2019-2020 数理统计期末答案

——独自完成，如有错漏，欢迎提出！

一、填空题

1. $\frac{2}{n}\sum_{i=1}^{n}X_i-1$. $\mu_1=E(\mathrm{X})=\frac{1}{2}(a+1), then\ a=2\mu_1-1. Thus,\ \hat{a}=2A_1-1=\frac{2}{n}\sum_{i=1}^{n}X_i-1.$

2. $\frac{n-1}{n}\sigma^2$. $D\left(\bar{X}-X_1\right)=D\left(\left(\frac{1}{n}-1\right)X_1+\frac{1}{n}\sum_{i=2}^{n}X_i\right)=\left(\frac{1}{n}-1\right)^2D(X_1)+\frac{1}{n^2}\sum_{i=2}^{n}D(X_i)=\frac{n-1}{n}\sigma^2.$

易错解答:

①$D\left(\bar{X}-X_1\right)=D\left(\bar{X}\right)+D(X_1)=\frac{\sigma^2}{n}+\sigma^2=\left(\frac{1}{n}+1\right)\sigma^2$.因为$\bar{X}$与$X_1$不相互独立

②$D\left(\bar{X}-X_1\right)=D\left(\frac{1}{n}\sum_{i=2}^{n}(X_i-X_1)\right)=\frac{1}{n^2}\sum_{i=2}^{n}D(X_i-X_1)$.因$X_i-X_1$与$X_j-X_1$不独立

3. 0.5. $F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}, then\ F_{0.9}(20,8)=\frac{1}{F_{0.1}(8,20)}=0.5.$

4. 0.15. 第一类错误：对判错(H_0成立时,样本值落入拒绝域. 弃真,一般较激进.)
第二类错误：错判对(H_0不成立时,样本值不落入拒绝域. 取伪,一般较保守)

5. $D\left(\hat{\theta}\right)\leqslant D\left(\hat{\beta}\right)$. 点估计量有效性定义.

6. (4.4, 5.6). $1-\alpha=0.95\Rightarrow\alpha=0.05.\ \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

$$\Rightarrow\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|<\mathrm{U}_{\alpha/2}\Rightarrow\bar{X}-\frac{\sigma}{\sqrt{n}}\mathrm{U}_{\alpha/2}<\mu<\bar{X}+\frac{\sigma}{\sqrt{n}}\mathrm{U}_{\alpha/2}\Rightarrow(4.4,5.6).$$

二、单选题

1. B. 统计量是且仅是样本的函数.

2. B. $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=F(n_1-1,n_2-1), then\ \frac{S_1^2/4}{S_2^2/5}=\frac{5S_1^2}{4S_2^2}\sim F(7,9).$

3. D. 解析如下：无偏估计即为数字特征法进行点估计.

①② 无偏估计: $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i,\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2.$

③④ 矩估计: $\begin{cases}\mu_1=E(X)=\mu\\ \mu_2=E(X^2)=D(X)+E(X)^2=\sigma^2+\mu^2\end{cases}.$

因样本k阶矩$A_k\left(A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\right)$是总体$k$阶矩的无偏估计量，

故用A_1,A_2分别代替μ_1,μ_2. Then $\begin{cases}\hat{\mu}=A_1=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}\\ \hat{\sigma}^2=A_2-A_1^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\end{cases}$

4. D. A.检验方差$\Rightarrow$检验均值 B.双边检验$\Rightarrow$单边检验 $CD.SS_E$组内， SS_A组间(效应间).

5. A. B.两处‘备择假设’$\Rightarrow$原假设 C.不变$\Rightarrow$变小 D.接受备择假设$\Rightarrow$接受原假设.

6. D. 区间估计: 估计一个范围,且知及其包含参数θ真值的可信程度.

三、解答题

1.(1) $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}\bar{x}^2\right)$

$=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2-2\bar{x}\cdot n\bar{x}+n\bar{x}^2\right)=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2.$

(2) $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{n}\sum_{i=1}^{n}(x_iy_i-x_i\bar{y}-\bar{x}y_i+\bar{x}\bar{y})=\frac{1}{n}\left(\sum_{i=1}^{n}x_iy_i-\bar{y}\sum_{i=1}^{n}x_i-\bar{x}\sum_{i=1}^{n}y_i+n\bar{x}\bar{y}\right)$

$=\frac{1}{n}\left(\sum_{i=1}^{n}x_iy_i-\bar{y}\cdot n\bar{x}-\bar{x}\cdot n\bar{y}+n\bar{x}\bar{y}\right)=\frac{1}{n}\left(\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}\right)=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\bar{x}\bar{y}.$

2.解: 记X为某段时间内开动的机床数，设所供电能最多同时供N台机床工作.

由题知，$X\sim B(100,0.8)$. $E(X)=80$, $D(X)=16$. 题目化为：求满足$P\{X\leqslant N\}\geqslant 95\%$的最小的$N$.

由$De\ Moivre-Laplace$定理知，$P\{X\leqslant N\}=P\left\{\frac{X-80}{\sqrt{16}}\leqslant\frac{N-80}{\sqrt{16}}\right\}=\Phi\left(\frac{N-80}{\sqrt{16}}\right)$.

令$\Phi\left(\frac{N-80}{4}\right)\geqslant 95\%=\Phi(1.65)$,则$\frac{N-80}{4}\geqslant 1.65$,得$N\geqslant 86.6$.因$N\in\mathbb{Z}^+$,故$N\geqslant 87$, $15\cdot N\geqslant 1305$.

于是，最少供应该车间1305个单位电能，才有95%的概率保证不致因供电不足而影响生产.

3.(1) 因$X_i\sim N(0,0.3^2)$, $i=1,2,\cdots,10$，则$\frac{X_i-0}{0.3}\sim N(0,1)$.

于是, $\sum_{i=1}^{3}X_i\sim N(0,0.27)$, $\sum_{i=4}^{10}\left(\frac{X_i-0}{0.3}\right)^2=\frac{1}{0.3^2}\sum_{i=4}^{10}X_i^2\sim\chi^2(7)$.

进而, $\frac{\sum_{i=1}^{3}X_i-0}{\sqrt{0.27}}\Big/\sqrt{\left(\frac{1}{0.3^2}\sum_{i=4}^{10}X_i^2\right)\Big/7}=\sqrt{\frac{7}{3}}\sum_{i=1}^{3}X_i\Big/\sqrt{\sum_{i=4}^{10}X_i^2}\sim t(7)$,知$a=\sqrt{\frac{7}{3}}$.

(2) $\sum_{i=1}^{10}\left(\frac{X_i-0}{0.3}\right)^2=\frac{1}{0.3^2}\sum_{i=1}^{10}X_i^2\sim\chi^2(10)$,则$P\left\{\sum_{i=1}^{10}X_i^2>1.44\right\}=P\left\{\frac{1}{0.3^2}\sum_{i=1}^{10}X_i^2>16\right\}$.

由于$\chi_{0.1}^2(10)=15.987$(题目数据有误),则$P\left\{\sum_{i=1}^{10}X_i^2>1.44\right\}=P\left\{\frac{1}{0.3^2}\sum_{i=1}^{10}X_i^2>16\right\}=0.1$.

4.矩估计: $\mu_1=E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_0^{+\infty} x\cdot\frac{x}{\theta^2}e^{-x^2/(2\theta^2)}\mathrm{d}x=-\int_0^{+\infty} x\mathrm{d}e^{-x^2/(2\theta^2)}$

$$=-xe^{-x^2/(2\theta^2)}\Big|_0^{+\infty}+\int_0^{+\infty} e^{-x^2/(2\theta^2)}\mathrm{d}x=\int_0^{+\infty} e^{-\frac{x^2}{2\theta^2}}\mathrm{d}x=\frac{1}{2}\sqrt{2\pi}\,\theta,\text{故}\,\theta=\mu_1\sqrt{\frac{2}{\pi}}.$$

因样本k阶矩$A_k\left(A_k=\frac{1}{n}\sum_{i=1}^n X_i^k\right)$是总体$k$阶矩的无偏估计量，故用$A_1$代替$\mu_1$，

得到参数θ的矩估计量为$\hat{\theta}=\sqrt{\frac{2}{\pi}}\cdot\frac{1}{n}\sum_{i=1}^n X_i=\sqrt{\frac{2}{\pi}}\cdot\bar{X}$.

$\left[\textit{Poisson}\text{积分}\int_{-\infty}^{+\infty} e^{-x^2}\mathrm{d}x=\sqrt{\pi}\text{.或用正态分布概率密度函数性质}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=1\right]$

极大似然估计量: 似然函数为$L(\theta)=\prod_{i=1}^n f(x_i)=\begin{cases}\prod_{i=1}^n\frac{x_i}{\theta^2}e^{-x_i^2/(2\theta^2)}, & x_i>0,i=1,2,\cdots,n.\\ 0, & else.\end{cases}$

当$L(\theta)>0$时，对数似然函数$\ln L(\theta)=\ln\left[\prod_{i=1}^n\frac{x_i}{\theta^2}e^{-x_i^2/(2\theta^2)}\right]=\sum_{i=1}^n\left(\ln x_i-2\ln\theta-\frac{x_i^2}{2\theta^2}\right)$.

令$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=\sum_{i=1}^n\left(-\frac{2}{\theta}+\frac{x_i^2}{\theta^3}\right)=0$，即$\frac{1}{\theta^3}\left(\sum_{i=1}^n x_i^2-2n\theta^2\right)=0$，

解得θ的极大似然估计值为$\hat{\theta}=\sqrt{\frac{1}{2n}\sum_{i=1}^n x_i^2}$,故参数$\theta$的极大似然估计量为$\hat{\theta}=\sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$.

5.由题总体服从正态分布$X\sim N(\mu,\sigma^2)$且参数μ,σ^2均未知,$n=10,\bar{x}=0.452\%,s=0.037\%$.

(1)依题需在显著性水平$\alpha=0.05$下检验假设: $H_0:\mu\geqslant\mu_0=0.5\%,H_1:\mu<\mu_0=0.5\%$.

由于σ^2未知,故采用t检验.取检验统计量$T=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$.又 $t_\alpha(n-1)=t_{0.05}(9)=1.8331$,

则拒绝域为$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\leqslant -t_\alpha(n-1)$.因$t$的观察值为$t=\frac{0.452\%-0.5\%}{0.037\%/\sqrt{10}}=-4.10241<-1.8331$,

落在拒绝域内，故在显著性水平$\alpha=0.05$下拒绝原假设H_0.

(2)依题需在显著性水平$\alpha=0.05$下检验假设: $H_0:\sigma\leqslant\sigma_0=0.03\%,H_1:\sigma>\sigma_0=0.03\%$.

由于μ未知,故采用χ^2检验.取检验统计量$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$.又$\chi_\alpha^2(n-1)=\chi_{0.05}^2(9)=16.919$,

则拒绝域为$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}\geqslant\chi_\alpha^2(n-1)$.因$\chi^2$的观察值为$\chi^2=\frac{(10-1)\cdot(0.037\%)^2}{(0.03\%)^2}=13.69<16.919$,

不落在拒绝域内，故在显著性水平$\alpha=0.05$下接受原假设H_0.

6.作出方差分析表如下：

方差来源	平方和	自由度	均方和	$F_{比}$
A	352.933	2	176.4665	$\overline{S_A}/\overline{S_E}=3.9157$
e	540.8	12	45.0667	
T	893.733	14		

以μ_1,μ_2,μ_3依次表示三台机器各自五天的日产量均值.在显著性水平α下检验假设:

$$H_0:\mu_1=\mu_2=\mu_3,H_1:\mu_1,\mu_2,\mu_3\text{不全相等}.$$

由$F_{0.05}(2,12)=3.89<F_{比}$,故在显著性水平$\alpha=0.05$下拒绝$H_0$,
认为三台机器各自五天的日产量均值有显著的差异,即认为三台机器的生产能力有显著的差异.

四、应用题

1.记X为掷一次骰子出现的点数.依题需在显著性水平$\alpha=0.05$

下检验假设H_0：X的分布律为$P\{X=k\}=\dfrac{1}{6}=0.1667,k=1,2,\cdots,6.$

在假设H_0成立的前提下，将X的全体可能取值分成6个两两

互不相交的子集$A_1,A_2,\cdots,A_6\,(A_i=\{X=i\},i=1,2,\cdots,6)$.列表如下(其中$n=120$)：

A_i	f_i	p_i	np_i	$f_i^2/(np_i)$
A_1	x	1/6	20	$x^2/20$
A_2	20	1/6	20	20
A_3	20	1/6	20	20
A_4	20	1/6	20	20
A_5	20	1/6	20	20
A_6	$40-x$	1/6	20	$(40-x)^2/20$

令$\chi^2=\displaystyle\sum_{i=1}^{6}\frac{f_i^2}{np_i}-n=0.1x^2-4x+40$.由分组数$k=6,r=0$,知

χ^2的自由度为$k-r-1=5,\chi_\alpha^2(k-r-1)=\chi_{0.05}^2(5)=11.070$.

要使原假设H_0被接受，需使$\chi^2<\chi_\alpha^2(k-r-1)$，解得$9.4786<x<30.5214$.

故当$x\in\{a|10\leqslant a\leqslant 30,a\in\mathbb{Z}\}$时，此骰子是均匀的假设可在显著性水平$\alpha=0.05$下被接受.

2.(1)计算数据:$n=6,\displaystyle\sum_{i=1}^{n}x_i=21,\sum_{i=1}^{n}y_i=79.6,\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=3.5,\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=13.2667,\sum_{i=1}^{n}x_i^2=91,$

$\displaystyle\sum_{i=1}^{n}y_i^2=1076.9,\sum_{i=1}^{n}x_iy_i=297.5.\ S_{xx}=\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2=17.5,S_{yy}=\sum_{i=1}^{n}y_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2=20.8733,$

$\displaystyle S_{xy}=\sum_{i=1}^{n}x_iy_i-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)=18.9.$故$\hat{b}=\dfrac{S_{xy}}{S_{xx}}=1.08,\hat{a}=9.4867$,得到回归方程$\hat{y}=9.4867+1.08x.$

(2)σ^2的无偏估计量为$\hat{\sigma}^2=\dfrac{1}{n-2}\left(S_{yy}-\hat{b}S_{xy}\right)=0.1153$,则$\hat{\sigma}=\sqrt{\hat{\sigma}^2}=0.3396.$

于是,由$|t|=\dfrac{|\hat{b}|}{\hat{\sigma}}\sqrt{S_{xx}}=13.3038>4.6041=t_{0.005}(4)=t_{\alpha/2}(n-2)$,知线性回归的效果是显著的.