**Exercici .- 1 Eliminar els identificadors del conjunt de dades (0.5p)**

Responeu a les següent qüestions:

a. Quins són els atributs identificadors en el conjunt de dades?

Els atributs que identifiquen una persona de manera única i inequívocament són el DNI i el número de la SS.

b. Descriu les comandes que has emprat, indicant breument la funció de cadascun dels paràmetres

que has utilitzat.

dades\_subset = dades[,c(3,4,5)]. Amb aquesta comanda el que fem es que creem una nova variable “dades\_subset” on hi guardarem el contingut de les columnes 3, 4 i 5 que es corresponen a CP, Edat i salari.

**Exercici 2.- Re-identificació (0.5p)**

Responeu a les següent qüestions:

a. Quins d’aquests atributs o combinacions poden conduir a la identificació única d’un registre del conjunt de dades?

Hem trobat que tant per Edat, CP i Salari hi ha persones que tenen valors únics i llavors qualsevol dels 3 atributs poden identificar de manera única a una persona concreta. A més a més també hem trobat que les combinacions CP-Edat, CP-Salari, Edat-Salari i CP-Edat-Salari identifiquen a una persona concretament.

b. Indica quines comandes has fet servir per obtenir aquesta informació, explicitant el significat de les comandes i dels seus paràmetres.

dades\_CP\_freq = table(dades\_subset$CP)

Amb aquesta comanda creem una nova taula a partir del subset que hem treballat en l’apartat anterior per a la columna CP. Com a sortida d’aquesta comanda rebrem una taula amb els diferents valors que pren el CP i la freqüència amb la que aquests valors surten.

Si ara volem fer una combinació, la comanda resultant es molt similar:

dades\_CP\_Edat\_freq = table(dades\_subset$CP,dades\_subset$Edat)

Ara tenim una taula que ens mesura per a la variable1(CP) i per a la variable2(Edat) les combinacions possibles i quants cops apareixen.

Un cop hem probat totes les combinacions i tots els camps que considerem que poden ser quasi-identificadors utilitzarem:

'1' %in% TAULA

on TAULA és el nom de la taula generada anteriorment com per exemple dades\_CP\_freq i ens mostrarà TRUE si hi ha cap individu que es pot identificar de manera única.

**Exercici 3.- Aplicació de soroll additiu (additive noise) (0.5p)**

Responeu a les següent qüestions:

a. Indica la comanda que has fet servir, junt amb els paràmetres emprats i el significat de cadascun d’ells.

Nosaltres ho hem fet en 2 passos:

1: edat\_soroll <- addNoise(dades\_subset,'Edat',20). Amb això creem un objecte de tipus sdcMicroObj-class en el qual guardarem el soroll que hem afegit a la variable edat, el primer parametre és el data frame a partir del qual es treballa, el segon parametre és la variable o variables a les quals es vol afegir soroll i el tercer parametre es la quantitat de soroll que afegim

2: dades.an = dades\_subset[,c(1,3)].

dades.an$Edat <- edat\_soroll$xm. En aquestes dues comandes el que fem es crear un data frame “dades.an” on guardem el Salari i el CP de “dades\_subset” i després afegim el camp xm del objecte sdcMicroObj-class “edat\_soroll” que hem creat abans al data frame “dades.an”.

b. Executa el següent codi adjunt i explica els resultats obtinguts, justificant la resposta.

TODO

**Exercici 4.- Aplicació del mètode d’intercanvi de rang additiu (rang swap) (0.5p)**

Responeu a les següent qüestions:

a. Indica la comanda que has fet servir, junt amb els paràmetres emprats i el significat de cadascun d’ells.

dades.rs <- rankSwap(dades\_subset,'Edat',P=10). primer parametre és el data frame a partir del qual es treballa, el segon parametre és la variable o variables a les quals es vol afegir soroll i el tercer parametre(P) es el percentatge amb el qual es fara el rank swap

b. Genera una gràfica similar a la que hem emprat en l’exercici anterior – apartat b), on es pugui veure de forma visual i ràpida la dispersió o alteració de l’atribut Edat desprésd’aplicar el mètode d’intercanvi de rang.

TODO

**Exercici 5.- Càlcul de la utilitat (0.5p)**

Responeu a les següent qüestions:

a. Utilitza la comanda dUtility de la llibreria sdcMicro per a estimar la pèrdua de informació que s’ha produït en generar la versió pertorbada de l’atribut Edat emprant els mètodes de soroll additiu i intercanvi de rang (generats en els dos exercicis anteriors). Indica la comanda que has fet servir, junt amb els paràmetres emprats i el significat de cadascun d’ells.

TODO

b. Comenta els resultats obtinguts i descriu el significat dels valors obtingut i com es calculen.

TODO