

Exercices d'économétrie

TD 3 : Le modèle de régression multiple.

Janvier, 2020

Exercice 3.1

Dans le domaine de la santé publique, un problème important pour les décideurs politiques est de déterminer les effets de la consommation de tabac par la mère durant la grossesse (**cigs**) sur la santé de son enfant. Le poids à la naissance (**bwght**) est une mesure de la santé infantile. Un poids à la naissance trop faible peut augmenter le risque de contracter différentes maladies. Il existe d'autres facteurs qui affectent le poids du bébé à la naissance. Comme ces facteurs sont susceptibles d'être corrélés avec la consommation de cigarettes, nous devons les prendre en compte. Par exemple, un revenu familial (**faminc**) plus élevé facilite l'accès aux soins avant la naissance et assure également une meilleure alimentation de la mère. Considérons l'équation suivante, qui tient compte du revenu comme variable de contrôle :

$$\text{bwght}_i = \beta_0 + \beta_1 \text{cigs}_i + \beta_2 \text{faminc}_i + u_i$$

- i Quel est le signe le plus probable pour β_2 ?
- ii Pensez-vous que **cigs** et **faminc** soient susceptibles d'être corrélés ? Expliquez pourquoi la corrélation pourrait être positive ou négative.
- iii En utilisant la base de données **bwght**, estimez l'équation par les MCO avec et sans **faminc**. L'ajout de **faminc** change-t-il de manière substantielle l'effet estimé de **cigs** sur **bwght** ?

Exercice 3.2

La base de données **ceosal2** contient un jeu de données portant sur 177 PDG aux États-Unis. Ces données peuvent être utilisées pour examiner les effets de la performance des entreprises sur le salaire des PDG.

- i Estimez un modèle expliquant le salaire annuel du PDG (**salary**) par le chiffre d'affaires (**sales**) et la valeur de marché (**mktval**) de son entreprise. Faites en sorte que ce modèle soit un modèle à élasticité constante pour les deux variables indépendantes.
- ii Ajoutez la variable **profits** au modèle utilisé dans la question précédente. Pourquoi ne doit-on pas inclure cette variable sous forme de logarithmique ? Considérez-vous que ces variables de performance des entreprises expliquent l'essentiel de la variation des salaires des PDG ?
- iii Ajoutez l'expérience du PDG (**ceoten**) comme variable explicative au modèle utilisé dans la question précédente. Toutes choses étant égales par ailleurs, quel est le « rendement » estimé d'une année supplémentaire d'expérience ?
- iv Calculez le coefficient de corrélation entre les variables $\log(\text{mktval})$ et **profits** dans l'échantillon. Ces variables sont-elles fortement corrélées ? Qu'est-ce que cela implique pour les estimateurs des MCO ?

Exercice 3.3

On utilise la base de données `meap93`.

- i Estimez un modèle :

$$\text{math10}_i = \beta_0 + \beta_1 \log(\text{expend}_i) + \beta_2 \text{lnchprg}_i + u_i$$

où `math10` est le pourcentage d'élèves réussissant le test de mathématiques, `expend` est la dépense par élève (\$), et `lnchprg` est le pourcentage d'élèves participant au « *shool lunch program* » utilisé comme une proxy de la pauvreté des élèves. Les signes des estimateurs sont-ils ceux que vous anticipiez ?

- ii Pouvez-vous interpréter l'estimateur de l'ordonnée à l'origine, $\hat{\beta}_0$?
- iii Reprenez l'estimation du modèle en éliminant la variable `lnchprg`. Comparez les résultats avec la première estimation. L'estimation de l'effet des dépenses est-elle plus grande ou plus petite que dans le premier modèle ?
- iv Calculez la corrélation entre $\log(\text{expend})$ et `lnchprg`. Pouvez-vous justifier son signe ?
- v Utilisez le résultat sur le calcul de la corrélation pour justifier la différence entre les deux estimations.

Exercice 3.4

On utilise la base de données `charity`.

- i Estimez le modèle suivant par les MCO :

$$\text{gift}_i = \beta_0 + \beta_1 \text{mailsyar}_i + \beta_2 \text{giftlast}_i + \beta_3 \text{propresp}_i + u_i$$

où `gift` est le montant des dons (en florin néerlandais), `mailsyar` est le nombre moyen de sollicitations envoyées par la poste sur l'année, `giftlast` le montant moyen du dernier don, et `propresp` le taux de réponse aux envois postaux. Comparez le R^2 de cette régression avec celui que nous obtiendrions en excluant les deux dernières variables (`giftlast` et `propresp`).

- ii Interprétez le coefficient associé à `mailsyar`. Est-il plus grand ou plus petit que le coefficient estimé à partir de la régression simple ?
- iii Interprétez le coefficient associé à `propresp`.
- iv Réestimez le modèle en ajoutant la variable `avggift` (montant moyen des dons effectués les années précédentes). Que devient l'effet estimé de `mailsyar` ?
- v Dans l'équation de la dernière estimation, qu'est-il arrivé au coefficient de `giftlast` ? Que se passe-t-il ?