

- Analyse de données -
TP 4 Analyse en composantes principales

1 Véhicules vendus en France

On dispose de données techniques sur 52 véhicules vendus en France (année modèle 1994).

Les variables sont : Puiss (puissance, en chevaux fiscaux), Cyl (cylindrée, en cm³), Long (longueur de la voiture), Larg (largeur de la voiture), Surf (surface de la voiture) Poids (poids total en kg), Vites (vitesse maximum en km/h), DepArr (temps, en secondes, pour parcourir 1000 m, départ arrêté), Conso (consommation moyenne aux 100 km, en litres d'essence ou gazole).

Les marques des véhicules sont abrégées Al (Alfa-Romeo), BM (BMW), Ci (Citroen), Fi (Fiat), Pe (Peugeot) et Re (Renault).

Les données sont dans le fichier `vehicules.txt`.

2 Questions

1. Quelles sont les couples de variables particulièrement peu corrélés entre elles ?
2. Quelle variable se distingue des autres ? Expliquez pourquoi.
3. On procède à un analyse en composantes principales. Quel est le nombre d'axes principaux que l'on doit garder pour l'analyse. Quelle part d'inertie totale sera alors représentée ?
4. Quelles sont les variables qui déterminent les deux premières composantes principales (précisez les critères utilisés) ?
5. Quels sont les individus qui déterminent les deux premiers axes principaux ? (précisez les critères utilisés)
6. Comment peut-on interpréter la première composante principale ? Et la seconde ?
7. Expliquer la signification et la définition de la qualité de la représentation des individus. Expliquer comment calculer la qualité de la représentation de Alpha 155 2.0 sur le premier axe principal.
8. Quels sont les 5 individus les moins bien représentés par le sous espace qu'on a décidé de conserver ?
9. Compléter l'analyse.