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Im函数等), 所有结果(画图除外)均由手动计算器计算得出。

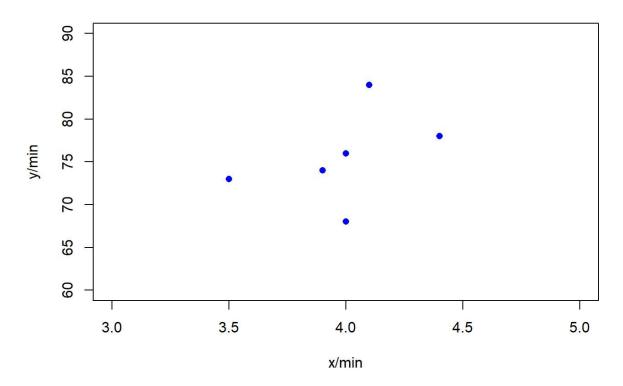
Q1

首先,我们计算出:

```
egin{aligned} ullet & ar{x} = rac{23.9}{6} \ ullet & ar{y} = 75.5 \ ullet & S_{xx} = \sum (x_i - ar{x})^2 pprox 0.428333 \ ullet & S_{xy} = \sum (x_i - ar{x})(y_i - ar{y}) = 3.25 \ ullet & S^2 = rac{\sum (y_i - \hat{y_i})^2}{n-2} = rac{1476.64}{4} \end{aligned}
```

(a)

Scatter Plot



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

$$\hat{eta}_1 = rac{S_{xy}}{S_{xx}} = rac{\sum (x_i - ar{x})(y_i - ar{y})}{\sum (x_i - ar{x})^2} = rac{\sum x_i y_i - n ar{x} ar{y}}{\sum x_i^2 - n ar{x}^2}$$

和

$$\hat{eta_0} = ar{y} - \hat{eta_1}ar{x}$$
 ,

代入数据 $ar{x}=rac{23.9}{6},\;ar{y}=75.5$ 可以得出

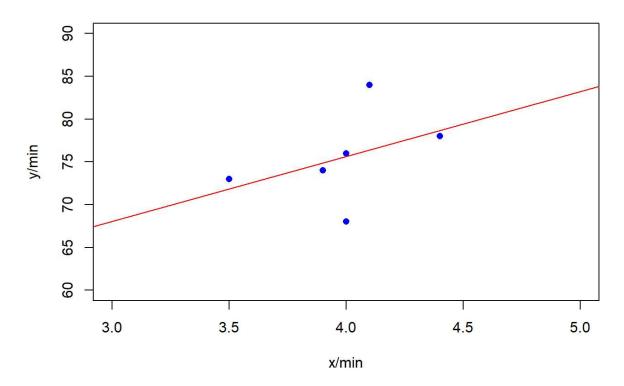
$$\hat{eta_1}pprox 7.59,~\hat{eta_0}pprox 45.27$$
 .

故回归方程为

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

```
x \leftarrow 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
    x1im = c(3, 5),
                               # 设置X轴的范围
    y1im = c(60, 90)
                                  # 设置Y轴的范围
abline(a = 45.27, b = 7.59, col = "red")
```

Scatter Plot



(c)

线性模型为

$$Y_i=eta_0+eta_1X_i+e_i, i=1,2,\ldots,n$$

其中 $\hat{eta_1}=7.59$, $\hat{eta_0}=45.27$ 。

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

$$H_0\colon eta_1=0 \leftrightarrow H_1:eta_1
eq 0$$
 .

由检验统计量

$$rac{\hat{eta}_1-eta_1}{S/\sqrt{S_{xx}}}\sim t(n-2)$$
,其中 $S^2=rac{\sum(y_i-\hat{y_i})^2}{n-2}$

若 $rac{eta_1-eta_1}{S/\sqrt{S_{xx}}}>t(rac{lpha}{2},n-2)$,则以显著水平lpha拒绝原假设;反之则接受原假设。 代入 $eta_1=0$ 和lpha=0.05得

$$rac{\hat{eta}_1 - eta_1}{S/\sqrt{S_{xx}}} = 0.911 < t(rac{lpha}{2}, n-2) = 2.7764$$

故不能拒绝原假设 $H_0:eta_1=0$,即接受原假设。

(e)

因为

$$rac{\hat{eta}_1-eta_1}{S/\sqrt{S_{xx}}}\sim t(n-2)$$
,其中 $S^2=rac{\sum(y_i-\hat{y_i})^2}{n-2}$

故 β_1 的 $(1-\alpha)100\%$ 的置信区间为

$$[\hat{eta_1}-t(lpha/2,n-2)rac{S}{\sqrt{S_{xx}}},~\hat{eta_1}+t(lpha/2,n-2)rac{S}{\sqrt{S_{xx}}}]$$

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)

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

代入数据得

$$R^2 pprox 0.1718$$

即有17.18%的方差是可以被模型解释的。

(g)

在给定 x_h 时,

$$y_h = \beta_o + \beta_1 x_h + e_h$$

我们用 $\hat{y_h}=\hat{eta_0}+\hat{eta_1}x_h$ 去估计 y_h ,得到 $\hat{y_h}=75.63$ 。

注意到:

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

因此有

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

由此我们可以构造 y_h 的 $(1-\alpha)100\%$ 置信区间为

$$[\hat{y_h} - t(rac{lpha}{2}, n-2)S\sqrt{1 + rac{1}{n} + rac{(x_h - ar{x})^2}{S_{xx}}}, \; \hat{y_h} + t(rac{lpha}{2}, n-2)S\sqrt{1 + rac{1}{n} + rac{(x_h - ar{x})^2}{S_{xx}}}]$$

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)

 $y_i=eta_0+eta_1x_i+e_i$,其中 $e_i\sim N(0,\sigma^2)$ 且 e_i 之间彼此独立,则有:

$$y_i \sim N(eta_0 + eta_1 x_i, \sigma^2)$$

我们要进行求解线性回归模型: $\hat{y_i} = \hat{eta_0} + \hat{eta_1} x_i$,下面进行MLE。

似然函数为

$$L(eta_0,eta_1.\sigma^2) = \prod Normal(y_i) = \prod (rac{1}{\sqrt{2\pi}\sigma}exp\{-rac{(y_i-eta_0-eta_1x_i)^2}{2\sigma^2}\})$$

取对数后可得

$$l(eta_0,eta_1.\,\sigma^2) = -rac{n}{2}ln(2\pi\sigma^2) - rac{1}{2\sigma^2}\sum (y_i - eta_0 - eta_1x_i)^2$$

对各个变量求偏导等于零得:

$$rac{\partial l}{\partial eta_0} = rac{1}{2\sigma^2} \sum 2(y_i - eta_0 - eta_1 x_i) = 0$$

即

$$eta_0 = ar{y} - eta_1 ar{x}$$
 $rac{\partial l}{\partial eta_1} = rac{1}{2\sigma^2} \sum 2(y_i - eta_0 - eta_1 x_i) x_i = 0$

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

$$eta_1 = rac{\sum x_i y_i - nar{x}ar{y}}{\sum x_i^2 - nar{x}^2} \ rac{\partial l}{\partial \sigma^2} = -rac{n}{2\sigma^2} + rac{\sum (y_i - eta_0 - eta_1 x_i)^2}{4\sigma^2} = 0$$

即

$$\sigma^2 = rac{\sum (y_i - eta_0 - eta_1 x_i)^2}{n}$$

综上,

$$eta_1^{MLE} = rac{\sum s_i y_i - nar{x}ar{y}}{\sum x_i^2 - nar{x}^2} \ eta_0^{MLE} = ar{y} - eta_1^{MLE}ar{x} \ eta_0^{2} = rac{\sum (y_i - eta_0^{MLE} - eta_1^{MLE} x_i)^2}{n}$$

(b)

$$eta_1^{MLE} = rac{\sum s_i y_i - nar{x}ar{y}}{\sum x_i^2 - nar{x}^2} = eta_1^{LSE}$$

$$eta_0^{MLE} = ar{y} - eta_1^{MLE}ar{x} = ar{y} - eta_1^{LSE}ar{x} = eta_0^{LSE}$$

$$\sigma_{MLE}^2 = rac{\sum (y_i - eta_0^{MLE} - eta_1^{MLE} x_i)^2}{n} = rac{\sum (y_i - \hat{y_i})^2}{n} = rac{n}{n-2} S^2$$
,其中 $S^2 = rac{\sum (y_i - \hat{y_i})^2}{n-2}$ 为 σ^2 的无偏估计量。