

2021-2022

M2 Expertise Sciences des Populations

ANALYSE DES DONNÉES LONGITUDINALES

Arno Muller

arno.muller@ined.fr

(Ined - Service Méthodes Statistiques)

Et vous ?

- Parcours
- Logiciel (aucun, R, SAS, les deux)
- Sujet(s) d'intérêt(s) (facultatif) (les points précédents aussi d'ailleurs)
- Projet professionnel ? (obligatoire) (nan je déconne) (mais imagine quand-même)

Plan du cours :

- 23 septembre :
 - *Présentation du cours*
 - *Introduction aux logiciels*
- 30 septembre :
 - *Présentation des différents types de données de l'AL*
 - *Manipulation des données sur R*
- 7 octobre – 21 octobre :
 - *Analyse de séquence (+ DM)*
- *Pause + Jour férié : Rattrapable si vous voulez*
- 18 novembre – 2 décembre :
 - *Modèle de durée (part 1) (+ DM facultatif)*
- 9 décembre – 13 janvier :
 - *Modèle de durée (le retour) (+ DM)*

Examen

- DM 1 : Analyse de séquences :
 - *Rendre pour le 13 novembre 2021*
 - *Application des TD : faire une analyse de séquence sur une base de donnée différentes*
- DM (facultatif) : Modèle de Durée Non Paramétrique
 - *Rendre pour le 16 décembre 2021*
 - *Application des TD sur une autre base, Kaplan-Meier*
- DM 2 ou exam final : Modèle de Durée
 - *Rendre pour le 20 janvier 2022*
 - *Application des TD : Modèle Kaplan-Meier + Modèle de Cox*

R

- Les avantages :
 - *Gratuit*
 - *Extrêmement adaptable : Package*
 - *Communauté : tout en ligne*
 - *Sortie graphique*
 - *Diffusion et valorisation de la recherche : Rmarkdown, Git, Shiny*
- Les inconvénients :
 - *Coût d'entrée*
 - *Sorties de tableaux pas sexy (je mens à moitié, possibilité de faire le lien avec Excel pour mise en page ou encore les fonctions comme kable dans Rmarkdown)*

TP 1 : Introduction à R

- Installer : <https://pbil.univ-lyon1.fr/CRAN/>
- Rstudio : <https://www.rstudio.com/products/rstudio/download/>
- Ouvrir Rstudio
- Créer un projet TP1 -> Créer un script et le sauvegarder
- Exploration de l'interface Rstudio
 - *Console* : $2+2 = ?$
 - *Script* : $2+2 = ?$
 - *Environnement*
 - *La 4^{ème} fenêtre*
 - Graphique / Help / Importer donnée / Package
- Importer les données Titanic.csv
 - *Le working directory*
 - *Les fonctions read* : ex : `read.csv()`
- Manipulation des données : tidyverse et dplyr

TP1 : Titanic

- *Objectif :*
 - *Peut-on vérifier que les femmes et les enfants ont bien été sauvé d'abord ?*

- *Exercice 1 : A partir d'un tableau croisé :*
 - *Quelle est la probabilité d'avoir survécu, pour un enfant ? pour un adulte ?*
 - *Quel est le rapport de risque de survivre pour un enfant par rapport à un adulte ?*
 - *Quel est l'odds ratio d'un enfant vs un adulte ?*

- *Exercice 2 :*
 - *Modéliser la probabilité de décéder, en fonction du sexe, de la classe et de l'âge.
On prendra en modalités de référence : enfant - homme – équipage*

- *Exercice 3 :*
 - *Calculer la probabilité qu'un enfant de sexe masculin de seconde classe ait survécu au naufrage*

- *Exercice 4 :*
 - *Entrer une interaction entre l'âge et la classe, et commenter les résultats*

Prochaine séance

- CM :
 - *Les différents types de données utilisés en analyse longitudinale*
- TD :
 - *Questions sur la séance précédente*
 - *Manipuler des jeux de données*
 - Format long / format larges
 - Merge
 - Moults fun