

TD 1 — Modèle de Poisson (R)

Année universitaire 2025–2026 ■ Parcours Économie de la santé & Développement durable

Pierre Beaucoral

Contexte

On étudie d'anciennes données reliant tabagisme et décès par cancer du poumon.

Variables : `age` (classes), `smoking status` (4 classes), `population` (centaines de milliers), `deaths` (décès annuels).

Préparation

Objectifs de ce TD

- Importer et préparer un tableau « comptages + exposition » (population à risque).
- Ajuster un GLM Poisson avec offset (log-exposition).
- Évaluer l'ajustement : déviance (vs modèle saturé) & Pearson.
- Comparer des modèles via tests de rapport de vraisemblance (LR).
- Interpréter en ratios de taux d'incidence (IRR) et produire des comptes attendus.

Import et manipulation des données

1. Importer les données `smoking_dat.xlsx`.
2. Décrire et commenter le dictionnaire des variables :
 - `age` : classes d'âge.
 - `smoking status` : 4 classes.
 - `population` : en centaines de milliers.
 - `deaths` : décès annuels.

3. Coder les variables `age` et `smoking status` en facteurs.
4. Justifier l'unité d'exposition retenue (population ou population $\times 100\,000$).

Estimations

Modèle de base

5. Ajuster un modèle Poisson log-linéaire avec effets de `smoking_status` et `age` et offset `log(exposure)`.
 - Q1 : Pourquoi utiliser des variables indicatrices plutôt que des codes numériques continus ?
6. Calculer la déviance du modèle ajusté.
 - Q2 : Interpréter la déviance et le p-value (DEV1).
7. Interpréter l'effet de l'âge sur la probabilité de décès.
 - Q3 : Que disent les coefficients d'âge en termes d'IRR ?

Ajustement du modèle

8. Réaliser les deux tests d'ajustement : déviance GOF et Pearson GOF.
 - Q4 : Justifier les degrés de liberté.
 - Q5 : Discuter les conditions d'application du test du χ^2 .

Comparaison de modèles

9. Ajuster un modèle sans la variable tabac et effectuer un test LR entre les deux modèles.
 - Q6 : Conclure sur l'impact de l'usage du tabac sur la probabilité de décès.

Variable binaire « cigarette »

10. Créer une variable binaire `cigarette_user` (=1 si l'individu fume des cigarettes, 0 sinon).
11. Ajuster un modèle avec `age + cigarette_user`.
12. Comparer ce modèle avec le modèle initial par un test LR.
 - Q7 : Le type de produit fumé influence-t-il différemment le taux de décès ?

Extensions (facultatif)

13. Calculer et présenter les IRR avec IC à 95 %.
14. Présenter les comptes observés vs attendus et commenter.
15. Vérifier la présence éventuelle de sur-dispersion et proposer, si nécessaire, un modèle adapté (Quasi-Poisson ou Négative Binomiale).