Economics 375: Introduction to Econometrics
Winter, 2016 Online Midterm

Student ID #: ANSWERS

Please do not place your name anywhere on this exam. I will give partial credit to students who show work. This is a closed book, closed neighbors, and closed note exam. In some cases, the correct answer to a question will require a hypothesis test (even if I don't specifically ask for one). When performing a hypothesis test, please list your null and alternative hypothesis, your test statistic, the critical value of the distribution, and your conclusions. Unless otherwise mentioned, all hypothesis tests should be performed at the 95% level of confidence. Partial credit will be given only if work is shown. The number of points each problem is worth appears in parenthesis.

1. A researcher gathered the following data measuring the wages (measured in dollars per hour) earned by employees and their tenure (the number of years they worked for their employer). The data gathered is:

<table>
<thead>
<tr>
<th>Wages</th>
<th>Tenure</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>36</td>
<td>23</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>32</td>
<td>25</td>
</tr>
</tbody>
</table>

The researcher found that $\Sigma Wages^2 = 4,028; \Sigma Tenure^2 = 1,688; \Sigma Wages \times Tenure = 2,512; \bar{Wages} = 25; \bar{Tenure} = 15; \sum (Wages - \bar{Wages})^2 = 278; \sum (Wages - Wages) = 0; \sum (Tenure - \bar{Tenure})^2 = 338; \Sigma Wages = 150; \Sigma Tenure = 90.$

a. The researcher is interested in estimating the regression: $Wages_i = \beta_0 + \beta_1 Tenure_i + \epsilon_i$. What are your estimates of $\beta_0$ and $\beta_1$? Interpret these estimates. (11)

$\hat{\beta}_1 = 7.75 \quad \hat{\beta}_0 = 13.37$

1 year = $77.50$
b. Does tenure impact wages? (I’ve reproduced a chart to help you organize data that might help you to answer this question) (12)

<table>
<thead>
<tr>
<th>Wages</th>
<th>Tenure</th>
<th>$C$</th>
<th>$C^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>5</td>
<td>4.75</td>
<td>22.57</td>
</tr>
<tr>
<td>24</td>
<td>18</td>
<td>-3.32</td>
<td>11.05</td>
</tr>
<tr>
<td>36</td>
<td>23</td>
<td>4.79</td>
<td>23.02</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>-1.57</td>
<td>2.477</td>
</tr>
<tr>
<td>18</td>
<td>11</td>
<td>-3.89</td>
<td>15.20</td>
</tr>
<tr>
<td>32</td>
<td>25</td>
<td>-4.75</td>
<td>56.25</td>
</tr>
</tbody>
</table>

\[ \sum = 74.9 \]

\[ G^2 = \frac{74.9}{6-1-1} = 18.725 \]

\[ SE(B_1) = \sqrt{\frac{18.725}{338}} = .235 \]

\[ H_0: B_1 = 0 \]
\[ H_n: B_1 \neq 0 \]

\[ t = \frac{.775 - 0}{.235} = 3.29 \]

\[ t_c, 4.95\% = 2.776 \]

Reject \( H_0 \) \( \bigcirc \)

\[ c. \text{ A colleague of mine claims that tenure is so important that firms raise wages by$1 per hour for every year of tenure. Evaluate this claim.} \] (8)

\[ H_0: B_1 = 1 \]
\[ H_n: B_1 \neq 1 \]

\[ t = \frac{.775 - 1}{.235} = -.957 \]

\[ t_c = 2.776 \]

Fail to Reject \( H_0 \) \( \bigcirc \)
d. I’ve been a Western employee for 16 years. Using the above regression, construct a 90% confidence interval for my hourly wage. (8)

\[
\hat{Y} = 13.37 + 1.775 \times 16 = 25.77
\]

\[
\sigma_f^2 = 18.725 \left[ 1 + \frac{1}{6} + \frac{(16-15)^2}{338} \right] = 21.90
\]

\[
25.77 \pm 2.132 \sqrt{21.9} = [15.79, 35.74]
\]

2. Is it possible to prove that a regression’s error terms are not correlated with its independent variables? If not, explain why not. If so, prove this. (8)

No. Error terms are never observed, only assumed.
3. First, a little explanation on baseball for those of you that are not sports fans. In each game a player has a certain number of "at-bats", i.e., chances to hit the ball. Sometimes the at-bat results in a "hit" which is good, sometimes it results in an "out," which is bad. Most players get hits in about 25-30% of at-bats. A player named Joe Dimaggio holds the record for longest "hitting streak" - he had at least one hit in each of 56 consecutive games. This is very difficult, and many baseball fans think that Dimaggio's record will never be broken. We are going to calculate the probability that a typical player would be able to do this. Suppose that a particular player has a career batting average of .300 (i.e., he has a 30% chance of getting a hit per at-bat), that he has four at-bats (chances to get a hit) per game, and that the outcome of each at bat is (statistically) independent of the others.

a. What is the probability this player will get at least one hit in their next game? (3)

\[ 1 - .7^4 = .7599 \]

b. What is the probability this player will get at least one hit in each of their next 56 games? (3)

\[ .7599^{56} = .0000002101 \]
4. In a study for Western, Dr. Carl Simpson, a former professor in the Sociology Department, commented on a third-party's work, "The analytic strategy used in this analysis is Ordinary Least Squares Regression. The dependent variable, salary, is continuous and essentially normally distributed, making regression an efficient and unbiased tool for estimating the unique effects of each factor tested."

Do you agree or disagree with the econometric content of this statement? Why or why not? (5)

Disagree. Distribution of dependent variables are not one of the classical assumptions.

5. Two variables Y and X are believed to be related by the following population regression equation:

\[ Y = B_0 + B_1X + \varepsilon \]

where \( \varepsilon \) is the usual random error with zero mean and a constant variance \( \sigma^2 \). To check this relationship one researcher takes a sample size of 8 and estimates \( B_0 \) and \( B_1 \) with OLS. A second researcher takes a different sample, again with size of 8, and also estimates \( B_0 \) and \( B_1 \) with OLS. The data they used and the results they obtained are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Researcher 1</th>
<th></th>
<th>Researcher 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>y</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>4.0</td>
<td>3</td>
<td>2.0</td>
<td>1</td>
</tr>
<tr>
<td>4.5</td>
<td>3</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>4.5</td>
<td>3</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>3.5</td>
<td>3</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td>4.5</td>
<td>4</td>
<td>11.5</td>
<td>10</td>
</tr>
<tr>
<td>4.5</td>
<td>4</td>
<td>10.5</td>
<td>10</td>
</tr>
<tr>
<td>5.5</td>
<td>4</td>
<td>10.5</td>
<td>10</td>
</tr>
<tr>
<td>5.0</td>
<td>4</td>
<td>11.0</td>
<td>10</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 1.875 + .750x \]

\[ (1.20) (.339) \]

\[ R^2 = .45 \]

\[ \sigma = .48 \]

\[ \hat{y} = 1.5 + .970x \]

\[ (.27) (.038) \]

\[ R^2 = .99 \]

\[ \sigma = .48 \]

where standard errors are in parenthesis below estimates of \( B_0 \) and \( B_1 \).

a. Explain why the standard error of \( B_1 \) for the first researcher is larger than the standard error of \( B_1 \) for the second researcher? (5)

\( \Sigma x \) is smaller for researcher \#1 than \#2.
A recent Southern Economic Journal article by Trost and Salehi-Isfahani examined the impact of completing homework on course grades in economics. Trost and Salehi-Isfahani randomly assigned students from many introductory classes into one of two groups. The first group were assigned homework and told to complete them, much as would happen in a standard class. The second group were given the same homeworks but told that they did not have to hand these in. Trost and Salehi-Isfahani then examined the grades earned by students on their final exams. Here are two of the OLS regressions Trost and Salehi-Isfahani estimated:

<table>
<thead>
<tr>
<th></th>
<th>Column A</th>
<th></th>
<th>Column B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Required</td>
<td>3.71</td>
<td>(1.49)</td>
<td>3.90</td>
<td>(1.62)</td>
</tr>
<tr>
<td>Enrollment</td>
<td>-.10</td>
<td>(.02)</td>
<td>-.08</td>
<td>(.02)</td>
</tr>
<tr>
<td>Male</td>
<td>-.72</td>
<td>(1.42)</td>
<td>-.81</td>
<td>(1.61)</td>
</tr>
<tr>
<td>Sophomore</td>
<td>-.22</td>
<td>(1.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>3.82</td>
<td>(2.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td>-5.15</td>
<td>(3.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Score</td>
<td>.74</td>
<td>(.12)</td>
<td>.81</td>
<td>(.13)</td>
</tr>
<tr>
<td>N</td>
<td>830</td>
<td></td>
<td>830</td>
<td></td>
</tr>
<tr>
<td>RSS</td>
<td>1,041</td>
<td></td>
<td>1,061</td>
<td></td>
</tr>
<tr>
<td>TSS</td>
<td>12,611</td>
<td></td>
<td>12,611</td>
<td></td>
</tr>
</tbody>
</table>

The dependent variable is the final exam test score (out of 100 possible points). Regression standard errors are in parenthesis.

A description of the independent variables are:

Homework Required: This variable is equal to zero if the observation is in the group that did not need to turn in the homework. It is equal to one if the observation is in the group that was required to turn in homework.

Enrollment: Is the size of the observation’s class.

Male is a variable equal to zero if the observation is a woman; it is equal to one if the observation is a man.

Sophomore, Junior, and Senior are variables equal to one if the observation is one of these three categories and equal to zero if not.

Math Score represents a basic math test that all students took prior to enrolling in the economics course. The Math Score varies from zero (the lowest possible) to 100 (the highest possible).
a. Using Column A, does requiring homework increase final exam scores? (8)

\[ H_0 : \beta_1 < 0 \]
\[ H_a : \beta_1 \geq 0 \]

\[ t = \frac{3.71 - 0}{1.49} = 2.49 \]
\[ t_{0.05} = 1.658 \ or \ 1.645 \]

Reject \( H_0 \).

b. Using Column A, do women perform better on economic exams than men? (8)

\[ H_0 : \beta_3 > 0 \]
\[ H_a : \beta_3 \leq 0 \]

\[ t = \frac{-0.72 - 0}{1.42} = -0.507 \]
\[ t_{0.05} = 1.658 \ or \ 1.645 \]

Fail to reject.
c. Using column A, do the seven independent variables explain a significant amount of the variation in final exam scores? (8)

\[ H_0: R^2 = 0 \]
\[ H_a: R^2 > 0 \]

\[
F = \frac{(12611 - 1041) / 7}{1041 / (830 - 7 - 1)} = 1305
\]

\[ F_{7,822} = 2.01 \]
\[ \text{Reject} \]

d. Does class standing (i.e. Sophomore, Junior, and Senior) impact final exam scores in this class? (8)

\[ H_0: \beta_4 = \beta_5 = \beta_6 = 0 \]
\[ H_a: \text{Not } H_0 \]

\[
F = \frac{(1061 - 1041) / 3}{1041 / (830 - 7 - 1)} = 5.26
\]

\[ F_{3,822} = 2.60 \]
\[ \text{Reject} \]