

L'idée de ce projet est d'utiliser la méthode de l'Analyse en Composante Principale (ACP) sur les données auto. Pour illustrer la méthode avec des résultats, choisissez les variables quantitatives notamment les variables : CYCL (Cylindre), PUISS (Puissance du moteur), LONG (Longueur), LARG (Largeur), POIDS (Poids à vide) et V.MAX (Vitesse Max). Vous avez un tableau de données de n individus (véhicules) et p variables (ici $p=6$)

1. Préparer les données en sélectionnant les variables citées ci-dessus.
2. Écrire un script, qui vérifie si une donnée est manquante, alors supprimer la ligne correspondante.
3. Écrire une fonction permettant de normaliser les données (centrage et réduction).
4. On va d'abord étudier la corrélation entre 2 variables 2 à 2.
 - a. Faire un script qui permet de dessiner les nuages de points en considérant les variables 2 à 2 (toutes les combinaisons possibles).
 - b. Écrire un script qui permet d'appliquer une régression simple sur ces variables prises 2 à 2. Afficher les résultats des régressions en dessinant la droite et en donnant l'équation ainsi que le coefficient de détermination.
 - c. Calculer la matrice de corrélation qui fournit un tableau pxp qui calcule la corrélation entre la variable i et la variable j .
5. Appliquer la méthode ACP (sous R, ou sous Matlab), en écrivant un script qui permet d'exploiter les résultats, en affichant le pourcentage d'information par axe principale.
6. Projeter les points sur les plans factoriels en utilisant les axes principaux (1,2), (1,3) et enfin (2,3).
7. Projeter dans un autre graphique les vecteurs variables sur ces mêmes plans factoriels.
8. Interpréter les résultats obtenus pour chaque plan factoriel en répondant aux questions suivantes :
 - a. Quelles sont les véhicules qui se ressemblent selon les caractéristiques choisies (les variables choisies), on parle de proximité des individus (véhicules)
 - b. Sur quelles variables sont fondées les ressemblances / dissemblances
 - c. Quelles sont les relations entre les variables
9. Projeter les points et les vecteurs variables sur l'espace à 3 dimensions intégrant les 3 axes principaux.
10. Si vous observez un nombre de groupes homogènes, appliquer la méthode de k-means pour déterminer les individus du même groupe.
11. **Question exotique** : rajouter d'autres variables quantitatives et refaites une ACP avec une interprétation des résultats sur le plan factoriel des deux premiers axes principaux. Appliquer un clustering.