

Pakiet mice

Martyna Majchrzak, Agata Makarewicz, Jacek Wiśniewski

26 03 2020



Wprowadzenie

- ▶ Multivariate Imputation by Chained Equations
- ▶ wielowymiarowa imputacja za pomocą równań łańcuchowych

Wykorzystanie

Pakiet mice zawiera funkcje służące do:

- ▶ sprawdzenia wzorca brakujących danych (-> md.pattern, ...)
- ▶ imputacji brakujących danych (wielokrotnie) (-> mice)
- ▶ diagnozowania jakości imputowanych wartości (jakie funkcje?)
- ▶ analizy każdego uzupełnionego zbioru danych (?)
- ▶ zebrania wyników powtarzanych analiz (-> pool)
- ▶ przechowywania i eksportowania imputowanych danych w różnych formatach (?)
- ▶ generowania symulowanych niekompletnych danych (-> ampute)
- ▶ (Incorporate custom imputation methods)?

Imputacja danych

Funkcja mice

W zależności od typu brakujących danych, funkcja mice przyjmuje jako parametr inne metody imputacji danych.

Dane podzielone są na 4 kategorie:

- ▶ dane numeryczne (ciągłe)
- ▶ dane binarne (dane typu factor z dwoma poziomami)
- ▶ nieuporządkowane dane katégoryczne (dane typu factor z więcej niż 2 poziomami)
- ▶ uporządkowane dane katégoryczne (dane typu factor z więcej niż 2 poziomami uporządkowanymi)

Dowolne dane

Niektóre metody imputacji możemy zastosować do każdego typu danych.

- ▶ pmm (predictive mean matching/ predykcyjne dopasowanie średniej)
- ▶ midastouch (weighted predictive mean matching/ ?)
- ▶ sample (losowa próbka)
- ▶ cart (drzewo klasyfikacyjne i regresji (?))
- ▶ rf (random forest/lasy losowe)
- ▶ 2lonly.pmm (Level-2 class predictive mean matching) <- ?

Dane numeryczne

- ▶ pmm (domyślna)
- ▶ mean (średnia)
- ▶ norm (Bayesian linear regression/regresja liniowa)
 - ▶ norm.nob (linear regression ignoring model error)
 - ▶ norm.boot (linear regression using bootstrap)
 - ▶ norm.predict (linear regression, predicted values)
- ▶ quadratic (imputation of quadratic terms)
- ▶ ri (random indicator for nonignorable data) #Nie do końca czaje te wszystkie 2l. coś tam, nw czy je chcemy
- ▶ 2l.norm (Level-1 normal heteroscedastic)
- ▶ 2l.lmer (Level-1 normal homoscedastic, lmer)
- ▶ 2l.pan (Level-1 normal homoscedastic, pan)
- ▶ 2lonly.mean (Level-2 class mean)
- ▶ 2lonly.norm (Level-2 class normal)

Dane binarne

- ▶ logreg (logistic regression/regresja logistyczna) (domyślna)
- ▶ logreg.boot (logistic regression with bootstrap)
- ▶ 2l.bin (Level-1 logistic, glmer)

Nieuporządkowane dane kategoryczne

- ▶ polyreg (Polytomous logistic regression) (domyślna)
- ▶ lda (liniowa analiza dyskryminacyjna)

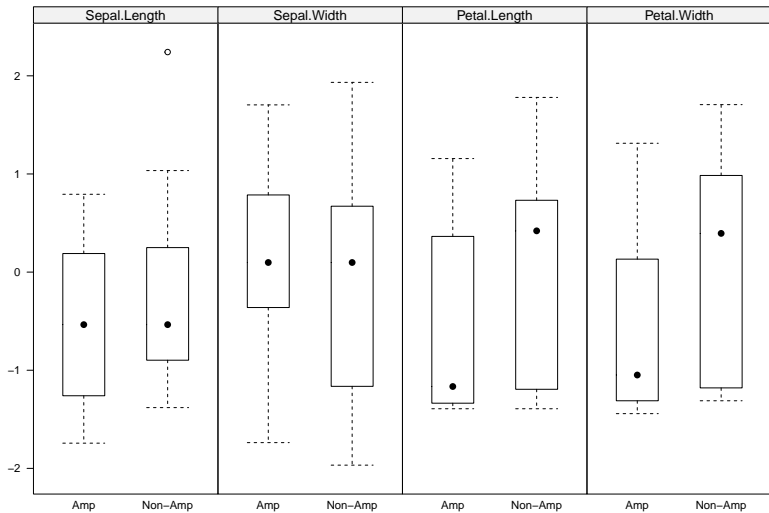
Uporządkowane dane kategoryczne

- ▶ polr (Proportional odds model) (domyślna)

Wykresy

- ▶ `xyplot`
- ▶ `md.pattern`
- ▶ `fluxplot`
- ▶ `densityplot`
- ▶ `stripplot`

Generowanie braków danych



Data distributions in pattern 1

Zebranie wyników analiz

Funkcja pool

Zbiory danych

- ▶ boys
- ▶ pattern
- ▶ brandsma