Homework 5 Answer Key

Problem 1:

A)

For part A we must use an F-test to test the joint significance of *brthord* and *sibs*.

H_0: $\beta_{sibs}=\beta_{brthord}=0$
H_A: H_0 is false

Formula for an F-stat: $((SSR_{restricted}-SSR_{unrestricted})/q)$ / $(SSR_{unrestricted}/(n-k-1))$
Plugging in numbers from the two regressions below we get:
$F_{stat} = ((123.6-122.7)/2) / (122.7/846) = 3.1$
Our table tells us that $F_{crit} = 3$
Our F-stat is greater than the critical F value so we may reject H_0.

Output for basic model:

Source	SS	df	MS
Model	23.2137901	3	7.73793002
Residual	123.602285	848	.145757411
Total	146.816075	851	.172521827

Number of obs	= 852
F(3, 848)	= 53.09
Prob > F	= 0.0000
R-squared	= 0.1581
Adj R-squared	= 0.1551
Root MSE	= .38178

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0750167	.0067488	11.12	0.000	.0617703	.088263
exper	.0137138	.0034858	3.93	0.000	.006872	.0205556
tenure	.0130759	.0026442	4.95	0.000	.0078861	.0182658
_cons	5.527082	.1151851	47.98	0.000	5.301001	5.753163

Output for extended model:

Source	SS	df	MS
Model	24.0698217	5	4.81396434
Residual	122.746253	846	.145090134
Total	146.816075	851	.172521827

Number of obs	= 852
F(5, 846)	= 33.18
Prob > F	= 0.0000
R-squared	= 0.1639
Adj R-squared	= 0.1590
Root MSE	= .38091

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0713002	.0069163	10.31	0.000	.0577251	.0848753
exper	.013549	.0034822	3.89	0.000	.0067143	.0203838
tenure	.0128036	.0026407	4.85	0.000	.0076206	.0179866
brthord	-.0130115	.0102188	-1.27	0.203	-.0330686	.0070457
sibs	-.006911	.0071817	-0.96	0.336	-.0210069	.007185
_cons	5.630831	.1227862	45.86	0.000	5.38983	5.871832

B)

Our basic model is not nested within the regression estimated below. Therefore an F-test would be inappropriate. Instead, we use Adjusted R^2 to compare the models. The basic model's Adjusted R^2 is greater so we prefer the basic logged model.

Output for extended model using wage:

Source	SS	df	MS
Model	21152128.5	5	4230425.7
Residual	117635777	846	139049.382
Total	138787905	851	163088.02

Number of obs	852
F(5, 846)	30.42
Prob > F	0.0000
R-squared	0.1524
Adj R-squared	0.1474
Root MSE	372.89

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	70.64326	6.770777	10.43	0.000	57.35376	83.93275
exper	13.25727	3.408938	3.89	0.000	6.566298	19.94823
tenure	7.644611	2.585103	2.96	0.003	2.570643	12.71858
brthord	-12.47506	10.0038	-1.25	0.213	-32.11025	7.160127
sibs	-8.107592	7.030565	-1.15	0.249	-21.90699	5.691806
_cons	-141.006	120.2029	-1.17	0.241	-376.9369	94.92501

C)

This question asks us to perform a full exclusionary F-test on the basic model.

H_0: $\beta_{educ}=\beta_{exper}=\beta_{tenure}=0$
H_A: H_0 is false

Stata has already printed out the F-stat for this test (53), which s much larger than our critical F-value of 3. We can reject H_0 because our F-stat is greater than the critical F-value.

D)

To test whether returns to education are affected by experience, we must add an interaction term to our regression: gen interaction = educ*exper. We then add this new variable to our basic model and test the significance of its coefficient. As we can see in the printout below 0 is not contained in the 95% confidence interval for $\beta_{interaction}$ so we can reject the null hypothesis that $\beta_{interaction}=0$.

Basic Model with interaction effect:

Source	SS	df	MS
Model	24.170512	4	6.04262799
Residual	122.645563	847	.144799956
Total	146.816075	851	.172521827

Number of obs	852
F(4, 847)	41.73
Prob > F	0.0000
R-squared	0.1646
Adj R-squared	0.1607
Root MSE	.38053

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0329329	.0177002	1.86	0.063	-.0018084	.0676742
exper	-.0378396	.0203549	-1.86	0.063	-.0777916	.0021124
tenure	.0134432	.0026393	5.09	0.000	.0082628	.0186236

interaction	.0039761	.0015469	2.57	0.010	.00094	.0070122
_cons	6.085645	.2457654	24.76	0.000	5.603265	6.568026

Problem 2

Using the Bwght.dta dataset from the website, we wish to predict the probability of smoking by expectant mothers. That is, we wish to test the following:

$$Smoke = \beta_0 + \beta_1 meduc + \beta_2 feduc + \beta_3 cigprice + u$$

where *Smoke* equals 1 if the mother smokes, and zero otherwise.

a. **Generate the variable, Smoke.**

. gen smoke=0

. replace smoke=1 if cigs>0

b. **Do cigarette prices affect the probability of smoking? Test this hypothesis, and interpret the parameter of interest.**

```
. reg smoke  motheduc fatheduc cigprice, robust

Linear regression                                      Number of obs =       1191
                                                           F(  3,  1187) =    24.34
                                                           Prob > F      =   0.0000
                                                           R-squared     =   0.0582
                                                           Root MSE      =   .33237
```

smoke	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
motheduc	-.0270209	.0052752	-5.12	0.000	-.0373706	-.0166712
fatheduc	-.0085852	.004566	-1.88	0.060	-.0175435	.0003731
cigprice	.0012866	.0009659	1.33	0.183	-.0006085	.0031818
_cons	.4349079	.1357851	3.20	0.001	.1685024	.7013134

A one-unit increase in cigarette price increases the probability of smoking, on average, by 0.0012 or 0.12%, holding other factors fixed. But this effect is counterintuitive and insignificant with a p-value as high as 0.183. Or looking at the 95% confidence level, it contains zero, which means that we can't reject the null that $\beta_3 = 0$.

c. **What is a potential problem in running this type of regression? Do we have this problem in this particular case?**

The biggest problem with Linear Probability Model is the possible unrealistic predictions. The predicted probability could be either less than zero or greater than one. We do have this problem in this particular case. For example, with a combination of cigprice (122.3), fatheduc(16) and motheduc(18), the probability of smoking will be -0.0314.