BRODY AND JAMIE'S COLLEAGUE: THE DIFFERENCE IN CONFIDENCE INTERVAL ESTIMATORS AND ESTIMATES

Kristen E. Roland and Jennifer J. Kaplan
Appalachian State University, Boone, NC, USA
Middle Tennessee State University, Murfreesboro, TN, USA
rolandke@appstate.edu

In an ever-increasing data-driven world, understanding and describing uncertainty around statistical results have become increasingly more important. Although describing uncertainty can take many forms, recent statements from professional associations (e.g., American Statistical Association) have suggested an increased use of confidence intervals. The calculation of a confidence interval (CI) is fairly straightforward, but the interpretation of a frequentist CI is not. This paper presents a case that demonstrates a potentially important connection among the CI estimator, estimate, and the coverage probability to gain a deep understanding of the meaning of the word confident in the interpretation of a CI. The case provides evidence of the connections among the coverage probability, long-run interpretation of a CI, and the interpretation of a CI that may support robust knowledge.

INTRODUCTION

Over the past few years, statistics educators have been reminded of the importance of data and statistical literacy. Social media has been inundated with evidence of misunderstandings of uncertainty associated with risk, modeling, and statistical analysis. Misunderstanding of statistics and uncertainty can be applied to the uncertainty expressed by weather predictions (e.g., that the cone used to express the variability of models predicting the path of a hurricane instead indicates the area affected by the hurricane), the difference between practical and statistical significance, and the margin of error around a point estimate (e.g., poll data for an election). Research communities have also begun to take notice of the misuse and misunderstanding of uncertainty. In recent years, the American Statistical Association (ASA) and American Psychological Association (APA) have urged researchers to communicate the uncertainty within findings with more transparency and less reliance on *statistical significance* (Wasserstein et al., 2019; Wilkinson, 1999). By openly communicating the uncertainty through standard errors, effect sizes, and interval estimates in published works, whether through formal or informal public venues, authors can demonstrate that they *Accept uncertainty*, are *Thoughtful* researchers, *Open* to positive practices in research work, and *Modest* in the limitations of their work (ATOM; Wasserstein et al., 2019).

Interval estimates, such as frequentist confidence intervals (CIs) and Bayesian percentile intervals, aid in the expression of naturally occurring sampling variability. Frequentist CIs, however, can be difficult to understand and communicate (Fidler, 2005; Morey et al., 2016; Roland, 2020). Many published works have focused on *misconceptions*[1] of CIs, hypothesis tests, and *p*-values (e.g., Belia et al., 2005; Crooks et al., 2019; Fidler, 2005). These have produced long lists of statements that study participants expressed or selected from closed-form questions about the definitional interpretations of CIs and confidence levels (C-Lvls; see delMas et al., 2007; Fidler, 2005) and relational characteristics of CIs (changes to CI width from changes in sample size and C-Lvl; see Canal & Gutiérrez, 2010; Fidler, 2005). Little work, however, has been done to identify what conceptualizations have led to these non-normative ideas about CIs.

It is hypothesized that the difficulty in understanding CIs is due, in part, to the definition of frequentist probability as the long-run behavior of a random process and the fundamental difference in probabilities associated with an *estimator* and an *estimate.* In short, an *estimator* is a function of a random variable used to estimate an unknown value of a parameter [e.g., the (point) *estimator* function $\bar{X} = \frac{\sum x_i}{n}$ to estimate μ] and, under certain conditions, the behavior of its outcomes can be well-modeled by a probability distribution (e.g., $\bar{X}$ can be modeled by a normal distribution if the central limit theorem is applicable). In contrast, an *estimate* is a realized value of a function of a random variable [e.g., if the sample was {1, 2, 3}, the (point) *estimate* would be $\bar{x} = \frac{1+2+3}{3} = 2$] used to estimate an unknown parameter (μ), which means it is an individual outcome from a random process (e.g., repeated sampling of fixed size *n* from a population) and does not have an associated frequentist probability. The

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

communication of uncertainty becomes difficult after a sample has been collected. It is, therefore, hypothesized that the difference between an *estimator* and an *estimate* is important for understanding the interpretation of a CI and C-Lvl. This paper presents a single case, Brody (a pseudonym, the use of the gendered pronoun is based on an assumption of outward projection of gender), from a larger research study ($n = 11$) aimed to uncover ideas of the range of knowledge about CIs and what conceptualizations might lead to non-normative and normative ideas about CIs. Several participants discussed the meaning of the word *confident* in their first interview, using a conceptualization identified as the Capture/Not Capture conceptualization. In this conceptualization, participants stated that the meaning of the word *confident* was different from *probability* in the interpretation of a CI because a calculated CI interval [CI *estimate*] either captures (with a probability of 1) or does not capture (with a probability of 0) the parameter of interest. Six sub-tasks on the second interview were designed to answer the following research question: Do individuals who hold the Capture/Not Capture conceptualization demonstrate knowledge about *why* the Capture/Not Capture conceptualization is true? Meaning, do participants have knowledge about the difference between a CI *estimator* and a CI *estimate*. This paper presents a single case about Brody's navigation through the six sub-tasks to demonstrate final connections needed to understand the Capture/Not Capture conceptualization.

THEORY OF CONFIDENCE INTERVALS

By definition, a frequentist CI *estimator* is an interval estimator with a corresponding measure of confidence defined by a *coverage probability*, $(1-\alpha)$. The coverage probability is the probability that the random interval will contain the unknown value of the parameter. Therefore, a probability statement for deriving a CI for an unknown population mean, using a *t*-interval (for a detailed derivation of this CI see Roland & Kaplan, 2022), can be written as:

$$P\left(\bar{X} - t^* * {}^{S}\!/_{\sqrt{n}} < \mu < \bar{X} + t^* * {}^{S}\!/_{\sqrt{n}}\right) = 1 - \alpha \qquad (1)$$

In this probability statement, it is assumed that there are random components (point *estimators*, e.g., $\bar{X}$ and *S*) and non-random but unknown components (parameters, e.g., μ) and known components critical values, e.g., t^*) that make the probability statement true for a given $(1-\alpha)$. Thus, the only truly random parts of Equation 1 are the point *estimators* ($\bar{X}$ and *S*). After a sample has been collected, the point *estimators* become point *estimates* and the interval becomes a CI estimate.

The probabilistic difference between the CI *estimator* and the CI *estimate* in statistics has large ramifications for the interpretations of a CI and C-Lvl. When discussing the CI *estimator*, it can be interpreted with the frequentist definition of probability, the long-run behavior of a random process: (a) approximately $100(1-\alpha)\%$ of all possible $100(1-\alpha)\%$ CIs should capture the unknown value of the parameter and/or (b) prior to collecting data, the probability of a $100(1-\alpha)\%$ CI capturing the value of the unknown parameter was $100(1-\alpha)\%$. In contrast, interpreting a CI estimate can be interpreted with a commonly taught frequentist interpretation[2]: "We are $100(1-\alpha)\%$ confident that the calculated interval contains the true value of the parameter." *Confident* in this sentence implies that the random process that was used to derive the formula for calculating a CI will capture the parameter $100(1-\alpha)\%$ of the time, rather than the probability that any one CI *estimate* contains the value of the unknown parameter.

THEORETICAL FRAMEWORK

The theoretical framework of the larger research project was based in the development of a formal concept image for the concept of CIs. A *concept image* is "the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes" (Tall & Vinner, 1981, p. 152). Tall and Vinner (1981) describe a *formal concept image* as the formally defined mathematical (statistical) concepts and a *personal concept image* as an individual's interpretation and coordination of concepts and images, which may or may not align with the formal concept image. Roland and Kaplan (2022) present a formal concept image for the concept of CIs, which was created from the theoretical definition of a CI. Of import to this paper are two statistical curricular concepts identified by the formal concept image: a CI *estimator* and a CI *estimate*. The case presented in this paper demonstrates a potentially important connection between these statistical curricular concepts and the interpretations of a CI and C-Lvl.

METHODS

The larger research project recruited introductory, intermediate, and senior statistics students and statistics master and doctoral graduate students from a large, research-focused institution in the Southeastern part of the United States. Brody was a graduating senior pursuing dual degrees in statistics and theology and had completed most of the course work for the statistics program, including the two-course mathematical statistics sequence. Brody engaged in two, one-hour, clinical interviews. The interviews were mostly conversational in nature and were designed to elicit the conceptualizations participants had about the interpretation of CIs and of C-Lvls. Because little was known about the connections students make when conceptualizing and reasoning about CIs, a thematic analysis (Braun & Clarke, 2012) was conducted in the larger research study to uncover themes among the participants' conceptualizations of CIs and C-Lvls.

The results presented in this paper are from four main tasks asked over the course of the two interviews. Task 1 focused on gathering an initial assessment each participant's general knowledge of CIs and C-Lvls through general questions centered around a main context: Two friends, Jamie [a recurring character in the interviews who is a newspaper reporter at a fictional university, Hill Top State University (HTSU)], and Alex. They had created a group playlist for a long car ride and were attempting to estimate the proportion of songs on the playlist that belonged to Jamie. Among other questions, the following two sub-task questions were posed:

- Task 1.1: How should Jamie and Alex interpret the 95% CI you just calculated?
- Task 1.2: Jamie and Alex cannot remember what the 95% represents in the calculation and the interpretation. How would you remind them what the 95% represents?

During conversations around the meaning of *confident* in Tasks 1.1 and 1.2, the Capture/Not Capture conceptualization was used by several participants. This prompted the creation of Tasks 2, 3, and 4, presented below, that were part of the second interview. These tasks were designed to elicit knowledge needed to understand the Capture/Not Capture conceptualization. Task 2 contained three statements that were similar to the correct interpretation of a CI: Jamie is 93% *confident* that the actual mean monthly rent for all students at HTSU is between \$705 and \$793. The difference in the three statements was the use of the words *confident* (Task 2.1), *sure* (Task 2.2), and *probability* (Task 2.3). Participants were asked to reflect on the meanings and correctness of the statements. Of these statements, only Task 2.1 is correct, with *probability* being incorrect because the sentence is referring to a CI estimate, and *sure* is typically synonymous with probability.

Task 3 was only presented to participants with advanced statistical knowledge and focused on the verbal description of a CI estimator. Participants were, again, asked to reflect on the meanings and correctness of the statements:

- Task 3.1: There is a 93% probability that the actual mean monthly rent for students at HTSU is within the interval $\bar{X} \pm \left(t_{n-1}^{*} \frac{S}{\sqrt{n}}\right)$.
- Task 3.2: The process used to generate confidence intervals will capture the actual mean monthly rent for students at HTSU approximately 93% of the time.

Tasks 3.1 and 3.2 are both correct, with Task 3.1 referring to Equation 1 in words rather than symbols and Task 3.2 being an alternative interpretation of a C-Lvl.

Task 4, titled *Jamie's Colleague*, focused on the knowledge needed to understand the Capture/Not Capture statement. Participants were asked to comment on Jamie's colleague's statement after reading the following vignette.

> Jamie talked to a fellow reporter about constructing 93% confidence intervals. Jamie's colleague said that prior to collecting his sample, there is a 93% probability that the confidence interval will capture the actual mean monthly rent of all HTSU students. The colleague continued the explanation by saying that once Jamie collected a sample, the probability of the [previously constructed 93% confidence] interval (\$705, \$793) actually containing the mean monthly rent of all HTSU students is now either 0 or 1.

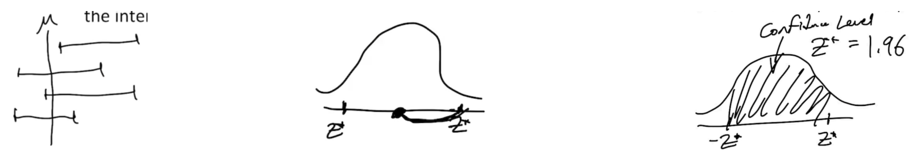
Jamie's colleague is correct, there is associated probability prior to collecting a sample and capture/not capture probability associated with a CI estimate. This task describes the reason why the Capture/Not Capture conceptualization is correct.

RESULTS

Brody demonstrated a well-connected conceptualization of CIs and C-Lvls. He demonstrated normative understanding of the interpretations of CIs and C-Lvls, which included how these interpretations connected to the ideas of coverage probability and the sampling distribution. Brody's responses to Tasks 1.1 and 1.2 included a normative interpretation of a CI. He described the meaning of the word, confident, in the interpretation of a CI as "a quantifiable measure of how certain you are that your estimate is correct, essentially." He explicitly indicated that confident did not mean probability because of the Capture/Not capture conceptualization: "So, the probability of that p [parameter] value being on your interval is either zero or one dependent on whether or not it is on your interval or it isn't. It's just that you don't know where it [the parameter] is." This belief that confident did not mean probability continued into the second interview to Task 2.2 where he said, "I think 93% sure sounds like it saying a 93% probability. … So, I think the confidence level shows more that that's not what that means." Thus, Brody has demonstrated fairly normative knowledge about the interpretation of a CI, but his definition of confident has been truncated by the belief that CIs do not have associated probability.

In his discussion of the interpretation of a C-Lvl during Task 1.2, Brody demonstrated normative conceptualizations that focused both on the long-run interpretation of a C-Lvl and how the C-Lvl connects to the margin of error. Figure 1(a) is similar to the image Brody drew in Interview 2 for Task 1.2, which demonstrates the long run interpretation of a C-Lvl that C-Lvl% of CI's will contain the actual parameter. Brody stated that C-Lvl% of statistics would be contained within a particular set of critical values [$\pm z^*$in Figure 1(b) and 1(c), which are similar to images Brody drew during Interview 1]. During Interview 2, he was able to use Figure 1(b) to demonstrate how the C-Lvl set the margin of error by describing how every statistic within the region in the middle of the two figures would produce a CI that contained the parameter of interest:

> I think of how we look at the distribution, we'd expect … 93% [of the statistics to fall] within these like z^* or t^*. And we'll have a margin of error that's this long [drawing the line connecting the dot with the z^*]. And so, 93% of the sample values will include whatever is at the center [pointing to the dot], right? And, obviously, approximately, if we do anything less than infinite sampling, and we're not going to get an exact representation [that exactly 93% of CIs will capture the parameter].



(a) (b) (c)

Figure 1. Brody's drawings used to visualize his explanations during Interview 2

While these conceptualizations and drawings are correct, Brody did not discuss the idea that there is coverage probability—that there is .93 probability of randomly selecting a sample that would produce an interval that would capture the parameter. This became evident when Brody was faced with Task 3.1 and 3.2. Immediately, Brody agreed with Task 3.1 and drew Figure 1(a) to confirm the correctness of the statement. When Brody read Task 3.2, however, he initially disagreed with the statement because of the word, probability. The interviewer asked Brody about what might be random in the statement and what pre/post sample collection would do to the randomness. Brody identified that $\bar{X}$ and S were random variables and identified the difference pre/post data collection would have on the values of $\bar{X}$ and S: "If you've already collected it, then … the $\bar{x}$ is no longer a random or it would still be a random variable, but you know the value of it. Whereas if you haven't yet, obviously, you don't know the value." The interviewer asked about the difference in probability, to which Brody said "you could look at probability differently on what is the probability of obtaining a sample monthly rent. … But I don't think it changes the probability for the actual mean monthly rent." In this statement, Brody is acknowledging there is probability associated with obtaining a sample, but states that the probability does not relate to the parameter of interest, which is correct.

Brody was struggling with his parallel conceptualizations: (a) the long-run interpretation of a C-Lvl, (b) the C-Lvl as the proportion of statistics within a given region of a sampling distribution, and

(c) the difference in randomness associated with pre/post data collection. *Jamie's Colleague* forced Brody to make the connection between these conceptualizations and the underlying coverage probability associated with the CI estimator. After reading the task, Brody began to connect the randomness associated with the estimator with a measure of probability. He eventually convinced himself that Jamie's Colleague is correct by stating:

> It's actually kind of makes sense … I think, is that like, instinctively you go, it's like it has to be 0 or 1. And the more I'm thinking now about before you take the sample, right? It's because before you take the sample, you don't know where the interval is, right? And so, I guess it would be true that it's not until you set those parameters [gestures to indicate the bounds of the interval, not statistical parameters] that it either is or it isn't. But if our sample is still up in the air, and it's random … if we haven't taken our sample yet, there is still a chance that it could be or that could not be and so I think I think your colleague's right I don't know. This is the most in depth of actually explored this, I haven't really thought about it. … Because I've always been taught that like, it's not a probability because it's set already and that like that's true. … But when it's not set, I think this [referring to *Jamie's Colleague*] is correct.

Although Brody had begun to associate probability with the estimator, he had not made the connection to the coverage probability and how it related to the C-Lvl. The interviewer then asked Brody to refer to his previously drawn diagrams [Figure 1(b) and 1(c)] to see if Brody could make the final connection between the probability associated with the CI estimator and the coverage probability. He continued:

> If 95% of these are in this range, that means 95% would give me intervals that include the true parameter. And since this is the sampling distribution for a sample mean of size, whatever we're using, there's a 95% chance that a random value chosen from this distribution would be on the interval [referring to the area in the sampling distribution that contains central 95% of the statistics]. And so yeah, there is a 95% chance that my sample yields a confidence interval that includes the true proportion, or true mean … But we can only say that before we do the sample.

Thus, Brody was able to connect his parallel conceptualizations to understand how the C-Lvl dictates the region of the sampling distribution where C-Lvl% of statistics are within, the long-run interpretation of the C-Lvl, the meaning of the word confident, and the coverage probability.

DISCUSSION

The connections that Brody made among his parallel conceptualizations were unique within the participants of the larger study and may hold an important place in developing a robust understanding of CIs and how to interpret CIs. The use of *Jamie's Colleague* appears to have helped Brody make the final connections between his understanding of the long-run behavior of frequentist probability and the coverage probability. This provides evidence of the importance of teaching the difference between a CI estimator and a CI estimate and providing more background about the derivation of critical values used in the calculation of a CI. This can be done using applets such as the Rossman-Chance Simulating Confidence Intervals Applet[3], which graphically displays the long-run behavior of a CI based on the C-Lvl [similar to Brody's Figure 1(a) diagram] and displays a sampling distribution, which can be used to develop ideas of margin of error [Brody's Figure 1(b) and 1(c)]. It can also be used to leverage the ideas within *Jamie's Colleague*. As a limitation to this study, the interviewer was not able to add to Brody's knowledge of CIs with formal statistical language such as estimator and estimate. It would have been useful to follow up with Brody about *how* he had developed his conceptualizations, as they were unusual in the larger study and fairly unique to him. Future studies will be needed to inform this issue.

Although the Capture/Not Capture conceptualization is, at face-value, helpful, there is concern that this general statement is truncating students' understanding of the meaning of the word, confident, in the interpretation of a CI. The concept of CIs and the Capture/Not Capture conceptualization are highly complex. Without fully discussing the estimator and estimate with respect to a CI and how the estimator is derived based on the coverage probability, instructors may be oversimplifying the nature of CIs to their students. The results of this case study are the first to connect the concerns of other researchers [see Morey and colleagues (2016)] that the word *confident* is problematic in the interpretation of a confidence interval with a potential source of confusion (i.e., the difference probability in the CI *estimator* and CI *estimate*) and a way to address the confusion. These connections can clearly be made in a mathematical statistics course but may also be introduced via applets and grounding the identification of critical values in the idea of coverage probability in lower-level courses.

More research is needed, however, because thus far these approaches have had mixed results for the authors.

NOTES

1. We prefer the words *conception* or *resource* to *misconception*. We use the word misconception to reflect the intentions and meanings of the original studies cited.
2. Morey et al. (2016) make an argument that the word confident should not be used in the interpretation of a CI (for more information see Kaplan & Roland, 2022).
3. http://www.rossmanchance.com/applets/2021/confsim/ConfSim.html

REFERENCES

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*(4), 389–396. https://doi.org/10.1037/1082-989X.10.4.389

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol. 2. pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Canal, G. Y., & Gutiérrez, R. B. (2010). The confidence intervals: A difficult matter, even for experts. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of Eighth International Conference on Teach Statistics (ICOTS8)*. ISI/IASE. https://iase-web.org/documents/papers/icots8/ICOTS8_C143_CANAL.pdf?1402524973

Crooks, N., Bartel, A., & Alibali, M. W. (2019). Conceptual knowledge of confidence intervals in psychology undergraduate and graduate students. *Statistics Education Research Journal*, *18*(1), 46–62. https://doi.org/10.52041/serj.v18i1.149

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

Fidler, F. (2005). *From statistical significance to effect estimation: Statistical reform in psychology, medicine and ecology* [Unpublished doctoral dissertation]. The University of Melbourne.

Kaplan, J. J., & Roland, K. E. (2022). Confidence means what?!?: Lexical ambiguity in the interpretation of confidence intervals. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the gap: Empowering and educating today's learners in statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS 11)*. ISI/IASE.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Roland, K. E. (2020). *Student conceptualization of the interpretation of the confidence interval and the confidence level: Identifying similarities and differences in student concept images of confidence intervals* [Doctoral dissertation, The University of Georgia]. ScholarWorksUGA. https://esploro.libs.uga.edu/esploro/outputs/9949365855202959

Roland, K. E., & Kaplan, J. J. (2022). Using concept images as a framework for the concept of confidence intervals. In S. S. Karunakaran & A. Higgins (Eds.), *Proceedings of the 24th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 490–497). RUME. http://sigmaa.maa.org/rume/RUME24.pdf

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169. https://doi.org/10.1007/BF00305619

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$." *The American Statistician*, *73*(Suppl. 1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594