PLSC 503: Problem Set 4

Thad Dunning

Department of Political Science

Yale University

Spring 2010

Due February 9, 2010

1 Theoretical/Conceptual Exercises

Before completing this problem set, it is recommended that you work Exercise Sets A and B in Freedman (2009, Chapter 5). You should also read and think about the discussion questions.

1. Suppose that you have an $n \times 1$ column vector of observations on a variable Y and an $n \times 1$ column vector of observations on a variable X. You convert each observation for each variable to standard units. Then you run a regression of standardized Y on standardized X. Show that the fitted slope coefficient on standardized X is r, the coefficient of correlation between Y and X.

2. Exercise 6.C.2 in Freedman (2009), that is, exercise C.2 in Chapter 5.

3. Exercise 6.C.3 in Freedman (2009).

4. Exercise 6.C.4 in Freedman (2009). Conduct a test of the hypothesis that $\beta_2 = \beta_1$.

5. Holland (1986) presents the slogan "No causation without manipulation." What does he mean by this? How convincing do you find his argument? In one or two paragraphs, discuss the merits and possible demerits of his argument about the status of individual attributes—such as race or gender—as causal variables.

2 Computer exercise

1. Data snooping

   (a) Generate 1000 observations of a standard normal random variable y.

   (b) Generate 1000 observations each of 50 standard normal random variables x_1 through x_{50}.

   The following loop syntax in Stata may be useful:

forvalues i=1/50{

gen x‘i’=invnorm(uniform())

}

In the newer versions of Stata, you'll replace invnorm(uniform()) with rnormal().

(c) Regress y on a constant and the 50 variables x_1 through x_{50}. Report your output.

(d) Keep the variables for which the estimated coefficients are significant at the 0.10 level. Regress y on these variables. Report your output.

(e) Repeat (a)-(d), this time screening at the 0.25 level in (d). Report your output, and compare your results to your results screening at the 0.10 level. What differences do you observe, and why?