

STATISTICAL REASONING WITH THE SAMPLING DISTRIBUTION

¹Bridgette L. Jacob and ²Helen M. Doerr

¹Onondaga Community College, Syracuse, New York USA

²Syracuse University, Syracuse, New York USA

jacobb@sunyocc.edu, hmdoerr@syr.edu

Statistical reasoning surrounding the sampling distribution is necessary for formal statistical inference. Our study of introductory statistics students aged 16-18 suggests that knowledge of the characteristics of a sampling distribution and experiences with generating sampling distributions do not provide a sufficient basis for this reasoning. Following instruction on the sampling distribution and its characteristics, the students in this study were given the opportunity to sample before drawing an informal conclusion based on a sampling distribution. The majority of the students took several samples and/or considered generating a second sampling distribution for comparison. These are not incorrect statistical notions; and they do provide insight into possible missing elements in instruction. The results are discussed in relation to the development of students' statistical reasoning throughout their introductory statistics course.

INTRODUCTION

Recent research efforts in statistics education have focused on informal statistical inference to understand how students begin to reason about data. Makar and Rubin (2009) defined informal inferential reasoning as generalizing about a population using sample data as evidence while recognizing the uncertainty that exists. This focus on informal inferential reasoning as students take part in informal statistical inference tasks has been under investigation for the last decade (Ben-Zvi, 2004; Pfannkuch, 2006; Pratt, Johnston-Wilder, Ainley, & Mason, 2008). While researchers are building definitions of informal inferential reasoning and frameworks for researching its development (Makar & Rubin, 2009; Pfannkuch, 2006; Zieffler, Garfield, DelMas, & Reading, 2008), exactly how informal inferential reasoning develops and how students demonstrate such reasoning across a range of contexts is still under investigation. The research reported here was part of a larger study designed to add to the understanding of that development by investigating how secondary students demonstrated informal inferential reasoning throughout a year-long course in introductory statistics in the United States. The focus of this paper is introductory statistics students' reasoning when asked to make an informal statistical inference in relation to a sampling distribution. This will be discussed with respect to their statistical reasoning prior to and following their work with the sampling distribution.

BACKGROUND

A conceptual understanding of the sampling distribution is necessary for formal statistical inference. Saldanha and Thompson (2002) designed a study to examine students' developing ideas about repeated sampling and the sampling distribution while they participated in instruction on the topic. Their study was based on the knowledge gained from studies that found students focused on the statistics of samples, the sample mean, for example, rather than on how these statistics were distributed. Therefore, instruction for their teaching experiment stressed two themes: "1) the random selection process can be repeated under similar conditions, and 2) judgments about sampling outcomes can be made on the basis of relative frequency patterns that emerge in collections of outcomes of similar samples" (p.259). However, the majority of the students still compared a single sample statistic to the population parameter rather than to the sampling distribution of all such statistics when asked to determine if it was unusual. Saldanha and Thompson claim that a conception of sample in which the distinctions among the population, the individual samples taken from the population, and the distribution of many such samples are made, will help students to understand why statisticians can be confident in inferring about the population based on data from a single sample.

In a research project spanning seven years, Chance, delMas, and Garfield (2004) used interactive software designed to assist students in making these distinctions among the population, samples, and the sampling distribution. Some of the common misconceptions held by the students

in their studies were: the belief that the sampling distribution should look like the population; predicting that sampling distributions for small and large sample sizes have the same variability; the belief that sampling distributions for large samples have more variability; and a lack of understanding that a sampling distribution is a distribution of sample statistics (p. 302).

These studies together demonstrate the difficulties students encounter when introduced to the sampling distribution. Many of these difficulties surround students' misunderstandings of the effects of sample size and, thus, the variability of the sampling distribution. In addition, once students have an initial perception of the sampling distribution, placing the sampling distribution in relation to the population and individual samples is problematic for students.

DESIGN AND METHODOLOGY

The students taking part in this study were enrolled in introductory statistics courses for college credit in their high schools. These students were either in their 11th or 12th grade year, 16 to 18 years of age. They had completed at least one of the two algebra courses required for high school graduation in New York State. Following the principles proposed by Goldin (2000) seven pairs of students that participated in a sequence of four task-based interviews, spanning their progress over their introductory statistics course. Each pair came from one of seven statistics classes taught by four different high school mathematics teachers from two high schools. These statistics classes met for approximately three and one-half hours each week for the 40-week school year beginning in September, 2011. The structure of the task-based interviews (Goldin, 2000) included explicit interview protocols allowing students to think about their responses without critiquing for correctness. The interview tasks were designed with multiple parts that increased in complexity.

When comparing distributions of class test scores in the first task-based interview, students primarily focused on the mean or median in drawing a conclusion as to which class scored better on a test (Jacob, 2013). For the majority of these students, as they compared symmetrical distributions with the same mean and median, the equitable measures of center were more important than the differences in variability in drawing a conclusion. Students noticed the differences in variability between the two distributions but the majority of the students did not use that variability in concluding which class had scored better.

Students exhibited a clear understanding of the law of large numbers as they tossed Monopoly houses in the second task-based interview (Jacob, 2013). They estimated the probability that a house would land upright by taking samples.

The third task-based interview involving sampling distributions consisted of three parts. In the first part, students predicted what sampling distributions would look like for a given population distribution and considered the variability of those sampling distributions. Given a tri-modal distribution and its mean, students chose the distributions that represented 500 samples of size 4 and size 16 from five possible graphs (Chance, delMas, & Garfield, 2004).

In the second part of the interview, students viewed the Random Rectangle simulation in *Fathom*, shown in Figure 1, which was designed to help them understand distinctions among the population distribution, individual sample distributions, and the distribution of sample mean areas. The population of rectangles, labelled with their corresponding areas, is on the left and the graph of the areas of the total population of rectangles is in the upper middle. The Sample of Rectangles graph (in the lower middle) displays a single random sample of rectangles. The Measures from Samples of Rectangles graph in the lower right displays the sampling distribution of the mean areas. Students were able to watch a demonstration that animated how each sample was taken from the population, graphed, and then the mean area from each sample was added to build the sampling distribution.

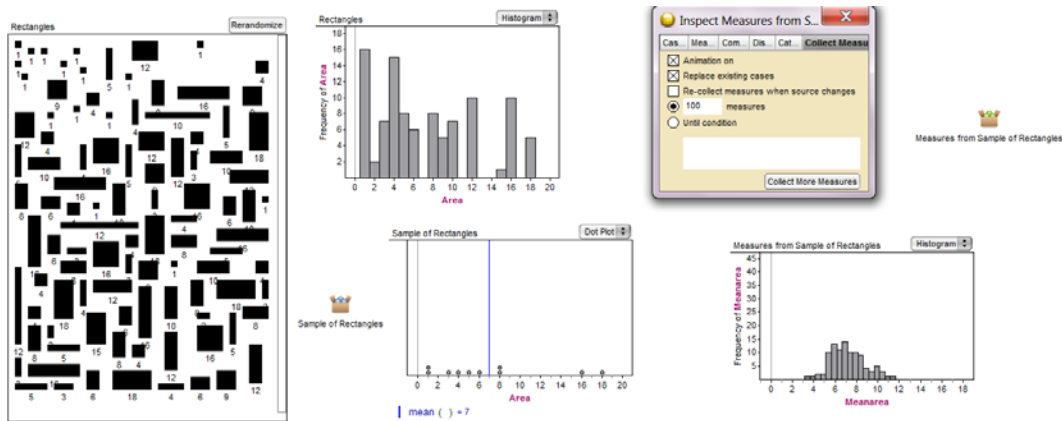


Figure 1: Screen shot of Random Rectangle simulation

The third part of the sampling distribution interview was influenced by the work of Saldanha and Thompson (2002) who found that even after instruction on the sampling distribution, students tended to compare the results from a sample to the distribution of the original population rather than to the sampling distribution. Therefore, students returned to the Monopoly houses and were asked to make an informal inference requiring them to situate a single sample in comparison to a sampling distribution generated from 200 samples.

In the fourth task-based interview on formal inference, the students were asked to construct a confidence interval for the proportion of red beads in a container of multi-colored beads and to conduct a formal hypothesis test. Similar to their sampling work with the Monopoly houses, they were able to draw a sample from the container with a paddle that captured 32 beads.

RESULTS

In this study, we found that most students had a base knowledge of the sampling distribution and its characteristics. In Part 1 of the sampling distribution task-based interviews when students were shown a tri-modal distribution, all seven pairs of interviewees chose sampling distribution graphs that were approximately normal in shape. Six of them also correctly identified the effect of sample size on the variability of the sampling distributions. Only one pair incorrectly identified the variability of the sampling distribution for a sample of size four; however, they did correctly identify the variability as less for the sample size of 16.

We also found that most students demonstrated an understanding of the probabilities and variability associated with the normality of these sampling distributions of mean areas. Following the *Fathom* demonstration the third interview, the students were shown three sampling distributions generated with this simulation for 100 samples of sizes five, 10, and 25 rectangles. When asked how the three distributions compared to one another, six of the pairs of students referred to these distributions as becoming more centered or having the same mean. Five of the pairs also referred to the decrease in variability as the sample size increased, as did this student:

Interviewer: So we went from a sample size of 5, then to 10, now to 25. So how about this one [of sample size 25]?

Student: This one's even more compact. The last one [of sample size 10] got all the way out to like 12. This one hasn't gone past 10 [referring to maximum mean area].

The remaining pair referred to the decrease in variability alone, mentioning the formula (σ/\sqrt{n}) for standard deviation in support of this decrease. Students were then asked what mean areas would be likely and which would be rare or unlikely for each of the sampling distributions. The students identified ranges of outcomes surrounding the peak of the approximately normal distributions as likely and those in the tails as rare.

In an effort to bring the characteristics of the sampling distribution and the concept of normality related to the sampling distribution together to make an informal inference, the students

were shown the sampling distribution in Figure 2. It was explained that this sampling distribution was generated by tossing 10 Monopoly houses 200 times, recording the proportion of houses landing upright. The students were then asked if they could determine whether the probability that a Monopoly hotel would land upright was the same as that for a house. The hotels were slightly larger with a rectangular rather than a square base like the houses. Both the houses and hotels were available for students to manipulate.

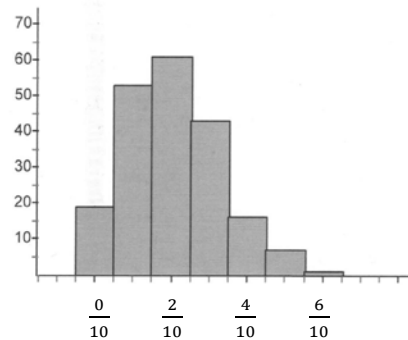


Figure 2: Sampling distribution for 200 tosses of 10 houses

Four of the seven pairs expressed that they would need to generate a sampling distribution exactly the same as the one they were shown for the houses with 200 tosses of 10 hotels. Another pair thought they would need to toss 32 hotels five times as they had with the 32 houses in the second task-based interview. Since time constraints did not allow for replicating the sampling distribution and there were not 32 hotels available, the pairs chose a variety of sampling methods. One pair tossed 10 hotels two times, another pair tossed 10 hotels three times, and a third pair tossed 10 hotels 10 times. Three other pairs also tossed 10 hotels 10 times, averaging their results to obtain a proportion for the hotels landing upright. The seventh pair tossed one hotel 10 times, choosing that method to reduce variability caused by the hotels bumping into one another as they were tossed. All of the pairs' tosses resulted in proportions that were at or close to the peak of the sampling distribution; however, their remarks demonstrated a variety of conclusions. At least one student from four of the seven pairs stated that the probability of a hotel landing upright was slightly different than that of a house. Additionally, four of these pairs expressed skepticism in the accuracy of their results due to their small number of tosses. They were not taking the natural variability of sample proportions into consideration even though they had just identified likely and rare outcomes with the sampling distributions of mean areas of rectangles. The majority of these students were not yet ready to draw a conclusion from a single sample of data using the variability of the sampling distribution.

However, their statements, many expressing a degree of certainty or uncertainty, provided evidence that they were at a point in their informal inferential reasoning when they might be able to consider this next level of reasoning. This could be seen when the following question was asked of the pair who tossed 10 hotels three times:

Interviewer: So is it [the results] enough less to say, do you think, that the probability is different?
 Student: How far away would it have to be? Like, I mean, I don't know, I think it would be a couple percentage points. You know just a little lower.

This student's question is at the foundation of formal statistical inference. We take the posing of this question as an indication that this student's understanding of what it means for these sample proportions to be relatively the same or slightly different is still developing.

The fourth task-based interview involved formal inference with the multi-colored beads. To form a confidence interval based on a sample, two pairs of students took 10 samples to find the sample proportion, \hat{p} , for their confidence interval, and one pair of students took three samples. These students were averaging to obtain their value for the sample proportion to use as their estimate. We take this as evidence that the students did not believe that one sample proportion

would be sufficient for constructing the confidence interval. Throughout the final interview, the pairs of students demonstrated their procedural knowledge in constructing confidence intervals and conducting hypothesis tests. However, when asked about the meaning of the percentage in the confidence interval, the pairs of students did not demonstrate an understanding that this value was based on the sampling distribution and its normality. For example, three of the pairs interpreted the 90% in their confidence interval as the chance the true population proportion of red beads would be in the interval; one pair interpreted it as the percentage of intervals constructed that would result in the exact same interval; two of the pairs stated that they did not know how to interpret the 90%; and the remaining pair was the only one to allude to 90% of multiple trials, but could not provide a complete explanation. When asked to define the p -value in their hypothesis test, six of the pairs described the procedure of comparing it to the significance level for drawing a conclusion. The remaining pair stated that the p -value was a probability, but both students were unsure of what this probability referred to or meant.

DISCUSSION AND CONCLUSIONS

The characteristics of the statistical reasoning students exhibited in the third interview on the sampling distribution had appeared in their reasoning in the first two task-based interviews. Some (but not all) of these characteristics carried through to the final interview on formal statistical inference. We take this as an indication that the students' reasoning had not yet fully developed so as to tackle the intricacies of drawing conclusions from the sampling distribution.

When students were asked to draw a conclusion based on a sampling distribution, the majority of them wanted to compare distributions by generating another sampling distribution for the proportion of hotels landing upright. Their statistical reasoning in this task resembled their prior reasoning when comparing a variety of distributions during the first task-based interview. In that interview, the students were primarily relying on comparisons of the measures of center. Therefore, comparing two sampling distributions was a viable option for determining if the hotels behaved similar to the houses. During the second interview, students took their own samples to approximate the proportion of houses landing upright when tossed, demonstrating an understanding of the law of large numbers. With this understanding, they also wanted to take many more samples to compare to the sampling distribution. A combination of these statistical reasonings likely presented a barrier for them in using the sampling distribution; they did not have confidence that they could draw a conclusion based on just one or even a small number of samples. This occurred after they had instruction on the sampling distribution in their regular classrooms and immediately after they exhibited understanding of the characteristics of the sampling distribution during the interview.

Our results suggest that the comparison of means provided these students with a higher degree of certainty than relying on the normality of the sampling distribution. Taking many samples with the hotels may have felt like a much more certain method for inferring about the probability that a hotel would land upright. While this was an attempt to reduce variability, which is an accurate intuition as well, using the known variability that exists in the sampling distribution of all such samples is a more efficient method. It is also possible that relying on a sampling distribution they did not create was viewed as another layer of uncertainty that they chose to avoid.

Whether students' difficulties in relying on the sampling distribution stemmed from not considering the sampling distribution a powerful tool generated from a large number of samples or whether they found more certainty in comparing measures of center, this prevented them from making the necessary connections for formal statistical inference as well. This same type of reasoning appeared during the final task-based interview on formal statistical inference when three of the student pairs took several samples for the proportion estimate to construct a confidence interval. This again was not incorrect statistical reasoning; they were trying to obtain an accurate estimate. It was evident, however, that they were primarily depending upon their procedural knowledge while completing the formal statistical inference task-based interview. Without a deep understanding of the power of the sampling distribution, relying on it for formal statistical inference is not possible. Therefore, as students learned the procedures for formal statistical inference, the role of the sampling distribution was likely not paramount. Students may have made little connection between the procedures for formal statistical inference and the underlying conception of the sampling distribution. All seven pairs were successful in creating a confidence

interval and conducting a hypothesis test based on their own samples. However, when asked about the meaning of the percentage in the confidence interval or the p -value, they did not demonstrate an understanding that these values relied on the sampling distribution and its normality. Without this connection, they could only depend upon their procedural knowledge.

It would be worthwhile to investigate how students would feel about the need for more samples in a task like the one with the Monopoly houses and hotels if they were able to generate another sampling distribution for comparison as many of the pairs initially wanted. Developing activities that allow introductory statistics students to explore their notions regarding samples and comparing them to the sampling distribution may support their informal inferential reasoning by helping them to understand that taking a larger number of samples will not necessarily provide more certainty in their conclusions; and this might better support their formal inferential reasoning. For the students in this study, it appeared that their statistical reasoning in working with the sampling distribution was still in development rather than entirely incorrect. Discussions about efficiency as well as accuracy in data collection might help students develop their statistical reasoning. Introductory statistics students likely do not have any practical experience working within budget or time constraints for using data to draw conclusions; therefore, creating a sampling distribution and/or taking many samples to compare and then make a decision may have seemed like the only certain method. Students likely need multiple experiences to help them understand that one sample can be enough to draw an inference about a population.

REFERENCES

- Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42-63. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ3\(2\)_BenZvi.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_BenZvi.pdf)
- 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). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Finzer, W. (2001). *Fathom Dynamic Data Software*. (Version 2.1) [Computer Software]. Emeryville, CA: Key Curriculum Press.
- Goldin, G. A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 517–545). Mahwah, NJ: Laurence Erlbaum Associates.
- Jacob, B. L. (2013). *The development of introductory statistics students' informal inferential reasoning and its relationship to formal inferential reasoning* (Doctoral dissertation). Retrieved from http://surface.syr.edu/tl_etd/245
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ8\(1\)_Makar_Rubin.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ8(1)_Makar_Rubin.pdf)
- Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27-45. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Pfannkuch.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Pfannkuch.pdf)
- Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, 7(2), 107-129. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Pratt.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Pratt.pdf)
- Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51(3), 257-270.
- Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Zieffler.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf)