Empirical Methods

Topic 1a:

Probabilistic and Statistical Foundations of the CLRM

Econometrics as regression I

- The most common tool in the econometrics toolkit is a (linear) regression.
- Regression is a statistical technique that attempts to "explain" movements in one variable, the *dependent variable*, denoted y_i,...
  - ...based on movements in a set of other variables, the *explanatory variables*, denoted $x_{i1}, x_{i2}, ..., x_{iK}$.
    - ⋆ As suggested, there are K explanatory variables
    - ⋆ I will write x_i to indicate the set of all of the x_{ik}
    - ⋆ (In the exercise sessions this week, Matteo will spend some time on notation)

Econometrics as regression II

In the introductory slides last lecture, I introduced three common uses of regression:

1. Descriptive analysis
   - Think of this as "sophisticated data summary"
2. Causal analysis
   - Trying to measure the *causal* effect of a particular x_{ik} on y_i
   - (IMO, the majority of academic and policy uses of econometrics have a causal question in mind)
3. Prediction
   - Trying to best predict y_i based on a (possibly very large) set of x_i...
     - ...without being particularly concerned about the impact of any one x_{ik}
   - (The majority of Big Data topics have a goal of predicting something)

Econometrics as regression III

- In this course, the default use of regression will be Use $\#$*
  - *
    - *

Slide Organization Aside

- (Look at the top of the slide)
  - (There is a structure to the lecture notes that is revealed there)
  - (The last slide reveals an outline for the notes) Outline

Probabilistic Foundations: The foundations

Latex Color = "LightGrey"

Probability and Statistics

- Regression relies on certain results from the mathematical disciplines of Probability and Statistics
  - Probability $\equiv$ The branch of mathematics concerned with the likelihood that an event will occur
  - Statistics $\equiv$ The branch of mathematics concerned with collecting and analyzing numerical data in large quantities, especially for the purpose of inferring properties of a population from those in a representative sample

Probability and Statistics Plan I

- In what follows, I will list - and provide examples of - a few of the basic topics in probability and statistics that we rely on extensively in econometrics
  - I will assume that (a) you have seen these things in previous courses and (b) therefore know them
- Of course, you may have forgotten them all!
  - As such, the first Exercise Sessions (check with MG: later this week) will go over some of the details I will skip today

Probability and Statistics Plan II

- All well and good, but...
  - If - after the review - you still feel you're below par on the covered topics
    - ⋆ Teach them to yourself (they're not hard)
    - ⋆ (or ask your TA for some help) *

Random Variables I

- In econometrics, we treat both y_i and x_i in a regression of y_i on x_i as *random variables*
  - $\equiv$ a variable whose possible values are numerical outcomes of (what we take to be) a random phenomenon
  - (where, remember, x_i is a set of K different x_{ik}'s)

Random Variables II

- Random variables come in several varieties
  - Univariate v Multivariate
    - Univariate = only one RV
    - Multivariate = more than one RV (possibly related)
  - Discrete v Continuous
    - Discrete = taking discrete values
    - Continuous = taking a range of values
- Regardless of type, RVs are defined by their probability density function (PDF)
  - And their associated cumulative distribution function (CDF)

Random Variables Example

- Let y_i be student i's final grade in Econ 21
- Let x_i be the number of hours student i spends working on the course over the semester
  - Call this "hours spent on the course"
  - (In this case, there would be only one variable in x_i, i.e. $K = 1$)
- The probability y_i take a particular value is denoted $P(y_i = y_r) = p_r$
  - e.g., $P(y_i = 5.5) = 0.124$ or 12.4%

Properties of RVs

- We often wish to summarize information about...
  1. The probability distribution of a single (i.e. univariate) RV
  2. The relationship within the (joint) probability distribution of two or more (i.e. multivariate) RVs

Properties of Univariate RVs

The most common properties of univariate RVs we care about are

- Its *mean* (or *Expected Value*), denoted $E(x_i)$ or Ex or μ_x
  - A measure of the center of a distribution
- Its *variance* (denoted $V(x_i)$ or Vx or σ_x^2)
  - A measure of the "spread" of a distribution around its mean
  - A second measure of spread is a RV's *standard deviation*
    - Given by the square root of the variance, therefore denoted σ_x

Note here units (as it helps us later with intuition):

- μ_x and σ_x are measured in whatever are the units of x_i
  - (whereas σ_x^2 is measured in the square of the units of x_i)

Linear Rules for Univariate RVs

- In econometrics, we often use the following rules relating means and variances of random variables
  - Let x_i be a random variable with mean Ex and variance Vx
  - Let $y_i = a + bx_i$
- Then

$$E(y_i) = a + bE(x_i) \qquad \begin{aligned} V(y_i) &= V(a + bx_i) \\ &= V(bx_i) \\ &= b^2 V(x_i) \end{aligned}$$

- In sum:

$$\boxed{Ey = a + bEX} \qquad \boxed{Vy = b^2 Vx} \qquad \boxed{\sigma_y = b\sigma_x}$$

Intuition for these formulas*

* = [Space for you to take notes from our classroom discussion]

See Figure in Class

Common Random Variables I

In econometrics, we often analyze the following RVs:

1. The most common is the Normal distribution
   - $x_i \sim N(\mu_x, \sigma_x^2)$
     - $\equiv x_i$ is distributed as a Normal RV with mean μ_x and variance σ_x^2
   - If $\mu_x = 0$ and $\sigma_x^2 = 1$, you get a "standard normal"
     - Denoted $z \sim N(0, 1)$
   - If $x_i \sim N(\mu_x, \sigma_x^2)$, then you can "standardize" it,
     - i.e. turn it into a standard normal as follows:

$$z = \frac{x - \mu_x}{\sigma_x} \sim N(0, 1)$$

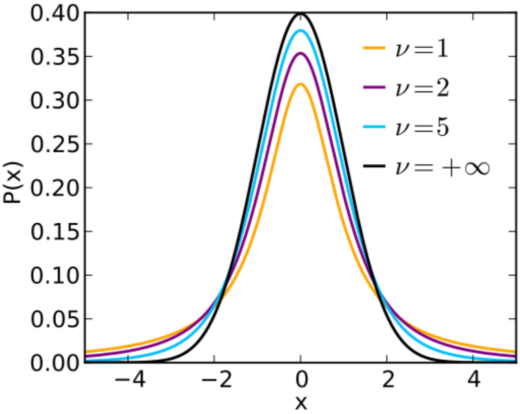
     - This formula is frequently used to form hypothesis tests in econometrics

Common Random Variables II

Other common RVs in econometrics:

2. The ("Student's") t-distribution with v degrees of freedom
   - Denoted $x_i \sim t_\nu$
     - ★ Similar to the Normal distribution but with "fatter tails"
   - If have N observations and K covariates in a regression, then have $v = N - K$ "degrees of freedom"

Student's t-distribution

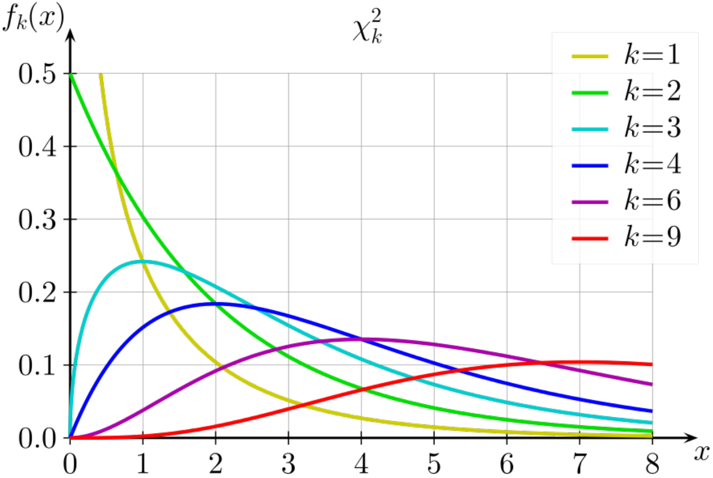


Common Random Variables III

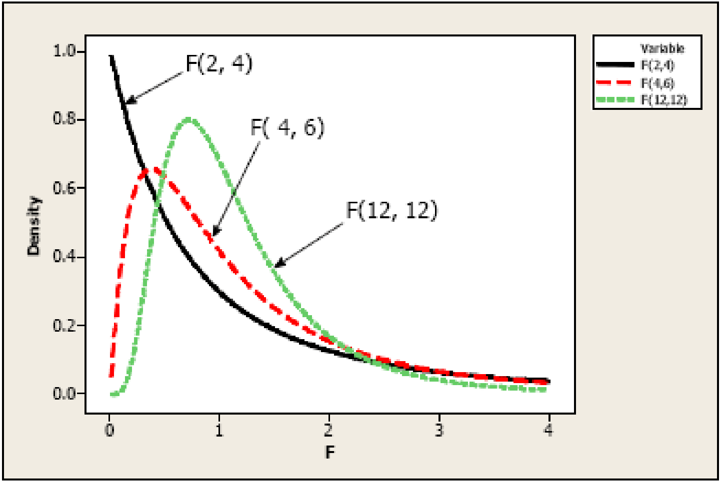
Other common RVs in econometrics, cont:

3. The χ^2 distribution (with k degrees of freedom)
    - In econometrics, k usually corresponds to the number of restrictions in the hypothesis we're testing
4. ("Snedecor's")* F-distribution
    - Denoted $x_i \sim F(v_1, v_2)$
    - With v_1 (numerator) d.o.f. and v_2 (denominator) d.o.f
    - Is the small-sample analog of the χ^2 distribution
        - Where v_1 corresponds to k, the number of restrictions, and v_2 corresponds to $N - K$, the number of degrees of freedom in the regression

χ^2 distribution



F distribution



Common Random Variables: Wrapup I

- In econometrics, "F is to χ^2 as t is to Normal"
- In other words:
  - We (mostly) rely on two kinds of tests in econometrics
    1. Tests of single hypotheses, e.g. $H_0 : \beta_1 = 0$
    2. Tests of multiple hypotheses, e.g. $H_0 : \beta_1 = 0, \beta_2 = 0, ..., \beta_k = 0$
- For each kind of test, there are two distributions one *could* use:
  - For (1), we use the Normal or t distribution
    - (Depending on whether you know σ^2 or have to estimate it)
  - For (2), we use the χ^2 or F distribution
    - (Ditto)

Common Random Variables: Wrapup II

- In this course I won't make a conceptual distinction between t and Normal random variables
  - Nor between F and χ^2 RVs
- Why not???*
  - ________________________________

Probabilistic Foundations: Multivariate RVs

Multivariate Random Variables I

- In econometrics, we almost always care about the relationship *between* random variables
  - e.g., if one has K random variables, $x_{i1}, \ldots, x_{iK}$, its probability distribution is called a multivariate probability distribution
    - Here, a K-variate probability distribution
- Each of the RVs in a multivariate distribution can be discrete or continuous
  - And the joint distribution has a (multivariate) PDF

Multivariate Random Variable Example

- The probability x_i and y_i take particular values is denoted $P(x_i = x_s, y_i = y_r) = p_{sr}$, e.g.
  - $P(x_i = male, y_i = 5.5) = 0.061$ or 6.1% and
  - $P(x_i = female, y_i = 5.5) = 0.063$ or 6.3%
- And similarly for all other combinations of gender and Econ 21 grades

Multivariate Random Variables II

- In econometrics, we rely on several concepts and properties of multivariate RVs that aren't present for univariate RVs:
    1. Marginal distributions
    2. Conditional distributions
    3. Covariance and correlation
    4. Properties of linear combinations of RVs

Marginal Probability Distributions

- Given a joint probability distribution of K random variables, one can always get a distribution of one of its components by adding up over all of the other components
  - This is its "marginal distribution"
- For example,

$$\begin{aligned} P(y_i = 5.5) &= \sum_s P(x_i = x_s, y_i = 5.5) \\ &= P(x_i = male, y_i = 5.5) + P(x_i = female, y_i = 5.5) \\ &= 0.061 + 0.063 \\ &= 0.124 \text{ or } 12.4\% \end{aligned}$$

Conditional Probability Distributions I

- We can also calculate the probability that one random variable (y_i) takes a certain value *given the other(s) have a particular value* ($x_i = x_s$).
  - This is a conditional probability distribution
  - (Note we're fixing the value of all K of the x's in x_i)
- For a discrete distribution,

$$P(y_i = y_r | x_i = x_s) = \frac{P(x_i = x_s, y_i = y_r)}{P(x_i = x_s)}$$

Conditional Probability Distribution: Example I

- We talked earlier about the probability distribution of a student's final grade in Econ 21, y_i.
- Consider three elements of x_i:
  - $x_{1i} \equiv$ a constant term, equal to the number one ("1") for every i
  - $x_{2i} \equiv$ student i's hours spent working on the course
  - $x_{3i} \equiv$ student i's gender

Conditional Probability Distribution: Example II

- While the unconditional probability, $P(y_i = 5.5)$ might equal 12.4%,
  - Conditional probabilities can be very different
- For example,
  - $P(y_i = 5.5 | x_{2i} \in [0, 20), x_{3i} = male)$ might equal 1.7%
  - $P(y_i = 5.5 | x_{2i} \in [100, 120), x_{3i} = female)$ might equal 38.2%
    - ★ In other words, spending time on the course is likely to help you get a better grade!
    - ★ (Regardless of your gender)

Conditional Probability Distributions II

- This example demonstrates that while (x_i, y_i) is a multivariate distribution...
  - ...$(y_i | x_i = x_s)$ is a *univariate* distribution
- It's actually a *set* of univariate distributions, one distribution for each possible outcome of x_i $(x_i = x_s)$
  - i.e. one distribution for each possible *combination* of outcomes for x_{1i}, x_{2i}, and x_{3i}

Means and variances of multivariate RVs

- As for any univariate random variable, we can calculate the mean and variance of
  - Any one of a set of multivariate random variables, (x_i, y_i),
    - $E(x_i)$ and/or $E(y_i)$
    - (Using the appropriate marginal distribution)
  - Any one of a set of conditional random variables, $(y_i|x_i = x_s)$
    - We denote this $E(y_i|x_i = x_s)$ or $E(y|x)$ (where the particular value for x is implicit - often the specific xs in our empirical anlaysis)
    - This is an expectation over the random variable y_i for given values of the random variable x_i

Covariance and Correlation

- Covariance measures the degree of (linear) association between two random variables, denoted $Cov(x_i, y_i)$ or σ_{xy}

$$\boxed{Cov(x_i, y_i) = \sigma_{xy} = E(x_i y_i) - ExEy}$$

  - Units: if x_i is measured in hours and y_i is measured in grade points, then σ_{xy} is measured in hours-grade points.

- Correlation, denoted ρ_{xy} or just ρ, provides a *unitless* measure of the relationship between x_i and y_i:

$$\boxed{\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}}$$

  - Even nicer: $\rho \in [-1, 1]$

Properties of multivariate random variables

- Properties of multivariate random variables we often use in econometrics:
  - Two random variables are *independent* if and only if their joint probability distribution can be written as the product of respective marginal distributions:
    - i.e., if $P(x_i = x_s, y_i = y_r) = P(x_i = x_s)P(y_i = y_r)$
  - Linear rules for combinations of random variables:

$$\boxed{\begin{gathered} E(x_i + y_i) = E(x_i) + E(y_i) \\ V(x_i + y_i) = V(x_i) + V(y_i) + 2Cov(x_i, y_i) \\ V(x_i - y_i) = V(x_i) + V(y_i) - 2Cov(x_i, y_i) \end{gathered}}$$

Statistical Foundations

Latex Color = "LightGrey"

Statistical Foundations Introduction

- OK, that's enough for Probabilistic Foundations.
  - How do we *use* these ideas in econometrics?
- Statistical Foundations of Regression
  - Regression as Conditional Mean Estimation
  - Populations versus Samples
  - The Population and Sample Regression Functions

Regression as Conditional Mean Estimation I

- Return to our example with y_i student i's grade in Econ 21 and x_i a set of possible explanatory variables
  - Including a constant term (x_{1i}), how many hours they spent on the course (x_{2i}), their gender (x_{3i}), and any other factors we might like to include
- $E(y_i|x_i)$ is then student i's expected grade given their gender and the number of hours they spent on the course
  - Quick Quiz: *True, False, or Uncertain:* $E(y_i|x_i)$ is a random variable.*
    - ⋆ Don't answer yet!

Regression as Conditional Mean Estimation II

- The answer to this is important enough to merit a bunch of slides.
  - (7 of them!)
- Some basic principles:
  - x_i and y_i are (individually) random variables
  - When we take the expectation of $y_i|x_i$, we "integrate out" the randomness associated with y_i
  - For example, y_i is random, but $E(y_i) = \mu_y$ is not.
    - ⋆ It's just a number
    - ⋆ We may not know what it is, but it's a number; it's not random

Regression as Conditional Mean Estimation III

- So when we take the expectation of y_i given x_i ($E(y_i|x_i)$), the randomness associated with y_i "goes away"
- But we're still left with x_i
  - And x_i is itself a random variable
  - So technically the answer to the question is "True"
  - *But...*

Regression as Conditional Mean Estimation IV

- *But?*
  - When we take expectations in this course, we will often take *conditional expectations*
    - [See the slide title :-)]
  - In particular, we will take expectations of things *conditional on* x_i
    - i.e. conditional on a particular value of x_i
  - And if we condition on a particular value of x_i, then $E(y_i|x_i)$ is *not* random
    - It's just the expectation of y_i at that particular value of x_i
    - (e.g. $E(y_i|x_{2i} \in [1000, 120), x_{3i} = female)$)

Regression as Conditional Mean Estimation V

- **Bottom Line:**
  - We will typically be taking the second view of things
    - i.e. expectations of $<$stuff$>$ conditional on a *particular value* of x_i
  - So you can largely think of $E(\cdot|x_i)$ as *non-random*
    - (When conditioning on a particular value of x_i!)

Regression as Conditional Mean Estimation VI

- Given the previous discussion, we can write $E(y_i|x_i) = f(x_i)$
  - Where $f(x_i)$ is some (possibly very flexible) function of x_i
- Regression simply decomposes any random variable y_i into its conditional mean given x_i and an error term, ϵ_i:

$$\boxed{y_i = E(y_i|x_i) + \epsilon_i}$$

  - This is *completely without loss of generality*
  - We are only saying that we can take any random variable, y_i, and decompose it into:
    1. The part that we can explain as a function of x_i, $E(y_i|x_i)$, and
    2. The part that we can't, ϵ_i
  - We call this the *Population Regression Function*

Regression as Conditional Mean Estimation VII

- So, what's in the error term, ϵ_i???
- There could be many things:
  - Intrinsic randomness in behavior
  - Measurement error in y_i and/or x_i
  - Omitted variables (e.g. someone's "diligence" in EC021)
- We will spend lots of time in this course thinking about what's in the error term...
  - ...and whether it can cause us problems interpreting our regression estimates as causal...
    - ★ ...but *not yet!*

Populations versus Samples

Latex Color = "LightGrey"

Populations versus Samples I

- You should interpret everything we've said up to this point to reflect our understanding of the *population* values of y_i...
  - i.e. the population of students taking (and getting a grade in) Econ 21
- ...and its relationship to the *population* values of x_i
  - i.e. the hours spent on the course and gender mix of this student population

Populations versus Samples II

- In Statistics, a population $\equiv$ a set of similar items or events which is of interest for some research question or experiment
  - One of the two main goals of Statistics is to estimate (i.e. learn something about) the features of this population
    - e.g., By how much will a student's grade increase if they spend 2 more hours/week on the course?
    - (Note: this is a causal statement about $\partial E(y_i|x_i)/\partial x_{2i}$)
    - (We'll need to state some assumptions which must be true before we can make causal statements)

Populations versus Samples III

- *How* do we learn something about a population of interest?
  - Could try to survey the whole population, but that can be expensive, impractical, etc.
    - (Do students from 10 years ago remember their grade and how much time they spent on the course?)
- Instead, we often draw a *sample* of elements from the population and try to infer features of the population from features of the sample
  - Sample $\equiv$ a subset of the population selected for analysis

Simple Random Samples I

- In practice, there are lots of ways to sample from a population
  - Random sampling, stratified sampling, importance sampling, etc.
- In this course, we will only look at the easiest: *random sampling*
- A simple random sample is a sample where
  1. "Each element of the population has an equal chance of being chosen"
  2. "The selection of one element has no effect on the probability of another element being selected"

Simple Random Samples II

- Simple random samples usually yield a set of random variables that are what we call *"independent and identically distributed"*
  - (Independence is just property (2) above)
  - (Identical distributions imply property (1) above)
- We denote this "i.i.d" or "iid"

Population objects often estimated in statistics I

- Population objects often estimated in statistics include:
  - The (population) mean, $E(x_i)$ or μ_x
    - Estimated by the sample mean: $\boxed{\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i}$
  - The variance, $V(x_i) = E[(x_i - \mu_x)^2]$ or σ_x^2
    - Estimated by the sample variance: $\boxed{s_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$
- Note:*

Population objects often estimated in statistics I

- Population objects often estimated in statistics, cont:
  - The covariance, $Cov(x_i, y_i) = E[(x_i - \mu_x)(y_i - \mu_y)]$ or σ_{xy}
    - Estimated by the sample covariance: $\boxed{s_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}$
  - The correlation coefficient, $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
    - Estimated by the sample correlation coefficient: $\boxed{r_{xy} = \frac{s_{xy}}{s_x s_y}}$

Populations versus Samples V

- Features of the population are *fixed*, i.e. *non-random* (but unknown)
  - e.g. $E(y_i)$ or $\partial E(y_i|x_i)/\partial x_{1i}$
- Features of the sample are *random* (but known)
  - e.g. the average of the y_i in a particular sample
    - (If you drew a new sample...
    - ...you'd likely get a different sample average)
  - Note:*

Distribution of the Sample Mean I

- Let me show you how sample averages are random while simultaneously demonstrating one of the fundamental principles of statistics
- We'll do so by:
  1. Taking the mean and variance of the sample mean
     - ⋆ (Only interesting if the sample mean is random)
  2. Introducing a property of averaging (as the sample mean does) that econometrics *repeatedly* relies on as we do so

Distribution of the Sample Mean II

- To help understand what we're doing, let's sample a few sample means...
- Bear with me... Let's take everyone in the first row...
  - In groups of twos, can you calculate your average birth day
    - i.e. if your birthday is January 15, then your birth day is 15; if your birthday is June 27, then your birth day is 27
  - Let's plot the distribution of average birth days when $N = 2$
    - (Each one is a sample mean)
    - (And let's fill in a few more datapoints)

Distribution of the Sample Mean II

- Let's now do the same thing when $N = 4$
  - (By averaging across groups of 4)
  - (Again let's fill in a few more datapoints)
- Do you notice anything different in the distribution of sample means when $N = 2$ versus when $N = 4$?
  - *
- We could go on (e.g. do the same thing for $N = 8$, etc.)...
  - ...but let's leave it there for now and derive a result that underlies the pattern we just saw...

Distribution of the Sample Mean II

- Suppose x_i, $i = 1, \dots, N$, is $i.i.d.$ with mean μ_x and variance σ_x^2
  - i.e., $x_i \sim i.i.d(\mu_x, \sigma_x^2)$
- $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. What is its mean?*

Distribution of the Sample Mean III

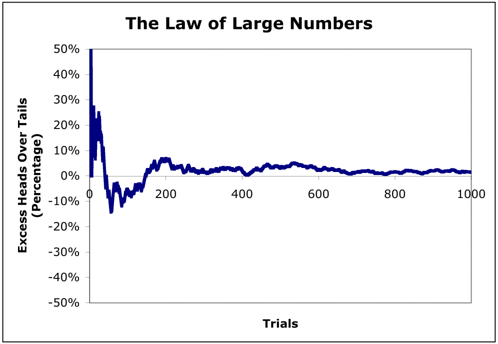
- Suppose x_i, $i = 1, \ldots, N$, is $i.i.d.$ with mean μ_x and variance σ_x^2
- $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. What is its variance?*

Distribution of the Sample Mean IV

- So $\boxed{E[\bar{x}] = \mu_x}$ and $\boxed{V[\bar{x}] = \frac{\sigma_x^2}{N} \ldots}$ Some intuition please?
- These results say two things:
  1. That the expected value of the sample mean equals the population mean
     - ★ Regardless of how big is each sample
  2. That the variance of the sample mean *decreases* as N increases
     - ★ This is the essence of "more data is always better"
     - ★ The more data you have (from the same population), the more accurate is your estimate of features of that population
     - ★ (In the sense the the variance of your estimator gets smaller)

Distribution of the Sample Mean IV

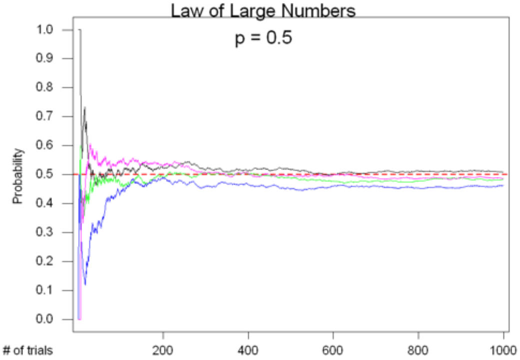
- Let's show the same things in a figure
- Suppose you calculated the probability of a "Head" after n "trials"



  - (Where the y-axis in the figure should go from 0% to 100%)

Distribution of the Sample Mean V

- A Question: Is the previous graph *representative*?
- Imagine starting over and doing the same again...



  - Getting 1st the black line; then the blue line; then the pink line...

Distribution of the Sample Mean VI

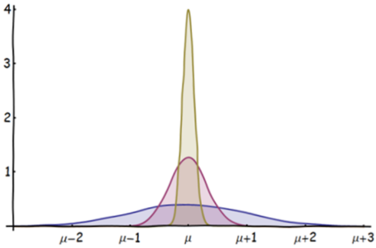
- The point of these figures:
  - It's always the case that the sample mean should be "close" to the population mean.
    - $E(\bar{x}) = \mu_x$
    - Q: But *how* close?
  - A: While it's quite possible to get an "outlier" sample mean of Heads for a small number of trials (flips) of a fair coin...
    - e.g. After 10, or 50, or even 100 flips
  - After a large number of trials, the sample mean is very likely to lie close to 0.50 (50%)
    - e.g. after 500, or even 200

Distribution of the Sample Mean VI

- The way we describe this property in the language of statistics is that...
  - The sample mean is very likely to be close to the population mean when the sample size is large
- Or, more accurately:
  - $\boxed{\text{The variance of the distn of the sample mean decreases with } N}$
    - ★ i.e., $\boxed{V[\bar{x}] = \frac{\sigma_x^2}{N}}$

Distribution of the Sample Mean VII

- In a figure, we can show $f(\bar{x})$ for different values of N as:

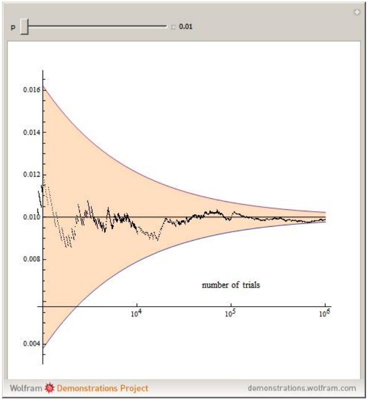


(Where $\mu = 0.5$ for us, and increments are 0.1, not 1)

  - For example, $N_{blue-ish} \sim 10$, $N_{purple-ish} \sim 100$, and $N_{brown-ish} \sim 1,000$
  - What happens in the limit? i.e. when $N \to \infty$?
    - ⋆ *

Distribution of the Sample Mean VIII

- My favorite figure combines *both* previous figures:



  - The curved lines show 95% confidence intervals for $\bar{x}$ as N gets large
    - ★ (As expected: they get tighter as N grows)

Distribution of the Sample Mean X

- The fact that the average of the sample mean converges on the population mean as the sample size, N, gets large...
  - i.e., as $N \to \infty$
  - ...underlies all of statistics (and thus all of econometrics)
- It is the result underlying what is called a "Law of Large Numbers", or LLN
  - I'll introduce LLNs later in the semester
    - (Probably) (OK, almost surely)
  - As well as its sibling, what is called a "Central Limit Theorem", or CLT
- But for now we march on...

Populations versus Samples VI

Comments about populations versus samples, cont:

- While one often only has one sample in a given analysis...
  - ...it is a conceptually useful task to imagine how your analysis would have changed had you had a different sample
  - (As we actually did have for birth days)
- Indeed, there is something called "bootstrapping" that relies on exactly this idea:
  - Construct estimates of a population value by creating multiple (what-are-called) *replication samples*
  - And estimate the distribution of (e.g.) a test statistic by looking at its actual distribution across all of these replication samples*

The Population and Sample Regression Functions

Our First Assumption

- The Classical Linear Regression Model (CLRM) is predicated on five key assumptions
  - It is now time for our first
- Assumption 1: Linearity
  - We assume that the population regression function...

$$y_i = E(y_i|x_i) + \epsilon_i$$

  - ... takes a linear form:

$$y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \epsilon_i$$

The Population Regression Function I

- In short, Assumption 1 states that we are assuming that the population regression function is linear:

$$\boxed{y_i = x_i'\beta + \epsilon_i}$$

- Let's parse that sentence a bit:

1. "...we are assuming...":
   - ★ We must be clear that we are making assumptions.
   - ★ Much of the time in this course will be spent trying to think through what could go wrong if our assumptions are wrong.
   - ★ (But first we show what goes *right* when our assumptions are *right!*)

The Population Regression Function II

- Parsing, cont.:
  2. "...the population regression function..."
     - ★ We said earlier that we could write $y_i = E(y_i|x_i) + \epsilon_i$ without loss of generality. This is true.
     - ★ There is a regression function in the population; another name for this is the (population) *data generating process*.
     - ★ It could be a very complicated, nonlinear, function, but...

The Population Regression Function III

- Parsing, cont.:
    3. "...is linear."
        - ★ But... we're assuming it's not. We're assuming it's linear.
        - ★ Thus we're assuming that all the various possible points for $E(y_i|x_i)$, which could lie all over the map, instead line up nicely in a line (hyperplane).
        - ★ (An alternative rationale is that we use a *linear approximation* to the true-but-possibly-nonlinear conditional mean, $E(y_i|x_i)$)

The Population Regression Function IV

See Figure in Class

The Point of Statistics

- Quick Quiz: *True, False, or Uncertain:* What are the two primary tasks of statistics?*
    - ______________________________
    - ______________________________
- We'll focus first on the first of these.
    - Imagine you have a population of interest...
    - Imagine you have a sample from that population...
    - Can we *estimate* the impact of studying on students' grades?

The Residual I

- We said the population regression function was $y_i = x_i'\beta + \epsilon_i$
- Suppose we had an estimate of β
  - Call this $\hat{\beta}$
- Using $\hat{\beta}$, we can also estimate y_i.
- We write our estimate of y_i as $\hat{y}_i$:

$$\hat{y}_i = x_i'\hat{\beta}$$

The Residual II

- Define the difference between the true y_i and our estimate of of it, $\hat{y}_i$, the *residual*, e_i:

$$
\begin{aligned}
e_i &= y_i - \hat{y}_i \\
&= y_i - x_i'\hat{\beta}
\end{aligned}
$$

  - e_i is sometimes written as $\hat{\epsilon}_i$
    - ★ (Intuitive as it is our estimate of ϵ_i for each observation in our dataset.)

Two ways to write y_i

- We now have two ways to decompose y_i into component parts:
  1. The *Population Regression Function*:

$$y_i = x_i'\beta + \epsilon_i$$

  2. And, now, what we'll call the $\boxed{\textit{Sample Regression Function}}$:

$$\boxed{y_i = x_i'\hat{\beta} + e_i}$$

Two ways to write y_i

- We now have two ways to decompose y_i into component parts:
  1. The *Population Regression Function*:

  $$y_i = x_i'\beta + \epsilon_i$$

  2. And, now, what we'll call the $\boxed{\textit{Sample Regression Function}}$:

  $$\boxed{y_i = x_i'\hat{\beta} + \hat{\epsilon}_i}$$

Population versus Sample Regression Functions I

- A useful bit of word association:
  - When I say "LHS", you should think "RHS":
    - ★ "Population" $\sim \begin{cases} \text{"Truth"} \\ \text{"The object of interest"} \\ \text{"Fixed but unknown"} \end{cases}$
    - ★ "Sample" $\sim \begin{cases} \text{"Our estimate of truth"} \\ \text{"Our estimate of the object of interest"} \\ \text{"Random but known"} \end{cases}$

Population versus Sample Regression Functions II

- The sample regression function is our best estimate of the (true) population regression function
  - If we had a different sample, we'd have a different sample regression function...
  - ...but the population regression function *wouldn't* change.
    - (See next slide)

The Population and Sample Regression Functions I

See Figure in Class

The Population and Sample Regression Functions II

- The goal of econometrics is...
  - ...to estimate the population regression function...
  - ...with the sample regression function
  - ...
  - AND
  - ...
  - USUALLY
  - ...
  - ...to learn what is the causal effect of particular x_{ik} on y_i

Alternative Approaches to Estimation

Latex Color = "LightGrey"

Alternative Approaches to Estimation I

- We've just said that we would estimate $\hat{\beta}$...
  - How???
- Statistics provides *many* different ways to estimate the parameters of a statistical process
  - (Our statistical process:
    - The (assumed-linear) population regression function $y_i = x_i'\beta + \epsilon_i$)

Alternative Approaches to Estimation II

- In Master's-level econometrics, you will see (at least) *three*:
  - Ordinary Least Squares (OLS)
    - The bread-and-butter of linear econometric methods
  - (Generalized) Method of Moments (GMM)
    - A more general framework that includes both OLS and Instrumental Variables (IV) estimation
  - Maximum Likelihood Estimation (MLE)
    - A less-general-but-more-powerful-if-its-assumptions-are-true framework commonly used in Limited Dependent Variable estimation
    - (We probably won't do any MLE this semester)
- We will only use OLS and IV in this course
  - But it's good for you to have a high-level overview of all the options

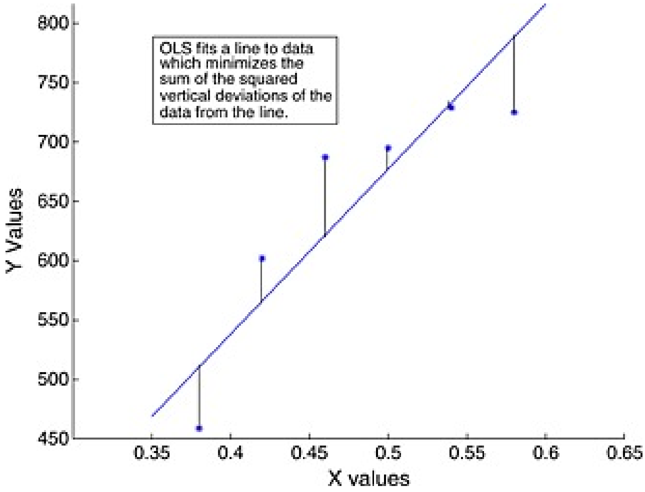
OLS Overview

- Ordinary least squares chooses $\hat{\beta}$ to minimize the sum of squared residuals, i.e.

$$\min_{\hat{\beta}} \sum_{i=1}^{N} (y_i - x_i' \hat{\beta})^2$$

- Note:
  - The sum of squared residuals is a scalar (i.e. just a number)
  - There are K elements in $\hat{\beta}$
  - Thus doing the standard calculus yields K equations in K unknowns
    - The first-order-condition for each element in $\hat{\beta}$ yields the K equations
    - The K elements in $\hat{\beta}$ are the K unknowns

OLS in a Picture



GMM Overview I

- The second method is called the (Generalized) Method of Moments
  - This approach has revolutionized econometrics by allowing researchers to flexibly estimate many econometric models that were previously intractable.
- The basic idea: Equate the moments of a statistical model with the actual moments in the sample.
  - But first: what's a moment?
  - Basically, the r^{th} moment of a random variable, x_i, is $E(x_i^r)$
    - ★ The r^{th} *central* moment of a random variable, x_i, is $E[(x_i - Ex)^r]$
  - Thus the mean is the 1st moment, the variance is the 2nd central moment, etc.
- In (cross-section) econometrics, we almost exclusively work with 1st and 2nd moments.

GMM Overview II

- I would normally now show you a figure to give you some intuition
- This is the best I could find:

See Figure in Class

GMM Overview III

- GMM Comments:
  - The MM (no "G") was developed in 1894 and specifically compared (only) the moments of a (univariate) statistical model with the same (univariate) moments in the data
    - e.g., estimating μ_x with $\bar{x}$
  - The (big-G) Generalization was to use *moment-like objects* relating multiple statistical distribution (e.g. covariance)
    - As we'll show later, OLS and IV can be derived under the assumption that the *covariance* between each of the x_i and the error term, ϵ_i, is zero

MLE Overview I

- The third method is called Maximum Likelihood Estimation (MLE)
- The basic idea:
  - MLE relies on the intuitive notion that the data we observe are more likely to be drawn from some distributions than others
- I would normally now show you a figure to give you some intuition...

See Figure in Class

MLE Overview II

- MLE works by
  1. Assuming the data are generated by a particular family of distributions (e.g. Normal) with unknown parameters β
  2. Choosing an estimate of β, $\hat{\beta}$, to maximize the likelihood of seeing the actual sample that we see
- At the Master's level, we use MLE most often with dummy-dependent variable models
  - e.g., Logit and Probit
- You will learn more as you take more econometrics...

How to Choose an Estimator

Latex Color = "LightGrey"

How to choose between estimators I

- OK so that's a bunch of estimators
- How should we choose between them???*
  - ____________________
  - ____________________

How to choose between estimators II

- In practice, there are lots of different theoretical properties of estimators to consider!
- We will look at only a few; they come in pairs
- The first element in the first pair is *Unbiasedness*
  - An estimator $\hat{\beta}$, is said to be unbiased if $E(\hat{\beta}) = \beta$

How to choose between estimators III

- The second element in the first pair is *Efficiency*
  - (Also known as *Minimum Variance*)
- Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be two estimators of β
  - If $V(\hat{\beta}_1) < V(\hat{\beta}_2)$, then $\hat{\beta}_1$ is said to be *more efficient* than $\hat{\beta}_2$.
    - ⋆ Technically, this definition of efficiency is for two unbiased estimators
    - ⋆ (It will be sufficient for most of the estimators we look at in this class)

How to choose between estimators IV

- Bias and Efficiency are included in what are called finite-sample (or small-sample) properties of estimators.
  - Evaluating these properties requires that you know the *actual distribution* of the data generating process
    - ★ e.g. that ϵ_i is distributed as a *Normal* distribution
  - This can be hard to know exactly.
  - (For us, will be true only for OLS, not IV or Panel data estimators)
    - ★ (In the CLRM, we assume $\epsilon_i \sim N(0, \sigma^2)$)

How to choose between estimators V

- People often also look at what are called asymptotic (or large-sample) properties of estimators.
- When one has large samples...
  - (or is-willing-to-*assume*-large-sample-results-apply-even-if-they-have-small-samples...)
  - ...it turns out much more general results are available for a wide variety of estimators.

How to choose between estimators VI

- This yields our *second* pair of theoretical properties of estimators
  - *Consistency* and *Asymptotic Efficiency*
    - (I'll introduce these later in the semester)
  - (Also *Asymptotic Normality*, another large-sample property)

The Last Slide

- The goal of these slides has been to
  - Highlight those parts of probability and statistics that we rely on most in econometrics
- We are now (finally) in a position to dig into econometrics proper
  - We'll start with the Classical Linear Regression Model (CLRM)

Coming up next!

Table of Contents

Back