

# Titre provisoire

Louis Bourges, Jean-Baptiste Lagrange-Dupuis et Luc Letonturier

14 mai 2023

Et ici on peut écrire ... et insérer des blocs de code qui s'exécutent, avec le code et le résultat qui s'affichent

```
a <- 2+2  
  
a  
  
## [1] 4
```

ou juste le résultat :

```
## [1] 6
```

ou totalement invisibles :

Et ensuite on peut citer les résultats : à première vue  $4 < 6$  mais je crois que c'est 8 qui est le plus grand.  
Un petit graphique :

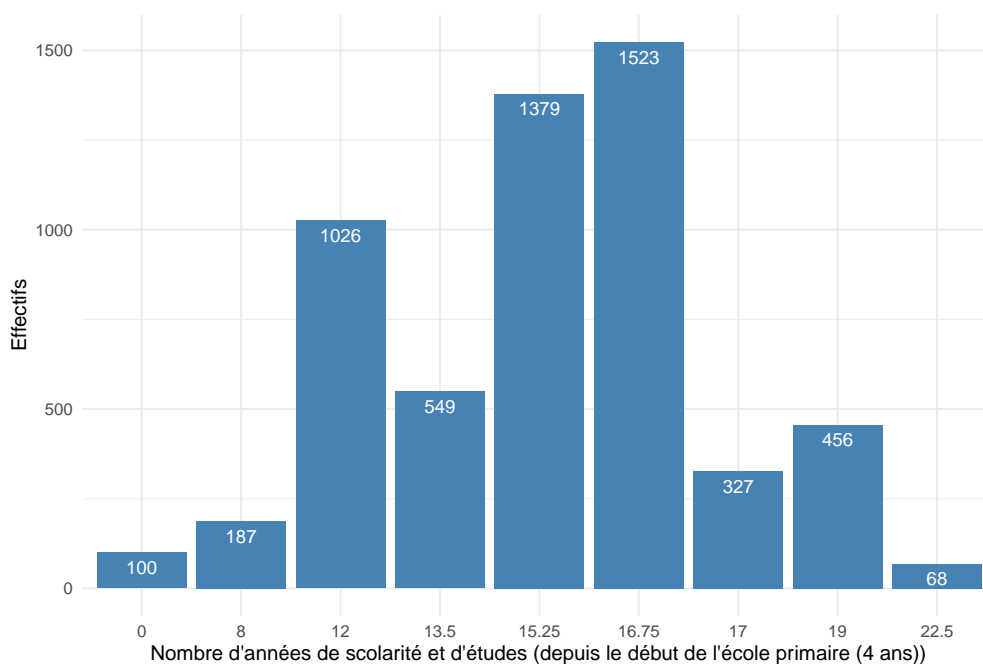


FIGURE 1 – Niveau d'éducation (avec diplôme) des individus de l'échantillon

```
## [1] 5615  
  
## [1] 0
```

```
% latex table generated in R 4.2.1 by xtable 1.8-4 package
% Sun May 14 12:07:52 2023
\begin{table}[ht]
\centering
\caption{Tableau des r sidus}
\label{tb:lm1}
\begin{tabular}{rrrrr}
\toprule
& Estimate & Std. Error & t value & Pr(>|t|) & \\
\midrule
(Intercept) & 5.6786 & 0.1095 & 51.84 & 0.0000 & \\
data\$age & 0.0072 & 0.0010 & 7.11 & 0.0000 & \\
data\$genre & -0.3039 & 0.0225 & -13.53 & 0.0000 & \\
data\$heures & 0.0142 & 0.0008 & 17.03 & 0.0000 & \\
data\$experience & 0.0024 & 0.0011 & 2.15 & 0.0316 & \\
data\$nbenfants & -0.0097 & 0.0096 & -1.01 & 0.3112 & \\
data\$education & 0.1083 & 0.0059 & 18.44 & 0.0000 & \\
\bottomrule
\end{tabular}
\end{table}
```

Ici c'est l'annexe

```
knit_hooks$set(source=listing, output=listing)

a <- 2+2

a

## [1] 4

b <-3+3

b

## [1] 6

c <- 4+4

c

## [1] 8

setwd("~/projet-econometrie-ensps")

library(ggplot2)

update_geom_defaults("text", list(size=30))

library(xtable)

###Importation des bases de donn es

library(readr)

library(ggplot2)
```

```

school <- read_delim("data_LISS/work-and-school.csv", delim = ";", escape_double =
  FALSE, trim_ws = TRUE, show_col_types = FALSE)

## Warning: One or more parsing issues, call 'problems()' on your data
## frame for details, e.g.:
##   dat <- vroom(...)
##   problems(dat)

background<-read_delim("data_LISS/avars_202207_EN_1.0p.csv", delim = ";",
  escape_double = FALSE, trim_ws = TRUE, show_col_types = FALSE)

###Fusion des bases de donn es

agregdata <- merge(background, school, by="nomem_encr" )

###Gestion des NA

#d tecti on

nrow(agregdata)

## [1] 5723

agregdata <- agregdata[!agregdata$oplzon==7,]#suppression de la cat gorie "autre
"

nrow(agregdata)

## [1] 5670

sum(is.na(agregdata$oplzon))

## [1] 0

sum(agregdata$oplzon==8)

## [1] 0

#aucun na      priori

###Pr -traitement des donn es et r criture des variables (a.e = ann es d'
tudes )

#On s'appuie d'abord sur la variable oplzon (niveau d' tudes "with diploma")

agregdata$educ <- agregdata$oplmet*100

agregdata$educ <- ifelse(agregdata$educ==800|agregdata$educ==900, 0,
  #pas d' tudes du tout : "Not yet completed any education
  " or "Not (yet) started any education"
  ifelse(agregdata$educ==100, 8,
    #primary de 4      12 ans => 8 ans d' tudes :
    primary school
    ifelse(agregdata$educ==200, 12,

```

```

#1st option : VMBO (4 ans, de 12      16) =>
8+4=12 a.e. : "vmbo (intermediate
secondary education, US: junior high
school)"
ifelse(agregdata$educ==300, 13.5,
#2nd option : HAVO (5 ans, de 12
17) or VWO (6 ans, de 12      18)
=> 8+5,5 = 13.5 a.e. : "havo/vwo
(higher secondary education/
preparatory university education,
US: senior high school)"
ifelse(agregdata$educ==400, 15.25,
#1st option : MBO (entre 1 et
4 ans, apr s VMBO, HAVO
ou VWO) => 12.75 + 2.5 =
15.25 a.e. : "mbo (
intermediate vocational
education, US: junior
college)"
ifelse(agregdata$educ==500,
16.75,
#2nd option : HBO (4
ans, apr s VMBO,
HAVO ou VWO) =>
12.75+4=16.75 a.e.
: "hbo (higher
vocational
education, US:
college)"
ifelse(agregdata$educ
==600, 17, -9))))))
)

#3rd option : WO (3 ans, apr s 1ere ann e HBO (13.75 a.e.) ou apr s VWO (14 a.e
.), licence) => 14+3=17 a.e. : "wo (university)"

nrow(agregdata)

## [1] 5670

sum(agregdata$educ==-9)

## [1] 55

agregdata <- agregdata[!agregdata$educ==-9,]

nrow(agregdata)

## [1] 5615

#Puis, pour distinguer entre licence, master et doctorat, on s'appuie sur la
variable cw22o005 :

sum(agregdata$cw22o005==27) #nombre de titulaires d'un PhD

## [1] 68

sum(agregdata$cw22o005==26) #nombre de titulaires d'un master (au maximum)

## [1] 456

```

```

sum(agregdata$cw22o005==25) #nombre de titulaires d'une licence (au maximum)

## [1] 117

sum(agregdata$cw22o005==27)+sum(agregdata$cw22o005==26)+sum(agregdata$cw22o005
==25) #nombre de dipl m s du sup rieur (au moins licence), selon cw22o005

## [1] 641

sum(agregdata$educ==18)

## [1] 0

sum((agregdata$cw22o005==27|agregdata$cw22o005==26|agregdata$cw22o005==25)&
agregdata$educ!=17)#nombre d'observations "bizarres"

## [1] 62

#On consid re que les r ponses la question cw22o005 sont plus fiables car
plus pr cises.

agregdata$educ2 <- ifelse(agregdata$cw22o005==25, 17,
#niveau licence => 18 ans d' tudes
ifelse(agregdata$cw22o005==26, 19,
#niveau master, le master dure 1, 2 ou 3 ans => 17
+ 2 = 19 a.e.
ifelse(agregdata$cw22o005==27, 22.5,
agregdata$educ))) #apr s le master, un
doctorat dure 3 4 ans => 19 + 3.5 = 22.5 a.e
.

g <- ggplot(agregdata, aes(x=factor(agregdata$educ2)), id=id, transition=TRUE)+
geom_bar(fill="steelblue")+
labs(x="Nombre d'ann es de scolarit et d' tudes (depuis le d but de l'
cole primaire (4 ans))", y = "Effectifs")+
geom_text(stat='count', aes(label=after_stat(count)), vjust=1.6, color="white",
size=7)+
theme_minimal(base_size = 22)

ggsave(g, filename = "figure/educ.pdf", width=15, height=10)

## Error in grDevices::pdf(file = filename, ..., version = version): cannot open file 'figure/educ.pdf'

###Cr ation d'une table avec les donn es pertinentes

data<-data.frame(agregdata$nomem_encr, agregdata$leeftijd, agregdata$geslacht,
agregdata$brutoink, agregdata$cw22o127, agregdata$cw22o134, agregdata$educ,
agregdata$aantalki)

names(data)<-c("identite", "age", "genre", "revenu", "heures", "experience", "
education", "nbenfants")

nrow(data)

## [1] 5615

```

```

sum(is.na(data$nbenfants))

## [1] 0

data <- data[!data$revenu<=0,]

data <- data[!is.na(data$heures),]

data <- data[!is.na(data$experience),]

data <- data[!is.na(data$education),]

data <- data[!data$experience==999,]

data <- data[!data$education==9,]

data <- data[!data$heures==999,]

data <- data[!data$genre==3,]

###R ecriture des variables :

data$genre <- ifelse(data$genre==1, 0, 1)

data$log_revenu <- log(data$revenu)

data$experience <- 2022 - data$experience

lm1 <- lm(data$log_revenu ~ data$age + data$genre + data$heures + data$experience
+ data$nbenfants + data$education)

print(xtable(summary(lm1), caption="Tableau des r sidus", label="tb:lm1"),
      booktabs=TRUE, caption.placement="top")

## % latex table generated in R 4.2.1 by xtable 1.8-4 package
## % Sun May 14 12:07:53 2023
## \begin{table}[ht]
## \centering
## \caption{Tableau des r sidus}
## \label{tb:lm1}
## \begin{tabular}{rrrrr}
## \toprule
## & Estimate & Std. Error & t value & Pr(>|t|) & \\
## \midrule
## (Intercept) & 5.6786 & 0.1095 & 51.84 & 0.0000 & \\
## data$age & 0.0072 & 0.0010 & 7.11 & 0.0000 & \\
## data$genre & -0.3039 & 0.0225 & -13.53 & 0.0000 & \\
## data$heures & 0.0142 & 0.0008 & 17.03 & 0.0000 & \\
## data$experience & 0.0024 & 0.0011 & 2.15 & 0.0316 & \\
## data$nbenfants & -0.0097 & 0.0096 & -1.01 & 0.3112 & \\
## data$education & 0.1083 & 0.0059 & 18.44 & 0.0000 & \\
## \bottomrule
## \end{tabular}
## \end{table}

```

