

Département Informatique
Analyse des données
Travaux pratiques
Séance 2

Étude de l'ACP sur les données IRIS

La séance de TP se déroulera en deux parties:

On commencera par une découverte de la syntaxe en lien avec l'acp et ensuite un travail personnel vous sera demandé vous amenant à comprendre les rouages de l'analyse en composantes factorielles. Pour répondre à certaines questions posées, il sera nécessaire de vous appuyer sur les notes du cours magistral.

Partie 1

Il vous est demandé de tester les deux scripts donnés. Tout d'abord le script_stat et le script acp_iris. Exécuter pas à pas le premier et le second en observant bien les effets dans la fenêtre graphique et la console de commande.

Leur maîtrise vous permettra de répondre aux questions de la partie suivante :

Partie 2

1. Donner et exécuter la commande qui permet de retrouver l'ensemble des valeurs fournies par l'ACP.
2. Examiner la matrice des corrélations entre les variables. Faites un commentaire.
3. Examiner le nuage de points des individus:
Donner l'expression et la valeur de l'inertie du premier plan factoriel.
Donner l'expression de l'inertie d'un point. Quels sont les individus dont l'inertie est la plus grande? Quels sont ceux dont l'inertie est la plus faible?
4. Donner le profil des valeurs propres et proposer le meilleur plan d'observation à retenir pour l'analyse. Justifier le nombre de composantes principales trouvées?
5. Donner l'expression et la valeur de l'inertie de ce nuage de points.
6. Donner l'expression et la valeur de l'indice ponctuel de qualité pour le point le mieux placé pour ce critère
7. Donner l'expression qui mesure la contribution d'un point sur un axe factoriel. Quels sont les individus qui contribuent le plus fortement à la formation du premier axe principal?
8. Donner deux individus qui sont bien représentés par leur première composante principale et deux les plus mal représentés par cette composante.

Département Informatique
Analyse des données
Travaux pratiques
Séance 2

Étude de l'ACP sur les données IRIS

9. Analyser de la même façon le deuxième puis le troisième axe principal (reprendre les questions 6,7 et 8)
10. Quelles sont les variables les plus fortement corrélées avec la première composante principale. Interpréter cet axe à l'aide de ces variables.
11. Faisant suite à la question précédente, donner une interprétation concernant le deuxième et troisième axe factoriel.
12. Proposez une/des interprétation(s) possible(s) suite à l'étude des graphiques obtenus dans R.