

TP : Prédiction du prix d'une voiture d'occasion

Jeu de données

Le jeu de données `Toyota.txt` contient 400 observations et les variables suivantes :

Variable	Description
Price	Prix de la voiture (en euros)
Age	Âge du véhicule (en mois)
KM	Kilométrage total (en kilomètres)
HP	Puissance (en chevaux)
MetColor	Métallisé (1 = oui, 0 = non)
Automatic	Transmission auto (1 = oui, 0 = non)
Doors	Nombre de portes
Weight	Poids du véhicule (en kg)

Objectif

Expliquer le prix d'une Toyota Corolla d'occasion en fonction de ses caractéristiques à l'aide d'un modèle de régression linéaire multiple.

Consignes

1. Importation et visualisation

Charger `Toyota.txt` dans R et présenter un aperçu du jeu de données.

2. Analyse univariée

Calculer les statistiques descriptives (moyenne, écart-type, etc.) pour chaque variable, et commenter.

3. Corrélations

Calculer et représenter la matrice de corrélations entre les variables numériques, et commenter.

4. Modèle complet

Ajuster le modèle complet.

- Vérifier les hypothèses du modèle (normalité des résidus, homoscedasticité, indépendance).

- Estimer et commenter les coefficients.

- Tester la significativité individuelle des paramètres (seuil 5%).

5. Qualité de l'ajustement

- Calculer R^2 .

- Tester la signification globale du modèle.

— Discuter de la pertinence d'un modèle linéaire.

6. Discussion

En vous appuyant sur les résultats obtenus, résumez brièvement ce que vous avez appris sur les facteurs qui influencent le prix d'une voiture.

D'après vous, ce modèle est-il suffisamment bon pour faire des prédictions approximatives ? Pourquoi ?