

Module II, exercice 2

Fichier de données: `ess2008.sav`
Sélectionner la France

Régression multiple

Nous allons chercher à expliquer le conservatisme des Français à partir du niveau de formation, de l'âge et de la religiosité.

1 Modèle théorique

- 1.1 Formuler l'hypothèse idoine.
- 1.2 Dessiner le schéma du modèle et le commenter.
- 1.3 Quels sont les indicateurs (variables) qui vous permettent de tester votre modèle ? Justifier le choix des indicateurs.
- 1.4 Quelle est l'équation de régression multiple correspondant à votre modèle ?

2 Analyse univariée

- 2.1 Faire un diagnostic des variables du modèle
- 2.2 Combien d'individus seront inclus dans le modèle ?

3 Tests empiriques du modèle

- 3.1 Faire un diagnostic des corrélations simples entre les variables du modèle (sens, force et significativité des corrélations, et éventuellement présence d'un problème de colinéarité).
- 3.2 Quel est le pouvoir explicatif du modèle ? Que peut-on dire de la significativité de notre modèle ?
- 3.3 Interpréter l'erreur moyenne de prédiction.
- 3.4 Interpréter le coefficient B_0 (constante).
- 3.5 Selon le modèle, si le nombre d'années d'études à plein-temps augmente de 10, toutes choses égales par ailleurs, de combien augmenterait ou diminuerait le niveau de conservatisme ?
- 3.6 Selon le modèle, si le nombre d'années d'études à plein-temps augmente de 2, l'âge augmente de 40 et le degré de religiosité de 5, de combien augmenterait ou diminuerait le niveau de conservatisme ?
- 3.7 Quels coefficients sont significatifs au seuil de 0.05 ? Que peut-on en conclure ?
- 3.8 Interpréter l'intervalle de confiance à 95% du coefficient B pour la variable *eduysr*.
- 3.9 Résumer le modèle à l'aide d'un schéma fléché en utilisant les coefficients standardisés. Ecrire l'équation du modèle avec les coefficients standardisés. Quelle est la variable qui

Module II, exercice 2

a l'importance relative la plus forte dans le modèle ? Et celle qui a le poids le plus faible ?

3.10 Est-ce que votre hypothèse est validée ? Justifier.

4 Etude des corrélations et diagnostic de multicollinéarité

4.1 Quel est la proportion de variance expliquée uniquement par *agea* ? Et par *eduys* ? Et par *rlgdgr* ?

4.2 De combien le R^2 diminuerait-il si on enlève *agea* de notre modèle ?

4.3 Observe-t-on un problème de multicollinéarité ? Justifier.