Economics 20 Prof. Patricia M. Anderson

Problem Set 2

(due at the beginning of class October 16, 2000)

Part I. For each of the following questions, choose the one best answer. Briefly explain your reasoning.

1 & 2) An economist is interested in the determinants of child labor, and based on the following population model, wants to test whether the literacy of the mother and father have equal effects.

$$\text{hours worked} = \beta_0 + \beta_1\text{age} + \beta_2\text{rural} + \beta_3\text{mother's literacy} + \beta_4\text{father's literacy} + \beta_5\text{female} + u$$

Thus, the economist wants to test the following: H_0: $\beta_3 = \beta_4$; H_1: $\beta_3 \neq \beta_4$, which can be rewritten in terms of the parameter θ, as H_0: $\theta = \beta_3 - \beta_4 = 0$; H_1: $\theta \neq 0$. The economist produces the following Stata output, where *tradhrs* and *mlit* are the variables for hours worked and mother's literacy respectively, while all other variables but X are as in the model above.

tradhrs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	1.724362	.0394592	43.70	0.000	1.647017	1.801708
rural	5.824115	.2389982	24.37	0.000	5.355645	6.292586
mlit	-.4916243	.4298869	-1.14	0.253	-1.334264	.3510151
X	-2.468815	.2486429	-9.93	0.000	-2.95619	-1.981439
female	1.67421	.222099	7.54	0.000	1.238864	2.109555
_cons	-11.28735	.4963216	-22.74	0.000	-12.26021	-10.31449

1. If this output provides a direct estimate of θ and its standard error, as well as a direct estimate of all but one of the β's, how must X be defined?

a) X is father's literacy
b) X is father's literacy – mlit
c) X is mlit – father's literacy
d) X is mlit + father's literacy

2. If this output provides a direct estimate of θ and its standard error, as well as a direct estimate of all but one of the β's, what is the estimate of θ?

a) –.49
b) –2.47
c) –1.98
d) –2.96

3. Consider the probability density function for a t-distribution. As degrees of freedom increase, the symmetric cutoff points that leave only 5% of the probability in the tails of the distribution:

a) move further apart (i.e. further away from zero)
b) move closer together (i.e. closer to zero)
c) go to positive and negative infinity
d) both go to zero

4. Suppose you want to test for whether you can predict house price based on assessments, without knowing house characteristics. You estimate $price = \beta_0 + \beta_1 assess + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + u$ using a sample of 125 houses. Then your best approach to answering the question is to:

a) check the overall F-test reported by Stata – if it is > 2.45 you definitely need the characteristics

b) check each of the p-values for β_2, β_3 and β_4 – if none are < 0.05, you definitely don't need the characteristics

c) check each of the p-values for β_2, β_3 and β_4 – if none are < 0.05, you definitely need the characteristics

d) estimate $price = \beta_0 + \beta_1 assess + u$ and calculate $((R^2_1 - R^2_2)/3) / ((1-R^2_1)/(120))$, where R^2_1 is the R^2 from the first regression and R^2_2 is the R^2 from this second regression. If it is > 2.68 you definitely need the characteristics.

5. Suppose you want to estimate a model that will allow you to test the effect of being a pot smoker (pot=1) and a drinker (drink=1) on the number of sexual partners (partners) an individual has had. The best model to estimate would be:

a) partners = $\beta_0 + \beta_1$pot + $\beta_2$drink + u

b) partners = $\beta_0 + \beta_1$pot *drink + u

c) partners = $\beta_0 + \beta_1$pot + $\beta_2$drink + $\beta_3$pot*drink + u

d) partners = $\beta_1$pot *drink + u

Part II. Stata Problems. See "Important Things to Know about Stata" on our website for more about installing and using Stata. This PS uses the Stata commands **desc, gen, tab (**and **tab** with the **gen()** option**)** and **reg**. All problems use driving.dta, which is a random sample of HS students in 2001.

1. a) Use the desc command to take a look at what the variables are. Note that many are dummy variables. In this data set, the categorical variables have been given value labels. Thus, when you use the sum command to look at the variables *gradecat* it ranges from 1 to 5. If you use the tab command, however, you see that each number represents a category. Look at the means of all the variables. Look at frequencies for the categorical variable and create a set of dummy variables for it.

b) Suppose that you are interested in the determinants of seatbelt wearing. You have in mind the following heuristic model: seat belt wearing = f(demographics, other car behaviors, school behaviors). Estimate the econometric model implied by this heuristic model, assuming that you want your comparison group to be white males who usually get F's and have driven drunk and rode with a drunk.

c) Are blacks significantly more or less likely than whites to wear seatbelts? What about Hispanics? What about people who ride with drunk drivers? Who are themselves drunk drivers? Does being a better student make you more or less likely to wear a seatbelt?

d) Are males or females more likely to wear seatbelts? If the data set had a dummy for male, instead of for female, what would the estimated coefficient on the male dummy be? (Assuming no hermaphrodites!) Create interactions of *female* and *drivedrunk* and of *female* and *ridedrunk* and add them to the regression. Is either significantly different from zero at the 5% level (assume a 2-sided test)? What is the effect of riding with a drunk for males? For females? What is the derivative of seatbelt wearing with respect to being female?