Exercise Session - Problem Set 10

Multiple Regression Analysis with Qualitative Information

Problem 1 (**Quadratics**, cf. Wooldridge, Exercise C6.2) Consider the model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u,$$

where *wage* is hourly wage, *educ* is years of education, and *exper* is years of workforce experience. Estimation via OLS gives

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.105932)}{0.127998} + \underset{(0.007468)}{0.090366} \cdot educ + \underset{(0.005197)}{0.041009} \cdot exper - \underset{(0.000116)}{0.000714} \cdot exper^2 \\ n &= 526, \quad R^2 = 0.3.\end{aligned}$$

(i) Is $exper^2$ statistically significant at the 1% level?

(ii) Using the approximation

$$\%\Delta\widehat{wage} \approx 100 \cdot (\widehat{\beta}_2 + 2\widehat{\beta}_3 exper)\Delta exper,$$

find the approximate return to the fifth year of experience. What is the approximate return to the twentieth year of experience.

(iii) At what value of *exper* does additional experience actually lower predicted wage?

(iv) How could you test whether additional experience has a negative effect on predicted wage for $exper > 25$?

Problem 2 (**Interaction Terms**, cf. Wooldridge, Exercise C6.3) Consider the model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ \cdot exper + u. \tag{1}$$

(i) Find the return of another year of education, holding *exper* fixed.

(ii) State the null hypothesis that the return to education does not depend on the level of $exper$. What could be a reasonable alternative hypothesis?

(iii) What is the interpretation of parameters β_1 and β_2 in model (1)?

(iv) How could you test a hypothesis about the return to education when $exper$ is equal to its sample average?

Problem 3 Suppose we have estimated

$$y = 10 + 2 \cdot x + 3 \cdot female,$$

where y is wage, x is education, and $female$ is one for females and zero for males.

(a) If we were to rerun this regression with the dummy redefined as two for females and one for males, what results would we get?

(b) If it were defined as one for females and minus one for males, what results would we get?

Problem 4 Suppose two researchers, A and B, with the same data, have run similar regressions, namely

- Researcher A:

$$y = \delta_0 + \delta_1 \cdot x + \delta_2 \cdot female + \delta_3 \cdot region + \delta_4 \cdot female \cdot region$$

- Researcher B:

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot female + \beta_3 \cdot region + \beta_4 \cdot female \cdot region,$$

where $female$ is one for females and zero for males, but researcher A has defined $region$ as one for north and zero for south, whereas researcher B has defined it the other way—zero for north and one for south. Researcher A gets an insignificant t value on the $female$ coefficient, but researcher B does not.

(a) In terms of the interpretation of the model, what hypothesis is A implicitly testing when looking at the significance of his t value?

(b) In terms of interpretation of the model, what hypothesis is B implicitly testing when looking at the significance of her t value?

(c) In terms of the parameters of her model, what null hypothesis would B have to test in order to produce a test of A's hypothesis?

Problem 5 Suppose we have obtained the following regression results:

$$y = 10 + 5 \cdot x + 4 \cdot female + 3 \cdot region + 2 \cdot female \cdot region,$$

where $female$ is one for females and zero for males, region is one for north and zero for south.

(a) What coefficient estimates would we get if we regressed y on an intercept, x, NF (one for northern females, zero otherwise), NM (one for northern males, zero otherwise), and SF (one for southern females, zero otherwise)?

Problem 6 A friend has added regional dummies to a regression, including dummies for all regions and regressing using a no–intercept option. Using t tests, each dummy coefficient estimate tests significantly different from zero, so he concludes that region is important.

(a) Why would he have used a no–intercept option when regressing?

(b) Has he used an appropriate means of testing whether region is important? If not, how would you have tested?