EVALUATING THE IMPACT OF SOCIAL PROGRAMS

CLASS #2

EDUC 1160

KOLAJO P. AFOLABI

SEPTEMBER 19, 2011

Today's Class

- Review: Statistical Inference
- Review: OLS Regression
- Experiments in a Potential Outcomes Framework
- Effect Sizes
- Tennessee's STAR class size experiment

Statistical Inference: A Review

- Given what we observe in our sample, how likely is it that the true value of this estimate in our population of interest is zero?

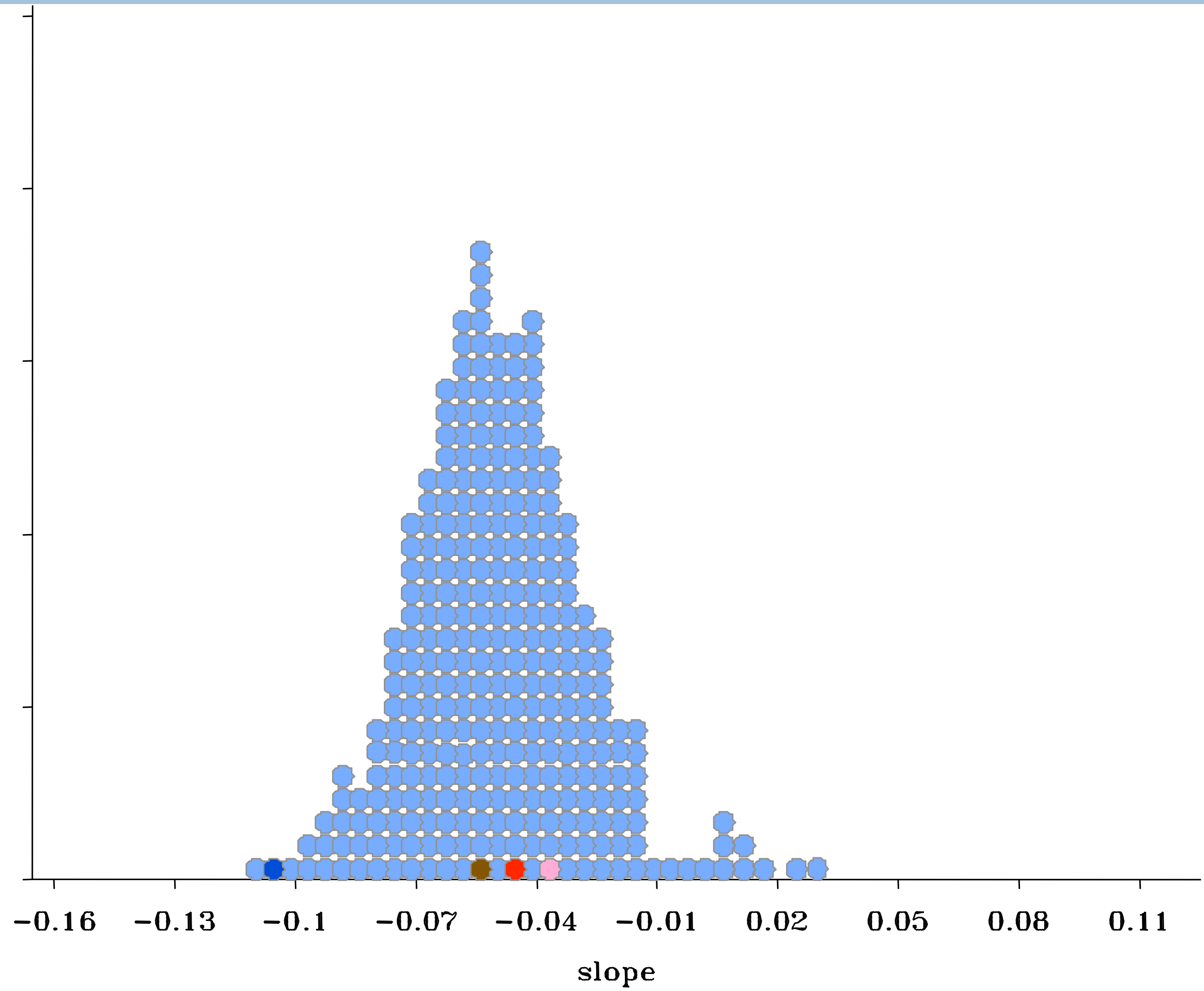
β_1 = True relationship between X and Y in the population

$\hat{\beta}_1$ = *Relationship between X and Y in sample*

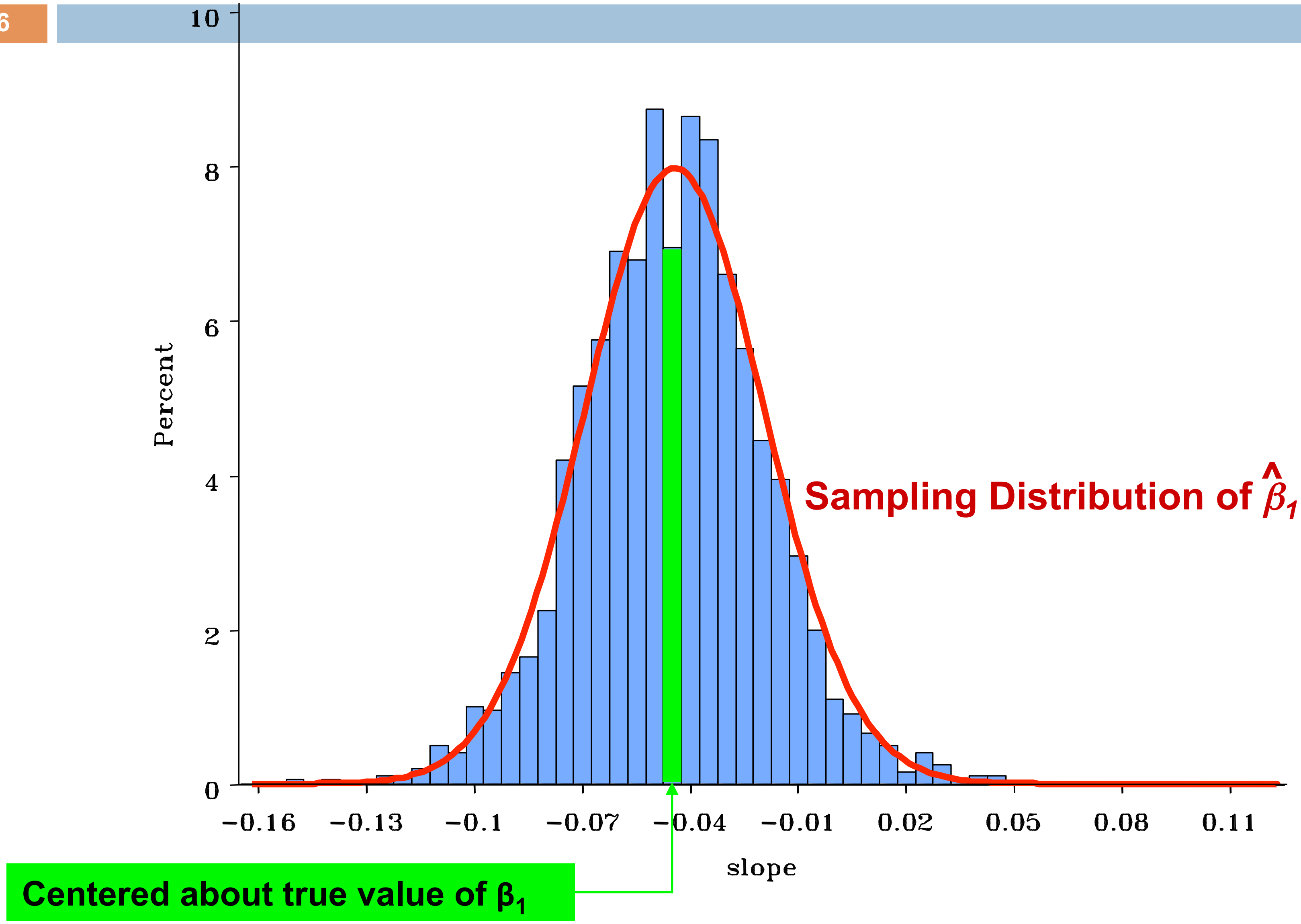
Desirable properties of an estimator

- Unbiased: the expected value of the estimator is equal to the parameter (where expected value means the "long-run average" – or the average from many repeated samples of the same size). Centered around the actual value of the population parameter.
- Precise: the values of the estimator from many repeated samples have the smallest variance of any estimator. Smallest standard error= best precision.
- Unbiasedness and precision are both claims about the sampling distribution of the estimator.

300 Repeated Random Samples

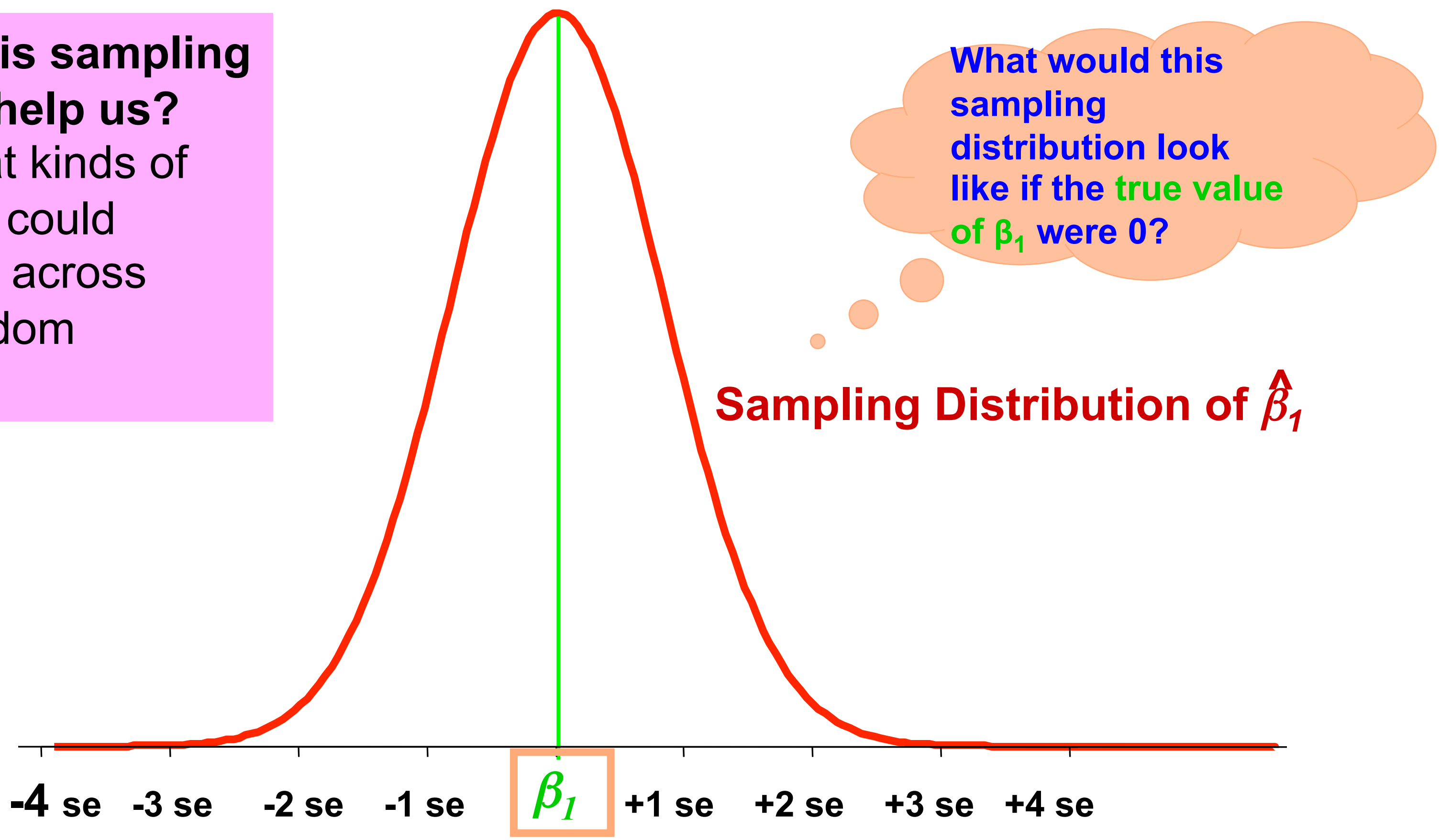


2000 Repeated Random Samples



Sampling Distribution

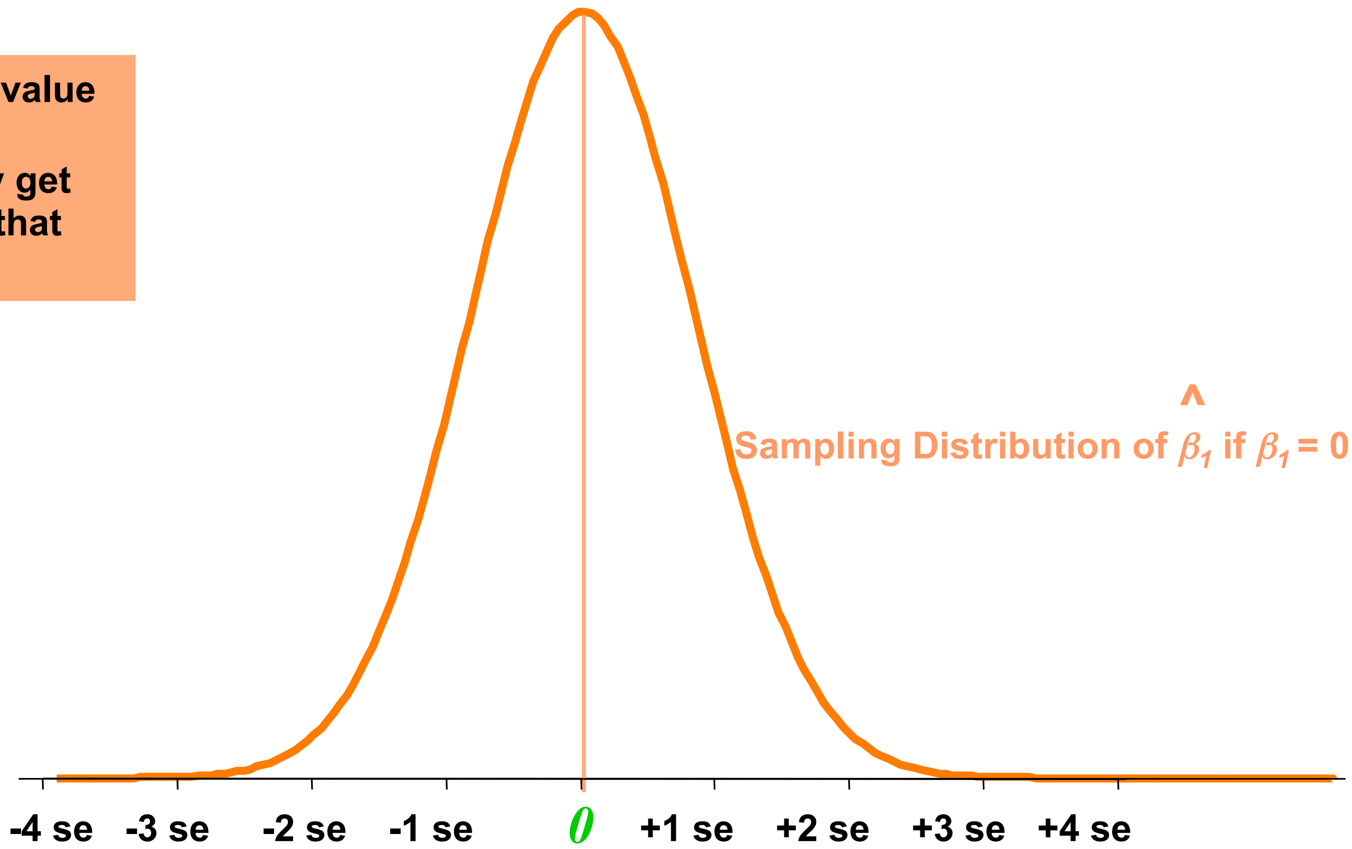
How does this sampling distribution help us?
It tells us what kinds of estimates we could expect to find across repeated random sampling



The standard deviation of a sampling distribution is known as a "**standard error**," commonly abbreviated as "**se**"

Sampling Distribution if $\beta_1 = 0$

Even if the true value of β_1 is 0, we will certainly get estimates of β_1 that are non-zero!!!



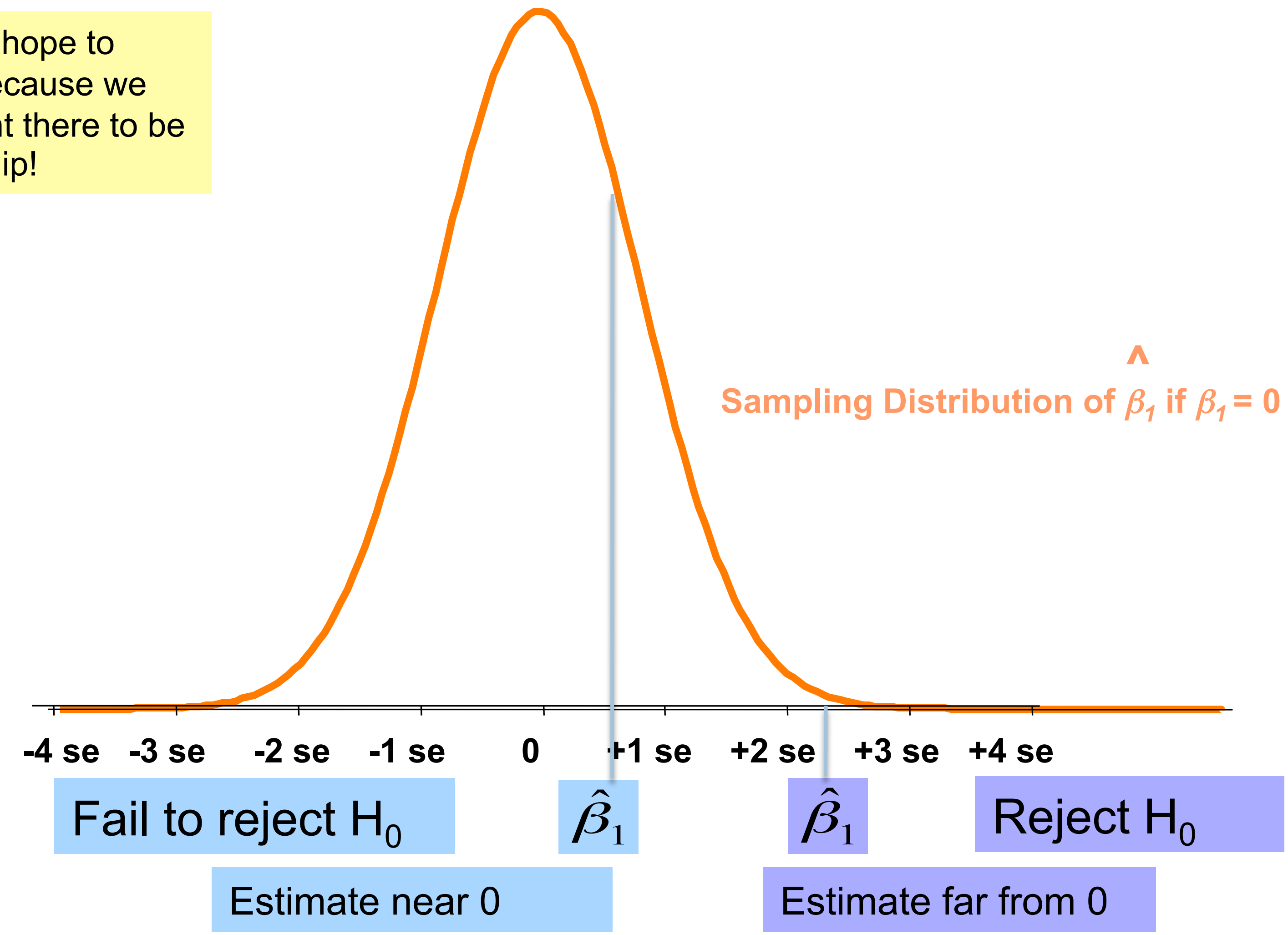
So…just because we found a non-zero estimated slope doesn't mean that it didn't come from a population in which the true slope **is** 0

Hypothesis Testing

- Null Hypothesis, H_0 ($\beta_1 = 0$)
  - There really is no relationship between X and Y in the population
- Alternative Hypothesis ($\beta_1 \neq 0$)
  - There really is a relationship between X and Y in the population
- Assuming the null hypothesis was true, how likely is it that we would have gotten the sample result we did?

Reject or Fail to Reject H_0?

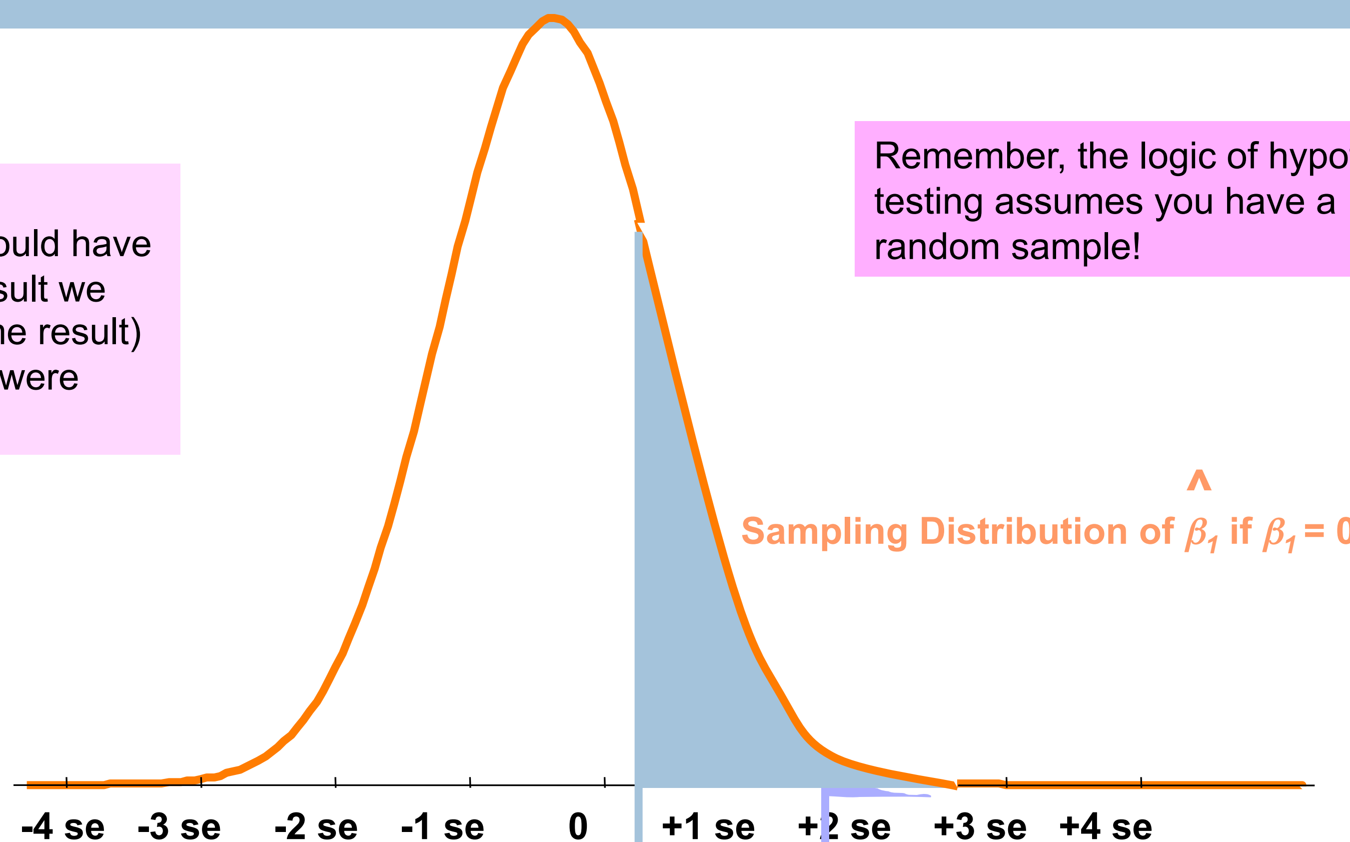
We usually hope to reject H_0 because we usually want there to be a relationship!



Hypothesis Testing and p-values

p-value:
The probability we would have gotten the sample result we did (or a more extreme result) if the null hypothesis were really true.

Remember, the logic of hypothesis testing assumes you have a random sample!



p is large, so it's possible that we would have gotten this result if the null were true

$\hat{\beta}_1$

$\hat{\beta}_1$

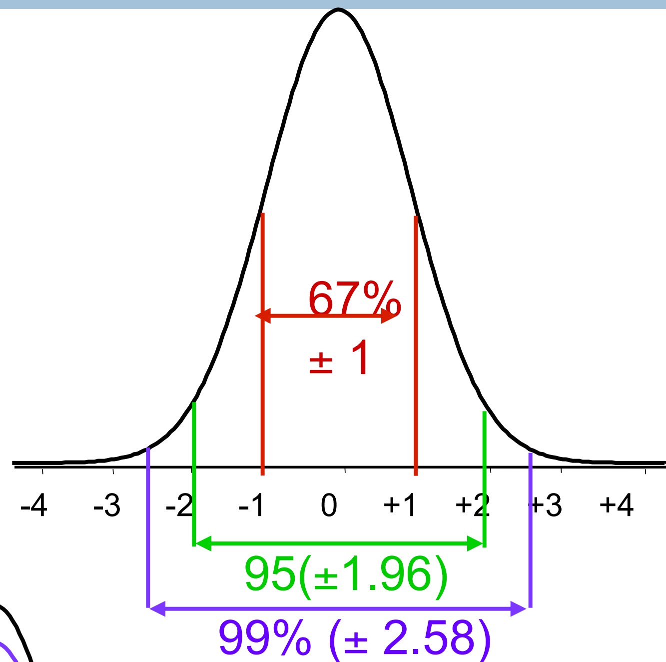
p is small, so it's unlikely that we would have gotten this result if the null were true

Fail to reject H_0

Reject H_0

The T-Distribution

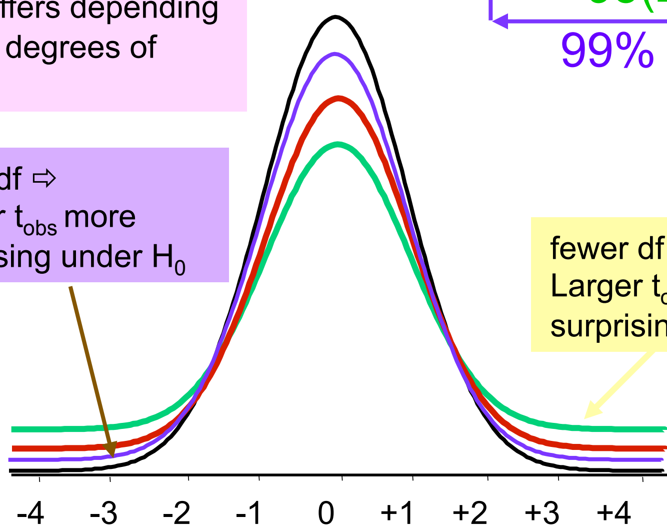
When df is huge, the t-dist^n is a standard normal dist^n



When df isn't huge, the t-dist^n differs depending upon its degrees of freedom

more df ⇨
Larger t_{obs} more surprising under H_0

fewer df ⇨
Larger t_{obs} less surprising under H_0



How large is large? Critical values of $t_{observed}$

df	Two-sided probability level, p		
	0.10	0.05	0.01
10	1.81	2.23	3.17
20	1.72	2.09	2.85
30	1.70	2.04	2.75
50	1.68	2.01	2.68
100	1.66	1.98	2.63
infinite	1.64	1.96	2.58

So…**if t_{obs} is "large" given the df,** it's very **unlikely** that we would have gotten a result this extreme (or more extreme) if H_0 is true, so we **reject H_0**

$$t_{obs} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{(n-2)df}$$

T-Statistics

$$t_{obs} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$$

If H_0: $\beta_1 = 0$

$$t_{obs} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

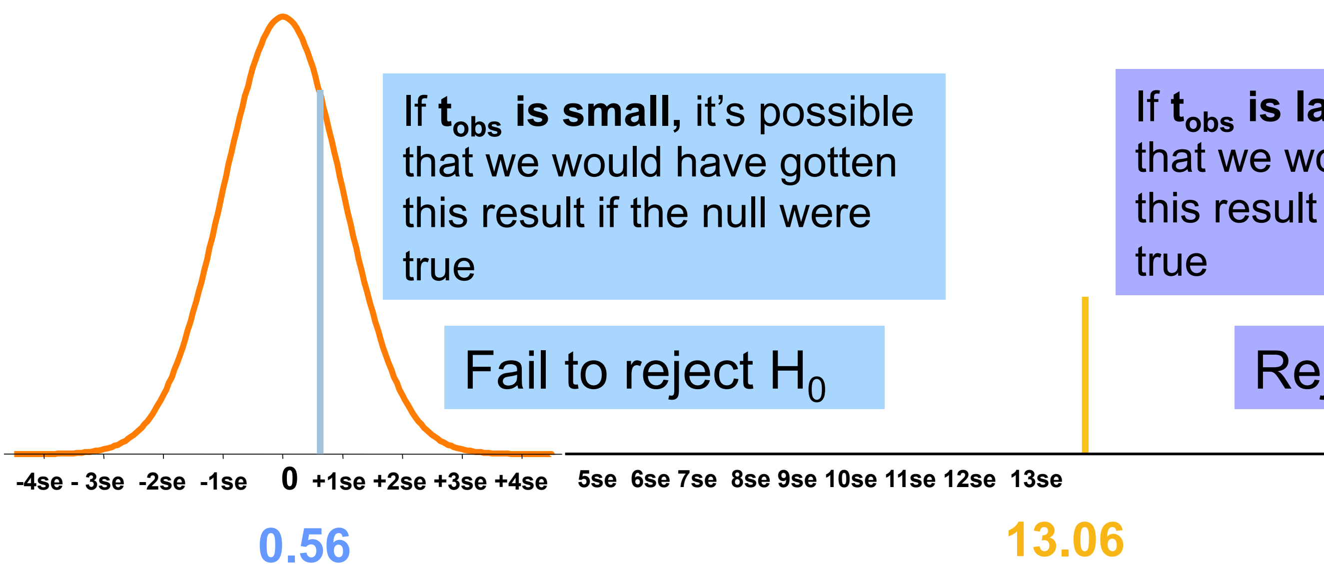
So t_{obs} tells us how many standard errors away from 0 our sample estimate is.

If t_{obs} **is small,** it's possible that we would have gotten this result if the null were true

If t_{obs} **is large**, it's unlikely that we would have gotten this result if the null were true

Fail to reject H_0

Reject H_0



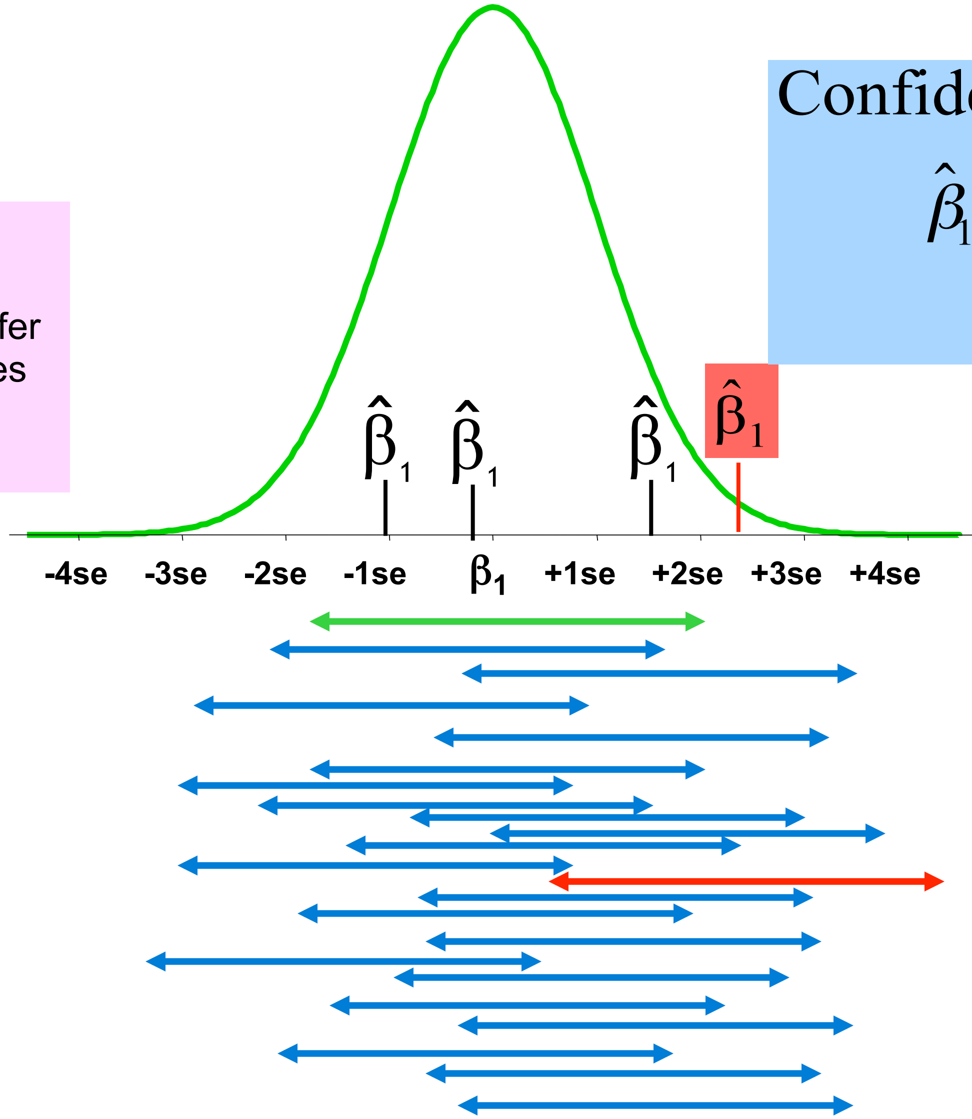
Confidence Intervals

Idea: Can we use this sampling distribution to construct intervals that offer a range of plausible values for the population parameter?

Confidence Interval for β_1

$$\hat{\beta}_1 \pm t_{n-2}[se(\hat{\beta}_1)]$$

$$\hat{\beta}_1 \pm 2se(\hat{\beta}_1)$$



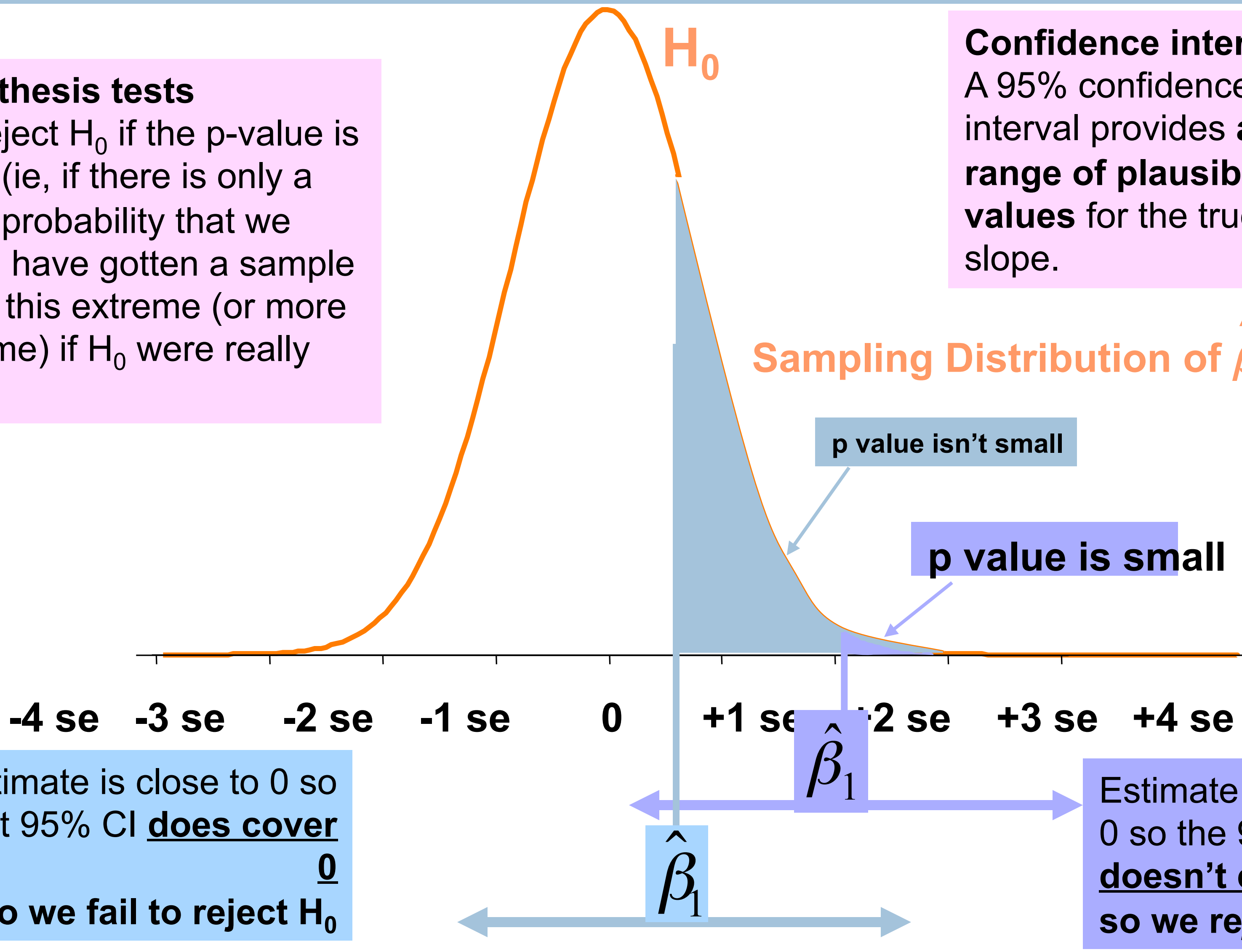
For every 20 intervals we construct, we estimate that an average of 1 won't cover the true value of β_1

Unfortunately, when we compute any 95% CI, we don't know whether its one of the lucky 95% that do cover the true value or the unfortunate 5% that don't

Hypothesis Testing & Confidence Intervals

Hypothesis tests
We reject H_0 if the p-value is small (ie, if there is only a small probability that we would have gotten a sample result this extreme (or more extreme) if H_0 were really true).

Confidence interval:
A 95% confidence interval provides **a range of plausible values** for the true slope.



Estimate is close to 0 so that 95% CI **<u>does cover 0</u>**
so we fail to reject H_0

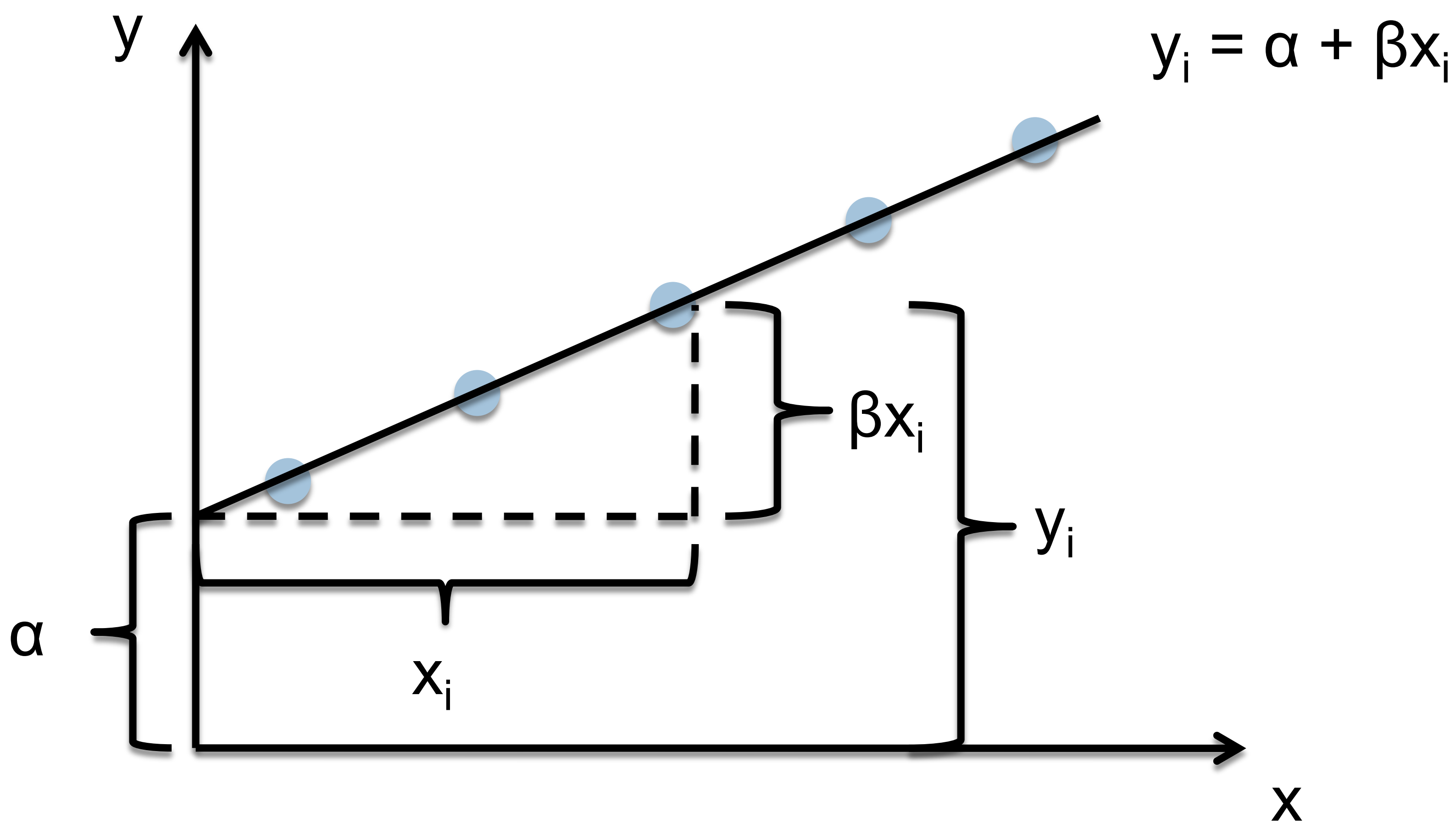
Estimate is far from 0 so the 95% CI **<u>doesn't cover 0</u>**
so we reject H_0

QUESTIONS?

Bivariate Linear Regression

- Assume y is some *outcome* of interest
- Assume x is observations of an *explanatory variable* of y
- We believe that x and y are *linearly* related so we model them as such
  - $y_i = \alpha + \beta x_i$

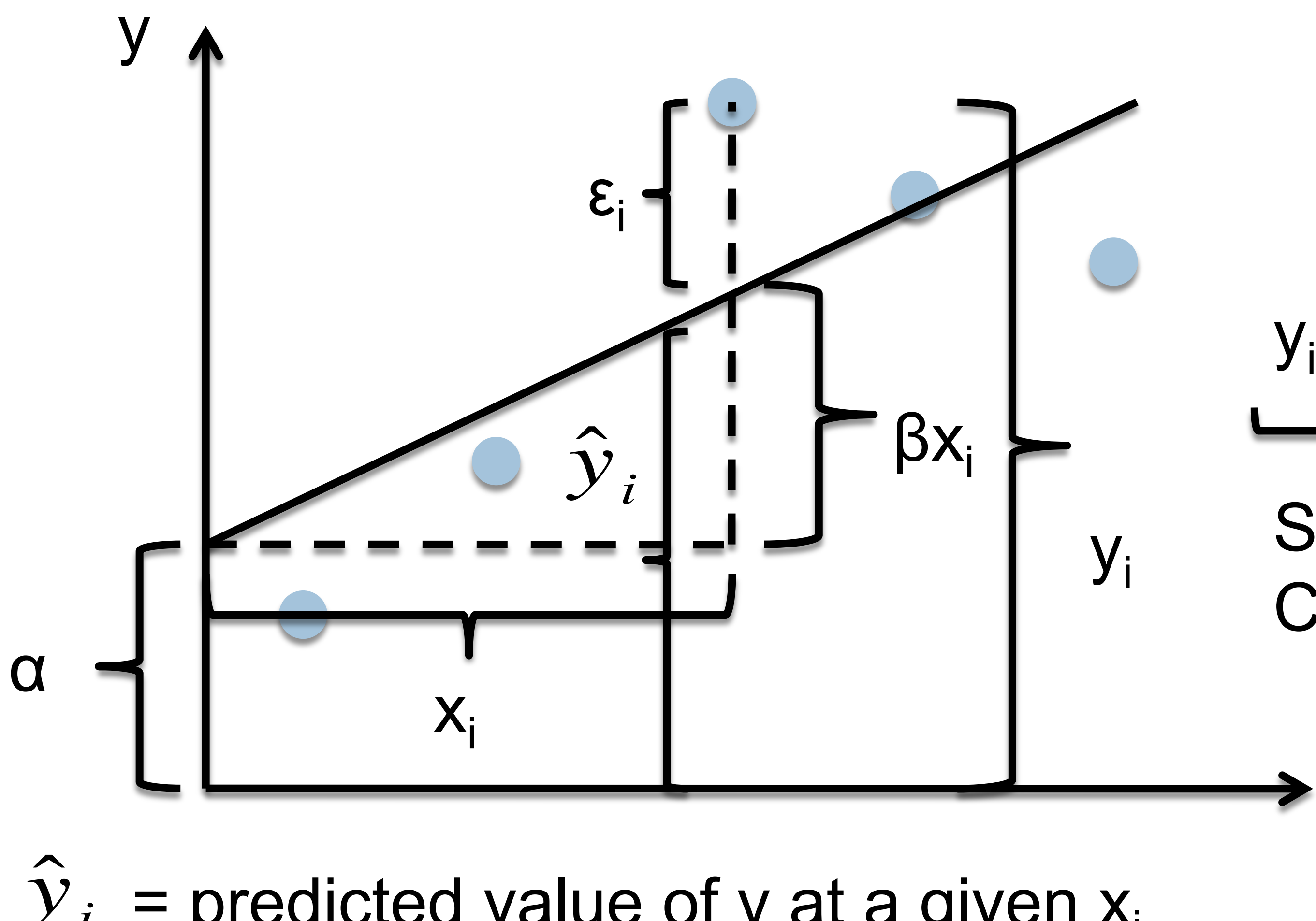
Bivariate Linear Regression



Bivariate Linear Regression

- Under certain conditions, the slope estimate, β, gives us the causal impact of x on y
- β,is the slope of the x/y line, so it is interpreted as a one unit change in x corresponds to a β unit change in y
- α represents the intercept, which is the average value (mean) of y when x=0

Bivariate Linear Regression



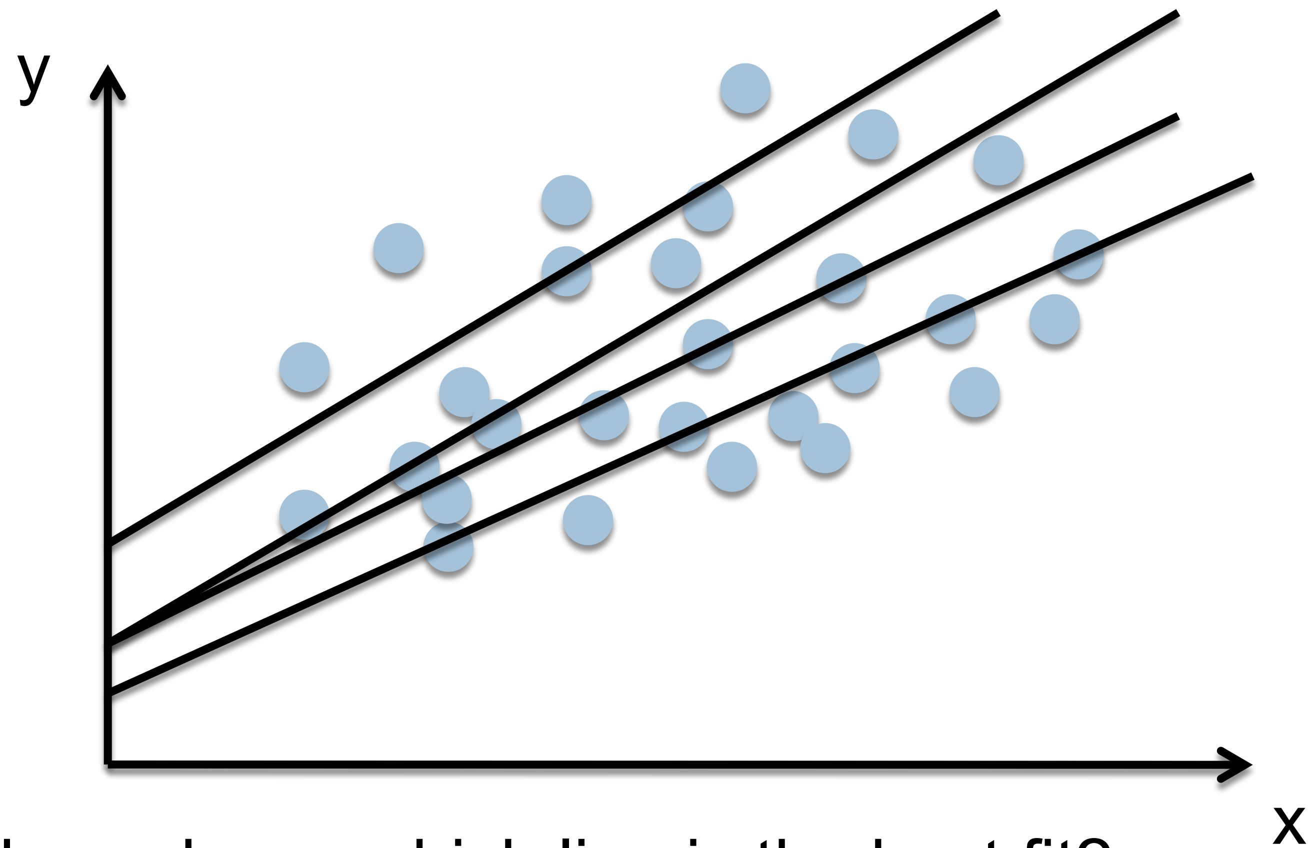
Stochastic Component

$$\underbrace{y_i = \alpha + \beta x_i}_{\text{Structural Component}} + \varepsilon_i$$

$\hat{y}_i$ = predicted value of y at a given x_i

ε_i = residual

Bivariate Linear Regression



- How do we know which line is the best fit?
- We minimize the sum of the squared errors between the actual values of y and the predicted values of y.

Ordinary Least Squares (OLS)

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\hat{y}_i = \alpha + \hat{\beta} x_i$$

$$\hat{\beta}^{OLS} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{Cov(X,Y)}{Var(X)}$$

- Estimates of α and β obtained using this criterion are called the Ordinary Least Squares (OLS) estimates
  - We are estimating the slope and intercept with the OLS estimator.
- OLS estimates of α and β are unbiased (if key assumptions are met) and relatively efficient
- In fact, the OLS estimator is the Best (that is, the most efficient) Linear Unbiased Estimator possible.

Variance & Standard Error

- Since we are getting an OLS estimate of the slope, there is a sampling distribution of the slope
- Using data from our sample, we can calculate the variance and standard error of our estimate

$$Var\left(\hat{\beta}^{OLS}\right) = \frac{\hat{\sigma}_{\varepsilon}^{2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{Var(e)}{nVar(X)}$$

$$S.E.\left(\hat{\beta}^{OLS}\right) = \sqrt{Var\left(\hat{\beta}^{OLS}\right)} = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^{2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{Var(e)}{nVar(X)}}$$

Standard Error

- We want our estimates to be both *unbiased* and *precise*
- The standard error of our estimate is a measure of its precision
- Also remember that the t-statistic is just the estimate divided by its standard error (assuming β=0)

$$t_{obs} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

Variance & Standard Error (2)

- You can also calculate the variance of your treatment estimator as

$$Var(\hat{\beta}_1) = \frac{\sigma^2_{Y-Treatment}}{N_{Treatnent}} - \frac{\sigma^2_{Y-Control}}{N_{Control}}$$

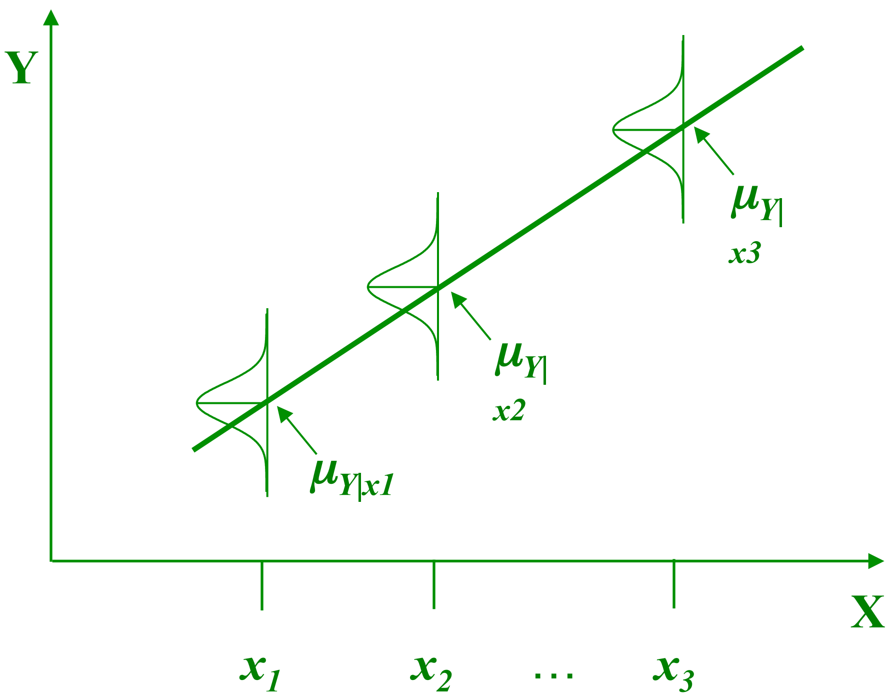
- Implications:
  - Variance increases with increase in σ^2
  - Variance decreases with N_T & N_C
  - Variance decreases with balance between N_T & N_C

OLS Assumptions

- **At each value of X, there is a distribution of Y.** These distributions have a mean $\mu_{Y|X}$ and a variance of $\sigma^2_{Y|X}$
- **X and e are uncorrelated**. Omitted variables that affect Y are uncorrelated with X.
- **Correct functional form.** The means of each of these distributions, the $\mu_{Y|X}$'s, may be joined by a straight line.
- **Homoscedasticity.** The variances of each of these distributions, the $\sigma^2_{Y|X}$'s, are identical.
- **Independence of observations.** Conditional on the values of X, the values of Y (the y_i's) are independent of each other.
- **Normality**. At each given value of X (at each x_i), the values of Y (the y_i's) are normally distributed.

OLS Assumptions

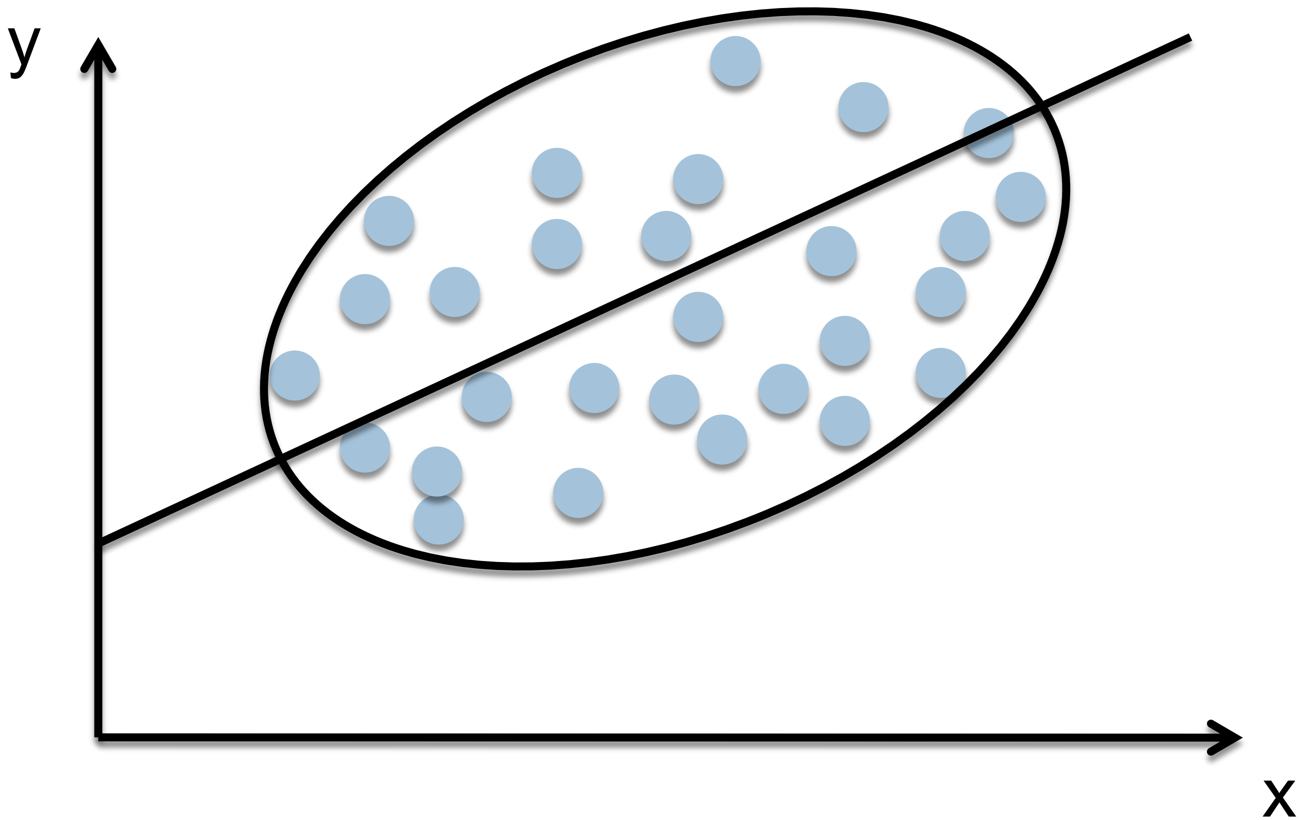
$$Y = \beta_0 + \beta_1 X + \varepsilon$$



- **At each value of X, there is a distribution of Y.** These distributions have a mean $\mu_{Y|X}$ and a variance of $\sigma^2_{Y|X}$
- **X and ε are uncorrelated**. Omitted variables that affect Y are uncorrelated with X.
- **Correct functional form.** The means of each of these distributions, the $\mu_{Y|X}$'s, may be joined by a straight line.
- **Homoscedasticity.** The variances of each of these distributions, the $\sigma^2_{Y|X}$'s, are identical.
- **Independence of observations.** Conditional on the values of X, the values of Y (the y_i's) are independent of each other.
- **Normality**. At each given value of X (at each x_i), the values of Y (the y_i's) are normally distributed.

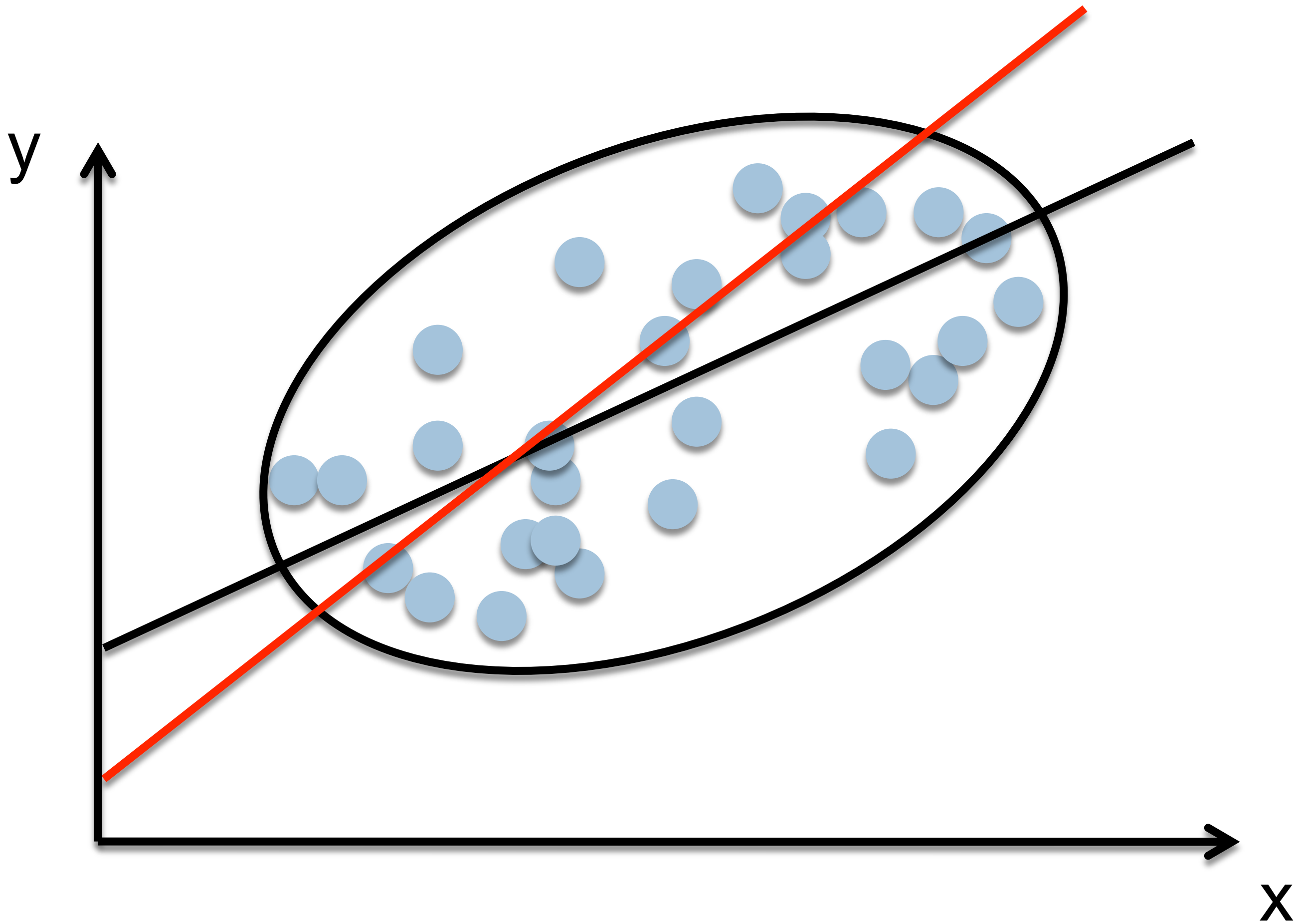
X and ε are uncorrelated

- Cov(X, e)=0



X and ε are uncorrelated

- Cov(X, e)>0



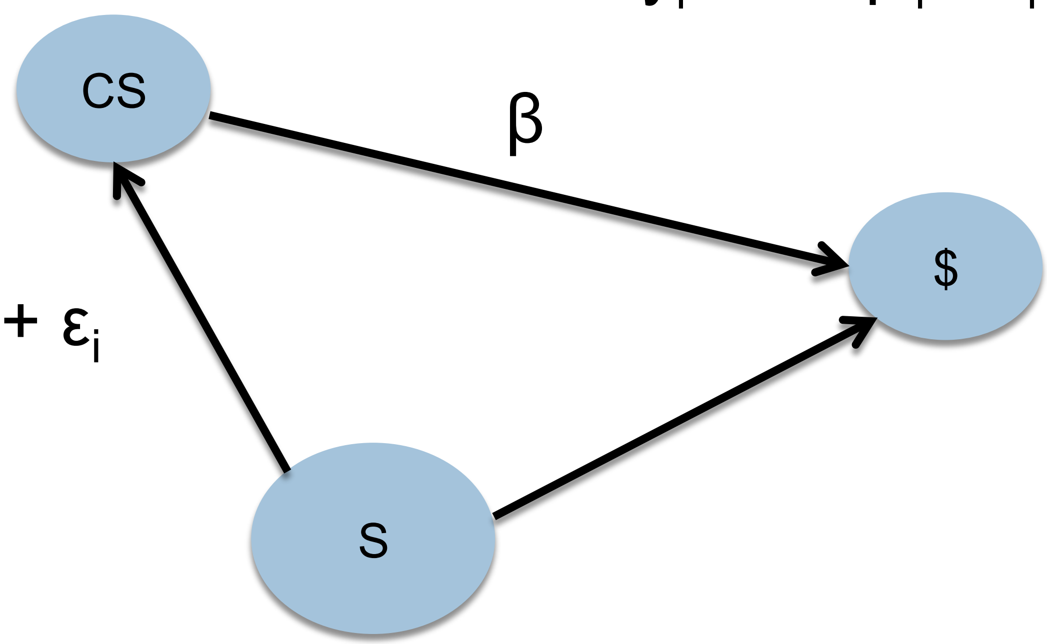
X and ε are uncorrelated

- Cov(X, e)>0
- If people with low values of X tend to have low ε's and high values of X tend to have high ε's, then OLS gives you a *biased* estimate of the slope
- It may be there is an *omitted variable* that affects Y and is correlated with X

Multivariate Regression

- Allows us to examine the effect of X on Y *controlling for other variables*
- Example: cognitive skills (CS), schooling (S), and adult earnings ($)

$$y_i = \alpha + \beta_1 CS_i + \beta_2 S_i + \varepsilon_i$$



$$y_i = \alpha + \beta_1 CS_i + \varepsilon_i$$

Regression & Experiments

- Why do we need regression in experiment?
- Experimental impact estimator

$$ATE = \bar{Y}_T - \bar{Y}_C$$

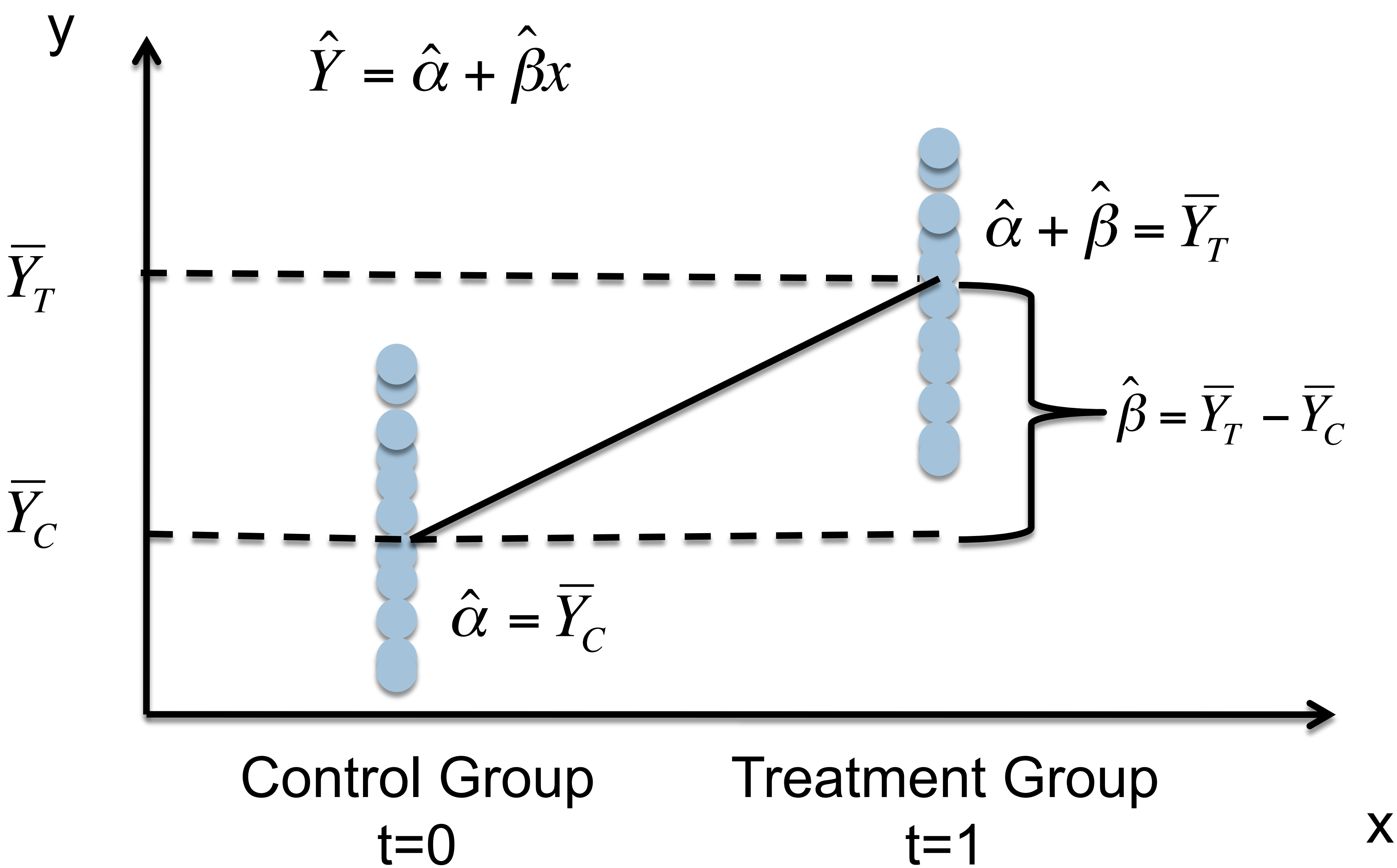
Or

$$\widehat{ATE} = \left(\frac{\sum_{i=1}^{n_1} Y_i}{n_i}\right) - \left(\frac{\sum_{i=1}^{n_0} Y_i}{n_0}\right)$$

- Variance of impact estimator

$$Var(\hat{\beta}_1) = \frac{\sigma^2_{Treatment}}{N_{Treatnent}} - \frac{\sigma^2_{Control}}{N_{Control}}$$

Regression & Experiments

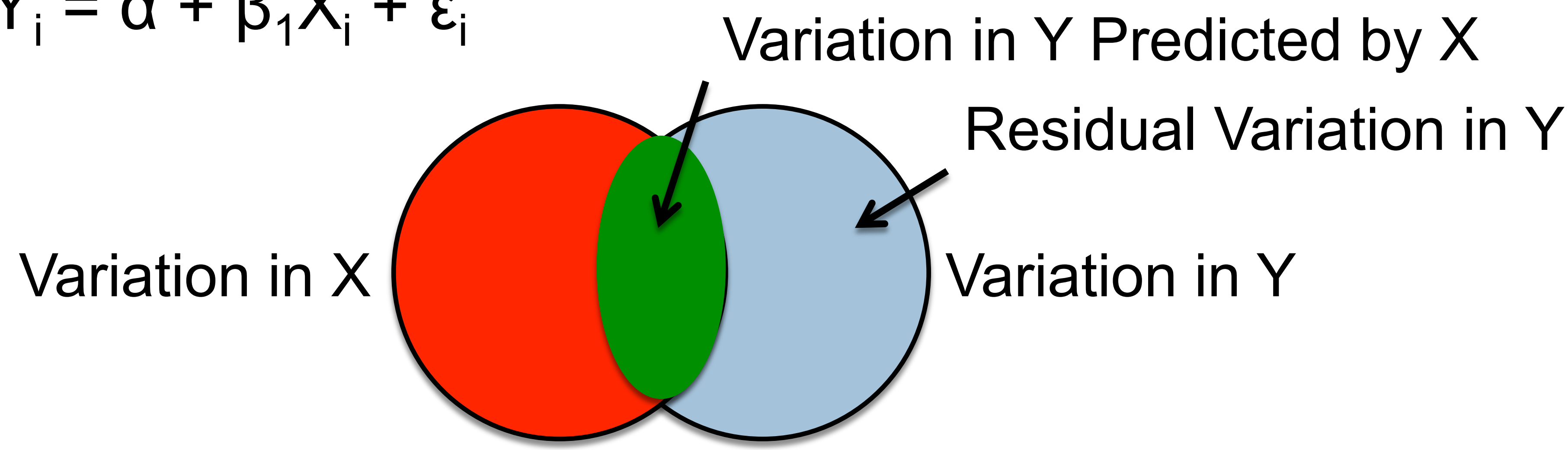


Regression & Experiments

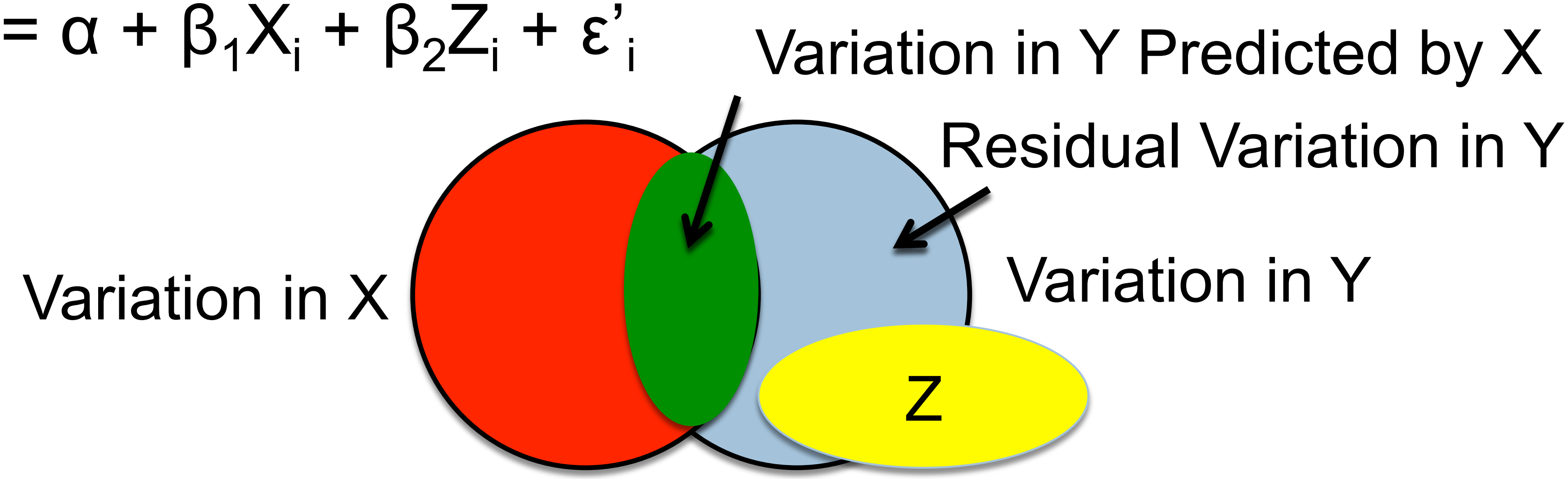
- Why do we “normally” (i.e., in observational studies) estimate program impacts with regression?
  - We want to control for observed differences between program participants and non-participants that are related to the outcome
- What should be the difference in observed (and non-observed) characteristics between program participants and non-participants in an experimental study?
  - On average, there should be none
- Then why use regression to “control” for differences that presumably aren‟t present?
  - 1.To “absorb” *residual variation and hence reduce the standard errors associated with the impact estimate*
  - 2.Also to control for chance differences in observed characteristics of the treatment and control groups

Regression & Experiments

- $Y_i = \alpha + \beta_1 X_i + \varepsilon_i$



- $Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + \varepsilon'_i$



Including Covariates

- The purpose of incorporating relevant covariates into an analysis of experimental data is to reduce residual variation, decrease standard errors, and increase statistical power.
- Include covariates which:
  - Do not vary over time
  - Are measured prior to random assignment (cannot include variables that may have been affected by participation in the experiment)
    - Need to be exogenous

Potential Outcomes Framework

- Also known as Rubin's Causal Model
- What we would like to do is observe and estimate a participant under treatment and under control
  - Not possible

	Outcome	
	Treatment	**Control**
Treatment	X	????
Control	????	X

QUESTIONS?
BREAK!

Potential Outcomes Framework

- Also known as Rubin's Causal Model
- $Y_i(1)$ – i^{th} participant under treatment
- $Y_i(0)$ – i^{th} participant under counterfactual/control
  - Each condition is "potentially" observable, but we cannot observe both conditions for the same participant
- Individual Treatment Effect (ITE)
  - $ITE_i = Y_i(1) - Y_i(0)$
    - Cannot calculate this
  - $ATE = E[\, Y_i(1) - Y_i(0) \,]$
    - Expectation of population average of individual treatment effects

Potential Outcomes Framework

- Rubin has shown that it is possible to estimate the ATE from experimental data

$$\widehat{ATE} = \left(\frac{\sum_{i=1}^{n_1} Y_i}{n_i}\right) - \left(\frac{\sum_{i=1}^{n_0} Y_i}{n_0}\right)$$

- Participants must be randomly assigned so that they are *equal on average in expectation with respect to the outcome*

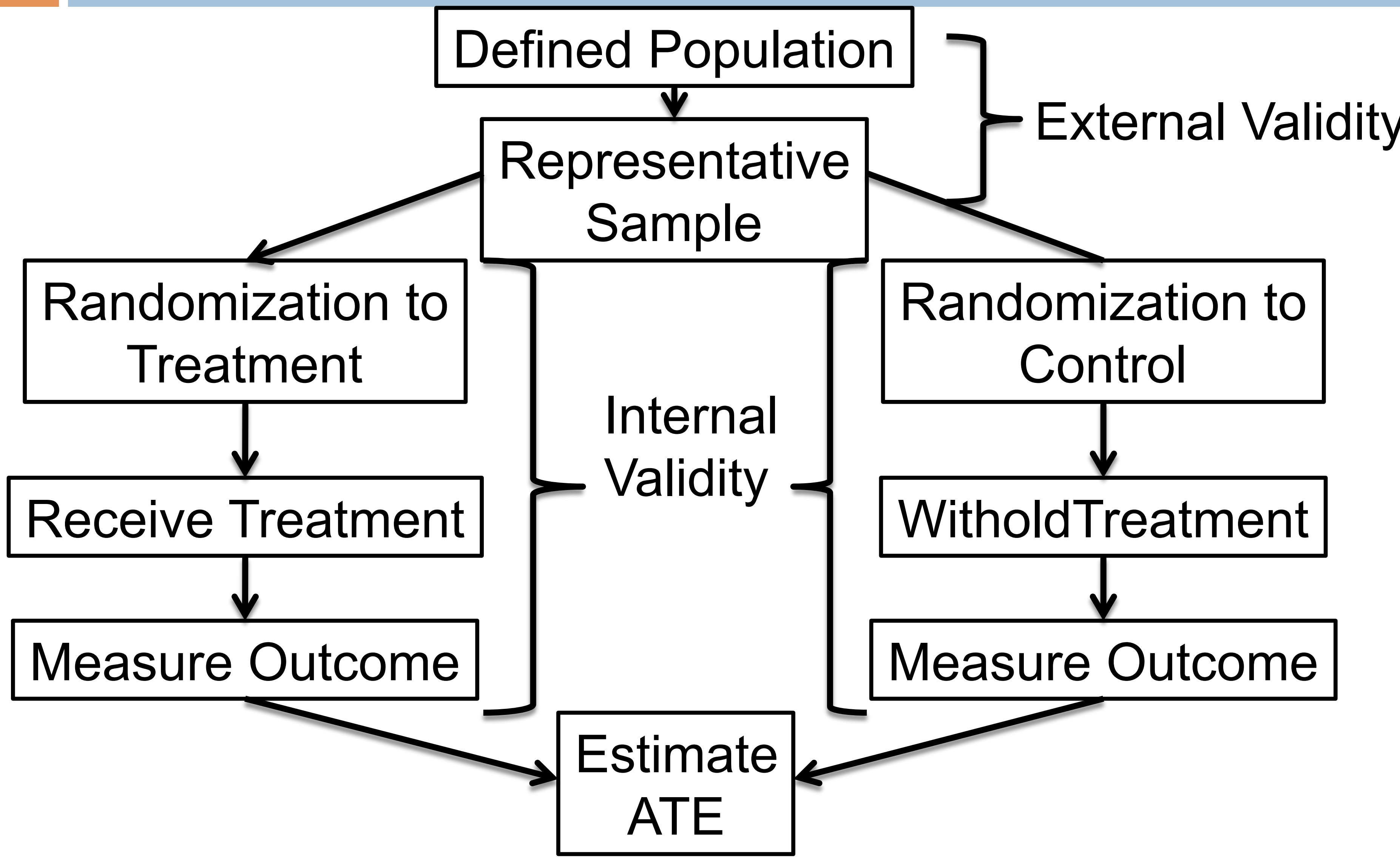
The Value of Experiments

- When assignment is random, all factors *other than treatment status* will tend to be distributed equally between participants in the treatment and control groups.
- Due to random sampling and random assignment, potential members of the treatment and control groups will be identical on all observed and unobserved characteristics **on average in the population.**

The Limits of Experiments

1. External validity may be low due to sample/site selection
2. Little information on why a program works
3. SUTVA (Stable-Unit-Treatment-Value-Assumption)-potential outcomes for each child cannot depend on the group to which other children were assigned (ex-peer groups)
4. Experiments take time and cost money
5. Implementation is often imperfect
6. Are not feasible when it is impossible to exclude the control group from treatment
7. You can't always answer the question of interest:
    1. Partial vs. general equilibrium effects
    2. Total vs. partial treatment effects

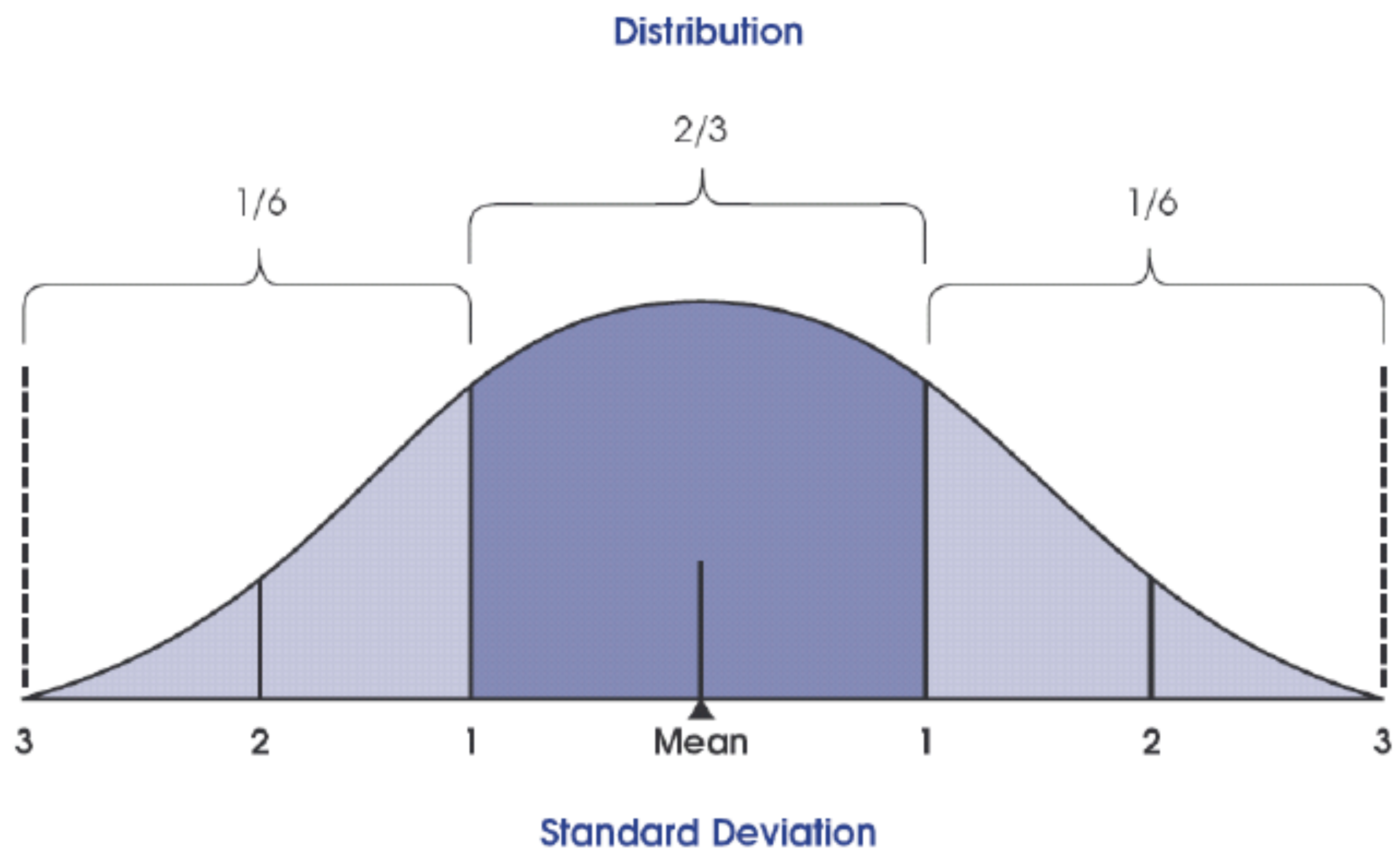
A Two-Group Randomized Experiment



Effect Sizes

- What problem do they solve?
  - Comparing treatment effects on outcomes measured in different units (usually different tests)
- How are they calculated?
  - Option 1: Divide estimated effect by one standard deviation of the outcome variable
  - Option 2: “Standardize” the outcome variable prior to estimation by subtracting its mean and dividing by its standard deviation (to produce a mean of zero and a standard deviation of 1)
- What do they mean? What is a “big” effect size

Effect Size



Calculating Effect Sizes

$$ES = \frac{Y_T - Y_C}{sd(Y)}$$

- The standard deviation of the outcome or of the outcome for the control is used
- Black-white test score gap in 8th grade (NAEP):
  - ~1 S.D.
- Difference in performance between 4th and 8th graders (NAEP):
  - ~1 S.D.
- A very rough rule of thumb: Effect sizes of 0.1 or more are worth talking about

Effect Size Interpretation

Effect Size	% of control group who would be below the average person in the treatment group
0	50%
.1	54%
.2	58%
.25	60%
.3	62%
.4	66%
.5	69%
.6	73%
.7	76%
.8	79%
.9	82%
1	84%