Exercices d'économétrie

TD 2 : Le modèle de régression linéaire simple.

Janvier, 2020

Exercice 2.1

On utilise la base de données `401k` pour cet exercice. Ces données cherchent à rendre compte de la participation des travailleurs au plan d'épargne-pension « 401k», un système de retraite par capitalisation très largement utilisé aux États-Unis. La variable `prate` représente le taux de participation des travailleurs. La variable `mrate` mesure la proportion de la contribution venant de l'employeur relativement à celle du salarié (si `mrate` est égal à 1/2 alors une contribution de 1$ du travailleur est complétée par une contribution de 0,5$ de l'employeur).

i Calculez le taux de participation moyen ainsi que la contribution relative moyenne de l'employeur dans l'échantillon.

ii Estimer le modèle de régression simple pour obtenir :

$$\widehat{\texttt{prate}}_i = \hat{\beta}_0 + \hat{\beta}_1 \texttt{mrate}_i$$

iii Interprétez les coefficients estimés.

iv Prédisez `prate` lorsque `mrate` est égal à 3,5.

v Quel pourcentage de la variation de `prate` est expliqué par `mrate` ?

Exercice 2.2

On utilise la base de données `sleep75` pour cet exercice. Ces données cherchent à rendre compte de l'arbitrage entre le temps passé à dormir (par semaine) et le temps passé à travailler (on ne compte que les activités rémunérées). Le sens de la causalité n'est pas évident, mais dans la suite nous considérerons la quantité de sommeil comme la variable dépendante, c'est-à-dire le modèle suivant :

$$\texttt{sleep}_i = \beta_0 + \beta_1 \texttt{totwrk}_i + u_i$$

`sleep` est le nombre de minutes de sommeil par semaine, `totwrk` le nombre de minutes travaillées par semaine.

i Estimez les paramètres et reportez le R^2. Interprétez la constante de la régression.

ii Calculez la variation du temps passé à dormir si la quantité de travail augmente de deux heures.

Exercice 2.3

On s'intéresse à l'investissement en Recherche et Développement des firmes de l'industrie chimique. La variable `rd` représente le montant des dépenses en R&D (en millions de dollars), la variable `sales` représente le montant des dépenses des firmes (toujours en million de dollar).

i Écrivez un modèle impliquant une élasticité constante entre les deux variables.

ii Estimez le modèle en utilisant la base de données `rdchem`. Quelle est l'élasticité estimée entre `rd` et `sales` ?

Exercice 2.4

Dans cet exercice nous n'utiliserons pas de données mais des données engendrées aléatoirement, afin d'illustrer les propriétés de l'estimateur des Moindres Carrés Ordinaires.

i Créez cinq cents observations pour une variable explicative, x_i, en tirant dans une loi uniforme entre 0 et 10. Utilisez la fonction texttrand dans le module Python `numpy.random`.

ii Générez cinq cents erreurs, u_i, en tirant dans une loi normale d'espérance nulle et de variance 36. Vous utiliserez la fonction `randn` du même module `numpy.random`. Calculez la moyenne et l'écart type des erreurs u_i. Que remarquez-vous ?

iii Créez la variable y selon le modèle :

$$y_i = b_0 + b_1 x_i + u_i$$

avec $\beta_0 = 1$ et $\beta_1 = 2$. Représentez graphiquement l'échantillon (utilisez le module `matplotlib`).

iv En utilisant les données engendrées dans les questions précédentes, estimez le modèle :

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

par les Moindres Carrés Ordinaires. Comparez $\hat{\beta}_0$ et $\hat{\beta}_1$ avec les vraies valeurs du modèles b_0 et b_1 (c'est-à-dire les valeurs de la constante et de la pente utilisées pour engendrer la variable expliquée y).

v Calculez les résidus estimés $\hat{e}$ et vérifiez que l'on a bien (aux erreurs numériques près) :

$$\sum_{i=1}^{500} \hat{e}_i = 0 \quad \text{et} \quad \sum_{i=1}^{500} x_i \hat{e}_i = 0$$

vi Calculez les deux mêmes sommes en utilisant les « vraies » erreurs plutôt que les résidus estimés. Commentez.

vii En utilisant une boucle engendrez 10 échantillons $(y_i, x_i)_{i=1}^{500}$ en tirant des erreurs différentes dans la même loi normale. Pour chaque échantillon estimez le modèle par les MCO et stockez les valeurs de $\hat{\beta}_1$ dans un vecteur. À la sortie de la boucle calculez la moyenne des $\hat{\beta}_1$.

viii Reprenez la question précédente avec 100 échantillons simulés, puis 1000 échantillons simulés. Comparez les moyennes de $\hat{\beta}_1$ calculées sur la base de 10, 100 et 1000 échantillons. Que remarquez vous ? À quoi correspond cette moyenne ?