Random is Random: Helping Students Distinguish Between Random Sampling and Random Assignment

Robert delMas
Elizabeth Brondos Fry
Department of Educational Psychology

Background and Motivation

According to statistics education recommendations (e.g., GAISE, 2016), students should understand the following about the role of randomness in study design:

- **Random sampling** tends to produce representative samples, allowing for **generalization** to a population.
- **Random assignment** tends to balance out confounding variables between groups, helping to enable **cause-and-effect** conclusions.

Background and Motivation

Some difficulties have been documented understanding these topics (e.g., Derry et al., 2000; Sawilowsky, 2004; Wagler & Wagler, 2013), such as:

- Confusion between random sampling and random assignment
- Disbelief that random assignment can help enable causal claims
- Believe larger samples are always better than smaller samples (regardless of method – i.e., biased sample)
- Believe unequal sample sizes do not allow for any conclusions

Research Question

A study design unit was created and implemented to answer the research question:

How does introductory statistics students' conceptual understanding of study design and conclusions change after participating in a learning unit designed to promote conceptual change in these areas?

Course and Audience

- Undergraduate, 3-credit introductory statistics course that fulfills general education mathematical thinking requirement at a large, Research 1 university (using CATALST curriculum; Garfield et al., 2012; Ziefﬂer et al., 2015)
    - Four sections: Three in-class and one online (30-45 students each)
    - Taught by advanced graduate students in statistics education
    - Engaged students in active learning and discovery, minimal lecture
- Study design unit lasted 2½ weeks, during second half of spring 2016 semester

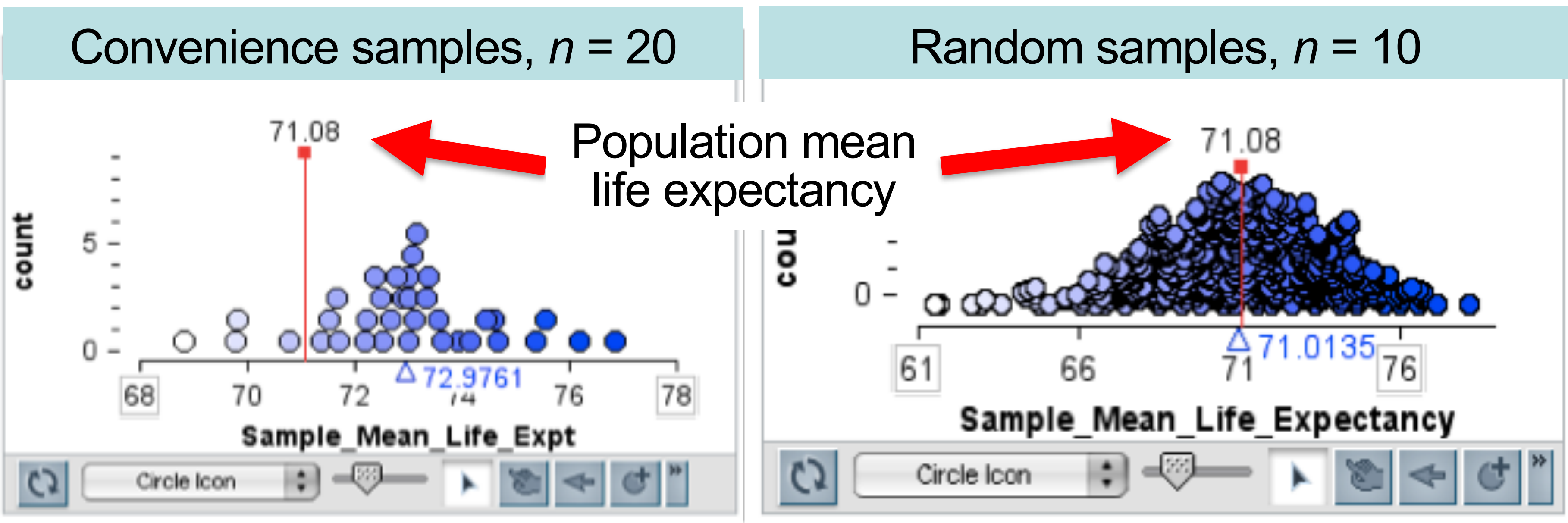
2½ Week Study Design Unit

Day	Topic	Activity name	Reading prior to activity
1	Sampling methods and unbiased estimation	*Sampling Countries*	None
2	Assignment to experimental groups and establishing causation	*Strength Shoe*	*Establishing Causation*
3	Observational studies	*Murderous Nurse*	*Scope of Inferences*
4	Study design and scope of inference	Group quiz	None
5	Distinguishing between random sampling/generalization and random assignment/causation	*Survey Incentives*	None

Day 1: Sampling Methods and Unbiased Estimation

Activity: "Sampling Countries"

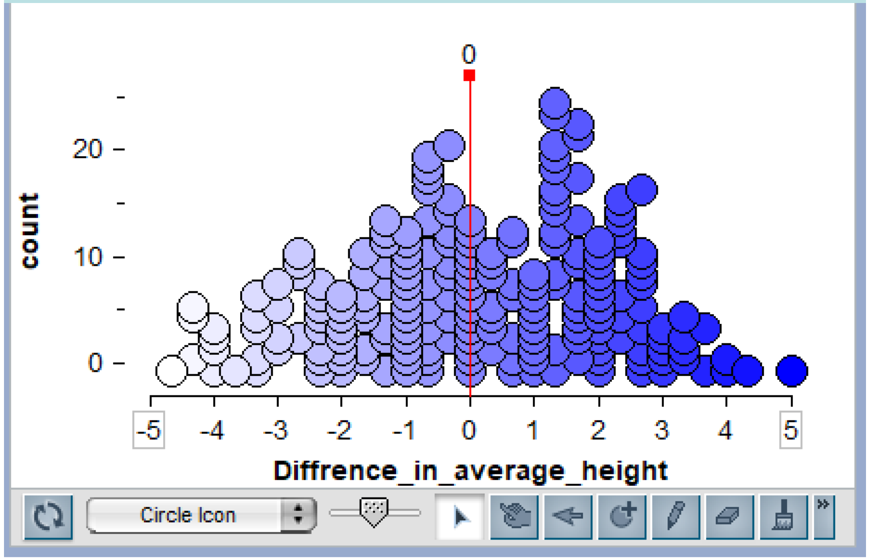
Students contrasted central tendency of average life expectancy from convenience samples (n = 20) with simple random samples (n = 10) of countries



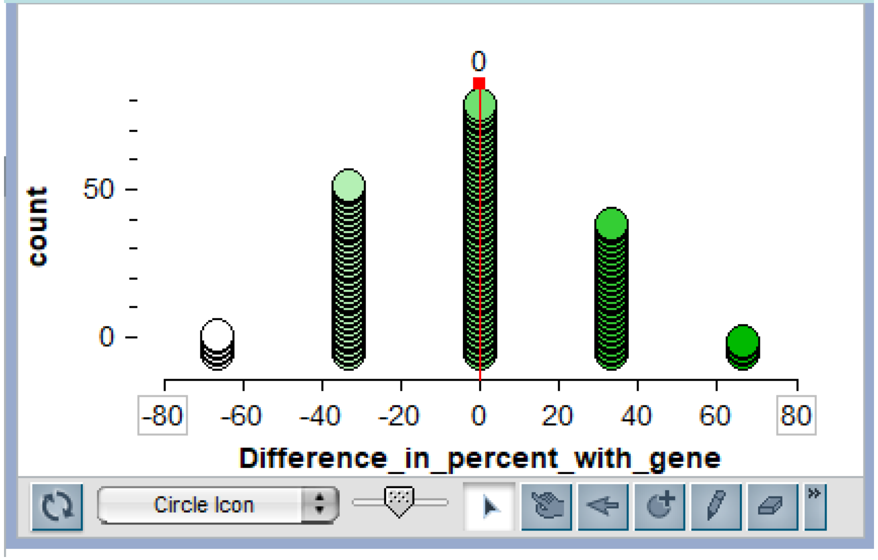
Day 2: Assignment to Experimental Groups and Establishing Causation

Activity: "Strength Shoe" (modified from Zieffler et al., 2015)
Students simulated random assignment to two groups and observed the distribution of group mean differences on several potential confounding variables

"Known" confounder: Height



"Unknown" confounder: Gene



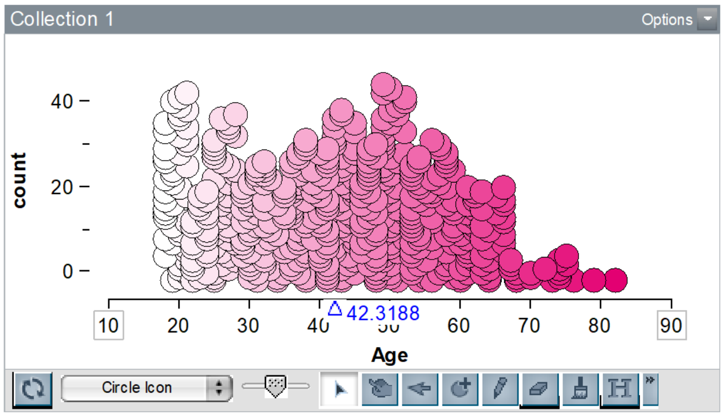
Day 5: Survey incentive - Context

- Mayor of a town wants to conduct a pilot study to see if giving a $20 incentive to complete a survey will increase response rates.
- Student is asked to play "statistical consultant" and conduct both random sampling and random assignment. Has to explain to the mayor the difference between random sampling and random assignment.

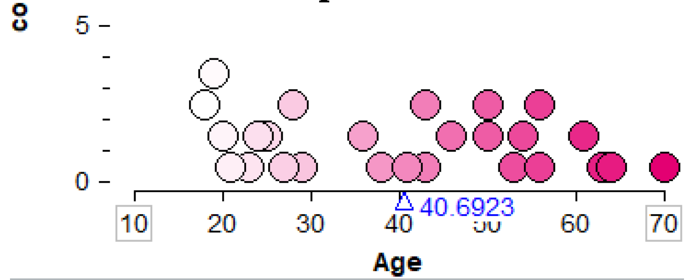
Survey Incentives: Part 1

- Students compare distribution of random samples to population
- Observe that when many samples are taken, sample means are centered at population mean

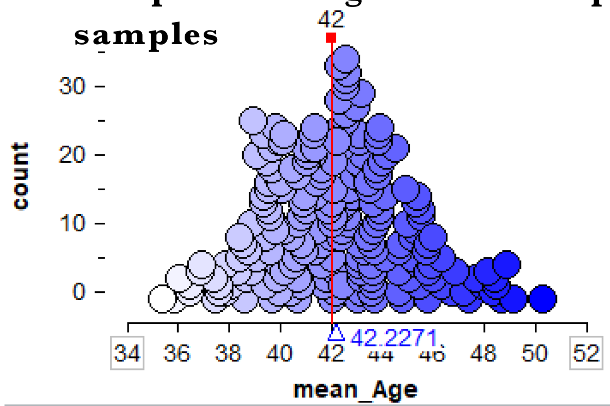
Age: Population distribution
Population Average = 42 years



Age distribution for one simple random sample

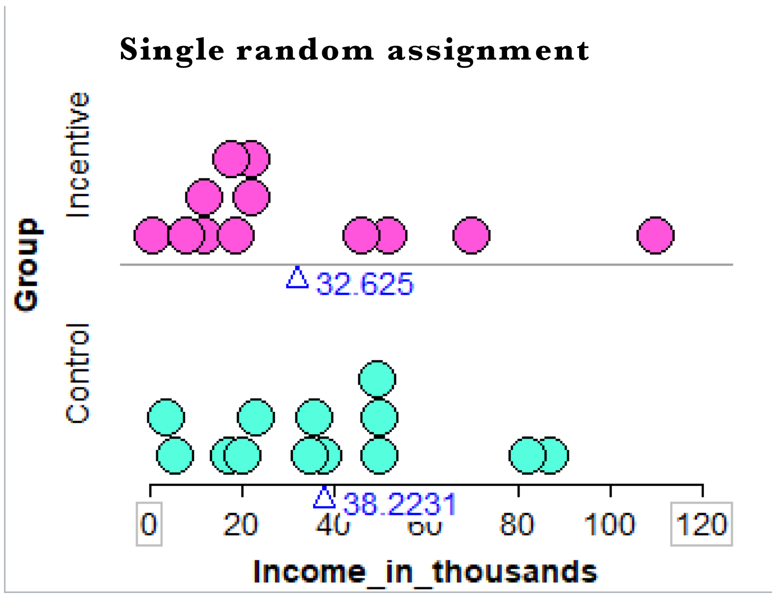


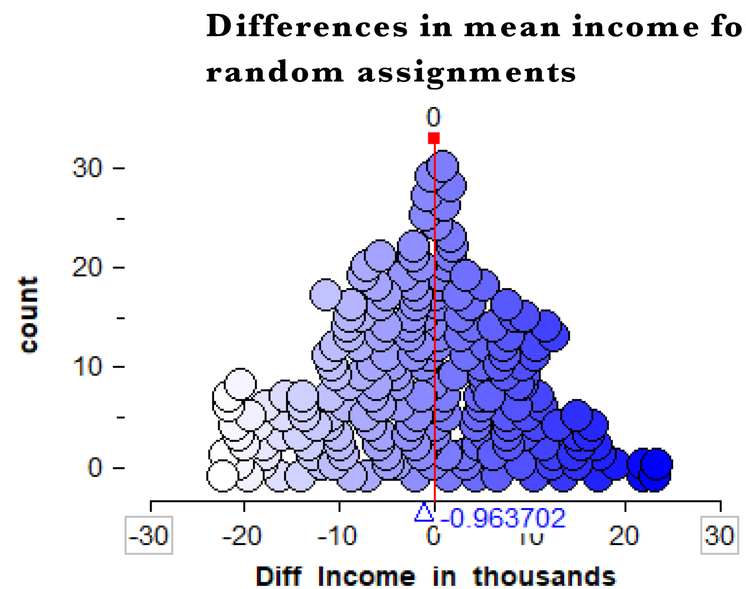
Sample mean age for 200 simple random samples



Survey Incentives: Part 2

- Students compare groups within a random assignment
- Observe that a single random assignment produces *similar* groups (but not identical) and group mean differences tend to balance out across many random assignments.





Assessment

Inferences from Design Assessment (IDEA)

- 22-item, forced-choice assessment, pretest & posttest completed by $n = 125$ students
  - 9 items on sampling/generalization (Sampling subscore)
  - 13 items on assignment/causation (Assignment subscore)
- Most items taken or modified from previous assessments (e.g., CAOS, delMas et al., 2007; ARTIST, Garfield et al., 2002)

Results: Overview

- IDEA changes in total score
- Items with high performance (pretest & posttest)
- Items with significant improvement
  - Comparisons to prior studies
- Item distractors related to misunderstandings

IDEA Changes in Score

Results from paired *t*-tests of differences in IDEA score (posttest-pretest) for $n = 125$ students

	Mean Diff.	SD	t	p	Cohen's d
Difference in total score (22 items)	3.30	2.94	12.57	<.001	1.12
Difference in sampling subscore (9 items)	1.75	1.79	10.97	<.001	0.98
Difference in assignment subscore (13 items)	1.55	1.87	9.29	<.001	0.83

High Performance Items

Students performed very well (88% or more correct on both pretest and posttest) on nine items related to learning outcomes such as…

- Identifying the sample
- Determining what type of study was conducted (observational or experimental)
- Understanding that random assignment is ideal for answering research questions about causation
- Distinguishing between statements that make causal claims and statements that make association-only claims

Possible explanation: Study design unit occurred during second half of semester, after students had worked with data from samples, and had learned about comparing treatment and control groups in experiments

Items with Most Improvement

After adjusting for multiple comparisons, $\alpha_c = .002$

Item	Measured learning outcome: Ability to understand...	Percent correct (n = 125)		McNemar's test p
		Pretest	Posttest	
16	Correlation does not imply causation	28.0	77.6	<.0001
18	Purpose of random assignment in an experiment	32.0	77.6	<.0001
3	What it means to make an appropriate generalization to a population, using sample data	23.2	63.2	<.0001
6	Small random sample is preferable to a larger sample gathered with a biased sampling method.	46.4	85.6	<.0001

Comparisons to Prior Studies

- The two IDEA items that showed the most improvement were slightly modified from items on the Comprehensive Assessment of Outcomes in Statistics (CAOS; delMas et al., 2007).
- Performance on IDEA was compared with performance in similar CAOS items from:
  - A national sample of 13,432 undergraduate introductory statistics students enrolled in U.S. universities (2005-2017)
  - Two samples of introductory statistics students at a small college (Tintle et al., 2012):
    - Randomization-based curriculum ($n = 76$)
    - Consensus curriculum ($n = 78$)

Comparisons to Prior Studies

Measured learning outcome: Understanding that correlation does not imply causation

Sample	Pretest % correct	Posttest % correct
National sample n = 13,432	50.3	57.1
Tintle et al. (2012; n = 76) Randomization-based curriculum	47.4	59.2
Tintle et al. (2012; n = 78) Consensus curriculum	57.7	62.8
IDEA	28.0	77.6

Comparisons to Prior Studies

Measured learning outcome: Understanding the purpose of random assignment in an experiment

Sample	Pretest % correct	Posttest % correct
National sample n = 13,432	9.2	16.9
Tintle et al. (2012; n = 76) Randomization-based curriculum	1.3	18.4
Tintle et al. (2012; n = 78) Consensus curriculum	7.7	14.1
IDEA	32.0	77.6

Other Items with Statistically Significant Improvement

Item	Measured learning outcome Ability to….	Percent correct (n = 125) Pretest	Post-test	McNemar's test p
1	Identify population to which inferences can be made, based on a sample	40.8	65.6	<.0001
5	Understand when sample estimates may be biased due to lack of a representative sample	70.4	86.4	.0005
21	Understand that assigning subjects to treatments as they walk into a room does not help balance out confounding variables	60.7	79.5	.0006
22	Recognize when a randomized experiment should be used for a particular research question	79.8	91.9	.0015

Distractor Analysis

Some difficulties have been documented understanding these topics (e.g., Derry et al., 2000; Sawilowsky, 2004; Wagler & Wagler, 2013), such as:

- **Confusion between random sampling and random assignment**
- Disbelief that random assignment can help enable causal claims
- **Believing larger samples are always better than smaller samples (regardless of method)**
- Believing unequal sample sizes in two groups do not allow for any conclusions

Distractors: Confusing Random Sampling with Random Assignment

Item	Misconception or Misunderstanding	Percent (n = 125)		McNemar's test p
		Pretest	Posttest	
16	The sample was randomly *selected*, so causation can be inferred	24.8	12.0	.0090
18	Purpose of random assignment: To ensure participants are likely to be representative of the larger population	40.0	14.4	<.0001

There was a decrease in these two confusions.

Even so, more than 10% chose these options on the posttest.

Distractors: Confusing Random Sampling with Random Assignment

Item	Misconception or Misunderstanding	Percent (n = 125)		McNemar's test p
		Pretest	Posttest	
16	The sample was randomly *selected*, so causation can be inferred	24.8	12.0	.0090
18	Purpose of random assignment: To ensure participants are likely to be representative of the larger population	40.0	14.4	<.0001
9	Cannot generalize due to lack of random *assignment*	9.6	23.2	.0046

There was an increase in the confusion that random assignment is needed to generalize to a population.

Distractors: Confusing Random Sampling with Random Assignment

Item	Misconception or Misunderstanding	Percent (n = 125)		McNemar's test p
		Pretest	Posttest	
16	The sample was randomly *selected*, so causation can be inferred	24.8	12.0	.0090
18	Purpose of random assignment: To ensure participants are likely to be representative of the larger population	40.0	14.4	<.0001
9	Cannot generalize due to lack of random *assignment*	9.6	23.2	.0046

However, on posttest: Less than 10% of students chose two out of three of these incorrect options, and 0% of students chose all three.

Distractors: Sample Size Misunderstandings

Item	Misconception or Misunderstanding	Percent (n = 125)		McNemar's test p
		Pretest	Posttest	
6	Larger sample size more likely to provide unbiased estimate than smaller sample (despite biased sampling method for larger sample)	16.0	0.8	<.0001
9	n = 100 is too small to make generalization claim	29.6	6.4	<.0001
16	Sample size of 1,000 is too small to allow causation to be inferred	35.2	7.2	<.0001

Misunderstandings about sampling and sample size decreased, with less than 10% choosing the above options on the posttest.

Summary of Results

- Overall, evidence of learning gains in concepts of study design and conclusions
- A small, but noticeable portion of students experience difficulties such as:
  - Confusion between random sampling and random assignment
  - Giving sample size more importance than sampling method

Limitations

- No random sampling
  - Cannot generalize to all introductory statistics students
- No random assignment or comparison to other curricula
  - All course sections are taught with same curriculum
- Pretest given just before unit and posttest given just after unit: Did not measure student knowledge at beginning or end of course
- IDEA instrument limitations
  - Reliability as measured by coefficient Omega Total: 0.63 on pretest, 0.79 on posttest; lower for sampling and assignment subscales
  - Nine items (out of 22 total) with high pretest & posttest performance had little variation.

Implications for Teaching

- Random is Random, but not always for the same purpose – easy to conflate the purposes of randomization in study design.
  - Idea of "random" central to both sampling and assignment to groups, but role of randomness is different
  - "Bias" can refer to bias in sampling, or researcher bias in assigning groups

Implications for Teaching

- Students need to make connections between study design concepts and other statistical concepts:
  - Sample size: Larger samples are not always better, but still important to learn how sample size affects results from inference methods
  - Simulation: Some students stated the random re-allocation performed in a randomization test supported a causal claim

THANK YOU!

For more details, see Elizabeth Fry dissertation:

http://iase-web.org/Publications.php?p=Dissertations

Course materials:

http://z.umn.edu/studydesign

Item #16

Researchers conducted a survey of 1,000 randomly selected adults in the United States and found a strong, positive, statistically significant correlation between income and the number of containers the adults reported recycling in a typical week.

Can the researchers conclude that higher income causes more recycling among U.S. adults? Select the best answer from the following options.

a) No, the sample size is too small to allow causation to be inferred.
b) **No, the lack of random assignment does not allow causation to be inferred.**
c) Yes, the statistically significant result allows causation to be inferred.
d) **Yes, the sample was randomly selected, so causation can be inferred.**

Pretest	Posttest
35.2	7.2
28.0	**77.6**
12.0	3.2
24.8	**12.0**

Item #18

A research study randomly assigned participants into two groups. One group was given Vitamin E to take daily. The other group received only a placebo pill. The research study followed the participants for eight years. After the eight years, the proportion of each group that developed a particular type of cancer was compared.

What is the primary reason that the study used random assignment?

a) **To ensure that the groups are likely to be similar in all respects except for the level of Vitamin E.**
b) To ensure that a person is not likely to know whether or not they are getting the placebo.
c) **To ensure that the study participants are likely to be representative of the larger population.**

Pretest	Posttest
32.0	**77.6**
28.0	8.0
40.0	**14.4**