Lab_3_HW

Ngoc Phan

September 25, 2018

Import .dta files into R

First, we need to import the *Happy Planet Index* dataset (read more about it here). The file is in Stata's *.dta* format. To import such files, install the package **readstata13** and use the command **read.dta13** as shown below.

```
#install.packages("readstata13")
library(readstata13)
happy <- read.dta13("happy_planet.dta")
```

Explore the data (1 point)

1. Present an overview of the data frame **happy**, showing the mean, median, minimum and maximum values of each variable.

Data Cleaning (3 points)

2. Obtain a subset of the data frame with observations from Africa, then assign it to a new object called **africa**.
3. Create a new variable, **newregion**, that takes value of "West", "Middle East", "Africa", "South Asia", "East Asia", "Eastern Europe/USSR", and "Latin America", based on existing variables in the dataset.
4. Drop variables other than **country**, **lifesat010**, **gdppercapitappp**, **newregion** from the data frame.

Bivariate Linear Regression (6 points)

5. Create a scatter plot of **lifesat010** against **gdppercapitapp**, using either *plot* or *ggplot*. Please label the axes appropriately.
6. Regress **lifesat010** on **gdppercapitappp**. Report the results (R output is fine, no need to use *stargazer* yet)
7. Interpret the coefficient on **gdppercapitapp**
8. Generate a new variable in **happy** that takes on the predicted values of **lifesat010** from the regression. Name it **yhat**.
9. Generate a new variable that takes on the value of the residuals from the regression. Name it **uhat**.
10. Plot the residuals against **gdppercapitappp**. Are Gauss-Markov assumptions 4 and 5 satisfied?