

Labo9: Tableau croisé - chi-carré

Visseho Adjiwanou, PhD.

20 March 2023

Questions

A partir de la base de données **cis_short_5percent_2015.csv**, répondez aux questions suivantes:

1. Créer une nouvelle variable **quintile_revenu** qui divise la distribution de revenu en 5 parties égales. Renommer les catégories de “Tres pauvre” à “Tres riche”
2. Quelle est la nature de cette variable?
3. Existe-t-il une association entre la region de résidence et le quintile de revenu?
4. Présenter un graphique pour montrer cette association.
5. Que constatez-vous aux questions 3 et 4 quand vient le temps de l’interprétation? Interprétez les résultats?
6. Quel est le degré de liberté de ce tableau?
7. Quelle est la valeur du chi-carré?
8. Est-ce que cette relation est significative au seuil 5%? Qu’en-est-il au seuil de signification de 10%?
9. Reprendre les étapes 1 à 8 cette fois-ci avec la variable **tercile_revenu**. Cette variable divise la distribution en 3 parties égales. Cette fois-ci également, garder les chiffres 1, 2 et 3 pour les catégories de la nouvelle variable. Finalement, prenez soin d’exclure les individus qui ont un revenu plus faible que 0, et ceux qui ont revenu de plus de 250000.
10. Que constatez-vous globalement en refaisant cet exercice. Quels sont les leçons apprises? les erreurs commises?

Question : échantillonnage

Prenez un échantillon de 6% de la base de données que vous nommez **base_6**.

1. Dresser le tableau de contingence entre la région et la variable **tercile_revenu**?
2. Est-ce qu’il existe une association entre ces deux variables
3. Est-ce que cette association est significative au seuil de 5%
4. Donnez quelques mesures de l’intensité de cette relation.