熊本雕 0404722

东北大学研究生院《应用数理统计》试题(2004秋)

$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$

说明：样本均值与样本方差分别定义为

$$\bar X=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar X)^2$$

共五题，满分80分，步骤尽量详细写出。

2.(1) $\boxed{\bar X^2-\frac{S^2}{n}=\hat{\mu^2}}$

$E(\bar X^2)=D(\bar X)+E^2(\bar X)=\frac{\sigma^2}{n}+\mu^2$

$E(S^2)=\sigma^2\frac{n-1}{n}$

$E(\hat{\mu^2})=E(\bar X^2-\frac{S^2}{n-1})=E(\bar X^2)-E(S^2)\frac{1}{n-1}=\frac{\sigma^2}{n}+\mu^2-\frac{1}{n-1}\cdot\frac{n-1}{n}\sigma^2=\frac{\sigma^2}{n}+\mu^2-\frac{\sigma^2}{n}=\mu^2$

$\therefore\ \hat{\mu^2}$ 为

一、(7分) 假定总体 X 分布为 $N(\mu,0.5)$，如果要以至少 0.9965 的概率保证偏差 $|\bar X-\mu|\le 0.01$，确定样本容量 n 应该至少取多大？

解. $X\sim N(\mu,0.5)$ $\bar X\sim N(\mu,\frac{0.5}{n})$ $\frac{\bar X-\mu}{\sqrt{0.5/n}}\sim N(0,1)$

$P\{|\bar X-\mu|\le 0.01\}=P\{|\frac{\bar X-\mu}{\sqrt{0.5/n}}|\le\frac{0.01}{\sqrt{0.5/n}}\}=2\Phi(\frac{0.01}{\sqrt{0.5/n}})-1\ge 0.9965$

$\Rightarrow\Phi(\frac{0.01}{\sqrt{0.5/n}})\ge\frac{1.9965}{2}=0.99825$ 即 $\frac{0.01}{\sqrt{0.5/n}}\ge u_{0.99825}\Rightarrow n\ge u^2_{0.99825}\times0.5\times10^4=$

$S^2=\frac{1}{n-1}\sum(x_i-\bar x)^2$

有样本方差都是指的修正样本方差 $S^2=\frac{1}{n}\sum(x-\bar x)^2$

二、(23分) 假定 $X_1,X_2,\ldots,X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的一组简单随机样本。

1.(8分) 利用样本均值和样本方差给出 μ^2 的一个无偏估计，并证明你的结论；

2. ①(5分) 如果 μ 是已知的，计算总体方差 σ^2 的极大似然估计；

② (5分) 证明或者举例反证这个极大似然估计是否是无偏估计；

③ (5分) 这个极大似然估计与样本方差比较，哪一个的方差更小？

$\frac{n}{n-1}[E(\bar X^2)-\bar X^2]$

$\bar X^2-\frac{1}{n}S^2$　② $\hat{\sigma^2}=\frac{1}{n}\sum(X_i-\mu)^2$

② $E(\hat{\sigma^2})=\sigma$

解. 1. $D(\bar X)=E(\bar X^2)-E^2(\bar X)$

$=E(\bar X^2)-\sigma^2$

$E^2(\bar X)=\mu^2=E(\bar X^2)-\sigma^2$

$\bar X^2-\frac{S^2}{n}$

$\hat{\mu^2}=\frac{1}{n-1}\sum x_i^2-S^2=\frac{n}{n-1}(\frac{1}{n}\sum_{i=1}^{n}x_i^2)-S^2$

$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x)^2=\frac{1}{n-1}[\sum_{i=1}^{n}x_i^2-n\bar x^2]$

$=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2-\frac{n}{n-1}\bar x^2$

2.① $L(\sigma^2)=\prod_{i=1}^{n}f(x_i,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

$\ln L(\sigma^2)=-\frac{n}{2}(\ln 2\pi+\ln\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

$\frac{\partial\ln L(\sigma^2)}{\partial\sigma^2}=-\frac{n}{2}\cdot\frac{1}{\sigma^2}+\frac{1}{2}\cdot\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0\Rightarrow\hat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$

③ $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$　$D(\chi^2)=2n$

即 $E(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2)=\sigma^2$ $\therefore\ \hat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$ 为 σ^2 的无偏估计

2.① $\frac{x_i-\mu}{\sigma}\sim N(0,1)$　$E(\frac{1}{n}\sum(x_i-\mu)^2)=\sigma^2$

$\frac{\sum(x_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

$\sum\frac{(x_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

$E[\frac{\sum(x_i-\mu)^2}{\sigma^2}]=n$

③ $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x)^2$，$\hat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$

2.③ $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$　$D(\chi^2)=2n$

$D(\frac{1}{n}\sum(x_i-\mu)^2)=2\sigma^4\frac{n}{n^2}=\frac{2\sigma^4}{n}$

$\therefore D(\frac{n-1}{\sigma^2}S^2)=2(n-1)$　即 $D(S^2)=\frac{2\sigma^4}{n-1}$

$\because\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$　故 $D(\frac{\sum(x_i-\mu)^2}{\sigma^2})=2n$

$D(\frac{n}{\sigma^2}\cdot\frac{1}{n}\sum(x_i-\mu)^2)=\frac{n^2}{\sigma^4}D(\frac{1}{n}\sum(x_i-\mu)^2)=2n$

$D(\hat{\sigma^2})=\frac{2\sigma^4}{n}<D(S^2)$

三、(?分) 假定 $X_1,X_2,\ldots,X_n$ 是来自正态总体 $N(\mu_1,\sigma^2)$ 的简单随机样本，$Y_1,Y_2,\ldots,Y_m$ 是来自另一个独立的正态总体 $N(\mu_2,\sigma^2)$ 的简单随机样本，μ_1、μ_2、σ^2 都未知，具体构造 $\mu_1-2\mu_2$ 的一个置信水平 $1-\alpha$ 的区间估计。

提示：正态分布的线性变换仍然服从正态分布。

记 $S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar X)^2$，$S_Y^2=\frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar Y)^2$

$\frac{(n-1)S_X^2}{\sigma^2}\sim\chi^2(n-1)$　$\frac{(m-1)S_Y^2}{\sigma^2}\sim\chi^2(m-1)$ 则 $\frac{1}{\sigma^2}[(n-1)S_X^2+(m-1)S_Y^2]\sim\chi^2(n+m-2)$

解. 由题意可知：$\bar X\sim N(\mu_1,\frac{\sigma^2}{n})$，$\bar Y\sim N(\mu_2,\frac{\sigma^2}{m})$

$2\bar Y\sim N(2\mu_2,\frac{4}{m}\sigma^2)$ 则 $\bar X-2\bar Y\sim N(\mu_1-2\mu_2,(\frac{1}{n}+\frac{4}{m})\sigma^2)$

$\frac{(\bar X-2\bar Y)-(\mu_1-2\mu_2)}{\sigma\sqrt{\frac1n+\frac4m}}\sim N(0,1)$，并且 $\frac{1}{\sigma^2}[(n-1)S_X^2+(m-1)S_Y^2]$ 相互独立，

$t\triangleq\frac{[(\bar X-2\bar Y)-(\mu_1-2\mu_2)]/(\sigma\sqrt{\frac1n+\frac4m})}{\frac{1}{\sigma}\sqrt{[(n-1)S_X^2+(m-1)S_Y^2]/(n+m-2)}}$

$=\frac{(\bar X-2\bar Y)-(\mu_1-2\mu_2)}{\sqrt{(n-1)S_X^2+(m-1)S_Y^2}}\cdot\sqrt{\frac{nm(n+m-2)}{m+4n}}\sim t(n+m-2)$

因为 $\bar X-2\bar Y$ 是 $\mu_1-2\mu_2$ 的良好无偏估计量，故通常 $|(\bar X-2\bar Y)-(\mu_1-2\mu_2)|$ 应很小，即存在正数 C，通常有 $|(\bar X-2\bar Y)-(\mu_1-2\mu_2)|<C$。

由于 $\bar X$，$\bar Y$ 分别为 μ_1，μ_2 的最小方差无偏估计量

因此 $\bar X-2\bar Y-(\mu_1-2\mu_2)$ 通常很小，考虑 $\mu_1-2\mu_2$ 的区间形式为 $\bar X-2\bar Y\pm C$ ....

$$1-\alpha = P\{|(\bar{X}-2\bar{Y})-(\mu_1-2\mu_2)| < c\} = P\left\{\left|\frac{(\bar{X}-2\bar{Y})-(\mu_1-2\mu_2)}{\sqrt{(n-1)S_X^2+(m-1)S_Y^2}}\sqrt{\frac{nm(n+m-2)}{m+4n}}\right| < \frac{c}{\underbrace{\sqrt{(n-1)S_X^2+(m-1)S_Y^2}}_{N}}\underbrace{\sqrt{\frac{nm(n+m-2)}{m+4n}}}_{M}\right\}$$

$$= \Phi\left(\frac{c}{N/M}\right) - \Phi\left(\frac{-c}{N/M}\right) = 2\Phi\left(\frac{c}{N/M}\right) - 1 \qquad \therefore \Phi\left(\frac{c}{N/M}\right) = 1-\frac{\alpha}{2}$$

$$\therefore \frac{c}{N/M} = t_{1-\frac{\alpha}{2}}(n+m-2) \text{ 即 } c = t_{1-\frac{\alpha}{2}}(n+m-2)\frac{N}{M} = \sqrt{(n-1)S_X^2+(m-1)S_Y^2}\cdot\sqrt{\frac{m+4n}{nm(n+m-2)}}\cdot t_{1-\frac{\alpha}{2}}(n+m-2)$$

0
6
0
7
2
2

$\therefore \mu_1-2\mu_2$ 的一个置信水平 $1-\alpha$ 的区间估计为 $\left[(\bar{X}-2\bar{Y}) \pm t_{1-\frac{\alpha}{2}}(n+m-2)\cdot\sqrt{(n-1)S_X^2+(m-1)S_Y^2}\cdot\sqrt{\frac{m+4n}{nm(n+m-2)}}\right]$

四. (15 分) 假定 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu_1, \sigma^2)$ 而 $Y_1, Y_2, \dots, Y_m$ 是来自另一个独立的 $N(\mu_2, \sigma^2)$ 的简单随机样本，μ_1、μ_2、σ^2 都未知，对于待检验的假设 H_0: $\mu_1=\mu_2 \leftrightarrow H_1$: $\mu_1 \neq \mu_2$，甲同学采用 t 分布来检验而乙同学则利用方差分析来处理，问甲、乙两同学的处理方法本质上是否相同？具体给出你的讨论过程。 本质：假设检验

提示：乙同学考虑的是单因素两个水平的方差分析 记 $S_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ $S_2^2=\frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar{Y})^2$ 则 $\frac{(n-1)S_1^2}{\sigma^2}\sim\chi^2(n-1)$ $\frac{(m-1)S_2^2}{\sigma^2}\sim\chi^2(m-1)$

解：甲同学：H_0: $\mu_1=\mu_2$，H_1: $\mu_1\neq\mu_2$

因为 $\bar{X}-\bar{Y}$ 是 $\mu_1-\mu_2$ 的最小方差无偏估计量

所以 H_0 成立时，H_0 的拒绝域为 $\{|\bar{X}-\bar{Y}|>c\}$

又 $\bar{X}\sim N(\mu_1, \frac{\sigma^2}{n})$，$\bar{Y}\sim N(\mu_2, \frac{\sigma^2}{m})$

$\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2, (\frac{1}{n}+\frac{1}{m})\sigma^2) = N(0, (\frac{1}{n}+\frac{1}{m})\sigma^2)$

故 $\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim N(0,1)$

$\frac{1}{\sigma^2}[(n-1)S_1^2+(m-1)S_2^2]\sim\chi^2(n+m-2)$ 且与 $\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$ 相互独立

$\therefore T = \frac{(\bar{X}-\bar{Y})/(\sigma\sqrt{\frac{1}{n}+\frac{1}{m}})}{\frac{1}{\sigma}\sqrt{(n-1)S_1^2+(m-1)S_2^2}/\sqrt{n+m-2}} = \frac{\bar{X}-\bar{Y}}{\sqrt{(n-1)S_1^2+(m-1)S_2^2}}\cdot\sqrt{\frac{nm(n+m-2)}{n+m}}\sim t(n+m-2)$

故 H_0 的拒绝域为 $\left\{|\bar{X}-\bar{Y}|\sqrt{\frac{nm(n+m-2)}{(n+m)[(n-1)S_1^2+(m-1)S_2^2]}} > t_{1-\frac{\alpha}{2}}(n+m-2)\right\}$

乙同学：H_0: $\mu_1=\mu_2$，H_1: $\mu_1\neq\mu_2$。 $Q_A = n\bar{X}^2+m\bar{Y}^2-\frac{(\sum X_i+\sum Y_i)^2}{n+m}$

$Q_e=(n-1)S_1^2+(m-1)S_2^2$ 则 $\frac{Q_A}{\sigma^2}\sim\chi^2(1)$，$\frac{Q_e}{\sigma^2}\sim\chi^2(n+m-2)$ 且 $\frac{Q_A}{\sigma^2}$ 与 $\frac{Q_e}{\sigma^2}$ 独立

$F \triangleq \frac{Q_A/\sigma^2}{Q_e/\sigma^2(n+m-2)} = \frac{(n+m-2)Q_A}{Q_e}\sim F(1, n+m-2)$

若 $\frac{(n+m-2)Q_A}{Q_e}$ 较大，说明水平之间的差异影响比随机误差影响大，故应拒绝 H_0，所以 H_0 的拒绝域为 $\left\{\frac{(n+m-2)Q_A}{Q_e}>c\right\}$

当 α 给定后，由 $\alpha = P\{F>c|H_0 \text{成立}\}$ 得：$1-\alpha=P\{F\le c|H_0\text{成立}\}$

$\therefore H_0$ 的拒绝域为 $\left\{\frac{(n+m-2)Q_A}{Q_e}>F_{1-\alpha}(1, n+m-2)\right\}$

从以上的分析可知：甲、乙两同学的处理方式本质上是相同的，都是先提出原假设，确定拒绝域，只不过在确定拒绝域中的常数 c 时，构造的统计量不一样。甲同学采用 t 分布，而乙同学采用 F 分布。

五. (20 分) 已知 y 关于 x 具有线性回归关系 $y=\beta_0+\beta_1 x+\varepsilon$，$\varepsilon\sim N(0,\sigma^2)$，现在有 6 组观测数据，

x	2	4	6	8	10	12
y	64	136	205	285	360	y_0

其中 y_0 是一个特殊的观察值。

1. (8 分) 分别把 β_0 与 β_1 的最小二乘估计表示成 y_0 的函数形式。

2. (12 分) 考虑对于回归系数的检验 H_0: $\beta_1=0 \leftrightarrow H_1$: $\beta_1\neq 0$，取检验水平 $\alpha=0.0474$，此时拒绝域为 $\{F_0>8\}$，这里 F_0 是某个服从 F 分布的检验统计量；问观察值 y_0 必须满足什么条件才能保证回归方程成立(即拒绝零假设)？

提示：t 分布的平方服从 F 分布，一些可能有用的计算结果如下

$\bar{x}=\frac{1}{6}\sum_{i=1}^{6}x_i=7$，$\sum_{i=1}^{6}(x_i-\bar{x})^2=70$，$\bar{y}=\frac{1}{6}\sum_{i=1}^{6}y_i=175+y_0$，

$\sum_{i=1}^{6}x_iy_i=7782+72y_0$，$\sum_{i=1}^{6}(y_i-\bar{y})^2=30y_0^2-2100y_0+91692 \triangleq L_{yy}$

1. 解：$\hat{\beta}_1=\frac{L_{xy}}{L_{xx}}$

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

$L_{xy}=\sum_{i=1}^{6}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{6}x_iy_i-\bar{y}\sum_{i=1}^{6}x_i-\bar{x}\sum_{i=1}^{6}y_i+6\bar{x}\bar{y}$

$=\sum_{i=1}^{6}x_iy_i-6\bar{x}\bar{y}=7782+72y_0-6\times7\times(175+y_0)=432+30y_0$

$L_{xx}=\sum_{i=1}^{6}(x_i-\bar{x})^2=70$

$\therefore \hat{\beta}_1=\frac{1}{70}(432+30y_0)$，$\hat{\beta}_0=175+y_0-7\times\frac{1}{70}(432+30y_0)$
$=131.8-2y_0$

2. 解：H_0: $\beta_1=0$ H_1: $\beta_1\neq 0$

$\because \hat{\beta}\sim N(\beta, \sigma^2 C)$ 即 $\hat{\beta}_i\sim N(\beta_i, \sigma^2 c_{ii})$，$i=0,1$

$\therefore H_0$ 成立时，$(\hat{\beta}_1/\sigma\sqrt{c_{11}})\sim N(0,1)$ 又 $\because \hat{\beta}$ 与 Q_e 独立，故 $\hat{\beta}_1$ 与 Q_e 独立且 $\frac{Q_e}{\sigma^2}\sim\chi^2(n-k-1)$ $(n=6, k=1)$

$\therefore H_0$ 成立时，$\frac{\hat{\beta}_1^2}{\sigma^2 c_{11}}\sim\chi^2(1)$

$F \triangleq \frac{\hat{\beta}_1^2/\sigma^2 c_{11}}{\frac{Q_e}{\sigma^2}\cdot\frac{1}{n-2}} = \frac{4\hat{\beta}_1^2}{Q_e c_{11}}\sim F(1,4)$

$\therefore H_0$ 的拒绝域为 $\{F>F_{1-\alpha}(1,4)\}$

$\therefore F=F_0$，$F_{1-\alpha}(1,4)=F_{0.9526}(1,4)=8$

$Q_e=L_{yy}-\frac{L_{xy}^2}{L_{xx}}=30y_0^2-2100y_0+91692-\frac{(432+30y_0)^2}{70}$
$=\frac{120}{7}(y_0-72.06)^2$

$U_R=\frac{30^2(y_0+14.4)^2}{70}$

$X'=\begin{bmatrix}1&1&1&1&1&1\\2&4&6&8&10&12\end{bmatrix}$

$X'X=\begin{bmatrix}6&42\\42&364\end{bmatrix}$ $C=(X'X)^{-1}=\frac{1}{420}\begin{bmatrix}364&-42\\-42&6\end{bmatrix}$ $\therefore c_{11}=\frac{1}{70}$

$\therefore \frac{4\hat{\beta}_1^2}{Q_e c_{11}}=\frac{4\times(\text{?})^2}{1}=\frac{4\times\frac{1}{70}(432+30y_0)^2}{\frac{1}{70}(30y_0^2-2100y_0+91692-\frac{(432+30y_0)^2}{70})}$

$=\frac{4\times(432+30y_0)^2}{(2100y_0^2-147000y_0+91692\times70-(432+30y_0)^2)}>8$

$\therefore F_0=\frac{U_R}{Q_e/n-k-1}=\frac{4U_R}{Q_e}=3\left(\frac{y_0+14.4}{y_0-72.06}\right)^2\sim F_{1-\alpha}(1,4)$

$\Rightarrow y_0<39.22$ 或 $y_0>208.65$