FOSTERING CHANGES IN CONFIDENCE INTERVALS INTERPRETATION

Luisa Andrade, Felipe Fernández and Ingrith Álvarez
Universidad Pedagógica Nacional, Bogotá, Colombia
landrade@pedagogica.edu.co, fjfernandez@pedagogica.edu.co

As literature has reported, it is usual that university students in statistics courses and even statistics teachers interpret the confidence level associated with a confidence interval as the probability that the parameter value will be between the lower and upper interval limits. To confront this misconception, a class activity has been designed with the aim to realize that this application of confidence level explicitly violates the basic laws of probability. We consider two non-overlapping confidence intervals, that could plausibly correspond to two random samples from the same population, where the probability of events within this interpretation contradicts the probability rule for disjoint events and the rule of monotonicity ($P[E] \leq P[F]$ if $E \subseteq F$). We use simulation to help students shift to a frequentist interpretation of confidence intervals.

INTRODUCTION

Identification of misconceptions associated with the interpretation of the confidence interval obtained by estimating the mean of a population based on a sample mean and its standard error has attracted the attention of many researchers (see e.g., Behar, 2007; Fidler & Cumming, 2005; Olivo, 2008; Olivo & Batanero, 2007; Kalinowski, 2010; Yañez & Behar, 2010; Salcedo et al., 2011, among others). In most of the work mentioned above it is reported that one of the major misconceptions associated with the interpretation of the confidence level is to think that confidence level is the probability that the population mean is within the lower and upper limits of the calculated interval. The proper interpretation, as researchers who have been studying it warn, must incorporate the probability frequentist perspective for the confidence interval generated, and not for the population mean. This means that the confidence level indicates the probability that one of the intervals produced in the process to generate confidence intervals contains the population mean. Then, as Behar (2007) says, the probability refers to the likelihood of the method for making intervals and not to the parameter; if sampling is repeated a sufficient number of times, the percentage of intervals generated containing the population mean is given by the confidence level as well. In summary, a 90% confidence interval is one of 90 of 100 possible intervals obtained under the same sampling conditions, which contain the population mean.

On the other hand, although literature points out that both students and experts reveal such misconception, the references to research design that propose the planning or implementation of instruction aimed at trying to overcome it, are rare. In this paper we want to show some parts of the class activities designed and the learning assumptions that underlie them. In Fernandez, Andrade and Álvarez (2013) is possible to find a more detailed discussion about the conceptual content, the methodology and the activities in themselves, developed in a research project.

CLASS ACTIVITIES

The planned activities were implemented along several consecutive class sessions, with college students who had taken a first course in statistical methods. The design of the instruction followed the guidelines suggested by the research design methodology, such as hypotheses formulation about how students' learning is expected to evolve throughout the instruction. The planned activities are divided into five parts, which will be briefly overviewed in this paper. Then the focus will be on the third part, which is the one we are interested in discussing.

The context for the class activities suggests a problem of measurement and statistical characterization of the IQs of a group of a thousand students who make up the target population. In the first part the students work around the difference between the notions of population and sample of individuals, as well as around data sets related to the context of the situation which describe a trait of the individuals. The second part revises random sampling of the given population and sample mean computation of IQs; this part aims at clarifying the difference between parameters, estimators and estimates. The third part addresses the interpretation of confidence level and questions its meaning in order to contribute to its re-conceptualization. The fourth part proposes a

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

manual simulation that aims at helping the students to verify the new interpretation of the confidence level, as well as to let them gain confidence in their own finding. Finally, in the last part, we appeal to computer simulation in order to strengthen the frequentist interpretation of confidence level.

WORKING AROUND THE INTERPRETATION OF CONFIDENCE LEVEL

The baseline scenario for the proposed work in this part of the instruction is that the students' common interpretation of the confidence level, once the interval is built, coincides with the misconception described above, i.e., that the confidence level is the probability that the population mean is within the lower and upper limits of the calculated interval. In other words, we assume that students consider that the probability implicitly refers to the confidence interval in itself as an event of the sample space, understanding the latter as the real number line.

The students start the work in this part of the instruction by interpreting the confidence level before the computation of the confidence interval. We believe that the frequentist approximation of the confidence level may be more visible to students at a time prior to the construction of the confidence interval, i.e. when the probability that constitutes the confidence level is established. Since there are no numbers yet that determine the interval, it is possible that students will not imagine the real number line as the sample space or the reference set.

Next, we look forward to confronting the students' failed idea of confidence level as they work with two samples generating non-overlapping confidence intervals. Then, students should make the interpretation of the two confidence levels, which we anticipate would fit the misconception and are identical for both intervals. Naturally students will see that the confidence intervals do not overlap and we expect that due to this fact, they will be questioned with the confidence level interpretation as a probability of the mean contained in the interval, 90% in this situation, being the same for both intervals.

Third part (work in pairs)

1. If you were to estimate the IQ mean of the population, by using a 90% confidence interval, describe your interpretation of the confidence level, before computing the interval.
2. For the next sample* of twenty students IQs, construct a 90% confidence interval for the population mean. Refer to it as I_1.

   149 129 119 130 97 128 129 107 98 122
   136 113 115 117 118 142 137 120 134 140
3. Describe your interpretation of the confidence level of this interval.
4. For the next sample of twenty students IQs, also obtained randomly and which is also thought as representative of the population data, construct a 90% confidence interval for the population mean. Refer to it as I_2.

   101 107 130 101 115 104 91 91 121 109
   104 125 113 98 110 119 102 92 111 120
5. Describe your interpretation of the confidence level of this interval.
6. Explain whether there is an inconsistency with the confidence level interpretations for the two built intervals (I_1 and I_2).

* The sample was obtained randomly and it is considered representative of the population data.

Since it is quite possible that students will not detect the inconsistency in the interpretation of the confidence level and will insist in the referred misconception, the work is then focused on representing both intervals on the real number line, expecting that this concrete and familiar mathematical structure will help to make the conflict more noticeable. In the first place, students will see that one interval is a subset of the complement of the other, and vice versa. In the second place, students should note -based on their knowledge of probability- that the maximum probability of the complement of the interval is 10% (since the confidence level is 90%), and that the probability, as interpreted by them, of the event related to the other interval that is a subset of such complement, was established as 90%. Finally, students should recognize that this is absurd, as it contradicts the rule of monotonicity which establishes that if $E \subset F$ then $P[E] \leq P[F]$; this way we expect that students will find out the conflict and realize the presence of something inappropriate in their idea of confidence level. Besides, we think that students could be aware that the previously considered probabilities, related to the intervals and their complements, must satisfy the probability rule of the union of disjoint events, which in this case will produce a probability greater than 1.

7. For the intervals built before, do the following:
   a. Plot the two intervals in the same real number line.
   b. Find out whether the intervals overlap. Identify the complement of each interval.
   c. Explicitly express the relationships between an interval and the complement of the other?
   d. Complete the following table by noting the intervals I_1 and I_2, the confidence level interpretations made on items 3 and 5 for these intervals, and the interpretation of the probabilities of the intervals complements, I_1^c and I_2^c.

Confidence interval	90% Confidence level interpretation	Interpretation of 10% associated to the interval complement

8. Considering the subset relationships found in item 7c, describe the relationship between the interpretation given to the 90% confidence level associated with the interval I_1 and the interpretation given to the 10% associated with the complement of the interval I_2.
9. Considering the subset relationships found in item 7c, describe the relationship between the interpretation given to the 90% confidence level associated with the interval I_2 and the interpretation given to the 10% associated with the complement of the interval I_1.
10. Based on the above, explain whether there is an inconsistency with the confidence level interpretations for the two built intervals (I_1 and I_2).
11. So finally, what is your pronouncement about the interpretation of the confidence level associated with the intervals (I_1 and I_2)?

In order to allow students to reassure their finding about an irregularity in the interpretation of confidence level, afterwards, we orally interact with students upon a prepared script depending on possible responses given to the table presented on item 7d. For example, the next script is used in case the students' responses would be the mentioned misconception.

Students' responses and interaction		
Confidence interval	90% Confidence level interpretation	Interpretation of 10% associated to the interval complement
I_1 = (118,79; 129,20)	*The population mean is in I_1 with a probability of 90%*	*The population mean is in I_1^c with a probability of 10%*
I_2 = (103,88; 112,11)	*The population mean is in I_2 with a probability of 90%*	*The population mean is in I_2^c with a probability of 10%*
The teacher questions the students: What are the events associated with the declared probabilities? The teacher asks about the probability rule of an event contained in another event, and requests them to express the relationship between the probabilities of those events. The teacher raises the question: What could then be the probability that the population mean is in the second interval obtained? Then the teacher asks if the above relationships are consistent.		

CONCLUSION

The relentless presence of the misconception about the notion of confidence level generates some hypotheses to be considered when addressing the instruction of confidence intervals. One of these assumptions is connected to students' difficulties conceptualizing probability notion and with the diversity of ways in which probabilities can be assigned. In the first case, students conceive probability as a numerical value, without being necessarily linked to a specific set of reference, i.e., to a sample space, when considering some situations that are not treated in textbooks; this conception can be associated to the intuitive interpretation for probability[1] which is related to personal opinions and beliefs. In the second case, recurrent use of classical probability

[1]In the literature, see e.g. Batanero (2005), different interpretations of probability are pointed out: the intuitive, the classical, the frequentist, the subjective and the axiomatic.

interpretation, although appropriate within certain contexts, leads to the extension of the equiprobability of outcomes to other situations where the frequentist interpretation makes more sense; for example, random experiments where there is not a natural symmetry and therefore the events generated may not have equal probability.

Another premise is related to the fact that variation study in statistics is limited; firstly, since class statistics work deal merely with variability of data sets; secondly, statistical inferences regarding a population based on the "induction" from a single sample, seems to truncate the possibility of thinking in the variation between samples. These, help to position the interpretation of confidence level as the probability with the sample space constituted by the population from which the sample is taken, and not formed by the set of all possible samples of a given size that can be taken from the population. Hence, to familiarize students with the frequentist interpretation, introductory descriptive statistics courses should include tasks that make variation perceptible, specially, variation between the values associated to estimators generated from different samples of the same size, i.e., tasks that allow students to account for the variation that is present in the estimations linked to the variation of the corresponding samples.

It is also important to consider that once the interval is computed, stating that the confidence level determines the number of intervals that contain the parameter, does not seem to shed much light upon the inference process being done. Consequently, the construction of only one confidence interval appears as a poorly accurate estimate, even as a useless estimate; this could lead to conclude that, when estimating a parameter, it would suffice to establish a confidence level and to indicate that it determines the percentage of intervals that contain the parameter, in case a good number of these would be built. In search of making sense and perceiving a practical utility of a confidence interval, instruction could propose situations where the built interval helps in decision making with uncertainty; e.g., when a desirable mean has been set, a confidence interval allows to make conclusions depending on whether the interval contains the mean or not. In addition, situations where computing a confidence interval is helpful in hypothesis tests; e.g., when conjecturing about the parameter before building the confidence interval, then computing it and checking whether the parameter belongs to the interval, thus rejecting or validating the conjecture.

REFERENCES

Batanero, C. (2005). Significados de la probabilidad en la educación secundaria. *RELIME, 8* (3), 247-263.

Behar, R. (2007). *¿Estamos buscando el ahogado aguas arriba? El caso de la estimación con intervalos de confianza*. Primer Encuentro Nacional de Educación Estadística, Bogotá.

Fernández, F. Andrade, L., & Álvarez, I. (2013). *Experimentos de enseñanza y razonamiento estadístico en situaciones de probabilidad e inferencia estadística*. Bogotá: Universidad Pedagógica Nacional.

Fidler, F., & Cumming, G. (2005). *Teaching Confidence Intervals: Problems and Potential Solutions*. Paper presented at the 55th Session of the World Congress of the International Statistical Institute, Sydney, Australia.

Kalinowski, P. (2010). Identifying misconceptions about confidence intervals. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8).* Ljubljana, Slovenia.

Olivo, R. (2008). *Significado de los intervalos de confianza para los estudiantes de ingeniería en México (doctoral thesis).* Granada: Universidad de Granada.

Olivo, R., & Batanero, C. (2007). Un estudio exploratorio de dificultades de comprensión de intervalo de confianza. *Revista Unión, 12,* 37-51.

Salcedo, A., Sarco, A., González, J., & Yañez, G. (2011*). Interpretación de intervalos de confianza por docentes en formación*. Caracas: Centro de Investigaciones Educativas (CIES).

Yañez, G., & Behar, R. (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 the Eighth International Conference on Teaching Statistics (ICOTS8).* Ljubljana, Slovenia.