

# 西安交通大学考试题数理统计 2000 年

## 一、填空

1. 设  $X_1, X_2, \dots, X_{10}$  是来自正态总体  $N(0, 4)$  的样本,

$$c \cdot \frac{\sum_{i=1}^5 X_i}{\sqrt{\sum_{j=1}^5 (X_j - \frac{1}{5} \sum_{i=1}^5 X_i)^2 + \sum_{i=6}^{10} X_i^2}} \sim t(m)$$

则  $c = 70$  ,  $m = 49$

2. 用  $\Phi(x)$  表示标准正态分布  $N(0, 1)$  的分布函数, 则  $\Phi(-x)$  与  $\Phi(x)$  的关系为

$\Phi(-x) = 1 - \Phi(x)$

3. 已知  $T \sim t(n)$ , 则  $T^2 \sim F(1, n)$

4. 设总体  $X$  的概率密度为  $f(x; \theta)$ , 则参数  $\theta$  估计的费歇 (Fisher) 信息量

$I(\theta) = -E \left[ \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right]$

## 二、选择题 (填 A,B,C,D, 有几个正确填几个, 若都不正确, 则填 E)

1. 设  $(X_1, X_2, \dots, X_{20})$  是来自正态总体  $N(\mu, \sigma^2)$  的样本, 统计量

$S = \sum_{i=1}^{10} (2X_{2i} - X_{2i-1})^2$ , 则

(C)

A.  $\frac{1}{5\sigma^2} S \sim \chi^2(9)$

B.  $\frac{1}{3\sigma^2} S \sim \chi^2(9)$

C.  $\frac{1}{5\sigma^2} S \sim \chi^2(10)$

D.  $\frac{1}{3\sigma^2} S \sim \chi^2(10)$

2. 独立地分别从两总体  $X$  和  $Y$  中抽得大小各为  $m$  和  $n$  的样本, 其样本均值分别为  $\bar{X}$  和  $\bar{Y}$ , 则  $D(\bar{X} - \bar{Y}) =$

(D)

A.  $\frac{D(X)}{m^2} - \frac{D(Y)}{n^2}$

B.  $\frac{D(X)}{m^2} + \frac{D(Y)}{n^2}$

C.  $\frac{D(X)}{m} - \frac{D(Y)}{n}$

D.  $\frac{D(X)}{m} + \frac{D(Y)}{n}$

3. 设  $X_1, X_2, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的样本, 其中  $\mu$  已知, 而  $\sigma^2$  未

知,则下列是统计量的是

(AD)

A.  $\bar{X} + X_6$

B.  $\frac{1}{\sigma^2} \sum_{i=1}^n X_i^2$

C.  $\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2$

D.  $(X_2 - \mu)^2 + 2X_8 + \mu$

4. 设  $F \sim F(6,8)$ , 则

(AD)

A.  $P\{F < F_{1-\alpha}(6,8)\} = \alpha$

B.  $P\{|F| > \frac{1}{F_{\alpha}(8,6)}\} = \alpha$

C.  $P\{|F| > F_{1-\frac{\alpha}{2}}(6,8)\} = \alpha$

D.  $P\{F < \frac{1}{F_{\alpha}(8,6)}\} = \alpha$

三、为比较 A、B 两型灯泡的寿命,随机抽取 A 型灯泡 5 只,测得平均寿命  $\bar{x}=1000$  (小时),标准差  $s_A=28$  (小时);随机抽取 B 型等泡 7 只,测得平均寿命  $\bar{y}=980$  (小时),标准差  $s_B=32$  (小时),设总体都是正态的,试在显著性水平之下 ( $\alpha=0.05$ ) 检验两总体寿命分布是否相同。

解:

依题意,需要检验假设为

$$H_{01} : \sigma_x = \sigma_y \leftrightarrow H_{02} : \sigma_x \neq \sigma_y$$

$$H_{02} : \mu_x = \mu_y \leftrightarrow H_{02} : \mu_x \neq \mu_y$$

i) 对于假设  $H_{01}$ , 由于  $\mu_x, \mu_y$  未知,则可选取

$$F = \frac{S_x^{*2}}{S_y^{*2}}$$

作为检验统计量,当  $H_{01}$  成立时,

$$F = \frac{S_x^{*2}}{S_y^{*2}} \sim F(n_x - 1, n_y - 1)$$

且当  $H_{01}$  成立时,  $F$  过分偏大将不利于  $H_{01}$  的成立,则其拒绝域的形式为

$$W = \{F \leq k_1 \cup F \geq k_2\}$$

对于给定的水平  $\alpha=0.05$ , 于是  $k_1 = F_{0.975}(n_x - 1, n_y - 1), k_2 = F_{0.025}(n_x - 1, n_y - 1)$

对于本题,

$$n_x = 5, S_x^{*2} = n_x \cdot s_A^2 / (n_x - 1) = 980, n_y = 7, S_y^{*2} = n_y \cdot s_B^2 / (n_y - 1) = 7 \times 32^2 / 6$$

则

$$F = \frac{S_x^{*2}}{S_y^{*2}} = \frac{980 \times 6}{7 \times 32^2} = 0.8203$$

查表得,  $F_{0.025}(4,6) = 6.23, F_{0.975}(4,6) = 1/F_{0.025}(6,4) = 1/9.2 = 0.0892$ 。由于  $0.0892 < F < 6.23$ , 则接受假设, 即认为  $\sigma_x = \sigma_y$ 。

ii) 对于假设  $H_{02}$ , 由于  $\sigma_x = \sigma_y$ , 且未知, 则可选取

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot \sqrt{\frac{(n_x - 1)S_x^{*2} + (n_y - 1)S_y^{*2}}{n_x + n_y - 2}}}$$

作为检验统计量, 当  $H_{02}$  成立时,

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot \sqrt{\frac{(n_x - 1)S_x^{*2} + (n_y - 1)S_y^{*2}}{n_x + n_y - 2}}} \sim t(n_x + n_y - 2)$$

且当  $H_{01}$  成立时,  $T$  过分偏大将不利于  $H_{02}$  的成立, 则其拒绝域的形式为

$$W = \{|T| \geq k\}$$

对于给定的水平  $\alpha = 0.05$ , 于是  $k = t_{0.025}(n_x - 1, n_y - 1)$

对于本题,

$$n_x = 5, S_x^{*2} = n_x \cdot s_A^2 / (n_x - 1) = 980, n_y = 7, S_y^{*2} = n_y \cdot s_B^2 / (n_y - 1) = 7 \times 32^2 / 6$$

则  $T = 1.0258$ , 查表得,  $T_{0.025}(10) = 2.2281$ 。由于  $1.0258 < T$ , 则拒绝假设, 即认为  $\mu_x \neq \mu_y$ 。

综上所述, 两总体寿命分布不相同。

四、想要考察特定一群人的收入与其花在书籍报纸上的支出有无关系, 把收入分成高、中、低三档, 书报上的支出分别为多、少两档。设随机抽查了 201 名, 结果如下表:

收入 \ 书报上支出	低	中	高
少	63	37	60
多	16	17	8

试在水平  $\alpha = 0.05$  之下检验收入与书报上支出有无关联。

解：

依题意，有两个指标，记  $X$  为书报上的支出情况，记  $Y$  为收入情况，且记  $A_1, A_2$  分别为书报上的支出多、少两档； $B_1, B_2, B_3$  分别为收入高、中、低三档。则需要检验的假设为  $H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j}, i=1,2; j=1,2,3 \leftrightarrow H_1: \text{至少存在某组}(i,j)$ ，使得  $p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j}$ 。其中  $p_{ij} = P\{X = A_i, Y = B_j\}, p_{i\cdot} = P\{X = A_i\}, p_{\cdot j} = P\{Y = B_j\}$ 。于是，

$$k = \frac{\sum_{j=1}^3 \sum_{i=1}^2 (n_{ij} - \frac{n_{i\cdot} n_{\cdot j}}{n})^2}{\frac{n_{i\cdot} n_{\cdot j}}{n}} = 6333.7$$

查表得， $\chi_{0.05}^2(2) = 5.991$ 。由于  $k > 5.991$ ，则拒绝假设，即认为收入与书报上支出有关联。

五、在硝酸钠（ $\text{NaNO}_3$ ）的溶解度试验中，测得在不同温度  $x$  ( $^{\circ}\text{C}$ ) 下，溶解于 9 份水中的硝酸钠份数  $y$  的数据如下表：

$x: 0 \quad 4 \quad 10 \quad 15 \quad 21 \quad 29 \quad 36 \quad 51 \quad 68$

$y: 66.7 \quad 71.0 \quad 76.3 \quad 80.6 \quad 85.7 \quad 92.9 \quad 99.4 \quad 113.6 \quad 125.1$

假设  $y$  与  $x$  之间有线性关系，在正态假定下，求  $y$  在  $x_0 = 25$  的置信度为 95% 的预测区间。

解：

依题意，可计算得

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0.8706$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 67.5078$$

则  $y$  与  $x$  之间有线性关系为  $y = 0.8706x + 67.5078$ 。

于是，当  $x_0 = 25$  时， $y_0 = 89.2728$ 。

$y$  置信度为 95% 的预测区间为

$$(y_0 \mp t_{0.025}(n-2) \cdot \sigma^* / \sqrt{1 + 1/n + (x_0 - \bar{x}) / \sum_{i=1}^n (x_i - \bar{x})})$$

则  $y$  在  $x_0 = 25$  的置信度为 95% 的预测区间为

六、设自一大批产品中随机抽出 200 个产品，发现其中 120 个是一等品，求这大批产品的一等品率的 95% 置信区间。

解：

$$U = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \sim N(0,1)$$

七、今有两台测量合金材料中某种金属含量的光谱仪，为鉴定他们的测量准确性有无显著差异，对 9 件含该金属分别为  $x_1, x_2, \dots, x_9$ （不等）的合金材料进行测量，第一台测量结果服从正态分布  $N(x_t + \delta_1, \sigma_1^2)$ ， $t = 1, 2, \dots, 9$ ；第二台测量结果服从  $N(x_t + \delta_2, \sigma_2^2)$ ， $t = 1, 2, \dots, 9$ 。测得的 9 对观测值如下：

第一台： 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00

第二台： 0.10 0.21 0.52 0.32 0.78 0.59 0.68 0.77 0.89

问能否认为第一台的测量值比第二台显著偏大（ $\alpha = 0.05$ ）？

八、设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本（ $n > 2$ ），总体的概率密度为

$$f(x; \lambda, a) = \begin{cases} \lambda e^{-\lambda(x-a)}, & \text{当 } x \geq a \\ 0, & \text{当 } x < a \end{cases} \quad (\text{参数 } \lambda > 0)$$

1. 设  $\lambda = \frac{1}{\sqrt{2}}$ ，试求参数  $a$  的最大似然估计。
2. 设  $a = 0$ ，试求参数  $\lambda$  的矩估计。
3. 设  $a = 0$ ，试推导  $2n\lambda\bar{X}$  服从的分布，其中  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 。
4. 设  $a = 0$ ，试计算  $E(\hat{\lambda})$ ，并求  $k$ ，使  $\hat{\lambda}^* = k\hat{\lambda}$  为  $\lambda$  的无偏估计。

解：

1. 先得到似然函数

$$L(a) = \prod_{i=1}^n \lambda e^{-\lambda(x_i - a)} = \lambda^n e^{-\lambda \sum_{i=1}^n (x_i - a)} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i + n\lambda a} \quad x_i \geq a$$

由于

- 2.

一. (本题满分 20 分)

填空题:

1. 设  $X_1, X_2, \dots, X_{10}$  是来自正态母体  $N(\mu, \sigma^2)$  的一个简单随机子样, 其中  $\mu, \sigma^2$  已知. 填充下列统计量的分布及其相应参数:

A.  $\frac{X_2 + \mu}{\sigma} + \frac{X_9 - 2\mu}{2\sigma} \sim \underline{N(\frac{3\mu}{2\sigma}, \frac{5}{4})}$

B.  $\frac{\sum_{i=1}^{10} (X_i - \mu)^2}{\sigma^2} \sim \underline{\chi^2(10)}$

C.  $\frac{2 \sum_{i=1}^6 (X_i - \mu)^2}{3 \sum_{i=7}^{10} (X_i - \mu)^2} \sim \underline{F(6, 4)}$

2. 设有一母体  $X$ , 其均值  $EX = \mu$ , 方差  $DX = \sigma^2$  以及四阶矩  $EX^4$  都存在,  $X_1, X_2, \dots, X_n$  是来自母体  $X$  的简单随机子样. 则  $\mu$  的无偏估计量为  $\bar{X}$ , 相合估计量为  $\bar{X}$ ;  $\sigma^2$  的无偏估计量为  $EX^2 - EX$ , 相合估计量为  $EX^2 - EX$ .

二. 选择题 (本题满分 20 分, 从 A~E 中选择一个完整的答案, 填入指定处)

1. 设  $X \sim N(0, 1)$ , 则  $P\{X > \cdot\} = 1 - a$  ( $0 < a < 1$ ). (C)

A.  $u_a$  B.  $-u_a$  C.  $u_{1-a}$  D. B 或 C E. A~D 的答案皆错

2. 设  $F \sim F(m, n)$ , 则  $P\{F > \cdot\} = 1 - a$  ( $0 < a < 1$ ). (E)

A.  $-F_a(m, n)$  B.  $F_{1-a}(m, n)$  C.  $F_a(m, n)$  D.  $F_a^{-1}(n, m)$  E. B 或 D

3. 设检验假设  $H: \theta = \theta_0$  的一个检验法则犯第一类错误的概率为  $P(I)$ , 检验的显著水平为  $\alpha$ , 则. (C)

A.  $P(I) = 1 - \alpha$  B.  $P(I) = \alpha / 2$  C.  $P(I) = \alpha$  D.  $P(I) \geq 1 - \alpha$  E. C 或 D

4. 设  $X_1, X_2, \dots, X_n$  是来自正态母体  $N(\mu, \sigma^2)$  的子样, 其中  $\mu$  未知,  $\sigma^2$  已知, 则是统计量. (D)

A.  $S^2 / \sigma^2$ ; B.  $(X_1 - \mu) / \sigma$ ; C.  $|X_1| + |X_2|$ ; D. A 和 C; E. A 和 B

5. 设母体  $X$  及  $Y$  的分布式任意的, 但分别是具有有限的非零方差, 记  $EX = \mu_1$ ,

$EY = \mu_2$ , 现独立地从两母体中各取一个子样, 子样容量分别是  $n_1$  和  $n_2$ . 在大子样下,

我们可以推出  $\mu_1 - \mu_2$  的置信概率近似为  $1 - \alpha$  的置信区间。这里所谓的大子样，一般是指。

(E)

A.  $n_1 \geq 50$ ; B.  $n_2 \geq 50$ ; C.  $n_1 + n_2 \geq 50$ ; D. A 且 B; E. A~D 的答案皆错

三. (本题满分 20 分)

设母体  $X$  的概率密度为

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (0 < \theta < +\infty)$$

1. 求  $\theta$  的矩估计量和最大似然估计量;
2. 用以上方法求得的估计量是否为  $\theta$  的无偏估计? 是否为  $\theta$  的相合估计?

解:

1.  $\theta$  的矩估计量和最大似然估计量均为  $\theta$
2. 显然用以上方法求得的估计量是  $\theta$  的无偏估计, 也为  $\theta$  的相合估计。

四. (本题满分 14 分)

已知某种设备的工作温度服从正态分布, 现对该温度作 10 次测量, 得数据 ( $^{\circ}\text{C}$ )

1250 1275 1265 1245 1260 1255 1270 1265 1250 1240

1. 求温度的母体均值  $\mu$  的 95% 置信区间;
2. 求温度的母体标准差  $\sigma$  的 95% 置信区间。

解:

1. 枢轴量

$$T = \frac{\bar{X} - \mu}{S^*/\sqrt{n}} \sim t(n-2)$$

2. 枢轴量

$$\chi^2 = \frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2(n-1)$$

五. (本题满分 14 分)

设有两个独立的来自不同正态母体的子样

$(-4.4, 4.0, 2.0, -4.8), (6.0, 1.0, 3.2, -4.0)$

问能否认为两个子样来自同一母体 ( $\alpha = 0.05$ ) ?

解:

(提示: 首先检验两母体的方差是否相同, 其次检验两母体的均值是否相同)

六. (本题满分 6 分)

设母体  $X \sim N(\mu, 1)$ , 希望检验假设  $H_0: \mu = 6 \leftrightarrow H_1: \mu = 7$ 。若从该母体中取出容量为 4 的简单随机子样, 并采用如下检验法则:

当  $\bar{X} \geq 7$  时, 拒绝  $H_0$ , 接受  $H_1$ ; 当  $\bar{X} < 7$  时, 接受  $H_0$ , 拒绝  $H_1$ 。求上述检验



法则犯第一、二类错误的概率。

解：

利用上述检验法则犯第一、二类错误的概率分别为

$$P_{\theta^1} = 1 - \Phi\left(\frac{7-6}{\sqrt{4}}\right) = 1 - \Phi(0.5)$$

$$P_{\theta^2} = \Phi\left(\frac{7-7}{\sqrt{4}}\right) = 0.5$$

七. (本题满分 6 分)

设  $t_{\alpha}(n), F_{\alpha}(m, n)$  分别表示  $t(n), F(m, n)$  分布相应的上侧分位数,

$$\text{求证: } [t_{\alpha/2}(n)]^2 = F_{\alpha}(1, n)$$

证明：

设  $X \sim N(0,1), Y \sim \chi^2(n)$  则,

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

于是, 对于给定显著水平  $\alpha$ , 有  $P\{|T| \geq t_{\alpha/2}\} = \alpha$ 。而

$$T^2 = \frac{X^2}{Y/n} \sim F(1, n)$$

于是,  $P\{|T^2| \geq [t_{\alpha/2}(n)]^2\} = \alpha = P\{|T^2| = F \geq F_{\alpha}(1, n)\}$

故  $[t_{\alpha/2}(n)]^2 = F_{\alpha}(1, n)$

一. (本题满分 14 分)

已知某零件的长度服从正态分布  $N(u, \sigma^2)$ , 其中  $\sigma^2 = 5.5 \text{ mm}^2$ , 从一大堆这种零件中随机抽取  $n$  个, 测量其长度。现用子样均值  $\bar{X}$  来估计母体均值  $u$ , 此时:

- (1) 若要估计量的标准差在  $1 \text{ mm}^2$  之下,  $n$  应取多大?
- (2) 若要估计误差的绝对值超过  $1 \text{ mm}$  的概率在 1% 以下,  $n$  应取多大?

解:

二. (本题满分 20 分)

判断下列命题的真伪并简述理由:

1. “统计量”与“估计量”是同一概念。(错, 用作估计的统计量称为估计量, 因此统计量的概念大一些)

2. “点估计”与“区间估计”的关系为: 前者是后者的一种特殊形式, 即区间估计的区间长度为一个点即为点估计。

3. 设母体  $X$  的均值和方差都存在,  $X_1, X_2, X_3$  为来自母体  $X$  的一个简单随机子样, 则  $\theta_1 = \frac{1}{3}(X_1 + X_2 + X_3)$  与  $\theta_2 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$  都是  $E(X)$  的无偏估计, 且  $\hat{\theta}_1$  比  $\hat{\theta}_2$  有效。  
(对, 利用方差比较, 越小越有效)

(4) 在一个确定的假设检验问题中, 其判断结果不但与其检验水平  $\alpha$  有关, 而且与抽到的子样有关。(对, 因此才会存在第一、二类错误)

四. (本题满分 14 分)

已知某种设备的工作温度服从正态分布, 现作十次测量, 得数据 ( $^{\circ}\text{C}$ )

1250 1275 1265 1245 1260 1255 1270 1265 1250 1240

- (1) 求温度的母体均值  $u$  的 95% 置信区间。
- (2) 求温度母体标准差  $\sigma$  的 95% 置信区间。

解:

(1) 枢轴量

$$T = \frac{\bar{X} - \mu}{S^*/\sqrt{n}} \sim t(n-1)$$

(2) 枢轴量

$$\chi^2 = \frac{(n-1)S^*}{\sigma^2} \sim \chi^2(n-1)$$

五. (本题满分 14 分)

设有两个独立的来自不同的正态母体的子样:

$$(-4.4, 4.0, 2.0, -4.8) \quad (6.0, 1.0, 3.2, -4.0)$$

问能否认为两个字样来自同一母体 ( $\alpha = 0.05$ ) ?

解:

(先检验两个子样的方差是否相同, 再证明两个子样的均值是否相同)

六. (本题满分 12 分)

下面的数据给出了三个地区人的血液中的胆固醇的含量

地区	测量值						
1	403	304	259	336	259	253	290
2	362	322	362	420	420	386	274
3	361	344	353	235	349	260	226

试用单因素方差分析法, 检验不同地区人的血液中胆固醇的平均量之间是否存在显著差别? ( $\alpha = 0.05$ )

解:

方差来源	平方和	自由度	均方和	F 值
因素 A (组间)		$r - 1$		
误差 E (组内)		$n - r$		
总和				

七. (本题满分 15 分)

在某乡镇, 随机地走访了十户居民加, 得其家庭月收入 ( $x$ ) 与日常开支 ( $y$ ) 的子样数据如下 (单位: 元)

$$\text{收入 } x: 820 \quad 930 \quad 1050 \quad 1300 \quad 1440 \quad 1500 \quad 1600 \quad 1800 \quad 2000 \quad 2700$$

$$\text{支出 } y: 750 \quad 850 \quad 920 \quad 1050 \quad 1200 \quad 1300 \quad 1300 \quad 1450 \quad 1560 \quad 2000$$

(1) 求日常开支  $y$  与家庭月收入  $x$  间的经验回归方程;

(2) 检验回归效果是否显著? ( $\alpha = 0.05$ )

(3) 对  $x_0 = 2200$  (元), 给出  $y$  的置信概率为 95% 的预测区间。

解:

(1)

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

(2) 检验假设为  $H_0: b = 0 \leftrightarrow H_1: b \neq 0$ , 检验统计量为

$$T = \frac{b}{\hat{\sigma}^* / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \sim t(n-2)$$

(3) 枢轴量

$$T = \frac{y_0 - \hat{y}_0}{\hat{\sigma}^* \cdot \sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}} \sim t(n-2)$$

八. (本题满分 6 分)

已知母体  $X$  为一个连续型随机变量,  $X$  的分布函数是  $F(x)$ , 设  $X_1, X_2, \dots, X_n$  是来自母体  $X$  的简单随机子样, 试证随机变量  $Y = -2 \sum_{i=1}^n \ln[F(X_i)]$  服从  $\chi^2(2n)$  分布。

# 西安交通大学研究生课程考试题（数理统计 2007）

附表：

标准正态分布的分布函数值： $\Phi(1.96) = 0.9750$

$t$  分布的上侧分位数：

$\alpha$ $n$	0.025	0.05
12	2.1788	1.7823
15	2.1314	1.7531
18	2.1009	1.7341

$\chi^2$  分布的上侧分位数：

$\alpha$ $n$	0.05	0.95
15	24.996	7.261

$F$  分布的上侧分位数： $F_{0.025}(9, 9) = 4.03$ ， $F_{0.05}(2, 12) = 3.89$ 。

一. 填空题（本题分值为 30）

- (1) 设  $X_1, \dots, X_n$  为 **i.i.d.**，其含义是\_\_\_\_\_。
- (2) 设  $U \sim N(0,1)$ ，若有  $P\{|U| < c\} = \alpha$  ( $0 < \alpha < 1$ )，则  $c =$ \_\_\_\_\_（用  $N(0,1)$  分布的上侧分位数符号表示）。
- (3) 设  $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$  为正态总体  $N(0, \sigma^2)$  的样本，若要

$$a \frac{\sum_{i=1}^n X_i^2}{\sum_{i=n+1}^{n+m} X_i^2} \sim F(b, c)$$

则  $a =$ \_\_\_\_\_， $b =$ \_\_\_\_\_， $c =$ \_\_\_\_\_。

- (4) 写出估计参数最常用的三种方法：  
\_\_\_\_\_，\_\_\_\_\_，\_\_\_\_\_。
- (5) 若参数假设问题  $H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1$  的拒绝域为  $W$ ，则该检验犯第 I 类错误的概率  $p_1 =$ \_\_\_\_\_，犯第 II 类错误的概率  $p_2 =$ \_\_\_\_\_。

二.（本题分值为 12）已知总体  $X$  的概率密度函数为

$$f(x; \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2} \exp\left\{-\frac{x-\theta_1}{\theta_2}\right\}, & x > \theta_1 \\ 0, & x < \theta_1 \end{cases}, \quad (-\infty < \theta_1 < +\infty, \theta_2 > 0)$$

设  $X_1, \dots, X_n$  是总体  $X$  的样本，求未知参数  $\theta_1, \theta_2$  的矩估计。

五.（本题分值为 12）

(1) 完成下列方差分析表中欠缺的项目：

方差来源	离差平方和	自由度	均方离差	$F$ 值
组间	2578.8		1289.4	
组内		12		
总和	6279.6			

(2) 问这是几个因素几种水平试验的方差分析表？

(3) 由上述方差分析表，检验各组均值是否有显著差异 ( $\alpha = 0.05$ )？

(4) 已知在因素的每一水平上进行等重复试验，且算得  $\bar{x}_1 = 87.2$ ， $\bar{x}_2 = 55.4$ ，求

$\mu_1 - \mu_2$  的 95% 置信区间

六. (本题分值为 6) 假设  $(x_i, y_i)$  满足线性回归关系：

$$y_i = a + bx_i + \varepsilon_i, \quad (i = 1, \dots, n)$$

其中  $\varepsilon_1, \dots, \varepsilon_n$  为 i.i.d. 且  $\varepsilon_i \sim N(0, \sigma^2)$ ， $x_1, \dots, x_n$  不全相同，试用极大似然法估计

参数  $a, b$ 。

七. (本题分值为 6) 设  $X_1, \dots, X_n$  是取自  $N(0, \sigma^2)$  的样本，其中  $\sigma > 0$  为未知参数。

(1) 问  $\sigma = \frac{1}{n} \sum_{i=1}^n |X_i|$  是否为  $\sigma$  的无偏估计？(若认为是  $\sigma$  的无偏估计，请给出证明；

若认为不是，对它作适当的修正，给出  $\sigma$  的无偏估计。)

(2) 针对 (1) 的讨论结果，求  $\sigma$  的无偏估计的 (有) 效率。

八. (本题分值为 5) 设  $X \sim N(\mu, 1)$ ，其中  $\mu$  为未知参数， $F(x)$  为  $X$  的分布函数。又

设常数  $c$  满足等式： $F(c) = 0.975$ 。先从总体  $X$  抽取一个样本，算得  $\bar{x} = 3.04$ ，求  $c$  的极大似然估计值。

九. (本题分值为 5) 设  $X_1, \dots, X_n$  为取自总体  $X$  的样本，已知总体  $X$  的分布函数  $F(x)$

为连续函数，证明  $F(X_{(1)}) \sim \beta(1, n)$ ，其中  $X_{(1)}$  是第一顺序统计量 (已知  $\beta(1, n)$  分

布的概率密度为  $f(x; 1, n) = \begin{cases} n(1-x)^{n-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$  )。

## 2009(上)《数理统计》考试题(A卷)及参考解答

### 一、填空题(每小题3分,共15分)

1, 设总体  $X$  和  $Y$  相互独立, 且都服从正态分布  $N(0, 3^2)$ , 而  $(X_1, X_2, \dots, X_9)$  和  $(Y_1, Y_2, \dots, Y_9)$  是分别来自  $X$  和  $Y$  的样本, 则  $U = \frac{X_1 + \dots + X_9}{\sqrt{Y_1^2 + \dots + Y_9^2}}$  服从的分布是\_\_\_\_\_.

解:  $t(9)$ .

2, 设  $\hat{\theta}_1$  与  $\hat{\theta}_2$  都是总体未知参数  $\theta$  的估计, 且  $\hat{\theta}_1$  比  $\hat{\theta}_2$  有效, 则  $\hat{\theta}_1$  与  $\hat{\theta}_2$  的期望与方差满足\_\_\_\_\_.

解:  $E(\hat{\theta}_1) = E(\hat{\theta}_2), D(\hat{\theta}_1) < D(\hat{\theta}_2)$ .

3, “两个总体相等性检验”的方法有\_\_\_\_\_与\_\_\_\_\_.

解: 秩和检验、游程总数检验.

4, 单因素试验方差分析的数学模型含有的三个基本假定是\_\_\_\_\_.

解: 正态性、方差齐性、独立性.

5, 多元线性回归模型  $Y = X\beta + \varepsilon$  中,  $\beta$  的最小二乘估计是  $\hat{\beta} =$ \_\_\_\_\_.

解:  $\hat{\beta} = (X'X)^{-1}X'Y$ .

### 二、单项选择题(每小题3分,共15分)

1, 设  $(X_1, X_2, \dots, X_n) (n \geq 2)$  为来自总体  $N(0, 1)$  的一个样本,  $\bar{X}$  为样本均值,  $S^2$  为样本方差, 则\_\_\_\_\_.

(A)  $n\bar{X} : N(0, 1)$ ;

(B)  $nS^2 : \chi^2(n)$ ;

(C)  $\frac{(n-1)\bar{X}}{S} : t(n)$ ;

(D)  $\frac{(n-1)X_1^2}{\sum_{i=2}^n X_i^2} : F(1, n-1)$ .

2, 若总体  $X : N(\mu, \sigma^2)$ , 其中  $\sigma^2$  已知, 当置信度  $1-\alpha$  保持不变时, 如果样本容量  $n$  增大, 则  $\mu$  的置信区间\_\_\_\_\_.

(A) 长度变大; (B) 长度变小; (C) 长度不变; (D) 前述都有可能.

3, 在假设检验中, 分别用  $\alpha$ ,  $\beta$  表示犯第一类错误和第二类错误的概率, 则当样本容量  $n$  一定时, 下列说法中正确的是\_\_\_\_\_.

(A)  $\alpha$  减小时  $\beta$  也减小;

(B)  $\alpha$  增大时  $\beta$  也增大;

- (C)  $\alpha, \beta$  其中一个减小, 另一个会增大; (D) (A) 和 (B) 同时成立.

4, 对于单因素试验方差分析的数学模型, 设  $S_T$  为总离差平方和,  $S_e$  为误差平方和,  $S_A$  为效应平方和, 则总有 A.

(A)  $S_T = S_e + S_A$ ;

(B)  $\frac{S_A}{\sigma^2} : \chi^2(r-1)$ ;

(C)  $\frac{S_A/(r-1)}{S_e/(n-r)} : F(r-1, n-r)$ ;

(D)  $S_A$  与  $S_e$  相互独立.

5, 在一元回归分析中, 判定系数定义为  $R^2 = \frac{S_{\text{回}}}{S_T}$ , 则 B.

(A)  $R^2$  接近 0 时回归效果显著;

(B)  $R^2$  接近 1 时回归效果显著;

(C)  $R^2$  接近  $\infty$  时回归效果显著;

(D) 前述都不对.

三、(本题 10 分) 设总体  $X : N(\mu_1, \sigma^2)$ 、 $Y : N(\mu_2, \sigma^2)$ ,  $(X_1, X_2, \dots, X_{n_1})$  和  $(Y_1, Y_2, \dots, Y_{n_2})$  分别是来自  $X$  和  $Y$  的样本, 且两个样本相互独立,  $\bar{X}$ 、 $\bar{Y}$  和  $S_X^2$ 、 $S_Y^2$  分别是它们的样本均值和样本方差, 证明

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : t(n_1 + n_2 - 2),$$

其中  $S_{\omega}^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}$ .

证明: 易知

$$\bar{X} - \bar{Y} : N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}), \quad U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : N(0, 1).$$

由定理可知

$$\frac{(n_1 - 1)S_X^2}{\sigma^2} : \chi^2(n_1 - 1), \quad \frac{(n_2 - 1)S_Y^2}{\sigma^2} : \chi^2(n_2 - 1).$$

由独立性和  $\chi^2$  分布的可加性可得

$$V = \frac{(n_1 - 1)S_X^2}{\sigma^2} + \frac{(n_2 - 1)S_Y^2}{\sigma^2} : \chi^2(n_1 + n_2 - 2).$$

由  $U$  与  $V$  得独立性和  $t$  分布的定义可得



$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{U}{\sqrt{V/(n_1 + n_2 - 2)}} : t(n_1 + n_2 - 2).$$

四、(本题 10 分) 已知总体  $X$  的概率密度函数为  $f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & \text{其它} \end{cases}$ , 其中未知参

数  $\theta > 0$ ,  $(X_1, X_2, \dots, X_n)$  为取自总体的一个样本, 求  $\theta$  的矩估计量, 并证明该估计量是无偏估计量.

解: (1)  $v_1 = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} \frac{1}{\theta} xe^{-\frac{x}{\theta}} dx = \theta$ , 用  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$  代替, 所以

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

(2)  $E(\hat{\theta}) = E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X) = \theta$ , 所以该估计量是无偏估计.

五、(本题 10 分) 设总体  $X$  的概率密度函数为  $f(x; \theta) = (1 + \theta)x^{\theta}, 0 < x < 1$ , 其中未知参数  $\theta > -1$ ,  $(X_1, X_2, \dots, X_n)$  是来自总体  $X$  的一个样本, 试求参数  $\theta$  的极大似然估计.

解:

$$L(\theta) = \begin{cases} (\theta + 1)^n (\prod_{i=1}^n x_i)^{\theta}, & 0 < x_i < 1 \\ 0, & \text{其它} \end{cases}$$

当  $0 < x_i < 1$  时,  $\ln L(\theta) = n \ln(\theta + 1) + \theta \sum_{i=1}^n \ln x_i$ , 令  $\frac{d \ln L(\theta)}{d\theta} = \frac{n}{\theta + 1} + \sum_{i=1}^n \ln x_i = 0$ , 得

$$\hat{\theta} = -1 - \frac{n}{\sum_{i=1}^n \ln x_i}.$$

六、(本题 10 分) 设总体  $X$  的密度函数为  $f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0, \end{cases}$  未知参数  $\lambda > 0$ ,

$(X_1, X_2, \dots, X_n)$  为总体的一个样本, 证明  $\bar{X}$  是  $\frac{1}{\lambda}$  的一个 UMVUE.

证明: 由指数分布的总体满足正则条件可得

$$I(\lambda) = -E\left[\frac{\partial^2}{\partial \lambda^2} \ln f(x; \lambda)\right] = -E\left(\frac{-1}{\lambda^2}\right) = \frac{1}{\lambda^2},$$

$\frac{1}{\lambda}$  的无偏估计方差的 C-R 下界为

$$\frac{[(\frac{1}{\lambda})']^2}{nI(\lambda)} = \frac{\left[\frac{-1}{\lambda^2}\right]^2}{n \frac{1}{\lambda^2}} = \frac{1}{n\lambda^2}.$$

另一方面

$$E(\bar{X}) = 1/\lambda, \quad \text{Var}(\bar{X}) = \frac{1}{n\lambda^2},$$

即  $\bar{X}$  得方差达到 C-R 下界, 故  $\bar{X}$  是  $\frac{1}{\lambda}$  的 UMVUE.

**七、(本题 10 分)** 合格苹果的重量标准差应小于 0.005 公斤. 在一批苹果中随机取 9 个苹果称重, 得其样本标准差为  $S = 0.007$  公斤, 试问: (1) 在显著性水平  $\alpha = 0.05$  下, 可否认为该批苹果重量标准差达到要求? (2) 如果调整显著性水平  $\alpha = 0.025$ , 结果会怎样?

参 考 数 据 :  $\chi_{0.025}^2(9) = 19.023$  ,  $\chi_{0.05}^2(9) = 16.919$  ,  $\chi_{0.025}^2(8) = 17.535$  ,

$\chi_{0.05}^2(8) = 15.507$  .

**解:** (1)  $H_0: \sigma^2 \leq 0.005$ ,  $\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(8)$ , 则应有:

$$P(\chi^2 > \chi_{0.05}^2(8)) = 0.005, \Rightarrow \chi_{0.05}^2(8) = 15.507,$$

具体计算得:  $\chi^2 = \frac{8 \times 0.007^2}{0.005^2} = 15.68 > 15.507$ , 所以拒绝假设  $H_0$ , 即认为苹果重量标准差指标未达到要求.

(2) 新设  $H_0: \sigma^2 \leq 0.005$ , 由  $\chi_{0.025}^2 = 17.535, \Rightarrow \chi^2 = \frac{8 \times 0.007^2}{0.005^2} = 15.68 < 17.535$ ,

则接受假设, 即可以认为苹果重量标准差指标达到要求.

**八、(本题 10 分)** 已知两个总体  $X$  与  $Y$  独立,  $X \sim (\mu_1, \sigma_1^2)$ ,  $Y \sim (\mu_2, \sigma_2^2)$ ,

$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  未知,  $(X_1, X_2, \dots, X_{n_1})$  和  $(Y_1, Y_2, \dots, Y_{n_2})$  分别是来自  $X$  和  $Y$  的样本, 求

$\frac{\sigma_1^2}{\sigma_2^2}$  的置信度为  $1-\alpha$  的置信区间.

**解:** 设  $S_X^2, S_Y^2$  分别表示总体  $X, Y$  的样本方差, 由抽样分布定理可知

$$\frac{(n_1-1)S_X^2}{\sigma_1^2} : \chi^2(n_1-1), \quad \frac{(n_2-1)S_Y^2}{\sigma_2^2} : \chi^2(n_2-1),$$

由  $F$  分布的定义可得

$$F = \frac{\frac{(n_1-1)S_X^2}{\sigma_1^2} / (n_1-1)}{\frac{(n_2-1)S_Y^2}{\sigma_2^2} / (n_2-1)} = \frac{S_X^2}{S_Y^2} \frac{\sigma_2^2}{\sigma_1^2} : F(n_1-1, n_2-1).$$

对于置信度  $1-\alpha$ ，查  $F$  分布表找  $F_{\alpha/2}(n_1-1, n_2-1)$  和  $F_{1-\alpha/2}(n_1-1, n_2-1)$  使得

$$P[F_{\alpha/2}(n_1-1, n_2-1) < F < F_{1-\alpha/2}(n_1-1, n_2-1)] = 1-\alpha,$$

即

$$P\left(\frac{S_X^2 / S_Y^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_X^2 / S_Y^2}{F_{\alpha/2}(n_1-1, n_2-1)}\right) = 1-\alpha,$$

所求  $\frac{\sigma_1^2}{\sigma_2^2}$  的置信度为  $1-\alpha$  的置信区间为  $\left(\frac{S_X^2 / S_Y^2}{F_{1-\alpha/2}(n_1-1, n_2-1)}, \frac{S_X^2 / S_Y^2}{F_{\alpha/2}(n_1-1, n_2-1)}\right)$ .

**九、(本题 10 分)** 试简要论述线性回归分析包括哪些内容或步骤.

**解:** 建立模型、参数估计、回归方程检验、回归系数检验、变量剔除、预测.

## 2009(上)《数理统计》考试题(B卷)及参考解答

### 一、填空题(每小题3分,共15分)

1, 设总体  $X$  服从正态分布  $N(0, 4)$ , 而  $(X_1, X_2, \dots, X_{15})$  是来自  $X$  的样本, 则

$$U = \frac{X_1^2 + \dots + X_{10}^2}{2(X_{11}^2 + \dots + X_{15}^2)}$$
 服从的分布是\_\_\_\_\_.

解:  $F(10, 5)$ .

2,  $\hat{\theta}_n$  是总体未知参数  $\theta$  的相合估计量的一个充分条件是\_\_\_\_\_.

解:  $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta, \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$ .

3, 分布拟合检验方法有\_\_\_\_\_与\_\_\_\_\_.

解:  $\chi^2$  检验、柯尔莫哥洛夫检验.

4, 方差分析的目的是\_\_\_\_\_.

解: 推断各因素对试验结果影响是否显著.

5, 多元线性回归模型  $Y = X\beta + \varepsilon$  中,  $\beta$  的最小二乘估计  $\hat{\beta}$  的协方差矩阵

$$\text{Cov}(\hat{\beta}) = \text{_____}.$$

解:  $\text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ .

### 二、单项选择题(每小题3分,共15分)

1, 设总体  $X \sim N(1, 9)$ ,  $(X_1, X_2, \dots, X_9)$  是  $X$  的样本, 则\_\_\_\_\_.

(A)  $\frac{\bar{X}-1}{3} \sim N(0, 1);$

(B)  $\frac{\bar{X}-1}{1} \sim N(0, 1);$

(C)  $\frac{\bar{X}-1}{9} \sim N(0, 1);$

(D)  $\frac{\bar{X}-1}{\sqrt{3}} \sim N(0, 1).$

2, 若总体  $X: N(\mu, \sigma^2)$ , 其中  $\sigma^2$  已知, 当样本容量  $n$  保持不变时, 如果置信度  $1-\alpha$  减小, 则  $\mu$  的置信区间\_\_\_\_\_.

(A) 长度变大; (B) 长度变小; (C) 长度不变; (D) 前述都有可能.

3, 在假设检验中, 就检验结果而言, 以下说法正确的是\_\_\_\_\_.

- (A) 拒绝和接受原假设的理由都是充分的;  
(B) 拒绝原假设的理由是充分的, 接受原假设的理由是不充分的;  
(C) 拒绝原假设的理由是不充分的, 接受原假设的理由是充分的;

(D) 拒绝和接受原假设的理由都是不充分的.

4, 对于单因素试验方差分析的数学模型, 设  $S_T$  为总离差平方和,  $S_e$  为误差平方和,  $S_A$  为效应平方和, 则总有 A.

(A)  $S_T = S_e + S_A$ ;

(B)  $\frac{S_A}{\sigma^2} : \chi^2(r-1)$ ;

(C)  $\frac{S_A/(r-1)}{S_e/(n-r)} : F(r-1, n-r)$ ;

(D)  $S_A$  与  $S_e$  相互独立.

5, 在多元线性回归分析中, 设  $\hat{\beta}$  是  $\beta$  的最小二乘估计,  $\hat{\varepsilon} = Y - X\hat{\beta}$  是残差向量, 则 B.

(A)  $E(\hat{\varepsilon}) = \mathbf{0}_n$ ;

(B)

$\text{Cov}(\hat{\varepsilon}) = \sigma^2 [I_n - X(X'X)^{-1}X']$ ;

(C)  $\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-p-1}$  是  $\sigma^2$  的无偏估计;

(D) (A)、(B)、(C) 都对.

三、(本题 10 分) 设总体  $X : N(\mu_1, \sigma^2)$ 、 $Y : N(\mu_2, \sigma^2)$ ,  $(X_1, X_2, \dots, X_{n_1})$  和  $(Y_1, Y_2, \dots, Y_{n_2})$  分别是来自  $X$  和  $Y$  的样本, 且两个样本相互独立,  $\bar{X}$ 、 $\bar{Y}$  和  $S_X^2$ 、 $S_Y^2$  分别是它们的样本均值和样本方差, 证明

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : t(n_1 + n_2 - 2),$$

其中  $S_\omega^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}$ .

证明: 易知

$$\bar{X} - \bar{Y} : N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}), \quad U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : N(0, 1).$$

由定理可知

$$\frac{(n_1 - 1)S_X^2}{\sigma^2} : \chi^2(n_1 - 1), \quad \frac{(n_2 - 1)S_Y^2}{\sigma^2} : \chi^2(n_2 - 1).$$

由独立性和  $\chi^2$  分布的可加性可得

$$V = \frac{(n_1 - 1)S_X^2}{\sigma^2} + \frac{(n_2 - 1)S_Y^2}{\sigma^2} : \chi^2(n_1 + n_2 - 2).$$

由  $U$  与  $V$  得独立性和  $t$  分布的定义可得

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{U}{\sqrt{V/(n_1 + n_2 - 2)}} : t(n_1 + n_2 - 2).$$

四、(本题 10 分) 设总体  $X$  的概率密度为  $f(x; \theta) = \begin{cases} \frac{1}{2\theta}, & 0 < x < \theta, \\ \frac{1}{2(1-\theta)}, & \theta \leq x < 1, \text{ 其中参} \\ 0, & \text{其他,} \end{cases}$

数  $\theta$  ( $0 < \theta < 1$ ) 未知,  $(X_1, X_2, \dots, X_n)$  是来自总体的一个样本,  $\bar{X}$  是样本均值, (1) 求参数  $\theta$  的矩估计量  $\hat{\theta}$ ; (2) 证明  $4\bar{X}^2$  不是  $\theta^2$  的无偏估计量.

解: (1)

$$E(X) = \int_{-\infty}^{+\infty} xf(x, \theta)dx = \int_0^{\theta} \frac{x}{2\theta} dx + \int_{\theta}^1 \frac{x}{2(1-\theta)} dx = \frac{1}{4} + \frac{\theta}{2},$$

令  $\bar{X} = E(X)$ , 代入上式得到  $\theta$  的矩估计量为  $\hat{\theta} = 2\bar{X} - \frac{1}{2}$ .

(2)

$$E(4\bar{X}^2) = 4E\bar{X}^2 = 4[D\bar{X} + (E\bar{X})^2] = 4\left[\frac{1}{n}DX + \left(\frac{1}{4} + \frac{1}{2}\theta\right)^2\right] = \frac{4}{n}DX + \frac{1}{4} + \theta + \theta,$$

因为  $D(X) \geq 0$ ,  $\theta > 0$ , 所以  $E(4\bar{X}^2) > \theta^2$ . 故  $4\bar{X}^2$  不是  $\theta^2$  的无偏估计量.

五、(本题 10 分) 设总体  $X$  服从  $[0, \theta]$  ( $\theta > 0$ ) 上的均匀分布,  $(X_1, X_2, \dots, X_n)$  是来自总体  $X$  的一个样本, 试求参数  $\theta$  的极大似然估计.

解:  $X$  的密度函数为

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta; \\ 0, & \text{其他,} \end{cases}$$

似然函数为

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_i < \theta, i=1, 2, \dots, n, \\ 0, & \text{其它} \end{cases}$$

显然  $\theta > 0$  时,  $L(\theta)$  是单调减函数, 而  $\theta \geq \max\{x_1, x_2, \dots, x_n\}$ , 所以

$\hat{\theta} = \max\{X_1, X_2, \dots, X_n\}$  是  $\theta$  的极大似然估计.

六、(本题 10 分) 设总体  $X$  服从  $B(1, p)$  分布,  $(X_1, X_2, \dots, X_n)$  为总体的样本, 证明  $\bar{X}$

是参数  $p$  的一个 UMVUE.

**证明：**  $X$  的分布律为

$$f(x; p) = p^x (1-p)^{1-x}, x=0,1.$$

容易验证  $f(x; p)$  满足正则条件，于是

$$I(p) = E \left[ \frac{\partial}{\partial p} \ln f(x; p) \right]^2 = \frac{1}{p(1-p)}.$$

另一方面

$$\text{Var}(\bar{X}) = \frac{1}{n} \text{Var}(X) = \frac{p(1-p)}{n} = \frac{1}{nI(p)},$$

即  $\bar{X}$  得方差达到 C-R 下界的无偏估计量，故  $\bar{X}$  是  $p$  的一个 UMVUE.

**七、(本题 10 分)** 某异常区的磁场强度服从正态分布  $N(\mu_0, \sigma^2)$ ，由以前的观测可知

$\mu_0 = 56$ . 现有一台新仪器，用它对该区进行磁测，抽测了 16 个点，得  $\bar{x} = 61, s^2 = 400$ ,

问此仪器测出的结果与以往相比是否有明显的差异 ( $\alpha = 0.05$ ). 附表如下:

t 分布表

n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$
14	1.3450	1.7613	2.1448
15	1.3406	1.7531	2.1315
16	1.3368	1.7459	2.1199

$\chi^2$  分布表

n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$
14	21.064	23.685	26.119
15	22.307	24.996	27.488
16	23.342	24.296	28.845

**解：** 设  $H_0: \mu = \mu_0 = 56$ . 构造检验统计量

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(15),$$

确定拒绝域的形式  $\left\{ |t| > t_{\frac{\alpha}{2}} \right\}$ . 由  $\alpha = 0.05$ , 定出临界值  $t_{\alpha/2} = t_{0.025} = 2.1315$ , 从而求出拒

绝域  $\left\{ |t| > 2.1315 \right\}$ .

$$\text{而 } n=16, \bar{x}=60, \text{ 从而 } |t| = \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| = \left| \frac{60 - 56}{20/\sqrt{16}} \right| = 0.8 < 2.1315, \text{ 接受假设 } H_0,$$

即认为此仪器测出的结果与以往相比无明显的差异.

**八、(本题 10 分)** 已知两个总体  $X$  与  $Y$  独立,  $X \sim (\mu_1, \sigma_1^2)$ ,  $Y \sim (\mu_2, \sigma_2^2)$ ,

$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  未知,  $(X_1, X_2, \dots, X_{n_1})$  和  $(Y_1, Y_2, \dots, Y_{n_2})$  分别是来自  $X$  和  $Y$  的样本, 求

$\frac{\sigma_1^2}{\sigma_2^2}$  的置信度为  $1-\alpha$  的置信区间.

**解:** 设  $S_1^2, S_2^2$  分别表示总体  $X, Y$  的样本方差, 由抽样分布定理知

$$P[F_{\alpha/2}(n_1-1, n_2-1) < F < F_{1-\alpha/2}(n_1-1, n_2-1)] = 1-\alpha,$$

则

$$P\left(\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}\right) = 1-\alpha,$$

所求  $\frac{\sigma_1^2}{\sigma_2^2}$  的置信度为  $1-\alpha$  的置信区间为  $\left(\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)}, \frac{S_1^2/S_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}\right)$ .

**九、(本题 10 分)** 试简要论述线性回归分析包括哪些内容或步骤.



## 2011-2012（下）研究生应用数理统计试题（A）

1 设  $X_1, X_2, \dots, X_n$  为正态总体  $X \sim N(\mu, \sigma^2)$  的样本，令  $d = \frac{1}{n} \sum_{i=1}^n |X_i - \mu|$ ，试证

$$E(d) = \sqrt{\frac{2}{\pi}} \sigma, \quad D(d) = \left(1 - \frac{2}{\pi}\right) \frac{\sigma^2}{n}. \quad (10 \text{ 分})$$

2 设总体  $X$  服从正态  $N(\mu, \sigma^2)$ ， $X_1, X_2, \dots, X_n$  为其样本， $\bar{X}$  与  $S^2$  分别为样本均值及方差。又

设  $X_{n+1}$  与  $X_1, X_2, \dots, X_n$  独立同分布，试求统计量  $Y = \frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}$  的分布。

（其中  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ）(10 分)

3 设总体  $X$  具有分布律

$X$	1	2	3
$p$	$\theta^2$	$2\theta(1-\theta)$	$(1-\theta)^2$

其中  $\theta (0 < \theta < 1)$  为未知参数，已知取得了样本值  $x_1 = 1, x_2 = 2, x_3 = 1$ ，求  $\theta$  的矩估计和最大似然估计。(10 分)

4 证明样本  $k$  阶原点矩  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$  是总体  $X$  的  $k$  阶原点矩  $\mu_k = E(X^k)$  的无偏估计量。

(10 分)

5 假定某商场某种商品的月销售量服从正态分布  $N(\mu, \sigma^2)$ ， $\mu, \sigma$  未知。为了决定商店对该商品的进货量，需对  $\mu$  作估计，为此，随机抽取若干月，其销售量分别为：64, 57, 49, 81, 76, 70, 59，求  $\mu$  的置信度为 0.95 的置信区间。(10 分)

6 一种元件，要求其使用寿命不得低于 1000（小时）。现在从一批这种元件中随机抽取 25 件，测得其寿命平均值为 950（小时）。已知该种元件寿命服从标准差  $\sigma = 100$ （小时）的正态分布，试在显著水平 0.05 下确定这批元件是否合格。(10 分)

7 某小学一年级共有三个班级，在一次数学考试中从三个班随机抽取 12, 15, 13 个学生的成绩。设学生成绩服从正态分布且方差相等，样本的方差分析表如下表 1 所示，问在显著性水平为 0.05 时，三个班的平均成绩有无显著差异？(10 分)

表 1 方差分析表

方差来源	平方和	自由度	均方差	F 值	显著性
因素 A	355.477				
误差	13429.498				
总和	13764.975				

8 某问题是一个四因素二水平试验，选用  $L_8(2^7)$  正交表，要考虑  $A \times B$ ，试验方案设计及试验结果见表 2。（15 分）

- (1) 各因素及交互作用的主次顺序（指标  $y$  越大越好）。
- (2) 试找最优工艺条件。
- (3) 在显著水平  $\alpha=0.05$  下，哪些因素的影响显著？

表 2

列号 试验号	A	B	A×B	C			D	数据 $y_i$
	1	2	3	4	5	6	7	
1	1	1	1	1	1	1	1	115
2	1	1	1	2	2	2	2	160
3	1	2	2	1	1	2	2	145
4	1	2	2	2	2	1	1	155
5	2	1	2	1	2	1	2	140
6	2	1	2	2	1	2	1	155
7	2	2	1	1	2	2	1	100
8	2	2	1	2	1	1	2	125
$I_j$	575	570	500	500	540	535	525	
$II_j$	520	525	595	595	555	560	570	
$R_j$	55	45	95	95	15	25	45	
$S_j$	378.1	253.1	1128.1	1128.1	28.1	78.1	253.1	

9 营业税税收总额  $y$  与社会商品零售总额  $x$  有关。为了利用社会商品零售总额预测税收总额，现收集了以下数据，见表 3。（15 分）

表 3

单位：亿元

序号	社会商业零售总额 $x$	营业税税收总额 $y$
1	142.08	3.93
2	177.30	5.96
3	204.68	7.85
4	242.88	9.82
5	316.24	12.50
6	341.99	15.55
7	332.69	15.79
8	389.29	16.39
9	453.40	18.45

- (1) 求营业税税收总额  $y$  与社会商品零售总额  $x$  的线性回归方程。
- (2) 在显著水平  $\alpha=0.05$  下检验回归方程的线性性。
- (3) 预测当社会商品零售总额  $x=300$  亿元时的营业税的平均税收总额。

附表：

## 2011-2012（下）研究生应用数理统计试题（A）

1 设  $X_1, X_2, \dots, X_n$  为正态总体  $X \sim N(\mu, \sigma^2)$  的样本，令  $d = \frac{1}{n} \sum_{i=1}^n |X_i - \mu|$ ，试证

$$E(d) = \sqrt{\frac{2}{\pi}} \sigma, \quad D(d) = \left(1 - \frac{2}{\pi}\right) \frac{\sigma^2}{n}. \quad (10 \text{ 分})$$

2 设总体  $X$  服从正态  $N(\mu, \sigma^2)$ ， $X_1, X_2, \dots, X_n$  为其样本， $\bar{X}$  与  $S^2$  分别为样本均值及方差。又

设  $X_{n+1}$  与  $X_1, X_2, \dots, X_n$  独立同分布，试求统计量  $Y = \frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}$  的分布。

（其中  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ）(10 分)

3 设总体  $X$  具有分布律

$X$	1	2	3
$p$	$\theta^2$	$2\theta(1-\theta)$	$(1-\theta)^2$

其中  $\theta (0 < \theta < 1)$  为未知参数，已知取得了样本值  $x_1 = 1, x_2 = 2, x_3 = 1$ ，求  $\theta$  的矩估计和最大似然估计。(10 分)

4 证明样本  $k$  阶原点矩  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$  是总体  $X$  的  $k$  阶原点矩  $\mu_k = E(X^k)$  的无偏估计量。

(10 分)

5 假定某商场某种商品的月销售量服从正态分布  $N(\mu, \sigma^2)$ ， $\mu, \sigma$  未知。为了决定商店对该商品的进货量，需对  $\mu$  作估计，为此，随机抽取若干月，其销售量分别为：64, 57, 49, 81, 76, 70, 59，求  $\mu$  的置信度为 0.95 的置信区间。(10 分)

6 一种元件，要求其使用寿命不得低于 1000（小时）。现在从一批这种元件中随机抽取 25 件，测得其寿命平均值为 950（小时）。已知该种元件寿命服从标准差  $\sigma = 100$ （小时）的正态分布，试在显著水平 0.05 下确定这批元件是否合格。(10 分)

7 某小学一年级共有三个班级，在一次数学考试中从三个班随机抽取 12, 15, 13 个学生的成绩。设学生成绩服从正态分布且方差相等，样本的方差分析表如下表 1 所示，问在显著性水平为 0.05 时，三个班的平均成绩有无显著差异？(10 分)

表 1 方差分析表

方差来源	平方和	自由度	均方差	F 值	显著性
因素 A	355.477				
误差	13429.498				
总和	13764.975				

8 某问题是一个四因素二水平试验，选用  $L_8(2^7)$  正交表，要考虑  $A \times B$ ，试验方案设计及试验结果见表 2。（15 分）

- (4) 各因素及交互作用的主次顺序（指标  $y$  越大越好）。  
 (5) 试找最优工艺条件。  
 (6) 在显著水平  $\alpha=0.05$  下，哪些因素的影响显著？

表 2

列号 试验号	A	B	A×B	C			D	数据 $y_i$
	1	2	3	4	5	6	7	
1	1	1	1	1	1	1	1	115
2	1	1	1	2	2	2	2	160
3	1	2	2	1	1	2	2	145
4	1	2	2	2	2	1	1	155
5	2	1	2	1	2	1	2	140
6	2	1	2	2	1	2	1	155
7	2	2	1	1	2	2	1	100
8	2	2	1	2	1	1	2	125
$I_j$	575	570	500	500	540	535	525	
$II_j$	520	525	595	595	555	560	570	
$R_j$	55	45	95	95	15	25	45	
$S_j$	378.1	253.1	1128.1	1128.1	28.1	78.1	253.1	

9 营业税税收总额  $y$  与社会商品零售总额  $x$  有关。为了利用社会商品零售总额预测税收总额，现收集了以下数据，见表 3。（15 分）

表 3

单位：亿元

序号	社会商业零售总额 $x$	营业税税收总额 $y$
1	142.08	3.93
2	177.30	5.96
3	204.68	7.85
4	242.88	9.82
5	316.24	12.50
6	341.99	15.55
7	332.69	15.79
8	389.29	16.39
9	453.40	18.45

- (1) 求营业税税收总额  $y$  与社会商品零售总额  $x$  的线性回归方程。  
 (2) 在显著水平  $\alpha=0.05$  下检验回归方程的线性性。  
 (3) 预测当社会商品零售总额  $x=300$  亿元时的营业税的平均税收总额。

附表：

# 西安交通大学研究生试卷

考试科目: 数 理 统 计

考试时间: 2008 年 1 月 8 日 时—— 时 考试方式: 闭卷

学 号: \_\_\_\_\_ 姓 名: \_\_\_\_\_ 成 绩 \_\_\_\_\_

## 一. 填空题(每空 2 分. 共 20 分)

1. 设总体  $X \sim N(\mu, \sigma^2)$ ,  $X_1, X_2, \dots, X_n$  是来自总体的简单样本,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \text{则统计量}$$

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \underline{\hspace{2cm}}, \quad \frac{\bar{X} - \mu}{S^* / \sqrt{n}} \sim \underline{\hspace{2cm}}, \quad \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \underline{\hspace{2cm}}$$

2. 设总体  $X \sim N(\mu, 1)$ ,  $X_1, X_2$  是来自总体的简单样本,  $\hat{a}_1 = \frac{2}{3}X_1 + \frac{1}{3}X_2$ ,

$\hat{a}_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2$ ,  $\hat{a}_3 = \frac{1}{2}X_1 + \frac{1}{2}X_2$  都是  $\mu$  的无偏估计量, 则最有效的是\_\_\_\_\_。

3. 设总体  $X \sim N(\mu, \sigma^2)$ , 其中  $\sigma^2$  已知, 为使总体均值  $\mu$  的置信度为  $1-\alpha$  的置信区间的长度不大于  $L$ , 则样本容量  $n$  至少应取\_\_\_\_\_。

4. 在一元方差分析中, 一次抽样后由  $n$  个子样值计算得  $F$  的数值, 对假设检验  $H_0: \mu_1 = \mu_2 = \dots = \mu_r = \mu$ , 按显著水平  $\alpha = 5\%$ , 对  $H_0$  的拒绝域是\_\_\_\_\_, 接受域是\_\_\_\_\_。

5. 对一元线性回归问题:  $\begin{cases} Y = \alpha + \beta X + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$ , 所谓线性关系的显著性检验,

注: 命题纸上一般不留答题位置。字、图清楚, 请勿超出边框, 以便复印。

是指检验假设  $H_0$ : \_\_\_\_\_, 若按显著水平  $\alpha$  拒绝了  $H_0$ , 就表示  
\_\_\_\_\_, 若接受  $H_0$ , 就表示\_\_\_\_\_。

## 二. 判断题(每题 2 分, 共 8 分)

1. 在对单个正态总体均值的假设检验中, 当总体方差已知时, 选用  $t$  检验法 ( )

1)假设  $H_0: \mu = \mu_0$  ( $\mu_0$  已知),  $H_1: \mu \neq \mu_0$ , 其中方差  $\sigma^2$  已知。

2)假设  $H_0: \mu = \mu_0$  ( $\mu_0$  已知),  $H_1: \mu \neq \mu_0$ , 其中方差  $\sigma^2$  未知。

六(15 分) 设母体  $X$  具有指数分布, 它的分布密度为

$$f(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{1}{\mu}x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

其中  $\mu > 0$ , 试问  $\bar{X}$  是不是  $\mu$  的优效估计(写出过程)。

七(14 分) 对于一元线性回归模型  $Y = \alpha + \beta X + \varepsilon$ , 已知  $n$  对试验值

$(x_i, y_i) (i = 1, 2, \dots, n)$ , 用最小二乘法给出  $\alpha$ 、 $\beta$  的估计值(写出过程)。

# 西安交通大学研究生课程考试题（数理统计 2007）

附表：

标准正态分布的分布函数值： $\Phi(1.96) = 0.9750$

$t$  分布的上侧分位数：

$\alpha \backslash n$	0.025	0.05
12	2.1788	1.7823
15	2.1314	1.7531
18	2.1009	1.7341

$\chi^2$  分布的上侧分位数：

$\alpha \backslash n$	0.05	0.95
15	24.996	7.261

$F$  分布的上侧分位数： $F_{0.025}(9, 9) = 4.03$ ， $F_{0.05}(2, 12) = 3.89$ 。

一. 填空题（本题分值为 30）

(1) 设  $X_1, \dots, X_n$  为 **i.i.d.**，其含义是\_\_\_\_\_。

(2) 设  $U \sim N(0,1)$ ，若有  $P\{|U| < c\} = \alpha$  ( $0 < \alpha < 1$ )，则  $c =$ \_\_\_\_\_（用  $N(0,1)$  分布的上侧分位数符号表示）。

(3) 设  $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$  为正态总体  $N(0, \sigma^2)$  的样本，若要

$$a \frac{\sum_{i=1}^n X_i^2}{\sum_{i=n+1}^{n+m} X_i^2} \sim F(b, c)$$

则  $a =$ \_\_\_\_\_,  $b =$ \_\_\_\_\_,  $c =$ \_\_\_\_\_。

(4) 写出估计参数最常用的三种方法：

\_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_。

(5) 若参数假设问题  $H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1$  的拒绝域为  $W$ ，则该检验犯第 I

类错误的概率  $p_1 =$ \_\_\_\_\_, 犯第 II 类错误的概率  $p_2 =$ \_\_\_\_\_。

二.（本题分值为 12）已知总体  $X$  的概率密度函数为

$$f(x; \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2} \exp\left\{-\frac{x-\theta_1}{\theta_2}\right\}, & x > \theta_1 \\ 0, & x < \theta_1 \end{cases}, \quad (-\infty < \theta_1 < +\infty, \theta_2 > 0)$$

设  $X_1, L, X_n$  是总体  $X$  的样本, 求未知参数  $\theta_1, \theta_2$  的矩估计。

五. (本题分值为 12)

(1) 完成下列方差分析表中欠缺的项目:

方差来源	离差平方和	自由度	均方离差	$F$ 值
组间	2578.8		1289.4	
组内		12		
总和	6279.6			

(2) 问这是几个因素几种水平试验的方差分析表? 、

(3) 由上述方差分析表, 检验各组均值是否有显著差异 ( $\alpha = 0.05$ ) ?

(4) 已知在因素的每一水平上进行等重复试验, 且算得  $\bar{x}_1 = 87.2$ ,  $\bar{x}_2 = 55.4$ , 求

$\mu_1 - \mu_2$  的 95% 置信区间

六. (本题分值为 6) 假设  $(x_i, y_i)$  满足线性回归关系:

$$y_i = a + bx_i + \varepsilon_i, \quad (i = 1, L, n)$$

其中  $\varepsilon_1, L, \varepsilon_n$  为 i.i.d. 且  $\varepsilon_i \sim N(0, \sigma^2)$ ,  $x_1, L, x_n$  不全相同, 试用极大似然法估计

参数  $a, b$ 。

七. (本题分值为 6) 设  $X_1, L, X_n$  是取自  $N(0, \sigma^2)$  的样本, 其中  $\sigma > 0$  为未知参数。

(1) 问  $\sigma = \frac{1}{n} \sum_{i=1}^n |X_i|$  是否为  $\sigma$  的无偏估计? (若认为是  $\sigma$  的无偏估计, 请给出证明;

若认为不是, 对它作适当的修正, 给出  $\sigma$  的无偏估计。)

(2) 针对 (1) 的讨论结果, 求  $\sigma$  的无偏估计的 (有) 效率。

八. (本题分值为 5) 设  $X \sim N(\mu, 1)$ , 其中  $\mu$  为未知参数,  $F(x)$  为  $X$  的分布函数。又

设常数  $c$  满足等式:  $F(c) = 0.975$ 。先从总体  $X$  抽取一个样本, 算得  $\bar{x} = 3.04$ ,

求  $c$  的极大似然估计值。

九. (本题分值为 5) 设  $X_1, L, X_n$  为取自总体  $X$  的样本, 已知总体  $X$  的分布函数  $F(x)$

为连续函数, 证明  $F(X_{(1)}) \sim \beta(1, n)$ , 其中  $X_{(1)}$  是第一顺序统计量 (已知  $\beta(1, n)$  分

布的概率密度为  $f(x; 1, n) = \begin{cases} n(1-x)^{n-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$  )。

试卷清晰度较差, 部分数据可能有误, 自己看着参考。若我看错了, 忘见谅!



这张试卷效果实在太差，很多内容看不太清，部分数据可能有误，但类型应该差不多，若我看错了，忘见谅！

## 西安交通大学研究生课程考试题（数理统计 2002）

一.（本题满分 14 分）

已知某零件的长度服从正态分布  $N(u, \sigma^2)$ ，其中  $\sigma^2 = 5.5 \text{ mm}^2$ ，从一大堆这种零件中

随机抽取  $n$  个，测量其长度。现用子样均值  $\bar{X}$  来估计母体均值  $u$ ，此时：

- (1) 若要估计量的标准差在  $1 \text{ mm}^2$  之下， $n$  应取多大？
- (2) 若要估计误差的绝对值超过  $1 \text{ mm}$  的概率在 1% 以下， $n$  应取多大？

二.（本题满分 20 分）

判断下列命题的真伪并简述理由：

1. “统计量”与“估计量”是同一概念。
2. “点估计”与“区间估计”的关系为：前者是后者的一种……………（瞅不清）
3. 设母体  $X$  的均值和方差都存在， $X_1, X_2, X_3$  为来自母体  $X$  的一个简单随机子样，则

$\theta_1 = \frac{1}{3}(X_1 + X_2 + X_3)$  与  $\theta_2 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$  都是  $E(X)$  的无偏估计，且  $\hat{\theta}_1$  比  $\hat{\theta}_2$  有效。

(4) 在一个确定的假设检验问题中，其判断结果不但与其检验水平  $\alpha$  有关，而且与抽到的子样有关。

四.（本题满分 14 分）

已知某种设备的工作温度服从正态分布，现作十次测量，得数据（ $^{\circ}\text{C}$ ）

1250 1275 1265 1245 1260 1255 1270 1265 1250 1240

- (1) 求温度的母体均值  $u$  的 95% 置信区间。
- (2) 求温度母体标准差  $\sigma$  的 95% 置信区间。

五.（本题满分 14 分）

设有两个独立的来自不同的正态母体的子样：

(-4.4, 4.0, 2.0, -4.8) (6.0, 1.0, 3.2, -4.0)

问能否认为两个字样来自同一母体（ $\alpha = 0.05$ ）？

六.（本题满分 12 分）

下面的数据给出了三个地区人的血液中的胆固醇的含量

地区	测量值						
1	403	304	259	336	259	253	290
2	362	322	362	420	420	386	274
3	361	344	353	235	349	260	226

试用单因素方差分析法，检验不同地区人的血液中胆固醇的平均量之间是否存在显著差别？（ $\alpha = 0.05$ ）

七. (本题满分 15 分)

在某乡镇, 随机地走访了十户居民, 得其家庭月收入 ( $x$ ) 与日常开支 ( $y$ ) 的子样数据如下 (单位: 元)

收入  $x$ : 820 930 1050 1300 1440 1500 1600 1800 2000 2700

支出  $y$ : 750 850 920 1050 1200 1300 1300 1450 1560 2000

- (1) 求日常开支  $y$  与家庭月收入  $x$  间的经验回归方程;
- (2) 检验回归效果是否显著? ( $\alpha = 0.05$ )
- (3) 对  $x_0 = 2200$  (元), 给出  $y$  的置信概率为 95% 的预测区间。

八. (本题满分 6 分)

已知母体  $X$  为一个连续型随机变量,  $X$  的分布函数是  $F(x)$ , 设  $X_1, X_2, \dots, X_n$  是来自

母体  $X$  的简单随机子样, 试证随机变量  $Y = -2 \sum_{i=1}^n \ln[F(X_i)]$  (看不清, 似乎是) 服从  $\chi^2(2n)$

分布。

一. (本题满分 20 分)

填空题:

1. 设  $X_1, X_2, \dots, X_{10}$  是来自正态母体  $N(\mu, \sigma^2)$  的一个简单随机子样, 其中  $\mu, \sigma^2$  已知. 填充下列统计量的分布及其相应参数:

A.  $\frac{X_2 + \mu}{\sigma} + \frac{X_9 - 2\mu}{2\sigma} \sim \underline{\hspace{2cm}} (\underline{\hspace{2cm}})$

B.  $\frac{\sum_{i=1}^{10} (X_i - \mu)^2}{\sigma^2} \sim \underline{\hspace{2cm}} (\underline{\hspace{2cm}})$

C.  $\frac{2 \sum_{i=1}^6 (X_i - \mu)^2}{3 \sum_{i=7}^{10} (X_i - \mu)^2} \sim \underline{\hspace{2cm}} (\underline{\hspace{2cm}})$

2. 设有一母体  $X$ , 其均值  $EX = \mu$ , 方差  $DX = \sigma^2$  以及四阶矩  $EX^4$  都存在,

$X_1, X_2, \dots, X_n$  是来自母体  $X$  的简单随机子样. 则  $\mu$  的无偏估计量为  $\underline{\hspace{2cm}}$ , 相

合估计量为  $\underline{\hspace{2cm}}$ ;  $\sigma^2$  的无偏估计量为  $\underline{\hspace{2cm}}$ , 相合估计量为  $\underline{\hspace{2cm}}$ .

二. (本题满分 20 分)

选择题 (从 A~E 中选择一个完整的答案, 填入指定处)

1. 设  $X \sim N(0, 1)$ , 则  $P\{X > \underline{\hspace{2cm}}\} = 1 - a$  ( $0 < a < 1$ ).

A.  $u_a$     B.  $-u_a$     C.  $u_{1-a}$     D. B 或 C    E. A~D 的答案皆错

2. 设  $F \sim F(m, n)$ , 则  $P\{F > \underline{\hspace{2cm}}\} = 1 - a$  ( $0 < a < 1$ ).

A.  $-F_a(m, n)$     B.  $F_{1-a}(m, n)$     C.  $F_a(m, n)$     D.  $F_a^{-1}(n, m)$     E. B 或 D

3. 设检验假设  $H: \theta = \theta_0$  的一个检验法则犯第一类错误的概率为  $P(I)$ , 检验的显著水平为  $\alpha$ , 则  $\underline{\hspace{2cm}}$ .

A.  $P(I) = 1 - \alpha$     B.  $P(I) = \alpha / 2$     C.  $P(I) = \alpha$     D.  $P(I) \geq 1 - \alpha$     E. C 或 D

4. 设  $X_1, X_2, \dots, X_n$  是来自正态母体  $N(\mu, \sigma^2)$  的子样, 其中  $\mu$  未知,  $\sigma^2$  已知, 则  $\underline{\hspace{2cm}}$  是统计量。

A.  $S^2 / \sigma^2$ ;    B.  $(X_1 - \mu) / \sigma$ ;    C.  $|X_1| + |X_2|$ ;    D. A 和 C;    E. A 和 B

5. 设母体  $X$  及  $Y$  的分布是任意的, 但分别是具有有限的非零方差, 记  $EX = \mu_1$ ,

$EY = \mu_2$ ，现独立地从两母体中各取一个子样，子样容量分别是  $n_1$  和  $n_2$ 。在大子样下，我们可以推出  $\mu_1 - \mu_2$  的置信概率近似为  $1 - \alpha$  的置信区间。这里所谓的大子样，一般是指\_\_\_\_\_。

A.  $n_1 \geq 50$ ; B.  $n_2 \geq 50$ ; C.  $n_1 + n_2 \geq 50$ ; D. A 且 B; E. A~D 的答案皆错

三. (本题满分 20 分)

设母体  $X$  的概率密度为

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (0 < \theta < +\infty)$$

1. 求  $\theta$  的矩估计量和最大似然估计量;
2. 用以上方法求得的估计量是否为  $\theta$  的无偏估计? 是否为  $\theta$  的相合估计?

四. (本题满分 14 分)

已知某种设备的工作温度服从正态分布，现对该温度作 10 次测量，得数据 ( $^{\circ}\text{C}$ )

1250 1275 1265 1245 1260 1255 1270 1265 1250 1240

1. 求温度的母体均值  $\mu$  的 95% 置信区间;
2. 求温度的母体标准差  $\sigma$  的 95% 置信区间。

五. (本题满分 14 分)

设有两个独立的来自不同正态母体的子样

$(-4.4, 4.0, 2.0, -4.8), (6.0, 1.0, 3.2, -4.0)$

问能否认为两个子样来自同一母体 ( $\alpha = 0.05$ ) ?

(提示: 首先检验两母体的方差是否相同, 其次检验两母体的均值是否相同)

六. (本题满分 6 分)

设母体  $X \sim N(\mu, 1)$ , 希望检验假设  $H_0: \mu = 6 \leftrightarrow H_1: \mu = 7$ 。若从该母体中取出容量为 4 的简单随机子样, 并采用如下检验法则:

当  $\bar{X} \geq 7$  时, 拒绝  $H_0$ , 接受  $H_1$ ; 当  $\bar{X} < 7$  时, 接受  $H_0$ , 拒绝  $H_1$ 。求上述

检验法则犯第一、二类错误的概率。

七. (本题满分 6 分)

设  $t_{\alpha}(n), F_{\alpha}(m, n)$  分别表示  $t(n), F(m, n)$  分布相应的上侧分位数,

求证:  $\left[ t_{\alpha/2}(n) \right]^2 = F_{\alpha}(1, n)$

(限时间、心情、眼力和水平所限, 可能有个别错误的地方, 忘海涵, 有错的地方可以指出来, 大家共同讨论一下)

## 西安交通大学考试题 数理统计 2000 年

### 一. 填空

1. 设  $X_1, X_2, \dots, X_{10}$  是来自正态总体  $N(0, 4)$  的样本,

$$c \cdot \frac{\sum_{i=1}^5 X_i}{\sqrt{\sum_{j=1}^5 (X_j - \frac{1}{5} \sum_{i=1}^5 X_i)^2 + \sum_{i=6}^{10} X_i^2}} \sim t(m)$$

则  $c =$  \_\_\_\_\_,  $m =$  \_\_\_\_\_

2. 用  $\Phi(x)$  表示标准正态分布  $N(0, 1)$  的分布函数, 则  $\Phi(-x)$  与  $\Phi(x)$  的关系为

$$\Phi(-x) = \underline{\hspace{2cm}}$$

3. 已知  $T \sim t(n)$ , 则  $T^2 \sim$  \_\_\_\_\_。

4. 设总体  $X$  的概率密度为  $f(x; \theta)$ , 则参数  $\theta$  估计的费歇 (Fisher) 信息量  $I(\theta) =$  \_\_\_\_\_。

### 二. 选择题 (填 A,B,C,D, 有几个正确填几个, 若都不正确, 则填 E)

1. 设  $(X_1, X_2, \dots, X_{20})$  是来自正态总体  $N(\mu, \sigma^2)$  的样本, 统计量

$$S = \sum_{i=1}^{10} (2X_{2i} - X_{2i-1})^2, \text{ 则 } \underline{\hspace{2cm}}$$

A.  $\frac{1}{5\sigma^2} S \sim \chi^2(9)$

B.  $\frac{1}{3\sigma^2} S \sim \chi^2(9)$

C.  $\frac{1}{5\sigma^2} S \sim \chi^2(10)$

D.  $\frac{1}{3\sigma^2} S \sim \chi^2(10)$

2. 独立地分别从两总体  $X$  和  $Y$  中抽得大小各为  $m$  和  $n$  的样本, 其样本均值分别为  $\bar{X}$  和  $\bar{Y}$ , 则  $D(\bar{X} - \bar{Y}) =$  \_\_\_\_\_。

A.  $\frac{D(X)}{m^2} - \frac{D(Y)}{n^2}$

B.  $\frac{D(X)}{m^2} + \frac{D(Y)}{n^2}$

C.  $\frac{D(X)}{m} - \frac{D(Y)}{n}$

D.  $\frac{D(X)}{m} + \frac{D(Y)}{n}$

3. 设  $X_1, X_2, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的样本, 其中  $\mu$  已知, 而  $\sigma^2$  未知, 则



七. 今有两台测量合金材料中某种金属含量的光谱仪, 为鉴定他们的测量准确性有无显著差异, 对 9 件含该金属分别为  $x_1, x_2, \dots, x_9$  (不等) 的合金材料进行测量, 第一台测量结果服从正态分布  $N(x_t + \delta_1, \sigma_1^2)$ ,  $t = 1, 2, \dots, 9$ ; 第二台测量结果服从  $N(x_t + \delta_2, \sigma_2^2)$ ,  $t = 1, 2, \dots, 9$ 。测得的 9 对观测值如下:

第一台	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
第二台	0.10	0.21	0.52	0.32	0.78	0.59	0.68	0.77	0.89

问能否认为第一台的测量值比第二台显著偏大 ( $\alpha = 0.05$ ) ?

(注: 此题甚不清晰, 数据可能有一两个有误,  $N(x_t + \delta_1, \sigma_1^2)$  和  $N(x_t + \delta_2, \sigma_2^2)$  也不是很清晰, 题意差不多, 知道方法就行。)

八. 设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本 ( $n > 2$ ), 总体的概率密度为

$$f(x; \lambda, a) = \begin{cases} \lambda e^{-\lambda(x-a)}, & \text{当 } x \geq a \\ 0, & \text{当 } x < a \end{cases} \quad (\text{参数 } \lambda > 0)$$

1. 设  $\lambda = \frac{1}{\sqrt{2}}$ , 试求参数  $a$  的最大似然估计。
2. 设  $a = 0$ , 试求参数  $\lambda$  的矩估计。
3. 设  $a = 0$ , 试推导  $2n\lambda\bar{X}$  服从的分布, 其中  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 。
4. 设  $a = 0$ , 试计算  $E(\hat{\lambda})$ , 并求  $k$ , 使  $\hat{\lambda}^* = k\hat{\lambda}$  为  $\lambda$  的无偏估计。(  $\chi^2(x)$  分布密度在写着在附录中, 但试卷上没有, 看书上的 )

(限时间、眼力和水平所限, 难免有些地方出错, 忘海涵, 谁发现有错的话可以指出来。)

## 2011-2012（下）研究生应用数理统计试题（A）

1 设  $X_1, X_2, \dots, X_n$  为正态总体  $X \sim N(\mu, \sigma^2)$  的样本，令  $d = \frac{1}{n} \sum_{i=1}^n |X_i - \mu|$ ，试证

$$E(d) = \sqrt{\frac{2}{\pi}} \sigma, \quad D(d) = \left(1 - \frac{2}{\pi}\right) \frac{\sigma^2}{n}. \quad (10 \text{ 分})$$

2 设总体  $X$  服从正态  $N(\mu, \sigma^2)$ ， $X_1, X_2, \dots, X_n$  为其样本， $\bar{X}$  与  $S^2$  分别为样本均值及方差。又

设  $X_{n+1}$  与  $X_1, X_2, \dots, X_n$  独立同分布，试求统计量  $Y = \frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}$  的分布。

解：由正态分布的特性可得

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad X_{n+1} \sim N(\mu, \sigma^2), \quad \frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1).$$

则有

$$X_{n+1} - \bar{X} \sim N\left(0, \frac{n+1}{n} \sigma^2\right),$$

从而

$$\frac{X_{n+1} - \bar{X}}{\sigma \sqrt{\frac{n+1}{n}}} \sim N(0, 1).$$

又  $\frac{X_{n+1} - \bar{X}}{\sigma \sqrt{\frac{n+1}{n}}}$  与  $\frac{(n-1)}{\sigma^2} S_n^2$  相互独立，从而

$$\frac{\frac{X_{n+1} - \bar{X}}{\sigma \sqrt{\frac{n+1}{n}}}}{\sqrt{\frac{(n-1)}{\sigma^2} S^2 / (n-1)}} = \sqrt{\frac{n}{n+1}} \frac{(X_{n+1} - \bar{X})}{S} \sim t(n-1).$$

（其中  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ）(10 分)

3 设总体  $X$  具有分布律

$X$	1	2	3
$p$	$\theta^2$	$2\theta(1-\theta)$	$(1-\theta)^2$

其中  $\theta (0 < \theta < 1)$  为未知参数，已知取得了样本值  $x_1 = 1, x_2 = 2, x_3 = 1$ ，求  $\theta$  的矩估计和最大似然估计。(10 分)

4 证明样本  $k$  阶原点矩  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$  是总体  $X$  的  $k$  阶原点矩  $\mu_k = E(X^k)$  的无偏估计量。

(10 分)

5 假定某商场某种商品的月销售量服从正态分布  $N(\mu, \sigma^2)$ ， $\mu, \sigma$  未知。为了决定商店对该商品的进货量，需对  $\mu$  作估计，为此，随机抽取若干月，其销售量分别为：64, 57, 49, 81，



76, 70, 59, 求  $\mu$  的置信度为 0.95 的置信区间。(10 分)

6 一种元件, 要求其使用寿命不得低于 1000 (小时)。现在从一批这种元件中随机抽取 25 件, 测得其寿命平均值为 950 (小时)。已知该种元件寿命服从标准差  $\sigma = 100$  (小时) 的正态分布, 试在显著水平 0.05 下确定这批元件是否合格。(10 分)

7 某小学一年级共有三个班级, 在一次数学考试中从三个班随机抽取 12,15,13 个学生的成绩。设学生成绩服从正态分布且方差相等, 样本的方差分析表如下表 1 所示, 问在显著性水平为 0.05 时, 三个班的平均成绩有无显著差异? (10 分)

表 1 方差分析表

方差来源	平方和	自由度	均方差	F 值	显著性
因素 A	355.477				
误差	13429.498				
总和	13764.975				

8 某问题是一个四因素二水平试验, 选用  $L_8(2^7)$  正交表, 要考虑  $A \times B$ , 试验方案设计及试验结果见表 2。(15 分)

- (1) 各因素及交互作用的主次顺序 (指标  $y$  越大越好)。  
(2) 试找最优工艺条件。  
(3) 在显著水平  $\alpha = 0.05$  下, 哪些因素的影响显著?

表 2

列号 试验号	A	B	A×B	C			D	数据 $y_i$
	1	2	3	4	5	6	7	
1	1	1	1	1	1	1	1	115
2	1	1	1	2	2	2	2	160
3	1	2	2	1	1	2	2	145
4	1	2	2	2	2	1	1	155
5	2	1	2	1	2	1	2	140
6	2	1	2	2	1	2	1	155
7	2	2	1	1	2	2	1	100
8	2	2	1	2	1	1	2	125
$I_j$	575	570	500	500	540	535	525	
$II_j$	520	525	595	595	555	560	570	
$R_j$	55	45	95	95	15	25	45	
$S_j$	378.1	253.1	1128.1	1128.1	28.1	78.1	253.1	

9 营业税税收总额  $y$  与社会商品零售总额  $x$  有关。为了利用社会商品零售总额预测税收总额,

现收集了以下数据，见表 3。（15 分）

表 3		单位：亿元
序号	社会商业零售总额 $x$	营业税税收总额 $y$
1	142.08	3.93
2	177.30	5.96
3	204.68	7.85
4	242.88	9.82
5	316.24	12.50
6	341.99	15.55
7	332.69	15.79
8	389.29	16.39
9	453.40	18.45

- （1）求营业税税收总额  $y$  与社会商品零售总额  $x$  的线性回归方程。
- （2）在显著水平  $\alpha = 0.05$  下检验回归方程的线性性。
- （3）预测当社会商品零售总额  $x = 300$  亿元时的营业税的平均税收总额。
- 附表：