

CODI TFG. Les trajectòries juvenils a Catalunya (1992-2012) des d'una perspectiva longitudinal: anàlisi de la transició escola-treball i la seva relació amb el temps de l'emancipació residencial.

Bruna Santanach

07-06-2023

Càrrega llibreries

```
library(writexl)
library(plyr)
library(haven)
library(tidyverse)
library(lubridate)
library(scales)
library(skimr)
library(ggplot2)
library(ggseqplot)
library(labelled)
library(TraMineR)
library(cluster)
library(kableExtra)
library(seqhandbook)
library(ade4)
library(cluster)
library(pvclust)
library(survival)
library(haven)
library(MASS)
library(survminer)
library(xfun)
library(forcats)
library(webshot2)
library(gt)
library(ggdendro)
library(factoextra)
```

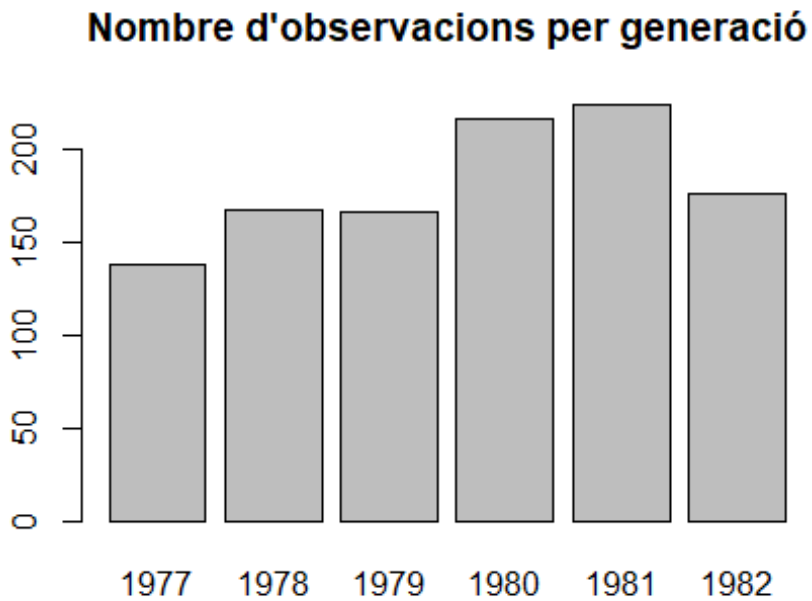
Càrrega de dades

Obrim les dades de l'EJC 12,i agafem les generacions nascudes entre al 1997 i 1982 (es a dir entre 30 i 35 anys al 2012), per poder analitzar trajectòries completes fins als 30.

```
EJC12<- read_sav("0_EJC12_amb vC_anonimitzada.sav")
EJC12$v1<-as.integer(format(EJC12$v1, format="%Y"))
EJC12<- EJC12 %>% filter(v1<=1982)
```

Tenim aquests casos per cohorts:

```
barplot(table(EJC12$v1), main="Nombre d'observacions per generació")
```



```
table(EJC12$v1)
```

```
##
## 1977 1978 1979 1980 1981 1982
## 138 167 166 217 224 176
```

Variables de gènere, lloc d'origen, nivell socioeconòmic familiar i generació (per utilitzar més és endavant)

```
variables<-EJC12[,c(1,6,10,11,12,916,911)]
colnames(variables)<-
c("id","municipi","sexe","cohort","lloc.naix","c.socioecon","n.estudis")
```

Municipi

```
attributes(variables$municipi)$labels
```

```
## Capital de comarca      Màxim rural      Rural intermedi      Urbà
##           1           2           3           4
```

```
variables$municipi<-as.factor(variables$municipi)
levels(variables$municipi)<-c("capital_Comarca", "max_rural",
"mitj_rural","urbà")#codifiquem
```

Sexe

```
attributes(variables$sexe)$labels
```

```
## Dona Home
##      1      2

variables$sexe<-as.factor(variables$sexe)
levels(variables$sexe)<-c("dona", "home")#codifiquem
```

Lloc naixement

```
attributes(variables$lloc.naix)$labels

##      A Catalunya A la resta d'Espanya      A l'estranger
##      1          2          3
##      No pertinent      Ns      Nc
##      97          98          99

variables$lloc.naix<-as.factor(variables$lloc.naix)
levels(variables$lloc.naix)<-c("Cat", "Esp", "altre_País")#codifiquem
```

Categoria socioeconòmica

```
attributes(variables$c.socioecon)$labels

##      Directors i gerents
##      1
##      Tècnics i Professionals
##      2
##      Ocupacions que requereixen mitjana qualificació
##      3
##      Ocupacions elementals
##      4
##      Ocupacions militars
##      5
##      Altres
##      6

variables$c.socioecon<-as.factor(variables$c.socioecon)
summary(variables$c.socioecon)

##      1      2      3      4      5 NA's
## 104 231 546 129      5    73

levels(variables$c.socioecon)<-c("directors_gerents", "tèncincs_professinals",
"ocup.quali.mitjana", "ocup.elementals", "ocup.militars")
```

Nivell estudis

```
attributes(variables$n.estudis)$labels

##      Estudis inferiors a obligatoris      Estudis obligatoris
##      1          2
##      Estudis secundaris postobligatoris      Estudis universitaris
##      3          4
##      No ho sap      No contesta
##      98          99
```

```
variables$n.estudis<-as.factor(variables$n.estudis)
levels(variables$n.estudis)<-c("no_obligatoris", "oblifatoris",
"secundaris_postoblig", "universitaris")#codifiquem
```

4.Anàlisi de seqüències multicanal de les transicions escola-treball

4.2. La codificació de trajectòries

- Construcció de la matriu de trajectòries educatives de l'EJC12

```
Educ_traj<-data.frame(EJC12$id, EJC12[,c(46:135)])

educ.names<-c("esdeveniment", "data.i", "data.f", "centre", "beca", "repetició",
"titulació", "abandonament", "canvi", "en.curs")

col.names<-c("id")
for (i in 0:8){
col.names<-c(col.names,(paste(educ.names," - E",i, sep="")))
}

colnames(Educ_traj)<-col.names
colnames(Educ_traj) <- gsub(" ", "", colnames(Educ_traj))
```

Passem les EJC12 a format llarg:

```
seq.long.e<-Educ_traj %>%
  dplyr::select(id, starts_with(c(educ.names[1:3]))) %>%
  pivot_longer(-id,names_to= c(".value", "Esdeveniment"), names_sep="-")

seq.long.e$data.i<-as.integer(format(seq.long.e$data.i, format="%Y"))
seq.long.e$data.f<-as.integer(format(seq.long.e$data.f, format="%Y"))
```

Eliminem aquells estats que no han queda omplerts en la graella

```
seq.long.e<-seq.long.e %>% filter(esdeveniment !=999)
```

Visualització:

```
seq.long.e[1:10,]

## # A tibble: 10 × 5
##       id Esdeveniment          esdeveniment data.i data.f
##   <dbl> <chr>          <dbl+lbl> <int> <int>
## 1     1 E0           1 [Educació obligatòria (ESO o EGB)]      1994  1994
## 2     1 E1           3 [FP1 / CF Grau Mig]                1994  1997
## 3     1 E2           0 [Escola taller, PGS, PQPI, escola d'adult...] 1997    NA
## 4     4 E0           1 [Educació obligatòria (ESO o EGB)]      1994  2000
## 5     4 E1           2 [BUP, COU o Batxillerat]              2000  2002
## 6     4 E2           4 [FP2 / CF Grau Superior]              2002  2004
## 7    45 E0           1 [Educació obligatòria (ESO o EGB)]      1995  1997
## 8    45 E1           4 [FP2 / CF Grau Superior]              1997  2001
## 9    47 E0           1 [Educació obligatòria (ESO o EGB)]      1993  1994
## 10   47 E1           2 [BUP, COU o Batxillerat]              1994  1996
```

Recodifiquem els nivells dels esdeveniments (explicat al word), a 8 nivells: “PQPI-escola d’adults”, “ESO”, “Batxillerat”, “FP1”, “FP2”, “Diplomatures-L·licenciatures”, “Postgrau-Masters-Doctorat” i “No estudiant”.

```

nivells.e<-attributes(seq.long.e$esdeveniment)$labels

for (i in 1:nrow(seq.long.e)){
  if (seq.long.e$esdeveniment[i]==6)seq.long.e$esdeveniment[i]<-5
  else{

if(seq.long.e$esdeveniment[i]==7|seq.long.e$esdeveniment[i]==9|seq.long.e$esdeveniment[i]==8) {seq.long.e$esdeveniment[i]<-6
}
  else{
    if (seq.long.e$esdeveniment[i]==10) seq.long.e$esdeveniment[i]<-7
  }
}
}

seq.long.e<- seq.long.e %>%
  mutate(esdeveniment=esdeveniment+1)
seq.long.e<-seq.long.e %>% filter(esdeveniment !=91011)

levels(as.factor(seq.long.e$esdeveniment))

## [1] "1" "2" "3" "4" "5" "6" "7" "8"

nivells.e<-c("PQPI o escola d'adults","ESO", "Batxillerat", "FP1","FP2",
"Diplomatures o Llicenciatures", "Postgrau, Masters o Doctorat", "No estudiant")

```

Convertim a matriu de seqüències per edats:

```

Edat<-filter(EJC12, id %in% seq.long.e$id)
Edat<-Edat[,c(1,11)]

edats.educatiu <- seqformat(seq.long.e, from = "SPELL", to = "STS",
  id = "id", begin = "data.i", end = "data.f", status = "esdeveniment", process
= TRUE, pdata=Edat, pvar=c("id", "v1"),limit=30)

edats.educatiu <- seqdef(edats.educatiu, alphabet = 1:8,
  labels = nivells.e)

edats.educatiu<-edats.educatiu[,-c(1:15)]
colnames(edats.educatiu)<-gsub("a","",as.character(colnames(edats.educatiu)))
alphabet(edats.educatiu) <- nivells.e
alphabet(edats.educatiu)

## [1] "PQPI o escola d'adults"      "ESO"
## [3] "Batxillerat"                "FP1"
## [5] "FP2"                        "Diplomatures o Llicenciatures"
## [7] "Postgrau, Masters o Doctorat" "No estudiant"

edats.educatiu[edats.educatiu=="%"]<- "No estudiant" ## seqüències faltant per la
dreta, ho podem fer, perquè els missings els codifica com a *

```

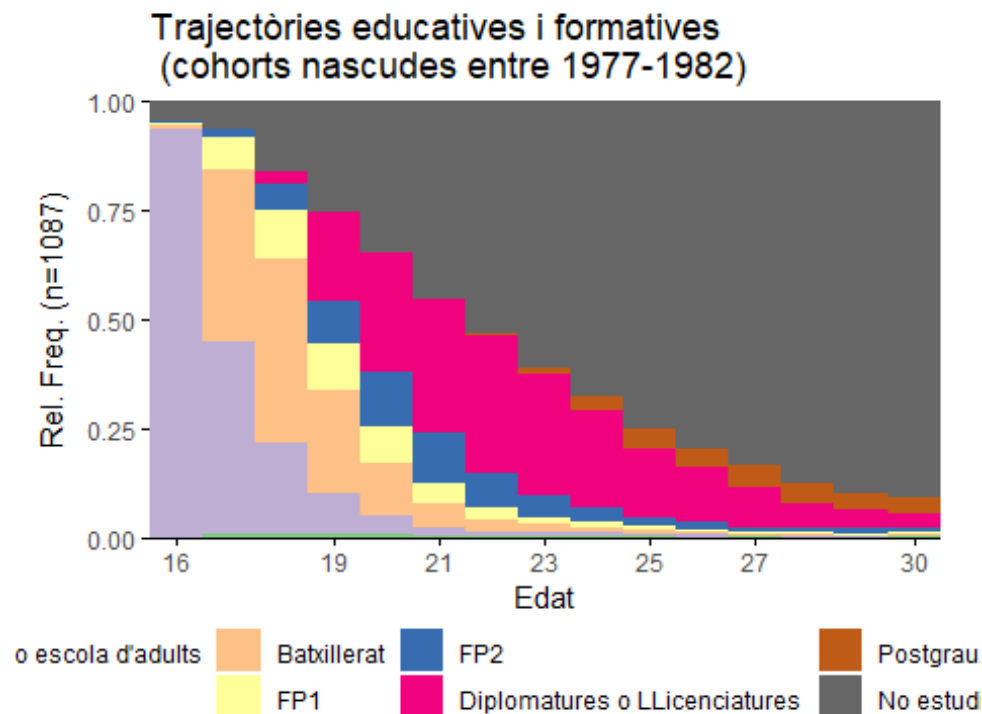
Visualització seqüència

```
edats.educatiu[1:5,]
```

```
## Sequence
## 1 FP1-FP1-FP1-FP1-No estudiant-No estudiant-No estudiant-No estudiant-No
estudiant-No estudiant-No estudiant-No estudiant-No estudiant-No estudiant-No
estudiant
## 4 ESO-ESO-ESO-ESO-ESO-ESO-Batxillerat-Batxillerat-FP2-FP2-FP2-No estudiant-
No estudiant-No estudiant-No estudiant
## 45 ESO-ESO-FP2-FP2-FP2-FP2-FP2-No estudiant-No estudiant-No estudiant-No
estudiant-No estudiant-No estudiant-No estudiant-No estudiant
## 47 ESO-Batxillerat-Batxillerat-Diplomatures o Llicenciatures-Diplomatures o
Llicenciatures-Diplomatures o Llicenciatures-No estudiant-No estudiant-No
estudiant-No estudiant-No estudiant-Diplomatures o Llicenciatures-Diplomatures o
Llicenciatures-No estudiant-No estudiant
## 55 ESO-ESO-ESO-Batxillerat-Batxillerat-Batxillerat-Diplomatures o
Llicenciatures-Diplomatures o Llicenciatures-Diplomatures o Llicenciatures-
Diplomatures o Llicenciatures-Diplomatures o Llicenciatures-Diplomatures o
Llicenciatures-No estudiant-No estudiant-No estudiant
```

Gràfic distribució dels estats per freqüències relatives de les trajectòries educatives (gràfic 1)

```
educ.plot<-ggseqdplot(edats.educatiu, ltext=names(nivells.e), border=F) +
  labs(title = "Trajectòries educatives i formatives \n (cohorts nascudes entre
1977-1982)", x="Edat")
educ.plot
```



- Construcció de la matriu de trajectòries laborals de l'EJC12

```
Lab_traj<-data.frame(EJC12$id, EJC12[,c(154:262)])
```

```
lab.names<-c("esdeveniment", "activitat_en_curs", "data.i", "data.f",
```

```

"ocupacio_1", "codi_ocupacio_1", "Ocupació_última", "codi_ocupació_ultima",
"situació_professional_primera", "situació_professional_última",
"altres_activitats", "activitats_simultànes")

col.names<-c("id")
for (i in 0:8){
col.names<-c(col.names,(paste(lab.names," - E",i, sep="")))
}

colnames(Lab_traj)<-col.names
colnames(Lab_traj) <- gsub(" ", "", colnames(Lab_traj))

```

Passem a format llarg

```

seq.long.l<-Lab_traj %>%
  dplyr::select(id,starts_with(c(lab.names[c(1,3,4,12)]))) %>%
  pivot_longer(-id,names_to= c(".value", "Esdeveniment"), names_sep="-")

seq.long.l$data.i<-as.integer(format(seq.long.l$data.i, format="%Y"))
seq.long.l$data.f<-as.integer(format(seq.long.l$data.f, format="%Y"))

seq.long.l<-seq.long.l %>%
  drop_na(data.i)%>%
  drop_na(esdeveniment)
seq.long.l$data.f[is.na(seq.long.l$data.f)]<-2012

```

Generem els diferents nivells:

canvis en les variables 1a activitat

```

seq.long.l$esdeveniment[seq.long.l$esdeveniment==7]<-6 #cura dels fills com a
tasques reproductives.
seq.long.l$esdeveniment[seq.long.l$esdeveniment==9]<-8 #no treball

seq.long.l<-seq.long.l%>% filter( esdeveniment !=13068777600)

```

canvis en les variables de 2na activitat, que la prendrem quan estat principi=estudiar

```

seq.long.l$activitats_simultànes[seq.long.l$activitats_simultànes==7] <-
6#ALTRES-NO ESTUDIANT
seq.long.l$activitats_simultànes[seq.long.l$activitats_simultànes==8] <-7
seq.long.l$activitats_simultànes[seq.long.l$activitats_simultànes==6]<-8
seq.long.l$activitats_simultànes[is.na(seq.long.l$activitats_simultànes)]<- 8
#no

```

Descartar la vairbale "estudiant" (=1) agafar 2a activitat en aquests casos.

```

for (i in 1:nrow(seq.long.l)){
  if (seq.long.l$esdeveniment[i]==1){
    seq.long.l$esdeveniment[i]<-seq.long.l$activitats_simultànes[i]
  }
}

```

```

}
}

#estudiar com a "no treball", tenint en compte que ja hem codificat la activitat
no principal
seq.long.l$esdeveniment[seq.long.l$esdeveniment==1]<-8 #no treball
seq.long.l<-seq.long.l %>% mutate(esdeveniment=esdeveniment-1)

```

Visualització

```

seq.long.l[1:10,]

## # A tibble: 10 × 6
##       id Esdeveniment esdeveniment data.i data.f
activitats_simultànes
##   <dbl> <chr>          <dbl> <int> <dbl>
<dbl+lbl>
## 1      1 E0              7  1994  1998 8 [Feinetes informals
d'acompa...
## 2      1 E1              1  1998  1999 2 [Pràctiques retribuïdes /
be...
## 3      1 E2              2  2000  2004 8 [Feinetes informals
d'acompa...
## 4      1 E3              2  2004  2004 8 [Feinetes informals
d'acompa...
## 5      1 E4              2  2004  2012 8 [Feinetes informals
d'acompa...
## 6      4 E0              6  1994  2004 7 [Altres situacions
d'inactiv...
## 7      4 E1              1  2004  2004 8 [Feinetes informals
d'acompa...
## 8      4 E2              4  2004  2005 8 [Feinetes informals
d'acompa...
## 9      4 E3              2  2005  2012 8 [Feinetes informals
d'acompa...
## 10     45 E0              1  1995  2001 2 [Pràctiques retribuïdes /
be...

```

Finalment tenim 7 nivells d'esdeveniments possibles de les trajectòries laborals, amb els següents nivells

```

nivells.l<- c("pràctiques", "treball mateixa empresa", "ocupacions precaries",
"atur", "treball reproductiu", "feinetes acompanyament estudis", "No treball")

```

Generem l'objecte seqüència per edats:

```

Edat<-filter(EJC12, id %in% seq.long.l$id)
Edat<-Edat[,c(1,11)]

edats.lab <- seqformat(seq.long.l, from = "SPELL", to = "STS",
  id = "id", begin = "data.i", end = "data.f", status = "esdeveniment", process
= TRUE, pdata=Edat, pvar=c("id", "v1"), limit=30)
## [!!] 1054 sequences truncated at limit 30 . Indexes in attribute 'issues'
## [>] converting SPELL data into 1088 STS sequences (internal format)

```



```

#edats.lab[is.na(edats.lab)]<-7
edats.lab <- seqdef(edats.lab, alphabet = 1:7,
  labels = nivells.l)
## [>] found missing values ('NA') in sequence data
## [>] preparing 1088 sequences
## [>] coding void elements with '%' and missing values with '*'
## [>] 7 distinct states appear in the data:
##      1 = 1
##      2 = 2
##      3 = 3
##      4 = 4
##      5 = 5
##      6 = 6
##      7 = 7
## [>] state coding:
##      [alphabet] [label] [long label]
##      1          1      1pràctiques
##      2          2      2treball mateixa empresa
##      3          3      3ocupacions precaries
##      4          4      4atur
##      5          5      5treball reproductiu
##      6          6      6feinetes acompanyament estudis
##      7          7      7No treball
## [>] 1088 sequences in the data set
## [>] min/max sequence length: 17/30
edats.lab<-edats.lab[,-c(1:15)]
colnames(edats.lab)<-gsub("a", "", as.character(colnames(edats.lab)))
alphabet(edats.lab) <- nivells.l

```

visualització seqüències creades

```

edats.lab[1:5,]

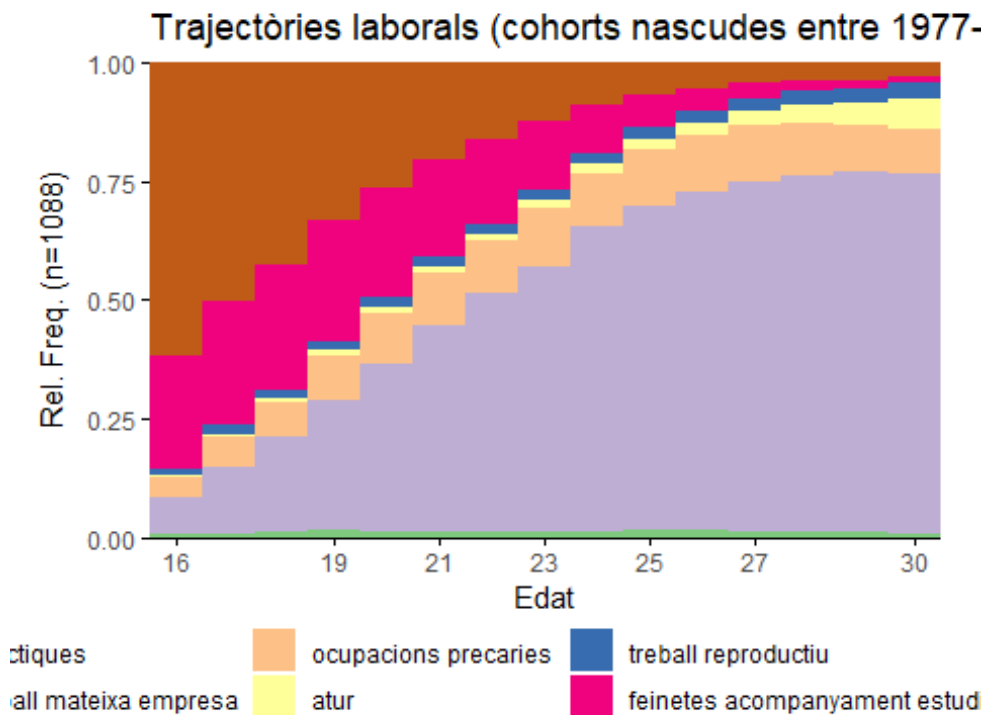
##      Sequence
## 1 No treball-No treball-No treball-No treball-pràctiques-pràctiques-treball
mateixa empresa-treball mateixa empresa-treball mateixa empresa-treball mateixa
empresa-treball mateixa empresa-treball mateixa empresa-treball mateixa empresa-
treball mateixa empresa-treball mateixa empresa
## 4 feinetes acompanyament estudis-feinetes acompanyament estudis-feinetes
acompanyament estudis-feinetes acompanyament estudis-feinetes acompanyament
estudis-feinetes acompanyament estudis-feinetes acompanyament estudis-feinetes
acompanyament estudis-feinetes acompanyament estudis-feinetes acompanyament
estudis-atur-treball mateixa empresa-treball mateixa empresa-treball mateixa
empresa-treball mateixa empresa
## 45 pràctiques-pràctiques-pràctiques-pràctiques-pràctiques-pràctiques-treball
mateixa empresa-treball mateixa empresa-treball mateixa empresa-treball mateixa
empresa-treball mateixa empresa-treball mateixa empresa-treball mateixa empresa-
treball mateixa empresa-treball mateixa empresa
## 47 ocupacions precaries-ocupacions precaries-ocupacions precaries-ocupacions
precaries-ocupacions precaries-ocupacions precaries-treball mateixa empresa-
treball mateixa empresa-treball mateixa empresa-treball mateixa empresa-treball
mateixa empresa-treball mateixa empresa-treball mateixa empresa-treball mateixa
empresa-treball mateixa empresa
## 55 No treball-No treball-No treball-No treball-No treball-No treball-No

```

```
treball-No treball-No treball-No treball-pràctiques-treball mateixa empresa-  
treball mateixa empresa-treball mateixa empresa-treball mateixa empresa
```

Gràfic distribució dels estats per freqüències relatives de les trajectòries laborals (grafic 1)

```
lab.plot<-ggseqdplot(edats.lab, ltext=names(nivells.e), border=F) +  
  labs(title = "Trajectòries laborals (cohorts nascudes entre 1977-1982)",  
x="Edat")  
lab.plot
```



4.3. Càlcul de distàncies per OM multicanal

Cal mirar per quins cassos (ID) tenim ambdues trajectòries, i tornar-los a convertir en un objecte "SEQDEF".

```
edats.educatiu<-edats.educatiu[rownames(edats.educatiu) %in%  
rownames(edats.lab),]  
edats.lab<-edats.lab[rownames(edats.lab) %in% rownames (edats.educatiu),]
```

Mirem si es troben associades les dues trajectòries, fent OM per cada canal i un test de mantel per mirar si la correlació entre les dues matrius de distàncies generades és significativa.

```
dist.educ<-seqdist(edats.educatiu, method = "OM", with.missing = T, sm="TRATE")  
## [>] including missing values as an additional state  
## [>] 1087 sequences with 9 distinct states  
## [>] Computing sm with seqcost using TRATE  
## [>] creating substitution-cost matrix using transition rates ...
```

```

## [>] computing transition probabilities for states PQPI o escola
d'adults/ESO/Batxillerat/FP1/FP2/Diplomatures o Llicenciatures/Postgrau, Masters
o Doctorat/No estudiant/* ...
## [>] generated an indel of type number
## [>] 471 distinct sequences
## [>] min/max sequence lengths: 15/15
## [>] computing distances using the OM metric
## [>] elapsed time: 0.34 secs
dist.lab<-seqdist(edats.lab, method = "OM", with.missing = T, sm="TRATE")
## [>] including missing values as an additional state
## [>] 1087 sequences with 8 distinct states
## [>] Computing sm with seqcost using TRATE
## [>] creating substitution-cost matrix using transition rates ...
## [>] computing transition probabilities for states pràctiques/treball mateixa
empresa/ocupacions precaries/atur/treball reproductiu/feinetes acompanyament
estudis/No treball/* ...
## [>] generated an indel of type number
## [>] 486 distinct sequences
## [>] min/max sequence lengths: 2/15
## [>] computing distances using the OM metric
## [>] elapsed time: 0.86 secs
dist.educ<-as.dist(dist.educ)
dist.lab<-as.dist(dist.lab)
mantel.rtest(dist.educ, dist.lab, nrepet = 1000)
## Warning in is.euclid(m1): Zero distance(s)
## Warning in is.euclid(m2): Zero distance(s)
## Monte-Carlo test
## Call: mantelnoneuclid(m1 = m1, m2 = m2, nrepet = nrepet)
##
## Observation: 0.1872007
##
## Based on 1000 replicates
## Simulated p-value: 0.000999001
## Alternative hypothesis: greater
##
##      Std.Obs   Expectation   Variance
## 13.5672416629 -0.0008291700 0.0001920749
cor.test(dist.educ, dist.lab)
##
## Pearson's product-moment correlation
##
## data: dist.educ and dist.lab
## t = 146.41, df = 590239, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.1847378 0.1896612
## sample estimates:
##      cor
## 0.1872007

```

P-valor significatiu, hi ha relació (taula 7)

Visualització matrius que es faran servir com a cost de substitució (opció “TRATE”, matrius segons les probabilitats de transició entre estats)

Per la matriu de les trajectòries educatives (taula 8)

```
subs.educ<-seqsubm(edats.educatiu, method = "TRATE")

## [>] creating substitution-cost matrix using transition rates ...

## [>] computing transition probabilities for states PQPI o escola
d'adults/ESO/Batxillerat/FP1/FP2/Diplomatures o Llicenciatures/Postgrau, Masters
o Doctorat/No estudiant ...

subs.educ<-as.data.frame(subs.educ)
subs.educ<-round(subs.educ,3)
subs.educ$nivells <- rownames(subs.educ)
subs.educ<-subs.educ[,c(9,1,2,3,4,5,6,7,8)]
subs.educ<-subs.educ %>%
  gt() %>%
  cols_width(nivells~px(250),
             `Diplomatures o Llicenciatures`~px(110),
             everything() ~ px(100)) %>%
  data_color(columns = 2:9,
             fn = col_numeric(palette = c("white","dodgerblue"),
                              domain = c(1.600,1.999)))

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

## Warning: Some values were outside the color scale and will be treated as NA

subs.educ<-subs.educ %>%
  tab_style_body(
    style = list(
      cell_text(color = "black"),cell_fill(color = "white")),
    values = c(0.000))

subs.educ<-subs.educ %>%
  tab_style_body(
    style = cell_fill(color = "darkblue"),
    values = c(2.000))

#subs.educ %>% gtsave("subs.educ.png", vwidth = 1200, path
="C:/Users/santa/Documents/UAB/TFG/TFG ESTADISTICA" )
subs.educ
```



```

subs.lab<-subs.lab %>%
  tab_style_body(
    style = list(
      cell_text(color = "black"),cell_fill(color = "white")),
    values = c(0.000))

subs.lab<-subs.lab %>%
  tab_style_body(
    style = cell_fill(color = "darkblue"),
    values = c(2.000))

getwd()

## [1] "C:/Users/santa/Documents/UAB/TFG/TFG ESTADISTICA"

#subs.lab %>% gtsave("subs.lab.png",vwidth = 1500, path
="C:/Users/santa/Documents/UAB/TFG/TFG ESTADISTICA" )
#seqsubm(edats.lab, method="TRATE")
subs.lab

```

nivells	pràctiques	treball mateixa empresa	ocupacions precaries	atur	treball reproductiu	feinetes acompanyament estudis	No treball
pràctiques	0.000	1.763	1.976	1.994	2.000	1.993	1.957
treball mateixa empresa	1.763	0.000	1.892	1.791	1.957	1.879	1.860
ocupacions precaries	1.976	1.892	0.000	1.954	1.986	1.982	1.955
atur	1.994	1.791	1.954	0.000	2.000	1.996	1.987
treball reproductiu	2.000	1.957	1.986	2.000	0.000	2.000	1.996
feinetes acompanyament estudis	1.993	1.879	1.982	1.996	2.000	0.000	1.953
No treball	1.957	1.860	1.955	1.987	1.996	1.953	0.000

Distàncies per OM multichannel

```

mcdist <- seqdistmc(channels=list(edats.educatiu, edats.lab),
  method="OM", sm =list("TRATE", "TRATE"), with.missing=T ) #TRATE-> Agafem
prob. de transició com a costos.

## [!!] 2 domains with 1087 sequences

## [!!] Some individuals have channels of different length. Shorter sequences
will be filled with missing values and corresponding channel with.missing set as
TRUE

## [>] building MD sequences of combined states... OK
## [>] including missing value as an additional state
## [>] computing substitution cost matrix for domain 1
## [>] including missing value as an additional state
## [>] computing substitution cost matrix for domain 2
## [>] computing MD substitution and indel costs with additive trick... OK

```

```
## [>] computing distances using additive trick ...
## [>] 1087 sequences with 60 distinct states
## [>] checking 'sm' (size and triangle inequality)
## [>] 935 distinct sequences
## [>] min/max sequence lengths: 15/15
## [>] computing distances using the OM metric
## [>] elapsed time: 1.57 secs
```

4.4. Tipificació de les trajectòries escola-treball: clústers jeràrquics per Ward

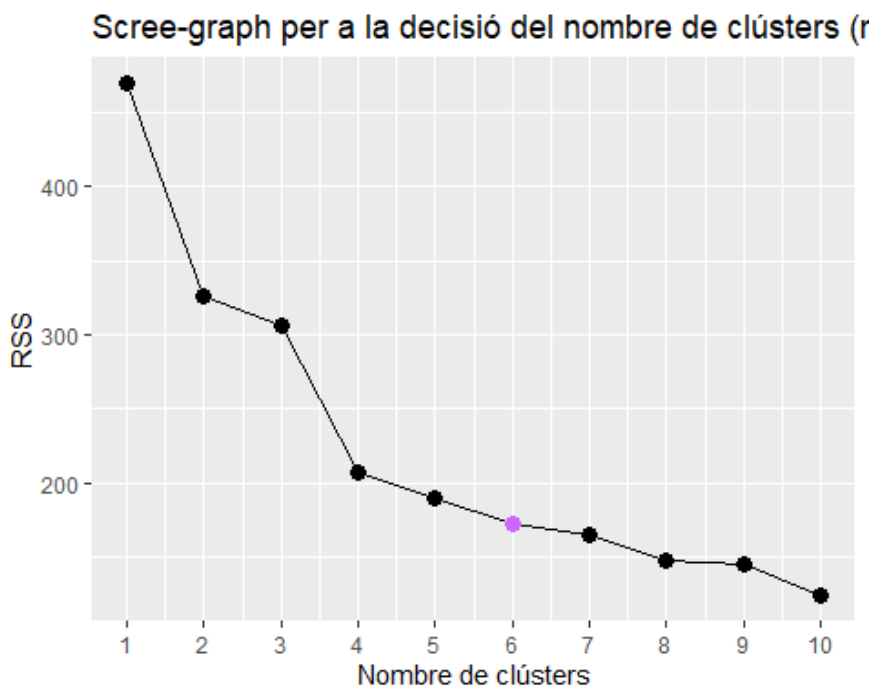
Generem els clústers per Ward

```
cluster<-agnes(mcdist, diss=T, method="ward")
```

Scree graph per la minimització de l'RSS (o WSS) com a criteri de decisió del numero de clusters (gràfic 2)

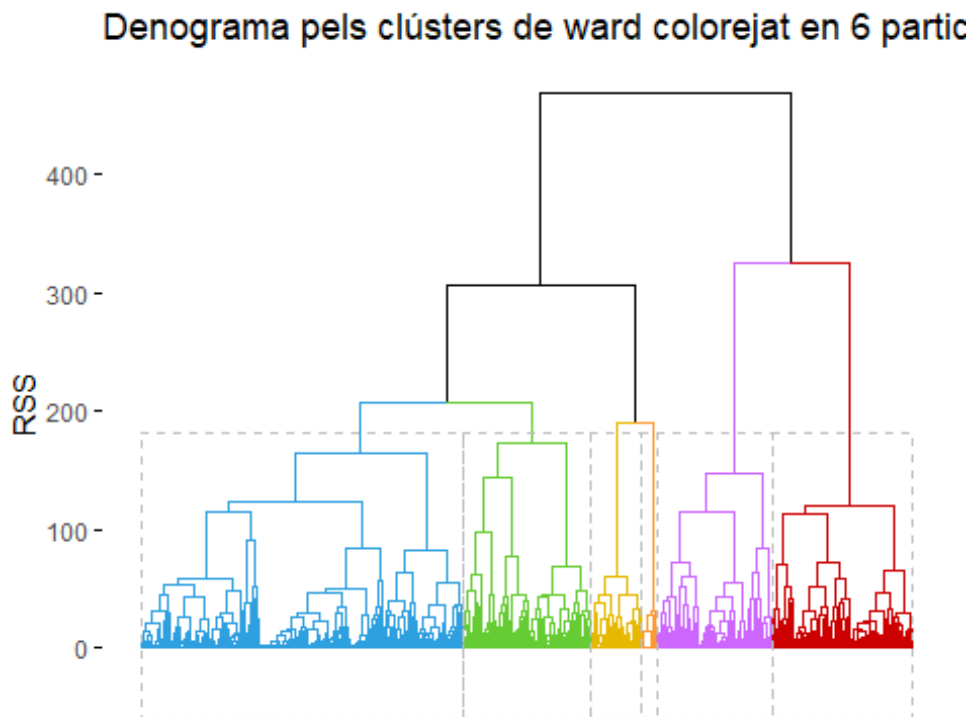
```
a<-sort(cluster$height, decreasing = TRUE)
a<-a[1:10]

qplot(x=c(1:10),y=a) +
  geom_line() +
  geom_point(size=3)+
  scale_x_continuous(name="Nombre de clústers",breaks=seq(0,10,1))+
  ylab("RSS") +
  ggtitle("Scree-graph per a la decisió del nombre de clústers
(minimització RSS o WSS)")+
  geom_point(aes(x=6,y=a[6]),
    color='#CC66FF',
    size=3)
```



Observem com quedaria el denograma amb sis particions (en colors) (gràfic 3)

```
dhc <- as.dendrogram(cluster)
fviz_dend(
  dhc,
  k = 6,
  rect = TRUE,
  k_colors = c("#2E9FDF", "#66CC33", "#E7B800", "#FF9933", "#CC66FF",
  "#CC0000"),
  show_labels = F, main = "Denograma pels clústers de ward colorejat en 6
particions", ylab = "RSS")
```



Scree graph per la minimització de l’RSS (o WSS) com a criteri de decisió del numero de clústers.

Tallem el denograma en sis clústers

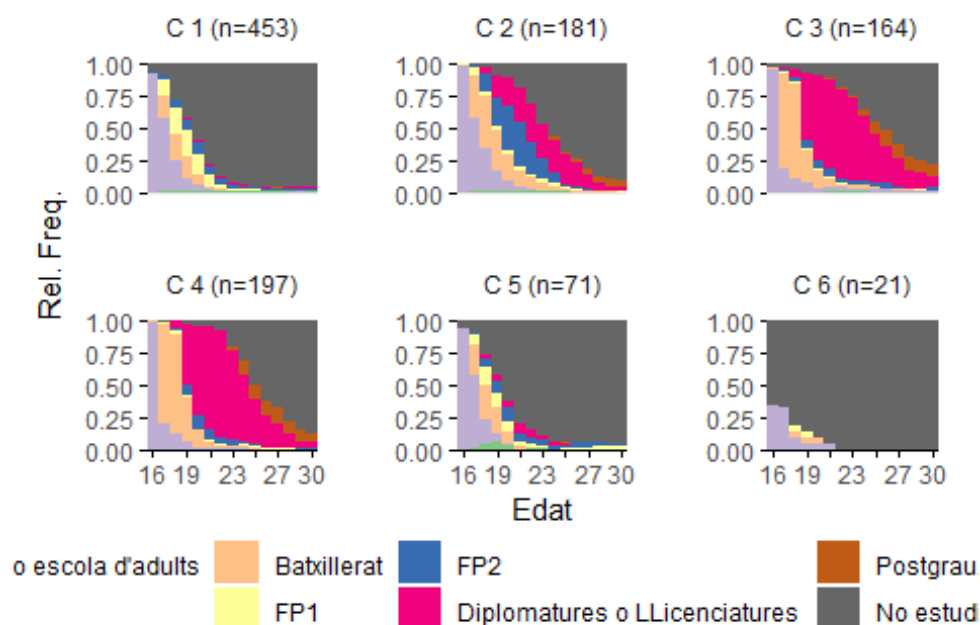
```
n.clus <- cutree(cluster, k = 6)
clusters <- factor(n.clus, labels = paste("C", 1:6))
```

4.5. Visualització i caracterització de les tipologies de trajectòries escola-treball

-Visualització de distribució dels estats per freqüència relativa, per clústers (gràfic 4)

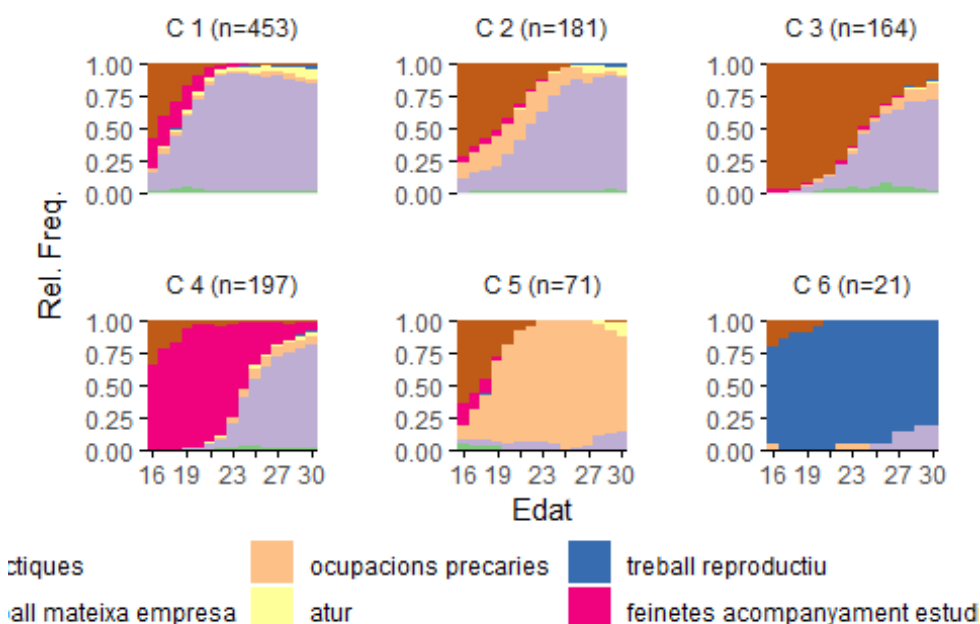
```
ggseqdplot(edats.educatiu, border=F, group=clusters)+
  labs(title = "Freqüència relativa dels estats en les trajectòries educatives
\n (distribució dels estats), per clústers", x="Edat")
```


Freqüència relativa dels estats en les trajectòries ec (distribució dels estats), per clústers



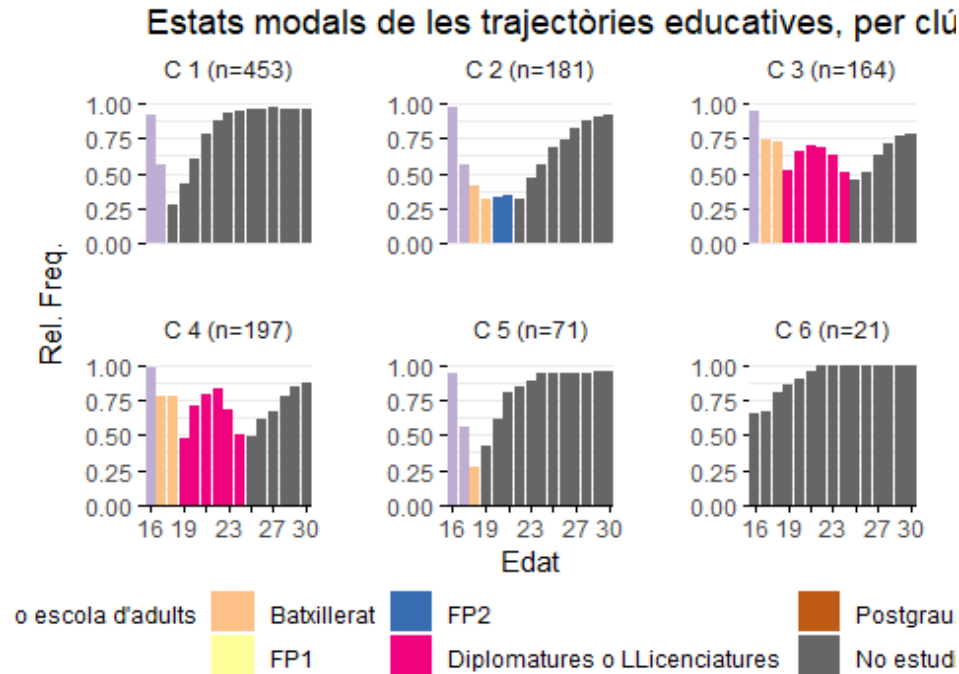
```
ggseqdplot(edats.lab, border=F, group=clusters) +  
  labs(title = "Freqüència relativa dels estats en les trajectòries laborals \n  
(distribució dels estats), per clústers", x="Edat")
```

Freqüència relativa dels estats en les trajectòries la (distribució dels estats), per clústers

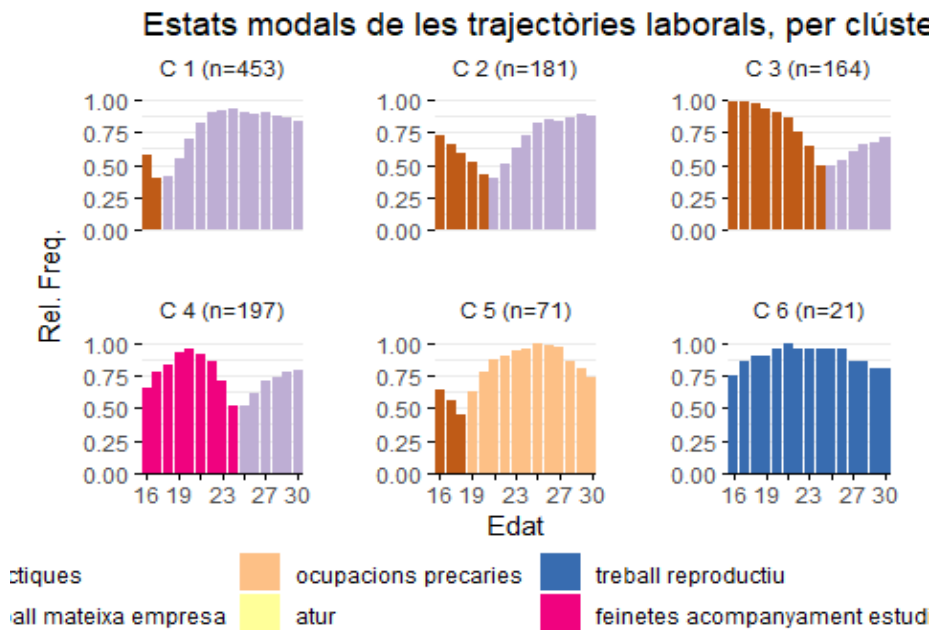


- Visualització dels estats modals per clústers (gràfic 5)

```
ggseqmsplot(edats.educatiu, group=clusters) +  
  labs(title = "Estats modals de les trajectòries educatives, per clústers",  
        x="Edat")
```



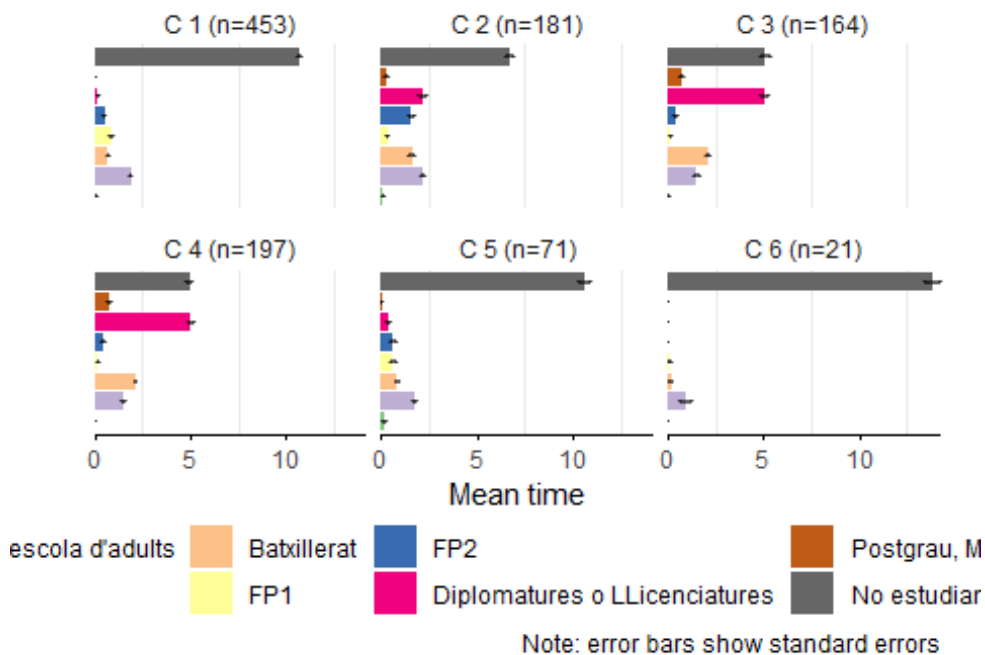
```
ggseqmsplot(edats.lab, group=clusters) +  
  labs(title = "Estats modals de les trajectòries laborals, per clústers",  
        x="Edat")
```



- Visualització temps mitjà en cada estat (gràfic 6)

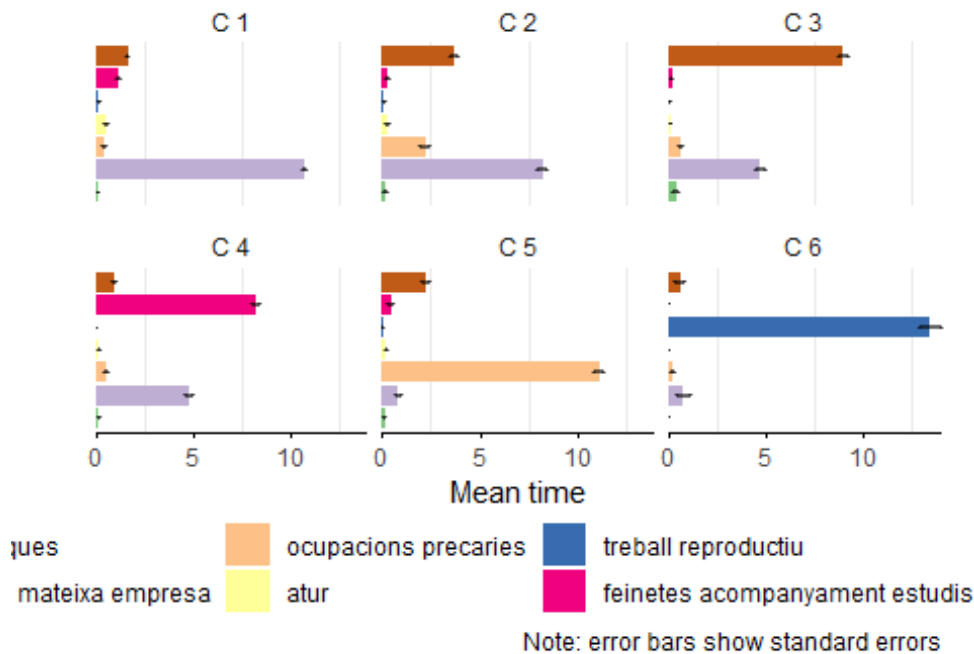
```
ggseqmtplot(edats.educatiu, no.n = F, error.bar = "SE", group = clusters) +
  coord_flip() +
  theme(axis.text.y=element_blank(),
        axis.ticks.y = element_blank(),
        panel.grid.major.y = element_blank())+
  labs(title = "Temps mitjà en cada estat per les trajectòries educatives, per
clústers")
```

Temps mitjà en cada estat per les trajectòries educatives



```
ggseqmtplot(edats.lab, no.n = T, error.bar = "SE", group = clusters) +
  coord_flip() +
  theme(axis.text.y=element_blank(),
        axis.ticks.y = element_blank(),
        panel.grid.major.y = element_blank())+
  labs(title = "Temps mitjà en cada estat per les trajectòries laborals, per
clústers")
```

Temps mitjà en cada estat per les trajectòries laborals,



4.6. Mesures de complexitat

Generem una llista amb les seqüències en cada cluster, per facilitar el càlcul mitjà de les mesures de complexitat en cada clúster.

```
#traj.educatvies
educatiu.c<-cbind(clusters,edats.educatiu)
llista.educ<-list(C1=data.frame(split(educatiu.c, educatiu.c$clusters)[1]),
                  C2=data.frame(split(educatiu.c, educatiu.c$clusters)[2]),
                  C3=data.frame(split(educatiu.c, educatiu.c$clusters)[3]),
                  C4=data.frame(split(educatiu.c, educatiu.c$clusters)[4]),
                  C5=data.frame(split(educatiu.c, educatiu.c$clusters)[5]),
                  C6=data.frame(split(educatiu.c, educatiu.c$clusters)[6]))

#canviem el nom

canvi.nom <- function(x) {
  names(x)<-c(16:30)
  return(x)
}
llista.educ<-lapply(llista.educ,`,-1)
llista.educ<-lapply(llista.educ, canvi.nom)
llista.educ<-lapply(llista.educ,seqdef)

#trajectòries laborals
lab.c<-cbind(clusters,edats.lab)
llista.lab<-list(C1=data.frame(split(lab.c, lab.c$clusters)[1]),
                 C2=data.frame(split(lab.c, lab.c$clusters)[2]),
                 C3=data.frame(split(lab.c, lab.c$clusters)[3]),
                 C4=data.frame(split(lab.c, lab.c$clusters)[4]),
```

```

C5=data.frame(split(lab.c, lab.c$clusters)[5]),
C6=data.frame(split(lab.c, lab.c$clusters)[6]))
#canviem el nom
canvi.nom <- function(x) {
  names(x)<-c(16:30)
  return(x)
}
llista.lab<-lapply(llista.lab,`,-1)
llista.lab<-lapply(llista.lab, canvi.nom)
llista.lab<-lapply(llista.lab,seqdef)

```

- Numero de canvis mitjà

```

canvi.seq <- function(x){
  mean(seqtransn(x))
}
#seq.laboral
trans.l<-lapply(llista.educ,canvi.seq) #grafics trajectories Laborals
trans.l<-as.data.frame(trans.l)

#seq.educativa
trans.e<-lapply(llista.lab,canvi.seq) #grafics trajectories Laborals
trans.e<-as.data.frame(trans.e)

```

-Índex de complexitat

```

complex <- function(x){
  mean(seqici(x))
}
#seq.laboral
complex.l<-lapply(llista.educ,complex) #grafics trajectories Laborals
complex.l<-as.data.frame(complex.l)

#seq.educativa
complex.e<-lapply(llista.lab,complex) #grafics trajectories Laborals
complex.e<-as.data.frame(complex.e)

```

Entropia

```

entropy <- function(x){
  mean(seqient(x))
}
#seq.laboral
entropy.l<-lapply(llista.educ,entropy) #grafics trajectories Laborals
entropy.l<-as.data.frame(entropy.l)

#seq.educativa
entropy.e<-lapply(llista.lab,entropy) #grafics trajectories Laborals
entropy.e<-as.data.frame(entropy.e)

```

Taula dels valors de les característiques de les trajectòries educatives

```

caracteristiques<-c("nombre transicions","complexity index","entropy")
caracteristiques.l<-rbind(trans.l,complex.l, entropy.l)
caracteristiques.l<-cbind(caracteristiques,caracteristiques.l)
caracteristiques.e<-rbind(trans.e, complex.e, entropy.e)

```

```

caracteristiques.e<-cbind(caracteristiques,caracteristiques.e)
caracteristiques.e

##      caracteristiques      C1      C2      C3      C4      C5
## 1 nombre transicions 1.5452539 1.6685083 1.3109756 1.6142132 1.3943662
## 2 complexity index 0.1695730 0.2032556 0.1601039 0.1990887 0.1663970
## 3 entropy 0.2665783 0.3562765 0.2848968 0.3511129 0.2844096
##      C6
## 1 0.57142857
## 2 0.07754181
## 3 0.15000343

caracteristiques.l

##      caracteristiques      C1      C2      C3      C4      C5
## 1 nombre transicions 1.8211921 2.8287293 3.1890244 3.5380711 2.1690141
## 2 complexity index 0.2012438 0.3211268 0.3396626 0.3873106 0.2262103
## 3 entropy 0.3157450 0.5171078 0.5110019 0.5988734 0.3363405
##      C6
## 1 0.52380952
## 2 0.06852144
## 3 0.12765157

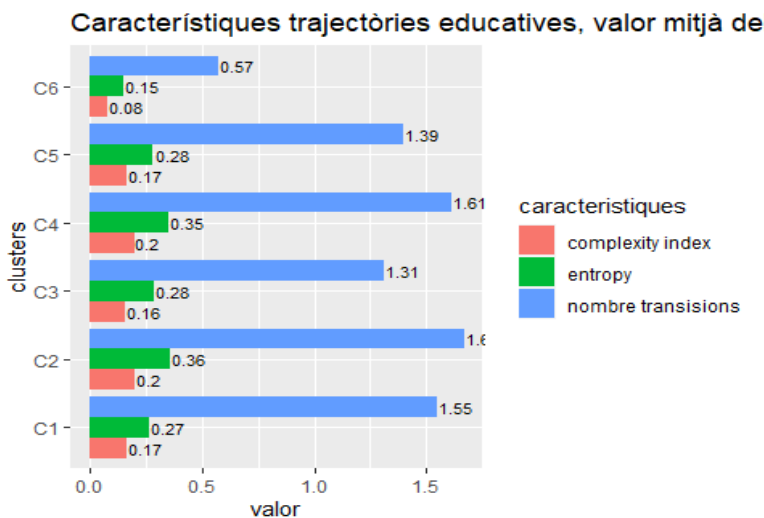
```

Visualització resumida de les mitjanes per clúster, trajectòries educatives (gràfic 7)

```

car.e<-caracteristiques.e%>%pivot_longer(!caracteristiques,
names_to="clusters",values_to="valor")
ggplot(car.e, aes(fill=caracteristiques, y=valor, x=clusters)) +
  geom_bar(position="dodge", stat="identity")+
  geom_text(aes(label = round(valor,2)), ,
  hjust = -0.05, size = 3,
  position = position_dodge(width = 0.9),
  inherit.aes = TRUE )+
  coord_flip()+
  ggtitle("Característiques trajectòries educatives, valor mitjà dels cústers")

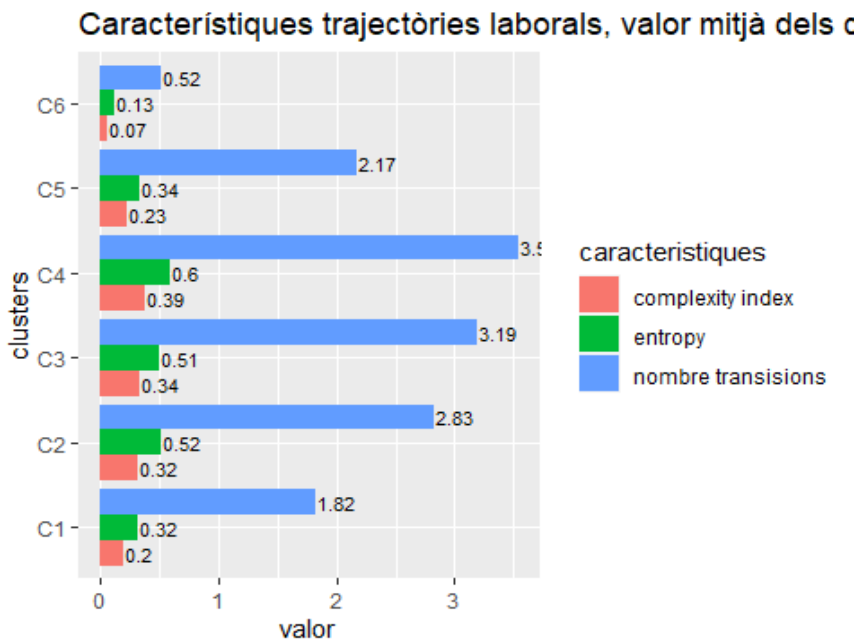
```



Visualització resumida de les mitjanes per clúster, trajectòries larboals (gràfic 7)

```
car.l<-caracteristiques.l%>%pivot_longer(!caracteristiques,
names_to="clusters",values_to="valor")

ggplot(car.l, aes(fill=caracteristiques, y=valor, x=clusters)) +
  geom_bar(position="dodge", stat="identity")+
  geom_text(aes(label = round(valor,2)), ,
  hjust = -0.05, size = 3,
  position = position_dodge(width = 0.9),
  inherit.aes = TRUE )+
  coord_flip()+
  ggtitle("Característiques trajectòries laborals, valor mitjà dels custers")
```



4.7. L'impacte de l'estratificació social en les trajectòries (models de regressió logística)

A partir d'una llista i una funció, generem una taula amb el p-valor i l'exp(coef) per sis models de regressió logística on la variable resposta binària és ser del cluster X contra no ser-ho. (taula 10 *s'exporta a excel)

```
variables.c<-variables[variables$id %in% rownames(educatiu.c), ]
```

```
model.clus<-list(c1= (clusters=="C 1"),
                 c2= (clusters=="C 2"),
                 c3= (clusters=="C 3"),
                 c4= (clusters=="C 4"),
                 c5= (clusters=="C 5"),
                 c6= (clusters=="C 6"))
```

#generem una funció que torna una taula amb el p-valor i l'exp(coef) on la variable resposta binària: ser del cluster X contra no ser-ho.

```
model<-function(x){
m1<-glm(x~ sexe +
```

```

lloc.naix + c.socioecon + n.estudis +
  cohort + municipi, data = variables.c)
a<-as.data.frame(summary(m1)$coefficients)[,c(1,4)]

a$odds<-exp(a$Estimate)
# B<-a[, -c(1)]
a<-a[, -c(1)]
a$`Pr(>|t|)`<-round(a$`Pr(>|t|)`,4)
#<-rownames(a)

return(a)
}

taules<-lapply(model.clus, model)

model.s<-cbind(taules[[1]], taules[[2]], taules[[3]], taules[[4]], taules[[5]],
taules[[6]])
colnames(model.s)<-c("p-value.c1", "exp(coef).c1", "p-value.c2",
"exp(coef).c2", "p-value.c3", "exp(coef).c3", "p-value.c4", "exp(coef).c4", "p-
value.c5", "exp(coef).c5", "p-value.c6", "exp(coef).c6")
model.s

##                p-value.c1 exp(coef).c1 p-value.c2
## (Intercept)          0.3057 7.568321e-09      0.7734
## sexehome             0.0000 1.133422e+00      0.8078
## lloc.naixEsp          0.5010 1.051414e+00      0.6689
## lloc.naixaltre_País   0.1295 1.060407e+00      0.7878
## c.socioeconcontencins_professionals 0.6659 1.024896e+00      0.3451
## c.socioeconocup.quali.mitjana      0.0426 1.113308e+00      0.8066
## c.socioeconocup.elementals          0.1241 1.109463e+00      0.4709
## c.socioeconocup.militars            0.2194 7.399186e-01      0.7132
## n.estudisoblifatoris                0.2710 1.128522e+00      0.3214
## n.estudissecundaris_postoblig       0.9983 9.997662e-01      0.2459
## n.estudisuniversitaris              0.5344 9.315858e-01      0.1860
## cohort                             0.3007 1.009588e+00      0.7772
## municipimax_rural                   0.3262 1.053596e+00      0.4300
## municipimitj_rural                  0.5950 9.789248e-01      0.6294
## municipiurbà                        0.0145 1.098843e+00      0.1995
##                exp(coef).c2 p-value.c3 exp(coef).c3
## (Intercept)        66.8114329      0.2788 2.670848e+06
## sexehome            1.0059164      0.5425 9.862670e-01
## lloc.naixEsp         0.9748535      0.8489 1.010683e+00
## lloc.naixaltre_País  0.9917139      0.5539 1.017275e+00
## c.socioeconcontencins_professionals 0.9579183      0.4296 9.668936e-01
## c.socioeconocup.quali.mitjana      0.9897010      0.0336 9.192503e-01
## c.socioeconocup.elementals          1.0396876      0.0135 8.824819e-01
## c.socioeconocup.militars            1.0746852      0.1129 7.474826e-01
## n.estudisoblifatoris                1.0909743      0.2408 9.080728e-01
## n.estudissecundaris_postoblig       1.1102449      0.6088 9.577565e-01
## n.estudisuniversitaris              1.1282257      0.9097 9.903638e-01
## cohort                             0.9979169      0.2891 9.927112e-01
## municipimax_rural                   0.9670134      0.3416 9.628595e-01
## municipimitj_rural                  1.0155780      0.5738 9.832742e-01
## municipiurbà                        1.0402813      0.0049 9.219707e-01
##                p-value.c4 exp(coef).c4 p-value.c5

```


## (Intercept)	0.2910	3.588467e+06	0.1080
## sexehome	0.0000	8.918359e-01	0.2138
## lloc.naixEsp	0.9865	9.990109e-01	0.1754
## lloc.naixaltre_País	0.0000	8.520975e-01	0.0246
## c.socioeconocup.quali.mitjana	0.1821	1.061328e+00	0.9847
## c.socioeconocup.elementals	0.8878	9.941743e-01	0.8199
## c.socioeconocup.militars	0.7585	9.838814e-01	0.7399
## n.estudisoblifatoris	0.0456	1.468364e+00	0.1489
## n.estudissecundaris_postoblig	0.7442	1.028454e+00	0.9075
## n.estudisuniversitaris	0.1770	1.126581e+00	0.7016
## cohort	0.1062	1.155328e+00	0.7192
## municipimax_rural	0.2980	9.925147e-01	0.1072
## municipimitj_rural	0.5314	9.742841e-01	0.2327
## municipiurbà	0.6219	9.846447e-01	0.1004
##	0.0033	9.151400e-01	0.0736
##	exp(coef).c5	p-value.c6	exp(coef).c6
## (Intercept)	2.818678e-07	0.8787	1.9899428
## sexehome	1.019573e+00	0.0031	0.9780223
## lloc.naixEsp	9.494208e-01	0.3391	1.0177547
## lloc.naixaltre_País	1.045710e+00	0.0000	1.0490644
## c.socioeconocup.quali.mitjana	1.000560e+00	0.5691	0.9920204
## c.socioeconocup.elementals	9.938316e-01	0.9538	0.9992438
## c.socioeconocup.militars	1.011583e+00	0.4342	0.9870384
## n.estudisoblifatoris	1.199509e+00	0.4493	0.9552051
## n.estudissecundaris_postoblig	1.006579e+00	0.0000	0.8640163
## n.estudisuniversitaris	9.780565e-01	0.0000	0.8536917
## cohort	9.791483e-01	0.0000	0.8492445
## municipimax_rural	1.007664e+00	0.9079	0.9997366
## municipimitj_rural	1.033145e+00	0.3365	1.0127063
## municipiurbà	1.034426e+00	0.6613	1.0043473
##	1.036041e+00	0.9360	1.0007641

```
model.s<-round(model.s, 3)
```

```
openxlsx::write.xlsx(model.s,"model.xlsx", rowNames=TRUE) #guardem a excel per
despres posar-la al document del treball.
```

6. Anàlisi de supervivència a la llar familiar

Codifiquem les matrius de la trajectòria residencial de l'EJC 2012

```
traj.hab<-EJC12[,c(1,401:521)]
noms12<- c("v62a0","v62a2", "v62b11", "v62b12")
traj.hab<-traj.hab %>%
  dplyr::select(id,starts_with(noms12))

noms.f.12<-c("situacio","actual","data.i","data.f")

contador<-rep(paste0(" - E",c(0:8)),4)
col.names12h<-paste0(rep(noms.f.12,1,each=9),contador)
col.names12h<-c("id",col.names12h) #vector de nous noms de columnes

colnames(traj.hab)<-col.names12h
colnames(traj.hab) <- gsub(" ", "", colnames(traj.hab))
```

```
#format Long:
long.hab<-traj.hab %>%
  pivot_longer(-id,names_to= c(".value", "Emancipat"), names_sep="-")
```

Passem els diferents estats a estar emancipat o no

```
long.hab$data.i<-as.integer(format(long.hab$data.i, format="%Y"))
long.hab$data.f<-as.integer(format(long.hab$data.f, format="%Y"))
```

```
#si la situació és actual(=1), data.f-> 2012
```

```
for (i in 1:nrow(long.hab)){
  if(!is.na(long.hab$actual[i])){
    if(long.hab$actual[i]==1){
      long.hab$data.f[i]<-2012
    }
  }
}
```

```
long.hab<-long.hab %>%
  drop_na(situacio)
long.hab<-long.hab[,c(-4)] #ja hem codificat si era actual
```

Codifiquem les diferents situacions com emancipat o no emancipat.

```
long.hab$situacio[long.hab$situacio==14|long.hab$situacio==15]<-NA #no sabem si
es troba emancipat o no
long.hab<-long.hab %>%
  drop_na(situacio)

for(i in 1:nrow(long.hab)){
  if
  (long.hab$situacio[i]==4|long.hab$situacio[i]==6|long.hab$situacio[i]==7|long.ha
b$situacio[i]==8|long.hab$situacio[i]==10|long.hab$situacio[i]==11|long.hab$situ
acio[i]==13){
    long.hab$situacio[i]<-1
  }
  else{ long.hab$situacio[i]<-0}
}
```

Convertim en una base de dades amb censura i el temps de la primera emancipació.

```
#dades fins la primera emancipació (fins al primer 1)
long.hab$situacio<-as.numeric(long.hab$situacio)
hab.1r<-long.hab %>%
  group_by(id) %>%
  filter(lag(cumsum(situacio=="1"), default=0)<1) %>%
  ungroup

hab.1r<-hab.1r[!duplicated(hab.1r$id, fromLast=T),] #ultima observacio per tenir
columna censura,
hab.1r<-hab.1r[,c(1,3,4)]
#censura--> tots surten de la base de dades al 2012 (tindran edats diferents)
for(i in 1:nrow(hab.1r)){
  if (hab.1r$situacio[i]==0) {
```

```

    hab.1r$data.i[i]=2012
  }
}

```

Afegim a les dades, a quin clúster pertanyen

```

id.clus<-cbind(rownames(educatiu.c), educatiu.c$clusters) #indiferent si es amb Lab,
colnames(id.clus)<-c("id","cluster")
hab.1r<-merge(hab.1r,id.clus, by="id" )
hab.1r<-merge(hab.1r,variables,by="id")

```

Convertim la data de la primera emancipació com a edat, a partir dels anys de naixemen dels individus.

```

hab.1r<-hab.1r %>% mutate(data.i=data.i-cohort)
colnames(hab.1r)[3]<-"temps"
hab.1r$cluster<-as.factor(hab.1r$cluster)

```

Corba de Kaplan-meyer per clústers (taula 11)

```

clus.fit <- survfit(Surv(temps, situacio)~as.factor(cluster), data=hab.1r)
clus.fit

```

```

## Call: survfit(formula = Surv(temps, situacio) ~ as.factor(cluster),
##      data = hab.1r)
##

```

	n	events	median	0.95LCL	0.95UCL
## as.factor(cluster)=1	453	391	25	24	25
## as.factor(cluster)=2	181	155	25	24	26
## as.factor(cluster)=3	164	135	25	24	26
## as.factor(cluster)=4	197	179	24	22	25
## as.factor(cluster)=5	71	44	25	23	NA
## as.factor(cluster)=6	21	18	21	19	29

Log-rank test, són diferents les corbes?

```

cluster.longrk<-survdif(Surv(temps, situacio)~as.factor(cluster), data=hab.1r)
cluster.longrk

```

```

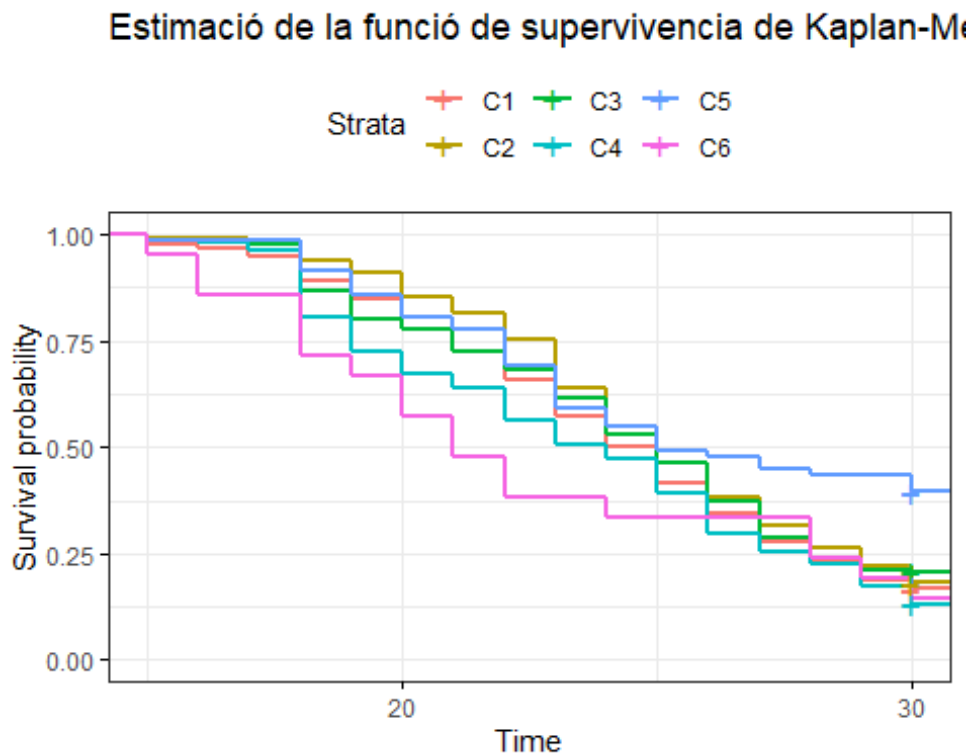
## Call:
## survdiff(formula = Surv(temps, situacio) ~ as.factor(cluster),
##      data = hab.1r)
##
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(cluster)=1 453      391      376.5      0.558      1.064
## as.factor(cluster)=2 181      155      164.8      0.578      0.794
## as.factor(cluster)=3 164      135      144.4      0.614      0.822
## as.factor(cluster)=4 197      179      149.3      5.896      7.928
## as.factor(cluster)=5  71       44       72.6     11.248     13.882
## as.factor(cluster)=6  21       18       14.4      0.892      1.021
##
##  Chisq= 22.4  on 5 degrees of freedom, p= 4e-04

```

El p valor és més petit de 0.05, així que la diferència entre les corbes es significativa amb un nivell de confiança del 95%.

Visualització de les corbes de supervivència (gràfic 8)

```
ggsurvplot(
clus.fit,
ggtheme = theme_bw(),
title = "Estimació de la funció de supervivència de Kaplan-Meier per clústers",
legend.labs=c("C1", "C2", "C3", "C4", "C5", "C6"),
xlim=c(15,30)
)
```



Model de COX, només amb la variable clústers (annex III)

```
cox.clus<-coxph(Surv(temps, situacio)~as.factor(cluster), data=hab.1r)
cox.clus

## Call:
## coxph(formula = Surv(temps, situacio) ~ as.factor(cluster), data = hab.1r)
##
##               coef exp(coef) se(coef)      z      p
## as.factor(cluster)2 -0.10143   0.90354  0.09493 -1.069 0.285277
## as.factor(cluster)3 -0.11133   0.89464  0.09993 -1.114 0.265231
## as.factor(cluster)4  0.15169   1.16380  0.09026  1.681 0.092856
## as.factor(cluster)5 -0.57351   0.56354  0.15935 -3.599 0.000319
## as.factor(cluster)6  0.17274   1.18856  0.24121  0.716 0.473893
##
## Likelihood ratio test=23.82 on 5 df, p=0.0002353
## n= 1087, number of events= 922
```

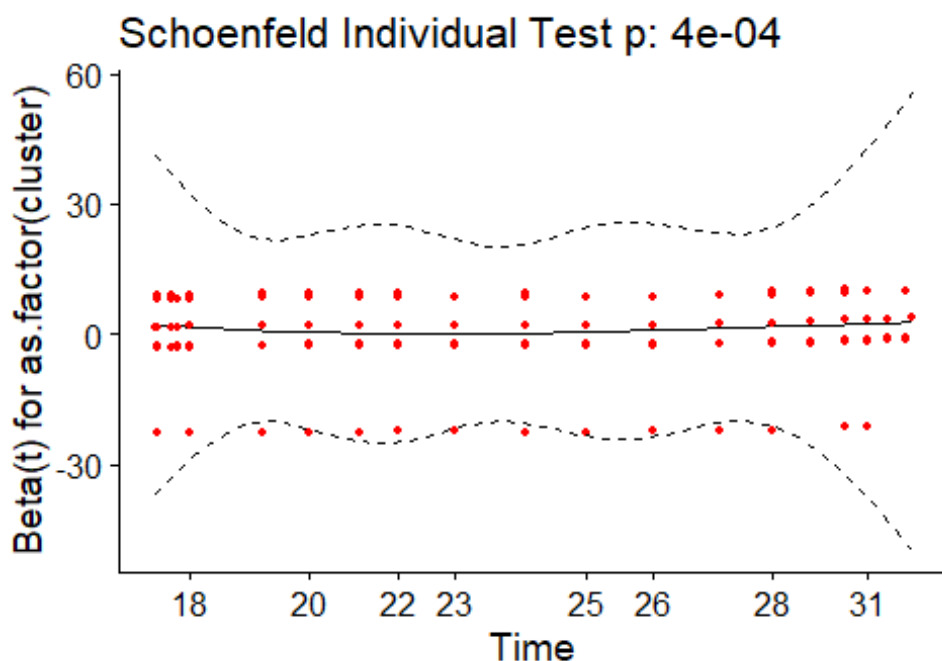
Comprovem si el supòsit de riscos proporcionals és compleix per mirar si és adequat el model:

```
test.ph<-cox.zph(cox.clus)
test.ph
```

```
##                chisq df      p
## as.factor(cluster) 22.8  5 0.00037
## GLOBAL             22.8  5 0.00037
```

```
ggcoxzph(test.ph)
```

Global Schoenfeld Test p: 0.0003688



Model de COX amb altres covairables (annex III)

```
cox.multi<-coxph(Surv(temps, situacio)~cluster+sexe+lloc.naix+
c.socioecon+n.estudis+cohort+municipi,data=hab.1r)
cox.multi
```

```
## Call:
```

```
## coxph(formula = Surv(temps, situacio) ~ cluster + sexe + lloc.naix +
##       c.socioecon + n.estudis + cohort + municipi, data = hab.1r)
```

```
##
```

	coef	exp(coef)	se(coef)	z	p
## cluster2	-0.18418	0.83178	0.09963	-1.849	0.064499
## cluster3	-0.19011	0.82687	0.10932	-1.739	0.082031
## cluster4	0.16817	1.18314	0.10178	1.652	0.098472
## cluster5	-0.62644	0.53449	0.17128	-3.657	0.000255
## cluster6	-0.41523	0.66019	0.30449	-1.364	0.172666
## sexehome	-0.22991	0.79460	0.07144	-3.218	0.001290
## lloc.naixEsp	0.22865	1.25691	0.16717	1.368	0.171372

```
## lloc.naixaltre_País      0.49222    1.63594    0.08881    5.543    2.98e-08
## c.socioeconòmic_professionals -0.13584    0.87298    0.12914    -1.052    0.292842
## c.socioeconocup.quali.mitjana -0.23483    0.79070    0.12024    -1.953    0.050824
## c.socioeconocup.elementals -0.30911    0.73410    0.15738    -1.964    0.049522
## c.socioeconocup.militars -1.67607    0.18711    0.72262    -2.319    0.020372
## n.estudisoblifatoris      0.02356    1.02384    0.24968    0.094    0.924819
## n.estudissecundaris_postoblig 0.06565    1.06785    0.25575    0.257    0.797416
## n.estudisuniversitaris     0.03855    1.03930    0.25861    0.149    0.881504
## cohort                    -0.01362    0.98647    0.02097    -0.650    0.515925
## municipimax_rural          -0.27226    0.76165    0.13203    -2.062    0.039191
## municipimitj_rural          0.12673    1.13511    0.09138    1.387    0.165474
## municipiurbà               0.16428    1.17854    0.08836    1.859    0.062986
##
## Likelihood ratio test=101.4 on 19 df, p=2.96e-13
## n= 997, number of events= 848
## (90 observations deleted due to missingness)
```

Refem el model, nivell estudis pares + cohort, no són significatius (annex III)

```
cox.multi<-coxph(Surv(temps,
situacio)~cluster+as.factor(sexe)+as.factor(lloc.naix)+
as.factor(c.socioecon)+as.factor(municipi),data=hab.1r)
cox.multi

## Call:
## coxph(formula = Surv(temps, situacio) ~ cluster + as.factor(sexe) +
##       as.factor(lloc.naix) + as.factor(c.socioecon) + as.factor(municipi),
##       data = hab.1r)
##
##
##               coef exp(coef) se(coef)
z
## cluster2      -0.17962    0.83559    0.09857 -
1.822
## cluster3      -0.18893    0.82784    0.10685 -
1.768
## cluster4       0.19111    1.21059    0.09951
1.920
## cluster5      -0.64186    0.52631    0.16849 -
3.809
## cluster6      -0.37187    0.68944    0.29190 -
1.274
## as.factor(sexe)home -0.22225    0.80072    0.07076 -
3.141
## as.factor(lloc.naix)Esp    0.17812    1.19497    0.16244
1.097
## as.factor(lloc.naix)altre_País    0.49310    1.63739    0.08655
5.697
## as.factor(c.socioecon)tèncincs_professionals -0.14758    0.86280    0.12733 -
1.159
## as.factor(c.socioecon)ocup.quali.mitjana -0.24071    0.78607    0.11706 -
2.056
## as.factor(c.socioecon)ocup.elementals -0.33320    0.71663    0.14992 -
2.223
## as.factor(c.socioecon)ocup.militars -1.47214    0.22943    0.59457 -
2.476
```

```
## as.factor(municipi)max_rural      -0.31277    0.73142    0.13150 -
2.378
## as.factor(municipi)mitj_rural      0.11139    1.11784    0.09084
1.226
## as.factor(municipi)urbà            0.17598    1.19242    0.08744
2.013
##                                     p
## cluster2                          0.068428
## cluster3                          0.077034
## cluster4                          0.054797
## cluster5                          0.000139
## cluster6                          0.202671
## as.factor(sexe)home                0.001685
## as.factor(lloc.naix)Esp            0.272851
## as.factor(lloc.naix)altre_País     1.22e-08
## as.factor(c.socioecon)tèncincs_professinals 0.246466
## as.factor(c.socioecon)ocup.quali.mitjana 0.039759
## as.factor(c.socioecon)ocup.elementals 0.026249
## as.factor(c.socioecon)ocup.militars 0.013288
## as.factor(municipi)max_rural      0.017384
## as.factor(municipi)mitj_rural     0.220112
## as.factor(municipi)urbà           0.044149
##
## Likelihood ratio test=106 on 15 df, p=9.196e-16
## n= 1014, number of events= 859
## (73 observations deleted due to missingness)
```

Mirem l'assumpció de riscos proporcionals del model (annex III)

```
test.ph<-cox.zph(cox.multi)
test.ph

##               chisq df      p
## cluster      20.72  5 0.00092
## as.factor(sexe) 6.83  1 0.00898
## as.factor(lloc.naix) 2.70  2 0.25883
## as.factor(c.socioecon) 5.59  4 0.23198
## as.factor(municipi) 7.47  3 0.05846
## GLOBAL       40.19 15 0.00042
```

Com que no es compleix per lloc naix i c.socioecon, hi posem strata (taula 13)

```
cox.multi<-coxph(Surv(temps,
situacio)~cluster+as.factor(sexe)+strata(as.factor(lloc.naix))+
strata(as.factor(c.socioecon))+as.factor(municipi),data=hab.1r)
cox.multi

## Call:
## coxph(formula = Surv(temps, situacio) ~ cluster + as.factor(sexe) +
##       strata(as.factor(lloc.naix)) + strata(as.factor(c.socioecon)) +
##       as.factor(municipi), data = hab.1r)
##
##               coef exp(coef) se(coef)      z      p
## cluster2      -0.17069    0.84308  0.09972 -1.712 0.086943
## cluster3      -0.17482    0.83961  0.10783 -1.621 0.104959
## cluster4       0.20307    1.22516  0.10052  2.020 0.043359
```

```
## cluster5          -0.63678    0.52899    0.16960 -3.755 0.000174
## cluster6          -0.25319    0.77632    0.29721 -0.852 0.394270
## as.factor(sexe)home -0.22872    0.79555    0.07148 -3.200 0.001376
## as.factor(municipi)max_rural -0.30142    0.73977    0.13246 -2.276 0.022872
## as.factor(municipi)mitj_rural  0.09896    1.10402    0.09198  1.076 0.281973
## as.factor(municipi)urbà      0.16102    1.17470    0.08865  1.816 0.069323
##
## Likelihood ratio test=58.87 on 9 df, p=2.21e-09
## n= 1014, number of events= 859
## (73 observations deleted due to missingness)
```

Tornem a mirar si els riscos són proporcionals

```
test.ph<-cox.zph(cox.multi)
test.ph

##              chisq df      p
## cluster      19.53  5 0.0015
## as.factor(sexe)   5.94  1 0.0148
## as.factor(municipi) 7.43  3 0.0593
## GLOBAL         29.66  9 0.0005
```

Més o menys, municipi poc significatiu, però ho deixem així, dona un bon test global.