

Political Science 100a/200a
Fall 2001
Problem Set 3

1. (a) Use Stata to create two independent random variables X and Y on 1000 observations with the following distributions:

$$\begin{aligned} X &\sim N(6, 3) \\ Y &\sim N(20, 8). \end{aligned}$$

(Show the Stata commands you used to report your work.)

- (b) Repeat (a), but now make the variables X and Y such that they are positively correlated with $\rho = .82$.
- (c) With the data you generated from part (b), use Stata to calculate the mean of Y conditional on X being in the bottom 25% of the X values. Repeat for X values within each of the next three quartiles. Fit a line by hand based on the expected value of Y for each quartile and the average value of X within each quartile, and estimate its slope. Compare this value to slope coefficient you estimate by regressing Y on X .
2. *Basketball*. The following questions concern two basketball players, Ariel and Brenda, who play on the same team. Suppose that the probability that Ariel makes any given shot is .44. Brenda's probability is .35. Assume that their chances are independent.

- (a) Suppose that in a certain game Ariel will take 25 shots. Estimate the probability that she makes 15 or more of them. First use the normal approximation, taking care about the endpoint (see FPP, p. 317). Then find the exact probability using the relevant probability distribution and Stata.
- (b) Now suppose that Ariel will take 50 shots in a given game. Estimate the probability that she will make 30 or more of them using the normal approximation. Explain why the result is different from (a).
- (c) Recall that Brenda's true probability of making a shot is .35. Using the normal approximation, estimate the probability that her observed shooting percentage in a series of games in which she takes 100 shots total is between .40 and .45.
- (d) A third player, Claire, is observed to have a field goal percentage of .50 in a series of games in which she takes 90 shots. Can we reject the hypothesis that her true, "underlying" probability of making a shot is .40 at the 1% level? at the 5% level?
- (e) Suppose that in a certain game Ariel will take 40 shots and Brenda will take 30. Assuming no dependencies, estimate the probability that the two of them *together* will make at least 35 shots. (Hint: use the fact that the sum of two normal distribution is a normal distribution.)
3. Suppose that 9 hand recounts were done of the presidential vote in Florida, and the results are -321, 1807, 1434, -822, -1256, 977, 1010, -43, 2738, where each number represents Bush's vote minus Gore's vote (so, for example, -321 means that in that recount, Gore was estimated to have received 321 more votes than Bush in Florida).

Can you reject the null hypothesis that Gore won the election against the alternative that Bush won the election (based on these numbers; leave aside the issue of confusing or illegal ballot forms). (You may assume that the magnitude of the errors arising in the counting process did not change appreciably in successive counts.) Hint: Be sure to have read and understood FPP chapter 26.

4. Download the data set **cddata.dta** from the course website. This contains a number of variables measured for countries in 1995 (or for some variables average values for the years 1995-99, or 1985-94). Public debate on the root causes of the spread of Islamic fundamentalism in heavily muslim countries often refers to socio-economic factors such as rapid population growth, poor economic performance, or a high percentage of young males in the population (e.g., Samuel Huntington, *Clash of Civilizations* makes this argument). This question asks you to compare average rates on one such variable between muslim-majority and other countries. Begin by creating a variable marking the states with 50% or greater muslim population (use the variable **mus**, which is percentage of population that is muslim).
 - (a) Suppose your (alternative) hypothesis is that Muslim majority countries have a significantly greater population share of young (15-24 year old) men (the variable name is **ymaleper**). Perform a t test to see if you can reject the relevant null hypothesis. Show your work (i.e., do not just reproduce the Stata output).
 - (b) Repeat the above, but restrict attention to countries in sub-Saharan Africa, Asia, and N. Africa/Middle East (use the variable **colony**, which marks these). Given a substantive interpretation of the difference in your results.
5. Still using **cddata.dta**, note the variable **cwar3**. This variable is 0 for countries that had no civil wars since 1945, 1 for countries that had one civil war, and 2 for countries that had more than one civil war in this period. Use a χ^2 test to see if you can reject the hypothesis that muslim-majority status is independent of the distribution of civil wars, as measured by this variable. Show the math (i.e., do not just report Stata's result), and interpret your result.
6. Draw a random sample of 25 countries from **cddata.dta** by using the **sample ...** command. Estimate the average percentage of population living in urban areas (**urbpop**) by country using this sample, and provide a 95% confidence interval. Compare your estimate based on the sample to the true population value. Then do the same for the muslim majority countries your 25 country sample. Compare your results to the true population value for these countries. Are your results less precise? If so, why?