Econometrics Exercise session for midterm preparation

Problem 1

Suppose that X is the number of free throws made by a basketball player out of two attempts and assume that the individual probabilities for each outcome of X are the following:

pr(x=0)=0.2; pr(x=1)=0.44 and pr(x=2)=0.36

i) Define the random variable.
ii) Draw the probability distribution associated to the above random variable.
iii) Calculate the expected value of the above random variable.
iv) Calculate the probability that the player makes at least one free throw

Problem 2

We have information about mortality rates (MORT=total mortality rate per 100,000 population) in a specific year for 51 States of the United States combined with information about potential determinants: INCC (per capita income by State in Dollars), POV (proportion of families living below the poverty line), EDU (proportion of population completing 4 years of high school), TOBC (per capita consumption of cigarettes by State) and AGED (proportion of population over the age of 65). Estimation results are presented in the following table:

OLS Estimation Results

Variable	Model 1 coefficients	Model 2 coefficients	Model 3 coefficients
Constant	194.747	531.608	-9.231
	(53.915)	(94.409)	(176.795)
Aged	5,546.56	5,024.38	5,311.4
	(445.727)	(358.218)	(334.415)
Incc		0.014	0.015
		(0.0038)	(0.0037)
Edu		-682.591	-285.715
		(114.812)	(152.926)
Pov			854.178
			(302.345)
Tobc			0.989
			(0.342)
n	51	51	51
Adjusted R squared	0.759	0.856	0.884
SSR	228,770.3	128,260.1	99,303.73

i) Interpret the slope coefficient in Model 1 and validate it at 1% significance level.

ii) Validate the joint significance of Model 2 in comparison to model 1 at 1% significance level?

iii) Comment on the effect of INCC on MORT in the second model. Why do you think is a positive and significant effect?

iv) In Model 3 we add two new explanatory variables: POV and TOBC. Test whether this inclusion helps to improve the quality of the model at 1% significance level. Is model 3 the best in terms of goodness-of-fit?

v) Are the effects of these two new variables the expected ones? Are they individually significant at 1% significance level?

vi) What about the individual significance of EDU in model 3 if compared with model 2? Why?

Problem 3

Suppose you are interested in studying the tradeoff between time spent sleeping and working and to look at other factors affecting sleep. You specify the following model:

$$sleep = \beta_0 + \beta_1 * totwrk + \beta_2 * educ + \beta_3 * age + u$$

where $sleep$ and $totwrk$ (total work) are measured in minutes per week and $educ$ and age are measured in years.

Suppose we estimated the following regression:

$$\widehat{sleep} = \underset{(112.28)}{3638.25} + \underset{(.017)}{0.148} * totwrk - \underset{(5.88)}{11.13} * educ + \underset{(1.45)}{2.2} * age$$

$n = 706,\ R^2 = .113$

where we report standard errors along with the estimates.

(i) Is either *educ* or *age* individually significant at the 5% level against a two-sided alternative? Show your work.

(ii) Dropping *educ* and *age* from the equation gives

$$\widehat{sleep} = \underset{(38.91)}{3586.38} + \underset{(.017)}{0.151} * totwrk$$

$$n = 706,\ R^2 = .103$$

Are *educ* and *age* jointly significant in the original equation at the 5% level? Justify your answer.

(iii) Does including *educ* and *age* in the model greatly affect the estimated tradeoff between sleeping and working?

(iv) Suppose that the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (i) and (ii)?

Problem 4

When estimating wage equations, we expect that young, inexperienced workers will have relatively low wages and that with additional experience their wages will rise, but then begin to decline after middle age, as the worker nears retirement. This lifecycle pattern of wages can be captured by introducing experience and experience squared to explain the level of wages. If we also include years of education, we have the equation:

$$Wage = \beta_0 + \beta_1 * Educ + \beta_2 * Exper + \beta_3 Exper^2 + u$$

a) What is the marginal effect of experience on wages?
b) What sign do you expect for each of the coefficients? Why?
c) Estimate the model using data ***cps_small.gdt***. Do the estimated coefficients have expecting signs?
d) Test the hypothesis that education has no effect on wages. What do you conclude?
e) Test the hypothesis that education and experience have no effect on wages. What do you conclude?
f) Include the dummy variable *black* in the regression. Interpret the coefficient and comment on its significance.
g) Include the interaction term of *black* and *educ*. Interpret the coefficient and comment on its significance.
h) Transform dependent variable in logarithmic form and estimate the equation. Interpret the coefficients.

Problem 5

consider a simple model to compare the returns to education at junior colleges and four-year colleges; for simplicity, we refer to the latter as "universities." The population includes working people with a high school degree, and the model is:

$$\log(wage) = \alpha_0 + \alpha_1 jc + \alpha_2 univ + \alpha_3 exper + u \qquad (1)$$

where

jc is number of years attending a two-year college, *univ* is number of years at a four-year college. *exper* is months in the workforce.

Note that any combination of junior college and four-year college is allowed, including *jc* =0 and *univ* = 0. Use the data ***twoyear.dta***

i) Test the hypothesis that $\alpha_1 = \alpha_2$. The hypothesis of interest is whether one year at a junior college is worth one year at a university.

(ii) The variable phsrank is the person's high school percentile. (A higher number is better. For example, 90 means you are ranked better than 90 percent of your graduating class.) Find the smallest, largest, and average phsrank in the sample.

(iii) Add phsrank to regression (2) and report the OLS estimates in the usual form. Is phsrank statistically significant? How much is 10 percentage points of high school rank worth in terms of wage?

(iii) Does adding phsrank to regression (2) substantively change the conclusions on the returns to two- and four-year colleges? Explain.