Midterm

Each problem is weighted equally out of 100 points. In order to get full credit, you must show any calculations used to arrive at your answers and completely answer the questions.

1. Suppose the classical linear regression model applies to \( y = \beta_0 + \beta_1x_1 + \beta_2x_2 + u \). If the data are cross-sectional at a point in time and \( x_2 \) does not vary in the cross-section, should you include \( x_2 \) anyway to avoid bias in your estimate of \( \beta_1 \)? Explain your reasoning.

2. Answer true, false, or uncertain for each statement below. Explain your choice.
   a) When sample size increases, \( t \)-statistics also increase.
   b) If the true model is \( y = \beta_0 + \beta_1x_1 + \beta_2x_2 + u \) but you estimate \( y = \beta_0 + \beta_1x_1 + u \), instead, your estimate of \( \beta_1 \) is more efficient.

3. In a regression of weight on height involving 51 students, 36 males and 15 females, the following regression results were obtained (\( t \)-statistics in parentheses):

   Model A: \( weight = -232.0655 + 5.5662height \)  
   \( (-5.2066) \quad (8.6246) \)

   Model B: \( weight = -122.9621 + 3.7402height + 23.8238male \)  
   \( (-2.5884) \quad (5.1613) \quad (4.0149) \)

   Model C: \( weight = -107.9508 + 3.5105height + 2.0073male + 0.3263maleht \)  
   \( (-1.2266) \quad (2.6087) \quad (0.0187) \quad (0.2035) \)

   \( weight \) is in pounds; \( height \) is in inches; \( male = 1 \) for male, 0 for female; \( maleht = male \cdot height \)

   In addition, the following correlation coefficients were calculated:

   \[
   \begin{array}{ccc}
   male & maleht \\
   height & 0.6276 & 0.6752 \\
   male & -- & 0.9971
   \end{array}
   \]

   a) Which regression would you choose between A and B? Why? If you choose A, but B is actually the correct model, what kind of error are you committing?
   b) What does the coefficient on \( male \) in B suggest? Why does it become insignificant in C?

4. In problem 3 above, which regression would you choose between B and C? Why? Why would the coefficient on \( height \) stay about the same, but the coefficient on \( male \) change so drastically?

5. For the equation \( y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + u \), describe how you would test the joint hypothesis \( \beta_1 = \beta_2 \) and \( \beta_3 = 1 \).
6. Suppose you have run a multiple regression for which the F-statistic for the equation is very highly statistically significant, and so are all of the t-statistics. But the R-squared is very low, only .05. How can this be? Is there a mistake in the statistical software?

7. The following equation is estimated using OLS on quarterly data for 1989 to 2008 inclusive (n=80):

\[
\hat{y} = 2.20 + 0.104x_1 + 3.48x_2 + 0.34x_3
\]

standard errors in parentheses, explained sum of squares =112.5, unexplained sum of squares =19.5

a) Calculate the value of \( R^2 \).

b) Calculate the value of \( \bar{R}^2 \).

c) Calculate the \( F \)-statistic and indicate whether or not the null hypothesis is rejected.

8. Suppose you know that \( y \) depends linearly on \( x \), but you are not sure whether or not \( y \) also depends on another variable \( z \). A friend of yours (who hasn’t taken Econ 385) suggests that you should regress \( y \) on \( x \) first, calculate the residuals, and then see whether the residuals are correlated with \( z \).

a) What is the problem with this testing procedure?

b) What should you do instead?

9. Suppose you decide to estimate \( \sigma^2 \), the variance for a population, using \( \hat{\sigma}^2 = (x_1 - \bar{x})^2 \)

(in contrast to the standard estimator, \( \hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \)).

a) Is \( \hat{\sigma}^2 \) unbiased? If it is biased, is it too large or too small?

b) Is \( \hat{\sigma}^2 \) consistent? Explain.

10. Here is the equation you estimated using OLS on problem set #5 for #C8.12(i):

\[
math4 = 91.93 - .449 \text{ lunch} - 5.40 \text{ lenroll} + 3.52 \text{ lexppp}
\]

regular standard errors in ( ), robust standard errors in [ ]

Answer true, false, or uncertain for each statement below for this equation. Explain your choice.

a) If there is heteroskedasticity and we use regular standard errors instead of robust standard errors, the probability of making type I error increases.

b) If there is heteroskedasticity and we use regular standard errors instead of robust standard errors, the probability of making type II error increases.