14 Oct 2025 - lec. 5

Exam 24/09/2024

EXERCISE 1

we observe $(x_i; y_i)$ for $i = 1, \dots, 16$

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

a) $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$

hence I need to compute the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$.

we know that

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

$\hat{\beta}_2 = \dfrac{s_{xy}}{s_x^2}$

hence we obtain

$$\hat{\beta}_2 = \frac{-134.1542}{131.0625} = -1.0236$$

$$\hat{\beta}_1 = \frac{1379}{16} + 1.0236 \cdot \frac{967}{16} = 148.0513$$

$$\hat{y}_i = 148.0513 - 1.0236\, x_i$$

b) interpretation of β_2

we take two women i and k of age x_i and x_k, such that $x_k = x_i + 1$.

Then, we compare their expected muscle mass

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

$\mathbb{E}[Y_k] = \beta_1 + \beta_2 x_k = \beta_1 + \beta_2 (x_i + 1)$

their difference is

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

In our case, $\hat{\beta}_2 = -1.02$. Hence, when age increases of 1 year, we expect a decrease of the muscle mass of 1.02 unit.

c) we need to test

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

We know that under H_0: $\hat{B}_2 \overset{H_0}{\sim} N(0, var(\hat{B}_2))$

unknown

$$\hat{B}_2 \overset{H_0}{\sim} N\left(0, \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)$$

Here we have the data, let's write the distribution explicitly

$$var(\hat{B}_2) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

and we know that $s_x^2 = \dfrac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$$\Rightarrow \sum_{i=1}^{n} (x_i - \bar{x})^2 = s_x^2 (n-1) = 131.06 \cdot 15 = 1965.9$$

Hence, under H_0,

$$\hat{B}_2 \overset{H_0}{\sim} N\left(0, \frac{\sigma^2}{1965.9}\right)$$

The variance of $\hat{B}_2$ is $var(\hat{B}_2) = \dfrac{\sigma^2}{1965.9}$

However σ^2 is unknown $\Rightarrow$ we estimate it

The unbiased estimate of σ^2 is $s^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 69.62$

We obtain an estimate of the variance of $\hat{B}_2$ as $\hat{var}(\hat{B}_2) = \dfrac{s^2}{1965.9} = \dfrac{69.62}{1965.9} = 0.035$

We need a pivotal quantity.

If we consider

$$\frac{\hat{B}_2 - 0}{\sqrt{\frac{\sigma^2}{1965.9}}} \overset{H_0}{\sim} N(0,1) \qquad \text{but we don't know } \sigma^2$$

If we substitute σ^2 with s^2, this is going to affect the distribution of the transformation because s^2 involves $(y_1, \dots, y_n) \rightarrow (Y_1, \dots, Y_n)$ are r.v.

We have seen that

mean under H_0

estimators: random!

$$T = \frac{\hat{B}_2 - 0}{\sqrt{\hat{var}(\hat{B}_2)}} = \frac{\hat{B}_2}{\sqrt{\frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}} \overset{H_0}{\sim} t_{n-2} = t_{14}$$

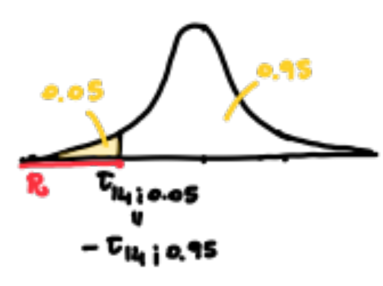
We have a one-side test: $H_1: \beta_2 < 0$

What values of t^{obs} point against H_0?

If $\beta_2 \ll 0 \rightarrow t^{obs} \ll 0$

we reject for large negative values of t^{obs}

$R = (-\infty; a)$



$\Rightarrow$ at a significance level 0.05,

$$R = (-\infty; -t_{14;0.95}) = (-\infty; -1.7613)$$

$$t^{obs} = \frac{\hat{\beta}_2}{\sqrt{\hat{var}(\hat{\beta}_2)}} = \frac{-1.0236}{\sqrt{0.03542}} = -5.4388 \Rightarrow t^{obs} \in R \Rightarrow \text{I reject } H_0$$

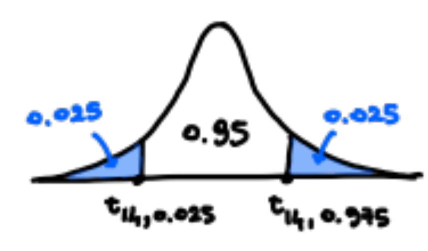
d) CI for β_2:

We have already computed the mle: $\hat{\beta}_2 = -1.0236$.

A 95% confidence interval is

$\hat{\Theta}$ such that $\mathbb{P}(\beta_2 \in \hat{\Theta}) = 0.95$

I use $T = \dfrac{\hat{B}_2 - \beta_2}{\sqrt{\hat{var}(\hat{B}_2)}} \sim t_{14}$



$$0.95 = \mathbb{P}(-t_{14,0.975} < T < t_{14,0.975})$$

$$0.95 = \mathbb{P}\left(\hat{B}_2 - \sqrt{\hat{var}(\hat{B}_2)} \cdot t_{14,0.975} < \beta_2 < \hat{B}_2 + \sqrt{\hat{var}(\hat{B}_2)} \cdot t_{14,0.975}\right)$$

with the data: $-1.0236 - 0.1882 \cdot 2.1448 < \beta_2 < -1.0236 + 0.1882 \cdot 2.1448$

$$\Rightarrow \beta_2 \in (-1.4272; -0.6199)$$

e) Let us consider woman A with $x_A = 38$.

Her predicted muscle mass is $\hat{y}_A = 148.0513 - 1.0236 \cdot 38 = 109.154$

Woman B has instead $x_B = 60$, we obtain

$\hat{y}_B = 148.0513 - 1.0236 \cdot 60 = 86.6353$

In the Gaussian linear model, prediction at a new value x_* is equivalent to estimating the mean $\mu_* = \mathbb{E}[Y_*] = \beta_1 + \beta_2 x_*$

For woman A, the estimator of μ_A is $\hat{M}_A = \hat{B}_1 + \hat{B}_2 x_A$

For woman B, the estimator of μ_B is $\hat{M}_B = \hat{B}_1 + \hat{B}_2 x_B$

Their distributions are:

$$\hat{M}_A \sim N\left(\mu_A, \sigma^2 \left(\frac{1}{n} + \frac{(x_A - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)\right)$$

$$\hat{M}_B \sim N\left(\mu_B, \sigma^2 \left(\frac{1}{n} + \frac{(x_B - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)\right)$$

The uncertainty of the estimate depends on the variance of the estimator

At the numerator, we have the square of the difference between the new point and the mean $\bar{x}$.

If this difference increases $\rightarrow$ the variance increases.

The mean of the observed x is $\bar{x} = 60.43$

Hence, the age of woman B is very close to the mean of observations.

On the contrary, for woman A we are extrapolating outside the range of observed ages.

$$(x_A - \bar{x})^2 < (x_B - \bar{x})^2$$

Hence, $\hat{y}_B$ has the largest uncertainty.

f) residuals $e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_i)$

$x_8 = 56, \quad y_8 = 80$

$\hat{y}_8 = 148.0513 - 1.0236 \cdot 56 = 90.7297$

hence $e_8 = 80 - 90.7297 = -10.7297$

The sum of the residuals in this case is zero

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = n\bar{y} - \sum_{i=1}^{n} (\hat{\beta}_1 + \hat{\beta}_2 x_i) = n\bar{y} - \sum_{i=1}^{n} (\bar{y} - \hat{\beta}_2 \bar{x} + \hat{\beta}_2 x_i)$$

$$= \cancel{n\bar{y}} - \cancel{n\bar{y}} + \cancel{n\hat{\beta}_2 \bar{x}} - \cancel{n\hat{\beta}_2 \bar{x}} = 0$$

g) $R^2 = \dfrac{SSR}{SST} = \dfrac{\text{regression sum of squares}}{\text{total sum of squares}} = \dfrac{SST - SSE}{SST} = 1 - \dfrac{SSE}{SST}$

it describes the proportion of variability of y explained by the model

we need: $SSE = \sum_{i=1}^{16} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{16} e_i^2$

since $s^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} e_i^2 \Rightarrow SSE = (16-2) \cdot s^2 = 14 \cdot 69.62 = 974.68$

then, $SST = \sum_{i=1}^{16} (y_i - \bar{y})^2$

$s_y^2 = \dfrac{1}{n-1} \sum_{i=1}^{16} (y_i - \bar{y})^2 \Rightarrow SST = (16-1) \cdot s_y^2 = 15 \cdot 202.2958 = 3034.437$

$R^2 = 0.6788$