Assignment1

12111603谭致恒

2023-09-23

本次作业没有直接使用R语言中的结果（如lm函数等），所有结果（画图除外）均由手动计算器计算得出。

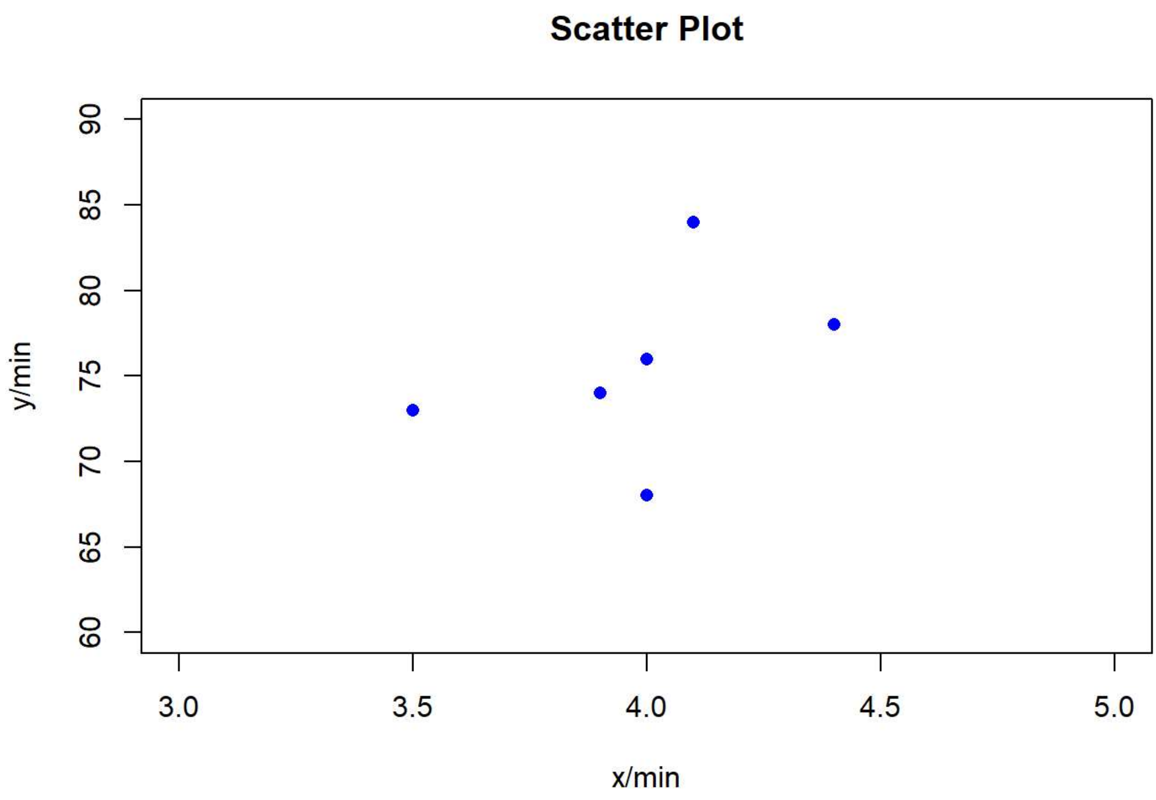
Q1

首先，我们计算出：

- $\bar{x} = \frac{23.9}{6}$
- $\bar{y} = 75.5$
- $S_{xx} = \sum(x_i - \bar{x})^2 \approx 0.428333$
- $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = 3.25$
- $S^2 = \frac{\sum(y_i - \hat{y_i})^2}{n-2} = \frac{1476.64}{4}$

(a)

```
x <- c(4.4,3.9,4,4,3.5,4.1)
y <- c(78,74,68,76,73,84)
plot(x, y,
     main = "Scatter Plot",        # 设置图的标题
     xlab = "x/min",        # 设置X轴标签
     ylab = "y/min",        # 设置Y轴标签
     pch = 16,                     # 设置点的形状
     col = "blue",                 # 设置点的颜色
     xlim = c(3, 5),               # 设置X轴的范围
     ylim = c(60, 90)                # 设置Y轴的范围
)
```



散点图显示出x和y之间有较强的线性关系。

(b)

由线性模型结论

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

和

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

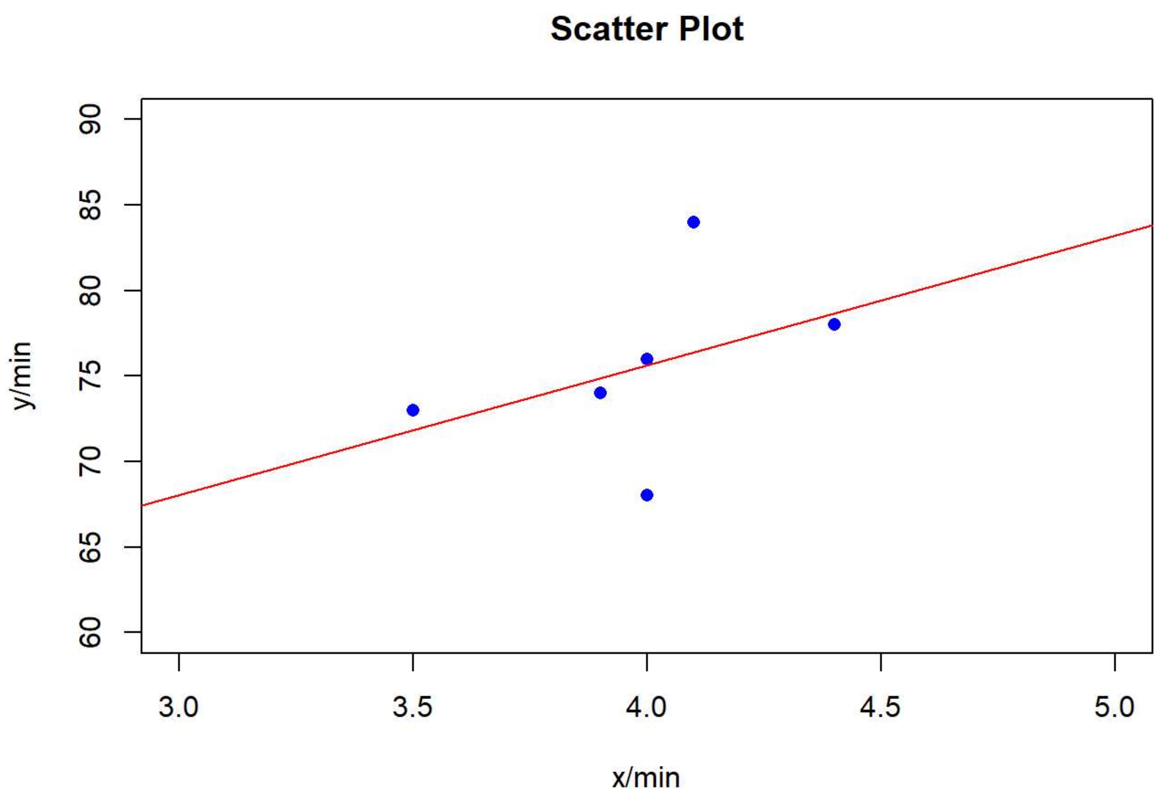
代入数据$\bar{x} = \frac{23.9}{6}$, $\bar{y} = 75.5$可以得出

$$\hat{\beta}_1 \approx 7.59,\ \hat{\beta}_0 \approx 45.27。$$

故回归方程为

$$\hat{y_i} = 7.59 x_i + 45.27$$

```
x <- c(4.4,3.9,4,4,3.5,4.1)
y <- c(78,74,68,76,73,84)
plot(x, y,
     main = "Scatter Plot",         # 设置图的标题
     xlab = "x/min",         # 设置X轴标签
     ylab = "y/min",         # 设置Y轴标签
     pch = 16,                      # 设置点的形状
     col = "blue",                  # 设置点的颜色
     xlim = c(3, 5),                # 设置X轴的范围
     ylim = c(60, 90)                 # 设置Y轴的范围
)
abline(a = 45.27, b = 7.59, col = "red")
```



(c)

线性模型为

$$Y_i = \beta_0 + \beta_1 X_i + e_i, i = 1, 2, \ldots, n$$

其中$\hat{\beta}_1 = 7.59$，$\hat{\beta}_0 = 45.27$。

所满足的假设为：

- y_i之间彼此独立。
- $Var(e_i) = \sigma^2$。
- $e_i \sim N(0, \sigma^2)$ (not necessary for obtaining LSE, but necessary for MLE)

(d)

建立假设检验

$$H_0\text{：} \beta_1 = 0 \leftrightarrow H_1 : \beta_1 \neq 0\text{。}$$

由检验统计量

$$\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} \sim t(n-2)\text{，其中}S^2 = \frac{\sum(y_i - \hat{y_i})^2}{n-2}$$

若$\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} > t(\frac{\alpha}{2}, n-2)$，则以显著水平$\alpha$拒绝原假设；反之则接受原假设。 代入$\beta_1 = 0$和$\alpha = 0.05$得

$$\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} = 0.911 < t(\frac{\alpha}{2}, n-2) = 2.7764$$

故不能拒绝原假设$H_0 : \beta_1 = 0$，即接受原假设。

(e)

因为

$$\frac{\hat{\beta}_1-\beta_1}{S/\sqrt{S_{xx}}}\sim t(n-2)\text{，其中}S^2=\frac{\sum(y_i-\hat{y_i})^2}{n-2}$$

故β_1的$(1-\alpha)100\%$的置信区间为

$$[\hat{\beta}_1-t(\alpha/2,n-2)\frac{S}{\sqrt{S_{xx}}},\ \hat{\beta}_1+t(\alpha/2,n-2)\frac{S}{\sqrt{S_{xx}}}]$$

代入数据可得β_1的95%置信区间为$[-15.53,30.71]$，即有95%的概率区间$[-15.53,30.71]$套住了β_1。

(f)

$$R^2=\frac{SSR}{SST}=\frac{S_{xy}^2}{S_{xx}S_{yy}}\text{且}R\text{的正负与}\hat{\beta}_1\text{相同}$$

代入数据得

$$R^2\approx 0.1718$$

即有17.18%的方差是可以被模型解释的。

(g)

在给定x_h时，

$$y_h=\beta_o+\beta_1x_h+e_h$$

我们用$\hat{y_h}=\hat{\beta}_0+\hat{\beta}_1x_h$去估计$y_h$，得到$\hat{y_h}=75.63$。

注意到：

$$y_h-\hat{y_h}\sim N(0,\sigma^2(1+\frac{1}{n}+\frac{(x_h-\bar{x})^2}{S_{xx}}))$$

因此有

$$\frac{y_h-\hat{y_h}}{S\sqrt{1+\frac{1}{n}+\frac{(x_h-\bar{x})^2}{S_{xx}}}}\sim t(n-2)$$

由此我们可以构造y_h的$(1-\alpha)100\%$置信区间为

$$[\hat{y_h}-t(\frac{\alpha}{2},n-2)S\sqrt{1+\frac{1}{n}+\frac{(x_h-\bar{x})^2}{S_{xx}}},\ \hat{y_h}+t(\frac{\alpha}{2},n-2)S\sqrt{1+\frac{1}{n}+\frac{(x_h-\bar{x})^2}{S_{xx}}}]$$

代入数据得[59.28, 91.98]

综上，y_h的预测值为75.63，y_h的95%置信区间为[59.28, 91.98]。

Q2

(a)

$y_i = \beta_0 + \beta_1 x_i + e_i$，其中$e_i \sim N(0, \sigma^2)$且$e_i$之间彼此独立，则有：

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

我们要进行求解线性回归模型：$\hat{y_i} = \hat{\beta_0} + \hat{\beta_1} x_i$，下面进行MLE。

似然函数为

$$L(\beta_0, \beta_1 . \sigma^2) = \prod Normal(y_i) = \prod(\frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\})$$

取对数后可得

$$l(\beta_0, \beta_1 . \sigma^2) = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

对各个变量求偏导等于零得：

$$\frac{\partial l}{\partial \beta_0} = \frac{1}{2\sigma^2} \sum 2(y_i - \beta_0 - \beta_1 x_i) = 0$$

即

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\frac{\partial l}{\partial \beta_1} = \frac{1}{2\sigma^2} \sum 2(y_i - \beta_0 - \beta_1 x_i) x_i = 0$$

即

$$\sum x_i y_i - n\bar{x}\beta_0 - \sum (x_i)^2 \beta_1 = 0$$

代入$\beta_0 = \bar{y} - \beta_1 \bar{x}$得

$$\beta_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum (y_i - \beta_0 - \beta_1 x_i)^2}{4\sigma^2} = 0$$

即

$$\sigma^2 = \frac{\sum (y_i - \beta_0 - \beta_1 x_i)^2}{n}$$

综上，

$$\beta_1^{MLE} = \frac{\sum s_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\beta_0^{MLE} = \bar{y} - \beta_1^{MLE} \bar{x}$$

$$\sigma_{MLE}^2 = \frac{\sum (y_i - \beta_0^{MLE} - \beta_1^{MLE} x_i)^2}{n}$$

(b)

$$\beta_1^{MLE} = \frac{\sum s_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \beta_1^{LSE}$$

$$\beta_0^{MLE} = \bar{y} - \beta_1^{MLE}\bar{x} = \bar{y} - \beta_1^{LSE}\bar{x} = \beta_0^{LSE}$$

$\sigma_{MLE}^2 = \dfrac{\sum(y_i - \beta_0^{MLE} - \beta_1^{MLE}x_i)^2}{n} = \dfrac{\sum(y_i - \hat{y_i})^2}{n} = \dfrac{n}{n-2}S^2$，其中$S^2 = \dfrac{\sum(y_i - \hat{y_i})^2}{n-2}$为$\sigma^2$的无偏估计量。