# SHEET - An Introduction to Statistical Learning
# Chapter 3 - Linear Regression

PLAYE Nicolas

4 December 2023

# 1 Linear Regression

## 1.1 Courses' Demonstrations

Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for $Y$ based on the $i$th value of $X_i$. Then $e_i = y_i - \hat{y}_i$ represents the $i$th residual - this is the difference between the $i$th observed response value and the $i$th response value that is predicted by our linear model. We define the residual sum of squares (RSS) as

$$\begin{aligned}\text{RSS} &= e_1^2 + e_2^2 + \cdots + e_n^2 \\ &= (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \cdots + (y_n - \hat{\beta}_0 - \hat{\beta}_n x_n)^2\end{aligned}$$

The least squares approach chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the RSS. Using some calculus, one can show that the minimizers are

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \qquad (3.4)\end{aligned}$$

where $\bar{y}$ is the sample mean, defined as

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Demonstration:

We search RSS as

$$f(\hat{\beta}_0, \hat{\beta}_1) = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

We get the minimum where

$$\begin{cases} \dfrac{\partial f}{\partial \hat{\beta}_0} = 0 \\ \dfrac{\partial f}{\partial \hat{\beta}_1} = 0 \end{cases}$$

$$\begin{cases} -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0 \\ \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \end{cases}$$

$$\begin{cases} \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0 \\ \sum_{i=1}^{n} x_i y_i - n\hat{\beta}_0 \bar{x} - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \end{cases}$$

$$\begin{cases} \hat{\beta}_0 = \bar{y} + \hat{\beta}_1 \bar{x} \\ 0 = \sum_{i=1}^{n} x_i y_i - n\hat{\beta}_0 \bar{x} - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \end{cases}$$

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ 0 = \sum_{i=1}^{n} x_i y_i - n(\bar{y} + \hat{\beta}_1 \bar{x})\bar{x} - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \end{cases}$$

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ 0 = \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x} + \hat{\beta}_1 (n\bar{x}^2 - \sum_{i=1}^{n} x_i^2) \end{cases}$$

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ 0 = \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x} + \hat{\beta}_1 \sum_{i=1}^{n} (x_i^2 - \bar{x}^2) \end{cases}$$

$$\begin{aligned} E[(X - E[X])(Y - E[Y])] &= E[XY - X * E[Y] - E[x] * Y + E[X] * E[Y]] \\ &= E[XY] - E[X * E[Y]] - E[E[x] * Y] + E[E[X] * E[Y]] \\ &= E[XY] - E[X] * E[Y] - E[x] * E[Y] + E[X] * E[Y] \\ &= E[XY] - E[X] * E[Y] \end{aligned}$$

$$\begin{aligned} E[(X - E[X])^2] &= E[X^2 - 2 * X * E[X] + E[X]^2] \\ &= E[X^2] - E[2 * X * E[X]] + E[E[X]^2] \\ &= E[X^2] - 2 * E[X] * E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned} \tag{1}$$

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{cases}$$

By developping we can have

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(y_i)(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We assume that the True relationship between X and Y takes the form

$$Y = f(X) + \varepsilon$$

for some unknown function f, where $\varepsilon$ is a mean-zero random error term. If f is to be approximated by a linear function then we can write the relationship as

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{3.5}$$

How accurate is the sample mean $\hat{\mu}$ as an estimate of $\mu$? In general, we answer this question by computing the standard error of $\hat{\mu}$, written as the standard error SE($\hat{\mu}$). A reasonable estimate is $\hat{\mu} = \bar{y}$. We have the well-known formula:

$$Var(\hat{\mu}) = SE(\hat{\mu}^2) = \frac{\sigma^2}{n} \tag{3.7}$$

where $\sigma$ is the standard deviation of each of the realizations $y_i$ of Y.

Demonstration:

We have

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mu_i$$

We can deduce

$$E[\bar{\mu}] = E\left[\frac{1}{n}\sum_{i=1}^{n} \mu_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[\mu_i] = \mu$$

$$Var(\hat{\mu}) = Var(\bar{y}) = E\left[(\bar{y} - E[\bar{y}])^2\right]$$

$$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i - E\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right]\right)^2\right]$$

$$= \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n} y_i - E[y_i]\right)^2\right]$$

We set $a_i = y_i - E[y_i]$

$$Var(\hat{\mu}) = \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n} a_i\right)^2\right] = \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{n} a_j\right)\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n} a_i^2 + 2\sum_{1\leq i<j\leq n} a_i a_j\right]$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} E[a_i^2] + 2\sum_{1\leq i<j\leq n} E[a_i a_j]\right)$$

We have $E[a_i] = E[y_i - E[y_i]] = E[y_i] - E[y_i] = 0$

$E[a_i a_j] = E[a_i]E[a_j]$ since $a_i$ and $a_j$ are independent

So $Var(\hat{\mu}) = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} E[(y_i - E[y_i])^2]$

$$Var(\hat{\mu}) = \frac{\sigma^2}{n}$$

In a similar vein, we can wonder how close $\hat{\beta}_0$ and $\hat{\beta}_0$ are to the true values $\beta_0$ and $\beta_1$ . To compute the standard errors associated with $\beta_0$ and $\beta_1$ , we use the following formulas:

$$SE(\hat{\beta}_0) = \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}] \tag{3.8}$$

$$SE(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.8}$$

where $\sigma^2 = Var(\varepsilon)$

Démonstration :

We have

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n w_i y_i \quad \text{with} \quad w_i = \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

We have too

$$\sum_{i=1}^n w_i = \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0$$

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

$$\sum_{i=1}^n w_i^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

We can deduce

$$\begin{aligned}
\mathbf{E}[\hat{\beta}_{\mathbf{1}}] &= E\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = E\left[\sum_{i=1}^n w_i y_i\right] \\
&= E\left[\sum_{i=1}^n w_i(\beta_1 x_i + \beta_0 + \varepsilon_i)\right] = E[\beta_1 \sum_{i=1}^n w_i x_i + \beta_0 \sum_{i=1}^n w_i + \varepsilon_i \sum_{i=1}^n w_i] = \\
&= E[\beta_1 * 1 + \beta_0 * 0 + \varepsilon_i * 0] = E[\beta_1] = \beta_{\mathbf{1}}
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}[\hat{\beta}_{\mathbf{0}}] &= E[\bar{y} - \hat{\beta}_1 \bar{x}] = E[\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_1 x_i)] = \frac{1}{n} \sum_{i=1}^n E[y_i - \hat{\beta}_1 x_i] \\
&= \frac{1}{n} \sum_{i=1}^n E[(\beta_1 x_i + \beta_0 + \varepsilon_i) - \hat{\beta}_1 x_i] = \frac{1}{n} \sum_{i=1}^n (E[(\beta_1] E[x_i] + E[\beta_0] + E[\varepsilon_i]) - E[\hat{\beta}_1] E[x_i]) \\
&= \frac{1}{n} \sum_{i=1}^n (\beta_1 x_i + \beta_0 + 0 - \beta_1 x_i) = \frac{1}{n} \sum_{i=1}^n \beta_0 = \beta_{\mathbf{0}}
\end{aligned}$$

Démonstration :

$$\begin{aligned}
\mathbf{Var}(\hat{\beta}_\mathbf{1}) &= Var(\sum_{i=1}^{n} w_i y_i) \\
\text{We have } Var(X+Y) &= E[(X+Y)^2] - (E[X+Y])^2 = E[X^2+2XY+Y^2] - (E[X]+E[Y])^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - (E[X]^2 + 2E[X]E[Y] + E[Y]^2) \\
&= E[X^2] + E[Y^2] - E[X]^2 - E[Y]^2 = Var(X) + Var(Y) \\
\text{So we have } Var(\hat{\beta}_1) &= \sum_{i=1}^{n} Var(w_i y_i) = \sum_{i=1}^{n} Var(\frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} y_i) \\
\text{Because } \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} &= Constante = B_i \\
\text{And Because } Var(BX) &= E[(B_i X - E[B_i X])^2] = E[B_i^2 (X - E[X])^2] = B_i^2 Var(X) \\
\text{We have } Var(\hat{\beta}_1) &= \sum_{i=1}^{n} B_i^2 Var(y_i) = \sum_{i=1}^{n} B_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} B_i^2 \\
&= \sigma^2 \sum_{i=1}^{n} (\frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2})^2 = \sigma^2 \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2)}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2} \\
Var(\hat{\beta}_1) &= \sigma^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2)}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2} \\
\mathbf{Var}(\hat{\beta}_\mathbf{1}) &= \frac{\boldsymbol{\sigma^2}}{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{x_i} - \bar{\mathbf{x}})^\mathbf{2}}
\end{aligned}$$

$$Var(\hat{\beta}_0) =$$

## 1.2 Answers of Exercises