

## PUBPOL 812: Statistics and Data Analysis for Policy-Makers

### Problem Set 4 - Due Friday, September 22, 11:55 pm

Turn in both the Word document write-up and your Stata .do file. Make sure your .do file includes a log statement and comments that explain what you are doing.

1. In Lab04, we generated a new composite variable averaging students' scores on the cheerfulness questions. We also looked at the distribution of cheerfulness. Now, I want to know how cheerfulness relates to Harry Potter houses. First, correlate cheerfulness with the different Harry Potter houses (ravenclaw, gryffindor, hufflepuff, & slytherin). These variables are calculations of what proportion of a student's responses on the Harry Potter quiz placed them in the Gryffindor versus Ravenclaw versus Hufflepuff versus Slytherin house. Which Harry Potter house(s) appears to be positively correlated with cheerfulness? Which Harry Potter house(s) appears to be negatively correlated with cheerfulness? Which Harry Potter house(s) appears to have no relationship with cheerfulness? In Stata, create scatterplots for cheerfulness and each Harry Potter house, with cheerfulness as the y-axis variable. Include all four graphs with your answer. *Note: if you are working in Windows, I suggest exporting the graphs to a .emf format.*
2. Instead of looking at a continuous measure of cheerfulness, I've decided that it makes more sense to just divide students into three broad groups. Divide students into thirds and label the groups "Low," "Medium," or "High" on your new variable, cheerfulness\_group. Don't forget to label the variable. Then, construct a bar graph showing the average ravenclaw, gryffindor, hufflepuff, and slytherin score for "Low," "Medium," and "High" cheerfulness. Make sure to include the bar graph with your answer. *Hint: if you do this correctly, you should be able to get 17 students in each group.*
3. Honestly, that bar graph didn't tell us that much. Instead of looking at each student's percentage Gryffindor, Ravenclaw, Hufflepuff, and Slytherin, let's just assign each student to a house and then examine cheerfulness scores by houses! The rule is, a student is assigned to a house if their percentage for that house is higher than their percentage for any other house. There are 5 students who have tied scores for multiple houses. Break the tie by assigning those students to the house that has fewer prior members. Tabulate the count of students by Harry Potter house. Then, create a box plot showing cheerfulness by Harry Potter houses. Also create a plot of Harry Potter house versus cheerfulness group to examine whether each house has majority low, middle, or high cheerfulness students (Hint: use `-catplot-`). Do these two graphs show us anything different than the correlations from Q1?
4. You role two 6-sided dice. (i) What is the probability that you role "snake eyes" (two 1s)? (ii) What is the probability that the sum of the two numbers is 4?

5. A mother and daughter were comparing their scores for the math GRE. The mother got 63 and the daughter got 72. The daughter was pleased that she had beaten her mother.

In the year when the mother took the exam, the mean mark was approximately Normally distributed with a mean of 50.4 and a standard deviation of 13.5.

In the year when the daughter took the exam, the mean mark was approximately Normally distributed with a mean of 58.0 and a standard deviation of 15.3.

- a. Comment on how the mother's and daughter's scores compare.
- b. In the year the daughter took the test, 86% of all college graduates took it, in the year the mother took the math GRE, only 41% of college graduates took the test. In one or 2 sentences, discuss how this affects your answer for (a).