

### Tutoriels de référence :

Ceux cités dans le document principal.

Voir également sous Google avec les mots clés « statistique descriptive r ».

### Questions :

On souhaite traiter le fichier « Autos.xls »

1. Installer, si ce n'est déjà fait, le package « xlsx ». Le charger ensuite ([library](#))
2. Charger le fichier « Autos.xlsx » ([read.xlsx](#))
3. Combien y a-t-il de variables dans le fichier ? Combien y a-t-il d'observations ? ([nrow](#), [ncol](#))
4. Isolez dans un nouveau data frame les véhicules correspondant aux caractéristiques « fuel type = gas » et « aspiration = std ». Combien d'observations correspondent à ces caractéristiques ? (**161 observations**). *A partir de maintenant, nous travaillerons uniquement sur ce sous-ensemble d'observations.*
5. Créez le graphique « nuage de points » avec en abscisse « engine size » et en ordonnée « price » ([plot](#)). Que constatez-vous ?
6. Calculez le coefficient de corrélation entre ces deux variables ([cor](#)).
7. Transformez les 2 variables en rangs ([rank](#)). Refaites le graphique puis calculez le coefficient de corrélation. Commentaires ?
8. Revenons sur les variables non transformées. Réalisez maintenant la régression linéaire simple de « price - Y » en fonction de « width - X » ( $Y = aX + b$ ) ([lm](#)). Quelles valeurs des coefficients a et b obtenez-vous ? Les afficher explicitement en accédant à la propriété « coefficients » de l'objet.
9. Faire afficher le coefficient de détermination  $R^2$  de la régression en appelant la fonction [summary\(\)](#). La régression est-elle satisfaisante ?
10. Créer le graphique « nuage de points » entre « width » (abscisse) et « price » (ordonnée). Ajouter la droite de régression dans le graphique ([abline](#))
11. Calculer les prédictions du modèle sur les 161 véhicules constituant la base filtrée ([predict](#)). Afficher le graphique avec en abscisse les valeurs observées de « price » et en ordonnée les valeurs prédites par le modèle. Tracer une droite sur la diagonale principale pour situer la qualité de la prédiction. *Attention, mettez les mêmes limites en abscisse et ordonnée pour que le graphique soit carré (cf. les options de [plot](#))*
12. Sur la même base, coder « num.of.doors » en variable binaire avec 1 quand num.of.doors == « four », et 0 quand num.of.doors == « two » ([ifelse](#))
13. On souhaite réaliser une régression linéaire multiple expliquant le « price » en fonction de « engine.size », « horsepower », « width » et la nouvelle variable nombre de portes recodée en 0/1. Reproduire les étapes (8, 9, 11). La régression est-elle de meilleure qualité ?
14. Rajouter les deux prédictions ([cbind](#)) à l'ensemble de données (le data frame avec 161 observations) et exporter le fichier au format « xlsx » ([write.xlsx](#)).