

Pol Sci 630: Problem Set 2

Anh Le

September 9, 2016

Note 1: It is absolutely essential that you show all your work, including intermediary steps, in your (mathematical) calculations and that you comment on your R code to earn full credit (you can comment on your R code both with the use of `#` in the R code and in the \LaTeX code). Showing all steps and commenting on code will also be required in future problem sets.

Note 2: Please submit a PDF file created through knitr containing all your answers to the problem set. knitr allows you to combine R code and \LaTeX code in one document, meaning that you can include both the answers to R programming and math problems. Also submit the source code that generates the PDF file (i.e. the .Rnw file).

Note 3: Make sure that the PDF files you submit do not include any references to your identity. The grading will happen anonymously. You can submit your answer at the following website: <http://ps630-f15.herokuapp.com/>

1 Expected Value and Its Properties

a.

Suppose that one word is to be selected at random from the sentence ‘the girl put on her beautiful red hat’. If X denotes the number of letters in the word that is selected, what is the value of $E(X)$?

b.

Suppose that one letter is to be selected at random from the 30 letters in the sentence given in Exercise 4. If Y denotes the number of letters in the word in which the selected letter appears, what is the value of $E(Y)$?

2 Plotting distribution

For this problem, you'll need to Google some R techniques (e.g. side-by-side / overlapping plot). Also, label the axes and the plots accordingly.

a.

Download a variable you are interested in, using `WDI`. BEFORE YOU PLOT, write down your expectation about the range of values and the distribution of the variable (yes, please, I mean it, it'll be fun). Then plot the histogram, density plot, boxplot, and normal quantile plot. Did your expectation match?

b.

Plot the density plots of that variable for Europe and Asia, 1) side by side (Hint: `par(mfrow=c(?, ?))`), and 2) overlapping in the same plot.

c.

Draw the scatterplot of that variable against another variable that you think are highly correlated. Do they turn out to be correlated? Is there a group of outliers perhaps that screw up the correlation?

d.

Label the point that represents your country (Hint: Tutorial) and color it red (Some Googling involved)