Comparison of Learning Outcomes for Simulation-Based and Traditional Curricula for Statistical Inference in a Designed Educational Experiment

Karsten Maurer

Iowa State University

June 4, 2015

Presentation Outline

- Statistical-Inference Curricula
- Experimental Design
- Data Description
- Model Based Analysis
  - Model Selection
  - Results
  - Model Assessment
- Discussion and Conclusions

Curriculum Study Introduction

Study Goal: Compare student learning outcomes for under two introductory curricula for statistical inference in a designed educational experiment

Two Curricula for Statistical Inference

- Traditional statistical inference curricula
- Simulation-based statistical inference curricula

Statistical Inference Curricula

Traditional Curriculum for Statistical Inference

- Result of Stat Ed reform movement of 1990's
- Summarized in GAISE College Report (2005)
- Statistical inference through theory-based approach
  - Normal model, CLT, formulas and tables
- Textbooks Examples
  - Introduction to the Practice Statistics (Moore, 2005)
  - Statistics: The Art and Science of Learning From Data (Agresti & Franklin, 2012)

Statistical Inference Curricula

Simulation-Based Curriculum for Statistical Inference

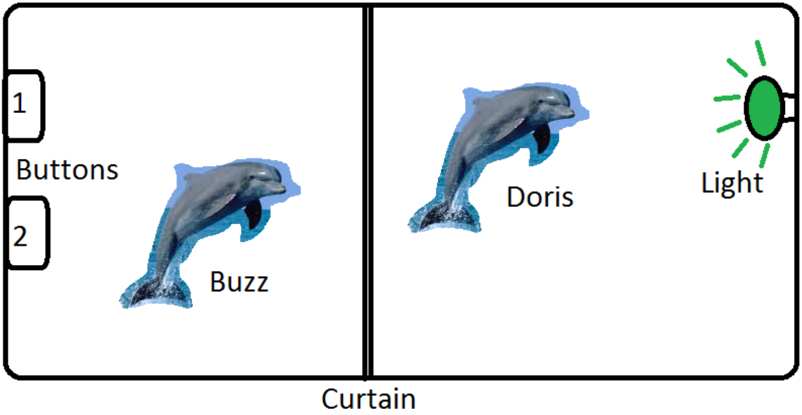
- Builds from Traditional curriculum for statistical inference
- Simulation/randomization to teach core concepts of inference
- Statistical inference introduced through simulation-based methods
    - Bootstrap confidence intervals
    - Simulation/randomization-based hypothesis tests
- Textbooks Examples
    - Statistics: Unlocking the Power of Data (Lock, Lock, Morgan, Lock & Lock, 2012)
    - Introduction to Statistical Investigations (Tintle, Chance, Cobb, Rossman, Roy, Swanson & VanderStoep, 2014)
- May elect to additionally teach theory-based methods

Simulation-Based Statistical Inference Classroom Example

Can Dolphins Communicate?

- Example from workshop section led by Alan Rossman
- Well known study from 1960s investigated if dolphins could communicate abstract ideas
- 2 trained dolphins: Buzz and Doris
- Trained:
    - Press button 1 when light blinking
    - Press button 2 when light steady.
    - Correct $\rightarrow$ get fish!

Simulation-Based Statistical Inference Classroom Example



Simulation-Based Statistical Inference Classroom Example

- Buzz pressed right button 15/16 times.
- Test H_o: Buzz guessing
- Simulate under H_o to make randomization distribution
- Using coins first then `Statkey` software
- Compare data to randomization distribution

Experimental Design: Subjects

Conducted within Stat 104: Introduction to Statistics

- Agricultural and biological science students
- (2 hour lecture + 2 hour lab) / week
- 16 week course

116 Students enrolled in two sections of Stat 104

- 4 Students dropped by week 2 of course
- 11 Students did not consent to data release
- 101 Students consented to data release

Experimental Design: Subjects

Conducted within Stat 104: Introduction to Statistics

- Agricultural and biological science students
- (2 hour lecture + 2 hour lab) / week
- 16 week course

116 Students enrolled in two sections of Stat 104

- 4 Students dropped by week 2 of course
- 11 Students did not consent to data release
- 101 Students consented to data release

Experimental Design: Random Assignment

Each student randomly assigned to inference curriculum

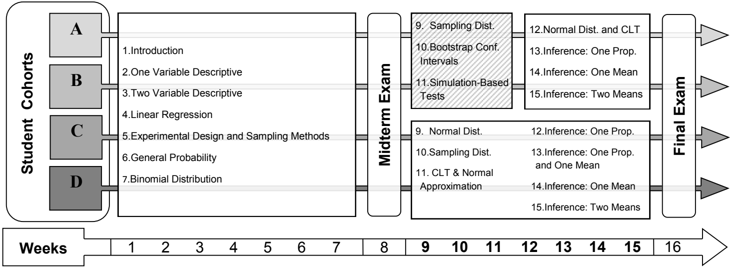
Resulted in four cohorts

- Cohort A: Section 1 / Simulation-Based
- Cohort B: Section 2 / Simulation-Based
- Cohort C: Section 1 / Traditional
- Cohort D: Section 2 / Traditional

101 students consented to release of course data for study

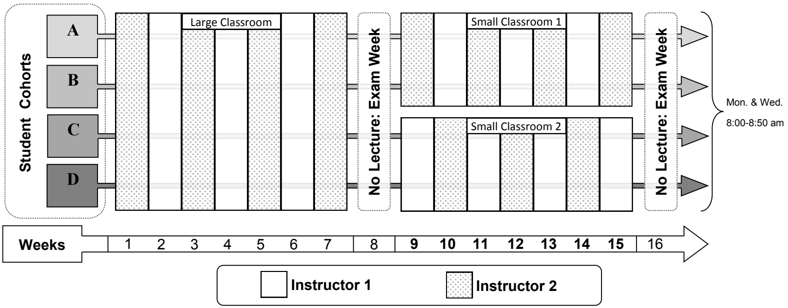
- 50 students from Traditional Curriculum
- 51 students from Simulation-Based Curriculum

Experimental Design: Treatments



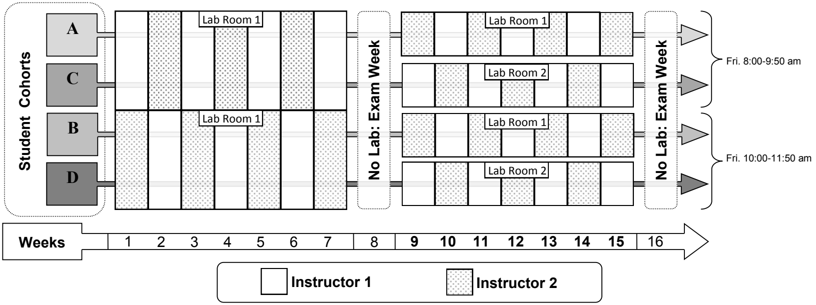
Curricula Schedules

Experimental Design: Treatment Administration



Lecture: Instructor and Room Schedule

Experimental Design: Treatment Administration



Lab: Instructor and Room Schedule

Data Description

Course Administration

- Enrollment Section
- Statistical Inference Curriculum

Pre-Treatment Measures

- Homework 1-7 Scores
- Lab 1-7 Scores
- Midterm Exam Score

Learning Outcomes: ARTIST scaled question sets

- Hypothesis Testing (HT) score out of 10
- Confidence Intervals (CI) score out of 10

Data Description

Course Administration

- Enrollment Section
- Statistical Inference Curriculum

Pre-Treatment Measures

- Homework 1-7 Scores
- Lab 1-7 Scores
- Midterm Exam Score

Learning Outcomes: ARTIST scaled question sets

- Hypothesis Testing (HT) score out of 10
- Confidence Intervals (CI) score out of 10

Data Description

Course Administration

- Enrollment Section
- Statistical Inference Curriculum

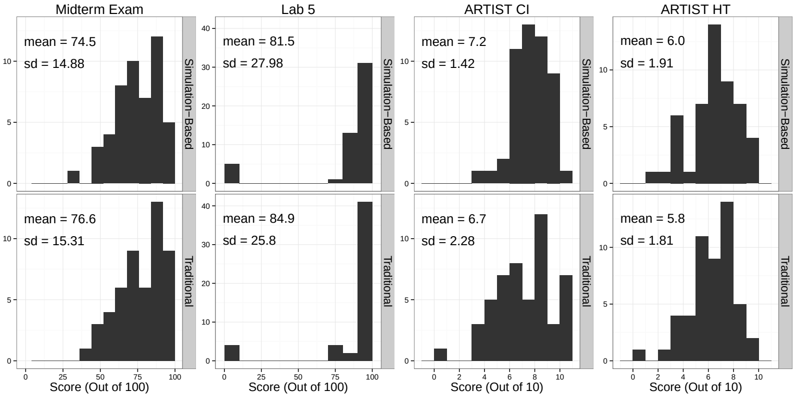
Pre-Treatment Measures

- Homework 1-7 Scores
- Lab 1-7 Scores
- Midterm Exam Score

Learning Outcomes: ARTIST scaled question sets

- Hypothesis Testing (HT) score out of 10
- Confidence Intervals (CI) score out of 10

Data Description



Histograms and summary statistics of scores by curriculum group

Model

Modeling approach based on data characteristics

- Bivariate responses: ARTIST scores for CI and HT
- Binary curricula indicator
- Continuous covariates: pre-treatment scores

Multivariate Analysis of Covariance (MANCOVA) Model

Model

Modeling approach based on data characteristics

- Bivariate responses: ARTIST scores for CI and HT
- Binary curricula indicator
- Continuous covariates: pre-treatment scores

Multivariate Analysis of Covariance (MANCOVA) Model

Model

Bivariate MANCOVA Model Structure

$$y_{ik} = \tau_k \mathbb{1}_{\{i \in T\}} + \beta_{0k} + \sum_{p=1}^{P} x_{ip}\beta_{pk} + \epsilon_{ik}, \qquad (1)$$

y_{ik}	k^{th} response ($k \in \{1, 2\}$) from student i, $1 \le i \le n$,
τ_k	treatment effect on response k, and
$\mathbb{1}_{\{i \in T\}}$	indicator function student i in Simulation-Based group
β_{0k}	common intercept for response k, and
β_{kp},	$1 \le p \le P$ model coefficients of P covariates.
x_{ip}	p^{th} pre-treatment covariate score of student i

Model

Bivariate MANCOVA Model Error Structure

$$\vec{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix} \overset{iid}{\sim} \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma^2_{11} & \sigma^2_{12} \\ \sigma^2_{21} & \sigma^2_{22} \end{bmatrix} \right)$$

Model Selection

Model Selection Process

- Bivariate MANCOVA model fit initially with all covariates
- Backward stepwise selection process based on AIC
- Additional covariates removed due to collinearity

Selected Parameters for ARTIST Model

- Midterm exam score
- Lab 5 score
- Curriculum treatment effect

Model Results: Bivariate Effects

	Pillai's Λ	Approx. F	Pr(>|F|)
Midterm	0.2109	12.8277	0.0000
Lab 5	0.0792	4.1261	0.0191
Treatment	0.0469	2.3605	0.0998

Pillai's Λ tests for overall covariate effects

Model Results: Univariate Effects

	Covariate Values	Estimate	95% Confidence Interval
Intercept	1	2.1053	(0.0046 , 4.2060)
Midterm	{0,1,...100}	0.0386	(0.0152 , 0.0620)
Lab 5	{0,1,...100}	0.0085	(-0.0046 , 0.0217)
Treatment	{0,1}	0.3050	(-0.3960 , 1.0059)

ARTIST Model coefficients for hypothesis testing topic score

Model Results: Univariate Effects

	Covariate Values	Estimate	95% Confidence Interval
Intercept	1	1.4648	(-0.5467 , 3.4763)
Midterm	{0,1,...100}	0.0477	(0.0253 , 0.0701)
Lab 5	{0,1,...100}	0.0183	(0.0057 , 0.0309)
Treatment	{0,1}	0.7146	(0.0435 , 1.3858)

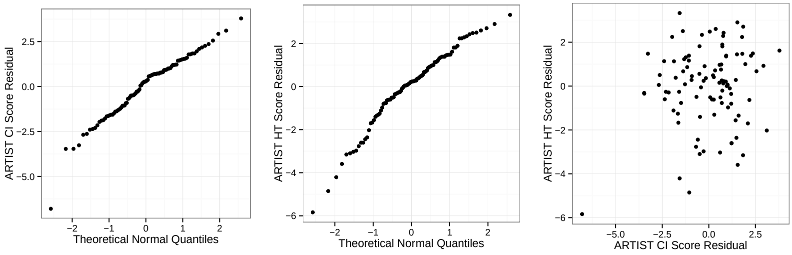
ARTIST Model coefficients for confidence interval topic score

Model Assessment

Model Assumptions Needing Assessment

- Normality of errors
- Linearity
- Constant variance
- Independence between students

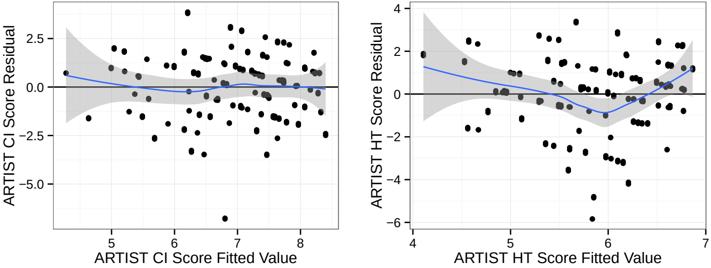
Model Assessment



Normal quantile plots (left & center)

Bivariate residuals scatterplot (right)

Model Assessment



ARTIST Model residual plots overlaid with Loess smoother

Model Assessment

Recall:

$$\vec{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix} \overset{iid}{\sim} \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix} \right)$$

Assumption of Independence

- Student score pairs are related: $\text{Cov}[\epsilon_{i1}, \epsilon_{i2}] = \sigma_{12}^2$
- Scores between students independent: $\text{Cov}[\epsilon_{ik}, \epsilon_{j\ell}] = 0, \; \forall i \neq j$

Impact of Violation of Independence

- Type I error rate when testing for curriculum treatment effect
- Bias in parameter estimation

Model Assessment

Recall:

$$\vec{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix} \overset{iid}{\sim} \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix} \right)$$

Assumption of Independence

- Student score pairs are related: $\text{Cov}[\epsilon_{i1}, \epsilon_{i2}] = \sigma_{12}^2$
- Scores between students independent: $\text{Cov}[\epsilon_{ik}, \epsilon_{j\ell}] = 0$, $\forall i \neq j$

Impact of Violation of Independence

- Type I error rate when testing for curriculum treatment effect
- Bias in parameter estimation

Model Assessment

Simulation Study

- Simulate responses under independence violation
- Assume MANCOVA model as generative model
  - Set coefficient values, β_{pk} for intercept, midterm and lab 5 based estimates from reduced MANCOVA
  - No curriculum treatment effect (i.e. $\tau_1 = \tau_2 = 0$)

Simulation Study

Assumption:

$$\text{Cov}[\epsilon_{ik}, \epsilon_{j\ell}] = 0, \quad \text{for all } i \neq j$$

Independence Violation:

- Stage 1:

$$\text{Cov}[\epsilon_{ik}, \epsilon_{j\ell}] = \eta\sigma^2_{k\ell}, \quad \text{where } \eta \in [0,1] \text{ for all } i \neq j, \text{ and } k, \ell \in \{1,2\}$$

- Stage 2:

$$\text{Cov}[\epsilon_{ik}, \epsilon_{j\ell}] = c\eta\sigma^2_{k\ell}, \quad \text{where } \eta \in [0,1] \text{ for all } i \neq j, \text{ and } k, \ell \in \{1,2\}$$

$$c = 1 \text{ for all i,j classmates}$$
$$c \in [0,1] \text{ for all i,j non-classmates}$$

Simulation Study

No Violation of Between Student Independence

$$
\begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{n2} \end{bmatrix} \sim N\left(\vec{0}, \left[\begin{array}{ccccc|ccccc}
\sigma^2_{11} & 0 & \cdots & 0 & 0 & \sigma^2_{12} & 0 & \cdots & 0 & 0 \\
0 & \sigma^2_{11} & \ddots & 0 & 0 & 0 & \sigma^2_{12} & \ddots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \ddots & \sigma^2_{11} & 0 & 0 & 0 & \ddots & \sigma^2_{12} & 0 \\
0 & 0 & \cdots & 0 & \sigma^2_{11} & 0 & 0 & \cdots & 0 & \sigma^2_{12} \\
\hline
\sigma^2_{12} & 0 & \cdots & 0 & 0 & \sigma^2_{22} & 0 & \cdots & 0 & 0 \\
0 & \sigma^2_{12} & \ddots & 0 & 0 & 0 & \sigma^2_{22} & \ddots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \ddots & \sigma^2_{12} & 0 & 0 & 0 & \ddots & \sigma^2_{22} & 0 \\
0 & 0 & \cdots & 0 & \sigma^2_{12} & 0 & 0 & \cdots & 0 & \sigma^2_{22}
\end{array} \right] \right)
$$

Simulation Study

Stage 1 Violation of Between Student Independence

$$\begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{n2} \end{bmatrix} \sim \mathrm{N}\left(\vec{0}, \left[\begin{array}{ccccc|ccccc}
\sigma^2_{11} & \eta\sigma^2_{11} & \cdots & \eta\sigma^2_{11} & \eta\sigma^2_{11} & \sigma^2_{12} & \eta\sigma^2_{12} & \cdots & \eta\sigma^2_{12} & \eta\sigma^2_{12} \\
\eta\sigma^2_{11} & \sigma^2_{11} & \ddots & \eta\sigma^2_{11} & \eta\sigma^2_{11} & \eta\sigma^2_{12} & \sigma^2_{12} & \ddots & \eta\sigma^2_{12} & \eta\sigma^2_{12} \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\eta\sigma^2_{11} & \eta\sigma^2_{11} & \ddots & \sigma^2_{11} & \eta\sigma^2_{11} & \eta\sigma^2_{12} & \eta\sigma^2_{12} & \ddots & \sigma^2_{12} & \eta\sigma^2_{12} \\
\eta\sigma^2_{11} & \eta\sigma^2_{11} & \cdots & \eta\sigma^2_{11} & \sigma^2_{11} & \eta\sigma^2_{12} & \eta\sigma^2_{12} & \cdots & \eta\sigma^2_{12} & \sigma^2_{12} \\
\hline
\sigma^2_{12} & \eta\sigma^2_{12} & \cdots & \eta\sigma^2_{12} & \eta\sigma^2_{12} & \sigma^2_{22} & \eta\sigma^2_{22} & \cdots & \eta\sigma^2_{22} & \eta\sigma^2_{22} \\
\eta\sigma^2_{12} & \sigma^2_{12} & \ddots & \eta\sigma^2_{12} & \eta\sigma^2_{12} & \eta\sigma^2_{22} & \sigma^2_{22} & \ddots & \eta\sigma^2_{22} & \eta\sigma^2_{22} \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\eta\sigma^2_{12} & \eta\sigma^2_{12} & \ddots & \sigma^2_{12} & \eta\sigma^2_{12} & \eta\sigma^2_{22} & \eta\sigma^2_{22} & \ddots & \sigma^2_{22} & \eta\sigma^2_{22} \\
\eta\sigma^2_{12} & \eta\sigma^2_{12} & \cdots & \eta\sigma^2_{12} & \sigma^2_{12} & \eta\sigma^2_{22} & \eta\sigma^2_{22} & \cdots & \eta\sigma^2_{22} & \sigma^2_{22}
\end{array} \right] \right)$$

Simulation Study

Stage 2 Violation of Between Student Independence

$$
\begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{n2} \end{bmatrix} \sim \mathrm{N} \left(\vec{0}, \left[\begin{array}{ccccc|ccccc}
\sigma^2_{11} & \eta\sigma^2_{11} & \cdots & c\eta\sigma^2_{11} & c\eta\sigma^2_{11} & \sigma^2_{12} & \eta\sigma^2_{12} & \cdots & c\eta\sigma^2_{12} & c\eta\sigma^2_{12} \\
\eta\sigma^2_{11} & \sigma^2_{11} & \ddots & c\eta\sigma^2_{11} & c\eta\sigma^2_{11} & \eta\sigma^2_{12} & \sigma^2_{12} & \ddots & c\eta\sigma^2_{12} & c\eta\sigma^2_{12} \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
c\eta\sigma^2_{11} & c\eta\sigma^2_{11} & \ddots & \sigma^2_{11} & \eta\sigma^2_{11} & c\eta\sigma^2_{12} & c\eta\sigma^2_{12} & \ddots & \sigma^2_{12} & \eta\sigma^2_{12} \\
c\eta\sigma^2_{11} & c\eta\sigma^2_{11} & \cdots & \eta\sigma^2_{11} & \sigma^2_{11} & c\eta\sigma^2_{12} & c\eta\sigma^2_{12} & \cdots & \eta\sigma^2_{12} & \sigma^2_{12} \\
\hline
\sigma^2_{12} & \eta\sigma^2_{12} & \cdots & c\eta\sigma^2_{12} & c\eta\sigma^2_{12} & \sigma^2_{22} & \eta\sigma^2_{22} & \cdots & c\eta\sigma^2_{22} & c\eta\sigma^2_{22} \\
\eta\sigma^2_{12} & \sigma^2_{12} & \ddots & c\eta\sigma^2_{12} & c\eta\sigma^2_{12} & \eta\sigma^2_{22} & \sigma^2_{22} & \ddots & c\eta\sigma^2_{22} & c\eta\sigma^2_{22} \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
c\eta\sigma^2_{12} & c\eta\sigma^2_{12} & \ddots & \sigma^2_{12} & \eta\sigma^2_{12} & c\eta\sigma^2_{22} & c\eta\sigma^2_{22} & \ddots & \sigma^2_{22} & \eta\sigma^2_{22} \\
c\eta\sigma^2_{12} & c\eta\sigma^2_{12} & \cdots & \eta\sigma^2_{12} & \sigma^2_{12} & c\eta\sigma^2_{22} & c\eta\sigma^2_{22} & \cdots & \eta\sigma^2_{22} & \sigma^2_{22}
\end{array} \right] \right)
$$

Simulation Study

Simulation Process: Repeat M times

1. Simulate an error vector, $\vec{\epsilon}^{\,(m)}$, under violated error structure.
2. Derive responses:

$$\begin{bmatrix} \vec{y}_1^{\,(m)} \\ \vec{y}_2^{\,(m)} \end{bmatrix} = \begin{bmatrix} X\vec{\beta}_1 \\ X\vec{\beta}_2 \end{bmatrix} + \vec{\epsilon}^{\,(m)}$$

3. Randomly permute treatment label to break association with covariates
4. Fit full ARTIST Model (assuming independence)
5. Record estimates, standard errors, etc.

Simulation Study

Stage 1:

- Repeat $M = 25,000$ simulations at each η in $\{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$

Stage 2:

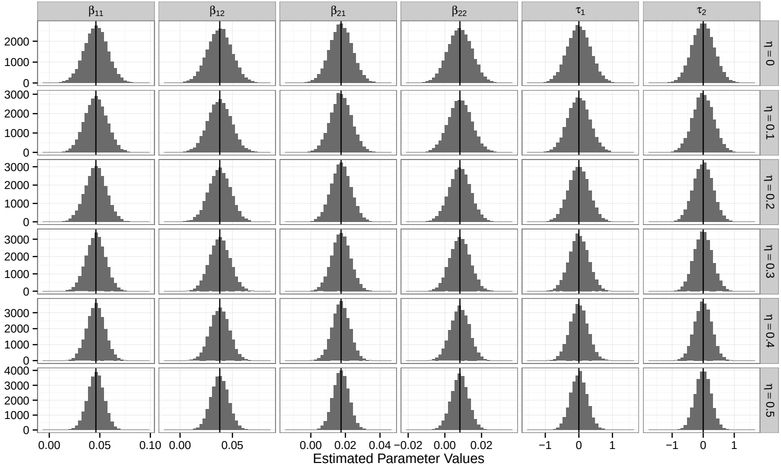
- Repeat $M = 10,000$ simulations at each combination of η in $\{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ and c in $\{0.9, 0.925, 0.95, 0.975, 1\}$

Simulation Study: Stage 1 Results

η	Pillai's Test $H_o: \tau_1 = \tau_2 = 0$	t-test (CI) $H_o: \tau_1 = 0$	t-test (HT) $H_o: \tau_2 = 0$
0.0	0.0499	0.0491	0.0494
0.1	0.0496	0.0522	0.0514
0.2	0.0491	0.0491	0.0468
0.3	0.0496	0.0500	0.0486
0.4	0.0517	0.0505	0.0500
0.5	0.0494	0.0514	0.0489

Type I error rates for tests for curriculum treatment effect

Simulation Study: Stage 1 Results



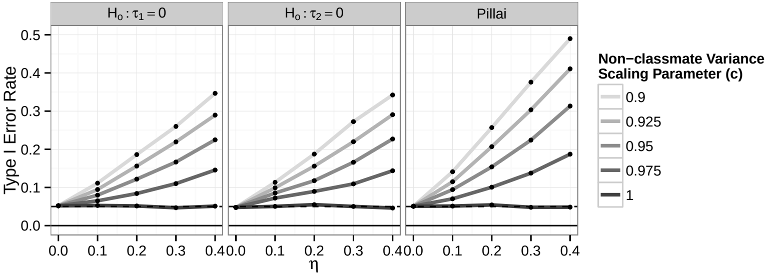
Parameter estimates from 25,000 simulations at each η.

Simulation Study: Stage 1 Results

η	s_{τ_1}	$\hat{\text{SE}}[\tau_1]$	s_{τ_2}	$\hat{\text{SE}}[\tau_2]$
0.0	0.3486	0.3484	0.3486	0.3576
0.1	0.3309	0.3326	0.3309	0.3395
0.2	0.3120	0.3128	0.3120	0.3169
0.3	0.2918	0.2918	0.2918	0.2987
0.4	0.2704	0.2716	0.2704	0.2777
0.5	0.2465	0.2475	0.2465	0.2537

Standard deviation for simulated treatment effect estimates, s_{τ_k}, and average standard errors for treatment effects, $\hat{\text{SE}}[\tau_k]$

Simulation Study: Stage 2 Results



Type I error rates for tests of treatment effects

Simulation Study: Stage 2 Results

Type I error rates increase as c increases

What does c represent?

- covariance disparity attributable to *non-treatment* factors
- generative model must carry no curriculum treatment effect

Cannot estimate c due to classroom confounding with treatment

Steps taken to minimize non-treatment differences in classroom

- Alternation of instruction
- Identical curricula weeks 1 to 8
- Physical classrooms similar

Discussion and Conclusions

Study Results:

- Simulation-based group scored significantly higher on CI
  - After controlling for midterm and lab 5 scores
  - 7% higher on ARTIST scale
- No significant difference in HT scores

Discussion Issues:

- Criticisms of ARTIST sets
- Representation
- Causality
- Treatment Complexity

Discussion and Conclusions

Study Results:

- Simulation-based group scored significantly higher on CI
  - After controlling for midterm and lab 5 scores
  - 7% higher on ARTIST scale
- No significant difference in HT scores

Discussion Issues:

- Criticisms of ARTIST sets
- Representation
- Causality
- Treatment Complexity

Acknowledgements

Co-investigator: Dennis Lock

Study Advisors: Dr. Heike Hofmann and Dr. Bob Stephenson

Thanks!

Thank you for attending!

Any questions?