f) Obtain a point estimate of the mean service time when $x = 5$ parts are repaired. Obtain a confidence interval with level 0.90.

g) Obtain the partition of the total sum of squares and coefficient R^2.

Solutions

a) To obtain the estimated regression model, we need to compute the estimates for our parameters (β_1 and β_2).

Starting from the Gaussian linear model as a form

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \qquad i = 1, \dots, 18$$

where $\varepsilon_i \sim N(0, \sigma^2)$ - see assumptions in ~~note~~ the lectures notes or in the previous exercise.

In this case, we know that

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{1098}{74.5} = 14.74$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} = \frac{1152}{18} - 14.74 \cdot \frac{81}{18} = -2.33$$

Then,

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i = -2.33 + 14.74\, x_i \qquad i = 1, \dots, 18$$

b) In the previous exercise, we obtained $\hat{\beta}_2 = 14.74$ and recalling

$\mathbb{E}[Y_i] = \beta_1 + \beta_2 x_i$

⇒ The mean of Y_i increases ~~by~~ 14.74 as the value of x_i increases ~~by~~ one unit.

Using the information about our variables:

⇒ The mean of the total number of minutes spent by the service person increases by 14.74 minutes as the value of x_i ~~increases~~ (the number of repairs) increases by one-unit.

While β_1 represents the intercept of our regression line. This value represents changes when there is not repairs in that call.

Given that data are collected from 18 call to ~~f~~ ~~person~~ perform maintenance service, an absence of repairs is not within the scope of the observation. The value of $\hat{\beta}_1$ doesn't contain relevant information but it is important as constant term of the regression line.

c) Let consider the following system of hypothesis

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

is a one-tail test.

The test statistic corresponds to

$$t_2^{oss} = \frac{\hat{\beta}_2}{\sqrt{\hat{Var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{n-2}$$

Hence,

$$t_2^{oss} \overset{*}{=} \frac{14.74}{0.52} = 28.34615$$

$$(*) \sqrt{\hat{Var}(\hat{\beta}_2)} = \sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \sqrt{\frac{19.85}{74.5}} = 0.5161811$$

where

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 - \hat{\beta}_2^2\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = \frac{1}{16}\left(16504 - (14.74)^2 \cdot 74.5\right) = 19.84774$$

P-value:

$$\alpha^{oss} = \mathbb{P}(t_{n-2} > 28.34615) = 2.091323 \cdot 10^{-15} < 0.05 \Rightarrow \text{reject } H_0$$

d) $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow 1-\frac{\alpha}{2} = 0.975$

The confidence interval is equal to

$$\left(\hat{\beta}_2 - t_{16,0.975}\sqrt{\hat{Var}(\hat{\beta}_2)},\ \hat{\beta}_2 - t_{16,0.975}\sqrt{\hat{Var}(\hat{\beta}_2)}\right)$$

Which corresponds to

$$(14.74 - 2.12 \cdot 0.5162,\ 14.74 + 2.12 \cdot 0.5162) = (13.6457, 15.83414)$$

e) Given the result of the previous test (ex 1c), It is not necessary to perform the test with an alternative hypothesis $\beta_2 \neq 0$.

Let's think what is the meaning of β_2: it is the coefficient associated to x which represents the number of repairs and affects the total numbers of minutes spent by service person. As the number of repairs increases, it could be weird that the mean of y decreases.

Given that it's reasonable ~~that~~ having $\beta_2 > 0$, considering a null hypothesis of the form $\beta_2 \neq 0$ could be redundant.

However, in that case the relationship between the p-values of these two test is:

$$2 \cdot \text{p-value}^{(\beta_2 > 0)} = \text{p-value}^{(\beta_2 = 0)}$$

f) We can see $x^* = 5$ as new data.

At the beginning, we need to find the estimate for y:

$$\hat{y}^* = \hat{\beta}_1 + \hat{\beta}_2 x^* = -2.33 + 14.74 \cdot 5 = 71.37$$

or, equivalently, we can also use

$$\hat{y}^* = \bar{y} + \hat{\beta}_2 (x^* - \bar{x}) = \frac{1152}{18} + 14.74 \cdot \left(\frac{81}{18} - 5\right) = 64 + 14.74(4.5 - 5) = 71.37$$

From the theoretical lectures, we know

$$\hat{y}^* \sim N\left(\underbrace{\beta_1 + \beta_2 x^*}_{\mu_*}, \frac{\sigma^2}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sigma^2\right)$$

and hence

$$\frac{\hat{y}^* - \mu_*}{s\sqrt{\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}} \sim t_{n-2}$$

We are ready to compute our confidence interval:

$$\left(\hat{y}^* - t_{n-2;\,1-\alpha/2}\, s\sqrt{\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)},\ \hat{y}^* + t_{n-2;\,1-\frac{\alpha}{2}}\, s\sqrt{\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}\right)$$

From the previous exercise (1.c) we know that

$$s = \sqrt{19.84774} = 4.45508$$

and $1 - \alpha = 0.90 \Rightarrow \alpha = 0.1 \Rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow t_{16,0.95} = 1.745884$

$$\frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{(5-4.5)^2}{74.5} = 0.003355705$$

Then, the confidence interval of 90% corresponds to

$$\left(71.37 - 1.746 \cdot \left(4.45508 \cdot \sqrt{\frac{1}{18} + 0.003355705}\right), 71.37 + 1.746\left(4.45508\sqrt{\frac{1}{18} + 0.0033557\text{0}}\right.\right.$$

$$= (71.37 - 1.888, 71.37 + 1.888) = (69.482, 73.258)$$

g) In the theoretical lectures, you saw

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{SST} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{SSR} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{SSE}$$

And, using the data, we find

$SST = 16504$ and $SSE = 321.39$

$SSR = SST - SSE = 16504 - 321.39 = 16182.61$

$$R^2 = \frac{SSR}{SST} = \frac{16182.61}{16504} = 0.981$$