USING RE-SAMPLING AND SAMPLING VARIABILITY IN AN APPLIED CONTEXT AS A BASIS FOR MAKING STATISTICAL INFERENCES WITH CONFIDENCE

Luis Saldanha[1] and Michael McAllister[2]
[1]Département de Mathématiques, Université du Québec à Montréal, Canada
[2]School of Mathematical and Statistical Sciences, Arizona State University, USA
saldanha.luis@uqam.ca, memcalli@asu.edu

We describe an instructional sequence that engaged a class of 9th grade students in making statistical inferences on the basis of distributions of a sample statistic. The sequence involved a scenario and tasks that entailed comparing samples of two types of organisms on a common attribute. Students engaged in: 1) making sense of the scenario and a TinkerPlots™ simulation that produced distributions of a sample statistic, 2) examining and interpreting a sequence of such distributions in relation to increases in sample size, and 3) drawing a conclusion about the attribute in the sampled population and assessing their confidence of the conclusion. We highlight aspects of students' understandings of what an empirical sampling distribution represented in terms of the scenario's context, and their abilities to track the multi-tiered re-sampling process that began with a population and culminated with distributions of the sample statistic.

INTRODUCTION AND BACKGROUND

The relevance of statistical decision making for an informed citizenry in our information and data-driven age makes statistical inference arguably one of the most important schemes of ideas to target in school mathematics instruction (Garfield & Ben-Zvi, 2008). The apparent logic of inference is that a characteristic of a population of interest—the whole of which is usually not directly accessible—can be inferred only indirectly by examining the same characteristic for a randomly drawn subset (a sample) of that population. In practice, such inferences are typically made on the basis of a single sample drawn from a population. Yet, randomly selected samples have outcomes (values of a particular statistic) that typically vary from sample to sample. As Rubin, Bruce and Tenney (1990) have argued, the ability to balance and coordinate these two seemingly antithetical ideas—that of individual sample representativeness and the idea of sample-to-sample variability—is key to a coherent understanding of inference. The challenge for instruction, as Rubin et al. (1990) frame it, is to have students unify these contrasting ideas of representativeness and variability into the single notion of a sampling distribution. We have echoed Rubin et al.'s argument in that "…we do not see how the normative practice of drawing an inference from an individual sample to a population can be understood deeply without reconciling the ideas of sample-to-sample variability and relative frequency patterns that emerge in collections of values of a sample statistic…" (Saldanha & Thompson, 2007, p.275).

Over the last fifteen years, the development of software designed specifically to support the learning of statistics with a focus on visualization (e.g., Finzer, 2012; Konold & Miller, 2011) has made the use of sampling simulations and the generation of associated sampling distributions increasingly common in statistics instruction at the high school level and beyond. Statistics educators and researchers have recommended and explored the use of simulation-based statistics instruction and curricula to support the development of students' understanding of inference (e.g. Garfield & Ben-Zvi, 2008; Zieffler & Garfield, 2007). Our paper reports on an effort in this vein involving a group of 9th-grade students' engagement with simulation-based instructional activities intended to support their ability to make distribution-based inferences.

THE STUDY AND INSTRUCTIONAL SEQUENCE

Participants, Instructional Context, and Data Corpus

We report on the second phase of an instructional sequence that engaged an intact class of 9th grade students in three 65-80 minute lessons in a school located in a suburb of a large city in the southwestern United States. Instructional activities involved the use of *TinkerPlots*™ software (Konold & Miller, 2011) to explore and analyze data sets in a first phase, and to simulate re-sampling and use the resulting sampling distributions as a basis for inferring a population

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

parameter's value in a second phase. Students had some prior knowledge of elementary statistical ideas and conventional inscriptions acquired in their previous coursework. For instance, most students understood a *sample* to be a "small part" of a larger group of items that could be used to indicate information about a characteristic of the latter. Students knew how to compute the arithmetic mean of a set of values, and they had a procedure for finding its median ("the middle value"). Students also demonstrated an ability to construct and use dot plots and histograms to compare and draw conclusions about two data sets. Students had not previously been exposed to *TinkerPlots*, nor had they previously engaged with re-sampling activities or been exposed to distributions of a sample statistic as products of repeated sampling.

The authors participated in conducting instruction and data collection; the first author oversaw and orchestrated the unfolding of the instructional sequence and class discussions that emerged, while the second author observed and took field notes. The class lessons and discussions were recorded with digital video cameras, and all students' written work on activity sheets and a final exam were recorded. Each student took part in a video-recorded exit interview conducted within one week of the end of the instructional sequence. Our overarching research goal was to gain insight into students' thinking and understandings of ideas promoted in instruction in relation to their engagement with that instruction.

The Instructional Sequence: Lessons 8-10

The phase of the instructional sequence reported here engaged students in an investigation designed to foster their ability to make inferences to a population by considering the variability amongst outcomes of samples of a common size chosen from that population, and by comparing the variability across distributions of a sample statistic generated from collections of such samples of different sizes. The investigation centered on two big statistical ideas: 1) random sampling can be used to draw conclusions about the sampled population, and 2) larger samples lead to sampling distributions that tend to be less variable and hence lead to more confident conclusions about the population. The context for the investigation revolved around the following opening scenario that involved testing whether a species of genetically engineered fish tends to grow longer than normal fish (Key Curriculum Press, 2011):

A fish farmer stocked a pond with a new type of genetically engineered fish. The company that supplied the new type claims that these fish will grow to be longer than normal fish. The farmer decided to test the company's claim by stocking the pond with 625 fish, some normal and some genetically engineered. When the fish were fully-grown the farmer caught a sample of 130 fish from the pond and measured the length of each fish in his sample.

TinkerPlots' sampling simulator tool was introduced and used throughout this part of the sequence to efficiently generate data and collections of random samples of various sizes from the population of fish, and to graphically display the distributions of a resulting sample statistic.

Lesson 8. This lesson introduced students to the *Fish Farmer* scenario shown above (adapted from Key Curriculum Press, 2011, and Konold, 2005), framing the issue as one of a skeptical fish farmer wanting to test the claim that a genetically engineered type of fish tends to grow longer than a normal type. In the opening part of the investigation, prior to divulging the fish farmer's approach, students were first asked to consider how the farmer might go about testing the claim. The ensuing class discussions focused on issues of data collection and selecting a representative sample, providing occasion for students to consider possible ways in which the farmer might proceed. Students then examined a *TinkerPlots* data file showing the lengths of the two types of fish in the farmer's sample of 130 fish. They explored the data using *TinkerPlots* graphical tools and techniques that they had learned in the preceding phase of the sequence (see Figure 1 for an example of such); students considered the group differences and what they suggested about the lengths of the two types of fish in the population on the basis of what they observed in this single random sample. The big statistical idea promoted in this lesson was that if a randomly selected sample is assumed to be representative of its parent population, then it can be used as a basis for making claims about that population.

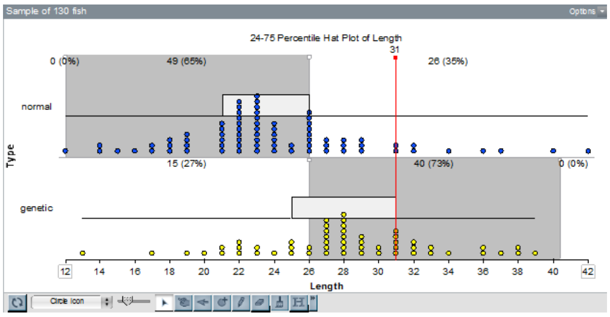
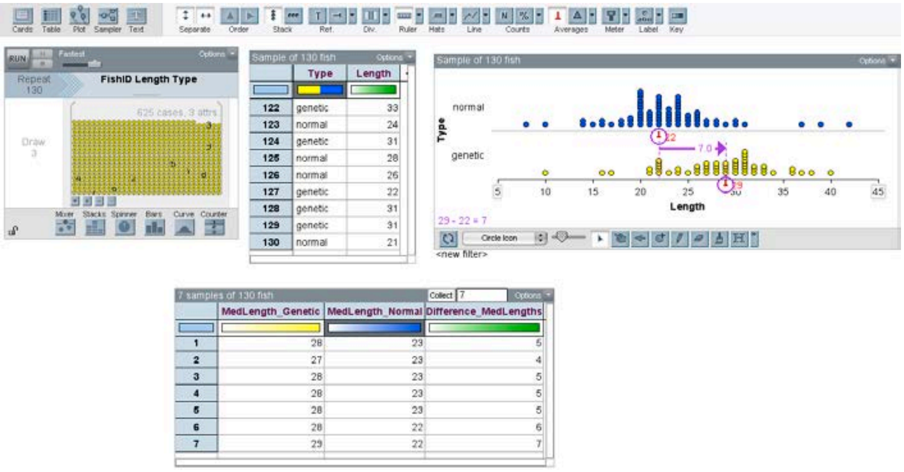


Figure 1. Two students' use of *TinkerPlots* tools to analyze a sample of 130 fish-lengths, separated into lengths of normal and genetically engineered fish in the sample.

Lesson 9. The subsequent lesson introduced students to the idea that sampling outcomes are expected to vary from sample to sample, and that such variability therefore poses a problem for making inferences to an underlying population on the basis of a single sample. At the same time, the lesson aimed to help students begin to understand that the variability amongst samples exhibits patterns that are predictable over the long run, and that such patterns can be discerned by analyzing collections of sampling outcomes. The lesson began by having students consider whether they would expect to obtain similar or different results if another sample of 130 fish were randomly drawn from the fish farmer's pond. Students were then prompted to reflect on whether a sample of 130 fish is big enough to make a confident claim about the sampled population, and to share and justify their intuitions about this question. The lesson then moved to a more systematic exploration of these questions by having students use *TinkerPlots*' sampling simulator to generate several samples of size 130, and then of size 15, from the simulated fish population (see Figure 2). Students recorded the median length of each type of fish in a sample, and the difference between these medians as a measure of the group differences (i.e., difference between the median length of two types of fish in a sample). Figure 2 displays the *TinkerPlots* set-up that the instructor and students used in this investigation; the simulator (left hand tool) used a mixer to represent the population of mixed fish. Each selected sample of 130 fish was represented in a case table displaying each fish's type and length. The case table was linked to an ordered and stacked dot plot showing the distribution of lengths of fish in the sample, separated by type and displaying the median of each type as well as the difference between medians using *TinkerPlots*' ruler tool. Each row of the table at the bottom of Figure 2 recorded the value of the three measures for a sample (median length of each type of fish and the difference between the medians) generated by running the simulation once.



Sample of 130 fish

	Type	Length
122	genetic	33
123	normal	24
124	genetic	31
125	normal	28
126	normal	26
127	genetic	22
128	genetic	31
129	genetic	31
130	normal	21

29 - 22 = 7

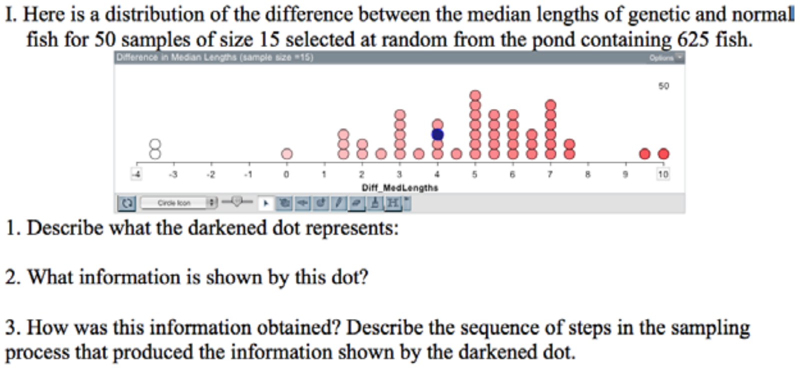
7 samples of 130 fish

	MedLength_Genetic	MedLength_Normal	Difference_MedLengths
1	28	23	5
2	27	23	4
3	28	23	5
4	28	23	5
5	28	23	5
6	28	22	6
7	29	22	7

Figure 2. A *TinkerPlots* simulation of sampling 130 fish from a population of 625 mixed fish. The table at the bottom shows the three measures recorded for seven trials of the sampling experiment.

Students explored the patterns in these measures generated by simulating the sampling experiment seven times; they identified similarities and differences among the resulting medians and difference in medians for the collection of seven samples, and they used their observations as a basis for proposing how this might help test the claim that genetically engineered fish tend to grow longer than normal fish in the larger population. Class discussions around this exploration showcased students' perceived patterns, culminating with a general consensus that the genetically engineered fish in the population were inferred to "between 4 and 7 centimeters longer" than the normal fish.

Lesson 10. The final lesson built on the activities and issues raised in Lesson 9 by having students examine the effects of sample size on the variability of the difference between median lengths of fish that they had previously explored only for seven samples. The lesson began with a demonstration and discussion of the *TinkerPlots* simulation of selecting 50 samples of size 15 from the simulated fish population, culminating with the presentation of a distribution of the difference in median lengths for each of the 50 simulated samples as shown in Figure 3.



I. Here is a distribution of the difference between the median lengths of genetic and normal fish for 50 samples of size 15 selected at random from the pond containing 625 fish.

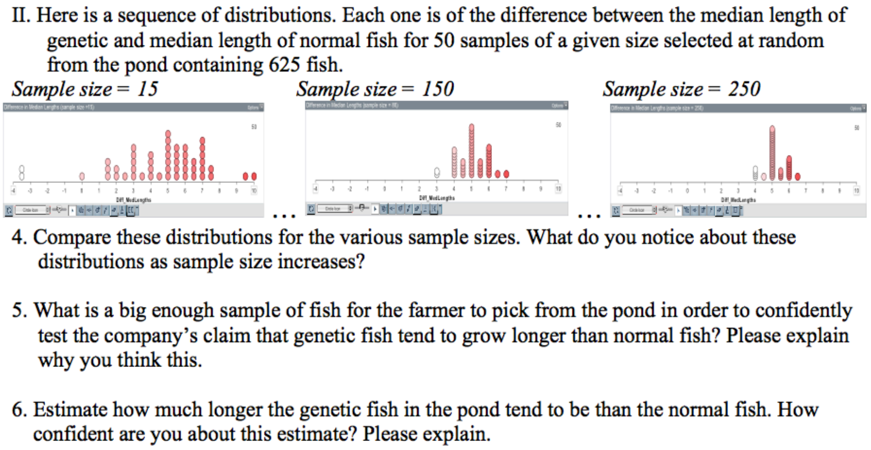
1. Describe what the darkened dot represents:

2. What information is shown by this dot?

3. How was this information obtained? Describe the sequence of steps in the sampling process that produced the information shown by the darkened dot.

Figure 3. Sampling distribution and accompanying prompts of the opening activity of Lesson 10.

Discussions around this demonstration centered on having students track and explain the process of how the dot plot of the sampling distribution resulted from the sampler in terms of the various intermediate objects produced by the simulation and shown in the *TinkerPlots* window, as displayed in Figure 2. This activity aimed to help students build and solidify their imagery of the repeated sampling process and their meaning for the resulting sampling distribution displayed in Figure 3. The accompanying activity prompts also assessed the strength and robustness of students' imagery by having them work backwards from a particular point on the dot plot and explain what it represented and the process that produced it. The activity prepared students for the subsequent part of Lesson 10, which required that they be able to decode and interpret a sequence of such dot plots coherently.



II. Here is a sequence of distributions. Each one is of the difference between the median length of genetic and median length of normal fish for 50 samples of a given size selected at random from the pond containing 625 fish.

Sample size = 15 *Sample size = 150* *Sample size = 250*

4. Compare these distributions for the various sample sizes. What do you notice about these distributions as sample size increases?

5. What is a big enough sample of fish for the farmer to pick from the pond in order to confidently test the company's claim that genetic fish tend to grow longer than normal fish? Please explain why you think this.

6. Estimate how much longer the genetic fish in the pond tend to be than the normal fish. How confident are you about this estimate? Please explain.

Figure 4. Task prompts and a subset of the distributions of the final activity of Lesson 10.

In the second part of Lesson 10 students examined and interpreted a sequence of five distributions of the difference in median lengths, each for 50 simulated samples of a different size drawn from the fish population. Students examined and interpreted these sampling distributions in relation to the increase in sample size. A subset of these distributions and the accompanying activity prompts are displayed in Figure 4.

The first prompt (Question 4) aimed to orient students' attention to the fact that the clustering of the sample statistic becomes increasingly compact (its variability decreases) with increasing sample size. A group discussion of this observation ensued which involved eliciting students' ideas about how to describe and measure the pattern of observed variability. This discussion was followed by prompting students to use this pattern as a basis for choosing a sufficiently large sample in order to confidently infer whether genetically engineered fish tend to grow longer than normal fish in the population (Question 5). The final prompt (Question 6) asked students to estimate *how much longer* genetic fish tend to grow than normal fish. These questions culminated in a group discussion about the trade-off between the competing interests of maximizing sampling accuracy and minimizing sample size.

Three days after Lesson 10, students took an exam designed to assess their thinking and conceptions of the ideas addressed in the instructional sequence. Students responded to a set of questions nearly identical to those from Lesson 10, but couched in a different story context (i.e., testing whether a genetically modified variety of cucumbers tended to grow longer than a normal variety). Follow-up interviews conducted within a week of the exam queried students' responses to this transfer task.

SUMMARY OF KEY FINDINGS AND CONCLUSION

Due to space considerations we only summarize a few salient aspects of our study regarding the influence of the instructional sequence on students' thinking *as a whole*, inferred from our initial analysis of the classroom discussions that emerged around the sequence:

- Class discussions proved to be productive vehicles for helping students notice and reflect on key ideas. The whole-group discussions around the orienting prompts and reflection questions of the activity sequence turned out to consistently provide rich opportunities for students to notice patterns of dispersion in the distributions that resulted from re-sampling simulations. As a particular case in point, the discussions around the activity of Lesson 9 (see Figure 2) highlighted several students' observations that "the samples [median lengths] don't vary by much" and that the median length of genetically engineered fish in a sample was larger than that of the normal fish in every sample chosen. This last was taken as evidence, and led to a consensus among students, that the genetically engineered fish in the population tended to grow longer (by between 4 and 7 centimeters) than normal fish.

- Students generally exhibited an appreciation of the tension between two seemingly competing ideas: although an individual random sample can be used to make an inference about the sampled population, repeating a sampling experiment shows that sampling outcomes (a statistic's values) vary from sample-to-sample. Significantly, a majority of the students seemed to readily appropriate the idea (promoted in instruction) of using a distribution of multiple sampling outcomes for samples of a particular size as a basis for informally assessing their level of confidence in their inference about the population.

We also examined students' written responses to the task prompts of Lesson 10, as well as their explanations of the responses to the transfer task that they provided in the exit interviews. These data indicated that many students tended to experience two common and related challenges:

- Difficulty understanding what the distributions of the sample statistic—difference between medians lengths—represented in terms of the scenario(s) in which they were embedded. Students often unwittingly interpreted distributions like those shown in Figure 3 as displaying either lengths or values of the median length of fish. Our prompts for elaboration during class discussions and in the exit interviews suggested the following source of their slippage:

- Difficulty keeping mental track of, and describing, the multi-tiered simulated re-sampling process (i.e., conceiving of the coordinated sequence of actions and resulting mathematical objects) that began with an identifiable simulated population, involved a random sampling process, and culminated in the generation of a distribution of the appropriate sample statistic.

These two challenges played out in students' ability (or inability, in a number of cases) to make a coherent and distribution-based inference about the sampled population: it was often the case that those students who experienced the above challenges most severely also tended to experience difficulties in either deciding on a big enough sample size for making a confident inference about the underlying population, or deciding on a level of confidence for a particular sample size. On the other hand, there was some evidence that the ability to hold in mind a sustained and clear image of the simulated re-sampling process and its various levels supported successful students' ability to make such decisions.

These findings underscore challenges regarding the design of instruction for supporting the development of informal distribution-based inferential reasoning, and they orient us to consider possible refinements of our instructional activities. One possible refinement currently within our sights is inspired by Harel's repeated reasoning instructional principle (Harel & Koichu, 2010). The refinement would consist of an elaboration of the activity shown in Figure 3 designed to have students practice imagining and explicitly identifying and distinguishing the processes and intermediate objects generated by the re-sampling scheme that produces the distributions of the sample statistic shown in Figures 3 and 4. The aim of such an elaboration would be to help students build stable and robust mental images of the re-sampling scheme and its products, so that they would be better positioned to hold the ensemble of coordinated ideas in mind as they attempt to make distribution-based inferences and quantify their confidence in such inferences.

REFERENCES

Finzer, W. (2012). *Fathom Dynamic Statistics (version 2.1)*. [Computer program]. Emeryville, CA: Key Curriculum Press.

Harel, G., & Koichu, B. (2010). An operational definition of learning. *The Journal of Mathematical Behavior*, 20(*3*), 115-124.

Key Curriculum Press (2011). Fish-Length Distributions. Retrieved May 2011, http://www.keypress.com/x26493.xml.

Konold, C. (2005). *Exploring data with TinkerPlots*. Emeryville, CA: Key Curriculum Press.

Konold, C., & Miller, C. D. (2011). *TinkerPlots: Dynamic Data Exploration (version 2.0)*. [Computer program]. Emeryville, CA: Key Curriculum Press.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning.* New York: Springer.

Rubin, A., Bruce, B., & Tenney, Y. (1990). *Learning about sampling: Trouble at the core of statistics.* Paper presented at the Third International Conference on Teaching Statistics, Dunedin, New Zealand.

Saldanha, L. A., & Thompson, P. W. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of student's engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education, 2*(3).

Zieffler, A., & Garfield, J. (2007). *Studying the role of simulation in developing students' statistical reasoning*. Paper presented at the International Statistical Institute, 56th session.