

# Prova a casa

Marc\_Cabanas

3 juliol de 2019

Primer de tot, carreguem les llibreries que farem servir per a la prova:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readxl)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.
2.1 --
```

```
## v tibble  2.0.1      v purrr   0.2.5
## v tidyr   0.8.2      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0
```

```
## -- Conflicts ----- tidyverse_conflict
s() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(foreign)
library(ggthemes)
library(knitr)
```

**1. En quins municipis hi havia més atur abans de la crisi? I després de la crisi? En quins indrets va créixer més l'atur? En quins menys?**

Per contestar aquesta pregunta seguirem els següents passos:

a. Carreguem les bases de dades

```
setwd("C:/Users/Marc Cabañas/Desktop/Datos/examen_final")
atur <- read_xlsx("municatatur.xlsx")
pob08 <- read_xlsx("municipiscat08_pob.xlsx")
pob12 <- read_xlsx("municipiscat12.xlsx")
```

b) Unim els dataframe en un de sol: (Per fer-ho, mirarem que les dades tinguin un nom en comú per poder unir-les mitjançant *merge*. En aquest cas, la variable serà *municipi*)

```
names(atur)
```

```
## [1] "codimun"      "municipi"      "desembre de 2012"
## [4] "desembre de 2008"
```

```
names(pob08)
```

```
## [1] "Codi"          "Literal"
## [3] "Homes. De 0 a 14 anys" "Homes. De 15 a 64 anys"
## [5] "Homes. De 65 anys i més" "Dones. De 0 a 14 anys"
## [7] "Dones. De 15 a 64 anys" "Dones. De 65 anys i més"
## [9] "Total. De 0 a 14 anys" "Total. De 15 a 64 anys"
## [11] "Total. De 65 anys i més"
```

```
names(pob12)
```

```
## [1] "Codi"          "Literal"
## [3] "Homes. De 0 a 14 anys" "Homes. De 15 a 64 anys"
## [5] "Homes. De 65 anys i més" "Dones. De 0 a 14 anys"
## [7] "Dones. De 15 a 64 anys" "Dones. De 65 anys i més"
## [9] "Total. De 0 a 14 anys" "Total. De 15 a 64 anys"
## [11] "Total. De 65 anys i més"
```

```
names(pob08)[2] ="municipi"
names(pob12)[2]="municipi"
names(pob08)[10]="Població_activa08"
names(pob12)[10]="Població_activa12"
names(atur)[3]="atur12"
names(atur)[4]="atur08"
```

```
names(pob08)
```

```
## [1] "Codi"          "municipi"
## [3] "Homes. De 0 a 14 anys" "Homes. De 15 a 64 anys"
## [5] "Homes. De 65 anys i més" "Dones. De 0 a 14 anys"
## [7] "Dones. De 15 a 64 anys" "Dones. De 65 anys i més"
## [9] "Total. De 0 a 14 anys" "Població_activa08"
## [11] "Total. De 65 anys i més"
```

```
names(pob12)
```

```
## [1] "Codi"          "municipi"
## [3] "Homes. De 0 a 14 anys" "Homes. De 15 a 64 anys"
## [5] "Homes. De 65 anys i més" "Dones. De 0 a 14 anys"
## [7] "Dones. De 15 a 64 anys" "Dones. De 65 anys i més"
## [9] "Total. De 0 a 14 anys" "Població_activa12"
## [11] "Total. De 65 anys i més"
```

```
pob08[3:9] <- list(NULL)
pob08[1] <-NULL
pob08[3] <- NULL
pob12[1] <- NULL
pob12[2:8] <- NULL
pob12[3] <- NULL
dades <- merge(pob08,pob12)
aturcat <- merge(dades,atur)
```

Ja hem unit totes les dades en un mateix dataframe anomenat **aturcat** on hi ha l'atur de cada any per cada municipi i la població activa d'aquests.

c. Ja tenim totes les dades en la mateixa base, ara busquem la taxa d'atur de 2008 i 2012:

```
atur08 <- aturcat$atur08
pob08 <- aturcat$Població_activa08
atur12 <- aturcat$atur12
pob12 <- aturcat$Població_activa12
atur08 <- as.numeric(atur08)
pob08 <- as.numeric(pob08)
atur12 <- as.numeric(atur12)
pob12 <- as.numeric(pob12)
aturcat$perc_08 <- (atur08/pob08)*100
aturcat$perc_12 <- (atur12/pob12)*100
```

Ara, endrecem de més gran a més petit:

```
aturcat <- aturcat[with(aturcat,order(-aturcat$perc_08)),]
```

I ara obtenim els 10 primers nombres en atur al 2008:

```
head(aturcat,10)
```

```
##          municipi Població_activa08 Població_activa12 codimun
## 928      Vilanova del Camí          8764          8576  08302
## 751 Santa Margarida de Montbui          6787          6509  08250
## 717      Sant Miquel de Fluvià           487           541  17175
## 145          Cabanabona                45            49  25060
## 458          Montesquiú                586           593  08131
## 465          Montmell, el            1013          1010  43090
## 254          Creixell               2150          2213  43050
## 80          Badia del Vallès          9651          9035  08904
## 544          Pineda de Mar          18229          17478  08163
## 556      Poble de Claramunt, la        1497          1449  08165
##      atur12 atur08 perc_08 perc_12
## 928   1697   1294 14.76495 19.787780
## 751   1390   1001 14.74878 21.355047
## 717    105    67 13.75770 19.408503
## 145     3     6 13.33333  6.122449
## 458    82    78 13.31058 13.827993
## 465   201   134 13.22804 19.900990
## 254   387   276 12.83721 17.487573
## 80   1910  1238 12.82769 21.140011
## 544   3187  2292 12.57337 18.234352
## 556    207   186 12.42485 14.285714
```

I els 10 més baixos:

```
tail(aturcat,10)
```

```
##          municipi Població_activa08 Població_activa12 codimun
## 280      Esterri de Cardós           51             54   25087
## 287          Fígols              23             24   08080
## 288      Fígols i Alinyà         166            161   25908
## 365          Ivorra              84             68   25114
## 446      Molsosa, la            85             76   25136
## 622      Riu de Cerdanya         90             90   25913
## 689 Sant Jaume de Frontanyà      20             21   08216
## 769      Savallà del Comtat       43             46   43143
## 819          Tiurana            44             44   25222
## 178      Canonja, la             0            3945   43907
##      atur12 atur08 perc_08   perc_12
## 280      2      0      0   3.703704
## 287      2      0      0   8.333333
## 288      3      0      0   1.863354
## 365      3      0      0   4.411765
## 446      2      0      0   2.631579
## 622      1      0      0   1.111111
## 689      1      0      0   4.761905
## 769      1      0      0   2.173913
## 819      0      0      0   0.000000
## 178     468      0     NaN  11.863118
```

Si els visualitzem, ens surt el següent:

```
library(xtable)
mesalt08 <- head(aturcat,10)
mesbaix <- tail(aturcat,10)
mesalt08[2:6] <- NULL
mesalt08[3:4] <- NULL
mesbaix [2:6] <- NULL
mesbaix[3:4] <- NULL
print(mesalt08,
      caption=" 10 municipis amb més atur al 2008")
```

```
##          municipi perc_08
## 928      Vilanova del Camí 14.76495
## 751 Santa Margarida de Montbui 14.74878
## 717      Sant Miquel de Fluvià 13.75770
## 145      Cabanabona 13.33333
## 458      Montesquiu 13.31058
## 465      Montmell, el 13.22804
## 254      Creixell 12.83721
## 80       Badia del Vallès 12.82769
## 544      Pineda de Mar 12.57337
## 556      Pobla de Claramunt, la 12.42485
```

I la taula dels 10 més baixos:

```
print(mesbaix,
      caption=" 10 municipis amb menys atur al 2008")
```

```
##          municipi perc_08
## 280      Esterri de Cardós    0
## 287          Fígols        0
## 288      Fígols i Alinyà    0
## 365          Ivorra        0
## 446      Molsosa, la        0
## 622      Riu de Cerdanya    0
## 689 Sant Jaume de Frontanyà 0
## 769      Savallà del Comtat 0
## 819          Tiurana        0
## 178      Canonja, la      NaN
```

Ara fem el mateix per a l'atur al 2012:

```
aturcat <- aturcat[with(aturcat,order(-aturcat$perc_12)),]
head(aturcat,10)
```

```
##          municipi Població_activa08 Població_activa12 codimun
## 751 Santa Margarida de Montbui      6787      6509    08250
## 80      Badia del Vallès          9651      9035    08904
## 759      Santa Oliva             2100      2189    43140
## 460      Montferri               218       252    43089
## 465      Montmell, el            1013      1010    43090
## 928      Vilanova del Camí       8764      8576    08302
## 122      Bisbal del Penedès, la   2289      2250    43028
## 717      Sant Miquel de Fluvià     487       541    17175
## 883      Vendrell, el          24625     24853    43163
## 18      Albinyana               1537      1536    43002
##      atur12 atur08 perc_08 perc_12
## 751    1390    1001 14.74878 21.35505
## 80     1910    1238 12.82769 21.14001
## 759     446     258 12.28571 20.37460
## 460      51      22 10.09174 20.23810
## 465     201     134 13.22804 19.90099
## 928    1697    1294 14.76495 19.78778
## 122     441     274 11.97029 19.60000
## 717     105      67 13.75770 19.40850
## 883    4608    3045 12.36548 18.54102
## 18      284     159 10.34483 18.48958
```

i els 10 més baixos:

```
tail(aturcat,10)
```

```
##          municipi Població_activa08 Població_activa12 codimun
## 138          Bovera          192          169    25056
## 507           Orpí          134          116    08152
## 478          Nalec           59           59    25145
## 119          Biosca          131          134    25055
## 654 Sant Agustí de Lluçanès          71           68    08195
## 622          Riu de Cerdanya          90           90    25913
## 670 Sant Esteve de la Sarga          98           93    25196
## 487           Odèn         200          178    25148
## 528          Pardines          96           89    17125
## 819          Tiurana          44           44    25222
##      atur12 atur08   perc_08   perc_12
## 138      3      1 0.5208333 1.7751479
## 507      2      8 5.9701493 1.7241379
## 478      1      1 1.6949153 1.6949153
## 119      2      2 1.5267176 1.4925373
## 654      1      3 4.2253521 1.4705882
## 622      1      0 0.0000000 1.1111111
## 670      1      4 4.0816327 1.0752688
## 487      1      2 1.0000000 0.5617978
## 528      0      1 1.0416667 0.0000000
## 819      0      0 0.0000000 0.0000000
```

I tornem a visualitzar en una taula:

```
mesatur12 <- head(aturcat,10)
mesatur12[2:7] <- NULL
mesatur12[3] <- NULL
print(mesatur12)
```

```
##          municipi   perc_12
## 751 Santa Margarida de Montbui 21.35505
## 80      Badia del Vallès 21.14001
## 759      Santa Oliva 20.37460
## 460      Montferri 20.23810
## 465      Montmell, el 19.90099
## 928      Vilanova del Camí 19.78778
## 122      Bisbal del Penedès, la 19.60000
## 717      Sant Miquel de Fluvià 19.40850
## 883      Vendrell, el 18.54102
## 18      Albinyana 18.48958
```

I els 10 més baixos al 2012:

```
mesbaix12 <- tail(aturcat,10)
mesbaix12[2:7] <- NULL
mesbaix12[3] <- NULL
print(mesbaix12)
```

```
##          municipi   perc_12
## 138          Bovera 1.7751479
## 507           Orpí 1.7241379
## 478          Nalec 1.6949153
## 119          Biosca 1.4925373
## 654 Sant Agustí de Lluçanès 1.4705882
## 622          Riu de Cerdanya 1.1111111
## 670 Sant Esteve de la Sarga 1.0752688
## 487           Odèn 0.5617978
## 528          Pardines 0.0000000
## 819          Tiurana 0.0000000
```

- d. Ara observem on ha crescut més l'atur entre 2008 i 2012: Per fer-ho, fem una nova variable que es digui *dif*.  
Mostrem els 10 municipis catalans on més va créixer l'atur.

```
aturcat$dif <- aturcat$perc_12-aturcat$perc_08
aturcat <- aturcat[with(aturcat,order(-aturcat$dif)),]
head(aturcat,10)
```

```
##          municipi Població_activa08 Població_activa12 codimun
## 757 Santa Maria de Miralles          90           88   08257
## 831 Torre de Fontaubella, la         81           80   43151
## 775 Senterada                     104          102   25202
## 150 Cabra del Camp                 739          748   43036
## 460 Montferri                     218          252   43089
## 944 Vinebre                       325          319   43177
## 329 Ginestar                      710          578   43067
## 274 Espunyola, l'                  164          154   08078
## 773 Senan                         39           40   43146
## 433 Menàrguens                    509          573   25134
##      atur12 atur08 perc_08 perc_12 dif
## 757      15      4 4.444444 17.04545 12.601010
## 831      13      3 3.703704 16.25000 12.546296
## 775      15      3 2.884615 14.70588 11.821267
## 150     137     60 8.119080 18.31551 10.196428
## 460      51     22 10.091743 20.23810 10.146352
## 944      48     16 4.923077 15.04702 10.123945
## 329      76     22 3.098592 13.14879 10.050197
## 274      21      6 3.658537 13.63636 9.977827
## 773       6      2 5.128205 15.00000 9.871795
## 433      83     24 4.715128 14.48517 9.770038
```

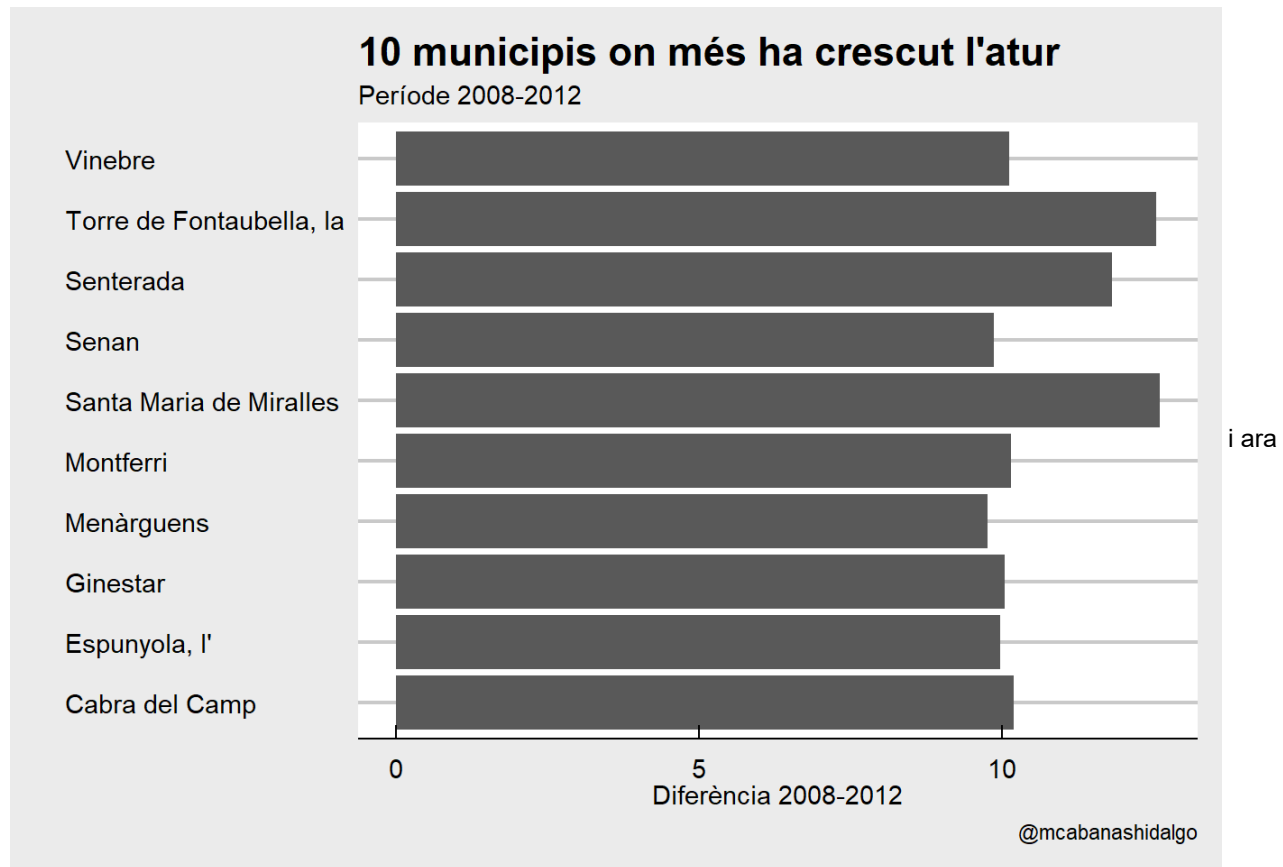
i els 10 on menys ha crescut:

```
tail(aturcat,10)
```

```
##          municipi Població_activa08 Població_activa12 codimun
## 654 Sant Agustí de Lluçanès          71           68   08195
## 670 Sant Esteve de la Sarga          98           93   25196
## 578 Pradell de la Teixeta          108          132   43115
## 901 Viladasens                     137          140   17216
## 507 Orpí                          134          116   08152
## 792 Sora                          120          131   08272
## 647 Sales de Llierca                96          107   17154
## 520 Palau de Santa Eulàlia          76           73   17119
## 145 Cabanabona                     45           49   25060
## 178 Canonja, la                     0          3945   43907
##      atur12 atur08 perc_08 perc_12 dif
## 654       1       3 4.225352 1.470588 -2.754764
## 670       1       4 4.081633 1.075269 -3.006364
## 578       8      10 9.259259 6.060606 -3.198653
## 901       4       9 6.569343 2.857143 -3.712200
## 507       2       8 5.970149 1.724138 -4.246011
## 792       3       8 6.666667 2.290076 -4.376590
## 647       4       8 8.333333 3.738318 -4.595016
## 520       2       6 7.894737 2.739726 -5.155011
## 145       3       6 13.333333 6.122449 -7.210884
## 178     468       0      NaN 11.863118      NaN
```

- e. Ara farem crearem dues variables *mesatur* i *menysatur* on apareguin els 10 municipis on més i menys ha pujat l'atur:

```
mesatur <- head(aturcat,10)
menysatur <- tail(aturcat,10)
ggplot(mesatur,aes(municipi,dif)) +
  geom_bar(stat="identity") +
  coord_flip() +
  theme_economist_white()+
  labs(x= "",
       y= "Diferència 2008-2012",
       title = "10 municipis on més ha crescut l'atur",
       subtitle = "Període 2008-2012",
       caption = "@mcabanashidalgo")
```



observem aquells on menys ha crescut l'atur:

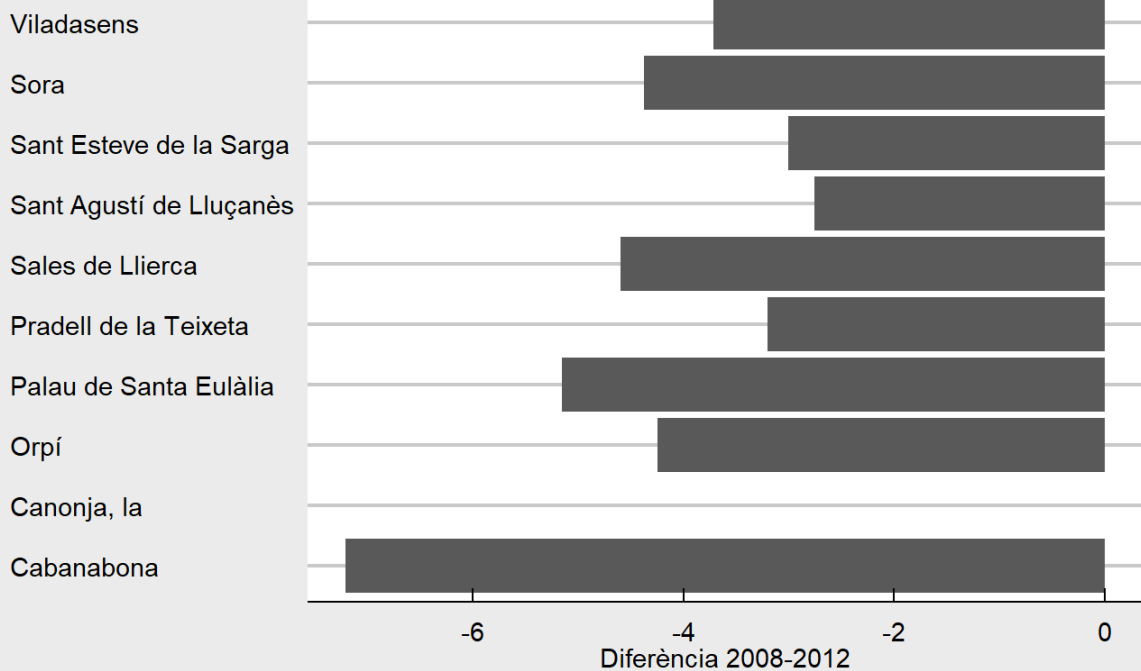
```
ggplot(menysatur,aes(municipi,dif)) +
  geom_bar(stat="identity") +
  coord_flip() +
  theme_economist_white()+
  labs(x= "",
       y= "Diferència 2008-2012",
       title = "10 municipis on menys ha crescut l'atur",
       subtitle = "Període 2008-2012",
       caption = "@mcabanashidalgo")
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



## 10 municipis on menys ha crescut l'atur

Període 2008-2012



2. Quins

@mcabanashidalgo

### factors poden explicar aquesta variació en els nivells i creixement de l'atur?

La segona part de la prova demana que investiguem quins són els possibles factors que afectin a la variació del atur. Per fer-ho, he buscat la xifra d'hotels al 2008 i al 2012, ja que la demanda hotelera, en època d'atur cau ja que hi ha menys diners per gastar, i per tant, menys vacances i pot comportar un tancament d'establiments, i la segona, el nombre de naixements en aquests dos anys, ja que normalment, en època d'atur es tenen menys nens.

##carreguem Les dades##

```
naix08 <- read_xlsx("C:/Users/Marc Cabañas/Desktop/Datos/examen_final/naix_08.xlsx")
naix12 <- read_xlsx("naix_12.xlsx")
hot08 <- read_xlsx("establiments_hot08.xlsx")
hot12 <- read_xlsx("establiments_hot12.xlsx")
names(naix08)[3] <- "Naix2008"
names(naix12)[3] <- "Naix2012"
names(hot08)[3] <- "Hotels08"
names(hot12)[3] <- "Hotels12"
hot08[1] <- NULL
hot12[1] <- NULL
naix08[1] <- NULL
naix12[1] <- NULL
```

Ara unim les variables per generar un sol dataframe

```
naixements <- merge(naix08, naix12)
hotels <- merge(hot08, hot12)
ex2 <- merge(naixements, hotels)
total <- merge(aturcat, ex2)
total$dif_naix <- (total$Naix2012 - total$Naix2008)
total$dif_hotel <- total$Hotels12 - total$Hotels08
total$perc_naix <- (total$dif_naix / total$Hotels08) * 100
total$perc_hotel <- (total$dif_hotel / total$Hotels08) * 100
```

Ara fem els descriptius de les variables explicatives:

```
##Naixements##
summary(total$dif_naix)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1014.00   -8.00   -1.00   -12.23    1.00   56.00
```

```
##Hotels##
summary(total$dif_hotel)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -8.00000  0.00000  0.00000  0.1066  0.00000  70.00000
```

I ara fem les regressions simples corresponents:

```
r1 <- lm(total$dif~total$dif_naix, data = total)
summary(r1)
```

```
##
## Call:
## lm(formula = total$dif ~ total$dif_naix, data = total)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.634   -1.361    0.108    1.333    9.184
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.422911   0.076811  44.563 < 2e-16 ***
## total$dif_naix -0.005416   0.001473  -3.677  0.00025 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.296 on 944 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.01412,    Adjusted R-squared:  0.01307
## F-statistic: 13.52 on 1 and 944 DF,  p-value: 0.0002496
```

La variable *naixements* demostra que allà on la xifra de naixements és 0, la variació és de 3,422, i si augmenta en un punt percentual, l'atur baixa en 0,005. L'Rquadrat és d'1,4%, és a dir, explica en aquesta proporció la variable.

```
r2 <- lm(total$dif~total$dif_hotel, data=total)
summary(r2)
```

```
##
## Call:
## lm(formula = total$dif ~ total$dif_hotel, data = total)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.7009  -1.3739   0.1075   1.3838   9.1110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.490028    0.075244  46.383  <2e-16 ***
## total$dif_hotel -0.004443    0.030811  -0.144    0.885
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.312 on 944 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  2.203e-05, Adjusted R-squared:  -0.001037
## F-statistic: 0.02079 on 1 and 944 DF,  p-value: 0.8854
```

En aquest cas, la el percentatge d'hotels que hi ha a cada municipi, modifica la variaició en 3,49 quan la xifra d'hotels és 0. L'R quadrat ajustat és de -0.1, per tant, no explica gaire la variació de l'atur.

**Conclusió** L'atur va afectar a tota la població catalana entre els anys 2008 i 2012 (encara que les seves conseqüències encara hi duren), però va fer-ho especialment allà on van decreixer els naixements.

Les dades utilitzades en aquest exercici són de l'IDESCAT.