Pol Sci 630: Problem Set 3 - Comparisons and Inference

Prepared by: Jan Vogler (jan.vogler@duke.edu)

Due Date: Tuesday, September 15th, 2015, 10 AM (Beginning of Class)

Note 1: It is absolutely essential that you show all your work, including intermediary steps, and comment on your R code to earn full credit. Showing all steps and commenting on code them will also be required in future problem sets.

Note 2: Please use a *single* PDF file created through knitr to submit your answers. 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. either .Rnw or .Rmd files)

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/`

R Programming

Problem 1

Do the following in R:

a) (2 points) Create a variable X that is a sequence from 1 to 1000 (intervals of 1). Then, create a variable Y that is linearly dependent on X. Now evaluate their covariance and correlation. Interpret the results. Then, create a second variable Y_2 that is linearly dependent on X but additionally has some normally distributed error. Now evaluate their covariance and correlation. Interpret the results. Finally, interpret the difference between the covariances and correlations.

b) (1 point) Write a function that returns the correlation between two vectors in R. For this purpose, do not use the covariance (*cov*) or the correlation (*cor*) functions that are built into R. Instead, please refer to the lecture and text book for the mathematical definitions of covariance and correlation and emulate those calculations in your function. Demonstrate that your function works by plugging in vectors and comparing it to the built-in *cor* function in R.

Bonus problem: (1 point) Copy the function you created for problem B. Now integrate two error messages. The first message should appear if you plug in two vectors of different lengths. The second message should appear if you plug in a non-numeric vector. Demonstrate that your function displays the correct error message by, first, plugging in two vectors of different lengths and, second, plugging in a vector consisting of characters.

Problem 2

Do the following in R:

a) (1 point) Create two vectors of length 50 with random draws from a Poisson distribution. The mean of the draws in the first vector should be 10, the mean of the draws in the second vector should be 12. Conduct a two-sample t-test with the *t.test* function for those two vectors in R.

Hint: If you don't know how to draw from a Poisson distribution, try to find out through a search engine.

b) (2 points) Write a function that allows you to perform a two-tail, two-sample t-test, using vectors as input (essentially a copy of the *t.test* function). In your calculation, assume that the two groups have different variances. The function is meant to return a t-value. Test the function you created by plugging in the two vectors from problem 2a and verify that you get the same result again.

Note: You're allowed to use built-in functions except for the *t.test* function itself.

Bonus problem: (1 point) Include an error message if any vector plugged in is not numeric.

Probability Theory: Covariance and Correlation

Problem 3

Do the following problems. Show every step.

a) (2 points) This problem is taken from Jim Pitman (1993) *Probability*: Let (X, Y) have uniform distribution on the four points $(-1, 0), (0, 1), (0, -1), (1, 0)$. Show that X and Y are uncorrelated. Then prove that they are not independent.

b) (2 points) This problem is taken from Jim Pitman (1993) *Probability*: Let X have uniform distribution on -1,0,1 and let $Y = X^2$. Are X and Y uncorrelated? Are X and Y independent? Show mathematically and explain carefully.

Problem 4

a) (2 points) This problem is taken from Jim Pitman (1993) *Probability*: Let X_1 and X_2 be the numbers on two independent fair six-sided die rolls, $X = X_1 - X_2$ and $Y = X_1 + X_2$. Show that X and Y are uncorrelated, but not independent.

Hint: For problem c, you may calculate the covariance in R or manually.

b) (2 points) Assume that d, e, f, and g are constants. Assume that Y_1 and Y_2 are random variables. Demonstrate mathematically that the following relationship holds:
$Cov(dY_1 + e, fY_2 + g) = df Cov(Y_1, Y_2)$.