PUBLIC POLICY 639: QUIZ 3 SOLUTIONS
Winter 2011

On pages 3 and 4 you will see variable definitions, summary statistics, and regression output from regressions of an infant's one-minute APGAR score (a general measure of health at birth) on mother's age and other controls.

1. (1 point) Interpret the coefficient on mother's age in column (2). Be specific and precise.

Each additional year of mother's age is associated with a 0.0052 point increase in APGAR scores, holding parental education constant.

2. (2 points) The coefficient on mother's age changes considerably between column (1) and column (2) and the R^2 (and adjusted R^2) increases. Explain why. Be specific and use the framework we discussed in class.

The regression in column (1) was suffering from omitted variable bias. Socioeconomic status (as measured by parental education) is positively related to age and to APGAR, so that failing to control for it causes us to overstate the relationship between mother's age and APGAR.

3. (3 points). Column (3) controls for the number of prenatal visits the parents made. If you knew from some other data source that older mothers tend to have more prenatal visits than younger ones (even after controlling for education), should the sign of the coefficient on # prenatal visits in column (3) be positive or negative? Explain.

   [Note: 1 point will be awarded for your answer of (+) or (-) and 2 points will be awarded for your explanation].

(+)

Since the coefficient on age decreases when we control for the number of prenatal visits, this means that regression (2) was suffering from positive OVB. If age and the number of visits are positively correlated, then visits and APGAR should also be positively correlated for the # of visits to create positive OVB.

4. (2 points) You are asked to test the hypotheses that there is a relationship between parental education (mother's education and father's education) and one-minute APGAR score, holding mother's age constant.

   a. Which regression in the table answers this question? Write out the <u>population</u> regression function corresponding to this regression. (0.5 point)

Regression 2

$$omaps_i = \beta_0 + \beta_1 \times mage_i + \beta_2 \times meduc_i + \beta_3 \times feduc_i + u$$

   b. How would you test this hypothesis? Note that a full answer will specify the null and alternative hypotheses, describe the test statistic and its critical value (full equation for the F statistic not needed), and interpret the results of the test. (1.5 points).

Null hypothesis : $H_0: \beta_2 = 0$ *and* $\beta_3 = 0$
Alternative hypothesis : $H_a: \beta_2 \neq 0$ *or* $\beta_3 \neq 0$

We would use a F-statistic to test this hypothesis and compare it to its critical value at the 5% level of significance, which is 3.0 since there are two restrictions tested. In Stata, we would use the following command right after running the regression.

```
test meduc feduc
```

It would give us the value of the F-statistic and the corresponding p-value.

5. (0.5 point) Under what assumptions is it appropriate to compute a valid F-statistic by comparing the R^2 from regressions with and without the covariates of interest?

When we assume the errors are homoscedastic we can utilize this technique.

6. (1.5 points) Interpret the R^2 and adjusted R^2 values from column (2). Additionally, explain when you should and should not expect to observe differences between these two statistics generally.

Model 2 has an R^2 value of 0.006, thus 0.6% of variation in infant's APGAR is accounted for by the model 2 regressors – specifically mother's age, mother's education, and father's education.

Model 2 has an adjusted R^2 value of 0.004, thus 0.4% of variation in infant's APGAR is accounted for by the model 2 regressors – specifically mother's age, mother's education, and father's education.

Given that the adjusted R^2 is a modified version of R^2 that does not necessarily increase when a new regressor is added, we shouldn't observe large differences in these values when sample sizes are large and when the number of regressors is small.

STATA OUTPUT

Variable Description

```
. desc omaps mage meduc feduc npvis;

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
omaps           byte   %10.0g                 one minute apgar score
mage            byte   %10.0g                 mother's age, years
meduc           byte   %10.0g                 mother's educ, years
feduc           byte   %10.0g                 father's educ, years
npvis           byte   %10.0g                 total number of prenatal visits
```

Summary Statistics

```
. summ omaps mage meduc feduc npvis;

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       omaps |       1728    8.394676    1.109738         0         10
        mage |       1728    29.53125    4.777645        16         44
       meduc |       1728    13.73958    2.102858         3         17
       feduc |       1728    13.93171    2.272407         3         17
       npvis |       1728    11.64468    3.652686         0         40
```

Regression output

Table 1. Results of OLS Regression of Infant Health on Mother's Age and Various Controls

	Dependent variable: One-minute APGAR Score		
Regressor	(1)	(2)	(3)
Mother's age (years)	0.0103**	0.0052	0.0042
	(0.005)	(0.006)	(0.006)
Mother's education (years)		0.0165	0.0147
		(0.017)	(0.016)
Father's education (years)		0.0220	0.0199
		(0.015)	(0.0151)
# Prenatal visits			
Constant	8.091***	7.708***	7.562***
	(0.1570)	(0.2250)	(0.2400)
Observations	1728	1728	1728
R-squared	0.002	0.006	0.01
Adj R-squared	0.001	0.004	0.008

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data comes from Jeff Wooldridge.

F-statistic critical values

$q \rightarrow$	2	3	4	5	6	7	8	9	10
5% critical value	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83
1% critical value	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32