PUBLIC POLICY 639: QUIZ 2 SOLUTIONS
Winter 2011

Bivariate Regression

On page 3 you will see variable definitions, summary statistics, and regression output from a regression of annual earnings on months spent in job training. The data is from a sample of participants in a federal job training program run in the late 1970s.

1. (1 point) Write the formula for the sample regression that we have run.

$earnings_i = \hat{\beta}_0 + \hat{\beta}_1 MonthsTraining_i + \hat{u}_i$ OR $\widehat{earnings_i} = \hat{\beta}_0 + \hat{\beta}_1 MonthsTraining_i$

2. (1 point) What is the value of $\hat{\beta}_1$ in the regression we have run? Interpret $\hat{\beta}_1$ in words. [Note that earnings are in $1,000]

One month in training is associated with a $243 increase in earnings. $\hat{\beta}_1 = 0.24$

3. (1 point) What is the value of $\hat{\beta}_0$ in the regression we have run? Interpret $\hat{\beta}_0$ in words. [Note that earnings are in $1000]

Based on our estimates, individuals with zero training are predicted to have earnings of $1,863. $\hat{\beta}_1 = 1.86$

4. (1 point) What is the predicted annual earnings for someone who participates in a ten-month job training program?

$\widehat{earnings_i} = \hat{\beta}_0 + \hat{\beta}_1 MonthsTraining_i$

$\widehat{earnings_i} = 1.86 + (.24) \times (10) = 1.86 + 2.43$ = *$4,289 or $4,290 depending on how your rounded*

5. (3 points) Test the hypothesis that there is a population relationship between months spent in training and annual earnings. Clearly state your null and alternative hypotheses, test statistic, and conclusion (in words).

H_0: $\beta_1 = 0$ H_a: $\beta_1 \neq 0$ (1 point) $t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{0.24}{0.11} = 2.16$ (1 point)

Since t = 2.16 with a p-value = .032, we reject the null hypothesis at an alpha level of .05. (1 point)

Conceptual

6. (1 point) What does the OLS estimator minimize?

These are all correct:

The sum of squared residuals (this is the most common usage)

The sum of squared mistakes (found in book, shouldn't use this term as no one uses it)

The sum of the squared errors

$$\sum \hat{\mu}_i^2 \quad OR \quad \sum \left(Y_i - \hat{Y}_i\right)^2$$

7. (1 point) Why would a researcher consider using the "robust" regression option in Stata?

The robust regression option in Stata is used if the variance of μ is dependent on a given level of X. If there is evidence of heteroskedasticity, then the robust option provides the heteroskedasticity-robust standard error of $\hat{\beta}_1$. *By not using the robust option, the estimated standard errors (using the homoskedasticity-only formula) are most likely wrong due to the violation of the OLS assumption.*

8. (1 point) You see we used the "robust" option in our regression. If we had forgotten to do so, what might we get wrong? Circle all that apply.

a. $\hat{\beta}_1$ - *No*

b. $SE\left(\hat{\beta}_1\right)$ - *Yes, might get this wrong*

c. Confidence interval for $\hat{\beta}_1$ **-** *Yes, might get this wrong*

d. Predicted test score for each of the data points - *No*

e. t-test for the null hypotheses that $\hat{\beta}_1 = 0$ **-** *Yes, might get this wrong*

Failing to use the robust option will only affect things related to the standard errors (e.g. confidence intervals, test statistics). The coefficient estimates (or predicted values) will not be affected.

STATA OUTPUT

Variable Description

```
. desc re78 mostrn

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
re78            float  %9.0g                  real earns., 1978, $1000s
mostrn          byte   %9.0g                  months in training
```

Summary Statistics

```
. sum re78 mostrn

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        re78 |       185    6.349145    7.867405          0    60.3079
      mostrn |       185    18.49189    4.911299          6         24
```

Regression output

```
. reg re78 mostrn, robust

Linear regression                                      Number of obs =     185
                                                       F(  1,   183) =    4.66
                                                       Prob > F      =  0.0322
                                                       R-squared     =  0.0229
                                                       Root MSE      =  7.7979

------------------------------------------------------------------------------
             |               Robust
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      mostrn |   .2426028   .1123987     2.16   0.032     .0208388    .4643668
       _cons |    1.86296   1.907693     0.98   0.330     -1.90094     5.62686
------------------------------------------------------------------------------
```