EXERCISE 1

$n = 20$ students

simple Gaussian LM : $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

a) assumption on ε_i : $\varepsilon_i \sim N(0, \sigma^2)$ iid for $i = 1, \dots, 20$

(normality, mean = 0, homoscedasticity, independence)

b) The MLE of (β_1, β_2) are

$$\hat\beta_1 = \bar y - \hat\beta_2 \bar x$$

$$\hat\beta_2 = \frac{cov(x,y)}{var(x)} = \frac{S_{xy}}{S_x^2}$$

First, we compute $\hat\beta_2$:

$$n \cdot cov(x,y) = \sum_{i=1}^n (x_i - \bar x)(y_i - \bar y) =$$

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

$$\Rightarrow n \cdot cov(x,y) = 257.66 - \cancel{20} \cdot \frac{100}{\cancel{20}} \cdot \frac{50}{20} = 7.66$$

$$n \cdot var(x) = \sum_{i=1}^n (x_i - \bar x)^2 = \sum_{i=1}^n x_i^2 - n \bar x^2$$

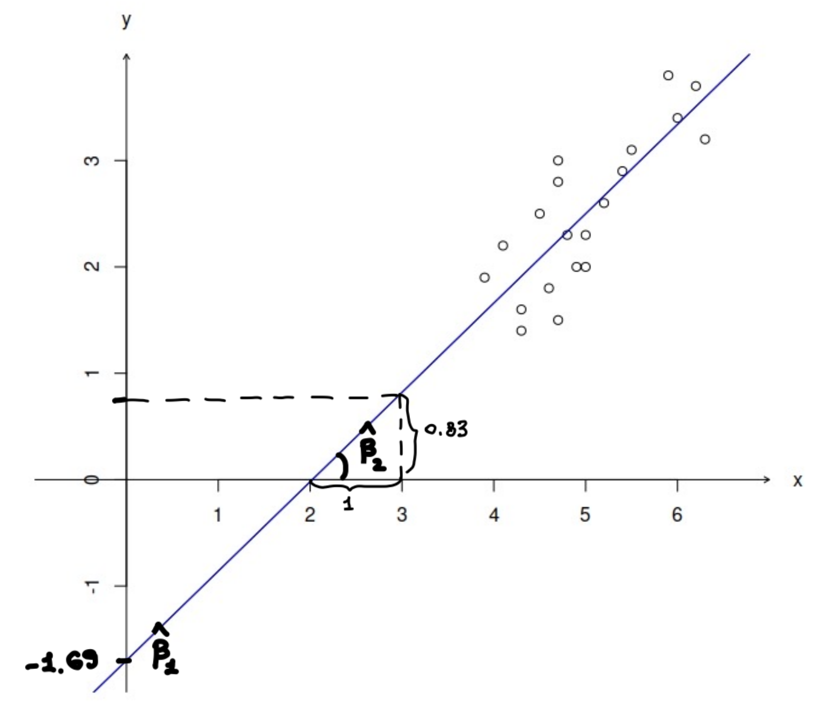
$$\Rightarrow n\ var(x) = 509.12 - 20 \cdot \left(\frac{100}{20}\right)^2 = 9.12$$

Hence $\hat\beta_2 = \frac{7.66}{9.12} = 0.8399$

Then, $\hat\beta_1 = \bar y - \hat\beta_2 \bar x = \frac{50}{20} - 0.8399 \cdot \frac{100}{20} = -1.6995$

- $\hat\beta_1$ is the mean GPA corresponding to an entrance test score equal to 0
- $\hat\beta_2$ is the expected change in the mean GPA if I increase the entrance test score of 1 unit.

c) $\hat y_i = -1.6995 + 0.8399\, x_i$

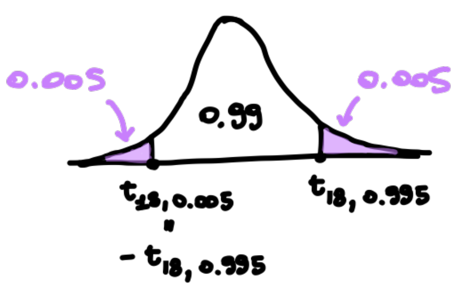


d) $\hat B_2(Y)$: is the random interval defined by

$$P(\beta_2 \in \hat B_2(Y)) = 0.99 \quad \forall \beta_2 \in \mathbb{R}$$

To obtain it I use the pivotal quantity

$$T = \frac{\hat\beta_2(Y) - \beta_2}{\sqrt{\widehat{var}(\hat\beta_2)}} \sim t_{18}$$



$$P(-t_{18,0.995} < T < t_{18,0.995}) = 0.99$$

$$P\left(-t_{18,0.995} < \frac{\hat\beta_2(Y) - \beta_2}{\sqrt{\widehat{var}(\hat\beta_2)}} < t_{18,0.995}\right) = 0.99$$

Hence $\beta_2 \in \hat\beta_2 \pm t_{18,0.995} \cdot \sqrt{\widehat{var}(\hat\beta_2)}$

$\beta_2 \in (0.8399 \pm 2.878 \cdot 0.1440) = (0.4254 ; 1.2543)$ it does not contain zero.

If the CI contains 0, it means that we do not reject the hypothesis $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$ using a significance level 0.01.

Hence, in this case, it would suggest that the entrance score is not a good measure for predicting the students' GPA

e) Yes, because $0 \notin CI$ in the previous point.

f) Prediction intervals get wider when we try to predict y far from the observed range of points x_i. In particular, the smallest interval is obtained at $x_i = \bar x$.

Since $\bar x = 5.0 = x_A$ (entrance test of student A)

while $x_B = 6.5$ is even outside of the observed range,

the prediction interval for student B will be wider.