

# Consignes pour le projet

E. Adjakossa, P. Barbillon, P.Gloaguen et E. Parent

2021/2022

## Consignes générales

Dans le cadre du module, nous vous demandons d'appliquer la démarche de modélisation statistique vue en cours pour répondre à des questions que vous souleverez sur un jeu de données.

- Ce projet se fera en binôme ou en trinôme sur le jeu de données présenté par Matthieu Authier;
- Les données sont à utiliser pour le projet et ne doivent pas être utilisées par ailleurs, nous vous faisons confiance sur ce point.
- Sur ce jeu de données, vous souleverez 2 à 3 questions d'ordre écologique ou biologique. Une conjecture de réponse à cette question pourra être faite grâce à l'exploration de données et des interprétations graphiques. Cependant, ces conjectures doivent impérativement s'accompagner d'une réponse argumentée par une modélisation statistique vue en cours (du type modèle linéaire et ses extensions, modèle linéaire généralisé, modèle mixte, etc).
- Les modèles devront être écrits de manière formelle et complète (y compris les notations).
- L'analyse des sorties R devra être détaillée, pour justifier de votre compréhension des rendus logiciels. En particulier, pour toute analyse de sorties de tests statistiques, vous devrez stipuler quelles sont les hypothèses testées.
- Le rendu de projet étant sous la forme d'un rapport, la qualité de la rédaction sera prise en compte. Pêle-mêle, cela implique de soigner la présentation des questions, la discussion des résultats, et la conclusion, de n'inclure que des figures pertinentes dans le propos du rapport, d'en soigner la lisibilité (labelisation des axes, légende, etc.), de citer ses sources, etc. . .
- Le projet est à rendre sur ecampus le vendredi 8 Janvier à 16h. Le format attendu de rendu est détaillé ci dessous.

## Format attendu des questions

Ce projet se veut être un exercice en autonomie. Nous attendons de vous que vous exploitiez la richesse des données mises à votre disposition. Nous vous demandons de vous focaliser sur deux à trois questions qui justifient la formulation de modèles statistiques pertinents. Ces questions peuvent être indépendantes, ou reliées entre elles, selon votre choix. Nous attendons de vous de nous convaincre de la pertinence de vos questions.

Il est évidemment encouragé de se documenter sur les données et les questions qui peuvent en émerger. Vos questions peuvent donc venir de recherches bibliographiques. Evidemment, dans ce cas, vous citerez vos sources.

## Traitement/modification des données

Lors de votre analyse, vous pouvez être amenés à

- N'utiliser qu'une sous partie des données;
- Créer/supprimer/modifier des colonnes/lignes dans le tableau initial;

- Gérer les données manquantes...

Toutes ces manipulations sont inhérentes à l'analyse de données et sont donc totalement autorisées. Cependant, toutes les modifications doivent être reproductibles par le correcteur. Concrètement, cela implique que le jeu de données initial est celui distribué en classe, et commun à tous, et que toutes les modifications effectuées doivent être consignées dans votre fichier `.Rmd` (voir ci dessous). Il n'est pas autorisé de modifier vos données sous Excel, d'enregistrer le résultat et de baser votre étude dessus.

Servez vous du tutoriel utilisé en TP1 pour cette étape!

## Format attendu de rendu

Le format de rendu de projet est double. Nous vous demandons:

- 1 fichier Rmarkdown (donc au format `.Rmd`). qui contiendra tous les codes utilisés pour obtenir les résultats du rapport final.
- La version compilée de ce fichier Rmd. C'est à dire un fichier au format `.html` ou au format `.pdf`. Les codes ne doivent pas nécessairement apparaître dans ce document compilé, qui sera la version que nous lirons (ce n'est pas nécessaires, car ils seront présents dans le fichier `.Rmd`).
- En compilant votre fichier `.Rmd`, nous devons avoir le même rapport que celui que vous nous rendez.

Dans le dossier du projet était donné un fichier `squelette_rapport_rmarkdown.Rmd`, qui pourra vous servir de base. De nombreuses aides sont disponibles sur internet, notamment:

- Sur le site de Rstudio
- Sur un site dédié

Nous vous demandons un rapport d'un maximum de 10 pages, tout inclus<sup>1</sup>

## Consignes de plagiat

Au cas où deux rapports de deux groupes présentent des redondances manifestes, la note attribuée sera 0 pour les deux binômes. Un plagiat caractérisé d'une source tierce sera également sanctionné de 0.

---

<sup>1</sup>Si vous rendez au format `'html'`, une pagination n'est pas possible. Dans ce cas, essayez de conserver cet ordre de grandeur, en compilant chez vous en pdf ou en word pour évaluer la taille.