学院
班级
学号
姓名

密 封 线

东北大学研究生考试试卷(A)

2010 — 2011 学年第 1 学期

课程名称：应用数理统计

总分	一	二	三	四	五	六	七	八

说明：1、共八题，尽量详细写出计算与证明过程，小数点后保留两位；

2、样本均值与样本方差分别定义为

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2;$$

3、分位点 Q_α 取为上分位点，即：$P(X>Q_\alpha)=\alpha$。

$\sigma^2=100$

一、（10分）假设 $X_1,\cdots,X_n$ 是来自总体 $N(5,100)$ 的一组简单随机样本，问样本容量 n 至少取多大才能保证 $P(X_{(n)}>5)\leq 0.99$？

解：$P(X_{(n)}>5)=1-P(X_{(n)}\leq 5)=1-P\{X_1<5,X_2<5\cdots X_n<5\}$

$1-P\{X_{(n)}<5\}$

∵ $X_1\cdots X_n$ 相互独立

∴ $P(X_{(n)}>5)=1-P(X_1\leq 5)$

且 $X\sim N(5,100)$

∴ $P\{X<5\}=P\{\frac{X-5}{10}<0\}=\Phi(0)=0.5$

∵ 要保证 $P(X_{(n)}>5)>0.99$

∴ $1-0.5^n>0.99$

∴ $0.5^n<0.01$

∴ $n\geq 7$

∴ n 至少为7时才能保证 $P(X_{(n)}>5)>0.99$

二、（10分）设总体 X 的分布如下：

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

现在随机抽取了10个样本，发现1出现2次，2出现4次，3出现4次；给出参数 θ 的矩估计。

解：$E(X)=\theta^2+4\theta(1-\theta)+3(1-\theta)^2=3-2\theta=\overline{X}$

∴ $\hat{\theta}=\frac{1}{2}(3-\overline{X})$

$\overline{X}=\frac{1}{10}(1\times 2+2\times 4+3\times 4)=2.2$

∴ $\hat{\theta}=0.4$

密 封 线

三、(15 分) 假设 $X_1,\dots,X_n$ 是来自总体 $N(\mu,\sigma^2)$ 的一组简单随机样本。

1、(10 分) 构造 2μ 的一个置信水平 $1-\alpha$ 的区间估计；

2、(5 分) 计算这个置信区间的平均长度（即区间长度的数学期望）。

解：1. 因为 $2\bar{X}$ 是 2μ 的最小方差无偏估计量，所以 $2\bar{X}$ 与 2μ 应非常接近，即 $|2\bar{X}-2\mu|$ 也较小，
设其为 C，那么 $|2\bar{X}-2\mu|<C$ 从而 $2\bar{X}-C<2\mu<2\bar{X}+C$

∴ 2μ 的置信区间为 $(2\bar{X}-C,\ 2\bar{X}+C)$

a. 当 σ^2 已知

∵ $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$ ∴ $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

∴ $1-\alpha=P\{2\bar{X}-C<2\mu<2\bar{X}+C\}=P\{\frac{2|\bar{X}-\mu|}{\sigma/\sqrt{n}}<\frac{C}{\sigma/\sqrt{n}}\}$

$=P\{-\frac{C}{2\sigma/\sqrt{n}}<\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\frac{C}{2\sigma/\sqrt{n}}\}=2P\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\frac{C}{2\sigma/\sqrt{n}}\}-1$

$=2(1-P\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\geq\frac{C}{2\sigma/\sqrt{n}}\})-1=1-2P\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\geq\frac{C}{2\sigma/\sqrt{n}}\}$

∴ $P\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\geq\frac{C}{2\sigma/\sqrt{n}}\}=\frac{\alpha}{2}$

∴ $C=\frac{2\sigma}{\sqrt{n}}\cdot u_{\frac{\alpha}{2}}$ 即 2μ 的置信水平 $1-\alpha$ 的置信区间为 $(2\bar{X}\pm\frac{2\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}})$

b. 当 σ^2 未知时 ?

∵ $\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ ~~且与是否已知无关~~

∴ $1-\alpha=P\{\frac{2|\bar{X}-\mu|}{S/\sqrt{n}}<\frac{C}{S/\sqrt{n}}\}=1-2P\{\frac{\bar{X}-\mu}{S/\sqrt{n}}\geq\frac{C}{2S/\sqrt{n}}\}$

∴ $P\{\frac{\bar{X}-\mu}{S/\sqrt{n}}\geq\frac{C}{2S/\sqrt{n}}\}=\frac{\alpha}{2}$

∴ $C=\frac{2S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)$ ∴ 2μ 的置信水平 $1-\alpha$ 的置信区间为 $(2\bar{X}\pm\frac{2S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1))$

2. 设统计量 Z 表示置信区间长度

a. σ^2 已知时   a. σ^2 已知时，区间长度为确定值

~~$Z=2\bar{X}+\frac{2\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}}$~~

b. σ^2 未知时，$Z=2\bar{X}+\frac{2S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)-(2\bar{X}-\frac{2S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1))=\frac{4S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)$

$E(Z)=\frac{4}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)E(S)$ ∵ $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ ∴ 统计量 $W=\frac{(n-1)S^2}{\sigma^2}$ 的概率密度 $f(x)=\begin{cases}\frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}x^{\frac{n-1}{2}-1}e^{-\frac{x}{2}}, & x>0\\ 0, & x\leq 0\end{cases}$

∴ $E[\frac{\sqrt{n-1}}{\sigma}\cdot S]=\int_0^{\infty}\sqrt{x}\cdot\frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}x^{\frac{n-1}{2}-1}e^{-\frac{x}{2}}dx=\int_0^{\infty}\frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx=\frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}\cdot(2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2}))$

$=\frac{\sqrt{2}\,\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$ ∴ $E(S)=\frac{\sigma}{\sqrt{n-1}}\cdot\sqrt{2}\cdot\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$ ∴ $E(Z)=\frac{4}{\sqrt{n}}\cdot\frac{\sigma}{\sqrt{n-1}}\cdot\sqrt{2}\,t_{\frac{\alpha}{2}}(n-1)\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$

四、(15 分) 设 $X_1,\dots,X_n$ 是来自几何分布：$P(X=k)=p(1-p)^{k-1}$，$k\geq 1$ 的一组简单随机样本，

1、(10 分) 求参数 p 的极大似然估计；

2、(5 分) 依据这个极大似然估计给出 $1/p$ 的一个无偏估计。

解：1. $L(p)=\prod_{i=1}^{n}p(1-p)^{x_i-1}=p^n\prod_{i=1}^{n}(1-p)^{x_i-1}$

$\ln L(p)=n\ln p+\ln(1-p)\sum_{i=1}^{n}(x_i-1)=n\ln p+(n\bar{x}-n)\ln(1-p)$

$\frac{\partial\ln L(p)}{\partial p}=\frac{n}{p}+\frac{n-n\bar{x}}{1-p}=0$ ∴ $\hat{p}=\frac{1}{\bar{x}}$

2. $E(\frac{1}{\hat{p}})=E(\bar{X})=E(X)=\sum_{k=1}^{n}kp(1-p)^{k-1}=\frac{1}{p}$ 几何分布的期望

∴ $\frac{1}{\hat{p}}$ 是 $\frac{1}{p}$ 的无偏估计

$p+2p(1-p)+3p(1-p)^2$

$p(1-p)+$

$P(|\bar{X}|>C)$

五. (15 分) $X_1, \dots, X_{10}$ 是来自总体 $B(1, p)$ 的 10 个简单随机样本。对于假设：

$$H_0: p = 0.3 \leftrightarrow H_1: p = 0.7 \quad (p \neq 0.3)$$

1、(10 分) 构造 H_0 的一个检验水平为 0.05 的否定域；

2、(5 分)，计算你的检验犯第二类错误的概率。

提示：$Y \sim B(10, 0.3)$ 的分布函数值 $p_k = P(Y \leq k)$ 如下：

k	0	1	2	3	4	5	6	7	8	9
p_k	0.0283	0.1493	0.3828	0.6496	0.8494	0.9527	0.9894	0.9984	0.9998	0.9999

解：1. 令 $Y = X_1 + \cdots + X_{10} = \sum_{i=1}^{10} X_i = 10\bar{X}$

∵ $X_1, \dots, X_{10} \sim B(1, p)$

∴ $Y \sim B(10, p)$

令 $P_0(\cdot) = P(\cdot \mid H_0真)$

由于 $\bar{X}$ 是 p 的无偏估计，所以通常 $\bar{X}$ 与 p 很接近，考虑到 H_1，所以 H_0 的否定域为 $\{\bar{X} > C\}$

∴ $\alpha = P_0(\bar{X} > C) = P_0(\frac{Y}{10} > C) = P_0(Y > 10C)$

∴ $P_0(Y \leq 10C) = 1 - \alpha = 1 - 0.05 = 0.95$

查表 $4 < 10C < 5$

∴ H_0 的水平为 0.05 的否定域为 $\{(x_1, \dots, x_n), \sum_{i=1}^{10} x_i \geq 5\}$.

2. $\beta = P\{接受 H_0 \mid H_1 为真\} = P_{p=0.7}(\sum_{i=1}^{10} X_i < 5) = C_{10}^0 0.7^0 0.3^{10} + C_{10}^1 0.7^1 0.3^9 + C_{10}^2 0.7^2 0.3^8 + C_{10}^3 0.7^3 0.3^7 + C_{10}^4 0.7^4 0.3^6$

$= 0.05$

六. (10 分) 把一个正四面体随机抛掷 40 次，每个面出现的次数分别是 8 次、12 次、7 次、13 次。问在检验水平 0.05 下能不能认为这个正四面体是均匀的？

不同概率的分布律，没有未知参数

解：用 $A_1A_2A_3A_4$ 分别表示正四面体的四个面，$V_1 \cdots V_4$ 表示了面出现的次数

假设：$H_0: p = 0.25$

$\frac{(V_i - np_i)^2}{np_i}$

A_i	V_i	np_i	V_i^2	V_i^2/np_i
A_1	8	10	64	6.4
A_2	12	10	144	14.4
A_3	7	10	49	4.9
A_4	13	10	169	16.9

$\sum 42.6$

$\chi^2_{0.05}(3) = 7.815 > \sum V_i^2/np_i - n = 2.6$　注意 怎样查表

∴ 接受 H_0，即认为四面体是均匀的。

易错

七.（10分）设用 3 台机器 A, B, C 制造同一种产品，假定各机器日产量服从正态分布且方差相等。对每台机器随机观测 5 天的日产量如下（单位：件）

A	41	48	41	57	49
B	65	57	54	72	64
C	45	51	56	48	48

问在显著性水平 $\alpha = 0.05$ 下各台机器的日产量是否有显著差异？

提示：三台机器数据的均值与样本方差分别为：$A: \bar{x}_1 = 47.2, s_1^2 = 44.2$

$B: \bar{x}_2 = 62.4, s_2^2 = 50.3$, $C: \bar{x}_3 = 49.6, s_3^2 = 17.3$;　　$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

全部 15 个数据的样本均值与样本方差为：$\bar{x} \doteq 53.07, s^2 = 79.64$.

解：设 A_1, A_2, A_3 分别表示机器 A.B.C 的日产量，则 $A_i \sim N(\beta_i, \sigma^2)$

假设 H_0: $\beta_1 = \beta_2 = \beta_3$

组间平方和 $CSS = \sum_{i=1}^{3} n_i(\bar{y}_i - \bar{y})^2 = 5\times(47.2-53.07)^2 + 5\times(62.4-53.07)^2 + 5(49.6-53.07)^2 = 5\times(34.4569 + 87.0489 + 12.0409) = 2218.29$

组内平方和 $RSS = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{3}(n_i - 1)S_i^2 = 4(44.2 + 50.3 + 17.3) = 447.2$　　$(\frac{1}{5}\times 5)\ 1=4$

构造统计量 $\frac{CSS/r-1}{RSS/n-r} \sim F(r-1, n-r)$

构造检验统计量 $F = \frac{(n-r)CSS}{(r-1)RSS} \sim F(r-1, n-r)$

因此，H_0 的拒绝域为 $\{F \geq F_\alpha(r-1, n-r)\}$

$F = \frac{(15-3)\times 2218.29}{(3-1)\times 447.2} = 29.76 > F_{0.05}(2, 12) = 3.89$

∴拒绝 H_0.

八.（15分）已知汽车产量 X（单位：万辆）与薄钢板的需求量 Y（单位：万吨）具有线性关系 $Y = \beta_0 + \beta_1 X + \varepsilon$，$\varepsilon \sim N(0, \sigma^2)$，现在有 5 组观测数据：

X	13.98	13.52	12.54	14.91	18.6
Y	191.8	199.37	217.19	302.62	303.99

1、(10分) 计算 β_0, β_1 的最小二乘估计；

2、(5分) 给出 $H_0: \beta_0 = 0 \leftrightarrow H_1: \beta_0 \neq 0$ 的检验水平为 0.05 的否定域。

提示：自变量 X 与因变量 Y 各自的样本均值与样本方差分别为

$\bar{X} = 14.71, s_X^2 = 5.46$; $\bar{Y} = 242.99, s_Y^2 = 3116.37$;

样本协方差为 $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 78.90$.　　看清重要信息

1. 解：$L_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 78.9 \times 5 = 394.5$　　不用算.

$L_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 4 \times 5.46 = 21.84$

$L_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 4 \times 3116.37 = 12465.48$

$\hat{\beta}_1 = \frac{L_{xy}}{L_{xx}} = 18.06$　　$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 242.99 - 18.06 \times 14.71 = -22.67$

2.

~~相关系数 $r^2 = \frac{L_{xy}^2}{L_{xx}L_{yy}} = \frac{394.5^2}{21.84 \times 12465.48} = 0.57$~~　　$\because \hat{\beta}_0 \sim N(\beta_0, \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}))$

$\frac{Q_e}{\sigma^2} \sim \chi^2(n-2)$

~~构造检验统计量 $F = \frac{(n-2)r^2}{1-r^2}$~~　　当 H_0 成立时 $\frac{\hat{\beta}_0}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}}} \sim N(0,1)$　　∴ $\frac{\hat{\beta}_0^2}{\sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}})} \sim \chi^2(1)$

H_0 — ∴统计量 $F = \frac{\chi^2/1}{Q_e/\sigma^2(n-2)} \sim F(1, n-2)$　　$\Rightarrow F = \frac{(\hat{\beta}_0/\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}})^2/1}{Q_e/\sigma^2(n-2)} \sim F(1, n-2)$

拒绝域为 $\left\{\frac{(n-2)\hat{\beta}_0^2 n(n-1)S_x^2}{Q_e\sum_{i=1}^{n}x_i^2} > F_{0.05}(1, n-2)\right\}$　　$= \frac{\hat{\beta}_0^2(n-2)}{(\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}})Q_e} \sim F(1, n-2)$

↓

$= \left\{\frac{3\times 22.67^2\times 5\times 4\times 5.46}{12465.48\times 0.43\times\sum x_i^2} > 10.13\right\}$

$C_{00} = \frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}$