ECO 4720 **Name:**
Exam #1

There are 3 questions with subparts. Please try to attempt everything. All the best!

1. The effect of class size on pupil achievement is a question of policy relevance, and thereby analyzed by several studies. One such analysis is interested in exploring how class size affects test scores for fifth graders. It relies on a data set containing information across 1000 schools and estimates the following simple regression:

$$\hat{score} = 75.32 - 0.031size.$$

Here, for any school:
score - average reading comprehension score (among fifth graders) varying between 0 and 100, and
size - average size of a class (for fifth graders).

a. Based on the simple regression above, what is the effect of increasing a school's average class size by 100 on the average reading score? Please explain the sign and magnitude of the impact.

Answer: Decreases score by 3.1.

b. Given that students are not randomly allocated to classes of different size, please provide an example of an unobserved characteristic that could be correlated with both the dependent and independent variables.

Answer: Teacher quality.

2. A number of studies explore the determinants of crime rate across countries. Suppose one such analysis relies on data across countries to estimate:

$$\log(crime) = \beta_0 + \beta_{\log(GNP)}\log(GNP) + \beta_{educ}educ + \beta_{urban}urban + \beta_{\log(police)}\log(police) + u.$$

The estimated equation is

$$\underset{}{\log(\hat{crime})} = \underset{(3.263)}{-7.67} + \underset{(0.319)}{1.061}\log(GNP) - \underset{(0.036)}{0.022}educ + \underset{(0.021)}{0.007}urban + \underset{(0.195)}{0.348}\log(police).$$

Here, the standard errors are in parentheses; the sample size $n = 65$ and $R^2 = 0.72$. Moreover,
crime: number of thefts as a percentage of population
GNP: per capita gross national product
educ: share of children enrolled in primary school
urban: percentage of population in urban areas, and
police: percentage of policemen in population

a. Suppose we have a two-tailed test with H_0: $\beta_{\log(GNP)} = 0$. For the 5% level of significance, what are the *t* distribution's critical values?

Answer: df = 65 – 4 – 1 = 60. Critical value = 2.

b. Given the estimated value of $\beta_{\log(GNP)}$, can we reject H_0 at the 5% level of significance? Explain by calculating the corresponding test statistic and comparing it against the *t* distribution's critical values.

Answer: Test statistic = 1.061/0.319 = 3.33. Reject H_0.

c. Given the information above, how much of the variation in log(*crime*) is explained by log(*GNP*), *educ*, *urban*, and log(*police*).

Answer: 72%.

3. Consider the data set "bwghtexam1" provided on AsULearn. Please download the data before proceeding with this question. Using the data, suppose the following multiple regression is estimated:

$$bwght = \beta_0 + \beta_1 mage + \beta_2 meduc + \beta_3 npvis + u.$$

The variables are
bwght: infant birth weight in grams
mage: mother's age in years
meduc: mother's education in years
npvis: total number of prenatal visits.

a. What is the average value of $bwght$ in the data?

Answer: 3402.685 grams.

b. What is the maximum value of $mage$ in the data?

Answer: 44 years.

c. Please use Stata to estimate the multiple regression model above and report the coefficient estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$?

Answer: $\hat{\beta}_1 = 2.037$, $\hat{\beta}_2 = 5.762$, and $\hat{\beta}_3 = 13.464$.

d. Given the regression results, what are the values of total, explained, and residual sum of squares?

Answer: SST = 578462996, SSE = 5248845.16, and SSR = 573214151.

e. From the regression results, what is the value of R^2?

Answer: 0.0091.

f. Suppose the null and alternative hypotheses are stated as H_0: $\beta_3 = 0$ and H_1: $\beta_3 \neq 0$; β_3 is the coefficient corresponding to $npvis$. Given your regression results, would you reject or fail to reject the null hypothesis at the 5% level of significance? Please explain in terms of the p-value.

Answer: Reject since p-value<0.05.

g. Next, suppose the null and alternative hypotheses are stated as H_0: $\beta_3 = 10$ and H_1: $\beta_3 \neq 10$; β_3 is the coefficient corresponding to $npvis$. Given your regression results, would you reject or fail to reject the null hypothesis at the 5% level of significance? Please explain in terms of the confidence interval.

Answer: Fail to reject since the 95% confidence interval contains 10.

h. Suppose one observation in the dataset has $bwght = 3060$ grams, $mage = 26$ years, $meduc = 12$ years, and $npvis = 12$ visits. Based on your regression coefficients, what is the predicted value of $bwght$, i.e., $\widehat{bwght}$?

Answer: Predicted value = 3106.791 + 2.037*26 + 5.762*12 + 13.464*12 = 3390.465.

i. For the observation from part (h) above, what is the value of the residual?

Answer: Residual = 3060 - 3390.465 = -330.465.

j. Suppose the null and alternative hypotheses are stated as H_0: $\beta_1 = 0, \beta_2 = 0$; and H_1: at least one of $\beta_1, \beta_2 \neq 0$. β_1 and β_2 are the coefficients corresponding to $mage$ and $meduc$, respectively. After estimating the regression models with and without H_0 imposed, what are the values of restricted and unrestricted R^2, respectively?

Answer: Restricted R^2 = 0.008; unrestricted R^2 = 0.009.

k. Using the values of restricted and unrestricted R^2 from part (j), construct the test statistic for the F test and calculate its numerical value.

Answer: Using the R^2 version of the formula for the F statistic, we have

$$F = \frac{(0.009 - 0.008)/2}{(1 - 0.009)/(1751 - 3 - 1)} = 0.881.$$

l. For the test in part (j), what is the critical value of the F statistic at the 5% level of significance? Given the test statistic from part (k), do you reject or fail to reject the null hypothesis at the 5% level of significance?

Answer: Numerator df = 2; denominator df = 1747; F statistic = 3. Fail to reject H_0.