Econometrics

Lecture 6

Nathaniel Higgins

ERS and JHU

10 October 2011

Review of last time

- Last time we talked about the multivariate OLS estimator for the first time
- The multivariate OLS estimator is a powerful tool
  - Estimates a coefficient associated with each independent variable (ceteris paribus relationships)
  - Coefficient estimates are *marginal effects*
- Some things to consider when dealing with the multivariate model
  - Omitted variables can bias estimates (of course, this is a problem with the bivariate model as well — it's just that we've seen the problem in action now)
  - Multicollinearity can be a problem

Review of last time

- Multivariate tool does a couple of things for us:
  1. It allows us to estimate a relationship we care about, while holding all other things constant (that's the "ceteris paribus" stuff from the previous slide)
  2. It allows us to deal with a major source of endogeneity: omitted variable bias
- Endogeneity
  - OVB is just one type of endogeneity. It happens to be a type that we can understand pretty easily.
  - Endog. is present any time an x-variable is related to the unobservable (the u term) in a regression equation
  - We don't like endog. because it biases our coefficient estimates (remember what *bias* means in this context)

- In general with OLS models: have to consider how the independent variables relate to the unobservable in the model (the error term)
- Omitted variables generate bias because the omitted variable is subsumed into u

- In general with OLS models: have to consider how the independent variables relate to the unobservable in the model (the error term)
- Omitted variables generate bias because the omitted variable is subsumed into u
- This is bad. When u is related to the x-variable(s), we accidentally attribute to x some movements that are really movements in u

- In general with OLS models: have to consider how the independent variables relate to the unobservable in the model (the error term)
- Omitted variables generate bias because the omitted variable is subsumed into u
- This is bad. When u is related to the x-variable(s), we accidentally attribute to x some movements that are really movements in u
- We don't know yet what to do about endogeneity, but we know to look out for it

Variance of the OLS estimators

- So now we know how to estimate a model if we are given data
- We know how to estimate a model with multiple independent variables, and we know (a little bit about) what to look out for
- But this is only half the battle
- We want to use our model for inference
- What does this mean?

Inference

- We want to say something about how *likely* our model was to spit out the estimate it did
- Every time we estimate our model on slightly different data, we get slightly different coefficient estimates
- The model estimates we get are totally specific to the sample of data that we get — if we get a different sample of data, we get different point estimates

Inference (step 1)

Variability of the estimator

- Eventually, we are going to want to know exactly how $\widehat{\beta}_j$ is distributed
  - (so we can say things like “we are 95% sure that the true value of β_j is in the interval blah)
- We have already said that the mean of the estimator is the true value

$$E(\widehat{\beta}_j) = \beta_j$$

- Eventually, we are going to want to know exactly how $\widehat{\beta}_j$ is distributed
  - (so we can say things like "we are 95% sure that the true value of β_j is in the interval blah)
- We have already said that the mean of the estimator is the true value

$$E(\widehat{\beta}_j) = \beta_j$$

- Now we want to know about the variance
- We want to take a look at the estimate we get using one sample of data and have some idea of how much this estimate might change if we were to apply the same estimator to another sample of data

- FIGURE with two estimates from random samples of data

- FIGURE with two estimates from random samples of data
- Once we have some idea of the variance of the estimator, we can use this to create test statistics (just like you did in statistics and in the stats review for this class); compare these test statistics to a table of critical values

Inference (step 1)
Variability of the estimator

- FIGURE with two estimates from random samples of data
- Once we have some idea of the variance of the estimator, we can use this to create test statistics (just like you did in statistics and in the stats review for this class); compare these test statistics to a table of critical values
- The variance of the estimator is intrinsically linked to the variance of u
- To start, we assume that the variance of u is constant for all levels of the x-variables
- This is known as “homoskedasticity”

- FIGURE with two estimates from random samples of data
- Once we have some idea of the variance of the estimator, we can use this to create test statistics (just like you did in statistics and in the stats review for this class); compare these test statistics to a table of critical values
- The variance of the estimator is intrinsically linked to the variance of u
- To start, we assume that the variance of u is constant for all levels of the x-variables
- This is known as “homoskedasticity”
- (Easiest way to understand homoskedasticity is to think of an example when it doesn’t hold: think of unobserved propensity to consume as income increases)
- The opposite of homoskedasticity is heteroskedasticity
- Anyway, we assume homoskedasticity to start

Gauss-Markov assumptions

- Homoskedasticity rounds-out the GM assumptions (summary on next slide)
- These are the “base” assumptions

Gauss-Markov assumptions

- Homoskedasticity rounds-out the GM assumptions (summary on next slide)
- These are the “base” assumptions
- Much of applied econometrics deals with figuring out how to proceed when these assumptions fail
- Of course, in reality they *always* fail a little
- In practice, we worry about them a lot when there is a compelling reason to think that one of the assumptions fails badly
- Then we look for solutions
- But as a first cut at almost every problem, we do the analysis assuming that the Gauss-Markov assumptions hold

Gauss-Markov assumptions

1. MLR.1 (p. 84)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

2. MLR.2 (p. 84)
   Our data (x,y) represent a random sample from the relevant population
3. MLR.3 (p. 85)

$$x_i \neq z * x_j$$

4. MLR.4 (p. 87)

$$E(u|x) = 0$$

5. MLR.5 (p. 94)

$$Var(u|x) = \sigma^2$$

Variance of the OLS estimators

- With the GM assumptions we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- The formula can be found on p. 95
- The formula has some intuitive content

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

Variance of the OLS estimators

- With the GM assumptions we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- The formula can be found on p. 95
- The formula has some intuitive content

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- The variance of our estimator increases when the numerator increases and decreases when the denominator increases
- The numerator is a constant that represents the true variance of the unobservables in our equation. This makes sense. As the unexplainable portion of y varies more, our estimator is less capable of picking out variation in y that is caused by x (think of a moving target)
- The denominator is made out of two related things (next slide)

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- Denominator is made of
  1. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
     - How much a whole column of x-data varies
     - More variation in a given x-variable makes estimation easier
     - If we want to estimate β_1, then we want x_1 to be varying like crazy — if we want to estimate the effect of education on earnings, then a sample made out of our class would be a good/bad (?) sample?

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- Denominator is made of
  1. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
     - How much a whole column of x-data varies
     - More variation in a given x-variable makes estimation easier
     - If we want to estimate β_1, then we want x_1 to be varying like crazy — if we want to estimate the effect of education on earnings, then a sample made out of our class would be a good/bad (?) sample? Bad. Everyone has a pretty similar educational background. We need to vary x to see how y reacts
  2. $(1 - R_j^2)$
     - see next slide

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- Denominator is made of
  1. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
     - see previous slide
  2. $(1 - R_j^2)$
     - $R_j^2 =$ The proportion of variation in x_j explained by all the other x-variables
     - Saying that the x-variables are highly correlated is like saying that the x-variables don't do a lot of varying independently (think using the Ballantine)
     - This is bad, because we don't have much variation in the x-variables to use to estimate each independent relationship

- For more precise estimates, we can do a few things
  1. Collect x-variables that explain as much of y as possible (this reduces the role of u, which should make the variation of u as small as possible)
  2. Make the sample size as big as possible (increase n whenever possible)
  3. Be aware of the co-variation of x-variables (multicollinearity) sidebar on next slide

- Multicollinearity = bad
- Multicollinearity is not something we can “cure”
- If two variables *belong* in a model together, but they co-vary a lot, what to do?
- Be aware of it
- Suppose that x_1 and x_2 co-vary a lot (they move together)
- You must realize that your point estimates of $\widehat{\beta_1}$ and $\widehat{\beta_2}$, and any inference you make about their values, are going to be weak
- In general, the more x_1 and x_2 co-vary, the weaker is our ability to reach conclusions about the true values of β_1 and β_2
- The only way to deal directly with multicollinearity without risk of any downside is to increase *n*

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

1. $\sigma^2 = Var(u|x)$
   - Estimate this value using residuals from our regression model: $\widehat{\sigma^2} = \frac{(\widehat{u_1^2} + \widehat{u_2^2} + \ldots + \widehat{u_n^2})}{(n-k-1)}$
2. $SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$
3. $R_j^2 = R^2$ from a regression of x_j on all the other x-variables (x_{-j})

- We estimate the variance of the OLS estimator

$$Var(\widehat{\beta}_j) = \frac{\widehat{\sigma^2}}{SST_j(1 - R_j^2)}$$

- ...then we take the square root of it

$$se(\widehat{\beta}_j) = \sqrt{Var(\widehat{\beta}_j)}$$

- ...and THIS is the value that we see in our regression output

Stata code

```
use lecture-05-dset-01.dta
reg y x1 x2
```

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) = 2106.00
       Model |  3573.6071     2  1786.80355           Prob > F      =  0.0000
    Residual | 82.2982026    97  .848435078           R-squared     =  0.9775
-------------+------------------------------           Adj R-squared =  0.9770
       Total |  3655.9053    99  36.9283364           Root MSE      =  .92111

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.063415   .2051654    10.06   0.000     1.656218    2.470611
          x2 |  -2.968643   .0457425   -64.90   0.000    -3.059429   -2.877857
       _cons |   4.741896   .6182503     7.67   0.000      3.51484    5.968952
------------------------------------------------------------------------------
```

- Under GM assumptions (MLR.1 – MLR.5) the OLS estimator is BLUE
- It is the Beast Linear Unibiased Estimator
- You know what “linear” and “unbiased” mean
- If we compare all estimators that are linear and unbiased, the OLS estimator has the lowest variance

- OK, so we know how to come up with a formula that estimates the variance of the OLS estimator

$$Var(\widehat{\beta}_j) = \frac{\widehat{\sigma^2}}{SST_j(1 - R_j^2)}$$

- ...and we know how to use this to calculate a thing called the *standard error* of the j^{th} coefficient estimate
- Now what?

- OK, so we know how to come up with a formula that estimates the variance of the OLS estimator

$$Var(\widehat{\beta}_j) = \frac{\widehat{\sigma^2}}{SST_j(1 - R_j^2)}$$

- ...and we know how to use this to calculate a thing called the *standard error* of the j^{th} coefficient estimate
- Now what?

- So what? Who cares?

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - Oh *that value*? We see that value all the time.

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - Oh *that value*? We see that value all the time.
- Knowing how likely a value is to have occurred by chance means that we can say intelligent things about how likely it is that our coefficient estimates were obtained because of something *other than chance*
  - Not a likely value just by chance, eh? Must be something deterministic going on here

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - Oh *that value*? We see that value all the time.
- Knowing how likely a value is to have occurred by chance means that we can say intelligent things about how likely it is that our coefficient estimates were obtained because of something *other than chance*
  - Not a likely value just by chance, eh? Must be something deterministic going on here
- Take our example dset from last time (dset1)

- When we ran a regression of y on x_1 and x_2, we obtained a point estimate of $\beta_1 = 2.063415$.
- Is this estimate “big?”

- When we ran a regression of y on x_1 and x_2, we obtained a point estimate of $\beta_1 = 2.063415$.
- Is this estimate "big?" I don't know either. "Big" is relative. Only context of the problem makes an estimate "big." (think of a returns-to-education example if you want to think about "big" in context)
- Is the coefficient estimate of 2.06 is an anomaly? Maybe the fact that we obtained 2.06 is just due to random fluctuations (rather than the fact that the true value of the coefficient is around 2.06). How likely is it that the true coefficient is actually 0?

- When we ran a regression of y on x_1 and x_2, we obtained a point estimate of $\beta_1 = 2.063415$.
- Is this estimate “big?” I don’t know either. “Big” is relative. Only context of the problem makes an estimate “big.” (think of a returns-to-education example if you want to think about “big” in context)
- Is the coefficient estimate of 2.06 is an anomaly? Maybe the fact that we obtained 2.06 is just due to random fluctuations (rather than the fact that the true value of the coefficient is around 2.06). How likely is it that the true coefficient is actually 0?
- This is the type of question we can start to answer now

- Let me ask you a question: if the point estimate for β_1 was 2.063415 and the standard error (a measure of variation) was 0.00000000000001, how likely do you think it is that the true value of β_1 is 0?

- Let me ask you a question: if the point estimate for β_1 was 2.063415 and the standard error (a measure of variation) was 0.00000000000001, how likely do you think it is that the true value of β_1 is 0?
- Not very
- That is a loose notion, though. The whole idea of test statistics is to put a finer set of bounds on when the standard error is "too big" for us to consider our estimate of β_j "significantly different" from 0 (or some other value we choose)

- We have always assumed that the *u* are random
- But we haven't assumed much else
- In order to determine the s.e. of the coefficient estimates, we assumed homoskedasticity
- We used this assumption to estimate σ^2, the variance of *u*
- In order to take the next step, we need to make an assumption about exactly how *u* is distributed

- We have always assumed that the u are random
- But we haven't assumed much else
- In order to determine the s.e. of the coefficient estimates, we assumed homoskedasticity
- We used this assumption to estimate σ^2, the variance of u
- In order to take the next step, we need to make an assumption about exactly how u is distributed
- In order to say how our estimator $\widehat{\beta}_j$ is distributed, we need to know how u is distributed
- In more advanced work, you might look at your estimates of u (the $\widehat{u}$ that you obtain when you estimate a model) and use them to make a guess of how u is really distributed
- In most cases you will assume that u is distributed in a certain way
- The most common assumption is that u is distributed normally

- We add to our set of working assumptions the *normality assumption*
- Thus our accumulated assumptions are MLR.1 – MLR.5 and the assumption that u is distributed normally
- Combining these assumptions gives us MLR.6

$$u \sim N(0, \sigma^2)$$

- We add to our set of working assumptions the *normality assumption*
- Thus our accumulated assumptions are MLR.1 – MLR.5 and the assumption that u is distributed normally
- Combining these assumptions gives us MLR.6

$$u \sim N(0, \sigma^2)$$

- We can operate an OLS model without MLR.6. Period.

- We add to our set of working assumptions the *normality assumption*
- Thus our accumulated assumptions are MLR.1 – MLR.5 and the assumption that u is distributed normally
- Combining these assumptions gives us MLR.6

$$u \sim N(0, \sigma^2)$$

- We can operate an OLS model without MLR.6. Period. But we need MLR.6 to use our basic tools of statistical inference

- We add to our set of working assumptions the *normality assumption*
- Thus our accumulated assumptions are MLR.1 – MLR.5 and the assumption that u is distributed normally
- Combining these assumptions gives us MLR.6

$$u \sim N(0, \sigma^2)$$

- We can operate an OLS model without MLR.6. Period. But we need MLR.6 to use our basic tools of statistical inference
- Our total set of assumptions so far (MLR.1 – MLR.6) are the assumptions of the Classic Linear Regression Model (or CLM, in Wooldridge's parlance)

- Under CLM assumptions

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

  - (the formula for $Var(\widehat{\beta}_j)$ was given earlier in the lecture)

- Under CLM assumptions

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

  - (the formula for $Var(\widehat{\beta}_j)$ was given earlier in the lecture)
- Knowing this allows us to conclude that

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- This expression is a really useful way to summarize what we know about the distribution of $\widehat{\beta}_j$
- We can use it to test hypotheses about β_j
- The most common null hypothesis is $\beta_j = 0$

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- This expression is a really useful way to summarize what we know about the distribution of $\widehat{\beta}_j$
- We can use it to test hypotheses about β_j
- The most common null hypothesis is $\beta_j = 0$
- This is akin to testing that the j^{th} x-variable has *no effect* on y
- Proving otherwise (i.e. proving that the x-variable does have some effect on y) is often a powerful conclusion

- Does education have any effect on earnings?
  - Rejecting the null bolsters arguments for spending money on education

- Does education have any effect on earnings?
  - Rejecting the null bolsters arguments for spending money on education
- Does access to micro-credit increase the (long-term) income of people in developing countries?
  - Rejecting the null argues in favor of micro-credit

- Does education have any effect on earnings?
  - Rejecting the null bolsters arguments for spending money on education
- Does access to micro-credit increase the (long-term) income of people in developing countries?
  - Rejecting the null argues in favor of micro-credit
- Is the squared term in an EKC regression equal to 0?
  - Rejecting the null suggests that eventually, increasing GDP further will cause pollution to decrease

- The most common null hypothesis is that the true value of β_j is zero

$$H_0 : \beta_j = 0$$

- The most common alternative hypothesis is

$$H_A : \beta_j \neq 0$$

- Recall again our regression results from lecture 5

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) = 2106.00
       Model |  3573.6071     2  1786.80355            Prob > F      =  0.0000
    Residual | 82.2982026    97  .848435078            R-squared     =  0.9775
-------------+------------------------------           Adj R-squared =  0.9770
       Total |  3655.9053    99  36.9283364            Root MSE      =  .92111

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.063415   .2051654    10.06   0.000     1.656218    2.470611
          x2 |  -2.968643   .0457425   -64.90   0.000    -3.059429   -2.877857
       _cons |   4.741896   .6182503     7.67   0.000      3.51484    5.968952
------------------------------------------------------------------------------
```

- What is $t_{\widehat{\beta_1}}$?

- Recall again our regression results from lecture 5

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) = 2106.00
       Model |  3573.6071      2  1786.80355           Prob > F      =  0.0000
    Residual | 82.2982026     97  .848435078           R-squared     =  0.9775
-------------+------------------------------           Adj R-squared =  0.9770
       Total |  3655.9053     99  36.9283364           Root MSE      =  .92111

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.063415   .2051654    10.06   0.000     1.656218    2.470611
          x2 |  -2.968643   .0457425   -64.90   0.000    -3.059429   -2.877857
       _cons |   4.741896   .6182503     7.67   0.000      3.51484    5.968952
------------------------------------------------------------------------------
```

- What is $t_{\widehat{\beta_1}}$?

$$\frac{2.063415}{0.2051654} = 10.06$$

- Recall again our regression results from lecture 5

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) = 2106.00
       Model |   3573.6071     2  1786.80355           Prob > F      =  0.0000
    Residual |  82.2982026    97  .848435078           R-squared     =  0.9775
-------------+------------------------------           Adj R-squared =  0.9770
       Total |   3655.9053    99  36.9283364           Root MSE      =  .92111

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.063415   .2051654    10.06   0.000     1.656218    2.470611
          x2 |  -2.968643   .0457425   -64.90   0.000    -3.059429   -2.877857
       _cons |   4.741896   .6182503     7.67   0.000      3.51484    5.968952
------------------------------------------------------------------------------
```

- What is $t_{\widehat{\beta}_1}$?

$$\frac{2.063415}{0.2051654} = 10.06$$

- To verify: `scalar t_b1 = 2.063415/0.2051654`

Testing null hypotheses in Stata

- Using `lecture-05-dset-01.dta`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$

Testing null hypotheses in Stata

- Using `lecture-05-dset-01.dta`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy.

Testing null hypotheses in Stata

- Using `lecture-05-dset-01.dta`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy. Take the t-statistic reported in the row labeled `x1` and compare it to the critical value on page 825 of Wooldridge

Testing null hypotheses in Stata

- Using `lecture-05-dset-01.dta`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy. Take the t-statistic reported in the row labeled `x1` and compare it to the critical value on page 825 of Wooldridge
- If the t-statistic is greater than the critical value, then we reject the null (we reject H_0)

Testing null hypotheses in Stata

- Using `lecture-05-dset-01.dta`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy. Take the t-statistic reported in the row labeled `x1` and compare it to the critical value on page 825 of Wooldridge
- If the t-statistic is greater than the critical value, then we reject the null (we reject H_0)
- To look up the proper critical value you need to know one thing: the degrees of freedom of the test (the number of observations we are basic the test statistic on)

Testing null hypotheses in Stata

- Using `lecture-05-dset-01.dta`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy. Take the t-statistic reported in the row labeled `x1` and compare it to the critical value on page 825 of Wooldridge
- If the t-statistic is greater than the critical value, then we reject the null (we reject H_0)
- To look up the proper critical value you need to know one thing: the degrees of freedom of the test (the number of observations we are basic the test statistic on)
- We are using one degree of freedom to estimate the intercept in the model, one to estimate β_1, and one to estimate β_2. We started out with 100 observations. How many are left to use for inference? 100 - 3 = 97.

- You start with 100 observations
- Then you estimate the value of β_0

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$

Degrees of freedom

What you talkin' 'bout?

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$
- What? Think of it this way:
- Suppose I had a sample composed of two observations (crappy dset, I know, but this is a toy example)
- I use this observation to estimate the population mean

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$
- What? Think of it this way:
- Suppose I had a sample composed of two observations (crappy dset, I know, but this is a toy example)
- I use this observation to estimate the population mean
- I give you the first observation: 7
- I give you my estimate of the mean: 8
- What was the other observation in my sample?

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$
- What? Think of it this way:
- Suppose I had a sample composed of two observations (crappy dset, I know, but this is a toy example)
- I use this observation to estimate the population mean
- I give you the first observation: 7
- I give you my estimate of the mean: 8
- What was the other observation in my sample?
- Seeeee?

Looking up the t-stat

- We know the degrees of freedom
- We decide on the significance level of the test (0.05 is very common, but there is absolutely, positively nothing special about it)
- We look up the critical value in our t-table
- If our t-stat is greater in absolute value than the critical t, then we reject the null
- For our test, the critical t is somewhere in the neighborhood of 1.7
- So we can reject our null like nobody's business

Looking up the t-stat

- We know the degrees of freedom
- We decide on the significance level of the test (0.05 is very common, but there is absolutely, positively nothing special about it)
- We look up the critical value in our t-table
- If our t-stat is greater in absolute value than the critical t, then we reject the null
- For our test, the critical t is somewhere in the neighborhood of 1.7
- So we can reject our null like nobody's business
- Can we tell anything from the fact that our t-stat was so damn big? Hold that thought

- The most common null hypothesis is that the true value of β_j is zero

$$H_0 : \beta_j = 0$$

- The most common alternative hypothesis is

$$H_A : \beta_j \neq 0$$

- This is a two-sided alternative (it could be either that $\beta_j < 0$ or $\beta_j > 0$)
- We could have one-sided alternatives too

$$H_A : \beta_j < 0$$

- Sometimes a one-sided test makes a great deal of sense
- The EKC gives one example

- Sometimes a one-sided test makes a great deal of sense
- The EKC gives one example
- In the EKC example, we know ahead of time that we are likely to see a negative value for GDP^2
- We still wish to test the null

$$H_0 : \beta_j = 0$$

- ...but now we posit that the only reasonable alternative is

$$H_A : \beta_j < 0$$

- ...because $\beta_j > 0$ is something we are ruling out before we run our test

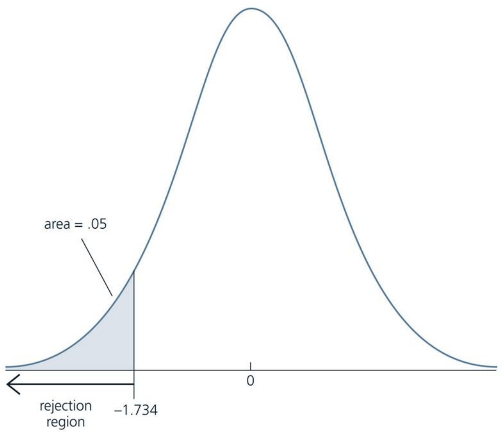
- Suppose that our estimate of β_2 in `lecture-05-dset-01.dta` was such an estimate (we knew it was going to be negative before we started — we just weren't sure how negative)

- Suppose that our estimate of β_2 in `lecture-05-dset-01.dta` was such an estimate (we knew it was going to be negative before we started — we just weren't sure how negative)
- Then we compare our t-statistic to the one-tailed t-stat

- If our t-stat is less than $-t_c$, then we reject H_0

FIGURE 4.3

5% rejection rule for the alternative H_1: $\beta_j < 0$ with 18 *df*.



- We know how to test null hypotheses that look like

$$H_0 : \beta_j = 0$$

- ...against alternative hypotheses that look like

$$H_A : \beta_j \neq 0$$

- We know how to test null hypotheses that look like

$$H_0 : \beta_j = 0$$

- ...against alternative hypotheses that look like

$$H_A : \beta_j \neq 0$$

or

$$H_A : \beta_j < 0$$

- We know how to test null hypotheses that look like

$$H_0 : \beta_j = 0$$

- ... against alternative hypotheses that look like

$$H_A : \beta_j \neq 0$$

or

$$H_A : \beta_j < 0$$

or

$$H_A : \beta_j > 0$$

- But sometimes we go crazy and want to test other nulls
- Always testing that $\beta_j = 0$ is pretty limiting
- Suppose we weren't impressed just by saying that education *matters* — suppose we wanted to say something about *how much* it mattered

- But sometimes we go crazy and want to test other nulls
- Always testing that $\beta_j = 0$ is pretty limiting
- Suppose we weren't impressed just by saying that education *matters* — suppose we wanted to say something about *how much* it mattered
- Maybe we are estimating a wage equation and we want to know the contribution of education

$$log(wage) = \beta_0 + \beta_1 educ + \ldots + \beta_k x_k + u$$

- So we estimate our coefficient on the education x-variable, which we are calling β_1
- This coefficient is our best guess of the effect of education on wages

$$log(wage) = \beta_0 + \beta_1 educ + \ldots + \beta_k x_k + u$$

- So we estimate our coefficient on the education x-variable, which we are calling β_1
- This coefficient is our best guess of the effect of education on wages
- Now we want to test a null

$$log(wage) = \beta_0 + \beta_1 educ + \ldots + \beta_k x_k + u$$

- So we estimate our coefficient on the education x-variable, which we are calling β_1
- This coefficient is our best guess of the effect of education on wages
- Now we want to test a null
- But the null that $\beta_1 = 0$ isn't satisfying. We want to test the null that β_1 = 0.1

$$log(wage) = \beta_0 + \beta_1 educ + \ldots + \beta_k x_k + u$$

- So we estimate our coefficient on the education x-variable, which we are calling β_1
- This coefficient is our best guess of the effect of education on wages
- Now we want to test a null
- But the null that $\beta_1 = 0$ isn't satisfying. We want to test the null that β_1 = 0.1
- If this null were true, it would be saying that as education increases by one year, wages increase by ...

$$log(wage) = \beta_0 + \beta_1 educ + \ldots + \beta_k x_k + u$$

- So we estimate our coefficient on the education x-variable, which we are calling β_1
- This coefficient is our best guess of the effect of education on wages
- Now we want to test a null
- But the null that $\beta_1 = 0$ isn't satisfying. We want to test the null that β_1 = 0.1
- If this null were true, it would be saying that as education increases by one year, wages increase by . . . 10%

- We have been motoring along, just dividing our coefficient estimates by their standard errors, getting t-stats, and comparing them to critical values we found in tables

- We have been motoring along, just dividing our coefficient estimates by their standard errors, getting t-stats, and comparing them to critical values we found in tables
- But remember where this all started

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- We have been motoring along, just dividing our coefficient estimates by their standard errors, getting t-stats, and comparing them to critical values we found in tables
- But remember where this all started

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- We have been dividing our estimated coefficient by its standard error because we have been testing the null that $\beta_j = 0$

- When we test the null that $\beta_j = 0$, what happens to the t-statistic?

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- When we test the null that $\beta_j = 0$, what happens to the t-statistic?

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- It turns into

$$\frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- When we test the null that $\beta_j = 0$, what happens to the t-statistic?

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- It turns into

$$\frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- To test more complicated hypotheses (like β_1 (education) = 0.1), we need to put β_j back in the formula

- Testing such a hypothesis is as simple as plugging the numbers into the formula
- Try

Stata code

```
reg y x1 x2
scalar t_b1 = (2.063415 - 0.1)/.2051654
```

- What do you get?

- Testing such a hypothesis is as simple as plugging the numbers into the formula
- Try

Stata code

```
reg y x1 x2
scalar t_b1 = (2.063415 - 0.1)/.2051654
```

- What do you get?
- I get a smaller t-statistic

- Testing such a hypothesis is as simple as plugging the numbers into the formula
- Try

Stata code

```
reg y x1 x2
scalar t_b1 = (2.063415 - 0.1)/.2051654
```

- What do you get?
- I get a smaller t-statistic
- This means: we are damn sure that the coefficient isn't 0 (big t-stat) and we are still very sure that the coefficient isn't 0.1 (but we are less sure than we were about the 0 thing)

Confidence intervals

- This experience might lead you to wonder . . .
- OK, so it's nice to know that something *isn't* 0
- I guess it's also nice to know that something *isn't* 0.1
- But where does this get us?

Confidence intervals

- This experience might lead you to wonder ...
- OK, so it's nice to know that something *isn't* 0
- I guess it's also nice to know that something *isn't* 0.1
- But where does this get us?
- Confidence intervals are here to help

Confidence intervals

- This experience might lead you to wonder ...
- OK, so it's nice to know that something *isn't* 0
- I guess it's also nice to know that something *isn't* 0.1
- But where does this get us?
- Confidence intervals are here to help
- We could keep increasing the number in our test, so that we tested 0, then 0.1, then (if it "passed" the test) 0.2, and so on

Confidence intervals

- This experience might lead you to wonder ...
- OK, so it's nice to know that something *isn't* 0
- I guess it's also nice to know that something *isn't* 0.1
- But where does this get us?
- Confidence intervals are here to help
- We could keep increasing the number in our test, so that we tested 0, then 0.1, then (if it "passed" the test) 0.2, and so on
- Doing this would eventually lead us to a value that we could NOT reject

Confidence intervals

- This experience might lead you to wonder . . .
- OK, so it's nice to know that something *isn't* 0
- I guess it's also nice to know that something *isn't* 0.1
- But where does this get us?
- Confidence intervals are here to help
- We could keep increasing the number in our test, so that we tested 0, then 0.1, then (if it "passed" the test) 0.2, and so on
- Doing this would eventually lead us to a value that we could NOT reject
- Then we'd say about this value (call it z_L): We cannot reject the null hypothesis that $\beta_j = z_L$
- There would be a bunch of values $z > z_L$ that we could not reject

Confidence intervals

- To test null hypotheses with non-zero hypothesized values
  - We used `scalar t_b1 = (2.063415 - 0.1)/.2051654`
  - Stata also has a shorthand way to do this: `lincom x1-0.1`

Confidence intervals

- To test null hypotheses with non-zero hypothesized values
  - We used `scalar t_b1 = (2.063415 - 0.1)/.2051654`
  - Stata also has a shorthand way to do this: `lincom x1-0.1`
- Try increasing the value a little bit to try to find z_L

Confidence intervals

- To test null hypotheses with non-zero hypothesized values
  - We used `scalar t_b1 = (2.063415 - 0.1)/.2051654`
  - Stata also has a shorthand way to do this: `lincom x1-0.1`
- Try increasing the value a little bit to try to find z_L
- I find that z_L is somewhere between 1.6 and 1.7

Confidence intervals

- To test null hypotheses with non-zero hypothesized values
  - We used `scalar t_b1 = (2.063415 - 0.1)/.2051654`
  - Stata also has a shorthand way to do this: `lincom x1-0.1`
- Try increasing the value a little bit to try to find z_L
- I find that z_L is somewhere between 1.6 and 1.7
- Now watch what happens when we increase our hypothesized value by a whole bunch: `lincom x1-4`

Confidence intervals

- To test null hypotheses with non-zero hypothesized values
  - We used `scalar t_b1 = (2.063415 - 0.1)/.2051654`
  - Stata also has a shorthand way to do this: `lincom x1-0.1`
- Try increasing the value a little bit to try to find z_L
- I find that z_L is somewhere between 1.6 and 1.7
- Now watch what happens when we increase our hypothesized value by a whole bunch: `lincom x1-4`
- Once again, we can reject the null hypothesis (this makes sense)

Confidence intervals

- There is a maximum hypothesized value that we can reject

Confidence intervals

- There is a maximum hypothesized value that we can reject
- I find that we cannot reject the null when the hypothesized value is greater than $z_U =$ somewhere between 2.4 and 2.5

Confidence intervals

- There is a maximum hypothesized value that we can reject
- I find that we cannot reject the null when the hypothesized value is greater than $z_U =$ somewhere between 2.4 and 2.5
- The t-stat enables us to reject a null hypothesis at a given level of significance

Confidence intervals

- There is a maximum hypothesized value that we can reject
- I find that we cannot reject the null when the hypothesized value is greater than $z_U =$ somewhere between 2.4 and 2.5
- The t-stat enables us to reject a null hypothesis at a given level of significance
- A confidence interval tells us the range of values for which we could not reject the null
- What to do with this?

Confidence intervals

- We've got a range of values
- Inside this range, we cannot reject the null at the given level of confidence (say this level of significance was 0.05)

Confidence intervals

- We've got a range of values
- Inside this range, we cannot reject the null at the given level of confidence (say this level of significance was 0.05)
- Outside this range, we reject the null
- What does it mean to reject the null?

Confidence intervals

- We've got a range of values
- Inside this range, we cannot reject the null at the given level of confidence (say this level of significance was 0.05)
- Outside this range, we reject the null
- What does it mean to reject the null?
- Rejecting the null says: with 95% confidence, we reject the hypothesized value of β_j

Confidence intervals

- We've got a range of values
- Inside this range, we cannot reject the null at the given level of confidence (say this level of significance was 0.05)
- Outside this range, we reject the null
- What does it mean to reject the null?
- Rejecting the null says: with 95% confidence, we reject the hypothesized value of β_j
- So we have a bunch of values that we can reject; and a small band of values that we can NOT reject with 95% confidence

Confidence intervals

- We've got a range of values
- Inside this range, we cannot reject the null at the given level of confidence (say this level of significance was 0.05)
- Outside this range, we reject the null
- What does it mean to reject the null?
- Rejecting the null says: with 95% confidence, we reject the hypothesized value of β_j
- So we have a bunch of values that we can reject; and a small band of values that we can NOT reject with 95% confidence
- This is the range of values that we are 95% sure that the true value of the coefficient is IN

Confidence intervals

- Finding the confidence interval in a single step (rather than gradually playing with the `lincom` function like I did) is straightforward
- A simple manipulation of the t-statistic formula does the trick (see p. 138)

Confidence intervals

- Finding the confidence interval in a single step (rather than gradually playing with the `lincom` function like I did) is straightforward
- A simple manipulation of the t-statistic formula does the trick (see p. 138)
- But of course, the CI is part of standard regression output
- Look again at the output from
  `reg y x1 x2`
- Compare to the values I found: I said z_L is between 1.6 and 1.7 and z_U is between 2.4 and 2.5

Confidence intervals

- Finding the confidence interval in a single step (rather than gradually playing with the `lincom` function like I did) is straightforward
- A simple manipulation of the t-statistic formula does the trick (see p. 138)
- But of course, the CI is part of standard regression output
- Look again at the output from
  `reg y x1 x2`
- Compare to the values I found: I said z_L is between 1.6 and 1.7 and z_U is between 2.4 and 2.5
- Check out the CI in the Stata output. Interpret.

p-values

- So far so good
- We have found ways to test null hypotheses (zero and non-zero, one-sided and two-sided)

p-values

- So far so good
- We have found ways to test null hypotheses (zero and non-zero, one-sided and two-sided)
- We have *always* had to set a level of significance (α) ahead of time
- Each of the tests is good for a specific value of α

p-values

- So far so good
- We have found ways to test null hypotheses (zero and non-zero, one-sided and two-sided)
- We have *always* had to set a level of significance (α) ahead of time
- Each of the tests is good for a specific value of α
- You might say: OK, so we can reject this null at a significance value of 0.05. I wonder if we can also reject the null at an even more *stringent value* of α

p-values

- So far so good
- We have found ways to test null hypotheses (zero and non-zero, one-sided and two-sided)
- We have *always* had to set a level of significance (α) ahead of time
- Each of the tests is good for a specific value of α
- You might say: OK, so we can reject this null at a significance value of 0.05. I wonder if we can also reject the null at an even more *stringent value* of α
- p-values are here to help

p-values

- p-values are a useful concept (and also part of standard regression output)
- A p-value is the minimum level of significance (the most stringent level) at which we can reject the null hypothesis

p-values

- p-values are a useful concept (and also part of standard regression output)
- A p-value is the minimum level of significance (the most stringent level) at which we can reject the null hypothesis
- We would come up with different p-values for each kind of null hypothesis
- The only one that comes as part of the standard regression output is the one that corresponds to the most basic hypothesis:

p-values

- p-values are a useful concept (and also part of standard regression output)
- A p-value is the minimum level of significance (the most stringent level) at which we can reject the null hypothesis
- We would come up with different p-values for each kind of null hypothesis
- The only one that comes as part of the standard regression output is the one that corresponds to the most basic hypothesis: $H_0 : \beta_j = 0$

p-values

- p-values are a useful concept (and also part of standard regression output)
- A p-value is the minimum level of significance (the most stringent level) at which we can reject the null hypothesis
- We would come up with different p-values for each kind of null hypothesis
- The only one that comes as part of the standard regression output is the one that corresponds to the most basic hypothesis: $H_0 : \beta_j = 0$
- Obtaining the p-value for a different hypothesis is easy: `lincom x1-2.4`, e.g.

p-values

- Computing p-values for levels of α other than 0.05 is straightforward as well:
  ```
  lincom x1-2.4, level(99)
  ```

p-values

- Computing p-values for levels of α other than 0.05 is straightforward as well:
  ```
  lincom x1-2.4, level(99)
  ```
- Computing p-values for one-sided tests isn't so straightforward, but it isn't impossibly hard, either
- If we have time: example on page 134

- p-value =
$P(|T| > 1.85) = 2 * P(T > 1.85) = 2 * (0.0359) = 0.0718$

- p-value = $P(|T| > 1.85) = 2 * P(T > 1.85) = 2 * (0.0359) = 0.0718$
- To obtain this using Stata, use the `ttail` function
- The `ttail` distribution gives the probability in the tails of the t-distribution

- p-value = $P(|T| > 1.85) = 2 * P(T > 1.85) = 2 * (0.0359) = 0.0718$
- To obtain this using Stata, use the `ttail` function
- The `ttail` distribution gives the probability in the tails of the t-distribution
- To find $P(T > 1.85)$ with 40 degrees of freedom, use `ttail(40,1.85)`

“Next” time

- Next time (which is actually in two lecture periods) we will move to testing compound hypotheses
- Then we will look to relax some of the assumptions that we have relied on to get where we are (MLR.1 – MLR.6)
- But the *real* next time is our review for the mid-term

Homework

•