Atelier mice

Vincent Audigier

16 Novembre 2021

Packages R

Installation des packages requis (inutile si déjà installés)

```
install.packages(c("micemd","FactoMineR","parallel","DescTools",
                   "VIM","mice", "mvtnorm", "funModeling","randomForest"))
```

Chargement des packages (indispensable)

```
library(mice)
library(micemd)
library(FactoMineR)
library(parallel)
library(DescTools)
library(VIM)
library(mvtnorm)
library(funModeling)
```

Importation des données

```
load("diabetes.Rdata")
```

Une description du jeu de données est disponible ici : https://github.com/niharikagulati/diabetesprediction

Analyse exploratoire

1. Explorer le lien entre les variables explicatives d'une part et entre la variable réponse et les variables explicatives d'autre part.
2. Explorer le dispositif des données manquantes. On pourra utiliser la fonction `aggr` du package VIM pour l'analyse univariée, `CramerV` du package DescTools pour l'analyse bivariée, `MCA` du package FactoMineR pour l'analyse multidimensionnelle.
3. Explorer la nature du mécanisme. On pourra utiliser les fonctions `marginmatrix` et `matrixplot` du package VIM pour les analyses bivariées.

Imputation multiple

1. Utiliser la fonction `mice` du package mice pour imputer le jeu de données. Quels sont les paramètres (nombre de tableaux imputés, nombre d'itérations, modèles conditionnels) utilisés par défaut ?
2. Vérifier la convergence de l'algorithme
3. A l'aide de la fonction `densityplot` du package mice, comparer les distributions des valeurs imputées et observées pour chacune des variables.
4. A l'aide de la fonction `marginmatrix`, comparer les distributions des valeurs imputées et observées pour le couple de variables `(Age, Skin.Thick)` sur le premier jeu de données imputé. On pourra utiliser la fonction `complete` pour obtenir ce tableau. Faire de même pour l'ensemble des couples de variables.
5. Vérifier l'ajustement du modèle d'imputation en utilisant la fonction `overimpute` du package `micemd`. On pourra paralléliser les calculs en spécifiant l'argument `nnodes`.
6. Proposer d'autres modèles d'imputation pour la variable `BMI`.
7. A l'aide de la fonction `with.mids`, ajuster un modèle de régression logistique sur chaque tableau imputé expliquant la variable `Outcome` à partir de l'ensemble des variables explicatives, puis agréger les résultats à l'aide de la fonction `pool`.
8. Comparer avec une analyse des cas-complets.
9. Effectuer une analyse de sensibilité par la méthode de l'ajustement delta sur la variable `Skin.Thick`. On pourra se reporter à la vignette dédiée (https://www.gerkovink.com/miceVignettes/Sensitivity_analysis/Sensitivity_analysis.html).