Introduction to Quantitative Research and Program Evaluation Methods

Dennis A. Kramer II, PhD
Assistant Professor of Education Policy
Director, Education Policy Research Center

UF | College of Education
UNIVERSITY of FLORIDA

Agenda for the Day

- Brief Intro
- Overview of Statistical Concepts
- Introduction to Research / Evaluation Methods
- Publicly Available Datasets

Learning Outcomes:

Participants completing this session will take away the following outcomes:

1. Strategies for accessing and managing publicly available higher education data.
2. Techniques for evaluating the efficacy of an education policy and/or program implementation.
3. Creative ways of framing and theorizing education and program evaluation research.
4. Promises and pitfalls of using research evidence in decision-making.

Brief Introduction

UF | College of Education
UNIVERSITY *of* FLORIDA

Introduction to Dr. Kramer

- **What do I do at UF:**
  - Assistant Professor of Education Policy
  - Director, Education Policy Research Center
  - Program Coordinator, Ph.D. in Higher Education Policy
  - Faculty Senator, UF Faculty Academic Senate
  - Member, University Assessment Committee
  - Academic Fellow, Office of Evaluation Sciences (DC)
    - Formerly: White House Behavioral Sciences Team

Introduction to Dr. Kramer

- **Prior positions:**
  - Visiting Assistant Professor of Higher Education, University of Virginia
  - Senior Research and Policy Analyst, Georgia Department of Education
  - Research and Policy Fellow, Knight Commission on Intercollegiate Athletics
  - Assistant Director, Univ. of Southern California's McNair Scholars Program

- **Education:**
  - Ph.D. Higher Education, Institute of Higher Education University of Georgia
  - M.Ed. Postsecondary Administration and Policy, University of Southern California
  - B.S. Clinical & Social Psychology, San Diego State University

The Research/Evaluation Process

UF | College of Education
UNIVERSITY of FLORIDA

Review of Research Concepts

UF College of Education
UNIVERSITY of FLORIDA

Overview of Research Approaches

- Lack of a single, appropriate methodological approach to study education

- Two major approaches
  - Quantitative
  - Qualitative

Overview of Research Approaches

- Differentiating characteristics
  - Goals
    - **Quantitative**: tests theory, establishes facts, shows relationships, predicts, or statistically describes
    - **Qualitative**: develops grounded theory, develops understanding, describes multiple realities, captures naturally occurring behavior
  - Research design
    - **Quantitative**: highly structured, formal, and specific
    - **Qualitative**: unstructured, flexible, evolving

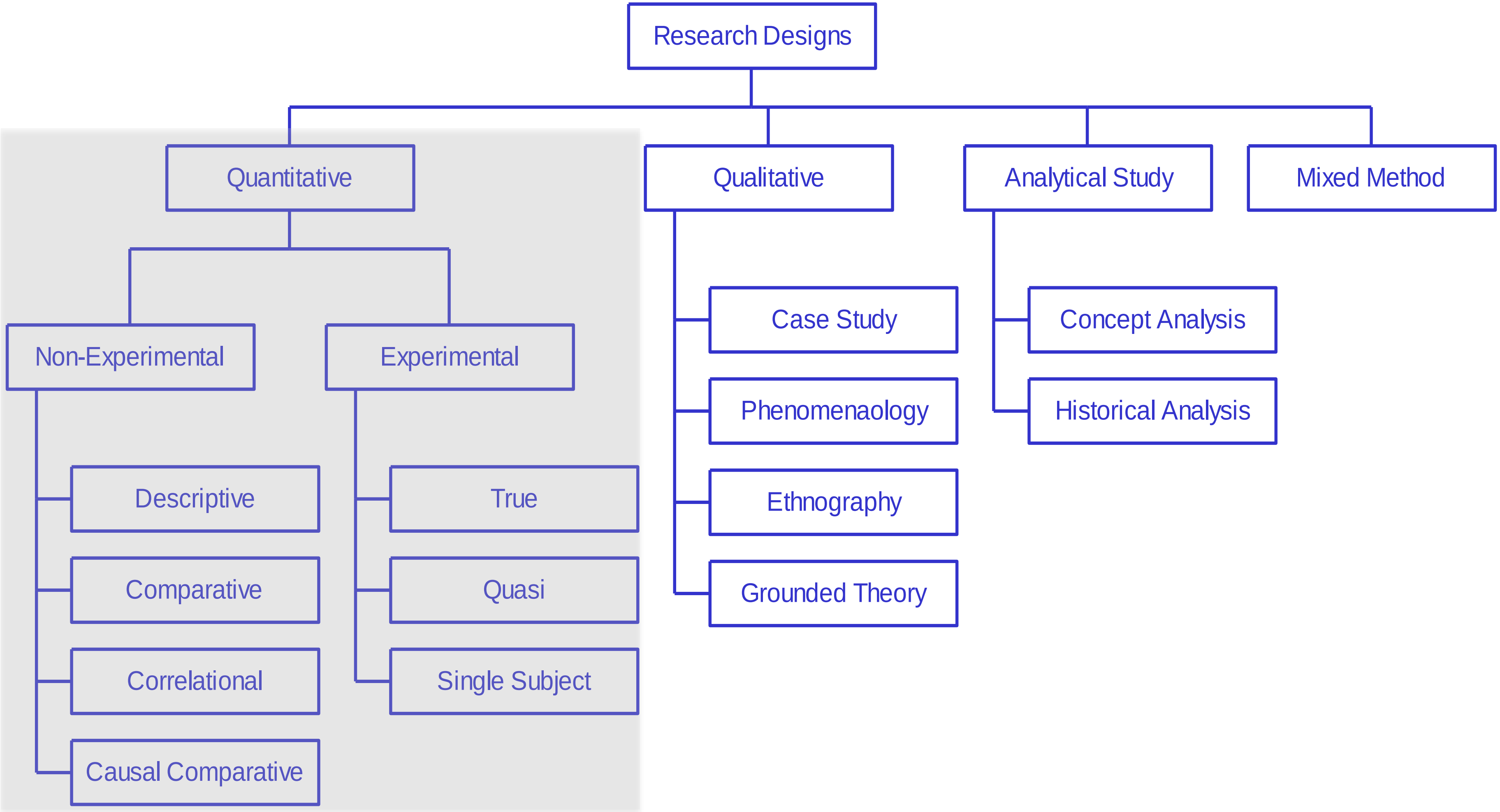
Overview of Research Approaches

- Differentiating characteristics
  - Participants
    - **Quantitative:** many participants representative of the groups from which they were chosen using probabilistic sampling techniques
    - **Qualitative:** few participants chosen using non-probabilistic sampling techniques for specific characteristics of interest to the researchers
  - Data, data collection, and data analysis
    - **Quantitative**: numerical data collected at specific times from tests or surveys and analyzed statistically
    - **Qualitative**: narrative data collected over a long period of time from observations and interviews and analyzed using interpretive techniques

Overview of Research Approaches

- Differentiating characteristics
  - Researcher's role
    - Quantitative: detached, objective observers of events
    - Qualitative: participant observers reporting participant's perspectives understood only after developing long-term, close, trusting relationships with participants
  - Context
    - Quantitative: manipulated and controlled settings
    - Qualitative: naturalistic settings

Types of Research Design



Quantitative Designs

- Differentiating the three types of experimental designs
  - True experimental
    - Random assignment of subjects to groups

{Not really experimental, but close}

  - Quasi-experimental
    - Non-random assignment of subjects to groups
  - Single subject
    - Non-random selection of a single subject

Quantitative Designs

- Differentiating the four types of non-experimental designs
  - Descriptive
    - Makes careful descriptions of the current situation or status of a variable(s) of interest
  - Comparative
    - Compares two or more groups on some variable of interest
  - Correlational
    - Establishes a relationship (i.e., non-causal) between or among variables
  - Ex-post-facto
    - Explores possible causes and effects among variables that cannot be manipulated by the researcher.

Correlation vs. Causation

- Correlation tells us two variables are related
- Types of relationship reflected in correlation
  - X causes Y or Y causes X (causal relationship)
  - X and Y are caused by a third variable Z (spurious relationship)
- In order to imply causation, a true experiment (or a really good quasi-experimental study) must be performed where subjects are randomly assigned (or approximated) to different conditions

Correlation vs. Causation

- Research has found that ice-cream sales and deaths are linked. As ice-cream sales goes up, so do drownings.
  - We can conclude that ice-cream consumption causes drowning, right?

- Why can't we conclude this?
- What are some possible alternative explanations?

Introduction to Research Analysis

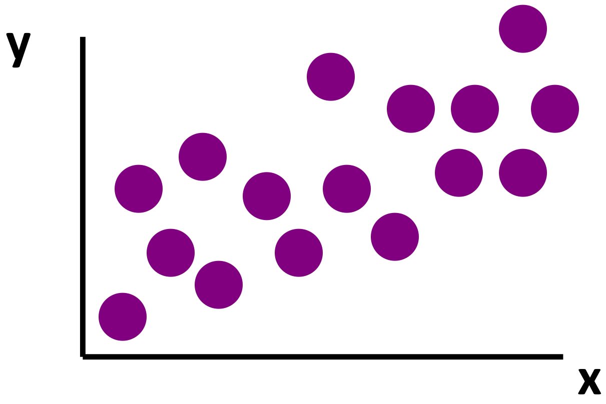
UF | College of Education
UNIVERSITY *of* FLORIDA

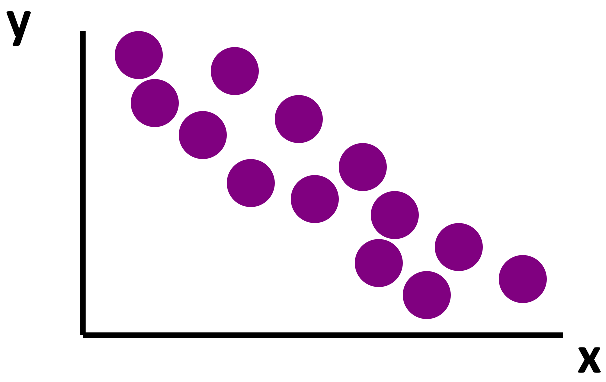
Scatter Plot and Correlation

- A scatter plot (or scatter diagram) is used to show the relationship between two variables
- Correlation analysis is used to measure strength of the association (linear relationship) between two variables
  - Only concerned with strength of the relationship
  - No causal effect is implied

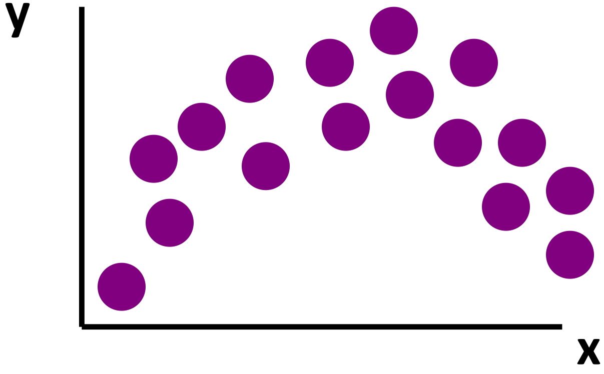
Scatter Plot Example

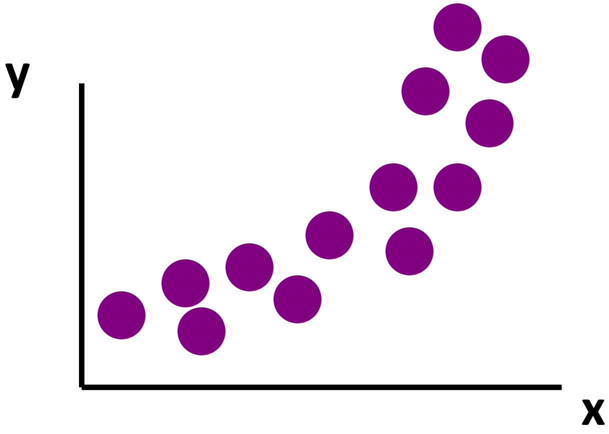
Linear relationships





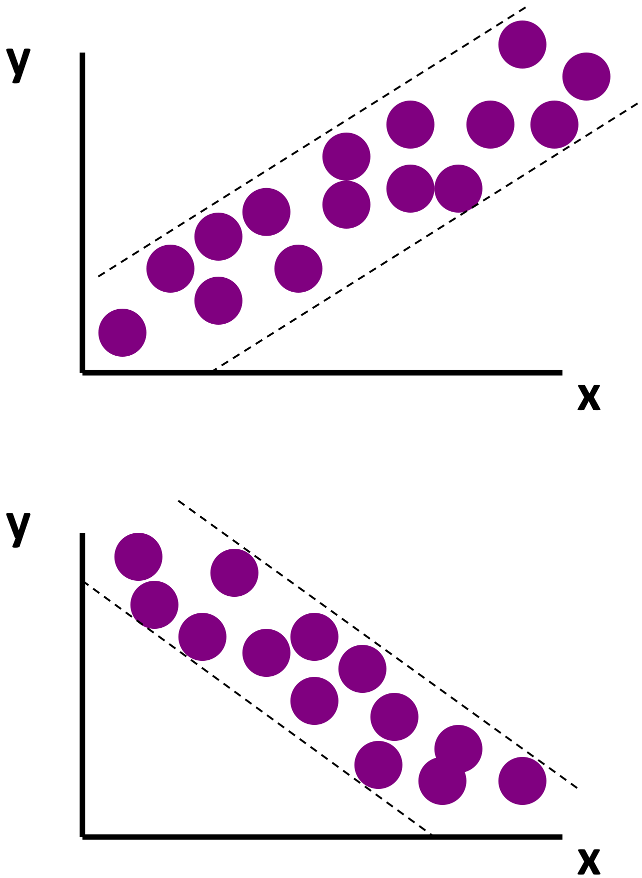
Non-linear / curvilinear relationships



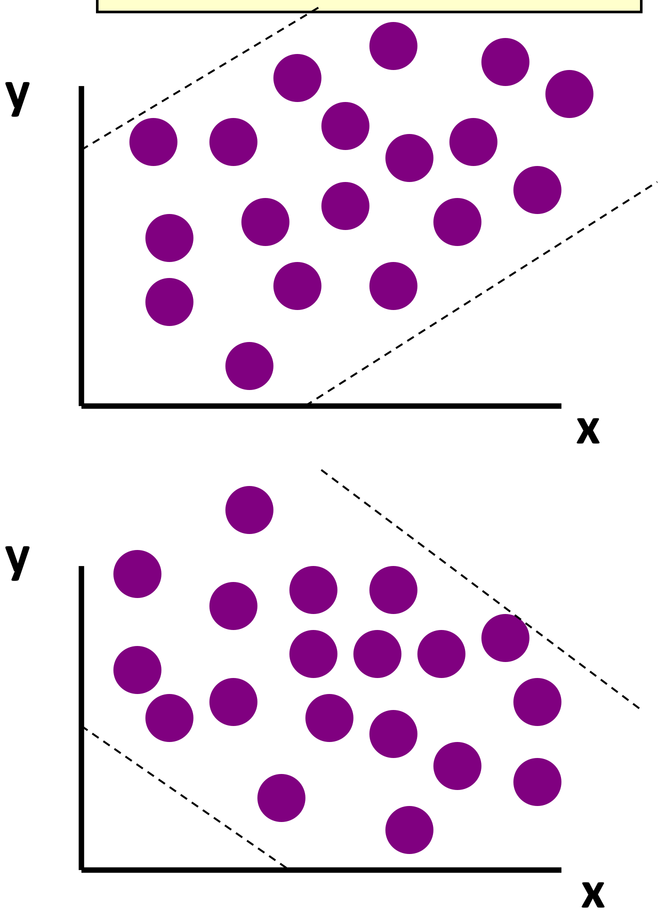


Scatter Plot Example

Strong relationships

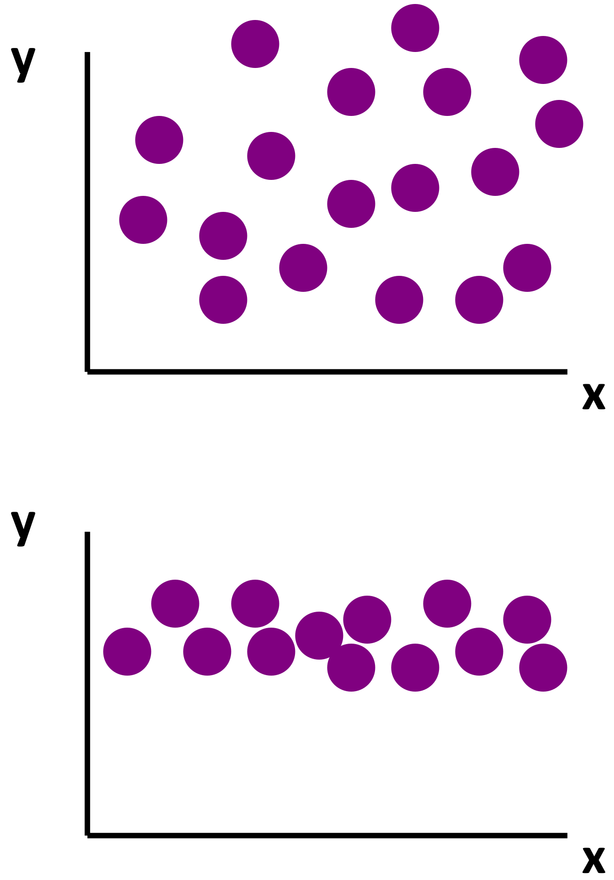


Weak relationships



Scatter Plot Example

No relationship



Correlation Coefficient

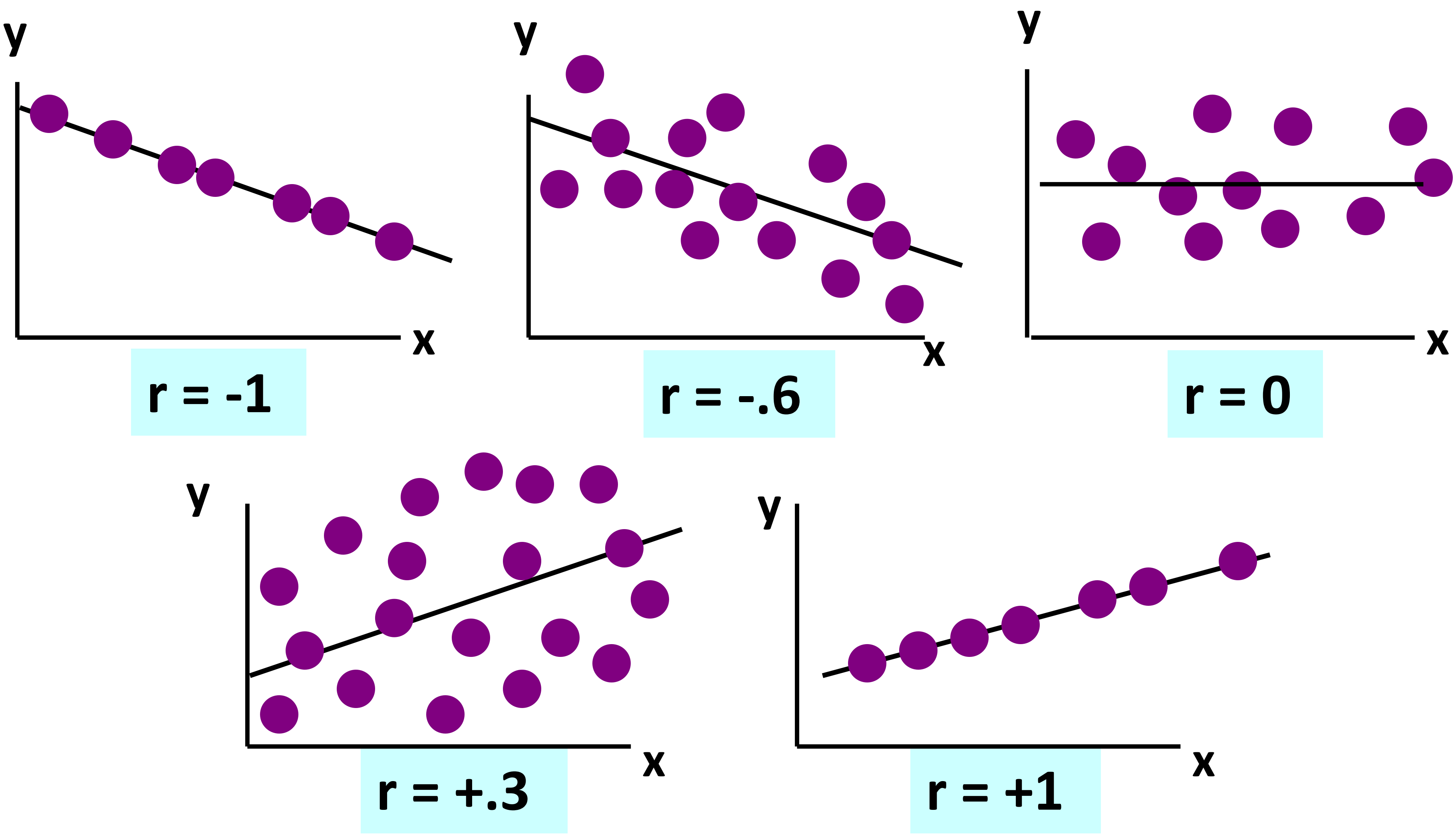
- The population correlation coefficient p (rho) measures the strength of the association between the variables

- The sample correlation coefficient r is an estimate of p and is used to measure the strength of the linear relationship in the sample observations

Correlation Coefficient

- Unit free
- Range between -1 and 1
- The closer to -1, the stronger the negative linear relationship
- The closer to 1, the stronger the positive linear relationship
- The closer to 0, the weaker the linear relationship

Examples of r Values (approximate)



Simple Linear Regression

UF College of Education
UNIVERSITY of FLORIDA

Two Main Objectives

- Establish is there is a **relationship** between two variables
  - More specifically, establish a **statistically significant** relationship between two variables
  - Examples: Income and spending; wage and gender; height and exam score.

- **Forecast** new observations
  - Can we use what we know about the relationship to forecast unobserved values?
  - Examples: What will our enrollment for next fall? How many incidents will be have in the residence hall next week?

Variable Roles

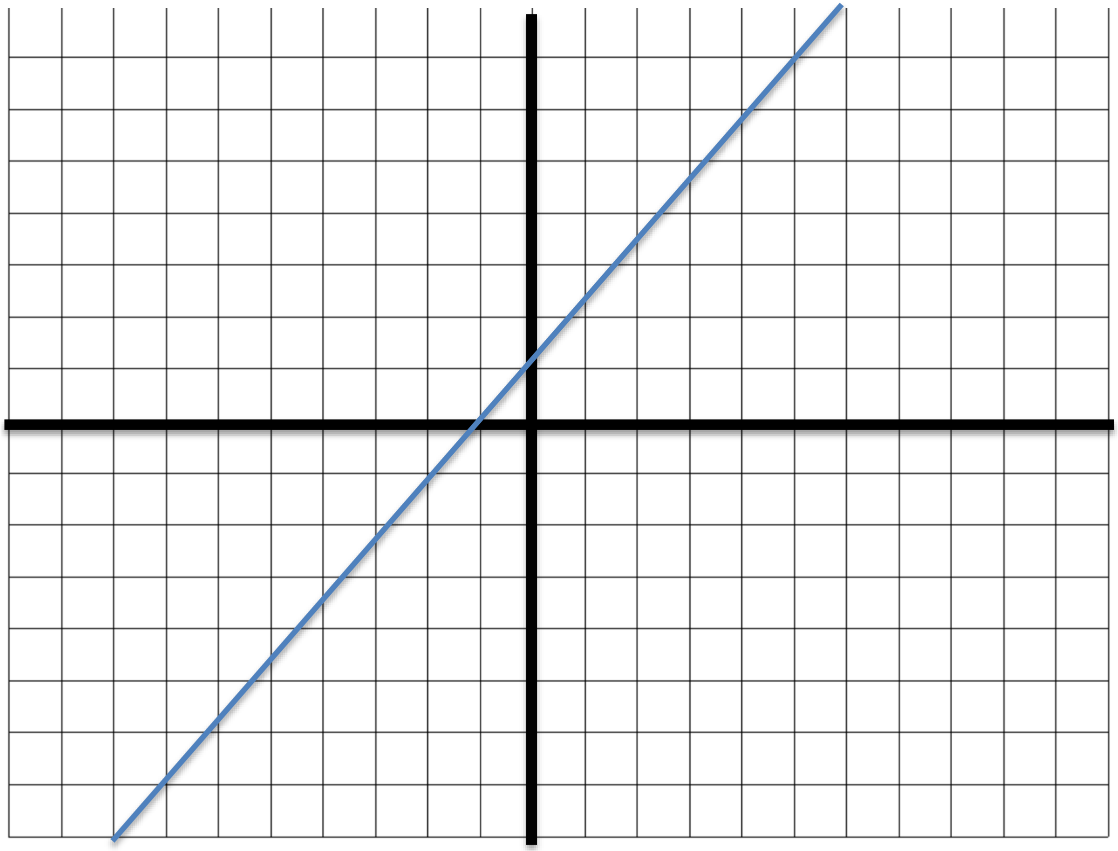
- **Dependent Variable**
  - This is the variable whose value we want to explain or forecast
  - Its value DEPENDS on something else
  - In most regression models this will be denoted by y.
- **Independent Variable**
  - This is the variable that explains variation in the dependent variable
  - Its value are independent
  - In most regression models this will be denoted by X.

The Magic: A Linear Equation

- You may remember one of these from high school
  - $y = a + bx$
  - $y = mx + b$
- In the stats/econ world we like to complicate things and we use a different notation.
  - $y = \beta_0 + \beta_1 x$
    - β_0 = constant
    - $\beta_1 x$ = independent variable / slope of x
- We call it linear because the equation represents a straight line in a bi-dimensional plot.

Linear Regression Example

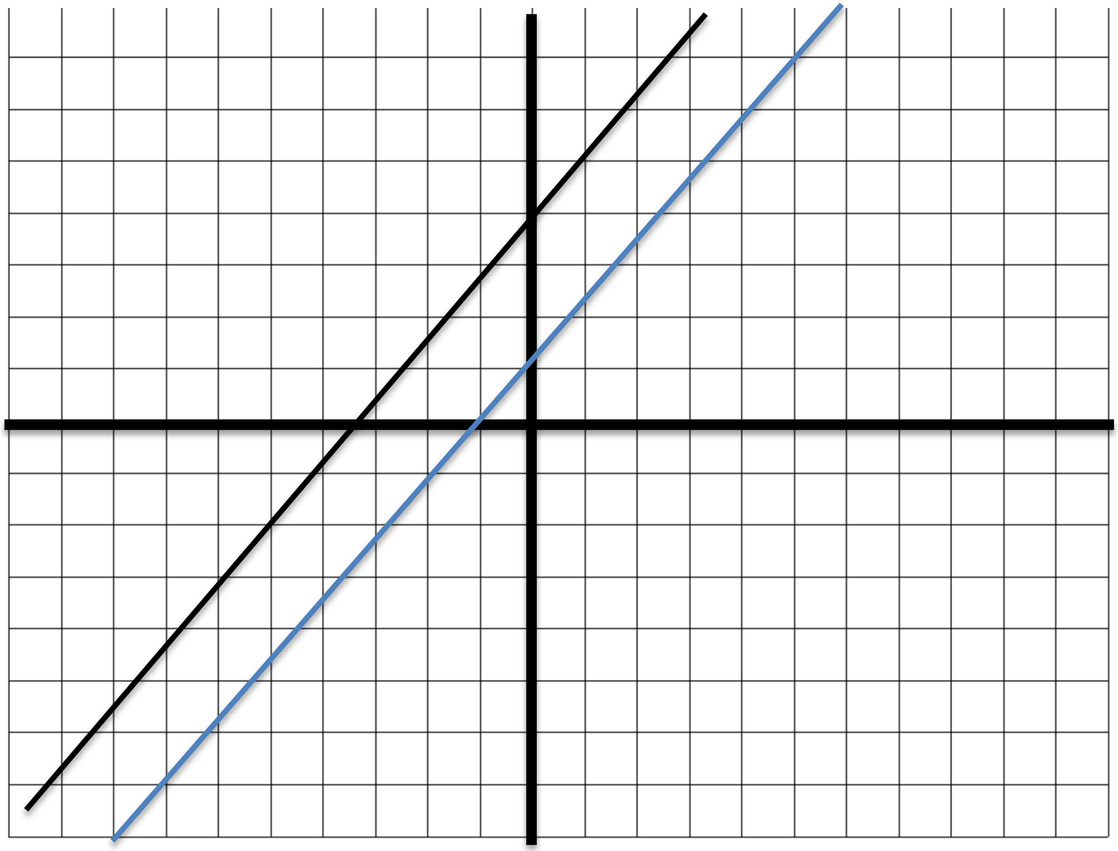
- $y = \beta_0 + \beta_1 x$
  - $y = 1 + 1x$



Linear Regression Example

- $y = \beta_0 + \beta_1 x$
  - $y = 1 + 1x$

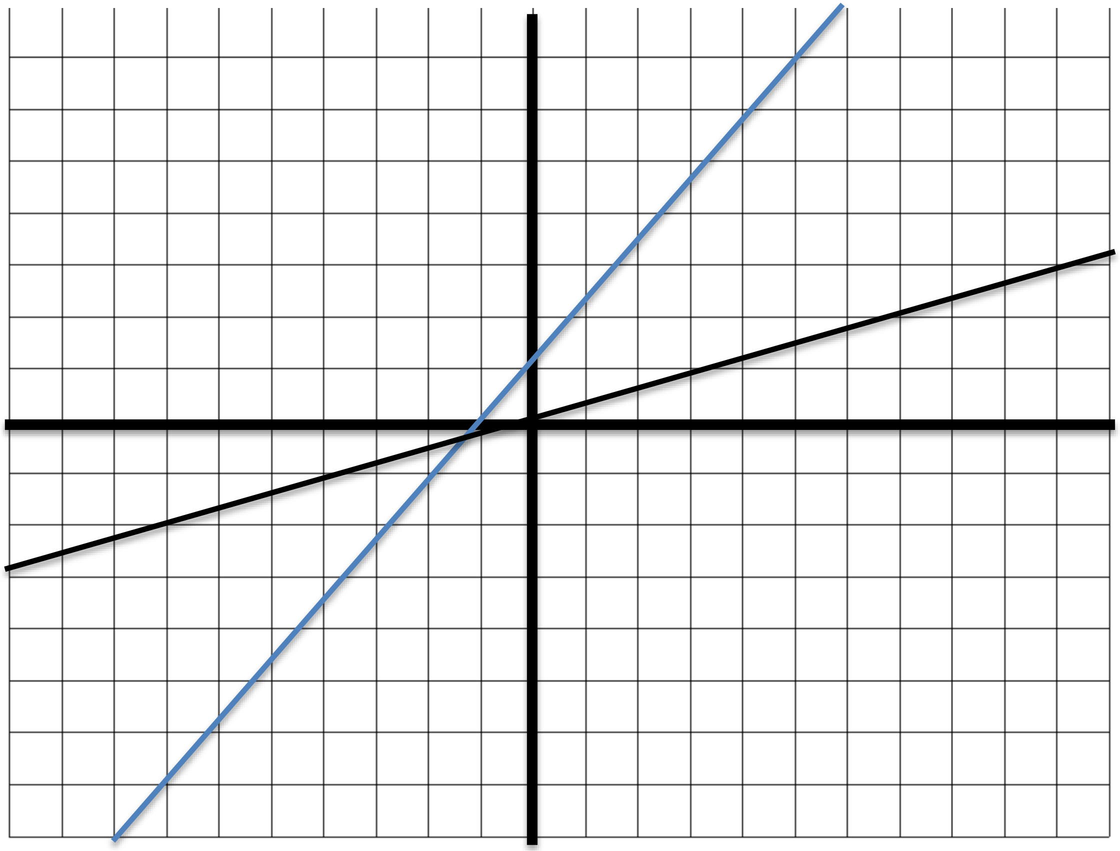
- What happens if the intercept changes from 1 to 4?
  - $y = 4 + 1x$



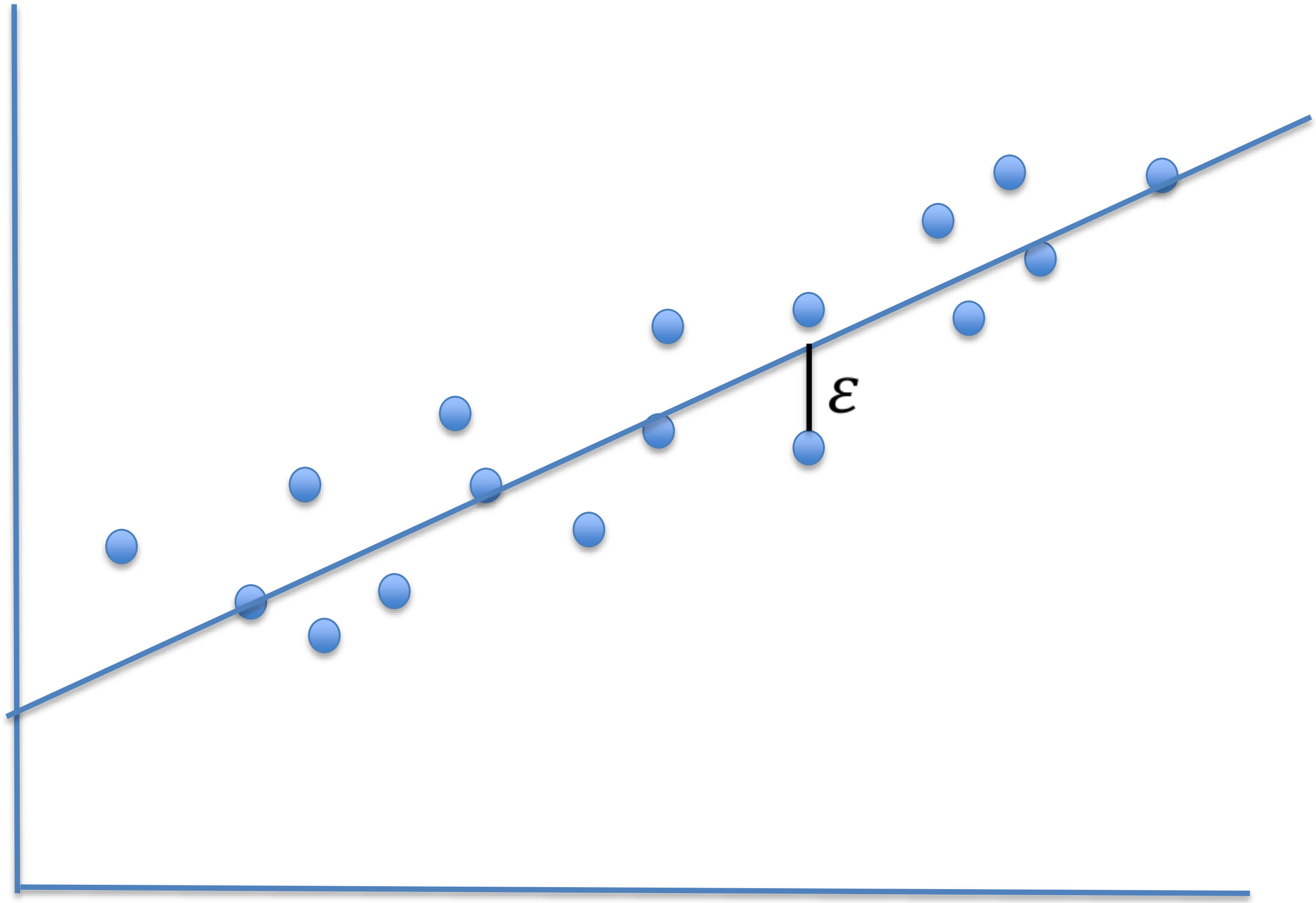
Linear Regression Example

- $y = \beta_0 + \beta_1 x$
  - $y = 1 + 1x$

- What happens if the slope changes from 1 to 0.3?
  - $y = 1 + 0.3x$



The World is Not Perfectly Linear



Simple Linear Regression Model is Now

- $y = \beta_0 + \beta_1 x + \varepsilon$
  - Where y is the dependent variable
  - x is the independent variable that explains y
  - β_0 is the constant or intercept
  - β_1 is x's slope or coefficient
  - ε is now our error term
    - We try to minimize our error

Statistically Significant Relationship

- General Rule: If zero (0) is outside of our 95% confident interval, we claim there is a statistically significant relationship.

- Formally, we reject the (null) hypothesis that there is no relationship or that 0 is a possible value for the slope.

- Since we reject the null hypothesis, we accept the alternate hypothesis that 0 is not a possible value for the slope.

Statistically Significant Relationship

- **Another General Rule**: if the p-value is below 5% (0.05), we can there is a statistically significant relationship.
  - This is used more than confidence intervals
- What are p-values
  - These values are reported as standard outputs in statistical software packages (STATA – yay!)
  - Roughly speaking, they represent the probability that we reject the null hypothesis when it is actually true. In other words, the probability that there is no relationship.

Oh the stars …

- Within academic journals you will see results that have some version of *** associated with it to denote a significant relationship:
  - + = $p<0.10$
    - Meaning you are 90% confident that there is a significant relationship greater or less than zero
  - * = $p<0.05$
    - Meaning we are 95% confident that there is a significant relationship greater or less than zero
  - ** = $p<0.01$
    - Meaning we are 99% confident that there is a significant relationship greater or less than zero.
  - *** = $p<0.001$
    - Meaning we are 99.9% confident that there is a significant relationship greater or less than zero

Key takeaways (from this section)

- Sampling induces uncertainty in our estimates
- We find that 95% confidence interval of a coefficient by computing two (2) standard errors above and below the point estimate of the coefficient.
- If the confidence interval includes zero, we say there is no statistically significant relationship. If it excludes zero then there is!!
- We can also check the p-values. If it is above 0.05 we say there is no statistically significant relationship. It it is below, then there is a statistically significant relationship.

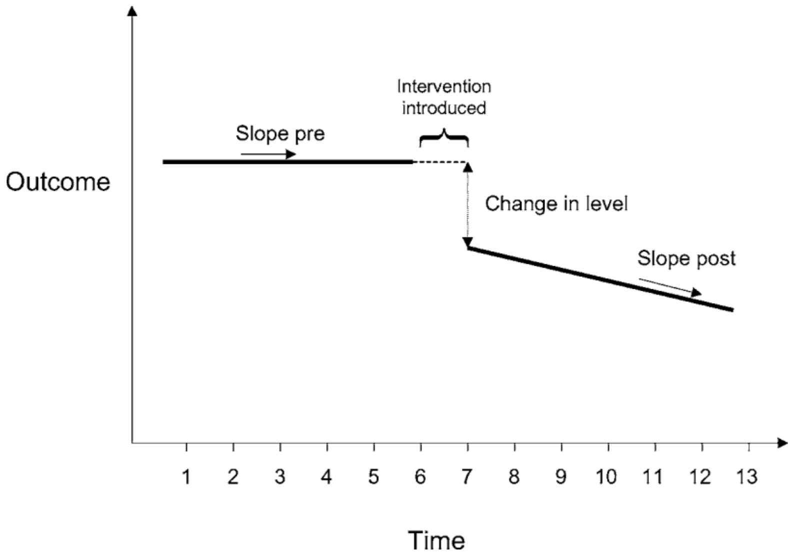
Overview of More Advanced Techniques

UF College of Education
UNIVERSITY of FLORIDA

Interrupted Time Series

- This design uses several waves of observation before and after the introduction of the independent (treatment) variable X.
- It is diagrammed as follows:

O1 O2 O3 O4 X O5 O6 O7 O8

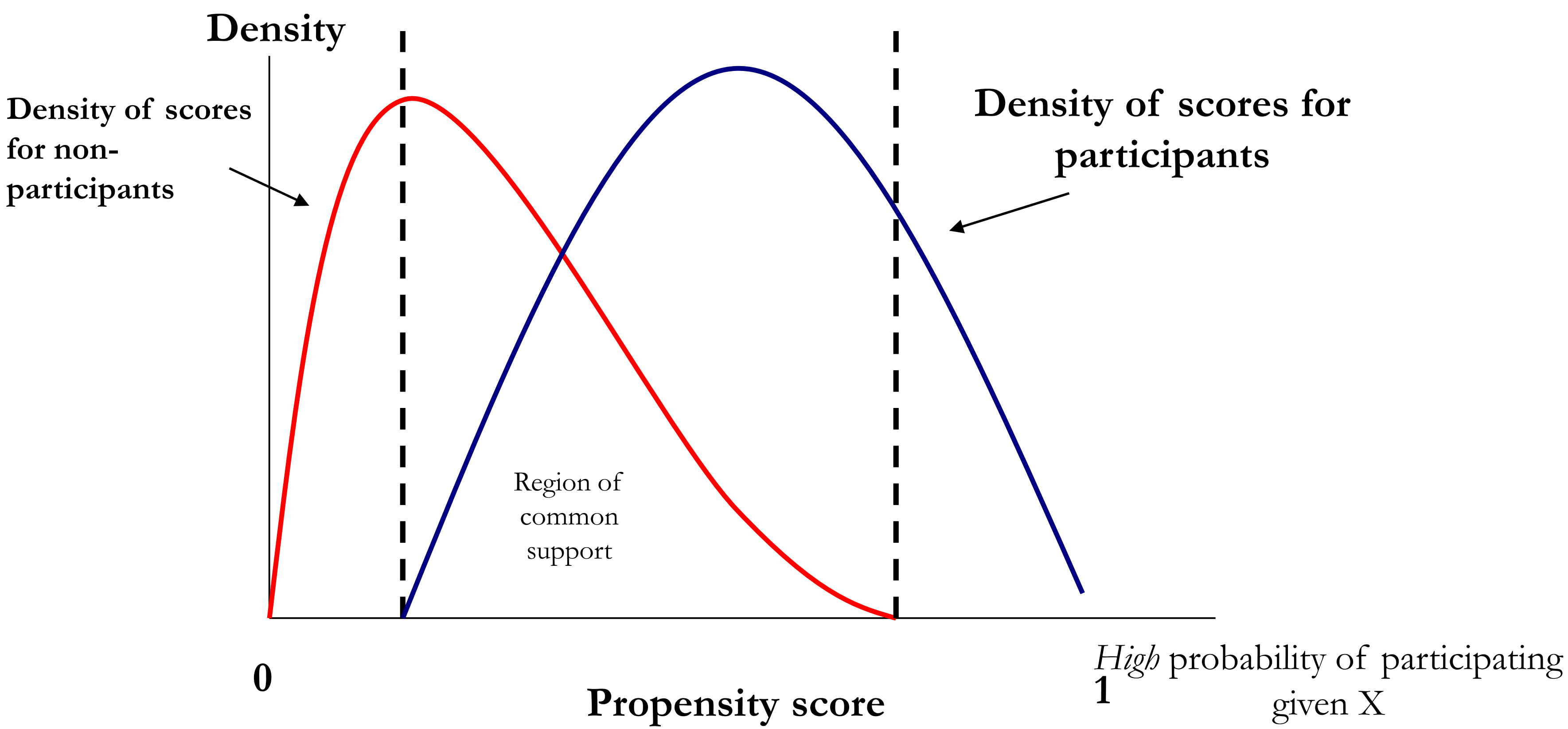


Propensity Score Matching

- *Propensity score matching:* match treated and untreated observations on the estimated probability of being treated (propensity score). **Most commonly used.**

- Match on the basis of the **propensity score**
  - $P(X) = Pr\ (d=1 | X)$
  - D indicates participation in project
  - Instead of attempting to create a match for each participant with exactly the same value of X, we can instead match on the probability of participation.

Propensity Score Matching



Propensity Score Matching

Steps for Score Matching

1. Need representative and comparable data for both treatment and comparison groups

2. Use a logit (or other discrete choice model) to estimate program participations as a function of observable characteristics

3. Use predicted values from logit to generate propensity score $p(x_i)$ for all treatment and comparison group members

Difference-in-Differences

(Comparative Interrupted Time Series)

- The simple DID is almost a cliché at this point:
  - 2 Groups
  - 2 Time Periods
  - One group is exposed to treatment between periods.
  - Design can avoid bias from special classes of omitted variables

Difference-in-Differences

(Comparative Interrupted Time Series)

- The classic DID estimator is the difference between two before – after differences.
  - Before after change observed in the treatment group.
  - Before after change observed in the control group.

- The idea is that the simple pre-post design may be biased because of unobserved factors that affect outcomes and that changed along with the treatment.
- If these unobserved factors also affected the control group, then double differencing can remove the bias and isolate the treatment effect.

Difference-in-Differences

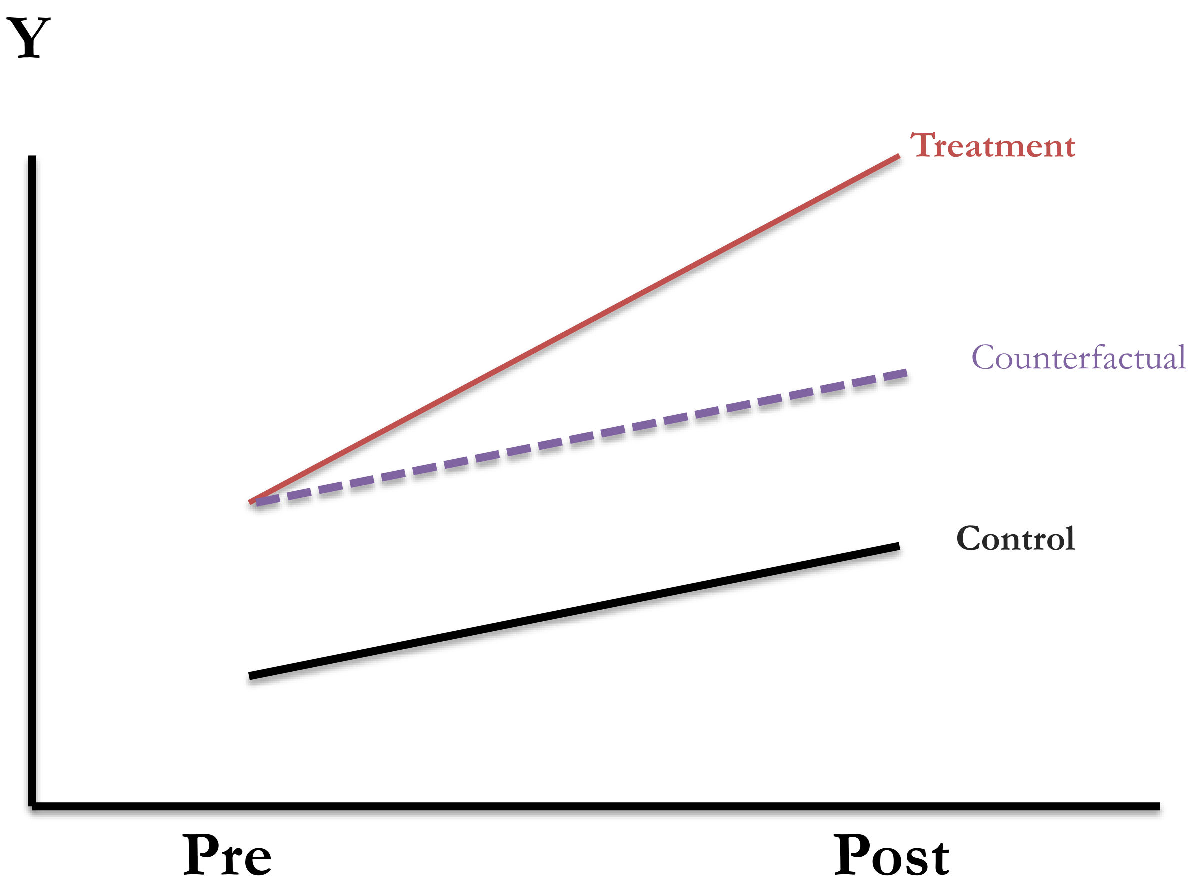
(Comparative Interrupted Time Series)

- The classic DID estimator is the difference between two before – after differences.
  - Before after change observed in the treatment group.
  - Before after change observed in the control group.

- The idea is that the simple pre-post design may be biased because of unobserved factors that affect outcomes and that changed along with the treatment.
- If these unobserved factors also affected the control group, then double differencing can remove the bias and isolate the treatment effect.

Difference-in-Differences

(Comparative Interrupted Time Series)

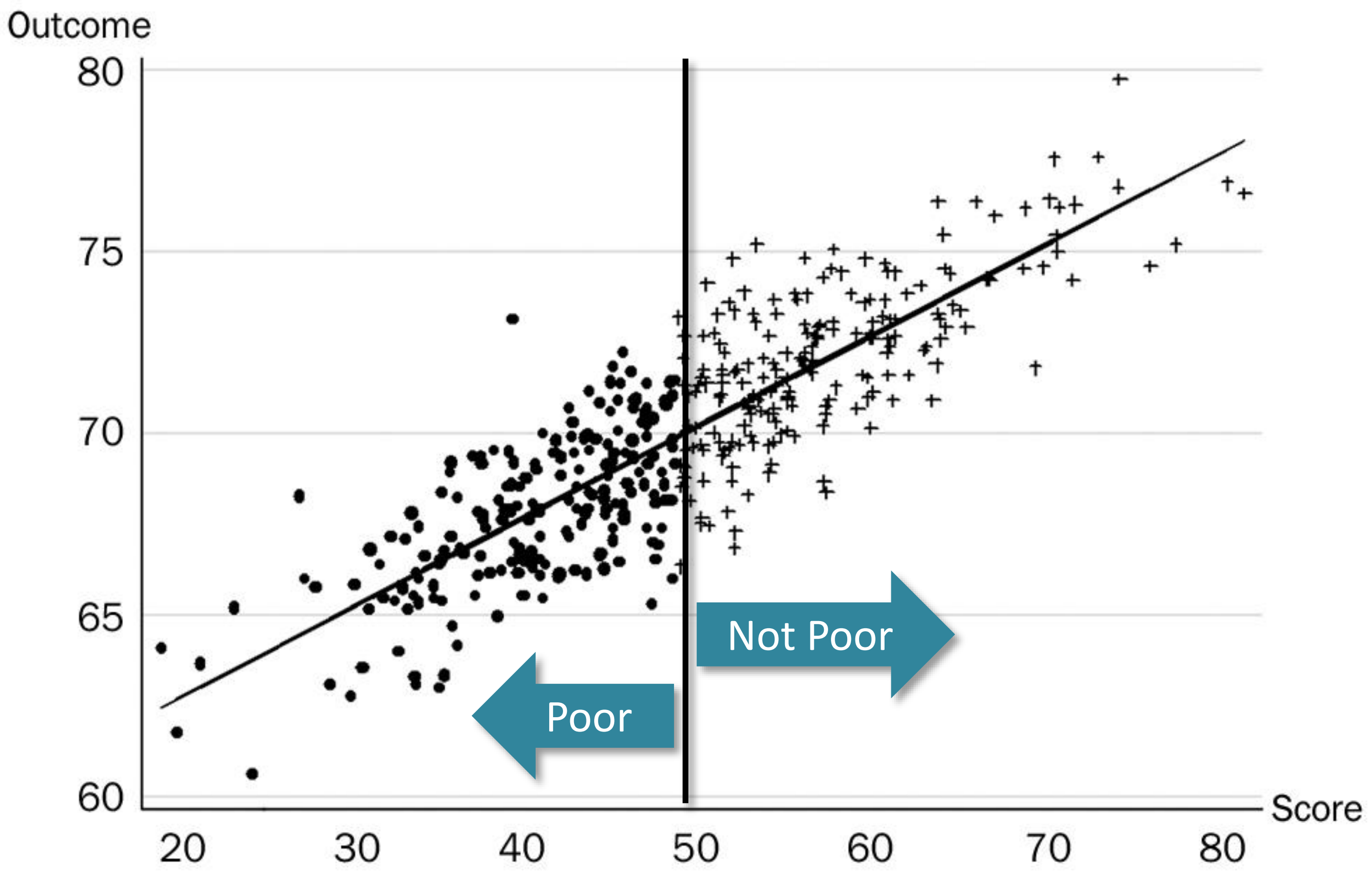


Regression Discontinuity

- A useful method for determining whether a program of treatment is effective
- Participants are assigned to program or comparison groups based on *a cutoff score* on a pretest
  - e.g. Evaluating new learning method to children who obtained low scores at the previous test.
    - Cutoff score = 50
    - The treatment group: children who obtained 0 to 50
    - The comparison group: children who obtained 51 to 100
- The program (treatment) can be given to those most in need

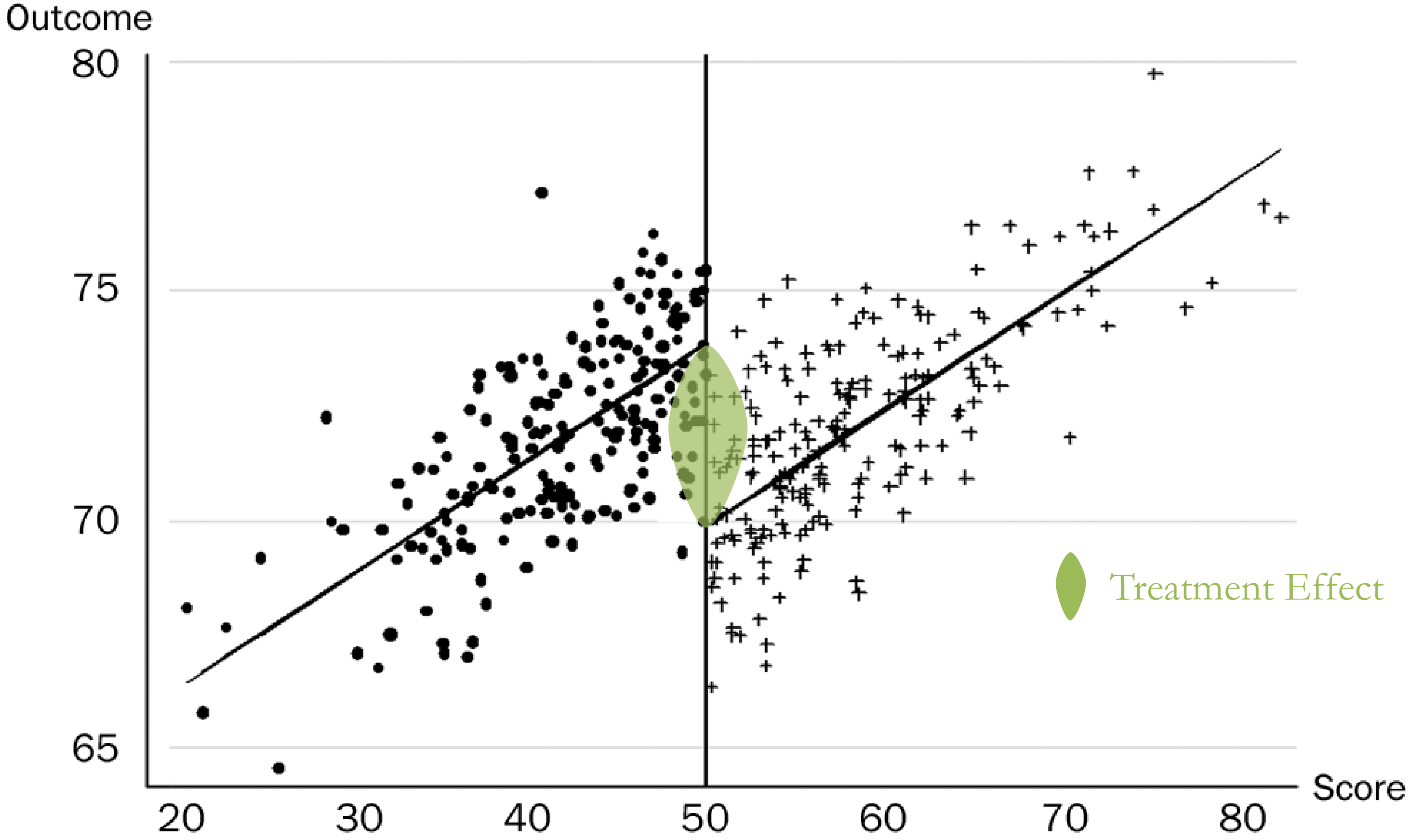
Regression Discontinuity

- **Baseline (prior to the treatment)**



Regression Discontinuity

- **Post Treatment**



Randomized Control Trials (RCTs)

- A randomized controlled trial (RCT) is a way of doing impact evaluation in which the population receiving the program or policy intervention is chosen at random from the eligible population, and a control group is also chosen at random from the same eligible population.
  - It tests the extent to which specific, planned impacts are being achieved.

- The distinguishing feature of an RCT is the random assignment of members of the population eligible for treatment to either one or more treatment groups or to the control group.
  - The effects on specific impact areas for the different groups are compared after set periods of time.

Randomized Control Trials (RCTs)

- The simplest RCT design has one treatment group (or 'arm') and a control group. Variations on the design are to have either:
  - **multiple treatment arms**, for example, one treatment group receives intervention A, and a second treatment group receives intervention B, or
  - a **factorial design**, in which a third treatment arm receives both interventions A and B
- In situations where an existing intervention is in use, it is more appropriate for the control group to continue to receive this, and for the RCT to show how well the new intervention compares to the existing one.

Selecting a method …

Design	When to use	Advantages	Disadvantages
Randomization	Whenever feasible When there is variation at the individual or community level	Gold standard Most powerful	Not always feasible Not always ethical
Regression Discontinuity	If an intervention has a clear, sharp assignment rule	Project beneficiaries often must qualify through established criteria	Only look at sub-group of sample Assignment rule in practice often not implemented strictly
Difference-in-Differences	If two groups are growing at similar rates Baseline and follow-up data are available	Eliminates fixed differences not related to treatment	Can be biased if trends change Ideally have 2 pre-intervention periods of data
Matching	When other methods are not possible	Overcomes observed differences between treatment and comparison	Assumes no unobserved differences (often implausible)

Level of Causality

Data Analysis Example

UF | College of Education
UNIVERSITY *of* FLORIDA

Data Example

- RQ – Interested in the effect of remediation course on English 101 performance.
- Intervention is assigned to students receiving below a 50 on the placement test
- Four years of data with only two (2) years in which the policy treatment was in place.
- You are tasked with advising institutional leaders on if the policy should remain in place

Data Example

- You have the following data points:
  - English 101 Grade
  - Placement Test Score
  - Race / Ethnicity
  - Gender
  - High School GPA
  - SAT Score

- Based on the conversation today, which of the following methods would you propose to use?
  - OLS / Linear Regression
  - Propensity Score Matching
  - Difference-in-Differences
  - Regression Discontinuity

Data Example

Table 1: Example Estimates

	OLS		PSM		DiD		RD	
	w/o covs	w/ covs	w/o covs	w/ covs	w/o covs	w/ covs	w/o covs	w/ covs
English 101 Grade	0.269 *	0.187	0.092	-0.127	26.082 ***	26.082 ***	12.487***	12.345***
	(0.131)	(0.163)	(0.231)	(0.182)	(0.215)	(0.215)	(0.682)	(0.684)
# of Observations	30,385	30,385	16,548	16,548	30,385	30,385	15,227	15,227
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes. robust standard errors in parentheses; + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001*

Institutional Data Sets

UF | College of Education
UNIVERSITY of FLORIDA

K-12 Dataset

- Elementary / Secondary Information System
  - The Elementary/Secondary Information System (ElSi) is an NCES web application that allows users to quickly view public and private school data and create custom tables and charts using data from the Common Core of Data (CCD) and Private School Survey (PSS).
  - https://nces.ed.gov/ccd/elsi/default.aspx?agree=0

Institutional Datasets

- ELSi Speed Challenge
  - I am going to place four (4) questions on the board that need to be answered by pulling data from the ELSi system.
  - The first person to bring a correct answer to ALL four (4) questions will not have to complete one (1) chapter from the Pollock book.
  - This is an individual exercise.

Institutional Datasets

- The Integrated Postsecondary Education Data System (IPEDS)
  - IPEDS is a system of interrelated surveys conducted annually by the National Center for Education Statistics (NCES), a part of the Institute for Education Sciences within the United States Department of Education. IPEDS consists of twelve interrelated survey components that are collected over three collection periods (Fall, Winter, and Spring) each year as described in the Data Collection and Dissemination Cycle. The completion of all IPEDS surveys is mandatory for all institutions that participate in, or are applicants for participation in, any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended. Statutory Requirements For Reporting IPEDS Data.
  - http://nces.ed.gov/ipeds/

Institutional Datasets

- Delta Cost Data
  - The Delta Cost Project uses publicly available data to clarify the often daunting world of higher education finance. Delta staff translate the data into formats that can be used for long-term analyses of trends in money received and money spent in higher education. Using four key metrics, researchers produce trend and other analytic reports and presentations that help policy makers understand what is happening in higher education finance.
  - http://www.deltacostproject.org/

Institutional Datasets

- Campus Safety and Security
  - The Campus Safety and Security Data Analysis Cutting Tool is brought to you by the Office of Postsecondary Education of the U.S. Department of Education. This analysis cutting tool was designed to provide rapid customized reports for public inquiries relating to campus crime and fire data. The data are drawn from the OPE Campus Safety and Security Statistics website database to which crime statistics and fire statistics (as of the 2010 data collection) are submitted annually, via a web-based data collection, by all postsecondary institutions that receive Title IV funding (i.e., those that participate in federal student aid programs). This data collection is required by the Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act and the Higher Education Opportunity Act.
  - http://ope.ed.gov/campussafety/#/

Institutional Datasets

- Intercollegiate Athletics
  - The Equity in Athletics Data Analysis Cutting Tool is brought to you by the Office of Postsecondary Education of the U.S. Department of Education. This analysis cutting tool was designed to provide rapid customized reports for public inquiries relating to equity in athletics data. The data are drawn from the OPE Equity in Athletics Disclosure Website database. This database consists of athletics data that are submitted annually as required by the Equity in Athletics Disclosure Act (EADA), via a Web-based data collection, by all co-educational postsecondary institutions that receive Title IV funding (i.e., those that participate in federal student aid programs) and that have an intercollegiate athletics program.
  - http://ope.ed.gov/athletics/#/

Institutional Datasets

- National Survey of Student Engagement
  - The National Survey of Student Engagement (NSSE) (pronounced: nessie) is a survey mechanism used to measure the level of student participation at universities and colleges in Canada and the United States as it relates to learning and engagement. The results of the survey help administrators and professors to assess their students' student engagement. The survey targets first-year and senior students on campuses. NSSE developed ten student Engagement Indicators (EIs) that are categorized in four general themes: academic challenge, learning with peers, experiences with faculty, and campus environment. Since 2000, there have been over 1,600 colleges and universities that have opted to participate in the survey. Additionally, approximately 5 million students within those institutions have completed the engagement survey. Overall, NSSE assesses effective teaching practices and student engagement in educationally purposeful activities. The survey is administered and assessed by Indiana University School of Education Center for Postsecondary Research.
  - http://nsse.indiana.edu/html/report_builder.cfm

IPEDS Activity

UF | College of Education
UNIVERSITY *of* FLORIDA

IPEDS Activity

- Go to http://nces.ed.gov/ipeds/
- We are going to walk/talk through how to extract data from IPEDS
  - This is the primary dataset for secondary data researchers within higher education
  - It has a wealth of information

Questions?

UF | College of Education
UNIVERSITY of FLORIDA

Contact Information

- If there is anything I can to help, please contact me

Dennis A. Kramer II, Ph.D.
Norman Hall 293
352.273.4315
dkramer@coe.ufl.edu