THE UNIVERSITY OF TENNESSEE KNOXVILLE

Economics 381, Fall 2016

Problem Set 2 – OLS and interpreting Stata

Due Date: September 6

Instructions: Complete the problem set answers in the spaces given. Take your time in crafting your answers to make sure they're precise.

Name (printed):

Some Stata output is shown below. The dependent variable is the number of thefts in different counties in Ireland, quarterly, from 2003 to 2016.

Source	SS	df	MS		
Model	3.0458e+09	5	609158283	Prob > F	0.0000
Residual	804305638	1,107	726563.359	R-squared	0.7911
				Adj R-squared	0.7902
Total	3.8501e+09	1,112	3462317.5	Root MSE	852.39

	Number of obs	=	1,113
	F(5, 1107)	=	838.41

theft_all	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
scaled_unemployment	57.46693	2.827878	20.32	0.000	51.91832	63.01554
ln_pop	1333.706	75.50478	17.66	0.000	1185.557	1481.854
_Iquarter_2	29.28673	71.64313	0.41	0.683	-111.2849	169.8584
_Iquarter_3	5.394129	71.65891	0.08	0.940	-135.2085	145.9967
_Iquarter_4	1.932871	71.65184	0.03	0.978	-138.6559	142.5216
_cons	-15989.12	877.1134	-18.23	0.000	-17710.12	-14268.13

The primary variable of interest is `scaled_unemployment`, which is the number of people (in thousands) who were unemployed in that county that quarter. The `ln_pop` variable is the natural log of the population of that county that quarter. Dummy variables for quarter 2 (Apr–Jun), quarter 3 (Jul–Sep), and quarter 4 (Oct–Dec) are also included. You should interpret those coefficients as relative to the 'excluded' category of quarter 1 (Jan–Mar).

Q1. Our regression formula, for each observation i, is:

$$Y_i = \hat{\alpha} + \hat{\beta}_1 \text{scaled\_unemp}_i + \hat{\beta}_2 \text{ln\_pop}_i + \hat{\beta}_3 \text{Q2}_i + \hat{\beta}_4 \text{Q3}_i + \hat{\beta}_4 \text{Q4}_i + \epsilon_i$$

which, after a tiny change, can be expressed as:

$$\epsilon_i = Y_i - \hat{\alpha} - \hat{\beta}_1 \text{scaled\_unemp}_i - \hat{\beta}_2 \text{ln\_pop}_i - \hat{\beta}_3 \text{Q2}_i - \hat{\beta}_4 \text{Q3}_i - \hat{\beta}_4 \text{Q4}_i$$

This is for any particular observation i. If we wanted to add the residuals up for all observations, the relevant equation would be:

$$\sum_{i=1}^{n} \epsilon_i = \sum_{i=1}^{n} Y_i - \hat{\alpha} - \hat{\beta}_1 \text{scaled\_unemp}_i - \hat{\beta}_2 \text{ln\_pop}_i - \hat{\beta}_3 \text{Q2}_i - \hat{\beta}_4 \text{Q3}_i - \hat{\beta}_4 \text{Q4}_i$$

Make one further adjustment to both sides of this equation so that it represents the problem OLS solves.

Q2. How many observations (n) are there included in this regression?

Q3. Including the constant, how many explanatory variables (k) are included in this regression?

Q4. In the top left-hand corner of the output is a column entitled `SS`. This stands for "sum of squares". The first number is labelled 'Model (SS)', and starts with 3.0458. The third number is labelled 'Total (SS)', and starts with 3.8501. Divide 3.0458 by 3.8501, and report your answer to four decimal points.

Q5. Report and interpret the R^2 for this regression.

Q6. In one sentence, precisely interpret the coefficient on `scaled_unemp`.

Q7. What would you expect to happen to the point estimate of `scaled_unemp` if the sample size doubled?

Q8. What would you expect to happen to the standard error of the estimate of `scaled_unemp` if the sample size doubled?

Q9. In one sentence, precisely interpret the coefficient on `ln_pop`.

Q10. In one sentence, precisely interpret the coefficient on `_Iquarter_3`.

Q11. Is it fair to conclude from this analysis that unemployment causes crime?