

TP Analyse de données

Séance 4 – Analyse en composantes principales

1. Un échantillon de données quantitatives

Nous allons étudier les données de l'échantillon `LifeCycleSavings`.

1. Consulter l'information fournie sur ces données à l'aide de la commande `help(LifeCycleSavings)`
2. Créer une variable `tab` contenant ces données. Une description sommaire en est fournie par la commande `summary(tab)`. La commande `plot(tab)` fournit une représentation graphique (d'intérêt limité cependant).
3. Créer une variable `moyenne` et une variable `variance` contenant la moyenne et la matrice de covariance des données. Quel est l'effet de la commande `X <- scale(tab, scale=sqrt(diag(variance)), center=moyenne) ?`
4. Déterminer les coefficients de corrélation à l'aide de la commande `cor(tab)` et interpréter les résultats.

2. La commande `prcomp`

1. Exécuter la commande `acp1<-prcomp(X)`
2. Vérifier qu'on obtient le même résultat directement à partir de la commande `acp<-prcomp(tab, scale=TRUE)`
3. Les nouvelles coordonnées sont stockées dans une variable `Y` à l'aide de la commande `Y <- acp$x`
Vérifier que l'on obtient le même résultat en exécutant "manuellement" les commandes
`X <- as.matrix(X)`
`U <- acp$rotation`
`Y1 <- X %*% U`
4. Vérifier que la matrice de covariance de `Y` est diagonale aux erreurs d'arrondi près.
5. Exécuter `plot(acp)`. En déduire que les deux premières composantes principales retiennent une part importante de l'inertie.
6. Afficher la projection des deux premières composantes avec `biplot(acp)`
Interpréter les résultats.

3. Applications

Effectuer une ACP des échantillons `USArrests` et `state.x77`