

TD 6 - Analyse en composantes principales II

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

1 Analyse de céréales de petit-déjeuner

Récupérer les données UScereal de la librairie MASS. L'étude proposée est similaire à celle du TD 5. Garder dans un tableau noté C les données quantitatives.

- **ACP.** Effectuer une ACP sur les données centrées réduites avec la commande `prcomp`, appeler le résultat `acpC`.
- **Valeurs propres.** Récupérer les valeurs propres dans un tableau que l'on nommera `vpC`. Représenter à l'aide de `barplot` le pourcentage de variance expliquée par chaque axe principal, puis le pourcentage de variance expliquée en fonction du nombre de composantes gardées. Combien de composantes conviendrait-il de garder pour expliquer 80% de la variance ? 90% ?
- **Vecteurs propres.** Visualiser à l'aide de `barplot` les coordonnées des axes principaux dans l'ancienne base. Expliquer l'information retenue par les deux premières composantes.
- Charger la librairie `ade4`. Etudier et exécuter les commandes :

```
CD=dudi.pca(C,nf=3,scan=FALSE);  
s.corcircle(CD$co ); title("Cercle des Corrélations")
```

Commenter le graphique obtenu. Calculer la matrice de corrélation de C et comparer avec ce qu'on observe sur le graphique.
- **Analyse des individus projetés.** Tracer à l'aide de `biplot` un graphique représentant les projections de chaque individu sur les deux premières composantes principales. Refaire au besoin le graphique en étiquetant les céréales à l'aide d'un numéro. Quelles céréales semblent être atypiques ?
- Tracer les projections des individus sur le plan des première et troisième composantes. Commenter
- **Analyse d'individus particuliers.** Calculer la qualité de la représentation des individus sur le premier, puis le second axe. Calculer la contribution de chaque individu à la construction du premier, puis du second axe. Positionner et étudier les céréales suivantes : 100% Bran, All-Bran, All-Bran with Extra Fiber, Corn Flakes, ainsi que deux autres céréales de votre choix. Penser à étudier les coordonnées dans la matrice `scale(C)`.

2 Etude de notes d'élèves

Récupérer les données `seconde` de la librairie `ade4`, et les nommer N.

- Effectuer une ACP sur les données réduites. Reproduire toute l'étude menée au paragraphe 1.
- **Analyse d'individus particuliers.** Etudier et comparer la position des individus 15 et 8 sur le biplot correspondant aux deux premières composantes. Mettre en relation avec les notes de ces étudiants. Etudier ces individus grâce aux autres outils développés. Penser à regarder la matrice `scale(N)`. Etudier également les individus 2, 9, 11 et 19. Que dire de l'individu 5 quant à la qualité de sa représentation sur les premiers axes et sa contribution à la construction des premiers axes ?

TD 6 - Analyse en composantes principales II

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

1 Analyse de céréales de petit-déjeuner

Récupérer les données `UScereal` de la librairie `MASS`. L'étude proposée est similaire à celle du TD 5. Garder dans un tableau noté `C` les données quantitatives.

- **ACP.** Effectuer une ACP sur les données centrées réduites avec la commande `prcomp`, appeler le résultat `acpC`.
- **Valeurs propres.** Récupérer les valeurs propres dans un tableau que l'on nommera `vpC`. Représenter à l'aide de `barplot` le pourcentage de variance expliquée par chaque axe principal, puis le pourcentage de variance expliquée en fonction du nombre de composantes gardées. Combien de composantes conviendrait-il de garder pour expliquer 80% de la variance ? 90% ?
- **Vecteurs propres.** Visualiser à l'aide de `barplot` les coordonnées des axes principaux dans l'ancienne base. Expliquer l'information retenue par les deux premières composantes.
- Charger la librairie `ade4`. Etudier et exécuter les commandes :

```
CD=dudi.pca(C,nf=3,scan=FALSE);  
s.corcircle(CD$co ); title("Cercle des Corrélations")
```

Commenter le graphique obtenu. Calculer la matrice de corrélation de `C` et comparer avec ce qu'on observe sur le graphique.
- **Analyse des individus projetés.** Tracer à l'aide de `biplot` un graphique représentant les projections de chaque individu sur les deux premières composantes principales. Refaire au besoin le graphique en étiquetant les céréales à l'aide d'un numéro. Quelles céréales semblent être atypiques ?
- Tracer les projections des individus sur le plan des première et troisième composantes. Commenter
- **Analyse d'individus particuliers.** Calculer la qualité de la représentation des individus sur le premier, puis le second axe. Calculer la contribution de chaque individu à la construction du premier, puis du second axe. Positionner et étudier les céréales suivantes : 100% *Bran*, *All-Bran*, *All-Bran with Extra Fiber*, *Corn Flakes*, ainsi que deux autres céréales de votre choix. Penser à étudier les coordonnées dans la matrice `scale(C)`.

2 Etude de notes d'élèves

Récupérer les données `seconde` de la librairie `ade4`, et les nommer `N`.

- Effectuer une ACP sur les données réduites. Reproduire toute l'étude menée au paragraphe 1.
- **Analyse d'individus particuliers.** Etudier et comparer la position des individus 15 et 8 sur le biplot correspondant aux deux premières composantes. Mettre en relation avec les notes de ces étudiants. Etudier ces individus grâce aux autres outils développés. Penser à regarder la matrice `scale(N)`. Etudier également les individus 2, 9, 11 et 19. Que dire de l'individu 5 quant à la qualité de sa représentation sur les premiers axes et sa contribution à la construction des premiers axes ?