

# PAC1 (Pràctica)

Josep Andreu Miralles

27/9/2020

## EXERCICI 1

- a) Calculeu els resums numèrics d'aquestes dades.
- b) Feu un histograma per explicar la variable.
- c) A partir dels resum numèrics i de l'histograma descriviu com són les dades.
- d) Feu la comanda `plot(nhtemp)`. Expliqueu el gràfic que surt. Comenteu si és un gràfic útil per aquestes dades

## Solució

- a) Carreguem l'arxiu i mitjançant l'instrucció “summary” calculem els resums numèrics.

```
summary(nhtemp)
```

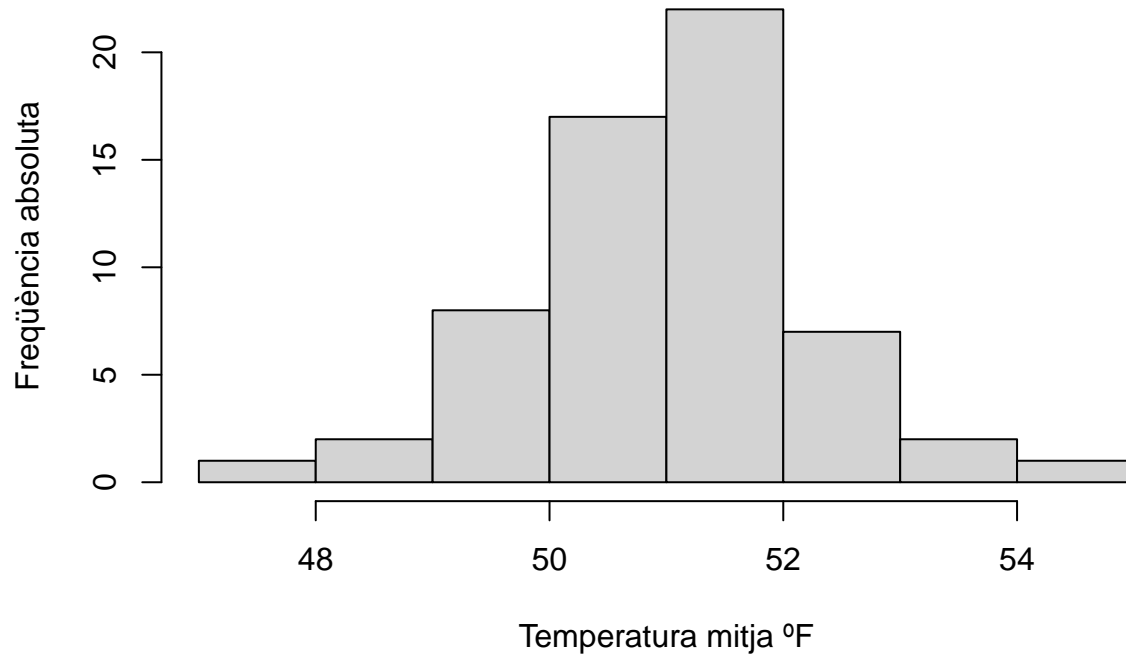
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  47.90   50.58   51.20   51.16   51.90   54.60
```

Els resums numèrics són 47.9, 50.575, 51.2, 51.16, 51.9, 54.6.

- b) Fem l'histograma mitjançant l'instrucció “hist” i en modifiquem el títol i el nom dels eixos.

```
hist(x = nhtemp, main = "Histograma de temperatures mitges a NH 1912 a 1971", xlab = "Temperatura mitja")
```

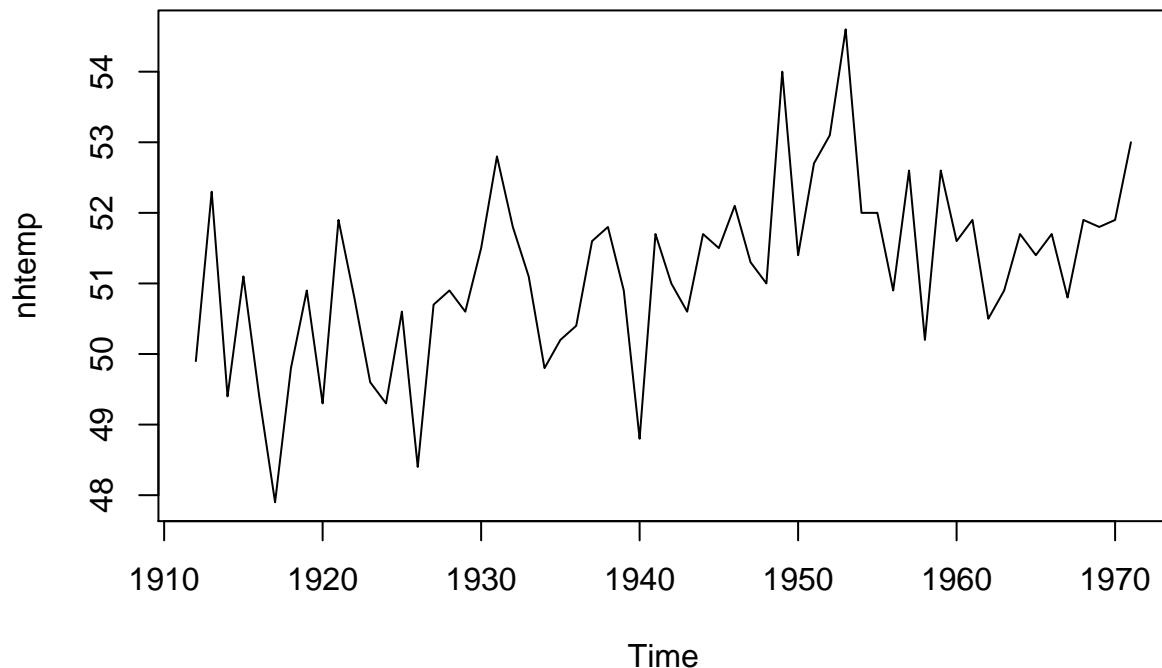
## Histograma de temperatures mitges a NH 1912 a 1971



c) Com podem veure, les classes són  $[47,48)$  ... fins a  $(54,55]$ . És un histograma centrat i unimodal, amb un únic pic corresponent a l'interval  $[51,52)$ , el comportament és lleugerament simètric, i no es detecten dades extremes ni classes buides.

d) Executem l'ordre "plot(nhtemp)".

```
plot(nhtemp)
```



El que s'obté és un gràfic de línies, on s'observa l'evolució de la temperatura (eix Y - ordenades) amb el pas dels anys (eix X - abscisses). És un gràfic que mostra clarament l'evolució de la temperatura mitja, i és útil per que ens serveix per detectar tendències i relacions amb altres gràfics de línies; en aquest cas podem detectar que hi ha hagut un lleuger increment en la temperatura mitja de New Haven, i amb un clar pic als anys 50s.

## EXERCICI 2

- Trobeu els resums numèrics (mitjana, mediana, quartils, desviació típica mostral) del nombre de contractes signats l'any 2019 i comenteu els resultats.
- Obteniu una taula de quants municipis de cada província hi ha a la base de dades. Feu un gràfic adient per mostrar la distribució. Comenteu els resultats.
- Feu un boxplot de la mitjana dels preus mitjans dels contractes signats l'any 2019 distingint per províncies i comenteu quines diferències s'hi aprecien.

## Solució

- Carreguem l'arxiu i busquem els resums numèrics demanats tenint en compte de despreciar els valors buits que han estat substituïts per NA per R amb la instrucció `na.rm=TRUE`.

```
dades <- read.csv("PIS_MUN.csv", sep = ";")
attach(dades)
resum.num <- cbind (mean(N2019,na.rm=TRUE),median(N2019,na.rm=TRUE),sd(N2019,na.rm=TRUE),t(quantile(N2019,
colnames(resum.num) <- c("Mitjana","Mediana","DTM","Mínim","Q1","Q2","Q3","Màxim")
resum.num
```

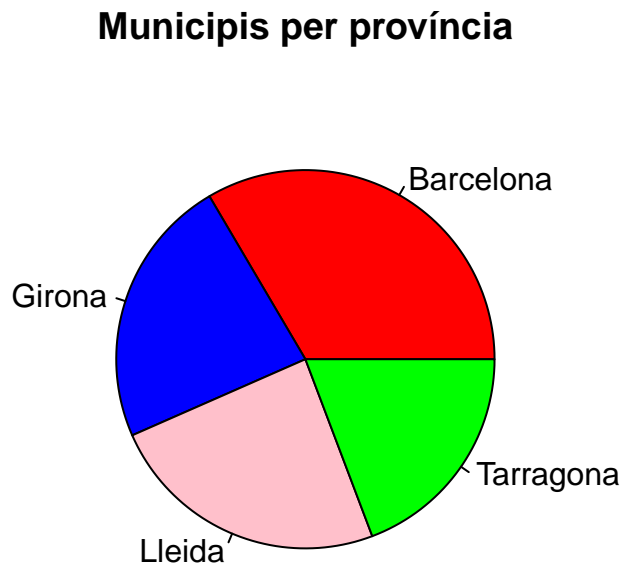
```
##      Mitjana Mediana      DTM Mínim Q1 Q2   Q3 Màxim
## [1,] 234.6222      12 1880.096     1  3 12 61.5 51294
```

El que veiem es que hi ha una dada que clarament distorsiona la mostra, i que per tan fa augmentar la mitjana però que no afecta la mediana.

- b) Aillem la variable província i creem una taula. El diagrama de sectors ens serveix per representar la proporció de municipis per províncies. El diagrama ens indica que la proporció de municipis més gran pertany a Barcelona, que Girona i Lleida tenen una proporció similar de municipis i que Tarragona és la que té una proporció més petita de municipis.

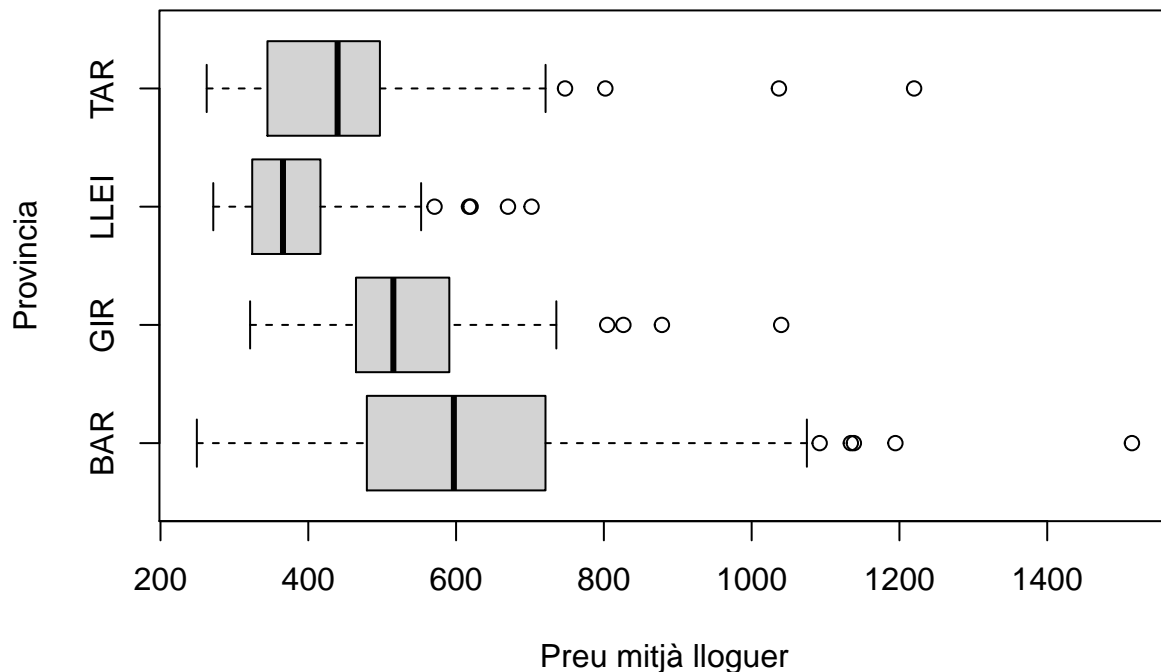
```
taula_prov <- factor(PROV)
table(taula_prov)

## taula_prov
## Barcelona   Girona   Lleida Tarragona
##         320       221       231       184
pie(table(taula_prov),col = c("red", "blue", "pink", "green"),main = "Municipis per província")
```



- c) Representem la comparació del diagrama de caixes en horitzontal.

```
boxplot(dades$M2019 ~ dades$PROV,xlab="Preu mitjà lloguer",ylab="Província",names=c("BAR","GIR","LLEI"),
```



### EXERCICI 3

Imagineu que volem estudiar a qui votaran els estudiants de la UOC a les properes eleccions. Podem fer-ho de diverses maneres:

- Enviant un missatge a través del campus a tots els estudiants i analitzant les respostes.
- Seleccionem 9 comarques a l'atzar i seleccionem a l'atzar 50 estudiants d'aquestes comarques als quals truquem per telèfon.
- Assignem un número a cada estudiant i truquem als que el seu número és de la forma 12324\*n fins que obtinguem 450 respostes.

Digueu quin tipus de mostreig fem servir en cada cas i quins són els seus inconvenients. Quin mostreig dels proposats escolliríeu i per què?

### Solució

- Podríem dir que no es tracta d'un mostreig, sinó d'un estudi exhaustiu, ja que es consulten tots els individus que formen part de la població objecte de l'estudi. Té l'inconvenient que la consulta requereix que l'estudiant respongui i se senti interpel·lat per tal que contesti.
- Es tracta d'un mostreig per conglomerats. Té l'inconvenient que les comarques són molts diferents les unes a les altres, i les triades poden no abarcar la totalitat de les diferents tipus de comarques, i llavors la mostra pot perdre representativitat. A més a més poden resultar triades comarques que tenen una població molt inferior o superior respecte altres que no han sortit triades i això faria que les característiques d'aquestes quedessin infrarepresentades o sobrerepresentades.
- Es tracta d'un mostreig sistemàtic. Té el problema que si existeix una certa periodicitat que coincideixi

amb l'interval triat.

Jo em decantaria per utilitzar el mostreig sistemàtic, ja que no depèn tan de la voluntat de resposta dels individus objecte del mostreig i tampoc ens podem trobar en el cas de tenir una mostra sobrerrepresentada o infrarepresentada segons la comarca de residència dels individus objecte de la mostra.