厦门大学《概率统计 I》试卷

_________学院_________系_________年级_________专业

主考教师：_ _ _ _ _试卷类型：(A 卷)

解题过程中可能需要用到如下数据：

$\Phi(1.65)=0.9500$，$\Phi(1.69)=0.9545$, $\Phi(2.75)=0.9970$, $\Phi(1.96)=0.9750$，

$\chi^2_{0.05}(4)=9.488$，$\chi^2_{0.05}(5)=11.070$，$t_{0.025}(7)=2.365$，$t_{0.05}(7)=1.895$，$t_{0.025}(8)=2.306$

$t_{0.025}(9)=2.262$，$t_{0.025}(16)=2.120$，$t_{0.05}(16)=1.746$，$t_{0.025}(17)=2.110$，$t_{0.05}(17)=1.740$

$F_{0.025}(3,4)=9.98$, $F_{0.025}(4,3)=15.10$. $F_{0.05}(9,7)=3.68$. $F_{0.05}(2,12)=3.89$，$F_{0.05}(2,13)=3.81$

$F_{0.025}(8,8)=4.43$，$F_{0.025}(9,9)=4.03$

分数	阅卷人

1、（12 分）设随机变量 X 和 Y 的联合概率密度为

$$f(x,y)=\begin{cases}8xy, & 0\le x\le y\le 1,\\ 0, & 其他,\end{cases}$$

求 $Cov(X,Y)$ 和 $D(X+Y)$.

解法 1：由于

$$f_X(x)=\begin{cases}4x(1-x^2), & 0\le x\le 1,\\ 0, & 其他,\end{cases}$$

$$f_Y(x)=\begin{cases}4y^3, & 0\le y\le 1,\\ 0, & 其他,\end{cases}$$

于是

$$EX=\int_0^1 x\cdot 4x(1-x^2)dx=\frac{8}{15},$$

$$EY=\int_0^1 y\cdot 4y^3dy=\frac{4}{5},\ E(XY)=\frac{4}{9}$$

从而 $Cov(X,Y)=E(XY)-EXEY=\dfrac{4}{225}$.

$$EX^2=\int_0^1 x^2\cdot 4x(1-x^2)dx=\frac{1}{3}$$

$$EY^2=\int_0^1 y^2\cdot 4y^3dy=\frac{2}{3},$$

从而 $DX=EX^2-(EX)^2=\dfrac{11}{225}$，同理 $DY=\dfrac{2}{75}$

故 $D(X+Y)=DX+DY+2Cov(X,Y)=\dfrac{1}{9}$.

解法 2：

$$EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(\mathrm{x},\mathrm{y})dxdy=\int_0^1\int_0^y 8x^2ydxdy=\frac{8}{3}\int_0^1 y^4dy=\frac{8}{15}$$

$$EY=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(\mathrm{x},\mathrm{y})dxdy=\int_0^1\int_0^y 8xy^2dxdy=4\int_0^1 y^4dy=\frac{4}{5}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(\mathrm{x},\mathrm{y})dxdy=\int_0^1\int_0^y 8x^2y^2dxdy=\frac{8}{3}\int_0^1 y^5dy=\frac{4}{9}$$

$$\mathrm{cov}(X,Y)=E(XY)-EX\cdot EY=\frac{4}{225}$$

$$EX^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2f(\mathrm{x},\mathrm{y})dxdy=\int_0^1\int_0^y 8x^3ydxdy=2\int_0^1 y^5dy=\frac{1}{3}$$

$$EY^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y^2f(\mathrm{x},\mathrm{y})dxdy=\int_0^1\int_0^y 8xy^3dxdy=4\int_0^1 y^5dy=\frac{2}{3}$$

$$DX=EX^2-(EX)^2=\frac{11}{225}$$

$$DY=EY^2-(EY)^2=\frac{2}{75}$$

$$D(X+Y)=DX+DY+2\,\mathrm{cov}(X,Y)=\frac{1}{9}$$

分数	阅卷人

2、(10分) 某工厂每月生产10000台液晶投影机，它的液晶片车间生产液晶片合格率为80%，为了以99.7%的可能性保证出厂的液晶投影机都能装上合格的液晶片，试问该液晶片车间每月至少应该生产多少片液晶片？（利用中心极限定理 ）

解：设每月应该生产n片液晶片，其中合格液晶片有X片，X服从二项分布$b\left(n, 0.8\right)$,因

$EX = np = 0.8n,\ \ Var\left(X\right) = np\left(1 - p\right) = 0.16n,$ 且n应大于10000，n很大，根据中心

极限定理知$\dfrac{X - 0.8n}{0.4\sqrt{n}} \sim N(0,1)$,

因$P\{X \geq 10000\} \approx 1 - \Phi(\dfrac{10000 - 0.8n}{0.4\sqrt{n}}) = \Phi(\dfrac{0.8n - 10000}{0.4\sqrt{n}}) \geq 0.997,$

则$\dfrac{0.8n - 10000}{0.4\sqrt{n}} \geq 2.75,$ 即$0.8n - 1.1\sqrt{n} - 10000 \geq 0,$ 解得$\sqrt{n} \geq 112.4930,$ 故

$n \geq 12654.67,$ 即n至少为12655.

分数	阅卷人

3、（10 分）**设两个相互独立的总体** $X \sim N(50,36)$, $Y \sim N(46,16)$, **从这两个总体分别抽取容量为 10 和 8 的样本．又记它们的样本均值分别为** $\bar{X}$, $\bar{Y}$, **样本方差分别为** S_1^2, S_2^2, **求下列概率：**

(1) $P(0 < \bar{X} - \bar{Y} < 8)$,　　(2) $P(S_1^2 / S_2^2 < 8.28)$.

解：（1）

$$\bar{X} \sim N(50, 36/10), \bar{Y} \sim N(46, 16/8),$$

$$E(\bar{X} - \bar{Y}) = 4, D(\bar{X} - \bar{Y}) = \frac{36}{10} + \frac{16}{8} = 5.6$$

$$\bar{X} - \bar{Y} \sim N(4, 5.6),$$

$$Z = \frac{\bar{X} - \bar{Y} - 4}{\sqrt{5.6}} \sim N(0,1),$$

$$P(0 < \bar{X} - \bar{Y} < 8) = P\left(-\frac{4}{\sqrt{5.6}} < Z < \frac{4}{\sqrt{5.6}}\right) = P(-1.69 < Z < 1.69) = 2\Phi(1.69) - 1 = 0.909$$

（2）

$$F = \frac{S_1^2 / 6^2}{S_2^2 / 4^2} = \frac{4}{9} \frac{S_1^2}{S_2^2} \sim F(9, 7),$$

$$P\left(\frac{S_1^2}{S_2^2} < 8.28\right) = P\left(\frac{4S_1^2}{9S_2^2} < 3.68\right) = 1 - P(F \geq 3.68) = 0.95$$

分数	阅卷人

4、 (11 分) 设$X_1, X_2, ..., X_n$是来自总体 $\boldsymbol{X}$ 的简单随机样本， $\boldsymbol{X}$ 的概率密度函数如下：

$$f(x)=\begin{cases}\dfrac{1}{\lambda^2}xe^{-\frac{x}{\lambda}}, & x\ge 0\\ 0, & x<0\end{cases}$$

其中未知参数$\lambda>0$.

（1）求λ的最大似然估计量$\hat{\boldsymbol{\lambda}}$ ；（2） 问$\hat{\boldsymbol{\lambda}}$是否为$\boldsymbol{\lambda}$的无偏估计量？并证明之。

解：(1) 当$x_1, x_2, ..., x_n$为样本值时， 似然函数为

$$L(\lambda)=\lambda^{-2n}(\prod_{i=1}^{n}x_i)\cdot e^{-\frac{\sum_{i=1}^{n}x_i}{\lambda}}$$

对数似然函数为

$$\ln L(\lambda)=-2n\log\lambda+\ln(\prod_{i=1}^{n}x_i)-\frac{\sum_{i=1}^{n}x_i}{\lambda},$$

令 $\dfrac{\partial}{\partial\lambda}\ln L(\lambda)=-2n/\lambda+\dfrac{\sum_{i=1}^{n}x_i}{\lambda^2}=0,$

可得最大似然估计值 $\hat{\lambda}=\dfrac{\sum_{i=1}^{n}x_i}{2n},$

故$\boldsymbol{\lambda}$的最大似然估计量为 $\hat{\lambda}=\dfrac{\sum_{i=1}^{n}X_i}{2n}$.

(2) 因为 $EX=\int_0^{+\infty}x\cdot\dfrac{1}{\lambda^2}xe^{-\frac{x}{\lambda}}dx=\int_0^{+\infty}\dfrac{1}{\lambda^2}x^2e^{-\frac{x}{\lambda}}dx=2\lambda,$

故 $E\hat{\lambda}=\dfrac{\sum_{i=1}^{n}EX_i}{2n}=\dfrac{EX}{2}=\lambda,$

因此λ的最大似然估计$\hat{\boldsymbol{\lambda}}$ 具有无偏性。

分数	阅卷人

5、（10 分）两批导线, 从第一批中抽取 4 根, 从第二批抽取 5 根, 测得其电阻值后, 得到样本均值和样本方差分别为:

$\bar{x}=0.1413,\ s_1^2=8.25\times10^{-6},\ \bar{y}=0.1392,\ \ s_2^2=5.20\times10^{-6}.$ 设两批导线的电阻分别服从正态分布 $X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2,\sigma_2^2)$, 且相互独立, 其中的参数均未知, 求这两批导线电阻的均值差 $\mu_1-\mu_2$（假定 $\sigma_1^2=\sigma_2^2$）以及方差比 σ_1^2/σ_2^2 置信水平为 0. 95 的置信区间.

解： $t_{0.025}(7)=2.36,\quad \bar{x}=0.1413, \bar{y}=0.1392,$

$$S_w^2=\frac{1}{4+5-2}\left(3\times8.25\times10^{-6}+4\times5.20\times10^{-6}\right)=6.507\times10^{-6}，\ S_w=2.55\times10^{-3}.$$

$\boldsymbol{\mu_1-\mu_2}$ 的置信区间： $\left(\bar{x}-\bar{y}\pm t_{\alpha/2}(7)S_w\sqrt{\frac{1}{4}+\frac{1}{5}}\right)=(-0.002,0.006)$

(2) $F_{0.025}(3,\ 4)=9.98, F_{0.975}(3,\ 4)=\frac{1}{15.10},$

$\frac{\sigma_1^2}{\sigma_2^2}$ 的置信区间：$\left(\frac{s_1^2}{s_2^2}\times\frac{1}{F_{0.025}(3,\ 4)},\frac{s_1^2}{s_2^2}\times\frac{1}{F_{0.975}(3,\ 4)}\right)=(0.159,23.96).$

分数	阅卷人

6、(12 分) 工厂从某种化学药品中提取有效成分，为了提高提取率，改革了提炼方法。现对同一质量的药品，分别抽测了采用新、旧方法的提取率（%）各 9 个样本，获得样本数据如下：旧方法的样本均值和样本方差分别为 $\bar{x}=76.4, s_1^2=3.3,$ 新方法的样本均值和样本方差分别为 $\bar{y}=79.6, s_2^2=2.1375,$ 假设新、旧方法的提取率均服从正态分布，

(1) 试检验 $H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$ ；

(2) 在（1）结论的基础上，试问新方法的提取率是否有明显提高？
（显著性水平 α =0.05）

解：(1)$H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$

$$0.226=\frac{1}{F_{0.025}(8,8)}=F_{0.975}(8,8)<\frac{s_1^2}{s_2^2}=\frac{3.3}{2.1375}=1.544<F_{0.025}(8,8)=4.43,$$

接受H_0，认为方差相等。

（2）$H_0:\mu_1\geq\mu_2 \leftrightarrow H_1:\mu_1<\mu_2$

$$s_w^2=\frac{8}{16}s_1^2+\frac{8}{16}s_2^2=2.719, s_w=1.65,$$

$$t_0=\frac{\bar{x}-\bar{y}}{s_w\sqrt{\frac{1}{9}+\frac{1}{9}}}=-4.114<-t_{0.05}(16)=-1.746,$$

拒绝H_0，认为新方法的提取率有明显提高。

分数	阅卷人

7、(15 分) 用显微镜来确定红血球数目，要花费很多时间又不确定。反之，血液中含有细胞的容积却很容易得到。为了掌握两个变量间的可能关系，抽取 10 条狗的血样。对每一血样测定了含有细胞的容积 x 和相应的红血球数 y，所得的资料如下表：

x(毫米^3)	45	42	56	48	42	35	58	40	39	50
y(百万个)	6.53	6.3	9.52	7.5	6.99	5.9	9.49	6.2	6.55	8.72

$$\sum_{i=1}^{10} x_i = 455, \sum_{i=1}^{10} y_i = 73.7, \sum_{i=1}^{10} x_i^2 = 21203, \sum_{i=1}^{10} x_i y_i = 3441.52, \sum_{i=1}^{10} y_i^2 = 560.32$$

求（1）求线性回归方程 $\hat{y} = \hat{a} + \hat{b}x$；

（2）假设检验 $H_0: b = 0 \leftrightarrow H_1: b \neq 0$（显著性水平 $\alpha = 0.05$）；

（3）对含有细胞的容积为 52 毫米^3 估计红血球数，并计算预测值的置信水平为 0.95 的预测区间。

解：

$$(1)\bar{x} = 45.5, \bar{y} = 7.37, S_{xy} = 88.17, S_{xx} = 500.5, S_{yy} = 17.151, \hat{b} = \frac{S_{xy}}{S_{xx}} = 0.176,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = -0.64, \hat{y} = -0.64 + 0.176x.$$

$$(2)Q_e = S_{yy} - \hat{b}S_{xy} = 17.151 - 0.176 * 88.17 = 1.633, \hat{\sigma}^2 = 0.204,$$

$$|t_0| = \left|\frac{0.176}{\sqrt{0.204}}\sqrt{500.5}\right| = 8.72 > t_{0.025}(8) = 2.306, \text{拒绝H}_0$$

$$(3)x = 52, \hat{y} = -0.64 + 0.176 * 52 = 8.51.$$

预测区间为：$(8.51 \pm t_{0.025}(8)\,\hat{\sigma}\sqrt{1 + \frac{1}{10} + \frac{(52-45.5)^2}{s_{xx}}}) = (7.377, 9.644)$.

分数	阅卷人

8、(10 分) 一个年级有 3 个班，现从各个班级中随机地抽取一些学生，记录其成绩（10 分制）如下：

1 班	2 班	3 班
8 9 8 7 6 6	7 9 7 6	7 8 8 9 7

试检验各班级的平均分是否有显著差异（显著性水平$\alpha = 0.05$）？

解：以μ_1, μ_2, μ_3分别表示三个班级的分数的均值，现检验假设
$H_0: \mu_1 = \mu_2 = \mu_3, H_1: \mu_1, \mu_2, \mu_3$不全相等。
现在$s = 3, n_1 = 6, n_2 = 4, n_3 = 5, n = 15,$

$S_T = \sum_{j=1}^{3} \sum_{i=1}^{n_j} x_{ij}^2 - \frac{T_{\cdot\cdot}^2}{15} = 852 - \frac{112^2}{15} = 15.7333,$

$S_A = \sum_{j=1}^{3} \frac{T_{\cdot j}^2}{n_j} - \frac{T_{\cdot\cdot}^2}{15} = \left(\frac{44^2}{6} + \frac{29^2}{4} + \frac{39^2}{5}\right) - \frac{112^2}{15} = 0.85,$

$S_E = S_T - S_A = 14.8833.$

S_T, S_A, S_E的自由度依次为 14，2，12，方差分析表如下

方差来源	平方和	自由度	均方	F 比
因素 误差	0.85 14.8833	2 12	0.425 1.2403	0.3427
总和	15.7333	14		

因$F_{0.05}(2,12) = 3.89 > 0.3427$, 故在显著性水平0.05下，认为各班的平均成绩无显著差异。

分数	阅卷人

9、(10 分) 一计算机程序用来产生在区间（0，10）均匀分布的随机变量的简单随机样本值（即产生区间（0，10）上的随机数），以下是相继得到的 250 个数据的分布情况。试取$\alpha = 0.05$检验这些数据是否来自均匀分布 U(0,10)的总体。亦即检验这一程序是否符合要求。（显著性水平$\alpha = 0.05$）

数据所在区间	0~2	2~4	4~6	6~8	8~10
频 数	38	55	54	41	62

解：要检验假设H_0:这些数据来自均匀分布 U(0,10)的总体。检验统计量为$\chi^2 = \sum_{i=1}^{5} \frac{f_i^2}{np_i} - n$，所需计算列表如下：

A_i	f_i	p_i	np_i	$f_i^2/(np_i)$
A_1	38	0.2	50	28.88
A_2	55	0.2	50	60.5
A_3	54	0.2	50	58.32
A_4	41	0.2	50	33.62
A_5	62	0.2	50	76.88

$\chi^2 = \sum_{i=1}^{5} \frac{f_i^2}{np_i} - n = 258.2 - 250 = 8.2$,检验的临界值为 $\chi^2{}_{0.05}(5-1) = 9.488$ 。因为$\chi^2 = 8.2 < 9.488$，所以样本值没有落入拒绝域，因此接受原假设，即认为这些数据来自均匀分布 U(0,10)的总体。这一程序符合要求。