

Interpréter les coefficients d'une régression linéaire

Pour des raisons pédagogiques, nous utiliserons une application de la régression linéaire par moindres carrés afin d'apprendre à interpréter les coefficients d'un modèle. Considérons l'exemple très classique dans lequel nous voudrions estimer l'impact du nombre d'années d'études d'un individu sur son salaire. Nous disposons de données en coupe (ie. l'unité d'observation est individuelle) avec N observations et nous avons comme variables :

Sal_i : le salaire de l'individu i

$Educ_i$: le nombre d'années d'études de l'individu i

Age_i : l'âge de l'individu

$Sexe_i$: le sexe de l'individu

1 Modèle niveau-niveau

Considérons le modèle linéaire suivant estimé par moindres carrés :

$$Sal_i = \beta_0 + \beta_1 Educ_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i \quad \forall i \in \{1; N\} \quad (1)$$

Avec ε_i le terme d'erreur.

Dans l'équation (1), le coefficient β_1 s'interprète comme l'**effet marginal** d'une année supplémentaire d'études $Educ_i$ sur le salaire Sal_i . Elle correspond à la variation de β_1 unités du salaire de l'individu induite par la variation d'une unité du niveau d'études *toutes choses égales par ailleurs*¹, c'est-à-dire en prenant en compte l'âge et le sexe de l'individu. Formellement, il s'agit de la dérivée partielle. En effet :

$$\frac{\partial Sal_i}{\partial Educ_i} = \frac{\partial(\beta_0 + \beta_1 Educ_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i)}{\partial Educ_i}$$

$$\Leftrightarrow \frac{\partial Sal_i}{\partial Educ_i} = \beta_1 \quad (2)$$

2 Modèle log-log

Considérons le même modèle que précédemment mais dans lequel la variable dépendante Sal_i et la variable indépendante $Educ_i$ sont exprimées en logarithme :

$$\ln(Sal)_i = \beta_0 + \beta_1 \ln(Educ)_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i \quad \forall i \in \{1; N\} \quad (3)$$

Avec ε_i le terme d'erreur.

Pour savoir comment interpréter le coefficient β_1 , il suffit d'étudier comme précédemment la dérivée partielle de Sal_i par rapport à $Educ_i$. Pour ce faire, nous pouvons réécrire l'équation (3) en réalisant une transformation exponentielle :

$$\begin{aligned} Sal_i &= e^{\beta_0 + \beta_1 \ln(Educ)_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i} \\ &= e^{\beta_1 \ln(Educ)_i} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i} \\ &= Educ_i^{\beta_1} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i} \end{aligned}$$

1. Vous pouvez éventuellement être pédant et utiliser la locution latine : *ceteris paribus sic stantibus*.

Nous pouvons ensuite dériver Sal_i par rapport à $Educ_i$:

$$\begin{aligned}\frac{\partial Sal_i}{\partial Educ_i} &= \frac{\partial (Educ_i^{\beta_1} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i})}{\partial Educ_i} \\ \Leftrightarrow \frac{\partial Sal_i}{\partial Educ_i} &= \beta_1 Educ_i^{\beta_1 - 1} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i} \\ \Leftrightarrow \frac{\partial Sal_i}{\partial Educ_i} &= \beta_1 \frac{Educ_i^{\beta_1} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i}}{Educ_i} \\ \Leftrightarrow \frac{\partial Sal_i}{\partial Educ_i} &= \beta_1 \frac{Sal_i}{Educ_i}\end{aligned}$$

En isolant β_1 , on obtient :

$$\boxed{\beta_1 = \frac{Educ_i}{\partial Educ_i} \frac{\partial Sal_i}{Sal_i}} \quad (4)$$

On reconnaît donc bien ici une **élasticité**. Elle peut être interprétée comme le changement de β_1 % du salaire induite par un changement du nombre d'années d'études d'un pourcent, toutes choses égales par ailleurs. Notons qu'il convient de parler d'élasticité partielle puisque la régression prend en compte le sexe et l'âge de l'individu.

3 Modèle Log-niveau

Considérons le modèle de régression avec la variable dépendante Sal_i en logarithme :

$$\ln(Sal)_i = \beta_0 + \beta_1 Educ_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i + \varepsilon_i \quad \forall i \in \{1; N\} \quad (5)$$

Avec ε_i le terme d'erreur.

De la même manière, on réalise une transformation exponentielle de l'équation (5) :

$$\begin{aligned}Sal_i &= e^{\beta_0 + \beta_1 Educ_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i} \\ \Leftrightarrow Sal_i &= e^{\beta_1 Educ_i} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i}\end{aligned}$$

On dérive ensuite Sal_i par rapport à $Educ_i$:

$$\begin{aligned}\frac{\partial Sal_i}{\partial Educ_i} &= \beta_1 e^{\beta_1 Educ_i} e^{\beta_0 + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i} \\ \Leftrightarrow \frac{\partial Sal_i}{\partial Educ_i} &= \beta_1 Sal_i\end{aligned}$$

Isolons β_1 :

$$\beta_1 = \frac{\partial Sal_i}{\partial Educ_i Sal_i}$$

Multiplions par 100 l'équation :

$$\boxed{100 \times \beta_1 = \frac{100 \frac{\partial Sal_i}{\partial Educ_i}}{Sal_i} = \frac{\% \Delta Sal_i}{\partial Educ_i}} \quad (6)$$

Ainsi, on peut interpréter $100 \times \beta_1$ comme le **changement en pourcentage du salaire lorsque le niveau d'études augmente d'une unité**, toutes choses égales par ailleurs : lorsque le niveau d'études augmente d'une unité, le salaire augmente donc de $100 \times \beta_1$ à âge et sexe fixés.

4 Modèle niveau-log

Enfin, considérons le cas où cette fois-ci la variable dépendante est en niveau et la variable indépendante est en logarithme :

$$Sal_i = \beta_0 + \beta_1 \ln(Educ)_i + \beta_2 Age_i + \beta_3 Sexe_i + \varepsilon_i \quad \forall i \in \{1; N\} \quad (7)$$

Avec ε_i le terme d'erreur.

Dérivons Sal_i par rapport à $Educ_i$:

$$\begin{aligned} \frac{\partial Sal_i}{\partial Educ_i} &= \frac{\beta_1}{Educ_i} \\ \Leftrightarrow \beta_1 &= \frac{\partial Sal_i}{\frac{\partial Educ_i}{Educ_i}} \end{aligned}$$

Divisons par 100 de part et d'autre l'équation :

$$\boxed{\frac{\beta_1}{100} = \frac{\partial Sal_i}{100 \frac{\partial Educ_i}{Educ_i}} = \frac{\partial Sal_i}{\% \Delta Educ_i}} \quad (8)$$

Cette fois-ci, $\frac{\beta_1}{100}$ s'interprète comme le **changement en unité du salaire par rapport à une augmentation d'un pourcent du niveau d'études**, toutes choses égales par ailleurs : lorsque le nombre d'années d'étude augmente d'un pourcent, le salaire augmente de $\frac{\beta_1}{100}$ unités à âge et sexe fixés.