GV300: Quantitative Political Analysis

Problem Set 3

Due Thursday, November 8, 9.45am on Faser

1. (15 marks) Expectation of random variables:
   (a) (10 marks) Show for arbitrary random variable X that $Var(X) = E[X^2] - E[X]^2$
   (b) (5 marks) Is it generally true that $E[f(x)] = f(E[x])$ for arbitrary random variable f? Show your reasoning.

2. (20 marks) Find the PMF, mean, and variance of the following random variable and show your computations: y is the number of successful military interventions in 7 peacekeeping mission where the probability that an intervention is successful is .2.

3. (20 marks) Using your preferred statistical software, demonstrate that your calculations in problem 2 are approximately correct. Please turn in your run file (or copy and paste the code you wrote into your answer sheet) as well as any plot or table you produced to answer this question.

4. (15 marks) Theresa May's cabinet is an unruly bunch. Let's say we record the number of departures (resignations) from her cabinet on a monthly basis. The table below gives the number of departures per month and the associated probability distribution:

Number of departures	0	1	2	3	4
Probability	.6	.2	.12	.05	.03

   Calculate mean and variance of the number of departures.

5. (15 marks) Look at the election returns from the 2009 Iranian presidential election contained in the data set `Iran Elections 2009.dta`. It holds six variables:

   - the *number of votes* for four candidates – these variables are labeled with the name of the candidates: *ahmadinejad*, *karroubi*, *mousavi*, *rezaee*)
   - the *number of valid votes*
   - the *vote totals* in 29 provinces of Iran.

   Look at the four *numbers of votes*-variable; how could we judge whether the election was rigged in favor of one candidate? Use your preferred statistical software and plot these variables in a way that illustrates your answer; submit the plot with your answer. This is a fun question and there is a good answer, debate with others, think creatively.

6. (15 marks) In problem set 1, you sketched your research project and listed your main variables of interest (dependent variable Y_i and main explanatory variable of interest T_i). Now, you should start thinking about how to model the relationship between T_i and Y_i more precisely. What else is out there in the world potentially affecting Y_i or T_i or both in a way that it could change your understanding of the relationship of interest between T_i and Y_i. That is, give a list of variables, observables X_i and unobservables U_i, which you think you would need to include in the statistical analysis of your study. Also, restate your dependent variable Y_i and main explanatory variable(s) T_i. Your knowledge which variables are in X_i and U_i could be drawn from the literature on your topic or your understanding

of what makes the world go round. For this exercise you should provide 3-4 variables each for the set of observables X_i and the set of unobservables U_i. You do not need to (but you certainly may) provide literature to back up your claims why a particular variable in X_i or U_i needs to be included when you specify the data generating process. This exercise tries to make you identify variables that should enter your statistical analysis related to your project later in the term.