DESIGN AND EVALUATION OF A HYPOTHETICAL LEARNING TRAJECTORY TO CONFIDENCE INTERVALS BASED ON SIMULATION AND REAL DATA

Santiago Inzunza
Universidad Autonoma de Sinaloa
sinzunza@uas.edu.mx

This article discusses the design an evaluation of a learning trajectory to introduce confidence intervals in an introductory university course from an informal perspective based on sampling simulation and real data. The trajectory was evaluated as part of a first improvement cycle with a group of 11 students (19-21 years) of International Studies in a Mexican university. The results show that it is possible to reason adequately with complex concepts that underlie confidence intervals, using data with real contexts and dynamic and interactive computer tools that allow real time visualization of the sampling. The survey context seems appropriate to introduce social science students to the study of confidence intervals, since they involve concepts that can be discussed in context with meaning for students.

INTRODUCTION

Various models for teaching statistics have been proposed in the recent years, which emphasize the development of statistical reasoning and thinking, and have as an integrating element, the use of real data, digital technologies and constructivist learning environments, among other pedagogical components (Cobb & McClain, 2004; Pfannkuch, Wild & Parsonage, 2012). In statistical inference, technology is considered a valuable tool to rethink the way in which its central ideas and concepts are presented, in this sense "we need to throw away the old notion that the normal approximation to a sampling distribution belongs at the center of our curriculum, and create a new curriculum whose center is the core logic of inference . . . I like to think of it as three Rs: randomize, repeat, reject. Randomize data production; repeat by simulation to see what's typical and what's not; reject any model that puts your data in its tail" (Cobb, 2007, p. 11-12). However, a lot research is still required, since the foundations of statistical inference and its methods involve a diversity of concepts that are complicated for students, even for many teachers. In this research we have designed and evaluated in a first cycle a hypothetical learning trajectory (HLT) to promote the learning and development of informal inferential reasoning on confidence intervals in survey scenarios with real data through sampling simulation.

CONCEPTUAL FRAMEWORK

Hypothetical Learning Trajectories and Informal Inferential Reasoning

A HLT consists of three main components: the learning objectives for students, the mathematical tasks to promote the learning, and hypotheses about the students' learning process (Simon, 2014). The conceptual analysis of confidence intervals from an informal perspective considers sampling and variability as central concepts in the design of learning tasks. In this sense, Dierdorp et al. (2016) identify five concepts that underlie sampling and that are important for inferential reasoning: randomness, sample size, distribution, intuitive confidence intervals and relationship between sample and population.

Meanwhile, Makar and Rubin (2009) identify three key characteristics of a statistical inference: a) a statement of generalization "beyond the data", b) Use of data as evidence to support this generalization, c) a probabilistic language that expresses some uncertainty about generalization. Once the central concepts have been identified as learning objectives, the next step is the design of tasks to promote learning and reasoning of the conceptual baggage that underlies the confidence intervals. This is not a trivial activity, because it requires an in-depth analysis of the relationships between these concepts to link them coherently and generate a learning trajectory.

Rubin, Hammerman and Konold (2006) point out that Informal Inferential Reasoning (IIR) is reasoning that involves ideas and relationships such as center, variability, sample size and bias control. Zieffler et al. (2008) define IIR as the way in which students use their informal statistical knowledge to make arguments to support inferences about populations. IIR "combines cognitive aspects of reasoning about data and chance as well as sociocultural aspects that take place in the

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

classroom and individual practices, dispositions and discourse" (Ben-Zvi, Gil & Appel, 2007, p.3).

Technology as a cognitive tool

The interactivity and multiplicity of dynamic visual representations available in many software tools, as well as the power of simulation to extract almost simultaneously, large number of samples from a population, can help at students to access the great ideas of statistical inference. Pea (1987) refers to this potential of technology as a "reorganizing metaphor", which when properly used, has the capacity to cause structural changes in the cognitive system of students, through a reorganization and transformation of the activities they perform with the representations and their transformations.

METHODOLOGY

In the design of the HLT, we adopted a conceptually holistic approach (Bakker & Derry, 2011). The content of each concept is articulated in its relations with other concepts and not as isolated concepts part by part. The conceptual basis of the HLT is sampling and its relationship with randomness, sample size, sampling distribution, intuitive confidence intervals and relationship between sample and population. The HLT are composed of four activities with the following sequence: populations, samples and sample variability, sampling distribution of a proportion, connection between sampling distribution and confidence intervals, calculation of a confidence interval for a proportion and exploration of margin error, confidence level and sample size.

Each activity is based on the question from a survey. The context of the first two activities was a survey to estimate the consumption of refreshing beverages by the Mexican population; it was found that 72% of the population consumed refreshing beverages regularly (P = 0.72). The third activity was based on commercial relationship between Mexico and the United States; when asked about the importance of the USA for Mexico, 79% of the respondents answered that the USA was very important (P = 0.79). The last activity had as context a survey on the legalization of the consumption of marijuana in Mexico; when asked if they agreed with the legalization, 77% of the respondents were against legalization (P = 0.77).

The research was carried out with 11 students of International Studies who took a basic course in probability and statistics. The students had little mathematical background in the subject, so we decided to focus the course towards the modeling and simulation of random events and sampling of populations using the computer environment provided by TinkerPlots (Konold & Miller, 2015). The researcher was the teacher of the subject and in the development of the activities he acted as a guide who participated only when necessary. The information collection instruments used were worksheets, software files, interviews with some students and the Assessment Inferential Reasoning in Statistics (AIRS), an evaluation questionnaire for inferential reasoning, applied at the end of the HLT.

RESULTS AND DISCUSSION

The population and its model (Sampler)

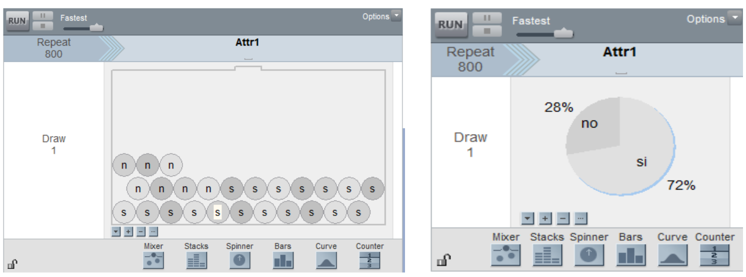

Figure 1. Two correct models (samplers) built by students in activity 1

Noll, Glover and Gebresenbet (2016) identify the construction of the model as the first stage of reasoning in the simulation of a statistical problem. In the context of TinkerPlots, consists in construct a device known as *Sampler* to appropriately model the problem. The construction of the model was a simple task for the students; they correctly expressed the value of P and the sample size in the *Sampler*, to then perform the first trial of the model. Figure 1 shows two models correctly constructed in activity 1, to represent 72% of the population that regularly consumes refreshing beverages.

The empirical sampling distribution

Once the simulator model is built, a sample is simulated and the statistic of interest is calculated. The *Collect Statistic* option of TinkerPlots repeats the previous process and accumulates the results to build the empirical sampling distribution (see Figure 2).

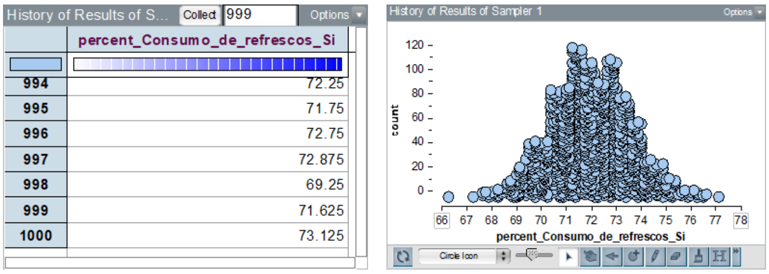

Figure 2. Empirical sampling distribution for the proportion of SI in the activity of consumption of refreshing beverages (n = 1000 samples)

The empirical sampling distribution allows visualizing intervals where most of the samples results are located and intervals where sample results is unlikely to occur. We proposed to the students visually identify a central interval for the population parameter (P). In all cases, the intervals they built contained the parameter. It was also proposed they estimate the probability that a sample would occur 6% beyond 72%, that is, less than 66% and more than 78%. Some answers of the students:

Nothing probable, the graph shows that in 500 repetitions of the event, never appers. Frida

It is very unlikely, the simulation was done and they did not appear such values. Nayeli

It is very unlikely, with the simulation we can verify that the parameter varies around 3% maximum, and 6% is a very far value. Jasmine.

The next step of the activity was to increase the sample size to 1,500 people and build the sampling distribution again. All students noticed that the difference between the center of the sampling distribution and the P parameter decreased considerably as the sample size increased. Let's see some answers.

There is a very minimal difference, almost zero, of 0.02%. Frida

It is very similar, it only varies by 0.0299%. Aglae

It is almost identical with only 0.165% variation of 72%. Mariangel

The visualization of the sampling distribution allowed the students to identify sample results with high or low probability of occurrence. Even Jasmine establishes a margin of 3% as a limit from which the probability of sample results decreases considerably, and erroneously she denominates parameter. The students also identified that the difference between the center of the sampling distribution and the population parameter they are almost the same, as the sample size increases, an important idea in parameter estimation.

Connecting sampling distributions with confidence intervals

The sampling distributions contain the information to make an inference. In this activity the students estimated a confidence interval for an unknown parameter, establishing a direct connection with the sampling distribution. This activity comes from a survey on the commercial relationship between the USA and Mexico; the parameter consists of the proportion of Mexicans who consider that the USA is important for Mexico.

The students estimated the parameter (P) visualizing the sampling distribution (see Figure 3) considering a confidence level approximated of 90%. The estimation in all cases was based on the mean of the sampling distribution, influenced by the previous activity where they observed the great approximation between the mean of sampling distribution and the parameter. There were no cases where another value was selected within the shaded area for the specified confidence level, which would also have been correct. Considering the theoretical calculation of the 90% confidence interval and taking the value of p = 0.79, the interval to P is (76.6, 81.4). A total of 8 students made an estimate very close to this interval using the *Ruler* option to shade the interval (see Figure 4).

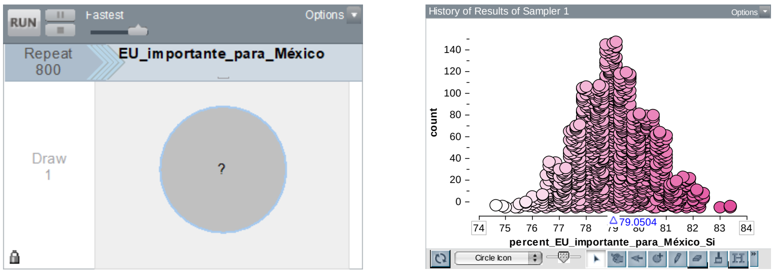

Figure 3. *Sampler* with hidden model and sampling distribution for 1000 samples of 800 people

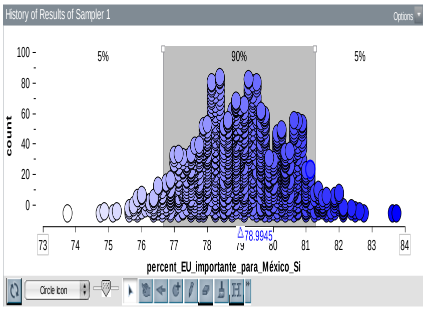

Figure 4. Sampling distribution with shaded area of 90% (Arlin)

The sampling distribution also served for estimate the margin of error. A total of 7 students provided a margin of error ranging between 2% and 3%, obtained by visualizing the difference between the mean of the sampling distribution to upper or lower limits of the interval. These results are similar with the results provided by the survey. For the case of the 95% confidence level, the students identified a slightly wider range than in the case of 90%. Some answers:

If the confidence level is increased, the amplitude of the interval increases. Arlin

The margin of error increases proportionally to the percentage of confidence level. María.

The margin of error increases of 90% to 95% confidence level. Frida

In general, students identified correctly that the amplitude (margin of error) of the interval increases as the level of confidence increases, which is an important concept in estimation of parameters.

Calculation of confidence intervals and exploration of their elements

The purpose of the last activity was to explore the concepts of margin of error, confidence level and effect of sample size. The context of the activity comes from a survey on the decriminalization of marijuana consumption in Mexico. The question we have chosen for the design of the activity refers to the opinion on the legalization of marijuana, in which 77% respond against, 20% in favor and 3% still do not know. For the exploration, it was required the spreadsheet option to enter the formulas of the interval elements for a proportion: $p = z\sqrt{\frac{p(1-p)}{n}}$. The repeated simulation of the sampling automatically generates the proportions of each sample (first column of the spreadsheet), the others columns (standard error, margin of error, lower limit, upper limit and capture or not capture the parameter) require the introduction of the corresponding formula (see Figure 5).

History of Results of Sampler 1

	prop_LEGALI...	Error_Estandar	Margen_de_Error	Límite_Inferior	Límite_Superior	Resultado
993	0.765	0.0149906	0.0293816	0.735618	0.794382	Cae dentro
994	0.79375	0.0143052	0.0280382	0.765712	0.821788	Cae dentro
995	0.7525	0.0152579	0.0299056	0.722594	0.782406	Cae dentro
996	0.78	0.0146458	0.0287058	0.751294	0.808706	Cae dentro
997	0.765	0.0149906	0.0293816	0.735618	0.794382	Cae dentro

Figure 5. Spreadsheet with the elements of a confidence interval

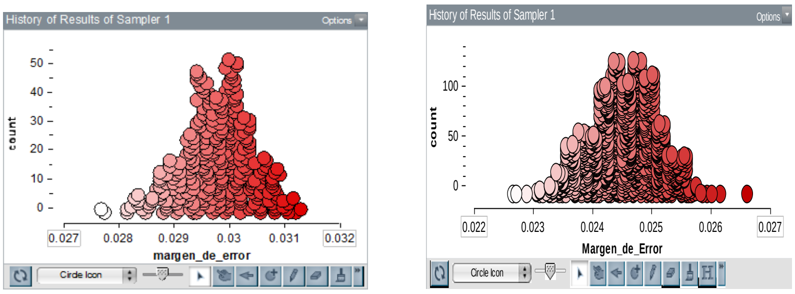

Figure 6. Graphs of margin of error for each interval constructed by Frida, for 95% and 90% confidence level respectively

Once the simulation of 500 samples was done, the margin of error for each interval was graphed, with the idea of exploring its distribution and visualizes the range of variability. The visual exploration contributed to that many students understand the variability of the sampling error, which in some cases may be almost zero, but in other cases may be greater than 3%, very close to the 3.5% reported in the survey. For example, Frida observed that for confidence level 95% the margin of error ranged between 2.6% and 3.1% approximately, but when the reliability was 90%, the margin of error did not exceed 2.7% for the 500 simulated samples (see Figure 6).

If the confidence level is reduced, there is a smaller margin of error. Frida

There is a maximum margin of error of 3.2%, very close to the survey. Nayeli

Confidence level increases the margin of error presented in the graphs. Aglae

The other component of an interval that we investigated was the confidence level. The students constructed a graph with the results of the last column (the interval capture or not capture the parameter). The aim was for the students to relate the confidence level (90% or 95%) with the percentage of intervals that capture the parameter, to understand the confidence level as the proportion of intervals in a repeated sampling that the parameter is captured (see Figure 7).

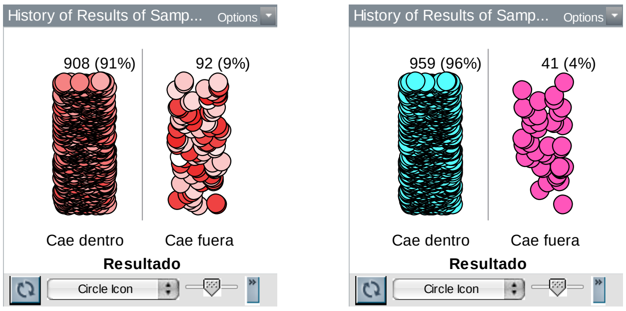

Figure 7. Number of intervals that capture /do not capture the parameter for 90% and 95% confidence level

Some representative answers of the students:

If the confidence level increases, more intervals capture the parameter. Frida

The greater the reliability, the more samples fall within the confidence interval. Marielena

The relationship between the confidence level and the percentage of samples is that the confidence level decides the percentage of samples that are within the margin of confidence. Aglae

CONCLUSIONS

The results show that it is possible to reason adequately with complex concepts that underlie confidence intervals from an informal perspective, using data with real contexts and dynamic and interactive computer tools that allow real time visualization of the sampling. The survey context seems appropriate to introduce social science students to the study of statistical inference, and in particular to confidence intervals, since they involve concepts such as margin of error, confidence level, sampling, sample size, uncertainty, among others, that can be discussed in context with meaning for students. The cognitive properties and the power of simulation and visualization of the software used, has allowed an empirical approach to abstract concepts, such as the sampling distributions, the margin of error and the confidence level of a confidence interval.

The results of the final evaluation, in which items of the AIRS test developed by Park (2012) were used, show groups of concepts that were more difficult than others for the students, as was the case distinction between population, sample and sampling distribution, properties of sampling distributions and confidence intervals, in which 5 or less of the 11 students answered correctly. However, there were concepts where the percentage of correct was greater than 70% as was the case of the expected value of a sample, construction of the model, the relationship between confidence interval and margin of error and relationship between sample size and sampling distribution. In accordance with the principles of HLT, these results will serve for a redesign of activity and modification of the sequence of activities of the HLT.

REFERENCES

Bakker, A. & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*(1-2), 5–26.

Ben-Zvi, D., Gil, E. & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In *Proceedings of 5th International Research Forum on Statistical Reasoning, Thinking and Literacy*. University of Warwick, UK.

Cobb, G. (2007). The introductory statistics course: a Ptolemaic curriculum? *Technology Innovations in Statistics Education*, *1*(1), 1–15.

Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In *The challenge of developing statistical literacy, reasoning and thinking* (pp. 375-395). Springer, Dordrecht.

Dierdorp, A., Bakker, A., van Maanen, J., & Eijkelhof, H. (2016). Supporting students to develop concepts underlying sampling and to shuttle between contextual and statistical spheres. In *The Teaching and Learning of Statistics* (pp. 37-48). Springer, Cham.

Konold, C. & Miller, C. (2015). *TinkerPlots. Dynamic Data Exploration.* Learn Troop.

Makar, K. & Rubin A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

Noll, J., Gebresenbet, M. & Demorest, E. (2016). A Modeling and Simulation Approach to Informal Inference: Successes and Challenges. In D. Ben-Zvi, & K. Makar (Eds.). *The Teaching and Learning of Statistics: International Perspectives*, Springer, 139-150.

Park, J. (2012). *Developing and Validating an Instrument to Measure College Students' Inferential Reasoning in Statistics: An Argument-Based Approach to Validation* (Doctoral dissertation). Universidad de Minnesota.

Pea, R. (1987). Cognitive technologies for mathematics education. In Schoenfeld, A. (Ed.). *Cognitive Science and Mathematics Education* (pp. 89-122). New Jersey: Lawrence Erlbaum Associates.

Pfannkuch, M., Wild, CH. & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM Mathematics Education*, *44*(7), 899-911.

Rubin, A., Hammerman, J. & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Proceeding of 7th International Conference on Teaching Statistics*. Salvador Bahía, Brazil. Voorburg: I.S.I.

Simon, M. (2014). Hypothetical Learning trajectories in Mathematics Education. In S. Lerman. (Ed.) *Encyclopedia of Mathematics Education*, 272-275.

Zieffler, A., Garfield, J., DelMas, R. & Reading, CH. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40–58.