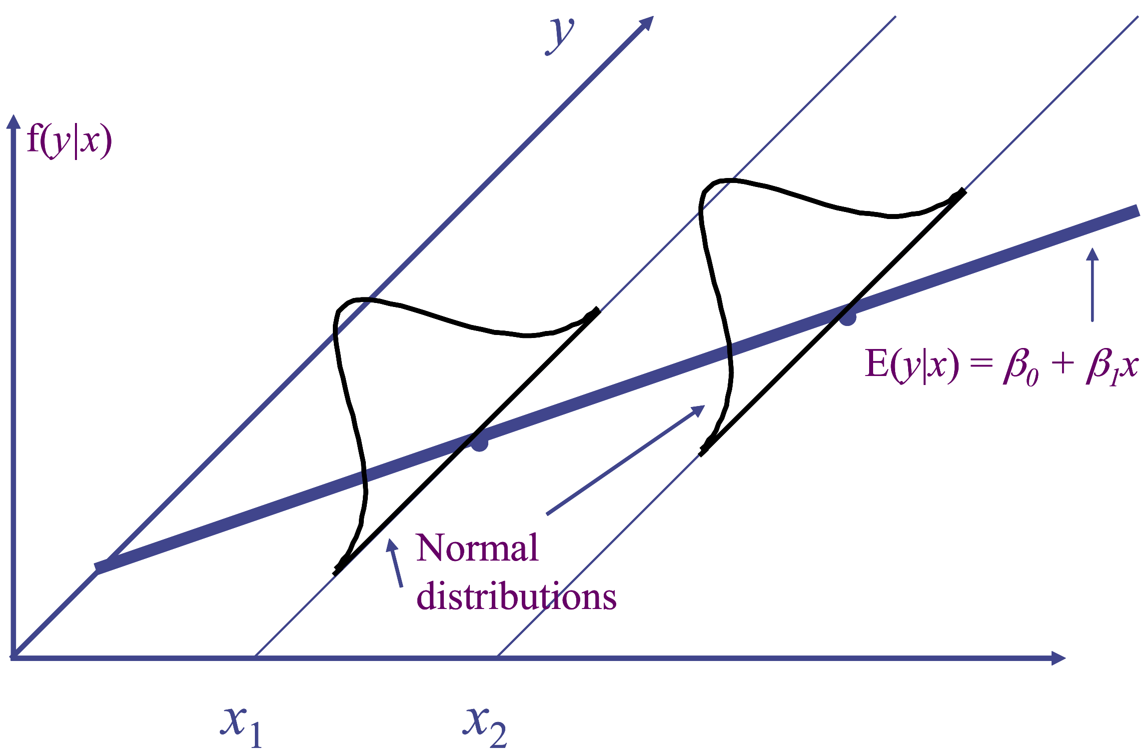
PPOL502-01, Spring 2016
Course Notes #7: Normality Assumption for *u*, Variance of OLS Estimates, One-Tail Tests, Confidence Intervals for Regression Coefficients and Predictions

---

So far, we've focused on the point estimates of regression coefficients, $\hat{\beta}_j$. We talked about statistical significance for these estimates in the simple regression case, but we haven't addressed the issue again in multiple regression. We do so now.

I. NORMALITY ASSUMPTION FOR *u* and TESTING FOR STAT SIG OF COEFF

- **Assumption MLR6.** *(normality of the error term u, conditional on **X**)*.
  The population error u is independent of the explanatory variables **X** and is normally distributed with zero mean and variance σ^2. That is, $u \sim N(0, \sigma^2)$,



- Adding MLR6 to assumptions MLR1 through MLR5 (the Gauss-Markov assumptions) forms a set of assumptions know as the **classical linear model (CLM)** assumptions, and expands OLS to a class of **minimum variance unbiased estimators** (i.e., they are not just the "best" among linear estimators in this sense).

- With this normality assumption for the error term, if we could know σ^2 with certainty, we could use *Z*-scores to test hypotheses about regression coefficients. However, this is seldom if ever the case. Instead, we have to estimate σ^2 with $\hat{\sigma}^2$.

- Now, under the CLM assumptions (MLR1 through MLR6):

$$\frac{(\hat{\beta}_j - \beta_j)}{s.e._{\hat{\beta}_j}} \sim t_{n-k-1}$$ where *k* is the number of *X*s in the model.

  Where
  H$_0$: $\beta_j = 0$
  H$_1$: $\beta_j \neq 0$

- The standard error of <u>each coefficient</u> is calculated as: $\sqrt{s.e._{\hat{\beta}_j}^2} = \sqrt{Var(\hat{\beta}_j)} = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}$

  where $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - k - 1}$

$$SST_j = \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2$$

  R_j^2 = the R-squared from the regression of X_j—where *j* indicates the *X* variable of interest—on **all the *other Xs*** in the original model of interest (i.e., all the *Xs* in the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_k X_k + u$)

Example:

ORIGINAL MODEL: Dependent Variable: CRIME1

```
. *** Original Model
. regress crime1 police criminal

      Source |       SS       df       MS              Number of obs =       9
-------------+------------------------------           F(  2,     6) =  880.19
       Model |  59.5431697     2  29.7715849           Prob > F      =  0.0000
    Residual |  .202943222     6   .03382387           R-squared     =  0.9966
-------------+------------------------------           Adj R-squared =  0.9955
       Total |  59.7461129     8  7.46826412           Root MSE      =  .18391

------------------------------------------------------------------------------
      crime1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      police |  -1.014086    .075082   -13.51   0.000    -1.197805   -.8303668
    criminal |   2.008263   .0791434    25.37   0.000     1.814606     2.20192
       _cons |  -.0192298   .1393464    -0.14   0.895    -.3601981    .3217384
------------------------------------------------------------------------------
```

How is the standard error on POLICE obtained?

1. calculate $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-k-1} = \dfrac{\sum (Y_i - \hat{Y}_i)^2}{n-k-1} = \dfrac{0.20294}{6} = 0.03382$

2. calculate $SST_j = \sum_{i=1}^{n} (X_{ij} - \overline{X}_j)^2 = 60$ (obtained from regression output below)

3. calculate $(1 - R_j^2) = 1 - 0.90 = 0.10$ (obtained from regression output below)

   where SST_j and R_j^2 are obtained from:

Dependent Variable: POLICE

```
. regress police criminal

      Source |       SS       df       MS              Number of obs =       9
-------------+------------------------------           F(  1,     7) =   63.00
       Model |          54     1          54           Prob > F      =  0.0001
    Residual |           6     7  .857142857           R-squared     =  0.9000
-------------+------------------------------           Adj R-squared =  0.8857
       Total |          60     8         7.5           Root MSE      =  .92582

------------------------------------------------------------------------------
      police |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    criminal |          1   .1259882     7.94   0.000     .7020853    1.297915
       _cons |          0   .7014724     0.00   1.000    -1.658719    1.658719
------------------------------------------------------------------------------
```

4. Put the pieces together:

$$s.e.^2_{\hat{\beta}_j} = Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1-R_j^2)} = \frac{.03382}{60*0.10} = \frac{0.03382}{6} = 0.005636667$$

5. To get the standard error, take the square root of the variance:

$$s.e._{\hat{\beta}_j} = \sqrt{s.e.^2_{\hat{\beta}_j}} = \sqrt{0.005636667} = 0.07508$$

ANOTHER EXAMPLE:

- Consider the model: Wages = *f*(IQ, Educ, KWW, motheduc, fatheduc) where:

```
  wage                         hourly wage in cents, 1976
  IQ                           IQ score
  KWW                          knowledge world of work score
  educ                         years of schooling, 1976
  fatheduc                     father's schooling
  motheduc                     mother's schooling


. summarize wage IQ educ KWW motheduc fatheduc, separator(0)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |      3010    577.2824    262.9583        100       2404
          IQ |      2061    102.4498    15.42376         50        149
        educ |      3010    13.26346    2.676913          1         18
         KWW |      2963    33.54067    8.611619          4         56
    motheduc |      2657    10.34814    3.179671          0         18
    fatheduc |      2320    10.00345    3.720737          0         18
```

Dependent Variable: wage

```
. regress wage IQ educ KWW motheduc fatheduc

      Source |       SS       df       MS              Number of obs =    1604
-------------+------------------------------           F(  5,  1598) =   38.56
       Model |  12214605.9     5  2442921.17           Prob > F      =  0.0000
    Residual |   101247813  1598  63359.0817           R-squared     =  0.1077
-------------+------------------------------           Adj R-squared =  0.1049
       Total |   113462418  1603  70781.2966           Root MSE      =  251.71

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |  -.1159857    .508114    -0.23   0.819    -1.112626    .8806542
        educ |   12.89259   3.431463     3.76   0.000     6.161945    19.62323
         KWW |   9.563279   .9609077     9.95   0.000     7.678507    11.44805
    motheduc |   1.835333   2.770523     0.66   0.508    -3.598907    7.269574
    fatheduc |   .2444579   2.367129     0.10   0.918    -4.398547    4.887463
       _cons |    82.9006   50.01291     1.66   0.098    -15.19721    180.9984
------------------------------------------------------------------------------
```

- Reminders:

  (1) the null hypothesis is H_0: $\beta_{fatheduc} = 0$

  it is **not** H_0: $\hat{\beta}_{fatheduc} = 0$

  Why is the second statement incorrect?

  (2) *p-values* are often useful and more informative than "reject" or "fail to reject" conclusions from a classical hypothesis test. Instead of saying that an observed value of the test statistic is significant or is not significant, many writers in the research literature prefer to report the exact probability of getting a value as extreme or more extreme (in absolute value for a two-sided test) than that observed if the null hypothesis is true. For example, p=.0002 associated with the t-value of 3.76 for the *educ* coefficient above means that the probability of getting a value as extreme as t = 3.76 in either direction (i.e., >3.76 or <-3.76) when the null hypothesis is true, is .0002 (.0001+.0001). The quantity p is referred to as the p-value of the test. The p-value for a test is the smallest value of alpha (probability of a type I error) for which the null hypothesis can be rejected.

  (3) For sufficiently large sample sizes, the Normal distribution can be used instead of the *t* distribution. E.g., in the example on p. 125, Wooldridge says: "Since we have 522 degrees of freedom, we can use the standard normal critical values."

  Why is this the case?

(4) There's a difference between statistical and substantive (or economic, or practical) significance.

How do you interpret statistical significance?

How do you interpret substantive significance?

What drives statistical significance?

What drives substantive significance?

II. TALKING ABOUT STATISTICAL SIGNIFICANCE OF COEFFICIENTS (OR ANYTHING ELSE)

Example 1: Rejecting the null hypothesis

- Above, in the education coefficient example, we tested the null hypothesis that H_0: $\beta_{educ} = 0$. We rejected this null hypothesis, with a p-value of 0.0002 (i.e., at the $\alpha = 0.0002$ level).

- The precise way of stating this result is to say:

  Reject the null hypothesis that there is no influence of education on wages (p = 0.0002), holding constant IQ, KWW, and mother's and father's education.

  It is unlikely that education has no effect on wages, holding the other variables in the model constant.

- The more informal way of stating this is:

  Education has an effect on wages, holding the other variables in the model constant.

Example 2: Failing to reject the null hypothesis

- How about testing the null hypothesis that H_0: $\beta_{fatheduc} = 0$. We fail to reject this null hypothesis (p= 0.9178)

- The precise way of stating this result is to say:

Fail to Reject the null hypothesis that there is no influence of father's education on wages (p = 0.9178), holding constant IQ, education, KWW, and mother's education.

Given the probability of a Type I error we are willing to accept (usually .05), we **can** conclude that father's education has no effect on wages, holding the other variables in the model constant.

- A more informal way of stating this is:

  Father's education has no effect on wages, holding the other variables in the model constant

- This informal interpretation actually makes a claim about something that the hypothesis test simply can't tell you. For example, what if the "true" slope parameter is $\beta_{fatheduc} = 0.002$? The distribution of coefficient estimate values if this were the "true" state of the world would be shifted slightly to the right of the distribution we would draw assuming a null of H_0: $\beta_{fatheduc} = 0$ (think Type II errors here!)

- O.K.: so what's the bottom line? It is o.k. if you use the more informal language in your write-ups: this is the language that journal articles and other analyses use. BUT, please keep in mind the distinction about what hypothesis testing is actually doing for you.

III. ONE-TAILED HYPOTHESIS TESTS

- Up to this point, we've focused on two-tailed hypothesis tests, that is, tests of the form:

  H_0: $\beta_1 = 0$
  H_1: $\beta_1 \neq 0$

- This is the default test that statistical packages typically produce and report p-values for.

- Sometimes, *before a regression or test is ever done*, the analyst has a strong prior expectation (from theory or prior work) about the *sign* of the coefficient for a particular variable of interest.

Example: Suppose you have a strong prior that the more years of education a person's father has, the higher the person's wages will be (conditional on the other *Xs* in the model). Then, you may want to run a one-tailed test of the form:

H_0: $\beta_{fatheduc} = 0$
H_1: $\beta_{fatheduc} > 0$

Let's say we still want to set a 95% confidence level. Now, however, instead of splitting the remaining 5% into the two tails of the sampling distribution, we load all 5% into the upper tail.

If using the classical form of the hypothesis test, to get the critical value associated with this, we just go to the *t*-table, look at the one-tailed test section, α=0.05, d.f.=*n-k-1*, (1604-5-1 = 1598) and get a *t-critical* value of +1.645.

Using the model with Wages = *f*(IQ, Educ, KWW, motheduc, fatheduc), what is the calculated t-statistic for this test?

What is the p-value?

Compare the calculated test statistic to the critical value:

One-tailed tests make it easier to reject the null hypothesis under certain conditions. What are those conditions, and why?

IV. CONFIDENCE INTERVALS FOR REGRESSION COEFFICIENTS

- Recall that you can also test hypotheses using confidence intervals. Plus, we may be interested in CIs for other reasons. Such a CI is calculated as:

  $\text{CI}(\beta_j) = \hat{\beta}_j \pm t * s.e._{\hat{\beta}_j}$

- Suppose you select the t-value to calculate a 95% confidence interval (i.e., $t = \pm 1.96$ for a sufficiently large sample size). What does it mean to say that $\hat{\beta}_j \pm t * s.e._{\hat{\beta}_j}$ is a 95% confidence interval?

  If we were to select repeated random samples from the population and calculate 95% confidence intervals around the estimate of $\hat{\beta}_j$, then approximately 95% of these intervals will contain the true population parameter β_j.

- In other words, we can say that the procedure we're using (taking simple random samples and calculating interval estimates) is such that there is a probability of 0.95 or 95 chances out of 100 that we will get a value of $\hat{\beta}_j$ such that the interval

$\hat{\beta}_j \pm t * s.e._{\hat{\beta}_j}$ contains the *true* regression slope parameter, β_j. There remains a 5% chance that we will not get such a value of $\hat{\beta}_j$, in which case the interval $\hat{\beta}_j \pm t * s.e._{\hat{\beta}_j}$ will *not* contain the *true* regression slope parameter β_j.

- ***************** Important Note ******************
  We *can't* say that a given interval contains the population coefficient β_j with probability 0.95: After an interval has been constructed around the sample coefficient [i.e., $\hat{\beta}_j \pm t * s.e._{\hat{\beta}_j}$], that interval actually either contains the population coefficient or it doesn't.

  The probability that the interval you construct contains β_j is either 1 or 0, depending on whether β_j is actually inside or outside the confidence interval. But we can never actually know this! All we can say is that with repeated construction of confidence intervals, 95% such intervals will contain the population parameter.

  ****
- Recall that when you use confidence intervals to test hypotheses, if the hypothesized value of the population parameter (usually 0 for regression coefficients) falls *within* the confidence interval that you construct from the parameter estimate, then you "fail to reject" the null hypothesis. If the hypothesized value of the population parameter falls outside of the interval you construct, then you "reject the null hypothesis": it is unlikely that you would have observed the coefficient value that you did, if in fact the null hypothesis were true.

- **95% confidence intervals are included in Stata regression output by default:**

```
Dependent Variable: wage

. regress wage IQ educ KWW motheduc fatheduc

      Source |       SS       df       MS              Number of obs =    1604
-------------+------------------------------           F(  5,  1598) =   38.56
       Model |  12214605.9     5  2442921.17           Prob > F      =  0.0000
    Residual |   101247813  1598  63359.0817           R-squared     =  0.1077
-------------+------------------------------           Adj R-squared =  0.1049
       Total |   113462418  1603  70781.2966           Root MSE      =  251.71

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |  -.1159857    .508114    -0.23   0.819    -1.112626    .8806542
        educ |   12.89259   3.431463     3.76   0.000     6.161945    19.62323
         KWW |   9.563279   .9609077     9.95   0.000     7.678507    11.44805
    motheduc |   1.835333   2.770523     0.66   0.508    -3.598907    7.269574
    fatheduc |   .2444579   2.367129     0.10   0.918    -4.398547    4.887463
       _cons |    82.9006   50.01291     1.66   0.098    -15.19721    180.9984
------------------------------------------------------------------------------
```

- The conclusion you make based on the confidence interval test is exactly the same conclusion you would have reached based on looking at the p-values above

- [Note: to make e.g.,99% confidence intervals, the command would be}:

```
. regress wage IQ educ KWW motheduc fatheduc, level(99)

      Source |       SS       df       MS              Number of obs =    1604
-------------+------------------------------           F(  5,  1598) =   38.56
       Model |  12214605.9     5  2442921.17           Prob > F      =  0.0000
    Residual |   101247813  1598  63359.0817           R-squared     =  0.1077
-------------+------------------------------           Adj R-squared =  0.1049
       Total |   113462418  1603  70781.2966           Root MSE      =  251.71

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [99% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |  -.1159857    .508114    -0.23   0.819    -1.426366    1.194394
        educ |   12.89259   3.431463     3.76   0.000     4.043155    21.74202
         KWW |   9.563279   .9609077     9.95   0.000     7.085185    12.04137
    motheduc |   1.835333   2.770523     0.66   0.508    -5.309594    8.980261
    fatheduc |   .2444579   2.367129     0.10   0.918    -5.860154     6.34907
       _cons |    82.9006   50.01291     1.66   0.098    -46.07817    211.8794
------------------------------------------------------------------------------
```

IV. CONFIDENCE INTERVALS FOR PREDICTIONS

- Public policy analysts are usually most concerned with estimates of coefficients because they are interested in the effects of a particular factor(s) (holding others constant) on some outcome of interest. However, there may be instances where you want to use the regression results for prediction purposes.

PREDICTION OF AVERAGES FOR SPECIFIC SET OF XS

- When you plug in particular values for each X and obtain a prediction for Y, you are obtaining an estimate of the *expected value for y* given the particular values of the independent variables ($X_1 \ldots X_k$).

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \ldots \hat{\beta}_k X_k$$

- As for regression coefficients, Yhat is an *estimate* based on a sample and there is a corresponding standard error that can be used to compute a Confidence Interval for that estimate. Because Yhat in multiple regression is a linear combination of the OLS coefficients, estimating a standard error is more complicated. Wooldridge provides a method where you can fit a model so that the intercept will be the prediction for a particular set of X values and the corresponding SE for the intercept can be used to compute a Confidence Interval for the prediction.

  WOOLDRIDGE EXAMPLE 6.5, PAGE 207

```
clear
use gpa2.dta
gen sat0     = sat     - 1200
gen hsperc0  = hsperc  - 30
gen hsize0   = hsize   - 5
gen hsizesq0 = hsize^2 - 25

regress colgpa sat0 hsperc0 hsize0 hsizesq0

. regress colgpa sat0 hsperc0 hsize0 hsizesq0

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sat0 |   .0014925   .0000652    22.89   0.000     .0013646    .0016204
     hsperc0 |  -.0138558    .000561   -24.70   0.000    -.0149557   -.0127559
      hsize0 |  -.0608815   .0165012    -3.69   0.000    -.0932328   -.0285302
    hsizesq0 |   .0054603   .0022698     2.41   0.016     .0010102    .0099104
       _cons |   2.700075   .0198778   135.83   0.000     2.661104    2.739047
------------------------------------------------------------------------------
```

- The same general logic and interpretation of Confidence Intervals applies for predictions as described above for regression coefficients.

- Can you interpret the 95% CI of (2.66, 2.74) in Wooldridge Example 6.5? How would you compute that CI from the Stata output above?

PREDICTION OF INDIVIDUAL UNITS

- We just discussed how to obtain a CI for the expected value (i.e., average) for a particular combination of X values. Now suppose you are interested in obtaining a prediction *for a particular unit* (i.e., person, firm, county) that is not in the sample.

- Although the predicted value is the same when predicting a value for an individual unit as when predicting the expected value across all units with the same characteristics, the error associated with the prediction for a specific unit is different.

- Intuitively, it makes sense that the standard error is *larger* when making a prediction for an individual unit than when predicting the expected value (average) for all units with the same X characteristics. A more statistical explanation is:

  If repeated samples were taken, there would be more variability in Y across individual units than across average Y's for units with the same characteristics.

- There are two sources of error associated with predictions for a specific unit not in the sample: Error associated with estimating the coefficients + Unexplained variation associated with the model [Var(yhat) + σ^2].

WOOLDRIDGE EXAMPLE 6.6, PAGE 209

- Can you interpret the 95% CI of (1.60, 3.80) in Wooldridge Example 6.6 (page 209)?

V. PREDICTING *Y* WHEN *LN(Y)* IS THE DEPENDENT VARIABLE.

-- Wooldridge pp. 212-214 covers this. We won't go to go into detail here, because you will seldom have to do this in life.

-- However, you should know that if you're ever in a situation where you need to make a prediction of Y from some model of the form:

$$\ln(Y) = \beta_0 + \sum \beta_k X_k + \varepsilon$$

you ***can't*** get an estimate of Y as

$$e^{\ln(Y)} = e^{\left(\hat{\beta}_0 + \sum \hat{\beta}_k X_k\right)}$$
$$Y = e^{\left(\hat{\beta}_0 + \sum \hat{\beta}_k X_k\right)}$$

-- Instead, you have to go through some additional steps to get the correct prediction of *Y* from this model (see Wooldridge).