第四章作业

2. 为确定广告费用 x 和销售额 y 的关系，现作一统计，得资料如表 10-8.

表 10-8

x(万元)	40	25	20	30	40	40	25	20	50	20	50	50
y(万元)	490	395	420	475	385	525	480	400	560	365	510	540

(1) 求销售额 y 对广告费 x 的线性回归方程；

(2) 对于显著性水平 $\alpha=0.05$，检验假设 $H_0: b=0$，其中 b 是回归直线的斜率；

(3) 当广告费 $x=35$ 时，求销售额 y 的预测值和预测区间（取 $\alpha=0.05$）.

解：(1) 设 x 与 y 之间的回归方程为

$$y = a + bx$$

$n = 12 \quad \sum_{i=1}^{n} x_i^2 = 15650 \quad \sum_{i=1}^{n} y_i^2 = 2610925$

$\sum_{i=1}^{n} x_i y_i = 196325 \quad \bar{x} = 34.167 \quad \bar{y} = 462.083$

$L_{xx} = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2 = 1641.667$

$L_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = 6070.833$

于是 $\hat{b} = \frac{L_{xx}}{L_{xy}} = 0.239 \quad \hat{a} = \bar{y} - \hat{b}\bar{x} = 453.920$

故回归方程为

$$\hat{y} = 453.920 + 0.239\hat{x}$$

(2) 取检验统计量

$$F_0 = \frac{S_R}{S_Q/n-2} \sim F(1, n-2)$$

$S_R = \sum_{i=1}^{n} (\hat{y} - \bar{y})^2 = \hat{b}^2 L_{xx}$

$S_Q = Q_{min}(a, b) = \sum_{i=1}^{n} (y - \hat{y})^2 = L_{yy} - \hat{b}^2 L_{xx}$

$\therefore F_0 = 0.0193$

$F_\alpha(1, n-2) = F_{0.05}(1, 10) = 4.964 > F_0$

故接受H_0，认为$b=0$

(3) $x_0=35$，代入 $\hat{y}=453.920+0.239\hat{x}$

得 $\hat{y}=462.285$

由(1)知 $\bar{x}=34.167$ $L_{xx}=1641.667$

$\hat{\sigma}=\sqrt{\frac{Q_{min}(a,b)}{n-2}}=\sqrt{\frac{L_{yy}-\hat{b}L_{xy}}{n-2}}=68.579$

$\delta_n=\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{L_{xx}}}=1.041$

$t_{0.025}(10)=2.2281$

于是$\hat{y}$的95%预测区间为

$(\hat{y}-\delta_n\hat{\sigma}t_{\frac{\alpha}{2}}(n-2),\ \hat{y}+\delta_n\hat{\sigma}t_{\frac{\alpha}{2}}(n-2))=(303.208,\ 621.351)$

4-1 设

$$\begin{cases} y_1=a+\varepsilon_1, \\ y_2=2a-b+\varepsilon_2, \\ y_3=a+2b+\varepsilon_3, \end{cases} \quad \varepsilon=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix}\sim N_3(0,\sigma^2 I_3).$$

(1) 试求参数a,b的最小二乘估计；

解：依题意得

$$Y=\begin{pmatrix}y_1\\ y_2\\ y_3\end{pmatrix}=\begin{pmatrix}1 & 0\\ 2 & -1\\ 1 & 2\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}+\varepsilon=X\beta+\varepsilon$$

由最小二乘估计

$Q(\beta)=\sum\varepsilon_i^2=\varepsilon'\varepsilon=(Y-X\beta)'(Y-X\beta)$

取Q_{min}时 $\hat{\beta}=(X'X)^{-1}X'Y$

$$=\left[\begin{pmatrix}1&2&1\\0&-1&2\end{pmatrix}\begin{pmatrix}1&0\\2&-1\\1&2\end{pmatrix}\right]^{-1}\begin{pmatrix}1&2&1\\0&-1&2\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{6}&0\\0&\frac{1}{5}\end{pmatrix}\begin{pmatrix}y_1+2y_2+y_3\\-y_2+2y_3\end{pmatrix}=\begin{pmatrix}\frac{1}{6}(y_1+2y_2+y_3)\\\frac{1}{5}(-y_2+2y_3)\end{pmatrix}$$

4-3 设 Y 与 x_1,x_2,x_3 有相关关系，其 8 组观测数据见表 4.5.

表 4.5 观测数据

序号	x_1	x_2	x_3	Y
1	38	47.5	23	66.0
2	41	21.3	17	43.0
3	34	36.5	21	36.0
4	35	18.0	14	23.0
5	31	29.5	11	27.0
6	34	14.2	9	14.0
7	29	21.0	4	12.0
8	32	10.0	8	7.6

(1) 设 $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon$，试求回归方程及决定系数 R^2 和均方误差 s^2；

解：依题意得

$$Y=\begin{bmatrix}66.0\\43.0\\36.0\\23.0\\27.0\\14.0\\12.0\\7.6\end{bmatrix}\quad X=\begin{bmatrix}1&38&47.5&23\\1&41&21.3&17\\1&34&36.5&21\\1&35&18.0&14\\1&31&29.5&11\\1&34&14.2&9\\1&29&21.0&4\\1&32&10.0&8\end{bmatrix}\quad \beta=\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\\\beta_3\end{bmatrix}\quad \varepsilon=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\\\varepsilon_6\\\varepsilon_7\\\varepsilon_8\end{bmatrix}$$

由python程序计算 $(X'X)$，$X'Y$

得 $\hat{\beta}=(X'X)^{-1}X'Y$

```
/Users/fanglunlin/Desktop/pythonPro/Statistics/.venv/bin/python /Users/fanglunlin/Desktop/pythonPro/Statistics/回归分析/多元回归分析.py
X_X1 =  [[ 3.06337137e+01 -1.00399334e+00 -2.21871186e-01  7.00513651e-01]
 [-1.00399334e+00  3.37616119e-02  7.04586031e-03 -2.44281802e-02]
 [-2.21871186e-01  7.04586031e-03  3.73119325e-03 -8.35862148e-03]
 [ 7.00513651e-01 -2.44281802e-02 -8.35862148e-03  2.56469086e-02]]

X_Y =  [ 228.6 8204.2 7102.2 3858.8]
beta = [-106.72667694    3.25178616    1.3313261    -0.67456599]


Process finished with exit code 0
```

故回归方程为

$$\hat{Y} = \hat{X}\hat{\beta}$$

$$TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = Y'Y - n\bar{Y}^2$$

$$(MSS)\ S^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{\beta}'X'Y - n\bar{Y}^2$$

$$R^2 = \frac{S^2}{MSS} = 0.991$$

```
TSS = 2624.5150000000003, MSS = 2600.6696796529586

R =  0.9909143897645691

Process finished with exit code 0
```

4-10 考虑洛河在某河段河水受污染情况. 考察的指标(因变量)有两个, Y_1 表示 BOD 浓度; ~~Y_2 表示氧亏浓度~~. 而 Y_1, Y_2 又与以下几个因素(自变量)有关:

x_1——初始断面的 BOD 浓度 L_0,

x_2——初始断面的氧亏浓度 C_0,

x_3——水温 T,

x_4——河流流量 Q,

x_5——排污口流量 g,

x_6——污水 BOD 浓度 l,

x_7——流过该河段所需时间 t.

共观测了 15 组数据(见表 4.6), 试用逐步回归或~~双重筛选逐步回归~~求出 BOD 浓度 Y_1、氧亏浓度 Y_2 与 $x_1, x_2, \cdots, x_7$ 的回归方程.

表 4.6 水污数据

序号	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y_1
1	6.88	−0.25	27.0	67.4784	1.1232	477.0	0.083	9.35
2	6.08	−2.21	27.5	47.7792	1.1232	193.0	0.083	12.30
3	2.14	−3.04	26.0	47.7792	1.1232	404.0	0.083	15.60
4	5.02	−0.73	26.0	85.6224	1.1232	363.0	0.073	5.88
5	7.89	−2.26	26.0	85.6224	1.1232	363.0	0.069	6.34
6	2.38	−1.65	15.0	149.0400	1.5552	428.0	0.104	4.00
7	1.86	−1.35	15.8	149.0400	1.5552	428.0	0.104	3.76
8	1.02	−2.12	17.1	149.4720	1.3824	428.0	0.104	3.98
9	1.22	−1.92	17.5	149.4720	1.3824	428.0	0.104	3.98
10	0.90	−0.27	17.0	362.8800	0.9936	202.0	0.104	2.78
11	2.58	−0.09	17.0	362.8800	0.9936	202.0	0.104	1.88
12	2.78	−1.17	13.5	326.5920	0.9936	114.0	0.104	2.56
13	2.10	−1.30	13.5	326.5920	0.9936	114.0	0.104	2.72
14	2.32	−0.60	14.5	364.6080	0.8640	57.3	0.104	1.64
15	2.96	−0.60	14.5	364.6080	0.8640	57.3	0.104	2.36

解：分别建立 $x_1, \cdots\cdots, x_7$ 与 Y_1 的一元回归方程

$$Y_1 = a_i + b_i x_i \quad (i = 1, \cdots, 7)$$

并计算7个回归方程中回归系数的统计量 F_i

x_1: F = 10.07960964317245

x_2: F = 0.08045105718331873

x_3: F = -16.452471634213293

x_4: F = -13.00000747454045

x_5: F = 0.011900484869357712

x_6: F = -13.000001110854177

x_7: F = 2.9902051236714423e-09

$|F_1|、|F_3|、|F_4|、|F_6| \geq F_{0.05}(1, n-2)$

故引入 $x_1、x_3、x_4、x_6$

$X = (1_n, x_1, x_3, x_4, x_6)' \qquad \beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)'$

$$Y_1 = X\beta + \varepsilon$$

$$\hat{\beta} = (X'X)^{-1}X'Y = (5.422,\ -0.874,\ 0.586,\ -0.0231,\ -0.0133)'$$

```
/Users/fanglunlin/Desktop/pythonPro/Statistics/.venv/bin/python /Users/fanglunlin/Desktop/pythonPro/Statistics/回归分析/多元回归分析.py
X_X1 =  [[ 1.28906512e+01  5.45622748e-02 -3.28159347e-01 -1.94590301e-02
  -9.38399262e-03]
 [ 5.45622748e-02  3.78602415e-02 -1.22194418e-02  5.72139494e-05
   1.64211366e-04]
 [-3.28159347e-01 -1.22194418e-02  1.29466133e-02  4.21296903e-04
   1.10912309e-04]
 [-1.94590301e-02  5.72139494e-05  4.21296903e-04  3.22560366e-05
   1.58448357e-05]
 [-9.38399262e-03  1.64211366e-04  1.10912309e-04  1.58448357e-05
   1.21223128e-05]]

X_Y =  [   79.13       308.4366    1779.636    10420.554528 26072.71    ]
beta = [ 5.422039   -0.87352522  0.58601865 -0.02314745 -0.01334948]

TSS = 234.69777333333332, MSS = 195.81515611490624

R =  0.8343289897207359
F =  8.833440379729808

Process finished with exit code 0
```