东北大学研究生院《应用数理统计》试题（2006）

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. 分位点统一为上分位点，即随机变量 X 的上 α 分位点 Q_α 是指：

$$P(X>Q_\alpha)=\alpha$$

共五题，满分 80 分，步骤尽量详细写出。

$U_{1-\alpha}$

一. （10 分）设总体 X 分布为 $N(\mu,0.04)$，如果要以至少 0.95 的概率保证偏差 $|\overline{X}-\mu|\le 0.1$，确定样本容量 n 应该至少取多大？

$X=e^{y}$

$\int_{-\infty}^{+\infty}f(x)dx=0$

二. （15 分）假定 $X_1,X_2,\cdots X_n$ 是来自总体 $X(X>0)$ 的简单随机样本，$Y=\ln X$ 服从正态分布 $N(\mu,1)$.

$EX=\int_{-\infty}^{+\infty}e^{\frac{2y-(y-\mu)^2}{2}}dy=$

$b=e^{\mu+\frac{1}{2}}$ 由 μ 知 u 的区间估计

人如从把 u 的区间估计代入

$EX=e^{\mu+\frac{1}{2}}$

$EX=e^{\mu+\frac{1}{2}}$

1. （5 分）求 μ 的置信度为 0.95 的区间估计；

2. （10 分）求 X 的数学期望 EX（记为 b）的置信度为 0.95 的区间估计。

$E(X)=E(e^{Y})=\int_{-\infty}^{+\infty}e^{y}f(y)dy$　　$\bar{x}\pm t(n-1)\frac{S}{\sqrt{n}}$

再做

三. （20 分）假定 $X_1,X_2,\cdots X_n$ 相互独立，其中 X_i 服从 $N(\beta\mu_i,\sigma^2)$ 分布，$i=1,\cdots n$.

其中 $\mu_1,\mu_2,\cdots\mu_n$ 为已知常数

$\hat{\beta}=\frac{\sum_{i=1}^{n}x_i\mu_i}{\sum_{i=1}^{n}\mu_i^2}$

估计量和极大似然估计量的函数

1. （10 分）试求 β,σ^2 的极大似然估计 $\hat{\beta},\hat{\sigma}^2$；

2. （5 分）$\hat{\beta}$ 是否为 β 的无偏估计；无偏

$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\beta}\mu_i)^2$

3. （5 分）$\hat{\sigma}^2$ 是否为 σ^2 的无偏估计，若不是请由此构造一个无偏估计。

$E(\hat{\sigma}^2)=\frac{n-1}{n}\sigma^2$ 故不是无偏估计　因：$E(\frac{n}{n-1}\hat{\sigma}^2)=\sigma^2$ 即 $\frac{n}{n-1}\hat{\sigma}^2$ 是 σ^2 的无偏估计

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

四. （15 分）某种配偶的后代按体格的属性分为三类，按照某种遗传模型其概率应分别为 p^2，$2p$（$1-p$）和（$1-p$）^{2}。今随机抽样得各类的数目分别为 n_1，n_2，n_3。试给出在水平 $\alpha=0.05$ 之下用 χ^2 检验法检验数据与模型是否相符的拒绝域。

未知参数 $l=1$　　χ　　$\chi^2_\alpha(3-1)$　　$t(n)$　　$\chi^2(n-1)$

$\chi^2_{0.05}(m-1-l)$

$=\chi^2_{0.95}(1)=3.841$

p 已知？未知？

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

[3] 2004 秋

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

其中 y_0 是一个特殊的观察值。

1. (12 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. (8 分) 若 $x^*=14$, 试给出 y^* 的置信度为 0.95 的预测区间。

1 此种情况不检验回归方程是否成立

一些可能有用的计算结果如下:

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

$$\sum_{i=1}^{6}x_i y_i=7782+72y_0, \sum_{i=1}^{6}\left(y_i-\bar{y}\right)^2=30y_0^2-2100y_0+91692$$

k表示 β_k 中非常数项,未知参数个数

预测区间:用途:因为 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x$ 用 $\frac{U_R}{Q_e/(n-2)}=\frac{U_R/k}{Q_e(n-k-1)}>F_{0.95}(k, n-k-1)$

若: $\frac{U_R/k}{Q_e/(n-k-1)}>F_{0.95}(k,n-k-1)$ 则 y 与 x 之间线性关系显著, 故可用上述回归方程进行预测。当 $x=x^*$ 时, y 的点预测值为 $\hat{y}_{x^*}$ 为:

$\hat{y}_{x^*}=\hat{\beta}_0+\hat{\beta}_1 x^*=$. 2、$\hat{\sigma}_e=\sqrt{\frac{Q_e}{n-k-1}}=$ $t_{0.975}(4)=$ $\bar{x}$

$L_{xx}=\cdots \Rightarrow \sqrt{1+G C G'}=\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{L_{xx}}}$

$\Rightarrow \hat{y}_1=\hat{y}_{x^*}-\hat{\sigma}_e t_{1-\alpha/2}(n-k-1)\sqrt{1+GCG'}$

$\hat{y}_2=\hat{y}_{x^*}+\hat{\sigma}_e t_{1-\alpha/2}(n-k-1)\sqrt{1+GCG'}$

其中 $U_R=\hat{\beta}_1^2 L_{xx}=\hat{\beta}_1 L_{xy}$ $Q_e=L_{yy}-U_R$

一元回归: β_0, β_1 时

多元回归: $U_R=\sum_{i=1}^{n}\left(\sum_{j=1}\hat{\beta}_j x_{ij}\right)^2$

则: $\hat{y}_1=47.2+4y_0-7.85\sqrt{y_0-72.061}$ $\frac{2.1682}{5}$

$\hat{y}_2=47.2+4y_0+7.85\sqrt{y_0-72.061}$ 131.5

故预测区间为: $(47.2+4y_0-7.85\sqrt{y_0-72.061}, 47.2+4y_0+7.85\sqrt{y_0-72.061})$