EXERCISE "COMPUTER REPAIR DATA"

(x_i, y_i) $i = 1, \dots, 14$ x_i = number of units to replace

y_i = minutes of intervention

a) we know that the MLE are

$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

$\hat{\beta}_2 = \frac{S_{xy}}{S^2_x}$

Hence, we need to compute S_{xy} and S^2_x (not given)

We have that $\sum_{i=1}^{14} (x_i - \bar{x})^2 = 114$ $\Rightarrow$ $S^2_x = \frac{1}{n-1} \sum_{i=1}^{14} (x_i - \bar{x})^2$

$= \frac{1}{13} \cdot 114 = 8.769$

$S_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

We do not have the quantities needed to compute it directly, however, we know that $R^2 = \rho^2_{xy}$ with ρ_{xy} correlation coefficient (only in the SIMPLE LM)

$\rho^2_{xy} = 0.984$ $\Rightarrow$ $|\rho_{xy}| = \sqrt{0.984} = 0.9919$

$|\rho_{xy}| = \frac{|S_{xy}|}{\sqrt{S^2_x \, S^2_y}} = \frac{|S_{xy}|}{\sqrt{8.769 \cdot 2392.951}} = 0.9919$ $S^2_y = \frac{1}{13} \cdot 31108.357 = 2392.951$

$\Rightarrow |S_{xy}| = 0.9919 \cdot \sqrt{20983.79} =$

$= 0.9919 \cdot 144.858 = 143.58$

It is reasonable to assume that if we have an increasing number of units to be repaired, the minutes of intervention increase $\rightarrow$ positive correlation

$\Rightarrow$ $S_{xy} = +143.58$

Now we have all the elements to compute $\hat{\beta}_2$

$\hat{\beta}_2 = \frac{143.58}{8.769} = 16.38$

and $\hat{\beta}_1 = 95.768 - 16.38 \cdot 6 = -2.51$

b) estimate of σ^2 is $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ this is biased

alternatively, the unbiased estimate is $s^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

We have $R^2 = 0.984$

and we know that $R^2 = \frac{SSR}{SST}$ where $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 31108.357$

$\Rightarrow$ $SSR = R^2 \cdot SST = 0.984 \cdot 31108.357 = 30610.62$

Finally, we know that $SST = SSE + SSR$ $\Rightarrow$ $SSE = 31108.357 - 30610.62$

$= 497.37$

where $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

we obtain $\hat{\sigma}^2 = \frac{497.37}{14} = 35.53$

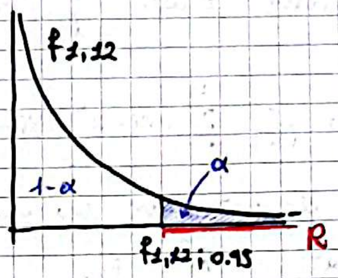
$s^2 = \frac{497.37}{12} = 41.47$

We need to test $\begin{cases} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{cases}$

the test statistic is

$$F = \frac{R^2}{1-R^2}(n-2) \overset{H_0}{\sim} F_{1,n-2} = F_{1,12}$$

its observed value is $f^{obs} = \frac{0.984}{1-0.984} \cdot 12 = 738$



the reject region is

$R = (f_{1,n-2;1-\alpha} ; +\infty)$

using a 5% significance level,

$R = (f_{1,12;0.95} ; +\infty) = (4.74 ; +\infty)$

$f^{obs} \in R$ we reject H_0: the model is useful for explaining the variability of y

c) significance of the coefficients

$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

the test statistic is $T = \frac{\hat{\beta}_1}{\sqrt{\hat{var}(\hat{\beta}_1)}} \overset{H_0}{\sim} t_{n-2} = t_{12}$

the observed value is

$t^{obs} = \frac{\hat{\beta}_1}{\sqrt{\hat{var}(\hat{\beta}_1)}} = \frac{-2.51}{4.014} = -0.6253$

the reject region is $R = (-\infty ; -t_{12;1-\frac{\alpha}{2}}) \cup (t_{12,1-\frac{\alpha}{2}} ; +\infty)$ if $\alpha = 0.05$

$= (-\infty ; -t_{12;0.975}) \cup (t_{12;0.975} ; +\infty)$

$= (-\infty ; -2.178) \cup (2.178 ; +\infty)$

$t^{obs} \notin R \Rightarrow$ we don't reject H_0 at a 5% level

$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$

test statistic $T = \frac{\hat{\beta}_2}{\sqrt{\hat{var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{12}$

the observed value of the test is

$$t^{obs} = \frac{\hat{\beta}_2}{\sqrt{\hat{var}(\hat{\beta}_2)}} = \frac{16.38}{0.604} = 27.119$$

the reject region is the same of the previous test,

$$R = (-\infty, -2.178) \cup (2.178, +\infty)$$

Here, however, $t^{obs} \in R$ $\Rightarrow$ we reject H_0

$\Rightarrow$ β_2 is significant

d) yes, the test about the significance of β_2 had already been performed in point b).

Indeed, testing $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$ is equivalent to testing $H_0: R^2 = 0$ vs $H_1: R^2 \neq 0$, in the simple linear model.

We can show it by noticing that

$$(t^{obs})^2 = 27.119^2 = 735.4 = f^{obs}$$ (up to numerical errors)

Scansionato con CamScanner

EXERCISE 4: BACTERIA MORTALITY DATA

$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$ iid

$n = 15$

a) t value refers to the observed value of the test statistic when testing

$$\begin{cases} H_0: \beta_r = 0 \\ H_1: \beta_r \neq 0 \end{cases} \qquad r = 1, 2.$$

We know that the test statistic is $T = \dfrac{\hat{\beta}_r - 0}{\sqrt{\hat{var}(\hat{\beta}_r)}} \overset{H_0}{\sim} t_{n-2} = t_{13}$

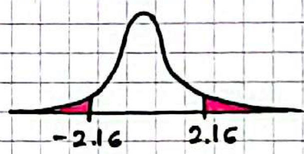
hence the observed value is

- for β_1: $\dfrac{\hat{\beta}_1}{\sqrt{\hat{va}(\hat{\beta}_1)}} = \dfrac{49.162}{22.76} = 2.16$

- for β_2: $\dfrac{\hat{\beta}_2}{\sqrt{\hat{va}(\hat{\beta}_2)}} = t^{obs} \Rightarrow \sqrt{\hat{va}(\hat{\beta}_2)} = \dfrac{\hat{\beta}_2}{t^{obs}} = \dfrac{-19.46}{-7.79} = 2.498$

The p-value for $\beta_1 = 0$ is

$$\begin{aligned} \alpha^{obs} &= \mathbb{P}_{H_0}(|T| \geq |t^{obs}|) \\ &= 2\,\mathbb{P}_{H_0}(T \geq |t^{obs}|) \\ &= 2\,\mathbb{P}_{H_0}(T \geq 2.16) \\ &= \underbrace{2(1 - \underbrace{\mathbb{P}_{H_0}(T \leq 2.16)}_{0.975})}_{0.025} \end{aligned}$$



$\Rightarrow \alpha^{obs} = 0.05$

b) $\begin{cases} H_0: R^2 = 0 \\ H_1: R^2 \neq 0 \end{cases}$ in the case of the simple linear model, is equivalent to testing the significance of β_2.

From the table, we know that the pvalue of the corresponding test is $\alpha^{obs} < 0.0001$. Hence, we reject H_0 in both tests.

c) The residuals have have a clear quadratic pattern

This is suggesting that the model was not correctly specified, and a quadratic component is needed:

$$Y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$$