Teaching Introductory Statistics Students Sampling Distributions Using Computer Versus Tactile Simulations

Advisor: Stacey Hancock,
Montana State University
email: stacey.hancock@montana.edu

Wendy Rummerfield,
University of California, Irvine
email: wrummerf@uci.edu

Background

Sampling Distribution: Probability distribution of a statistic obtained through a large number of samples drawn from a specific population.
Problem: The concept of sampling distributions is very difficult for many introductory statistics students to comprehend. Students often are able to perform rote calculations, but are unable to grasp the underlying concepts and motivation behind such procedures and may be unable to apply their knowledge of statistics outside of the classroom.
Goal: Compare two popular methods for teaching sampling distributions - computer simulation methods (CSM's) alone, or a mix of tactile simulations followed by computer simulations - to discover which pedagogical method leads to enhanced understanding of this notoriously difficult topic.

Literature Review

The Problem:

- Disconnect between probability models and statistical inference [1] [2].
- Students often confuse distributions of samples with distributions of sample statistics and have difficulty understanding sampling variability [1].

Differing Views:

Pro Computer Simulations

- Statistical software packages and online applets allow students to quickly perform simulations and work with real data without advanced knowledge of statistics [2].
- Students get immediate feedback on their predictions and can confront their false conceptions.

Pro Tactile Simulations

- Since students are actively collecting the data themselves they cannot be passive observers and will better understand the underlying population [4]
- Students are more likely to believe their simulations reflect reality and can overcome their misconceptions [4].

The Controversy:

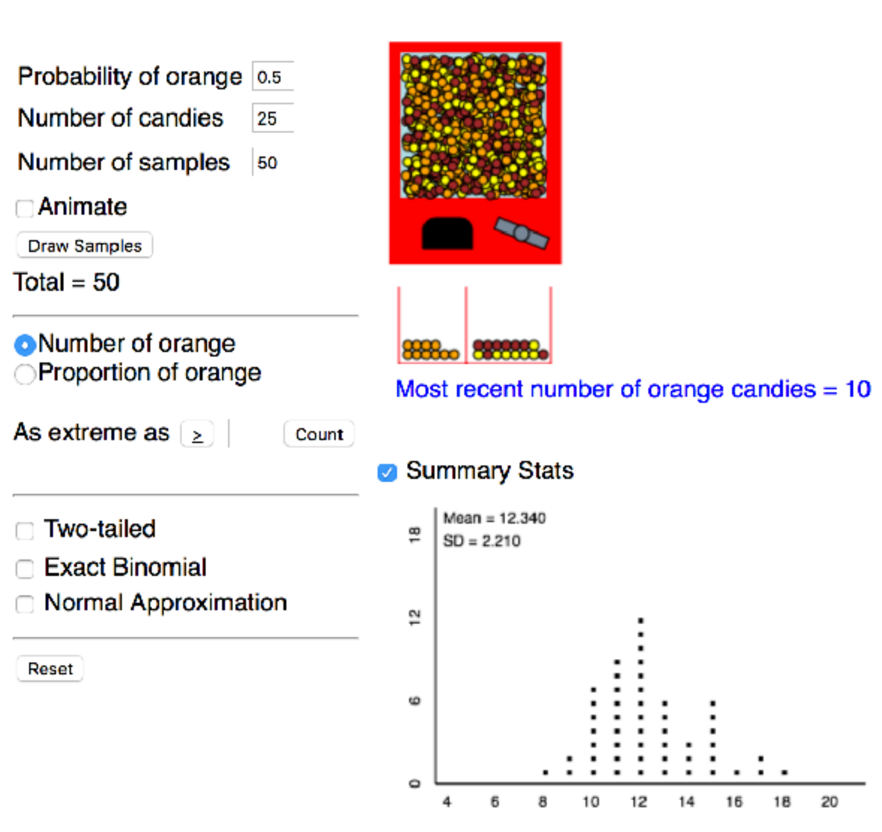
- CSM's can lead students astray by suggesting that "bigger samples are better" [5], while [3] claims using a large enough sample size can conceal misleading effects and help students understand randomness.
- [1] concluded CSM's did not lead to any elevated comprehension of the material while [4] found that hands-on activities lead to little improvement in performance over time.

General Consensus:

- Students should be actively involved in data collection and analysis.
- Having students predict the results, then use simulations to verify or contradict their predictions is an effective pedagogical method.
- Technology can facilitate learning if activities are adequately designed to promote a deeper understanding of concepts, using CSM's only as tools.
- Performing hands-on simulations prior to CSM's is widely accepted, but has not been thoroughly tested.

Rossman/Chance Applet

Reese's Pieces



Source of the Data

This experiment was performed in Dr. Stacey Hancock's Basic Statistics course offered at UC Irvine in the Fall quarter of 2015. As depicted in the diagram, one of two activity types-CSM's alone or tactile simulations preceding CSM's-was assigned to each discussion group. All TA's had the opportunity to lead both activity types, each of which focused on the same set of goals. Students filled out preliminary questions asking them to predict the outcomes. After performing the activity, students answered follow-up questions for reflection. The experiment was performed in a sequence of three discussions throughout the quarter.

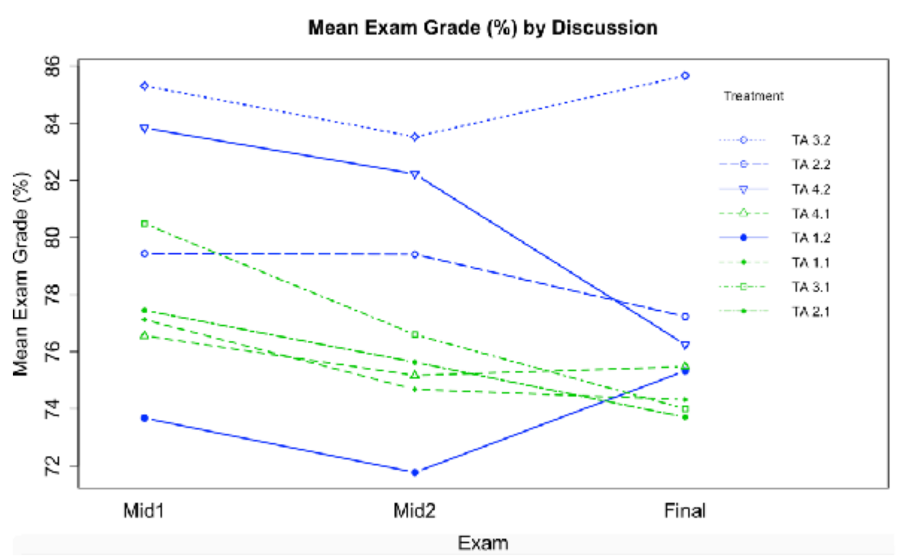
Lecture A (208 Students)	Lecture B (~211 Students)
Disc 1.A (TA 1) Fri: 11:00 - 11:50 CSM - 54 Students	Disc 1.B (TA 3) Fri: 1:00 - 1:50 CSM - 53 Students
Disc 2.A (TA 1) Fri: 12:00 - 12:50 Tactile - 52 Students	Disc 2.B (TA 3) Fri: 2:00 - 2:50 Tactile - 55 Students
Disc 3.A (TA 2) Fri: 2:00 - 2:50 Tactile - 50 Students	Disc 3.B (TA 4) Fri: 3:00 - 3:50 Tactile - 53 Students
Disc 4.A (TA 2) Fri: 3:00 - 3:50 CSM - 52 Students	Disc 4.B (TA 4) Fri: 4:00 - 4:50 CSM - 50 Students

Number of Disc's Attended	0	1	2	3
Student Count	9	23	172	182

Descriptive Statistics

	Comp (n = 90) mean (sd) or n (%)	Tactile (n = 92) mean (sd) or n (%)
Midterm 1	80 (11.9)	78 (14.2)
Midterm 2	77.9 (13.9)	76.2 (15.3)
Final	75.3 (14.9)	77.2 (14.9)
Major		
not STEM	45 (51%)	43 (48%)
STEM	44 (49%)	47 (52%)
Gender		
Female	61 (68%)	53 (58%)
Male	29 (32%)	39 (42%)

Only includes students who participated in all 3 activities and exams



Modeling

Objective:

- Determine the differences between mean exam scores for each activity type (CSM or tactile simulations prior to CSM).
- Determine the effect of the activities on retention of the material over time.
- Hypothesis - Students who participate in the tactile activities before computer simulations will show more improvement over time.

A Priori Model

- Potential Confounders: TA, major (STEM or non-STEM), and gender.
- Determine if the effect of treatment on mean exam scores changes with major (STEM majors, for instance, might respond better to CSM's).
- Model - analyze our data using a Linear Mixed Effects Model to:
  - account for correlation between repeated measures on the same subject,
  - model within and between subject variability,
  - deal with missing data.

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 X_i + \beta_2 T1_{ij} + \beta_3 T2_{ij} + \beta_4 X_i * T1_{ij} + \beta_5 X_i * T2_{ij} + \beta_6 * I_{[TA2]_i} + \beta_7 * I_{[TA3]_i} + \beta_8 * I_{[TA4]_i} + \beta_9 * I_{[Male]_i} + \beta_{10} * I_{[STEM]_i} + \epsilon_{ij}$$

where $b_{0i} \overset{iid}{\sim} N(0,\sigma_b^2)$, $\epsilon_{ij} \overset{iid}{\sim} N(0,\sigma_\epsilon^2)$ for $i = 1, \cdots, 211$ and $j = 1,2,3$

$$X_i = \begin{cases} 1, \text{ tactile} \\ 0, \text{ computer} \end{cases} \quad T1_{ij} = \begin{cases} 1 \text{ midterm 2} \\ 0 \text{ else} \end{cases} \quad T2_{ij} = \begin{cases} 1 \text{ final} \\ 0 \text{ else} \end{cases}$$

and the indicators for TA, Male, and STEM have the following structure:

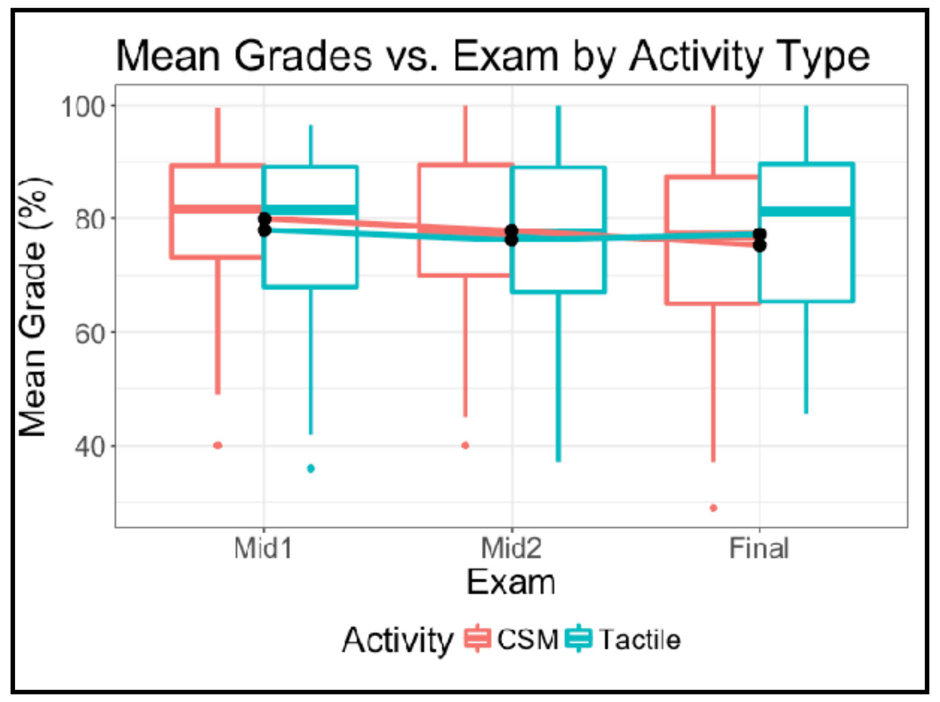
$$I_{[Z]_i} = \{1,\ Z_i \in Z; 0\ Z_i \notin Z$$

Model Selection

The following table is a summary of our final model obtained via linear mixed effects modeling with a random intercept:

	Value	Std. Error	CI.95.low	CI.95.high	t-value	p-value
(Intercept)	74.67	2.32	70.16	79.19	32.17	0.0000
Tactile	-2.43	2.02	-6.38	1.52	-1.20	0.2304
Mid2	-2.08	1.19	-4.40	0.24	-1.75	0.0811
Final	-4.63	1.19	-6.95	-2.31	-3.89	0.0001
TA2	-0.97	2.40	-5.66	3.73	-0.40	0.6880
TA3	3.04	2.45	-1.75	7.83	1.24	0.2160
TA4	1.17	2.61	-3.93	6.26	0.45	0.6554
Male	4.20	1.89	0.51	7.90	2.22	0.0277
STEM	6.24	1.83	2.67	9.82	3.41	0.0008
Tactile: Mid2	0.33	1.68	-2.93	3.59	0.20	0.8424
Tactile: Final	3.87	1.68	0.61	7.13	2.31	0.0214

As expected, the indicator for tactile was not significant as the treatment was not assigned until after the first midterm. The estimated coefficient for the final reflects a substantial decrease in mean exam scores from midterm 1 to the final for the CSM group. The interaction between treatment and exam was only statistically significant for the final exam. This is evidence that the treatment did not have a notable effect on the mean exam score until after the students participated in all three activities. The increase in mean exam score from midterm 1 to the final exam is between 0.61 and 7.13 points (out of 100) higher for those students in the tactile group, with 95% confidence. In summary, students in the tactile group are associated with higher retention of the material over time compared to the computer group.



Discussion

The study suggests that students who participated in tactile simulation activities when learning sampling distributions had a deeper understanding of the material compared to those who only performed computer simulations. With midterm 1 as our baseline measurement before the intervention took place, we found that the effect of each of the treatments was not immediate, but those in the tactile group showed better retention of the material.

Future Work

- Control for student demographics such as gender, race, and GPA.
- Analyze further assessment measurements collected throughout the quarter (e.g., student reflections, sampling distribution-specific exam questions).
- For generalizability, we want to perform the experiment across a broader range of schools.

References

[1] Chance, B., delMas, R., and Garfield, J. (2004), Reasoning About Sampling Distributions, in *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, eds. D. Ben-Zvi and J. Garfield, Dordrecht, The Netherlands: Kluwer, pp. 295-323.
[2] Garfield, J., Le, L., Zieffler, A. et al. Educ Stud Math (2015) 88: 327. doi:10.1007/s10649-014-9541-7
[3] Lane, D., *Simulations of the Sampling Distribution of the Mean Do Not Necessarily Mislead and Can Facilitate Learning*, Journal of Statistics Education **v.23, n.2**, 2015.
[4] Pfaff, T. and Weinberg, A., *Do Hands-On Activities Increase Student Understanding?: A Case Study*, Journal of Statistics Education **v.17, n.3**, 2009.
[5] Watkins, A., Bargagliotti, A., Franklin, C., *Simulation of the Sampling Distribution of the Mean Can Mislead*, Journal of Statistics Education **v.22, n.3**, 2014.d