问该估计量是否为无偏估计量？说明理由。（15 分）

$E(X) = \bar{X}$

$E(X) = \frac{\theta}{2}$

$\therefore \hat{\theta} = 2\bar{X}$ 即为矩估计量

又 $\because E(2\bar{X}) = E(\hat{\theta}) = \theta$

$\therefore \hat{\theta} = 2\bar{X}$ 是无偏估计量

对任意的$\alpha=\begin{bmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_p\end{bmatrix}$，有

$$\alpha'\Sigma\alpha=\alpha'E\left[(X-E(X))(X-E(X))'\right]\alpha=E\left[\alpha'(X-E(X))(X-E(X))'\alpha\right]$$

$$=E\left[(\alpha'(X-E(X)))^2\right]\geq 0$$

所以$\Sigma\geq 0$….10 分

六、解:

$$A\xi=\lambda\xi$$

$$A^2\xi=A(A\xi)=A(\lambda\xi)=\lambda A\xi=\lambda^2\xi$$

又因为A是对称幂等矩阵，所以$A^2\xi=A\xi$

从而$\lambda^2\xi=\lambda\xi$,即$(\lambda^2-\lambda)\xi=0$, ...（10 分）

又$\xi\neq 0$，所以$\lambda^2-\lambda=0$

即：$\lambda=0$或1

七、解:

解:（1）(X,Y)关于X和 Y的边缘概率密度分别为

$$f_X(x)=\begin{cases}\int_0^2\frac{1}{8}(x+y)dy=\frac{1}{4}(x+1) & 0\leq x\leq 2,\\ 0 & others.\end{cases}$$

$$f_Y(y)=\begin{cases}\int_0^2\frac{1}{8}(x+y)dx=\frac{1}{4}(y+1) & 0\leq y\leq 2,\\ 0 & others.\end{cases}$$ ....2 分

$$E(X)=\int_0^2\frac{x}{4}(x+1)dx=\left(\frac{x^3}{12}+\frac{x^2}{8}\right)\Big|^2_{x=0}=\frac{7}{6},$$

由X与Y的对称性同理可求得 $E(Y)=\frac{7}{6}$.

$$E(XY)=\int_0^2\int_0^2 xy\frac{1}{8}(x+y)dxdy=\frac{4}{3}.$$

扫描全能王 创建

$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{[illegible]}{3}\ \ldots\ 36\ \ldots\ 36$

$$D(X)=E(X^2)-E^2(X)=\int_0^2\frac{1}{4}x^2(x+1)dx-\frac{49}{36}=\frac{5}{3}-\frac{49}{36}=\frac{11}{36}.$$

同理 $D(Y)=\frac{11}{36}.$ …………………2 分

$$\rho_{XY}=\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=-\frac{1}{11}.$$ …………………1 分

八、解： （1）由矩估计得：$E(X)=\overline{X}$，又因为 X 在区间 $[0,\theta]$ 上服从均匀分布，所以

$E(X)=\frac{\theta}{2}$，故 θ 的矩估计量为

$\hat{\theta}=2\overline{X}$. …………………3 分

由于 $E(\hat{\theta})=E(2\overline{X})=\theta$，所以 θ 的矩估计量 $\hat{\theta}=2\overline{X}$ 是无偏估计量。…………….2 分

山东大学 多元统计分析 课程试卷（B）卷　　课程号 28230460　　2022-2023 学年 1 学期

题号	一	二	三	四	五	六	七	八	总分	阅卷人
得分										

得分	阅卷人

一、试述逐步回归的基本思想.（10 分）

得分	阅卷人

二、试述主成分分析的主要思想.（10 分）

得分	阅卷人

三、设三维随机向量 $X \sim N_3(u, 2I_3)$，已知

$u=\begin{bmatrix}2\\0\\0\end{bmatrix}$，$A=\begin{bmatrix}0.5 & -1 & 0.5\\ -0.5 & 0 & -0.5\end{bmatrix}$，$d=\begin{bmatrix}1\\2\end{bmatrix}$，试求 $Y=AX+d$ 的分布.（10 分）

$Y \sim N(Au+d, A2I_3A')$

$Au+d=\begin{bmatrix}0.5 & -1 & 0.5\\ -0.5 & 0 & -0.5\end{bmatrix}\begin{bmatrix}2\\0\\0\end{bmatrix}+\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}1\\-1\end{bmatrix}+\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}$

$A2I_3A'=2\begin{bmatrix}0.5 & -1 & 0.5\\ -0.5 & 0 & -0.5\end{bmatrix}\begin{bmatrix}0.5 & -0.5\\ -1 & 0\\ 0.5 & -0.5\end{bmatrix}=2\begin{bmatrix}1.5 & -0.5\\ -0.5 & 0.5\end{bmatrix}=\begin{pmatrix}3 & -1\\ -1 & 1\end{pmatrix}$

得分	阅卷人

四、设三个总体 G_1, G_2 和 G_3 的分布分别为：$N(2, 0.5^2)$，$N(0, 2^2)$ 和 $N(3, 1^2)$.

试问：按距离判别准则，样品 $x=2.5$ 应判归哪一类？（10 分）

$D_1=\frac{(2.5-2)^2}{0.5^2}=1$

$D_2=\frac{(2.5)^2}{2^2}=(1.25)^2$

$D_3=\frac{(2.5-3)^2}{1^2}=0.25$

贝叶斯

$D_1=\frac{(2.5-2)^2}{0.5^2}+d_1=\frac{(2.5-2)^2}{0.5^2}+\ln(0.5^2)=$

得分	阅卷人

五、试证：随机向量 X 的协方差阵 $D(X)=\Sigma$ 是对称非负定阵（15 分）

得分	阅卷人

、六（本题 10 分）设一个仓库中共有 10 箱同样规格的产品，已知这 10 箱产品中依次有 5，3，2 箱是甲、乙、丙三个厂生产的，又知，甲、乙、丙三厂生产的该产品的次品率依次为 $\frac{1}{10},\frac{1}{15},\frac{1}{20}$。现从这 10 箱产品中任取一箱，再从取得的这箱中任取一件产品，求取得正品的概率；假如取到这件产品是正品，问所取的那箱产品是甲厂生产的概率是多少？（15 分）

得分	阅卷人

七、已知随机变量 X 的密度为 $f(x)=\begin{cases} ax+b, & 0<x<1 \\ 0, & 其他 \end{cases}$，

且 $P(X>\frac{1}{2})=\frac{5}{8}$，（1）求 a,b；（2）计算 $P(\frac{1}{4}<X\le\frac{1}{2})$。（15 分）

得分	阅卷人

八、设总体 X 在区间 $[0,\theta]$ 上服从均匀分布，其中 θ 是未知参数。

$X_1,X_2,\cdots,X_n$ 是总体 X 的一组样本，$x_1,x_2,\ldots x_n$ 是 X 的一组样本观测值。

（1）求 θ 的矩估计，试问该估计量是否为无偏估计量？说明理由.（15 分）

扫描全能王 创建

22-23(1) 多元统计分析试题卷(801 课 B)答案

一、1. 逐个引入自变量，每次引入对 Y 影响最显著的自变量，并对方程中的老变量逐个进行检验，把变为不显著的变量逐个从方程中剔除掉。最终得到的方程中即不漏掉对 Y 影响显著的变量，又不包含对 Y 影响不显著的变量。--10 分

二、主成分分析是将多个指标化为少数几个综合指标的一种统计分析方法.

在实际问题中,研究多指标的问题是经常遇到的问题.由于变量个数太多,并且彼此之间存在着一定的相关性,势必增加分析问题的复杂性.

主成分分析就是设法把原来的多个指标重新组合成较少几个新的互不相关的综合变量来代替原来的变量;而且这几个综合变量又能够尽可能多地反映原来变量的信息.

。。。。。。。。。。。。。。 10 分

三、解:

解:

$$Y \sim N_2\left(Au+d, A2I_3A'\right) \text{……………5 分}$$

其中 $Au+d=\begin{bmatrix}2\\1\end{bmatrix}$，$A2I_3A'=\begin{bmatrix}3 & -1\\-1 & 1\end{bmatrix}$……………10 分

四、

五、证明:

因为 $\text{cov}(X_i, X_j)=\text{cov}(X_j, X_i)$，所以 $\Sigma=\Sigma'$。

对任意的 $\alpha=\begin{bmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_p\end{bmatrix}$，有

$$\alpha'\Sigma\alpha=\alpha' E\left[(X-E(X))(X-E(X))'\right]\alpha=E\left[\alpha'(X-E(X))(X-E(X))'\alpha\right]$$

$$=E\left[(\alpha'(X-E(X)))^2\right]\geq 0$$

所以 $\Sigma\geq 0$....10 分

六、解：设 A_1, A_2, A_3 分别表示“任取一箱，该箱是甲、乙、丙厂生产的”，则

$P(A_1)=\frac{5}{10}, P(A_2)=\frac{3}{10}, P(A_3)=\frac{2}{10}$……………………………………（3 分）

设 B 表示“任取一箱，再任取一件，该产品是正品”

由全概率公式，$P(B)=\sum_{i=1}^{3}P(A_i)P(B\mid A_i)$……………………………………（7 分）

$$=\frac{5}{10}\times\left(1-\frac{1}{10}\right)+\frac{3}{10}\times\left(1-\frac{1}{15}\right)+\frac{2}{10}\times\left(1-\frac{1}{20}\right)=0.92$$ ………………………（10 分）

已知是正品，该产品来自于甲厂的概率

$$P(A_1\mid B)=\frac{P(A_1)P(B\mid A_1)}{\sum_{i=1}^{3}P(A_i)P(B\mid A_i)}=\frac{P(A_1)P(B\mid A_1)}{P(B)}$$ ………………………（13 分）

$$=\frac{\frac{5}{10}\times\left(1-\frac{1}{10}\right)}{0.92}=\frac{45}{92}\approx 0.489$$ ………………………（15 分）

七、解：

由 $\int_{-\infty}^{+\infty} f(x)dx=\int_0^1(ax+b)dx=\frac{a}{2}+b=1$，且 $P(X>\frac{1}{2})=\int_{\frac{1}{2}}^{1}(ax+b)dx=\frac{3}{8}a+\frac{b}{2}=\frac{5}{8}$

得： $a=1,b=\frac{1}{2}$. ………………………5 分

(2) $P(\frac{1}{4}<X\le\frac{1}{2})=\int_{\frac{1}{4}}^{\frac{1}{2}}(x+1/2)dx=\frac{7}{32}$ ………………………5 分

八、解： （1）由矩估计得：$E(X)=\overline{X}$，又因为 X 在区间 $[0,\theta]$ 上服从均匀分布，所以

$E(X)=\frac{\theta}{2}$，故 θ 的矩估计量为

$\hat{\theta}=2\overline{X}$. ………………………3 分

由于 $E(\hat{\theta})=E(2\overline{X})=\theta$，所以 θ 的矩估计量 $\hat{\theta}=2\overline{X}$ 是无偏估计量。………………2 分