

Supplementary source code

This file provides source code so that, combined with the provided input data, readers should be able to reproduce the results showed in the article.

```
##//CHILEAN DATA BASES//##
```

#Library

```
library(magrittr)
library(vctrs)
library(hsm)
library(readxl)
library(rlang)
library(dplyr)
library(dplyr)
library(doBy)
library(tidyr)
library(remotes)
library(foreign)
library(data.table)
library(ggplot2)
library(psychometric)
library(devtools)
library(psc1)
library(psc1)
library(writexl)
library(lme4)
library(robustbase)
library(croostable)
library(gmodel)
library(car)
library(tidyverse)
library(haven)
library(dplyr)
library(plyr)
library(purrr)
library("prais")
```

##CHOLECYSTECOMY DATA BASE

#Import the hospital data base from 2002 to 2019 (public data base)

```
setwd("D:/Dropbox/Mi PC (DESKTOP-QINV2QK)/Documents/Doctorado/Cole_mortalidad_chile/Egresos/Egresos
2002_2019_csv")
temp = list.files(pattern="*.csv")
myfiles = lapply(temp, read.csv2)
```

#Review the structure of every year

```
str(myfiles[1])
table(myfiles[[1]]$ANO_EGRESO)
```

#Drop the "Extranjero" observation from the variable REGION_RESIDENCIA since 2011 to 2019

```
for(i in 1:19){myfiles[[i]] <-filter(myfiles[[i]],REGION_RESIDENCIA!="Extranjero")}
for(i in 1:19){myfiles[[i]]$REGION_RESIDENCIA<-as.integer(myfiles[[i]]$REGION_RESIDENCIA)}
```

#Append all the data base in one data frame

```
hospital <- do.call("rbind", myfiles)
```

#Keep only the diagnosis related to GB disease, codes K80 to K83 from CIE-10

#Import de codes K80-K83 in a csv file

```
cie_cole<-read.csv("cie_to_select_03062021.csv",sep=";")
str(cie_cole)
cie_cole=cie_cole%>%
rename(codigo=DIAG1)
cie_cole<-rename(cie_cole,codigo=DIAG1)
```

#Filter the hospital data base whit the GB disease codes

```
GBD<-hospital[hospital$DIAG1 %in% cie_cole$DIAG1 , ]
```

#From the GBD data base keep the persons wich go under colecistectomy, correspondig to the variable INTERV_Q=1

```
str(GBD$INTERV_Q)
cole<-filter(GBD, INTERV_Q==1)
cole$INTERV_Q<-as.integer(cole$INTERV_Q)
```

#Review the variables of interest for the study

##SEXO

```
table(cole$SEXO)
cole<-filter(cole, SEXO<3)
```

##AGE

```
table(cole$EDAD_CANT)
```

##REGION

```
table(cole$REGION_RESIDENCIA)
cole<-filter(cole, cole$REGION_RESIDENCIA<99)
```

#Recode the REGION_RESIDENCIA variable, to obtain 13 regions instead of 16 regions

```
cole$reg2<-cole$REGION_RESIDENCIA
cole$reg2<-replace(cole$reg2,cole$REGION_RESIDENCIA==15,1)
cole$reg2<-replace(cole$reg2,cole$REGION_RESIDENCIA==16,8)
cole$reg2<-replace(cole$reg2,cole$REGION_RESIDENCIA==14,10)
table(cole$reg2)
```

#Collapse the cole data base using the reg2 variable

```
str(cole)
cole<- summaryBy(INTERV_Q ~SEXO + ANO_EGRESO+ AGE_CANT+reg2,FUN=sum, data=cole)
```

#Rename de INTERV_Q to persons without GB (persons_wo_GB)

```
cole=cole%>%
  rename(people_wo_GB=INTERV_Q.sum)
```

#Generate the missing ages and replace whit 0 the number of persons_wo_GB

```
cole<- cole %>%
  complete(AGE_CANT,nesting(ANO_EGRESO,reg2,SEXO),fill=list(people_wo_GB=0))
```

#Rename some variables

```
str(cole)
cole<-cole%>%
  rename(year=ANO_EGRESO)
```

```
cole<-cole%>%
  rename(age=AGE_CANT)
```

```
cole<-cole%>%
  rename(gender=SEXO)
```

#MORTALITY DATA BASE

#Import the mortality data base (publica data base)

```
mortality_1990_2018<-read.csv("DEF_1990-2018.csv",sep=";")
str(mortality_1990_2018)
```

#Keep the years of interest from 2007 to 2018

```
mortality_2007_2018<-filter(mortality_1990_2018, ANO_DEF>2006)
table(mortality_2007_2018$ANO_DEF)
```

#Explore the variable of interest

```
#SEXO
table(mortality_2007_2018$SEXO)
mortality_2007_2018<-filter(mortality_2007_2018, SEXO<3)
```

#AGE

```
table(mortality_2007_2018$AGE_CANT)
mortality_2007_2018<-filter(mortality_2007_2018, AGE_CANT<200)
```

#REGION

```
table(mortality_2007_2018$REG_RES)
mortality_2007_2018<-filter(mortality_2007_2018, REG_RES<99)
```

#Recode the REG_RES variable, leave 13 regions instead of 16

```
#cole$REGION_RESIDENCIA<-as.integer(cole$REGION_RESIDENCIA)
mortality_2007_2018$reg2<-mortality_2007_2018$REG_RES
mortality_2007_2018$reg2<-replace(mortality_2007_2018$reg2,mortality_2007_2018$REG_RES==15,1)
mortality_2007_2018$reg2<-replace(mortality_2007_2018$reg2,mortality_2007_2018$REG_RES==16,8)
mortality_2007_2018$reg2<-replace(mortality_2007_2018$reg2,mortality_2007_2018$REG_RES==14,10)
table(mortality_2007_2018$reg2)
```

#Gen a counting variable

```
mortality_2007_2018$cases<-1
```

#Collapse the mortality_2007_2018 data base according to the new

```
# variable region, reg2
mortality_2007_2018_reg2<- summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+DIAG1,FUN=sum,
data=mortality_2007_2018)
str(mortality_2007_2018_reg2)
```

#Rename the variable cases

```
mortality_2007_2018_reg2=mortality_2007_2018_reg2%>%
  rename(cases=cases.sum)
```

#Generate the missing ages and replace whit 0 the number

of cases

```
mortality_2007_2018_reg2<- mortality_2007_2018_reg2 %>%
  complete(AGE_CANT,nesting(ANO_DEF,reg2,SEXO,DIAG1),fill=list(cases=0))
table(mortality_2007_2018_reg2$AGE_CANT)
```

```

#Import the mortality data base from 2016 to 2021 (publica data base)
mortality_2016_2021<-read.csv("DEFUNCIONES_FUENTE_DEIS_2016_2021_25022021.csv",sep=";")

#keep only 2019
mortality_2019<-filter(mortality_2016_2021, ANO_DEF==2019)

#Explore the variable of interest
#SEXO
mortality_2019$SEXO<-1
mortality_2019$SEXO<-replace(mortality_2019$SEXO,mortality_2019$GLOSA_SEXO=="Mujer",2)
mortality_2019$SEXO<-replace(mortality_2019$SEXO,mortality_2019$GLOSA_SEXO=="Indeterminado",9)
mortality_2019<-filter(mortality_2019,SEXO<3)

#AGE
table(mortality_2019$edad_CANT)

#REGION
#Gen a numeric variable to identify the region and keep 13 regions
table(mortality_2019$REG)
mortality_2019$reg2<-1
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Antofagasta",2)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Atacama",3)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Coquimbo",4)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Valparaíso",5)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="Del Libertador B. O'Higgins",6)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="Del Maule",7)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="Del Biobío",8)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Ñuble",8)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De La Araucanía",9)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Los Lagos",10)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Los Ríos",10)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Aisén del Gral. C. Ibañez del Campo",11)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="De Magallanes y de La Antártica Chilena",12)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="Metropolitana de Santiago",13)
mortality_2019$reg2<-replace(mortality_2019$reg2,mortality_2019$REG=="Ignorada",99)
mortality_2019<-filter(mortality_2019, reg2<99)

#Gen a counting variable
mortality_2019$cases<-1

#Collpase the mortality_2019 data base according to the variable reg2
mortality_2019_reg2<- summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+DIAG1,FUN=sum, data=mortality_2019)
str(mortality_2019_reg2)
#Rename the variable cases
mortality_2019_reg2=mortality_2019_reg2%>%
  rename(cases=cases.sum)

#Generate the missing ages and replace whit 0 the number of cases
mortality_2019_reg2<- mortality_2019_reg2 %>%
complete(AGE_CANT, nesting(ANO_DEF, reg2, SEXO, DIAG1), fill=list(cases=0))
table(mortality_2019_reg2$AGE_CANT)
table(mortality_2019_reg2$cases)

#Append the two mortality collapses data bases
head(mortality_2007_2018_reg2)
head(mortality_2019_reg2)
mortality_2019_reg2$SEXO<-as.integer(mortality_2019_reg2$SEXO)
mortality_2007_2019<-rbind(mortality_2007_2018_reg2,mortality_2019_reg2)

##GENERAL MORTALITY
#Create the GENERAL MORTALITY base from the mortality_2007_2019 base
general_mortality<-summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2,FUN=sum, data=mortality_2007_2019)
str(general_mortality)
general_mortality=general_mortality%>%
  rename(cases=cases.sum)

#Add the chilean population according to year, region, age and gender (publica data base)
population_2002_2035<-read.csv("ine_estimaciones-y-proyecciones-2002-2035_base-2017_region_base.csv")
colnames(population_2002_2035)

#Reshape the population base to long format
population_2002_2035_long<-reshape(population_2002_2035,
  direction = "long",
  varying = list(names(population_2002_2035)[4:37]),
  v.names = "population",
  idvar = c("Region", "Sexo", "Age"),
  timevar = "Year",
  times = 2002:2035)
colnames(population_2002_2035_long)

#Explore the variables of interest

#REGION
#Recode the variable REGION
table(population_2002_2035_long$Region)
population_2002_2035_long$reg2<-population_2002_2035_long$Region
table(population_2002_2035_long$reg2)

```

```

population_2002_2035_long$reg2<-
replace(population_2002_2035_long$reg2,population_2002_2035_long$Region==15,1)
population_2002_2035_long$reg2<-
replace(population_2002_2035_long$reg2,population_2002_2035_long$Region==16,8)
population_2002_2035_long$reg2<-
replace(population_2002_2035_long$reg2,population_2002_2035_long$Region==14,10)
table(population_2002_2035_long$reg2)

```

#SEX

```
table(population_2002_2035_long$Sexo)
```

#AGE

```
table(population_2002_2035_long$Age)
```

#YEAR

```
table(population_2002_2035_long$Year)
```

#Keep the year of interest

```

population_2007_2019<-filter(population_2002_2035_long, Year>2006)
population_2007_2019<-filter(population_2007_2019, Year<2020)
table(population_2007_2019$Year)

```

#Collapse the population_2007_2019 data base according to the new # region variable reg2

```
population_2007_2019<- summaryBy(population~ Sexo + Year+ Age +reg2,FUN=sum, data=population_2007_2019)
```

#Rename some variables

```

population_2007_2019<-population_2007_2019%>%
  rename(year=Year)
population_2007_2019<-population_2007_2019%>%
  rename(age=Age)
population_2007_2019<-population_2007_2019%>%
  rename(gender=Sexo)
population_2007_2019<-population_2007_2019%>%
  rename(population=population.sum)

```

#Merge the general mortality with the population_2007_2019

#Rename the variables in the general mortality data base

```

str(general_mortality)
general_mortality<-general_mortality%>%
  rename(age=AGE_CANT)
general_mortality<-general_mortality%>%
  rename(gender=SEXO)
general_mortality<-general_mortality%>%
  rename(year=ANO_DEF)
table(general_mortality$gender)

```

#Merging the two data base

```

general_mortality<-merge(general_mortality,population_2007_2019, all.x=F, all.y = F)
table(general_mortality$age)

```

#Generate the ratio between cases and population (mortality)

```

general_mortality$mortality<-general_mortality$cases/general_mortality$population
str(general_mortality$mortality)

```

#Generate 1-mortality (one_mortality)

```

general_mortality$one_mortality<-1-general_mortality$mortality
str(general_mortality)

```

#CANCER DATA BASE

```

#Generate the mortality data base for different cancer types, the are some cancer
#codes that are the union of individual cancer types codes, for that reason 4 list of
#cancer codes were generated
#All the cancer code from ICD-10(CIE-10) where saved in a .csv file

```

```
cie_cancer<-read.csv("cie_cancer.csv")
```

#Filter the cancer diagnosis from de mortality_2007_2019 data base

```

cancer<-mortality_2007_2019[mortality_2007_2019$DIAG1 %in% cie_cancer$DIAG1,]
str(cancer$DIAG1)
str(cancer)

```

#keep the cancer diagnosis of interest and assign a numeric code

#1.First list

```

cancer_1<-cancer
cancer_1$code<-4
cancer_1<-filter(cancer_1,DIAG1!="C23X")
table(cancer_1$DIAG1, cancer_1$cases)

```

#Collapse the data base according to the code assigned

```

str(cancer_1$code)
cancer_1<-summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+code,FUN=sum, data=cancer_1)
table(cancer_1$code, cancer_1$cases.sum)

```

#2. Second list

```
cancer_2<-cancer  
cancer_2$code<-0
```

```
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C180",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C181",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C182",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C183",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C184",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C185",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C186",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C187",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C188",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C189",18)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C19x",20)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C20x",20)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C210",21)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C211",21)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C212",21)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C218",21)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C220",22)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C221",26)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C240",24)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C241",24)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C248",24)  
cancer_2$code<-replace(cancer_2$code, cancer$DIAG1=="C249",24)
```

#Collapse the data base according to the code assigned

```
str(cancer_2$code)  
table(cancer_2$code, cancer_2$AGE_CANT, cancer_2$cases.sum)  
table(cancer_2$code, cancer_2$DIAG1)  
cancer_2<-summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+code,FUN=sum, data=cancer_2)  
table(cancer_2$code)
```

#Eliminate the cases which the codes are 0

```
cancer_2<-filter(cancer_2, code!=0)  
table(cancer_2$code)
```

#3.Third list

```
cancer_3<-filter(cancer, DIAG1=="C241")  
cancer_3$code<-241  
table(cancer_3$DIAG1, cancer_3$code)  
table(cancer_3$DIAG1, cancer_3$cases, cancer_3$AGE_CANT)
```

#Drop the variable DIAG1

```
head(cancer_3)  
cancer_3<-summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+code,FUN=sum, data=cancer_3)
```

#4.Four list

```
cancer_4<-cancer  
cancer_4$code<-0
```

```
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C000",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C001",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C002",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C003",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C004",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C005",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C006",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C008",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C009",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C01x",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C020",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C021",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C022",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C023",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C024",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C028",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C029",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C030",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C031",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C039",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C040",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C041",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C048",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C049",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C050",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C051",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C052",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C058",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C059",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C060",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C061",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C062",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C068",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C069",3)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C07x",7)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C080",7)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C081",7)  
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C088",7)
```



```

cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C913",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C914",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C915",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C916",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C917",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C918",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C919",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C920",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C921",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C922",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C923",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C924",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C925",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C926",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C927",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C928",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C929",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C930",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C931",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C933",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C937",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C939",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C940",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C942",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C943",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C944",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C946",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C947",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C950",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C951",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C957",91)
cancer_4$code<-replace(cancer_4$code, cancer$DIAG1=="C959",91)

```

#Collapse the data base according to the code assigned

```

table(cancer_4$code)
cancer_4<-summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+code,FUN=sum, data=cancer_4)

```

#Eliminate the cases which the codes are 0

```

cancer_4<-filter(cancer_4, code!=0)
table(cancer_4$code)

```

#5. Sixth list

```

cancer_5$code<-replace(cancer_5$code, cancer$DIAG1=="C211",27)
cancer_5$code<-replace(cancer_5$code, cancer$DIAG1=="C240",27)
cancer_5$code<-replace(cancer_5$code, cancer$DIAG1=="C241",27)
cancer_5$code<-replace(cancer_5$code, cancer$DIAG1=="C248",27)
cancer_5$code<-replace(cancer_5$code, cancer$DIAG1=="C249",27)

```

#Collapse the data base according to the code assigned

```

table(cancer_5$code)
cancer_5<-summaryBy(cases~ SEXO + ANO_DEF+ AGE_CANT+reg2+code,FUN=sum, data=cancer_5)

```

#Eliminate the cases which the codes are 0

```

cancer_5<-filter(cancer_5, code!=0)
table(cancer_5$code)

```

#Append the 5 cancer data base to create a unique data whit the cancer of interest

```

cancer_list<-rbind(cancer_1, cancer_2, cancer_3, cancer_4, cancer_5)
table(cancer_list$code)

```

#Rename the variables of the cancer_list data base

```

str(cancer_list)
cancer_list=cancer_list%>%
  rename(cases=cases.sum)

cancer_list=cancer_list%>%
  rename(gender=SEXO)

cancer_list=cancer_list%>%
  rename(year=ANO_DEF)

cancer_list=cancer_list%>%
  rename(age=AGE_CANT)

cancer_list=cancer_list%>%
  rename(region=reg2)

```

#Change the base to wide format

```

str(cancer_list)
table(cancer_list$code)
cancer_list$code<-as.character(cancer_list$code)
cancer_list_wide<-reshape(as.data.frame(cancer_list),
  timevar = "code",
  idvar = c("gender", "year","age","region"),
  v.names="cases",
  direction = "wide")

```

```
#Generate the 3 decades of interest from cole data base:  
#2007-2016, 2008-2017 and 2009-2018
```

```
mylist<-list()  
for(j in 0:2){  
  mylist[[1+j]] <-filter(cole, (cole$age>29 & cole$age<71 & year==2007+j)|  
    (cole$age>30 & cole$age<72 & year==2008+j)|  
    (cole$age>31 & cole$age<73 & year==2009+j)|  
    (cole$age>32 & cole$age<74 & year==2010+j)|  
    (cole$age>33 & cole$age<75 & year==2011+j)|  
    (cole$age>34 & cole$age<76 & year==2012+j)|  
    (cole$age>35 & cole$age<77 & year==2013+j)|  
    (cole$age>36 & cole$age<78 & year==2014+j)|  
    (cole$age>37 & cole$age<79 & year==2015+j)|  
    (cole$age>38 & cole$age<80 & year==2016+j))  
}  
}
```

```
#Merge the general mortality to each element of the list
```

```
for(j in 0:2) {  
  mylist[[j+1]]<-merge(mylist[[j+1]],general_mortality, all.x=T, all.y = F)  
}
```

```
#Generate and indicator to correlative ages among years
```

```
for(j in 0:2) {  
  mylist[[j+1]]$id<-0  
  for (i in 0:40){  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==30+i & mylist[[j+1]]$year==2007+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==31+i & mylist[[j+1]]$year==2008+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==32+i & mylist[[j+1]]$year==2009+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==33+i & mylist[[j+1]]$year==2010+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==34+i & mylist[[j+1]]$year==2011+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==35+i & mylist[[j+1]]$year==2012+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==36+i & mylist[[j+1]]$year==2013+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==37+i & mylist[[j+1]]$year==2014+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==38+i & mylist[[j+1]]$year==2015+j,i+1)  
    mylist[[j+1]]$id<-replace(mylist[[j+1]]$id,mylist[[j+1]]$age==39+i & mylist[[j+1]]$year==2016+j,i+1)  
  }  
}
```

```
#Order the data base using the id generated to correct the cholecystectomy for the general mortality
```

```
for(j in 0:2) {  
  mylist[[j+1]]<-mylist[[j+1]][order(mylist[[j+1]]$gender,mylist[[j+1]]$reg2,mylist[[j+1]]$id,-  
mylist[[j+1]]$year),]  
}
```

```
#Calculate the cumulative product between the number of colecistectomys and  
#and the 1-mortality to obtain an adjusted value.
```

```
for(j in 0:2) {  
  mylist[[j+1]]<- mylist[[j+1]]%>%  
  group_by(gender, reg2, id) %>%  
  mutate(adjusted=people_wo_GB*cumprod(one_mortality))  
}
```

```
#Order the data base
```

```
head(mylist[[1]]$adjusted)  
for(j in 0:2) {  
  mylist[[j+1]]<-  
  mylist[[j+1]][order(mylist[[j+1]]$gender,mylist[[j+1]]$reg2,mylist[[j+1]]$id,mylist[[j+1]]$year),]  
}
```

```
head(mylist[[1]])
```

```
#Calculate the cumulative sum of the adjusted value for each decade
```

```
for(j in 0:2) {  
  mylist[[j+1]]<-mylist[[j+1]]%>%  
  group_by(id,gender,reg2) %>% mutate(csum = cumsum(adjusted))  
}
```

```
#Calculate the cumulative sum of people without gallbladder
```

```
for(j in 0:2) {
  mylist[[j+1]]<-mylist[[j+1]]%>%
  group_by(id,gender,reg2) %>% mutate(csum_2 = cumsum(people_wo_GB))
}
```

#Data base for age, whit cole corrected for mortality

#Generate the number of year without GB

```
for(j in 0:2) {
  mylist[[j+1]]$years_wo_GB <- 2017+j-mylist[[j+1]]$year
}
table(mylist[[1]]$years_wo_GB)
table(mylist[[2]]$years_wo_GB)
table(mylist[[3]]$years_wo_GB)
str(mylist[[1]])
```

#Rename some variables

```
for(j in 0:2) {
  mylist[[j+1]]=mylist[[j+1]]%>%
  rename(total_people_wo_GB_adjusted=csum) #Rename the cumulative sum
  mylist[[j+1]]=mylist[[j+1]]%>%
  rename(people_wo_GB_adjusted=csum_2) #Rename the cumulative sum 2
}
```

#Bind to a data frame

```
decadas_year <- do.call("rbind", mylist)
```

#Remove some variable

```
str(decadas_year)
decadas_year<-decadas_year[,c(-6,-7,-8,-9,-10)]
```

#Recode age and year

```
decadas_year$age2<-decadas_year$age+decadas_year$years_wo_GB
decadas_year$year2<-decadas_year$year+decadas_year$years_wo_GB
decadas_year<-decadas_year[,c(-1,-2)]
```

#Rename some variables

```
decadas_year=decadas_year%>% rename(age=age2)
decadas_year=decadas_year%>%rename(year=year2)
decadas_year=decadas_year%>%rename(people_wo_gb_ad=adjusted)
```

#Calculate the percentage of people without GB

```
decadas_year=decadas_year%>%rename(region=reg2)
decadas_year<-merge(decadas_year,population_2007_2019, all.x=T, all.y = F)
decadas_year$percentage_people_wo_gb<-(decadas_year$people_wo_gb/decadas_year$population)*100
```

#Reshape the population base to wide format

```
decadas_year<-decadas_year[,(-8)]
```

```
decadas_year_wide<-reshape(as.data.frame(decadas_year),
  timevar = "years_wo_GB",
  idvar = c("age", "year", "region", "gender", "population"),
  v.names=c("people_wo_gb", "percentage_people_wo_gb"),
  direction = "wide")
```

```
decadas_year_wide<-merge(decadas_year_wide,cancer_list_wide,all.x=T, all.y = F)
```

#Merge the cancer data base to the wide people wo GB over years

```
str(cancer_list_wide)
table(cancer_list_wide$gender)
table(decadas_year$gender)
str(decadas_year$gender)
decadas_year_wide<-merge(decadas_year_wide,cancer_list_wide,all.x=T, all.y = F)
```

#Keep only 10 year without GB

```
year_10_long<-filter(decadas_year,years_wo_GB==10)
table(year_10_long$years_wo_GB)
```

#Drop regions with <10 death for GB

```
table(year_10_long$region)
year_10_long<-filter(year_10_long,region!=11)
year_10_long<-filter(year_10_long,region!=12)
```

#Drop 2019

```
year_10_long<-filter(year_10_long,year!=2019)
```

#Merge with cancer mortality data base

```

table(cancer_list$gender)
table(year_10_long$gender)

year_10_long<-merge(year_10_long,cancer_list, all.x=T, all.y=F,by=c("gender","year","age","region"))

#Classification of years
colnames(year_10_long)
year_10_long$years_b<-4

#Generate age category
table(year_10_long$age)
str(year_10_long)
year_10_long$age4<-year_10_long$age
year_10_long$age4<-cut(year_10_long$age4, c(39,45,50,55,60,65,70,75,80),
  labels=c("40-45", "46-50", "51-55", "56-60", "61-65", "66-70", "71-75", "76-80"))

#Generate a numeric id for age category (age5)
year_10_long$age5<-1
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>45 & year_10_long$age<51),2)
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>50 & year_10_long$age<56),3)
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>55 & year_10_long$age<61),4)
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>60 & year_10_long$age<66),5)
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>65 & year_10_long$age<71),6)
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>70 & year_10_long$age<76),7)
year_10_long$age5<-replace(year_10_long$age5, (year_10_long$age>75 & year_10_long$age<81),8)

table(year_10_long$age5,year_10_long$age4)

#Calculate the quartiles
summary(year_10_long$percentage_people_wo_gb)
year_10_long$q<-1
year_10_long$q<-replace(year_10_long$q,year_10_long$percentage_people_wo_gb>0.1801 &
year_10_long$percentage_people_wo_gb<0.3389,2)
year_10_long$q<-replace(year_10_long$q,year_10_long$percentage_people_wo_gb>0.3388 &
year_10_long$percentage_people_wo_gb<0.5457,3)
year_10_long$q<-replace(year_10_long$q,year_10_long$percentage_people_wo_gb>0.5456 ,4)
table(year_10_long$q)

#Summary the cases acording cancer code

year_10_long_c<-summaryBy(cases+population~ years_b + age4+age5 +q+gender+code ,FUN=sum,
data=year_10_long)
head(year_10_long_c)
str(year_10_long_c)

year_10_long_c$code<-as.factor(year_10_long_c$code)

year_10_long_c=year_10_long_c%>%
  rename(cases=cases.sum)

year_10_long_c=year_10_long_c%>%
  rename(population=population.sum)

str(year_10_long_c)
table(year_10_long_c$code)
###DATA BASE PEOPLE WITHOUT GB AFTER 10 YEARS###

##Drop regions with <10 death for GB
decades_reg<-decadas_year_wide
table(decades_reg$region)
decades_reg<-filter(decades_reg,region!=11)
decades_reg<-filter(decades_reg,region!=12)
table(decades_reg$region)

#Recode age in 8 categories.

str(decades_reg)
decades_reg$age4<-decades_reg$age
decades_reg$age4<-cut(decades_reg$age4, c(39,45,50,55,60,65,70,75,80),
  labels=c("40-45", "46-50", "51-55", "56-60", "61-65", "66-70", "71-75", "76-80"))

table(decades_reg$age,decades_reg$age4)

#Generate a numeric id for age category (age5)
decades_reg$age5<-1
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>45 & decades_reg$age<51),2)
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>50 & decades_reg$age<56),3)
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>55 & decades_reg$age<61),4)
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>60 & decades_reg$age<66),5)
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>65 & decades_reg$age<71),6)
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>70 & decades_reg$age<76),7)
decades_reg$age5<-replace(decades_reg$age5, (decades_reg$age>75 & decades_reg$age<81),8)

table(decades_reg$age5)
table(decades_reg$age4,decades_reg$age5)

```

#Recode year

```
table(decades_reg$year)
decades_reg<-filter(decades_reg,year!=2019)
decades_reg$years<-3
```

#Second year recode

```
decades_reg$years_b<-4
table(decades_reg$year,decades_reg$years_b)
```

#Recode ancestry

```
decades_reg$ancestry<-3
decades_reg$ancestry<-replace(decades_reg$ancestry,decades_reg$region>0 & decades_reg$region<4 ,2)
decades_reg$ancestry<-replace(decades_reg$ancestry,decades_reg$region>7 & decades_reg$region<11,1)
table(decades_reg$region,decades_reg$ancestry)
```

```
decades_reg$ancestry_label<-"other"
decades_reg$ancestry_label<-replace(decades_reg$ancestry_label,decades_reg$ancestry==2, "aymara")
decades_reg$ancestry_label<-replace(decades_reg$ancestry_label,decades_reg$ancestry==1, "mapuche")
table(decades_reg$ancestry_label)
table(decades_reg$ancestry_label,decades_reg$ancestry)
```

##Generate the quartiles

```
summary(decades_reg$percentage_people_wo_gb.10)
decades_reg$q<-1
decades_reg$q<-replace(decades_reg$q,decades_reg$percentage_people_wo_gb.10>0.1801 &
decades_reg$percentage_people_wo_gb.10<0.3389,2)
decades_reg$q<-replace(decades_reg$q,decades_reg$percentage_people_wo_gb.10>0.3388 &
decades_reg$percentage_people_wo_gb.10<0.5457,3)
decades_reg$q<-replace(decades_reg$q,decades_reg$percentage_people_wo_gb.10>0.5456 ,4)
table(decades_reg$q)
```

#Generate a data base short name

```
df_10<-decades_reg
df_10$id<-10
```

```
#Generate 5 YEAR OF PEOPLE WO GB o from cole data base:  
# to 2012-2018, initial codes then repeated from line 850 to 931  
table(cole$age, cole$gender)
```

```
mylist_2<-list()  
for(j in 0:6){  
  mylist_2[[j+1]] <-filter(cole, (cole$age>34 & cole$age<76 & year==2007+j)|  
                             (cole$age>35 & cole$age<77 & year==2008+j)|  
                             (cole$age>36 & cole$age<78 & year==2009+j)|  
                             (cole$age>37 & cole$age<79 & year==2010+j)|  
                             (cole$age>38 & cole$age<80 & year==2011+j))  
}  
}
```

```
#Merge the general mortality to each element of the list
```

```
for(j in 0:6) {  
  mylist_2[[j+1]]<-merge(mylist_2[[j+1]],general_mortality, all.x=T, all.y = F)  
}  
table(mylist_2[[1]]$age)  
table(mylist_2[[2]]$age)  
head(mylist_2[[1]])
```

```
#Generate and indicator to group correlative ages among years
```

```
for(j in 0:6) {  
  mylist_2[[j+1]]$id<-0  
  for (i in 0:40){  
    mylist_2[[j+1]]$id<-replace(mylist_2[[j+1]]$id,mylist_2[[j+1]]$age==35+i &  
mylist_2[[j+1]]$year==2007+j,i+1)  
    mylist_2[[j+1]]$id<-replace(mylist_2[[j+1]]$id,mylist_2[[j+1]]$age==36+i &  
mylist_2[[j+1]]$year==2008+j,i+1)  
    mylist_2[[j+1]]$id<-replace(mylist_2[[j+1]]$id,mylist_2[[j+1]]$age==37+i &  
mylist_2[[j+1]]$year==2009+j,i+1)  
    mylist_2[[j+1]]$id<-replace(mylist_2[[j+1]]$id,mylist_2[[j+1]]$age==38+i &  
mylist_2[[j+1]]$year==2010+j,i+1)  
    mylist_2[[j+1]]$id<-replace(mylist_2[[j+1]]$id,mylist_2[[j+1]]$age==39+i &  
mylist_2[[j+1]]$year==2011+j,i+1)  
  }  
}  
}
```

```
#Keep only 5 years without GB
```

```
year_5_long<-filter(year_5,years_wo_GB==5)  
table(year_5_long$years_wo_GB)
```

```
year_5_long=year_5_long%>% rename(people_wo_gb=people_wo_GB_adj)  
str(year_5_long)
```

```
#Join with cancer mortality data base
```

```
year_5_long<-merge(year_5_long,cancer_list, all.x=T, all.y = F,by=c("gender","year","age","region"))
```

```
#Drop regions with <10 death for GB
```

```
year_5_long<-filter(year_5_long,region!=11)  
year_5_long<-filter(year_5_long,region!=12)
```

```
#Second classification of years
```

```
table(year_5_long$year)  
year_5_long$years_b<-4  
year_5_long$years_b<-replace(year_5_long$years_b,year_5_long$year<2014 ,2)  
year_5_long$years_b<-replace(year_5_long$years_b,year_5_long$year>2013 & year_5_long$year<2017 ,3)
```

```
#Generate age category
```

```
table(year_5_long$age)  
str(year_5_long)  
year_5_long$age4<-year_5_long$age  
year_5_long$age4<-cut(year_5_long$age4, c(39,45,50,55,60,65,70,75,80),  
  labels=c("40-45", "46-50", "51-55", "56-60", "61-65", "66-70", "71-75", "76-80"))
```

```
#Generate a numeric id for age category (age5)
```

```
year_5_long$age5<-1  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>45 & year_5_long$age<51),2)  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>50 & year_5_long$age<56),3)  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>55 & year_5_long$age<61),4)  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>60 & year_5_long$age<66),5)  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>65 & year_5_long$age<71),6)  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>70 & year_5_long$age<76),7)  
year_5_long$age5<-replace(year_5_long$age5,(year_5_long$age>75 & year_5_long$age<81),8)
```

```
table(year_5_reg$age5,year_5_reg$age4)
```

```
#Calculate quartiles
```

```
summary(year_5_long$percentage_people_wo_gb.5)
year_5_long$q<-1
year_5_long$q<-replace(year_5_long$q,year_5_long$percentage_people_wo_gb.5>0.2744 &
year_5_long$percentage_people_wo_gb.5<0.4424,2)
year_5_long$q<-replace(year_5_long$q,year_5_long$percentage_people_wo_gb.5>0.4423 &
year_5_long$percentage_people_wo_gb.5<0.6472,3)
year_5_long$q<-replace(year_5_long$q,year_5_long$percentage_people_wo_gb.5>0.6471 ,4)
table(year_5_long$q)
```

#Summary the cases acording cancer code

```
year_5_long_c<-summaryBy(cases+population~ years_wo_GB + age4+age5 +q+gender+code ,FUN=sum,
data=year_5_long)
head(year_5_long_c)
str(year_5_long_c)

year_5_long_c$code<-as.factor(year_5_long_c$code)
```

#Drop regions with <10 death for GB

```
year_5_reg<-year_5_wide
table(year_5_reg$region)
year_5_reg<-filter(year_5_reg,region!=11)
year_5_reg<-filter(year_5_reg,region!=12)
table(year_5_reg$region)
```

#Recode ancestry

```
str(year_5_reg)
year_5_reg$ancestry<-3
year_5_reg$ancestry<-replace(year_5_reg$ancestry,year_5_reg$region>0 & year_5_reg$region<4 ,2)
year_5_reg$ancestry<-replace(year_5_reg$ancestry,year_5_reg$region>7 & year_5_reg$region<11,1)
table(year_5_reg$region,year_5_reg$ancestry)
year_5_reg$ancestry_label<-"other"
year_5_reg$ancestry_label<-replace(year_5_reg$ancestry_label,year_5_reg$ancestry==2,"aymara")
year_5_reg$ancestry_label<-replace(year_5_reg$ancestry_label,year_5_reg$ancestry==1,"mapuche")
table(year_5_reg$ancestry_label)
table(year_5_reg$ancestry_label,year_5_reg$ancestry)
```

#Generate age category

```
table(year_5_reg$age)
str(year_5_reg)
year_5_reg$age4<-year_5_reg$age
year_5_reg$age4<-cut(year_5_reg$age4, c(39,45,50,55,60,65,70,75,80),
                    labels=c("40-45", "46-50", "51-55", "56-60", "61-65", "66-70", "71-75", "76-80"))
```

```
table(year_5_reg$age,year_5_reg$age4)
```

#Generate a numeric id for age category (age5)

```
year_5_reg$age5<-1
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>45 & year_5_reg$age<51),2)
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>50 & year_5_reg$age<56),3)
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>55 & year_5_reg$age<61),4)
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>60 & year_5_reg$age<66),5)
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>65 & year_5_reg$age<71),6)
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>70 & year_5_reg$age<76),7)
year_5_reg$age5<-replace(year_5_reg$age5,(year_5_reg$age>75 & year_5_reg$age<81),8)
```

```
table(year_5_reg$age5,year_5_reg$age4)
```

#Group years according to de data base year_5_reg

```
year_5_reg$years_b<-4
year_5_reg$years_b<-replace(year_5_reg$years_b,year_5_reg$year<2014 ,2)
year_5_reg$years_b<-replace(year_5_reg$years_b,year_5_reg$year>2013 & year_5_reg$year<2017 ,3)
table(year_5_reg$years_b)
table(year_5_reg$year,year_5_reg$years_b)
```

##Generate quartiles

```
summary(year_5_reg$percentage_people_wo_gb.5)
year_5_reg$q<-1
year_5_reg$q<-replace(year_5_reg$q,year_5_reg$percentage_people_wo_gb.5>0.2744 &
year_5_reg$percentage_people_wo_gb.5<0.4424,2)
year_5_reg$q<-replace(year_5_reg$q,year_5_reg$percentage_people_wo_gb.5>0.4423 &
year_5_reg$percentage_people_wo_gb.5<0.6472,3)
year_5_reg$q<-replace(year_5_reg$q,year_5_reg$percentage_people_wo_gb.5>0.6471 ,4)
table(year_5_reg$q)
```

#Generate GES age

```
str(year_5_reg)
year_5_reg$age6<-year_5_reg$age
year_5_reg$age6<-cut(year_5_reg$age6, c(39,55,65,80),
                    labels=c("40-55", "56-65", "66-80"))
```

Generate a numeric id for age GES

```
year_5_reg$age7<-1
year_5_reg$age7<-replace(year_5_reg$age7,(year_5_reg$age>55 & year_5_reg$age<66),2)
year_5_reg$age7<-replace(year_5_reg$age7,(year_5_reg$age>65 & year_5_reg$age<81),3)
table(year_5_reg$age6,year_5_reg$age7)
```

#Generate a data base short name

```
df_5<-year_5_reg
df_5$id<-5
```



```
##### FOR GENDER
```

```
list_2<-list(df_5,df_10)
```

```
#Collapse by sum the data base
```

```
for(j in 1:2) {  
  list_27[[j]]<- summaryBy(cases.23+population~ years_b + age4+age5 +q+gender ,FUN=sum, data=list_27[[j]])  
}
```

```
#Complete the cases of GBC and population by age, gender and year to have the all observation per quartile (GBC=0 and population=2)
```

```
list_2[[1]]<- list_2[[1]]%>%  
  complete(q,nesting(age5,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
list_2[[2]]<- list_2[[2]]%>%  
  complete(q,nesting(age5,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
#Generate age category to complete the new variables generates in the previous line
```

```
for(j in 1:2) {  
  list_2[[j]]$age6<-1  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==1,"40-45")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==2,"46-50")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==3,"51-55")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==4,"56-60")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==5,"61-65")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==6,"66-70")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==7,"71-75")  
  list_2[[j]]$age6<-replace(list_2[[j]]$age6,list_2[[j]]$age5==8,"76-80")  
}
```

```
#Complete the indicator por year without GB
```

```
list_2[[1]]$id<-5  
list_2[[2]]$id<-10
```

```
#Calculate the mortality adjusted rate
```

```
for(j in 1:2) {  
  list_2[[j]] <- list_2[[j]]%>% group_by(age5, gender,years_b)%>%  
    mutate(pop.sum = sum(population.sum))  
  
  list_2[[j]] <- list_2[[j]]%>% group_by(q, gender,years_b)%>%  
    mutate(pop.total = sum(pop.sum))  
  
  list_2[[j]]$proportion<-list_2[[j]]$pop.sum/list_2[[j]]$pop.total  
  
  list_2[[j]]$rate<-list_2[[j]]$cases.23.sum/list_2[[j]]$population.sum #rate cruda  
  list_2[[j]]$rate_adj<-list_2[[j]]$rate*list_2[[j]]$proportion*list_2[[j]]$pop.total  
  list_2[[j]]<- list_2[[j]]%>% group_by(q,gender)%>%  
    mutate(rate_adj.sum.q = sum(rate_adj))  
  
  list_2[[j]]<- list_2[[j]]%>% group_by(q,gender)%>%  
    mutate(cases.23.total = sum(cases.23.sum))  
}
```

```
#Call the list to a data frame
```

```
df_2<- do.call("rbind", list_2)  
str(df_2)  
df_2<-df_2[c(-2,-4:-8,-10:-14)]
```

```
#Reshape data frame to a wide position
```

```
df_2_w<-reshape(as.data.frame(df_2),  
  timevar = "q",  
  idvar = c("id", "gender"),  
  v.names=c("rate_adj.sum.q", "cases.23.total"),  
  direction = "wide")
```

```
#FOR ANCESTRY( REGION/ Zones of Chile)
```

```
#Mapuche= South  
#Aymara= North  
#Other=Centre
```

```
list_3<-list(df_5,df_10)
```

#Collapse by sum the data base

```
for(j in 1:2) {  
  list_3[[j]]<- summaryBy(cases.23+population~ ancestry+years_b + age4+age5 +q+gender ,FUN=sum,  
data=list_3[[j]])  
}
```

Complete the cases of GBC and population by ancestry, age, gender and year to have the all observation per quartile (GBC=0 and population=2)

```
list_3[[1]]<- list_3[[1]]%>%  
  complete(q,nesting(ancestry,age5,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
list_3[[2]]<- list_3[[2]]%>%  
  complete(q,nesting(ancestry,age5,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

#Generate age category to complete the new variables generates in the previous line

```
for(j in 1:2) {  
  
  list_3[[j]]$age6<-1  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==1,"40-45")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==2,"46-50")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==3,"51-55")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==4,"56-60")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==5,"61-65")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==6,"66-70")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==7,"71-75")  
  list_3[[j]]$age6<-replace(list_3[[j]]$age6,list_3[[j]]$age5==8,"76-80")  
  
}
```

#Complete the indicator por year without GB

```
list_3[[1]]$id<-5  
list_3[[2]]$id<-10
```

#Calculate the mortality adjusted rate

```
for(j in 1:2) {  
  
  list_3[[j]] <- list_3[[j]]%>% group_by(ancestry,age5, gender,years_b)%>%  
    mutate(pop.sum = sum(population.sum))  
  
  list_3[[j]] <- list_3[[j]]%>% group_by(q, ancestry,gender,years_b)%>%  
    mutate(pop.total = sum(pop.sum))  
  
  list_3[[j]]$proportion<-list_3[[j]]$pop.sum/list_3[[j]]$pop.total  
  
  list_3[[j]]$rate<-list_3[[j]]$cases.23.sum/list_3[[j]]$population.sum  
  list_3[[j]]$rate_adj<-list_3[[j]]$rate*list_3[[j]]$proportion*list_3[[j]]$pop.total  
  
  list_3[[j]]<- list_3[[j]]%>% group_by(q,ancestry)%>%  
    mutate(rate_adj.sum.q = sum(rate_adj))  
  
  list_3[[j]]<- list_3[[j]]%>% group_by(q,ancestry)%>%  
    mutate(cases.23.total = sum(cases.23.sum))  
}
```

#Call the list to a data frame

```
df_3<- do.call("rbind", list_3)  
str(df_3)  
df_3<-df_3[c(-3:-9,-11:-15)]
```

#Reshape data frame to a wide position

```
df_3_w<-reshape(as.data.frame(df_3),  
  timevar = "q",  
  idvar = c("id", "ancestry"),  
  v.names=c("rate_adj.sum.q", "cases.23.total"),  
  direction = "wide")
```

#FOR GES AGE CATEGORY

```
#GES : 40-49/50-59/60-69/70-80/
```

```
#Generate age categories
```

```
###For 10 years without GB
```

```
df_10$age9<-df_10$age  
df_10$age9<-cut(df_10$age9, c(39,49,59,69,80),  
                labels=c("40-49","50-59", "60-69", "70-80"))
```

```
table(df_10$age,df_10$age9)
```

```
#Generate and indicator for each category
```

```
df_10$age10<-1  
df_10$age10<-replace(df_10$age10,(df_10$age>49 & df_10$age<60),2)  
df_10$age10<-replace(df_10$age10,(df_10$age>59 & df_10$age<70),3)  
df_10$age10<-replace(df_10$age10,(df_10$age>69 & df_10$age<81),4)
```

```
table(df_10$age9,df_10$age10)
```

```
###For 5 years without GB
```

```
df_5$age9<-df_5$age  
df_5$age9<-cut(df_5$age9, c(39,49,59,69,80),  
                labels=c("40-49","50-59", "60-69", "70-80"))
```

```
table(df_5$age,df_5$age9)
```

```
#Generate and indicator for each category
```

```
df_5$age10<-1  
df_5$age10<-replace(df_5$age10,(df_5$age>49 & df_5$age<60),2)  
df_5$age10<-replace(df_5$age10,(df_5$age>59 & df_5$age<70),3)  
df_5$age10<-replace(df_5$age10,(df_5$age>69 & df_5$age<81),4)
```

```
#Join the data bases
```

```
list_4<-list(df_5,df_10)
```

```
#Collapse by sum the data base
```

```
for(j in 1:2) {  
  list_4[[j]]<- summaryBy(cases.23+population~ list_4[[2]]$years_b + age9+age10 +q+gender ,FUN=sum,  
  data=list_4[[j]])  
}
```

```
#Complete the cases of GBC and population by age, gender and year to have the all observation  
per quartile (GBC=0 and population=2)
```

```
list_4[[1]]<- list_4[[1]]%>%  
  complete(q,nesting(age9,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
list_4[[2]]<- list_4[[2]]%>%  
  complete(q,nesting(age9,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
#Generate age category to complete the new variables generates in the previous line
```

```
for(j in 1:2) {  
  list_4[[j]]$age11<-1  
  list_4[[j]]$age11<-replace(list_4[[j]]$age11,list_4[[j]]$age9=="50-59",2)  
  list_4[[j]]$age11<-replace(list_4[[j]]$age11,list_4[[j]]$age9=="60-69",3)  
  list_4[[j]]$age11<-replace(list_4[[j]]$age11,list_4[[j]]$age9=="70-80",4)  
}
```

```
#Complete the indicator por year without GB
```

```
list_4[[1]]$id<-5  
list_4[[2]]$id<-10
```

```
#Calculate the mortality adjusted rate
```

```
for(j in 1:2) {  
  list_4[[j]] <- list_4[[j]]%>% dplyr::group_by(age11, gender,)%>%  
    mutate(pop.sum = sum(population.sum)  
  
  list_4[[j]] <- list_4[[j]]%>% dplyr::group_by(q, age11)%>%  
    mutate(pop.total = sum(pop.sum))  
  
  list_4[[j]]$proportion<-list_4[[j]]$pop.sum/list_4[[j]]$pop.total  
  
  list_4[[j]]$rate<-list_4[[j]]$cases.23.sum/list_4[[j]]$population.sum  
  list_4[[j]]$rate_adj<-list_4[[j]]$rate*list_4[[j]]$proportion*list_4[[j]]$pop.total  
  list_4[[j]]<- list_4[[j]]%>% dplyr::group_by(q,age11)%>%  
    mutate(rate_adj.sum.q = sum(rate_adj))
```

```
list_4[[j]]<- list_4[[j]]%>% dplyr::group_by(q,age11)%>%
  mutate(cases.23.total = sum(cases.23.sum))
}
```

#Call the list to a data frame

```
df_4<- do.call("rbind", list_4)
str(df_4)
```

```
df_4<-df_4[c(-3:-7,-10:-14)]
```

#Reshape data frame to a wide position

```
df_4_w<-reshape(as.data.frame(df_4),
  timevar = "q",
  idvar = c("id", "age9"),
  v.names=c("rate_adj.sum.q","cases.23.total"),
  direction = "wide")
```

##FOR CANCER : AFTER 10 YEAR WITHOUT GB

```
#Generate a list with each cancer code
list_5<-split(year_10_long_c,year_10_long_c$code, drop = FALSE)
```

#Complete the cases and population by age, gender and year to have the all observation per quartile (GBC=0 and population=2)

```
for(n in 1:40){
list_5[[n]]<- list_5[[n]]%>%
  complete(q,nesting(age5,gender,years_b,code),fill=list(cases=0,population=1))
}
```

#Generate age category to complete the new variables generates in the previous line

```
for(n in 1:40) {
  list_5[[n]]$age6<-1
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==1,"40-45")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==2,"46-50")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==3,"51-55")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==4,"56-60")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==5,"61-65")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==6,"66-70")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==7,"71-75")
  list_5[[n]]$age6<-replace(list_5[[n]]$age6, list_5[[n]]$age5==8,"76-80")
}
```

#Calculate the mortality adjusted rate

```
for(n in 1:40) {
  list_5[[n]] <- list_5[[n]]%>% group_by(age5, gender,years_b)%>%
    mutate(pop.sum = sum(population))
  list_5[[n]] <- list_5[[n]]%>% group_by(q, gender,years_b)%>%
    mutate(pop.total = sum(pop.sum))
  list_5[[n]]$proportion<-list_5[[n]]$pop.sum/list_5[[n]]$pop.total #Razon entre populationes
  list_5[[n]]$rate<-list_5[[n]]$cases/list_5[[n]]$population
  list_5[[n]]$rate_adj<-list_5[[n]]$rate*list_5[[n]]$proportion*list_5[[n]]$pop.total
  list_5[[n]]<- list_5[[n]]%>% group_by(q)%>%
    mutate(rate_adj.sum.q = sum(rate_adj))
  list_5[[n]]<- list_5[[n]]%>% group_by(q)%>%
    mutate(cases.total = sum(cases))
}
```

#Drop some variables

```
for(n in 1:40) {
  list_5[[n]] <-list_5[[n]][ c(-2:-4,-6,-7,-9:-14)]
}
```

#Reshape to a wide format

```
list_6<-list()
```

```

for(n in 1:40) {
  list_6[[n]]<-reshape(as.data.frame(list_5[[n]]),
    timevar = "q",
    idvar = c("code"),
    v.names=c("rate_adj.sum.q", "cases.total"),
    direction = "wide")
}

```

#Call the list to a data frame

```
df_5<- do.call("rbind", list_6)
```

```
#####
###CALCULATE THE STANDARIZED MORTALITY RATE###
```

FOR THE ALL COUNTRY

#GENERATING ADJUSTED MORTALITY RATES BY YEAR, GENDER AND AGE GROUPS

```
list_1<-list(df_5,df_10)
```

#Collapse by sum the data base

```
for(j in 1:2) {
  list_1[[j]]<- summaryBy(cases.23+population~ years_b + age4+age5 +q+gender ,FUN=sum, data=list_1[[j]])
}

```

#Complete the cases of GBC and population by age, gender and year to have the all observation per quartile (GBC=0 and population=2)

```
list_1[[1]]<- list_1[[1]]%>%
  complete(q,nesting(age5,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
list_1[[2]]<- list_1[[2]]%>%
  complete(q,nesting(age5,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

#Generate age category to complete the new variables generates in the previous line

```
for(j in 1:2) {
  list_1[[j]]$age6<-1
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==1,"40-45")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==2,"46-50")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==3,"51-55")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==4,"56-60")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==5,"61-65")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==6,"66-70")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==7,"71-75")
  list_1[[j]]$age6<-replace(list_1[[j]]$age6, list_1[[j]]$age5==8,"76-80")
}

```

```

head(list_1[[1]])
head(list_1[[2]])
head(list_1[[3]])

```

#Complete the indicator por year without GB

```
list_1[[1]]$id<-5
list_1[[2]]$id<-10
```

#Calculate the mortality adjusted rate

```

for(j in 1:2) {
  list_1[[j]] <- list_1[[j]]%>% group_by(age5, gender,years_b)%>%
    mutate(pop.sum = sum(population.sum))#Accumulated population by age category

  list_1[[j]] <- list_1[[j]]%>% group_by(q, gender,years_b)%>%
    mutate(pop.total = sum(pop.sum))#Total population grouped by quartiles

  list_1[[j]]$proportion<-list_1[[j]]$pop.sum/list_1[[j]]$pop.total #Ratio between populations

  list_1[[j]]$rate<-list_1[[j]]$cases.23.sum/list_1[[j]]$population.sum #Crude rate
  list_1[[j]]$rate_adj<-list_1[[j]]$rate*list_1[[j]]$proportion*list_1[[j]]$pop.total #Adjusted rate

  list_1[[j]]<- list_1[[j]]%>% group_by(q)%>%
    mutate(rate_adj.sum.q = sum(rate_adj))# Total adjusted rate pooled for each quartile
}

```

```
list_1[[j]]<- list_1[[j]]%>% group_by(q)%>%
mutate(cases.23.total = sum(cases.23.sum))# Total number of GBC death pooled for each quartile
```

```
#Call the list to a data frame
```

```
age_gender_years_b<- do.call("rbind", list_1)
str(age_gender_years_b)
```

```
#Reshape data frame to a wide position
```

```
df_1<-reshape(as.data.frame(age_gender_years_b),
              timevar = "q",
              idvar = c("years_b", "age5", "age6", "gender", "id"),
              v.names=c("rate_adj.sum.q", "cases.23.sum", "cases.23.total",
"population.sum", "pop.sum", "pop.total", "proportion", "rate", "rate_adj"),
              direction = "wide")
```

```
str(df_1)
```

```
#Calculate the Standardized Mortality Rates 1 (SMR1)
```

```
df_1$q2_q1<-df_1$rate_adj.sum.q.2/df_1$rate_adj.sum.q.1
table(df_1$q2_q1)
```

```
#Calculate 95% CI for SMR1
```

```
#log of SMR1
```

```
df_1$ln_rr<-log(df_1$q2_q1)
table(df_1$ln_rr)
```

```
#Standard Error (SE) of ln RR
```

```
df_1$SEln_rr<-sqrt((1/df_1$rate_adj.sum.q.2)+(1/df_1$rate_adj.sum.q.1))
table(df_1$SEln_rr)
```

```
# 95% CI limits
```

```
df_1$lnRRlow=df_1$ln_rr-1.96*df_1$SEln_rr
table(df_1$lnRRlow)
```

```
df_1$lnRRupp=df_1$ln_rr+1.96*df_1$SEln_rr
table(df_1$lnRRupp)
```

```
#log of the limits for 95% CI
```

```
df_1$RRlow=exp(df_1$lnRRlow)
table(df_1$RRlow)
```

```
df_1$RRupp=exp(df_1$lnRRupp)
table(df_1$RRupp)
```

```
#Calculate the standardized Mortality Rates 2 (SMR2)
```

```
df_1$deciles_q1<-df_1$rate_adj.sum.q.3/df_1$rate_adj.sum.q.1
table(df_1$deciles_q1)
```

```
#Calculate 95% CI for SMR1
```

```
#log of SMR2
```

```
df_1$ln_rr.3<-log(df_1$deciles_q1)
table(df_1$ln_rr.3)
```

```
#Standard Error (SE) of ln RR
```

```
df_1$SEln_rr.3<-sqrt((1/df_1$rate_adj.sum.q.3)+(1/df_1$rate_adj.sum.q.1))
table(df_1$SEln_rr.3)
```

```
# 95% CI limits
```

```
df_1$lnRRlow.3=df_1$ln_rr.3-1.96*df_1$SEln_rr.3
table(df_1$lnRRlow.3)
```

```
df_1$lnRRupp.3=df_1$ln_rr.3+1.96*df_1$SEln_rr.3
table(df_1$lnRRupp.3)
```

```
#log of the limits for 95% CI
```

```
df_1$RRlow.3=exp(df_1$lnRRlow.3)
table(df_1$RRlow.3)
```

```
df_1$RRupp.3=exp(df_1$lnRRupp.3)
table(df_1$RRupp.3)
```

```
#Calculate the standardized Mortality Rates 3 (SMR3)
```

```
df_1$q4_q1<-df_1$rate_adj.sum.q.4/df_1$rate_adj.sum.q.1
table(df_1$q4_q1)
```

```
#Calculate 95% CI for SMR3
```

```
#log of SMR3
```

```
df_1$ln_rr.4<-log(df_1$q4_q1)
table(df_1$ln_rr.4)
```

#Standard Error (SE) of ln RR

```
df_1$SEln_rr.4<-sqrt((1/df_1$rate_adj.sum.q.4)+(1/df_1$rate_adj.sum.q.1))
table(df_1$SEln_rr.4)
```

95% CI limits

```
df_1$lnRRlow.4=df_1$ln_rr.4-1.96*df_1$SEln_rr.4
table(df_1$lnRRlow.4)
```

```
df_1$lnRRupp.4=df_1$ln_rr.4+1.96*df_1$SEln_rr.4
table(df_1$lnRRupp.4)
```

#log of the limits for 95% CI

```
df_1$RRlow.4=exp(df_1$lnRRlow.4)
table(df_1$RRlow.4)
```

```
df_1$RRupp.4=exp(df_1$lnRRupp.4)
table(df_1$RRupp.4)
```

*Repeat this analysis to all variables for 10 and 5 years without GB

#####

CALCULATE DECILES FOR PERFORMANCE REGRESSION ANALYSIS AND 95% CI STIMATION

##FOR THE ALL COUNTRY 10 YEAR WITHOUT GB

Determine the deciles based on the number of GBC death (manually determination) and choose #the corresponding percentage of people without GB.

```
df_10$d<-1
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.1801 &
df_10$percentage_people_wo_gb.10<0.2527,2)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.2525 &
df_10$percentage_people_wo_gb.10<0.339,3)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.3383 &
df_10$percentage_people_wo_gb.10<0.413002,4)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.4123 &
df_10$percentage_people_wo_gb.10<0.46349,5)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.46331 &
df_10$percentage_people_wo_gb.10<0.49560,6)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.495535 &
df_10$percentage_people_wo_gb.10<0.54599,7)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.54551 &
df_10$percentage_people_wo_gb.10<0.60753,8)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.60748 &
df_10$percentage_people_wo_gb.10<0.690766,9)
```

```
df_10$d<-replace(df_10$d,df_10$percentage_people_wo_gb.10>0.6902,10)
```

#Collapse by sum the data base

```
df_10$deciles<- summaryBy(cases.23+population~ age4+age5 +d+gender ,FUN=sum, data=df_10)
```

##DECIL DETERMINATION FOR WOMEN

```
df_10$gender<- replace(df_10$gender,df_10$gender==2,0)
table(df_10$gender)
df_10_m<-df_10[df_10$gender==0,]
```

```
df_10_m$deciles_m<-1
```

```
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.2984 &
df_10_m$percentage_people_wo_gb.10<=0.4156, 2)
```

```
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.4156 &
df_10_m$percentage_people_wo_gb.10<= 0.4615,3)
```

```
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.4615 &
df_10_m$percentage_people_wo_gb.10<=0.4861,4)
```

```
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.4861 &
df_10_m$percentage_people_wo_gb.10<=0.5231,5)
```

```
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.5231 &
df_10_m$percentage_people_wo_gb.10<=0.5635,6)
```

```
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.5635 &
df_10_m$percentage_people_wo_gb.10<=0.6057,7)
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.6057 &
df_10_m$percentage_people_wo_gb.10<=0.6563,8)
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.6563 &
df_10_m$percentage_people_wo_gb.10<=0.7352,9)
df_10_m$deciles_m<-replace( df_10_m$deciles_m, df_10_m$percentage_people_wo_gb.10>0.7352,10)
```

#Collapse by sum the data base

```
df_10_m<- summaryBy(cases.23+population~years_b+ age4+age5 +deciles_m,FUN=sum, data=df_10_m)
```

##Complete the cases of GBC and population by age and year to have the all observation per decil (GBC=0 and population=1)

```
df_10_m<- df_10_m%>%
  complete(deciles_m,nesting(age5,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

##DECIL DETERMINATION FOR MEN

```
df_10_h<-df_10[df_10$gender==1,]
df_10_h$deciles_h<-1
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.1151 &
df_10_h$percentage_people_wo_gb.10<=0.1760, 2)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.1760 &
df_10_h$percentage_people_wo_gb.10<=0.2115,3)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.2115 &
df_10_h$percentage_people_wo_gb.10<=0.2358,4)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.2358 &
df_10_h$percentage_people_wo_gb.10<=0.2792,5)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.2792 &
df_10_h$percentage_people_wo_gb.10<=0.3080,6)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.3080 &
df_10_h$percentage_people_wo_gb.10<=0.3327,7)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.3327 &
df_10_h$percentage_people_wo_gb.10<=0.3986,8)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.3986 &
df_10_h$percentage_people_wo_gb.10<=0.4542,9)
df_10_h$deciles_h<-replace( df_10_h$deciles_h, df_10_h$percentage_people_wo_gb.10>0.4542,10)
```

#Collapse by sum the data base

```
df_10_h<- summaryBy(cases.23+population~years_b+ age4+age5 +deciles_h,FUN=sum, data=df_10_h)
```

##Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```
df_10_h<- df_10_h%>%
  complete(deciles_h,nesting(age5,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

##DECIL DETERMINATION FOR MAPUCHES

```
table(df_10$ancestry)
df_10_mapu<-df_10[df_10$ancestry==1,]
df_10_mapu$deciles_mapu<-1
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.1332 &
df_10_mapu$percentage_people_wo_gb.10<=0.1974, 2)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.1974 &
df_10_mapu$percentage_people_wo_gb.10<= 0.2496,3)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.2496 &
df_10_mapu$percentage_people_wo_gb.10<=0.3233,4)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.3233 &
df_10_mapu$percentage_people_wo_gb.10<=0.4130,5)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.4130 &
df_10_mapu$percentage_people_wo_gb.10<=0.4794,6)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.4794 &
df_10_mapu$percentage_people_wo_gb.10<=0.5252,7)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.5252 &
df_10_mapu$percentage_people_wo_gb.10<=0.5857,8)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.5857 &
df_10_mapu$percentage_people_wo_gb.10<=0.6834,9)
df_10_mapu$deciles_mapu<-replace( df_10_mapu$deciles_mapu, df_10_mapu$percentage_people_wo_gb.10>0.6834,10)
```

#Collapse by sum the data base

```
df_10_mapu<- summaryBy(cases.23+population~years_b+ age4+age5 +gender+deciles_mapu,FUN=sum,
data=df_10_mapu)
```

##Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```
df_10_mapu<- df_10_mapu%>%
  complete(deciles_mapu,nesting(age5,years_b,gender),fill=list(cases.23.sum=0,population.sum=1))
```


#DECIL DETERMINATION FOR AYMARA

```
table(df_10$ancestry)
df_10_aymara<-df_10[df_10$ancestry==2,]
df_10_aymara$deciles_aymara<-1
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.1455 & df_10_aymara$percentage_people_wo_gb.10<=0.2581, 2)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.2581 & df_10_aymara$percentage_people_wo_gb.10<= 0.3554,3)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.3554 & df_10_aymara$percentage_people_wo_gb.10<=0.4046,4)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.4046 & df_10_aymara$percentage_people_wo_gb.10<=0.5049,5)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.5049 & df_10_aymara$percentage_people_wo_gb.10<=0.5847,6)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.5847 & df_10_aymara$percentage_people_wo_gb.10<=0.7352,7)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.7352 & df_10_aymara$percentage_people_wo_gb.10<=0.8389,8)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.8389 & df_10_aymara$percentage_people_wo_gb.10<=0.9761,9)
df_10_aymara$deciles_aymara<-replace( df_10_aymara$deciles_aymara,
df_10_aymara$percentage_people_wo_gb.10>0.9761,10)
```

#Collapse by sum the data base

```
df_10_aymara<- summaryBy(cases.23+population~years_b+gender+ age4+age5 +deciles_aymara,FUN=sum,
data=df_10_aymara)
```

#Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```
df_10_aymara<- df_10_aymara%>%
complete(deciles_aymara,nesting(age5,years_b,gender),fill=list(cases.23.sum=0,population.sum=1))
```

#DECIL DETERMINATION FOR OTHER

```
table(df_10$ancestry)
df_10_otra<-df_10[df_10$ancestry==3,]
df_10_otra$deciles_otra<-1
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.2568 &
df_10_otra$percentage_people_wo_gb.10<=0.33007, 2)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.33007 &
df_10_otra$percentage_people_wo_gb.10<= 0.4106,3)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.4106 &
df_10_otra$percentage_people_wo_gb.10<=0.4557,4)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.4557 &
df_10_otra$percentage_people_wo_gb.10<=0.4805,5)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.4805 &
df_10_otra$percentage_people_wo_gb.10<=0.522,6)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.522 &
df_10_otra$percentage_people_wo_gb.10<=0.572965,7)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.572965 &
df_10_otra$percentage_people_wo_gb.10<=0.6162,8)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.6162 &
df_10_otra$percentage_people_wo_gb.10<=0.6903,9)
df_10_otra$deciles_otra<-replace( df_10_otra$deciles_otra, df_10_otra$percentage_people_wo_gb.10>0.6903,10)
```

#Collapse by sum the data base

```
df_10_otra<- summaryBy(cases.23+population~years_b+gender+ age4+age5 +deciles_otra,FUN=sum,
data=df_10_otra)
```

#Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```
df_10_otra<- df_10_otra%>%
complete(deciles_otra,nesting(age5,years_b,gender),fill=list(cases.23.sum=0,population.sum=1))
```

#DECIL DETERMINATION FOR AGE RANGE 50-59

```
df_10_age2<-df_10[df_10$age10==2,]
df_10_age2$deciles_age2<-1
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.19481 &
df_10_age2$percentage_people_wo_gb.10<=0.2444,2)
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.2444 &
df_10_age2$percentage_people_wo_gb.10<=0.38945,3)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.38945 &
df_10_age2$percentage_people_wo_gb.10<=0.4633,4)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.4633 &
df_10_age2$percentage_people_wo_gb.10<=0.5482,5)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.5482 &
df_10_age2$percentage_people_wo_gb.10<=0.6019,6)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.6019 &
df_10_age2$percentage_people_wo_gb.10<=0.6486,7)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.6486 &
df_10_age2$percentage_people_wo_gb.10<=0.7140,8)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.7140 &
df_10_age2$percentage_people_wo_gb.10<=0.792,9)
```

```
df_10_age2$deciles_age2<-replace( df_10_age2$deciles_age2, df_10_age2$percentage_people_wo_gb.10>0.792,10)
```

```
#Collapse by sum the data base
```

```
df_10_age2<- summaryBy(cases.23+population~years_b+gender+age9+age10 +deciles_age2,FUN=sum,
data=df_10_age2)
```

```
#Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)
```

```
df_10_age2<- df_10_age2%>%
complete(deciles_age2,nesting(age10,gender,years_b),fill=list(cases.23.sum=0,population.sum=1))
```

```
#Complete age label category
```

```
df_10_age2$age9<-replace(df_10_age2$age9,df_10_age2$age10==2,"50-59")
```

#DECIL DETERMINATION FOR AGE RANGE 60-69

```
df_10_age3<-df_10[df_10$age10==3,]  
df_10_age3$deciles_age3<-1  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.1854 &  
df_10_age3$percentage_people_wo_gb.10<=0.244, 2)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.244 &  
df_10_age3$percentage_people_wo_gb.10<=0.3121, 3)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.3121 &  
df_10_age3$percentage_people_wo_gb.10<=0.4042, 4)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.4042 &  
df_10_age3$percentage_people_wo_gb.10<=0.4607, 5)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.4607 &  
df_10_age3$percentage_people_wo_gb.10<=0.49554, 6)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.49554 &  
df_10_age3$percentage_people_wo_gb.10<=0.5372, 7)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.5372 &  
df_10_age3$percentage_people_wo_gb.10<=0.60495, 8)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.60495 &  
df_10_age3$percentage_people_wo_gb.10<=0.674, 9)  
df_10_age3$deciles_age3<-replace( df_10_age3$deciles_age3, df_10_age3$percentage_people_wo_gb.10>0.674, 10)
```

#Collapse by sum the data base

```
df_10_age3<- summaryBy(cases.23+population~years_b+gender+age9+age10 +deciles_age3, FUN=sum,  
data=df_10_age3)
```

#Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```
df_10_age3<- df_10_age3%>%  
complete(deciles_age3, nesting(age10, gender, years_b), fill=list(cases.23.sum=0, population.sum=1))
```

#Complete age label category

```
df_10_age3$age9<-replace(df_10_age3$age9, df_10_age3$age10==3, "60-69")
```

#DECIL DETERMINATION FOR AGE 70-80

```
df_10_age4<-df_10[df_10$age10==4,]  
df_10_age4$deciles_age4<-1  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.2257 &  
df_10_age4$percentage_people_wo_gb.10<=0.3118, 2)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.3118 &  
df_10_age4$percentage_people_wo_gb.10<=0.364, 3)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.364 &  
df_10_age4$percentage_people_wo_gb.10<=0.4234, 4)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.4234 &  
df_10_age4$percentage_people_wo_gb.10<=0.4701, 5)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.4701 &  
df_10_age4$percentage_people_wo_gb.10<=0.4953, 6)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.4953 &  
df_10_age4$percentage_people_wo_gb.10<=0.5341, 7)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.5341 &  
df_10_age4$percentage_people_wo_gb.10<=0.59, 8)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.59 &  
df_10_age4$percentage_people_wo_gb.10<=0.664, 9)  
df_10_age4$deciles_age4<-replace( df_10_age4$deciles_age4, df_10_age4$percentage_people_wo_gb.10>0.664, 10)
```

#Collapse by sum the data base

```
df_10_age4<- summaryBy(cases.23+population~years_b+gender+age9+age10 +deciles_age4, FUN=sum,  
data=df_10_age4)
```

#Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```
complete(deciles_age4, nesting(age10, gender, years_b), fill=list(cases.23.sum=0, population.sum=1))
```

#Complete age label category

```
df_10_age4$age9<-replace(df_10_age4$age9, df_10_age4$age10==4, "70-80")
```

#####

DECILES FOR CANCER CASE 10 YEARS WITOUTH GB

#Select cancer codes

```
year_10_ca<-filter(year_10_long,  
  (year_10_long$code==15)|  
  (year_10_long$code==16)|  
  (year_10_long$code==18)|  
  (year_10_long$code==22)|  
  (year_10_long$code==27)|  
  (year_10_long$code==24)|  
  (year_10_long$code==25)|  
  (year_10_long$code==4))
```

#Past the cancer codes to a list

```
list_7<-split(year_10_ca,year_10_ca$code, drop = FALSE)
```

#DECIL DETERMINATION

#OESOPHAGUS

```
list_7[[1]]$deciles<-1  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.137 &  
list_7[[1]]$percentage_people_wo_gb<=0.196,2)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.196 &  
list_7[[1]]$percentage_people_wo_gb<= 0.235 ,3)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.235 &  
list_7[[1]]$percentage_people_wo_gb<=0.276,4)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.276 &  
list_7[[1]]$percentage_people_wo_gb<=0.3198926,5)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.3198926 &  
list_7[[1]]$percentage_people_wo_gb<=0.3721,6)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.371 &  
list_7[[1]]$percentage_people_wo_gb<=0.4368,7)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.4368 &  
list_7[[1]]$percentage_people_wo_gb<=0.5001,8)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.5001 &  
list_7[[1]]$percentage_people_wo_gb<=0.595,9)  
list_7[[1]]$deciles<-replace( list_7[[1]]$deciles, list_7[[1]]$percentage_people_wo_gb>0.595,10)
```

#STOMACH

```
list_7[[2]]$deciles<-1  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.1287 &  
list_7[[2]]$percentage_people_wo_gb<=0.194,2)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.194 &  
list_7[[2]]$percentage_people_wo_gb<= 0.233,3)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.233 &  
list_7[[2]]$percentage_people_wo_gb<=0.2792,4)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.2792 &  
list_7[[2]]$percentage_people_wo_gb<=0.3144,5)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.3144 &  
list_7[[2]]$percentage_people_wo_gb<=0.36257,6)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.36257 &  
list_7[[2]]$percentage_people_wo_gb<=0.4302,7)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.4302 &  
list_7[[2]]$percentage_people_wo_gb<=0.500,8)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.500 &  
list_7[[2]]$percentage_people_wo_gb<=0.5953,9)  
list_7[[2]]$deciles<-replace( list_7[[2]]$deciles, list_7[[2]]$percentage_people_wo_gb>0.5953,10)  
table(list_7[[2]]$deciles)
```

#COLON

```
list_7[[3]]$deciles<-1
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.1695 &
list_7[[3]]$percentage_people_wo_gb<=0.2303,2)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.2303 &
list_7[[3]]$percentage_people_wo_gb<=0.2852,3)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.2852 &
list_7[[3]]$percentage_people_wo_gb<=0.3313,4)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.3313 &
list_7[[3]]$percentage_people_wo_gb<=0.4046,5)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.4046 &
list_7[[3]]$percentage_people_wo_gb<=0.4546,6)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.4546 &
list_7[[3]]$percentage_people_wo_gb<=0.5016,7)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.5016 &
list_7[[3]]$percentage_people_wo_gb<=0.5672,8)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.5672 &
list_7[[3]]$percentage_people_wo_gb<=0.663,9)
list_7[[3]]$deciles<-replace( list_7[[3]]$deciles, list_7[[3]]$percentage_people_wo_gb>0.663,10)
```

#LIVER

```
list_7[[4]]$deciles<-1
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.16924 &
df_10_mod$percentage_people_wo_gb.10<=0.21761,2)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.21761 &
df_10_mod$percentage_people_wo_gb.10<=0.2672,3)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.2672 &
df_10_mod$percentage_people_wo_gb.10<=0.3082,4)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.3082 &
df_10_mod$percentage_people_wo_gb.10<=0.35576,5)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.35576 &
df_10_mod$percentage_people_wo_gb.10<=0.4184,6)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.4184 &
df_10_mod$percentage_people_wo_gb.10<=0.47501,7)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.47501 &
df_10_mod$percentage_people_wo_gb.10<=0.52822,8)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.52822 &
df_10_mod$percentage_people_wo_gb.10<=0.62844,9)
list_7[[4]]$deciles<-replace( list_7[[4]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.62844,10)
table(list_7[[4]]$deciles)
```

#BILE DUCT

```
list_7[[5]]$deciles<-1
list_7[[5]]$deciles<-1
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.1703 &
df_10_mod$percentage_people_wo_gb.10<=0.2465,2)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.2465 &
df_10_mod$percentage_people_wo_gb.10<=0.29101,3)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.29101 &
df_10_mod$percentage_people_wo_gb.10<=0.34785,4)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.34785 &
df_10_mod$percentage_people_wo_gb.10<=0.4184,5)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.4184 &
df_10_mod$percentage_people_wo_gb.10<=0.4696,6)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.4696 &
df_10_mod$percentage_people_wo_gb.10<=0.5144,7)
```

```
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.5144 &
df_10_mod$percentage_people_wo_gb.10<=0.568,8)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.568 &
df_10_mod$percentage_people_wo_gb.10<=0.65795,9)
list_7[[5]]$deciles<-replace( list_7[[5]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.65795,10)
```

EXTRAHEPATIC BILE DUCT

```
list_7[[6]]$deciles<-1
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.1705 &
list_7[[6]]$percentage_people_wo_gb<=0.2525, 2)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.2525 &
list_7[[6]]$percentage_people_wo_gb<=0.2898,3)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.2898 &
list_7[[6]]$percentage_people_wo_gb<=0.3476,4)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.3476 &
list_7[[6]]$percentage_people_wo_gb<=0.4184,5)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.4184 &
list_7[[6]]$percentage_people_wo_gb<=0.4701,6)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.4701 &
list_7[[6]]$percentage_people_wo_gb<=0.5263,7)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.5263 &
list_7[[6]]$percentage_people_wo_gb<=0.5727,8)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.5727 &
list_7[[6]]$percentage_people_wo_gb<=0.6652,9)
list_7[[6]]$deciles<-replace( list_7[[6]]$deciles, list_7[[6]]$percentage_people_wo_gb>0.6652,10)

list_7[[6]]<-list_7[[6]]%>%
  group_by(deciles) %>% mutate(total_cases = sum(cases))

table(list_7[[6]]$deciles,list_7[[6]]$total_cases)
```

INTRAHEPATIC BILE DUCT

```
list_7[[7]]$deciles<-1
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.16946 &
df_10_mod$percentage_people_wo_gb.10<=0.2369,2)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.2369 &
df_10_mod$percentage_people_wo_gb.10<=0.29745,3)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.29745 &
df_10_mod$percentage_people_wo_gb.10<=0.352,4)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.352 &
df_10_mod$percentage_people_wo_gb.10<=0.4196,5)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.4196 &
df_10_mod$percentage_people_wo_gb.10<=0.46792,6)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.46792 &
df_10_mod$percentage_people_wo_gb.10<=0.50468,7)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.50468 &
df_10_mod$percentage_people_wo_gb.10<=0.5636,8)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.5636 &
df_10_mod$percentage_people_wo_gb.10<=0.62844,9)
list_7[[7]]$deciles<-replace( list_7[[7]]$deciles, df_10_mod$percentage_people_wo_gb.10>0.62844,10)
```

PANCREAS

```

list_7[[8]]$deciles<-1
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.1788 &
list_7[[8]]$percentage_people_wo_gb<=0.2353, 2)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.2353 &
list_7[[8]]$percentage_people_wo_gb<=0.2938,3)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.2938 &
list_7[[8]]$percentage_people_wo_gb<=0.3478,4)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.3478 &
list_7[[8]]$percentage_people_wo_gb<=0.41453,5)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.41453 &
list_7[[8]]$percentage_people_wo_gb<=0.4683,6)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.4683 &
list_7[[8]]$percentage_people_wo_gb<=0.508,7)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.508 &
list_7[[8]]$percentage_people_wo_gb<=0.5818,8)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.5818 &
list_7[[8]]$percentage_people_wo_gb<=0.671,9)
list_7[[8]]$deciles<-replace( list_7[[8]]$deciles, list_7[[8]]$percentage_people_wo_gb>0.671,10)

```

ALL CITE EXCEPT 23

```

list_7[[9]]$deciles<-1
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.1694 &
list_7[[9]]$percentage_people_wo_gb<=0.22693, 2)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.22693 &
list_7[[9]]$percentage_people_wo_gb<= 0.2815,3)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.2815 &
list_7[[9]]$percentage_people_wo_gb<=0.3232,4)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.3232 &
list_7[[9]]$percentage_people_wo_gb<=0.3896,5)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.3896 &
list_7[[9]]$percentage_people_wo_gb<=0.4376,6)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.4376 &
list_7[[9]]$percentage_people_wo_gb<=0.4914,7)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.4914 &
list_7[[9]]$percentage_people_wo_gb<=0.5548,8)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.5548 &
list_7[[9]]$percentage_people_wo_gb<=0.642,9)
list_7[[9]]$deciles<-replace( list_7[[9]]$deciles, list_7[[9]]$percentage_people_wo_gb>0.642,10)

```

#Collapse by sum the data base

```

list_8<-list()
for(j in 1:9) {
  list_8[[j]]<-summaryBy(cases+population~ + age4+age5 +deciles+gender+code ,FUN=sum, data=list_7[[j]])
}

```

#Complete the cases of GBC and population by age, gender and year to have the all observation per decil (GBC=0 and population=1)

```

for(n in 1:9){
  list_8[[n]]<- list_8[[n]]%>%
  complete(deciles,nesting(age5,gender,code),fill=list(cases.sum=0,population.sum=1))
}

```

##Complete age label category

```

for(n in 1:9) {
  list_8[[n]]$age6<-1
  list_8[[n]]$age6<-replace(list_8[[n]]$age6, list_8[[n]]$age5==1,"40-45")
  list_8[[n]]$age6<-replace(list_8[[n]]$age6, list_8[[n]]$age5==2,"46-50")
  list_8[[n]]$age6<-replace(list_8[[n]]$age6, list_8[[n]]$age5==3,"51-55")
  list_8[[n]]$age6<-replace(list_8[[n]]$age6, list_8[[n]]$age5==4,"56-60")
  list_8[[n]]$age6<-replace(list_8[[n]]$age6, list_8[[n]]$age5==5,"61-65")
  list_8[[n]]$age6<-replace(list_8[[n]]$age6, list_8[[n]]$age5==6,"66-70")
}

```

```
list_8[[n]]$age6<-replace(list_8[[n]]$age6,list_8[[n]]$age5==7,"71-75")
list_8[[n]]$age6<-replace(list_8[[n]]$age6,list_8[[n]]$age5==8,"76-80")
}
```

#Rename some variables

```
for(n in 1:9) {
  list_8[[n]]= list_8[[n]] %>% #Rename variable reg2
  rename(cases=cases.sum)

  list_8[[n]]= list_8[[n]] %>% #Rename variable reg2
  rename(population=population.sum)
}
```

#Calculate the mortality adjusted rate

```
for(n in 1:9) {

  list_8[[n]] <- list_8[[n]]%>% group_by(age5, gender)%>%
  mutate(pop.sum = sum(population))

  list_8[[n]] <- list_8[[n]]%>% group_by(deciles, gender)%>%
  mutate(pop.total = sum(pop.sum))

  list_8[[n]]$proportion<-list_8[[n]]$pop.sum/list_8[[n]]$pop.total

  list_8[[n]]$rate<-list_8[[n]]$cases/list_8[[n]]$population
  list_8[[n]]$rate_adj<-list_8[[n]]$rate*list_8[[n]]$proportion*list_8[[n]]$pop.total
  list_8[[n]]<- list_8[[n]]%>% group_by(deciles)%>%
  mutate(rate_adj.sum.q = sum(rate_adj))#rate ajustada total agrupada para cada cuartil

  list_8[[n]]<- list_8[[n]]%>% group_by(deciles)%>%
  mutate(cases_total = sum(cases))#rate ajustada total agrupada para cada cuartil
}
```

#Drop some variables

```
for(n in 1:9) {
  list_8[[n]] <-list_8[[n]][ c(-2,-3,-5,-6,-7,-8,-9:-13)]
}
```

#Reshape the data frame to a wide format

```
list_9<-list()

for(n in 1:9) {
  list_9[[n]]<-reshape(as.data.frame(list_8[[n]]),
    timevar = "deciles",
    idvar = c("code"),
    v.names=c("rate_adj.sum.q", "cases_total"),
    direction = "wide")
}
```

#Call the list to a data frame

```
df_cancer_10<- do.call("rbind", list_9)
```

#SMR CALCULATATION, REGESSION ANALYSIS AND 95%CI STIMATION FOR THE ALL COUNTRY

#Complete the cases of GBC and population by age and gender to have the all observation per decil (GBC=0 and population=1)

```
df_10_deciles<- df_10_deciles%>%
  complete(d,nesting(age5,gender),fill=list(cases.23.sum=0,population.sum=1))
```

#Generate age category to complete the new variables generates in the previous line

```
df_10_deciles$age6<-1
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==1,"40-45")
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==2,"46-50")
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==3,"51-55")
```



```
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==4,"56-60")
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==5,"61-65")
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==6,"66-70")
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==7,"71-75")
df_10_deciles$age6<-replace(df_10_deciles$age6,df_10_deciles$age5==8,"76-80")
```

#Calculate the mortality adjusted rate

```
df_10_deciles <- df_10_deciles%>% dplyr::group_by(age5, gender)%>%
  dplyr::mutate(pop.sum = sum(population.sum))
df_10_deciles <- df_10_deciles%>% dplyr::group_by(d, gender)%>%
  dplyr::mutate(pop.total = sum(pop.sum))
df_10_deciles$proportion<-df_10_deciles$pop.sum/df_10_deciles$pop.total
df_10_deciles$rate<-df_10_deciles$cases.23.sum/df_10_deciles$population.sum
df_10_deciles$rate_adj<-df_10_deciles$rate*df_10_deciles$proportion*df_10_deciles$pop.total
df_10_deciles<- df_10_deciles%>% dplyr::group_by(deciles)%>%
  dplyr::mutate(rate_adj.sum.q = sum(rate_adj))
df_10_deciles<- df_10_deciles%>% dplyr::group_by(deciles)%>%
  dplyr::mutate(cases.23.total = sum(cases.23.sum))

df_10_deciles<- df_10_deciles %>%
dplyr::rename(cases_total=cases.23.total)
```

#Reshape data frame to a wide position

```
df_10_deciles<-df_10_deciles[,c(-4)]

df_10_deciles_w<-reshape(as.data.frame(df_10_deciles),
  timevar = "d",
  idvar = c("age5", "age6", "gender"),
  v.names=c("rate_adj.sum.q", "cases.23.sum", "cases.23.total",
"population.sum", "pop.sum", "pop.total", "proportion", "rate", "rate_adj"),
  direction = "wide")
```

#Calculate the Standardized Mortality Rates 1 (SMR1)

```
df_10_deciles_w$d_2<-df_10_deciles_w$rate_adj.sum.d.2/df_10_deciles_w$rate_adj.sum.d.1
table(df_10_deciles_w$d2_d1)
```

#Calculate 95% CI for SMR1

#log of SMR1

```
df_10_deciles_w$ln_rr<-log(df_10_deciles_w$d.1)
table(df_10_deciles_w$ln_rr)
```

#Standard Error (SE) of ln RR

```
df_10_deciles_w$SEln_rr<-sqrt((1/df_10_deciles_w$rate_adj.sum.d.2)+(1/df_10_deciles_w$rate_adj.sum.d.1))
table(df_10_deciles_w$SEln_rr)
```

95% CI limits

```
df_10_deciles_w$lnRRlow=df_10_deciles_w$ln_rr-1.96*df_10_deciles_w$SEln_rr
table(df_10_deciles_w$lnRRlow)
```

```
df_10_deciles_w$lnRRupp=df_10_deciles_w$ln_rr+1.96*df_10_deciles_w$SEln_rr
table(df_10_deciles_w$lnRRupp)
```

#log of the limits for 95% CI

```
df_10_deciles_w$RRlow_2=exp(df_10_deciles_w$lnRRlow)
table(df_10_deciles_w$RRlow)
```

```
df_10_deciles_w$RRupp_2=exp(df_10_deciles_w$lnRRupp)
table(df_10_deciles_w$RRupp)
```

*Repeat calculations for:

```
d3/d1
d4/d1
d5/d1
d6/d1
d7/d1
d8/d1
d9/d1
d10/d1
```

REGRESION ANALISYS

#Rename cases

```
df_10_deciles_w <- df_10_deciles_w %>%
  rename(cases_1=cases_total.1) %>%
  rename(cases_2=cases_total.2) %>%
  rename(cases_3=cases_total.3) %>%
  rename(cases_4=cases_total.4) %>%
  rename(cases_5=cases_total.5) %>%
  rename(cases_6=cases_total.6) %>%
  rename(cases_7=cases_total.7) %>%
  rename(cases_8=cases_total.8) %>%
  rename(cases_9=cases_total.9) %>%
  rename(cases_10=cases_total.10)
```

#Resume the data base

```
resume <- c("cases_1","cases_2" ,"d_2","RRlow_2","RRupp_2",
           "cases_3","d_3","RRlow_3","RRupp_3",
           "cases_4","d_4","RRlow_4","RRupp_4",
           "cases_5","d_5","RRlow_5","RRupp_5",
           "cases_6","d_6","RRlow_6","RRupp_6",
           "cases_7","d_7","RRlow_7","RRupp_7",
           "cases_8","d_8","RRlow_8","RRupp_8",
           "cases_9","d_9","RRlow_9","RRupp_9",
           "cases_10","d_10", "RRlow_10","RRupp_10")
```

```
df_10_deciles_resume <- df_10_deciles_w [, resume ]
```

#Reshape the data base to long

```
df_10_deciles_long <- pivot_longer(
  data = df_10_deciles_resume,
  cols = starts_with(c("cases_", "d_", "RRlow_", "RRupp_")),
  names_to = c(".value", "decil"),
  names_sep = "_"
)
```

#Rename d

```
df_10_deciles_long <- df_10_deciles_long %>%
  rename(smr=d)
```

#Generate a decil data base

```
summary(df_10$percentage_people_wo_gb.10)
```

```
decil<-1:10
```

```
decil_range_lower_all<-c(0, 0.1801, 0.2525, 0.3383, 0.4123, 0.46331, 0.495535, 0.54551, 0.60748, 0.6902)
```

```
decil_range_upper_all<-c(0.1801, 0.2525, 0.3383, 0.4123, 0.46331, 0.495535, 0.54551, 0.60748, 0.6902, 1.8056)
```

```
decil_all <- data.frame(decil, animal, decil_range_lower_all, decil_range_upper_all)
```

#Join the deciles values

```
df_10_deciles_long<-merge(df_10_deciles_long, decil_all, all.x=T, all.y = F)
```

#Calculate the range decil mean

```
df_10_deciles_long$mean_d<- (df_10_deciles_long$decil_range_lower_all+
  df_10_deciles_long$decil_range_upper_all)/2
```

#Calculate the standard deviation

```
df_10_deciles_long$sd<-( df_10_deciles_long$RRlow- df_10_deciles_long$RRupp)/(2*1.96)
```

#Calculate regression

```
lm<-lm(smr~mean_d, data= df_10_deciles_long)
summary(lm)
```

#CALCULATE DE 95%CI

#Generate 1000 samples from a normal distribution excluding the first and last decil

```
smr<- cuartil_10_deciles$smr
sd<- cuartil_10_deciles$sd
mean_decil<- df_10_deciles_long$mean_d
```

```
list<-list()
```

```
for(i in 1:8){
  list[[i]]<- data.frame(
```

```

    smr= rnorm(1000, mean=smr[[i]],sd=sd[[i]])
    list[[i]]$id_exp=sd(1,1000,length=1000)
    list[[i]]$d_media=rep(mean_decil[i])
}

```

#Past the list to a data frame

```

sim_all<-do.call("rbind", list)
sim_all<- sim[order(sim$id_exp),]

```

#Pass the sim data frame to a list

```

sim_list <- split(sim, f = sim$id_exp)

```

#Run the regression for each sample

```

mfun<- function(data) lm(smr~d_media,data=data)
results<-list()
results<-lapply(sim_list, FUN=mfun)

```

#Obtained regression coef

```

coef_all<-ldply(coef_all,function(x)coef(x))

coef_all <-coef_all[order(coef_upper$q_media),]
coef_all$ID <- seq_along(coef_upper [,1])

```

**The values for the IC were the coefficient in the position 25 and 975

*Repeat this analysis for gender, ancestry(or regions) and age range.

**In the case of 5 years without GB, deciles were manually calculated using the number of deaths by GB (cases.23) grouped by: All country, sex, ancestry and age category, then the corresponding percentage of persons without GB after 5 years was chosen as decile. The procedure to obtain the SMR, the regression coefficient and the 95%CI estimate was the same as shown above.

#####

INTERRUPTED TIME SERIES ANALISYS##

##CHOLECYSTECTOMY DATA BASE

```

variable INTERV_Q=1
str(GBD$INTERV_Q)
cole_its<-filter(GBD, INTERV_Q==1)
cole_its$INTERV_Q<-as.integer(cole_its$INTERV_Q)

```

#Review the variables of interest for the study

```

##SEXO
table(cole_its$SEXO)
cole_its<-filter(cole_its, SEXO<3)

```

##AGE

```

table(cole_its$AGE_CANT)

```

##REGION

```

table(cole_its$REGION_RESIDENCIA)
cole_its<-filter(cole_its, cole_its$REGION_RESIDENCIA<99)

```

#Recode the REGION_RESIDENCIA variable, to obtain 13 regions instead of 16 regions

```

cole_its$reg2<-cole_its$REGION_RESIDENCIA
cole_its$reg2<-replace(cole_its$reg2,cole_its$REGION_RESIDENCIA==15,1)
cole_its$reg2<-replace(cole_its$reg2,cole_its$REGION_RESIDENCIA==16,8)
cole_its$reg2<-replace(cole_its$reg2,cole_its$REGION_RESIDENCIA==14,10)
table(cole_its$reg2)

```

#Collapse the cole_its data base using the reg2 variable

```

str(cole_its)

```

```

cole_its$cases<-1
cole_its<- summaryBy(cases ~SEXO + ANO_EGRESO+ AGE_CANT+reg2+ INTERV_Q,FUN=sum, data=cole_its)

```

#Rename de INTERV_Q to persons without GB (persons_wo_GB)

```

cole_its=cole_its%>%
  rename(interv=INTERV_Q) %>%
  rename(year=ANO_EGRESO) %>%
  rename(age=AGE_CANT) %>%
  rename(gender=SEXO) %>%
  rename(cases=cases.sum)

```

##Drop regions with <10 death for GB

```

cole_its<-filter(cole_its,reg2!=11)
cole_its<-filter(cole_its,reg2!=12)

```

```
##Keep until 2018
```

```
cole_its<-filter(cole_its,year<2019)
```

```
#Recode ancestry
```

```
cole_its$ancestry<-3  
cole_its$ancestry<-replace(cole_its$ancestry,cole_its$region>0 & cole_its$region<4 ,2)  
cole_its$ancestry<-replace(cole_its$ancestry,cole_its$region>7 & cole_its$region<11,1)  
cole_its$ancestry_label<-"other"  
cole_its$ancestry_label<-replace(cole_its$ancestry_label,cole_its$ancestry==2,"aymara")  
cole_its$ancestry_label<-replace(cole_its$ancestry_label,cole_its$ancestry==1,"mapuche")  
table(cole_its$ancestry_label)  
table(cole_its$ancestry_label,cole_its$ancestry)
```

```
#Write to a dta format to work in STATA
```

```
write.dta(cole_its, "cole_its.dta")
```

```
##GBC DATA BASE
```

```
gbc_its<-filter(mortality_1990_2018, ANO_DEF>2001)
```

```
#Explore the variable of interest
```

```
#SEXO
```

```
table(cancer_GBC$SEXO)  
gbc_its<-filter(gbc_its, SEXO<3)
```

```
#AGE
```

```
table(gbc_its$AGE_CANT)  
gbc_its<-filter(gbc_its, AGE_CANT<200)
```

```
#REGION
```

```
table(gbc_its$REG_RES)  
gbc_its<-filter(gbc_its, REG_RES<99)
```

```
#Recode the REG_RES variable, leave 13 regions instead of 16
```

```
gbc_its$reg2<-gbc_its$REG_RES  
gbc_its$reg2<-replace(gbc_its$reg2,gbc_its$REG_RES==15,1)  
gbc_its$reg2<-replace(gbc_its$reg2,gbc_its$REG_RES==16,8)  
gbc_its$reg2<-replace(gbc_its$reg2,gbc_its$REG_RES==14,10)  
table(gbc_its$reg2)
```

```
#Gen a counting variable
```

```
gbc_its$cases<-1  
str(gbc_its$cases)
```

```
#Collapase the gbc_its data base according to the new  
# variable region, reg2
```

```
#Keep only the GBC death
```

```
gbc_its<-filter(gbc_its, DIAG1=="C23X")
```

```
gbc_its<- summaryBy(cases~ SEXO + ANO_DEF+ + AGE_CANT+reg2+DIAG1,FUN=sum, data=gbc_its)
```

```
#Rename the variable cases
```

```
gbc_its=gbc_its%>%  
  rename(cases=cases.sum) %>%  
  rename(year=ANO_EGRESO) %>%  
  rename(age=AGE_CANT) %>%  
  rename(gender=SEXO) %>%
```

```
#Write to a dta format to work in STATA
```

```
write.dta(gbc_its, "gbc_its.dta")
```

**STATA

//CALCULATE THE COLECYSTECTOMY AGE ADJUSTED RATE AND CREATE THE VARIABLES FOR ITS ANALISYS

// FOR THE ALL COUNTRY

// FIRST OBTAIN THE POPULATIONS DATA BASES

**Obtain the population the general population, from a public data base

```
clear all
use population_2002_2035.dta
```

**Recode region

```
table region
gen reg2=region
tab reg2
drop if region==.
replace reg2= 1 if reg2==15
replace reg2=10 if reg2==14
replace reg2=8 if reg2==16
table reg2
drop if reg2==11 | reg2==12
```

**Keep the year of interest

```
keep if year>2001 & year<2019
```

```
rename (sexo age poblacon) (gender age population)
```

**Collapase the population data base by sum considering all categories

```
collapse (sum) population, by(year gender age reg2)
save population_2002_2019_reg2,replace
```

**Collapse the data base just for year and age

```
collapse (sum) population, by(year age )
```

```
save population_2002-2018_year.dta,replace
```

**The reference population from 2002 will be considered for calculate the standardized cholecystectomy rate

```
keep if year==2002
rename pobulation pob_ref
save population_2002_year.dta,replace
```

**Calculate the total population of 2002

```
gen id=1
collapse (sum) pob_ref, by(id)
rename pob_ref pob_total
save population_2002_total_year,replace
```

//CALCULATE THE STANDARIZED COLEICISTECTOMY RATE

**Collapse the data base for the all country by age

```
clear all
use cole_its
collapse (sum) cases, by(year age)
save cole_2002_2018_year,replace
```

**Merge the all population from 2002 to 2018

```
merge 1:1 year age using population_2002-2018_year
drop _merge
drop if cases==.
drop if population==.
```

**Merge the reference population of 2002

```
merge m:1 age using year_ref_2002_year.dta
drop _merge
drop if cases==.
gen id=1
```

**Merge the total population from 2002

```
merge m:m id using population_2002_total_year
drop _merge
drop if cases==.
```

```
save cole_2002-2018_year.dta, replace
```

```
**Calculate the age adjusted mortality rate
```

```
*Calculate the crude rate
```

```
gen crude_rate=(cases/population)*100000
```

```
*Multiply the crude rate with the reference population
```

```
gen crude_rate_pref=crude_rate*pob_ref
```

```
*Sum of the quotients between the crude rate and the reference population per year
```

```
bysort year: egen total_rate=total(crude_rate_pref)
```

```
* Divide the total rate by the total reference population to obtain the adjusted rate
```

```
gen rate_adj=total_rate/pob_total
```

```
*Keep the maximum value of rate adjusted
```

```
collapse(max) rate_adj, by(year)
```

```
save cole_est_2002-2018_year, replace
```

```
*Repeat the rate adjusted in other 3 variables , just to work in a loop in Rstudio
```

```
rename (rate_adj) (rate_adj1)
```

```
gen rate_sdj2= rate_adj1
```

```
gen rate_sdj3= rate_adj1
```

```
gen id="all"
```

```
gen id_2="cole" /*In the case of GBC data bases id_2=gbc */
```

```
#Generate the variables por ITS, "program start" (d), "program impact"(p), and "earthquake".
```

```
gen d=0
```

```
replace d=1 if year>2006
```

```
gen p=0
```

```
replace p=[_n-1]-4 if year>2006
```

```
gen earthquake =0
```

```
replace earthquake =1 if year>2009
```

```
save base_1.dta
```

```
//BY GENDER
```

```
**Sum the population data base by year age gender
```

```
clear all
```

```
use population_2002_2019_reg2
```

```
collapse (sum) population, by(year age gender)
```

```
**The reference population from 2002 will be considered for calculate the standardized
```

```
cholecystectomy rate
```

```
keep if year==2002
```

```
rename population pob_ref
```

```
tab year
```

```
save year_ref_2002_year_gender.dta,replace
```

```
**Calculate the total reference population of 2002
```

```
gen id=1
```

```
clear all
```

```
collapse(sum) pob_ref, by(gender)
```

```
rename pob_ref pob_total
```

```
save year_ref_2002_total_year_gender.dta,replace
```

```
//CALCULATE THE STANDARIZED COLECISTECTOMY RATE
```

```
**Collapse the data base for the all country by age
```

```
clear all
```

```
use cole_its
```

```
collapse (sum) cases, by(year gender age)
```

```
save cole_2002_2018_gender,replace
```

```
**Merge the all population from 2002 to 2018
```

```
merge 1:1 year age sexo using population_2002-2018_year_sexo
```

```
drop _merge
```

```
drop if cases==.
```

```
drop if population==.
```

```
**Merge the reference population of 2002
```

```
merge m:1 age sexo using year_ref_2002_year_sexo.dta
```

```
drop _merge
```

```
drop if cases==.
```

****Merge the total population from 2002**

```
merge m:m sexo using year_ref_2002_total_year_sexo.dta
drop _merge
drop if cases==.
```

****Calculate the age adjusted mortality rate**

***Calculate the crude rate**

```
gen crude_rate=prop*100000
```

***Multiply the crude rate with the reference population**

```
gen crude_rate_pref=( cases/population) *pob_ref
```

***Sum of the quotients between the crude rate and the reference population per year**

```
bysort year gender: egen total_rate=total(crude_rate_ref)
```

*** Divide the total rate by the total reference population to obtain the adjusted rate**

```
gen rate_adj=total_rate/pob_total
```

***Keep the maximum value of rate adjusted by year and gender**

```
collapse(max) rate_est, by(year gender)
save cole_rate_est_2002-2018
```

***Reshape the data base in wide format**

```
reshape wide rate_est, i(year gender)
```

***Repeat the rate adjusted in other 1 variables , just to work in a loop in Rstudio**

```
gen rate_adj3= rate_adj1 /*Men*/
gen id=gender
rename (rate_adj0) (rate_adj2) /*Women*/
gen id_2="cole" /*In the case of GBC data bases id_2=gbc */
```

#Generate the variables por ITS, "program start" (d), "program impact"(p), and "earthquake".

```
gen d=0
replace d=1 if year>2006

gen p=0
replace p=[_n-1]-4 if year>2006

gen earthquake =0
replace earthquake =1 if year>2009
```

```
save base_2.dta
```

//BY ANCESTRY

****Recode the region variable**

```
clear all
use population_2002_2035_reg2.dta
gen ancestry=3
```

```
replace ancestry=2 if reg2>0 & reg2<4
replace ancestry=1 if reg2>7 & reg2<11
```

```
label define ancestry_label 1"Mapuche" 2"Aymara" 3"Others"
label values ancestry ancestry_label
```

```
tab reg2 ancestry
tab ancestry
```

****Obtain the reference population**

```
keep if year==2002
rename population pob_ref
tab year
tab ancestry
save year_ref_2002_year_ancestry.dta,replace
```

****Obtain the total reference population**

```
clear all
use year_ref_2002_year_ancestry.dta
collapse(sum) pob_ref, by(ancestry)
tab ancestry
```

```
rename pob_ref pob_total
save year_ref_2002_total_year_ancestry.dta,replace
```

```
//CALCULATE THE STANDARIZED COLECISTECTOMY RATE
```

```
**Collapse the data base for the all country by age
```

```
clear all  
use cole_its  
collapse (sum) cases, by(year ancestry age)  
save cole_ancestry_2002_2018,replace
```

```
**Merge the all population from 2002 to 2018
```

```
merge 1:m year ancestry age using population_2002-2018_year_ancestry  
drop _merge  
drop if cases==.  
drop if population==.
```

```
**Merge the reference population of 2002
```

```
merge m:1 ancestry age using year_ref_2002_year_ancestry.dta  
drop _merge  
drop if cases==.
```

```
**Merge the total population from 2002
```

```
merge m:m ancestry using year_ref_2002_total_year_ancestry.dta  
drop _merge  
drop if cases==.
```

```
save cole_ancestry_2002_2018,replace
```

```
**Calculate the age adjusted mortality rate
```

```
gen rate_b=( cases/population) *100000
```

```
*Multiply the crude rate with the reference population
```

```
gen rate_b_pref=rate_b*pob_ref
```

```
*Sum of the quotients between the crude rate and the reference population per year
```

```
bysort year ancestry: egen total_rate=total(rate_b_pref)
```

```
* Divide the total rate by the total reference population to obtain the adjusted rate
```

```
gen rate_adj=total_rate/pob_total  
save cole_est_2002-2018_year_ancestry, replace
```

```
*Reshape the data base in wide format
```

```
reshape wide rate_adj, i(year) j(ancestry)
```

```
/* rate_adj1=Mapuche or South*/  
/* rate_adj2=Aymara or North*/  
/* rate_adj3=Others or Center*/
```

```
gen id=ancestry  
gen id_2="cole" /*In the case of GBC data bases id_2=gbc */
```

```
#Generate the variables por ITS, "program start" (d), "program impact"(p), and "earthquake".
```

```
gen d=0  
replace d=1 if year>2006
```

```
gen p=0  
replace p=[_n-1]-4 if year>2006
```

```
gen earthquake =0  
replace earthquake =1 if year>2009
```

```
save base_3.dta
```

```
//BY AGE GES CATHEGORY
```

```
**Gen age GES cathegory
```

```
clear all  
use population_2002_2035_reg2.dta
```

```
recode age 0/34=1 35/49=2 50/100=3 , gene(age_ges)  
label define agec 1"<35" 2"35-49" 3">49"  
label values age_ges agec
```

```
**Keep the year of interest
```

```
keep if year>2001 & year<2019
```

```
**Collapase the population data base by sum
```

```
collapse (sum) population, by( year gender age_ges)
```

```
save pob_2002_2018_year_ges,replace
```

```
*Keep the reference population
```

```
tab year  
keep if year==2002  
rename population pob_ref
```



```
save pob_ref_2002_year_ges.dta, replace
```

```
*Keep the total reference population
```

```
clear all  
keep if year==2002  
collapse (sum) population, by( year age_ges)  
use population_2002_2019_reg2  
rename population pob_total  
gen id=1  
drop year  
save pob_total_2002_ges,replace
```

```
//CALCULATE THE STANDARIZED COLECISTECTOMY RATE
```

```
**Collapse the data base for the all country by age
```

```
clear all  
use cole_its
```

```
**Recode GES ages
```

```
recode age 0/34=1 35/49=2 50/100=3 , gene(age_ges)  
label define agec 1"<35" 2"35-49" 3">49"  
label values age_ges agec  
set more off  
tab age age_ges
```

```
**Collapse the data base
```

```
collapse (sum) cases, by (year gender age_ges)  
save cole_2002_2018_gender,replace
```

```
**Merge the all population from 2002 to 2018
```

```
merge 1:1 age year using pob_2002_2018_year_ges  
drop _merge  
drop if cases==.  
drop if population==.
```

```
**Merge the reference population of 2002
```

```
merge m:1 age using pob_ref_2002_year_ges  
drop _merge  
drop if cases==.  
drop if population==.
```

```
**Merge the total population of 2002
```

```
gen id=1  
merge m:m id using pob_total_2002_age  
drop id _merge
```

```
**Calculate the age adjusted mortality rate
```

```
*Calculate the crude rate
```

```
gen crude_rate=(cases/population)*100000
```

```
*Multiply the crude rate with the reference population
```

```
gen crude_rate_pref=crude_rate*pob_ref
```

```
*Sum of the quotients between the crude rate and the reference population per year
```

```
bysort year age_ges: egen total_rate=total(crude_rate_pref)
```

```
* Divide the total rate by the total reference population to obtain the adjusted rate
```

```
gen rate_adj=total_rate/pob_total
```

```
*Keep the maximun value of rate adjusted
```

```
collapse(max) rate_adj, by(year age_ges)  
save cole_est_2002-2018_year_age_ges, replace
```

```
*Reshape the data base in wide format
```

```
reshape wide rate_adj, i(year) j(age_ges)  
gen id=age_ges  
gen id_2="cole" /*In the case of GBC data bases id_2=gbc */
```

```
/* rate_adj1= <35*/  
/* rate_adj2= 35-49*/  
/* rate_adj2= >49*/
```

```
#Generate the variables por ITS, "program start" (d), "program impact"(p), and "earthquake".
```

```
gen d=0  
replace d=1 if year>2006
```

```
gen p=0  
replace p=[_n-1]-4 if year>2006
```

```
gen earthquake =0
```

```
replace earthquake =1 if year>2009
```

```
save base_4.dta
```

```
**Repeat the same procedure to create the GBC age adjusted mortality rate for ITS analysis, from gbc_its
```

```
-----  
##RSTUDIO
```

```
#Calculate the ITS
```

```
#Call the data .dta bases (base_1-base_8) into a list including cholecystectomy's  
and GBC data bases created in STATA.
```

```
setwd(".")  
temp = list.files(pattern="*.dta")  
myfiles = lapply(temp, read_dta)
```

```
#Check for the Autocorrelation Durbin Watson test (acceptable between 1.5 y 2.5)
```

```
mfun4 <- function(data) durbinwatsonTest (lm(rate_adj1~year+d+p+ earthquake,data=data))  
mfun5 <- function(data) durbinwatsonTest (lm(rate_adj2~year+d+p+ earthquake,data=data))  
mfun6 <- function(data) durbinwatsonTest (lm(rate_adj3~year+d+p+ earthquake,data=data))
```

```
#Run the autocorrelation test
```

```
##For adjusted rate 1  
results4<-list()  
results4<-lapply(myfiles, FUN=mfun4)
```

```
##For adjusted rate 2  
results4<-list()  
results4<-lapply(myfiles, FUN=mfun4)
```

```
##For adjusted rate 3  
results4<-list()  
results4<-lapply(myfiles, FUN=mfun4)
```

```
## Autocorrelation was detected in Mapuche (South), Aymara (North), age <35 and age 35-49 for  
cholecystectomies adjusted rate Prais winsten regression was applied for obtained regression  
coeficientes ( year, d and p) and 95IC%
```

```
adj_rate<- do.call("rbind", myfiles)  
cole_reg<- adj_rate[adj_rate$id=="ancestry" & adj_rate$id_2=="cole",]  
cole_age<- adj_rate[adj_rate$id=="age" & adj_rate$id_2=="cole",]
```

```
summary(prais_winsten(rate_adj1~year+d+p+ earthquake, data = cole_reg, index = "year"))#Mapuche  
summary(prais_winsten(rate_adj2~year+d+p+ earthquake, data = cole_reg, index = "year"))#Aymara  
summary(prais_winsten(rate_adj1~year+d+p+ earthquake, data = cole_age, index = "year"))#age <35  
summary(prais_winsten(rate_adj1~year+d+p+ earthquake, data = cole_age, index = "year"))#age 35-49
```

```
**For the 95%IC run the Prais(command prais) regression in STATA.  
STATA EXAMPLE FOR MAPUCHE COLESCHYSTECTOMY ADJUSTED RATE
```

```
clear all  
use base_3  
  
tsset year  
prais rate_adj1 year d p earthquake
```

```
#Create a function to calculate the effect of GES for each adjusted rate
```

```
mfun1 <- function(data) lm(rate_adj1~year+d+p+ earthquake,data=data)  
mfun2 <- function(data) lm(rate_adj2~year+d+p+ earthquake,data=data)  
mfun3 <- function(data) lm(rate_adj3~year+d+p+ earthquake,data=data)  
fun<-c(mfun_1,mfun_2,mfun_3)  
  
coef_1<-function(data) coef(lm(rate_adj1~year+d+p+ earthquake,data=data))  
coef_2<-function(data) coef(lm(rate_adj2~year+d+p+ earthquake,data=data))  
coef_3<-function(data) coef(lm(rate_adj3~year+d+p+ earthquake,data=data))  
coef_4<-c(coef_1,coef_2,coef_3)
```

```
results<-list()  
coef<-list()  
p<-list()  
ci<-list()  
r<-list()  
ci<-list()  
ci_wide<-list()  
year_2006_results<-list()  
  
for(i in 1:3){
```

```

results[[i]]<-lapply(myfiles, FUN=fun[[i]])
coef[[i]]<-list()
coef[[i]] <- ldply(myfiles, coef_4[[i]])
p[[i]] <- map(results[[i]], ~ summary(.x)$coefficients[,4])
ci[[i]]<-map(results[[i]], ~ confint(.x))
p[[i]]<-do.call(rbind,p[[i]])
ci[[i]]<-do.call(rbind,ci[[i]])
r[[i]] <- map(results[[i]], ~ summary(.x)$r.squared)
r[[i]]<-do.call(rbind,r[[i]])
p[[i]]<-as.data.frame(p[[i]])

##Rename p-values
str(p[[i]])
p[[i]]<-p[[i]]%>%
  rename(pv_intercept="(Intercept)")
p[[i]]<-p[[i]]%>%
  rename(pv_d=d)
p[[i]]<-p[[i]]%>%
  rename(pv_p=p)
p[[i]]<-p[[i]]%>%
  rename(pv_year=year)
p[[i]]<-p[[i]]%>%
  rename(pv_earthquake = earthquake)
str(p[[i]])

##Rename
r[[i]]<-as.data.frame(r[[i]])
str(r[[i]])

r[[i]]<-r[[i]]%>%
  rename(r=v1)

##Calculate 95%IC
ci[[i]]<-as.data.frame(ci[[i]])
ci[[i]]$id<-rep(c(1:8),each=5)
ci[[i]]$coef<-rep(c("intercept","year","d","p","earthquake"),times=4)

str(ci[[i]])
ci[[i]]<-ci[[i]]%>%
  rename(ic_1="2.5 %")

ci[[i]]<-ci[[i]]%>%
  rename(ic_2="97.5 %")

ci_wide[[i]]<-reshape(as.data.frame(ci[[i]]),
  timevar = "coef",
  idvar = c("id","id_2"),
  v.names=c("ic_1", "ic_2"),
  direction = "wide")

year_2006_results[[i]]<-cbind(coef[[i]],p[[i]],ci_wide[[i]],r[[i]])
year_2006_results[[i]]$id_rate<-id[[i]]
}

```

#Call the list to a data frame

```

year_2006_results_2 <- do.call("rbind", year_2006_results)
colnames(year_2006_results_2)

```

```

year_2006_out_resume<-year_2006_tasa_outA[, c(23,11,22,2,14,15,7,3,16,17,8,4,18,19,9,5,20,21,10)]

```

```

year_2006_out_resume <-year_2006_out_resume[order(year_2006_out_resume$id), ]

```

```

year_2006_out_resume <-filter(year_2006_out_resume ,id!=1|id_rate!=2)
year_2006_out_resume <-filter(year_2006_out_resume ,id!=1|id_rate!=3)
year_2006_out_resume <-filter(year_2006_out_resume ,id!=2|id_rate!=3)
year_2006_out_resume <-filter(year_2006_out_resume ,id!=5|id_rate!=2)
year_2006_out_resume <-filter(year_2006_out_resume ,id!=5|id_rate!=3)
year_2006_out_resume <-filter(year_2006_out_resume ,id!=6|id_rate!=3)

```

****Also eliminate the result of the regression with autocorrelation**

#PREDICT AGE ADJUSTED CHOLECYSTECTOMY AND GBC RATE FOR SENCE 2019 TO 2022 ESTRATIFATONING FOR THE VARIABLES

#For ancestry (regions)

#CHOLECYSTECTOMY

```

str(cole_reg)

```

****In the case of autocorrelation use STATA to predict the values**

STATA EXAMPLE FOR MAPUCHE COLESCHYSTECTOMY ADJUSTED RATE

```

clear all
use base_3

```

```

tsset year
prais rate_adj1 year d p earthquake
predict ptasa_1 if e(sample)
summarize if e(sample)
margins,atmeans at(year=2019 d=1 p=14 earthquake=0)
margins,atmeans at(year=2020 d=1 p=15 earthquake=0)
margins,atmeans at(year=2021 d=1 p=16 earthquake=0)
margins,atmeans at(year=2022 d=1 p=17 earthquake=0)

```

#Predicted values in R

```

cole_o<-lm(rate_adj3~year+d+p+earthquake,data=cole_reg)
str(cole_o)
predict(cole_o, newdata =data.frame(year= 2019, d=1, p=14, earthquake=0),interval = "confidence")
predict(cole_o, newdata =data.frame(year= 2020, d=1, p=15, earthquake=0),interval = "confidence")
predict(cole_o, newdata =data.frame(year= 2021, d=1, p=16, earthquake=0),interval = "confidence")
predict(cole_o, newdata =data.frame(year= 2022, d=1, p=17, earthquake=0),interval = "confidence")

```

#GBC MORTALITY

```

cole_reg<- adj_rate[adj_rate$id=="ancestry" & adj_rate$id_2=="gbc",]

```

```

gbc_aymara<-lm(rate_adj2~year+d+p+earthquake,data=gbc_reg)
predict(gbc_aymara, newdata =data.frame(year= 2019, d=1, p=14, earthquake=0),interval = "confidence")
predict(gbc_aymara, newdata =data.frame(year= 2020, d=1, p=15, earthquake=0),interval = "confidence")
predict(gbc_aymara, newdata =data.frame(year= 2021, d=1, p=16, earthquake=0),interval = "confidence")
predict(gbc_aymara, newdata =data.frame(year= 2022, d=1, p=17, earthquake=0),interval = "confidence")

```

```

gbc_mapuche<-lm(rate_adj1~year+d+p+earthquake,data=gbc_reg)
predict(gbc_mapuche<-, newdata =data.frame(year= 2019, d=1, p=14, earthquake=0),interval = "confidence")
predict(gbc_mapuche<-, newdata =data.frame(year= 2020, d=1, p=15, earthquake=0),interval = "confidence")
predict(gbc_mapuche<-, newdata =data.frame(year= 2021, d=1, p=16, earthquake=0),interval = "confidence")
predict(gbc_mapuche<-, newdata =data.frame(year= 2022, d=1, p=17, earthquake=0),interval = "confidence")

```

```

gbc_other<-lm(rate_adj3~year+d+p+earthquake,data=gbc_reg)
predict(gbc_other, newdata =data.frame(year= 2019, d=1, p=14, earthquake=0),interval = "confidence")
predict(gbc_other, newdata =data.frame(year= 2020, d=1, p=15, earthquake=0),interval = "confidence")
predict(gbc_other, newdata =data.frame(year= 2021, d=1, p=16, earthquake=0),interval = "confidence")
predict(gbc_other, newdata =data.frame(year= 2022, d=1, p=17, earthquake=0),interval = "confidence")

```

##Repeat the code for predict the cholecystectomy and GBC adjusted rate for all country, gender, age.