



UNIVERSITÉ CLERMONT-AUVERGNE IUT D'AURILLAC

Projet pour la SAE 3-03 autour de l'étude de l'évolution du poids

Auteurs :

Valentin EVENO

Viny Presty NAKAVOUA

NSALOUMOUNA

Angélique RIVIÈRE

Professeurs référents :

Paul Marie GROLLEMUND

Chapitre 1

Approches utilisées lors du pré-traitement

1.1 Segment introductif

Dans ce projet, nous avons étudié des jeux de données provenant d'un protocole expérimentale sur un régime combiné à une thérapie comportementale dans l'objectif de perdre du poids se déroulant en deux étapes.

- Dans une première étape, avant le début de la thérapie et du régime, le sujet de l'expérimentation se pèse plusieurs fois par jour sur plusieurs jours, dans cette première phase, on dispose d'un total de 114 entrées.
- Puis une seconde phase, cette fois-ci avec le régime et la thérapie, qui dure deux semaines pendant laquelle le sujet n'effectue qu'une seule pesée par jour. Sur cette deuxième phase, on dispose du poids de l'individu ainsi que le pourcentage de matière grasse, pour cette partie, on a obtenu 14 entrées.

1.2 Pré-traitement du jeu de donnée

Après avoir importé et visualisé une première fois les données, nous avons transcrit les jeux de donnée des deux phases en deux data frame distincts. Nous avons pu remarquer plusieurs problèmes sur les variables qu'il faut résoudre et ajuster lors du pré-traitement.

1.2.1 Variable jour

D'abord pour ce qui est de la variable jour, ces données étaient en format de date français "jour/mois/année". Nous les avons transcrites en format date de rstudio qui correspond à "année-mois-jour" pour être plus facilement utilisables par le logiciel.

1.2.2 Variable poids

Ensuite, pour la variable poids, on a remarqué qu'il y avait des valeurs manquantes, on a donc choisit de supprimer les jours où il n'y avait pas d'observations. On a considéré qu'estimer ces poids manquants avec des méthodes d'imputation comme celles par la moyenne ou la médiane ne serait pas judicieux.

1.2.3 Variable heure

Enfin pour la variable heure, on avait des chaines de caractère, pour que ces données puissent être traitables, nous les avons mises à un format "H :M" temps de Rstudio. Puis pour des questions d'analyses et d'interprétation, on a converties la variable temps au format décimal.

On a aussi remarqué l'heure de la pesée au 20 mai 2020 (index 48), la pesée a été faite à minuit ce qui rendait les graphiques moins cohérents, on l'a alors remplacé par 24.

concernant les valeurs manquantes, on a utilisé le package zoo de R qui nous a permis de les imputer par la méthode d'interpolation. l'objectif était vraiment d'éviter de perdre davantage de données et là en l'occurrence, l'interpolation était une méthode appropriée.

Chapitre 2

Transformations du jeu de donnée

2.1 Lissage par régression locale : le poids en fonction de l'heure

Après avoir pré-traités les données, nous pouvons désormais les étudier, ainsi dans un premier temps nous allons étudier comment le poids varie en fonction de l'heure. Nous utilisons la fonction loess qui permet de modéliser les données à l'aide d'une régression locale avec ici un lissage de paramètre égale à 0,3 .

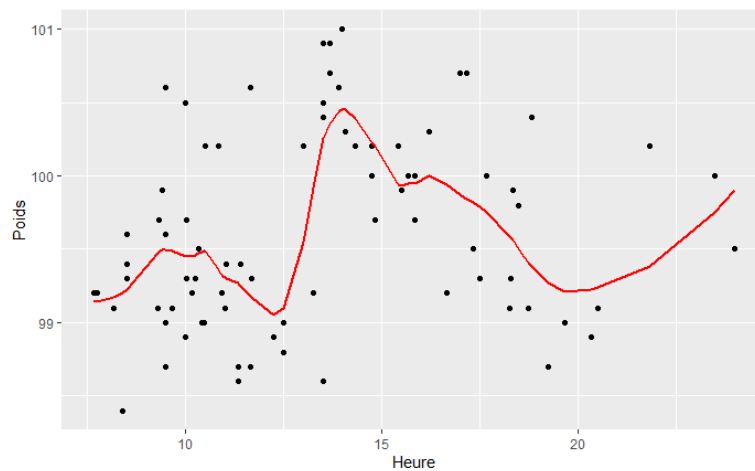


FIGURE 2.1 – représentation d'une régression locale avec un lissage à 0.3

2.2 Calcul de l'évolution du poids sur le temps long en phase 2

Maintenant nous devons déterminer l'évolution du poids pendant la seconde phase indépendamment des variations dans la journée.

Pour cela on va utiliser le lissage obtenu précédemment. En effet, l'idée ici est de rendre les données comparables (donc sur une même échelle). Pour cela, on va déterminer le poids à midi de ce sujet pendant tous les jours de la seconde phase.

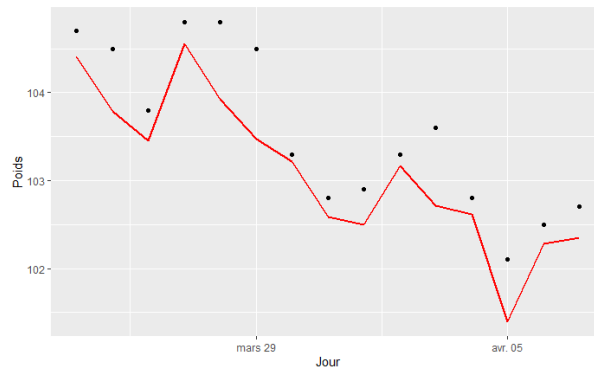


FIGURE 2.2 – évolution du poids en phase 2 indépendamment de l’heure

2.3 Evolution de la quantité de matière grasse

Pour calculer l’évolution de la quantité de matière grasse lors de cette seconde phase de 14 jours, on a simplement convertir le pourcentages de matières grasses en kilogrammes.

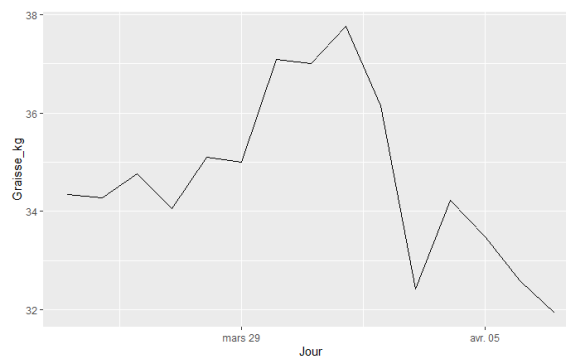


FIGURE 2.3 – représentation de l’évolution de la quantité de matière grasse en phase 2

On observe sur ce graphique que la quantité de matière grasse subit une augmentation entre la fin du mois de mars jusqu’au 2 avril puis on remarque une baisse assez importante d’environ 5.8246 kilogrammes de matière grasse sur les données observables. La quantité de matière grasse perdue sur le temps long du régime, donc en prenant la première et dernière valeur obtenue, est de 2.4019 kilogrammes en passant de 34.3416 kilos de matière grasse au 24 mars 2021 à 31.9397 kilos le 7 avril, date de fin du régime.

2.4 Prédiction sur la matière grasse perdue

Maintenant nous devons effectuer une prédiction sur le jeu de donnée de la phase 2 pour déterminer combien de matière grasse l’individu aurait perdu si il avait continuer son régime 3 jours de plus.

On observe pas de tendance sur ces données, donc on va modéliser notre prédiction en utilisant deux régressions linéaires. On sépare d’abord notre jeu de données en deux échantillons :

- Un premier échantillon du 24 mars au 1 avril.
- Un deuxième échantillon du 2 avril au 7 avril.

Ensuite sur chacun de ces deux échantillons, nous avons appliqué des régressions linéaires, pour la prédiction de perte de la quantité de matière grasse, on va utiliser la régression linéaire effectuée sur l’échantillon 2 puisque c’est la plus récent et donc la plus logique pour déterminer des prévisions à court terme.

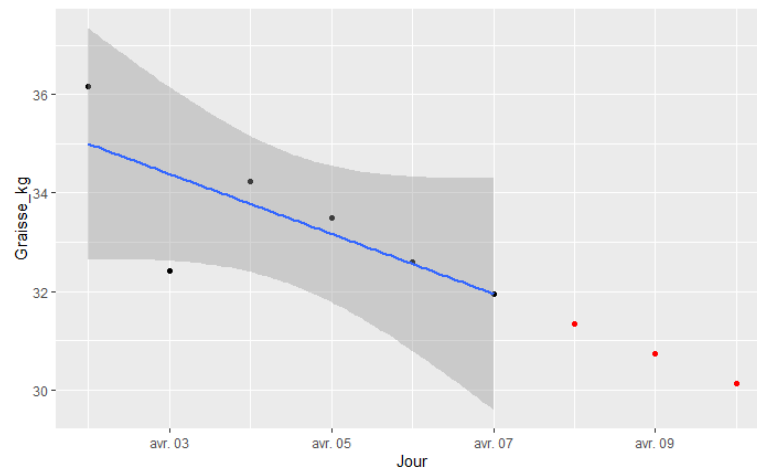


FIGURE 2.4 – prédiction de la quantité de matière grasse perdue avec 3 jours de régime en plus

Si la personne avait poursuivi le régime durant 3 jours supplémentaires, selon notre prédiction, elle aurait perdu environ 4.218229 kilogrammes contre seulement 2.4019 kilos qu'elle a perdu en mettant un terme à son régime le 7 avril.