

Curs d'introducció a R

Exercicis

Eudald Correig i Fraga

11 de desembre de 2018

Pots fer tots els exercicis en aquest mateix document. Recorda mirar la documentació (?nom_de_la_funció) i en els documents linkats en la classe 0.

Part I: manipulació de dades

1. Importa el fitxer que es diu "Heart.csv".
 - Visualitza la base de dades (recorda, View()).
 - Elimina la primera columna, que no serveix per res.
 - Mira com són les variables i posa-les en el format correcte (numèric, factorial o caràcter).
 - Resulta que no volem el MaxHR, si no que volem el normal, que és el 70% del màxim (ja sé que no però com si sí.)
 - Fes una nova variable que digui que si colesterol és dolent o no (és a dir, que sigui 1 a partir de 200 i 0 si no.)
 - Fes un histograma de les edats
 - Fes un gràfic de caixes comparant el MaxHR entre homes i dones
 - Fes un gràfic de punts comparant l'edat i el RestBP
 - Fes un nou dataframe on hi hagi només dones (Sex = 0)
 - En aquest nou dataframe, calcula la mitja d'edat, de colesterol i fes una taula de freqüències dels nivells de "ChestPain"

Per més exercicis:

Aquí, pàgina 52

Aquí

Part II: Estadística bàsica

2. Seguint amb l'anterior dataframe:
 - Les distribucions de les variables contínues són normals?
 - Hi ha diferències significatives entre el MaxHR entre homes i dones
 - I entre AHD sí o no?
 - Hi ha correlació entre RestBP i MaxHR?
 - Hi ha diferència significativa entre el Colesterol segons el Thal? Alerta que hi ha NAs!
 - Hi ha diferència significativa del RestBP segons el ChestPain?
 - Hi ha diferències significatives entre l'AHD segons el sexe?

Part III: Models predictius

3. Encara amb el mateix dataframe:
 - Fes un model predictiu d'MaxHR segons totes les altres variables. (**Nota:** alerta amb les dues variables que potser has creat a la primera part dels exercicis, correlacionen molt amb les originals i et poden donar problemes, millor esborrar-les.) També hi ha un parell de variables amb NA's, s'ha de fer alguna cosa abans de continuar.
 - Fes un resum del model i intenta fer gràfics per explicar-lo.

- Fes un model predictiu d'AHD segons totes les altres variables.
- També, intenta dibuixar.
- *Ampliació*: Fes selecció de model utilitzant el lasso.