Test 2

M. Schonhoeft

6/26/2020

# Test 2

Repo can be found here. <https://github.com/mitchelly00/test_2.git>.

#### Question 1 cleaning the global enviroment

rm(list = ls())

#### Question 2 loading the inequality package

library(rio)  
inequality\_data<- import("inequality.xlsx", which=1)

#### Question 3 This is a Cross-sectional dataset

* It is cross sectional because as you can see in the summary that the only year is 2015

summary(inequality\_data)

## iso2c country inequality\_gini year   
## Length:203 Length:203 Min. :25.40 Min. :2015   
## Class :character Class :character 1st Qu.:31.55 1st Qu.:2015   
## Mode :character Mode :character Median :35.75 Median :2015   
## Mean :36.81 Mean :2015   
## 3rd Qu.:41.12 3rd Qu.:2015   
## Max. :59.10 Max. :2015   
## NA's :123

#### Question 4 finding the gini using subset

subset(inequality\_data,country=="Denmark")

## iso2c country inequality\_gini year  
## 40 DK Denmark 28.2 2015

subset(inequality\_data,country=="Sweden")

## iso2c country inequality\_gini year  
## 174 SE Sweden 29.2 2015

#### Question 5 finding the gini of Brazil

subset(inequality\_data,country=="Brazil")

## iso2c country inequality\_gini year  
## 13 BR Brazil 51.9 2015

#### Question 6

It is better to have a **low** inquality\_gini score as seen by the lower Scandinavian countries.

#### Question 7 Quick peak

head(inequality\_data)

## iso2c country inequality\_gini year  
## 1 AL Albania 32.9 2015  
## 2 AM Armenia 32.4 2015  
## 3 AT Austria 30.5 2015  
## 4 BY Belarús 25.6 2015  
## 5 BE Belgium 27.7 2015  
## 6 BZ Belize NA 2015

#### Question 8 taking of the accent

accent.remove<- function(s){  
 #1character subsitutions  
 old1<- "ú"  
 new1<- "u"  
 s1 <- chartr(old1,new1,s)  
 }  
  
inequality\_data$country <- accent.remove(inequality\_data$country)

Quick peak to make sure it is gone

head(inequality\_data)

## iso2c country inequality\_gini year  
## 1 AL Albania 32.9 2015  
## 2 AM Armenia 32.4 2015  
## 3 AT Austria 30.5 2015  
## 4 BY Belarus 25.6 2015  
## 5 BE Belgium 27.7 2015  
## 6 BZ Belize NA 2015

#### Question 9 Sort the data to see which are the lowest scores

The head comand allows us to see what the top 5 countries are

inequality\_data = inequality\_data[order(inequality\_data$inequality\_gini),]  
head(inequality\_data)

## iso2c country inequality\_gini year  
## 161 SI Slovenia 25.4 2015  
## 190 UA Ukraine 25.5 2015  
## 4 BY Belarus 25.6 2015  
## 39 CZ Czech Republic 25.9 2015  
## 92 XK Kosovo 26.5 2015  
## 160 SK Slovak Republic 26.5 2015

#### Question 10 the mean inequaility\_gini score

na.rm is used to take out missing values

mean(inequality\_data$inequality\_gini,na.rm = TRUE)

## [1] 36.81375

#### Question 11 using ifelse command

inequality\_data$low\_high = NA  
inequality\_data$low\_high[inequality\_data$inequality\_gini >= 36.81]=0  
inequality\_data$low\_high[inequality\_data$inequality\_gini< 36.81]=1  
  
ifelse(test = inequality\_data$low\_high==1,  
 yes = "low\_inequality",no = "high\_inequality")

## [1] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [5] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [9] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [13] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [17] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [21] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [25] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [29] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [33] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [37] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [41] "low\_inequality" "low\_inequality" "low\_inequality" "low\_inequality"   
## [45] "low\_inequality" "low\_inequality" "high\_inequality" "high\_inequality"  
## [49] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [53] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [57] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [61] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [65] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [69] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [73] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [77] "high\_inequality" "high\_inequality" "high\_inequality" "high\_inequality"  
## [81] NA NA NA NA   
## [85] NA NA NA NA   
## [89] NA NA NA NA   
## [93] NA NA NA NA   
## [97] NA NA NA NA   
## [101] NA NA NA NA   
## [105] NA NA NA NA   
## [109] NA NA NA NA   
## [113] NA NA NA NA   
## [117] NA NA NA NA   
## [121] NA NA NA NA   
## [125] NA NA NA NA   
## [129] NA NA NA NA   
## [133] NA NA NA NA   
## [137] NA NA NA NA   
## [141] NA NA NA NA   
## [145] NA NA NA NA   
## [149] NA NA NA NA   
## [153] NA NA NA NA   
## [157] NA NA NA NA   
## [161] NA NA NA NA   
## [165] NA NA NA NA   
## [169] NA NA NA NA   
## [173] NA NA NA NA   
## [177] NA NA NA NA   
## [181] NA NA NA NA   
## [185] NA NA NA NA   
## [189] NA NA NA NA   
## [193] NA NA NA NA   
## [197] NA NA NA NA   
## [201] NA NA NA

#### Question 12

#### Question 13

names<-c("The World Bank", "African Development Bank","the Bill and Melinda Gates Foundation")

#### Question 14

I think there is a correlation between inflation and inequaility. If there is higher inflation then the government is doing a poor job of regulating the economy which would cause inequality to increase.

#### Question 15 inserting the data

library(WDI)  
inflation\_ds =WDI(country = "all",indicator = "NY.GDP.DEFL.KD.ZG",  
 # indicator from web  
 start = 2015, end = 2015,   
 extra = FALSE, cache = NULL)

#### Question 16 the new name of the variable

library(data.table)  
setnames(inflation\_ds,"NY.GDP.DEFL.KD.ZG","inflation")

#### Question 17 merging the data frames

library(tidyverse)

## -- Attaching packages ----------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.1 v purrr 0.3.4  
## v tibble 3.0.1 v dplyr 1.0.0  
## v tidyr 1.1.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts -------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::between() masks data.table::between()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::first() masks data.table::first()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::last() masks data.table::last()  
## x purrr::transpose() masks data.table::transpose()

inflation\_ds$year=as.numeric(inflation\_ds$year)  
  
merged\_df = left\_join(x=inequality\_data,y=inflation\_ds,  
 by=c("iso2c","year"))

take away the country.x

merged\_df<-  
 merged\_df %>%   
 select(-c("country.x")) %>%   
 dplyr::rename("country"="country.y")

#### Question 18 removing the NAs

merged\_df=na.omit(merged\_df, select=c("inequality\_gini","inflation"))  
summary(merged\_df)

## iso2c inequality\_gini year low\_high   
## Length:80 Min. :25.40 Min. :2015 Min. :0.000   
## Class :character 1st Qu.:31.55 1st Qu.:2015 1st Qu.:0.000   
## Mode :character Median :35.75 Median :2015 Median :1.000   
## Mean :36.81 Mean :2015 Mean :0.575   
## 3rd Qu.:41.12 3rd Qu.:2015 3rd Qu.:1.000   
## Max. :59.10 Max. :2015 Max. :1.000   
## country inflation   
## Length:80 Min. :-4.6206   
## Class :character 1st Qu.: 0.7577   
## Mode :character Median : 1.8488   
## Mean : 3.1996   
## 3rd Qu.: 3.8277   
## Max. :38.8817

#### Question 19 only keeping data greater than 30

data\_greater\_30 <-  
 merged\_df%>%  
 dplyr::filter(!(inequality\_gini > 30))

#### Question 20 using data\_greater\_30 to see which countries have ai in their name

grep("ai", data\_greater\_30$country)

## [1] 2

#### Question 21

sapply(data\_greater\_30$inequality\_gini, sum)

## [1] 25.4 25.5 25.6 25.9 26.5 26.5 26.8 26.8 27.0 27.1 27.5 27.7 28.2 28.2 29.0  
## [16] 29.2 29.4

#### question 22

library(labelled)

## Warning: package 'labelled' was built under R version 4.0.2

var\_label(data\_greater\_30)<-list(`country`="country" ,  
 `year` = "year" ,  
 `inequality\_gini` =  
 "gini inequaility index" ,  
 `iso2c` = "ISO-2 country code" ,  
 `inflation` = "percent inflation of every country" ,  
 `country` = "country")

#### Question 23

library(rio)  
export(data\_greater\_30,"clean\_dataset.dta")