linear model:

$$Y = \beta_0 + \beta_1 X + \underbrace{\varepsilon}_{\text{noise}}$$

assumptions:

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$

  $\underline{\mathbb{V}[\varepsilon_i] = \sigma^2}$

- $\forall i, j \in \{1, \ldots, n\}$ $\underline{\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0}$

Matricial writting

$$\underline{Y} = \mathbb{X}\beta + U$$

$$\underline{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad U = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\mathbb{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$\hat{\beta}_n = ({}^t\mathbb{X}\mathbb{X})^{-1}\,{}^t\mathbb{X}\,\underline{Y}$$

least square solution

Rk: $({}^t\mathbb{X}\mathbb{X})^{-1} \neq \mathbb{X}^{-1}({}^t\mathbb{X})^{-1}$

because in general $\mathbb{X}^{-1}$ does not exist

$$\mathbb{E}[\hat{\beta}_n] = \beta$$

$$V[\hat{\beta}_n] = \sigma^2 ({}^t X X)^{-1}$$

<u>proof:</u>

$$V[\hat{\beta}_n] = \sigma^2 ({}^t X X)^{-1}$$

$$\hat{\beta}_n = \left({}^t\mathbb{X}\mathbb{X}\right)^{-1}\left({}^t\mathbb{X}\underline{Y}\right)$$

$$\underline{Y} = \mathbb{X}\beta + U$$

$$\Rightarrow \hat{\beta}_n = \left({}^t\mathbb{X}\mathbb{X}\right)^{-1}\,{}^t\mathbb{X}\left(\mathbb{X}\beta + U\right)$$

$$= \underbrace{\left({}^t\mathbb{X}\mathbb{X}\right)^{-1}\left({}^t\mathbb{X}\mathbb{X}\right)}_{I}\beta + \left({}^t\mathbb{X}\mathbb{X}\right)^{-1}\,{}^t\mathbb{X}U$$

$$= \beta + \left({}^t\mathbb{X}\mathbb{X}\right)^{-1}\,{}^t\mathbb{X}U$$

$$V[\hat{\beta}_n] = \mathbb{E}\left[\left(\hat{\beta}_n - \mathbb{E}[\hat{\beta}_n]\right) \times {}^t\left(\hat{\beta}_n - \mathbb{E}[\hat{\beta}_n]\right) \right]$$

$$= \mathbb{E}\left[\left(\hat{\beta}_n - \beta\right) \times {}^t\left(\hat{\beta}_n - \beta\right) \right]$$

$$= \mathbb{E}\left[\underbrace{({}^t\mathbb{X}\mathbb{X})^{-1}\,{}^t\mathbb{X}}_{\text{fixed}} \underbrace{U \times {}^tU}_{\text{random}} \underbrace{\mathbb{X}\,({}^t\mathbb{X}\mathbb{X})^{-1}}_{\text{fixed}} \right]$$

$$= ({}^t\mathbb{X}\mathbb{X})^{-1}\,{}^t\mathbb{X}\, \boxed{\mathbb{E}[U \cdot {}^tU]}\, \mathbb{X}\,({}^t\mathbb{X}\mathbb{X})^{-1}$$

$$\mathbb{E}\left[U \cdot {}^tU\right] = \mathbb{E}\left[\left(U - \mathbb{E}[U]\right) \times {}^t\left(U - \mathbb{E}[U]\right)\right]$$

since $\mathbb{E}[U] = \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}\right\} n \text{ rows}$

$$= V[U]$$

$$V[U] = \begin{pmatrix} V[\varepsilon_1] & cov(\varepsilon_1, \varepsilon_2) & \cdots & cov(\varepsilon_1, \varepsilon_n) \\ & \ddots & & \\ & & \ddots & \\ & & & V[\varepsilon_n] \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & 0 \\ & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix} = \sigma^2 I_n$$

$$V[\hat{\beta}_n] = ({}^t\mathbb{X}\mathbb{X})^{-1}\,{}^t\mathbb{X}\,(\sigma^2 I_n)\,\mathbb{X}\,({}^t\mathbb{X}\mathbb{X})^{-1}$$

$$= \sigma^2 ({}^t\mathbb{X}\mathbb{X})^{-1} \underbrace{({}^t\mathbb{X}\mathbb{X})({}^t\mathbb{X}\mathbb{X})^{-1}}_{I = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{\text{In the simple linear regression}}}$$

$$= \sigma^2 ({}^t\mathbb{X}\mathbb{X})^{-1}$$

Rk: $\hat{\beta}_n = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$

$$V[\hat{\beta}_n] = \begin{pmatrix} V[\hat{\beta}_0] & cov(\hat{\beta}_0, \hat{\beta}_1) \\ cov(\hat{\beta}_0, \hat{\beta}_1) & V[\hat{\beta}_1] \end{pmatrix}$$

$$\Rightarrow V[\hat{\beta}_0] = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \overline{X}_n)^2}, \quad V[\hat{\beta}_1] = \frac{\sigma^2}{\sum (X_i - \overline{X}_n)^2}$$

$$\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \bar{x}_n / \sum (X_i - \bar{X}_n)^2$$

Those variances are theoretical quantities since they depend on σ^2 which is unknown

$\rightarrow$ we need to estimate σ^2

Difficulty:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$\Rightarrow \mathbb{E}[Y_i] = \beta_0 + \beta_1 X_i$$

$\Rightarrow$ the Y_i are not identically distributed.

but $V[Y_i] = \sigma^2$

if we consider

$$Z_i = Y_i - (\beta_0 + \beta_1 X_i)$$

$$\hookrightarrow E[Z_i] = 0$$

$$V[Z_i] = \sigma^2$$

but we do not have access to the Z_i

definition: Residuals

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i \quad \text{when } \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

prediction of Y for individual i

Property:

$$E[\hat{\varepsilon}_i] = 0$$

Proposition:

An unbiased estimator of σ^2 is given by:

$$\hat{\sigma}_n^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

Prevision : main aim of linear model

If we have a new observation for the variable X, new observation denoted x_{n+1}, we would like to get the associated value for Y, denoted y_{n+1}.

↳ we assume that we have:

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1}$$

with $\mathbb{E}[\varepsilon_{n+1}] = 0$, $\mathbb{V}[\varepsilon_{n+1}] = \sigma^2$

and $\forall k \in \{1, \dots, n\}$, $\mathrm{cov}(\varepsilon_{n+1}, \varepsilon_k) = 0$

One way to predict y_{n+1} is to use

$$\hat{y}^{P}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$$

this is to say that this is a prevision since x_{n+1} has not been used in the construction of $\hat{\beta}_0$ and $\hat{\beta}_1$

Proposition

- $$V\left[\hat{y}^{P}_{n+1}\right] = \sigma^2\left(\frac{1}{n} + \frac{(x_{n+1} - \bar{x}_n)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}\right)$$

- prevision error : $\hat{\varepsilon}^{P}_{n+1} = y_{n+1} - \hat{y}^{P}_{n+1}$

$$\mathbb{E}\left[\hat{\varepsilon}^{P}_{n+1}\right] = 0$$

$$V\left[\hat{\varepsilon}^{P}_{n+1}\right] = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x}_n)^2}{\sum(x_i - \bar{x}_n)^2}\right)$$

Rk: We have

$$V\left[\hat{\varepsilon}^{P}_{n+1}\right] = \sigma^2 + V\left[\hat{y}^{P}_{n+1}\right]$$

why?

$$\hat{y}^{P}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{n+1}$$

$\rightarrow$ one error of estimation, when we consider that the model is the reality

$$\hat{\varepsilon}^{p}_{n+1} = y_{n+1} - \hat{y}^{p}_{n+1}$$

$$= \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1} - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{n+1}\right)$$

the error if the model is the reality

quantifi- -cation of the gap between the reality and the model

For next time: prove the
last proposition

We just have one more thing to see with just the minimal assumptions on the noise.

This is : <u>how to quantify the quality of the linear model ?</u>

Geometrical interpretation of linear model

Representation of the variables

$\mathbb{R}^n$: the space of the variables

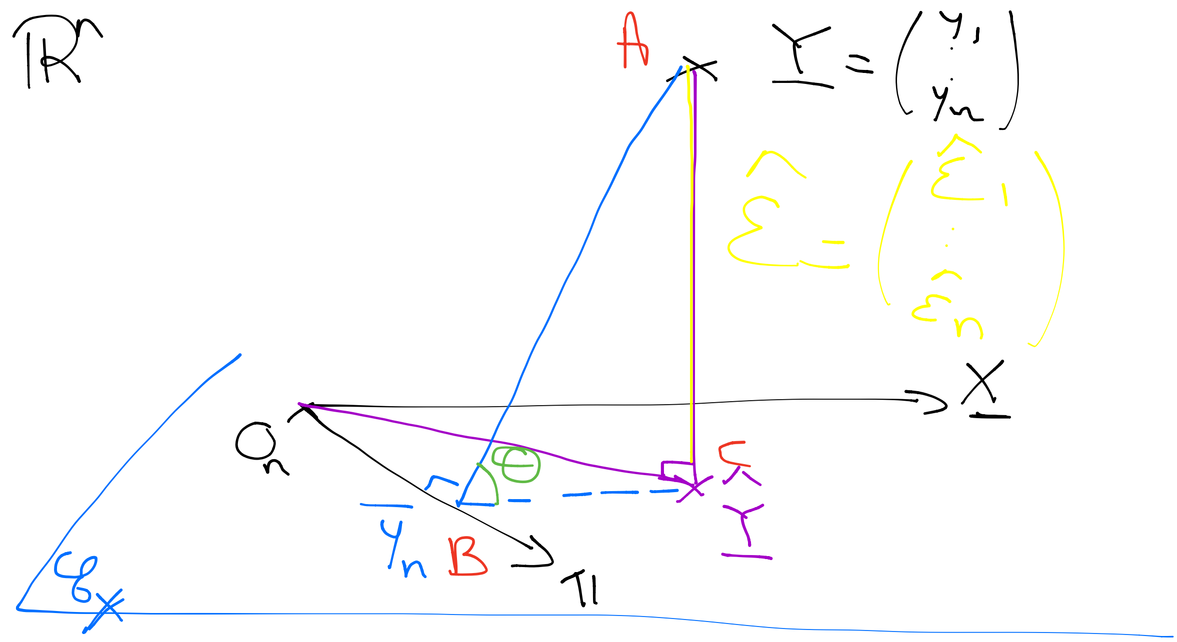
$\mathcal{E}_{\mathbb{X}}$: linear subspace generated by the columns of $\mathbb{X}$

(simple linear model,

$\mathcal{E}_{\mathbb{X}}$ is generated by $\underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbb{1}}$ and $\underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}}_{\underline{X}}$)

$\underline{\dim \mathcal{E}_{\mathbb{X}} = 2}$

$\mathbb{R}^n$



$$\hat{\underline{Y}} = \hat{\beta}_0 \mathbb{1} + \hat{\beta}_1 \underline{X} \Rightarrow \hat{\underline{Y}} \in \mathcal{E}_X$$

because of the least square criterion, $\hat{\underline{Y}}$ is the point de $\mathcal{E}_{\underline{\underline{X}}}$ which is the closest from $\underline{Y}$.

$\Rightarrow$ $\hat{\underline{Y}}$ is the projection of $\underline{Y}$ on $\mathcal{E}_{\underline{\underline{X}}}$

<u>Rk</u>: the projection of $\underline{Y}$ on the linear subspace generated by $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is $\overline{Y}_n \times \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

the projection of $\underline{\hat{Y}}$ on the same linear subspace is also $\overline{Y}_n \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

Thanks to Pythagore, we have:

$$\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}_{\substack{\text{total variance}\\\text{of } y\\(SST)}} = \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}_{\substack{\text{residual}\\\text{variance}\\(SSR)}} + \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_n)^2}_{\substack{\text{variance}\\\text{explained}\\\text{by the}\\\text{model}\\(SSE)}}$$

Coefficient of determination : R^2

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y}_n)^2}{\sum (y_i - \bar{y}_n)^2} = \frac{SSE}{SST}$$

<u>properties:</u>

- $0 \leq R^2 \leq 1$
- $R^2 = \cos^2 \theta$
- If $R^2 = 1 \Rightarrow \theta = 0 \Rightarrow \underline{Y} \in \mathcal{E}_X$

  $\Rightarrow$ the linear model is the reality

- If $R^2 = 0 \Rightarrow \sum (\hat{Y}_i - \overline{Y}_n)^2 = 0$

$\Rightarrow \hat{Y}_i = \overline{Y}_n$

$\Rightarrow$ the linear model is totaly wrong

- If R^2 is near 0, it means that $\underline{Y}$ is quite in $(\mathcal{E}_X)^{\perp}$

$\hookrightarrow$ the linear model is not correct

If R^2 is near 1, then $\underline{Y}$ is quite in

$\mathcal{E}_X \Rightarrow$ the linear model

is correct