厦门大学《概率统计I》课程
期末考试卷

_______学院_____系_____年级_____专业

主考教师：______　　试卷类型：（A卷）

可能用到的数值: 标准正态分布的分布函数值: $\Phi(1.84)=0.9671, \Phi(2)=0.9772$;
上侧分位数: $z_{0.05}=1.645, z_{0.01}=2.326$, $\chi^2_{0.10}(3)=6.251, \chi^2_{0.10}(4)=7.779, F_{0.05}(2,27)=3.35$, $F_{0.05}(2,28)=3.34, F_{0.05}(3,27)=2.96, F_{0.025}(2,27)=4.24, F_{0.025}(27,2)=39.46, t_{0.05}(27)=1.7033$, $t_{0.05}(28)=1.7011, t_{0.025}(3)=3.182, t_{0.025}(4)=2.776$.

1.

分数	阅卷人

(12分) 假设甲乙两人同时去办某项业务. 现在有两个窗口A, B可选择, 对应为两个业务员A, B. 两个业务员办理该项业务的时间(单位:分钟)为随机变量, 相互独立, 期望分别为8.8, 10, 方差分别为30.4, 25(窗口屏幕显示). 现在窗口A已有10名人员排队(包含1名刚刚开始办理的, 下同), 窗口B已有9名人员排队. 采用中心极限定理近似, 考虑以下问题:

(i) 甲先选择, 甲要求120分钟内必须开始办理他的业务. 试分别求甲选择窗口A和B可以实现其要求的概率.

(ii) 甲按上述要求选择了概率大的那个窗口, 乙选择另外一个窗口. 试求甲比乙先开始的概率有多大.

解: 设 X_i, Y_i 分别表示窗口A和B第i个排队人员的业务办理时间

$$\sum_{i=1}^{10} X_i \overset{近似}{\sim} N(88, 304), \quad \sum_{i=1}^{9} Y_i \overset{近似}{\sim} N(90, 225)$$

(i) $$P\left(\sum_{i=1}^{10} X_i \le 120\right) = P\left(\frac{\sum_{i=1}^{10} X_i - 88}{\sqrt{304}} \le \frac{120-88}{\sqrt{304}}\right) \approx \Phi(1.84) = 0.9671$$

$$P\left(\sum_{i=1}^{9} Y_i \le 120\right) = P\left(\frac{\sum_{i=1}^{9} Y_i - 90}{\sqrt{225}} \le \frac{120-90}{\sqrt{225}}\right) \approx \Phi(2) = 0.9772$$

(ii) 甲选择了B, 乙选择了A. $\sum_{i=1}^{10} Y_i - \sum_{i=1}^{9} X_i \overset{近似}{\sim} N(2, 23^2)$

$$P\left(\sum_{i=1}^{10} Y_i \le \sum_{i=1}^{9} X_i\right) = P\left(\sum_{i=1}^{10} Y_i - \sum_{i=1}^{9} X_i \le 0\right) = P\left(\frac{\sum_{i=1}^{10} Y_i - \sum_{i=1}^{9} X_i - 2}{23} \le \frac{0-2}{23}\right)$$

$$\approx \Phi(-0.09) = 1-\Phi(0.09) = 1-0.5359 = 0.4641$$

2.

分数	阅卷人

(13分) 随机抽取200 只某种电子元件进行寿命试验, 测得元件的样本平均寿命(单位: h)为325 h, 频数分布为下列表格:

元件寿命	≤ 200	$(200,300]$	$(300,400]$	$(400,500]$	>500
频数	94	25	22	17	42

试检验元件的寿命是否服从指数分布. ($\alpha = 0.10$)

解：建立假设 H_0：X服从参数为θ的指数分布

θ的最大似然估计为 $\hat{\theta} = \bar{x} = 325$.

利用X的密度函数

$$f_X(x) = \begin{cases} \frac{1}{325} e^{-\frac{x}{325}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

容易计算

$\hat{p}_1 = \int_0^{200} f(x)\,dx = 0.4596$，$\hat{p}_2 = \int_{200}^{300} f(x)\,dx = 0.1431$

$\hat{p}_3 = \int_{300}^{400} f(x)\,dx = 0.1052$，$\hat{p}_4 = \int_{400}^{500} f(x)\,dx = 0.0774$

$\hat{p}_5 = 1 - \hat{p}_1 - \hat{p}_2 - \hat{p}_3 - \hat{p}_4 = 0.2147$

这里 $n_1 = 94$，$n_2 = 25$，$n_3 = 22$，$n_4 = 17$，$n_5 = 42$，$n = 200$.

计算可得 $\chi^2 = \left(\sum_{i=1}^{5} \frac{n_i^2}{n\hat{p}_i}\right) - n = 0.7186 < \chi^2_{0.10}(3) = 6.251$

因此接受H_0，即认为元件的寿命服从指数分布.

3.

分数	阅卷人

(12分) 设$X_1,X_2,\ldots,X_{n_1}$ 和$Y_1,Y_2,\ldots,Y_{n_2}$是分别来自正态总体$N(\mu_1,\sigma^2)$,$N(\mu_2,\sigma^2)$的独立样本, σ^2未知. 试求$\mu_1+3\mu_2$的置信区间(置信水平为$1-\alpha$).

解: $\bar{X} \sim N(\mu_1, \frac{\sigma^2}{n_1})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma^2}{n_2})$

$$\Rightarrow \bar{X}+3\bar{Y} \sim N(\mu_1+3\mu_2, \frac{\sigma^2}{n_1}+\frac{9\sigma^2}{n_2})$$

$$\Rightarrow U = \frac{\bar{X}+3\bar{Y}-(\mu_1+3\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{9}{n_2}}} \sim N(0,1)$$

$$V = \frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$$

设$S_W=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$, 因U与V相互独立

$$\Rightarrow \frac{U}{\sqrt{V/(n_1+n_2-2)}} = \frac{\bar{X}+3\bar{Y}-(\mu_1+3\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{9}{n_2}}} \sim t(n_1+n_2-2)$$

$$\Rightarrow P\left[\left|\frac{\bar{X}+3\bar{Y}-(\mu_1+3\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{9}{n_2}}}\right| \leq t_{\frac{\alpha}{2}}(n_1+n_2-2)\right]=\alpha$$

$\Rightarrow \mu_1+3\mu_2$的置信区间为$\left(\bar{X}+3\bar{Y} \pm t_{\frac{\alpha}{2}}(n_1+n_2-2)S_W\sqrt{\frac{1}{n_1}+\frac{9}{n_2}}\right)$.

4.

分数	阅卷人

(16分) 假设某保险公司在过去的年份家庭用车得到如下的年索赔数据(索赔额单位为万元): 家庭用车司机主要为丈夫时(组别A)的样本数据, 样本容量为10, 样本均值为3, 样本方差为0.4; 家庭用车司机主要为妻子时(组别B)的样本数据, 样本容量为10, 样本均值为4, 样本方差为1; 家庭用车司机为夫妻轮流时(组别C) 的样本数据, 样本容量为10, 样本均值为5, 样本方差为4.

(i)试用方差分析的方法(请列出单因素方差分析表)分析司机情况是否对索赔额有显著影响(显著性水平$\alpha=0.05$)?

(ii)分析司机主要为妻子时的索赔额是否显著大于司机主要为丈夫时的索赔额?
(设$H_0:\mu_B\le\mu_A, H_1:\mu_B>\mu_A$, 显著性水平$\alpha=0.05$)

解：(i) $\bar{X}_{\cdot 1}=3$, $\bar{X}_{\cdot 2}=4$, $\bar{X}_{\cdot 3}=5$, $n_1=n_2=n_3=10$,

$$n=30,\quad \bar{X}=\frac{1}{n}\sum_{j=1}^{3} n_i\bar{X}_{\cdot j}=\frac{1}{3}\sum_{j=1}^{3}\bar{X}_{\cdot j}=4$$

$$S_A=\sum_{j=1}^{3} n_i(\bar{X}_{\cdot j}-\bar{X})^2=10\left[1^2+0^2+1^2\right]=20$$

$$\bar{S}_E=\frac{S_E}{n-3}=\frac{\sum_{i=1}^{3}\sum_{i=1}^{10}(X_{ij}-\bar{X}_{\cdot j})^2}{27}=\frac{\sum_{i=1}^{3}9\cdot\frac{1}{9}\sum_{i=1}^{10}(X_{ij}-\bar{X}_{\cdot j})^2}{27}$$

$$=\frac{1}{3}\sum_{i=1}^{3}\frac{1}{9}\sum_{i=1}^{10}(X_{ij}-\bar{X}_{\cdot j})^2=\frac{1}{3}\left[0.4+1+4\right]=1.8$$

方差分析表

方差来源	平方和	自由度	均方	F比
因素A	20	2	10	5.56
误差	50.4	27	1.8	
总和	70.4	29		

F比 > 临界值 $F_{0.05}(2,27)=3.35$

拒绝H_0，即司机情况对索赔额有显著影响.

(ii) 枢轴量 $T=\dfrac{\bar{X}_{\cdot 2}-\bar{X}_{\cdot 1}-(\mu_B-\mu_A)}{\sqrt{\bar{S}_E\left(\frac{1}{n_2}+\frac{1}{n_1}\right)}}\sim t(n-s)$

$$t_0=\frac{4-3}{\sqrt{1.8\cdot\left(\frac{1}{10}+\frac{1}{10}\right)}}=\frac{1}{\sqrt{0.36}}=\frac{1}{0.6}=1.67$$

临界值 $t_{0.05}(27)\doteq 1.7033$

不能拒绝H_0. 即认为司机主要为妻子时的索赔额没有显著大于司机主要为丈夫时的索赔额.

5.

分数	阅卷人

(13分) 考察某校大学女生进入大学学习一年体重变化情况. 在九月份份开学初随机抽取9名新生女生, 测量其体重(单位为公斤), 设为随机变量X_i, 一年后, 继续测量她们的体重, 设为Y_i. 数据如下:

大一开学初	40	50	55	42	40	52	55	53	50
大二开学初	42	51	55	41	38	49	51	48	44

假设$D_i = X_i - Y_i, i = 1, 2, \ldots, 9$相互独立, 且服从同一分布$N(\mu_D, 3^2)$. 试检验一年的大学学习, 女生体重是否显著变轻了, 此时的右边检验设为如下形式:

$$H_0 : \mu_D = 0, H_1 : \mu_D > 0.$$

(i) 设显著性水平$\alpha = 0.5$, 能否拒绝H_0?

(ii) 设显著性水平$\alpha = 0.1$, 能否拒绝H_0?

(iii) 试求此时Z检验的p值.

解: 枢轴量 $Z = \dfrac{\bar{D} - \mu_D}{3/\sqrt{n}} \sim N(0,1)$

检验统计量 $Z_0 = \dfrac{\bar{D}}{3/\sqrt{n}} = \bar{D}$

拒绝域为 $Z_0 = \bar{D} > z_\alpha$

由 D_i 的值: $-2, -1, 0, 1, 2, 3, 4, 5, 6$

Z_0 观察值为 $z_0 = 2$

(i) $z_{0.05} = 1.645$, 由 $z_0 > z_{0.05}$, 拒绝 H_0

(ii) $z_{0.01} = 2.326$, 由 $z_0 < z_{0.05}$, 接受 H_0

(iii) p值 $= P\{Z_0 > z_0 \mid H_0\text{为真}\}$

$= 1 - \Phi(2) = 1 - 0.9772 = 0.0228$

6.

分数	阅卷人

(14分)某医院用光电比色计检验尿汞时, 得尿汞含量x (mg/l) 与消光系数读数Y的结果如下:

尿汞含量(x_i)	2	4	6	8	10
消光系数(y_i)	64	138	205	285	360

(i) 建立Y对x的一元线性回归方程$\hat{Y}=\hat{a}+\hat{b}x$, 并给出σ^2的估计;

(ii) 在$\alpha=0.05$下检验回归效果是否显著?

解: (i) 由表中数据计算可得

$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=6$, $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=210.4$, $S_{xx}=\sum_{i=1}^{n}x_i^2-n\bar{x}^2=40$,

$S_{xy}=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}=1478$, $S_{yy}=\sum_{i=1}^{n}y_i^2-n\bar{y}^2=54649.2$

$\Rightarrow \hat{b}=\frac{S_{xy}}{S_{xx}}=36.95$, $\hat{a}=\bar{y}-\hat{b}\bar{x}=-11.3$

$\Rightarrow$ Y对x的回归方程为 $\hat{Y}=\hat{a}+\hat{b}x=-11.3+36.95x$

σ^2的估计为 $\hat{\sigma}^2=\frac{S_{yy}-\hat{b}S_{xy}}{n-2}=12.37$

(ii) 进行回归效果显著性检验

$H_0: b=0$; $H_1: b\neq 0$.

检验统计量的观察值

$$t=\frac{\hat{b}}{\hat{\sigma}}\sqrt{S_{xx}}=\frac{36.95}{\sqrt{12.37}}\sqrt{40}=66.45$$

由于 $|t|=66.45>t_{0.025}(3)=3.182$,

因此拒绝H_0, 也即认为回归效果显著.

7.

分数	阅卷人

(20分) 已知离散型总体

$$X \sim \begin{pmatrix} -1 & 0 & 1 \\ \theta & 1-3\theta & 2\theta \end{pmatrix}$$

未知参数$\theta \in (0,1/3)$. 样本$X_1, X_2, \ldots, X_n$来自总体X. 现在得到θ的两个估计: $\hat{\theta}_1 = \bar{X}, \hat{\theta}_2 = \frac{\sum_{i=1}^n X_i^2}{3n}$.

(i) 试分别从$EX, E(X^2)$角度证明$\hat{\theta}_1, \hat{\theta}_2$均为$\theta$的矩估计.

(ii) 证明$\hat{\theta}_1, \hat{\theta}_2$均为$\theta$的无偏估计; 并且判别哪个更有效?

(iii) 证明$\hat{\theta}_2$也为θ的最大似然估计.

解: (i) $EX = -\theta + 2\theta = \theta \Rightarrow \theta = EX \Rightarrow \hat{\theta}_1 = \bar{X}$

$E(X^2) = \theta + 2\theta = 3\theta \Rightarrow \theta = \frac{1}{3}E(X^2) \Rightarrow \hat{\theta}_2 = \frac{\sum_{i=1}^n X_i^2}{3n}$

(ii) $E(\hat{\theta}_1) = E(\bar{X}) = EX = \theta$

$E(\hat{\theta}_2) = \frac{1}{3n} E\left(\sum_{i=1}^n X_i^2\right) = \frac{1}{3n} \cdot n\, E(X^2) = \frac{1}{3} \cdot 3\theta = \theta$

因此均为无偏估计

$X^2 \sim \begin{pmatrix} 0 & 1 \\ 1-3\theta & 3\theta \end{pmatrix} \Rightarrow D(X^2) = 3\theta(1-3\theta)$

$D(\hat{\theta}_1) = D(\bar{X}) = \frac{DX}{n} = \frac{E(X^2)+(EX)^2}{n} = \frac{3\theta+\theta^2}{n} = \frac{\theta\cdot(3+3\theta)}{3n}$

$D(\hat{\theta}_2) = D\left(\frac{\sum_{i=1}^n X_i^2}{3n}\right) = \frac{1}{9n^2} \cdot n\, D(X^2) = \frac{3\theta(1-3\theta)}{9n} = \frac{\theta(1-3\theta)}{3n}$

因 $3+3\theta > 1-3\theta$, 有 $D(\hat{\theta}_1) > D(\hat{\theta}_2)$

因此 $\hat{\theta}_2$ 更有效.

(iii) 设 n_1 表示 $x_1, \cdots, x_n$ 中"-1"的个数

n_2 表示 $x_1, \cdots, x_n$ 中"0"的个数

n_3 表示 $x_1, \cdots, x_n$ 中"1"的个数

于是 $L(\theta) = \prod_{i=1}^{n} P\{X_i = x_i\} = \prod_{i=1}^{n} P\{X = x_i\}$

$$= \theta^{n_1}(1-3\theta)^{n_2}(2\theta)^{n_3}$$

$$\Rightarrow \ln L(\theta) = n_1 \ln\theta + n_2 \ln(1-3\theta) + n_3 \ln(2\theta)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0 \Rightarrow n_1 \cdot \frac{1}{\theta} + n_2 \cdot \frac{1}{1-3\theta} \cdot (-3) + n_3 \frac{1}{2\theta} \cdot \theta = 0$$

$$\Rightarrow \frac{3n_2}{1-3\theta} = \frac{n_1+n_3}{\theta} \Rightarrow 3n_2\theta = n_1+n_3 - 3(n_1+n_3)\theta$$

$$\Rightarrow \theta = \frac{n_1+n_3}{3(n_1+n_2+n_3)} = \frac{n_1+n_3}{3n}$$

而 $n_1 + n_3 = \sum_{i=1}^{n} x_i^2$

因此得到唯一的驻点，也即最大似然估计

$$\hat{\theta}_2 = \frac{\sum_{i=1}^{n} x_i^2}{3n}$$