*Module 3, Week 2, Paper and Pencil Assignment 6*

Focus on the last output from your Week 6 gretl assignment to complete this Week 6 paper and pencil assignment. I’ve coped and pasted the gretl output for that last model below for your convenience. If you got something different you should go back through your gretl assignment to determine where/how you got off. If it was a simple thing like entering an incorrect variable name or something more serious like a misunderstanding of the course material!

#Build a model using the independent variables RM, AGE, TAX and PTRATIO

? ols CMEDV 0 RM AGE TAX PTRATIO --vcv

Model 8: OLS, using observations 1-374

Dependent variable: CMEDV

Coefficient std. error t-ratio p-value

----------------------------------------------------------------------------------------------------

Const −41.5580 2.48352 −16.73 3.59e-047 \*\*\*

RM 10.6594 0.308388 34.56 9.32e-118 \*\*\*

AGE −0.0282967 0.00723543 −3.911 0.0001 \*\*\*

TAX −0.00941443 0.00254816 −3.695 0.0003 \*\*\*

PTRATIO 0.168757 0.0109340 15.43 8.12e-042 \*\*\*

Mean dependent var 24.70294

Sum squared resid 5204.009

R-squared 0.801674

F(4, 369) 372.8935

Log-likelihood −1023.041

Schwarz criterion 2075.703

S.D. dependent var 8.387353

S.E. of regression 3.755397

Adjusted R-squared 0.799524

P-value(F) 3.5e-128

Akaike criterion 2056.081

Hannan-Quinn 2063.872

Covariance matrix of regression coefficients:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| const | RM | AGE | TAX | PTRATIO |  |
| 6.16786 | -0.717855 | -0.00367959 | -0.00310496 | -0.0189146 | const |
|  | 0.095103 | 2.68636e-04 | 1.81649e-04 | 0.00207286 | RM |
|  |  | 5.23514e-05 | -4.45934e-06 | 9.65437e-06 | AGE |
|  |  |  | 6.49312e-06 | 8.97430e-06 | TAX |
|  |  |  |  | 1.19552e-04 | PTRATIO |

1. You are interested in what factors are driving the home values. You specify the following linear regression model:

= home value

= average number of rooms in owner-occupied dwellings per census tract

= proportion of owner-occupied units built prior to 1940

= full-value property-tax rate per USD 10,000

= Pupil-teacher ratio by census tract

Remember from the Midterm Exam that there is a bit of misinterpretation. The data were actually acquired per census tract but in many places the documentation says “per town”.

You have a random sample of 374 census tracts and estimate the model above using least squares producing the following results. We did that in gretl already. That output is above. Standard errors for each estimated coefficient are in parentheses. All have been rounded to two decimal places. Ordinarily I would change the sign from all plusses to the correct sign if a term is negative. I’ve put negative coefficients in parentheses here so you can more directly tie the values back to your gretl output.

(2.48) (0.31) (0.01) (0.00) (0.01)

First, interpret each coefficient of the regression model.

1. The intercept value, -41.56, represents where the regression line would intercept the y-axis at x=0. The question is, “What is that home value in whole USD dollars (no cents) \_\_\_\_\_?” If you need to be sure to include the appropriate sign for this value.  (Hint: think about the axes of the 2-D regression line and what they represent.)
2. Is the home value in Question 1 a realistic home value? Yes or No?
3. How much do home values increase for an increase of an additional (one) room? Enter whole USD dollars (no cents)?
4. How much do home values increase for each year beyond 1940?
5. How much do home values increase for each 10,000 USD increase in the tax rate?
6. It makes sense that, in questions 4 and 5, home values actually decrease or have a negative slope coefficient as the age of the home increases and/or the property-tax rate increases. True or False?
   1. True
   2. False
7. Home values \_\_\_\_\_\_\_\_\_ when K-12 Pupil-teacher ratios increase. Enter either increase or decrease.
8. Analogous to simple linear regression, if the data contain substantially more data points than the number of parameters (independent variables) the R-squared value for a multivariable linear regression model indicates how well the model fits the data. Yes or No?

Keep in mind the formula given in your textbook in Section 4.8 (right before the exercises begin), i.e.

Here, or observations and

Go back to your gretl output for these values!

1. Enter the number of observations (in this last model with limited number of independent variables RM, AGE, TAX and PTRATIO).
2. Enter the number of independent variables.
3. The adjusted R-squared value is \_\_\_\_\_\_\_\_\_. (Hint: Be careful here because it seems pretty simple. I calculated it incorrectly the first time and caught that when I checked it against the gretl output!)
4. Consider an F-test to verify the overall utility of our multivariable linear regression model for home values. Based on the gretl output our P-value is 3.5 e-128 or incredibly small, very near zero. Therefore we cannot reject the null hypothesis. This test has proven that the model does not have overall utility. (Hint: if this is confusing read through your second textbook Section 4.6 including Example 4.3) True or False?
5. Given the results of an F-test verifying a multivariable linear regression model’s overall utility we can also conclude that the model is the best model that can be built. (Hint: this is covered in the second textbook same reference pages as for question 12.) True or False?
6. Conduct the appropriate test to evaluate whether or not the “major variable” in our multivariable linear regression model for home values is statistically significant. The value of the statistic for that test is \_\_\_\_\_\_\_\_\_. (Hint: this is discussed in your second textbook in Section 4.7. However, in reality you will probably never do this by hand. gretl already did this computation for you/us… What does gretl give for the associated P-value?)

If you do attempt to compute this by hand use a hypothesis test as follows:

The value of the “major variable” coefficient is: 10.66, its sd error is 0.31. The t-test statistic is given by:

1. Predict the home-value in whole USD (no cents) for a home built in 1950 with (average) 6 rooms, an “assessed” home value of 100,000 USD, and a pupil-teacher ratio of 20:1.
2. Is the result you computed for question 15 more than the mean of the dependent variable in our current multivariable linear regression model? Yes or No. (Hint: look through your gretl output!)