数理统计复习提纲(茆诗松版)

1 统计量与抽样分布

- 统计量的概念
  ◇ 样本的函数，且不含有任何参数

- χ^2-分布, t-分布, F-分布的定义
  ◇ $\{X_i\}_1^n \overset{i.i.d}{\sim} N(0,1)$, $X = \sum\limits_{i=1}^{n} X_i^2 \sim \chi^2(n)$
  ◇ $X\sim N(0,1), Y\sim\chi^2(n)$,且$X,Y$独立，则$T = \frac{X}{\sqrt{Y/n}}\sim t(n)$
  ◇ $X\sim\chi^2(m), Y\sim\chi^2(n)$,且$X,Y$独立，则$T = \frac{X/m}{Y/n}\sim F(m,n)$

- χ^2-分布, t-分布, F-分布的性质, α分位点查表;
  ◇ χ^2-分布($X\sim\chi^2(n)$): 图像；$E(X) = n$,$D(X) = 2n$；可加性；分位点($P(X \le \chi^2_\alpha(n)) = \alpha$)
  ◇ t-分布($X\sim t(n)$): 图像；分位点($P(X \le t_\alpha(n)) = \alpha$)；$t_{1-\alpha}(n) = -t_\alpha(n)$
  ◇ F-分布($X\sim F(m,n)$): 图像；$\frac{1}{X} = F(n,m)$；可加性；分位点($P(X \le F_\alpha(m,n)) = \alpha$)；$F_{1-\alpha}(m,n) = \frac{1}{F_\alpha(n,m)}$

- 正态总体八种抽样分布:

$\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$	σ已知时, 对μ的推断
$\dfrac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$	σ未知时, 对μ的推断
$\sum \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$	μ 已知时, 对σ的推断
$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	μ未知时, 对σ的推断
$\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1}{m}+\frac{\sigma_2}{n}}} \sim N(0,1)$	σ_1,σ_2 已知时, 对$\mu_1-\mu_2$的推断
$\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$	$\sigma_1=\sigma_2$且未知时, 对$\mu_1-\mu_2$的推断
$\dfrac{(X_i-\mu_1)^2/\sigma_1^2}{(Y_i-\mu_2)^2/\sigma_2^2} \sim F(m,n)$	μ_1和μ_2已知时, 对$\dfrac{\sigma_1}{\sigma_2}$的推断
$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(m-1,n-1)$	μ_1和μ_2未知时, 对$\dfrac{\sigma_1}{\sigma_2}$的推断

2 参数点估计

- 无偏估计; 无偏估计的有效性; 相合估计的概念;
  ◇ $E(\hat{\theta}) = \theta$

$\diamond$ $E(\hat{\theta}_1)=\theta$, $E(\hat{\theta}_2)=\theta$, 若$D(\hat{\theta}_1)<D(\hat{\theta}_2)$
$\diamond$ $\hat{\theta}\xrightarrow{P}\theta$，即任意$\varepsilon>0$, $\lim_{n\to\infty}P(|\hat{\theta}_n-\theta|<\varepsilon)=1$ 或者$\lim_{n\to\infty}P(|\hat{\theta}_n-\theta|\geq\varepsilon)=0$

- 矩估计方法与最大似然估计法.

例1：设简单随机样本$X_1,X_2,\cdots,X_n$为取自密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{2x}{\theta^2} & x\in[0,\theta]\\ 0 & \text{其它}\end{cases}$$

的总体. 求

1. θ的矩估计量$\hat{\theta}_M$;
2. θ的最大似然估计量$\hat{\theta}_L$.

1答:

$$\begin{aligned}\mathrm{E}(X) &= \int_0^\theta \frac{2x^2}{\theta^2}=\frac{2}{3}\theta\\ \theta &= \frac{3}{2}\mathrm{E}(X)\\ \hat{\theta}_M &= \frac{3}{2}\overline{X}\end{aligned}$$

2答：似然函数为:

$$\begin{aligned}L(\theta) &= \begin{cases}\dfrac{2^n}{\theta^{2n}}\prod_{i=1}^n x_i & x_1,\cdots,x_n\leq\theta\\ 0 & \text{其它}\end{cases}\\ &= \begin{cases}\dfrac{2^n}{\theta^{2n}}\prod_{i=1}^n x_i & \theta\geq x_{(n)}\\ 0 & \theta<x_{(n)}\end{cases}\end{aligned}$$

其中$x_{(n)}=\max(x_1,\cdots,x_n)$. 当$\theta\geq x_{(n)}$时, $L(\theta)$是θ的单调递减函数, 而当$\theta<x_{(n)}$时, $L(\theta)=0$, 于是$L(\theta)$在$x_{(n)}$处达到最大值, 即得$\hat{\theta}_L=x_{(n)}$, θ的最大似然估计量为$\hat{\theta}_L=X_{(n)}$

3 正态总体参数的区间估计与假设检验

- μ, σ^2的置信度为$1-\alpha$区间估计:

  1. 从上表中选择合适的$T(X_1,\cdots,X_n;\theta)$;
  2. 根据$T(X_1,\cdots,X_n;\theta)$的分布, 确定分位点$c_{1-\frac{\alpha}{2}}$, $c_{\frac{\alpha}{2}}$, 使得$P\{c_{1-\frac{\alpha}{2}}\leq T(X_1,\cdots,X_n;\theta)\leq c_{\frac{\alpha}{2}}\}=1-\alpha$. (注：当$T(X_1,\cdots,X_n;\theta)$的密度关于$Y$-轴对称时, $c_{1-\frac{\alpha}{2}}=-c_{\frac{\alpha}{2}}$);
  3. 解不等式：$c_{1-\frac{\alpha}{2}}\leq T(X_1,\cdots,X_n;\theta)\leq c_{\frac{\alpha}{2}}$, 即得$\theta$的置信度为$1-\alpha$区间估计.

- 显著性水平为α的μ, σ^2的假设检验:

  1. 从上表中选择合适的随机变量, 其中的θ取H_0中等号后的数值: $T(X_1,\cdots,X_n;\theta_0)$;
  2. 根据H_0和H_1的形式，判断$T(X_1,\cdots,X_n;\theta_0)$何时对$H_0$有利，何时不利, 比如越大对$H_1$越有利, 则拒绝域的形式为$W=\{T|T(X_1,\cdots,X_n;\theta)>c\}$;

3. 根据$T(X_1,\cdots,X_n;\theta_0)$的分布, 确定$c$使得$P\{T\in W|\theta=\theta_0\}=\alpha$.

4. 将样本观测值带入T 得到T_0，若$T_0\in W$，则$RejH_0$，否则，$AccH_0$.

- 显著性水平α, 两类错误的概念

  ◇ $P(T\in W|H_0)=\alpha$—第一类错误

  ◇ $P(T\in \bar{W}|H_1)=\beta$—第二类错误

例2. 设某种职业的年收入$\mathbb{X}\sim N(\mu,\sigma^2)$, 现从该种职业人群中随机抽取一组容量为36的样本, 算得样本均值$\bar{x}=6.5$(万元), 样本标准差$s=1.0$(万元).

1. 求总体标准差σ的置信度为0.95的区间估计;

2. 在显著性水平$\alpha=0.05$下, 能否认为该种职业的平均年收入超过了6万元.

1答: 总体标准差σ的置信度为0.95的区间估计为:

$$\begin{aligned}
& \left[s\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}}, s\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}}\right] \\
= & \left[\sqrt{\frac{35}{53.2033}}, \sqrt{\frac{35}{20.5694}}\right] \\
= & [0.8111,\ 1.3044]
\end{aligned}$$

2答: 构造假设:

$$H_0:\ \mu\le 6 \quad H_1:\ \mu>6$$

的拒绝域为:

$$\begin{aligned}
W &= \left\{\boldsymbol{x}|\bar{x}>6+\frac{s}{\sqrt{n}}t_\alpha(n-1)\right\} \\
&= \left\{\boldsymbol{x}|\bar{x}>6+\frac{1}{\sqrt{36}}1.6896\right\} \\
&= \{\boldsymbol{x}|\bar{x}>6.2816\}
\end{aligned}$$

由于$\bar{x}=6.5\in W$, 因此拒绝H_0, 即可以认为该种职业的平均年收入超过了6万元.

4 分布拟合检验

- 一般步骤：

- H_0:$X\sim F(x)$

  1. 将总体X的取值范围分成k个互不相交的区间$A_1,\ldots,A_k$;

  2. 计算样本观察值落入每个小区间$(A_i,A_{i+1}]$的频数$f_i, i=1,\ldots,k-1$.

  3. 计算H_0为真时，总体X落入区间$(A_i,A_{i+1}]$ 的概率$p_i=P(A_i<X\le A_{i+1}), i=1, ldots, k-1$.
     (注意，若$F(x)$ 中含有r 个未知数，则需要先用MLE 估计未知数，然后再计算p_i.)

  4. 构造检验统计量$K=\sum_{i=1}^{k}\frac{(f_i-np_i)^2}{np_i}$, 由Pearson定理知道，$K\sim\chi^2(k-1)$.
     (注意，若$F(x)$ 中含有r 个未知数，则需要先用MLE 估计这r个未知数，此时，$K\sim\chi^2(k-1-r)$.)

5. 拒绝域$W = K > \chi^2_\alpha(k-1-r)$. 若$K$ 的观察值$K_0 \in W$, 则$Rej H_0$, 否则，$Acc H_0$.

- 掌握具体检验方法(例子). 掷一颗骰子60次，结果如下：

点数	1	2	3	4	5	6
次数	7	8	12	11	9	13

试在显著性水平为下检验这颗骰子是否均匀。

解： H_0：$F(x)$是离散分布$P(X=k)=\frac{1}{6}$，$k=1,\cdots,6$。

组频率：

k	1	2	3	4	5	6
f_k	0.117	0.133	0.2	0.184	0.15	0.216

$\chi^2 = \sum_{i=1}^{k}\frac{n}{p_i}(n_i/n-p_i)^2 = 2.780$，

拒绝域$W = \{\chi^2 = \sum_{i=1}^{k}\frac{n}{p_i}(n_i/n-p_i)^2 \ge \chi^2_\alpha(k-1)\} = \{\chi^2 \ge \chi^2_{0.05}(5) = 11.07\}$

所以接受。

5 单因子方差分析

单因子方差分析步骤：

步骤：假设 H_0：$a_1=a_2=\cdots=a_r=0$ vs H_1：$a_1, a_2, \cdots, a_r$不全等于 0，

统计量 $F=\frac{S_A/f_A}{S_e/f_e}=\frac{MS_A}{MS_e}\sim F(r-1, n-r)$，

显著水平α，右侧拒绝域 $W=\{f\ge f_{1-\alpha}(r-1, n-r)\}$，

计算f，并作出判断.

这是 F 检验法.

单因子方差分析数据：

在单因子方差分析中通常将试验数据及基本计算结果写成表格形式

因子水平	试验数据				和	和的平方	平方和
A_1	Y_{11}	Y_{12}	$\cdots$	Y_{1m}	T_1	T_1^2	$\sum Y_{1j}^2$
A_2	Y_{21}	Y_{22}	$\cdots$	Y_{2m}	T_2	T_2^2	$\sum Y_{2j}^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_r	Y_{r1}	Y_{r2}	$\cdots$	Y_{rm}	T_r	T_r^2	$\sum Y_{rj}^2$
Σ					T	$\sum_{i=1}^{r} T_i^2$	$\sum_{i=1}^{r}\sum_{j=1}^{m} Y_{ij}^2$

方差分析表：

通常列成方差分析表：

来源	平方和	自由度	均方和	F比
因子	S_A	$f_A=r-1$	$MS_A=S_A/f_A$	$F=MS_A/MS_e$
误差	S_e	$f_e=n-r$	$MS_e=S_e/f_A$	
总和	S_T	$f_T=n-1$		

记

$$T_i=\sum_{j=1}^{m}Y_{ij}\text{，}\quad T=\sum_{i=1}^{r}T_i=\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}\text{，}$$

可得

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{m}(Y_{ij}-\overline{Y})^2=\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}^2-n\overline{Y}^2=\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}^2-\frac{1}{n}\left(\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}\right)^2=\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}^2-\frac{1}{n}T^2\text{，}$$

$$S_A=m\sum_{i=1}^{r}(\overline{Y}_{i\cdot}-\overline{Y})^2=m\left[\sum_{i=1}^{r}\overline{Y}_{i\cdot}^2-r\overline{Y}^2\right]=m\sum_{i=1}^{r}\left(\frac{1}{m}\sum_{j=1}^{m}Y_{ij}\right)^2-mr\left(\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}\right)^2=\frac{1}{m}\sum_{i=1}^{r}T_i^2-\frac{1}{n}T^2\text{，}$$

$$S_e=S_T-S_A=\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}^2-\frac{1}{m}\sum_{i=1}^{r}T_i^2\text{．}$$

例题：

例　在饲料养鸡增肥的研究中，现有三种饲料配方：A_1,A_2,A_3，为比较三种饲料的效果，特选 24 只相似的雏鸡随机均分为三组，每组各喂一种饲料，60 天后观察它们的重量．实验结果如下表所示：

饲料	鸡重/g							
A_1	1073	1009	1060	1001	1002	1012	1009	1028
A_2	1107	1092	990	1109	1090	1074	1122	1001
A_3	1093	1029	1080	1021	1022	1032	1029	1048

在显著水平 $\alpha=0.05$ 下检验这三种饲料对雏鸡增重是否有显著差别．

解：假设 H_0：$a_1=a_2=a_3=0$　vs　H_1：a_1,a_2,a_3 不全等于 0，

统计量 $F=\dfrac{S_A/f_A}{S_e/f_e}=\dfrac{MS_A}{MS_e}\sim F(r-1,n-r)$，平方和

显著水平 $\alpha=0.05$，$n=24$，$r=3$，$m=8$，右侧拒绝域 $W=\{f\ge f_{0.95}(2,21)\}=\{f\ge 3.47\}$，

试验数据计算表

因子水平	试验数据 Y_{ij}								T_i	T_i^2	$\sum_{j=1}^{m}Y_{ij}^2$
A_1	1073	1009	1060	1001	1002	1012	1009	1028	8194	67141636	8398024
A_2	1107	1092	990	1109	1090	1074	1122	1001	8585	73702225	9230355
A_3	1093	1029	1080	1021	1022	1032	1029	1048	8354	69789316	8728984
总和									25133	210633177	26357363

计算可得

$$S_A=\frac{1}{m}\sum_{i=1}^{r}T_i^2-\frac{1}{n}T^2=\frac{1}{8}\times 210633177-\frac{1}{24}\times 25133^2=9660.0833\text{，}$$

$$S_e=\sum_{i=1}^{r}\sum_{j=1}^{m}Y_{ij}^2-\frac{1}{m}\sum_{i=1}^{r}T_i^2=26357363-\frac{1}{8}\times 210633177=28215.875\text{，}$$

方差分析表

来源	平方和	自由度	均方和	F 比
因子	9660.0833	2	4830.0417	3.5948
误差	28215.875	21	1343.6131	
总和	37875.9583	23		

有 F 比 $f=3.5948\in W$，

故拒绝 H_0，接受 H_1，可以认为这三种饲料对雏鸡增重有显著差别，

并且检验的 p 值 $p=P\{F\ge 3.5948\}=1-0.9546=0.0454<\alpha=0.05$．

6　一元线性回归分析

一元线性回归模型：

一元线性回归模型

$$\begin{cases} Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \cdots, n; \\ \text{各}\varepsilon_i\text{相互独立，且服从相同的正态分布}N(0, \sigma^2). \end{cases}$$

根据观测数据(x_i, y_i)，对参数β_0, β_1作出估计，得到$\hat{\beta}_0, \hat{\beta}_1$，取

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x ,$$

称为 Y 关于 x 的经验回归函数，也称为回归方程．若给定 x 的值 x_0，可得$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$，称为随机变量 Y 在 x_0 处的回归值或预测值．

模型参数的最小二乘估计：

对于 n 组观测数据(x_i, y_i)，$i = 1, 2, \cdots, n$，总的误差平方和

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 ,$$

选取β_0, β_1的值，使得 Q 达到最小．令 Q 关于β_0, β_1的偏导数等于 0，得

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i) \cdot (-1) = 0; \\ \dfrac{\partial Q}{\partial \beta_1} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i) \cdot (-x_i) = 0. \end{cases}$$

称为正规方程组，经过整理，可得

$$\begin{cases} n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

故取β_0, β_1的最小二乘估计为

$$\begin{cases} \hat{\beta}_1 = \dfrac{l_{xY}}{l_{xx}}; \\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x}. \end{cases}$$

线性回归模型的三种检验：

一．F 检验

步骤：假设 H_0：$\beta_1 = 0$ vs H_1：$\beta_1 \neq 0$，

统计量$F = \dfrac{S_R}{S_e/(n-2)} \sim F(1, n-2)$，

显著水平α，右侧拒绝域 $W = \{f \geq F_{1-\alpha}(1, n-2)\}$，

计算f，并作出判断．

计算公式：$S_R = \hat{\beta}_1^2 l_{xx}$，$S_T = l_{yy}$，$S_e = S_T - S_R = l_{yy} - \hat{\beta}_1^2 l_{xx}$．

二．T检验

因 $\hat{\beta}_1 \sim N\left(\beta_1, \dfrac{\sigma^2}{l_{xx}}\right)$， $\dfrac{S_e}{\sigma^2} \sim \chi^2(n-2)$，且 $\hat{\beta}_1 = \sqrt{\dfrac{S_R}{l_{xx}}}$ 与 S_e 相互独立，有 $\dfrac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{l_{xx}}} \sim N(0,1)$，则根据 t 分布的定义可知：

$$T = \frac{\dfrac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{l_{xx}}}}{\sqrt{\dfrac{S_e}{\sigma^2}\Big/(n-2)}} = \frac{(\hat{\beta}_1 - \beta_1)\sqrt{l_{xx}}}{\sqrt{\dfrac{S_e}{n-2}}} = \frac{(\hat{\beta}_1 - \beta_1)\sqrt{l_{xx}}}{\hat{\sigma}} \sim t(n-2).$$

步骤：假设 H_0：$\beta_1 = 0$ vs H_1：$\beta_1 \neq 0$，

统计量 $T = \dfrac{\hat{\beta}_1\sqrt{l_{xx}}}{\hat{\sigma}} \sim t(n-2)$，其中 $\hat{\sigma} = \sqrt{\dfrac{S_e}{n-2}}$，

显著水平 α，双侧拒绝域 $W = \{|t| \geq t_{1-\alpha/2}(n-2)\}$，

计算 t，并作出判断.

注意到 $T^2 = \dfrac{\hat{\beta}_1^2 l_{xx}}{\hat{\sigma}^2} = \dfrac{S_R}{S_e/(n-2)} = F$，可见 T 检验与 F 检验本质上是一致的.

三．相关系数检验

对应于总体相关系数 $\mathrm{Corr}(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \dfrac{E\{[X-E(X)][Y-E(Y)]\}}{\sqrt{E[X-E(X)]^2}\sqrt{E[Y-E(Y)]^2}}$，定义样本相关系数 $r = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$，有 $|r| \leq 1$，且当 $|r| = 1$ 时，X_i 与 Y_i 具有完全的线性关系，即存在常数 a、b，使得 $Y_i = aX_i + b$. 如果 $|r|$ 越接近 1，表明 X_i 与 Y_i 的线性关系越强；如果 $|r|$ 越接近 0，表明 X_i 与 Y_i 的线性关系越弱.

对于假设 H_0：$\beta_1 = 0$，可将拒绝域取为 $W = \{|r| \geq c\}$ 的形式.

因 $r = \dfrac{l_{xY}}{\sqrt{l_{xx}} \cdot \sqrt{l_{YY}}}$，则 $r^2 = \dfrac{l_{xY}^2}{l_{xx} \cdot l_{YY}} = \dfrac{\hat{\beta}_1^2 l_{xx}}{l_{YY}} = \dfrac{S_R}{S_T} = \dfrac{S_R}{S_R + S_e} = \dfrac{\dfrac{S_R}{S_e/(n-2)}}{\dfrac{S_R}{S_e/(n-2)} + (n-2)} = \dfrac{F}{F + (n-2)}$，可见相关系数检验（$r$ 检验）与 F 检验本质上是一致的.

为了方便使用，根据 F 分布的分位数，可得 $|r|$ 的分位数

$$r_{1-\alpha}(n-2) = \sqrt{\frac{F_{1-\alpha}(n-2,1)}{F_{1-\alpha}(n-2,1) + n - 2}}.$$

步骤：假设 H_0：$\beta_1 = 0$ vs H_1：$\beta_1 \neq 0$，

统计量 $r = \dfrac{l_{xY}}{\sqrt{l_{xx}} \cdot \sqrt{l_{YY}}}$，

显著水平 α，拒绝域 $W = \{|r| \geq r_{1-\alpha}(n-2)\}$，

计算样本相关系数 r，并作出判断.

例题：

例　为了考察某企业产量与成本的关系，调查获得 5 组数据：

产量（吨）	25	28	30	32	35
成本（万元）	384	395	412	417	430

求：（1）产量与成本的线性回归方程；（2）对回归方程作显著性检验（$\alpha=0.01$）；（3）回归系数β_1与β_0，误差方差σ^2以及产量为 40 时平均成本 $E(Y_0)$ 的置信区间（$\alpha=0.01$）；（4）产量为 40 时成本 Y_0 的预测区间（$\alpha=0.01$）.

解：（1）根据试验数据得出计算表：

试验数据计算表

$\Sigma x_i=150$	$n=12$	$\Sigma y_i=2038$
$\overline{x}=30$		$\overline{y}=407.6$
$\Sigma x_i^2=4558$	$\Sigma x_i y_i=61414$	$\Sigma y_i^2=832014$
$n\overline{x}^2=4500$	$n\overline{x}\,\overline{y}=61140$	$n\overline{y}^2=830688.8$
$l_{xx}=58$	$l_{xy}=274$	$l_{yy}=1325.2$
	$\hat{\beta}_1=l_{xy}/l_{xx}=4.7241$ $\hat{\beta}_0=\overline{y}-\hat{\beta}_1\overline{x}=265.8759$	

故回归方程为 $\hat{Y}=265.8759+4.7241x$；

（2）F 检验：假设 H_0：$\beta_1=0$　vs　H_1：$\beta_1\neq 0$，

统计量 $F=\dfrac{S_R}{S_e/(n-2)}\sim F(1,n-2)$，

显著水平 $\alpha=0.01$，$n=5$，$F_{1-\alpha}(1,n-2)=F_{0.99}(1,3)=34.12$，右侧拒绝域 $W=\{f\geq 34.12\}$，

因 $S_R=\hat{\beta}_1^2 l_{xx}=4.7241^2\times 58=1294.4138$，$S_T=l_{yy}=1325.2$，有 $S_e=S_T-S_R=30.7862$，

则 $f=\dfrac{1294.4138}{30.7862/3}=126.1358\in W$，

故拒绝 H_0，接受 H_1，回归方程显著；

方差分析表

来源	平方和	自由度	均方和	F 比
回归	1294.4138	1	1294.4138	126.1358
残差	30.7862	3	10.2621	
总和	1325.2	4		

（注：检验的 p 值为 $p=P\{F\geq 126.1358\}=0.0015$）；

或 r 检验：假设 H_0：$\beta_1 = 0$ vs H_1：$\beta_1 \neq 0$，

统计量 $r = \dfrac{l_{xY}}{\sqrt{l_{xx}} \cdot \sqrt{l_{YY}}}$，

显著水平 $\alpha = 0.01$，$n = 5$，$r_{1-\alpha}(n-2) = r_{0.99}(3) = 0.959$，拒绝域 $W = \{|r| \geq 0.959\}$，
因 $l_{xx} = 0.018567$，$l_{xy} = 2.4675$，$l_{yy} = 345.0625$，

则 $r = \dfrac{l_{xy}}{\sqrt{l_{xx}} \cdot \sqrt{l_{yy}}} = \dfrac{274}{\sqrt{58} \times \sqrt{1325.2}} = 0.9883 \in W$，

故拒绝 H_0，接受 H_1，回归方程显著；

（3）因 $n = 5$，$\alpha = 0.01$，有 $\bar{x} = 30$，$\hat{\sigma} = \sqrt{\dfrac{S_e}{n-2}} = \sqrt{\dfrac{30.7862}{3}} = 3.2034$，

$t_{1-\alpha/2}(n-2) = t_{0.995}(3) = 5.8409$，

$\chi^2_{\alpha/2}(n-2) = \chi^2_{0.005}(3) = 0.0717$，$\chi^2_{1-\alpha/2}(n-2) = \chi^2_{0.995}(3) = 14.8603$，

故回归系数 β_1 的 0.99 置信区间为

$$\left[\hat{\beta}_1 \pm t_{1-\alpha/2}(n-2) \cdot \frac{\hat{\sigma}}{l_{xx}}\right] = \left[4.7241 \pm 5.8409 \times \frac{3.2034}{58}\right] = [4.4015, 5.0467]$$

$= [4.4015, 5.0467]$，

回归系数 β_0 的 0.99 置信区间为

$$\left[\hat{\beta}_0 \pm t_{1-\alpha/2}(n-2) \cdot \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}}}\right] = \left[265.8759 \pm 5.8409 \times 3.2034 \times \sqrt{\frac{1}{5} + \frac{30^2}{58}}\right]$$

$= [191.6961, 340.0556]$；

误差方差 σ^2 的 0.99 置信区间为

$$\left[\frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}(n-2)}, \frac{(n-2)\hat{\sigma}^2}{\chi^2_{\alpha/2}(n-2)}\right] = \left[\frac{3 \times 3.2034^2}{14.8603}, \frac{3 \times 3.2034^2}{0.0717}\right] = [2.0717, 429.3753]；$$

产量 $x_0 = 40$（吨）时平均成本 $E(Y_0)$ 的 0.99 置信区间为

$$\left[\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{1-\alpha/2}(n-2) \cdot \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}\right]$$

$$= \left[265.8759 + 4.7241 \times 40 \pm 5.8409 \times 3.2034 \times \sqrt{\frac{1}{5} + \frac{(40-30)^2}{58}}\right]$$

$= [428.8867, 480.7960]$；

（4）产量为 40 时成本 Y_0 的预测区间为

$$\left[\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{1-\alpha/2}(n-2) \cdot \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}\right]$$

$$= \left[265.8759 + 4.7241 \times 40 \pm 5.8409 \times 3.2034 \times \sqrt{1 + \frac{1}{5} + \frac{(40-30)^2}{58}}\right]$$

$= [422.8453, 486.8374]$.