

Solució de la PAC1

Estadística - Primavera 2020 - UOC

Pregunta 1 (40 punts)

- a) Quants contractes de lloguer es van signar a Catalunya l'any 2018?
- b) A quantes ubicacions (municipis o districtes) es van signar més de 6000 contractes de lloguer l'any 2018.
- c) A quantes ubicacions es van signar més contractes l'any 2018 que l'any 2017? Quina proporció representa sobre el total?

Nota: podeu consultar el *Mòdul 1. L'entorn estadístic R* (p. 17) i l'apartat 2.7 *Bases de dades*.

Solució

- a) Calculem

```
t2018<-sum(dades$N2018)
t2018

## [1] 166410
```

I per tant es van signar 166410 contractes de lloguer l'any 2018.

- b) Calculem

```
t2018_6000<-length(dades$N2018[dades$N2018>6000])
t2018_6000

## [1] 2
```

i per tant a 2 ubicacions es van signar més de 6000 contractes de lloguer l'any 2018.

- c) Calculem

```
m2018<-length(dades$N2018[dades$N2018>dades$N2017])
m2018

## [1] 500

p2018<-m2018/length(dades$N2018)
p2018

## [1] 0.5230126
```

i per tant es van signar més contractes l'any 2018 que el 2017 a 500 ubicacions, el que representa el 52.3% de les ubicacions

Pregunta 2 (20 punts)

Considerem la variable que ens dona la diferència entre el nombre de contractes de lloguer signats l'any 2018 i l'any 2017 (a la que anomenem DIF1817). Calculeu el seus resums numèrics (amb *summary*) i la seva desviació típica mostral i comenteu els resultats, en particular la diferència entre la mitjana i la mediana.

Solució

Creem la nova variable i fem els càlculs

```
DIF1817<-dades$N2018-dades$N2017  
SUM1817<-summary(DIF1817)  
SUM1817
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
## -370.00  -1.00    1.00   12.66    6.00 1489.00
```

```
sd(DIF1817)
```

```
## [1] 75.74925
```

La mitjana és 12.66, força més alta que la mediana (1). Això és degut a que hi ha valors positius molt allunyats (com el màxim que és 1489) que afecten a la mitjana però no a la mediana.

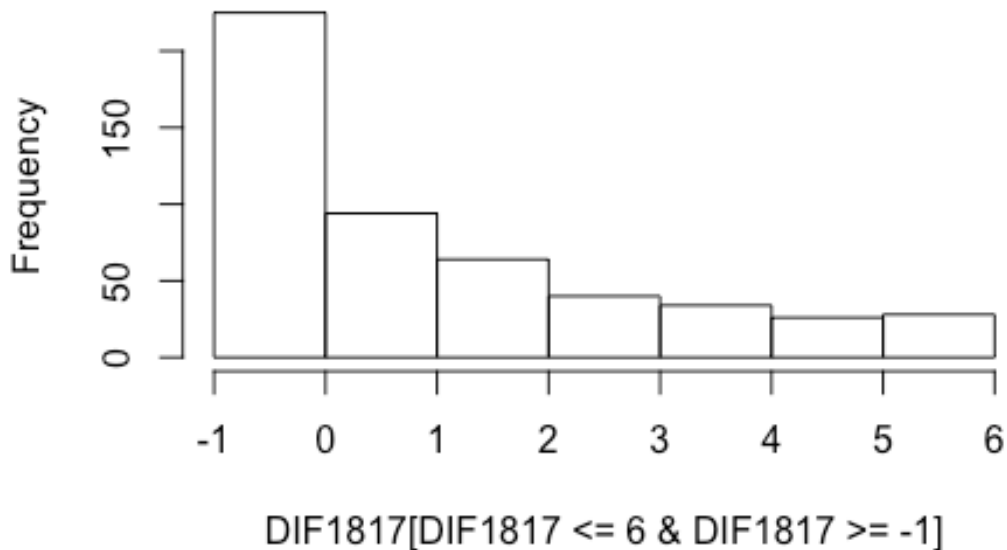
Pregunta 3 (20 punts)

Dibuixeu un histograma dels valors de la variable DIF1817 però usant només els valors que es troben entre el primer i el tercer quartil (incloent aquests valors). Comenteu el resultat.

Solució

```
hist(DIF1817[DIF1817<=6 & DIF1817>=-1],breaks=5)
```

Histogram of DIF1817[DIF1817 <= 6 & DIF1817 >= -



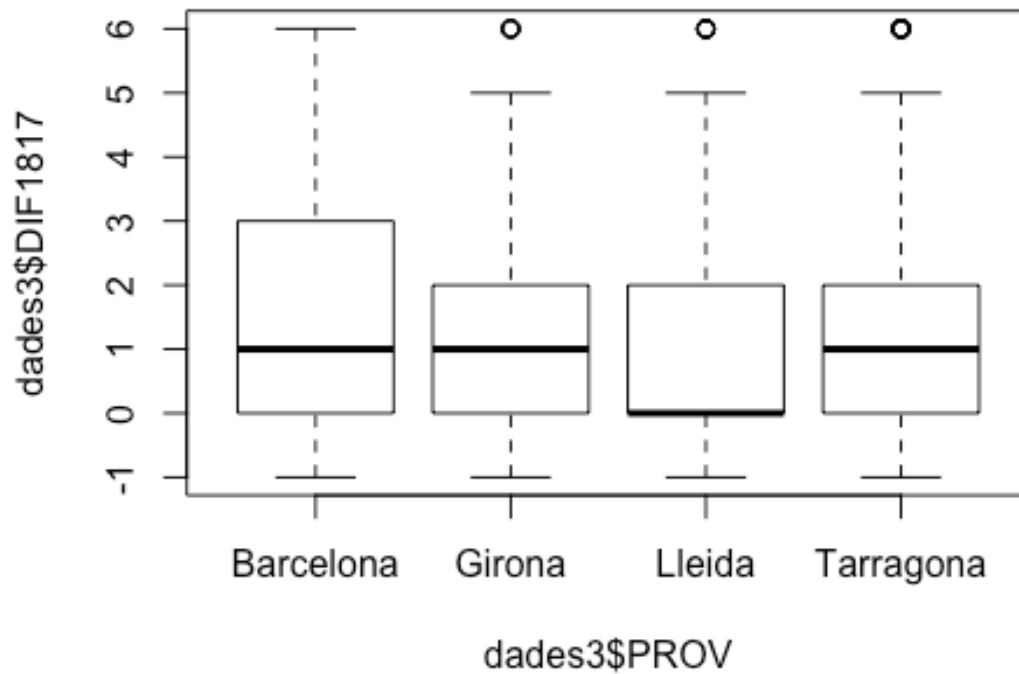
És unimodal i força asimètric, amb la moda a l'interval $[-1,0]$.

Pregunta 4 (20 punts)

Feu un boxplot de la variable DIF1817, usant com a l'apartat anterior els valors que es troben entre el primer i el tercer quartil (incloent aquests valors) i distingint per províncies; comenteu quines diferències s'hi aprecien. Indicació: en primer lloc podeu afegir la variable DIF1817 a la *data.frame* anomenada *dades* amb la funció *cbind* i després fer un *subset*.

Solució

```
dades2<-cbind(dades,DIF1817)
dades3<-subset(dades2,DIF1817<=6 & DIF1817>=-1)
boxplot(dades3$DIF1817~dades3$PROV)
```



Barcelona és la que té més dispersió als valors positius; les distribucions de Girona i Tarragona són molt similars i la de Lleida és la més asimètrica amb la mediana situada al 0.