

TD Analyse en composantes principales

Ce TD utilise le package FactoMineR

1.

Vous commencez par la lecture du fichier contenant les vins français

Remarquer la prise en compte du fait que la première colonne n'est pas une variable mais désigne le nom du vin.

2.

Commençons par "regarder" les données : vous devez remarquer deux variables qualitatives, et les autres variables quantitatives.

Que pouvez vous dire des variables quantitatives (des notes données par des experts en œnologie).

3.

Remarquez aussi la commande `fix()` permettant d'explorer les données, mais n'apportez pas de modifications.

4.

Pour ceux qui ont chargé le package `corrplot`, visualisez la matrice de covariance des données numériques

5.

Après avoir fait un premier examen des données, effectuez une analyse en composante principale des données

Appelez la fonction effectuant l'analyse en composantes principales

Attention : les deux première colonnes ne désignent pas des variables numériques.

Commencez par interpréter librement les deux graphiques

Essayez de caractériser les deux premiers axes principaux, ici les étudiants font une

Ensuite continuez par interpréter les résultats affichés :

l'attribut "Eigenvalues"

l'information "Individuals" : éléments, Dist, Dim.n cos2, Contr est la contribution de la variable dans la formation de la dimension :

```
plot(Vins.PCA$ind$coord[,1]^2,Vins.PCA$ind$contrib[,1],pch=19)
```

l'information "Variables" : éléments Dim.n et cos2, pour visualiser les angles, on peut, si nécessaire, refaire : `plot.PCA(Vins.PCA,choix="var")`

l'information "Supplementary categories"

éventuellement regardez la documentation

Examinez les variances expliquées par les axes principaux

6.

Trouvez le moyen de visualiser les variances expliquées

Remarque : vous devez retrouver le nombre de variables numériques avec la commande :
`sum(Vins.PCA$eig[,1])`

7.

Examinez les coordonnées des variables sur les axes principaux

Regardez la somme des carrés des coordonnées des composantes de VARS,
que remarquez-vous ?

8.

Examinez les coordonnées des individus sur les 5 premiers axes principaux

9.

Prenez maintenant le temps d'examiner les variables sur leur graphe : appréciation en séance

Examinez les individus et commenter librement ...

Dessinez, sur le premier plan principal, avec trois différentes couleurs, les vins Bourgueil, Chinon et Saumur

Pouvons-nous les séparer à l'œil ?

10.

Maintenant nous effectuons la même chose avec des vins italiens

Avec les mêmes types de commentaires

Quelle différence notable observe-t-on ?

il y a 3 classes

les individus sont répartis selon les classes

Faites une exploration classique

11.

Faites l'analyse en composantes principales, et commentez les axes, laissez les étudiants faire des commentaires

12. Si on a le temps, on regarde les valeurs propres, les variables et les individus

13.

Il semble que les trois classes sont bien séparées

14.

Nous allons maintenant effectuer une segmentation hiérarchique des vins italiens.

Faites-la avec 3 segments et la commande HCPC fournie dans le package.

15.

Bien examiner les trois figures obtenues : deux dendogramme, dont un dans l'espace et une représentation dans le premier plan principal, prenez le temps de regarder tout cela

Examinez aussi cet objet : avec R cela doit être un réflexe, explorer les objets obtenus

16.

Cherchez où se trouve la description des segments

17.

La matrice de confusion

C'est la comparaison entre les segments trouvés et le classement initial des experts

Pour cela, fabriquer une matrice de confusion,

Vous pouvez utiliser la commande table

18.

Dernière partie du TD, régression linéaire sur composantes principales.

Nous partons de l'exemple du TD régression : Vehicules,

Pour garder le résultat en mémoire, nous refaisons la régression linéaire du premier TD

Puis une analyse en composante principale

19.

Commencer par interpréter les deux graphiques de la composante principale/

On peut recommencer à examiner la capacité à séparer en 3 classes pour voir si les trois classes de véhicules sont reconnaissables et faire une matrice de confusion

Examinez le résultat avec un regard critique

Et construisez la matrice de confusion

20.

Maintenant, on va effectuer des régressions sur les composantes principales, à vous de jouer !!

Commencer par identifier les individus puis faire la régression

Faire plusieurs modèles, suivant le nombre de composantes principales

Essayer, librement pour les étudiants, de réduire le nombre de composantes principales permettant d'expliquer le prix du véhicule

21. Et si vraiment il reste du temps !!!

Données Boston Essayez de reprendre la régression du premier cours, mais cette fois sur les composantes principales

Ici nous laissons les étudiants essayer plusieurs modèles ...