Problem Set 3

Your Name

This problem set is due at 8:30 am on 9/26

Please upload both Rmd and PDF files on Sakai

PDF file should contain both codes and outputs

This week we will be working with a dataset(`week3.dta`) that offers an alternative measure of citizen welfare, based on life satisfaction, life expectancy, and ecological footprint. The ranking is called the Happy Planet Index (HPI) and is available on this website (http://happyplanetindex.org/countries). The index is highly provocative, providing an alternative understanding of the goals of economic policy.

Variable defintions:

- `hpirank`: Happy Planet Index
- `country`: Country
- `region`: Region
- `lifeexpyears`: Average Life Expectancy
- `wellb010`: Average Wellbeing (0-10)
- `gdppercapitappp`: GDP per capita ($PPP)
- `gini`: GINI Index
- `popul`: Population

Hypothesis Testing (3 Points)

- Test the null hypothesis that the average life expectancy of the sample is equal to 60 years (H0: mean `lifeexpyears`= 60). (You may use the `t.test` command in this and following questions).
- Is the life expectancy of Europe countries (hint: `region == "Europe"`) significantly greater than the other countries in the sample?
- Is the life expectancy of countries in the Americas significantly greater than the life expectancy of countries in Sub Saharan Africa?(hint: filter data first and then perform t-test)

Visual analysis of scatter plot (2 Points)

- Create a scatter plot depicting the relationship between `wellb010` and `gdppercapitappp`.
- Draw a regression line through that scatter plot using `ggplot()`.
- Eyeballing the line what do you think is its intercept? Its slope?
- Label the observations that represent potential outlier(s) in the plot.
- Drop the outlier(s), and then add another fitted line using the truncated dataset. Does the slope change?

Bivariate Linear Regression (5 points)

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- BY HAND: Perform a linear regression of the above function, where Y is `wellb010` and X is `gdppercapitappp`. Show β_1, β_0, TSS, ESS, SSR, R^2.
- What are the slope and intercept of the equation?
- Interpret the results by filling in the blank to this sentence: "A $1000 increase in GDP per capita leads to a _____ point change in the predicted well-being index."
- Verify your regression results using R's regress command `lm()`
- Export your regression using `stargazer()`
- Add a fitted line using the value of β_1 and β_0 computed by hand to the previous scatterplot. Does this fitted line overlap the fitted line plotted by `ggplot`?
- Can this finding be trusted? Why or why not?