

7.6 Setè lliurament

1. Treballarem amb el data frame del fitxer “heart.csv” que pots trobar al Moodle, extret de Kaggle.com, de diferents mesures preses en pacients de cardiologia, relacionades amb el risc cardiovascular.
 - (a) Fes una descripció ràpida del fitxer i defineix com a factors les variables que sigui necessari. Recorda que hi ha accions que et seran més ràpides de programar utilitzant `apply`, `sapply` o `lapply`.
 - (b) Crea les següents variables EN EL `data.frame` (han de ser part del `data.frame`)
 - Crea un factor anomenat `edat.gr` amb la variable `Age` amb els grups “joves” (si edat entre 20 i 35), “adults” (si edat entre 36 i 65) i “grans” (si edat més gran que 66).
 - Crea un factor anomenat `colest.nivell` amb la variable `Cholesterol` amb les categories:
 Normal si el nivell de colesterol és menor que 200
 Alt si el nivell de colesterol és més gran o igual que 200 i menor que 250
 Molt alt si el nivell de colesterol és més gran o igual que 250
 - (c) Estudia el nivell de colesterol en els grups definits pel factor `edat.gr`. Repeteix l'estudi comparatiu del nivell de colesterol en els dos grups definits per la variable `Sex`. Quin factor influeix més en el nivell de colesterol?
 - (d) La variable `FastingBS` mesura el nivell de sucre en dejú en mg/dl (en el fitxer, els nivells ≤ 120 s'han convertit en 0). Compara aquesta variable en els grups definits per la variable `HeartDisease` (0: normal, 1: malalt de cor). Hi ha relació entre el nivell de sucre en dejú i ser malalt de cor? Diries que tenir nivell alt de sucre és un risc per aquesta malaltia?
 - (e) Compara la variable nivell de colesterol en persones malaltes del cor i persones que no ho són. Creus que el nivell de colesterol influeix en la malaltia del cor? Influeix més o menys que el nivell de sucre en sang?
 - (f) Comprova que les persones grans tenen la variable `FastingBS` més alta. Fes un sub `data.frame` de les persones grans i compara el nivell de sucre en sang entre malalts de cor i no malalts. És també influent?
 - (g) Feu un gràfic de punts de la variable `RestingBP` per grups d'edat (variable `edat.gr`, tal com s'explica al final de la pàgina 74, amb els punts de color blau i de mida un 50% del que surt per defecte).
 - (h) Descriu el prototip de malalt del cor i de no malalt.
2. Treballarem amb el fitxer `TRC.sav` que conté informació d'una sèries de pacients als quals han sigut implantats d'un marcapàs al seu cor. També hem mesurat un conjunt de variables al moment de ser implantats (`CF` = classe funcional (II, III o IV), `CREA` = valor de la creatinina en sang, `DTSVE` = diàmetre sistòlic del ventricle esquerra, `DTDVE` = diàmetre diastòlic del ventricle esquerra. També tenim la variables que ens indica si el pacient ha mort al cap d'un any posterior a l'operació (`MORT-12m`).
 - (a) Importa el fitxer `TRC.sav`
 - (b) Calcula el coeficient de variació mediana V_{Md} (desviació mediana dividit per la mediana) de totes les variables numèriques del `data.frame` `TRC` utilitzant la funció `sapply`. Quina és més dispersa?
 - (c) Explora la influència de cadascun dels factors `SEXE` i `MORT-12m` en les variables `DTSVE` i `DTDVE`. Fes gràfics de violí per representar aquesta relació.