

TP1

Traitement de données – ACP

Vous devez réaliser une analyse en composantes principales sur les 2 jeux de données

Exercice 1 : ACP sur les données Températures

Les données donnent les températures moyennes relevées dans 35 grandes villes Européennes ; elles se trouvent dans le fichier excel : Temperatures.xlsx.

1. Observer très rapidement ces données et regarder si des états se différencient des autres etc. (Rq: ne pas passer trop de temps sur cette question).
2. Donner les moyennes et les variances des différentes variables. Donnez également la variance totale du nuage de points.
3. Donner les distributions des différentes variables.
4. Calculer les corrélations entre les variables et utiliser une représentation graphique pour visualiser les coefficients de corrélations. Commenter ces résultats.
5. Réaliser l'ACP
6. Tracer l'évolution de la variance expliquée en fonction du nombre de facteur. Calculer la somme des variances.
7. Quel est le pourcentage de variance rapporté par les 2 premiers axes ?
8. Représenter les Villes suivant les 2 premiers axes principaux.
9. Que représentent les 2 composantes principales ? Pour répondre à cette question vous analyserez le cercle des corrélations entre les variables de départ et les composantes principales.
10. Formulez des conclusions sur les différents Villes.

Exercice 2 : ACP sur les données Criminalités

Les données se trouvent dans le fichier excel : Criminalite.xlsx. Les données reprennent différents taux de criminalité (exprimés en pour 100000 habitants) des 50 états des Etats-Unis. 7 taux ont été retenus: meurtres, viols, vols à main armée, agressions, vols avec effraction, vol à la tire, infractions au code de la route.

1. Observer très rapidement ces données et regarder si des états se différencient des autres etc. (Rq: ne pas passer trop de temps sur cette question)
2. Donner les moyennes et les variances des différentes variables. Donnez également la variance totale du nuage de points.
3. Regarder les distributions des différentes variables.
4. Calculer les corrélations entre les variables et utiliser une représentation graphique pour visualiser les coefficients de corrélations. Commenter ces résultats.

Analyse en Composantes Principales (ACP) : vous réaliserez deux ACP différentes:

- une ACP sur les données centrées
- une ACP sur les données centrées et réduites

Attention : le pré-traitement est à réaliser sur la matrice de données avant de réaliser l'ACP (donc avant la diagonalisation de la matrice de variances-covariances et avant la projection des données dans la base de axes principaux)

5. Pré-traiter les données (il sera intéressant pour vous de regarder alors les variances des variables mais également la matrice de variances-covariances et celle de corrélation)
6. Réaliser l'ACP
7. Tracer l'évolution des variances expliquées en fonction du nombre de facteurs. Calculer la somme des variances.
8. Quel est le pourcentage de variance rapporté par les 2 premiers axes ?
9. Représenter les Etats suivant les 2 premiers axes principaux.
10. Que représentent les 2 composantes principales ? Pour répondre à cette question vous analyserez le cercle des corrélations entre les variables de départ et les composantes principales.
11. Formulez des conclusions sur les différents Etats.
12. Quel est l'intérêt de réaliser l'ACP sur les données centrées réduites ?