Econ 140 - Spring 2016
Section 5

GSI: Fenella Carpena

February 18, 2016

This note explains how to read Stata output. In this example, we consider a regression of wages on education. The variables are *educ* (years of schooling) and *wage* (hourly wage measured in 1976 dollars).

```
. reg wage educ

      Source |       SS       df       MS              Number of obs =    3010
-------------+------------------------------           F(  1,  3008) =  301.64
       Model |  18962671.5     1  18962671.5           Prob > F      =  0.0000
    Residual |   189100858  3008  62865.9769           R-squared     =  0.0911
-------------+------------------------------           Adj R-squared =  0.0908
       Total |   208063530  3009  69147.0688           Root MSE      =  250.73

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   29.65544   1.707507    17.37   0.000     26.30744    33.00344
       _cons |   183.9487   23.10395     7.96   0.000     138.6476    229.2499
------------------------------------------------------------------------------
```

- The command `reg wage educ` tells us that this is a regression of the variable `wage` on the variable `educ`. So the regression model we have is $wage_i = \beta_0 + \beta_1 * educ_i + u_i$
- Top right: contains information about model fit
  - `Number of obs` tells us that N = 3010.
  - `F(1, 3008)` tells us that the $F-statistic$, which we will cover in the multivariate linear regression model, Chapter 8.
  - `Prob > F` is the p-value associated with the above $F - statistic$. This is again covered in the multivariate linear regression model, Chapter 8.
  - `R-squared` is the R-squared of the regression. Since it is 0.0911, this tells us that the variable `educ` explains about 9.1 percent of the variation in `wage`.
  - `Adj R-squared` is the adjusted R-squared, which will be covered in the multivariate linear regression model, chapter 7.
  - `Root MSE` is the root mean-squared error, i.e. it is the sample standard deviation of the error term. So in this class, the Root MSE is the same as the SER. Note that SER $= \sqrt{\frac{SSR}{n-2}} = \sqrt{\frac{189100858}{3010-2}} =$ 250.73, as we have talked about in class.
- Top left table: contains information about TSS, ESS, and SER
  - The `Source` column breaks down the variance of the outcome variable into 3 sources: Model, Residual, and Total. The Total variance is partitioned into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Residual).

- The `SS` column means "sum of squares." So TSS = 208063530, ESS = 18962671.5, SSR = 189100858. Note that TSS = ESS + SSR as we have shown in section.
- The `df` column means "degrees of freedom." The model has 1 degree of freedom since we have 1 regressor, and the residual has $n-2$ degrees of freedom since we are estimating two parameters (β_0 and β_1).
- The `MS` column means "mean squares," it is the sum of squares divided by their respective degrees of freedom.

• Bottom table: has the parameters estimated in the regression

- The first line on the top left corner of the bottom table tells us that `wage` is outcome variable in this regression.
- $\hat{\beta}_0 = 183.9487$, $\hat{\beta}_1 = 29.65544$, $SE(\hat{\beta}_0) = 23.10395$, $SE(\hat{\beta}_1) = 1.707507$.
- The column with `t` is the t-statistic for the two-sided hypothesis test that the true population value is 0. So, for $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$, the t-statistic $= \frac{29.65544}{1.707507} = 17.37$. Similarly, for $H_0 : \beta_0 = 0$, $H_1 : \beta_0 \neq 0$, the t-statistic $= \frac{183.9487}{23.10395} = 7.96$.
- The column with $P > |t|$ is the p-value for the above test. So the p-value for testing $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$ is 0.000. And the p-value for testing $H_0 : \beta_0 = 0$, $H_0 : \beta_1 \neq 0$ is also 0.000.
- The last column gives us the 95% confidence interval, i.e., $\hat{\beta}_1 \pm 1.96 * SE(\hat{\beta}_1) = (26.30744, 33.00344)$. Similarly for $\hat{\beta}_0$.