EXPLORING SIMULATION-BASED INFERENCE IN A HIGH SCHOOL COURSE

Bruce Dunham
Department of Statistics, University of British Columbia, Canada
b.dunham@stat.ubc.ca

Although inferential concepts are typically introduced in courses at high school, the approaches taught are usually the methodologies in introductory classes at university level. There is much research to support that learners have difficulty with classical frequentist inference and that a better understanding of inferential concepts can be obtained via an introduction using simulation-based methods. A new course available to high schools in British Columbia, Canada, incorporates several novel aspects, a key feature being the reliance on "intuitive," simulation-based inference. We describe the pedagogical approaches adopted in this course and how students appeared to have learned from their experiences.

INTRODUCTION

Given the ample evidence that university undergraduates have difficulty mastering concepts in traditional statistical inference (e.g., Castro Sotos et al., 2007; Chance et al., 2004), it is surprising that the curricula of statistics courses at the high school level typically include classical inferential methods (see, for example, curricula for AP Statistics (AP Central, 2022) and A-level Statistics (Cambridge International Assessment, 2022)). Alternative approaches to inference, based around simulation of sampling distributions, have been advocated in recent years and evidence of improved learning from such methods has accumulated (see, as examples, Beckman et al., 2017; Hildreth et al., 2018; Tintle et al., 2012). Because there are also resources available for facilitating simulation-based instruction such as *StatKey* (Lock et al., 2017) and the *Introduction to Statistical Investigations* applets (Tintle et al., 2015), there are strong arguments that an introduction to inference at school level should be based around simulations of sampling distributions rather than the use of theoretical results.

Described here is a new course, Statistics 12, at high school level in British Columbia (BC), Canada. The development and aims of the course are briefly reviewed. The author was involved in co-teaching a class of Statistics 12 in 2021, and details are provided on the course structure, materials, pedagogy, assessment, student engagement, and learning.

STATISTICS 12 BACKGROUND

In 2015, The BC Ministry of Education announced plans for new high school graduation requirements and curricula, including a proposal for a new grade 12 statistics course. At the time there was no statistics course at school level in the province. The author collaborated with local educators and officials in the province to develop documentation describing the new course (BC Ministry of Education, 2018). The main themes for the new course are as follows:

1. Appreciating the role of statistics in research, decision making, and policy.
2. Understanding types of research and evidence – what makes a good study?
3. Exploring, describing, modeling, and explaining variation.
4. Conducting inference based on simulation.
5. Communicating statistical ideas.

Recognising that local teachers required support if they were to feel comfortable handling the new course, in the following years the author ran various workshops for local educators both in person and online. The workshops introduced teachers to simulation-based inference (SBI) and provided sample activities and assessments. An article motivating the new course, and instruction in teaching statistics in general, appeared in a journal for local mathematics teachers (Dunham, 2018).

In the fall of 2021, the author co-taught a class of Statistics 12 with an experienced teacher at Centennial High School, Coquitlam, BC. Detailed learning outcomes, in-class activities, and assessments were developed. The course is described in what follows.

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

STATISTICS 12 PEDAGOGY

The course ran over an eighteen-week semester, with the class meeting daily for approximately 75 minutes. The final enrolment was twenty, mostly grade 12 students for whom the class counted as a mathematics elective. Efforts were made to use evidence-based pedagogy: classes were activity-based, whereby after a brief introduction students worked in small groups on an activity provided. Students were encouraged to hand in their work at the end of class, and the work was returned with both written and verbal feedback the following class. Both teachers were present to support the students in each class. Students were expected to write reflections on their engagement with the learning outcomes for the topic encountered each week. The students were directed to engage with online materials developed at CK-12 Foundation (2022) for learning outside of class.

The activities centered around real case studies. Inferential ideas for a proportion, a mean, a difference in two proportions, a difference in two means, two categorical variables, matched pairs, and regression were explored via both physical simulations (using dice, cards, coins, and tacks) and online applets (available at the online repository StatSpace (https://statspace.elearning.ubc.ca) plus those of Tintle et al., 2015, which includes examples on which some activities were based). The emphasis was on exploring a study starting from the research question and culminating in communicating conclusions in context. Each activity was expected to take about one hour in class, and students were permitted to work through the activities at their own pace rather than skip activities for classes they either missed or did not engage with.

Most classes started with around ten minutes of direct instruction. Usually this related to ideas previously explored in activities and, where appropriate, how those ideas were to be extended in the activity for that class. Any common errors and misconceptions arising from the previous activity were also discussed.

Students were given course credit for the work they submitted on the activities and their weekly reflections, although these formative assessments were not graded. A handful of additional activities, similar to those worked in class, were set as components of the course homework. There were five in-class tests spaced throughout the semester, including a cumulative final exam. Each test was two-stage (45 minutes individual, 25 minutes in groups). Practice tests were provided before a test, each including questions assessing the relevant learning outcomes.

Two group projects progressed throughout the semester. The first required the students to conduct an observational research study, the second required the groups to design and conduct an experiment. Students were assigned to different groups of four for each project. Each group was required to submit a detailed proposal outlining their research question(s) and suggested methodology before embarking on gathering data. Students presented their projects to the class in the final week of term. Each student submitted their own written report of their group's work after a peer review of an initial draft. Students were expected to apply SBI to their data to explore plausible answers to their research questions.

ASSESSMENT OF LEARNING

The extent to which students mastered learning outcomes relating to sampling variability and inferential concepts was explored via (a) the correctness of application and interpretation of SBI in the group projects and (b) responses to CAOS (delMas et al., 2007) items 32–35 presented on the final test.

There were ten group projects completed overall, five observational studies and five experiments. Pleasingly, all groups attempted SBI in their analyses via the online tools they had used in the course, with only one group given substantive guidance (that was ignored by the students in their analysis). In total, seven of the ten analyses were deemed to be broadly correct in their applications of SBI and the inferences obtained. Errors arose in other cases. One group compared just the mean of the distribution of correlations obtained by repeatedly sampling their bivariate data with the observed correlation. In a study estimating a proportion based on a large sample and comparing with a proportion found in a published study, the group were unable to decide whether their observed proportion was consistent with the published value and instead reported requiring more data. Regrettably, one group appeared to have entered their data into the correct Tintle et al. (2015) applet but neglected to press "Use Data" and so conducted SBI on data that were irrelevant to them. In their analysis, that group also ignored the pairedness that had been sensibly incorporated into their design.

Lastly, one group correctly performed the appropriate simulation-based hypothesis test but concluded that an "empirical P-value" (not a term used in the course, incidentally) of 20% was "surprising" and therefore cast doubt on their null hypothesis, reflecting that there had been little discussion of how far into the tails of a simulated sampling distribution a test statistic should fall before we deem its value "surprising."

The CAOS test items were modified slightly by the addition of a request to "Explain your answer." The individual student responses are compared here to post-course percentages reported by delMas et al. (2007) from 763 students in higher education. On item 32 (on sampling variation), 30% of the students were correct compared to the 17% reported by delMas et al.; however, this item appears flawed, noting that three of the six correct responses were accompanied by a faulty argument while one incorrect response was supported by correct reasoning. Issues stem from use of the terms "sampling error" ("sampling variation" would be preferable) and the nebulous "almost identical." On item 33 (on interpreting an empirical distribution), 65% were correct (compared to 40%), on item 34 (on the distribution of a sample), 65% were correct (compared to 65%), and on item 35 (distribution of the sample mean), 45% were correct (compared to 44%) although one correct response included a dubious explanation. The group attempts at these four CAOS test items gave percentages correct of 20%, 60%, 80%, and 80% respectively.

CONCLUSIONS

Although the students found much of the material challenging, most persevered with the in-class activities to the extent that class attempts at established concept inventory questions at the end of the course were comparable (indeed, slightly better) than results previously reported for undergraduate students. Moreover, all students were able to attempt to apply SBI to data they had obtained in group projects without instructor support. Although a minority of groups did struggle with implementing or interpreting their analyses, there was clear evidence that the students were able to transfer inferential concepts to a novel setting.

Much effort was made to implement evidence-based pedagogy such as activity-based learning, an approach with which the author has much experience in undergraduate education. The relatively small class was consistently supported by two able teachers, and the person-hours devoted to the course (particularly in reading and giving feedback on student work) was in excess of what could typically be expended. The SBI ideas were introduced using physical simulations and then explored further using established and freely available online tools that, to a discipline expert at least, offer a way for learners to apply and appreciate interval estimation and hypothesis testing without having to engage with difficult mathematics.

In some respects, both learning and engagement were somewhat disappointing. Perhaps it related to the COVID-19 pandemic that some students struggled to maintain their effort during the fall term. By the winter break, around a third of the class were well behind schedule with the activities, and working consistently during class time appeared to be very effortful for them. Hopes that nearly all students would, given the resources utilized for the course, "knock it out of the park'' when it came to understanding statistical inference proved to be optimistic. The conclusion was that although learners obviously must be active to attain meaningful, lasting learning, such work may exceed the effort level some students are willing (or perhaps able) to expend.

Another factor was that, although encouraged to bring laptops and tablets to school, most students relied solely on their cell phones by which to interact with the online applets throughout the course. All adopted applets worked on phones but due to small screen size, the controls and inputs could be fiddly. More seriously, it is suspected that seeing the desired aspects of the simulated sampling distributions becomes more difficult when using the applets on a phone rather than on a laptop or tablet.

That said, most students appeared to genuinely enjoy the course and the learning materials. The students seemed to particularly appreciate the personal contact and feedback each class. All students obtained at least a partial appreciation of statistical inference, which is, perhaps, the best that can be expected for an introductory class at high school level. A few students clearly demonstrated a good level of understanding of SBI by the end of the course.

With regards to reducing workload for both students and teachers in the course, although the learning journals were evidently helpful for some students, many seemed to struggle reflecting on the learning outcomes. Perhaps a more efficient way to encourage students to link learning outcomes to activities could be explored, such as via online quizzes.

It is hoped Statistics 12 can become a model for high school statistics classes. To that end, all activities developed for the course will be made freely available on StatSpace this year.

ACKNOLWEDGEMENTS

The author wishes to thank staff and students at Centennial High School, Coquitlam, BC, especially teacher Ms. Anita Do and principal Mr. Anthony Ciolfitto. The online resources developed by Profs Mike Whitlock and Nathan Tintle (and collaborators) were integral to this work. The project would not have been possible without the generous support of UBC's Teaching and Learning Enhancement Fund.

REFERENCES

AP Central. (2022). *AP Statistics course and exam description*. College Board. https://apcentral.collegeboard.org/courses/ap-statistics

Beckman, M. D., delMas, R. C., & Garfield, J. (2017). Cognitive transfer outcomes for a simulation-based introductory statistics curriculum. *Statistics Education Research Journal*, *16*(2), 419–440. https://doi.org/10.48550/arXiv.1707.06537

British Columbia's Ministry of Education. (2018). *Statistics 12 curriculum*. https://curriculum.gov.bc.ca/curriculum/mathematics/12/statistics

Cambridge International Assessment. (2022). *Syllabus Cambridge International Assessment and A Level Mathematics 9709.* https://www.cambridgeinternational.org/Images/415060-2020-2022-syllabus.pdf

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review,* *2*, 98–113. https://doi.org/10.1016/j.edurev.2007.04.001

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–323). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2278-6_13

CK-12 Foundation. (2022). *CK-12 probability and statistics concepts*. https://www.ck12.org/book/ck-12-advanced-probability-and-statistics-concepts/

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28– 58. https://doi.org/10.52041/serj.v6i2.483

Dunham, B. (2018). Why statistics? *Vector, The official Journal of the BC Association of Mathematics Teachers, 59*(2), 10–13. https://www.bcamt.ca/wp-content/uploads/Vector-2018-592-online.pdf

Hildreth, L. A., Robison-Cox, J., & Schmidt, J. (2018). Comparing student success and understanding in introductory statistics under consensus and simulation-based curricula. *Statistics Education Research Journal*, *17*(1), 103–120. https://doi.org/10.52041/serj.v17i1.178

Lock, R., Frazer Lock, P., Lock Morgan, K., Lock, E., & Lock, D. (2017). *StatKey* [Computer software]. https://www.lock5stat.com/StatKey/

Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2015). *Introduction to statistical investigations: Applets and data sets for iPad and other mobile devices* [Applets, Data sets]. http://www.isi-stats.com/isi/applets.html

Tintle, N., Topliff, K., VanderStoep, J., Holmes, V., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, *11*(1), 21–40. https://doi.org/10.52041/serj.v11i1.340