Exam 2 PDF

RLong

6/26/2020

clear the environment

rm(list=ls(all=TRUE))

new project, select working directory

load the inequality dataset and save the data frame

library(rio)  
inequality\_data = import("inequality.xlsx")

#this is a cross sectional dataset because it is analyzes data from one year, 2015.

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

summary(inequality\_data) #see the year column in the console

## 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

#subset command to provide the gini scores for denmark and sweden

subset(inequality\_data$inequality\_gini, "Denmark") subset(inequality\_data$inequality\_gini, “Sweden”)

#6 it is better to have low inequality\_gini scores, as seen with Denmark and Sweden.

#Taking a quick peak at the data frame

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

#writing a function called accent.remove to remove the accent on Belarus

inequality\_data$country[inequality\_data$country=="Belarús"] = "Belarus"

#head command again to show that I removed the accent.

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

#sort the data by lowest inequality\_gini scores

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

#mean of the inequality\_gini score

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

## [1] 36.81375

#creating dummy variables

low\_inequality <- ifelse(inequality\_data$inequality\_gini < 36.81, 1, 0)  
high\_inequality <- ifelse(inequality\_data$inequality\_gini > 36.81, 1, 0)

#run a cross tab using the dummy variables

library(doBy)  
summaryBy(low\_inequality ~ high\_inequality, data=inequality\_data,   
 FUN = c(mean, length))

## low\_inequality.mean low\_inequality.length  
## 1 NA 203

#write a for loop that prints three actors

people\_reduce <- c('The World Bank', 'African Development Bank', 'Bill and Melinda Gates Foundation')  
for ( i in people\_reduce){  
 print(i)  
}

## [1] "The World Bank"  
## [1] "African Development Bank"  
## [1] "Bill and Melinda Gates Foundation"

#the variable that is correlated with inequality is poverty. Multidimensional poverty headcount. I picked it because it has the Gini index and says Poverty and equity data

#import the variable into R

library(WDI)  
multi\_poverty=WDI(country = "all",  
 indicator = c("SI.POV.MDIM"), # indicator from web  
 start = 2015, end = 2015, extra = FALSE, cache = NULL)

#merge the new variable into other dataset

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::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::order\_by() masks doBy::order\_by()

merged\_df = left\_join(x=inequality\_data,  
 y=multi\_poverty,  
 by=c("country", "year"))

#remove the missing data on the basis of inequality\_gini

merged\_df<- na.omit(merged\_df, select=c("inequality\_gini", "multi\_poverty"))

#keeping the data with inequality\_gini scores greater than 30

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

#count how many countries have the sequence “ai” in their name

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

## integer(0)

#apply family to take the sum of inequality\_gini

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

#label your variables

library(labelled)

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

var\_label(merged\_df) <- list(`country` = "Country",  
 `year` = "year",  
 `inequality\_gini` = "Gini Index for Inequality",  
 `SI.POV.MDIM` = "MultiDimensional Poverty Measure")

#save the labelled data frame as a Stata

library(rio)

export(merged\_df, file = “final\_data.dta”)

[link to the git hub repo](https://github.com/RolongAlong/exam2.git)

ssh key [git@github.com](mailto:git@github.com):RolongAlong/exam2.git