Minimisation pb: $\underset{f \in G}{\operatorname{argmin}} \sum_{i=1}^{n} \ell(y_i - f(x_i))$ où (y_i, x_i) observed.

$\ell(u) = u^2$: quadratic loss.
$f(u) = |u|$: absolute loss.

Regression model: $Y = \underbrace{\beta_1}_{\text{intercept}} + \underbrace{\beta_2}_{\text{regression}} X + \varepsilon$.

$\Rightarrow \left| \begin{array}{l} y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \\ i = 1, \dots, n \end{array} \right.$ where (x_i) known fixed values.
(β_i) model parameters. unknown
(ε_i) realizations of unknown r.v.
(y_i) observations

OLS (Ordinary Least Squares) estimators: $(\hat{\beta}_1, \hat{\beta}_2) = \underset{(\beta_1, \beta_2) \in \mathbb{R} \times \mathbb{R}}{\operatorname{argmin}} S(\beta_1, \beta_2)$.

$$S(\beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2 = \| Y - \beta_1 \mathbb{1} - \beta_2 X \|^2$$

where: $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \Big\}$ n-times, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

Assumption: H_1: $X \notin \mathbb{R}\mathbb{1}$.

Define: $L(\beta) = Y - \beta_1 \mathbb{1} - \beta_2 X$ where $\beta = (\beta_1, \beta_2)$. Then: $S(\beta) = \| L(\beta) \|^2$.

$\forall \alpha \in [0,1]$, $\forall \beta, \beta' \in \mathbb{R}^2$

$$L((1-\alpha)\beta + \alpha\beta') = (1-\alpha) L(\beta) + \alpha L(\beta').$$

$$S((1-\alpha)\beta + \alpha\beta') = \| L((1-\alpha)\beta + \alpha\beta') \|^2 = \| (1-\alpha) L(\beta) + \alpha L(\beta') \|^2$$

$$\leq \{ (1-\alpha) \| L(\beta) \| + \alpha \| L(\beta') \| \}^2$$ (inégalité triang. pour $\|.\|$ et $\nearrow$ de $u \mapsto u^2$ sur $\mathbb{R}^+$).

$$\leq (1-\alpha) \underbrace{\| L(\beta) \|^2}_{S(\beta)} + \alpha \underbrace{\| L(\beta') \|^2}_{S(\beta')}.$$ (since: $u \mapsto u^2$ convex.)

Equality iff: $L(\beta) = L(\beta')$ $\Leftrightarrow$ $(\beta_1 - \beta_1')\mathbb{1} + (\beta_2 - \beta_2')X = 0$ impossible by H0

Hence: S strictly convex.

$\Rightarrow \hat\beta = (\hat\beta_1, \hat\beta_2)$ are obtained by: $\frac{\partial S}{\partial \beta_1}(\hat\beta) = \frac{\partial S}{\partial \beta_2}(\hat\beta) = 0$

$$\begin{cases} \frac{\partial S}{\partial \beta_1}(\hat\beta) = -2\sum_{i=1}^{n}\left(y_i - \hat\beta_1 - \hat\beta_2 x_i\right) = 0. \Rightarrow \bar y = \hat\beta_1 + \hat\beta_2 \bar x \\ \frac{\partial S}{\partial \beta_2}(\hat\beta) = -2\sum_{i=1}^{n} x_i\left(y_i - \hat\beta_1 - \hat\beta_2 x_i\right) = 0 \Rightarrow \overline{x.y} = \bar x\,\hat\beta_1 + \overline{x^2}\,\hat\beta_2 \end{cases}$$

where: $\bar x = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar y = \frac{1}{n}\sum_{i=1}^{n} y_i$

ie: $\overline{x.y} - \bar y.\bar x = \hat\beta_2\left(\overline{x^2} - (\bar x)^2\right)$.

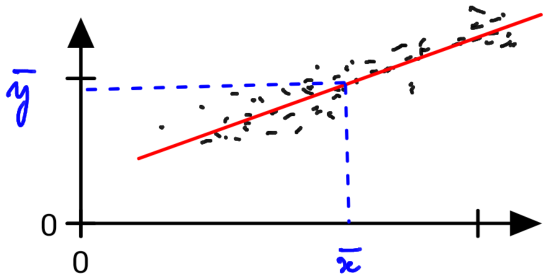
ie: $$\boxed{\hat\beta_2} = \frac{\overline{x.y} - \bar x.\bar y}{\overline{x^2} - \bar x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n}\frac{x_i}{n}\right)^2} = \boxed{\frac{\sum_{i=1}^{n}(x_i-\bar x)(y_i-\bar y)}{\sum_{i=1}^{n}(x_i-\bar x)^2}}.$$

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

(noting: $\sum_i (x_i - \bar x) = 0$, $\sum_i (y_i - \bar y) = 0$.)

Note the denom $\neq 0$ by H1 (which is eq to: $\exists\, i,j$ s.t $x_i \neq x_j$).

$$\boxed{\hat\beta_1 = \bar y - \hat\beta_2 \bar x}$$



Regression line: $y = \hat\beta_1 + \hat\beta_2 x$.

note that: $(\bar x, \bar y) \in$ regression line.

Def 1.2: . With $(x_i, y_i)_{1\le i\le n}$, we estimate $\hat\beta_1, \hat\beta_2$.

From one x_i, we can estimate $\hat y_i = \hat\beta_1 + \hat\beta_2 x_i$: $\hat y_i$ is an estimated value.

From a new x^*, we can predict: $\hat y^* = \hat\beta_1 + \hat\beta_2 x^*$: $\hat y^*$ is a predicted value.

2 questions: (1) $\hat\beta$ unbiased?

(2) $\hat\beta$ of minimal variance among estimators of a given class?

Second assumption: $\boxed{H2: \forall i \in [1:n],\ \mathbb{E}[\varepsilon_i] = 0,\ Var(\varepsilon_i) = \sigma^2. \quad \forall i \neq j,\ Cov(\varepsilon_i, \varepsilon_j) = 0.}$

Déf 1.3: biais de l'estimateur $\hat{\beta}$ de β : $\mathbb{E}(\hat{\beta}) - \beta$.

Prop 1.1: $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ unbiased estimator of $\beta = (\beta_1, \beta_2)$

Prop 1.2: $Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$; $Var(\hat{\beta}_1) = \frac{\sigma^2 \sum_i x_i^2}{n \left(\sum_{i=1}^n (x_i - \bar{x})^2\right)}$; $Cov(\hat{\beta}_1, \hat{\beta}_2) = -\frac{\sigma^2 \bar{x}}{\sum_i (x_i - \bar{x})^2}$.

Rq: Small variance for $\hat{\beta}_1, \hat{\beta}_2$ when:

- σ^2 small : observ. close to the reg. line.
- $\sum (x_i - \bar{x})^2$: large.
- $|\bar{x}|$ small indeed: $\frac{\sigma^2 \sum_i x_i^2 / n}{\sum_i (x_i - \bar{x})^2} = \frac{\sigma^2 \left(\sum \frac{x_i^2}{n} - \bar{x}^2\right)}{\sum_i (x_i - \bar{x})^2} + \frac{\sigma^2 \bar{x}^2}{\sum (x_i - \bar{x})^2}$

$$= \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum_i (x_i - \bar{x})^2}.$$

if $\bar{x} \geq 0$, $Cov(\hat{\beta}_1, \hat{\beta}_2) \leq 0$ ie: if $\hat{\beta}_1 \nearrow$, $\hat{\beta}_2 \searrow$

Gauss Markov: $Var(\hat{\beta}_i) \leq Var(\tilde{\beta}_i)$ where $\tilde{\beta}_i = u^T Y$ and : $\mathbb{E}(\tilde{\beta}_i) = \beta_i$ for $i = 1, 2$.

Déf 1.4: $\hat{\varepsilon}_i = y_i - \hat{y}_i$: residuals. (ie: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$).

Prop 1.3: $\sum_{i=1}^n \hat{\varepsilon}_i = 0$.

Prop 1.4: $\hat{\sigma}^2 = \sum_{i=1}^n \frac{\hat{\varepsilon}_i^2}{n-2}$ unbiased estimator of σ^2.