ANSWER KEY

ECONOMICS 3950
Dr. R. Mueller
Spring 2008
Midterm Examination
March 6, 2008

Instructions: Students should answer all of the questions in Parts A, B and C. The value of each question is listed in parentheses following the question. You should show all of your work since partial credit is given for incomplete (but correct) answers. This examination is worth 240 points and counts as 30 per cent of your final grade.

NOTE: *You are free to use one 8 ½" x 11" handwritten sheet of formulas (one side only) and non-programmable handheld calculators in completing this examination.*

PART A: Long Answer Questions (120 points total)

1. (70 points total) Real housing prices in Calgary are a linear function of the square footage of the home. The results of the linear regression are as follows:

```
      Source |       SS       df       MS              Number of obs =    4682
-------------+------------------------------           F(  1,  4680) = 5283.09
       Model |  3.1126e+12     1  3.1126e+12           Prob > F      =  0.0000
    Residual |  2.7573e+12  4680   589170219           R-squared     =  0.5303
-------------+------------------------------           Adj R-squared =  0.5302
       Total |  5.8700e+12  4681  1.2540e+09           Root MSE      =   24273

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sqft |   75.57386   1.039747    72.68   0.000     73.53547    77.61225
       _cons |   59.86265   1547.726     0.04   0.969    -2974.408    3094.134
------------------------------------------------------------------------------
```

where price is the price of the home, and sqft is the number of square feet in the house.

(a) Interpret the constant term. How much faith do you put in this estimate? Why? (10 points)

Answer: Not much emphasis should be put on this estimate for two reasons: (1) a house of zero square feet doesn't seem too realistic, and (2) the t-statistic implies that the coefficient is not significant at any reasonable level.

(b) Assuming a house size of 2000 square feet, what does the point estimate say about the predicted price of this house? (5 points)

Answer: Actually, it says nothing because it can't talk. But what I mean to say is that the estimated housing price would be $151,207.58.

(c) How confident are you in the coefficient estimate on sqft? (5 points)

Answer: This is very accurately measured according to our estimates. The estimated standard error is very small which is why the 95% confidence interval is also small.

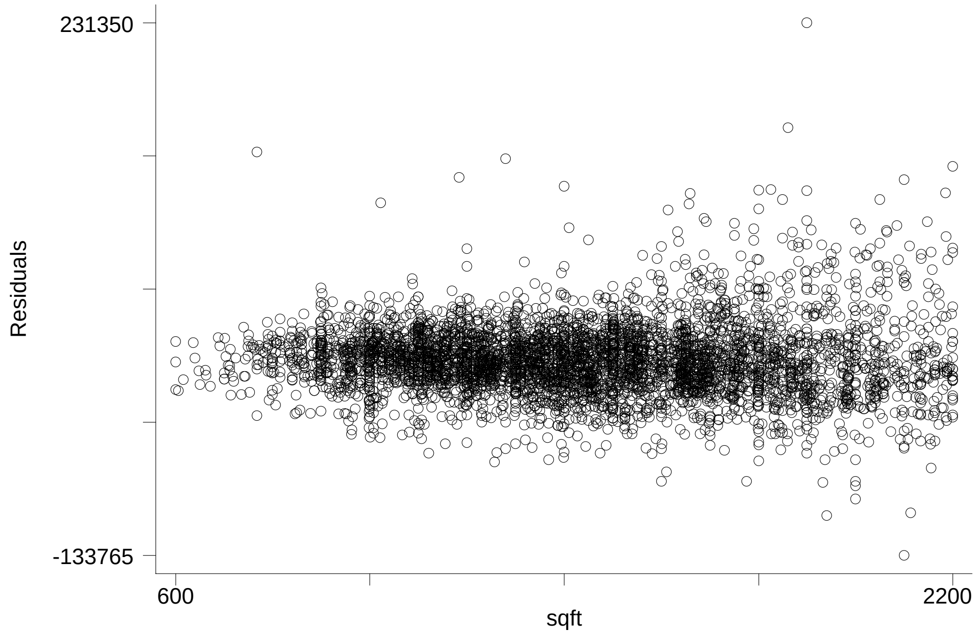
(d) How much of the variation in housing prices is explained by the square footage in this model? (5 points)

Answer: The R^2 value is .5303 meaning that about 53% of the variation of housing prices about the mean is explained by square footage.

(e) One of the assumptions of the OLS model is that $\text{var}(\varepsilon_t) = \sigma^2$. In other words, the variance of the error term is constant (or homeskedastic). How realistic is this assumption in the current model? (10 points)

Answer: This is not very realistic. We would likely see more variation amongst housing prices at the higher end of the housing market, and thus the variance of the error term would certainly be higher for larger houses.

(f) The following is a plot of the residuals of the estimated model against square footage. Explain what these mean in the context of part (e) above. (15 points)



Answer: The residuals are simply an estimate of the (unknown) error terms. Thus, the above plot shows that the assumption of homoskedasticity likely does not hold in the case of this model since the variance of the residual term is higher as housing size increases.

(g) What other variables might be important in determining housing prices? (10 points)

Answer: Other variables such as neighbourhood, number of bathrooms, upgrades, size of lot, age of the dwelling, etc. would also be very important in determining housing prices.

(h) Test the null hypothesis (at the 95% level) that the coefficient on sqft is equal to 80, against the alternative hypothesis that it is not equal to 80. (10 points)

Answer: The value of 80 falls outside of the 95% confidence interval, so obviously we could reject the null.

ANSWER KEY

2. (50 points total) The following estimates a model of the proportion of students at the state level (and the District of Columbia) who enroll (ENROLL) at private schools in the United States. The explanatory variables are as follows (all are at the state level):

CATHOL	Proportion of the population that is Catholic
PUPIL	Pupil-to-teacher ratio for public schools
WHITE	Proportion of population that is white
ADMEXP	Proportion of education expenditures for administration at public schools
REV	Per pupil education revenue for public schools
MEMNEA	Proportion of public school teachers who are members of the National Education Association
INCOME	Per capita household income (in thousands)
COLLEGE	Proportion of the population that has completed at least four years of college (i.e., university)

(a) Comment on the expected sign of each of the following coefficients (20 points):

CATHOL *Likely positive*

PUPIL *Positive since a smaller number is better as a measure of public school quality*

WHITE *Could go either way depending on preferences, but likely positive*

ADMEXP *Likely positive since less money for programs*

REV *Likely negative, again as a measure of quality of public schools*

MEMNEA *Negative because it's a measure of quality*

INCOME *Positive, shows ability to pay*

COLLEGE *Positive, shows demand for quality education by parents*

(b) The following model (MODEL 1) shows the results for the unrestricted regression. The second model (MODEL 2), shows the results when the least significant coefficients were dropped one at a time until the remaining p-values were no greater than .10. Using the ESS from each model, perform a Wald test (with a = .05) to see if eliminating these variables is justified. Carefully state your null and alternative hypotheses and your test statistic. (15 points)

```
MODEL 1: OLS estimates using the 51 observations 1-51
Dependent variable: ENROLL

      VARIABLE      COEFFICIENT      STDERROR       T STAT    2Prob(t > |T|)

   0)    const           0.2883        0.0988        2.917       0.005651 ***
   2)   CATHOL           0.2104        0.0524        4.017       0.000239 ***
   3)    PUPIL          -0.0024        0.0032       -0.755       0.454485
   4)    WHITE          -0.1717        0.0503       -3.412       0.001436 ***
   5)   ADMEXP          -0.1420        0.1373       -1.035       0.306753
   6)      REV           0.0026        0.0113        0.228       0.820512
   7)   MEMNEA           0.0138        0.0265        0.521       0.605055
   8)   INCOME           0.0050        0.0046        1.085       0.284222
   9)  COLLEGE          -0.4920        0.2383       -2.065       0.045147 **

Mean of dep. var.             0.096  S.D. of dep. variable            0.052
Error Sum of Sq (ESS)        0.0620  Std Err of Resid. (sgmahat)     0.0384
Unadjusted R-squared          0.545  Adjusted R-squared               0.458
F-statistic (8, 42)         6.28929  p-value for F()               0.000024
Durbin-Watson stat.           2.244  First-order autocorr. coeff     -0.137

MODEL 2: OLS estimates using the 51 observations 1-51
Dependent variable: ENROLL

      VARIABLE      COEFFICIENT      STDERROR       T STAT    2Prob(t > |T|)

   0)    const           0.1688        0.0577        2.927       0.005298 ***
   2)   CATHOL           0.2221        0.0490        4.531       0.000042 ***
   4)    WHITE          -0.1520        0.0421       -3.607       0.000760 ***
   8)   INCOME           0.0068        0.0036        1.891       0.064909 *
   9)  COLLEGE          -0.5007        0.2277       -2.199       0.032923 **

Mean of dep. var.             0.096  S.D. of dep. variable            0.052
Error Sum of Sq (ESS)        0.0658  Std Err of Resid. (sgmahat)     0.0378
Unadjusted R-squared          0.518  Adjusted R-squared               0.476
F-statistic (4, 46)         12.3368  p-value for F()               0.000001
Durbin-Watson stat.           2.113  First-order autocorr. coeff     -0.062
```

Answer: Here we use a Wald test where the ESS_U =.0620, EES_R =.0658, k-m = 4 and n-k = 42. Therefore we have F = [(0.0658 – 0.0620)/4]/[0.0620/42] =.6435. Right away we can tell that the calculated F-statistic will not be significantly different from zero and therefore we cannot reject the null hypothesis that these variables jointly have zero effect on ENROLL. (Note: the critical F = 2.61)

(c) Now use the R^2 values from the above regressions and perform the same Wald test. Do your answers differ? Why or why not? Again, show all your work. (15 points)

Now the F-statistic is 0.623, but we can still reject the null. The difference is due to the rounding of the ESS in the two models.

ANSWER KEY

PART B: Short Answer Question (20 points total)

1. The following is a partial computer output table from a simple linear regression of the relationship between years of education (yearsed) and annual earnings (in thousands of dollars).

```
      Source |       SS       df       MS              Number of obs =     294
-------------+------------------------------           F(  1,   292) =   65.83
       Model |  36057.2348     1  36057.2348           Prob > F      =  0.0000
    Residual |  159947.615   292  547.765806           R-squared     =  0.1840
-------------+------------------------------           Adj R-squared =  0.1812
       Total |   196004.85   293  668.958533           Root MSE      =  23.404

------------------------------------------------------------------------------
        earn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsed |   4.602427   .5672679     8.11   0.000     3.485975    5.718879
       _cons |  -32.07989   7.906112    -4.06   0.000    -47.64008    -16.5197
------------------------------------------------------------------------------
```

Fill in the blanks on this table. Note that for a=.05 and 292 d.f., t*=1.96 on a two-tailed test. (20 points)

Answer: See above.

ANSWER KEY

PART C: True-False-Uncertain Questions (100 points total). Answer each of the following TRUE-FALSE-UNCERTAIN questions by filling in the space with a "T" if you believe the answer to be true, "F" if you believe it to be false, and "U" if you believe the answer is uncertain. Give a one-sentence explanation justifying your answer.

1. The least squares estimators of β_1 and β_2 are said to be BLUE because they are the best unbiased estimators. ______

   FALSE or UNCERTAIN . . . they are best only in the class of linear and unbiased estimators (i.e., have the smallest variance of all linear, unbiased estimators).

2. The independent variable X must be obtained from experimental data to be considered fixed (i.e., nonstochastic). _____

   FALSE . . . it is also considered fixed in repeated sampling.

3. A random variable is a variable whose value is not known until it is observed. _____

   TRUE . . . we may know the possible values of the RV, but we don't know its exact value until it is observed.

4. OLS estimators are themselves random variables. _____

   TRUE . . . since they are functions of RVs, namely the RV Y.

5. The residuals from an OLS regression can be defined as $\hat{\mu}_t = \hat{y}_t - \alpha - \beta x_t$. _____

   FALSE . . . $\hat{\mu}_t = y_t - \hat{y}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$ *is how we define the residual. In other words, the difference between that actual and fitted values. Recall, we don't know the values of* α *and* β*, only their estimated values.*

6. If we have two models with the same R^2 values but different dependent variables, then we can say that each model is equal in terms of explaining the effects of the independent variables on the dependent variable.

   FALSE . . . We can't begin to compare the two models unless the dependent variable is the same in each case.

7. The addition of independent variables to an econometric model will always increase the adjusted R^2 value of the new model.

   UNCERTAIN . . . it could increase or decrease.

8. If a redundant independent variable is included in a regression model, the variances of other regression coefficients will be unbiased but inefficient.

   TRUE . . . The variances of the other regression coefficients will be both unbiased and inefficient.

9. When near perfect multicollinearity exists between two variables, the result is that OLS estimators are biased.

   FALSE . . . OLS estimators are unbaised, but have higher standard errors and therefore lower t-statistics.

10. Multicollinearity is easily detected by looking at the correlation coefficients between independent variables.

    FALSE or UNCERTAIN . . . multicollinearity exists when there is a linear relationship between two or more independent variables. Thus, simple correlation coefficients may not be able to detect multicollinearity.