东北大学研究生院《应用数理统计》试题(2009A)

说明：1、样本均值与样本方差分别定义为

$$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k,\quad S^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\overline{X})^2;$$

2、分位点Q_α取为上分位点，即：$P(X>Q_\alpha)=\alpha$。

共七题，满分100分，不必抄题，可能有用的一些数据已给出．计算过程中的数据及最终结果都保留小数点后两位；步骤尽量详细写出。

一（10分）、已知X_1，X_2是来自正态总体$N(0,0.08)$的简单随机样本，计算概率：$P\{(X_1+X_2)^2<648(X_1-X_2)^2\}$。 $\binom{X_1+X_2}{X_1-X_2}\sim N(0,\Sigma)$. $\Sigma=\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}0.08&0\\0&0.08\end{pmatrix}\begin{pmatrix}1&1\\1&-1\end{pmatrix}=\begin{pmatrix}0.4^2&0\\0&0.4^2\end{pmatrix}$

提示：考虑F分布 $\therefore X_1+X_2$ 与 X_1-X_2 $\sim N(0,0.4^2)$. (X_1+X_2) 与 (X_1-X_2) 是独立的.

$\therefore (X_1+X_2)^2/(X_1-X_2)^2=(X_1+X_2/0.4)^2/(X_1-X_2/0.4)^2\sim F(1,1)$.

故 $P=P(F(1,1)<648)=0.975$ ✓

二（15分）、假定$X_1,X_2,\ldots,X_n$是从如下总体中抽取的n个简单随机样本： $Var(\ )$

$EX=1/\lambda$.

$\therefore \hat\lambda_1=1/\bar x$ ✓

$$f(x)=\lambda e^{-\lambda x},\quad \lambda>0,\ x>0$$ $L(\theta)=\lambda^n e^{-n\lambda\bar x}$

$EX=\int_0^{\infty}x\lambda e^{-\lambda x}dx$

1（10分）、给出参数λ的矩估计$\hat\lambda_1$、极大似然估计$\hat\lambda_2$； $\therefore \hat\lambda_2=1/\bar x$ ✓

$\boxed{F(x)=1-\lambda e^{-\lambda x}}$

2（5分）、如果丢失了样本的具体观察值，只保留"n个样本里有n_0个落在区间$(c,+\infty)$内"这一信息；c为已知正常数，应该如何估计参数λ？

提示：利用频率估计概率． $\because P(X>c)=e^{-\lambda c}$ ，用 n_0/n 估计 $P(X>c)$ $P(X>c)=e^{-\lambda c}=\frac{n_0}{n}$;

$\therefore -\hat\lambda_3 c=\ln n_0/n$ 得 $\hat\lambda_3=\frac{1}{c}\ln n/n_0$ ✓

三（10分）、假定$X_1,X_2,\ldots,X_{16}$是来自正态总体$N(0,\sigma^2)$的简单随机样本，构造参数σ^2的置信水平0.95的区间估计。 μ(为已知)

$\chi^2_{\frac{\alpha}{2}}(n)$. $\chi^2_{\frac{\alpha}{2}}(n-1)$ $X_i/\sigma\sim N(0,1)$，$\therefore \sum_{i=1}^{16}X_i^2/\sigma^2\sim\chi^2(16)$

$P\{\chi^2_{0.975}(16)<\sum X_i^2/\sigma^2<\chi^2_{0.025}(16)\}=0.95$ 28.85

$\therefore$ 区间为：$(\sum X_i^2/28.85,\ \sum X_i^2/6.91)$ 即：$(0.0347\sum X_i^2,\ 0.1447\sum X_i^2)$ ✓ $(0.03\sum X_i^2,\ 0.14\sum X_i^2)$

H_0成立时 $\frac{\sigma_2^2 C_0}{\sigma_1^2}\geq 1$. 故取

四（15分）、假定$X_1,X_2,\ldots,X_{n_1}$是来自正态总体$N(\mu_1,\sigma_1^2)$的简单随机样本，$Y_1,Y_2,\ldots,Y_{n_2}$是来自另一个独立正态总体$N(\mu_2,\sigma_2^2)$的简单随机样本，$\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$都未知，$C_0$是已知的某个正常数。具体构造如下假设

$P(F(n_1-1,n_2-1)>C)=\alpha$

则得出 $P_{I}\leq\alpha$.

$C=F_\alpha(n_1-1,n_2-1)$.

$$H_0:\sigma_1^2\leq C_0\sigma_2^2\leftrightarrow H_1:\sigma_1^2>C_0\sigma_2^2$$

则 $Z_i=\sqrt{C_0}Y_i\sim N(\sqrt{C_0}\mu_2,C_0\sigma_2^2)$

记 $t_i=\sqrt{C_0}y_i$. $1\leq i\leq n_1$

$\boxed{\bar t=\frac{1}{n_2}\sum_{i=1}^{n_2}t_i=\sqrt{C_0}\bar y,\ S_t^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(t_i-\bar t)^2=C_0 S}$

域为：

$S_1^2/S_t^2>F_\alpha(n_1-1,n_2-1)$

的检验水平α的拒绝域。令 $Z_i=\sqrt{C_0}Y_i$;

$S_1^2/S_t^2>C_0F_\alpha(n_1-1,n_2-1)$

提示：考虑功效函数 $\therefore t_i$ 服从 $N(\sqrt{C_0}\mu_2,\ C_0\sigma_2^2)$. 有 $\frac{(n_2-1)S_t^2}{C_0\sigma_2^2}\sim\chi^2(n_2-1)$ ✓

$\frac{S_1^2/\sigma_1^2}{S_t^2/(C_0\sigma_2^2)}\sim F(n_1-1,n_2-1)$. i.e. $\frac{S_1^2}{S_t^2}\cdot\frac{C_0\sigma_2^2}{\sigma_1^2}\sim F(n_1-1,n_2-1)$. 设当 S_1^2/S_t^2 偏大时否定 H_0

$P(H_0)=P(S_1^2/S_t^2>C)=P(\frac{S_1^2}{S_t^2}\cdot\frac{C_0\sigma_2^2}{\sigma_1^2}>C\cdot\frac{C_0\sigma_2^2}{\sigma_1^2})=P(F(n_1-1,n_2-1)>C\cdot\frac{C_0\sigma_2^2}{\sigma_1^2})$ ……

相关 P118-119.

$\frac{S_X^2}{S_Y^2}>C_0F_\alpha(n_1-1,n_2-1)$

二项分布: $C_2^0(1-p)^2$ $C_2^1(1-p)p$ $C_2^2p^2$

五. (15 分) 对二项总体 $B(2,p)$ 随机抽取的 100 个样本中，有 22 个样本取值为 0，52 个样本取值为 1，26 个样本取值为 2。在检验水平 0.05 下能不能认为总体参数 $p=0.5$？如果把检验水平换成 0.75 呢？

提示：考虑 χ^2 检验

$\chi^2_{0.05}(2)=5.991$. $\chi^2_{0.75}(2)=0.575$.

七. (20 分) 已知 y 关于 x 具有线性回归关系 $y=\beta_0+\beta_1x+\varepsilon,\varepsilon\sim N(0,\sigma^2)$，现在有 8 组观测数据：

x	1	4	7	9	11	15	20	25
y	6	36	25	85	160	140	180	200

1. (12 分) 计算 β_0、β_1、σ^2 的最小二乘估计；

2. (8 分) 在水平 0.05 下检验回归方程成立。

提示：自变量 x 与因变量 y 各自的样本均值、样本方差如下：

$\bar{x}=11.5$, $s_x^2=65.71$, $\bar{y}=104$, $s_y^2=5750.57$；

样本协方差为 $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=497.75$.

$\hat{\beta}_1=\frac{L_{xy}}{L_{xx}}=\frac{8\times497.75}{7\times65.71}=8.657$.

$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\cdot\bar{x}=4.41$ 4.45

$\hat{\sigma}^2=\frac{1}{6}(7\times5750.57-\hat{\beta}_1\times8\times497.75)=961.65$. 963.96.

$\frac{1}{n-2}(L_{yy}-\hat{\beta}_1L_{xy})$;

$r^2=\frac{(8\times497.75)^2}{7\times65.71\times7\times5750.57}=0.856375$

$\therefore F=6\times\frac{r^2}{1-r^2}=35.775$. 35.759.

$F=\frac{(n-2)r^2}{(1-r^2)}$

$L_{xx}=(n-1)\cdot s_x^2$

$L_{yy}=(n-1)\cdot s_y^2$

$L_{xy}=n\cdot$ 协方差

六. (15 分) 随机从三台机器生产的产品中抽检部分样本数据如下：

9 甲：12，15，16，16，17，18，19，19，21；$\bar{x}_1=17$, $s_1^2=7$

10 乙：13，14，15，17，17，17，19，20，22，23；

8 丙：16，17，17，18，18，18，19，23；

假定各机器产品数据服从正态且方差相等，问在水平 0.025 下这三台机器产品的平均数据是否具有显著差异？

提示：三台机器数据的样本均值、样本方差分别为：甲：$\overline{x_1}=17,s_1^2=7$；

乙：$\overline{x_2}=17.7,s_2^2=10.9$； 丙：$\overline{x_3}=18.25,s_3^2=4.5$；

全部 27 个数据的样本均值与样本方差为：$\bar{x}=17.63,s^2=7.40$

$TSS=(n-1)\cdot s^2$

$RSS=(n_1-1)\cdot s_1^2+(n_2-1)s_2^2+(n_3-1)s_3^2$

$CSS=TSS-RSS$

$\because TSS=26\times7.4=192.4$

$RSS=8\times7+9\times10.9+7\times4.5=185.6$

$\therefore CSS=6.8$

$F_{0.025}(2,25)=4.29$

$F=\frac{CSS/2}{RSS/25}=\frac{3.4}{7.424}=0.431$.

$F_{0.05}(1,6)=5.99$

东北大学研究生院《应用数理统计》试题(2009B)

说明：1、样本均值与样本方差分别定义为

$$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k,\quad S^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\overline{X})^2;$$

2、分位点Q_α取为上分位点，即：$P(X>Q_\alpha)=\alpha$。

共七题，满分100分，不必抄题，可能有用的一些数据已给出．计算过程中的数据及最终结果都保留小数点后两位；步骤尽量详细写出。

$X\sim N(0,1)\Rightarrow E|X|=\sqrt{\frac{2}{\pi}}$ $\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x}{2}}$

$X\sim N(0,\sigma^2)\Rightarrow E|X|=\sqrt{\frac{2}{\pi}}\sigma$;

一 (10分)、假定$X_1,X_2,\dots,X_n$是来自总体$N(\mu,9)$的一组简单随机样本，样本容量n应该至少多大才能保证$E|\overline{X}-\mu|\le 0.01$?

$\frac{\sqrt{n}(\bar{x}-\mu)}{3}\sim N(0,1)$ $\bar{x}\sim N(\mu,\frac{9}{n})$, $\therefore E\frac{\sqrt{n}|\bar{x}-\mu|}{3}=\sqrt{\frac{2}{\pi}}$ $\therefore E|\bar{x}-\mu|=3\sqrt{\frac{2}{n\pi}}$ $\therefore 9\times\frac{2}{n\pi}\le\frac{1}{400}$

$n\ge\frac{2}{\pi}\times 9\times 400\approx 2292$

二 (15分)、假定$X_1,X_2,\dots,X_n$是从如下总体中抽取的n个简单随机样本：

$uv=udv+vdu$

$$f(x)=\lambda e^{-\lambda x},\quad \lambda>0,\ x>0$$

$E\overline{X}=\frac{1}{\lambda}$. $\Rightarrow\hat{\lambda}_{矩}=\frac{1}{\bar{x}}$

1 (10分)、给出参数λ的矩估计、极大似然估计；

$L(x,\theta)=\lambda^n e^{-n\lambda\bar{x}}$ $\hat{\lambda}_{极}=\frac{1}{\bar{x}}$

2 (5分)、如果丢失了样本的具体观察值，只保留"n个样本里有n_0个$(n_0>0)$落在区间$(0,c)$内"这一信息，c为已知正常数。应该如何估计参数λ?

$P(a<X<b)=\int_a^b f(x)dx$

提示：利用频率估计概率。 $P(X>c)=e^{-\lambda c}$; 令$\frac{n-n_0}{n}=P(X>c)=e^{-\lambda c}$

得 $\hat{\lambda}_3=\frac{1}{c}\ln\left(\frac{n}{n-n_0}\right)$

三 (10分)、假定$X_1,X_2,\dots,X_{16}$是来自正态总体$N(0,\sigma^2)$的简单随机样本，构造参数σ^2的置信水平$\underline{0.99}$的区间估计。

$X_i/\sigma\sim N(0,1)$; $\sum X_i^2/\sigma^2\sim\chi^2(16)$; $P\left\{\chi^2_{0.995}(16)<\frac{\sum_{i=1}^{16}X_i^2}{\sigma^2}<\chi^2_{0.005}(16)\right\}=0.99=1-\alpha$

$1-\alpha=0.99$ $\alpha=0.01$

四 (15分)、假定$X_1,X_2,\dots,X_{n_1}$是来自正态总体$N(\mu_1,\sigma_1^2)$的简单随机样本，

$Y_1,Y_2,\dots,Y_{n_2}$是来自另一个独立正态总体$N(\mu_2,\sigma_2^2)$的简单随机样本，

$\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$都未知，$C_0$是已知的某个正常数。具体构造如下假设

$$H_0:\sigma_1^2\le C_0\sigma_2^2\ \leftrightarrow\ H_1:\sigma_1^2>C_0\sigma_2^2$$

的检验水平α的拒绝域。(提示：考虑功效函数)。 $Z_i=\sqrt{C_0}Y_i\sim N(\sqrt{C_0}\mu_2,C_0\sigma_2^2)$;

五. (15 分) 对二项总体 $B(2, p)$ 随机抽取的 100 个样本中，有 21 个样本取值为 0，53 个样本取值为 1，26 个样本取值为 2。在检验水平 0.05 下能不能认为总体参数 $p=0.5$？如果把检验水平换成 0.75 呢？

提示：考虑 χ^2 检验 这种分布 $\frac{0}{1/4}\ \frac{1}{1/2}\ \frac{2}{1/4}$ 　$T^2=\frac{1}{100}[4\times21^2+2\times53^2+4\times26^2]-100$

$\chi^2_{0.05}(2)=5.991$ 　$\chi^2_{0.75}(2)=0.575$ 　$=0.86$

所以是水平 0.05 下 接受 H_0

水平 0.75 下 拒绝 H_0

六. (15 分) 随机从三台机器生产的产品中抽检部分样本数据如下：

9 甲：12，15，16，16，17，18，19，19，21；

10 乙：13，14，15，17，17，17，19，20，22，23；

8 丙：16，17，17，18，18，18，19，23；

假定各机器产品数据服从正态且方差相等，问在水平 0.025 下这三台机器产品的平均数据是否具有显著差异？

提示：三台机器数据的样本均值、样本方差分别为：甲：$\overline{x}_1=17, s_1^2=7$；

乙：$\overline{x}_2=17.7, s_2^2=10.9$； 　丙：$\overline{x}_3=18.25, s_3^2=4.5$；

全部 27 个数据的样本均值与样本方差为：$\overline{x}=17.63, s^2=7.40$ 　拒绝域为 $F>F_{0.025}(2,25)$ $=4.29$

$TSS=26\times7.4=192.4$

$RSS=8\times7+9\times10.9+7\times4.5=185.6$ 　$\therefore CSS=6.8$ 　$F=\frac{CSS/2}{RSS/25}=0.431$. 故接受原假设。

七. (20 分) 已知 y 关于 x 具有线性回归关系 $y=\beta_0+\beta_1x+\varepsilon, \varepsilon\sim N(0,\sigma^2)$，

现在有 8 组观测数据：

x	1	4	7	9	11	15	20	25
y	6	36	25	85	160	140	180	200

1. (12 分) 计算 β_0、β_1、σ^2 的最小二乘估计；

2. (8 分) 在水平 0.05 下检验回归方程成立。

提示：自变量 x 与因变量 y 各自的样本均值、样本方差如下：　拒绝域为 $F>F_{0.05}(1,6)=5.99$

$$\overline{x}=11.5,\ s_x^2=65.71;\ \ \overline{y}=104,\ s_y^2=5750.57;$$

样本协方差为 $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=497.75$。

$\hat{\beta}_1=\frac{L_{xy}}{L_{xx}}=\frac{8\times497.75}{7\times65.71}=8.657$. 　$r^2=\frac{(8\times497.75)^2}{7\times65.71\times7\times5750.57}=0.856375$

$\hat{\beta}_0=\overline{y}-\hat{\beta}_1\cdot\overline{x}=4.41$

$\hat{\sigma}^2=\frac{1}{6}\times(7\times5750.57-\hat{\beta}_1\times8\times497.75)=961.65$ 　$\therefore F=6\times\frac{r^2}{1-r^2}=35.775$.

故拒绝原假设，认为回归方程成立。

$r^2=\frac{L_{xy}}{L_{xx}L_{yy}}=$