华东理工大学 2010－2011 学年第一学期

研究生《数理统计》课程期末考试试卷 2011.01

开课学院： 理学院， 考试形式：闭卷， 所需时间 120 分钟

考生姓名：__________学号：____________ 学院___________ 任课教师 ___朱坤平______

题序	一	二	三	四	五	六	七	总 分
得分								
评卷人								

附表 $\chi^2_{0.025}(8)=2.180$, $\chi^2_{0.975}(8)=17.535$，$\chi^2_{0.95}(3)=7.815$，$\chi^2_{0.95}(1)=3.841$

$t_{0.975}(3)=3.1824$， $t_{0.975}(8)=2.306$， $t_{0.975}(4)=2.776$

$F_{0.95}(3,12)=3.49$ $F_{0.975}(2,9)=5.71$ $F_{0.95}(2,9)=4.26$

一． 选择题（每小题 4 分，共 36 分）

1．设总体的期望 μ 和方差 σ^2 均未知，从总体中抽取了一个容量为 n 的样本 $(X_1,X_2,\cdots,X_n)$，则下述选项中可以作为总体的期望 μ 和方差 σ^2 的无偏估计量的选项是（ A ）

(A) X_1 和 $\frac{1}{\mathrm{n}-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ (B) $\overline{X}$ 和 $\frac{1}{\mathrm{n}-1}\sum_{i=1}^{n}X_i^2-\overline{X}^2$

(C) $\overline{X}$ 和 $\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(X_i-\mu)^2$ (D) $\overline{X}$ 和 $\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(X_i-\overline{X})^2$

2．5 名评委对某歌手的打分分别是：63， 65， 70， 71， 95，根据打分，代表该歌手水平最合理的指标应是这些分值的（ B ）.

(A) 均值； (B) 中值； (C) 方差； (D) 众数

3．设总体期望为 μ，方差为 σ_0^2，$(X_1,X_2,\cdots,X_\mathrm{n})$ 为总体的一个容量为 n 的样本，$\overline{\mathrm{X}}$ 为样本均值，则（ D ）.

（A） 当 n 充分大时，$\overline{X}$ 近似服从正态分布 N(μ, σ_0^2)；

（B） 当 n 充分大时，$\overline{X}$ 的取值收敛于总体期望 μ；

（C） 因总体分布未知，无论 n 多大，$\overline{X}$ 都未必可视为服从正态分布；

(D) 当 n 充分大时，$\overline{X}$ 近似服从正态分布 $N(\mu,\ \sigma_0^2/n)$

4．设总体 $\xi \sim N(1,2^2)$，$(X_1, X_2, \cdots, X_{10})$ 是 ξ 的样本，

$Y = (X_1-1)^2 + (X_2-1)^2 + \ldots + (X_{10}-1)^2$，则下述选项正确的是（ C ）.

（A） $Y \sim \chi^2(10)$； （B） $Y \sim N(10,\ 40)$；

（C） $\dfrac{Y}{4} \sim \chi^2(10)$； （D） $\dfrac{Y}{2} \sim \chi^2(10)$

5. 不考虑交互作用的正交试验，若问题中有 4 个因子,每个因子都是 2 个水平，应选取的正交表是 (B).

（A） $L_4(2^3)$； （B） $L_8(2^7)$； （C） $L_9(3^4)$； （D） $L_{16}(2^{15})$

6. 设总体 $\xi \sim N(\mu, \sigma_0^2)$，其中 σ_0^2 已知，$(X_1, X_2, \cdots, X_n)$ 是 ξ 的样本，总体期望

μ 的置信水平为 $1-\alpha$ 的置信区间的长度记为 L，则**错误**的选项是 （ C ）。

（A） L 与样本容量 n 有关； （B） L 与置信水平 1-α 有关；

（C） L 与样本 $(X_1, X_2, \cdots, X_n)$ 的取值有关； （D） L 与总体方差 σ_0^2 有关.

7．显著性水平 α 下的某假设检验，原假设 H_0，则（ A ）.

（A） 犯第一类错误的概率一定不超过 α；
（B） 犯第二类错误的概率一定为 1-α；
（C） 犯第一类错误的概率一定为 α；
（D） 要么犯第一类错误，要么犯第二类错误,二者必居其一

8．多元线性回归模型 $Y = X\beta + e$，其中 $e \sim N(0, \sigma^2 I)$，关于 β 的最小二乘估计 $\hat{\beta}$，下述**错误**的选项是（ C ）。

（A） $\hat{\beta} = (X'X)^{-1}X'Y$ （B） $E(\hat{\beta}) = \beta$

（C） $\hat{\beta} \sim N(\beta,\ \sigma^2 I)$ （D） $\hat{\beta}$ 与残差平方和 SS_e 相互独立

9. 根据变元的 n 组观测值来求 m 元线性回归的复相关系数，下述选项正确的是 (A)

（A） $R = \sqrt{\dfrac{SS_R}{SS_T}}$ （B） $R = \sqrt{\dfrac{SS_e}{n-2}}$

（C） $R = \sqrt{\dfrac{SS_e}{n-m-1}}$ （D） $R = \sqrt{1 - \dfrac{SS_R}{SS_T}}$

二． (本题 10 分) 立邦牌油漆的干燥时间$\xi \sim N(\mu,\sigma^2)$，随机抽取 9 个样品，测得干燥时间（单位：小时）的样本均值为 6.2, **修正样本标准差**为 0.6928, 分别求μ,σ^2的置信水平为 95%的置信区间。

解: (1) μ的置信水平为 $1-\alpha$ 的置信区间为:

$$[\ \overline{X}-t_{1-\alpha/2}(n-1)\frac{S^*}{\sqrt{n}},\ \ \overline{X}+t_{1-\alpha/2}(n-1)\frac{S^*}{\sqrt{n}}\ ]$$

=[6.2-2.306*0.6928/3, 6.2+2.306*0.6928/3]=[5.6675, 6.7325]

(2) 由样本数据得到$n=9$，$s_{n-1}^2=0.48$，对于$\alpha=0.05$，自由度为 8，有

$\chi^2_{0.025}(8)=2.180,\quad \chi^2_{0.975}(8)=17.535$，所以

$$\frac{(n-1)S_{n-1}^2}{\chi^2_{0.975}(n-1)}=\frac{8\times0.48}{17.535}=0.2190;\qquad \frac{(n-1)S_{n-1}^2}{\chi^2_{0.025}(n-1)}=\frac{8\times0.48}{2.180}=1.7615$$

故σ^2的 95%的置信区间为[0.2190, 1.7615]

三. (本题 10 分) 设$(X_1,...,X_n)$是取自总体ξ的一个简单随机样本ξ的密度函数为

$$p(x)=\begin{cases} e^{-(x-\theta)}, & x\ge\theta \\ 0, & x<\theta \end{cases}$$ 其中$\theta>0$为未知参数，

（1） 求θ的矩法估计量$\hat{\theta}$, 并说明$\hat{\theta}$是否为θ的无偏估计?

（2） 求θ的极大似然估计.

解：(1)先计算$E\xi=\int_\theta^{+\infty}xe^{-(x-\theta)}dx=(-xe^{-(x-\theta)})\Big|_\theta^{+\infty}+\int_\theta^{+\infty}e^{-(x-\theta)}dx=\theta+1$

由于$E\xi=\overline{X}$，得到$\hat{\theta}=\overline{X}-1$

因 $\mathrm{E}\hat{\theta}=E(\overline{X}-1)=E\overline{X}-1=E\xi-1=(\theta+1)-1=\theta$， 故$\hat{\theta}=\overline{X}-1$是$\theta$无偏估计。

(2) 对于一组观测值$(x_1,x_2,\cdots x_n)$，设$x_1,\cdots,x_n\ge\theta$，此时似然函数

$$L(\theta)=\prod_{i=1}^n p(x_i)=\prod_{i=1}^n(e^{-(x_i-\theta)})$$

两边取对数，得对数似然函数 $\ln L(\theta)=-\sum_{i=1}^n x_i+n\vartheta$

分别关于θ求导，可得$\frac{d\ln L(\theta)}{d\theta}=n>0$ $\ln L(\theta)$关于θ严格单调递增，所以$\ln L(\theta)$的

极大值应在θ取值的右面的边界点上取到，故极大似然估计为$\hat{\theta}=\min\limits_{1\le i\le n} x_i$，

四.（本题 10 分）对某种合金材料的熔点作了四次测试，根据 4 次的测试数据算得样本均值为$\overline{X}=1267$(度)，修正样本标准差$S^*=3.65$ (度). 设合金材料的熔点服从正态分布，在显著性水平 $\alpha=5$ % 下：

(1) 能否认为该种合金的熔点符合厂家所公布的 1260 度？
(2) 能否认为该种合金熔点的标准差不超过 2 度？

解：由样本得 $\overline{X}=1267$，.

(1) 要检验的假设为 $H_0:\mu=1260$，$H_1:\mu\ne 1260$)

检验用的统计量 $T=\dfrac{\overline{X}-\mu_0}{S^*/\sqrt{n}}\sim t(n-1)$，

拒绝域为 $|T|\ge t_{1-\frac{\alpha}{2}}(n-1)=t_{0.975}(3)=3.1824$.

$|T|=\dfrac{1267-1260}{3.65/\sqrt{4}}=3.836>3.1824$，落在拒绝域内，

故拒绝原假设H_0，即不能认为结果符合公布的数字 1260^0C.

(2) 要检验的假设为 $H_0:\sigma\le 2$，$H_1:\sigma>2$

检验用的统计量 $\chi^2=\dfrac{(n-1)S^{*2}}{\sigma_0^2}\sim\chi^2(n-1)$，

拒绝域 $\chi^2>\chi^2_{1-\alpha}(n-1)=\chi^2_{0.95}(3)=7.815$

$\chi^2=40/4=10>7.815$,落在拒绝域内，故拒绝原假设H_0，即不能认为测定值的标准差不超过 2^0C.

五. (本题 10 分) 把一枚硬币连抛 100 次，结果出现了 40 次正面向上,60 次反面向上，在显著性水平显著性水平 $\alpha=5\%$ 下，能否认为这枚硬币是均匀的?

解：假设硬币是均匀的，令 X=0 表示反面向上,否则,X=1，即:

$$H_0:X\sim\begin{bmatrix}0 & 1\\ 0.5 & 0.5\end{bmatrix}$$

$$\chi^2=\frac{1}{n}\sum_{i=1}^{r}\frac{n_k^2}{p_k}-n\sim\chi^2(r-1);\quad \chi^2=\frac{1}{100}\left(\frac{60^2}{0.5}+\frac{40^2}{0.5}\right)-100=4$$

$\chi^2=4>\chi^2{}_{1-\alpha}(r-1)$=3.841，故拒绝原假设,认为该硬币不均匀.

六. (本题 14 分) 抽查 6 家企业, 根据产量 x_i (台) 与单位成本 y_i (万元)的统计数据得:

$$\sum x_i = 360, \quad \sum x_i^2 = 25000, \quad \sum y_i = 55, \quad \sum y_i^2 = 565, \quad \sum x_i y_i = 2860$$

（1）求单位成本与产量的相关系数;

（2）求单位成本关于产量的回归方程;

（3）求线性回归的残差平方和 SS_e 及估计的标准差 $\hat{\sigma}$;

(4) 在显著性水平 $\alpha = 0.05$ 下检验单位成本与产量是否有线性相关关系.

解：1） $\bar{x} = 60,$ $\bar{y} = 9.1667,$ $L_{xx} = \sum x_i^2 - n\bar{x}^2$=3400,

$L_{yy} = 60.8333$ $L_{xy} = \sum x_i y_i - n\bar{x}\ \bar{y}$= - 440,

$$r = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = -0.9675$$

2) $y = \beta_0 + \beta_1 x + \varepsilon,$ $\varepsilon \sim N(0, \sigma)$

$$\hat{\beta}_1 = \frac{L_{xy}}{L_{xx}} = -0.1294,$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 16.931$$

回归方程为 y = 16.931-0.1294x

3) $SS_e = L_{yy} - \hat{\beta}_1 L_{xy} = 3.8973$

$$\hat{\sigma} = \sqrt{\frac{SS_e}{n-2}} = 0.9871$$

4）$H_0 : \beta_1 = 0$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}} \sqrt{L_{xx}} \sim t(n-2)$$

$$T = \frac{-0.1294 - 0}{0.9871} \sqrt{3400} = -7.6439$$

$$|T| > t_{0.975}(4) = 2.7764$$

故拒绝原假设,即认为单位成本与产量有统计的线性相关关系.

七.（本题 10 分）为了研究一天中的不同工作时间对工作效率的影响,随机抽取 12 人,等分成三组,A 组做早班，B 组做晚班，C 组做夜班 ,分别记录他们完成同一种工作的完工时间,数据如下:

组别	完	工	时	间
A 早班	5.2	5.6	5.8	5.4
B 晚班	5.4	4.9	6.1	6.6
C 夜班	6.1	5.8	5.9	7.2

试利用方差分析的方法，在显著性水平 $\alpha = 0.05$ 下分析不同的班次对工作效率是否有显著性影响?

解：方差分析的前提是：假设不同班次的完工时间服从正态分布,且方差相等,即 $\xi_i \sim N(\mu_i, \sigma^2)$，i=1,2,3.

检验班次对工作效率是否有影响,相当于检验：$H_0 : \mu_1 = \mu_2 = \mu_3$

方差分析：单因素方差分析

SUMMARY

组	计数	求和	平均	方差
行 1	4	22	5.5	0.066667
行 2	4	23	5.75	0.563333
行 3	4	25	6.25	0.416667

方差分析

差异源	SS	df	MS	F	F crit
组间	1.166667	2	0.583333	1.671975	4.256492
组内	3.14	9	0.348889		
总计	4.306667	11			

F < F crit = 4.26，故 接受原假设，即在显著性水平 0.05 下认为不同的班次对工作效率无显著性影响.