

TD Régression Linéaire

1. Lecture des données

Commencez par vous placer dans le bon répertoire avec R

Lisez les données à partir du fichier Vehicules_TD_2023.csv

2. Examen des données

2.1

Prenez le temps d' "observer" en direct les données :

Y a-t-il des valeurs manquantes ?

Y a-t-il des valeurs aberrantes ?

2.2

Faites l'exercice qui a été fait en cours d'examiner les données avec les mêmes fonctions que celles du cours (si vous avez chargé le package corrplot utilisez corrplot).

2.3

Regardez la matrice de corrélation, essayez de la visualiser

3. Un peu d'esprit critique : quelle question peut-on se poser ?

Les trois catégories de véhicules sont-elles "visualisables"

Examinez les centres de gravité des différentes catégories

4. La régression linéaire (enfin) : Effectuez maintenant la régression linéaire de la variable PRIX par rapport aux autres variables numériques

4.1

Vous allez fabriquer un « modèle » qui sera un objet de R.

Exploration de chacun des attributs, pour exploitation ultérieure et vérification de quelques formules du cours

4.2 Commençons par les coefficients

Seriez-vous capable de les retrouver à partir de la formule matricielle du cours ?

4.3

Etes vous capables de calculer la prévision d'une valeur (par exemple d'une valeur associée aux données fournies) ?

... et de comparer avec les valeurs prévues calculées par R ?

Dessinez le graphe prix / valeurs prévues

4.4

Maintenant on va regarder les résidus

Vérifiez en direct qu'il s'agit bien de résidus (un exemple suffit)

Les résidus ont-ils un comportement de loi normale, vous avez le droit d'être subjectif

4.5

Maintenant, vous allez choisir des "individus" différents avec des paramètres plausibles

Voire les prévisions que donnent le modèle !

A vous de jouer !

5. Imprécision des coefficients

Avant d'interpréter les autres éléments du modèle, vous allez, en groupe

constater que les coefficients sont imprécis

Pour cela, vous allez chacun de vous reprendre le tableau de données et

supprimer une ligne différente, puis refaire le modèle linéaire sur le nouveau tableau

et comparer entre vous les différentes valeurs des coefficients.

6. Caractère significatif des coefficients

Maintenant, on va examiner le caractère significatif des coefficients. Examinez chaque composante du résultat `summary(Modele)` et commentez.

Si vous deviez choisir un modèle avec moins de variables, comment feriez-vous ?

Faites au moins un essai en classe. Vous pouvez en faire plus soyez simplement un peu critique !

7. Ensemble d'apprentissage et ensemble de test

Faites une régression en séparant un ensemble d'apprentissage et en ensemble de test.

Comparez le R^2 calculé par R avec l'erreur quadratique que vous avez obtenue

Commentez !

8. Si vous avez le temps

Même exercice avec d'autres données je vous propose `census_data`, dans le fichier `Census_data_fr.csv`, et construisez un modèle permettant de prédire la variable `Life.Exp`.

Essayer de réduire le nombre de variables permettant d'expliquer le modèle.