

TP1 - INFO831 :

Régression Linéaire Simple et Multiple

Antoine Bralet

Jeudi 2 Février 2023

1 Introduction

L'objectif de ce TP évalué est de mettre en pratique la méthode de régression linéaire simple et multiple. Bien sûr, au delà de l'implémentation c'est aussi et surtout l'interprétation des résultats qui va être évaluée. **Le rapport est à rendre au plus tard le Jeudi 9 Février 2023 sur le Moodle.** Le rapport peut être rédigé soit sous format .Rmd soit avec une archive .zip qui contient à la fois le code en R et le rapport en .pdf.

La Section 2 permet d'organiser correctement le jeu de données utilisé dans la Section suivante. Le jeu de données est disponible via le fichier "covidDataFrance2021.csv".

Les attentes sur ce TP sont la mise en place des fonctions vues en TD et surtout une interprétation poussée permettant d'expliquer les résultats et les comparaisons des modèles mis en place.

Les questions bonus ne sont pas nécessaires pour avoir la note maximale, ce sont essentiellement des questions de réflexion.

2 Préparation du jeu de données

Avant de pouvoir les exploiter, les données Covid doivent être pré-traitées :

- Récupérer les données dans un data frame.
- Remplacer les données manquantes (NA) par 0.
- Diviser le jeu de données en deux sous-jeux de données : le premier contiendra toutes les données cumulées. Le second contiendra toutes les données journalières et hebdomadaires. En plus de ceci, les deux jeux doivent contenir la date, le taux de positivité et le nombre de tests par cas.
- Mettre la date en attribut de ligne des data frame et la supprimer des données.
- Visualiser quelques combinaisons de variables qui vous semblent pertinentes.
- Mettre en place une fonction qui prend en argument le nom du fichier et qui renvoie les deux jeux de données d'intérêt dans une liste.

3 Régression Linéaire

On cherche ici à expliquer le nombre de personnes en réanimation (`weekly_icu_admissions`) à partir des données quotidiennes et hebdomadaires.

- Avant de toucher au code, rappeler en une phrase en quoi consiste une régression linéaire.

3.1 Régression Simple

- Mettre en place une régression linéaire simple en prenant `weekly_hosp_admissions` comme variable explicative.
- Faire de même en prenant `new_deaths` comme variable explicative puis `new_tests`.
- Vérifier numériquement que les valeurs de β_0 et β_1 correspondent bien aux valeurs obtenues.
- Quel modèle vous semble le plus pertinent ? Expliquer en vous appuyant sur des figures et sur des chiffres.
- Les variables explicatives des modèles non retenus peuvent-elles être considérées comme inutiles ? Si oui, peut-on les modifier pour les rendre plus intéressantes ? Si non, mettre en place un modèle de régression démontrant leur pertinence.

3.2 Régression Multiple

- Mettre en place une régression linéaire multiple avec toutes les variables quantitatives.
- Quelles sont les variables statistiquement significatives ?
- Trouver le modèle qui semble avoir le meilleur équilibre nombre de variables vs fiabilité du modèle en se basant sur les R2 et les R2 ajustés.
- Mettre en place une régression linéaire multiple sur les données cumulées. Rechercher le meilleur modèle et expliquer la démarche. Appuyez vous sur les R2 et R2 ajustés.

Bonus : Peut-on considérer que le modèle entraîné sur ces données est généralisable ? Si oui, pourquoi ? Si non, comment y remédier ?