

Exercice 4: Modélisation

Visseho Adjiwanou, PhD.

09 June 2021

PARTIE A: Déterminant du poids à la naissance

Vous travaillez avec la base de données `weight.dta` pour analyser le déterminant de la survie des enfants en Tanzanie. L'une des conditions de la survie des enfants est leur poids à la naissance. En utilisant cette base de données, nous voulons comprendre les déterminants du poids à la naissance en utilisant le niveau d'éducation de la mère, son statut au travail, l'indice de richesse du ménage et le sexe de l'enfant.

REmarque: faite attention à l'extension de la base de données et utiliser la commande appropriée pour la charger.

Question 0

- Quelle est la nature des variables `m19` et `v012`?
- Représenter graphiquement la relation entre les variables `m19` et `v012`.

Question 1

Exécutez la régression linéaire suivante: - Variable dépendante: poids à la naissance (`m19`) - Variable indépendante: âge actuel des femmes (`v012`)

Question 2

Calculez le poids de naissance prédit, nommé `p_weight`

Question 3

Comment interprétez-vous ce résultat?

Question 4

Présentez dans le même graphique le nuage de points de `m19` et `v012` et la droite de régression.

Question 5

Un de vos amis pense que l'effet de l'âge actuel des femmes sur le poids à la naissance est trop faible. Quel est le problème avec la variable âge des femmes?

Question 6

Le problème peut également dépendre de la variable dépendante. En effet, le poids à la naissance est renseignée de deux manières différentes, à partir de l'information dans les carnets de santé dans enfants, ou à partir d'une estimation faite par la femme. Vous vous en doutez bien que l'information dans le carnet de santé est plus crédible. La variable m19a vous donne des informations sur la manière dont l'information est collectée. Exécutez à nouveau la régression faite au 1) dans les deux groupes distincts de m19a. Est-ce qu'effet de l'âge change dans les deux cas? Interprétez vos résultats.

Question 7

Il y a plusieurs manières de tester la pertinence des hypothèses formulées pour l'utilisation de la méthode des moindres carrées pour estimer les paramètres des modèles de régression linéaire. Rappelons les hypothèses qui concernent les termes d'erreur. Et comme vous le voyez, la plupart de la vérification des hypothèses vont concerner les termes d'erreur. Raju Rimal présente les graphiques de diagnostics dans ce post. Documentez-vous sur l'ensemble de ses graphiques et dites et appliquer les à votre modèle estimé 1). des informations additionnelles peuvent être trouvées ici: https://wiki.qcbs.ca/r_atelier4, section 2.

Question 8

Un de vos amis a dit qu'il serait préférable de considérer le logarithme du poids à la naissance comme la variable dépendante.

- Calculez la variable logm19 représentant le logarithme de m19.
- Exécutez la régression: régressez logm19 sur v012
- Comment interprétez-vous le coefficient de v012