**Economics Honors Exam 2010 Solutions: Econometrics, Question 8**

**This question consists of two subquestions, each worth 30 points.**

*The answer keys below are suggested solutions, and students may propose alternative answers. As long as they provide valid arguments and well-founded equations, (partial) credits may be awarded at the grader's own judgement.*

1. **(30 points)** In this question we will look at the relation between the labor earnings and lottery winnings. Using data from a study by Imbens, Rubin and Sacerdote (2001) on the lottery prize, earnings after winning the lottery (earn), earning prior to winning the lottery (priorearn), age and the number of lottery tickets bought (tixbot), we find the following results for a regression of earnings on the lottery prize, age and prior earnings:

$$\begin{array}{rcccccccc} \widehat{\text{earn}}_i & = & 17.48 & - & 0.107 \cdot \text{prize}_i & - & 0.26 \cdot \text{age}_i & + & 0.79 \cdot \text{priorearn}_i \\ & & (2.41) & & (0.055) & & (0.04) & & (0.06) \end{array}$$

The lottery prize and earnings variables are measured in thousands of dollars.

a. **(7 points)** Construct a 95% confidence interval for the effect of the lottery prize on earnings.

**Answer:**

$[-0.107 - 1.96 \times 0.055, -0.107 + 1.96 \times 0.055] = [-.215, .001]$

**Points: This is worth 7 points**

2 points for noting the mean is $-0.107$.

2 points for noting the standard deviation is 0.055.

2 points for using 1.96.

1 point for a correct answer. (Students do not have to calculate the final numerical result $[-.215, .001]$ to be given full credit.)

b. **(7 points)** Predict earnings for a 50-year old individual who won a lottery prize of $20K$, and who was making 15K prior to winning the lottery.

**Answer:**

$17.48 - 0.107 \times 20 - 0.26 \times 50 + 0.79 \times 15 = 14.19$

**Points: This is worth 6 points**

3 points for using the regression equation.

1 point for the item $-0.107 \times 20$.

1 point for the item $-0.26 \times 50$.

1 point for the item $0.79 \times 15$.

1 point for a correct answer 14.19.

c. **(4 points)** Suppose we are concerned that there is a variable missing from this regression. This omitted variable is positively related to earnings, and positively related to the lottery prize. What is the sign of the bias on the coefficient on prize from omitting this variable? In other words, without this variable we estimated the effect of the prize to be -0.107. If we included this omitted variable in the regression, the effect of prize on earnings would be $b$. What is the sign of $-0.107 - b$?

**Answer:**

The coefficient in the regression with this omitted variable included would be smaller (more negative) than -0.107 because of the omitted variable bias formula.

**Points: This is worth 4 points**

4 points for a correct answer.

Partial credit for explanation (but wrong answer) at the grader's own judgement.

d. **(6 points)** Suppose we think that the effect of the lottery prize differs depending on the value of prior earnings. In order to analyze this we estimate the regression

$$\widehat{\text{earn}}_i = 15.07 - 0.013\text{prize}_i - 0.26\text{age}_i + 0.99\text{priorearn}_i - 0.0078\text{priorearn}_i \times \text{prize}_i$$

| | | | | |
|---|---|---|---|---|
| (2.76) | (0.077) | (0.04) | (0.13) | (0.0045) |

Test whether the coefficient on the interaction is equal to zero at the 5% level, against the alternative that it differs from zero.

**Answer:**

The t-statistic is $-0.0078/0.0045 = -1.75$, which is less than 1.96 in absolute value, so we do not reject the hypothesis.

**Points: This is worth 6 points**

2 points for calculating the t-statistic using the correct equation $-0.0078/0.0045$.

1 point for getting the t-statistic correct ($-1.75$).

1 point for comparing the t-statistic with 1.96.

2 points for rejecting the hypothesis.

e. **(6 points)** Predict the effect of an additional dollar of prize money for someone with prior earnings equal to 10K.

**Answer:**

$-0.013 - 0.0078 \times 10 = -0.091$

**Points: This is worth 6 points**

2 points for the item $-0.013$.

2 points for the interaction term $-0.0078 \times 10$.

2 points for getting the correct answer ($-0.091$).

2. **(30 points)** Suppose we are interested in estimating the effect of a job training program on earnings. We have data, for voluntary job training program, on earnings after the training program ($Y_i$), an indicator variable for whether the individual participated in the training $W_i$, and a variable measure earnings prior to the training ($X_i$).

a. **(9 points)** How would you test the null hypothesis that average earnings are the same for individuals who received the training and individuals who did not receive the training, against the alternative that they are different, at the 10% level?

**Answer:**

the t-statistic is

$$T = \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{S_0^2/N_0 + S_1^2/N_1}}$$

where

$$N_1 = \sum_{i=1}^{N} W_i \qquad N_0 = \sum_{i=1}^{N} (1 - W_i)$$

$$\overline{Y}_1 = \frac{1}{N_1} \sum_{i=1}^{N} W_i Y_i \qquad \overline{Y}_0 = \frac{1}{N_0} \sum_{i=1}^{N} (1 - W_i) Y_i$$

$$S_0^2 = \frac{1}{N_0 - 1} \sum_{i=1}^{N} (1 - W_i)(Y_i - \overline{Y}_0)^2$$

$$S_1^2 = \frac{1}{N_1 - 1} \sum_{i=1}^{N} W_i (Y_i - \overline{Y}_1)^2$$

compare the absolute value of the t-statistic to 1.645.

**Points: This is worth 9 points**

2 points for getting the t-statistic formula correct.

6 points for explaining the items in the t-statistic formula correctly. (Full or partial credits for verbal explanations at the grader's own judgement.)

1 point for comparing the absolute value of the t-statistic to 1.645.

b. **(6 points)** How would you construct a 95% confidence interval for the difference in average earnings between the individuals who took part in the training program and those who did not?

**Answer:**

$$95\%CI = \left(\overline{Y}_1 - \overline{Y}_0 - 1.96\sqrt{S_0^2/N_0 + S_1^2/N_1}, \overline{Y}_1 - \overline{Y}_0 + 1.96\sqrt{S_0^2/N_0 + S_1^2/N_1}\right)$$

**Points: This is worth 6 points**

2 points for noting the mean is $\overline{Y}_1 - \overline{Y}_0$.

2 points for noting the standard deviation is $\sqrt{S_0^2/N_0 + S_1^2/N_1}$.

2 points for using 1.96.

c. **(5 points)** What do you expect the sign of the correlation between pre-program earnings and post-program earnings to be?

**Answer:**

positive: earnings tend to be persistent.

**Points: This is worth 5 points**

3 points for answering "positive".

2 points for explaining it correctly.

Partial credit at the grader's own judgement.

d. **(5 points)** What do you expect the sign of the correlation between pre-program earnings and program participation to be?

**Answer:**

positive: participants tend to be more motivated, and were doing better already.

**Points: This is worth 5 points**

3 points for answering "positive".

2 points for explaining it correctly.

Partial credit at the grader's own judgement.

e. **(6 points)** Based on the previous two answers, what do you expect for the sign of the difference between $\beta_1$ and $\gamma_1$ in the following two regressions?

$$Y_i = \beta_0 + \beta_1 W_i + \varepsilon_i$$

and

$$Y_i = \gamma_0 + \gamma_1 W_i + \gamma_2 X_i + \varepsilon_i$$

**Answer:**

the same as the product of the two correlations, so positive if both previous answers were positive, or both were negative, and negative if the two previous answers had different signs.

**Points: This is worth 6 points**

3 points for noting that the sign depends on the product of the two correlations.

3 points for the correct explanation.

Partial credit at the grader's own judgement.