An Active Approach to Statistical Inference Using Randomization Methods

Todd Swanson[1], Nathan Tintle[1], Jill VanderStoep[1], Vicki-Lynn Holmes[1,2] and Brooke Quisenberry[3]

[1]Dept. of Mathematics, [2]Dept. of Education, [3]Dept. of Psychology, Hope College, Holland, Michigan

Introduction

For more than a decade the popular algebra-based introductory statistics course has had a generally accepted consensus curriculum focused on asymptotic distributions. While the statistics education reform movement, culminating in the GAISE guidelines (GAISE 2005), has greatly improved pedagogy, there has been little serious re-thinking about the core content of the curriculum. Cobb summarized the consensus approach to statistical inference by saying "our [consensus] curriculum is needlessly complicated because we put the normal distribution, as an approximate sampling distribution for the mean, at the center of the curriculum, instead of putting the core logic of inference at the center." (Cobb 2007). Instead of spending weeks on probability and sampling distributions, students can use a computationally intensive (computer-based) method called *permutation testing* or *randomization* to learn statistical inference more easily.

Recent projects (Rossman et al., West and Woodard) are developing modules to motivate statistical inference through randomization methods. Pilot testing of these modules and the resulting assessment data confirm that statistical inference is better learned by students using a randomization approach than traditional approaches (Holcomb et al. 2010a, Holcomb et al. 2010b, Tintle 2010, Tintle et al. 2010).

Instead of just including modules in a traditional course, we are developing a curriculum for introductory statistics that fully embraces the randomization approach. A preliminary version of this curriculum was compiled into a textbook titled *An Active Approach to Statistical Inference* during the summer of 2009. This in-house publication has since been used in all the introductory statistics courses at Hope College and continues to be revised.

Content

Since we believe that inference should be the core of an introductory statistics course, we begin with inference on the first day of the course and teach it throughout the entire semester. It is important to note, however, the core-logic of statistical inference is more than just obtaining a p-value. The core-logic of statistical inference encompasses the whole scientific process of data analysis and the logic that it is founded upon. For that reason, our students are exposed to real research throughout the course through activities, exercises, case studies, research papers, and their own projects.

Main differences between our randomization curriculum and traditional ones.

- We spend the bulk of the time visiting and re-visiting the core-logic of statistical inference as demonstrated by the randomization method.
- There is growing exposure to descriptive statistics at the K-12 level and many of these topics are well understood by students before taking statistics at college (CAOS; delMas et al. 2007). Therefore, we spend limited time teaching descriptive statistical methods and instead include time to review and reinforce the proper use of descriptive statistical methods through hands-on real data analysis experiences.
- Since traditional probability and sampling distributions are unnecessary to understand the core logic of statistics, we eliminate their explicit coverage.
- Power analysis, which is typically difficult for students, receives a more prominent place in our curriculum. We present an intuitive approach to power by looking at the relationships between power and sample size, standard deviation, difference in population proportions or means, etc. We think this helps students better understand the core logic of statistical inference.
- Confidence intervals are presented after inferential techniques instead of vice versa. We demonstrate how tests of significance can be used to create ranges of plausible values (more commonly known as confidence intervals) for the population parameter.

Table of Contents

<u>Unit 1. Introduction to inferential statistics using randomization methods.</u>

Chapter 1: Introduction to Statistical Inference: One proportion. Students are introduced to the process of conducting a test of significance. Flipping coins and applets are used to model the null and their results are used to determine p-values.

Chapter 2: Comparing Two Proportions: Randomization Method. Students are shown what explanatory and response variables are and the randomization method is introduced by comparing two proportions using playing cards and Fathom.

Chapter 3: Comparing Two Means: Randomization Method. Cards and Fathom are used again to aid in understanding the process of comparing two means. Also discussed are type I/type II errors and observational studies/experiments.

Chapter 4: Correlation and Regression: Randomization Method. Scatterplots, correlation, and regression are reviewed and the randomization method is used to test correlation. The meaning of r-squared is also introduced.

<u>Unit 2. Revisiting statistical inference using asymptotic methods, confidence intervals and statistical power.</u>

Chapter 5: Correlation and Regression: Revisited. Using inference on correlation, we transition to using traditional methods by showing how a sampling distribution can be used to model the randomization distributions. Confidence intervals are introduced as a range of plausible values for a population parameter. Power is introduced and is shown how it relates to sample size, significance level, and population correlation.

Chapter 6: Comparing Means: Revisited. Standard deviation, normal distributions, and t-distributions are discussed. The independent samples t test is introduced and we see how it is related to the randomization method. Related confidence intervals and power topics are included. The traditional paired-data t test and ANOVA are also introduced.

Chapter 7: Comparing Proportions: Revisited. The traditional test for comparing two proportions is introduced. Power is explored in relationship to the difference in population proportions, sample size, significance level, and size of the two proportions. The chi-square test for association is introduced.

Chapter 8: Tests of a Single Mean and Proportion. We revisit the test of a single proportion, exploring the exact binomial and normal approximation methods. We extend these ideas to tests of a single mean. We also introduce chi-squared goodness of fit test.

Pedagogy

In addition to changes to the content of the course, we have also significantly changed our pedagogical approach from passive (e.g. listening to lectures) to active learning which engages the full range of senses. After a brief introduction to a concept by the instructor, an activity is assigned in which the students learn or receive reinforcement on the concept. Students are involved in tactile learning experiences like shuffling cards and flipping coins to estimate their own p-values, using computer based simulations, collecting data, running experiments, and using computer software to help interpret results. Students also complete in-depth projects where they design a study, collect data, and present their results in both oral and written form. Overall, we advocate utilizing a small amount of lecture, but engage and strengthen different learning processes by way of a variety of active, self-discovery learning experiences.

In addition to the activities, each chapter includes exercises, a case study, and questions about a research article that the students read. The GAISE guidelines argue that statistics courses should utilize real data. We go a step further and argue that statistics courses should use real data *that matters*. Statistics should be less of a course in which students see cute but impractical illustrations of statistics in use, and more about examples where statistics is used to make decisions that have health, monetary or other implications impacting hundreds, thousands or millions of people. Our approach has two-fold benefit, first in improving their statistical literacy and second by motivating students that statistics is the inter-disciplinary language of research.

Assessment

The Comprehensive Assessment of Outcomes in Statistics (CAOS; delMas et al. 2007) represents the first standard, comprehensive assessment instrument for the algebra-based introductory statistics course. Students in our randomization course took this pre- and post-test in the Fall of 2009 (n = 202). These results were compared with students that took our traditional course in the Fall of 2007 (n = 198) and those from a national representative sample (n = 768).

Overall, learning gains were significantly higher for students that took the randomization course when compared to either those that took the traditional course at Hope or the national sample. More can be understood, however, when we look at the results question by question. Seven items showed significantly better student performance using the new curriculum as demonstrated by significantly higher posttest scores and statistically significant student learning gains. (See table below.)

CAOS identified four questions where the new curriculum performed worse than the traditional and national data. These questions involved one question from each of the following areas: sampling variability and the law of large numbers, boxplot interpretation, data collection and design, as well as one question involving a particular misinterpretation of the p-value. The first topic isn't taught in the new curriculum and the second topic is not explicitly taught in the new curriculum. These results will lead us to curricular improvements in subsequent versions of our textbook.

Items where students in the randomization curriculum performed significantly better than the traditional curriculum

Item Description (Topic)	Curriculum	% of Students Correct			McNemar's Test p[1]	Fisher's Exact Test p[2]
		Pre	Post	Diff		
Understanding of the purpose of randomization in an experiment (Data collection and design)	National	8.5	12.3	3.8	**0.013**	**0.003**
	Hope (Trad)	4.6	9.7	5.1	0.076	**0.003**
	Hope (Rand)	3.5	20.8	17.3	**<0.001**	-
Understanding that low p-values are desirable in research studies (Tests of significance)	National	49.9	68.5	18.6	**<0.001**	**<0.001**
	Hope (Trad)	56.9	85.6	28.7	**<0.001**	**<0.001**
	Hope (Rand)	56.9	96.0	39.1	**<0.001**	-
Understanding that correlation does not imply causation (Data collection and design)	National	54.6	52.6	-2.0	0.404	**0.021**
	Hope (Trad)	52.1	54.4	2.3	0.640	0.154
	Hope (Rand)	44.1	61.9	17.8	**<0.001**	-
Understanding that no statistical significance does not guarantee that there is no effect (Tests of significance)	National	63.1	64.4	1.3	0.630	**<0.001**
	Hope (Trad)	66.2	72.7	6.5	0.130	**0.003**
	Hope (Rand)	65.2	85.1	19.9	**<0.001**	-
Ability to recognize a correct interpretation of a p-value (Tests of significance)	National	46.8	54.5	7.7	**0.005**	0.172
	Hope (Trad)	36.1	41.0	4.9	0.402	**<0.001**
	HR	42.3	60.0	17.7	**0.002**	-
Ability to recognize an incorrect interpretation of a p-value. Specifically, probability that a treatment is not effective. (Tests of significance)	National	53.1	58.6	5.5	**0.044**	**<0.001**
	Hope (Trad)	59.8	68.6	8.8	0.085	**0.012**
	Hope (Rand)	58.9	79.7	20.8	**<0.001**	-
Understanding of how to simulate data to find the probability of an observed value (Probability)	National	20.4	19.5	-0.9	0.713	**<0.001**
	Hope (Trad)	20.0	20.0	0.0	1.000	**0.006**
	Hope (Rand)	20.0	32.2	12.2	**0.009**	-

1. McNemar's test comparing pre-test and post-test percentages within each curriculum. Equivalent to seeing if there is evidence that the difference in % correct is different pretest to posttest.
2. Fisher's exact test comparing post-test percentages from the traditional curriculum to the new curriculum (National vs. Hope (Rand) and Hope (Trad) vs. Hope (Rand)).

Literature cited

Cobb, G. (2007) "The introductory statistics course: a Ptolemaic curriculum?" *Technology Innovations in Statistics Education*. 1(1).
delMas, R., Garfield, J., Ooms, A., and Chance, B. (2007) "Assessing students' conceptual understanding after a first course in statistics" *Statistics Education Research Journal* 6(2):28-58.
Guidelines for Assessment and Instruction in Statistics Education (GAISE) (2005). http://www.amstat.org/education/gaise/ (accessed May 13, 2010.)
Holcomb, J., Chance, B. Rossman, A., Tietjen, E., and Cobb, G. (2010a), "Introducing Concepts of Statistical Inference via Randomization Tests," *Proceedings of the 8th International Conference on Teaching Statistics.*
Holcomb, J., Chance, B., Rossman, A., & Cobb, G., (2010b). "Assessing student learning about statistical inference," *Proceedings of the 8th International Conference on Teaching Statistics.*
Rossman, A., Chance, B., Cobb, G. and Holcomb, J. (2008) DUE-CCLI 0633349; "Concepts of Statistical Inference: A randomization-based curriculum." http://statweb.calpoly.edu/csi/ (accessed May 13, 2010.)
Tintle, N., VanderStoep J., Holmes V-L., Quisenberry B., and Swanson T. (2010). "Development and assessment of a preliminary randomization-based introductory statistics curriculum" Submitted to *Journal of Statistics Education*, April 2010. Draft version accessed http://math.hope.edu/aasi, May 10, 2010.
Tintle, N (2010). "Innovations in statistics education yield improved quantitative preparation for life sciences students" Submitted to *CBE−Life Sciences Education*, April 2010.
West, W. and Woodard, R., (2009) DUE 0817262; INCIST: Improving National Acceptance of Computing Intensive Statistical Techniques.

Acknowledgments

Funding for this project has come from a Howard Hughes Medical Institute Undergraduate Science Education program grant to Hope College, the Great Lakes College Association Pathways to Learning Collegium, and the Teagle Foundation. We also thank George Cobb, Allan Rossman, Beth Chance and John Holcomb for being willing to have us adapt materials from their project. Lastly we acknowledge the many departments at Hope College for their willingness to support and embrace radical changes to our statistics courses.

Further information

For further information about this project please contact Nathan Tintle at tintle@*hope.edu or visit www.math.hope.edu/aasi*.