

# Residus Catalunya

Genís Bosch, Bernat Armengol

5/10/2021

## 1. Descripció del dataset

Treballarem amb el dataset d'estadístiques de residus municipals proporcionat per la Generalitat de Catalunya a través del portal “dades obertes Catalunya”. Aquest dataset recull les estadístiques de residus municipals de Catalunya. Incorpora les quantitats de recollida selectiva per a les diferents fraccions de residus, i també els totals de recollides en termes absoluts i per càpita.

Complementari a aquest dataset hi haurà la taula de municipis participants en el model de recollida selectiva porta a porta, que ens permetrà separar en dos grups les dades de residus municipals, obtenint així per una banda els municipis amb servei porta a porta i per altre els municipis amb servei de recollida selectiva a vorera.

Amb aquests datasets podrem investigar l'eficiència del mètode porta a porta i avaluar si aquest aporta una clara millora sobre la tradicional recollida selectiva de vorera.

## 2-3. Integració, neteja i selecció de dades d'interès a analitzar

Hem obtingut dos datasets: Un des de la pàgina web de dades obertes de Catalunya, en el que apareixen les dades de la recollida selectiva de tots els pobles de Catalunya. Aquest es pot trobar a l'enllaç: <https://analisi.transparenciacatalunya.cat/Medi-Ambient/Estad-stiques-de-residus-municipals/69zu-w48s>

L'altre, obtingut a partir d'una taula HTML a la web de Porta a Porta (<https://www.portaaporta.cat/ca/municipis.php>), la qual hem copiat i enganxat en un full de càlcul Excel, per a després passar a CSV i obrir amb R.

En fusionar les dues taules per nom de municipi, hem trobat una quarantena de municipis escrits diferents, que hem hagut d'editar manualment en el fitxer origen. Com que les dades del fitxer “Estadístiques de residus municipals” contenen varies entrades per a cada poble (tantes com anys registrats), hem editat el fitxer “PortaAPorta.csv”, on només apareixen els domicilis que participen en la iniciativa Porta a Porta i solament ho fan un cop.

```
library(stringr)
residus <- read.csv("./Estad_stiques_de_residus_municipals.csv", encoding = "UTF-8")
portaAporta <- read.csv("./PortaAPorta.csv", encoding = "UTF-8")

portaAporta$Municipis <- tolower(portaAporta$Municipis)
residus$Municipi <- tolower(residus$Municipi)

residusR <- merge(x = residus, y = portaAporta, by.x = "Municipi", by.y="Municipis", all.x = TRUE)
pap <- residusR[!is.na(residusR$X..població.servida.amb.el.PaP),]
```

```

residusR$Municipi <- str_to_title(residusR$Municipi)

con<-file('./residus_catalunya.csv',encoding="UTF-8")
write.csv(residusR,file=con)

data <- read.csv("./residus_catalunya.csv", encoding = "UTF-8")
#head(data)

```

Ja amb els datasets fusionats, veiem que totes les poblacions que no es troben en la llista de porta a porta han obtingut NAs per a les noves columnes. Aquest resultat ens és convenient, a excepció de la columna que conté el percentatge de població coberta pel mètode de recollida selectiva porta a porta. Així doncs, canviem les files les quals aquesta columna està en NA amb un 0, indicant així que no hi ha porta a porta per a ningú.

Per altra banda, per als pobles que pertanyen a la iniciativa porta a porta, a partir de la data d'inici de recollida porta a porta (que és diferent per cada municipi) avaluem si l'any en qüestió el municipi estava utilitzant la recollida selectiva de vorera o la porta a porta, i li assignem el seu percentatge de població coberta en cas de porta a porta, un 0 en cas de vorera i un NA en cas del mateix any, ja que les dades són anuals i no podríem determinar les porcions de residus que s'han tractat amb un mètode i l'altre dins l'any de canvi de mètode. Aquestes entrades amb NA no les tindrem en compte per a l'anàlisi.

Aprofitem també per a reduir la informació de la fracció de residus que es recullen en el porta a porta, quedant-nos només amb el nombre de materials diferents que es recullen i obtenint així una columna de tipus enter que pot variar entre 0 i 5.

```

library(stringr)
data$X..població.servida.amb.el.PaP <- as.numeric(data$X..població.servida.amb.el.PaP)

for (i in 1:nrow(data)) {
  if(is.na(data$X..població.servida.amb.el.PaP[i]) || data$X..població.servida.amb.el.PaP[i] == 'NA'){
    data$X..població.servida.amb.el.PaP[i] = 0
  }
  else {
    if (as.integer(str_sub(data$X.U.FEFF.Data.d.inici[i], -4, -1)) > data$Any[i]){
      data$X..població.servida.amb.el.PaP[i] = 0
    }
    else if (as.integer(str_sub(data$X.U.FEFF.Data.d.inici[i], -4, -1)) == data$Any[i]){
      data$X..població.servida.amb.el.PaP[i] = NA
    }
  }
}

data <- data[which(!is.na(data$X..població.servida.amb.el.PaP)),]

for (i in 1:nrow(data)) {
  if (!is.na(data$Fraccions.recollides.PaP[i])){
    data$Fraccions.recollides.PaP[i] = as.integer(str_sub(data$Fraccions.recollides.PaP[i], 1, 2))
  }
}

data$Piles <- as.numeric(gsub(",", ".", data$Piles))

#head(data)

```

Decidim utilitzar les dimensions Municipi, Any, Comarca, Població i Percentatge de Població amb Porta

a Porta (PaP per a descriure les files i poder fer les seleccions dels grups, i com a mètriques d'interès ens quedem amb les tones de residus de paper i cartró, vidre, envasos i piles obtingudes anualment, així com els kg de residus per habitant i per any de recollida selectiva, els kg de residus per habitant i per any totals, la proporció de residus de recollida selectiva sobre recollida total, el total en tones de residus restants no recollits selectivament, el model de recollida selectiva i les fraccions (materials) que es recullen amb el servei Porta a Porta(en cas de tenir-lo).

```
interestData <- subset(data, select = c(Municipi, Any, Comarca, Població, Paper.i.cartró, Vidre,
                                       Envasos.lleugers,Piles, Kg.hab.any.recollida.selectiva,
                                       R.S....R.M....total, Suma.Fracció.Resta,
                                       Kg...hab...any, Model.RS.FORM.domiciliaria,
                                       X..població.servida.amb.el.PaP, Fraccions.recollides.PaP))

colnames(interestData) <- c("Municipi", "Any", "Comarca", "Població", "paper_cartro", "vidre",
                           "envasos", "piles", "kg_hab_any_recollida", "rs_rm", "suma_fraccio_resta",
                           "kg_hab_any", "model_rs", "percentatge_poblacio_pap", "fraccions_pap" )

con<-file('./residus_catalunya_net.csv',encoding="UTF-8")
write.csv(interestData,file=con)

data <- read.csv("./residus_catalunya_net.csv", encoding = "UTF-8")
#(data)
```

Per a poder tenir una comparació entre municipis, decidim obtenir les variables quantitatives per càpita en comptes de totals, pel que aquestes les dividim entre el nombre d'habitants.

A més, en aquesta etapa veiem que apareixen entrades “No Territorialitzables”, que decidim no tenir en compte per a l'anàlisi.

Finalment, hem creat una variable nova ‘proporcio’ que representa la proporció de kgs de recollida respecte els kgs de rebuig total per càpita i any, en la qual fem la neteja de les dades corresponent. Aquesta variable tindrà un paper molt important en l'estudi.

```
data$paper_cartro <- data$paper_cartro * 1000/data$Població
data$vidre <- data$vidre * 1000/data$Població
data$envasos <- data$envasos * 1000/data$Població
data$piles <- data$piles * 1000/data$Població
data$suma_fraccio_resta <- data$suma_fraccio_resta * 1000/data$Població

data <- data[which(!is.na(data$Municipi)),]
data <- data[which(!(data$Municipi == "No Territorialitzable")),]

data$proporcio <- data$kg_hab_any_recollida/data$kg_hab_any
data <- data[which(!is.na(data$proporcio)),]

head(data)
```

##	X	Municipi	Any	Comarca	Població	paper_cartro	vidre
## 1	1	Abella De La Conca	2000	Pallars	Jussà	186	33.11828 33.978495
## 2	2	Abella De La Conca	2001	Pallars	Jussà	186	18.27957 26.666667
## 3	3	Abella De La Conca	2002	Pallars	Jussà	179	29.10615 22.458101
## 4	4	Abella De La Conca	2003	Pallars	Jussà	183	36.66667 6.939891
## 5	5	Abella De La Conca	2004	Pallars	Jussà	187	29.57219 25.828877
## 6	6	Abella De La Conca	2005	Pallars	Jussà	183	32.89617 25.191257

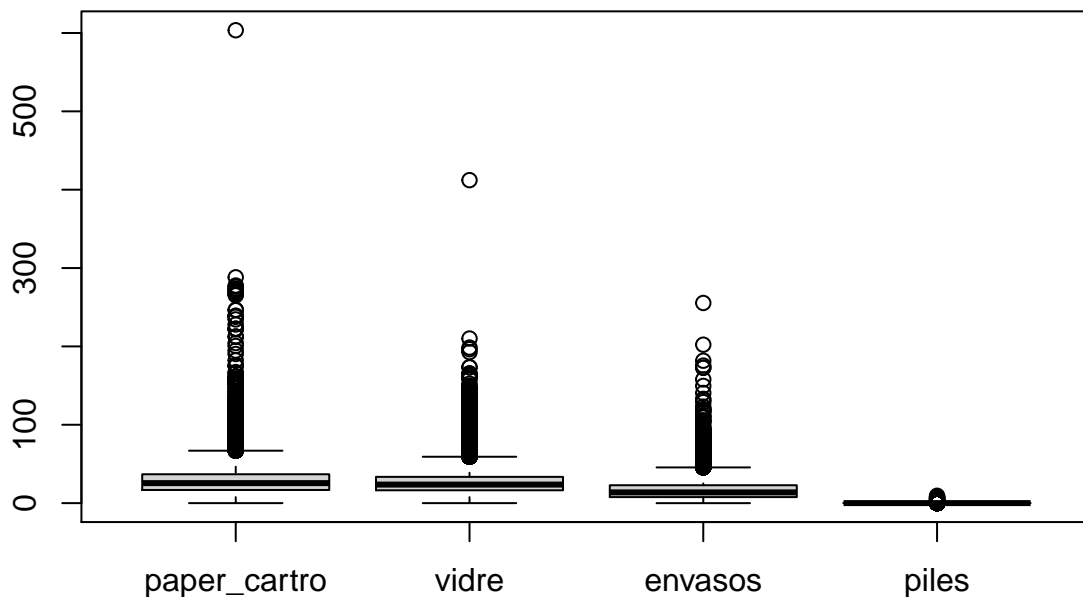
```
##      envasos      piles kg_hab_any_recollida rs_rm suma_fraccio_resta kg_hab_any
## 1 3.225806 0.05376344      79.53 13.65      503.1720      582.69
## 2 3.064516 0.05376344      64.42 15.14      361.1828      425.58
## 3 3.575419 0.05586592      93.48 19.33      390.0559      483.53
## 4 4.972678 0.05464481      77.78 17.58      364.5355      442.34
## 5 4.866310 0.10695187     127.75 26.84      348.2888      476.04
## 6 8.360656 0.10928962     121.49 23.97      385.2459      506.74
##      model_rs percentatge_poblacio_pap fraccions_pap proporcio
## 1      <NA>          0          NA 0.1364877
## 2      <NA>          0          NA 0.1513699
## 3      <NA>          0          NA 0.1933282
## 4      <NA>          0          NA 0.1758376
## 5      <NA>          0          NA 0.2683598
## 6      <NA>          0          NA 0.2397482
```

### 3.2 Identificació i tractament de valors extrems

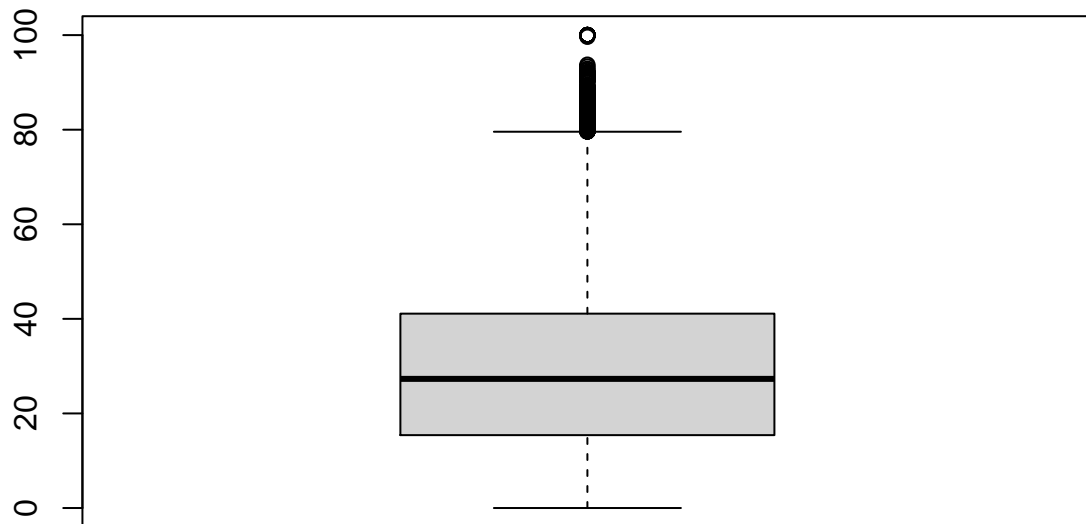
En aquest apartat, primerament analitzarem totes les mètriques d'interès quantitatives mitjançant el gràfic boxplot. D'aquesta manera, podem fer un primer anàlisi visual per identificar valors extrems.

A més a més, mitjançant la funció summary fem un breu anàlisi de les principals característiques descriptives per cada una d'elles.

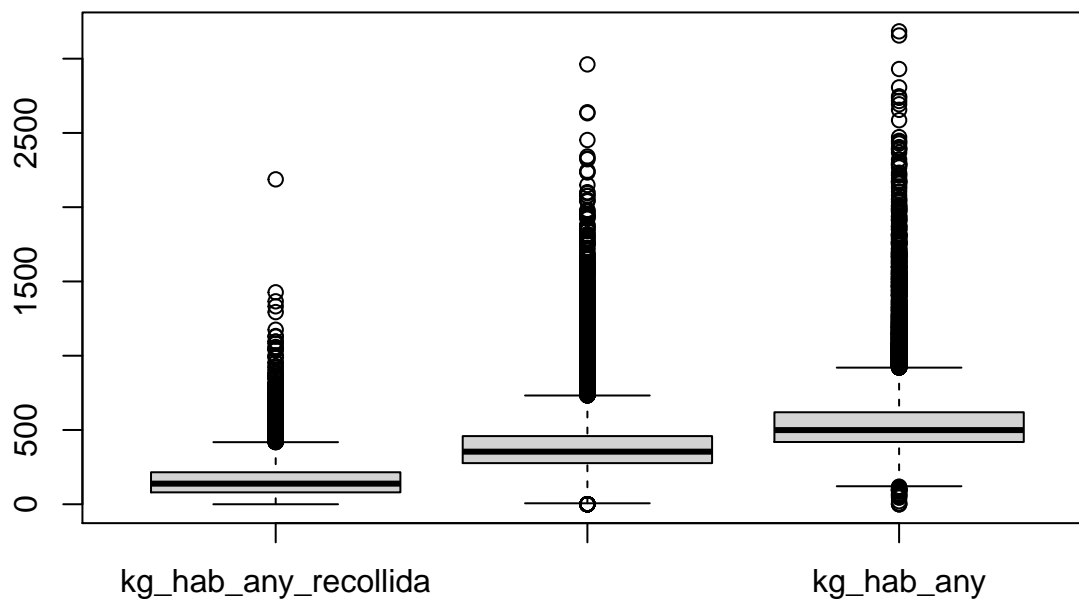
```
boxplot(x = subset(data, select = c("paper_cartro", "vidre", "envasos", "piles")))
```



```
boxplot(x = data$rs_rm)
```



```
boxplot(x = subset(data, select = c("kg_hab_any_recollida", "suma_fraccio_resta", "kg_hab_any")))
```



```
summary(data)
```

```
##           X           Municipi           Any           Comarca
## Min.      :    1   Length:18718   Min.      :2000   Length:18718
## 1st Qu.: 4732   Class :character   1st Qu.:2004   Class :character
## Median : 9466   Mode  :character   Median :2009   Mode  :character
## Mean    : 9476
## 3rd Qu.:14215
## Max.    :18946
##
##      Població      paper_cartro      vidre      envasos
## Min.      :    19   Min.      : 0.00   Min.      : 0.00   Min.      : -0.03808
## 1st Qu.:    311   1st Qu.: 16.76   1st Qu.: 16.38   1st Qu.:  7.68300
## Median :    905   Median : 25.50   Median : 23.64   Median : 13.69229
## Mean     :   7653   Mean    : 29.78   Mean     : 27.69   Mean     : 16.52489
## 3rd Qu.:   3499   3rd Qu.: 36.84   3rd Qu.: 33.50   3rd Qu.: 22.82721
## Max.     :1636762   Max.     :603.49   Max.      :412.20   Max.      :255.42857
##
##      piles      kg_hab_any_recollida      rs_rm      suma_fraccio_resta
## Min.      :0.00000   Min.      : 0.0   Min.      : 0.00   Min.      : 0.0
## 1st Qu.:0.00000   1st Qu.:  80.1   1st Qu.: 15.41   1st Qu.: 276.7
## Median :0.03472   Median : 138.7   Median : 27.31   Median : 354.1
## Mean     :0.06296   Mean     : 161.4   Mean     : 30.12   Mean     : 397.9
## 3rd Qu.:0.08047   3rd Qu.: 215.3   3rd Qu.: 41.10   3rd Qu.: 458.9
## Max.     :9.37143   Max.     :2187.4   Max.      :100.00   Max.      :2961.6
```

```
##
##      kg_hab_any      model_rs      percentatge_poblacio_pap fraccions_pap
## Min.      : 0.01      Length:18718      Min.      : 0.000      Min.      :2.00
## 1st Qu.: 419.03      Class :character      1st Qu.: 0.000      1st Qu.:4.00
## Median : 499.27      Mode  :character      Median : 0.000      Median :4.00
## Mean      : 559.27      Mean      : 7.416      Mean      :4.09
## 3rd Qu.: 619.32      3rd Qu.: 0.000      3rd Qu.:5.00
## Max.      :3183.92      Max.      :100.000      Max.      :5.00
##                                     NA's      :14526
##      proporcio
## Min.      :0.0000
## 1st Qu.:0.1541
## Median :0.2731
## Mean      :0.3012
## 3rd Qu.:0.4110
## Max.      :1.0000
##
```

```
out <- boxplot.stats(data$envasos)$out
negatiu <- min(data$envasos)
if (negatiu %in% out) {
  print("Fora del boxplot")
} else {
  print("Dins el boxplot")
}
```

```
## [1] "Dins el boxplot"
```

```
#Eliminem el negatiu
data <- data[which(data$envasos >= 0), ]

maxs <- data[data$paper_cartro == max(data$paper_cartro),]
maxs <- rbind(maxs, data[data$envasos == max(data$envasos),])
maxs <- rbind(maxs, data[data$vidre == max(data$vidre),])
maxs <- rbind(maxs, data[data$piles == max(data$piles),])
maxs <- rbind(maxs, data[data$kg_hab_any_recollida == max(data$kg_hab_any_recollida),])
maxs
```

```
##           X           Municipi Any      Comarca Població paper_cartro
## 13982 14138   Sant Martí D'albars 2011    Osona      109    603.48624
## 10973 11095  Pobla De Cérvoles, La 2008  Garrigues    245    17.75510
## 14318 14476   Sant Pere Pescador 2009  Alt Empordà    2029    49.67472
## 10558 10678      Pedret I Marzà 2011  Alt Empordà    175    29.94286
## 10557 10677      Pedret I Marzà 2010  Alt Empordà    173    38.84393
##           vidre   envasos      piles kg_hab_any_recollida rs_rm
## 13982  14.77064  29.54128 0.000000000      662.02 66.22
## 10973  30.24490 255.42857 0.000000000      304.13 46.09
## 14318 412.20306  65.76639 0.004928536      527.94 27.51
## 10558  40.51429  19.42857 9.371428571      102.81 19.79
## 10557  23.64162  17.34104 1.676300578      2187.45 69.31
##      suma_fraccio_resta kg_hab_any model_rs percentatge_poblacio_pap
## 13982      337.7064      999.72      <NA>      0
## 10973      355.7551      659.89      <NA>      0
```

```
## 14318      1390.8379    1918.78    <NA>          0
## 10558      416.8000     519.63    <NA>          0
## 10557      968.5549    3156.01    <NA>          0
##      fraccions_pap proporcio
## 13982      NA 0.6622054
## 10973      NA 0.4608798
## 14318      NA 0.2751436
## 10558      NA 0.1978523
## 10557      NA 0.6931062
```

Primer de tot, hem agrupat les variables quantitatives segons les seves unitats de mesura, així no només comprovem outliers sinó que també podem fer estudis comparatius entre elles.

Generalment, totes les variables es comporten de forma semblant, tenen el bigoti superior més allargat que l'inferior i només tenen valors extrems en el costat superior de la caixa, exceptuant la mètrica `kg_hab_any`.

Adicionalment, hem detectat en diferents variables un valor màxim significativament alt comparant amb la resta de registres, així doncs, hem mostrat aquests registres per poder detectar si es tractava d'un error o eren dades legítimes. Finalment, hem conclòs que aquests registres tenien sentit i els contemplarem dins la mostra a estudiar.

Per altra banda, hem vist un valor mínim negatiu de tones d'envasos anuals per càpita, la qual cosa no té sentit, considerem aquest registre com a valor perdut i no el tindrem en compte en l'estudi.

## 4. Anàlisi de les dades

### 4.1 Selecció dels grups de dades que es volen analitzar/comparar (planificació dels anàlisis a aplicar)

Com ja hem comentat en el primer apartat 'Descripció del dataset', un dels objectius d'aquest estudi és avaluar l'eficàcia del mètode porta a porta. Per fer-ho, hem dividit la mostra en dos subgrups: Pobles PaP, utilitzen el mètode de recollida porta a porta en tot el municipi, i Pobles No Pap, no utilitzen cap recollida porta a porta. En aquests subgrups realitzarem un test sobre la mitjana de proporció de recollida selectiva respecte la total.

Per altra banda, dividirem la mostra en dos subgrups diferents: pobles petits i pobles grans. Per aquests subgrups farem un contrast de mitjanes sobre quins tipus de poble generen més residus de recollida selectiva per càpita i any. A més a més, crearem models de regressió lineal multidimensional per explicar la variable '`kg_hab_any_recollida`' i finalment, veurem gràficament quins tipus de residus generen més els pobles petits i els pobles grans de forma comparativa.

```
poblesPetits <- data[which(data$Població < 5000),]
poblesGrans  <- data[which(data$Població >= 5000),]

poblesPaP <- data[which(data$percentatge_poblacio_pap == 100 ),]
poblesNoPaP <- data[which(data$percentatge_poblacio_pap == 0 ),]

c(nrow(poblesPetits),nrow(poblesGrans),nrow(poblesPaP),nrow(poblesNoPaP))
```

```
## [1] 14808 3909 1037 17221
```



## 4.2 Comprovació de la normalitat i homogeneïtat de la variància.

En aquest apartat, hem de comprovar la normalitat i la homogeneïtat de la variància per les dues variables d'interès: 'kg\_hab\_any\_recollida' i 'proporció' per tots els grups que hem generat en l'apartat anterior.

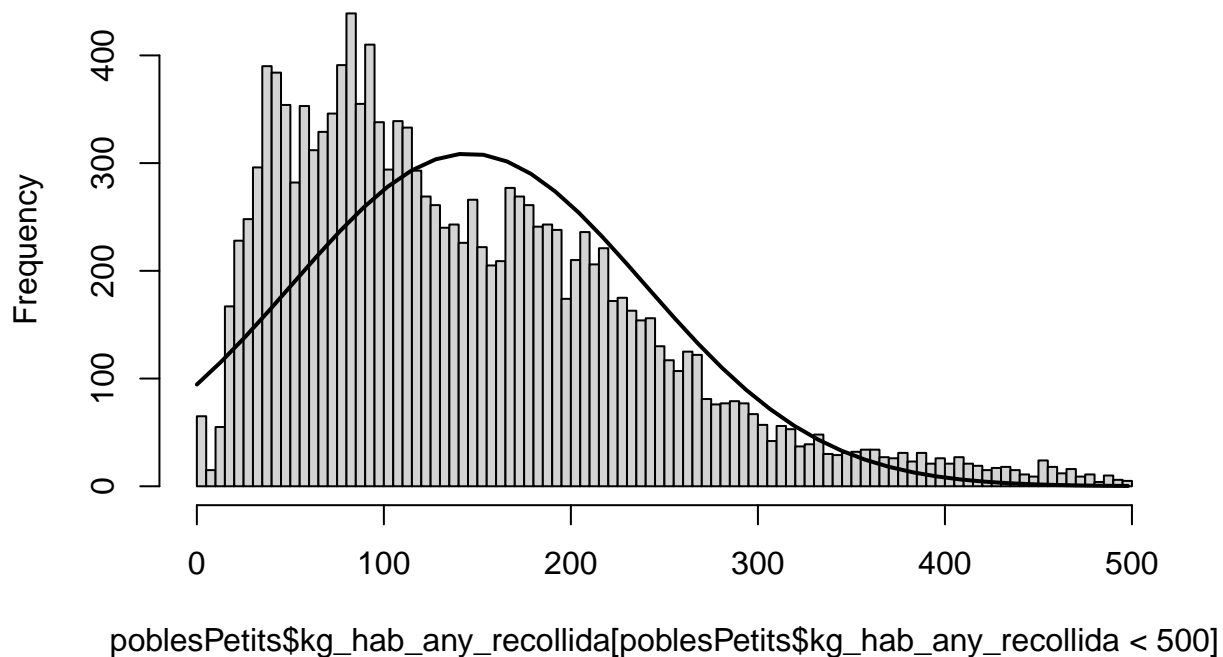
Per tant, hem creat una funció per analitzar la normalitat visualment i seguidament hem aplicat els tests per poder comprovar aquestes propietats numèricament.

```
library(normtest)
library(nortest)

histogramaNormalitat <- function(r, h) {
  xfit <- seq(min(r), max(r), length = 40)
  yfit <- dnorm(xfit, mean = mean(r), sd = sd(r))
  yfit <- yfit * diff(h$mids[1:2]) * length(r)
  return (lines(xfit, yfit, col = "black", lwd = 2))
}

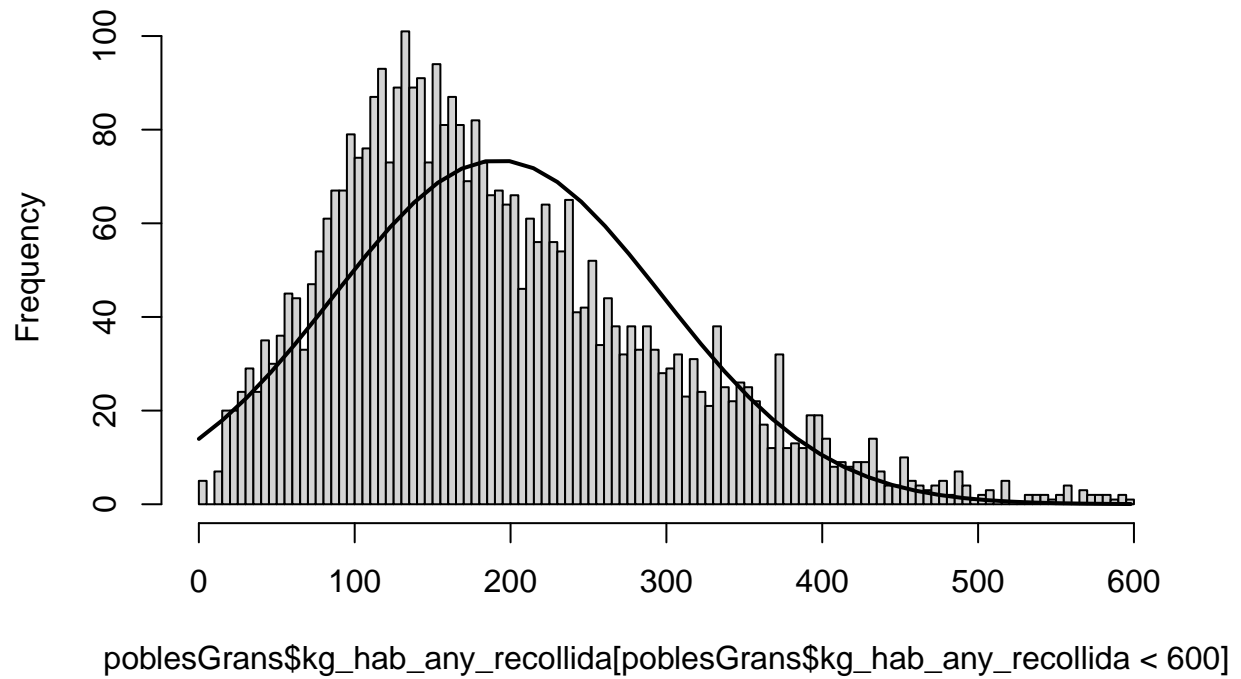
histogramaNormalitat(poblesPetits$kg_hab_any_recollida[poblesPetits$kg_hab_any_recollida < 500],
  hist(poblesPetits$kg_hab_any_recollida[poblesPetits$kg_hab_any_recollida < 500],
    breaks = 100, main = "Histograma de kg_hab_any_recollida en pobles petits"))
```

**Histograma de kg\_hab\_any\_recollida en pobles petits**



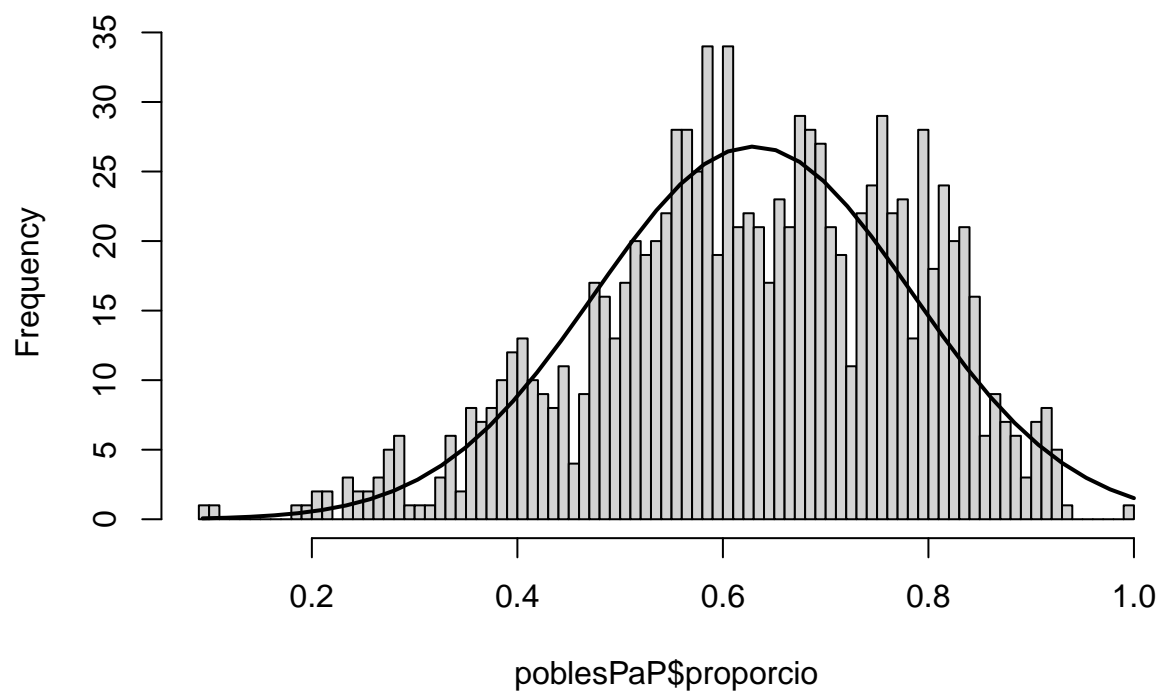
```
histogramaNormalitat(poblesGrans$kg_hab_any_recollida[poblesGrans$kg_hab_any_recollida < 600],
  hist(poblesGrans$kg_hab_any_recollida[poblesGrans$kg_hab_any_recollida < 600],
    breaks = 100, main = "Histograma de kg_hab_any_recollida en pobles grans"))
```

## Histograma de kg\_hab\_any\_recollida en pobles grans



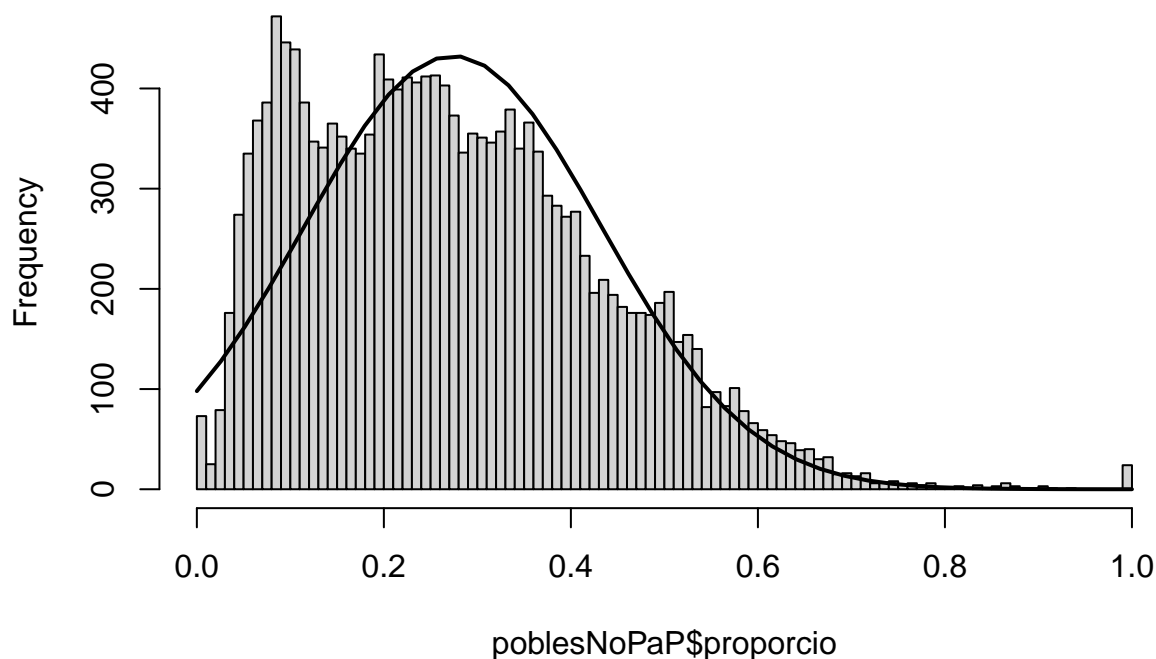
```
histogramaNormalitat(poblesPaP$proporcio,  
                      hist(poblesPaP$proporcio,  
                           breaks = 100, main = "Histograma de kg_hab_any_recollida en pobles PaP"))
```

## Histograma de kg\_hab\_any\_recollida en pobles PaP



```
histogramaNormalitat(poblesNoPaP$proporcio,  
                      hist(poblesNoPaP$proporcio,  
                           breaks = 100, main = "Histograma de kg_hab_any_recollida en pobles No Pap"))
```

## Histograma de kg\_hab\_any\_recollida en pobles No Pap



```
lillie.test(poblesPetits$kg_hab_any_recollida)
```

```
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  poblesPetits$kg_hab_any_recollida  
## D = 0.1029, p-value < 2.2e-16
```

```
lillie.test(poblesGrans$kg_hab_any_recollida)
```

```
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  poblesGrans$kg_hab_any_recollida  
## D = 0.089381, p-value < 2.2e-16
```

```
lillie.test(poblesPaP$proporcio)
```

```
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  poblesPaP$proporcio  
## D = 0.047997, p-value = 7.115e-06
```

```
lillie.test(poblesNoPaP$proporcio)
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: poblesNoPaP$proporcio  
## D = 0.052343, p-value < 2.2e-16
```

Tot i que a primera vista algun histograma podia semblar que seguia una distribució normal, ho refutem mitjançant el test de Lilliefors ja que el p-value és inferior a 0.05 per tant rebutgem la hipòtesi nul·la de normalitat.

No obstant, en conseqüència de la grandària de la mostra, podem aplicar el Teorema del Límit Central i suposa normalitat.

```
# Test de Homoscedasticitat de la variància
```

```
var.test(poblesPetits$kg_hab_any_recollida, poblesGrans$kg_hab_any_recollida)
```

```
##  
## F test to compare two variances  
##  
## data: poblesPetits$kg_hab_any_recollida and poblesGrans$kg_hab_any_recollida  
## F = 0.90452, num df = 14807, denom df = 3908, p-value = 6.639e-05  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.8602132 0.9504044  
## sample estimates:  
## ratio of variances  
## 0.9045166
```

```
var.test(poblesPaP$proporcio, poblesNoPaP$proporcio)
```

```
##  
## F test to compare two variances  
##  
## data: poblesPaP$proporcio and poblesNoPaP$proporcio  
## F = 0.94466, num df = 1036, denom df = 17220, p-value = 0.2171  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.8660031 1.0341206  
## sample estimates:  
## ratio of variances  
## 0.9446649
```

Seguidament, hem aplicat el test de homoscedasticitat de la variància i hem arribat a diferents conclusions:

Primer de tot, en la divisió de la mostra en pobles petits i pobles grans, el test per la variable 'kg\_hab\_any\_recollida' conclouem que hi ha heteroscedasticitat ja que el p-value és inferior a 0.05 i per tant rebutgem la hipòtesi nul·la d'igualtat de variàncies.

En canvi, en la divisió de la mostra en pobles porta a porta i pobles no porta a porta, el test per la variable 'proporcio' conclouem que hi ha homoscedasticitat ja que el p-value és superior a 0.05 i no podem rebutjar la hipòtesi nul·la.

Aquests resultats son necessaris per poder aplicar els tests estadístics correctament.

### 4.3 Aplicació de proves estadístiques per comparar els grups de dades.

En aquest apartat, primerament volem contrastar les nostres suposicions inicials.

Per un banda, pensem que els pobles amb la gestió de recollida selectiva porta a porta és més eficaç que la recollida selectiva de vorera tradicional.

$$H_0 : \mu_{\text{PaP}} = \mu_{\text{NoPaP}}$$

$$H_1 : \mu_{\text{PaP}} > \mu_{\text{NoPaP}}$$

Per realitzar aquest contrast, aplicarem el test de la t-Student a la variable ‘proporcio’ comentada anteriorment ja que la variància és desconeguda, com hem definit les hipòtesis l’alternative és “greater” i finalment, com podem veure en l’apartat anterior hi ha homoscedasticitat.

```
t.test(x=poblesPaP$proporcio, y=poblesNoPaP$proporcio, alternative= "greater",
      var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: poblesPaP$proporcio and poblesNoPaP$proporcio
## t = 70.218, df = 18256, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.3477822      Inf
## sample estimates:
## mean of x mean of y
## 0.6298926 0.2737677
```

Com que p-value és inferior a 0.05, hem de rebutjar la hipòtesi nul·la i per tant, la nostra suposició era correcta perquè acabem de veure que la mitjana de proporció de recollida selectiva respecte el total en els pobles amb mètode porta a porta és superior que en els pobles amb el mètode tradicional.

Per altra banda, com ja hem comentat en l’apartat 4.1, trobem interessant la comparativa de ‘kg\_hab\_any\_recollida’ entre pobles grans i pobles petits. Així doncs, definim aquest contrast d’hipòtesis on creiem que es genera més residus de recollida selectiva per càpita i any en pobles grans que en els pobles petits:

$$H_0 : \mu_{\text{grans}} = \mu_{\text{petits}}$$

$$H_1 : \mu_{\text{grans}} > \mu_{\text{petits}}$$

Per poder realitzar aquest contrast, aplicarem el test t-Student ja que la variància és desconeguda, l’alternative serà “greater” per definició del contrast d’hipòtesis i finalment, com hem pogut veure en l’apartat anterior, hi ha heteroscedasticitat.

```
t.test(x=poblesGrans$kg_hab_any_recollida, y=poblesPetits$kg_hab_any_recollida, alternative = "greater",
      var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: poblesGrans$kg_hab_any_recollida and poblesPetits$kg_hab_any_recollida
```

```
## t = 21.168, df = 5908.2, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 41.16253      Inf
## sample estimates:
## mean of x mean of y
## 196.6960 152.0648
```

Com que p-value és inferior a 0.05, hem de rebutjar la hipòtesi nul·la i per tant, la nostra suposició era correcta perquè acabem de veure que la mitjana de residus de recollida per habitant i any és superior en pobles grans que en pobles petits.

## FALTA EXPLICAR TOTA LA PART DE LES REGRESSIONS LINEALS MULTIDIMENSIONALS

```
data$Any <- as.factor(data$Any)

lineal <- lm(kg_hab_any_recollida ~ Any + Població + Comarca + percentatge_poblacio_pap, data = data)
lineal.summary <- summary(lineal)
lineal.summary
```

```
##
## Call:
## lm(formula = kg_hab_any_recollida ~ Any + Població + Comarca +
##     percentatge_poblacio_pap, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -244.23  -50.06  -13.08   29.40  2002.82
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.990e+01  5.203e+00   5.746 9.29e-09 ***
## Any2001         7.286e+00  4.155e+00   1.753 0.079550 .
## Any2002         1.463e+01  4.165e+00   3.511 0.000447 ***
## Any2003         2.224e+01  4.158e+00   5.349 8.92e-08 ***
## Any2004         4.158e+01  4.168e+00   9.976 < 2e-16 ***
## Any2005         5.515e+01  4.176e+00  13.207 < 2e-16 ***
## Any2006         7.635e+01  4.172e+00  18.299 < 2e-16 ***
## Any2007         8.774e+01  4.159e+00  21.098 < 2e-16 ***
## Any2008         9.808e+01  4.168e+00  23.532 < 2e-16 ***
## Any2009        1.102e+02  4.180e+00  26.356 < 2e-16 ***
## Any2010        1.232e+02  4.171e+00  29.530 < 2e-16 ***
## Any2011        1.175e+02  4.166e+00  28.205 < 2e-16 ***
## Any2012        1.174e+02  4.167e+00  28.180 < 2e-16 ***
## Any2013        1.124e+02  4.169e+00  26.956 < 2e-16 ***
## Any2014        1.188e+02  4.173e+00  28.480 < 2e-16 ***
## Any2015        1.276e+02  4.183e+00  30.491 < 2e-16 ***
## Any2016        1.312e+02  4.174e+00  31.421 < 2e-16 ***
## Any2017        1.441e+02  4.181e+00  34.456 < 2e-16 ***
## Any2018        1.638e+02  4.250e+00  38.538 < 2e-16 ***
## Any2019        1.773e+02  4.204e+00  42.168 < 2e-16 ***
## Població         5.387e-06  1.478e-05   0.364 0.715529
## ComarcaAlt Empordà 3.156e+01  5.014e+00   6.295 3.14e-10 ***
## ComarcaAlt Penedès 2.557e+01  5.842e+00   4.377 1.21e-05 ***
```

```

## ComarcaAlt Urgell      4.536e+01  6.381e+00   7.109 1.21e-12 ***
## ComarcaAlta Ribagorça  3.297e+01  1.245e+01   2.648 0.008113 **
## ComarcaAnoia          -2.096e+01  5.633e+00  -3.721 0.000199 ***
## ComarcaBages          3.625e+01  5.587e+00   6.488 8.91e-11 ***
## ComarcaBaix Camp      7.850e+01  5.807e+00  13.519 < 2e-16 ***
## ComarcaBaix Ebre      5.020e+01  6.956e+00   7.217 5.53e-13 ***
## ComarcaBaix Empordà   9.003e+01  5.536e+00  16.264 < 2e-16 ***
## ComarcaBaix Llobregat 4.411e+01  5.728e+00   7.700 1.42e-14 ***
## ComarcaBaix Penedès   1.475e+01  6.948e+00   2.122 0.033844 *
## ComarcaBarcelonès    -1.756e+01  1.197e+01  -1.467 0.142358
## ComarcaBerguedà      2.680e+01  5.706e+00   4.697 2.66e-06 ***
## ComarcaCerdanya      -2.692e+00  6.576e+00  -0.409 0.682234
## ComarcaConca de Barberà 7.518e+01  6.153e+00  12.219 < 2e-16 ***
## ComarcaGarraf        1.031e+02  9.342e+00  11.035 < 2e-16 ***
## ComarcaGarrigues     -5.080e+01  6.021e+00  -8.436 < 2e-16 ***
## ComarcaGarrotxa      3.787e+01  6.222e+00   6.086 1.18e-09 ***
## ComarcaGironès       4.605e+01  5.847e+00   7.876 3.55e-15 ***
## ComarcaMaresme       1.095e+02  5.698e+00  19.212 < 2e-16 ***
## ComarcaMoianès       7.924e+01  1.389e+01   5.704 1.19e-08 ***
## ComarcaMontsià       6.408e+01  7.296e+00   8.783 < 2e-16 ***
## ComarcaNoguera       -3.408e+01  5.734e+00  -5.944 2.83e-09 ***
## ComarcaOsona         6.711e+01  5.164e+00  12.998 < 2e-16 ***
## ComarcaPallars Jussà  1.244e+01  6.956e+00   1.788 0.073718 .
## ComarcaPallars Sobirà 7.334e+01  6.845e+00  10.714 < 2e-16 ***
## ComarcaPla d'Urgell   1.534e+01  6.689e+00   2.294 0.021804 *
## ComarcaPla de l'Estany 3.110e+01  7.525e+00   4.133 3.59e-05 ***
## ComarcaPriorat       -1.682e+01  6.060e+00  -2.776 0.005510 **
## ComarcaRibera d'Ebre  3.130e+01  6.938e+00   4.511 6.48e-06 ***
## ComarcaRipollès      5.002e+01  6.370e+00   7.853 4.29e-15 ***
## ComarcaSegarra       3.523e+01  6.183e+00   5.697 1.24e-08 ***
## ComarcaSegrià        -4.391e+01  5.483e+00  -8.008 1.23e-15 ***
## ComarcaSelva         4.435e+01  5.917e+00   7.496 6.87e-14 ***
## ComarcaSolsonès      -1.490e+01  6.817e+00  -2.186 0.028840 *
## ComarcaTarragonès    5.670e+01  6.201e+00   9.144 < 2e-16 ***
## ComarcaTerra Alta    -3.887e+01  7.387e+00  -5.262 1.44e-07 ***
## ComarcaUrgell        -9.493e-01  6.284e+00  -0.151 0.879930
## ComarcaVal d'Aran     8.034e+01  8.034e+00   9.999 < 2e-16 ***
## ComarcaVallès Occidental 5.075e+01  6.087e+00   8.337 < 2e-16 ***
## ComarcaVallès Oriental 6.823e+01  5.342e+00  12.771 < 2e-16 ***
## percentatge_poblacio_pap 9.087e-01  2.922e-02  31.098 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 90.28 on 18654 degrees of freedom
## Multiple R-squared:  0.3891, Adjusted R-squared:  0.3871
## F-statistic: 191.6 on 62 and 18654 DF,  p-value: < 2.2e-16

```

```

lineal_petits <- lm(kg_hab_any_recollida ~ Any + Població + Comarca + percentatge_poblacio_pap,
                    data = poblesPetits)

lineal_petits.summary <- summary(lineal_petits)
lineal_petits.summary

```

```
##
```



```
## Call:
## lm(formula = kg_hab_any_recollida ~ Any + Població + Comarca +
##     percentatge_poblacio_pap, data = poblesPetits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -215.46  -47.48  -13.31   27.01  2047.52
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.803e+04  2.569e+02  -70.192  < 2e-16 ***
## Any              9.033e+00  1.280e-01   70.590  < 2e-16 ***
## Població        1.867e-03  7.423e-04    2.514  0.011932 *
## ComarcaAlt Empordà  1.504e+01  5.060e+00    2.973  0.002954 **
## ComarcaAlt Penedès  1.623e+01  6.036e+00    2.689  0.007166 **
## ComarcaAlt Urgell   4.210e+01  6.395e+00    6.583  4.75e-11 ***
## ComarcaAlta Ribagorça 3.269e+01  1.211e+01    2.699  0.006970 **
## ComarcaAnoia       -2.189e+01  5.827e+00   -3.756  0.000173 ***
## ComarcaBages       2.420e+01  6.063e+00    3.991  6.62e-05 ***
## ComarcaBaix Camp    6.153e+01  6.037e+00   10.192  < 2e-16 ***
## ComarcaBaix Ebre    4.089e+01  7.632e+00    5.358  8.56e-08 ***
## ComarcaBaix Empordà 8.319e+01  5.761e+00   14.442  < 2e-16 ***
## ComarcaBaix Llobregat 4.593e+01  9.457e+00    4.857  1.21e-06 ***
## ComarcaBaix Penedès 7.894e+00  7.557e+00    1.045  0.296208
## ComarcaBerguedà     2.896e+01  5.717e+00    5.065  4.14e-07 ***
## ComarcaCerdanya    -3.282e+00  6.597e+00   -0.498  0.618830
## ComarcaConca de Barberà 7.591e+01  6.156e+00   12.332  < 2e-16 ***
## ComarcaGarraf       8.253e+01  1.431e+01    5.766  8.26e-09 ***
## ComarcaGarrigues   -5.033e+01  6.025e+00   -8.354  < 2e-16 ***
## ComarcaGarrotxa    3.747e+01  6.240e+00    6.005  1.96e-09 ***
## ComarcaGironès     4.753e+01  6.081e+00    7.815  5.85e-15 ***
## ComarcaMaresme     1.404e+02  8.480e+00   16.562  < 2e-16 ***
## ComarcaMoianès     6.334e+01  1.415e+01    4.477  7.62e-06 ***
## ComarcaMontsià     6.080e+01  8.637e+00    7.040  2.00e-12 ***
## ComarcaNoguera     -3.320e+01  5.738e+00   -5.786  7.34e-09 ***
## ComarcaOsona       6.685e+01  5.266e+00   12.696  < 2e-16 ***
## ComarcaPallars Jussà 1.065e+01  7.000e+00    1.521  0.128278
## ComarcaPallars Sobirà 7.705e+01  6.720e+00   11.467  < 2e-16 ***
## ComarcaPla d'Urgell 1.600e+01  6.747e+00    2.372  0.017728 *
## ComarcaPla de l'Estany 3.204e+01  7.634e+00    4.197  2.72e-05 ***
## ComarcaPriorat     -1.137e+01  5.963e+00   -1.906  0.056664 .
## ComarcaRibera d'Ebre 2.798e+01  6.933e+00    4.036  5.47e-05 ***
## ComarcaRipollès    4.955e+01  6.379e+00    7.767  8.55e-15 ***
## ComarcaSegarra     4.213e+01  6.251e+00    6.740  1.65e-11 ***
## ComarcaSegrià     -4.252e+01  5.575e+00   -7.627  2.56e-14 ***
## ComarcaSelva       3.624e+01  6.629e+00    5.466  4.67e-08 ***
## ComarcaSolsonès    -1.591e+01  6.851e+00   -2.322  0.020233 *
## ComarcaTarragonès  3.984e+01  6.790e+00    5.868  4.50e-09 ***
## ComarcaTerra Alta  -3.311e+01  7.226e+00   -4.582  4.65e-06 ***
## ComarcaUrgell     -1.576e+00  6.383e+00   -0.247  0.804960
## ComarcaVal d'Aran   6.791e+01  8.113e+00    8.370  < 2e-16 ***
## ComarcaVallès Occidental 6.686e+01  1.023e+01    6.536  6.52e-11 ***
## ComarcaVallès Oriental 9.947e+01  6.268e+00   15.870  < 2e-16 ***
## percentatge_poblacio_pap 8.003e-01  3.295e-02  24.291  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 87.18 on 14764 degrees of freedom
## Multiple R-squared:  0.4027, Adjusted R-squared:  0.401
## F-statistic: 231.5 on 43 and 14764 DF,  p-value: < 2.2e-16

lineal_grans <- lm(kg_hab_any_recollida ~ Any + Població + Comarca + percentatge_poblacio_pap,
                  data = poblesGrans)

lineal_grans <- summary(lineal_grans)
lineal_grans
```

```
##
## Call:
## lm(formula = kg_hab_any_recollida ~ Any + Població + Comarca +
##     percentatge_poblacio_pap, data = poblesGrans)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-320.38	-51.51	-9.44	36.84	675.43

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1.285e+04	5.342e+02	-24.058	< 2e-16	***
Any	6.490e+00	2.655e-01	24.441	< 2e-16	***
Població	-9.094e-06	1.527e-05	-0.595	0.551603	
ComarcaAlt Empordà	1.946e+02	1.899e+01	10.248	< 2e-16	***
ComarcaAlt Penedès	8.883e+00	1.979e+01	0.449	0.653632	
ComarcaAlt Urgell	6.129e+01	2.644e+01	2.318	0.020484	*
ComarcaAnoia	-7.954e+01	1.864e+01	-4.267	2.02e-05	***
ComarcaBages	-8.848e+00	1.773e+01	-0.499	0.617818	
ComarcaBaix Camp	8.897e+01	1.878e+01	4.737	2.24e-06	***
ComarcaBaix Ebre	5.890e+00	1.957e+01	0.301	0.763444	
ComarcaBaix Empordà	5.331e+01	1.820e+01	2.930	0.003414	**
ComarcaBaix Llobregat	-2.326e+01	1.708e+01	-1.362	0.173373	
ComarcaBaix Penedès	-3.949e+01	1.991e+01	-1.984	0.047365	*
ComarcaBarcelonès	-7.724e+01	2.004e+01	-3.854	0.000118	***
ComarcaBerguedà	-6.625e+01	2.452e+01	-2.702	0.006929	**
ComarcaCerdanya	-4.066e+01	2.644e+01	-1.538	0.124126	
ComarcaConca de Barberà	1.247e+01	2.686e+01	0.464	0.642554	
ComarcaGarraf	4.696e+01	1.957e+01	2.399	0.016470	*
ComarcaGarrigues	-1.035e+02	2.644e+01	-3.916	9.16e-05	***
ComarcaGarrotxa	-1.513e+01	2.644e+01	-0.572	0.567135	
ComarcaGironès	-3.631e+01	1.926e+01	-1.885	0.059534	.
ComarcaMaresme	2.496e+01	1.716e+01	1.454	0.145939	
ComarcaMoianès	9.436e+01	4.444e+01	2.123	0.033803	*
ComarcaMontsià	1.722e+00	1.895e+01	0.091	0.927588	
ComarcaNoguera	-1.045e+02	2.644e+01	-3.952	7.87e-05	***
ComarcaOsona	-2.290e+00	1.851e+01	-0.124	0.901558	
ComarcaPallars Jussà	-1.581e+01	2.644e+01	-0.598	0.549942	
ComarcaPla d'Urgell	-6.820e+01	2.644e+01	-2.580	0.009923	**
ComarcaPla de l'Estany	-4.222e+01	2.644e+01	-1.597	0.110366	
ComarcaRibera d'Ebre	8.145e+01	2.899e+01	2.810	0.004982	**

```
## ComarcaRipollès      1.066e+01  2.644e+01   0.403 0.686767
## ComarcaSegarra      -1.250e+02  2.382e+01  -5.248 1.62e-07 ***
## ComarcaSegrià       -1.182e+02  1.993e+01  -5.930 3.29e-09 ***
## ComarcaSelva        -8.154e+00  1.787e+01  -0.456 0.648273
## ComarcaSolsonès     -4.025e+01  2.644e+01  -1.523 0.127959
## ComarcaTarragonès   2.955e+01  1.851e+01   1.597 0.110444
## ComarcaUrgell       -4.811e+01  2.182e+01  -2.205 0.027531 *
## ComarcaVal d'Aran    1.680e+02  2.899e+01   5.795 7.36e-09 ***
## ComarcaVallès Occidental -2.203e+01  1.726e+01  -1.277 0.201745
## ComarcaVallès Oriental -2.755e+01  1.713e+01  -1.608 0.107936
## percentatge_poblacio_pap 1.419e+00  6.401e-02  22.161 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.16 on 3868 degrees of freedom
## Multiple R-squared:  0.4007, Adjusted R-squared:  0.3945
## F-statistic: 64.66 on 40 and 3868 DF,  p-value: < 2.2e-16
```

```
#table(data$percentatge_poblacio_pap)
#table(poblesPetits$percentatge_poblacio_pap)
#table(poblesGrans$percentatge_poblacio_pap)
```

En els tests de comparació de mitjanes, veiem si la diferència és significativa. En canvi, al fer la regressió lineal fem un model que intenti explicar els valors de kg\_hab\_any\_recollida per a cada conjunt, però aquest pot ser diferent i diverses variables poden afectar més o menys en cada grup.

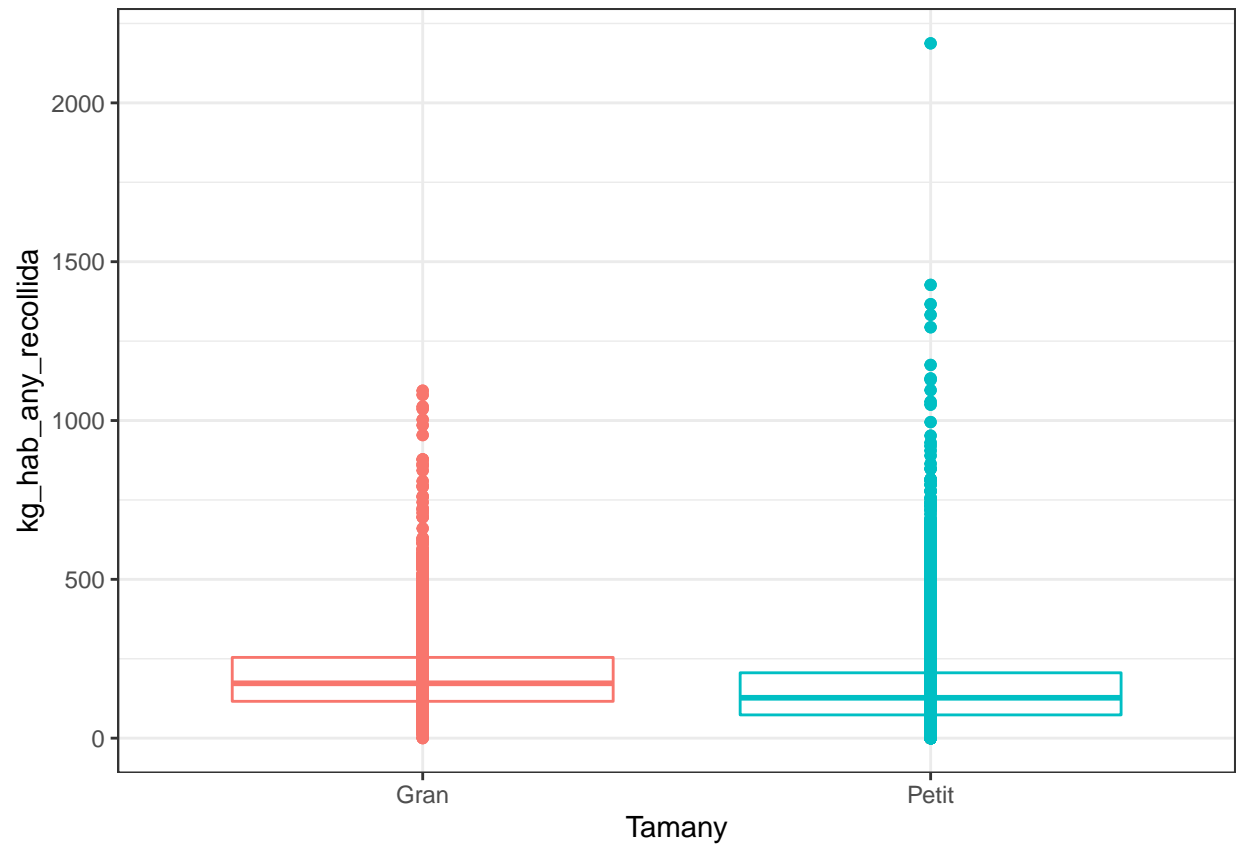
## 5. Representació dels resultats a partir de taules i gràfiques

```
library(ggplot2)

dataPlotTamany <- rbind(poblesPetits, poblesGrans)

tamany <- (dataPlotTamany$Població >= 5000)
dataPlotTamany$Tamany <- as.factor(ifelse(tamany==TRUE, "Gran", "Petit"))

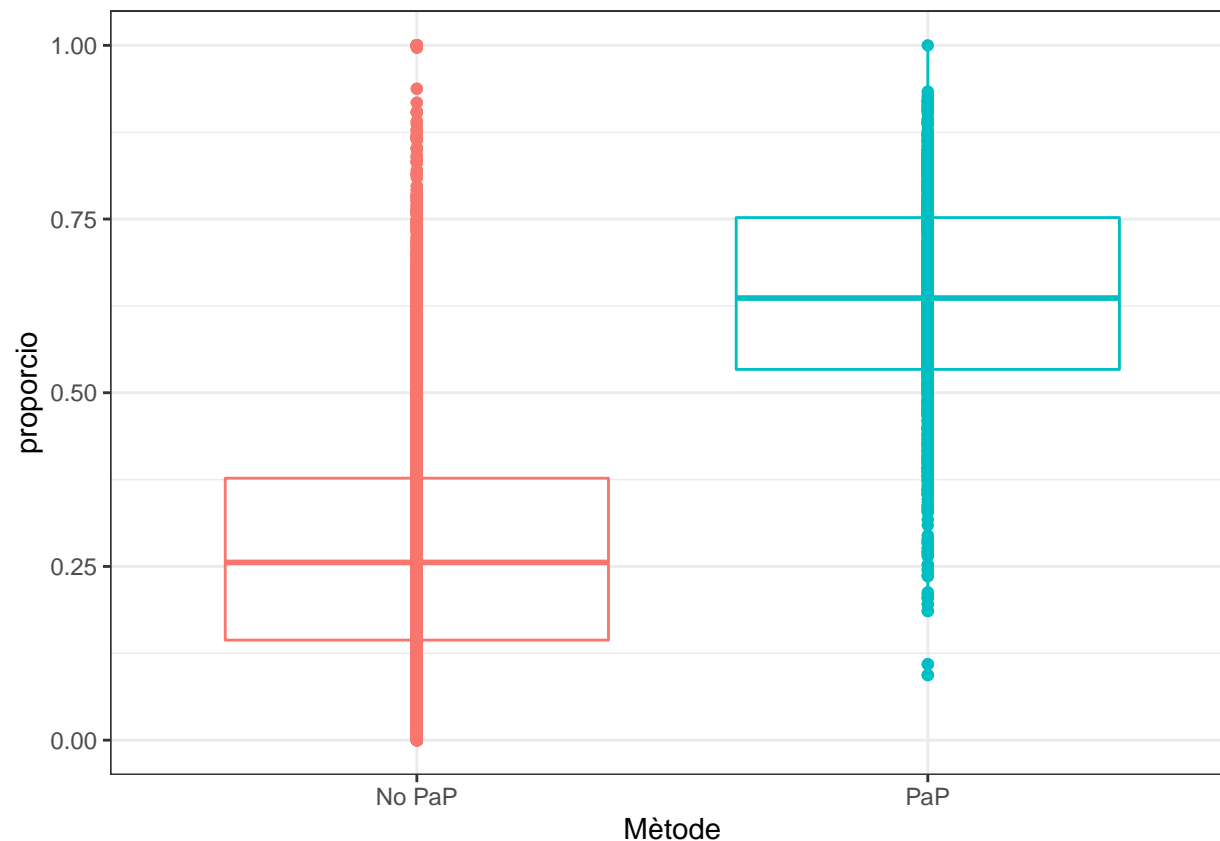
ggplot(data = dataPlotTamany, aes(x = Tamany, y = kg_hab_any_recollida, colour = Tamany)) +
  geom_boxplot() +
  geom_point() +
  theme_bw() +
  theme(legend.position = "none")
```



```
dataPlot <- rbind(poblesPaP, poblesNoPaP)

Mètode <- (dataPlot$percentatge_poblacio_pap == 100)
dataPlot$Mètode <- as.factor(ifelse(Mètode==TRUE, "PaP", "No PaP"))

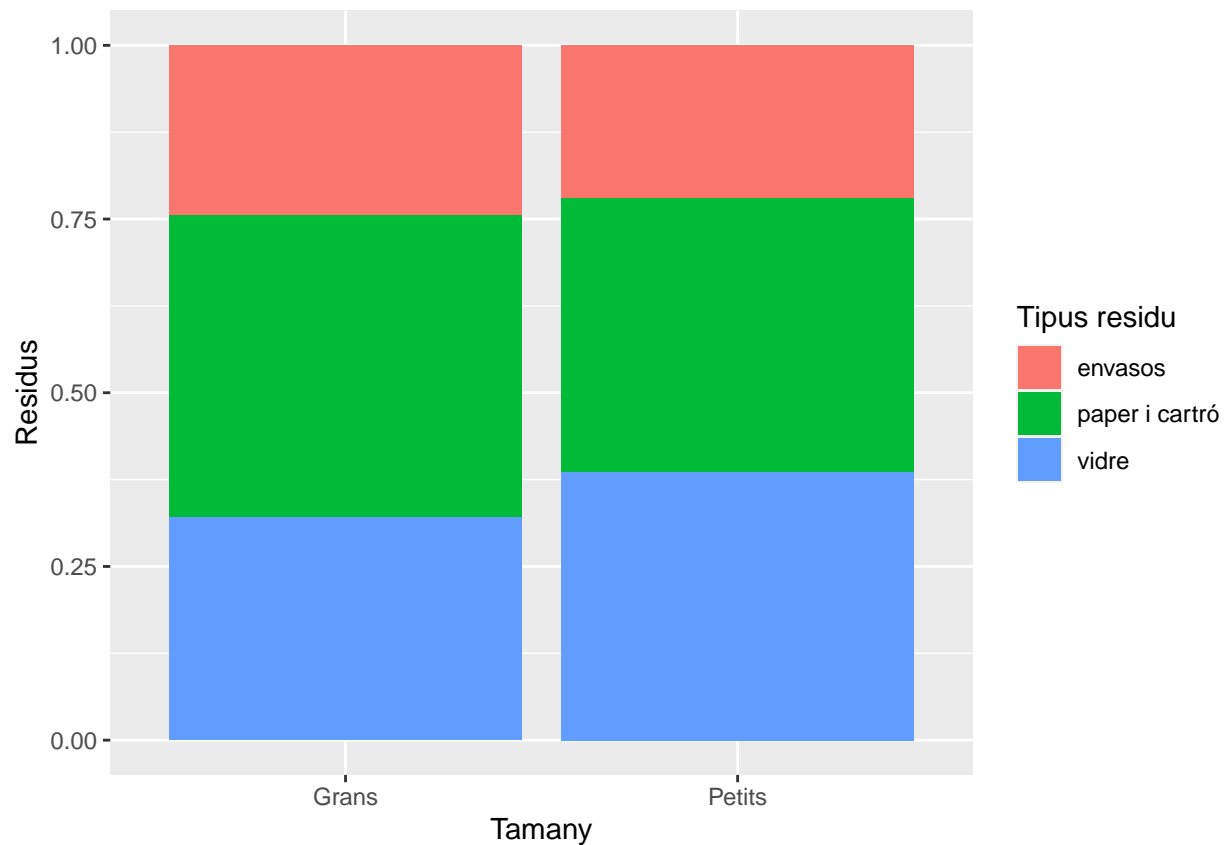
ggplot(data = dataPlot, aes(x = Mètode, y = proporcio, colour = Mètode)) +
  geom_boxplot() +
  geom_point() +
  theme_bw() +
  theme(legend.position = "none")
```



```
# create a dataset
specie <- c(rep("Petits" , 3) , rep("Grans" , 3) )
condition <- rep(c("vidre" , "paper i cartró" , "envasos") , 2)
value <- c(mean(poblesPetits$vidre), mean(poblesPetits$paper_cartro),
           mean(poblesPetits$envasos),
           mean(poblesGrans$vidre), mean(poblesGrans$paper_cartro),
           mean(poblesGrans$envasos))

dataToPlot <- data.frame(specie,condition,value)

# Stacked + percent
ggplot(dataToPlot, aes(fill=condition, y=value, x=specie)) +
  geom_bar(position="fill", stat="identity") +
  xlab("Tamany") +
  ylab("Residus") +
  labs(fill = "Tipus residu")
```



```
print("Taula de percentages (vidre, paper i cartró, envasos):")
```

```
## [1] "Taula de percentages (vidre, paper i cartró, envasos):"
```

```
print("Pobles petits:")
```

```
## [1] "Pobles petits:"
```

```
prop.table(c(mean(poblesPetits$vidre), mean(poblesPetits$paper_cartro),
              mean(poblesPetits$envasos)))
```

```
## [1] 0.3857475 0.3953890 0.2188636
```

```
print("Pobles grans:")
```

```
## [1] "Pobles grans:"
```

```
prop.table(c(mean(poblesGrans$vidre), mean(poblesGrans$paper_cartro),
              mean(poblesGrans$envasos)))
```

```
## [1] 0.3218574 0.4345222 0.2436203
```

## 6. Resolució del problema

## 7. Contribucions

En la següent taula apareixen les firmes dels participants conforme han contribuït en tots els àmbits del projecte.

<b>Contribucions</b>	<b>Firma</b>
Investigació prèvia	GB, BA
Redacció de les respostes	GB, BA
Desenvolupament codi	GB, BA