L’àmbit municipal i el vot a Vox

Ruben Pueyo i David Donaire

2023-03-08

logit <- function(p,eps = 0.001){  
 p[p==0] <- eps  
 p[p==1] <- 1-eps  
 return(log(p/(1-p)))  
}

###Funcions per anàlisi bivariat/propensions

errStd\_bin <- function(x) { sqrt(mean(x)\*(1-mean(x)) / length(x))}  
minInt\_bin <- function(x) { mean(x) - 1.96 \*errStd\_bin(x) }  
maxInt\_bin <- function(x) { mean(x) + 1.96 \*errStd\_bin(x) }

# 1. Introducció

Els discursos de la formació d’extrema dreta Vox se centren, en la lluita contra l’avortament, l’ecologisme, la immigració, els moviments feministes i els moviments LGTBIQ+. En el present treball veurem si hi ha una correlació entre els municipis amb una major taxa d’immigració o avortament, entre altres variables independents, i el vot a la formació de Santiago Abascal. És a dir, sense entrar en si suposen pors racionals o no, si hi ha una correlació en una major presència d’aquests fenòmens als quals Vox mostra un clar rebuig i una major suport en les localitats més afectades per aquests suposats problemes.

La metodologia utilitzada per a analitzar la nostra pregunta de recerca és completament quantitativa. Durem a terme una regressió múltiple per a identificar les correlacions entre el suport municipal a la formació verda i les diferents variables independents que puguin explicar parcialment la diferència de suport entre municipis.

Per poder realitzar el nostre estudi, s’ha realitzar una recopilació de dades mitjançant el padró municipal, la estadística municipal de l’atur de l’INE, un número d’indicadors econòmics del municipi i els resultats electorals municipals per al Congrés dels diputats. L’atur s’ha pres a data de novembre de 2019, donat que aquest any es van a dur a terme dues eleccions en pocs mesos, i la situació personal d’atur a aquelles dates podria haver afectat el vot.

data <- read.csv("C:/Users/rpuey/Documents/datosfinal.txt")

data <- rename(data, pob\_t = totalpoblacion, mun = municipio)  
names(data)

## [1] "mun" "Votosenblanco"   
## [3] "Votosnulos" "PSOE"   
## [5] "PP" "Cs"   
## [7] "ERC" "VOX"   
## [9] "JUNTS" "PNV"   
## [11] "PODEMOS\_IU" "EH\_Bildu"   
## [13] "BNG" "CUP"   
## [15] "atur" "aturhomesmenors24"   
## [17] "aturhomes25\_44" "aturhomesmes45"   
## [19] "aturdonesmenors24" "aturdones25\_44"   
## [21] "aturdonesmajors45" "aturagricultura"   
## [23] "aturindustria" "aturconstruccio"   
## [25] "aturserveis" "americana"   
## [27] "asiatica" "europeossinespanyoles"   
## [29] "pob\_t" "espaÃ.oles"   
## [31] "totalunioneuropea" "ambosexostotalamericana"  
## [33] "totaloceaniayapatridas" "totalhombres"   
## [35] "totalextranjeros" "totalafrica"   
## [37] "totalasia" "totalmujeres"   
## [39] "mujeresespanyolas" "hombresespanyoles"   
## [41] "menores" "edad20\_29"   
## [43] "edad30\_39" "edad40\_49"   
## [45] "edad50\_59" "edad60\_69"   
## [47] "edad70\_79" "mayores80"

data <- mutate(data,   
 esp = espaÃ.oles/pob\_t,  
 euro = europeossinespanyoles/pob\_t,   
 asia = totalasia/pob\_t,   
 ame = americana/pob\_t,   
 UE = totalunioneuropea/pob\_t,   
 oce\_apa = totaloceaniayapatridas/pob\_t,   
 afr = totalafrica/pob\_t,   
 h = totalhombres/pob\_t,   
 m = totalmujeres/pob\_t,   
 m19 = menores/pob\_t,  
 e20\_29 = edad20\_29/pob\_t,   
 e30\_39 = edad30\_39/pob\_t,   
 e40\_49 = edad40\_49/pob\_t,   
 e50\_59 = edad50\_59/pob\_t,  
 e60\_69 = edad60\_69/pob\_t,   
 e70\_79 = edad70\_79/pob\_t,  
 M80 = mayores80/pob\_t)  
  
data <- mutate(data, esp = espaÃ.oles/pob\_t,euro = europeossinespanyoles/pob\_t, asia = totalasia/pob\_t, ame = americana/pob\_t, UE = totalunioneuropea/pob\_t, oce\_apa = totaloceaniayapatridas/pob\_t, afr = totalafrica/pob\_t, h = totalhombres/pob\_t, m = totalmujeres/pob\_t, m19 = menores/pob\_t,e20\_29 = edad20\_29/pob\_t, e30\_39 = edad30\_39/pob\_t, e40\_49 = edad40\_49/pob\_t, e50\_59 = edad50\_59/pob\_t,e60\_69 = edad60\_69/pob\_t, e70\_79 = edad70\_79/pob\_t,M80 = mayores80/pob\_t)

padro <- readRDS(file = "data/SEPE/df\_sepe\_hist\_m\_padro.RDS")  
congres <- readRDS(file="data/electorals/d0911\_esp\_con\_2019\_mun.RData")  
congres <- aggregate(congres[,colnames(congres)[-c(1:5)]],list(id\_mun = substr(congres$id\_sscc, 1, 5)), sum, na.rm = TRUE)  
suma <- merge(congres, padro)  
data1 <- suma[suma$ANY == 2019,]

data1 <- rename(data1, mun = id\_mun)

data1 <- mutate(data1, v\_blanc = n\_vot\_blanco/n\_escrutinio\_censo, v\_nul = n\_vot\_nulos/n\_escrutinio\_censo, PSOE\_p = PSOE/n\_escrutinio\_censo, PP\_p = PP/n\_escrutinio\_censo, Cs\_p = Cs/n\_escrutinio\_censo, ERC\_p = ERC/n\_escrutinio\_censo, VOX\_p = VOX/n\_escrutinio\_censo, JUNTS\_p = JUNTS/n\_escrutinio\_censo, PNV\_p = PNV/n\_escrutinio\_censo, PODEMOS\_p = PODEMOS\_IU/n\_escrutinio\_censo, Bildu\_p = EH\_Bildu/n\_escrutinio\_censo, BNG\_p = BNG/n\_escrutinio\_censo, CUP\_p = CUP/n\_escrutinio\_censo)

data$mun <- sprintf("%05d", as.numeric(data$mun))

d <- merge(data, data1, by = c("mun"))

d1 <- data.frame(d$mun, d$v\_blanc, d$v\_nul, d$PSOE\_p, d$PP\_p, d$Cs\_p, d$ERC\_p, d$VOX\_p, d$JUNTS\_p, d$PNV\_p, d$PODEMOS\_p, d$Bildu\_p, d$BNG\_p, d$CUP\_p, d$t\_atur, d$t\_atur\_m24, d$t\_atur\_25\_44, d$t\_atur\_M45, d$t\_atur\_h, d$t\_atur\_d, d$pob\_t, d$esp, d$euro, d$asia, d$ame, d$UE, d$oce\_apa, d$afr,d$h, d$m, d$m19,d$e20\_29, d$e30\_39, d$e40\_49, d$e50\_59, d$e60\_69,d$e70\_79, d$M80)  
  
d1 <- rename(d1, mun= d.mun, v\_blanc= d.v\_blanc, v\_nul= d.v\_nul, PSOE= d.PSOE\_p, PP= d.PP\_p, Cs= d.Cs\_p, ERC= d.ERC\_p, VOX= d.VOX\_p, JUNTS= d.JUNTS\_p, PNV= d.PNV\_p, PODEMOS= d.PODEMOS\_p, Bildu= d.Bildu\_p, BNG= d.BNG\_p, CUP= d.CUP\_p, atur= d.t\_atur, atur\_m24= d.t\_atur\_m24, atur\_25\_44= d.t\_atur\_25\_44, atur\_M45= d.t\_atur\_M45, atur\_h= d.t\_atur\_h, atur\_d= d.t\_atur\_d, pob\_t= d.pob\_t, esp= d.esp, euro= d.euro, asia= d.asia, ame= d.ame, UE= d.UE, oce\_apa= d.oce\_apa, afr= d.afr,h= d.h, m= d.m, m19= d.m19,e20\_29= d.e20\_29, e30\_39= d.e30\_39, e40\_49= d.e40\_49, e50\_59= d.e50\_59, e60\_69= d.e60\_69,e70\_79= d.e70\_79, M80= d.M80)

names(d1)

## [1] "mun" "v\_blanc" "v\_nul" "PSOE" "PP"   
## [6] "Cs" "ERC" "VOX" "JUNTS" "PNV"   
## [11] "PODEMOS" "Bildu" "BNG" "CUP" "atur"   
## [16] "atur\_m24" "atur\_25\_44" "atur\_M45" "atur\_h" "atur\_d"   
## [21] "pob\_t" "esp" "euro" "asia" "ame"   
## [26] "UE" "oce\_apa" "afr" "h" "m"   
## [31] "m19" "e20\_29" "e30\_39" "e40\_49" "e50\_59"   
## [36] "e60\_69" "e70\_79" "M80"

Renda

df01\_W\_2019 <- readRDS(file="data/ine/dadesRENTA\_01\_2019.rds")  
df09\_W\_2019 <- readRDS(file="data/ine/dadesRENTA\_09\_2019.rds")  
df10\_W\_2019 <- readRDS(file="data/ine/dadesRENTA\_10\_2019.rds")  
  
dades\_sscc<-merge(df01\_W\_2019,df09\_W\_2019)  
dades\_sscc <- merge(dades\_sscc,df10\_W\_2019)  
  
dades\_socio\_mun <- dades\_sscc %>%   
 mutate(id\_mun = substr(id\_sscc,1,5)  
 ,Pob\_renta = ifelse(is.na(Renta\_neta\_media\_por\_persona\_),NA,Población)  
 ,Renta\_Pob = ifelse(is.na(Población),NA,Renta\_neta\_media\_por\_persona\_)  
 ,Pob\_gini = ifelse(is.na(Índice\_de\_Gini),NA,Población)  
 ,Gini\_Pob = ifelse(is.na(Población),NA,Índice\_de\_Gini)  
 ,Pob\_edad = ifelse(is.na(Edad\_media\_de\_la\_población),NA,Población)  
 ,Edad\_Pob = ifelse(is.na(Población),NA,Edad\_media\_de\_la\_población)  
 ,tam\_hog\_Pob = ifelse(is.na(Población),NA,Tamaño\_medio\_del\_hogar)  
 ,Por\_hog\_uni\_Pob = ifelse(is.na(Población),NA,Porcentaje\_de\_hogares\_unipersonales)  
 ) %>%   
 group\_by(id\_mun) %>%   
 summarise(Renta\_neta\_persona = sum(Renta\_Pob\*Pob\_renta,na.rm=TRUE)/sum(Pob\_renta,na.rm=TRUE)  
 ,Gini\_medio = median(Índice\_de\_Gini,na.rm=TRUE)  
 ,Edad\_media = sum(Edad\_Pob\*Pob\_edad,na.rm=TRUE)/sum(Pob\_edad,na.rm=TRUE)  
 ,Poblacion = sum(Población,na.rm=TRUE)  
 ,tam\_hog\_medio = sum(tam\_hog\_Pob\*Pob\_edad,na.rm=TRUE)/sum(Pob\_edad,na.rm=TRUE)  
 ,por\_hog\_uni = sum(Por\_hog\_uni\_Pob\*Pob\_edad,na.rm=TRUE)/sum(Pob\_edad,na.rm=TRUE)  
 ) %>%   
 ungroup() %>%   
 filter(Poblacion>0)  
  
d1 <- merge(d1,dades\_socio\_mun,by.x="mun",by.y="id\_mun")  
d1 <- na.omit(d1)

## 1.1. Rellevància i pertinència

Particularment a nosaltres no ens interessa fer recerca sense un motiu o interès més enllà de l’avanç de les ciències polítiques. Volem que el treball serveixi per a entendre millor per què la gent es decideix a votar per la formació d’extrema dreta a Espanya. També que aquest treball pugui ser utilitzat per alcaldies per a prevenir una expansió de l’extrema dreta a nivell local. Després dels resultats es podrà deduir si les pors als quals fa referència Vox són demandes insatisfetes pels governs locals a la població més afectada.

## 1.2. Pregunta d’investigació i hipòtesis

Pregunta d’investigació: hi ha una correlació entre variables independents i el vot a VOX en eleccions municipals a Espanya?

Variable dependent: vot a VOX en les municipals

5 variables independents claus: 1. Atur 2. Percentatge d’immigrants 3. Edat 4. Sexe, 5. Percentatge de vot a partits nacionalistes

# 2. Variables

## 2.1. Anàlisi univariant

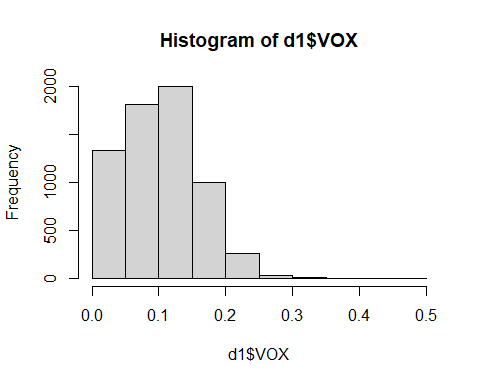
### 2.1.1. Percentatge de vot a Vox (Variable dependent)

summary(d1$VOX)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.05788 0.10198 0.10178 0.13939 0.45045

Podem veure que hi han municipis on VOX no va treure cap vot, i el màxim de vots que va treure va ser un 45,05%. La mitjana de vots va ser de 10,20%. És important analitzar la distribució de la variable.

hist(d1$VOX)



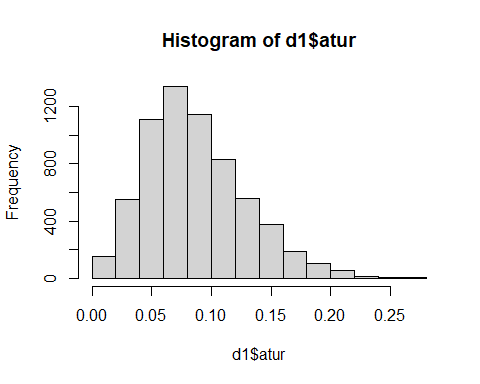
### 2.1.2. Atur (Variable independent)

#### Atur total

summary(d1$atur)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.05723 0.08112 0.08729 0.11221 0.27778

hist(d1$atur)

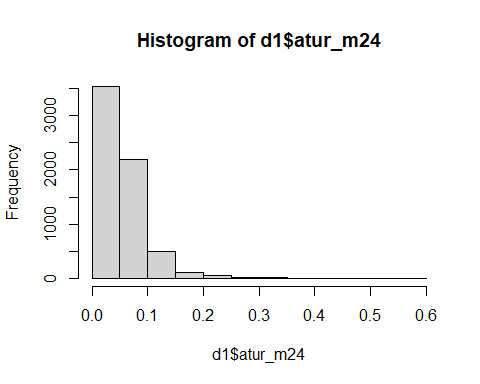


#### Atur de les persones menors de 24 anys

summary(d1$atur\_m24)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.02427 0.04571 0.05336 0.07217 0.58333

hist(d1$atur\_m24)

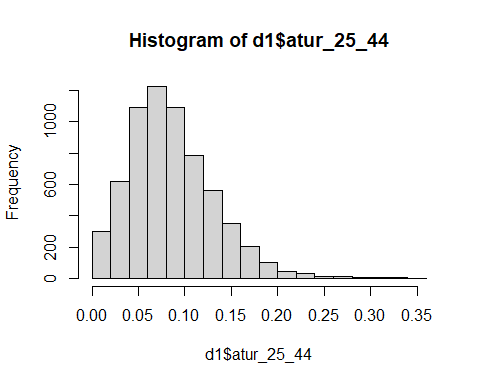


#### Atur de les persones entre 25 i 44 anys

summary(d1$atur\_25\_44)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.05342 0.07979 0.08567 0.11193 0.34677

hist(d1$atur\_25\_44)

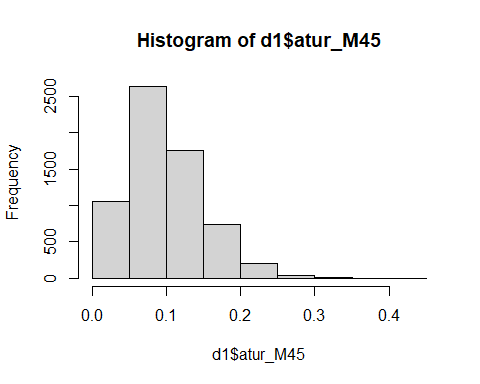


#### Atur de les persones majors de 45 anys

summary(d1$atur\_M45)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.06251 0.09161 0.09842 0.12746 0.42460

hist(d1$atur\_M45)



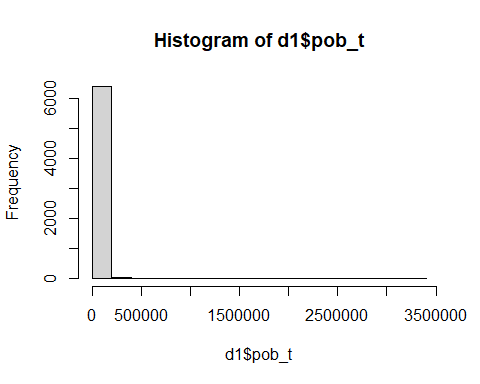
### 2.1.3. Percentatge d’immigrants

#### Població total

summary(d1$pob\_t)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 62 289 858 7136 3387 3266126

hist(d1$pob\_t)

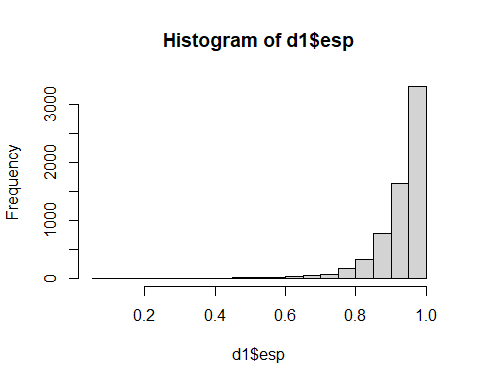


#### Espanya

summary(d1$esp)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.09291 0.90495 0.95236 0.92949 0.97866 1.00000

hist(d1$esp)

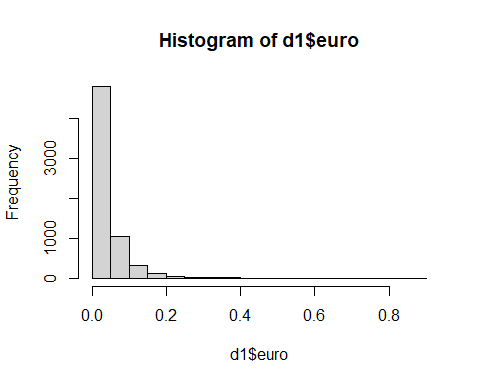


#### Europeus

summary(d1$euro)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000000 0.008825 0.022645 0.040836 0.050749 0.878378

hist(d1$euro)

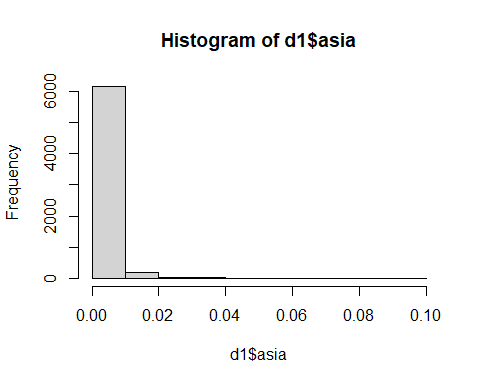


#### Asiàtics

summary(d1$asia)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000000 0.000000 0.000000 0.002046 0.002132 0.098901

hist(d1$asia)

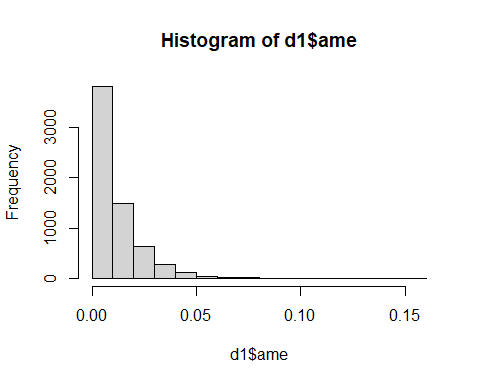


#### Americans

summary(d1$ame)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000000 0.003030 0.007752 0.011325 0.015874 0.154550

hist(d1$ame)

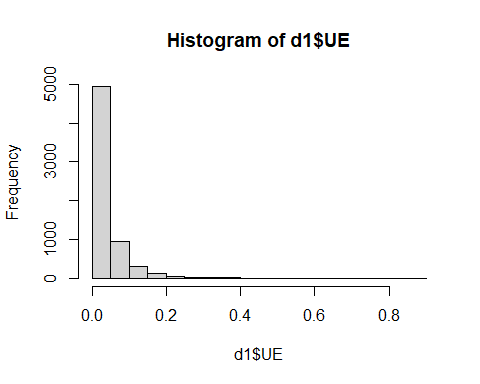


#### Unió Europea

summary(d1$UE)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.00800 0.02047 0.03808 0.04635 0.87331

hist(d1$UE)

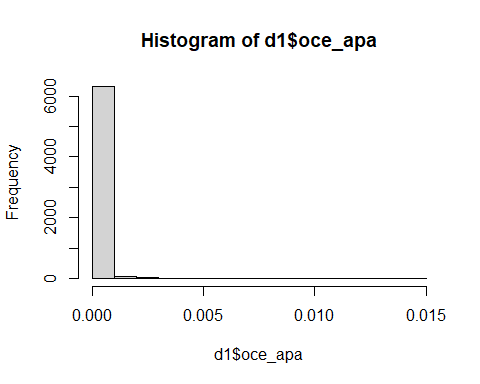


#### Oceania i apàtrides

summary(d1$oce\_apa)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000e+00 0.000e+00 0.000e+00 7.676e-05 0.000e+00 1.442e-02

hist(d1$oce\_apa)



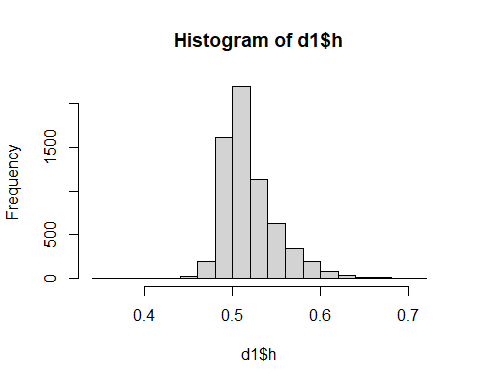
### 2.1.4. Sexe

#### Homes

summary(d1$h)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.3457 0.4985 0.5117 0.5190 0.5325 0.7021

hist(d1$h)

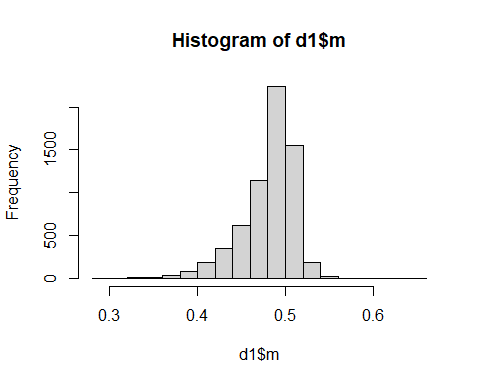


#### Dones

summary(d1$m)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.2979 0.4675 0.4883 0.4810 0.5015 0.6543

hist(d1$m)



## 2.2. Anàlisi bivariant

### 2.2.1. Vot a Vox i atur

d2 <- d1 %>% filter(atur > 0, VOX > 0)

cor(d2$atur,d2$VOX)

## [1] 0.1996308

cor(log(d2$atur), logit(d2$VOX))

## [1] 0.2530929

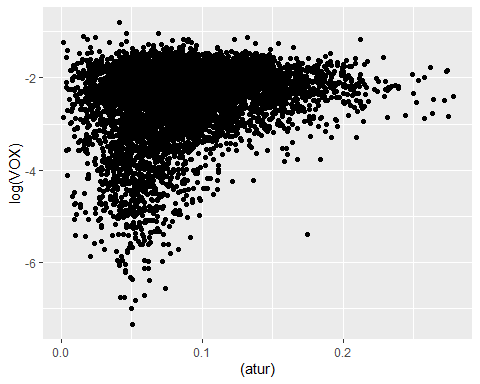
cor((d2$atur),logit(d2$VOX))

## [1] 0.2607664

cor(d2$atur, log(d2$VOX))

## [1] 0.2638057

d2 %>% ggplot(aes(x = (atur), y = log(VOX))) +  
 geom\_point()

 ### Propensió de l’atur

quantile(d2$atur, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.001282051 0.052897247 0.072099694 0.092767837 0.121471869 0.277777778

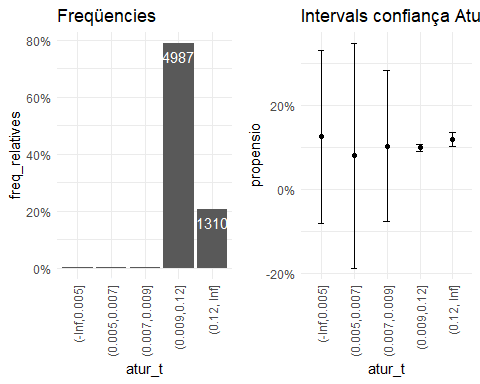
d2 <- d2 %>%   
 mutate(atur\_t = cut(atur, c(-Inf, 0.005, 0.007, 0.009, 0.12, Inf )))  
  
d2 %>% count(atur\_t)

## atur\_t n  
## 1 (-Inf,0.005] 10  
## 2 (0.005,0.007] 4  
## 3 (0.007,0.009] 11  
## 4 (0.009,0.12] 4987  
## 5 (0.12, Inf] 1310

PROPEN <- d2 %>%   
 group\_by(atur\_t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN

## # A tibble: 5 x 6  
## atur\_t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.005] 10 -0.0798 0.126 0.331 0.00158   
## 2 (0.005,0.007] 4 -0.186 0.0804 0.347 0.000633  
## 3 (0.007,0.009] 11 -0.0765 0.104 0.284 0.00174   
## 4 (0.009,0.12] 4987 0.0911 0.0994 0.108 0.789   
## 5 (0.12, Inf] 1310 0.102 0.119 0.137 0.207

p1 <- PROPEN %>%   
 ggplot(aes(atur\_t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p2 <- PROPEN %>%   
 ggplot(aes(atur\_t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança Atur")  
  
plot\_grid(p1, p2, ncol=2)



\*\*\* Queda pendent estudis posteriors sobre el valor 0

#### 2.2.1.1. Vot a Vox i atur dels menors de 24 anys

cor(d2$atur\_m24,d2$VOX, use = "pairwise.complete.obs")

## [1] 0.1700317

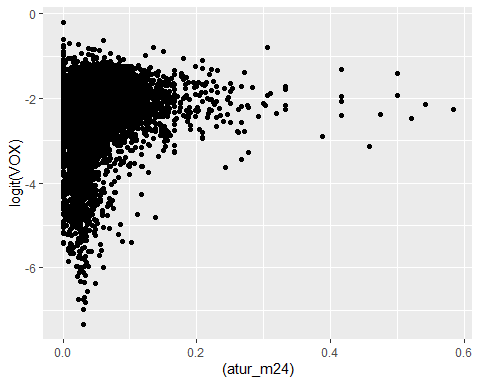
cor(logit(d2$atur\_m24),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.17627

cor(d2$atur\_m24,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.2152908

d2 %>% ggplot(aes(x = (atur\_m24), y = logit(VOX)))+  
 geom\_point()



quantile(d2$atur\_m24, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.00000000 0.02011162 0.03703704 0.05555556 0.08041996 0.58333333

d2 <- d2 %>%   
 mutate(atur\_m24t = cut(atur\_m24, c(-Inf, 0.02, 0.04, 0.05, 0.08, Inf )))  
  
d2 %>% count(atur\_m24t)

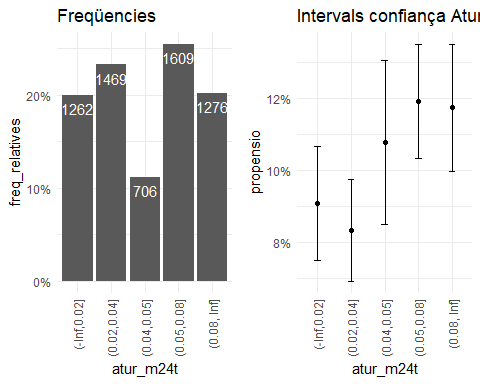
## atur\_m24t n  
## 1 (-Inf,0.02] 1262  
## 2 (0.02,0.04] 1469  
## 3 (0.04,0.05] 706  
## 4 (0.05,0.08] 1609  
## 5 (0.08, Inf] 1276

###Propensió atur dels menors de 24 anys

PROPEN4 <- d2 %>%   
 group\_by(atur\_m24t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN4

## # A tibble: 5 x 6  
## atur\_m24t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.02] 1262 0.0750 0.0909 0.107 0.200  
## 2 (0.02,0.04] 1469 0.0691 0.0833 0.0974 0.232  
## 3 (0.04,0.05] 706 0.0849 0.108 0.131 0.112  
## 4 (0.05,0.08] 1609 0.103 0.119 0.135 0.255  
## 5 (0.08, Inf] 1276 0.0998 0.117 0.135 0.202

p7 <- PROPEN4 %>%   
 ggplot(aes(atur\_m24t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p8 <- PROPEN4 %>%   
 ggplot(aes(atur\_m24t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança Atur menors 24 anys")  
  
plot\_grid(p7, p8, ncol=2)



#### 2.2.1.2. Vot a Vox i atur de persones entre 25 y 44 anys

cor(d2$atur\_25\_44,d2$VOX)

## [1] 0.1761566

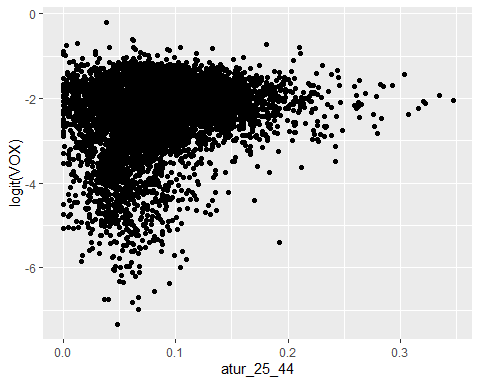
cor(logit(d2$atur\_25\_44),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.1881515

cor(d2$atur\_25\_44,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.2264125

d2 %>% ggplot(aes(x = atur\_25\_44, y = logit(VOX)))+  
 geom\_point()

 ###Propensió atur de persones entre 25 y 44 anys

quantile(d2$atur\_25\_44, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.00000000 0.04888453 0.07038067 0.09123261 0.12205172 0.34677419

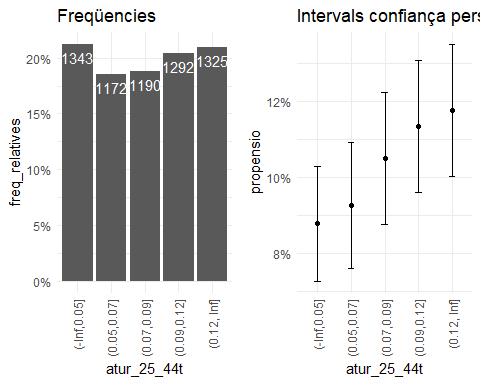
d2 <- d2 %>%   
 mutate(atur\_25\_44t = cut(atur\_25\_44, c(-Inf, 0.05, 0.07, 0.09, 0.12, Inf )))  
d2 %>% count(atur\_25\_44t)

## atur\_25\_44t n  
## 1 (-Inf,0.05] 1343  
## 2 (0.05,0.07] 1172  
## 3 (0.07,0.09] 1190  
## 4 (0.09,0.12] 1292  
## 5 (0.12, Inf] 1325

PROPEN5 <- d2 %>%   
 group\_by(atur\_25\_44t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN5

## # A tibble: 5 x 6  
## atur\_25\_44t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.05] 1343 0.0728 0.0879 0.103 0.212  
## 2 (0.05,0.07] 1172 0.0761 0.0927 0.109 0.185  
## 3 (0.07,0.09] 1190 0.0877 0.105 0.123 0.188  
## 4 (0.09,0.12] 1292 0.0963 0.114 0.131 0.204  
## 5 (0.12, Inf] 1325 0.100 0.118 0.135 0.210

p9 <- PROPEN5 %>%   
 ggplot(aes(atur\_25\_44t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p10 <- PROPEN5 %>%   
 ggplot(aes(atur\_25\_44t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança persones en atur entre 25 i 44 anys")  
  
plot\_grid(p9, p10, ncol=2)



#### 2.2.1.3. Vot a Vox i atur de majors de 45 anys

cor(d2$atur\_M45,d2$VOX)

## [1] 0.1844595

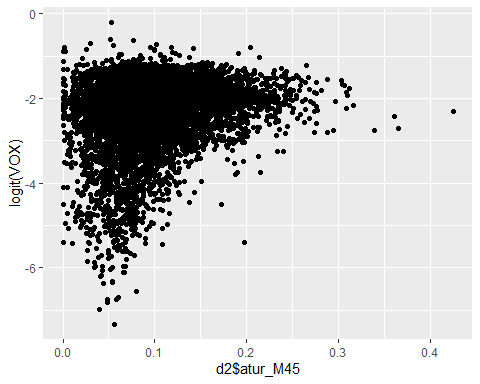
cor(logit(d2$atur\_M45),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.2010924

cor(d2$atur\_M45,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.2369323

d2 %>% ggplot(aes(x = d2$atur\_M45, y = logit(VOX)))+  
 geom\_point()

 ###Propensió atur de majors de 45 anys

quantile(d2$atur\_M45, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.00000000 0.05663666 0.08101405 0.10453481 0.13938552 0.42460317

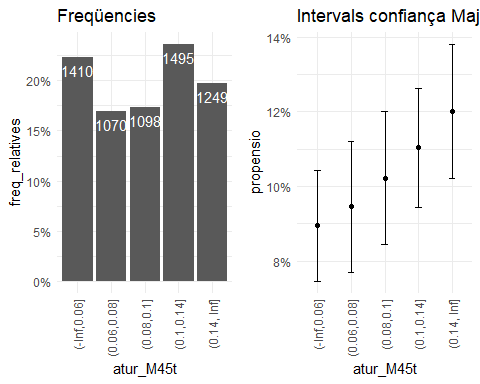
d2 <- d2 %>%   
 mutate(atur\_M45t = cut(atur\_M45, c(-Inf, 0.06, 0.08, 0.10, 0.14, Inf )))  
d2 %>% count(atur\_M45t)

## atur\_M45t n  
## 1 (-Inf,0.06] 1410  
## 2 (0.06,0.08] 1070  
## 3 (0.08,0.1] 1098  
## 4 (0.1,0.14] 1495  
## 5 (0.14, Inf] 1249

PROPEN6 <- d2 %>%   
 group\_by(atur\_M45t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN6

## # A tibble: 5 x 6  
## atur\_M45t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.06] 1410 0.0746 0.0895 0.104 0.223  
## 2 (0.06,0.08] 1070 0.0770 0.0946 0.112 0.169  
## 3 (0.08,0.1] 1098 0.0844 0.102 0.120 0.174  
## 4 (0.1,0.14] 1495 0.0945 0.110 0.126 0.236  
## 5 (0.14, Inf] 1249 0.102 0.120 0.138 0.198

p11 <- PROPEN6 %>%   
 ggplot(aes(atur\_M45t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p12 <- PROPEN6 %>%   
 ggplot(aes(atur\_M45t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança Majors 45 anys")  
  
plot\_grid(p11, p12, ncol=2)

 ### 2.2.2. Vot a Vox i edat

d2 %>%  
 mutate(logit\_VOX = logit(VOX)   
 ,m29 = m19 + e20\_29) %>%   
 select(m29, e30\_39, e40\_49, e50\_59, e60\_69, e70\_79, M80, logit\_VOX, VOX) %>%   
 cor()

## m29 e30\_39 e40\_49 e50\_59 e60\_69  
## m29 1.00000000 0.66214699 0.60603067 -0.32742005 -0.69254006  
## e30\_39 0.66214699 1.00000000 0.45267834 -0.39357321 -0.44005331  
## e40\_49 0.60603067 0.45267834 1.00000000 -0.33107348 -0.57429405  
## e50\_59 -0.32742005 -0.39357321 -0.33107348 1.00000000 0.17842565  
## e60\_69 -0.69254006 -0.44005331 -0.57429405 0.17842565 1.00000000  
## e70\_79 -0.81087686 -0.64637619 -0.58319247 0.03717047 0.53724211  
## M80 -0.84211908 -0.69296448 -0.69620043 0.19716161 0.44411637  
## logit\_VOX -0.02997110 0.03861837 -0.03480903 0.05532392 -0.06755104  
## VOX 0.04932834 0.10020407 0.03843048 0.02491326 -0.10859528  
## e70\_79 M80 logit\_VOX VOX  
## m29 -0.81087686 -0.84211908 -0.02997110 0.04932834  
## e30\_39 -0.64637619 -0.69296448 0.03861837 0.10020407  
## e40\_49 -0.58319247 -0.69620043 -0.03480903 0.03843048  
## e50\_59 0.03717047 0.19716161 0.05532392 0.02491326  
## e60\_69 0.53724211 0.44411637 -0.06755104 -0.10859528  
## e70\_79 1.00000000 0.68689894 0.01500536 -0.06287652  
## M80 0.68689894 1.00000000 0.04142835 -0.03900435  
## logit\_VOX 0.01500536 0.04142835 1.00000000 0.90285519  
## VOX -0.06287652 -0.03900435 0.90285519 1.00000000

d2$logit\_VOX <- logit(d2$VOX)  
d2$m29 <- d2$m19 + d2$e20\_29

### 2.2.2.1 Vot a Vox i menors 29 anys

cor(d2$m29,d2$VOX)

## [1] 0.04932834

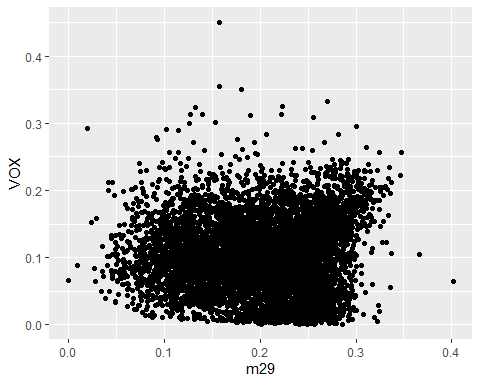
cor(logit(d2$m29),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.0422594

cor(d2$m29,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.0299711

d2 %>% ggplot(aes(x = m29, y = VOX))+  
 geom\_point()

 ### Propensión del vot a Vox i menors 29 anys

quantile(d2$m29, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.0000000 0.1383616 0.1831601 0.2229591 0.2545142 0.4017094

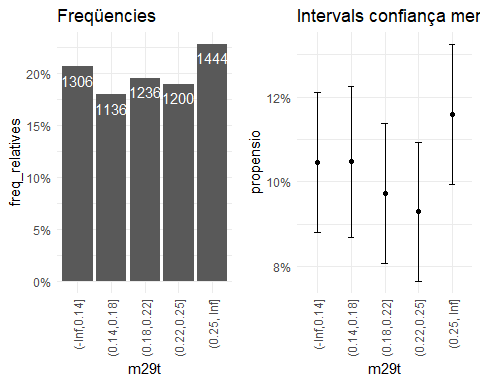
d2 <- d2 %>%   
 mutate(m29t = cut(m29, c(-Inf, 0.14, 0.18, 0.22, 0.25, Inf )))  
  
d2 %>% count(m29t)

## m29t n  
## 1 (-Inf,0.14] 1306  
## 2 (0.14,0.18] 1136  
## 3 (0.18,0.22] 1236  
## 4 (0.22,0.25] 1200  
## 5 (0.25, Inf] 1444

PROPEN9 <- d2 %>%   
 group\_by(m29t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN9

## # A tibble: 5 x 6  
## m29t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.14] 1306 0.0880 0.105 0.121 0.207  
## 2 (0.14,0.18] 1136 0.0869 0.105 0.123 0.180  
## 3 (0.18,0.22] 1236 0.0807 0.0972 0.114 0.196  
## 4 (0.22,0.25] 1200 0.0765 0.0930 0.109 0.190  
## 5 (0.25, Inf] 1444 0.0994 0.116 0.132 0.228

p14 <- PROPEN9 %>%   
 ggplot(aes(m29t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p15 <- PROPEN9 %>%   
 ggplot(aes(m29t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança menors 29 anys")  
  
plot\_grid(p14, p15, ncol=2)



### 2.2.2.2 Vot a Vox i edat 30 - 39

cor(d2$e30\_39,d2$VOX)

## [1] 0.1002041

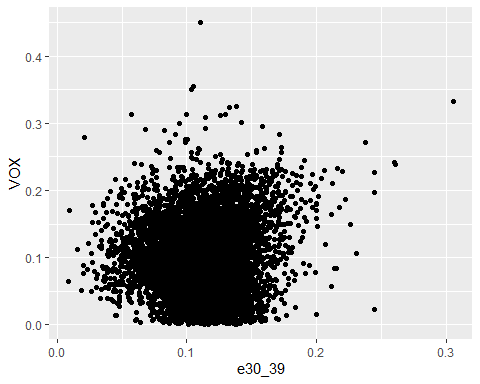
cor(logit(d2$e30\_39),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.01334474

cor(d2$e30\_39,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.03861837

d2 %>% ggplot(aes(x = e30\_39, y = VOX))+  
 geom\_point()

 ### Propensión del vot a Vox i edat 30 - 39 anys

quantile(d2$e30\_39, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.008849558 0.090909091 0.109189426 0.122479598 0.135470018 0.305743243

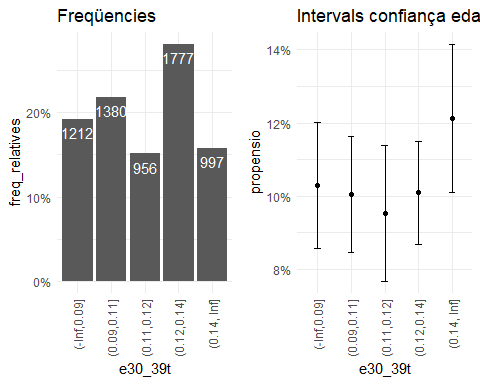
d2 <- d2 %>%   
 mutate(e30\_39t = cut(e30\_39, c(-Inf, 0.09, 0.11, 0.12, 0.14, Inf )))  
  
d2 %>% count(e30\_39t)

## e30\_39t n  
## 1 (-Inf,0.09] 1212  
## 2 (0.09,0.11] 1380  
## 3 (0.11,0.12] 956  
## 4 (0.12,0.14] 1777  
## 5 (0.14, Inf] 997

PROPEN10 <- d2 %>%   
 group\_by(e30\_39t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN10

## # A tibble: 5 x 6  
## e30\_39t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.09] 1212 0.0859 0.103 0.120 0.192  
## 2 (0.09,0.11] 1380 0.0845 0.100 0.116 0.218  
## 3 (0.11,0.12] 956 0.0767 0.0954 0.114 0.151  
## 4 (0.12,0.14] 1777 0.0869 0.101 0.115 0.281  
## 5 (0.14, Inf] 997 0.101 0.121 0.142 0.158

p16 <- PROPEN10 %>%   
 ggplot(aes(e30\_39t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p17 <- PROPEN10 %>%   
 ggplot(aes(e30\_39t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança edat 30 - 39")  
  
plot\_grid(p16, p17, ncol=2)

 ### 2.2.2.3 Vot a Vox i edat 40 - 49

cor(d2$e40\_49,d2$VOX)

## [1] 0.03843048

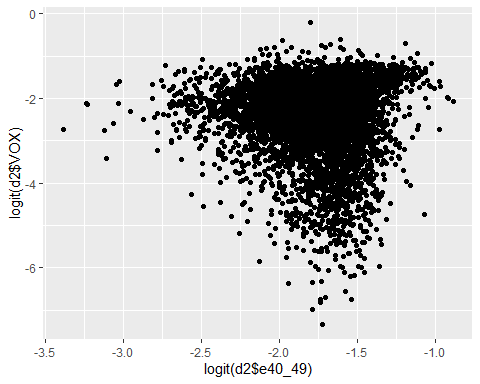
cor(logit(d2$e40\_49),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.04312601

cor(d2$e40\_49,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.03480903

d2 %>% ggplot(aes(x = logit(d2$e40\_49), y = logit(d2$VOX)))+  
 geom\_point()

 ### Propensión del vot a Vox i edat 40 - 49 anys

quantile(d2$e40\_49, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.0326087 0.1206069 0.1392173 0.1558442 0.1735694 0.2924528

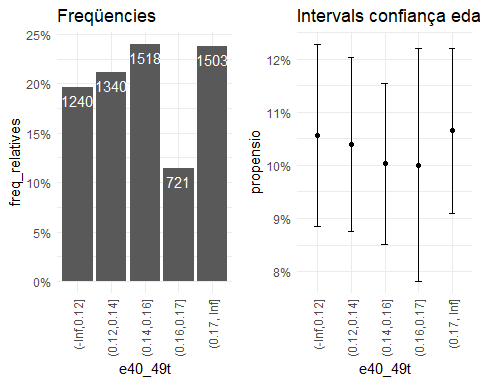
d2 <- d2 %>%   
 mutate(e40\_49t = cut(e40\_49, c(-Inf, 0.12, 0.14, 0.16, 0.17, Inf )))  
  
d2 %>% count(e40\_49t)

## e40\_49t n  
## 1 (-Inf,0.12] 1240  
## 2 (0.12,0.14] 1340  
## 3 (0.14,0.16] 1518  
## 4 (0.16,0.17] 721  
## 5 (0.17, Inf] 1503

PROPEN11 <- d2 %>%   
 group\_by(e40\_49t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN11

## # A tibble: 5 x 6  
## e40\_49t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.12] 1240 0.0886 0.106 0.123 0.196  
## 2 (0.12,0.14] 1340 0.0876 0.104 0.120 0.212  
## 3 (0.14,0.16] 1518 0.0852 0.100 0.115 0.240  
## 4 (0.16,0.17] 721 0.0782 0.100 0.122 0.114  
## 5 (0.17, Inf] 1503 0.0909 0.106 0.122 0.238

p18 <- PROPEN11 %>%   
 ggplot(aes(e40\_49t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p19 <- PROPEN11 %>%   
 ggplot(aes(e40\_49t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança edat 40 - 49")  
  
plot\_grid(p18, p19, ncol=2)

 ### 2.2.2.4 Vot a Vox i edat 50 - 59

cor(d2$e50\_59,d2$VOX)

## [1] 0.02491326

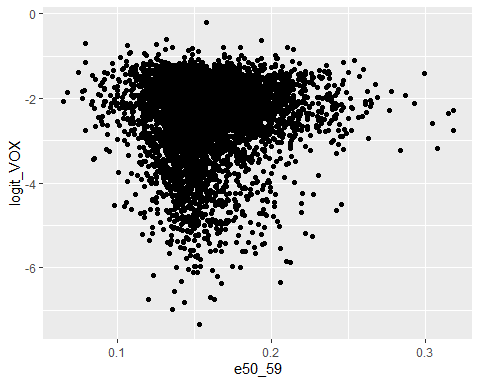
cor(logit(d2$e50\_59),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.04888718

cor(d2$e50\_59,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.05532392

d2 %>% ggplot(aes(x = e50\_59, y = logit\_VOX))+  
 geom\_point()

 ### Propensión del vot a Vox i edat 50 - 59 anys

quantile(d2$e50\_59, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.06451613 0.13962120 0.15091651 0.16166223 0.17819997 0.31782946

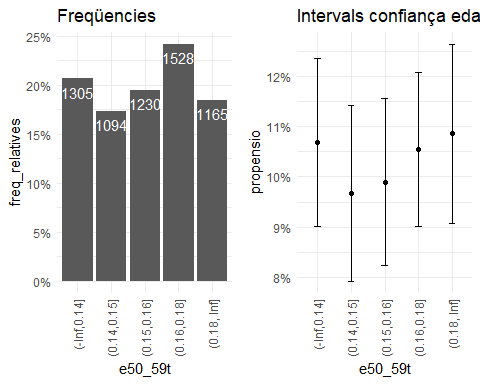
d2 <- d2 %>%   
 mutate(e50\_59t = cut(e50\_59, c(-Inf, 0.14, 0.15, 0.16, 0.18, Inf )))  
  
d2 %>% count(e50\_59t)

## e50\_59t n  
## 1 (-Inf,0.14] 1305  
## 2 (0.14,0.15] 1094  
## 3 (0.15,0.16] 1230  
## 4 (0.16,0.18] 1528  
## 5 (0.18, Inf] 1165

PROPEN12 <- d2 %>%   
 group\_by(e50\_59t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN12

## # A tibble: 5 x 6  
## e50\_59t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.14] 1305 0.0901 0.107 0.124 0.206  
## 2 (0.14,0.15] 1094 0.0792 0.0967 0.114 0.173  
## 3 (0.15,0.16] 1230 0.0823 0.0990 0.116 0.195  
## 4 (0.16,0.18] 1528 0.0901 0.106 0.121 0.242  
## 5 (0.18, Inf] 1165 0.0907 0.109 0.126 0.184

p20 <- PROPEN12 %>%   
 ggplot(aes(e50\_59t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p21 <- PROPEN12 %>%   
 ggplot(aes(e50\_59t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança edat 50 - 59")  
  
plot\_grid(p20, p21, ncol=2)

 ### 2.2.2.5 Vot a Vox i edat 60 - 69

cor(d2$e60\_69,d2$VOX)

## [1] -0.1085953

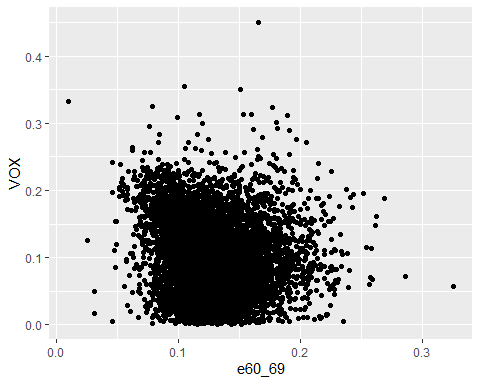
cor(logit(d2$e60\_69),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.08871025

cor(d2$e60\_69,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.06755104

d2 %>% ggplot(aes(x = e60\_69, y = VOX))+  
 geom\_point()

 ### Propensión del vot a Vox i edat 60 - 69 anys

quantile(d2$e60\_69, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.01013514 0.10260121 0.11643488 0.13232400 0.15302338 0.32500000

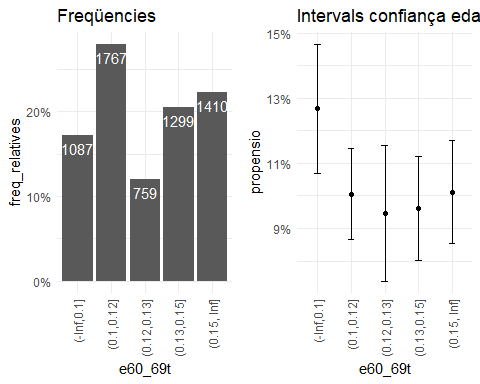
d2 <- d2 %>%   
 mutate(e60\_69t = cut(e60\_69, c(-Inf, 0.10, 0.12, 0.13, 0.15, Inf )))  
  
d2 %>% count(e60\_69t)

## e60\_69t n  
## 1 (-Inf,0.1] 1087  
## 2 (0.1,0.12] 1767  
## 3 (0.12,0.13] 759  
## 4 (0.13,0.15] 1299  
## 5 (0.15, Inf] 1410

PROPEN13 <- d2 %>%   
 group\_by(e60\_69t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN13

## # A tibble: 5 x 6  
## e60\_69t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.1] 1087 0.107 0.127 0.147 0.172  
## 2 (0.1,0.12] 1767 0.0866 0.101 0.115 0.280  
## 3 (0.12,0.13] 759 0.0738 0.0946 0.115 0.120  
## 4 (0.13,0.15] 1299 0.0801 0.0961 0.112 0.205  
## 5 (0.15, Inf] 1410 0.0854 0.101 0.117 0.223

p22 <- PROPEN13 %>%   
 ggplot(aes(e60\_69t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p23 <- PROPEN13 %>%   
 ggplot(aes(e60\_69t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança edat 60 - 69")  
  
plot\_grid(p22, p23, ncol=2)

 ### 2.2.2.6 Vot a Vox i edat 70 - 79

cor(d2$e70\_79,d2$VOX)

## [1] -0.06287652

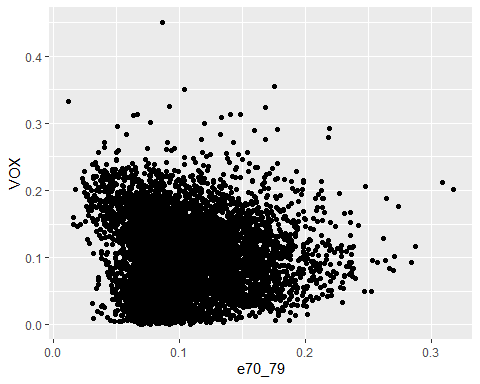
cor(logit(d2$e70\_79),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.0180569

cor(d2$e70\_79,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.01500536

d2 %>% ggplot(aes(x = e70\_79, y = VOX))+  
 geom\_point()

 ### Propensión del vot a Vox i edat 70 - 79 anys

quantile(d2$e70\_79, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.01182432 0.07373114 0.08966878 0.10921034 0.13737152 0.31775701

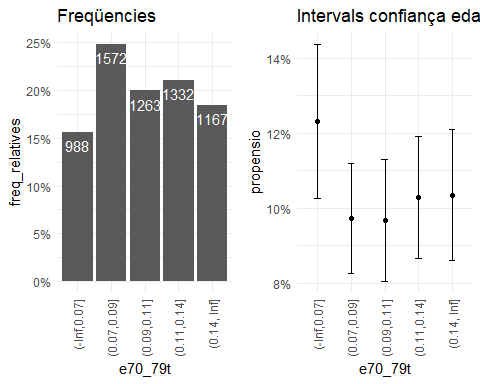
d2 <- d2 %>%   
 mutate(e70\_79t = cut(e70\_79, c(-Inf, 0.07, 0.09, 0.11, 0.14, Inf )))  
  
d2 %>% count(e70\_79t)

## e70\_79t n  
## 1 (-Inf,0.07] 988  
## 2 (0.07,0.09] 1572  
## 3 (0.09,0.11] 1263  
## 4 (0.11,0.14] 1332  
## 5 (0.14, Inf] 1167

PROPEN14 <- d2 %>%   
 group\_by(e70\_79t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN14

## # A tibble: 5 x 6  
## e70\_79t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.07] 988 0.103 0.123 0.144 0.156  
## 2 (0.07,0.09] 1572 0.0826 0.0972 0.112 0.249  
## 3 (0.09,0.11] 1263 0.0805 0.0968 0.113 0.200  
## 4 (0.11,0.14] 1332 0.0866 0.103 0.119 0.211  
## 5 (0.14, Inf] 1167 0.0860 0.103 0.121 0.185

p24 <- PROPEN14 %>%   
 ggplot(aes(e70\_79t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p25 <- PROPEN14 %>%   
 ggplot(aes(e70\_79t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança edat 70 - 79")  
  
plot\_grid(p24, p25, ncol=2)

 ### 2.2.2.7 Vot a Vox i edat majors 80

cor(d2$M80,d2$VOX)

## [1] -0.03900435

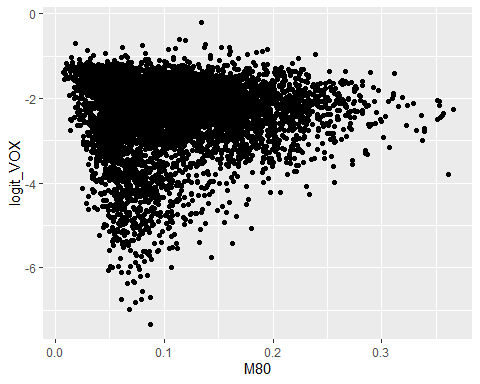
cor(logit(d2$M80),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.0001963107

cor(d2$M80,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.04142835

d2 %>% ggplot(aes(x = M80, y = logit\_VOX))+  
 geom\_point()

 ### Propensión del vot a Vox i majors 80 anys

quantile(d2$M80, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 0.006830601 0.056900548 0.080481155 0.112274866 0.156250000 0.365591398

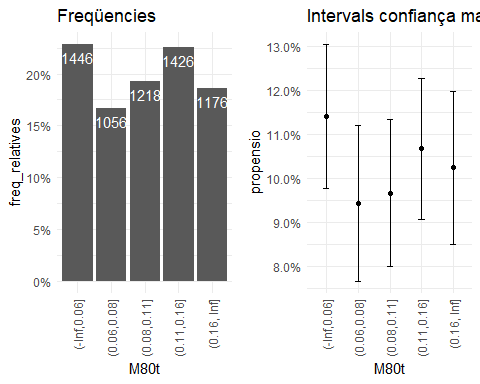
d2 <- d2 %>%   
 mutate(M80t = cut(M80, c(-Inf, 0.06, 0.08, 0.11, 0.16, Inf )))  
  
d2 %>% count(M80t)

## M80t n  
## 1 (-Inf,0.06] 1446  
## 2 (0.06,0.08] 1056  
## 3 (0.08,0.11] 1218  
## 4 (0.11,0.16] 1426  
## 5 (0.16, Inf] 1176

PROPEN15 <- d2 %>%   
 group\_by(M80t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN15

## # A tibble: 5 x 6  
## M80t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,0.06] 1446 0.0976 0.114 0.130 0.229  
## 2 (0.06,0.08] 1056 0.0767 0.0943 0.112 0.167  
## 3 (0.08,0.11] 1218 0.0801 0.0967 0.113 0.193  
## 4 (0.11,0.16] 1426 0.0906 0.107 0.123 0.226  
## 5 (0.16, Inf] 1176 0.0851 0.102 0.120 0.186

p26 <- PROPEN15 %>%   
 ggplot(aes(M80t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p27 <- PROPEN15 %>%   
 ggplot(aes(M80t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança majors 80")  
  
plot\_grid(p26, p27, ncol=2)



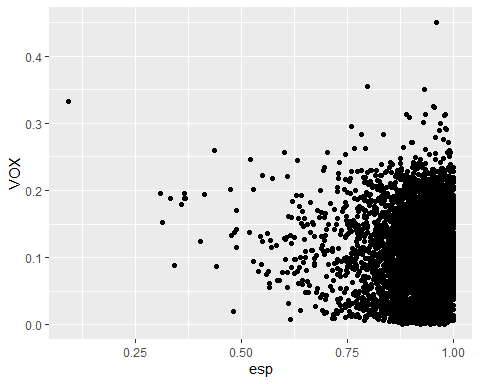
### 2.2.3. Vot a Vox i nacionalitat

d2 %>%  
 mutate(logit\_VOX = logit(VOX)) %>%   
 select(esp, euro, asia, ame, UE, oce\_apa, h, m, afr, logit\_VOX, VOX) %>%   
 cor()

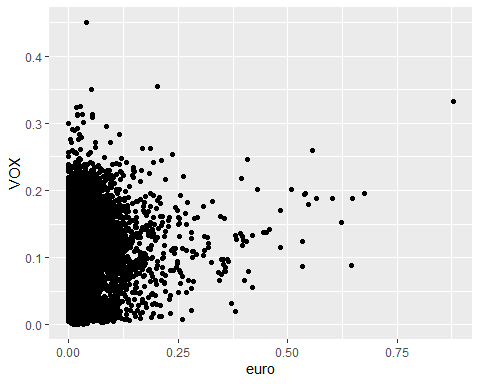
## esp euro asia ame UE  
## esp 1.00000000 -0.89061867 -0.35065482 -0.43835776 -0.87428196  
## euro -0.89061867 1.00000000 0.20042273 0.21294312 0.99298718  
## asia -0.35065482 0.20042273 1.00000000 0.32158037 0.17325282  
## ame -0.43835776 0.21294312 0.32158037 1.00000000 0.19164927  
## UE -0.87428196 0.99298718 0.17325282 0.19164927 1.00000000  
## oce\_apa -0.09535770 0.08959280 0.04951214 0.08914346 0.08528363  
## h 0.04660553 -0.02016117 -0.11837655 -0.16058520 -0.01087058  
## m -0.04660553 0.02016117 0.11837655 0.16058520 0.01087058  
## afr -0.56567846 0.17916211 0.18141695 0.22502723 0.16490349  
## logit\_VOX -0.01910603 0.06962309 -0.06095673 -0.07863587 0.08177046  
## VOX -0.06806519 0.09526608 -0.04590083 -0.02699317 0.10619207  
## oce\_apa h m afr logit\_VOX  
## esp -0.095357699 0.04660553 -0.04660553 -0.565678461 -0.01910603  
## euro 0.089592805 -0.02016117 0.02016117 0.179162106 0.06962309  
## asia 0.049512135 -0.11837655 0.11837655 0.181416948 -0.06095673  
## ame 0.089143465 -0.16058520 0.16058520 0.225027228 -0.07863587  
## UE 0.085283632 -0.01087058 0.01087058 0.164903494 0.08177046  
## oce\_apa 1.000000000 -0.04139945 0.04139945 0.006305305 -0.09591881  
## h -0.041399448 1.00000000 -1.00000000 0.011764574 0.10782588  
## m 0.041399448 -1.00000000 1.00000000 -0.011764574 -0.10782588  
## afr 0.006305305 0.01176457 -0.01176457 1.000000000 -0.04433022  
## logit\_VOX -0.095918814 0.10782588 -0.10782588 -0.044330220 1.00000000  
## VOX -0.065303996 0.12804547 -0.12804547 0.007056704 0.90285519  
## VOX  
## esp -0.068065191  
## euro 0.095266085  
## asia -0.045900826  
## ame -0.026993173  
## UE 0.106192070  
## oce\_apa -0.065303996  
## h 0.128045471  
## m -0.128045471  
## afr 0.007056704  
## logit\_VOX 0.902855191  
## VOX 1.000000000

### 2.2.3.1 Vot a Vox i espanyols

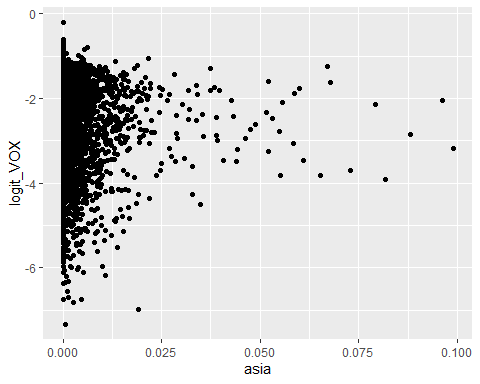
d2 %>% ggplot(aes(x = esp, y = VOX))+  
 geom\_point()

 ### 2.2.3.2 Vot a Vox i nacionalitat europea

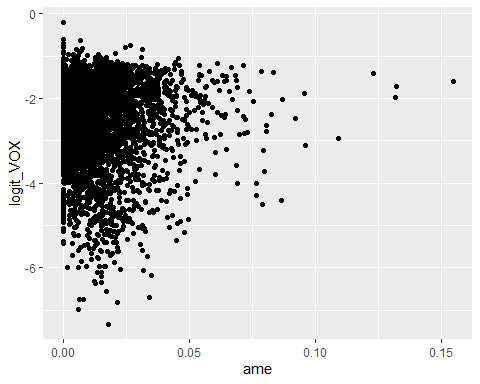
d2 %>% ggplot(aes(x = euro, y = VOX))+  
 geom\_point()

 ### 2.2.3.3 Vot a Vox i nacionalitat asiàtica

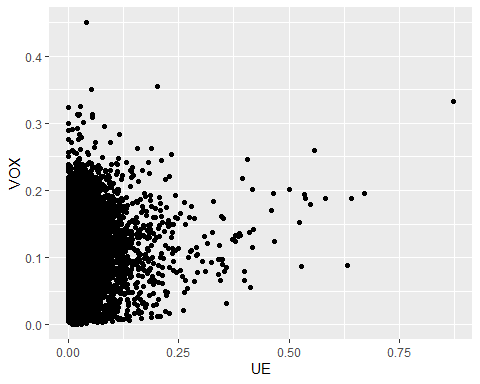
d2 %>% ggplot(aes(x = asia, y = logit\_VOX))+  
 geom\_point()

 ### 2.2.3.4 Vot a Vox i nacionalitat americana

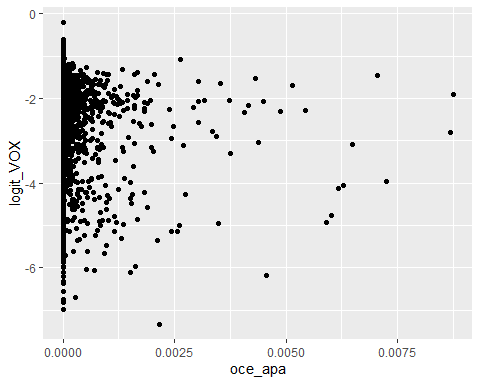
d2 %>% ggplot(aes(x = ame, y = logit\_VOX))+  
 geom\_point()

 ### 2.2.3.5 Vot a Vox i UE

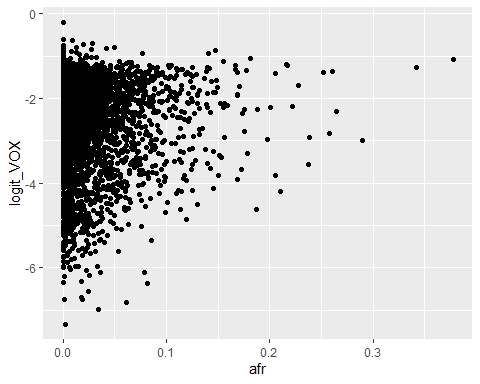
d2 %>% ggplot(aes(x = UE, y = VOX))+  
 geom\_point()

 ### 2.2.3.6 Vot a Vox i oceania i apàtrides

d2 %>% ggplot(aes(x = oce\_apa, y = logit\_VOX))+  
 geom\_point()

 ### 2.2.3.7 Vot a Vox i oceania i africans

d2 %>% ggplot(aes(x = afr, y = logit\_VOX))+  
 geom\_point()



### 2.2.4. Vot a Vox i renta neta per persona

d3 <- d2 %>% filter(VOX > -2)

names(d2)

## [1] "mun" "v\_blanc" "v\_nul"   
## [4] "PSOE" "PP" "Cs"   
## [7] "ERC" "VOX" "JUNTS"   
## [10] "PNV" "PODEMOS" "Bildu"   
## [13] "BNG" "CUP" "atur"   
## [16] "atur\_m24" "atur\_25\_44" "atur\_M45"   
## [19] "atur\_h" "atur\_d" "pob\_t"   
## [22] "esp" "euro" "asia"   
## [25] "ame" "UE" "oce\_apa"   
## [28] "afr" "h" "m"   
## [31] "m19" "e20\_29" "e30\_39"   
## [34] "e40\_49" "e50\_59" "e60\_69"   
## [37] "e70\_79" "M80" "Renta\_neta\_persona"  
## [40] "Gini\_medio" "Edad\_media" "Poblacion"   
## [43] "tam\_hog\_medio" "por\_hog\_uni" "atur\_t"   
## [46] "atur\_m24t" "atur\_25\_44t" "atur\_M45t"   
## [49] "logit\_VOX" "m29" "m29t"   
## [52] "e30\_39t" "e40\_49t" "e50\_59t"   
## [55] "e60\_69t" "e70\_79t" "M80t"

cor(d2$Renta\_neta\_persona,d2$VOX)

## [1] -0.2914367

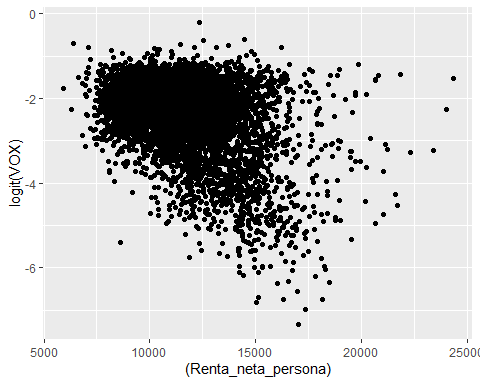
cor(d2$Renta\_neta\_persona,logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.3907623

cor(log(d2$Renta\_neta\_persona),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.378493

d2 %>% ggplot(aes(x = (Renta\_neta\_persona), y = logit(VOX)))+  
 geom\_point()

 ######Propensió Renta neta persona

quantile(d2$Renta\_neta\_persona, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 5898.000 9621.834 10857.252 11920.775 13319.076 24315.911

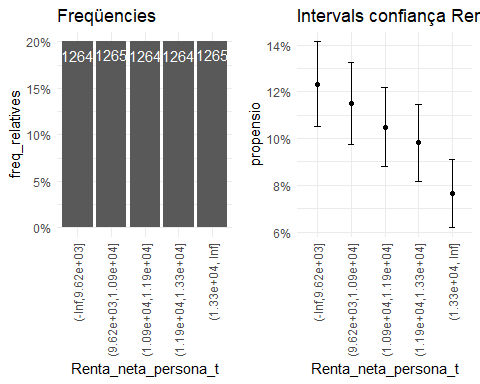
d2 <- d2 %>%   
 mutate(Renta\_neta\_persona\_t = cut(Renta\_neta\_persona, c(-Inf, 9621, 10857.252, 11920.775, 13319.076, Inf )))  
  
d2 %>% count(Renta\_neta\_persona\_t)

## Renta\_neta\_persona\_t n  
## 1 (-Inf,9.62e+03] 1264  
## 2 (9.62e+03,1.09e+04] 1265  
## 3 (1.09e+04,1.19e+04] 1264  
## 4 (1.19e+04,1.33e+04] 1264  
## 5 (1.33e+04, Inf] 1265

PROPEN2 <- d2 %>%   
 group\_by(Renta\_neta\_persona\_t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN2

## # A tibble: 5 x 6  
## Renta\_neta\_persona\_t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,9.62e+03] 1264 0.105 0.123 0.141 0.200  
## 2 (9.62e+03,1.09e+04] 1265 0.0975 0.115 0.133 0.200  
## 3 (1.09e+04,1.19e+04] 1264 0.0879 0.105 0.122 0.200  
## 4 (1.19e+04,1.33e+04] 1264 0.0818 0.0982 0.115 0.200  
## 5 (1.33e+04, Inf] 1265 0.0618 0.0764 0.0910 0.200

p3 <- PROPEN2 %>%   
 ggplot(aes(Renta\_neta\_persona\_t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p4 <- PROPEN2 %>%   
 ggplot(aes(Renta\_neta\_persona\_t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança Renta neta persona")  
  
plot\_grid(p3, p4, ncol=2)



### 2.2.5. Vot a vox i índex de Gini

cor(d2$Gini\_medio,d2$VOX)

## [1] 0.07386005

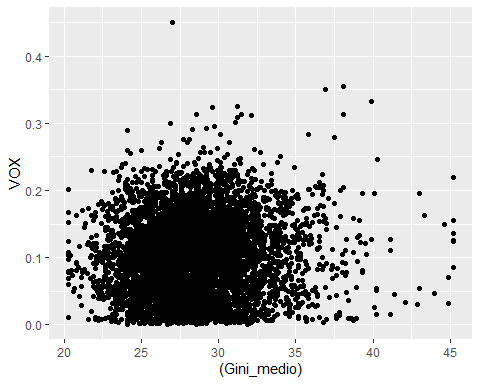
cor((d2$Gini\_medio),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.03778914

cor(log(d2$Gini\_medio),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.03996421

d2 %>% ggplot(aes(x = (Gini\_medio), y = VOX))+  
 geom\_point()

 ###Propensió gini medio

quantile(d2$Gini\_medio, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 20.3 25.8 27.4 28.8 30.7 45.2

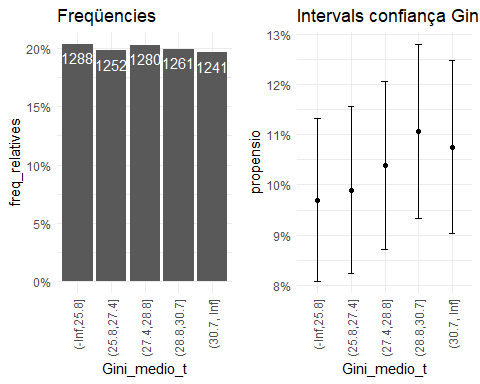
d2 <- d2 %>%   
 mutate(Gini\_medio\_t = cut(Gini\_medio, c(-Inf, 25.8, 27.4, 28.8, 30.7, Inf )))  
  
d2 %>% count(Gini\_medio\_t)

## Gini\_medio\_t n  
## 1 (-Inf,25.8] 1288  
## 2 (25.8,27.4] 1252  
## 3 (27.4,28.8] 1280  
## 4 (28.8,30.7] 1261  
## 5 (30.7, Inf] 1241

PROPEN3 <- d2 %>%   
 group\_by(Gini\_medio\_t) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN3

## # A tibble: 5 x 6  
## Gini\_medio\_t freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,25.8] 1288 0.0808 0.0970 0.113 0.204  
## 2 (25.8,27.4] 1252 0.0825 0.0990 0.116 0.198  
## 3 (27.4,28.8] 1280 0.0871 0.104 0.121 0.202  
## 4 (28.8,30.7] 1261 0.0933 0.111 0.128 0.199  
## 5 (30.7, Inf] 1241 0.0903 0.108 0.125 0.196

p5 <- PROPEN3 %>%   
 ggplot(aes(Gini\_medio\_t , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p6 <- PROPEN3 %>%   
 ggplot(aes(Gini\_medio\_t, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança Gini mitjà")  
  
plot\_grid(p5, p6, ncol=2)



### 2.2.6. Vot a Vox i tamany mig de la llar

cor(d2$tam\_hog\_medio,d2$VOX)

## [1] -0.02738314

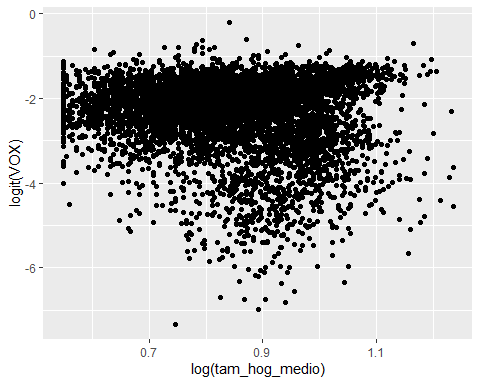
cor((d2$tam\_hog\_medio),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.1037847

cor(log(d2$tam\_hog\_medio),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] -0.1046679

d2 %>% ggplot(aes(x = log(tam\_hog\_medio), y = logit(VOX)))+  
 geom\_point()

 ###Propensió tamany mig de la llar

quantile(d2$tam\_hog\_medio, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 1.730000 2.080000 2.265085 2.430000 2.611497 3.440000

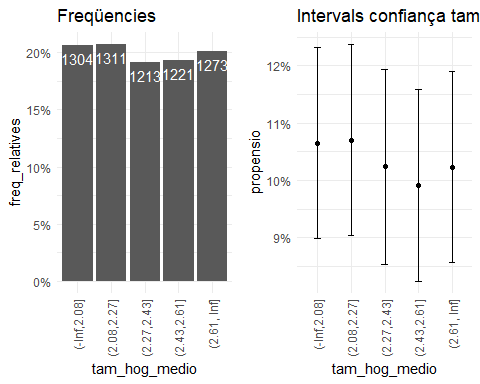
d2 <- d2 %>%   
 mutate(tam\_hog\_medio = cut(tam\_hog\_medio, c(-Inf, 2.08, 2.27, 2.43, 2.61, Inf )))  
  
d2 %>% count(tam\_hog\_medio)

## tam\_hog\_medio n  
## 1 (-Inf,2.08] 1304  
## 2 (2.08,2.27] 1311  
## 3 (2.27,2.43] 1213  
## 4 (2.43,2.61] 1221  
## 5 (2.61, Inf] 1273

PROPEN7 <- d2 %>%   
 group\_by(tam\_hog\_medio) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN7

## # A tibble: 5 x 6  
## tam\_hog\_medio freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,2.08] 1304 0.0898 0.107 0.123 0.206  
## 2 (2.08,2.27] 1311 0.0903 0.107 0.124 0.207  
## 3 (2.27,2.43] 1213 0.0853 0.102 0.119 0.192  
## 4 (2.43,2.61] 1221 0.0824 0.0991 0.116 0.193  
## 5 (2.61, Inf] 1273 0.0857 0.102 0.119 0.201

p10 <- PROPEN7 %>%   
 ggplot(aes(tam\_hog\_medio , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p11 <- PROPEN7 %>%   
 ggplot(aes(tam\_hog\_medio, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança tamany hogar mitjà")  
  
plot\_grid(p10, p11, ncol=2)



### 2.2.7. Vot a Vox i llars unipersonals

class((d2$por\_hog\_uni))

## [1] "numeric"

cor(d2$por\_hog\_uni,d2$VOX)

## [1] 0.03709287

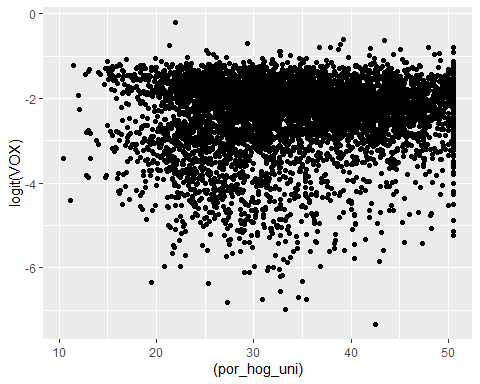
cor((d2$por\_hog\_uni),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.09487254

cor(log(d2$por\_hog\_uni),logit(d2$VOX), use = "pairwise.complete.obs")

## [1] 0.09039445

d2 %>% ggplot(aes(x = (por\_hog\_uni), y = logit(VOX)))+  
 geom\_point()

 ### 2.2.8. Propensió del vot a VOX i llars unipersonals

quantile(d2$por\_hog\_uni, seq(0,1,0.2))

## 0% 20% 40% 60% 80% 100%   
## 10.40000 26.09400 31.04068 36.37524 42.90000 50.50000

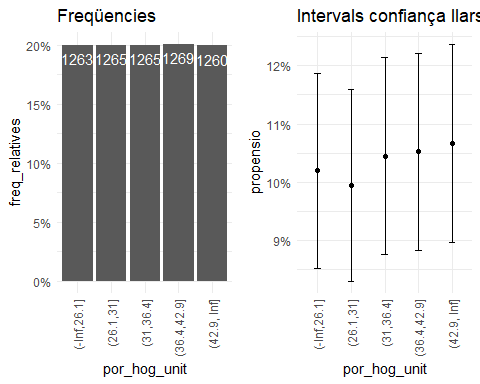
d2 <- d2 %>%   
 mutate(por\_hog\_unit = cut(por\_hog\_uni, c(-Inf, 26.09, 31.04, 36.38, 42.9, Inf )))  
  
d2 %>% count(por\_hog\_unit)

## por\_hog\_unit n  
## 1 (-Inf,26.1] 1263  
## 2 (26.1,31] 1265  
## 3 (31,36.4] 1265  
## 4 (36.4,42.9] 1269  
## 5 (42.9, Inf] 1260

PROPEN8 <- d2 %>%   
 group\_by(por\_hog\_unit) %>%   
 summarise(freq\_absolutes = n()  
 ,I\_minim = minInt\_bin(VOX )  
 ,propensio = mean(VOX )  
 ,I\_maxim = maxInt\_bin(VOX ))%>%   
 mutate(freq\_relatives = freq\_absolutes / sum(freq\_absolutes))  
PROPEN8

## # A tibble: 5 x 6  
## por\_hog\_unit freq\_absolutes I\_minim propensio I\_maxim freq\_relatives  
## <fct> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 (-Inf,26.1] 1263 0.0853 0.102 0.119 0.200  
## 2 (26.1,31] 1265 0.0830 0.0995 0.116 0.200  
## 3 (31,36.4] 1265 0.0876 0.104 0.121 0.200  
## 4 (36.4,42.9] 1269 0.0883 0.105 0.122 0.201  
## 5 (42.9, Inf] 1260 0.0896 0.107 0.124 0.199

p12 <- PROPEN8 %>%   
 ggplot(aes(por\_hog\_unit , freq\_relatives)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = freq\_absolutes), vjust=1.6, color="white")+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Freqüencies")  
  
p13 <- PROPEN8 %>%   
 ggplot(aes(por\_hog\_unit, propensio)) +  
 geom\_point(size = 1.5) +  
 geom\_errorbar(aes(ymin =I\_minim, ymax = I\_maxim, width = .2))+  
 scale\_y\_continuous(labels = percent)+  
 theme\_minimal()+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))+  
 labs(title="Intervals confiança llars unipersonals")  
  
plot\_grid(p12, p13, ncol=2)

 ## 2.3. Anàlisi multivariant

Primer fem un model amb totes les variables, per veure quines són significatives i quines no. Utilitzarem només les generals.

model0 <- lm(logit(VOX) ~ atur + euro + asia + ame + oce\_apa + afr + h + m + m19 + Renta\_neta\_persona + esp + pob\_t + UE + e20\_29 + e30\_39 + e40\_49 + e50\_59 + e60\_69 + e70\_79 + M80 + Gini\_medio + tam\_hog\_medio + por\_hog\_uni, data=d2)  
summary(model0)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur + euro + asia + ame + oce\_apa +   
## afr + h + m + m19 + Renta\_neta\_persona + esp + pob\_t + UE +   
## e20\_29 + e30\_39 + e40\_49 + e50\_59 + e60\_69 + e70\_79 + M80 +   
## Gini\_medio + tam\_hog\_medio + por\_hog\_uni, data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6012 -0.4009 0.1134 0.5006 3.0031   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.725e+00 9.381e-01 -6.103 1.11e-09 \*\*\*  
## atur 1.331e+00 3.057e-01 4.353 1.37e-05 \*\*\*  
## euro -5.775e+00 1.455e+00 -3.970 7.26e-05 \*\*\*  
## asia -3.597e+00 1.899e+00 -1.895 0.05817 .   
## ame -4.986e-01 9.160e-01 -0.544 0.58620   
## oce\_apa -1.234e+02 2.382e+01 -5.181 2.27e-07 \*\*\*  
## afr -2.355e+00 3.876e-01 -6.076 1.31e-09 \*\*\*  
## h 4.436e+00 3.883e-01 11.425 < 2e-16 \*\*\*  
## m NA NA NA NA   
## m19 -2.152e+00 1.156e+00 -1.862 0.06269 .   
## Renta\_neta\_persona -1.503e-04 6.145e-06 -24.451 < 2e-16 \*\*\*  
## esp NA NA NA NA   
## pob\_t 9.484e-07 1.845e-07 5.139 2.84e-07 \*\*\*  
## UE 6.791e+00 1.514e+00 4.485 7.42e-06 \*\*\*  
## e20\_29 3.414e+00 1.172e+00 2.914 0.00358 \*\*   
## e30\_39 6.292e+00 9.759e-01 6.447 1.22e-10 \*\*\*  
## e40\_49 6.403e+00 1.077e+00 5.943 2.95e-09 \*\*\*  
## e50\_59 3.079e+00 1.070e+00 2.877 0.00403 \*\*   
## e60\_69 -4.930e-01 9.923e-01 -0.497 0.61933   
## e70\_79 6.429e-01 1.007e+00 0.639 0.52303   
## M80 2.733e+00 9.669e-01 2.826 0.00472 \*\*   
## Gini\_medio 2.298e-02 3.545e-03 6.483 9.68e-11 \*\*\*  
## tam\_hog\_medio(2.08,2.27] -1.165e-01 3.643e-02 -3.197 0.00139 \*\*   
## tam\_hog\_medio(2.27,2.43] -2.154e-01 4.771e-02 -4.516 6.42e-06 \*\*\*  
## tam\_hog\_medio(2.43,2.61] -3.069e-01 5.898e-02 -5.204 2.01e-07 \*\*\*  
## tam\_hog\_medio(2.61, Inf] -2.361e-01 7.345e-02 -3.215 0.00131 \*\*   
## por\_hog\_uni -9.168e-03 2.833e-03 -3.237 0.00122 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.756 on 6297 degrees of freedom  
## Multiple R-squared: 0.2461, Adjusted R-squared: 0.2432   
## F-statistic: 85.64 on 24 and 6297 DF, p-value: < 2.2e-16

###A partir d’aquí, anem traient aquelles variables que no són significatives i afegim les dades de l’atur.

model01 <- lm(logit(VOX) ~ atur + euro + asia + oce\_apa + afr + h + m + Renta\_neta\_persona + esp + pob\_t + UE + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d2)  
summary(model01)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur + euro + asia + oce\_apa + afr +   
## h + m + Renta\_neta\_persona + esp + pob\_t + UE + e20\_29 +   
## e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6118 -0.3974 0.1170 0.4926 3.0665   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.345e+00 9.699e-01 -7.573 4.17e-14 \*\*\*  
## atur 1.557e+00 3.021e-01 5.153 2.64e-07 \*\*\*  
## euro -5.077e+00 1.791e+00 -2.835 0.00459 \*\*   
## asia -3.402e+00 2.245e+00 -1.515 0.12981   
## oce\_apa -1.290e+02 2.394e+01 -5.389 7.35e-08 \*\*\*  
## afr -2.142e+00 1.018e+00 -2.104 0.03543 \*   
## h 5.008e+00 3.546e-01 14.121 < 2e-16 \*\*\*  
## m NA NA NA NA   
## Renta\_neta\_persona -1.493e-04 5.985e-06 -24.948 < 2e-16 \*\*\*  
## esp 7.201e-01 9.129e-01 0.789 0.43024   
## pob\_t 9.910e-07 1.852e-07 5.351 9.05e-08 \*\*\*  
## UE 6.985e+00 1.520e+00 4.595 4.41e-06 \*\*\*  
## e20\_29 3.137e+00 5.444e-01 5.762 8.69e-09 \*\*\*  
## e30\_39 5.352e+00 5.128e-01 10.436 < 2e-16 \*\*\*  
## e40\_49 5.569e+00 4.709e-01 11.827 < 2e-16 \*\*\*  
## e50\_59 3.299e+00 4.392e-01 7.512 6.64e-14 \*\*\*  
## M80 3.759e+00 3.527e-01 10.659 < 2e-16 \*\*\*  
## Gini\_medio 2.206e-02 3.464e-03 6.366 2.07e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7598 on 6305 degrees of freedom  
## Multiple R-squared: 0.2375, Adjusted R-squared: 0.2355   
## F-statistic: 122.7 on 16 and 6305 DF, p-value: < 2.2e-16

model02 <- lm(logit(VOX) ~ atur + euro + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + UE + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d2)  
summary(model02)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur + euro + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + UE + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6012 -0.3981 0.1166 0.4933 3.0549   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.608e+00 2.714e-01 -24.344 < 2e-16 \*\*\*  
## atur 1.559e+00 3.019e-01 5.164 2.50e-07 \*\*\*  
## euro -6.578e+00 1.426e+00 -4.613 4.05e-06 \*\*\*  
## oce\_apa -1.309e+02 2.389e+01 -5.479 4.45e-08 \*\*\*  
## afr -2.958e+00 3.750e-01 -7.887 3.63e-15 \*\*\*  
## h 5.067e+00 3.525e-01 14.373 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.507e-04 5.935e-06 -25.396 < 2e-16 \*\*\*  
## pob\_t 9.086e-07 1.820e-07 4.993 6.12e-07 \*\*\*  
## UE 7.732e+00 1.490e+00 5.190 2.17e-07 \*\*\*  
## e20\_29 3.028e+00 5.419e-01 5.588 2.39e-08 \*\*\*  
## e30\_39 5.271e+00 5.093e-01 10.351 < 2e-16 \*\*\*  
## e40\_49 5.500e+00 4.671e-01 11.776 < 2e-16 \*\*\*  
## e50\_59 3.354e+00 4.386e-01 7.647 2.36e-14 \*\*\*  
## M80 3.760e+00 3.524e-01 10.671 < 2e-16 \*\*\*  
## Gini\_medio 2.127e-02 3.412e-03 6.234 4.83e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7601 on 6307 degrees of freedom  
## Multiple R-squared: 0.2367, Adjusted R-squared: 0.235   
## F-statistic: 139.7 on 14 and 6307 DF, p-value: < 2.2e-16

model03 <- lm(logit(VOX) ~ atur + atur\_m24 + atur\_25\_44 + atur\_M45 + atur\_h + atur\_d + euro + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + UE + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d2)  
summary(model03)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur + atur\_m24 + atur\_25\_44 + atur\_M45 +   
## atur\_h + atur\_d + euro + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + UE + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5816 -0.3843 0.1051 0.4844 3.0694   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.613e+00 2.860e-01 -23.125 < 2e-16 \*\*\*  
## atur 1.915e+01 5.499e+00 3.483 0.000500 \*\*\*  
## atur\_m24 1.675e+00 3.625e-01 4.620 3.91e-06 \*\*\*  
## atur\_25\_44 -1.824e-01 1.003e+00 -0.182 0.855714   
## atur\_M45 1.176e+00 1.469e+00 0.801 0.423324   
## atur\_h -1.403e+01 2.624e+00 -5.348 9.20e-08 \*\*\*  
## atur\_d -7.180e+00 2.021e+00 -3.553 0.000384 \*\*\*  
## euro -6.329e+00 1.416e+00 -4.469 7.99e-06 \*\*\*  
## oce\_apa -1.277e+02 2.369e+01 -5.390 7.28e-08 \*\*\*  
## afr -2.946e+00 3.731e-01 -7.896 3.36e-15 \*\*\*  
## h 5.108e+00 3.785e-01 13.495 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.514e-04 5.923e-06 -25.555 < 2e-16 \*\*\*  
## pob\_t 9.609e-07 1.806e-07 5.320 1.07e-07 \*\*\*  
## UE 7.555e+00 1.479e+00 5.109 3.34e-07 \*\*\*  
## e20\_29 2.339e+00 5.442e-01 4.298 1.75e-05 \*\*\*  
## e30\_39 4.961e+00 5.063e-01 9.798 < 2e-16 \*\*\*  
## e40\_49 5.272e+00 4.651e-01 11.334 < 2e-16 \*\*\*  
## e50\_59 3.546e+00 4.376e-01 8.104 6.36e-16 \*\*\*  
## M80 3.785e+00 3.555e-01 10.648 < 2e-16 \*\*\*  
## Gini\_medio 2.535e-02 3.409e-03 7.438 1.16e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7535 on 6302 degrees of freedom  
## Multiple R-squared: 0.2505, Adjusted R-squared: 0.2482   
## F-statistic: 110.9 on 19 and 6302 DF, p-value: < 2.2e-16

model04 <- lm(logit(VOX) ~ atur + atur\_m24 + atur\_h + atur\_d + euro + oce\_apa + afr + h + Renta\_neta\_persona + UE + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d2)  
summary(model04)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur + atur\_m24 + atur\_h + atur\_d +   
## euro + oce\_apa + afr + h + Renta\_neta\_persona + UE + pob\_t +   
## e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5801 -0.3831 0.1065 0.4851 3.0711   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.663e+00 2.851e-01 -23.372 < 2e-16 \*\*\*  
## atur 2.121e+01 4.556e+00 4.655 3.30e-06 \*\*\*  
## atur\_m24 1.449e+00 2.543e-01 5.698 1.26e-08 \*\*\*  
## atur\_h -1.445e+01 2.609e+00 -5.539 3.17e-08 \*\*\*  
## atur\_d -7.464e+00 2.015e+00 -3.704 0.000214 \*\*\*  
## euro -6.181e+00 1.414e+00 -4.370 1.26e-05 \*\*\*  
## oce\_apa -1.277e+02 2.369e+01 -5.390 7.30e-08 \*\*\*  
## afr -2.908e+00 3.721e-01 -7.815 6.42e-15 \*\*\*  
## h 5.158e+00 3.775e-01 13.665 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.500e-04 5.893e-06 -25.459 < 2e-16 \*\*\*  
## UE 7.421e+00 1.478e+00 5.022 5.26e-07 \*\*\*  
## pob\_t 9.653e-07 1.806e-07 5.347 9.28e-08 \*\*\*  
## e20\_29 2.365e+00 5.412e-01 4.371 1.26e-05 \*\*\*  
## e30\_39 4.994e+00 5.062e-01 9.866 < 2e-16 \*\*\*  
## e40\_49 5.331e+00 4.637e-01 11.497 < 2e-16 \*\*\*  
## e50\_59 3.464e+00 4.355e-01 7.956 2.10e-15 \*\*\*  
## M80 3.841e+00 3.536e-01 10.863 < 2e-16 \*\*\*  
## Gini\_medio 2.507e-02 3.403e-03 7.367 1.97e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7537 on 6304 degrees of freedom  
## Multiple R-squared: 0.2499, Adjusted R-squared: 0.2479   
## F-statistic: 123.6 on 17 and 6304 DF, p-value: < 2.2e-16

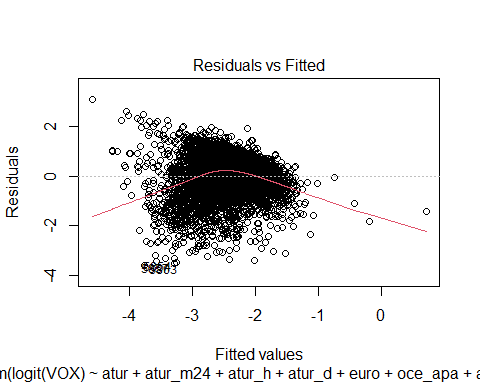
###Tot sembla indicar que el millor model on totes les variables introduides són significatives i amb un adjust de 24’99% és el model 04. Les variables escollides són les que presentaven una major correlació amb la variable vot a VOX.

model05 <- lm(VOX ~ atur + atur\_m24 + atur\_h + atur\_d + euro + oce\_apa + afr + h + Renta\_neta\_persona + UE + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d2)  
summary(model05)

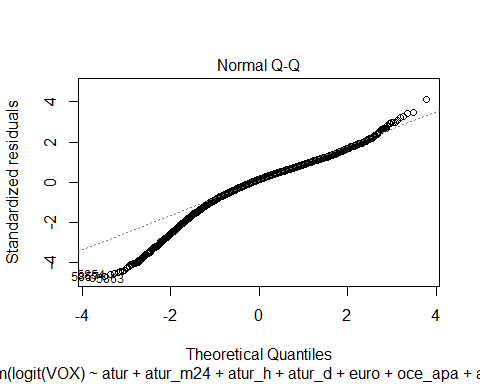
##   
## Call:  
## lm(formula = VOX ~ atur + atur\_m24 + atur\_h + atur\_d + euro +   
## oce\_apa + afr + h + Renta\_neta\_persona + UE + pob\_t + e20\_29 +   
## e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.15573 -0.03782 -0.00317 0.03140 0.32571   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.183e-01 1.905e-02 -11.458 < 2e-16 \*\*\*  
## atur 1.301e+00 3.045e-01 4.271 1.97e-05 \*\*\*  
## atur\_m24 8.839e-02 1.700e-02 5.200 2.06e-07 \*\*\*  
## atur\_h -9.686e-01 1.744e-01 -5.554 2.90e-08 \*\*\*  
## atur\_d -4.089e-01 1.347e-01 -3.036 0.002408 \*\*   
## euro -4.651e-01 9.453e-02 -4.920 8.86e-07 \*\*\*  
## oce\_apa -5.966e+00 1.584e+00 -3.767 0.000167 \*\*\*  
## afr -1.239e-01 2.487e-02 -4.984 6.41e-07 \*\*\*  
## h 3.791e-01 2.523e-02 15.026 < 2e-16 \*\*\*  
## Renta\_neta\_persona -7.255e-06 3.939e-07 -18.418 < 2e-16 \*\*\*  
## UE 5.562e-01 9.876e-02 5.632 1.86e-08 \*\*\*  
## pob\_t 4.186e-08 1.207e-08 3.469 0.000526 \*\*\*  
## e20\_29 1.578e-01 3.617e-02 4.362 1.31e-05 \*\*\*  
## e30\_39 3.435e-01 3.383e-02 10.154 < 2e-16 \*\*\*  
## e40\_49 3.315e-01 3.099e-02 10.697 < 2e-16 \*\*\*  
## e50\_59 1.746e-01 2.910e-02 5.998 2.11e-09 \*\*\*  
## M80 1.850e-01 2.363e-02 7.829 5.71e-15 \*\*\*  
## Gini\_medio 1.863e-03 2.275e-04 8.191 3.12e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05037 on 6304 degrees of freedom  
## Multiple R-squared: 0.1976, Adjusted R-squared: 0.1955   
## F-statistic: 91.33 on 17 and 6304 DF, p-value: < 2.2e-16

###El model 05 és el mateix, pero sense transformar la variable VOX a logit. Es pot apreciar que l’adjust cau força. ###A continuació realitzem una revisió dels residus de la regressió:

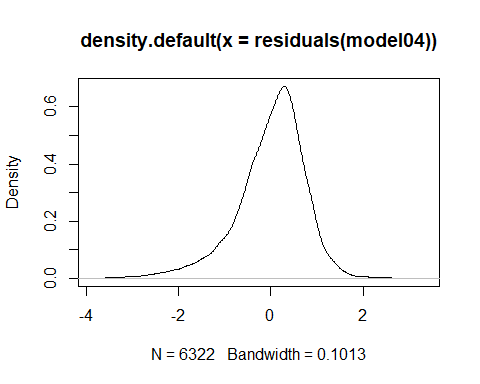
plot(model04,1)



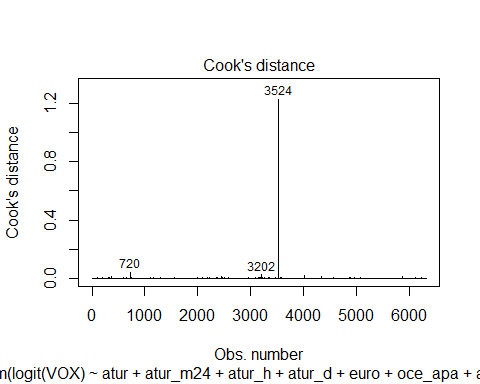
plot(model04,2)



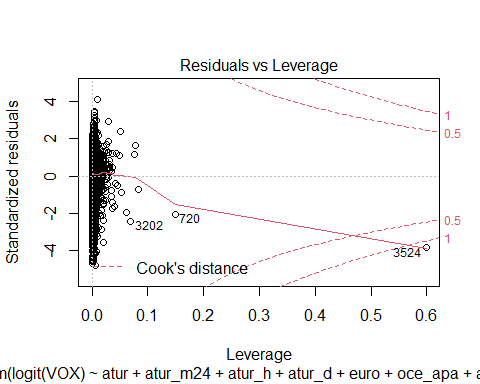
plot(density(residuals(model04)))



plot(model04,4)



plot(model04,5)

 ###Es pot observar una gran heterocedasticitat amb outliers que provoquen una cua important a l’esquerra. Per tant, intentarem fer el model més fiable traient aquests outliers. Abans utilitzarem el factor d’inflació de varianza per detectar colinealitats.

vif(model04)

## atur atur\_m24 atur\_h atur\_d   
## 403.510894 1.516618 94.367698 150.296126   
## euro oce\_apa afr h   
## 75.268657 1.030434 1.237889 1.477987   
## Renta\_neta\_persona UE pob\_t e20\_29   
## 1.892231 74.375277 1.042703 1.803918   
## e30\_39 e40\_49 e50\_59 M80   
## 2.344595 2.542211 1.481927 4.729273   
## Gini\_medio   
## 1.285293

###D’entrada, es pot observar que l’atur, atur d’homes, atur de dones i població europea junt amb UE presenten un registre molt elevat, i per tant, les traurem del model.

model06 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d2)  
summary(model06)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6227 -0.3940 0.1165 0.4929 3.1285   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.427e+00 2.612e-01 -24.604 < 2e-16 \*\*\*  
## atur\_m24 1.438e+00 2.238e-01 6.424 1.43e-10 \*\*\*  
## oce\_apa -1.270e+02 2.389e+01 -5.318 1.08e-07 \*\*\*  
## afr -3.120e+00 3.702e-01 -8.428 < 2e-16 \*\*\*  
## h 4.929e+00 3.491e-01 14.118 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.620e-04 5.020e-06 -32.274 < 2e-16 \*\*\*  
## pob\_t 9.005e-07 1.817e-07 4.956 7.38e-07 \*\*\*  
## e20\_29 2.856e+00 5.428e-01 5.262 1.47e-07 \*\*\*  
## e30\_39 5.214e+00 5.101e-01 10.222 < 2e-16 \*\*\*  
## e40\_49 5.422e+00 4.672e-01 11.604 < 2e-16 \*\*\*  
## e50\_59 3.345e+00 4.383e-01 7.631 2.68e-14 \*\*\*  
## M80 3.522e+00 3.480e-01 10.121 < 2e-16 \*\*\*  
## Gini\_medio 2.727e-02 3.177e-03 8.584 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7611 on 6309 degrees of freedom  
## Multiple R-squared: 0.2344, Adjusted R-squared: 0.2329   
## F-statistic: 160.9 on 12 and 6309 DF, p-value: < 2.2e-16

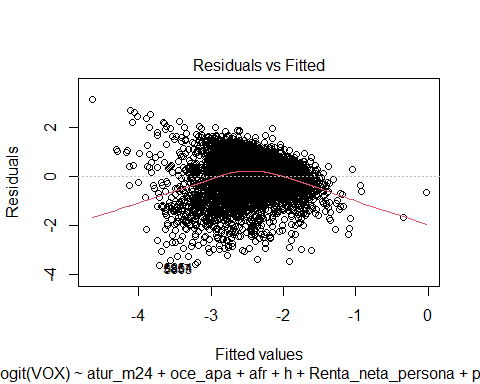
###L’adjust ens baixa una mica, però totes les variables continuen sent significatives.

vif(model06)

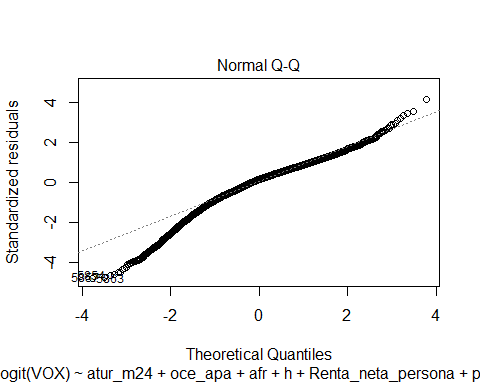
## atur\_m24 oce\_apa afr h   
## 1.151776 1.027042 1.201430 1.239657   
## Renta\_neta\_persona pob\_t e20\_29 e30\_39   
## 1.346041 1.035429 1.779052 2.334623   
## e40\_49 e50\_59 M80 Gini\_medio   
## 2.530596 1.472063 4.491197 1.098123

### En el model 06 no tenim problemes de multicolinealitat.

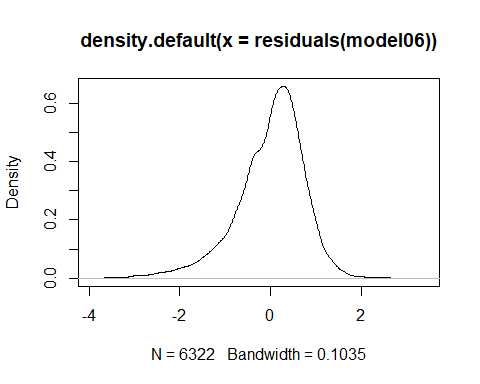
plot(model06,1)



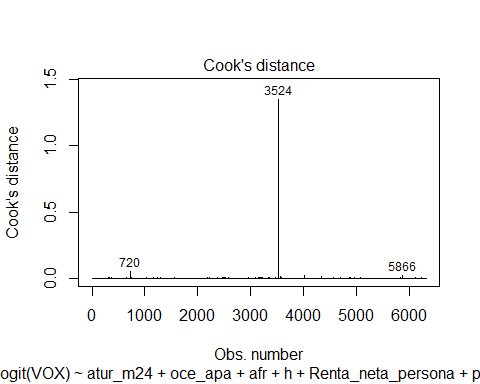
plot(model06,2)



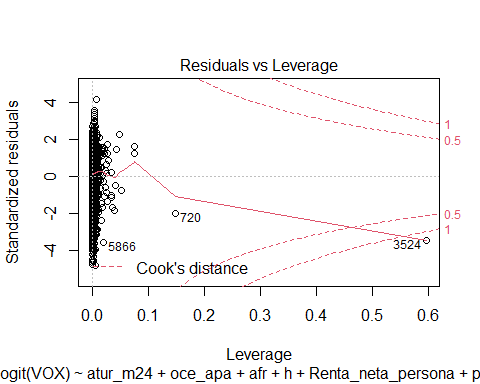
plot(density(residuals(model06)))



plot(model06,4)



plot(model06,5)

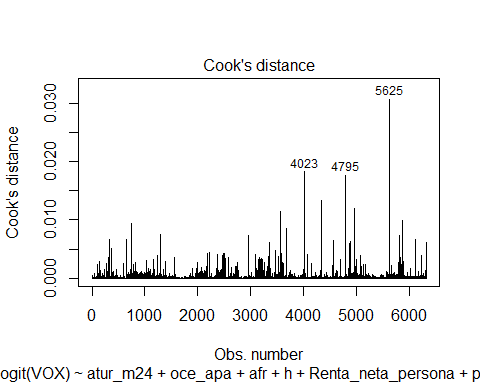
 ###Per contra, els residus continuen tenint el mateix problema. Anem a modelar els diferents outliers.

d4 <- d2  
d4=d4[-c(3524,720,5866),]

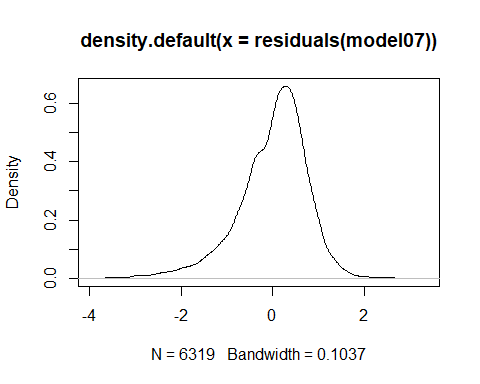
model07 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model07)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6258 -0.3974 0.1157 0.4917 3.0259   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.528e+00 2.610e-01 -25.010 < 2e-16 \*\*\*  
## atur\_m24 1.407e+00 2.232e-01 6.302 3.13e-10 \*\*\*  
## oce\_apa -1.139e+02 2.405e+01 -4.738 2.21e-06 \*\*\*  
## afr -3.175e+00 3.692e-01 -8.600 < 2e-16 \*\*\*  
## h 5.163e+00 3.506e-01 14.727 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.633e-04 5.012e-06 -32.582 < 2e-16 \*\*\*  
## pob\_t 2.589e-06 3.569e-07 7.256 4.47e-13 \*\*\*  
## e20\_29 2.789e+00 5.412e-01 5.153 2.64e-07 \*\*\*  
## e30\_39 5.210e+00 5.085e-01 10.246 < 2e-16 \*\*\*  
## e40\_49 5.416e+00 4.657e-01 11.629 < 2e-16 \*\*\*  
## e50\_59 3.386e+00 4.370e-01 7.748 1.08e-14 \*\*\*  
## M80 3.612e+00 3.472e-01 10.403 < 2e-16 \*\*\*  
## Gini\_medio 2.647e-02 3.169e-03 8.352 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7587 on 6306 degrees of freedom  
## Multiple R-squared: 0.2372, Adjusted R-squared: 0.2357   
## F-statistic: 163.4 on 12 and 6306 DF, p-value: < 2.2e-16

plot(model07,4)



plot(density(residuals(model07)))



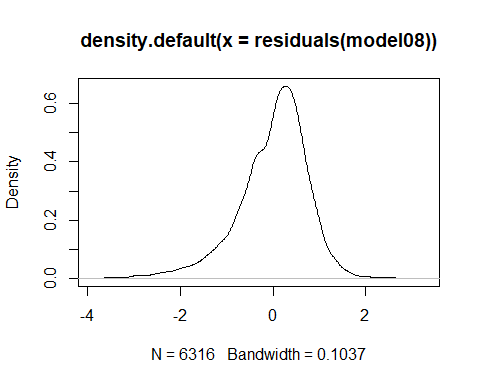
###Traiem els nous outliers. A partir d’ara ens centrarem en anar traient observacions atípiques del model per tal de reduir la heterocedasticitat.

d4 <- d2  
d4=d4[-c(3524,720,5866,4023,4795,5625),]

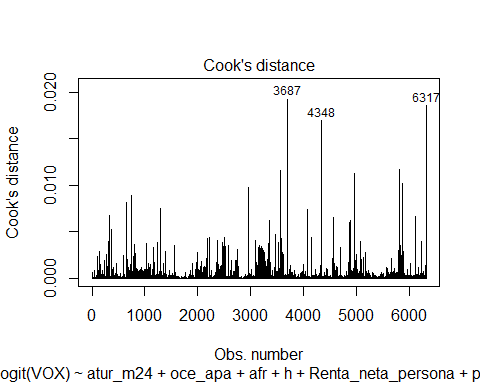
model08 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model08)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5989 -0.3973 0.1147 0.4916 2.9960   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.540e+00 2.611e-01 -25.053 < 2e-16 \*\*\*  
## atur\_m24 1.406e+00 2.231e-01 6.300 3.18e-10 \*\*\*  
## oce\_apa -1.252e+02 2.464e+01 -5.082 3.85e-07 \*\*\*  
## afr -3.191e+00 3.691e-01 -8.644 < 2e-16 \*\*\*  
## h 5.187e+00 3.512e-01 14.769 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.634e-04 5.012e-06 -32.606 < 2e-16 \*\*\*  
## pob\_t 3.037e-06 4.070e-07 7.463 9.59e-14 \*\*\*  
## e20\_29 2.785e+00 5.410e-01 5.147 2.73e-07 \*\*\*  
## e30\_39 5.183e+00 5.083e-01 10.197 < 2e-16 \*\*\*  
## e40\_49 5.425e+00 4.656e-01 11.653 < 2e-16 \*\*\*  
## e50\_59 3.390e+00 4.369e-01 7.760 9.83e-15 \*\*\*  
## M80 3.626e+00 3.470e-01 10.449 < 2e-16 \*\*\*  
## Gini\_medio 2.645e-02 3.168e-03 8.350 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7583 on 6303 degrees of freedom  
## Multiple R-squared: 0.2381, Adjusted R-squared: 0.2367   
## F-statistic: 164.2 on 12 and 6303 DF, p-value: < 2.2e-16

plot(density(residuals(model08)))



plot(model08,4)

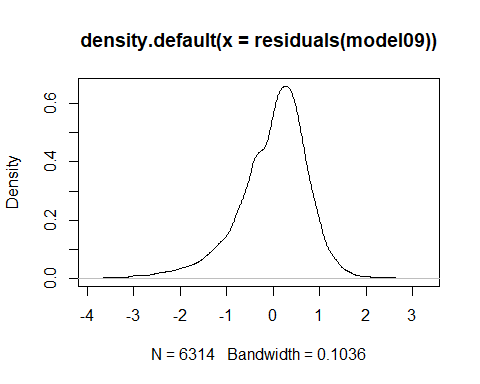


d4 <- d2  
d4=d4[-c(3524,720,5866,4023,4795,5625,3687,4348,5866),]

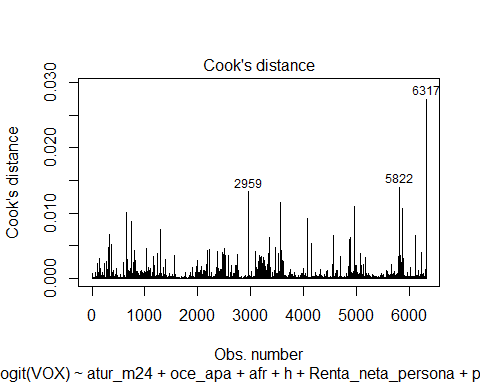
model09 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model09)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5806 -0.3962 0.1138 0.4913 2.9836   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.538e+00 2.611e-01 -25.037 < 2e-16 \*\*\*  
## atur\_m24 1.391e+00 2.232e-01 6.233 4.87e-10 \*\*\*  
## oce\_apa -1.365e+02 2.568e+01 -5.315 1.10e-07 \*\*\*  
## afr -3.199e+00 3.691e-01 -8.668 < 2e-16 \*\*\*  
## h 5.196e+00 3.515e-01 14.779 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.636e-04 5.013e-06 -32.629 < 2e-16 \*\*\*  
## pob\_t 3.237e-06 4.264e-07 7.590 3.67e-14 \*\*\*  
## e20\_29 2.776e+00 5.409e-01 5.132 2.95e-07 \*\*\*  
## e30\_39 5.164e+00 5.083e-01 10.158 < 2e-16 \*\*\*  
## e40\_49 5.418e+00 4.655e-01 11.640 < 2e-16 \*\*\*  
## e50\_59 3.390e+00 4.368e-01 7.762 9.72e-15 \*\*\*  
## M80 3.616e+00 3.471e-01 10.418 < 2e-16 \*\*\*  
## Gini\_medio 2.647e-02 3.168e-03 8.357 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7581 on 6301 degrees of freedom  
## Multiple R-squared: 0.2387, Adjusted R-squared: 0.2372   
## F-statistic: 164.6 on 12 and 6301 DF, p-value: < 2.2e-16

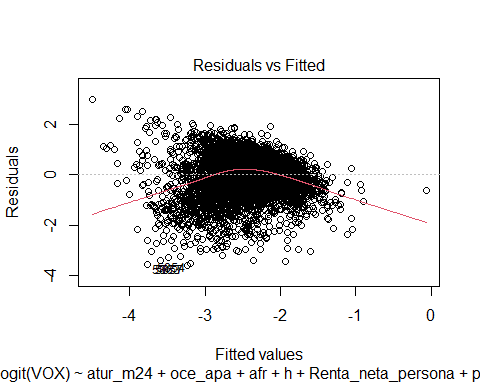
plot(density(residuals(model09)))



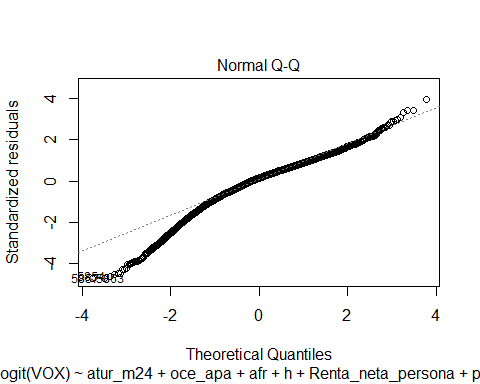
plot(model09,4)



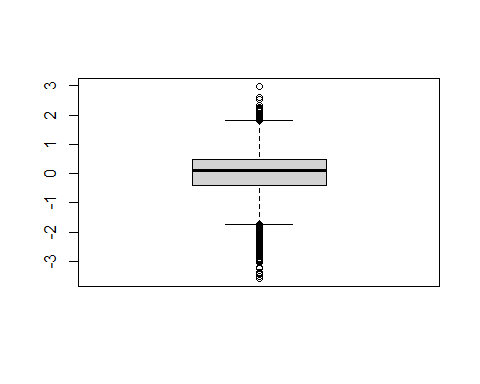
plot(model09,1)



plot(model09,2)

 ###Encara tenim un seriós problema d’heterocedasticitat.

outliers<-boxplot(model09$residuals)



outliers$out

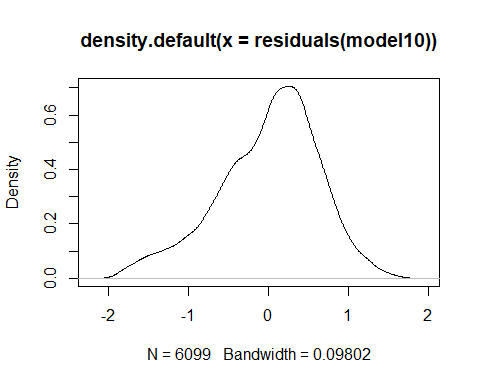
## 56 408 484 588 712 724 736 739   
## -2.008674 -1.991315 -1.753062 -1.872503 -1.924539 -2.493088 -2.325423 -2.210058   
## 775 797 824 829 831 834 843 936   
## -1.950399 -2.454395 -2.126791 -3.041865 -2.320865 -1.829373 -1.883199 -1.952595   
## 940 972 1112 1169 1299 1405 1466 1993   
## -2.001135 -1.786375 1.960141 2.060881 -1.767335 -1.902761 1.829808 -2.406042   
## 1994 2018 2041 2048 2054 2058 2062 2068   
## -1.904841 -1.801853 -2.005511 -1.913675 -1.785727 -2.049914 -1.757800 -1.751231   
## 2075 2086 2090 2097 2098 2117 2127 2150   
## -1.739187 -1.772712 -2.040661 -1.833781 -2.607517 -1.930990 -2.077104 -2.273873   
## 2157 2162 2171 2173 2178 2221 2368 2376   
## -2.905485 -2.097278 -2.187163 -2.888888 -2.125733 -2.375190 1.850758 1.974584   
## 2466 2468 2469 2471 2472 2474 2476 2477   
## -3.242317 -2.078863 -2.125439 -2.970011 -2.543138 -2.178562 -1.985998 -3.257647   
## 2478 2479 2480 2481 2482 2490 2491 2492   
## -2.451143 -1.997680 -2.908626 -2.779123 -2.003776 -2.399816 -2.598719 -1.785837   
## 2494 2498 2499 2500 2502 2503 2505 2507   
## -3.047256 -3.052718 -1.808629 -2.572192 -1.912672 -1.955119 -2.834604 -2.149841   
## 2508 2510 2514 2516 2519 2520 2527 2587   
## -1.789872 -1.938583 -1.804950 -3.390390 -1.980192 -2.005120 -2.245584 -3.454850   
## 2633 2643 2715 2725 3082 3085 3092 3095   
## 2.048705 1.943959 2.190977 1.875057 -2.821544 -2.432554 -2.723154 -1.835894   
## 3097 3100 3108 3112 3114 3120 3123 3130   
## -2.973498 -2.283750 -2.836698 -3.004227 -2.128112 -1.810026 -2.218147 -1.853189   
## 3131 3134 3136 3137 3141 3142 3155 3158   
## -1.766479 -2.525580 -1.809401 -2.050728 -2.271170 -2.616400 -2.163678 -1.760667   
## 3159 3165 3166 3168 3169 3178 3181 3182   
## -2.019022 -1.863040 -2.362914 -2.908840 -1.928267 -2.371520 -2.249808 -2.948216   
## 3183 3198 3202 3207 3208 3211 3217 3220   
## -1.960345 -2.149198 -2.513372 -2.613707 -2.282529 -2.238062 -2.445156 -3.399175   
## 3221 3225 3227 3230 3231 3239 3245 3249   
## -1.773332 -1.749313 -2.293923 -2.233312 -2.499134 -2.361738 -1.852810 -2.006036   
## 3265 3266 3270 3295 3324 3348 3358 3371   
## -2.571606 -2.678151 -2.459933 -1.905538 -2.283493 -2.043198 -3.051789 -1.849752   
## 3471 3525 3556 3565 3586 3603 3610 3615   
## 2.591625 2.323565 2.983603 2.170155 2.590151 2.278013 2.008779 2.183964   
## 3620 3797 4563 4702 4873 4888 4896 4907   
## 1.921871 -1.730128 -2.182691 2.238095 -2.373923 -2.976121 -1.874967 -1.800732   
## 4918 4927 4934 4949 4960 4963 4966 4968   
## -2.361077 -1.928767 -2.038323 -2.161212 -1.942828 -2.231519 -1.936873 -1.949143   
## 4969 4971 4977 4978 4979 4983 5005 5012   
## -2.113689 -2.108801 -2.950868 -2.323409 -1.979862 -2.352097 -1.812178 -1.803637   
## 5015 5028 5035 5661 5713 5793 5810 5811   
## -2.056929 -2.128595 -1.762735 1.858411 1.962768 1.935536 -2.542707 -2.288027   
## 5812 5814 5819 5820 5823 5824 5826 5827   
## -2.668941 -1.821512 -2.848580 -2.416910 -2.669647 -1.976802 -2.660365 -2.437122   
## 5828 5830 5832 5834 5837 5838 5846 5847   
## -2.010941 -1.749531 -2.728980 -2.127275 -2.526544 -2.586215 -1.893564 -2.258359   
## 5848 5849 5852 5853 5854 5855 5856 5857   
## -2.377197 -1.904770 -2.047843 -2.009532 -3.508915 -2.094426 -2.934171 -3.580632   
## 5860 5861 5862 5863 5867 5873 5881 5884   
## -2.447896 -2.204589 -1.794497 -3.574204 -2.503981 -3.209515 -2.707160 -2.347612   
## 5886 5888 5895 5896 6109 6235 6277   
## -1.906808 -1.937545 -1.739724 -1.866960 -2.142059 2.520019 -1.801886

d4 <- d2  
d4=d4[-c(3524,720,5866,4023,4795,5625,3687,4348,5866,56,408,484,588,712,724,736,739,775, 797,824,829,831,834,843,936,940,972,1112,1169,1299,1405,1466,1993,1994,2018,2041,2048,2054,2058,2062,2068,2075,2086,2090,2097,2098,2117,2127,2150,2157,2162,2171,2173,2178,2221,2368,2376,2466,2468,2469,2471,2472,2474,2476,2477,2478,2479,2480,2481,2482,2490,2491,2492,2494,2498,2499,2500,2502,2503,2505,2507,2508,2510,2514,2516,2519,2520,2527,2587,2633,2643,2715,2725,3082,3085,3092,3095,3097,3100,3108,3112,3114,3120,3123,3130,3131,3134,3136,3137,3141,3142,3155,3158,3159,3165,3166,3168,3169,3178,3181,3182,3183,3198,3202,3207,3208,3211,3217,3220,3221,3225,3227,3230,3231,3239,3245,3249,3265,3266,3270,3295,3324,3348,3358,3371,3471,3525,3556,3565,3586,3603,3610,3615,3620,3797,4563,4702,4873,4888,4896,4907,4918,4927,4934,4949,4960,4963,4966,4968,4969,4971,4977,4978,4979,4983,5005,5012,5015,5028,5035,5661,5713,5793,5810,5811,5812,5814,5819,5820,5823,5824,5826,5827,5828,5830,5832,5834,5837,5838,5846,5847,5848,5849,5852,5853,5854,5855,5856,5857,5860,5861,5862,5863,5867,5873,5881,5884,5886,5888,5895,5896,6109,6235,6277),]

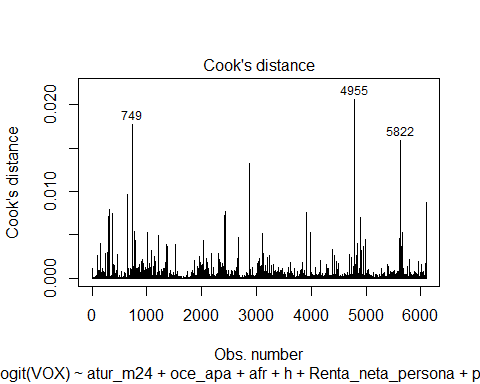
model10 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model10)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.91417 -0.39657 0.08589 0.43753 1.68182   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.712e+00 2.235e-01 -30.030 < 2e-16 \*\*\*  
## atur\_m24 1.277e+00 1.900e-01 6.722 1.95e-11 \*\*\*  
## oce\_apa -1.361e+02 2.177e+01 -6.254 4.27e-10 \*\*\*  
## afr -2.726e+00 3.130e-01 -8.708 < 2e-16 \*\*\*  
## h 5.566e+00 3.004e-01 18.529 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.416e-04 4.422e-06 -32.033 < 2e-16 \*\*\*  
## pob\_t 2.215e-06 3.581e-07 6.187 6.54e-10 \*\*\*  
## e20\_29 2.039e+00 4.617e-01 4.416 1.02e-05 \*\*\*  
## e30\_39 5.090e+00 4.332e-01 11.751 < 2e-16 \*\*\*  
## e40\_49 4.938e+00 3.973e-01 12.428 < 2e-16 \*\*\*  
## e50\_59 3.568e+00 3.720e-01 9.593 < 2e-16 \*\*\*  
## M80 2.846e+00 2.957e-01 9.626 < 2e-16 \*\*\*  
## Gini\_medio 2.643e-02 2.728e-03 9.686 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6336 on 6086 degrees of freedom  
## Multiple R-squared: 0.2507, Adjusted R-squared: 0.2492   
## F-statistic: 169.7 on 12 and 6086 DF, p-value: < 2.2e-16

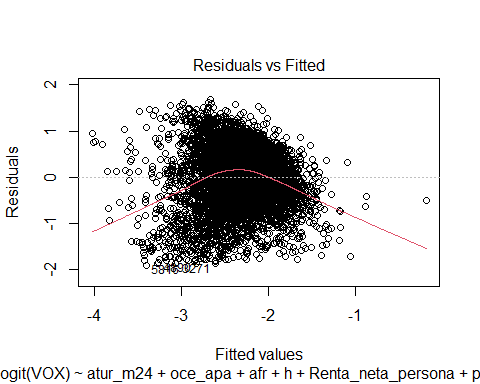
plot(density(residuals(model10)))



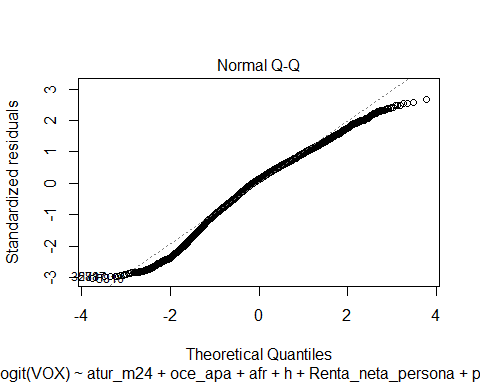
plot(model10,4)



plot(model10,1)



plot(model10,2)



vif(model10)

## atur\_m24 oce\_apa afr h   
## 1.148985 1.030979 1.203478 1.264662   
## Renta\_neta\_persona pob\_t e20\_29 e30\_39   
## 1.351582 1.101557 1.800902 2.363068   
## e40\_49 e50\_59 M80 Gini\_medio   
## 2.567086 1.474970 4.588795 1.100276

library(tseries)

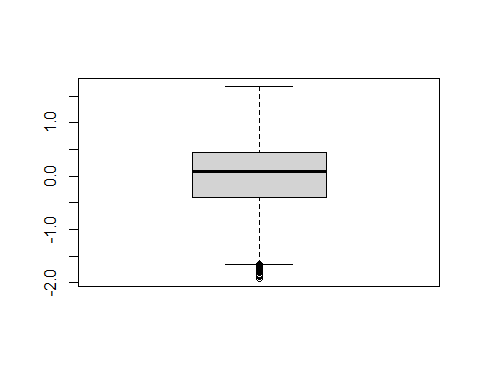
## Warning: package 'tseries' was built under R version 4.1.3

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

jarque.bera.test(model10$residuals)

##   
## Jarque Bera Test  
##   
## data: model10$residuals  
## X-squared = 242.94, df = 2, p-value < 2.2e-16

outliers<-boxplot(model10$residuals)



outliers$out

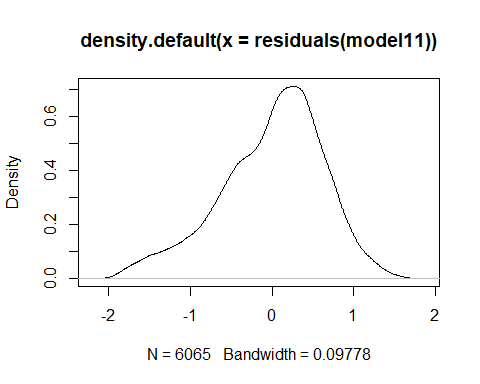
## 391 547 748 751 760 783 836 852   
## -1.793784 -1.665990 -1.648191 -1.701178 -1.793513 -1.847486 -1.781671 -1.805348   
## 862 957 987 1999 2004 2006 2038 2061   
## -1.785511 -1.803915 -1.744140 -1.780542 -1.751379 -1.728622 -1.876181 -1.788600   
## 2064 2081 2107 2119 2123 2124 2126 2143   
## -1.699379 -1.689232 -1.723650 -1.763017 -1.717237 -1.657222 -1.788643 -1.760984   
## 2149 2174 2177 2470 2475 2486 2487 2497   
## -1.773615 -1.810185 -1.724114 -1.706220 -1.850554 -1.778753 -1.807178 -1.697128   
## 2511 2512 3099 3101 3106 3118 3132 3160   
## -1.654146 -1.655856 -1.733247 -1.797350 -1.800336 -1.755970 -1.712189 -1.684935   
## 3161 3164 3200 3228 3256 3271 3650 4438   
## -1.844109 -1.773038 -1.673918 -1.664620 -1.832185 -1.887433 -1.720376 -1.703121   
## 4547 4575 4890 4917 4941 4974 5808 5816   
## -1.719699 -1.793241 -1.877679 -1.668225 -1.739550 -1.773276 -1.775770 -1.914166   
## 5833 5843 5877 5882 5887 5890 5894 5899   
## -1.811402 -1.741263 -1.676624 -1.663176 -1.873711 -1.686701 -1.736713 -1.765554

d4 <- d2  
d4=d4[-c(3524,720,5866,4023,4795,5625,3687,4348,5866,56,408,484,588,712,724,736,739,775, 797,824,829,831,834,843,936,940,972,1112,1169,1299,1405,1466,1993,1994,2018,2041,2048,2054,2058,2062,2068,2075,2086,2090,2097,2098,2117,2127,2150,2157,2162,2171,2173,2178,2221,2368,2376,2466,2468,2469,2471,2472,2474,2476,2477,2478,2479,2480,2481,2482,2490,2491,2492,2494,2498,2499,2500,2502,2503,2505,2507,2508,2510,2514,2516,2519,2520,2527,2587,2633,2643,2715,2725,3082,3085,3092,3095,3097,3100,3108,3112,3114,3120,3123,3130,3131,3134,3136,3137,3141,3142,3155,3158,3159,3165,3166,3168,3169,3178,3181,3182,3183,3198,3202,3207,3208,3211,3217,3220,3221,3225,3227,3230,3231,3239,3245,3249,3265,3266,3270,3295,3324,3348,3358,3371,3471,3525,3556,3565,3586,3603,3610,3615,3620,3797,4563,4702,4873,4888,4896,4907,4918,4927,4934,4949,4960,4963,4966,4968,4969,4971,4977,4978,4979,4983,5005,5012,5015,5028,5035,5661,5713,5793,5810,5811,5812,5814,5819,5820,5823,5824,5826,5827,5828,5830,5832,5834,5837,5838,5846,5847,5848,5849,5852,5853,5854,5855,5856,5857,5860,5861,5862,5863,5867,5873,5881,5884,5886,5888,5895,5896,6109,6235,6277, 143,161,323,367,440,1027,1842,1885,1896,2365,2373,2391,2412,2416,2434,2462,3511,3515,3606,3654,3760,3984,4019,4169,4304,4326,4619,4627,5217,5573,5615,5790,6090,6274 ),]

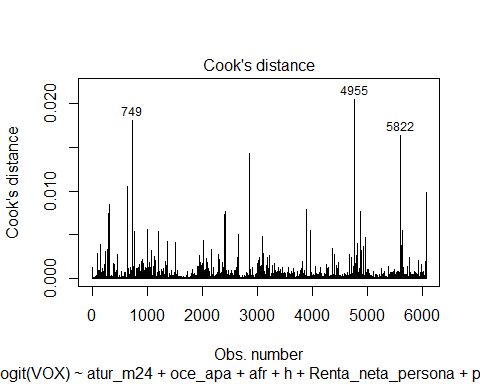
model11 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model11)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9041 -0.3940 0.0891 0.4371 1.5896   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.671e+00 2.226e-01 -29.967 < 2e-16 \*\*\*  
## atur\_m24 1.366e+00 1.886e-01 7.239 5.10e-13 \*\*\*  
## oce\_apa -1.369e+02 2.165e+01 -6.324 2.73e-10 \*\*\*  
## afr -2.737e+00 3.120e-01 -8.773 < 2e-16 \*\*\*  
## h 5.519e+00 3.008e-01 18.348 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.419e-04 4.388e-06 -32.339 < 2e-16 \*\*\*  
## pob\_t 2.264e-06 3.548e-07 6.381 1.88e-10 \*\*\*  
## e20\_29 1.911e+00 4.596e-01 4.157 3.26e-05 \*\*\*  
## e30\_39 5.077e+00 4.313e-01 11.771 < 2e-16 \*\*\*  
## e40\_49 4.989e+00 3.964e-01 12.585 < 2e-16 \*\*\*  
## e50\_59 3.612e+00 3.708e-01 9.742 < 2e-16 \*\*\*  
## M80 2.821e+00 2.943e-01 9.586 < 2e-16 \*\*\*  
## Gini\_medio 2.556e-02 2.714e-03 9.416 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6275 on 6052 degrees of freedom  
## Multiple R-squared: 0.2556, Adjusted R-squared: 0.2541   
## F-statistic: 173.2 on 12 and 6052 DF, p-value: < 2.2e-16

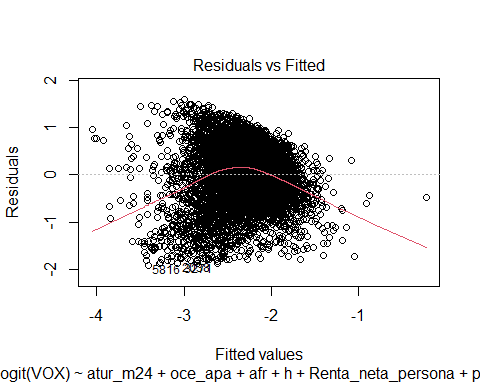
plot(density(residuals(model11)))



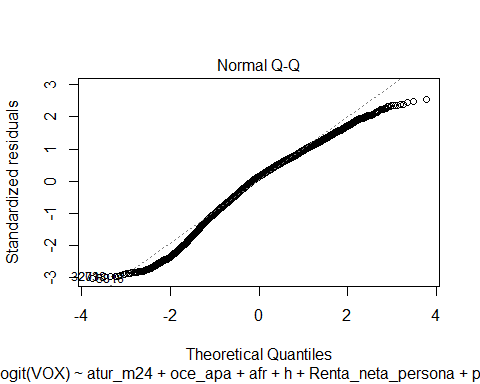
plot(model11,4)



plot(model11,1)



plot(model11,2)



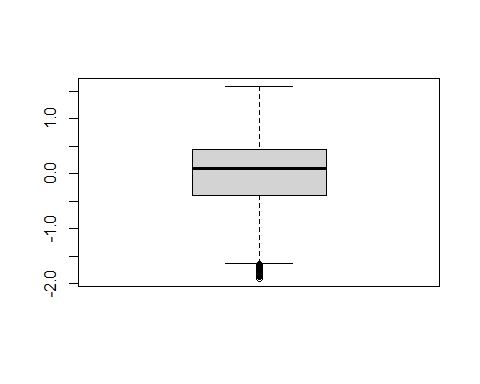
vif(model11)

## atur\_m24 oce\_apa afr h   
## 1.149290 1.031964 1.202144 1.272176   
## Renta\_neta\_persona pob\_t e20\_29 e30\_39   
## 1.350956 1.101913 1.803464 2.373505   
## e40\_49 e50\_59 M80 Gini\_medio   
## 2.582591 1.484101 4.615269 1.100474

library(tseries)  
jarque.bera.test(model11$residuals)

##   
## Jarque Bera Test  
##   
## data: model11$residuals  
## X-squared = 274.32, df = 2, p-value < 2.2e-16

outliers<-boxplot(model11$residuals)



outliers$out

## 391 547 751 760 783 836 852 862   
## -1.794205 -1.660574 -1.685568 -1.787892 -1.817358 -1.782791 -1.798113 -1.775977   
## 957 987 1999 2004 2006 2038 2061 2064   
## -1.790390 -1.731014 -1.767741 -1.741018 -1.720373 -1.872487 -1.767222 -1.694211   
## 2081 2107 2119 2123 2124 2126 2143 2149   
## -1.682188 -1.715098 -1.750856 -1.707894 -1.648751 -1.777081 -1.749668 -1.766617   
## 2174 2177 2470 2475 2486 2487 2497 2511   
## -1.803458 -1.710869 -1.701869 -1.841812 -1.772076 -1.806009 -1.690922 -1.650619   
## 2512 3099 3101 3106 3118 3132 3160 3161   
## -1.649857 -1.722869 -1.777989 -1.788214 -1.740282 -1.703465 -1.669510 -1.829637   
## 3164 3200 3228 3256 3271 3650 4438 4547   
## -1.762632 -1.659127 -1.655658 -1.819213 -1.878348 -1.720974 -1.690106 -1.728508   
## 4575 4890 4917 4941 4974 5807 5808 5816   
## -1.794949 -1.859574 -1.653919 -1.724120 -1.755780 -1.643850 -1.770750 -1.904135   
## 5833 5843 5877 5882 5887 5890 5894 5899   
## -1.802489 -1.733374 -1.672092 -1.658167 -1.865863 -1.683488 -1.727684 -1.761616

d4 <- d2  
d4=d4[-c(3524,720,5866,4023,4795,5625,3687,4348,5866,56,408,484,588,712,724,736,739,775, 797,824,829,831,834,843,936,940,972,1112,1169,1299,1405,1466,1993,1994,2018,2041,2048,2054,2058,2062,2068,2075,2086,2090,2097,2098,2117,2127,2150,2157,2162,2171,2173,2178,2221,2368,2376,2466,2468,2469,2471,2472,2474,2476,2477,2478,2479,2480,2481,2482,2490,2491,2492,2494,2498,2499,2500,2502,2503,2505,2507,2508,2510,2514,2516,2519,2520,2527,2587,2633,2643,2715,2725,3082,3085,3092,3095,3097,3100,3108,3112,3114,3120,3123,3130,3131,3134,3136,3137,3141,3142,3155,3158,3159,3165,3166,3168,3169,3178,3181,3182,3183,3198,3202,3207,3208,3211,3217,3220,3221,3225,3227,3230,3231,3239,3245,3249,3265,3266,3270,3295,3324,3348,3358,3371,3471,3525,3556,3565,3586,3603,3610,3615,3620,3797,4563,4702,4873,4888,4896,4907,4918,4927,4934,4949,4960,4963,4966,4968,4969,4971,4977,4978,4979,4983,5005,5012,5015,5028,5035,5661,5713,5793,5810,5811,5812,5814,5819,5820,5823,5824,5826,5827,5828,5830,5832,5834,5837,5838,5846,5847,5848,5849,5852,5853,5854,5855,5856,5857,5860,5861,5862,5863,5867,5873,5881,5884,5886,5888,5895,5896,6109,6235,6277, 143,161,323,367,440,1027,1842,1885,1896,2365,2373,2391,2412,2416,2434,2462,3511,3515,3606,3654,3760,3984,4019,4169,4304,4326,4619,4627,5217,5573,5615,5790,6090,6274,391,547,751,760,783,836,852,862,957,987,1999,2004,2006,2038,2061,2064,2081,2107,2119,2123,2124,2126,2143,2149,2174,2177,2470,2475,2486,2487,2497,2511,2512,3099,3101,3106,3118,3132,3160,3161,3164,3200,3228,256,3271,3650,4438,4547,4575,4890,4917,4941,4974,5807,5808,5816,5833,5843,5877,5882,5887,5890,5894,5899,5822,4955,749,2495,6317,2465),]

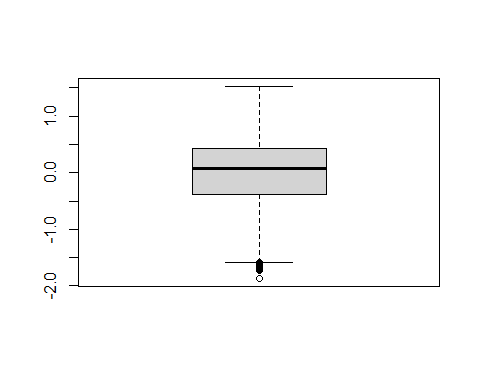
model12 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model12)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.86599 -0.38400 0.08229 0.42237 1.52436   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.628e+00 2.152e-01 -30.799 < 2e-16 \*\*\*  
## atur\_m24 1.351e+00 1.819e-01 7.427 1.27e-13 \*\*\*  
## oce\_apa -1.003e+02 2.251e+01 -4.454 8.60e-06 \*\*\*  
## afr -2.645e+00 3.018e-01 -8.763 < 2e-16 \*\*\*  
## h 5.485e+00 2.910e-01 18.848 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.313e-04 4.282e-06 -30.655 < 2e-16 \*\*\*  
## pob\_t 2.241e-06 3.750e-07 5.975 2.43e-09 \*\*\*  
## e20\_29 1.754e+00 4.442e-01 3.949 7.95e-05 \*\*\*  
## e30\_39 4.956e+00 4.164e-01 11.902 < 2e-16 \*\*\*  
## e40\_49 4.706e+00 3.825e-01 12.304 < 2e-16 \*\*\*  
## e50\_59 3.404e+00 3.577e-01 9.517 < 2e-16 \*\*\*  
## M80 2.533e+00 2.838e-01 8.924 < 2e-16 \*\*\*  
## Gini\_medio 2.565e-02 2.630e-03 9.755 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6029 on 5982 degrees of freedom  
## Multiple R-squared: 0.2432, Adjusted R-squared: 0.2417   
## F-statistic: 160.2 on 12 and 5982 DF, p-value: < 2.2e-16

library(tseries)  
jarque.bera.test(model12$residuals)

##   
## Jarque Bera Test  
##   
## data: model12$residuals  
## X-squared = 233.7, df = 2, p-value < 2.2e-16

outliers<-boxplot(model12$residuals)



outliers$out

## 212 747 748 777 779 841 914 929   
## -1.620756 -1.606460 -1.686469 -1.612480 -1.604714 -1.627782 -1.661571 -1.673358   
## 966 2005 2014 2017 2036 2040 2085 2095   
## -1.610096 -1.609632 -1.653674 -1.631402 -1.594501 -1.616163 -1.677199 -1.674270   
## 2467 2483 2488 2493 2517 2522 2523 2526   
## -1.684343 -1.700295 -1.680323 -1.663433 -1.729587 -1.625714 -1.652933 -1.661954   
## 3079 3117 3128 3146 3147 3241 3256 3369   
## -1.613031 -1.627177 -1.621191 -1.618432 -1.593670 -1.623539 -1.865988 -1.604954   
## 4965 5027 5825 5865 5878 5892 5898   
## -1.601552 -1.631624 -1.652396 -1.690226 -1.615640 -1.651736 -1.595284

d4 <- d2  
d4=d4[-c(3524,720,5866,4023,4795,5625,3687,4348,5866,56,408,484,588,712,724,736,739,775, 797,824,829,831,834,843,936,940,972,1112,1169,1299,1405,1466,1993,1994,2018,2041,2048,2054,2058,2062,2068,2075,2086,2090,2097,2098,2117,2127,2150,2157,2162,2171,2173,2178,2221,2368,2376,2466,2468,2469,2471,2472,2474,2476,2477,2478,2479,2480,2481,2482,2490,2491,2492,2494,2498,2499,2500,2502,2503,2505,2507,2508,2510,2514,2516,2519,2520,2527,2587,2633,2643,2715,2725,3082,3085,3092,3095,3097,3100,3108,3112,3114,3120,3123,3130,3131,3134,3136,3137,3141,3142,3155,3158,3159,3165,3166,3168,3169,3178,3181,3182,3183,3198,3202,3207,3208,3211,3217,3220,3221,3225,3227,3230,3231,3239,3245,3249,3265,3266,3270,3295,3324,3348,3358,3371,3471,3525,3556,3565,3586,3603,3610,3615,3620,3797,4563,4702,4873,4888,4896,4907,4918,4927,4934,4949,4960,4963,4966,4968,4969,4971,4977,4978,4979,4983,5005,5012,5015,5028,5035,5661,5713,5793,5810,5811,5812,5814,5819,5820,5823,5824,5826,5827,5828,5830,5832,5834,5837,5838,5846,5847,5848,5849,5852,5853,5854,5855,5856,5857,5860,5861,5862,5863,5867,5873,5881,5884,5886,5888,5895,5896,6109,6235,6277, 143,161,323,367,440,1027,1842,1885,1896,2365,2373,2391,2412,2416,2434,2462,3511,3515,3606,3654,3760,3984,4019,4169,4304,4326,4619,4627,5217,5573,5615,5790,6090,6274,391,547,751,760,783,836,852,862,957,987,1999,2004,2006,2038,2061,2064,2081,2107,2119,2123,2124,2126,2143,2149,2174,2177,2470,2475,2486,2487,2497,2511,2512,3099,3101,3106,3118,3132,3160,3161,3164,3200,3228,256,3271,3650,4438,4547,4575,4890,4917,4941,4974,5807,5808,5816,5833,5843,5877,5882,5887,5890,5894,5899,5822,4955,749,212 ,747,748,777,779,841,914,929,966,2005,2014,2017,2040,2085,2095,2465,2467,2483,2488,2493,2517,2522,2523,2526,3079,3117,3128,3146,3147,3241,3256,3369,4932,4965,5027,5825,5865,5878,5892),]

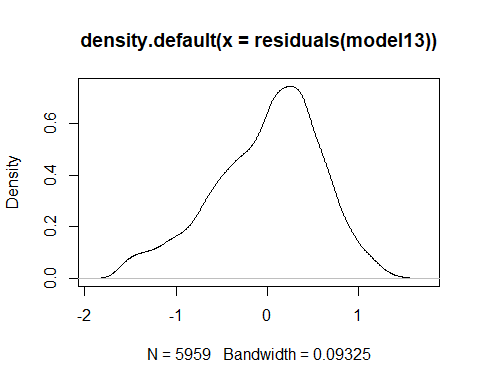
model13 <- lm(logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona + pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio, data=d4)  
summary(model13)

##   
## Call:  
## lm(formula = logit(VOX) ~ atur\_m24 + oce\_apa + afr + h + Renta\_neta\_persona +   
## pob\_t + e20\_29 + e30\_39 + e40\_49 + e50\_59 + M80 + Gini\_medio,   
## data = d4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.64146 -0.37692 0.07804 0.41427 1.46822   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.490e+00 2.112e-01 -30.733 < 2e-16 \*\*\*  
## atur\_m24 1.321e+00 1.782e-01 7.417 1.36e-13 \*\*\*  
## oce\_apa -1.009e+02 2.163e+01 -4.664 3.17e-06 \*\*\*  
## afr -2.685e+00 2.956e-01 -9.080 < 2e-16 \*\*\*  
## h 5.349e+00 2.854e-01 18.745 < 2e-16 \*\*\*  
## Renta\_neta\_persona -1.237e-04 4.229e-06 -29.255 < 2e-16 \*\*\*  
## pob\_t 1.968e-06 3.400e-07 5.790 7.41e-09 \*\*\*  
## e20\_29 1.589e+00 4.355e-01 3.649 0.000265 \*\*\*  
## e30\_39 4.874e+00 4.086e-01 11.928 < 2e-16 \*\*\*  
## e40\_49 4.436e+00 3.752e-01 11.823 < 2e-16 \*\*\*  
## e50\_59 3.116e+00 3.513e-01 8.870 < 2e-16 \*\*\*  
## M80 2.325e+00 2.783e-01 8.352 < 2e-16 \*\*\*  
## Gini\_medio 2.542e-02 2.581e-03 9.849 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.59 on 5946 degrees of freedom  
## Multiple R-squared: 0.232, Adjusted R-squared: 0.2305   
## F-statistic: 149.7 on 12 and 5946 DF, p-value: < 2.2e-16

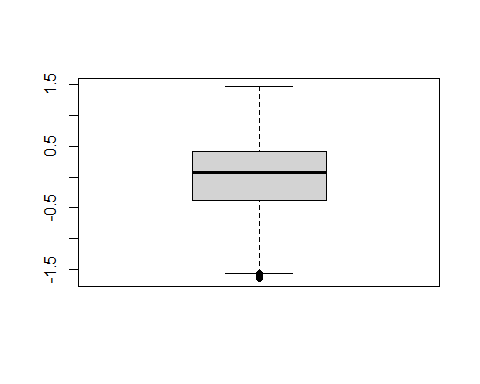
jarque.bera.test(model13$residuals)

##   
## Jarque Bera Test  
##   
## data: model13$residuals  
## X-squared = 221.48, df = 2, p-value < 2.2e-16

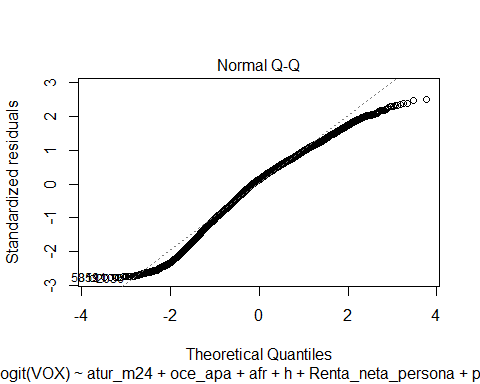
plot(density(residuals(model13)))



outliers<-boxplot(model13$residuals)



plot(model13,2)

 ###Podem comprobar que el jarque test encara ens dona heterocedasticitat, tot i haver normalitzat molt més la distribució. Per una altra banda, a mida que traiem outliers, tornen a aparéixer de nous, i el nostre model va perdent adjust. ###2.4. Anàlisi de components principals

library(FactoMineR)

## Warning: package 'FactoMineR' was built under R version 4.1.3

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.1.3

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

d4 <- d2 %>%  
 select(PSOE, VOX, PP, Cs, ERC, JUNTS, PNV, PODEMOS, Bildu, BNG, CUP) %>%   
 scale()  
  
head(d4)

## PSOE VOX PP Cs ERC JUNTS  
## [1,] 0.53087991 0.5145228 0.9450895 -0.86959251 -0.3787958 -0.3530381  
## [2,] 1.36852053 0.9498829 0.6237310 -0.44402846 -0.3787958 -0.3530381  
## [3,] -0.09416569 0.8730360 0.1609107 1.26225015 -0.3787958 -0.3530381  
## [4,] 1.37724952 0.8999373 0.1732129 0.03914441 -0.3787958 -0.3530381  
## [5,] 0.25368270 1.1772157 0.7536820 0.17776425 -0.3787958 -0.3530381  
## [6,] 1.14307477 -0.3214335 0.9870852 -0.12875747 -0.3787958 -0.3530381  
## PNV PODEMOS Bildu BNG CUP  
## [1,] -0.1544426 -0.3580243 -0.1477363 -0.2007384 -0.3395195  
## [2,] -0.1544426 -0.9302877 -0.1477363 -0.2007384 -0.3395195  
## [3,] -0.1544426 0.1788668 -0.1477363 -0.2007384 -0.3395195  
## [4,] -0.1544426 -0.7812781 -0.1477363 -0.2007384 -0.3395195  
## [5,] -0.1544426 -0.5186323 -0.1477363 -0.2007384 -0.3395195  
## [6,] -0.1544426 -0.8275554 -0.1477363 -0.2007384 -0.3395195

round(cov(d4),3)

## PSOE VOX PP Cs ERC JUNTS PNV PODEMOS Bildu BNG  
## PSOE 1.000 0.199 0.177 0.159 -0.546 -0.546 -0.240 0.034 -0.242 -0.041  
## VOX 0.199 1.000 0.381 0.421 -0.508 -0.492 -0.261 -0.014 -0.254 -0.201  
## PP 0.177 0.381 1.000 0.196 -0.575 -0.543 -0.247 -0.347 -0.244 0.124  
## Cs 0.159 0.421 0.196 1.000 -0.301 -0.319 -0.241 0.133 -0.235 -0.145  
## ERC -0.546 -0.508 -0.575 -0.301 1.000 0.911 -0.059 -0.038 -0.056 -0.076  
## JUNTS -0.546 -0.492 -0.543 -0.319 0.911 1.000 -0.055 -0.077 -0.052 -0.071  
## PNV -0.240 -0.261 -0.247 -0.241 -0.059 -0.055 1.000 0.059 0.884 -0.031  
## PODEMOS 0.034 -0.014 -0.347 0.133 -0.038 -0.077 0.059 1.000 0.038 -0.249  
## Bildu -0.242 -0.254 -0.244 -0.235 -0.056 -0.052 0.884 0.038 1.000 -0.030  
## BNG -0.041 -0.201 0.124 -0.145 -0.076 -0.071 -0.031 -0.249 -0.030 1.000  
## CUP -0.518 -0.471 -0.526 -0.301 0.877 0.882 -0.052 -0.046 -0.050 -0.068  
## CUP  
## PSOE -0.518  
## VOX -0.471  
## PP -0.526  
## Cs -0.301  
## ERC 0.877  
## JUNTS 0.882  
## PNV -0.052  
## PODEMOS -0.046  
## Bildu -0.050  
## BNG -0.068  
## CUP 1.000

pca\_municipi <- princomp(d4)  
  
summary(pca\_municipi)

## Importance of components:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5  
## Standard deviation 2.0116999 1.4567310 1.2242715 1.00751567 0.88276846  
## Proportion of Variance 0.3679615 0.1929455 0.1362798 0.09229531 0.07085486  
## Cumulative Proportion 0.3679615 0.5609071 0.6971869 0.78948219 0.86033705  
## Comp.6 Comp.7 Comp.8 Comp.9 Comp.10  
## Standard deviation 0.76467424 0.68494473 0.4134241 0.34027995 0.330787766  
## Proportion of Variance 0.05316538 0.04265668 0.0155406 0.01052807 0.009948896  
## Cumulative Proportion 0.91350243 0.95615911 0.9716997 0.98222777 0.992176671  
## Comp.11  
## Standard deviation 0.293330890  
## Proportion of Variance 0.007823329  
## Cumulative Proportion 1.000000000

attributes(pca\_municipi)

## $names  
## [1] "sdev" "loadings" "center" "scale" "n.obs" "scores" "call"   
##   
## $class  
## [1] "princomp"

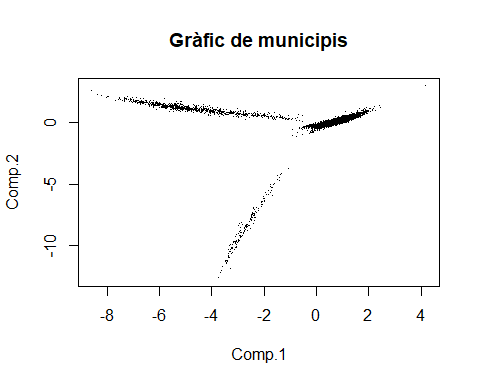
Z <- pca\_municipi$scores  
  
head(Z)

## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## [1,] 0.9589929 0.05514251 -0.6213888 -0.006050514 0.9763747 -0.6540425  
## [2,] 1.3551393 0.25490359 -0.5892182 0.051964765 1.2983580 0.2465315  
## [3,] 1.1187818 0.32725785 0.7050148 -0.730232317 -0.6058269 0.3376023  
## [4,] 1.3028934 0.27611909 -0.1862439 0.122340787 0.9679038 0.6284226  
## [5,] 1.2741620 0.31701412 -0.2155537 -0.764529718 0.3918164 -0.1086075  
## [6,] 1.0630134 0.14253373 -0.7966518 0.302246781 1.0036547 0.4763879  
## Comp.7 Comp.8 Comp.9 Comp.10 Comp.11  
## [1,] 0.05082424 0.25403290 0.010450776 0.173316683 0.07932420  
## [2,] 0.62615743 0.44244280 -0.010690054 0.306523263 0.12831734  
## [3,] 0.14052879 0.03041935 -0.006778470 0.007981414 -0.02738470  
## [4,] 0.71571061 0.31460478 -0.015348122 0.211975854 0.08023127  
## [5,] 0.40842257 0.25494417 -0.008640725 0.175021563 0.07344313  
## [6,] -0.59787471 0.24590537 0.001259457 0.151820950 0.05924092

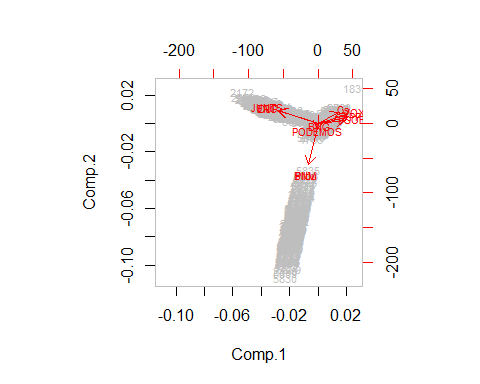
round(cov(Z),3)

## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10  
## Comp.1 4.048 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000  
## Comp.2 0.000 2.122 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000  
## Comp.3 0.000 0.000 1.499 0.000 0.000 0.000 0.000 0.000 0.000 0.000  
## Comp.4 0.000 0.000 0.000 1.015 0.000 0.000 0.000 0.000 0.000 0.000  
## Comp.5 0.000 0.000 0.000 0.000 0.779 0.000 0.000 0.000 0.000 0.000  
## Comp.6 0.000 0.000 0.000 0.000 0.000 0.585 0.000 0.000 0.000 0.000  
## Comp.7 0.000 0.000 0.000 0.000 0.000 0.000 0.469 0.000 0.000 0.000  
## Comp.8 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.171 0.000 0.000  
## Comp.9 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.116 0.000  
## Comp.10 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.109  
## Comp.11 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000  
## Comp.11  
## Comp.1 0.000  
## Comp.2 0.000  
## Comp.3 0.000  
## Comp.4 0.000  
## Comp.5 0.000  
## Comp.6 0.000  
## Comp.7 0.000  
## Comp.8 0.000  
## Comp.9 0.000  
## Comp.10 0.000  
## Comp.11 0.086

plot(pca\_municipi$scores[,1:2], pch =".", main ="Gràfic de municipis")



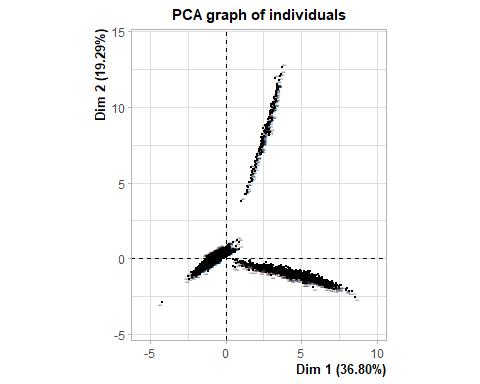
biplot(pca\_municipi, cex=0.7, col=c("grey","red"))



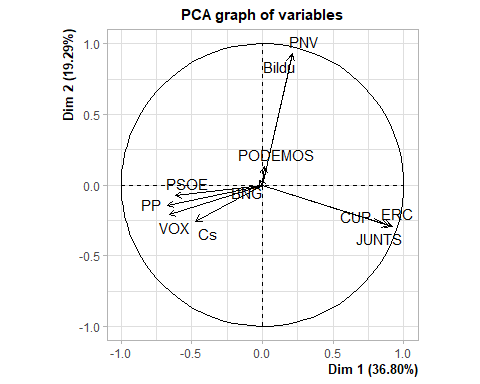
respca\_municipi <- PCA(d4, graph=FALSE)  
summary(respca\_municipi)

##   
## Call:  
## PCA(X = d4, graph = FALSE)   
##   
##   
## Eigenvalues  
## Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7  
## Variance 4.048 2.122 1.499 1.015 0.779 0.585 0.469  
## % of var. 36.796 19.295 13.628 9.230 7.085 5.317 4.266  
## Cumulative % of var. 36.796 56.091 69.719 78.948 86.034 91.350 95.616  
## Dim.8 Dim.9 Dim.10 Dim.11  
## Variance 0.171 0.116 0.109 0.086  
## % of var. 1.554 1.053 0.995 0.782  
## Cumulative % of var. 97.170 98.223 99.218 100.000  
##   
## Individuals (the 10 first)  
## Dist Dim.1 ctr cos2 Dim.2 ctr cos2 Dim.3 ctr  
## 1 | 1.672 | -0.959 0.004 0.329 | -0.055 0.000 0.001 | -0.621 0.004  
## 2 | 2.167 | -1.355 0.007 0.391 | -0.255 0.000 0.014 | -0.589 0.004  
## 3 | 1.701 | -1.119 0.005 0.433 | -0.327 0.001 0.037 | 0.705 0.005  
## 4 | 1.954 | -1.303 0.007 0.445 | -0.276 0.001 0.020 | -0.186 0.000  
## 5 | 1.670 | -1.274 0.006 0.582 | -0.317 0.001 0.036 | -0.216 0.000  
## 6 | 1.886 | -1.063 0.004 0.318 | -0.143 0.000 0.006 | -0.797 0.007  
## 7 | 2.757 | -0.608 0.001 0.049 | 0.241 0.000 0.008 | 0.012 0.000  
## 8 | 1.450 | -0.784 0.002 0.292 | -0.078 0.000 0.003 | 0.045 0.000  
## 9 | 1.133 | -0.889 0.003 0.615 | -0.128 0.000 0.013 | 0.453 0.002  
## 10 | 1.482 | -1.156 0.005 0.608 | -0.203 0.000 0.019 | -0.109 0.000  
## cos2   
## 1 0.138 |  
## 2 0.074 |  
## 3 0.172 |  
## 4 0.009 |  
## 5 0.017 |  
## 6 0.179 |  
## 7 0.000 |  
## 8 0.001 |  
## 9 0.159 |  
## 10 0.005 |  
##   
## Variables (the 10 first)  
## Dim.1 ctr cos2 Dim.2 ctr cos2 Dim.3 ctr cos2   
## PSOE | -0.618 9.439 0.382 | -0.075 0.262 0.006 | 0.086 0.490 0.007 |  
## VOX | -0.661 10.804 0.437 | -0.210 2.078 0.044 | 0.219 3.185 0.048 |  
## PP | -0.674 11.220 0.454 | -0.143 0.967 0.021 | -0.441 12.952 0.194 |  
## Cs | -0.472 5.500 0.223 | -0.258 3.126 0.066 | 0.396 10.438 0.156 |  
## ERC | 0.916 20.733 0.839 | -0.291 3.981 0.085 | 0.037 0.093 0.001 |  
## JUNTS | 0.911 20.526 0.831 | -0.292 4.015 0.085 | 0.001 0.000 0.000 |  
## PNV | 0.209 1.081 0.044 | 0.928 40.567 0.861 | 0.024 0.038 0.001 |  
## PODEMOS | 0.012 0.004 0.000 | 0.129 0.779 0.017 | 0.764 38.985 0.584 |  
## Bildu | 0.209 1.074 0.043 | 0.924 40.208 0.853 | 0.014 0.013 0.000 |  
## BNG | -0.022 0.012 0.000 | 0.035 0.058 0.001 | -0.712 33.777 0.506 |

plot(respca\_municipi,choix="ind", cex=.1)



plot(respca\_municipi,choix="var")



Z <- predict(respca\_municipi, newdata=d4)$coord[,1:2]  
colores <- c("red", "blue", "green")  
  
plot(Z, pch=".", col= colores, main="Municipis segons 3 clusters")

