Empirical Methods

Topic 1d:

The CLRM: Goodness of Fit and Hypothesis Testing

The Classical Linear Regression Model Goodness of Fit and Hypothesis Testing

CLRM Hypothesis Testing: Intro I

- Time for a change in focus
  - Our decks up to this point have been focused on *estimation* of the CLRM
    - ★ And the assumptions under which it has nice properties
- Now we turn to two related topics:
  - Goodness-of-fit
  - (Interval estimation and) Hypothesis testing

CLRM Hypothesis Testing: Intro II

- Why these next?
- Recall the three purposes of regression we started off with:
    1. Data summary
    2. Measuring causal effects
    3. Prediction

CLRM Hypothesis Testing: Intro III

- For prediction, we're interested in using the model to predict something
  - Thus understanding how well it fits the data is important
- For measuring causal effects, we're often interested in the (a) sign and (b) magnitude of particular $\hat{\beta}_k$
  - And we need hypothesis testing to tell us whether the signs and magnitudes we've estimated are "statistically significant"
    - i.e., can we reject the hypothesis that a particular $\beta_k = 0$?
- Interestingly:
  - For prediction, we're often *not* interested in signs and magnitudes of $\hat{\beta}_k$
  - For causal effects, we're often *not* (as) interested in how well we fit the data

Goodness of Fit

Latex Color = "LightGrey"

Goodness of Fit: Intro

- When evaluating the fit of a regression, there are several classic measures:
  - R^2
  - "Adjusted R^2"
- These are of only limited usefulness however
  - As they only predict fit *within* a given sample
- I will therefore also briefly introduce a popular method for evaluating fit using out-of-sample prediction
  - Called Cross Validation
  - (Widely used in time series and "Big Data" applications)

R^2 and Adjusted R^2

Latex Color = "LightCyan"

R^2 I

- R^2, also known as the "coefficient of determination," is a measure of the overall (within-sample) fit of a regression
  - What does this mean???
- Well, the thing we're trying to explain is y_i
  - And we have a measure of how well we fit y_i with our fitted value $\hat{y}_i = x_i'\hat{\beta}$
  - And we have a measure of how well we failed to fit y_i with our residual, e_i $(\Leftrightarrow \hat{\epsilon}_i)$

R^2 II

- R^2 uses these three objects $(y_i, \hat{y}_i, e_i)$ to construct a measure of fit
- *What* do we fit?
  - Well, we showed in the previous deck that OLS will always perfectly fit the mean of y_i
    - i.e., $\hat{\beta}_1$ (the constant) will be chosen such that $\bar{\hat{y}} = \bar{y}$
    - (So that's not much of a test)
- What about the *variability* in y_i?
  - That could be useful.
  - We'd like to know how much of the variance of y_i can be explained by our model
    - As measured by $\hat{y}_i$

TSS, ESS, and RSS: Preamble

- A brief aside on some notation that will be useful:
  - Recall we said last time that including a constant term "mean-deviates" all of the other variables in an econometric model
    - ★ (Top01c, Slide 29)
- Define two mean-deviates that are useful in what follows:
  - Let $\bar{y} = \frac{1}{N}\sum y_i$ and $\bar{\hat{y}} = \frac{1}{N}\sum \hat{y}_i$ be the sample means of y_i and $\hat{y}_i$
  - And let their "mean-deviates" be

$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{\hat{y}}_i = \hat{y}_i - \bar{\hat{y}}$$

TSS, ESS, and RSS I

- To measure the variability in y_i explained by our model, define three useful objects:

1. The *Total Sum of Squares (TSS)* as the sum of the squared $\tilde{y}_i$:

$$\begin{aligned} TSS &= \sum \tilde{y}_i^2 && \text{in summation notation} \\ &= \tilde{y}'\tilde{y} && \text{in matrix notation} \end{aligned}$$

2. The *Explained Sum of Squares (ESS)* as the sum of the squared $\tilde{\hat{y}}_i$:

$$\begin{aligned} ESS &= \sum \tilde{\hat{y}}_i^2 && \text{in summation notation} \\ &= \tilde{\hat{y}}'\tilde{\hat{y}} && \text{in matrix notation} \end{aligned}$$

3. The *Residual Sum of Squares (RSS)* as the sum of the squared e_i:

$$\begin{aligned} RSS &= \sum e_i^2 && \text{in summation notation} \\ &= e'e && \text{in matrix notation} \end{aligned}$$

TSS, ESS, and RSS II

- I'll show you how we use these momentarily, but before I do...
  - ...a comment on "naming conventions"
- Unfortunately, different econometrics textbooks both call - *and* abbreviate - these objects differently
  - e.g., some might say "sum of squared residuals" instead of "residual sum of squares"
  - And denote it "SSR" instead of "RSS"
- That's all fine:
  - Just be aware of it...
    - ...and do the mental translation if you look up this topic in a textbook
  - (I have to admit I can never keep them straight and may - unintentionally - use both!)

TSS, ESS, and RSS III

- It turns out that you can show that

$$\boxed{\begin{array}{rcl} \tilde{y}'\tilde{y} & = & \tilde{\hat{y}}'\tilde{\hat{y}} + e'e \\ TSS & = & ESS + RSS \end{array}}$$

  - As long as there is a constant term in the regression
    - (Yet another reason for always including a constant term)

R^2 III

$$\boxed{TSS = ESS + RSS}$$

- This relationship is useful as it gives us two equivalent ways to define R^2:

$$\begin{aligned} R^2 &= \frac{\tilde{\hat{y}}'\tilde{\hat{y}}}{\tilde{y}'\tilde{y}} = \frac{ESS}{TSS} \\ &= 1 - \frac{e'e}{\tilde{y}'\tilde{y}} = 1 - \frac{RSS}{TSS} \end{aligned}$$

(as long as there is a constant term in the regression)

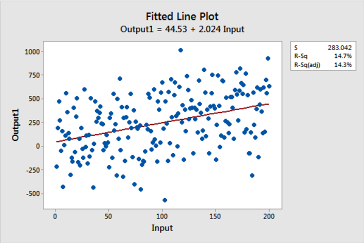
R^2 IV

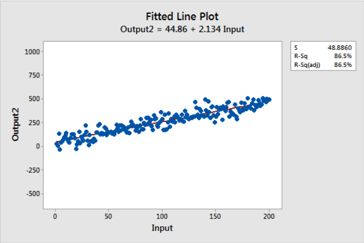
$$\boxed{\begin{aligned} R^2 &= \frac{ESS}{TSS} \\ &= 1 - \frac{RSS}{TSS} \end{aligned}}$$

- OK, those are the formulas
    - How about some intuition???

R^2 V

- If your regression line fits "more" of the data, then you will have a high R^2





  - The figure on the left has a relatively low R^2 of 0.143...
  - The figure on the right has a relatively high R^2 of 0.865...

R^2 VI

- The thought experiment underlying the use of R^2 is this:
  - Suppose you got a new x_i and predicted y_i for it $\rightarrow \hat{y}_i$
  - Do you think you'll get close to the true y_i?
  - Ideally: Yes ... if your R^2 is high
- This thought experiment also shows why R^2 is most relevant in a *prediction* context

Dangers of R^2 I

- There are many dangers associated with giving too much weight to a particular model's R^2
- These include:
  1. Different types of data imply different "typical" values of R^2:
     - ★ Cross-section data often have low R^2 (e.g. less than 0.20)
     - ★ Time-series data often have high R^2 (e.g. more than 0.80)
     - ★ Why?*

Dangers of R^2 II

Dangers of R^2, cont:

2. Different functional forms for y_i can change R^2
   - E.g., using $\log(y_i)$ instead of y_i in a regression will often increase the R^2 for that regression
     - (*regardless* of whether the true data generating process is better approximated with $\log(y_i)$)
     - (This is usually because $\log(y_i)$ simply has less variability than y_i)

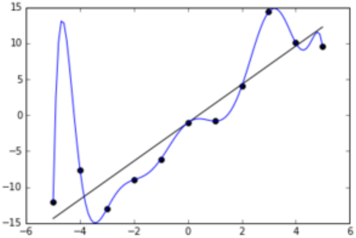
Dangers of R^2 III

Dangers of R^2, cont:

3. Adding explanatory variables to a model *always* increases R^2
   - Thus if the goal is to maximize R^2, one simply keeps adding covariates until you can *perfectly predict* your dependent variable
     - Namely if you have 100 observations and run a regression of y_i on $x_i'\beta$ with 100 elements in x_i', you will *perfectly predict* y_i
     - *Regardless* of whether all the covariates "make sense" or not
   - This is called *overfitting* your data

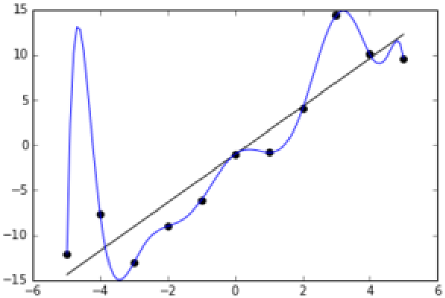
Dangers of Overfitting I

- Here is an example of overfitting:



  - The black line is the typical fitted OLS regression line (with $R^2 < 1$)
  - The blue line adds nonlinearities in x_i
    - ★ And perfectly predicts the data ($R^2 = 1$)

Dangers of Overfitting II



- But... if I were to approach you with a new value of x_i - for example $x_i = -4.4$ - and asked you to predict y_i...
  - Would you be more likely to believe the prediction from the black or blue line?

Adjusted R^2 I

- The dangers of overfitting make some researchers prefer an alternative measure of R^2 called *Adjusted* R^2
  - Denoted $\bar{R}^2$
- We said before that we could write R^2 as

$$R^2 \;=\; 1 - \frac{e'e}{\tilde{y}'\tilde{y}} \;=\; 1 - \frac{RSS}{TSS}$$

- Adjusted R^2 adjusts this formula for the number of "degrees of freedom" in each of the RSS and TSS

$$\bar{R}^2 \;=\; 1 - \frac{e'e/(N-K)}{\tilde{y}'\tilde{y}/(N-1)}$$

Adjusted R^2 II

$$\boxed{\bar{R}^2 \quad = \quad 1 - \frac{e'e/(N-K)}{\tilde{y}'\tilde{y}/(N-1)}}$$

- The Adjusted R^2 penalizes a regression for including variables that don't contribute (much) (if anything) to the overall explanatory power of the regression
  - In particular, if adding a covariate x_{ik} *didn't* reduce (much) the sum of squared residuals
  - Then $\bar{R}^2 \downarrow$
- Formally, you can show that $\bar{R}^2$ will increase whenever the absolute value of the t-statistic of the parameter associated with an included regressor (β_k) is greater than one in absolute value

Model Selection

Latex Color = "LightCyan"

Model Selection I

- The use of a degrees-of-freedom adjustment for $\bar{R}^2$ suggests that both it (and regular) R^2 can somehow be used for model selection
- This is correct...
  - ...both when the goal of the econometric analysis is prediction...
  - ...as well as to determine if one or more of a model's variables are individually significant
    - (i.e. for causal analysis)

Model Selection II

- There are two types of model selection methods:
  1. Those that rely on a model's *in-sample* properties
  2. Those that rely on a model's *out-of-sample* properties
- $\bar{R}^2$ relies on a model's in-sample properties
  - As do related methods in economics like the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)
    - ★ That formally rely on "information theory"
    - ★ (A type of analysis that combines math, stats, and other fields)
  - (I tell you these so that you've heard of them at least once...
    - ★ ...in case you run across them in your future endeavors)

Model Selection III

- If the goal of the model is *prediction*, then typically we are thinking about predicting well *out-of- sample*
  - i.e. Amazon.com wants to know if you'll buy product Y...
  - ...and uses data on people "like you"...
    - ★ i.e., people with values of X like you
    - ★ i.e., same gender, same age, same location, etc.
  - ...to predict if you'll buy product Y at price p_Y.
  - This is *out-of-sample* prediction

Model Selection IV

- This issue is particularly common in the analysis of Big Data
  - Where prediction is a very common application
    - (e.g., the Amazon.com example)
- The analysis of Big Data falls into the domain of several related disciplines
  - Statistics
  - Computer Science
    - The combination of which yields Computational Statistics and/or Machine Learning and/or Data Science
  - Economics (esp. Econometrics)

Model Selection V

- When evaluating a model, it is conventional among those who work on Big Data to *divide one's data* into separate datasets for
  1. Training
     - ★ (i.e. Estimation)
  2. Validation
     - ★ (i.e. Model selection)
  3. Testing
     - ★ (i.e. Evaluating how well the model has performed)
- (Tho the 2nd and 3rd steps are often combined)

Cross-Validation I

- Perhaps the most common method for model selection based on out-of-sample performance
  - ...at least among econometricians...
- ...is *cross-validation*

Cross-Validation II

k-fold cross validation involves the following steps:

1. Divide the data into k roughly equal subsets ("folds")
   - Index them by $s = 1, ..., k$
2. For each s:
   1. Fit your model using the $k - 1$ subsets *other than s*
   2. Predict your model on subset s and measure your loss, L_s
      - e.g. using the RSS, $L_s = e_s' e_s$
   3. Repeat for all other subsets
3. Evaluate the overall loss of the model
   - e.g. with $\bar{L} = \frac{1}{k} \sum L_s$
4. Select the model with smallest overall loss

Cross-Validation III

- Common choices for k include 10, 5, and the sample size minus 1
  - The latter being called "leave-one-out cross-validation"
- The so-called "test-train" cycle...
  - ...and cross-validation in particular...
  - ...are common techniques in the analysis of Big Data

Goodness-of-fit Conclusions I

- Other methods for model selection exist, e.g.
  - Ridge regression, "Least absolute shrinkage and selection operator"
    - Or "lasso" Picture
    - These are sometimes called "shrinkage estimators"
  - Principal components analysis (PCA) and Factor analysis
    - These are called "variable reduction methods"
- It's a big world out there...
  - ...and R^2 covers just the *very* beginnings of it re: goodness-of-fit and model selection
  - (And is growing in importance as an empirical undertaking)

Goodness-of-fit Conclusions II

- So...
- ...if prediction is your goal, *don't use* R^2!
  - Use cross-validation or some other technique based on out-of-sample model performance
- ...if causal analysis is your goal, *still don't use* R^2*!*
  - And in the rest of these slides we'll cover topics more relevant for that purpose
- Notably hypothesis testing of one or more elements of β
  - Next!

Interval Estimation and Hypothesis Testing

Interval Estimation and Hypothesis Testing Intro I

- OK, where are we:
  - We have an estimate of β, $\hat{\beta}$
    - (From OLS of course)
    - (With nice properties... under our assumptions)
  - We care about (at least) one element of the population parameter, β
    - e.g. β_k
  - We want to know something about β_k...
    - (Is it big or small?)
    - ...based on our estimate of $\hat{\beta}_k$

Interval Estimation and Hypothesis Testing Intro II

There are two ways we do so:

1. Interval Estimation
   - The basic idea: create an interval within which we expect β_k to live
   - How?
     1. Treat $\hat{\beta}_k$ as the best estimate of β_k
     2. Use (an estimate of) $V(\hat{\beta}_k)$ to construct lower and upper bounds within which we believe β_k lies with high probability.

Interval Estimation and Hypothesis Testing Intro III

Two ways we do so, cont.

2. Hypothesis Testing

   - The basic idea: ask "Is our estimate, β_k, 'compatible' with some specific value of the population parameter, β_k?"
   - How?
     1. Suppose β_k is a specific value (e.g. $\beta_k = 2$).
     2. Evaluate the probability of seeing the $\hat{\beta}_k$ we see under this assumption.
     3. If unlikely, reject the hypothesis that β_k is the specified value.

Interval Estimation and Hypothesis Testing Intro IV

- Q: Can we do these? A: Easy-peasy!
- We showed two decks ago (Top01b, Slide 66) that:

$$\boxed{\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})}$$

- Thus we already know the whole *distribution* of $\hat{\beta}$, i.e.
  - We know it's normally distributed:
    - ★ $N(\beta, \sigma^2(X'X)^{-1})$
  - We know its mean is β:
    - ★ $N(\beta, \sigma^2(X'X)^{-1})$
  - We know its variance is $\sigma^2(X'X)^{-1}$:
    - ★ $N(\beta, \sigma^2(X'X)^{-1})$

Interval Estimation and Hypothesis Testing Intro V

$$\boxed{\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})}$$

- We therefore know the distribution of any one of its elements, $\hat{\beta}_k$)
  - We know it's normally distributed... right?*
  - We know its mean is ______________________
  - We know its variance is ______________________
    - ______________________

Estimating $V(\hat{\beta})$

Latex Color = "LightCyan"

Estimating $V(\hat{\beta})$ I

- Therefore it's a simple matter to evaluate whether (e.g.) $\beta_k = 2$
  - Right?
- Well... almost
- While we may know *the formula* for $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$...
  - ...we can't actually *calculate* it...
    - ...as we don't yet know σ^2!
- So what to do?
  - *Estimate* σ^2 of course!
  - $\rightarrow$ An estimate of $V(\hat{\beta}) = \hat{V}(\hat{\beta})$

Estimating $V(\hat{\beta})$ II

- In practice, there are two dominant methods of estimating $V(\hat{\beta})$:
  1. "Use the formula"
     - (The "normal" way)
     - (Requires estimating σ^2 in $\sigma^2(X'X)^{-1}$)
  2. "Bootstrapping"
     - Estimate $V(\hat{\beta})$ using "resampling methods"
     - (More complicated logistically...
     - ...but will enhance your intuition)
- I'll show you the former...
  - ...and leave you to look up the other if you ever need it
  - (Which often happens in more complicated models)

Estimating σ^2 I

- So what should be our estimate of σ^2?
- Well where does it come from?
  - [Hint: $\epsilon_i \sim (0, \sigma^2)$]
    - (i.e. σ^2 is the variance of ϵ_i)
- Do we have an estimate of ϵ_i?
  - Sure! $e_i = \hat{\epsilon}_i = y_i - x_i'\beta$

Estimating σ^2 II

- OK, then let's just take the sample average of the e_i as our estimate of σ^2
  - Denote this s^2 or $\hat{\sigma}^2$
  - (Both are common)
- Indeed, a good estimate of σ^2 is

$$\begin{aligned} s^2_{MLE} &= \frac{1}{N}\sum e_i^2 && \text{in summation notation} \\ &= \frac{e'e}{N} && \text{in matrix notation} \end{aligned}$$

  - Indeed, this is the Maximum Likelihood estimate of σ^2
    - ★ (Hence I've labeled it s^2_{MLE})

Estimating σ^2 III

- In practice, we'll use a slightly different formula that adjusts for the degrees of freedom in our model

$$\boxed{s^2 = \frac{e'e}{N-K}}$$

  - Why? Because $Es^2_{MLE} = \frac{N-K}{N}\sigma^2$
    - i.e. s^2 is unbiased and s^2_{MLE} is not
    - (Not a big deal: both are consistent (the large-sample analog of unbiased))

Estimating σ^2 IV

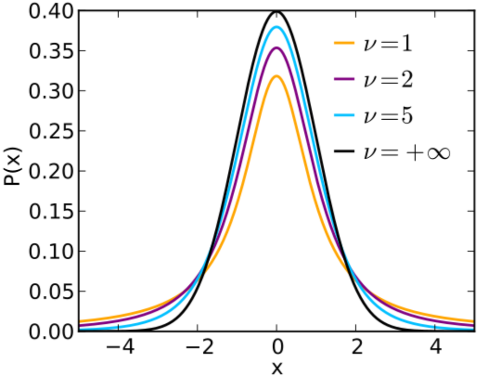
- We're now all set if we want to "use the formula" approach to do interval estimation and hypothesis testing
- We know:

$$\boxed{\begin{array}{ccl} \hat{\beta} & \sim & N(\beta, \sigma^2(X'X)^{-1}) \\ s^2 & = & \frac{e'e}{N-K} \end{array}}$$

Why t tests?

- Remember one point I brought up in a few weeks ago (Top01a, Slide24):
  - While $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ suggests doing interval estimation and hypothesis testing using a *Normal* distribution...
  - ...the fact that we have to estimate σ^2 with s^2 adds uncertainty...
  - ...so that we will do our IE and HT using the t distribution
    - ★ This is temporary: when we get to Instrumental Variables (IV) estimation next week, we will rely on large-sample results
    - ★ And in large samples $t_{N-K} \rightarrow N(0,1)$ Figure
    - ★ And $F_{q,N-K} \rightarrow \chi^2_q$

Student's t-distribution Redux



- Where ν = the degrees of freedom (i.e. $N - K$)
- t_∞ is the same as a standard Normal distribution

Back

Types of Hypotheses to Tests

Latex Color = "LightCyan"

Types of Hypotheses to Test I

OK, we want to test some hypotheses.

- What *types* of hypotheses shall we test?
  - The tools one uses depends on the answer to this question
- A typical taxonomy has three types of tests:
  1. Tests of a single hypothesis involving a single coefficient
     - e.g., $\beta_k = 2$

Types of Hypotheses to Test II

Types of hypotheses, cont:

- Typical taxonomy, cont:
  2. Tests of a single hypothesis involving multiple coefficients
     - ★ e.g., $\beta_j = \beta_k$
  3. Tests of multiple hypotheses involving either a single or multiple coefficients
     - ★ e.g., $\beta_j = 0$ *and* $\beta_k = 2$
     - ★ or $\beta_j = 0$ *and* $\beta_k = \beta_l$

Types of Hypotheses to Test III

Types of hypotheses, cont:

- For the first two types of tests (of single hypotheses), we use t (z) tests
- For the third type of tests (of multiple hypotheses), we use F (χ^2) tests
- We'll start with Type (1): single tests of single hypotheses
  - Finally (!): Q: How do we construct a confidence interval???
    - A: Remember what you did in the exercise session!
    - (I'll summarize it here)

Tests of a Single Hypothesis involving a Single Coefficient

Tests of a single hypothesis involving a single coefficient

- I'm always uncertain how much detail on these topics to show you
- So: I'll assume you've seen everything in this subsection before
  - Plus Matteo covered this in the exercise sessions
- *If that's not enough*, pull out one of the econometrics textbooks listed in the syllabus
  - They all have a review chapter on statistics (or almost all)
  - That will include a review of the basics of interval estimation and hypothesis testing
  - (At the level of testing a single hypothesis involving a single coefficient)
- I've kept some review slides in the deck that I won't cover, noted by *

The Basics of Interval Estimation I

- An interval estimate of a population parameter, β_k, consists of two bounds within which we expect β_k to lie, i.e.
  - $LB \leq \beta_k \leq UB$
    - Where LB and UB are the lower and upper bounds
- The probability that β_k lies within the provided interval estimate is called the confidence coefficient
  - It is denoted $1 - \alpha$
    - Where α is the *significance level* of the confidence interval
- Usually $\alpha = 0.05 \Rightarrow$ a 95% confidence interval
  - $P(LB \leq \beta_k \leq UB) = 1 - \alpha = 95\%$

The Basics of Interval Estimation II*

- How? Three steps:
  1. Find the lower and upper bounds for a t distribution with $N - K$ degrees of freedom (denoted "dof" or "df") for a given α
     - ★ We will call these lower/upper bounds "critical values"
     - ★ (And use them again in hypothesis testing)
  2. "Standardize" the distribution of the population parameter
     - ★ (Recall Topic 1a, Slide 17)
  3. Plug (2) into (1) to obtain a confidence interval for β_k

Hypothesis Testing: Intro

- In econometrics, rather than seeking a range of values for a population parameter, we often want to know if its consistent with a specific value.
    - Example: In an antitrust case, an econometrician estimates the cross-price elasticity of demand for the products of two firms that wish to merge (i.e. how much the quantity of one good responds to a change in price of another).
        - ★ If this cross-elasticity is greater than 2.5, the government will find the merger to be anti-competitive and sue to block it!
- You'll see lots more examples:
    - In the slides
    - On the problem sets
    - On the exam

The Basics of Hypothesis Testing I

There are four steps (in two flavors) to constructing a hypothesis test

1. Explicitly define two opposing hypotheses
   - The *Null Hypothesis* and the *Alternative Hypothesis*
   - Denoted H_0 and H_1
     - (Where H_1 is sometimes also written H_A)
2. Construct a test statistic under the null hypothesis
   - Usually the standardized distribution of our estimate of a population parameter
   - (I denote our test statistic $\tilde{t}$)

- The next steps depend on which of two paths we take

The Basics of Hypothesis Testing II

If we take what I call the "critical value" path:

3. Select a level of significance, α, and calculate its critical value
   - (As we did for confidence intervals a few slides ago)
4. Compare (the absolute value of) our test statistic to the critical value and...
   - ...either (a) reject or (b) "fail to reject" H_0
     - Note we never "accept" H_0, we just "fail to reject" it.
   - We use the absolute value as either either large positive or large negative values of the test statistic can cause you to reject

The Basics of Hypothesis Testing III

If we take what I call the “p-value” path:

3. Calculate the probability of getting a value of the test statistic as big as it is (in absolute value) if the null hypothesis is true
    - This is the hypothesis test’s *p-value*
4. Compare this probability (p-value) to a desired level of significance...
    - ... and either (a) reject or (b) fail to reject H_0

- Recall the “beer tax” exam question from the exercise session
    - It showed how to do both approaches

One-sided hypothesis tests

- We sometimes concern ourselves with one-sided alternatives when testing hypotheses.
  - If you are testing $H_0 : \beta_k < 0$ against $H_1 : \beta_k > 0$, then...
    - Only values of $\hat{\beta}_k$ *larger* than 0 would ever cause you to reject H_0
    - (Obviously)
  - Similarly if you are testing $H_0 : \beta_k > 0$, then only values of $\hat{\beta}_k$ smaller than 0 would ever cause you to reject H_0
- While obvious, students often forget this feature once they standardize $\hat{\beta}_k$ to calculate test statistics!

Intuition for Hypothesis Tests

Latex Color = "LightCyan"

Intuition for Hypothesis Testing I

- Before we jump into other - more complicated - hypothesis tests, let me give you some intuition for just what it is we are doing when we calculate a confidence interval or run a hypothesis test
  - I'll introduce it via the t-tests we've just covered...
    - ...but it applies more generally to all of hypothesis testing

Intuition for Hypothesis Testing II

- Recall a few things you knew before we got to hypothesis testing:
  - We can estimate $\hat{\beta}$ using OLS
  - Under the CLRM assumptions,
    - ★ $\hat{\beta}_k$ measures the causal effect of x_{ik} on y_i (Good!)
    - ★ $\hat{\beta}$ is unbiased ($E(\hat{\beta}) = \beta$)
    - ★ $V(\hat{\beta})$ is efficient among linear unbiased estimators
    - ★ $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$
    - ★ We can estimate σ^2 with $s^2 = \frac{1}{N-K}e'e$
    - ★ We can therefore estimate $V(\hat{\beta})$ with $\hat{V}(\hat{\beta}) = s^2(X'X)^{-1}$
- Not bad!

Intuition for Hypothesis Testing III

- The most common use of regression is to measure *causal effects*
- We therefore want to know about a particular β_k
  - Or perhaps a few particular elements of β, i.e. β_j and β_k
- Why?
  - Because if we've identified causal effects, then if I - as a business-person or policymaker - am able to change the value of x_{ik}...
    - ...I want to know what will be the change in the expected value of y_i!
  - (And there are an infinity of questions of this nature)

Intuition for Hypothesis Testing IV

More generally, for a causal effects analysis, we want to know the answers to two key questions:

1. Is β_k *economically* significant?
   - i.e. if I change x_{ik} by one unit, will I get an *economically meaningful* effect on y_i?
     - ★ This is evaluated by looking at the *magnitude* of $\hat{\beta}_k$

Intuition for Hypothesis Testing V

Two key questions, cont.:

2. Is β_k *statistically* significant?
   - i.e. my estimate of β_k is $\hat{\beta}_k$, but I understand that this is just an *estimate*
     - In another sample I would get another estimate
   - How can I be sure that the true β_k isn't somehow very different from my current estimate $\hat{\beta}_k$?
     - Answer: Hypothesis testing

Intuition for Hypothesis Testing VI

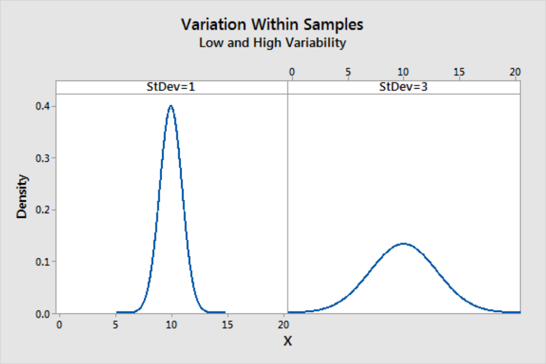
- *How* does hypothesis testing do this?
  - i.e., What's the *intuition*???
- Well, as we established earlier, not only do we obtain an estimate of β_k with $\hat{\beta}_k$
  - We also obtain an estimate of its *variance*, $\hat{V}(\hat{\beta}_k)$
- And we know that variance measures the *spread* in a distribution

Intuition for Hypothesis Testing VII

- So if $\hat{V}(\hat{\beta}_k)$ is *small*...
  - We know that the true β_k is likely to be very close to our estimate $\hat{\beta}_k$
    - (i.e. we will have a small confidence interval for β_k)
- And if $\hat{V}(\hat{\beta}_k)$ is *big*...
  - We know that the true β_k may be far from our estimate $\hat{\beta}_k$
    - (i.e. we will have a large confidence interval for β_k)

Intuition for Hypothesis Testing VIII

- Suppose $\hat{\beta} = 10$ and consider two possible distributions for $\hat{V}(\hat{\beta})$
  - (In the left figure, this distribution is centered around $\hat{\beta} = 10$...
  - ...but we're going to adjust that momentarily)



Intuition for Hypothesis Testing IX

- Hypothesis testing merely *formalizes* the idea that when $\hat{V}(\hat{\beta})$ is small, the true β is likely to lie close to $\hat{\beta}$
- It does so by asking "Could β_k be a *particular* value?"
  - Usually a value we care about, e.g. 0
- Quick Quiz: Why often zero???*
  1. ______________________________
  2. ______________________________

Intuition for Hypothesis Testing X

- Hypothesis testing formalizes this intuition using a particular sequence of steps:
  1. It says, OK, suppose H_0 is true
     - ★ i.e. suppose β_k really is zero (or seven, or whatever)
  2. It then asks, how likely is it that I would get the value of $\hat{\beta}_k$ that I get...
     - ★ ...if the *true* β_k is zero (or seven, or whatever)?
     - ★ (Indeed this likelihood is the p-value for the test)

Intuition for Hypothesis Testing XI

- For example, in the earlier figure (Figure)...
  1. Suppose β_k really is equal to 7
  2. How likely is it that I would get a value of $\hat{\beta}_k = 10$ if the true β_k is 7?
- To answer this, imagine estimating $\hat{\beta}_k$ many times on many different samples
  - If the true β_k is 7, there should be a distribution of $\hat{\beta}_k$ centered on 7
  - (So imagine shifting the distribution on the earlier slide so that it is centered on 7)

Intuition for Hypothesis Testing XII

- And so then:
  - How likely is it that I would get a value of $\hat{\beta}_k = 10$ if the true β_k is 7?
- The answers depend on our estimate of $\hat{V}(\hat{\beta}_k)$
  - In the first graph? (Figure) Not very likely (using "eye-conometrics", less than 1%)
    - ★ Thus reject H_0: β_k is unlikely to truly be equal to 7
  - In the second graph? Reasonably likely (maybe 20%?)
    - ★ Thus fail to reject H_0: β_k really could equal 7

Intuition for Hypothesis Testing XIII

- Wait! I'm not usually testing 7 versus 10, I'm always testing something around 1.96!
- But this is no big deal!
  - It is simply because we always *standardize* our tests
    - ★ If we didn't, we'd have to calculate critical values all the time depending on our specific distribution of $\hat{\beta}_k$
    - ★ (Which would be a pain in the neck)

Intuition for Hypothesis Testing XIV

- When we standardize our test we can know instantly whether the test statistic is above the "regular" critical values for common levels of significance:
  - For $\alpha = 0.05$, $\bar{t}_{two-sided} = 1.96$
  - For $\alpha = 0.10$, $\bar{t}_{two-sided} = 1.65$
    - ★ Noting also that for $\alpha = 0.05$, $\bar{t}_{one-sided} = 1.65$
  - For $\alpha = 0.01$, $\bar{t}_{two-sided} = 2.33$
    - ★ Easier to remember 3 numbers than have to calculation test-specific critical values all the time!

Intuition for Hypothesis Testing XV

- But don't focus on the standardization...
  - ...focus on the conceptualization
- Two steps:
  1. Assume $H_0 : \beta_k = val$ is true
  2. Ask how likely we'd get a value of $\hat{\beta}_k$ if indeed the true β_k is *val*
- If unlikely, e.g. if unlikely that $\beta_k = 0$,

Then whatever is the economic significance of $\hat{\beta}_k$, we can also have (statistical) confidence that this economic significance is "real"

(And *that's* often the goal of econometrics!)

Tests of a Single Hypothesis involving Multiple Coefficients

Other Hypothesis Tests Intro

- Recall the three types of hypothesis tests from slides 54 and 55 above. We began with
  - (1) Tests of a single hypothesis involving single coefficients
- It's time now to consider the other possibilities:
  - (2) Tests of a single hypothesis involving multiple coefficients
    - e.g., $\beta_j = \beta_k$
  - (3) Tests of multiple hypotheses involving a single or multiple coefficients
    - e.g., $\beta_j = 0$ *and* $\beta_k = 2$
    - or $\beta_j = \beta_k = \beta_l = \beta_m$

Single hypothesis involving multiple coefficients I

- Tests of single hypotheses involving multiple coefficients...
  - ...are still t tests
    - ★ They are just t tests where deriving the distribution of the test statistic is a little more complicated

Single hypothesis involving multiple coefficients II

- Consider a regression of wages on both education and experience:

$$\begin{aligned} wage_i &= x_i'\beta + \epsilon_i \\ &= \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i'\tilde{\beta} + \epsilon_i \end{aligned}$$

  where we've split out the variables we care most about - $educ_i$ and exp_i - from a general set of covariates x_i

  - (Labeling $\beta_2 = \beta_{educ}$ and $\beta_3 = \beta_{exp}$) and
  - (Labeling the remaining covariates and parameters with ˜)

- Suppose Assumption 2 (Mean-zero Error) holds so that we can make causal statements

Single hypothesis involving multiple coefficients III

$$wage_i = \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i$$

- Further suppose you are a policymaker and are wondering whether education or experience is more important for someone's wage
  - (Long-run wage? Lifetime earnings? Let's abstract from these - important - questions...)
- To address this question, you obtain a sample of data and run the above regression
- Quick Quiz: How would you test your assertion?*
  - ______________________________

Single hypothesis involving multiple coefficients IV

- Recall Step 2: "Construct a test statistic under the null hypothesis"
  - When we were testing $\beta_1 = 2$, we used the distribution of $\hat{\beta}_1$ to construct our test statistic
  - What distribution shall we use to construct *this* test statistic?*
    - ____
- This is a good general rule:
  - Set up your hypothesis tests so that you have
    - Functions of parameters on the left-hand-side $(\beta_{educ} - \beta_{exp})$
    - ...equaling $(=)$...
    - Known values on the right-hand-side (0)

Single hypothesis involving multiple coefficients V*

- So what then is the distribution of $\hat{\beta}_{educ} - \hat{\beta}_{exp}$?
- Easy. We can use our rules for linear combinations of random variables...
  - (Topic01a, Slide 36)
  - [Did you think I told you all that stuff for nothing? :-)]
- ...to derive this distribution
  - Tho before I do, note that

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad \text{is a } K \times K \text{ } \textit{matrix}$$

$$\Rightarrow \quad V(\hat{\beta}_k) = \sigma^2(X'X)^{-1}_{kk}$$

$$\text{and} \quad Cov(\hat{\beta}_j, \hat{\beta}_k) = \sigma^2(X'X)^{-1}_{jk}$$

Single hypothesis involving multiple coefficients VI*

Let's derive the distribution of $\hat{\beta}_j - \hat{\beta}_k$:

$$
\begin{aligned}
E[\hat{\beta}_j - \hat{\beta}_k] &= E(\hat{\beta}_j) - E(\hat{\beta}_k) \\
&= \beta_j - \beta_k \\
V[\hat{\beta}_j - \hat{\beta}_k] &= V(\hat{\beta}_j) + V(\hat{\beta}_k) - 2Cov(\hat{\beta}_j, \hat{\beta}_k) \\
&= \sigma^2 (X'X)^{-1}_{jj} + \sigma^2 (X'X)^{-1}_{kk} - 2\sigma^2 (X'X)^{-1}_{jk} \\
\Rightarrow \quad \hat{\beta}_j - \hat{\beta}_k &\sim N(\beta_j - \beta_k, \sigma^2((X'X)^{-1}_{jj} + (X'X)^{-1}_{kk} - 2(X'X)^{-1}_{jk}))
\end{aligned}
$$

Single hypothesis involving multiple coefficients VII*

We can now derive our test statistic for $\hat{\beta}_{educ} - \hat{\beta}_{exp}$ under H_0:

$$
\begin{aligned}
& \hat{\beta}_{educ} - \hat{\beta}_{exp} \quad \sim \quad N(\beta_{educ} - \beta_{exp}, \sigma^2((X'X)^{-1}_{2,2} + (X'X)^{-1}_{3,3} - 2(X'X)^{-1}_{2,3})) \\
\Rightarrow \quad & \hat{\beta}_{educ} - \hat{\beta}_{exp} \quad \sim \quad N(0, \sigma^2((X'X)^{-1}_{2,2} + (X'X)^{-1}_{3,3} - 2(X'X)^{-1}_{2,3})) \qquad \text{under } H_0 \\
& \Rightarrow \quad \frac{\hat{\beta}_{educ} - \hat{\beta}_{exp} - 0}{\sqrt{\sigma^2(X'X)^{-1}_{2,2} + \sigma^2(X'X)^{-1}_{3,3} - 2\sigma^2(X'X)^{-1}_{2,3}}} \quad \sim \quad N(0,1) \\
& \Rightarrow \quad \frac{\hat{\beta}_{educ} - \hat{\beta}_{exp} - 0}{\sqrt{s^2(X'X)^{-1}_{2,2} + s^2(X'X)^{-1}_{3,3} - 2s^2(X'X)^{-1}_{2,3}}} \quad \sim \quad t_{N-K}
\end{aligned}
$$

Single hypothesis involving multiple coefficients VIII*

And we're in a position to finish up:

- Step 3a: Select a level of significance and calculate its critical value
  - Let's use (as usual) $\alpha = 0.05$ and calculate the critical value for a (one-sided? two-sided?) test with $N - K = 616 - 6 = 610$ df
    - ★ (Two-sided reasonable as we don't have any particular reason to think one parameter is likely to be larger than the other)
    - ★ $\rightarrow \bar{t} = 1.964$
  - Plug in the necessary values to calculate it:

See Stata Example on OLAT

$$\frac{\hat{\beta}_{educ} - \hat{\beta}_{exp} - 0}{\sqrt{s^2(X'X)^{-1}_{2,2} + s^2(X'X)^{-1}_{3,3} - 2s^2(X'X)^{-1}_{2,3}}} \sim t_{N-K}$$

$$\Rightarrow \quad \tilde{t} = \frac{0.98399 - 0.47425}{\sqrt{0.02378445 + 0.00570468 - 2 * .000002457}} = \frac{0.50974}{0.17171} = 2.969$$

Single hypothesis involving multiple coefficients IX*

Finishing up,cont:

- Step 4a: Compare our test statistic with the critical value and...
  - Reject as $|2.969| > 1.964$
    - (Verify you reach the same conclusions under the p-value approach)
    - (With $p = 0.0031$)

Tests of Multiple Hypotheses

Latex Color = "LightCyan"

Multiple hypotheses involving single coefficients I

- Tests of multiple hypotheses involving single coefficients...
  - ...are F tests
    - (Which you may recall is the small-sample analog to χ^2 tests)
    - (From Topic01a, Slides 20-23)

Multiple hypotheses involving single coefficients II

- Conducting tests of multiple hypotheses is analogous to the general intuition I just gave:
  1. Assume H_0 is true
     - ★ Now (only) slightly more complicated as need to set more than one parameter to a particular value
     - ★ e.g. $\beta_j = 0$ *and* $\beta_k = 0$
  2. Ask how likely we'd get the values of $\hat{\beta}_j$ and $\hat{\beta}_k$ that we actually got if indeed H_0 is true
- The challenge when testing multiple hypotheses is...
  - ...what do we evaluate?

Multiple hypotheses involving single coefficients III

- What do you think? What should we evaluate???*
  - ______________________________
  - ______________________________
  - ______________________________

The F test I

- This is exactly what an F-test does!
- Let me show you in the context of our regression of wages on both education and experience:

$$
\begin{aligned}
wage_i &= x_i'\beta + \epsilon_i \\
&= \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i'\tilde{\beta} + \epsilon_i
\end{aligned}
$$

  - It turns out only one of our four steps to a hypothesis test is different relative to the t tests we showed earlier

The F test II

Remember our four steps to a hypothesis test

1. Step 1: Specify H_0 and H_1:
   - e.g., $H_0 : \beta_{educ} = 0$ and $\beta_{exp} = 0$ v $H_1 : \beta_{educ} \neq 0$ *or* $\beta_{exp} \neq 0$
     - i.e. neither education nor experience impacts wages
     - (Unrealistic I know, but I want to show you the tools here)
     - (We'll test more realistic hypotheses soon...)
   - Note the "or":
     - If *either* $\beta_{educ} \neq 0$ OR $\beta_{exp} \neq 0$, we will reject H_0
     - So even if *one part* of H_0 is true, that's not enough: *both parts* have to be true

The F test III

Our four steps, cont:

1. Step 2: Construct a test under the null hypothesis
    - For a t-test, this was the standardized distribution of $\hat{\beta}_k$
        - Or perhaps the standardized distribution of a (single) function of multiple $\hat{\beta}_k$
    - It's different for the F test

The F test IV

Step 2, cont:

- The test statistic for an F test relies on estimating two models:
  1. The Restricted model
     - i.e. the model that imposes H_0
     - For us, the model that imposes $\beta_{educ} = 0$ and $\beta_{exp} = 0$
  2. The Unrestricted model
     - i.e. the model that doesn't impose H_0
     - For us, this is just the regular model letting β_{educ} and β_{exp} be free parameters

The F test V

Step 2, cont:

- So for our regression,
  - The unrestricted model (U) is our original model:

$$\text{Unrestricted:} \quad wage_i = \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i$$

  - The restricted model (R) is:

$$\begin{aligned}\text{Restricted:} \quad wage_i &= \beta_1 + \underbrace{\beta_{educ}}_{=0} educ_i + \underbrace{\beta_{exp}}_{=0} exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i \\ &= \beta_1 + \tilde{x}_i' \tilde{\beta} + \epsilon_i\end{aligned}$$

The F test VI

$$\boxed{\begin{aligned} \text{Unrestricted:} \quad wage_i &= \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i \\ \text{Restricted:} \quad wage_i &= \beta_1 + \tilde{x}_i' \tilde{\beta} + \epsilon_i \end{aligned}}$$

- The intuition of an F test is straightforward:
  - If H_0 is true, then imposing that $\beta_{educ} = 0$ and $\beta_{exp} = 0$ is the *right thing to do*
    - i.e., it was a mistake to include them in our unrestricted model
  - Then estimating the restricted model should fit the data just as well as the unrestricted model
  - Which we evaluate by comparing the fit of each:
    - Using either R^2 or the sum of squared residuals (SSR)
  - If H_0 is true, then $R_U^2 \approx R_R^2$ ($SSR_U \approx SSR_R$)
    - And if they are *not* (statistically) equal, then H_0 *isn't* true

The F test VII

Step 2, cont:

- Formally the F statistic is defined as:

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(N-K)}$$

where

- SSR_R (SSR_U) is the sum of squared residuals from the Restricted (Unrestricted) model
  - i.e. $SSR_R = e_R' e_R$, $SSR_U = e_U' e_u$
- q is the number of restrictions in H_0
  - For us, two: $\beta_{educ} = 0$ and $\beta_{exp} = 0$

The F test VIII

Step 2, cont:

- Using the formula for R^2 (with a constant!), you can show that the F statistic can equivalently be defined as

$$F = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(N-K)}$$

where R_R^2 and R_U^2 are defined as you think they should be

The F test IX

$$\boxed{\begin{aligned} F &= \frac{(SSR_R - SSR_U)/q}{SSR_U/(N-K)} \\ &= \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(N-K)} \end{aligned}}$$

- An important question: can the F statistic be negative?*
  - ______________________________

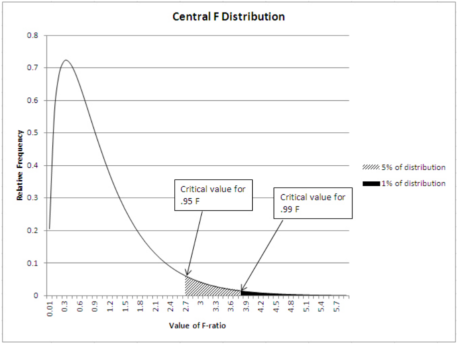
The F test X

- Under our assumptions, you can show that our statistic is distributed as an F distribution with q "numerator" degrees of freedom and $N - K$ "denominator" degrees of freedom
  - (Well, you can show it if you go get a PhD)
    - ★ (As the ratio of two independent χ^2 distributions is distributed as an F)
    - ★ (And you'll have to trust me that that's what we have!)

The F test XI

Our four steps, continued (belatedly):

- Step 3a: Select a level of significance and calculate its critical value
  - Since the F is always positive, there are *only* one-sided tests for an F



The F test XII

Our four steps, continued (belatedly):

- Step 4a: Compare (the absolute value of) our test statistic to the critical value and either (a) reject or (b) "fail to reject" H_0
  - With analogous versions of steps (3b) and (4b) if one wishes to take the "p-value" approach

See Stata Example on OLAT

The F test Comments I

- We've only shown how to do an F test when imposing that $\beta_k = 0$
  - But suppose in our example you wanted to test $\beta_{educ} = \beta_{exp} = 0.30$
    - ★ i.e. $\beta_{educ} = 0.30$ and $\beta_{exp} = 0.30$
- Following the same logic as earlier, the unrestricted model is the same and the restricted model is only slightly different:

$$\begin{aligned} \text{Unrestricted:} \quad wage_i &= \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i \\ \text{Restricted:} \quad wage_i &= \beta_1 + 0.30 educ_i + 0.30 exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i \end{aligned}$$

The F test Comments II

$$\text{Unrestricted:} \quad wage_i = \beta_1 + \beta_{educ} educ_i + \beta_{exp} exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i$$

$$\text{Restricted:} \quad wage_i = \beta_1 + 0.30 educ_i + 0.30 exp_i + \tilde{x}_i' \tilde{\beta} + \epsilon_i$$

- But... *how do you run the second regression?!?**
  - ____

See Stata Example on OLAT

The F test Comments III

- One last thing of interest:
  - When $q = 1$, we normally use a t test but could also do an F test
  - If you do, you'll find that your F-statistic is the *square* of your t-statistic
    - (And you'll have the *identical* p-value for your test)
    - (Basically two sides of the same coin)

See first Stata Example (testing $\beta_{educ} = \beta_{exp}$) on OLAT

Table of Contents