



# Practical Methods for Designing Research Studies and Analyzing Data in Educational Settings

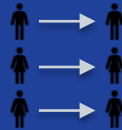
Whitney Zimmerman

Assistant Teaching Professor, Dept. of Statistics, Penn State





Randomized Experiments



Pre-Test / Post-Test



Quasi-Experiments



Propensity Scoring

# Measuring Intervention Impact

# Randomized Experiments

## Experiment + Randomization

A research study where the researchers randomly assign participants to different treatment conditions



Treatment

Control

# Randomized Experiments



Causal conclusions



Still need to worry  
about sampling issues  
and threats to validity

# Randomized Experiments



Pros

Causal conclusions

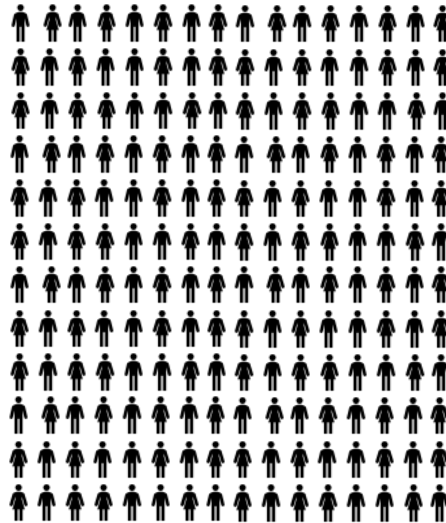


Cons

Still need to worry about sampling issues and threats to validity

Sampling

Population



Sample



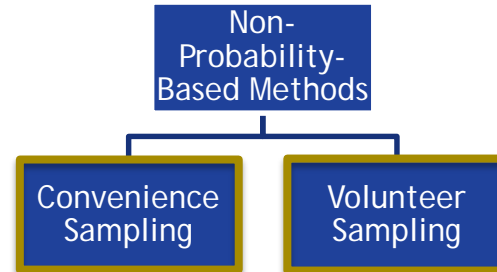
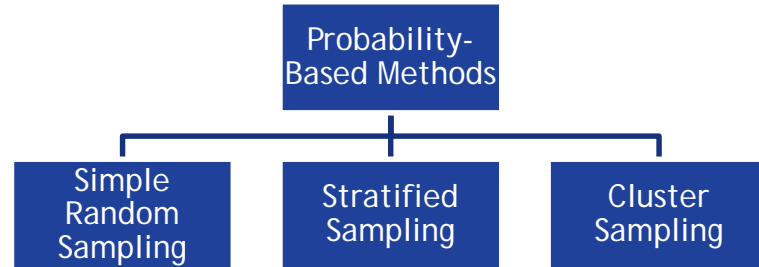
# Randomized Experiments



Causal conclusions



Still need to worry about sampling issues and threats to validity



# Randomized Experiments



Causal conclusions



Still need to worry  
about sampling issues  
and threats to validity

## Threats to Validity

Diffusion of Treatment  
Hawthorne Effect  
Novelty Effect  
Experimental Mortality

# Randomized Experiments



Causal conclusions



Still need to worry  
about sampling issues  
and threats to validity

Threats to Validity

**Diffusion of Treatment**



[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)



# Randomized Experiments



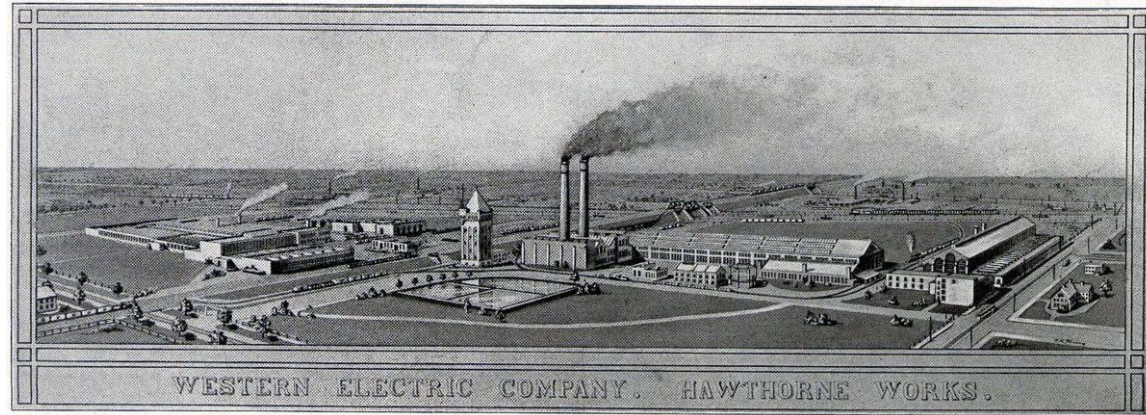
Causal conclusions



Still need to worry  
about sampling issues  
and threats to validity

## Threats to Validity

### Hawthorne Effect



[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

# Randomized Experiments



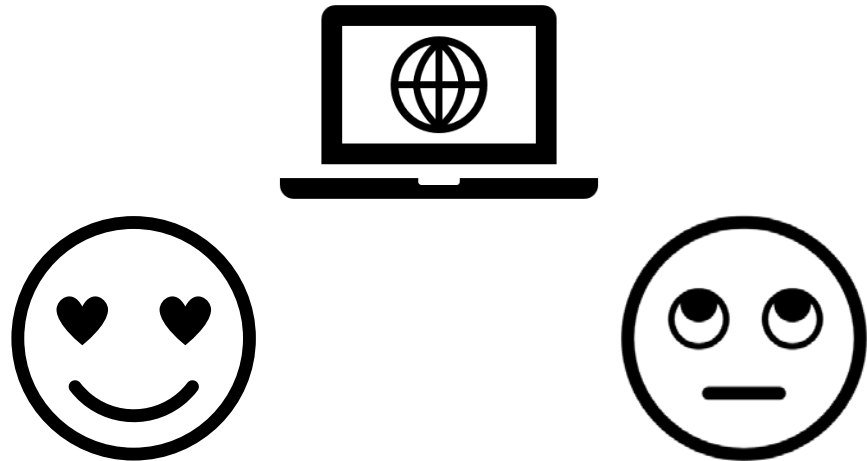
Causal conclusions



Still need to worry  
about sampling issues  
and threats to validity

## Threats to Validity

Novelty Effect



# Randomized Experiments



Causal conclusions

Threats to Validity

**Experimental Mortality**



Still need to worry  
about sampling issues  
and threats to validity

# Randomized Experiments



Causal conclusions

Works well with shorter studies

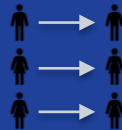


Still need to worry about sampling issues and threats to validity

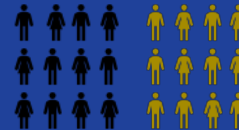
More challenging with longer studies and those not in laboratory-type settings



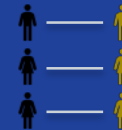
Randomized Experiments



Pre-Test / Post-Test



Quasi-Experiments

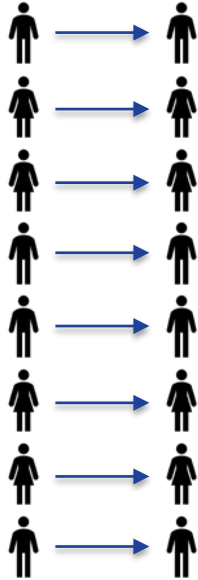


Propensity Scoring

# Measuring Intervention Impact

# Pre-Test / Post-Test

## Repeated Measures



### Pros

Greater statistical power

Smaller sample size

Logistically straight forward

### Cons

No control group

History

Maturation

Testing effect

# Pre-Test / Post-Test

## Solomon Four Group Design

Group #1	Pre-Test	Treatment	Post-Test
Group #2	Pre-Test		Post-Test
Group #3		Treatment	Post-Test
Group #4			Post-Test

### Pros

Control groups allow for comparisons

Can determine if there is a testing effect

### Cons

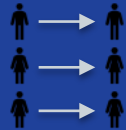
More complicated

Still need to worry about sampling issues and threats to validity

Randomization required for a causal conclusion



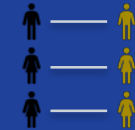
Randomized Experiments



Pre-Test / Post-Test



Quasi-Experiments



Propensity Scoring

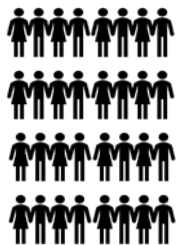
# Measuring Intervention Impact



# Quasi-Experimental Design

## Lacks Randomization

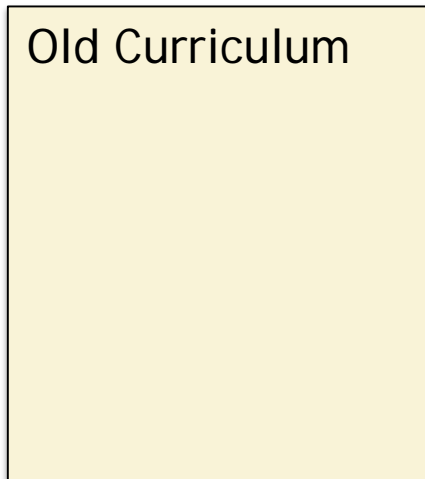
§ 001



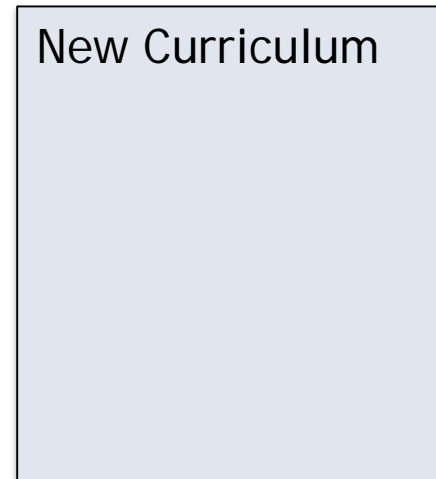
§ 002



Old Curriculum



New Curriculum



# Quasi-Experimental Design

## Pros

Logistically easy to implement

Consistent with real-life:  
“Ecological validity”

Can avoid potential ethical  
issues

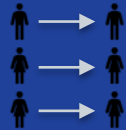
## Cons

Still need to worry about  
sampling issues and threats to  
validity

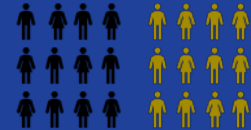
Cannot make casual  
conclusions



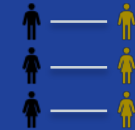
Randomized Experiments



Pre-Test / Post-Test



Quasi-Experiments




Propensity Scoring


# Measuring Intervention Impact

# Propensity Score Analysis

Research question: Does using our tutoring service improve final exam scores?



In my class, the better performing students are the ones who use the tutoring services.



Traditional-aged students probably have more time than adult learners to use the tutoring services and they usually do better.

Used Tutoring



$\bar{x} = 92$

Declined



$\bar{x} = 88$

# Propensity Score Analysis

**Propensity score:** Probability of receiving a treatment given baseline characteristics

## Covariates

GPA	3.5 GPA
Age	24 years old
Employment status	Work part-time
Gender	Woman
Credits completed	32 credits completed
Credits currently enrolled	Enrolled in 12 credits
Online courses completed	4 online courses completed

Used Tutoring



Declined



Propensity score = 0.75

# Propensity Score Matching

Used Tutoring



Declined



# Propensity Score Analysis



Observational data can be analyzed in a way that “mimics” a randomized experiment (Austin, 2011)

Stronger conclusions

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observation studies. *Multivariate Behavioral Research*, 46(3), 399-424.



Requires data from more variables

Some baseline similarities are necessary

Does not guarantee comparable groups

More complicated



# Practical Methods for Designing Research Studies and Analyzing Data in Educational Settings

Whitney Zimmerman

waz107@psu.edu

<https://www.linkedin.com/in/waz107/>