# Exam 2 Answers

JP Reppeto

6/26/2020

```
knitr::opts_chunk$set(echo = TRUE)
```

**Extremely crunched for time so please pardon my numbering on here 1-2:**

```
rm(list=ls(all=TRUE))

library(rio)
inequality_data <- import("inequality.xlsx")
```

**3:** As this dataset takes information from a specific moment in time (inequality ratings in 2015), we consider it cross-sectional. As the head shows, these are specific observations from a specific moment.

```
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
```

**4-5:**

```
#subsetting Denmark, Sweden, and Brazil
subset(inequality_data, country == "Sweden")
```

```
##     iso2c country inequality_gini year
## 174    SE  Sweden            29.2 2015
```

```
subset(inequality_data, country == "Denmark")
```

```
##    iso2c country inequality_gini year
## 40    DK Denmark            28.2 2015
```

```r
subset(inequality_data, country == "Brazil")
```

```
##    iso2c country inequality_gini year
## 13    BR  Brazil            51.9 2015
```

**6:** It is better to have *lower* gini scores.

**7-9:**

```r
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
```

```r
#remove the accent function
remove.accents <- function(s) {
  #single letter
  old1 <- "ú"
  new1 <- "u"
  s1 <- chartr(old1, new1,s)
}

inequality_data$country <- remove.accents(inequality_data$country)

#checking head
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
```

```r
#nice

#sorting by low ineq_gni
lowineq <- inequality_data[order(inequality_data$inequality_gini),]

#head of new data
head(lowineq)
```

```
##     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
```

**10-12:**

```
#mean function and mean gini
mean_function <- function(x)
{sum(x) / length(x)}

mean_function(inequality_data$inequality_gini)
```

```
## [1] NA
```

```
#ifelse dummy variables
inequality_data$high_inequality <- ifelse(inequality_data$inequality_gini >= 36.81, 1, 0)
inequality_data$low_inequality <- ifelse(inequality_data$inequality_gini < 36.81, 1, 0)

#crosstabbing our variables
library(doBy)

summaryBy(high_inequality ~ low_inequality,
          data = inequality_data, FUN= c(mean, length))
```

```
##   low_inequality high_inequality.mean high_inequality.length
## 1              0                    1                     34
## 2              1                    0                     46
## 3             NA                   NA                    123
```

**14-17:**

```
#importing wdi data
remotes::install_github('vincentarelbundock/WDI')
```

```
## Skipping install of 'WDI' from a github remote, the SHA1 (5b516c96) has not changed since last instal
##   Use 'force = TRUE' to force installation
```

```
library(WDI)

adult_literacy = WDI(country = "all",
                     indicator = c("SE.ADT.LITR.ZS"),
                     start = 2015, end = 2015, extra = FALSE, cache = NULL)

#renaming

library(data.table)

setnames(adult_literacy, "SE.ADT.LITR.ZS", "Literacy rate, adult total")

#merging the variable to ineq 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::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 dplyr::order_by()  masks doBy::order_by()
## x purrr::transpose() masks data.table::transpose()
```

```
merged_df = left_join(inequality_data, adult_literacy)
```

```
## Joining, by = c("iso2c", "country", "year")
```

```
by= c("iso2c", "country", "year")
```