

Assignment 3

Tori Dykes and Mansi Wadhwa

Saturday, November 05, 2016

Loading packages

```
library(stargazer) # for summary statistics and regression tables
library(magrittr)  # for 'piping': more readable code
library(ggplot2)   # the ggplot2 package provides nice function for plotting
library(arm)       # for the sim() function to simulate model estimates
library(interplot) # for plotting interactions
library(dplyr)     # for data manipulation
```

Data from Eurostats Labour Force adjusted series

Loading the data

```
d <- read.csv('ESS1-6e01_1_F1.csv') # ESS
e <- read.csv('Under 25 unemp.csv') # Eurostat Data
```

Since we want to focus on the youth, we need to restrict both datasets to observations from the relevant age group only. Summarising age variables for both datasets and subsetting the dataset accordingly

```
df <- subset(d, d$agea <= 25 & d$agea >= 18)
eu <- subset(e, e$AGE == "Less than 25 years")

summary(df$agea)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  18.00   20.00   22.00   21.53   24.00   25.00
```

```
summary(eu$AGE)
```

```
## Less than 25 years      Total
##              1170              0
```

Checking number of observations for each country in each round of the survey. This counts the number of entries for DE in round 1 and round 2 of the survey.

```
sum(with(df, cntry == "DE" & essround == 1))
```

```
## [1] 296
```

```
sum(with(df, cntry == "DE" & essround == 2))
```

```
## [1] 304
```

```
# Number of observations for Germany in all rounds
sum(with(df, cntry == "DE"))
```

```
## [1] 1818
```

Printing the number of observations for each of the 6 ESS rounds

```
for(i in 1:6){
x <- sum(with(df, essround == i))
print(x)
}
```

```
## [1] 4782
## [1] 5712
## [1] 4724
## [1] 5802
## [1] 5796
## [1] 5505
```

[Link to metadata for EU data: http://ec.europa.eu/eurostat/cache/metadata/en/une_esms.htm “The unemployment rate is the number of people unemployed as a percentage of the labour force. The labour force is the total number of people employed and unemployed. In the database, unemployment rates can be downloaded by choosing the unit”PC_ACT“, Percentage of Active Population.”]

In order to merge the datasets, the EU data needs to be modified further. Firstly, it only needs to contain one measure of the youth unemployment, namely youth unemployed as a percentage of active population. It currently contains three different measures.

Secondly, we need only the data for years for which we have the corresponding ESS survey values. The Eurostat data contains observations from 2006 to 2015 while the ESS rounds have data for 7 rounds: 2002, 2004, 2006, 2008, 2010, 2012, 2014.

```
eu <- subset(eu, eu$UNIT == "Percentage of active population")
eu <- subset(eu, eu$TIME == 2006 | eu$TIME == 2008 | eu$TIME == 2010 | eu$TIME == 2012 | eu$TIME == 2014)
```

Thirdly, the two datasets contain different countries (as shown below). While data from Eurostats contains 39 geographical levels, the ESS data contains 32 levels/countries. We need one common variable to be able to merge these two datasets together.

```
unique(df$cntry)
```

```
## [1] AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IL IS IT LT LU NL
## [24] NO PL PT RU SE SI SK TR UA
## 32 Levels: AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IL IS ... UA
```

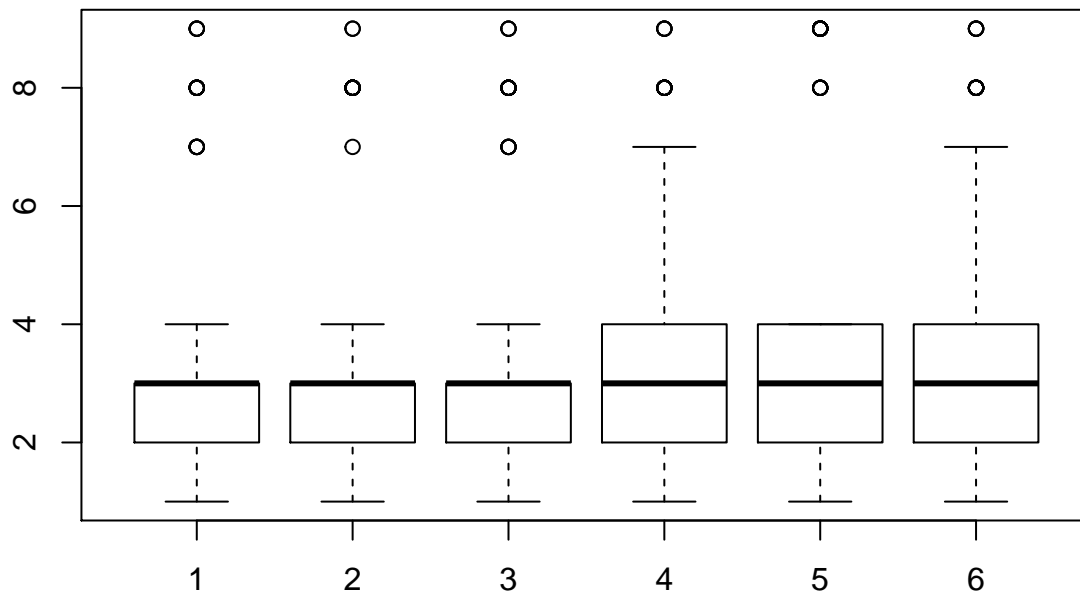
```
unique(eu$GEO)
```

```
## [1] European Union (28 countries)
## [2] European Union (27 countries)
## [3] European Union (25 countries)
## [4] Euro area (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17-2013, EA18-2014, EA19)
```

```
## [5] Euro area (19 countries)
## [6] Euro area (18 countries)
## [7] Belgium
## [8] Bulgaria
## [9] Czech Republic
## [10] Denmark
## [11] Germany (until 1990 former territory of the FRG)
## [12] Estonia
## [13] Ireland
## [14] Greece
## [15] Spain
## [16] France
## [17] Croatia
## [18] Italy
## [19] Cyprus
## [20] Latvia
## [21] Lithuania
## [22] Luxembourg
## [23] Hungary
## [24] Malta
## [25] Netherlands
## [26] Austria
## [27] Poland
## [28] Portugal
## [29] Romania
## [30] Slovenia
## [31] Slovakia
## [32] Finland
## [33] Sweden
## [34] United Kingdom
## [35] Iceland
## [36] Norway
## [37] Turkey
## [38] United States
## [39] Japan
## 39 Levels: Austria Belgium Bulgaria Croatia Cyprus ... United States
```

Plotting a boxplot to see how political interest varies in each ESS round starting 2002 to 2014

```
boxplot(df$polint ~ df$essround)
```



shows that
not very informative in my opinion since it the polintr var is not continuous

Variables for unemployment: - uempla (Doing last 7 days: unemployed, actively looking for job) - uempli (Doing last 7 days: unemployed, not actively looking for job) - uemp3m(Ever unemployed and seeking work for a period more than three months) - uemp12m (Any period of unemployment and work seeking lasted 12 months or more) - uemp5yr (Any period of unemployment and work seeking within last 5 years)

So,

Boxplot of uemp12m with ESS round shows weird results. `boxplot(dfuemp12m dfessround)`