STUDENTS' INFORMAL INFERENTIAL REASONING WHEN WORKING WITH THE SAMPLING DISTRIBUTION

Bridgette Jacob and Helen M. Doerr

Syracuse University

Introductory statistics students struggle with the complexities of the sampling distribution, particularly when asked to draw a conclusion based on a sample of data. This study investigates the informal statistical reasoning of seven pairs of secondary students as they collect a sample of data in an attempt to draw a conclusion based on the related sampling distribution. We present analysis of this task and the results of a pre/posttest assessment of their informal inferential reasoning with the sampling distribution.

INTRODUCTION

Recent research efforts in statistics education have focused on informal statistical inference to understand how students begin to reason about data they encounter. Informal inferential reasoning is thought to develop as it brings together the underlying concepts of descriptive statistics, probability, and the sampling distribution to infer about populations before the procedures of formal inference are introduced. 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; Watson and Moritz, 1999). 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. This research was designed to add to the understanding of that development by investigating how secondary students demonstrated informal inferential reasoning at the end of a year-long course in introductory statistics in the United States. In particular, we were interested in two questions: (1) what knowledge of the sampling distribution do students exhibit and (2) could students make an informal statistical inference by situating a single sample in relation to the related sampling distribution.

BACKGROUND

Even with instructional activities designed to help students develop an understanding of the complexities of the sampling distribution such as those implemented in the study by Saldanha and Thompson (2002), students are likely to experience difficulty in drawing informal conclusions with the sampling distribution. The majority of secondary statistics students in Saldanha and Thompson's study compared a single sample statistic to the population parameter rather than to the sampling distribution when asked to determine if it was unusual. Many of the difficulties students experience with the sampling distribution surround their misunderstandings of the effects of sample size and thus, the variability of the sampling distribution. Chance, delMas, and Garfield (2004) found this when using interactive software and paper-and-pencil tasks designed to assist students in making the distinctions between the population, samples, and the sampling distribution. For the introductory statistics students in their study, placing the sampling distribution in relation to the population and individual samples was problematic. The difficulties students have with the complexities of the sampling distribution and drawing conclusions with them have the potential to plague them further as they attempt to navigate formal statistical inference.

The definition proposed by Makar and Rubin (2009) was used to determine if students reported in this study were demonstrating informal inferential reasoning. The larger study, of which the sampling distribution task reported here is a part, was modified from a task framework for research on informal inferential reasoning developed by Zieffler, Garfield, delMas, and Reading (2008).

DESIGN AND METHODOLOGY

This research was part of a larger research project that examined the relationship between students' informal inferential reasoning and their formal inferential reasoning as this developed over a year-long course in introductory statistics at the upper secondary level. In this paper, we focus on students' reasoning about the sampling distribution as they completed a task designed to engage them in informal inferential reasoning. This task was the third in a sequence of four task-based interviews (Goldin, 2000) completed by seven pairs of students. We will also discuss their responses to the sampling distribution questions on a pre/posttest they completed as part of the larger study. The posttest included the same informal inferential reasoning questions as the pretest as well as formal inferential reasoning questions.

Setting and Participants

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, and had completed at least the first two courses of the three mathematics courses required for high school graduation. These students came from one of eight 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.

A pair of students from each of the eight classes was asked to take part in the task-based interviews; seven of the student pairs completed the study. The pairs of students represented a range of prior achievements in mathematics as judged by their classroom teachers.

Task-based Interviews

The structure of the task-based interviews followed the principles and techniques proposed by Goldin (2000). These included task-based interviews that (1) were designed specifically to answer the research questions, (2) included tasks with appropriate content for students' to grasp, (3) were structured based on key statistical concepts that gave students a variety of ways to demonstrate their understanding, (4) included an explicit interview protocol that allowed students to think about their responses without critiquing the correctness of their responses, and (5) involved students in free problem solving while they interacted with another student. The interview tasks were designed with multiple parts that increased in complexity.

Three classroom activities were implemented in each of the classes prior to the task-based interviews to provide the students with experiences in informal inferential reasoning. The three activities took place in the following order in conjunction with the progression of the class curriculum: (1) comparing distributions of data (Watson & Moritz, 1999); (2) a sampling and probability exploration (Konold et al., 2011); and (3) a sampling distribution activity.

Following each of these three classroom activities, the task-based interviews were conducted with the seven pairs of students. These interviews began with a recall of the classroom activity to gain insight into what the students learned from the activity. They were then asked to complete another problem in that same topic area to probe how their informal inferential reasoning was developing. A fourth task-based interview focused on formal statistical inference.

The focus of the research reported here involved the third task-based interview. This interview task 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. The classroom activity and task in this study were designed to support what they called a "multiplicative" conception of sample where the distinction between the population distribution, the distribution of a single sample taken from the population, and the distribution of the sample statistics of many samples is understood. The interview tasks included a task used by Chance, delMas, and Garfield (2004) in which students identified sampling distributions based on a population distribution and answered questions about the variability of these sampling distributions. This interview task culminated with an informal inference task which required students to situate a single sample statistic in comparison to a related sampling distribution of many samples.

Pre/posttests

The study began with a pretest assessment to measure students' informal inferential reasoning. The study concluded with a posttest which contained both informal (identical to pretest) and formal statistical inference questions. There were two items that specifically addressed informal inferential reasoning with the sampling distribution. Students' responses to these items are included in this analysis.

DATA ANALYSIS

The task-based interviews were video-recorded, transcribed, and coded for common themes in students' understanding of key statistical concepts. For the Sampling Distribution task-based interviews, we analyzed students' responses to determine the extent to which they (1) distinguished between the sampling distribution and the population distribution, (2) recognized the effects of sample size on the variability of the sampling distribution, and (3) drew an informal conclusion by situating a single sample in relation to the corresponding sampling distribution.

Students' responses to the sampling distribution items on the pretest and posttest were compared to determine if students improved on these items. Each item had two parts and students' responses to both parts were analyzed together to determine if students displayed consistent reasoning.

RESULTS

We will first report on the findings from the task-based interview on sampling distributions. This is followed by the results of the four pre/posttest sampling distribution questions for the seven pairs of students.

Sampling Distribution Task-based Interview

Part 1 of the third task-based interview involving sampling distributions had students predicting what sampling distributions would look like for a given population distribution and considering the variability of those sampling distributions. Given a tri-modal distribution and its mean (Chance, delMas, & Garfield, 2004), students chose the distributions that represented 500 samples of size 4 and size 16 from five possible graphs.

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. This was an indication that these students had a base knowledge of the sampling distribution and its characteristics.

To begin the second part of the task-based interview, students viewed the Random Rectangle simulation in *Fathom*, shown in Figure 1. 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 below that in the center 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.

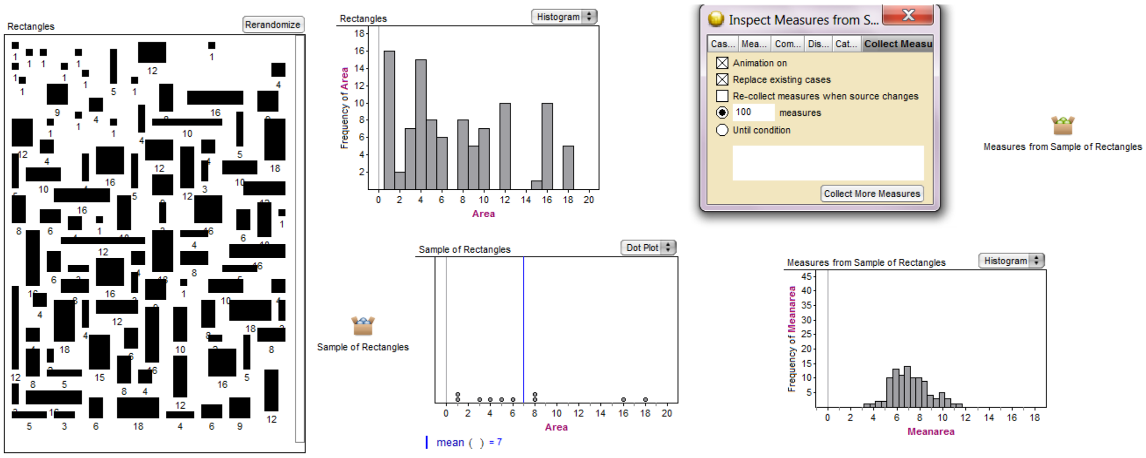

Figure 1: Screen shot of Random Rectangle simulation

Following this demonstration, the students were shown three sampling distributions generated from this simulation for 100 samples of sizes five, 10, and 25 rectangles. When asked about how all three distributions compared to one another, six of the pairs referred to these distributions as becoming more centered or having the same mean with five of them also referring 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 $\left(\frac{\sigma}{\sqrt{n}}\right)$ 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 had no difficulty in identifying ranges of outcomes surrounding the peak of the approximately normal distributions as likely and those in the tails as rare. They demonstrated an understanding of the probabilities and variability associated with the normality of these sampling distributions of mean areas.

In an effort to bring the previous concepts of normality and variability related to the sampling distribution together to make an informal inference, the interview concluded with the sampling distribution in Figure 2. In the second interview of the larger study, students tossed small plastic houses to approximate the probability that a house would land upright when tossed. Students were shown this sampling distribution which was generated from 200 samples of 10 houses tossed, recording the proportion of houses landing upright. The interviewees were then asked if they could determine whether the probability that a 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; and both the houses and hotels were available for students to manipulate.

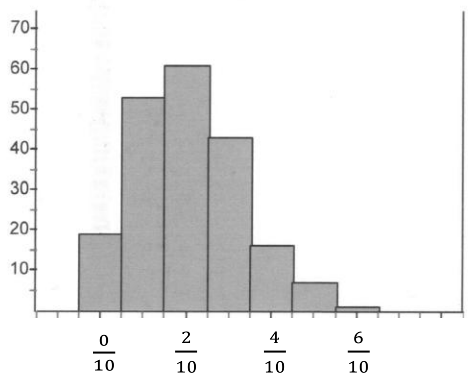

Figure 2: Sampling distribution for 200 tosses of 10 houses

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 houses in the second task-based interview.

Since time constraints did not allow for replicating the sampling distribution, all of the pairs tossed 10 hotels. Table 1 displays the number of tosses by each pair and some of their concluding remarks. Three of the pairs tossing 10 times averaged their 10 tosses to obtain a proportion for the hotels landing upright.

Number of Tosses	Students' Concluding Remarks
10 (averaged)	"Like we had more two's and it looks like this one has more two's. So I feel like it would have the same probability as the house." "I'm still doubtful. What we found was about 25%. So about a fourth of the time it'll land upright, if not a little bit more than that. And for this [the sampling distribution] we have like 20%, less than 20%, so that's just me doing math in my head and I just don't think it's likely. Plus we only did 10 trials."
2	"I think you'd have to try probably more times, many more times, but, as it looks right now it's about the same." "Maybe a little bit less."
1 hotel 10 times	"Yeah, it was pretty similar. Between 1 and 2 [houses landing upright out of 10] so I think that [their results] still validates that that's relatively the same."
3	"I'm figuring it's not going to be that far away. I think it's going to be roughly the same. It's maybe just a little bit less because it's weighted differently."
10 (averaged)	"So it would probably look normal [sampling distribution for hotels] but not as variable as the other one [sampling distribution for houses]." "I'd say they're very similar, the two, yeah."
10	"We didn't do nearly enough. I mean you did this 200 times, we did this 10 times so like you can't really say like, oh look what we did really quick and that refutes that."

	"Then I'd say it's different, but not by a lot."
10 (averaged)	"You gotta take more samples. …but with one trial, I think regardless of the outcome, you can't really compare that to what you got from this population [referring to sampling distribution]. It may fit into what you have seen. Like right here, this value right here, like 1, 2, [referring to peak in sampling distribution] ours was close so we could say yeah, it does compare similarly but I'm not going to bet my life on it."

Table 1: Number of tosses of hotels and students' concluding remarks

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 was slightly different. 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. Four of these pairs expressed their skepticism in the accuracy of their results due to their small number of tosses. The majority of these students were not yet ready to draw a conclusion from a single sample of data based on 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 we asked the following question of the pair who did three trials:

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 students' question lays the foundation for formal statistical inference. This may provide an opportune time to explore more deeply what it means for these sample proportions to be relatively the same or slightly different.

Sampling Distribution Questions on the Pre/posttest

There were two items on the pre/posttests involving drawing conclusions from a single sample based on the corresponding sampling distribution. The first item asked students to draw conclusions directly from a graph of the sampling distribution. For the second item, students were shown the population distribution and given the population mean. They were then asked to draw a conclusion based on the results of a random sample of size 50.

The first pre/posttest item had students drawing informal inferences based on a sampling distribution for the proportion of heads expected when a fair coin is balanced on its edge 10 times. Marked on the sampling distribution in Figure 3 were the results of 0.7 heads and 0.9 heads from two different samples.

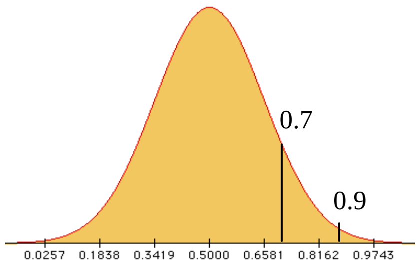

Figure 3: Sampling distribution for proportion of heads

In the first part of this item, students were asked if it was reasonable to conclude that the coin was fair with a sample proportion of 0.7 heads. Eight of the students answered correctly on the pretest that this result, which was between 1 and 1.5 standard deviations from the mean, was reasonable. This improved to 10 students answering correctly on the posttest; however, one student changed a correct answer on the pretest to incorrect on the posttest.

In the second part of the item, students were asked if it was reasonable to conclude that the coin was unfair with a sample proportion of 0.9 heads. Nine of the students answered correctly on the pretest and the posttest that this result, which was over 2 standard deviations from the mean, was reasonable. Two students changed their correct answers on the pretest to incorrect on the posttest.

Only six students answered both parts of this item correctly and one student answered both parts incorrectly. Of the seven other students, three concluded that both results indicated that the coin was unfair while the other four students concluded that both results indicated that the coin was fair. These responses are consistent with reasoning students displayed when working on the hotel task. This item, with the sample results of 0.7 and 0.9 clearly marked on the sampling distribution graph, provided further evidence that most of the students were not able to appropriately use that sample data as evidence as they were not fully considering the probabilities and variability associated with the sampling distribution.

For the second item on the pre/posttests shown in Figure 4, the students were shown a left-skewed distribution of exam scores for a particular exam. The average exam score for this population was 74 out of 100 points.

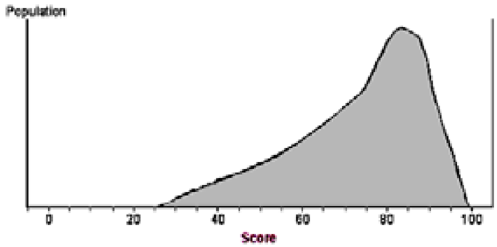

Figure 4: Population distribution of exam scores

In the first part of this item, the students were asked if a current group of 50 students with an average of 78 points did better on average than expected for this exam. Five students answered this part correctly on the pretest and seven answered correctly on the posttest. However, three students changed their responses from correct on the pretest to incorrect on the posttest.

The second part of this item asked the students if this higher sample average could just be due to chance. Nine students answered correctly that this higher score could be due to chance on the pretest and 10 answered correctly on the posttest. One student changed their answer from correct on the pretest to incorrect on the posttest.

Taking both parts of this item together, six of the students responded correctly and were consistent in their reasoning by answering that this score could not be considered better than what could be expected and that the higher sample average score was due to chance. Three other students were consistent in their incorrect reasoning by answering that this higher sample average could be considered better than what could be expected and that it was not due to chance. The remaining five students displayed inconsistencies in their reasoning. Four of these five answered incorrectly that this higher sample average could be considered better than what could be expected but also that it was due to chance. The incorrect and inconsistent responses to these two questions was further evidence that the majority of these students experienced difficulty in drawing conclusions based on the relationship between a population and the corresponding sampling distribution including the variability associated with it.

DISCUSSION AND CONCLUSIONS

Overall the students had general knowledge of the sampling distribution. They knew it took on the shape of a normal distribution and they made references to the decrease in variability as the sample size increased. They also identified sample data values that would be considered likely and rare based on probabilities associated with the normality of sampling distributions. However, when it came to making a decision about the hotels in the last part of the interview, they were not completely prepared to reference what they knew. The majority of them did not use the probabilities and variability associated with the normal distribution in drawing their conclusions and were not willing to compare a single sample proportion to the sampling distribution. Their responses to the posttest items provided evidence that these concepts were still creating difficulties for students at the completion of their introductory statistics course.

Based on students' responses during the houses/hotels sampling distribution task, the majority of them were expressing the tension created by results that would be likely but were also different as well as results derived from small samples. Referring to Makar and Rubin's (2009) definition of informal statistical inference, students were inferring informally; however, for the majority of them, the level of uncertainty they interpreted in the data they collected was preventing them from appropriately using their data as evidence. Fully exploring this tension and uncertainty with activities that give students the opportunity to collect a sample of data and compare it to the sampling distribution relative to the known probabilities associated with any normal distribution may help them understand how a single sample can be used to draw a conclusion 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

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.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N. J., & Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning, 13*(1), 68-86. doi:10.1080/10986065.2011.538299

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

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

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

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257-270.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*(2), 145–168.

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