

TD 5 - Analyse en composantes principales

But du TD : comprendre et visualiser les divers outils procurés par l'ACP.

1 Etude des données painters

Référence : *Principal Component Analysis*, I.T. Jolliffe, Springer Series in Statistics, 1986.

Récupérer les données `painters` de la librairie `MASS` (rappel : `library(MASS)` pour accéder à la librairie) et les renommer `peintres`. Garder dans un tableau nommé `peint` les variables quantitatives.

- **ACP sur données centrées réduites** : Analyser la commande `prcomp`, et en particulier, l'option `scale`. Réaliser une ACP sur les données centrées réduites. Nommer `acpeint` le résultat de l'ACP.
- **Etude des valeurs propres** : Etudier la commande `barplot`. Visualiser à l'aide de `barplot` les valeurs propres associées à chaque axe. Représenter le pourcentage de variance expliquée par chaque axe, puis calculer et visualiser le pourcentage de variance expliquée en fonction du nombre de composantes observées. Commenter. Retrouver ces résultats en utilisant la commande `summary` sur `prcomp` (voir l'aide de `prcomp`).

- **Etude des vecteurs propres** : Comprendre et utiliser les commandes suivantes :

```
layout(matrix(1:4,2,2))
for (i in 1:4)
{barplot(acpeint[2]$rotation[,i], main=paste("Composante",i))
}}
```

Analyser et expliquer en détail l'information retenue par les deux premières composantes.

- A l'aide de la commande `biplot`, visualiser la projection des individus sur les deux premiers axes principaux. Numérotter les peintres de 1 à 54 et représenter les individus par ce numéro. Pour cela, se servir du paramètre `xlabs` de la commande `biplot`. Calculer la corrélation des variables `Composition` et `Expression`. Commenter et relier ce résultat à la position respective des deux axes correspondants sur le graphique précédent. Faire de même pour les variables `Drawing` et `Colour`.
- **Analyse globale des individus projetés** : Visualiser les individus projetés sur les deux premiers axes principaux en représentant chaque individu par son courant de peinture (`School`). Quelles conclusions peut-on tirer de ce graphique sur l'appartenance à des écoles de peinture ? Représenter sur un graphique, dans la même fenêtre graphique, les projections des individus sur le premier et le troisième axes principaux (consulter l'aide de `biplot.princomp`).
- **Analyse d'individus particuliers** : Calculer la qualité de la représentation des individus sur les deux premiers axes. Pour disposer des données exprimées sur la nouvelle base des axes principaux, utiliser l'option `retx=TRUE` de la commande `prcomp`. Quels sont les individus les mieux représentés, les moins bien représentés ? Calculer la contribution des individus sur les deux premiers axes principaux. Que dire du peintre Fr. Penni ? Positionner et commenter 3 individus en particulier.

2 Etude des données crabs

Charger les données `crabs` dans la librairie `MASS`. Les renommer `crabes`. Dans un tableau de données appelé `crab`, stocker les données quantitatives. Reprendre dans le paragraphe précédent les questions

- ACP sur données centrées réduites.
- Etude des valeurs propres.
- Etude des vecteurs propres. Calculer la matrice des corrélations et comparer l'information qu'elle apporte avec la position des axes représentant les variables, sur le graphique représentant les individus projetés sur les deux premiers axes principaux.
- Analyse globale des individus projetés. Ici, remplacer la variable `School`, "courant de peinture", par la variable `sp`, puis par la variable `sex`. Représenter les individus projetés sur les deux premiers axes principaux, par leur espèce, puis par leur sexe. Peut-on remarquer quelque chose ? Répéter le processus avec les projections sur les premier et troisième axes principaux. Que remarque-t'on ?
- Analyse d'individus particuliers.