
LNG-1100 : Méthodes expérimentales et analyse de données

Les variables binaires et le nettoyage des données

Guilherme D. Garcia

fr.gdgarcia.ca

10



Plan de la séance

1. Révision : nettoyage
2. Figures pour les données binaires
3. Combinaison des scripts à partir de `source()`

Nettoyage

- Importez et nettoyez le fichier `questionnaire2.csv` (séance 9)
 - N'oubliez pas d'enlever la colonne qui indique le nom des participants
 - 👉 Réponses : consultez le fichier dans le projet
-
- Si vous avez déjà complété l'exercice :
 - Enchaînez les opérations dans une seule séquence de fonctions en utilisant le ►

Pratique

Partie II

Répondez aux questions du chapitre du livre du cours

- Travaillez dans vos équipes

Interprétation des patrons pertinents

- Créez quelques figures pour explorer les patrons des données

Pratique

Partie IV

- On vient de créer des variables binaires/dichotomisées avec `if_else()`
- Cette fonction nous permet de recoder n'importe quelle variable catégorielle

Pratique : le retour aux villes

1. Créer une nouvelle colonne, `langue`, qui dichotomise la langue officielle de la ville
2. Est-ce qu'il y a un effet de `langue` sur la `note`?

Questionnaire 3

Questions?

- Nettoyage de données + interprétation des patrons pertinents
- 5 questions; durée 2h50

Prochaine séance

Régression logistique

- Les séances 9 et 10 examinent une variable de réponse binaire : 0 ou 1
- ☞ Ce type de données ne peut pas être modélisé avec une régression linéaire (pas continu)

La semaine prochaine : chapitre 11 du livre numérique du cours

- On explorera un nouveau type de modèle : régression logistique (Garcia 2021, chap. 7, Larmarange 2023, chap. 22)

Références I

Guilherme D GARCIA : *Data visualization and analysis in second language research*. Routledge, New York NY, 2021.

Joseph LARMARANGE : *Introduction à l'analyse d'enquêtes avec R et RStudio*. 2023. Available at <https://larmarange.github.io/analyse-R/analyse-R.pdf>.