Econ 300 Spring 2020
Problem set 3

Instructions: Please write/type your answers and hand in a hard copy at the beginning of the class. The due date of the problem set is 03/02/2020. Total points available: 115 (Graded out of 100 points).

For STATA questions, submit the codes and prints of the screenshots of the results window.

1. (10 points) A researcher estimated the following model:

$$\log(earnings) = \beta_0 + \beta_1 age + \beta_2 educ + u$$

The output is given below:

```
. reg log_earnings age educ
```

Source	SS	df	MS	Number of obs	=	17,870
				F(2, 17867)	=	1812.75
Model	1320.8859	2	660.442948	Prob > F	=	0.0000
Residual	6509.53444	17,867	.364332817	R-squared	=	0.1687
				Adj R-squared	=	0.1686
Total	7830.42034	17,869	.438212566	Root MSE	=	.6036

log_earnings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0071698	.0004505	15.91	0.000	.0062868	.0080529
educ	.1007288	.0017117	58.85	0.000	.0973737	.1040838
_cons	8.904056	.0307154	289.89	0.000	8.84385	8.964261

(a) Interpret the coefficient of age. Make sure to use right units.

(b) Interpret the coefficient of education. Make sure to use right units.

2. (15 points) A researcher estimated the following model:

$$earnings = \beta_0 + \beta_1 \log(height) + u$$

The output is given below:

```
. reg earnings log_height
```

Source	SS	df	MS	Number of obs	=	17,870
				F(1, 17868)	=	195.97
Model	1.4052e+11	1	1.4052e+11	Prob > F	=	0.0000
Residual	1.2812e+13	17,868	717040022	R-squared	=	0.0108
				Adj R-squared	=	0.0108
Total	1.2953e+13	17,869	724863544	Root MSE	=	26778

earnings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_height	47363.66	3383.411	14.00	0.000	40731.85	53995.47
_cons	-152165.4	14219.83	-10.70	0.000	-180037.6	-124293.1

(a) Interpret the coefficient of log height. Make sure to use right units.

Another researcher suggested that probably below model is more useful.

$$\log(earnings) = \beta_0 + \beta_1 \log(height) + u$$

(b) (**BONUS)** When is this model more useful?

```
. reg log_earnings log_height
```

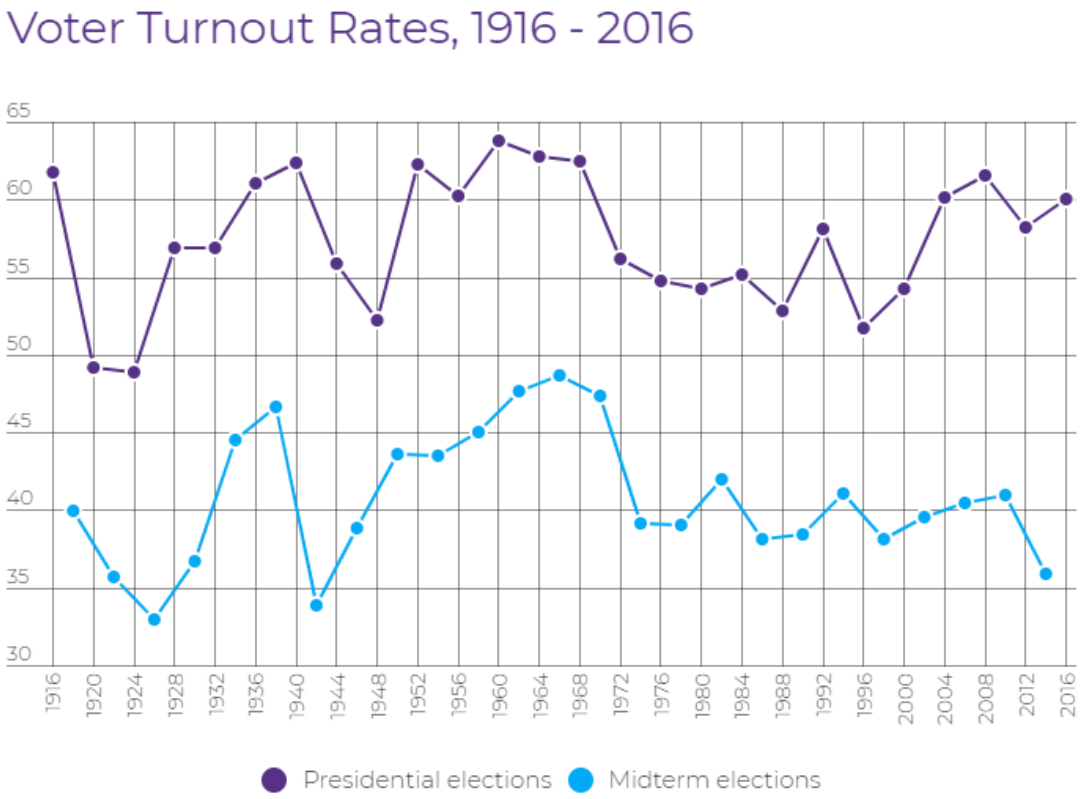
Source	SS	df	MS	Number of obs	=	17,870
				F(1, 17868)	=	243.18
Model	105.140281	1	105.140281	Prob > F	=	0.0000
Residual	7725.28006	17,868	.432352813	R-squared	=	0.0134
				Adj R-squared	=	0.0134
Total	7830.42034	17,869	.438212566	Root MSE	=	.65754

log_earnings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_height	1.29559	.0830811	15.59	0.000	1.132743	1.458437
_cons	5.116462	.349174	14.65	0.000	4.432047	5.800876

(c) The results are in the above table. Interpret the coefficient of log height. Make sure to use right units.

3. (10 points) This figure shows the history of voter turnout rates. The voter turnout rate for presidential elections was 63.8% in 1960 and 54.2% in 2000.
   a) Explain the changes between 1960 and 2000 in percentage points

   b) Explain the changes between 1960 and 2000 in percentage terms.



4. (15 points) Suppose you're interested in estimating the effect of family income on health status. Your original model is given as follows: health index $= \beta_0 + \beta_1$family income $+ u$

   (a) How would you change the model if you think that the effect of family income on health can be different for immigrant and non-immigrant?

   (b) In the modified model in (a), how would you test whether there is a differential effect of family income on health status depending on immigrant status? State the null hypothesis and alternative hypothesis.

   (c) How would you change the model if you think that the effect of family income on health can be different for people with different ages?

(d) In the modified model in (c), how would you test whether there is a differential effect of family income on health status depending on ages? State the null hypothesis and alternative hypothesis.

TABLE 8.3 Nonlinear Regression Models of Test Scores

Dependent variable: average test score in district; 420 observations.

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Student–teacher ratio (STR)	−1.00** (0.27)	−0.73** (0.26)	−0.97 (0.59)	−0.53 (0.34)	64.33** (24.86)	83.70** (28.50)	65.29** (25.26)
STR^2					−3.42** (1.25)	−4.38** (1.44)	−3.47** (1.27)
STR^3					0.059** (0.021)	0.075** (0.024)	0.060** (0.021)
% English learners	−0.122** (0.033)	−0.176** (0.034)					−0.166** (0.034)
% English learners ≥ 10%? (Binary, $HiEL$)			5.64 (19.51)	5.50 (9.80)	−5.47** (1.03)	816.1* (327.7)	
$HiEL \times STR$			−1.28 (0.97)	−0.58 (0.50)		−123.3* (50.2)	
$HiEL \times STR^2$						6.12* (2.54)	
$HiEL \times STR^3$						−0.101* (0.043)	
% Eligible for subsidized lunch	−0.547** (0.024)	−0.398** (0.033)		−0.411** (0.029)	−0.420** (0.029)	−0.418** (0.029)	−0.402** (0.033)
Average district income (logarithm)		11.57** (1.81)		12.12** (1.80)	11.75** (1.78)	11.80** (1.78)	11.51** (1.81)
Intercept	700.2** (5.6)	658.6** (8.6)	682.2** (11.9)	653.6** (9.9)	252.0 (163.6)	122.3 (185.5)	244.8 (165.7)

5. (20 points) Suppose we have the above table. Note that HiEL is defined as a binary variable that takes a value 1 if there are more than 10% of English learners.
   a) How would you interpret the coefficient of HiEL*STR in regression in column (3)?

   b) Is the hypothesis that the effect of student-teacher ratio is different depending on the share of English learners supported from the data?

c) How would you test this hypothesis? State the null hypothesis and explain how you made the decision.

d) What is the difference between regression (3) and (4)? (4) reduces concerns on which type of bias. Why?

6. (45 points) [STATA] Using the dataset TeachingRatings to carry out the following exercises.
    a. Estimate a regression of Course_Eval on Beauty, Intro, OneCredit, Female, Minority, and NNEnglish.
    b. Modify the regression in (a) so that the effect of Beauty on Course_Eval is different for men and women. Is the male-female difference in the effect of Beauty statically significant?
    c. Draw the scatter plot and fitted lines for the relationship between Beauty and Course-Eval, both for men and women (on the same graph).
    d. (15 points) Professor Smith is a man. He has cosmetic surgery that increases his beauty index from one standard deviation below and the average to one standard deviation above the average. What is his value of Beauty before surgery? After the surgery? Using the regression in (b), predict the increase in his course evaluation.
    e. (15 points) Repeat (d) for Professor Jones, who is a woman.

- Useful STATA commands
    - To generate an interaction term: generate *variablename* = **expression**
    - To plot scatter and fitted lines: twoway (scatter y x) (lfit y x)
    - If you want to give a condition: if expression, Ex. tw (scatter y x if z==1) (lfit y x if z==0)