Using Informal Inferential Reasoning to Develop Formal Concepts: Analyzing an Activity

Aaron Weinberg
Ithaca College

Emilie Wiesner
Ithaca College

Thomas J. Pfaff
Ithaca College

Abstract

Inferential reasoning is a central component of statistics. Researchers have suggested that students should develop an informal understanding of the ideas that underlie inference before learning the concepts formally. This paper presents a hands-on activity that is designed to help students in an introductory statistics course draw informal inferences about a bag of bingo chips and connect these ideas to the formal T-test and confidence interval. This activity is analyzed using a framework and recommendations drawn from the research literature.

1. Introduction

Drawing inferences from data is becoming an increasingly important component of teaching, both in recommendations for K-12 education (National Council of Teachers of Mathematics, 2000, 2009) and in undergraduate studies. As Pratt and Ainley (2008) noted, "inference is a foundational area in statistics" (p. 3). However, students at all levels have difficulty with the ideas related to inference. Garfield and Ahlgren (1988) described the views of college faculty members, who think that most students who take introductory statistics have a poor conception of statistical concepts; instructors believe that, while college students may be able to complete formulaic problems, they have little sense of "what the rationale is or how concepts can be applied in new situations" (p. 46). These views are supported by the research literature, which reveals that students have poor understandings of sampling distributions and difficulty understanding the logic behind hypothesis tests and the meaning of related concepts such as null hypotheses and p-values (Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007). In order to help students understand the concepts of formal statistical inference, researchers have advocated developing students' informal inferential reasoning. For example, Ben-Zvi (2006) noted: "Integration and cultivation of informal inference and informal argumentation seem to be essential in constructing students' statistical knowledge and reasoning in rich learning contexts" (p. 2). This suggests that educators should create pedagogical tools that will help students develop informal inferential reasoning and connect this reasoning to the formal concepts. This article describes and analyzes an activity aimed at these goals that is designed to be used in an introductory undergraduate statistics course.

2. Informal and Formal Inferential Reasoning

Pfannkuch (2007) described informal inference as "the drawing of conclusions from data that is based mainly on looking at, comparing, and reasoning from distributions of data" (p. 149). Similarly, Zieffler, Garfield, delMas, and Reading (2008) described it as "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (p. 44); and as "a process that includes:

- Reasoning about possible characteristics of a population (e.g., shape, center) based on a sample of data;

- Reasoning about possible differences between two populations based on observed differences between two samples of data (i.e., are differences due to an effect as opposed to just due to chance?); and

- Reasoning about whether or not a particular sample of data (and summary statistic) is likely (or surprising) given a particular expectation or claim" (p. 45).

In contrast, formal statistical reasoning generally involves working with theoretical constructs such as *p*-values and confidence intervals. For example, formal reasoning for a one-sample *T*-test would involve understanding the relationship between a sample mean, a null hypothesis, and the associated sampling distribution (Zieffler et al. 2008, p. 45).

Several researchers (e.g. Mooney, 2002; Jones, Langrall, Mooney, & Thornton, 2005) have described aspects of informal inferential reasoning, and others (e.g. Pfannkuch, 2005, 2006a, 2006b, 2007; Rubin, Hammerman, & Konold, 2006; Watson, 2008) have constructed frameworks for understanding this reasoning. However, there is no universal agreement about the components of inferential reasoning. The activity described here builds on the framework of Zieffler et al. (2008). Their framework suggests that tasks that are designed to develop students' informal inferential reasoning must challenge students to:

1. Make judgements, claims, or predictions about a population based on samples, but not using formal statistical procedures and methods (e.g. *p*-value, *t* tests);

2. Draw on, utilize, and integrate prior knowledge (formal and informal) to the extent that this knowledge is available; and

3. Articulate evidence-based arguments for judgments, claims, and predictions about populations based on samples (pp. 46-47)

Although an informal understanding of the ideas underlying statistical inference is an important precursor to understanding formal reasoning, Rubin, Hammerman, and Konold (2006) noted that there is no research that shows how this development occurs (p. 1). Pfannkuch (2006a) described current research as just beginning to describe the pathway from informal to formal inferential reasoning. Schwartz, Sears, and Chang (2007) suggested that students' informal knowledge should be developed through the use of activities that motivate and lay the foundation for later formal instruction. Other researchers have recommended the use of hands-on sampling activities to develop fundamental ideas of informal inferential reasoning by focusing on distributions of collections of samples (e.g. Saldanha & Thompson, 2002).

Activities that involve sampling may provide students with the opportunity to develop informal ideas and connect them with formal ones. However, the way the activity is implemented is important. In a study that used simulations to help students understand aspects of sampling distributions, delMas, Garfield, and Chance (1999) noted that simply

pointing out important features of the distributions did not improve their students' conceptual understanding, and the activities needed to be structured to help students connect their conjectures, observations, and the concepts. Similarly, Lipson (2002) used a simulation of the sampling process and found that the sampling alone failed to link the empirical and theoretical aspects; this linking needed to be done by the student. Gnanadesikan, Scheaffer, Watkins and Witmer (1997) and Watson (2002) noted that cognitive conflict and reflection—such as when students make conjectures, explain solutions, and reflect on their results—can help students make connections and develop ideas about inference.

3. Situating the Activity

The activity is designed to help students begin to understand concepts related to the T-test and confidence interval; it uses the recommendations of Zieffler et al. (2008) to develop informal reasoning, connect this reasoning to formal ideas, and encourage students to conjecture, explain, and reflect. Specifically, the goals of the activity are to have students make inferences about populations by drawing samples, introduce them to the mechanics of confidence intervals and hypothesis testing, and help them develop an understanding of the following ideas:

- *Samples vary naturally*: Different samples from the same population will likely produce different statistics. Consequently, differences between samples, or between a sample and a hypothesized value, may be due to chance. Hypothesis tests and confidence intervals are tools for helping us measure the amount of variation we expect to see and to draw conclusions about the parameter. Sometimes the natural variation between samples results in constructing a confidence interval that doesn't contain the parameter or a p-value that leads to drawing an incorrect inference.

- *The meaning of a hypothesis test*: A hypothesis test investigates how likely a statistic is given a particular hypothesis about the parameter, and the p-value measures this likelihood. The farther the sample statistic is from the hypothesized parameter, the smaller the p-value is, and the more likely we are to believe that our hypothesis may have been incorrect.

- *The meaning of a confidence interval*: The sample statistic will likely not be equal to the population parameter, but most statistics will be close to the parameter. Specifically, if most sample statistics are within a certain distance from the population parameter, then the parameter is likely to be within this same distance of most statistics. The sample statistic lies at the center of the T-confidence interval, and the width of the interval varies with the level of confidence, the sample size, and the amount of

variation within the sample. Due to sampling variability, the interval may not contain the parameter; when it does, the parameter may be anywhere within the interval.

Ideally, students in the course would have several weeks to develop the informal ideas related to confidence intervals and hypothesis tests before connecting them to formal concepts. However, introductory courses may have a significant service component, which requires the instructor to cover numerous topics during the semester or quarter. Consequently, this activity is designed as a collection of nine "modular" sub-activities: there is one warm-up homework, three informal in-class worksheets, a follow-up informal homework, two formal in-class worksheets, and two formal homework worksheets (these can be found in Appendix A). The informal activities require approximately one 50-minute class period and the formal activities (and related discussion) require approximately two 50-minute periods. The modular nature enables the instructor to omit either the confidence interval or hypothesis test component depending on the time available and the goals of the course; the instructor can also streamline the activities by omitting questions that don't address their course goals (e.g., questions dealing with the effects of sample size).

The activity is intended to be used in a class with at least 25 students, but the instructor can supplement the class-generated data with simulated data. These worksheets are designed to be used as the basis for small-group or whole-class discussion and reflection. In addition, the instructor has the option of using the out-of-class reflections and predictions as the basis of an in-class discussion (discussion prompts can be found in Appendix B).

The activity assumes that students have already developed an understanding of several statistical concepts. They need to understand the process of taking a sample from a population and understand what is represented by a histogram of sample means. Related to this, students need to understand the difference between a parameter and a statistic. Finally, students need to understand what standard deviation measures and that the mean of a large sample is more likely to be close to the parameter than the mean of a small sample (i.e., the law of large numbers).

4. The Activity

The worksheets for the homework and in-class components can be found in Appendix A. Suggested discussion prompts can be found in Appendix B and examples of class data and graphs can be found in Appendix C.

In the activity, students investigate the average winnings for a lottery through repeated sampling; their goal is to determine whether or not they would expect to break even in the

long run if they play the game. To create the lottery, each student receives an opaque bag filled with colored bingo chips, with each color corresponding to a different dollar amount. Students are told that it costs $5 to draw a chip from the bag, but they will win the dollar amount corresponding to the color of the chip they draw. The only stipulations are that they aren't allowed to look in the bag and they need to draw chips with replacement.

The activity begins with a warm-up homework in which students make predictions about the class' results. In the following class period, students draw samples and make informal conclusions about the lottery. Then, the class discusses the underlying ideas of confidence intervals and hypothesis tests; these ideas are informal, and are based on the students' observations and predictions. Following this discussion, students complete a homework set in which they use the class' samples to further investigate these informal ideas. After this informal introduction, the class is introduced to the formal ideas of confidence intervals and p-values through in-class discussion and homework.

4.1 Informal Component

In the warm-up homework, students are given a series of questions that prompt them to describe the population of interest, sample, parameter, and statistic. They are asked to think about the amount of variation they might see between samples, and then predict how large and how small they would reasonably expect a 40-chip average to be if the parameter is equal to $5. Then, students are asked to predict what the histogram of the class' sample averages will look like.

In class, the students discuss their predictions and answers with each other in small groups and the instructor leads a class discussion. Then, each student samples 40 chips (with replacement) and computes the sample mean and standard deviation. Based on this empirical result, the students draw an informal conclusion about the lottery and compare their results with their classmates to get an intuitive sense of the variation in the class' data.

Next, either in small groups or in a whole-class discussion, the students explore the (informal) ideas of a confidence interval and hypothesis test, and they connect these ideas to their intuition and experience. Students use their own statistic to predict a range of values that they think will contain the parameter, compare the width of this interval to their predictions from the warm-up, and think through how their interval might change if they altered the sample size, sample variance, or increased their confidence. Students then use the histogram they described in the warm-up homework to estimate the likelihood of getting their statistic if they assumed that the parameter was $5.

To set-up the homework, the students enter their 40-chip means into a class spreadsheet

and make a histogram (see Appendix C for an example).[1] After completing the homework, the instructor leads a class discussion to help them compare their ideas. The homework enables students to further investigate and reflect on the informal ideas from class using the collective class results:

- *Samples vary naturally*: Students describe the variation in the class' averages, the shape of the distribution, and explain why it makes sense for the sample means to be clustered relatively close to the parameter, making the distribution roughly symmetric and unimodal. Then, they describe how the histogram would change if the sample size were to increase.

- *The meaning of a hypothesis test*: Students estimate which of the class' statistics would be most and least likely to occur if the parameter was equal to $5. Then they describe which of these statistics would lead them to question the assumption that the parameter was $5.

- *The meaning of a confidence interval*: Students compare their statistic to the parameter and estimate an empirical 80%[2] confidence interval. The students then use the width of this interval to make new intervals that are centered at each person's statistic and determine which of these intervals would contain the parameter.

4.2   Formal Component and Connections

Following the informal introduction, students begin investigating the confidence intervals and hypothesis tests formally. Throughout this process, the instructor helps them make connections between the formal concepts and their previous predictions, reflections, and discussions.

The instructor introduces students to the techniques and vocabulary of confidence intervals and shows the students how to use a calculator or computer to compute an 80% theoretical confidence interval and the margin of error. Then the students compare this interval to their previous estimates. The students describe the meaning of the margin of error, investigate how changing the sample size, sample variance, or confidence level affects the interval,

---

[1]It is essential to have enough data points so that the histogram will look approximately Normal and so there will be enough students who draw an incorrect conclusion. If there are fewer than 25 students in the class, it may be useful for the instructor to include extra data points from a simulation, for students to run multiple trials, or to combine results from several class sections.

[2]In order to facilitate a productive discussion of the meaning of "n% confidence," it is helpful to have approximately $100 - n$% of the class' intervals fail to contain the parameter. If the class size is not large, using 80% confidence will usually produce enough intervals so that students believe that roughly 20% of the intervals do indeed fail to contain the parameter.

and explain the meaning of 80% confidence (i.e., if 80% of the sample statistics are within a certain distance of the parameter, then the parameter is within this same distance of 80% of the statistics). Then, the instructor introduces the idea of a formal null hypothesis and p-value and shows how to compute these; students compare these formal ideas to their previous discussions and use their p-value to draw a conclusion about the lottery.

After the in-class activity and discussion, the students add their confidence intervals and p-values to the class' spreadsheet. The instructor makes a graph of the confidence intervals and makes the data and graphs available to the students (see Appendix C for an example).

The students complete a homework assignment and discuss their answers in class. These questions ask students to describe and analyze the class' data and reflect on the following ideas:

- *Samples vary naturally*: Students look at the class' confidence intervals and discuss whether any of them seem problematic, distinguishing between computational issues (e.g., an interval that looks too wide or narrow) and natural variation. The students also discuss how some of their classmates' statistics result in unusually low or high p-values; some students will have (correctly) drawn the incorrect conclusion about the lottery, and all students are asked to discuss whether these students made a mistake.

- *The meaning of a hypothesis test*: Students examine the relationship between their classmates' statistics and the hypothesized mean and then describe what the p-value seems to indicate about this relationship. Then they examine what the p-value indicates about the relationship between the confidence intervals and the hypothesized mean.

- *The meaning of a confidence interval*: Students determine the percentage of the confidence intervals that contain the parameter and where the parameter tends to be located inside the intervals. Then, they describe where the statistic tends to be located in the intervals, how changing the confidence level affects the intervals, and how this affects whether the intervals contain the parameter.

After completing the homework, the students can participate in a class discussion in which the instructor helps them solidify the connections between their informal ideas and the formal concepts. The discussion will help students reconcile the differing p-values and confidence intervals by discussing what they mean and how they are affected by sampling variation. In particular, the students can discuss Type I and Type II errors and their practical ramifications.

5. Analysis

This activity follows Zieffler et al.'s (2008) framework for informal inference. Students create distributions through hands-on sampling and then make judgments, claims, and predictions about the underlying population. For example, students establish a range of values that they think a $5 lottery bag might produce in a sample, and then draw an informal conclusion about the composition of chips in the bag based on their empirical sample in class. Students build on prior informal knowledge to make these predictions and are prompted to draw on their more formal understanding of concepts such as the law of large numbers to make predictions about the variability in the class' distribution. By reflecting on their "minimum" and "maximum" values and the predicted variation, they draw conclusions and articulate reasons for these conclusions based on the samples.

Following Schwartz et al.'s (2007) recommendation, this activity develops the students' informal knowledge in a way that motivates the subsequent formal instruction. The students are prompted to make a conclusion about their original assumption (that the bag pays $5), estimate the likelihood of getting their sample statistic based on this assumption, and predict a range of values that the true average might lie within; these are the informal equivalents of a hypothesis test and confidence interval.

Similarly, Harel (1998) suggested that students should see an intellectual need for a concept before they formally encounter it. This *Necessity Principle* facilitates the construction of new knowledge and motivates studying the subject. This activity develops intellectual need by having students compare their predictions to their data, discuss their predictions with their classmates, and reflect on their results. Through this process, they confront the variation in their predictions and their data and see the formal concepts are not as arbitrary abstract ideas, but as tools that can be used to quantify variation, resolve debates, make inferences from the data and draw conclusions.

Lipson (2002) and delMas et al. (1999) recommended that activities should be structured to help students connect their conjectures and observations to the statistical concepts. The activity described here encourages students to link informal ideas, their own data, and the class' data with formal concepts. After this initial linking, it later reinforces the formal instruction by having students continue to seek out patterns in the data and conceptualize the formal ideas in terms of this data. For example, students use their intuition and previous experience to predict a range of statistics that a particular parameter might produce, then they collect data and use the class' results to investigate this interval empirically; after a formal introduction, the students compute and discuss the meaning of confidence intervals, then they examine the class' collection of intervals to develop a deeper understanding of the theoretical ideas and start to connect the formal and informal ideas. The activity connects

ideas related to hypothesis tests in a similar way. By finding and reflecting on patterns in their data, the students are more likely to make these important connections, making the ideas more meaningful.

In addition to helping students connect informal and formal ideas about statistical inference, the design of this activity helps lay the foundation for discussions about repeated sampling and sampling variation. The underlying population—the chips in the bag—is tangible to the students. When they draw samples, they experience the natural variation that occurs when multiple samples are drawn from the same population; this in turn helps them to think of their individual sample as being drawn from a distribution of samples that exists in the class' aggregated data. In addition, the students "own" their data, which fosters productive class discussions about the different conclusions they will draw—and the errors they may make—based on their own samples.

6. Extensions

The activity described here is designed to help students begin to understand the concepts that underlie the T-test and confidence interval. In doing so, it necessarily incorporates other important ideas that are fundamental to statistical inference. For example, the activity incorporates ideas of probability models, sampling variation, sampling distributions, and errors. Consequently, it can be used by instructors as a springboard to discuss other important statistical concepts in more detail. Since the introductory activities require relatively little prior statistical knowledge, the activity could also be used early in the semester as a way to informally explore ideas such as sampling variability, the law of large numbers, and expected value.

In addition to discussing other statistical concepts, the basic format of the activity—gathering data, exploring ideas informally, and making connections to formal concepts—can be extended to other statistical tests and techniques. For example, to explore the concepts of a two-sample T-test, students could be given two bags of colored chips and asked to decide if one lottery has a higher payout than the other. To investigate inference for proportions, students could be given a bag with two colors of bingo chips and asked to determine if the bag has an even split (a cover story might involve polling for an election to determine if one candidate holds a lead over the other). These designs also lend themselves to discussions of power and effect size, since the instructor can control sample size, expected values, and differences in means/proportions. By exploring these fundamental ideas through the use of multiple activities, the students continue to develop their informal reasoning while strengthening the connections to the formal techniques.

7. Appendix A: Worksheets

7.1   Warm-Up Homework Worksheet

Playing the Lottery: Warm-Up

We will be playing the following version of the lottery in class (but without money!). You pay $5 to draw a colored chip from a bag, and your prize depends on the color you draw:

white	green	purple	blue	orange	pink
$0	$1	$5	$10	$15	$25

Your goal is to determine whether or not you expect to break even (i.e., recoup your $5 payment exactly) in the long run. To do this, you will play the game 40 times and analyze your results.

1. What is the population in this situation? What is the sample?

2. What is the parameter in this situation?

3. What is the statistic? Explain how you will compute the statistic.

4. Why is it important to replace the chip after each draw?

5. Do you expect the average payout for your sample to exactly match the average payout for the lottery? Why or why not? If you don't, how far from the actual lottery average do you think your sample average could be? Would you expect it to be above or below the lottery average?

6. Assume for a moment that the long-run average payout is really $5.

    (a) How likely is it for your 40-chip sample to have an average of exactly $5?
        Very Unlikely   Very Likely
                    1   2   3   4   5
    (b) If you draw 40 chips (with replacement), what is the smallest you think your sample average might reasonably be? Call this amount "minimum."

    (c) If you draw 40 chips (with replacement), what is the *largest* you think your sample average might reasonably be? Call this amount "maximum."

7. Each person in the class will compute their sample average; we will plot the class' averages in a histogram or a dotplot.

(a) How will you label the horizontal axis of the histogram/dotplot? What about the vertical axis?

(b) If you make a dotplot, what will each dot represent?

(c) Based on your "minimum" and "maximum" values, roughly how wide do you think the histogram/dotplot will be? Is this connected to your "minimum" and "maximum" values?

(d) Based on other problems you've done, what do you think the shape of this histogram/dotplot will be? Explain your answer.

(e) Assuming that the long-run average payout is $5, sketch a histogram of sample averages.

7.2 In-Class Worksheet: Data Collection and Initial Reflection

Playing the Lottery: Data Collection

Your goal is to determine whether it is a good idea to play the lottery (that is, whether or not you would break even after playing numerous times). You pay $5 to draw a colored chip from a bag, and your prize depends on the color you draw:

white	green	purple	blue	orange	pink
$0	$1	$5	$10	$15	$25

Play the game 40 times, replacing your chip after each turn. Record your winnings below:

1. Do you believe you could break even in the long run? Why or why not?

2. What are the values of $\bar{y}$ (the average) and s (the standard deviation) for your sample?

3. Compare your results with your classmates. How many people think that they would break even in the long run? How much variation does there seem to be in the class' averages? Do you find this surprising? Explain why or why not.

7.3   In-Class Worksheet: Informal Confidence Intervals

Playing the Lottery: Interval Estimates

1. Give a range of values that you think the long-run average payout is likely to fall between (e.g. "I think my average winnings in the long run will be between a and b because..."). Explain your reasoning.

2. Where is $\bar{y}$ located in your interval (e.g. right at the end, near the top, etc.)?

3. How does s affect the width of your interval?

4. Would you change the width of your interval if:

    (a) you had sampled 100 chips instead of 40 chips?

    (b) the standard deviation of your sample was half as large?

    (c) you wanted to be more certain that your interval would contain the parameter?

    Explain your reasoning.

5. Now think about the histogram that you sketched in your homework. Based on this, find a range of values that contains roughly 80% of the hypothetical data points. Compare and contrast the width of this interval to the interval you described above.

7.4   In-Class Worksheet: Informal Hypothesis Testing

Playing the Lottery: Testing Hypotheses

1. In terms of the parameter, what would it mean for you to expect to break even if you played the lottery many times? Call this value of the parameter the "null value."

2. If the long-run average payout was $5, how likely do you think it would be for somebody to draw 40 chips and get *your* sample average?
                Very Unlikely   Very Likely
                       1   2   3   4   5

3. Assuming the long-run average payout is $5, estimate the probability of getting your 40-chip average or an average that is even further from $5. Use the histogram that you sketched in your homework to estimate this value.

4. Based on your answers, do you think the parameter (the long-run average payout) is equal to the null value? Explain your reasoning.

7.5   Homework Worksheet: Informally Looking at Class Data

Playing the Lottery: Analyzing Class Data

1. In the warm-up, you predicted the shape of the histogram of the class' averages. Compare your prediction to the actual shape. Write a short explanation for why it makes sense for this histogram to have this shape.

2. If everybody in class sampled 100 chips instead of 40 chips, how would this histogram change? Explain your reasoning.

3. Imagine that the long-run average payout was $5. Which people had statistics that would be most likely to occur in this case? Which people had statistics that would be least likely to occur?

4. Which people in class had statistics that would make you think that the long-run average payout *wasn't* $5? Explain your reasoning.

5. Again, based on the histogram of averages, give an estimate for the actual long-run average payout. Explain how you made this estimate.

6. The bag contained 30% white chips, 26% green chips, 22% purple chips, 14% blue chips, 6% orange chips, and 2% pink chips.[3]

    (a) Sketch a histogram of a probability model for the possible outcomes from playing this game. Make sure to identify what the random variable X represents.

    (b) Compare and contrast this histogram to the histogram of the class' averages. How are the shapes, centers, and spreads similar and different? Why does this make sense?

    (c) The expected value—the parameter—for this distribution is $4.16 and the standard deviation is $3.63. Based on this, explain why you should or shouldn't play this game.

7. Based on the class' histogram:

    (a) Find a range of values that contains roughly 80% of the data points and is centered at the expected value ($4.16).

    (b) Compare and contrast the width of this interval to the intervals you created in class.

---

[3]The instructor can easily adjust these percentages to make the expected value above $5 or within more or fewer standard errors of $5. In addition, the instructor can have the students compute the expected value and standard deviation by hand.

(c) Use this width to make an interval that is centered at your statistic. Is the parameter contained in this interval? Which people in class would produce an interval that did contain the parameter? What percentage of the class does this represent?

7.6 In-Class Worksheet: Formal Confidence Intervals

Playing the Lottery: Interval Estimates Part 2

1. Use your calculator to construct an 80% confidence interval for the population mean.

2. Compare and contrast the width and center of this interval to the intervals you estimated on the previous worksheets.

3. Based on your confidence interval, compute your margin of error. Write a sentence that explains what "margin of error" means. Then describe how the confidence interval is related to your statistic and your margin of error.

4. Using your calculator, explore how the interval would change if:

    (a) you had sampled 100 chips instead of 40 chips?
    (b) the standard deviation of your sample was half as large?
    (c) you used a higher degree of confidence.

    How do your answers compare to the predictions you made earlier?

5. On your homework, you computed a range of values that contained roughly 80% of the data points. Compare and contrast this to your 80% confidence interval. Discuss what "80% Confidence" means and how these two intervals are related to each other.

7.7 In-Class Worksheet: Formal Hypothesis Testing

Playing the Lottery: Testing Hypotheses Part 2

1. Write appropriate null and alternative hypotheses, in symbols and in words.

2. Use your calculator to compute your p-value for the hypothesis test. What do you get?

3. On the in-class and homework worksheets, you estimated how likely it would be for you to get your 40-chip average (or an average that was further from the mean/expected value) if the parameter was equal to $5. Compare this "likelihood" to your p-value.

4. Based on this comparison, discuss what a *p*-value means.

5. Based on your *p*-value, what do you conclude about the bag? Does this match your initial conclusion?

7.8 Homework Worksheet: Formal Confidence Intervals

Playing the Lottery: Confidence Interval Homework

1. Which students had confidence intervals that contained the parameter? What percent of the class does this represent?

2. Based on your observations, where is the parameter (μ) generally located in relation to the confidence intervals (e.g. near the top, right in the middle, etc.)? Describe any patterns you notice (if any).

3. Each person in class computed a statistic ($\bar{y}$). In general, where does their statistic lie in relation to their confidence interval (how many were in the top half, the bottom half, outside, etc.)?

4. Are there any confidence intervals that you think were incorrectly computed? If so, which ones are they and why do you think they are incorrect? If not, what characteristics would you look for to find an incorrect interval?

5. Explain what "80% confidence" means. Without computing a new interval, how do you think your confidence interval would change if you increased your confidence level? Why do you think it would change in this way?

6. Now pick two other students' data. For each person, compute a 70% confidence interval and a 90% confidence interval. Compare the 70% and 90% intervals to the original 80% interval. What stays the same? What changes?

7. How would changing the confidence level affect whether or not each person's interval contained the parameter?

7.9 Homework Worksheet: Formal Hypothesis Testing

Playing the Lottery: Hypothesis Testing Homework

1. Each person in class computed a *p*-value. What percent of the class' *p*-values caused them to reject the null hypotheses?

2. Compare each person's statistic ($\bar{y}$) to the hypothesized parameter (μ_0) and then look at their p-value. Describe what the p-value tends to tell you about the relationship between someone's statistic and the hypothesized parameter.

3. Explain what a p-value means. Does this explanation seem to match what you observed in the class' data? Explain your answer.

4. The expected payout from playing the game was *not* \$5, so you wouldn't break even if you played the game. However, some people didn't reject the null hypothesis (because their p-values weren't small enough). Why do you think this happened? Did they do something wrong?

5. What relationship do you notice between the p-values and whether or not the confidence interval contains \$5? Describe this relationship and explain why it makes sense.

8. Appendix B: Suggested Discussion Prompts

8.1 Warm-Up Homework

- What are the relationships between the population, parameter, sample, and statistic? Why is the statistic of interest to us?

- Why might somebody care about finding the expected winnings? What are some similar situations in which you might collect data to answer a question like this?

- Why is it important to sample with replacement? How might this in-class activity be different from a real-life scenario?

- What were everybody's estimations for the "minimum" and "maximum" values? How did you come up with your numbers?

- What might the distribution of individual chip values look like? What will the histogram of the class' 40-chip averages probably look like (i.e. the shape, center, and spread)? Why will it look like this, and why might it be different from the population distribution?

8.2 Follow-Up Informal Homework

- How did the class' histogram compare to the histogram we predicted earlier and to the probability distribution of chip values?

- How does increasing the sample size change the histogram? Why does this make sense?

- How did you use the class' histogram to compute intervals that contained 80% of the class' 40-chip averages?

- Compare and contrast each person's interval to this 80% interval.

- If each person made an interval that was centered at their statistic but had the 80% width, how many people in class would have an interval that contained the parameter? Discuss why this makes sense.

- Discuss the following idea: "If 80% of the statistics are within a certain distance of the parameter, then the parameter is within this same distance of 80% of the statistics." How is this idea connected to the intervals that were discussed previously?

- If you didn't know the value of the parameter, how might you produce an interval to estimate it with reasonable certainty?

- How would the sample size, sample variation, and level of confidence affect your interval?

- How would you use these results to decide whether you would break even in the long run?

8.3 Follow-Up Formal Homework

- What patterns did you find in the location of the parameter within each person's interval? What about the location of the statistic?

- What does an 80% confidence interval tell you?

- What does 80% confidence mean in terms of the class' data? How did this appear in the graph of the class' confidence intervals?

- Why are the sample means and confidence intervals not all the same? How much variation would you naturally expect to see?

- Did anybody compute an incorrect confidence interval? What would you look for to determine this, and why?

- How would these confidence intervals change if we used a sample size of 100? What about if we used 95% confidence? Why does this make sense?

- Who would have made an incorrect decision based on their confidence interval? Why would they have done this?

- Explain what a p-value means. How is it related to the statistic and the null hypothesis?

- How might the p-value change if we used a sample size of 100?

- Who would have made an incorrect decision based on their p-value? Why would they have done this?

- What are some relationships you noticed between the p-value and the confidence intervals? Why does this relationship make sense?
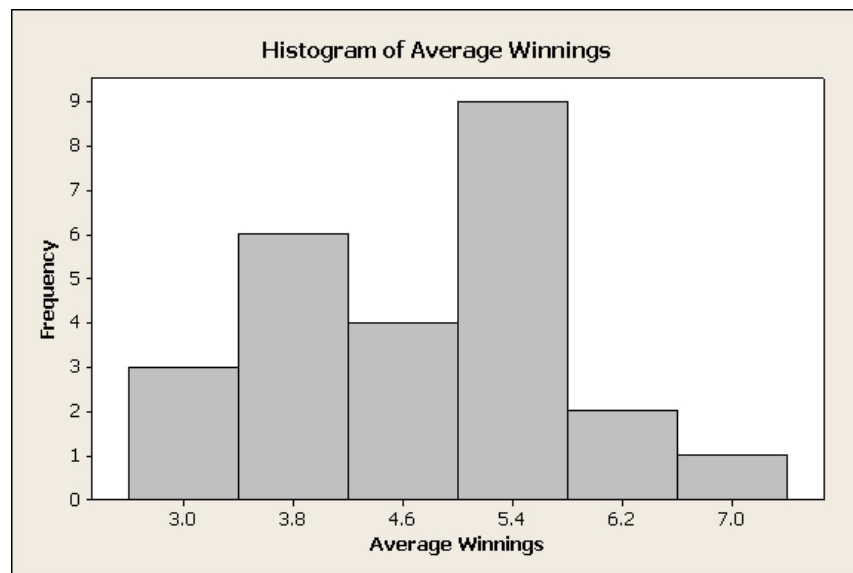
9. Appendix C: Example Class Data

Figure 1.   Sample Histogram of Sample Means

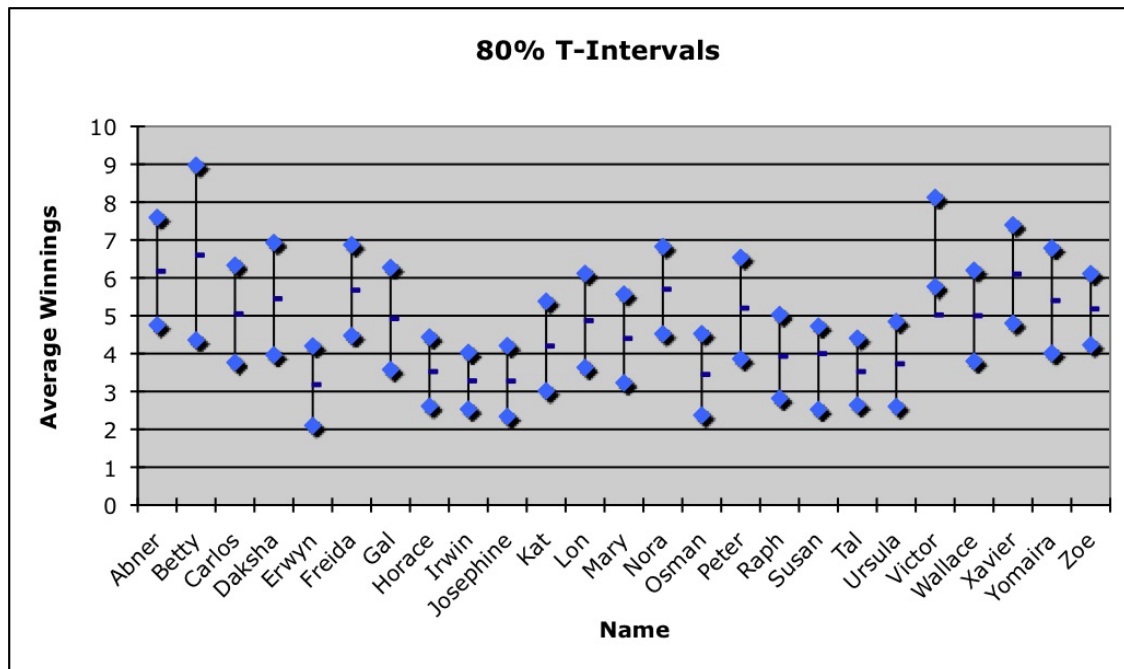| Name	p-value	lower CI - 80%	upper CI - 80%	Mean
Abner	0.2875	4.7547	7.5953	6.175
Betty	0.335168	4.3592	8.9742	6.6
Carlos	0.9598	3.7666	6.3334	5.05
Daksha	0.6947	3.967	6.933	5.45
Erwyn	0.004899	2.1	4.2	3.18
Freida	0.4673	4.4764	6.8736	5.675
Gal	0.942529	3.5775	6.2725	4.925
Horace	0.04153	2.6127	4.4373	3.525
Irwin	0.004821	2.5337	4.0304	3.28
Josephine	0.020974	2.3403	4.2097	3.275
Kat	0.382994	3.018	5.382	4.2
Lon	0.896	3.6337	6.1163	4.87
Mary	0.509	3.23	5.57	4.4
Nora	0.44966	4.5227	6.8273	5.7
Osman	0.067778	2.3744	4.5256	3.45
Peter	0.847	3.8607	6.5393	5.2
Raph	0.212	2.82	5.03	3.93
Susan	0.111043	2.5257	4.7243	4
Tal	0.035335	2.6431	4.4069	3.525
Ursula	0.146	2.6038	4.8462	3.730
Victor	0.979	8.1248	5.7697	5.020
Wallace	0.680	3.8	6.2	5.000
Xavier	0.276	4.8024	7.3976	6.100
Yomaira	0.709	4.0115	6.7885	5.400
Zoe	0.810	4.23	6.11	5.180

Figure 2.   Sample Class Data

Figure 3.   Sample Graph of Confidence Intervals

References

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7(3)*. Retrieved from http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm

Garfield, J., & Ahlgren, A. (1988, January). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19(1)*, 44-63.

Gnanadesikan, M., Scheaffer, R., Watkins, A., & Witmer, J. (1997). An activity-based statistics course. *Journal of Statistics Education, 5(2)*. Retrieved from http://www.amstat.org/publications/jse/v5n2/gnanadesikan.html

Harel, G. (1998). Two dual assertions: The first on learning and the second on teaching (or vice versa). *The American Mathematical Monthly 105*, 497-507.

Jones, G. A., Langrall, C. W., Mooney, E. S., & Thornton, C. A. (2005). Models of development in statistical reasoning. In D. Ben-Zvi & J. Garfield, *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 97-117). Dordrecht, The Netherlands: Kluwer.

Lipson, K. (2002). The role of computer based technology in developing understanding of the concept of sampling distribution. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics. Developing a statistically literate society*. Foorburg, The Netherlands: International Statistical Institute.

Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning*, 4(1), 23-63.

National Council for Teachers of Mathematics. (2000). *Principles and standards for school*

mathematics. Reston: Author.

National Council for Teachers of Mathematics. (2009). *Focus in High School Mathematics: Reasoning and Sense Making in Statistics and Probability*. Reston: Author.

Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). New York: Springer.

Pfannkuch, M. (2006a). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M. (2006b, November). 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/SERJ5(2)_Pfannkuch.pdf

Pfannkuch, M. (2007, October). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education, 2(3)*. Retrieved from http://www.iejme.com/032007/d3.pdf

Pratt, D., & Ainley, J. (2008, November). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal, 7(2)*, 3-4. Retrieved from http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Pratt_Ainley.pdf

Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

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

Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. In M. Lovett & P. Shah (Eds.), *Proceedings of the 33rd Carnegie Symposium on Cognition: Thinking with Data*. Mahweh, NJ: Erlbaum.

Sotos, A. E. C., 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.

Watson, J. M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics, 51(3)*, 225-256.

Watson, J. M. (2008, November). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal, 7(2)*, 59-82. Retrieved from http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Watson.pdf

Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008, November). 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

---

Aaron Weinberg
Ithaca College
953 Danby Rd.
Ithaca, NY, 14850
aweinberg@ithaca.edu

Emilie Wiesner
Ithaca College
953 Danby Rd.
Ithaca, NY, 14850
ewiesner@ithaca.edu

Thomas J. Pfaff
Ithaca College
953 Danby Rd.
Ithaca, NY, 14850
tpfaff@ithaca.edu

Volume 18 (2010) | Archive | Index | Data Archive | Resources | Editorial Board |
Guidelines for Authors | Guidelines for Data Contributors |
Guidelines for Readers/Data Users | Home Page |
Contact JSE | ASA Publications|