厦门大学《概率统计I》课程试卷

_学院_系 ______年级___专业

主考教师： 试卷类型：（A卷）

本试卷中可能用到的分位数：

（1）关于正态分布：$z_{0.025}=1.96$，$z_{0.05}=1.65$，$\Phi(1)=0.8413$；

（2）关于t分布：$t_{0.025}(4)=2.776$；

（3）关于χ^2分布：$\chi^2_{0.05}(2)=5.991$；

（4）关于F分布：$F_{0.025}(58,40)=1.8079$，$F_{0.025}(40,58)=1.7530$。

1.

分数	阅卷人

(10分) 设$X_1,X_2,...,X_n$独立同分布，来自总体X，$Y_1,Y_2,...,Y_m$独立同分布，来自总体Y，已知这两个总体相互独立，并且$E(X)=\mu$, $D(X)=\sigma^2$, $E(Y)=\mu$, $D(Y)=2\sigma^2$, 求常数a使得$\hat{\mu}=a\bar{X}+(1-a)\bar{Y}$是最有效估计。

计算估计量的方差

$$D(\hat{\mu})=D(a\bar{X}+(1-a)\bar{Y})=a^2\frac{\sigma^2}{n}+(1-a)^2\frac{2\sigma^2}{m} \qquad (4分)$$

求导，

$$\frac{dD(\hat{\mu})}{da}=\frac{2a\sigma^2}{n}-\frac{4(1-a)\sigma^2}{m}=0 \qquad (2分)$$

解得

$$a=\frac{2n}{2n+m} \qquad (2分)$$

验证

$$\frac{d^2D(\hat{\mu})}{da^2}=\frac{2\sigma^2}{n}+\frac{4\sigma^2}{m}>0 \qquad (2分)$$

所以，当$a=\frac{2n}{2n+m}$时，$\hat{\mu}$是最有效估计

2.

分数	阅卷人

(10分) 设$Y_1, Y_2, ..., Y_n$独立同分布，有概率密度

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

其中$\lambda > 0$，计算参数θ, λ的矩估计。

$$E(X) = \int_{\theta}^{\infty} x\frac{1}{\lambda}e^{-(x-\theta)/\lambda}dx = \theta + \lambda \qquad (2分)$$

$$E(X^2) = \int_{\theta}^{\infty} x^2\frac{1}{\lambda}e^{-(x-\theta)/\lambda}dx = 2\lambda^2 + 2\theta\lambda + \theta^2 \qquad (2分)$$

$$D(X) = E(X^2) - [E(X)]^2 = \lambda^2$$

所以

$$\begin{cases} \theta + \lambda = \bar{X} \\ \lambda^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \end{cases} \qquad (4分)$$

得到估计量

$$\hat{\theta} = \bar{X} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \qquad (1分)$$

$$\hat{\lambda} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (1分)$$

3.

分数	阅卷人

(10分) 设$Y_1, Y_2, ..., Y_n$独立同分布，均服从对数正态分布，有概率密度

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\right\}, \ y > 0,$$

计算参数μ, σ^2的最大似然估计。

数据$Y_1, Y_2, ..., Y_n$的似然函数为

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma Y_i} \exp\left\{-\frac{(\ln Y_i - \mu)^2}{2\sigma^2}\right\} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \frac{1}{\prod_{i=1}^{n} Y_i} \exp\left\{-\sum_{i=1}^{n} \frac{(\ln Y_i - \mu)^2}{2\sigma^2}\right\}$$

$$\ln L(\mu, \sigma^2) = -n \ln\left(\sqrt{2\pi}\sigma\right) - \sum_{i=1}^{n} \ln Y_i - \sum_{i=1}^{n} \frac{(\ln Y_i - \mu)^2}{2\sigma^2} \qquad (4分)$$

对对数似然函数求偏导，

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \sum_{i=1}^{n} \frac{\ln Y_i - \mu}{\sigma^2} = 0 \qquad (2分)$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(\ln Y_i - \mu)^2}{2\sigma^4} = 0 \qquad (2分)$$

所以，极大似然估计

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln Y_i \qquad (1分)$$

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (\ln Y_i - \hat{\mu})^2 \qquad (1分)$$

4.

分数	阅卷人

(20分) 设0.5，1.25，0.8，2.0来自总体X，已知$Y = \ln X \sim N(\mu, 1)$，

（1）计算X的数学期望$b = EX$；

（2）求μ的置信度为95%的置信区间；

（3）求b的置信度为95%的置信区间。

(1)

$$b = E(X) = E(e^Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^y e^{-\frac{(y-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mu+t} e^{-\frac{t^2}{2}}$$
$$= e^{\mu+\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-1)^2}{2}} = e^{\mu+\frac{1}{2}} \qquad (6分)$$

(2) 由于$Y \sim N(\mu, 1)$，μ的置信度为95%的置信区间为

$$\left(\bar{Y} - z_{0.025}\frac{1}{\sqrt{n}}, \bar{Y} + z_{0.025}\frac{1}{\sqrt{n}}\right) \qquad (5分)$$

代入数据，$z_{0.025} = 1.96$，$n = 4$，$\bar{y} = \frac{1}{4}(\ln 0.5 + \ln 1.25 + \ln 0.8 + \ln 2.0) = 0$，

μ的置信度为95%的置信区间为

$$(-0.98, 0.98) \qquad (2分)$$

(3)由于$b = e^{\mu+\frac{1}{2}}$，并且

$$P\left\{\bar{Y} - z_{0.025}\frac{1}{\sqrt{n}} \le \mu \le \bar{Y} + z_{0.025}\frac{1}{\sqrt{n}}\right\} = 0.95 \qquad (2分)$$

所以

$$P\left\{\exp\left(\bar{Y} - z_{0.025}\frac{1}{\sqrt{n}} + \frac{1}{2}\right) \le \exp\left(\mu + \frac{1}{2}\right) \le \exp\left(\bar{Y} + z_{0.025}\frac{1}{\sqrt{n}} + \frac{1}{2}\right)\right\} = 0.95 \qquad (3分)$$

代入数据，b的置信度为95%的置信区间为

$$(e^{-0.48}, e^{1.48}) \qquad (2分)$$

5.

分数	阅卷人

(10分) X, Y两个渔场在初春放养相同的鱼苗，但是采用不同的喂养方式。入冬时，从X渔场打捞出59 条鱼，从Y渔场打捞出41条鱼，分别计算出他们的平均重量和样本标准差

$$\overline{X}_n = 0.59,\ S_1 = 0.2,\ \overline{Y}_m = 0.62,\ S_2 = 0.21$$

在正态假设和显著性水平$\alpha = 0.05$下，能否认为两个渔场的鱼苗重量方差相等？

考虑假设检验问题

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \qquad (2分)$$

检验统计量为

$$F = \frac{S_1^2}{S_2^2} \qquad (3分)$$

假设检验的拒绝域为

$$\left\{S_1^2/S_2^2 > F_{0.025}(58,40)\text{或者}S_1^2/S_2^2 < F_{0.975}(58,40)\right\} \qquad (3分)$$

带入数据$S_1^2/S_2^2 = 0.2^2/0.21^2 = 0.9070$，$F_{0.025}(58,40) = 1.8079$，

$$F_{0.975}(58,40) = 1/F_{0.025}(40,58) = 1/1.7530 = 0.5704 \qquad (1分)$$

所以不拒绝原假设

(1分)

6.

分数	阅卷人

(10分) 一本书共有300页。在该书的第一稿中，每页的打印错误数相互独立，都服从参数为6的泊松分布；在第二稿中，每个打印错误相互独立地以概率0.8被订正；在第三稿中，第二稿的打印错误被相互独立地以概率0.9被订正。三稿完成后交付印刷，估计这本书的打印错误大于等于30个的概率。

用X表示该书第一稿每页的打印错误；则$X \sim \pi(6)$

用Y表示该书第二稿每页的打印错误，

$$\begin{aligned} P\{Y=k\} &= \sum_{s=k}^{\infty} P\{X=s\} \begin{pmatrix} s \\ k \end{pmatrix} 0.2^k 0.8^{s-k} \\ &= \sum_{s=k}^{\infty} \frac{6^s e^{-6}}{s!} \frac{s!}{k!(s-k)!} 0.2^k 0.8^{s-k} \\ &= \frac{e^{-6} 1.2^k}{k!} \sum_{s=k}^{\infty} \frac{4.8^{s-k}}{(s-k)!} = \frac{e^{-1.2} 1.2^k}{k!} \end{aligned} \qquad (3分)$$

所以，则$Y \sim \pi(1.2)$。

同理用Z表示该书第三稿每页的打印错误，则

$$Z \sim \pi(0.12) \qquad (2分)$$

假设300页书中每页错误数为$Z_1, Z_2, \dots Z_{300}$，$Z_i \sim \pi(0.12)$，并且

$$EZ_i = DZ_i = 0.12 \qquad (2分)$$

本书打印错误数为$\sum_{i=1}^{300} Z_i$，根据中心极限定理，

$$P\left\{\sum_{i=1}^{300} Z_i \geq 30\right\} = P\left\{\frac{\sum_{i=1}^{300} Z_i - 300*0.12}{\sqrt{300*0.12}} \geq \frac{30-300*0.12}{\sqrt{300*0.12}}\right\} \approx 1-\Phi(-1) = 0.8413 \qquad (3分)$$

7.

分数	阅卷人

(15分) 某公司的工会对职工参加体育活动的情况进行了抽样调查，情况如下：

	A每天锻炼	B每周锻炼	C很少锻炼
人数	9	16	28
平均体重/kg	71	74	73.2

如果已知这三类人的体重方差都是32，在显著性水平0.05之下，

（1）能否认为A类人的体重显著小于C类人的体重？

（2）根据方差分析构造关于这三类人的体重有无显著的差异的检验统计量，以检验这三类人的体重有无显著的差异。

(1)考虑假设检验问题

$$H_0 : \mu_1 \geq \mu_3, \quad H_1 : \mu_1 < \mu_3 \qquad (2分)$$

检验统计量为

$$z_{13} = \frac{\bar{X}_A - \bar{X}_C}{\sqrt{\frac{1}{n_1} + \frac{1}{n_3}}\sigma} \qquad (3分)$$

假设检验的拒绝域为

$$\left\{ \frac{\bar{X}_A - \bar{X}_C}{\sqrt{\frac{1}{n_1} + \frac{1}{n_3}}\sigma} < -z_{0.05} \right\} \qquad (3分)$$

带入数据$z_{13} = \frac{71-73.2}{\sqrt{\frac{1}{9}+\frac{1}{28}}\sqrt{32}} = -1.015$，$z_{0.05} = 1.65$，

所以不拒绝原假设，不能认为A类人的体重显著小于C类人的体重

(2分)

(2) 考虑假设检验问题

$$H_0 : \mu_1 = \mu_2 = \mu_3, \quad H_1 : \mu_1, \mu_2, \mu_3\text{不全相等} \qquad (1分)$$

记$\bar{X}$为所有职工体重的平均，则

$$\bar{X} = \frac{1}{9+16+28}(9 * \bar{X}_A + 16 * \bar{X}_B + 28 * \bar{X}_C) = 73.07$$

定义组间平方和

$$S = 9 * (\bar{X}_A - \bar{X})^2 + 16 * (\bar{X}_B - \bar{X})^2 + 28 * (\bar{X}_C - \bar{X})^2 \qquad (2分)$$

在原假设之下，$\frac{S}{\sigma^2} \sim \chi^2(2)$，所以假设检验问题的拒绝域为

$$\left\{\frac{S}{32} > \chi^2_{0.05}(2)\right\} \qquad (1分)$$

带入数据$S = 52.8757$，$\frac{S}{32} = 1.65 < \chi^2_{0.05}(2) = 5.991$，

所以不拒绝原假设，这三类人的体重无显著的差异

(1分)

8.

分数	阅卷人

(15分) 很多人关系比萨斜塔的倾斜程度，下面是1975-1986年比萨斜塔的部分测量记录，其中的倾斜值指测量时塔尖的位置与原始位置的距离。简化数据，表中只给出倾斜值小数点后面第2至第4位的值。对于以下原始数据

年份x	75	77	80	82	84	86
倾斜值y	642	...	...	...	...	...

为了方便计算，给出对于原始数据的简单计算结果：

$S_{xx}=\sum_{i=1}^{n}X_i^2-n\bar{X}^2=87.3333$, $S_{yy}=\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2=6892$, $S_{xy}=\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}=764$,

$\bar{X}=80.6667$，$\bar{Y}=689$，

（1）计算年份x和倾斜值y的样本相关系数；

（2）建立回归直线方程；

（3）给定置信水平0.95，计算1981年倾斜值的预测区间。

(1)计算

$$\hat{\sigma}_X^2=\frac{S_{xx}}{n-1}=17.4667,\quad \hat{\sigma}_Y^2=\frac{S_{yy}}{n-1}=1378.4,\quad \widehat{Cov}(X,Y)=\frac{S_{xy}}{n-1}=152.8$$

$$\hat{r}=\frac{152.8}{\sqrt{17.4667}\sqrt{1378.4}}=0.9848 \qquad (5分)$$

(2)根据最小二乘估计

$$\hat{b}=\frac{S_{xy}}{S_{xx}}=\frac{764}{87.3333}=8.7481,$$

$$\hat{a}=\bar{Y}-\hat{b}\bar{X}=689-8.7481*80.6667=-16.6804 \qquad (5分)$$

(3) 给定$x_0=81$，置信水平0.95的预测区间为

$$(\hat{a}+\hat{b}x_0-t_{0.025}(n-2)\sigma_0, \hat{a}+\hat{b}x_0+t_{0.025}(n-2)\sigma_0) \qquad (3分)$$

其中$\sigma_0=\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}}$，代入数据，

$$\hat{\sigma}^2=\frac{S_{yy}-\hat{b}^2S_{xx}}{n-2}=52.1119 \qquad (1分)$$

$$\sigma_0=\sqrt{52.1119}*\sqrt{1+\frac{1}{6}+\frac{(81-80.6667)^2}{S_{xx}}}=7.8015$$

$t_{0.025}(4)=2.776$，所以，预测区间为

$$(670.2587,\ 713.5727) \qquad (1分)$$