

Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

OLS #6 (le modèle linéaire multiple - 1)

Régression Linéaire

Nous supposons que les données collectées suivent le modèle suivant

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i,$$

où

- ▶ x_{ij} sont des nombres déterministes (connus). β_0 représente la constante (intercept dans les logiciels). On notera souvent $x_{i0} = 1$.
- ▶ $\beta_j, j = 0, 1, \dots, k$ paramètres réels à estimer. On pose $p = k + 1$
- ▶ les variables ε_i sont des fluctuations aléatoires (erreur de mesures, mauvaise spécification du modèle, ...).

On peut reformuler le modèle en:

$$\underbrace{\mathbf{Y}}_{(n \times 1)} = \underbrace{\mathbf{X}}_{(n \times p)} \underbrace{\boldsymbol{\beta}}_{(p \times 1)} + \underbrace{\boldsymbol{\varepsilon}}_{(n \times 1)}$$

Régression Linéaire

On peut reformuler le modèle en: $\underbrace{\mathbf{Y}}_{(n \times 1)} = \underbrace{\mathbf{X}}_{(n \times p)} \underbrace{\boldsymbol{\beta}}_{(p \times 1)} + \underbrace{\boldsymbol{\varepsilon}}_{(n \times 1)}$ où

$1 \leq p = 1 + k \leq n$ et

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & \dots & x_{nk} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$\boldsymbol{\beta} \in \mathbb{R}^p$ et $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^\top$. Lorsque $p = 2$, ce modèle s'appelle un modèle de régression linéaire **simple**. On notera

- ▶ \mathbf{x}_j le vecteur de taille $(n \times 1)$ des n observations de la j ème covariable.
- ▶ \mathbf{x}'_i le vecteur de taille $(1 \times p)$ des valeurs des p covariables pour l'individu i .
- ▶ \mathbf{Y} vecteur réponse; $\boldsymbol{\varepsilon}$ vecteur aléatoire (centré sans perte de généralité).

Régression Linéaire

Ainsi, la matrice \mathbf{X} appelée matrice de design (ou de plan d'expérience) s'écrit également: $\mathbf{X} = (\mathbf{x}_0, \dots, \mathbf{x}_k) = \begin{pmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_n \end{pmatrix}$.

\mathcal{H}_1 : La matrice de design \mathbf{X} est de plein rang.

Puisque $p \leq n$, $\mathcal{H}_1 \Rightarrow \text{rg}(\mathbf{X}) = p$, $\mathbf{X}^\top \mathbf{X}$ de taille (p, p) est symétrique, définie positive et donc **inversible**.

Régression Linéaire

Comment estimer β ?

L'idée consiste à minimiser une certaine norme de $\mathbf{Y} - \mathbf{X}\beta$.

Plusieurs méthodes pourraient être envisagées: tenter de minimiser

$$\blacktriangleright \sum_{i=1}^n (Y_i - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 = \|\mathbf{Y} - \mathbf{X}\beta\|^2.$$

(méthode des moindres carrés)

$$\blacktriangleright \sum_{i=1}^n \rho(Y_i - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}), \rho(\cdot) \geq 0, \rho(0) = 0, \text{ e.g.}$$
$$\rho(t) = |t|, |t|\mathbf{1}(|t| \leq M), \dots$$

(M-estimation)

$$\blacktriangleright \sum_{i=1}^n d_i^2, \text{ où } d_i \text{ est la distance la plus proche du point}$$
$$(x_{i1}, \dots, x_{ip}, y_i) \text{ sur le plan d'équation } \beta_1 x_1 + \cdots + \beta_p x_p$$

(régression orthogonale)

Régression Linéaire

On appelle estimateur des moindres carrés ordinaires (noté MCO ou OLS) $\widehat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ la valeur suivante:

$$\widehat{\boldsymbol{\beta}} = \underset{\beta_0, \beta_1, \dots, \beta_k}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ik})^2 = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2.$$

Sous l'hypothèse \mathcal{H}_1 , l'estimateur des MCO existe et vaut

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Par calcul matriciel, en calculant le gradient et la matrice hessienne de $S(\boldsymbol{\beta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \quad \text{et} \quad \frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X}.$$

Régression Linéaire Simple

L'estimateur des MCO du modèle de régression linéaire simple $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (sous l'hypothèse que $\widehat{\text{Var}}(\mathbf{x}) \neq 0$) est donné par

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(\mathbf{x}, \mathbf{Y})}{\widehat{\text{Var}}(\mathbf{x})} \quad \text{et} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

où $\widehat{\text{Cov}}(\mathbf{a}, \mathbf{b}) = (n - 1)^{-1} \sum (a_i - \bar{a})(b_i - \bar{b})$

Régression Linéaire

On suppose que l'on a un modèle homoscédastique

\mathcal{H}_2 : Les erreurs sont centrées, de même variance et non corrélées
 $\Leftrightarrow \mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ et $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbb{I}_n$.

Sous les hypothèses $\mathcal{H}_1 - \mathcal{H}_2$, l'estimateur $\hat{\boldsymbol{\beta}}$ des MCO est un estimateur sans biais de $\boldsymbol{\beta}$ et $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \mathbb{E}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}) = \mathbb{E}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top [\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}])$$

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}}_{=\mathbf{I}} + \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}(\boldsymbol{\varepsilon})}_{=0} = \boldsymbol{\beta}.$$

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \mathbb{E}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top [\boldsymbol{\varepsilon}][\boldsymbol{\varepsilon}]^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1})$$

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \underbrace{\mathbb{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top)}_{=\sigma^2 \mathbb{I}_n} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$$

Régression Linéaire, Gauss-Markov

Si $p = 2$, les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ sont sans biais et

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, \quad \text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2},$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{x}}{\sum (x_i - \bar{x})^2}$$

Sous les hypothèses $\mathcal{H}_1 - \mathcal{H}_2$, l'estimateur des MCO est optimal (sans biais et de variance minimale) dans la classe des estimateurs linéaires sans biais de β .

Régression Linéaire

Si n augmente, $\text{Var}(\hat{\beta})$ dépend de $\mathbf{X}^\top \mathbf{X}$, donc sans hypothèse supplémentaire, il est impossible d'établir une convergence pour $\hat{\beta}$.

\mathcal{H}_3 : La matrice de design \mathbf{X} est telle que lorsque $n \rightarrow \infty$

$$\frac{1}{n}(\mathbf{X}^\top \mathbf{X}) \rightarrow \mathbf{Q}$$

où \mathbf{Q} est une matrice définie positive.

Sous les hypothèses, $\mathcal{H}_1 - \mathcal{H}_3$ l'estimateur des MCO $\hat{\beta}$ est convergent en moyenne quadratique

Aspects Computationnels

```
1 > data(apartments, package = "DALEX")
2 > n = nrow(apartments)
3 > X = matrix(c(rep(1,n), apartments$construction.year,
4               apartments$surface,
5               apartments$no.rooms),n,4)
6 > Y = matrix(apartments$m2.price,n,1)
7 > head(cbind(Y,X))
8      [,1] [,2] [,3] [,4] [,5]
9 [1,] 5897    1 1953   25    1
10 [2,] 1818    1 1992  143    5
11 [3,] 3643    1 1937   56    2
12 [4,] 3517    1 1995   93    3
13 [5,] 3013    1 1992  144    5
14 [6,] 5795    1 1926   61    2
```

Ici $n = 1,000$, $\mathbf{Y} \in \mathbb{R}^n$ et \mathbf{X} est une matrice ($n \times (1 + 3)$).

Aspects Computationnels

$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ est l'estimateur par moindres carrés

```
1 > solve(t(X)%*%X)%*%(t(X)%*%Y)
2           [,1]
3 [1,] 6295.7094722
4 [2,] -0.8829108
5 [3,] -9.3826890
6 [4,] -80.6138913
```

$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ et $\hat{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$. On note

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum \hat{\epsilon}_i^2$$

```
1 > ypred = X %*% solve(t(X)%*%X)%*%(t(X)%*%Y)
2 > residus = Y-ypred
3 > (sigma2 = sum(residus^2)/(n-4))
4 [1] 611175.7
5 > sqrt(sum(residus^2)/(n-4))
6 [1] 781.7772
```

Aspects Computationnels

$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$ peut être estimée par $\hat{\sigma}^2(\mathbf{X}^\top \mathbf{X})^{-1}$

```
1 > (sum(residus^2)/(n-4))*solve(t(X)%*%X)
2           [,1]           [,2]           [,3]           [,4]
3 [1,] 3550207 -1.807629e+03 152.91300074 -3277.950843
4 [2,] -1807 9.214743e-01 -0.07991224 1.170977
5 [3,] 152 -7.991224e-02 2.56217569 -64.046489
6 [4,] -3277 1.170977e+00 -64.04648859 1922.299635
7 > sqrt(diag((sum(residus^2)/(n-4))*solve(t(X)%*%X)))
8 [1] 1884.19946 0.95993 1.60067 43.84403
```

Comme auparavant, on obtient directement ces informations via

```
1 > reg = lm(m2.price~construction.year+surface+no.rooms
2           , data=apartments)
3 > summary(reg)
```

Aspects Computationnels

On retrouve les $\hat{\beta}_j$, et un estimateur des $\text{Var}(\hat{\beta}_j)$

```
1 > summary(reg)
2
3 Coefficients:
4             Estimate Std. Error t value Pr(>t)
5 (Intercept)  6295.7095  1884.1995   3.341 0.000865 ***
6 const.year   -0.8829    0.9599  -0.920 0.357920
7 surface      -9.3827    1.6007  -5.862 6.22e-09 ***
8 no.rooms     -80.6139   43.8440  -1.839 0.066264 .
9
10 Residual standard error: 781.8, 996 degrees of freedom
11 Multiple R-squared:  0.2588, Adjusted R-squared:  0.2566
12 F-statistic: 115.9 on 3 and 996 DF, p-value: < 2.2e-16
```

et un estimateur de σ , l'écart-type des résidus ε_i .

Aspects Computationnels

Note On a un estimateur de $\text{Var}(\hat{\beta})$ avec

```
1 > vcov(reg)
2           (Intercept) const.year   surface   no.rooms
3 (Intercept)   3550207   -1.80e+03  152.91300 -3277.9508
4 construc.year   -1807    9.21e-01  -0.07991   1.1709
5 surface         152    -7.99e-02   2.56217  -64.0464
6 no.rooms       -3277    1.17e+00 -64.04648  1922.2996
```