# Introduction à l'économétrie
# Le modèle linéaire multiple

*Ce cours vous est proposé par Olivier Baron, Maître de conférences, Université de Bordeaux et par AUNEGe, l'Université Numérique en Économie Gestion.*

## Activités

**Attention** : ceci est la version corrigée de l'activité.

### Exercice 1

#### Consigne

On dispose de *n* observations sur une variable dépendante *y* et on régresse cette variable sur une constante, de sorte que le modèle qu'on cherche à estimer s'écrit :

$y_i = \beta_0 + \varepsilon_i$ i = 1 …… *n*

**1 – Écrire ce modèle sous forme matricielle**

**2 – Montrer que le paramètre estimé $\widehat{\beta}_0$ est égal à la moyenne empirique de la variable Y.**

**3 – Montrer que le coefficient de détermination associé à cette régression est nul.**

#### Correction

**1 – Écrire ce modèle sous forme matricielle**

Sous forme matricielle le modèle s'écrit : $\underset{(n,1)}{y} = \underset{(n,1)}{X} . \underset{(1,1)}{\beta_0} + \underset{(n,1)}{\varepsilon}$ où la matrice X est un vecteur colonne à n lignes composé uniquement de 1 : $\underset{(n,1)}{X} = \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$

**2 – Montrer que le paramètre estimé $\widehat{\beta}_0$ est égal à la moyenne empirique de la variable Y.**

On sait que le paramètre estimé $\hat{\beta}_0$ est égal à $(X'X)^{-1}X'y$.

La matrice X étant une matrice colonne, sa transposée X' est une matrice ligne et donc :

$$X'X = (1 \dots \dots 1).\begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix} = n$$

On déduit que $(X'X)^{-1} = \frac{1}{n}$ et donc que :

$$\hat{\beta}_0 = \frac{1}{n}(1 \dots \dots 1).\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

**3 – Montrer que le coefficient de détermination associé à cette régression est nul.**

La valeur prédite de y étant constante puisqu'égale à $\hat{\beta}_0$ et donc à $\bar{y}$, sa variance est nulle. Le coefficient de détermination, rapport entre la variance de $\hat{y}$ et celle de y, est, par conséquent, lui aussi nul.

# Exercice 2

## Consignes

Soit un modèle à 2 variables explicatives de la forme :

$$y_i = \beta_0 + \beta_1.x_{1,i} + \beta_2.x_{2,i} + \varepsilon_i$$

On dispose pour estimer ce modèle de 5 observations sur la variable dépendante $y$ et les deux variables explicatives $x_1$ et $x_2$. Les observations sont les suivantes :

$y_1 = 3;\ y_2 = 1;\ y_3 = 8;\ y_4 = 3;\ y_5 = 5$

$x_{1,1} = 3;\ x_{1,2} = 1;\ x_{1,3} = 5;\ x_{1,4} = 2;\ x_{1,5} = 4$

$x_{2,1} = 5;\ x_{2,2} = 4;\ x_{2,3} = 6;\ x_{2,4} = 4;\ x_{2,5} = 6$

**1 – Présenter le modèle à régresser sous forme matricielle et les matrices qui le composent.**

**2 – Estimer les coefficients $\boldsymbol{\beta_0, \beta_1}$ et $\boldsymbol{\beta_2}$ du modèle par calcul matriciel.**

**3 – En déduire l'expression de $\underset{(5,1)}{\hat{y}}$, vecteur des valeurs ajustées de y, et celle de $\underset{(5,1)}{\hat{\varepsilon}}$, vecteur des résidus estimés.**

**4 – Vérifier que le point moyen appartient à l'hyperplan de régression.**

**5 – Vérifier que la moyenne des valeurs ajustées est égale à la moyenne de la variable dépendante et que la moyenne des résidus estimés est nulle.**

**6 – Calculer le coefficient de détermination associé à cette estimation.**

**7 – Calculer le coefficient de détermination ajusté.**

**8 – Donner une estimation de la variance de la perturbation.**

**9 – En notant $X$ la matrice des variables explicatives définie à la question 1 de l'exercice, on donne $\underset{(3,3)}{(X'X)^{-1}} = \begin{pmatrix} \mathbf{26,7} & \mathbf{4,5} & \mathbf{-8} \\ \mathbf{4,5} & \mathbf{1} & \mathbf{-1,5} \\ \mathbf{-8} & \mathbf{-1,5} & \mathbf{2,5} \end{pmatrix}$. En déduire la valeur des écarts-types estimés des coefficients estimés calculés à la question 2 de l'exercice.**

## Correction

**1 – Présenter le modèle à régresser sous forme matricielle et les matrices qui le composent.**

**Le modèle s'écrit sous forme matricielle : $\underset{(5,1)}{\boldsymbol{y}} = \underset{(5,3)}{\boldsymbol{X}} . \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(5,1)}{\boldsymbol{\varepsilon}}$ avec :**

$$\underset{(5,1)}{y} = \begin{pmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{pmatrix}$$

$$\underset{(5,3)}{X} = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix}$$

$$\underset{(3,1)}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\underset{(5,1)}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$$

**2 – Estimer les coefficients $\boldsymbol{\beta_0}, \boldsymbol{\beta_1}$ et $\boldsymbol{\beta_2}$ du modèle par calcul matriciel.**

Par définition des MCO, le vecteur $\underset{(3,1)}{\hat{\beta}}$ qui minimise la somme des carrés des résidus estimés doit satisfaire le système des équations normales, soit :

$$\underset{(3,5)}{X'} \cdot \underset{(5,3)}{X} \cdot \underset{(3,1)}{\hat{\beta}} = \underset{(3,5)}{X'} \cdot \underset{(5,1)}{y}$$

$$\Leftrightarrow \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 1 & 5 & 2 & 4 \\ 5 & 4 & 6 & 4 & 6 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 1 & 5 & 2 & 4 \\ 5 & 4 & 6 & 4 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} 5\hat{\beta}_0 + 15\hat{\beta}_1 + 25\hat{\beta}_2 = 20 \\ 15\hat{\beta}_0 + 55\hat{\beta}_1 + 81\hat{\beta}_2 = 76 \\ 25\hat{\beta}_0 + 81\hat{\beta}_1 + 129\hat{\beta}_2 = 109 \end{cases}$$

Pour résoudre ce système linéaire, on utilise la méthode du Pivot (Gauss-Jordan). Le but est d'obtenir, à partir du système initial, un système triangulaire équivalent, plus simple à résoudre. On retranche 3 fois la première équation à la deuxième et 5 fois la première équation à la troisième. Ces opérations ont pour conséquences d'éliminer l'inconnue $\hat{\beta}_0$ des équations 2 et 3. On obtient :

$$\begin{cases} 5\hat{\beta}_0 + 15\hat{\beta}_1 + 25\hat{\beta}_2 = 20 \\ 10\hat{\beta}_1 + 6\hat{\beta}_2 = 16 \\ 6\hat{\beta}_1 + 4\hat{\beta}_2 = 9 \end{cases}$$

Pour éliminer l'inconnue $\hat{\beta}_1$ de l'équation 3, on retranche à cette dernière $\frac{3}{5}$ de fois la nouvelle équation 2, ce qui conduit au système :

$$\begin{cases} 5\hat{\beta}_0 + 15\hat{\beta}_1 + 25\hat{\beta}_2 = 20 \\ 10\hat{\beta}_1 + 6\hat{\beta}_2 = 16 \\ \frac{2}{5}\hat{\beta}_2 = -\frac{3}{5} \end{cases}$$

Le système ainsi obtenu admet les mêmes solutions que le système initial et est plus simple à résoudre puisqu'il suffit de remonter la chaine d'équations. On trouve :

$$\hat{\beta}_2 = -\frac{3}{2}; \quad \hat{\beta}_1 = \frac{5}{2}; \quad \hat{\beta}_0 = 4 \text{ soit } \underset{(3,1)}{\hat{\beta}} = \begin{pmatrix} 4 \\ \frac{5}{2} \\ -\frac{3}{2} \end{pmatrix}.$$

**3 – En déduire l'expression de $\underset{(5,1)}{\hat{y}}$, vecteur des valeurs ajustées de y, et celle de $\underset{(5,1)}{\hat{\varepsilon}}$, vecteur des résidus estimés.**

Le vecteur des valeurs ajustées de *y* est par définition égal à : $\underset{(5,1)}{\hat{y}} = \underset{(5,3)}{X} \cdot \underset{(3,1)}{\hat{\beta}}$. Compte tenu du résultat obtenu à la question précédente, on obtient :

$$\underset{(5,1)}{\hat{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix} \begin{pmatrix} 4 \\ 2,5 \\ -1,5 \end{pmatrix} = \begin{pmatrix} 4 \\ 0,5 \\ 7,5 \\ 3 \\ 5 \end{pmatrix}$$

La connaissance du vecteur $\underset{(5,1)}{\hat{y}}$ permet de déduire le vecteur des résidus estimés $\underset{(5,1)}{\hat{\varepsilon}}$ puisque, par définition : $\underset{(5,1)}{\hat{\varepsilon}} = \underset{(5,1)}{y} - \underset{(5,1)}{\hat{y}}$ . On obtient donc :

$$\underset{(5,1)}{\hat{\varepsilon}} = \underset{(5,1)}{y} - \underset{(5,1)}{\hat{y}} = \begin{pmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{pmatrix} - \begin{pmatrix} 4 \\ 0,5 \\ 7,5 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 0,5 \\ 0,5 \\ 0 \\ 0 \end{pmatrix}$$

**4 – Vérifier que le point moyen appartient à l'hyperplan de régression.**

L'hyperplan de régression est ici le plan d'équation : $\hat{y} = 4 + 2,5.x_1 - 1,5.x_2$

Pour répondre à la question posée, il suffit de vérifier que les coordonnées du point moyen satisfont l'équation précédente. Le point moyen, G, a pour composantes :

$G = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{y} \end{pmatrix}$ avec $\bar{x}_1 = 3$ et $\bar{x}_2 = 5$.

En substituant ces valeurs dans l'équation du plan de régression on a :

$4 + 2,5 * 3 - 1,5 * 5 = 4 = \bar{y}$

Les coordonnées du point moyen vérifient l'équation du plan de régression et donc ce point appartient au dit plan.

**5 – Vérifier que la moyenne des valeurs ajustées est égale à la moyenne de la variable dépendante et que la moyenne des résidus estimés est nulle.**

La vérification est immédiate :

$\bar{\hat{y}} = \frac{1}{5}\sum_{i=1}^{5}\hat{y}_i = 4$ et $\bar{\hat{\varepsilon}} = \frac{1}{5}\sum_{i=1}^{5}\hat{\varepsilon}_i \quad = 0$

**6 – Calculer le coefficient de détermination associé à cette estimation.**

Le coefficient de détermination est une mesure de la qualité de l'ajustement réalisé. Par définition, le coefficient de détermination représente la part de la variance de la variable dépendante qui est expliquée par le modèle, soit dans le cas présent :

$$R^2 = \frac{\sum_{i=1}^{5}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{5}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{5}(\hat{\varepsilon}_i)^2}{\sum_{i=1}^{5}(y_i - \bar{y})^2}$$

Nous connaissons $\sum_{i=1}^{5}(\hat{\varepsilon}_i)^2 = 1 + 0{,}25 + 0{,}25 = 1{,}5$

Il reste donc à évaluer :

$\sum_{i=1}^{5}(y_i - \bar{y})^2 = (3-4)^2 + (1-4)^2 + (8-4)^2 + (3-4)^2 + (5-4)^2 = 28$

On en déduit que : $R^2 = 1 - \frac{1{,}5}{28} \approx 0{,}946.$

Le modèle explique 94,6% de la variance de la variable dépendante.

**7 – Calculer le coefficient de détermination ajusté.**

Le coefficient de détermination ajusté, noté $\bar{R}^2$, est calculé pour palier certaines limites du coefficient de détermination. Par définition :

$$\bar{R}^2 = 1 - \frac{\frac{\sum_{i=1}^{n}(\hat{\varepsilon}_i)^2}{n-k}}{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}} = 1 - \frac{\frac{1{,}5}{2}}{\frac{28}{4}} = 1 - \frac{0{,}75}{7} = 0{,}893$$

**8 – Donner une estimation de la variance de la perturbation.**

On a vu que la quantité $\frac{SCR}{n-k} = \frac{\underset{(1,5)}{\hat{\varepsilon}'}\underset{(5,1)}{\hat{\varepsilon}}}{5-3} = \frac{\sum_{i=1}^{5}\hat{\varepsilon}_i{}^2}{2}$ est un estimateur sans biais de la variance de la perturbation, $\sigma^2$. On déduit immédiatement que :

$\hat{\sigma}^2 = \frac{1{,}5}{2} = 0{,}75.$

**9 – En notant $X$ la matrice des variables explicatives définie à la question 1 de l'exercice, on donne $\underset{(3,3)}{(X'X)^{-1}} = \begin{pmatrix} 26,7 & 4,5 & -8 \\ 4,5 & 1 & -1,5 \\ -8 & -1,5 & 2,5 \end{pmatrix}$. En déduire la valeur des écarts-types estimés des coefficients estimés calculés à la question 2 de l'exercice.**

On a montré que, sous les hypothèses $H_1$, $H_2$, $H_3$ et $H_4$, la matrice de variances-covariances du vecteur $\hat{\beta}$ s'écrivait :

$\underset{(3,3)}{\text{Var}}\,\hat{\beta} = \sigma^2.(X'X)^{-1}$. Dans cette matrice, les variances des coefficients estimés s'y trouvent sur la diagonale principale. Mais, comme on l'a vu, dans l'expression précédente figure $\sigma^2$, qui est un paramètre inconnu. Dans la pratique, $\underset{(3,3)}{\text{Var}}\,\hat{\beta}$ est donc inconnue et doit donc être estimée. Cela revient à estimer $\sigma^2$ et on utilise pour cela l'estimateur $\hat{\sigma}^2$, présenté à la question précédente. On déduit alors :

$\widehat{Var}\hat{\beta}_0 = \hat{\sigma}^2.26{,}7 = 0{,}75 * 26{,}7 = 20{,}025 \Rightarrow \hat{\sigma}(\hat{\beta}_0) = \sqrt{20{,}025} = 4{,}475$

$\widehat{Var}\hat{\beta}_1 = \hat{\sigma}^2.1 = 0{,}75 * 1 = 0{,}75 \Rightarrow \hat{\sigma}(\hat{\beta}_1) = \sqrt{0{,}75} = 0{,}866$

$\widehat{Var}\hat{\beta}_2 = \hat{\sigma}^2.2{,}5 = 0{,}75 * 2{,}5 = 1{,}875 \Rightarrow \hat{\sigma}(\hat{\beta}_2) = \sqrt{1{,}875} = 1{,}369$

## Complément

À partir des données de l'exercice, on peut réaliser l'estimation en utilisant un logiciel économétrique (ici XLSTAT). On retrouve la plupart des résultats obtenus « manuellement » :

Statistiques descriptives :

| Variable | Observations | Obs. avec données manquantes | Obs. sans données manquantes | Minimum | Maximum | Moyenne | |
|---|---|---|---|---|---|---|---|
| Y | 5 | 0 | 5 | 1,000 | 8,000 | 4,000 | $\bar{y}$ |
| X1 | 5 | 0 | 5 | 1,000 | 5,000 | 3,000 | $\bar{x}_1$ |
| X2 | 5 | 0 | 5 | 4,000 | 6,000 | 5,000 | $\bar{x}_2$ |

Matrice de corrélation :

| | X1 | X2 | Y |
|---|---|---|---|
| X1 | **1** | 0,949 | 0,956 |
| X2 | 0,949 | **1** | 0,850 |
| Y | 0,956 | 0,850 | **1** |

**Régression de la variable Y :**

Coefficients d'ajustement (Y) :

| | | |
|---|---|---|
| Observations | 5 | *n* |
| Somme des poids | 5 | |
| DDL | 2 | *n-k* |
| $R^2$ | 0,946 | $R^2$ |

| | | |
|---|---|---|
| R² ajusté | 0,893 | $\bar{R}^2$ |
| MCE | 0,750 | $\hat{\sigma}^2$ |
| RMCE | 0,866 | $\hat{\sigma}$ |

Analyse de la variance (Y) :

| Source | DDL | Somme des carrés | Moyenne des carrés | F | Pr > F | |
|---|---|---|---|---|---|---|
| Modèle | 2 | 26,500 | 13,250 | 17,667 | 0,054 | |
| Erreur | 2 | 1,500 | 0,750 | | | $\sum_{i=1}^{5} \hat{\varepsilon}_i^{\,2}$ |
| Total corrigé | 4 | 28,000 | | | | $\sum_{i=1}^{5} (y_i - \bar{y})^2$ |

Paramètres du modèle (Y) :

| Source | Valeur | Erreur standard | t | Pr > \|t\| | | |
|---|---|---|---|---|---|---|
| Constante | 4,000 | 4,475 | 0,894 | 0,466 | $\hat{\beta}_0$ | $\hat{\sigma}(\hat{\beta}_0)$ |
| X1 | 2,500 | 0,866 | 2,887 | 0,102 | $\hat{\beta}_1$ | $\hat{\sigma}(\hat{\beta}_1)$ |
| X2 | -1,500 | 1,369 | -1,095 | 0,388 | $\hat{\beta}_2$ | $\hat{\sigma}(\hat{\beta}_2)$ |

# Références

**Comment citer ce cours ?**

Introduction à l'économétrie, Olivier Baron, AUNEGe (http://aunege.fr), CC – BY NC ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).