习题五

5.1 试证明一元线性回归模型中参数β_0和β_1的最小二乘估计就是参数的极大似然估计.

证明 因$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N\left(0, \delta^2\right)$，故$y_i \sim N\left(\beta_0 + \beta_1 x_i, \delta^2\right)$。

似然函数为

$$L\left(\beta_0, \beta_1, \delta^2\right) = \prod_{i=1}^{n} P\left(y_i\right) = \prod_{i=1}^{n} \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{\left[y_i - \left(\beta_0 + \beta_1 x_i\right)\right]^2}{2\delta^2}}$$

其中$P\left(y_i\right)$为y_i的密度函数。

$$\ln L\left(\beta_0, \beta_1, \delta^2\right) = \sum_{i=1}^{n}\left(\ln \frac{1}{\delta\sqrt{2\pi}} - \frac{\left(y_i - \beta_0 - \beta_1 x_i\right)^2}{2\delta^2}\right)$$

$$\begin{cases} \dfrac{\partial \ln L\left(\beta_0, \beta_1, \delta^2\right)}{\partial \beta_0} = 0 \\ \dfrac{\partial \ln L\left(\beta_0, \beta_1, \delta^2\right)}{\partial \beta_1} = 0 \\ \dfrac{\partial \ln L\left(\beta_0, \beta_1, \delta^2\right)}{\partial \delta^2} = 0 \end{cases} \Rightarrow \begin{cases} \hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\delta^2} = \dfrac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 \end{cases}$$

即β_0, β_1的极大似然估计与最小二乘估计相同。

5.2 试证明变元$\left(x, y\right)$的一组观测值的样本相关系数就是把$\left(x, y\right)$视为二维随机变量时，随机变量x和y相关系数的矩法估计.

证明 二维随机变量$\left(x, y\right)$的相关系数为

$$r = \frac{\operatorname{cov}\left(x, y\right)}{\sqrt{Dx}\sqrt{Dy}} = \frac{Exy - ExEy}{\sqrt{Ex^2 - \left(Ex\right)^2}\sqrt{Ey^2 - \left(Ey\right)^2}}$$

但因$\left(x, y\right)$的分布未知，上述公式中Exy, Ex, Ey, Ex^2, Ey^2均不可求。但根据大数定理，可用样本矩来估计总体矩，于是：

$$\hat{r}=\frac{\frac{1}{n}\sum x_i y_i-\overline{x}\,\overline{y}}{\sqrt{\overline{x^2}-\left(\overline{x}\right)^2}\sqrt{\overline{y^2}-\left(\overline{y}\right)^2}}=\frac{\frac{1}{n}\sum\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\frac{1}{n}L_{xx}}\sqrt{\frac{1}{n}L_{yy}}}=\frac{L_{xy}}{L_{xx}L_{yy}}\ 。$$

5.3 某种钢材的强度 y（单位：kg/mm^2）与它的含碳量 x（单位：%）有关，现测得数据如下：

含碳量 x_i	0.08	0.10	0.12	0.14	0.16
强度 y_i	41.8	42.0	44.7	45.1	48.9

设有 $y_i=\beta_0+\beta_1 x_i+\varepsilon_i$ ， $\varepsilon_i\sim N(0,\sigma^2)$ ， $i=1,2,\cdots,5$ ， $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_5$ 相互独立。

求：（1）β_0,β_1 的最小二乘估计 $\hat{\beta}_0,\hat{\beta}_1$ ；

（2）残差平方和 SS_e ，估计的标准差 $\hat{\sigma}$ ，样本相关系数 r 。

解 $n=5$ ， $\overline{x}=0.12$ ， $L_{xx}=0.004$ ， $\overline{y}=44.5$ ， $L_{yy}=33.3$ ，

$$L_{xy}=\sum_{i=1}^{n}x_i y_i-n\,\overline{x}\,\overline{y}=27.046-5\times0.12\times44.5=0.346\ 。$$

（1） $\hat{\beta}_1=\frac{L_{xy}}{L_{xx}}=\frac{0.346}{0.004}=86.5$ ，

$\hat{\beta}_0=\overline{y}-\hat{\beta}_1\overline{x}=44.5-86.5\times0.12=34.12$ 。

所以，回归方程为 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x=34.12+86.5x$ 。

（2） $SS_e=L_{yy}-\hat{\beta}_1 L_{xy}=33.3-86.5\times0.346=3.371$ ，

$\hat{\sigma}=\sqrt{\frac{SS_e}{n-2}}=\sqrt{\frac{3.371}{5-2}}=1.060$ ， $r=\frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}}=\frac{0.346}{\sqrt{0.004\times33.3}}=0.9480$ 。

5.4 对工件表面作腐蚀刻线试验，测得蚀刻时间 x（单位：秒）和蚀刻深度 y（单位：μm）的数据如下：

蚀刻时间 x_i	20	30	40	50	60
蚀刻深度 y_i	13	16	17	20	23

设有 $y_i=\beta_0+\beta_1 x_i+\varepsilon_i$ ， $\varepsilon_i\sim N(0,\sigma^2)$ ， $i=1,2,\cdots,5$ ， $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_5$ 相互独立。

（1）求 β_0,β_1 的最小二乘估计 $\hat{\beta}_0,\hat{\beta}_1$ ；

（2）求残差平方和 SS_e ，估计的标准差 $\hat{\sigma}$ ，样本相关系数 r ；

（3）检验 H_0： $\beta_1=0$ （显著水平 $\alpha=0.05$ ） 。

解 $n=5$ ， $\bar{x}=40$ ， $L_{xx}=1000$ ， $\bar{y}=17.8$ ， $L_{yy}=58.8$ ，

$$L_{xy}=\sum_{i=1}^{n}x_i y_i-n\,\bar{x}\,\bar{y}=3800-5\times 40\times 17.8=240 \text{ 。}$$

（1） $\hat{\beta}_1=\dfrac{L_{xy}}{L_{xx}}=\dfrac{240}{1000}=0.24$ ，

$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=17.8-0.24\times 240=8.2$ 。

所以，回归方程为 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x=8.2+0.24x$ 。

（2） $SS_e=L_{yy}-\hat{\beta}_1 L_{xy}=58.8-0.24\times 240=1.200$ ，

$$\hat{\sigma}=\sqrt{\frac{SS_e}{n-2}}=\sqrt{\frac{1.2}{5-2}}=0.6325 \text{ ，}$$

$$r=\frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}}=\frac{240}{\sqrt{1000\times 58.8}}=0.9897 \text{ 。}$$

（3）用 t 分布检验：

$$T=\frac{\hat{\beta}_1}{\hat{\sigma}}\sqrt{L_{xx}}=\frac{0.24}{0.6325}\sqrt{1000}=12.0 \text{ 。}$$

对 $\alpha=0.05$ ，查 t 分布的分位数表，可得 $t_{1-\alpha/2}(n-2)=t_{0.975}(3)=3.1824$ ，因为 $|T|=|12.0|=12.0>3.1824$ ，所以拒绝 H_0： $\beta_1=0$ ，说明自变量 x 与因变量 y 之间有显著的统计线性相关关系。

用 F 分布检验：

$$F=\frac{L_{yy}-SS_e}{SS_e/(n-2)}=\frac{58.8-1.2}{1.2/(5-2)}=144.0 \text{ 。}$$

对 $\alpha=0.05$ ，查 F 分布的分位数表，可得 $F_{1-\alpha}(1,n-2)=F_{0.95}(1,3)=10.1$ ，因为 $F=144.0>10.1$ ，所以结论也是拒绝 H_0： $\beta_1=0$ 。

5.5 在研究钢线的含碳量 x（单位：%）与电阻 y（单位：μΩ）的关系时，测得数据如下：

含碳量 x_i	0.10	0.30	0.40	0.55	0.70	0.80	0.95
电阻 y_i	15.0	18.0	19.0	21.0	22.6	23.8	26.0

设有 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ， $\varepsilon_i \sim N(0, \sigma^2)$， $i = 1, 2, \cdots, 7$， $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_7$ 相互独立。

（1）求 β_0, β_1 的最小二乘估计 $\hat{\beta}_0, \hat{\beta}_1$ ；

（2）求残差平方和 SS_e ，估计的标准差 $\hat{\sigma}$ ，样本相关系数 r ；

（3）求 β_0, β_1 的置信水平为 95% 的置信区间 ；

（4）检验 H_0： $\beta_1 = 0$ （显著水平 $\alpha = 0.05$） 。

解 $n = 7$ ， $\bar{x} = 0.5428571$ ， $L_{xx} = 0.5321429$ ， $\bar{y} = 20.77143$ ， $L_{yy} = 84.03429$ ，

$$L_{xy} = \sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 85.61 - 7 \times 0.5428571 \times 20.77143 = 6.678572 \text{ 。}$$

（1） $\hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}} = \dfrac{6.678572}{0.5321429} = 12.55034$ ，

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 20.77143 - 12.55034 \times 0.5428571 = 13.95839 \text{ 。}$$

所以，回归方程为 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 13.95839 + 12.55034x$ 。

（2） $SS_e = L_{yy} - \hat{\beta}_1 L_{xy} = 84.03429 - 12.55034 \times 6.678572 = 0.21594$ ，

$$\hat{\sigma} = \sqrt{\frac{SS_e}{n-2}} = \sqrt{\frac{0.21594}{7-2}} = 0.20782 \text{ ，}$$

$$r = \frac{L_{xy}}{\sqrt{L_{xx} L_{yy}}} = \frac{6.678572}{\sqrt{0.5321429 \times 84.03429}} = 0.99871 \text{ 。}$$

（3） 对 $1 - \alpha = 0.95$，查 t 分布的分位数表可得 $t_{1-\alpha/2}(n-2) = t_{0.975}(5) = 2.5706$，

$$t_{1-\alpha/2}(n-2)\frac{\hat{\sigma}}{\sqrt{L_{xx}}} = 2.5706 \times \frac{0.20782}{\sqrt{0.5321429}} = 0.73233 \text{ 。}$$

$$\underline{\theta} = \hat{\beta}_1 - t_{1-\alpha/2}(n-2)\frac{\hat{\sigma}}{\sqrt{L_{xx}}} = 12.55034 - 0.73233 = 11.818 \text{ ，}$$

$$\overline{\theta}=\hat{\beta}_1+t_{1-\alpha/2}(n-2)\frac{\hat{\sigma}}{\sqrt{L_{xx}}}=12.55034+0.73233=13.283\ 。$$

所以 β_1 的水平为 95% 的置信区间为 $[11.818,13.283]$ 。

$$t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}}=2.5706\times0.20782\times\sqrt{\frac{1}{7}+\frac{0.5428571^2}{0.5321429}}=0.44589,$$

$$\underline{\theta}=\hat{\beta}_0-t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}}=19.95839-0.44589=13.513,$$

$$\overline{\theta}=\hat{\beta}_0+t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}}=19.95839+0.44589=14.404。$$

所以 β_0 的水平为 95% 的置信区间为 $[7.165,8.495]$ 。

（4）用 t 分布检验：

$$T=\frac{\hat{\beta}_1}{\hat{\sigma}}\sqrt{L_{xx}}=\frac{12.55034}{0.20782}\sqrt{0.5321429}=44.054\ 。$$

对 $\alpha=0.05$，查 t 分布的分位数表，可得 $t_{1-\alpha/2}(n-2)=t_{0.975}(5)=2.5706$，因为 $|T|=|44.054|=44.054>2.5706$，所以拒绝 H_0：$\beta_1=0$ ，说明自变量 x 与因变量 y 之间有显著的统计线性相关关系。

用 F 分布检验：

$$F=\frac{L_{yy}-SS_e}{SS_e/(n-2)}=\frac{84.03429-0.21594}{0.21594/(7-2)}=1940.8\ 。$$

对 $\alpha=0.05$ ，查 F 分布的分位数表，可得 $F_{1-\alpha}(1,n-2)=F_{0.95}(1,5)=6.61$ ，因为 $F=1940.8>6.61$ ，所以结论也是拒绝 H_0：$\beta_1=0$。

5.6 在一系列不同温度 x（单位：℃）下，观测硝酸钠在 100ml 水中溶解的重量 y（单位：g），得数据如下：

温度 x_i	0	4	10	15	21	29	36	51	68
重量 y_i	66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

设有 $y_i=\beta_0+\beta_1x_i+\varepsilon_i$ ，$\varepsilon_i\sim N(0,\sigma^2)$，$i=1,2,\cdots,9$，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_9$ 相互独立。

（1）求 β_0,β_1 的最小二乘估计 $\hat{\beta}_0,\hat{\beta}_1$ ；

（2）残差平方和 SS_e ，估计的标准差 $\hat{\sigma}$ ，样本相关系数 r ；

（3）检验 H_0： $\beta_1 = 0$ （显著水平 $\alpha = 0.05$） 。

解 $n = 9$ ， $\bar{x} = 26$ ， $L_{xx} = 4060$ ， $\bar{y} = 90.1444$ ， $L_{yy} = 3083.98$ ，

$$L_{xy} = \sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 24628.6 - 9 \times 26 \times 90.1444 = 3534.8 \text{ 。}$$

（1） $\hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}} = \dfrac{3534.8}{4060} = 0.870640$ ，

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 90.1444 - 0.870640 \times 26 = 67.5078 \text{ 。}$$

所以，回归方程为 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 67.5078 + 0.870640x$ 。

（2） $SS_e = L_{yy} - \hat{\beta}_1 L_{xy} = 3083.98 - 0.870640 \times 3534.8 = 6.4426$ ，

$$\hat{\sigma} = \sqrt{\frac{SS_e}{n-2}} = \sqrt{\frac{6.4426}{9-2}} = 0.95936 \text{ ，}$$

$$r = \frac{L_{xy}}{\sqrt{L_{xx} L_{yy}}} = \frac{3534.8}{\sqrt{4060 \times 3083.98}} = 0.99895 \text{ 。}$$

（3）用 t 分布检验：

$$T = \frac{\hat{\beta}_1}{\hat{\sigma}} \sqrt{L_{xx}} = \frac{0.870640}{0.95936} \sqrt{4060} = 57.826 \text{ 。}$$

对 $\alpha = 0.05$，查 t 分布的分位数表，可得 $t_{1-\alpha/2}(n-2) = t_{0.975}(7) = 2.3646$，因为 $|T| = |57.826| = 57.826 > 2.3646$，所以拒绝 H_0： $\beta_1 = 0$ ，说明自变量 x 与因变量 y 之间有显著的统计线性相关关系。

用 F 分布检验：

$$F = \frac{L_{yy} - SS_e}{SS_e/(n-2)} = \frac{3083.98 - 6.4426}{6.4426/(9-2)} = 3343.8 \text{ 。}$$

对 $\alpha = 0.05$ ，查 F 分布的分位数表，可得 $F_{1-\alpha}(1, n-2) = F_{0.95}(1,7) = 5.59$ ，因为 $F = 3343.8 > 5.59$ ，所以结论也是拒绝 H_0： $\beta_1 = 0$。

5.7 设 $y_i=\beta_0+\beta_1x_i+\varepsilon_i$ ，$\varepsilon_i\sim N(0,\sigma^2)$，$i=1,2,\cdots,n$，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 相互独立，$\hat{\beta}_0,\hat{\beta}_1$ 是 β_0,β_1 的最小二乘估计。证明：$\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)=0$ 的充分必要条件是 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=0$。

证 因为

$$\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)=\mathrm{Cov}(\bar{y}-\hat{\beta}_1\bar{x},\hat{\beta}_1)=\mathrm{Cov}(\bar{y},\hat{\beta}_1)-\mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_1)\bar{x}$$

$$=0-D(\hat{\beta}_1)\bar{x}=-\frac{\sigma^2}{L_{xx}}\bar{x}\ 。$$

（其中用到 **定理 5.2** $\mathrm{Cov}(\bar{y},\hat{\beta}_1)=0$ 和 **定理 5.1** $D(\hat{\beta}_1)=\frac{\sigma^2}{L_{xx}}$ ）

所以，当 $\bar{x}=0$ 时，有 $\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)=-\frac{\sigma^2}{L_{xx}}\bar{x}=0$；

反过来，由于$\sigma>0$，所以当 $\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)=-\frac{\sigma^2}{L_{xx}}\bar{x}=0$ 时，必有 $\bar{x}=0$。

5.8 具有重复试验的一元线性回归是指自变量 x 的每个不同取值 $x=x_i$ 都对因变量 y 作 m_i 次重复观测，记观测值为 $y_{i1},y_{i2},\ldots,y_{im_i}$，设 x 有 r 个观测值 $x_1,x_2,\ldots,x_r$，而 $\sum_{i=1}^{r}m_i=n$，于是重复试验的一元线性回归模型可表示为 $y_{ij}=\alpha+\beta x_i+\varepsilon_{ij}$，其中 $i=1,2,\ldots,r;\ j=1,2,\ldots,m_i;\ \varepsilon_{ij}\sim N(0,\sigma^2)$，试求$\alpha$和$\beta$的最小二乘估计.

解

$$\bar{x}=\frac{1}{n}\left(\underbrace{x_1+\ldots+x_1}_{m_1}+\underbrace{x_2+\ldots+x_2}_{m_2}+\ldots+\underbrace{x_r+\ldots+x_r}_{m_r}\right)=\frac{1}{n}\sum_{i=1}^{r}m_ix_i$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{m_i}y_{ij},\ \overline{x^2}=\frac{1}{n}\sum_{i=1}^{r}m_ix_i^2,\ \overline{xy}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{m_i}x_iy_{ij}$$

于是：$\hat{\beta}=\frac{L_{xy}}{L_{xx}}=\frac{\frac{1}{n}L_{xy}}{\frac{1}{n}L_{xx}}=\frac{\overline{xy}-\bar{x}\bar{y}}{\overline{x^2}-(\bar{x})^2}$；$\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}$ 。

5.9 设 $y_i=\beta_0+\beta_1x_i+\beta_2(3x_i^2-2)+\varepsilon_i$，$\varepsilon_i\sim N(0,\sigma^2)$，$i=1,2,3$，$\varepsilon_1,\varepsilon_2,\varepsilon_3$ 相互独立，$x_1=-1$，$x_2=0$，$x_3=1$。

（1）写出矩阵 X，X^TX 和 $(X^TX)^{-1}$；

（2）求 β_0,β_1,β_2 的最小二乘估计；

（3）证明 $\beta_2=0$ 时，β_0,β_1 的最小二乘估计与 $\beta_2\neq 0$ 时的最小二乘估计相同。

解 （1）$X=\begin{bmatrix}1&-1&1\\1&0&-2\\1&1&1\end{bmatrix}$，$X^TX=\begin{bmatrix}3&0&0\\0&2&0\\0&0&6\end{bmatrix}$，$(X^TX)^{-1}=\begin{bmatrix}1/3&0&0\\0&1/2&0\\0&0&1/6\end{bmatrix}$。

（2）$$\begin{bmatrix}\hat\beta_0\\\hat\beta_1\\\hat\beta_2\end{bmatrix}=\hat\beta=(X^TX)^{-1}X^TY$$

$$=\begin{bmatrix}1/3&0&0\\0&1/2&0\\0&0&1/6\end{bmatrix}\begin{bmatrix}1&1&1\\-1&0&1\\1&-2&1\end{bmatrix}\begin{bmatrix}y_1\\y_1\\y_3\end{bmatrix}=\begin{bmatrix}\dfrac{y_1+y_2+y_3}{3}\\\dfrac{-y_1+y_3}{2}\\\dfrac{y_1-2y_2+y_3}{6}\end{bmatrix}。$$

即有

$$\hat\beta_0=\frac{y_1+y_2+y_3}{3}，\hat\beta_1=\frac{-y_1+y_3}{2}，\hat\beta_2=\frac{y_1-2y_2+y_3}{6}。$$

（3）**（证法一）** $\beta_2=0$ 时，模型成为 $y_i=\beta_0+\beta_1x_i+\varepsilon_i$，$\varepsilon_i\sim N(0,\sigma^2)$，$i=1,2,3$，

$$X=\begin{bmatrix}1&-1\\1&0\\1&1\end{bmatrix}，X^TX=\begin{bmatrix}3&0\\0&2\end{bmatrix}，(X^TX)^{-1}=\begin{bmatrix}1/3&0\\0&1/2\end{bmatrix}，$$

$$\begin{bmatrix}\hat\beta_0\\\hat\beta_1\end{bmatrix}=(X^TX)^{-1}X^TY=\begin{bmatrix}1/3&0\\0&1/2\end{bmatrix}\begin{bmatrix}1&1&1\\-1&0&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{bmatrix}\dfrac{y_1+y_2+y_3}{3}\\\dfrac{-y_1+y_3}{2}\end{bmatrix}，$$

即有

$$\hat\beta_0=\frac{y_1+y_2+y_3}{3}，\hat\beta_1=\frac{-y_1+y_3}{2}，$$

β_0,β_1 的最小二乘估计与 $\beta_2\neq 0$ 时的最小二乘估计相同。

（证法二） $\beta_2 = 0$ 时，模型成为 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$，$\varepsilon_i \sim N(0, \sigma^2)$，$i = 1, 2, 3$，按照一元线性回归的计算公式，有

$$\bar{x} = \frac{-1+0+1}{3} = 0 \quad ， \quad L_{xx} = (-1-0)^2 + (0-0)^2 + (1-0)^2 = 2 \quad ，$$

$$\bar{y} = \frac{y_1 + y_2 + y_3}{3} \quad ， \quad L_{xy} = \sum_{i=1}^{3} x_i y_i - n\bar{x}\,\bar{y} = (-1)\cdot y_1 + 0 \cdot y_2 + 1 \cdot y_3 = -y_1 + y_3 \quad ，$$

$$\hat{\beta}_1 = \frac{L_{xy}}{L_{xx}} = \frac{-y_1 + y_3}{2} \quad ， \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \bar{y} = \frac{y_1 + y_2 + y_3}{3} \quad 。$$

β_0, β_1 的最小二乘估计与 $\beta_2 \neq 0$ 时的最小二乘估计相同。

5.10 为了考察某种植物的生长量 y（单位：mm）与生长期的日照时间 x_1（单位：小时）以及气温 x_2（单位：℃）的关系，测得数据如下：

日照时间 x_{1i}	269	281	262	275	278	282	268	259	275	255
气 温 x_{2i}	30.1	28.7	29.0	26.8	26.8	30.7	22.9	26.0	27.3	30.3
生长量 y_i	122	131	116	111	117	137	111	108	119	108

日照时间 x_{1i}	272	273	274	273	284	262	285	278	272	279
气 温 x_{2i}	26.5	29.8	28.3	24.4	30.1	24.9	25.6	24.9	24.8	30.7
生长量 y_i	125	132	136	128	138	76	130	127	123	133

设 $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$ ，$\varepsilon_i \sim N(0, \sigma^2)$，$i = 1, 2, \cdots, 20$ ，$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{20}$ 相互独立。

求：（1）$\beta_0, \beta_1, \beta_2$ 的最小二乘估计 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ ；

（2）残差平方和 SS_e ，估计的标准差 $\hat{\sigma}$ ，多重相关系数 r ；

（3）检验 H_0：$\beta_1 = \beta_2 = 0$ （显著水平 $\alpha = 0.05$） ；

（4）分别检验 H_{01}：$\beta_1 = 0$ 和 H_{02}：$\beta_2 = 0$（显著水平 $\alpha = 0.05$） 。

解 利用可作多元线性回归的计算机软件，求得：

（1）$\hat{\beta}_0 = -247.867$ ，$\hat{\beta}_1 = 1.15423$ ，$\hat{\beta}_2 = 1.98298$，所以，回归方程为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = -247.867 + 1.15423 x_1 + 1.98298 x_2 \text{ 。}$$

（2）残差平方和 $SS_e = 1542.59$ ，估计的标准差 $\hat{\sigma} = 9.5258$ ，多重相关系数 $r = 0.77921$ 。

（3）检验 H_0：$\beta_1 = \beta_2 = 0$ 的统计量 $F = 13.14$ ，对$\alpha = 0.05$，查F 分布表，可得分位数$F_{1-\alpha}(m, n-m-1) = F_{0.95}(2,17) = 3.59$，因为 $F = 13.14 > 3.59$ ，所以拒绝 H_0：$\beta_1 = \beta_2 = 0$ ，说明自变量 x_1, x_2 与因变量 y 之间有显著的统计线性相关关系。

（4）检验H_{01}：$\beta_1 = 0$ 的统计量 $F_1 = 18.96$ ，对 $\alpha = 0.05$，查 F 分布表，可得分位数 $F_{1-\alpha}(1, n-m-1) = F_{0.95}(1,17) = 4.45$，因为 $F_1 = 18.96 > 4.45$ ，所以拒绝 H_{01}：$\beta_1 = 0$，说明自变量 x_1 与因变量 y 统计线性相关 ；

检验H_{02}：$\beta_2 = 0$ 的统计量 $F_2 = 4.74$ ，对 $\alpha = 0.05$，查 F 分布表，可得分位数 $F_{1-\alpha}(1, n-m-1) = F_{0.95}(1,17) = 4.45$，因为 $F_2 = 4.74 > 4.45$ ，所以拒绝 H_{02}：$\beta_2 = 0$，说明自变量 x_2 也与因变量 y 统计线性相关 。

5.11 多元线性回归模型中，若先根据变量y和$x_i\ (i = 1, 2, \ldots, m)$的观测值 $(y_1, y_2, \ldots, y_n)^T$ 和$(x_{1i}, x_{2i}, \ldots, x_{ni})^T$ 对变量“标准化”，即令

$x_i^* = \dfrac{x_i - \bar{x}_i}{\sqrt{L_{ii}}}\ (i = 1, 2, \ldots, m)$；$y^* = \dfrac{y - \bar{y}}{\sqrt{L_{yy}}}$；$\hat{y}^* = \dfrac{\hat{y} - \bar{y}}{\sqrt{L_{yy}}}$，其中

$L_{ii} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2$；$\bar{x}_i = \dfrac{1}{n}\sum_{k=1}^{n} x_{ki}$；$L_{yy} = \sum_{k=1}^{n} (y_i - \bar{y})$.

此时再求y^*关于$x_i^*(i = 1, 2, \cdots, m)$的回归称为标准回归.

（1）证明标准回归方程的常数项为零，即 $\hat{y}^* = \sum_{i=1}^{m} d_i x_i^*$

（2）证明标准回归的总离差平方和 $\tilde{SS}_T = \sum_{i=1}^{n} (y_i^* - \bar{y}^*)^2 = 1.$

证明（1）设 y 关于 $x_1, x_2, \ldots, x_m$ 的线性回归方程为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_m x_m$$

于是有

$$\frac{\hat{y} - \bar{y}}{\sqrt{L_{yy}}} = \frac{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_m x_m - \bar{y}}{\sqrt{L_{yy}}}$$

即

$$\hat{y}^* = \frac{\hat{\beta}_0 + \hat{\beta}_1 \left(x_1^* \sqrt{L_{11}} + \bar{x_1} \right) + \ldots + \hat{\beta}_m \left(x_m^* \sqrt{L_{mm}} + \bar{x_m} \right) - \bar{y}}{\sqrt{L_{yy}}}$$

$$= \frac{\hat{\beta}_0 + \hat{\beta}_1 \bar{x_1} + \ldots + \hat{\beta}_m \bar{x_m} - \bar{y}}{\sqrt{L_{yy}}} + \hat{\beta}_1 \sqrt{\frac{L_{11}}{L_{yy}}} x_1^* + \ldots + \hat{\beta}_m \sqrt{\frac{L_{mm}}{L_{yy}}} x_m^*$$

$$= 0 + d_1 x_1^* + \ldots + d_m x_m^*$$

注：上式用到 $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x_1} + \ldots + \hat{\beta}_m \bar{x_m}$

此外，从上式还可得 $d_i = \hat{\beta}_i \sqrt{\frac{L_{ii}}{L_{yy}}}, i = 1, 2, \ldots, m$

（2）$\tilde{SS}_T = \sum_{i=1}^{n} \left(y_i^* - \overline{y^*} \right)^2 = \sum_{i=1}^{n} \left(y_i^* - 0 \right)^2 = \sum_{i=1}^{n} \left(\frac{y_i - \bar{y}}{\sqrt{L_{yy}}} \right)^2$

$$= \frac{1}{L_{yy}} \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 = \frac{L_{yy}}{L_{yy}} = 1 。$$

5.12 在不同的温度 x（单位：℃）下，观察平均每只红铃虫的产卵数 y（单位：个），得到数据如下：

温度 x_i	21	23	25	27	29	32	35
产卵数 y_i	7	11	21	24	66	115	325

设产卵数 y 与温度 x 之间，近似有下列关系：

$$y = \alpha \, \mathrm{e}^{\beta x} \quad ,$$

求常系数 α, β 的估计值。

解 回归方程为 $\hat{y} = \hat{\alpha}\, \mathrm{e}^{\hat{\beta} x}$ ，对方程两边同时取对数，得到

$$\ln \hat{y} = \ln \hat{\alpha} + \hat{\beta}\, x \ ,$$

令 $y^* = \ln y$， $\beta_0 = \ln \alpha$，它就化成了一个一元线性回归方程

$$\hat{y}^* = \hat{\beta}_0 + \hat{\beta}\, x \ 。$$

求得 β_0, β 的估计 $\hat{\beta}_0, \hat{\beta}$ 后，α 的估计，可以通过 $\hat{\alpha} = \mathrm{e}^{\hat{\beta}_0}$ 求得。

作为广义线性回归求解，在不加权的情况下，用计算机软件解得：

$$\hat{\beta}_0 = -3.849175\,, \ \hat{\beta} = 0.272026 \ , \ SS_e = 1537.66 \ ,$$

$$\hat{\alpha} = \mathrm{e}^{\hat{\beta}_0} = 0.0212973 \ ;$$

作为广义线性回归求解，在加权的情况下，用计算机软件解得：

$$\hat{\alpha} = 0.0100311 \ , \ \hat{\beta} = 0.296470 \ , \ SS_e = 506.640 \ ;$$

作为非线性回归求解，用计算机软件解得：

$$\hat{\alpha} = 0.00695936 \ , \ \hat{\beta} = 0.306934 \ , \ SS_e = 472.047 \ 。$$

5.13 某零件上有一条曲线，可以近似看作是一条抛物线 $y = \beta_0 + \beta_1 x + \beta_2 x^2$ 。为了在数控机床上加工这一零件，在曲线上测得 11 个点的坐标 (x_i, y_i) 数据如下：

x_i	0	2	4	6	8	10	12	14	16	18	20
y_i	0.6	2.0	4.4	7.5	11.8	17.1	23.3	31.2	39.6	49.7	61.7

求这条抛物线的函数表达式。

解 回归方程为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \ 。$$

令 $x_1 = x$， $x_2 = x^2$ ，原来的回归方程化成了下列形式：

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \ 。$$

利用可作多元线性回归的计算机软件，求得：

$$\hat{\beta}_0 = 1.01049 \ , \ \hat{\beta}_1 = 0.197110 \ , \ \hat{\beta}_2 = 0.140326 \ , \ SS_e = 1.23134 \ 。$$

5.14 猪的毛重 W（单位：kg）与它的身长 L（单位：cm），肚围 R（单位：cm）之间，近似有下列关系：

$$W = \alpha\, L^{\beta_1} R^{\beta_2} \quad ,$$

其中，$\alpha,\ \beta_1, \beta_2$ 都是常系数。现在对 14 头猪，测得它们的身长、肚围和毛重数据如下：

身长 L_i	41	45	51	52	59	62	69	72	78	80	90	92	98	103
肚围 R_i	49	58	62	71	62	74	71	74	79	84	85	94	91	95
毛重 W_i	28	39	41	44	43	50	51	57	63	66	70	76	80	84

求常系数 α, β_1, β_2 的估计值。

解 回归方程为 $\hat{W} = \hat{\alpha}\, L^{\hat{\beta}_1} R^{\hat{\beta}_2}$ ，对方程两边同时取对数，得到

$$\ln\hat{W} = \ln\hat{\alpha} + \hat{\beta}_1 \ln L + \hat{\beta}_2 \ln R \quad ,$$

令 $y = \ln W$， $\beta_0 = \ln\alpha$， $x_1 = \ln L$， $x_2 = \ln W$，它就化成了一个多元线性回归方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \quad 。$$

求得 β_1, β_2 的估计 $\hat{\beta}_1, \hat{\beta}_2$ 后，α 的估计，可以通过 $\hat{\alpha} = \mathrm{e}^{\hat{\beta}_0}$ 求得。

作为广义线性回归求解，在不加权的情况下，用计算机软件解得：

$\hat{\alpha} = 0.157551$ ， $\hat{\beta}_1 = 0.526691$ ， $\hat{\beta}_2 = 0.840853$ ， $SS_e = 33.1720$ ；

作为广义线性回归求解，在加权的情况下，用计算机软件解得：

$\hat{\alpha} = 0.196052$ ， $\hat{\beta}_1 = 0.609360$ ， $\hat{\beta}_2 = 0.709291$ ， $SS_e = 31.4119$ ；

作为非线性回归求解，用计算机软件解得：

$\hat{\alpha} = 0.189740$ ， $\hat{\beta}_1 = 0.611679$ ， $\hat{\beta}_2 = 0.714209$ ， $SS_e = 31.2833$ 。

5.15 热敏电阻器的电阻 y（单位：Ω）与温度 x（单位：℃）之间，近似有下列关系：

$$y = \alpha \exp\left(\frac{\beta}{x+\gamma}\right) \quad ,$$

其中，α, β, γ 都是常系数。现对 16 个热敏电阻器，测得温度 x 和电阻 y 的数据如下：

温度 x_i	50	55	60	65	70	75	80	85
电阻 y_i	34780	28610	23650	19630	16370	13720	11540	9744

温度 x_i	90	95	100	105	110	115	120	125
电阻 y_i	8266	7030	6005	5147	4427	3820	3307	2872

求常系数 α, β, γ 的估计值。

解 利用可作非线性回归的计算机软件，求得：

$$\hat{\alpha} = 0.00561861 \ ， \ \hat{\beta} = 6180.32 \ ， \ \hat{\gamma} = 345.199 \ ， \ SS_e = 100.694 \ 。$$

5.16 对某种蔬菜的生长期 x（单位：日）和平均每株蔬菜的质量 y（单位：g）进行观测，得到一组数据如下：

生长期 x_i	9	14	21	28	42	57	63	70	79
重量 y_i	8.93	10.80	18.59	22.33	39.35	56.11	61.73	64.62	67.08

设生长期 x 与平均每株蔬菜的质量 y 之间，近似有下列关系：

$$y = \frac{\alpha}{1 + \beta \, \mathrm{e}^{-\gamma x}} \quad ，$$

求常系数 α, β, γ 的估计值。

解 利用可作非线性回归的计算机软件，求得：

$$\hat{\alpha} = 72.4622 \ ， \ \hat{\beta} = 13.7093 \ ， \ \hat{\gamma} = 0.0673592 \ ， \ SS_e = 8.05652 \ 。$$