

Tutoriel sur R

François Husson

UP de mathématiques appliquées - l'institut Agro



Journées d'études en statistique – SFdS 2021

- ① Visualisation
- ② Imputation par ACP
- ③ Imputation par ACM
- ④ Imputation par AFDM

Visualiser le dispositif de données avec les fonctions du package VIM (facultatif)

- Récupérer ou importer le jeu de données ozone depuis <http://factominer.free.fr/missMDA/ozoneNA.csv>
- utiliser les fonction `aggr`, `matrixplot` et `marginplot` du package VIM pour visualiser le dispositif de données manquantes
- Construire une matrice de présence-absence et faire une ACM pour visualiser les associations de données manquantes entre variables

Jeu de données en écologie

Données Glopnet : 2494 espèces décrites par 6 variables quantitatives ([données](#), [lignes de code](#))

- LMA (leaf mass per area)
- LL (leaf lifespan)
- Amass (photosynthetic assimilation)
- Nmass (leaf nitrogen)
- Pmass (leaf phosphorus)
- Rmass (dark respiration rate)

et 1 variable qualitative : le biome (macro-écosystème)

Wright IJ, et al. (2004). The worldwide leaf economics spectrum. *Nature*, 428 :821.

www.nature.com/nature/journal/v428/n6985/extref/nature02403-s2.xls

- ① Visualisation
- ② Imputation par ACP
- ③ Imputation par ACM
- ④ Imputation par AFDM

Imputation simple par ACP et ACP sur tableau incomplet

- ① Récupérer ou importer le jeu de données ozone depuis
<http://factominer.free.fr/missMDA/ozoneNA.csv>
- ② Sélectionner les variables quanti uniquement
- ③ Estimer le nombre de composantes pour imputer
- ④ Imputer
- ⑤ Faire l'ACP à partir du jeu de données complété (les variables quali du jeu initiales seront utilisées après les avoir concaténées au jeu complété)

Imputer par ACP et faire une ACP avec Factoshiny

- ① Utiliser la fonction `Factoshiny` du package `Factoshiny` sur le jeu de donnée

Imputation multiple ACP avec mice et missMDA

- ① Utiliser la fonction MIPCA (du package missMDA) pour faire de l'imputation multiple
- ② Utiliser les tableaux imputés obtenu par MIPCA pour faire la régression du maximum d'ozone en fonction des autres variables sur le tableau incomplet

- ① Visualisation
- ② Imputation par ACP
- ③ Imputation par ACM**
- ④ Imputation par AFDM

ACM sur tableau incomplet

- ❶ Charger le jeu de données `vnf` : `data(vnf)`
- ❷ Faire l'ACM en ajoutant une modalité NA pour chaque variable ayant au moins un NA (ce qui est fait par défaut par la fonction MCA)
- ❸ Imputer le tableau de données par la fonction `imputeMCA` après avoir estimé le nombre de composantes pour imputer. Faire l'ACM en utilisant le tableau disjonctif imputé

- ① Visualisation
- ② Imputation par ACP
- ③ Imputation par ACM
- ④ Imputation par AFDM

Imputation par AFDM et AFDM sur tableau incomplet

- ① Reprendre le jeu de données ozone
- ② Imputer par AFDM
- ③ Faire l'AFDM sur le tableau en utilisant le tableau disjonctif complété
- ④ Imputer par forêts aléatoire avec la fonction `missForest` package `missForest`