

1 Données

2 Réalisation de l'ACP

Code ▼

TD Analyse en Composantes Principales

Marie Fourcot

09/02/2022

1 Données

Pour ce TD, nous allons utiliser les mêmes données que pour le TD de classification non supervisée.

Pour rappel, les données proviennent d'une étude de nutrition chez la souris. Elles ont été publiées par Martin et al. en 2007 et sont disponibles dans le package `mixOmics`, je vous les ai mises sur Moodle vu que nous n'allons pas nous servir de ce package par ailleurs.

Pour 40 souris, nous disposons :

- des données d'expression de 120 gènes recueillies sur membrane nylon avec marquage radioactif,
- des mesures de 21 acides gras hépatiques.

Par ailleurs, les 40 souris sont réparties selon deux facteurs :

- Génotype (2 modalités) : les souris sont soit de type sauvage (wt) soit génétiquement modifiées (PPAR) ; 20 souris dans chaque cas.
- Régime (5 modalités) : 4 souris de chaque génotype sont soumises à chaque régime alimentaire. Ces différents régimes sont :
 - ref : le régime de référence avec du maïs et de l'huile de colza
 - coc : un régime saturé en acides gras avec de l'huile de coco hydrogénée
 - sun : un régime riche en omega 6 avec de l'huile de tournesol
 - lin : une régime riche en omega 3 avec de l'huile de lin
 - fish : à base d'huile de poisson

1 Données

2 Réalisation de l'ACP

L'objectif de ce TD est d'explorer les possibilités de l'ACP pour obtenir des représentations lisibles des souris et des gènes. Puis d'interpréter cette ACP. Pour cela nous allons utiliser le package `FactoMineR` et le package `factoextra`.

2 Réalisation de l'ACP

Nous allons nous concentrer comme lors du précédent TD sur les données génétiques.

1. Quelles vont être nos variables actives? Quelles variables supplémentaires vous semblent pertinentes ?
2. Faut-il centrer et/ou réduire les données ?
3. Utilisez la fonction `PCA()` du package `FactoMineR` pour réaliser une ACP sur nos données génétiques.
PCA est l'acronyme anglais pour ACP.
Et interprétez les graphes.

2.1 Inertie

4. Utilisez la fonction `fviz_eig()` et commentez le résultat.

2.2 Étude des individus

5. Utilisez la fonction `fviz_pca_ind()` du package `factoextra` pour créer un graphe des individus coloré en fonction de la variable `genotype`, et interprétez le graphe obtenu.
N'hésitez pas à consulter l'aide de la fonction.
6. Faire de même pour la variable `regime`.
Gardez en tête la réponse à la question précédente.
7. Étudiez les individus contribuant beaucoup. Pour cela, tracez le graphe de contribution des individus avec la fonction `fviz_contrib()`.
Puis un graphe des individus avec les individus contribuant le plus, étudiez l'effet du

régime alimentaire.

1 Données

2 Réalisation de l'ACP

2.3 Étude des variables

8. Tracez le graphe des contributions des variables et le graphe des variables avec uniquement les gènes qui contribuent particulièrement.
9. Utilisez la fonction `dimdescrip()` et analysez le résultat.
10. Pouvons-nous utiliser le coefficient de corrélation pour toutes les variables? Pourquoi?