

La colestiramina és una resina d'intercanvi aniònic que, combinada amb recomanacions de dieta, es fa servir com a tractament contra la hipercolesterolèmia, excés de colesterol LDL ("dolent") en sang. En un estudi sobre la seva eficàcia es va fer un seguiment de 164 pacients que patien hipercolesterolèmia. El grau d'acompliment del tractament per part d'aquests pacients va ser molt desigual, n'hi va haver que van seguir estrictament les recomanacions, però també n'hi va haver que les van seguir molt escassament.

Al fitxer "colestiramina.RDa" hi ha les dades d'aquest estudi. Amb aquest fitxer al vostre directori de treball, feu:

```
load(file = "colestiramina.RDa")
```

per carregar aquestes dades a la vostra sessió de R. Obtindreu l'objecte "cholost" que és un "data.frame" amb 2 columnes. La primera (variable "z") correspon al grau d'acompliment del tractament per part de cada pacient, la segona (variable "y") indica la millora experimentada en el seu nivell de colesterol. Totes dues variables estan expressades en tants per cent. Valors negatius de la segona columna indiquen un empitjorament, el nivell de colesterol va pujar. En principi, un esperaria que a més acompliment del tractament hi hagués més millora, i viceversa. Això s'hauria de traduir en que, per exemple, el coeficient de correlació de Pearson entre aquestes variables fos positiu.

Comproveu, amb algun mètode gràfic, que difícilment es podria considerar que aquestes variables segueixen una distribució normal, especialment la primera.

Qüestions plantejades:

- 1) Obté els intervals de confiança bootstrap—t, bootstrap—t simetritzat, bootstrap percentil i BCa pel coeficient de correlació de Pearson. Compara'ls amb l'interval de confiança paramètric normal que obtindries sobre les mateixes dades. Corroboren els resultats obtinguts la hipòtesi que més acompliment del tractament es correspon amb més millora al nivell de colesterol, i viceversa?
- 2) A l'apartat anterior es suposa que els càlculs per a l'obtenció dels intervals els has realitzat pel teu compte, sense utilitzar cap paquet R especialitzat en bootstrap. Busca una llibreria R que tingui funcions per realitzar càlculs bootstrap, i aplicant-les calcula els mateixos intervals de confiança bootstrap. Comprova si coincideixen amb els que has obtingut a l'apartat anterior.

Al Campus virtual de l'assignatura, Tema 4, hi trobaràs el document "No paramètrica - Mètodes de suavitzat (Prof. P. Delicado)". A la pàgina 101 d'aquest document, apartat 3.6.4, es proposa una prova d'hipòtesis no paramètrica general per intentar establir la possible dependència entre dues variables aleatòries. La prova es basa en un estadístic obtingut a partir de la raó de versemblances i fa servir estimacions nucli de la densitat. Al mateix document s'explica com calcular el p-valor mitjançant un enfoc de permutacions. A l'script R "Prova d'independència basada en estimadors nucli", també al Campus virtual, es porta a la pràctica aquest test.

- 3) Amb les dades de colestiramina, obté el p-valor associat al contrast i a l'estadístic de test anteriors (raó de versemblances i estimacions nucli de la densitat) però sota un

enfoc bootstrap (i no de permutacions). La simulació bootstrap ha de ser “semiparamètrica”, és a dir: aprofitant que hem estimat la funció de densitat de les variables implicades, hem de generar els valors d’acord amb aquestes funcions de densitat “nucli”. (Les principals dificultats que hauràs de resoldre per portar-ho a la pràctica seran com generar remostres bootstrap d’acord amb la hipòtesi nul·la (independència entre acompanyament i millora) i com generar dades bootstrap però a partir de la densitat estimada –no, per exemple, a partir de la distribució empírica.)

Material a entregar:

- a) Script R amb el codi dels càlculs realitzats.
- b) Document Pdf o Word explicant i justificant els càlculs realitzats, i explicant els resultats obtinguts i les conclusions que se’n treuen.