**Name:**

**Student Number:**

**Midterm 1**

**ECO3151**

**Winter 2011**

**Instructions:**

1. Print your name and student number at the top of this midterm
2. No programmable calculators
3. You can answer in pencil or pen
4. This midterm consists of 8 short answer questions
5. Total marks: 50

**Question 1a (pink copy), Question 1b (white copy)**
a) Show that $\sum_{i=1}^{n}(w_i - \bar{w})(x_i - \bar{x}) = \sum_{i=1}^{n}(w_i - \bar{w})x_i$. (5 marks)

$$\begin{aligned}
\sum_{i=1}^{n}(w_i - \bar{w})(x_i - \bar{x}) &= \sum_{i=1}^{n}[(w_i - \bar{w})x_i - (w_i - \bar{w})\bar{x}] \\
&= \sum_{i=1}^{n}(w_i - \bar{w})x_i - \sum_{i=1}^{n}(w_i - \bar{w})\bar{x} \\
&= \sum_{i=1}^{n}(w_i - \bar{w})x_i - \bar{x}\sum_{i=1}^{n}(w_i - \bar{w}) \\
&= \sum_{i=1}^{n}(w_i - \bar{w})x_i - \bar{x} \cdot 0 \\
&= \sum_{i=1}^{n}(w_i - \bar{w})x_i
\end{aligned}$$

**Question 1b (pink copy), Question 1a (white copy)**

Show that $R^2 = r_{x,y}^2$ where $r_{x,y}^2 = \frac{[\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}$ (5 marks)

$$\begin{aligned}
R^2 &= \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \\
&= \frac{\sum_{i=1}^{n}(\hat{\beta}_1(x_i - \bar{x}))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \\
&= \hat{\beta}_1^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \\
&= \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \\
&= \frac{[\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}
\end{aligned}$$

**Question 2 (pink copy) Question 6 (white copy)**

You are interested in exploring whether living close to work has an impact on the number of hours worked. More precisely, you believe that those who live close to their work can work longer hours because they are spending less time commuting. Your co-author comes into your office and tells you that he has found a very large dataset that has millions of observations, and that as a result, you will be able to precisely measure the impact of distance to work on hours worked. Do you agree with your co-author. Discuss. (5 marks)

It is true that the estimate will be more precise. However, You may not be precisely estimating what you care about. You need the causality assumption to hold, you also need a random sample,... If MLR1-4 hold then you know you will have an unbiased estimator of the parameter of interest.

**Question 3**

Does unbiasedness of the OLS estimator imply that the estimate will be close to the true parameter of interest? Briefly explain. (5 marks)

All it means is that on average your guess is correct. An individual guess (from one sample) could far from the true population parameter.

**Question 4 (pink copy)**

As a researcher you are interested in the impact of W on Z. You are provided with the following information

| W | Z |
|---|---|
| 10 | 30 |
| 15 | 30 |
| 20 | 90 |

Calculate $\hat{\beta}_0$ and $\hat{\beta}_1$. Show your work. (5 marks)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(w_i - \bar{w})(z_i - \bar{z})}{\sum_{i=1}^{n}(w_i - \bar{w})^2} = 6$$
$$\hat{\beta}_0 = \bar{z} - \hat{\beta}_1 \bar{w} = -40$$

**Question 4 (white copy)**

As a researcher you are interested in the impact of Z on W. You are provided with the following information

| W | Z |
|---|---|
| 10 | 30 |
| 15 | 30 |
| 20 | 90 |

Calculate $\hat{\beta}_0$ and $\hat{\beta}_1$. Show your work. (5 marks)

$$\hat{\beta}_1 = \frac{\sum_{i=1} n(w_i - \bar{w})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} = 0.125$$
$$\hat{\beta}_0 = \bar{w} - \hat{\beta}_1 \bar{z} = 8.75$$

**Question 5**

Explain (in words) the importance of the random sampling assumption. (5 marks)

It means that it is representative of the population of interest. If, for example, you care about the return to education of Canadian workers that are 20 to 64 years of age, you want your sample to be representative of this population of interest. Otherwise you are not estimating your parameter of interest.

**Question 6 (pink copy) Question 2 (white copy)**

A clear distinction was made between $u_i$ and $\hat{u}_i$. What is that distinction? You cannot exclusively rely on equations. You must explain your answer in words. (5 marks)

$u$ is the error term in the model (which is not observed). All the factors that affect the dependent variable but are not in the explanatory variables, are in the $u$. $\hat{u}$ is the residual (which is observed). It is the part that the OLS line cannot account for.

**Question 7**

A sample consists of 1,129 women where $kids_i$ is the number of kids the woman $i$ has had, and $feduc_i$ is her level of education (in years). Given the information found in the log file (which is on the last page of the midterm)

a) Write the population model. (2 marks)

$$kids = \beta_0 + \beta_1 feduc + u \tag{1}$$

b) Interpret the estimate of $\hat{\beta}_1$. (3 marks)

An increase in one year of education will result in the mother having 0.06 less kids, ceteris paribus. (pink copy)

An increase in one year of education will result in the mother having 0.05 less kids, ceteris paribus. (white copy)

c) In this course, the professor has stressed the importance of the zero conditional mean assumption. What is it? Do you think it holds in this case? Justify your answer. (5 marks)

$$E(u|feduc) = 0$$

It means that the average u is the same (zero) for all levels of mother's education. Need a discussion of whether you believe it holds.

d) Provide two additional variables that should be included in the model. Write out the new population model. Make sure to clearly define your variables, and explain why they should be included. (5 marks)

See class answer.