

4 Régression Linéaire Multiple I

Exercice 1

Une régression multiple inclut deux régresseurs:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U.$$

- (a) Quelle est l'espérance de la variation de Y si X_1 augmente de 3 unités et X_2 ne change pas?
- (b) Quelle est l'espérance de la variation de Y si X_2 baisse de 5 unités et X_1 ne change pas?
- (c) Quelle est l'espérance de la variation de Y si X_1 augmente de 3 unités et X_2 baisse de 5 unités?
- (d) Expliquez pourquoi il est difficile d'estimer avec grande précision l'effet partiel de X_1 , en maintenant constant X_2 , si X_1 et X_2 sont fortement corrélés.

Exercice 2

Les résultats suivants ont été obtenus en utilisant les données de la mise à jour de l'Enquête sur la population de 1998 (CPS). La base de données contient des informations sur 4.000 employés à temps plein. Le niveau d'éducation le plus élevé atteint par chaque travailleur est soit un diplôme d'études secondaires soit une licence universitaire. L'âge des travailleurs est entre 25 et 34 ans. La base de données contient également des informations sur la région du pays où réside la personne, son état civil et le nombre d'enfants qu'elle a.

SHM: salaire moyen horaire (en dollars de 1998).

Université = variable binaire (1 pour les titulaires d'une licence universitaire, 0 pour un diplôme d'études secondaires).

Femme = variable binaire (1 pour les femmes, 0 pour les hommes)

Age = âge (en années)

NE = variable binaire (1 si la région est le Nord-Est, 0 sinon)

CO = variable binaire (1 si la région est le Centre-Ouest, 0 sinon)

S = variable binaire (1 si la région est le Sud, 0 sinon)

O = variable binaire (1 si la région est l'Ouest, 0 sinon)

- (a) Calculez le \bar{R}^2 pour chacune de ces régressions.

Utilisez les résultats de la colonne (1):

- (b) Est-ce que les titulaires d'une licence gagnent plus en moyenne que ceux qui n'ont qu'un diplôme d'études secondaires? Combien de plus?
- (c) Est-ce que les hommes gagnent plus en moyenne que les femmes? Combien de plus?

Utilisez les résultats de la colonne (2):

- (d) L'âge est-il un facteur important pour expliquer les revenus? Expliquez.

Table 1: Variable Dépendante: salaire horaire moyen (SHM)

| Régresseur | (1) | (2) | (3) |
|-------------------------------|-------|-------|-------|
| Université (X1) | 5,46 | 5,48 | 5,44 |
| Femme (X2) | -2,64 | -2,62 | -2,62 |
| Age (X3) | | 0,29 | 0,29 |
| NE (X4) | | 0,69 | |
| CO (X5) | | 0,60 | |
| S (X6) | | -0,27 | |
| Constante | 12,69 | 4,40 | 3,75 |
| Statistiques de la régression | | | |
| SER | 6,27 | 6,22 | 6,21 |
| R^2 | 0,176 | 0,190 | 0,194 |
| n | 4.000 | 4.000 | 4.000 |

(e) Sally a 29 ans et elle est titulaire d'une licence. Betsy a 34 ans et elle est également titulaire d'une licence. Faites une prédiction des salaires de Sally et de Betsy.

Utilisez les résultats de la colonne (3):

(f) Y a-t-il des différences régionales importantes?

(g) Pourquoi a-t-on omis la variable explicative Ouest dans la régression? Que se passerait-il si on l'incluait?

(h) Juanita est une femme du Sud de 28 ans, titulaire d'une licence universitaire. Jennifer est une femme du Centre-Ouest de 28 ans, titulaire d'une licence universitaire. Calculez l'espérance de l'écart de salaire entre Juanita et Jennifer.

Exercice 3

Les données suivantes ont été recueillies auprès d'un échantillon aléatoire de 220 maisons vendues dans une région en 2003. Soient la variable Prix (prix de la vente en milliers de \$), la variable CC (le nombre de chambres à coucher), la variable SDB (le nombre de salles de bain), la variable SURF (la taille de la maison en pieds carrés), la variable PARC (taille de la parcelle en pieds carrés), la variable AGE (l'âge de la maison en années), et la variable ETAT (une variable binaire égale à 1 si l'état général de la maison peut être qualifié de mauvais). L'estimation de la régression fournit les résultats suivants:

$$\text{Prix} = 119,2 + 0,485\text{CC} + 23,4\text{SDB} + 0,156\text{SURF} + 0,002\text{PARC} + 0,090\text{AGE} - 48,8\text{ETAT},$$

$$R^2 = 0,72, \text{SER} = 41,5.$$

- (a) Supposez qu'un propriétaire transforme une partie d'un salon en une salle de bains. Quelle est l'augmentation prévue de la valeur de la maison?
- (b) Supposez que le propriétaire ajoute une nouvelle salle de bain à son domicile, ce qui augmente la taille du logement de 100 pieds carrés. Quelle est l'augmentation de la valeur de la maison?
- (c) Quelle est la perte de valeur si un propriétaire laisse sa maison se détériorer jusqu'à ce que l'état général puisse être qualifié de mauvais?
- (d) Calculez le R^2 de la regression.

SOLUTIONS:

2. a) 0,175, 0,189 y 0,193 respectivement; b) 5,46 \$/heure de plus en moyenne; c) 2,64\$/heure de plus en moyenne; e) 15; 67 et 17; 12 respectivement; h) -0,87.
3. a) 23:400\$; b) 39:000\$; c) 48:800\$; d) 0,727.