Exercise Session 5

Problem 1

Suppose that you have a sample of n individuals who apart from their mother tongue (Czech) can speak English, German, or are trilingual (i.e., all individuals in your sample speak in addition to their mother tongue at least one foreign language). You estimate the following model:

$$wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 DM + \beta_5 Germ + \beta_6 Engl + \varepsilon \ ,$$

where

$educ$ . . . years of education
IQ . . . IQ level
$exper$ . . . years of on-the-job experience
DM . . . dummy, equal to one for males and zero for females
$Germ$ . . . dummy, equal to one for German speakers and zero otherwise
$Engl$ . . . dummy, equal to one for English speakers and zero otherwise

(a) Explain why a dummy equal to one for trilingual people and zero otherwise is not included in the model.

(b) Explain how you would test for discrimination against females (in the sense that *ceteris paribus* females earn less than males). Be specific: state the hypothesis, give the test statistic and its distribution.

(c) Explain how you would measure the payoff (in terms of wage) to someone of becoming trilingual given that he can already speak (i) English, (ii) German.

(d) Explain how you would test if the influence of on-the-job experience is greater for males than for females. Be specific: specify the model, state the hypothesis, give the test statistic and its distribution.

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

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:

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

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***

a) 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?

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