

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER BISKRA

Faculté des Sciences Exactes et Sciences de la Nature et de la Vie
Département de Mathématiques



Première Année Master

Notes de Cours

Analyse de Données

Chapitre 2 : Régression linéaire multiple
(Séance 5)

Auteur des notes :

Dr. Sana BENAMEUR

Année universitaire : 2020-2021

Proposition 2.2.

Les matrices H et P sont symétriques et du rang $(p + 1)$ et $(n - p - 1)$ respectivement, on dit alors qu'elles sont idempotentes.

Démonstration.

$$\begin{aligned}
 H^2 &= \left[X (X^t X)^{-1} X^t \right] \left[X (X^t X)^{-1} X^t \right] \\
 &= X (X^t X)^{-1} X^t X (X^t X)^{-1} X^t \\
 &= X (X^t X)^{-1} X^t \\
 &= H
 \end{aligned}$$

$$\text{rang}(H) = \text{tra}(H) = p + 1.$$

$$P^2 = (I_n - H)^2 = I_n^2 + H - 2H = I_n - H = P$$

$$\text{rang}(P) = \text{tra}(P) = \text{tra}(I_n) - \text{tra}(H) = n - (p + 1) = n - p - 1.$$

□

Théorème 2.1.

Les vecteurs $\hat{\beta}$ et $\hat{\varepsilon}$ ne sont pas corrélés entre eux

Démonstration.

$$\text{cov}(\hat{\beta}, \hat{\varepsilon}) = \mathbb{E} \left[\left(\hat{\beta} - \mathbb{E}(\hat{\beta}) \right) (\hat{\varepsilon} - \mathbb{E}(\hat{\varepsilon}))^t \right]$$

On a

$$\begin{aligned}
 \mathbb{E}(\hat{\varepsilon}) &= \mathbb{E}(Y - \hat{Y}) \\
 &= \mathbb{E}(X\beta + \varepsilon - X\hat{\beta}) \\
 &= X\beta + \mathbb{E}(\varepsilon) - X\mathbb{E}(\hat{\beta}) \\
 &= 0
 \end{aligned}$$

Donc

$$\begin{aligned}
 \text{cov}(\hat{\beta}, \hat{\varepsilon}) &= \mathbb{E} \left[(\hat{\beta} - \beta) \hat{\varepsilon}^t \right] \\
 &= \mathbb{E} \left[(X^t X)^{-1} X^t \varepsilon \hat{\varepsilon}^t \right] \\
 &= \mathbb{E} \left[(X^t X)^{-1} X^t \varepsilon (PY)^t \right]
 \end{aligned}$$

$$\begin{aligned}
PY &= P(X\beta + \varepsilon) \\
&= PX\beta + P\varepsilon \\
&= \left[I_n - X(X^tX)^{-1}X^t \right] X\beta + P\varepsilon \\
&= X\beta - X\beta + P\varepsilon \\
&= P\varepsilon
\end{aligned}$$

On obtient donc

$$\begin{aligned}
\text{cov}(\hat{\beta}, \hat{\varepsilon}) &= \mathbb{E} \left[(X^tX)^{-1} X^t \varepsilon (P\varepsilon)^t \right] \\
&= \mathbb{E} \left[(X^tX)^{-1} X^t \varepsilon \varepsilon^t P \right] \\
&= \mathbb{E} \left[(X^tX)^{-1} X^t \varepsilon \varepsilon^t \left(I_n - X(X^tX)^{-1}X^t \right) \right] \\
&= (X^tX)^{-1} X^t \sigma_\varepsilon^2 I_n \left(I_n - X(X^tX)^{-1}X^t \right) \\
&= \sigma_\varepsilon^2 (X^tX)^{-1} X^t - \sigma_\varepsilon^2 (X^tX)^{-1} X^t X (X^tX)^{-1} X^t \\
&= 0
\end{aligned}$$

□

Considérons maintenant la variance de l'erreur du modèle, on a σ_ε^2 est la variance de tous les perturbations ε_i , ($i = \overline{1, n}$)

Proposition 2.3 (Biais de $\hat{\sigma}_\varepsilon^2$).

La variance de l'erreur σ_ε^2 est estimée sans biais par

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

Théorème 2.2.

Le vecteur des résidus $\hat{\varepsilon}$ vérifié

- i) $\mathbb{E}[\hat{\varepsilon}] = 0$
- ii) $\mathbb{E}[\varepsilon - \hat{\varepsilon}] = 0$
- iii) $\text{Var}(\hat{\varepsilon}) = \sigma^2 P$

Démonstration. i) et ii) sont déjà faites

iii)

$$\begin{aligned}
Var(\hat{\varepsilon}) &= \mathbb{E}[\hat{\varepsilon}\hat{\varepsilon}^t] \\
&= \mathbb{E}[(PY)(PY)^t] \\
&= \mathbb{E}[(P\varepsilon)(P\varepsilon)^t] \\
&= \mathbb{E}[P\varepsilon\varepsilon^t P^t] \\
&= \sigma_\varepsilon^2 P
\end{aligned}$$

□

2.3 Qualité d'ajustement

De même que pour la régression linéaire simple, on a une égalité triangulaire généralisée :

$$\begin{aligned}
\sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
&= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2 \\
SCT &= SCE + SCR
\end{aligned}$$

La qualité d'ajustement est jugée par le coefficient de détermination R^2 , défini par :

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\hat{\varepsilon}^t \hat{\varepsilon}}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

L'ajustement soit parfaite dès que R^2 se rapproche de 1 ($R^2 \simeq 1$).

Il est possible de prendre de plus en considération le coefficient de détermination ajusté $\overline{R^2}$, défini par :

$$\overline{R^2} = 1 - \frac{\frac{1}{n-p-1} \sum_{i=1}^n \hat{\varepsilon}_i^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{n-1}{n-p-1} (1 - R^2).$$

2.4 Lois des Estimateurs

Reprenons le modèle de régression linéaire multiple

$$Y = X\beta + \varepsilon.$$

On ajoutera l'hypothèse de normalité des erreurs

$$\varepsilon \sim \mathcal{N}_n(0, \sigma_\varepsilon^2 I_n),$$

$\mathcal{N}_n(.,.)$: désigne la loi normale dans \mathbb{R}^n (la loi normale multivariée). Ainsi qu'il suffit que les erreurs soient indépendantes et identiquement distribuées (i.i.d).

Y est donc de loi normale dans \mathbb{R}^n

$$Y \sim \mathcal{N}_n(X\beta, \sigma_\varepsilon^2 I_n)$$

Théorème 2.3.

- i) $\hat{\beta} \sim \mathcal{N}_{p+1}(\beta, \sigma_\varepsilon^2 (X^t X)^{-1})$
- ii) $(n - p - 1) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-p-1}^2.$

2.5 Tests et Intervalles de Confiances

On a d'après le théorème 2.3 et si σ_ε^2 est connue. Notons V_j est le $j^{\text{ème}}$ terme diagonal de la matrice $(X^t X)^{-1}$, alors

$$\frac{\hat{b}_j - b_j}{\sigma_\varepsilon \sqrt{V_j}} \sim \mathcal{N}(0, 1)$$

Si σ_ε^2 est inconnue, on déduit que la statistique

$$\frac{\hat{b}_j - b_j}{\hat{\sigma}_\varepsilon \sqrt{V_j}} \sim \mathcal{T}_{n-p-1}$$

où \mathcal{T}_{n-p-1} la loi de Student à $(n - p - 1)$ degrés de liberté (dll). Ce qui permet d'obtenir une intervalle de confiance pour tout b_j et d'effectuer des tests d'hypothèses.

2.5.1 Intervalle de Confiance

Puisque $\frac{\hat{b}_j - b_j}{\hat{\sigma}_\varepsilon \sqrt{V_j}} \sim \mathcal{T}_{n-p-1}$. Alors

$$P \left(\frac{|\hat{b}_j - b_j|}{\hat{\sigma}_\varepsilon \sqrt{V_j}} < t \right) = 1 - \alpha.$$

L'intervalle de confiance au niveau de confiance $(1 - \alpha)$ pour chaque coefficient b_j du modèle, est donnée comme suit :

$$b_j = \hat{b}_j \mp t \hat{\sigma}_\varepsilon \sqrt{V_j}.$$

où t étant le fractile d'ordre $(1 - \alpha/2)$ de la loi de Student à $(n - p - 1)$ d.l.

2.5.2 Test de Student

Pour chaque paramètre b_j , les hypothèses à tester sont les suivantes :

$$\begin{cases} H_0 : b_j = b, \\ H_1 : b_j \neq b. \end{cases} \quad j = \overline{0, p}, b \in \mathbb{R}$$

Si α est le niveau de signification du test, $\alpha \in]0, 1[$, on accepte H_0 si

$$\frac{|\hat{b}_j - b|}{\hat{\sigma}_\varepsilon \sqrt{V_j}} \leq t_{1-\frac{\alpha}{2}},$$

Dans le cas où $b = 0$, on dit que l'on procède au test de Student de signification de b_j .