EXPERTS AND STUDENTS` CONCEPTIONS REGARDING CONFIDENCE INTERVALS

Roberto Behar Gutiérrez

Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali
roberto.behar@correounivalle.edu.co

Gabriel Yáñez Canal

Universidad Industrial de Santander, Bucaramanga
gyanez@uis.edu.co

Abstract

We present the results of a research whose purpose was to find out what a sample of experts (statisticians and statistics university professors) and university students understood exactly by confidence intervals. To this end, a questionnaire was answered by 41 experts and 297 students. The results show that both, students and professors, possess misconceptions regarding confidence intervals. The conception that these intervals contain sample means or single values of the population instead of possible parameter values, and the interpretation of significance levels as a measure of certainty, without any frequency referent, were found to be the most generalized misconceptions.

Keywords: Population, Sample, Confidence intervals, Misconceptions, Significance level, Certainty.

Resumen

Se presentan los resultados de una investigación cuyo propósito es encontrar lo que una muestra de expertos (estadísticos y profesores universitarios de estadística) y estudiantes universitarios entienden por intervalos de confianza. Con este fin se aplicó un cuestionario que fue respondido por 41 expertos y 297 estudiantes. Los resultados muestran que ambos, estudiantes y profesores poseen concepciones equivocadas sobre el significado de la estimación por intervalos de confianza. La concepción que esos intervalos contienen la media muestral o valores particulares de la población en lugar de posibles valores del parámetro poblacional, y la interpretación del nivel de significación como una medida de certidumbre, sin referente alguno sobre la frecuencia, fueron las más frecuentes concepciones erradas.

Palabras clave: Población, Muestra, Intervalos de confianza, Concepciones erradas, Nivel de significación, Certidumbre.

1. Introduction

The confidence intervals comprise of the statistical inference and are very useful to estimate population parameters from a sample of values of the population. The CIs contain a set of possible values of the searched parameter and have a confidence measurement associated that is responsible for the probability of finding the searched value among them.

Although it is certain that the CI in principle can help to surpass some of the errors that can be committed when the hypothesis tests and p-values are used to decide on certain hypothesis around the searched parameter (Fidler, 2005), it is also certain the fact that the students have bad conceptions around them. (Bower, 2003; Fidler, 2005). A very usual one is to interpret CIs as a statistical descriptive object, in the sense of containing possible values of the sample mean or of being a range of individual values ("95% of the population values are in the interval"), and not as a statistical inferential object that provides information about the value of a population parameter (Fidler, 2005).

But not only students misconstrue CIs. In an e-mail survey carried out by Cumming, Williams and Fidler (2004), a group of researchers and authors of scientific papers were asked to estimate 10 future sample means from a given sample mean and a 95% CI. 78% of the subjects (105 of 134) placed 9 of the means inside the interval. They made it clear that they had tried to adjust the number of future means within the interval to the confidence level (95%). That is, the confidence level is understood as the probability that the CI contain the values of the sample means.

Concerning the relations between interval size with confidence level and the sample size, it has been reported that students are not clear on the direct or inverse relations between the interval width, the sample size and the confidence level. In fact, Fidler (2005) informs that only 16% of 180 examined students answered that increasing the sample size decreases the interval width, and 73% answered that the 90% CI is wider than the 95% CI .

With this outlook of difficulties we ask ourselves, what about the professors? More precisely, what meaning do professors and experts have regarding confidence intervals? To provide an answer, we conducted a research that allows us to know the exact nature of the conceptions that statistics professors have about confidence intervals. Part of the results of this research are presented in this paper.

In order to interpret the obtained results we use *conceptual analysis* (Glassersfeld, 1995) which attempts to build a model or conceptual scheme that is compatible with the observed outcomes. In other words, it tries to endow a person with a conceptual structure which, acting in accordance with the subject, responds in a similar fashion to the way this person did. Therefore, our explanations are but hypotheses about the motives that drive people to answer like they did, put forward for corroboration or disproof by further studies who would look deeper into the logic of our results.

2. Methodology

In order to contrast the results we identified two target populations: one of ``experts", comprised of professionals devoted to statistics, or its teaching, and senior statistics students. A second population is constituted by the students: active engineering or business administration students who have formally studied statistical inference. These students had previously taken two statistics courses on a three-hours-per-week schedule during two academic semesters.

Given the fact that is not easy to get a professor or professional to participate in an activity in which their knowledge is evaluated, it was decided to offer the participants some incentives: a set of materials and articles on the teaching of statistics, granted to the first persons to show up at the event and fill out the test. 41 experts did so anonymously and voluntarily.

The sample of students is made up of Engineering and Business Administration undergraduate students of several universities in Cali (Colombia). Most of the 297 students that partook in the study received classes from the professors who answered the test. We used a test with 24 closed items. In this paper we consider only 14 of these, written in True or False fashion. The items are as related to the conceptual nature of the confidence intervals as they are to the relations of their size and confidence level, the sample size, and the population variability. Items connected with the definition of confidence intervals and their respective percentages of wrong answers in the two populations are shown in Table 1.

The statement for the first set of items is as follows:

For a set of weights (in pounds) a 95% confidence interval was constructed for the mean using a random sample. This turned out to be (42; 48). Answer true or false, according to your criteria, questions of Table 1.

Table 1. *Items and percentages of wrong answers related with definition of CIs.*

		Percentages of Wrong Answers	
	Items	***Experts***	***Students***
1	*95% of weights are between 42 and 48 pounds (F)*	*28,7*	*50,2*
2	*Most of weights are between 42 and 48 pounds (F)*	*57,4*	*84,2*
3	*The probability that the interval includes the sample mean is 95% (F)*	*36,2*	*52,9*
4	*If 200 CIs of the same process are generated, approximately 10 of them will not contain the population mean (T)*	*51,1*	*64,3*
5	*The probability that the interval includes the population mean is 95%*	*40,4*	*56,2*
6	*A CI for a mean always contains the sample mean (T)*	*31,9*	*30,6*

Additionally, we included a problem dealing with the comparison of two populations in a meaningful context:

Do cows that listen music show an increased production of milk? An experiment was carried out to answer this question; a set of milking cows was divided at random into two groups. Music was played for one of the groups; the other group (control group) did not listen to any music. The mean increase in milk production (cows with music - cows without music) was 2.5 liters/cow during the time of evaluation. A 95% confidence interval for the mean difference was (1.5; 3.5) liters/cow.

Answer true or false, according to your criteria question of Table 2.

Table 2. *Questions and wrong answer percentages for the problem of milking cows.*

	Items	Percentages of Wrong Answers	
		Experts	**Students**
7	*95% of cows increased their production between 1.5 and 3.5 litre /cow (F).*	31,9	54,5
8	*We are 95% sure that the mean increase in production in the sample is 2.5 litre/ cow (F).*	42,6	44,2
9	*We are 95% sure that the true mean increase in production for all cows is between 1.5 and 3.5 litre/cow (T).*	14,9	35,7

Next are the questions that sought to know the understanding about the relationships between the ICs size and the sample size, confidence level and the standard deviation of population.

The items and the percentages of wrong answers are shown in the Table 3.

Table 3. *Percentages of wrong answers to questions aimed at the relation between CI widths and some concepts that define them.*

	Items	Percentages of Wrong Answers	
		Experts	Students
10	*If we keep the sample size fixed, the confidence interval becomes wider when we increase the confidence level (T).*	36,2	47,8
11	*If we fix the confidence level, the confidence interval becomes narrower when we increase the sample size (T).*	17,0	48,5
12	*If the standard deviation of the population increases, the width of the confidence interval decreases (F).*	8,5	43,4
13	*The width of the interval can be made smaller by reducing the confidence level (T).*	46,8	48,5
14	*The interval's width can be reduced by reducing the standard deviation of the population (T).*	10,6	38,7

3. Results and Discussion

From now on (*n*) will denote the *n*-th item, as numbered in the tables.

On the first item, roughly a 30% of the experts and half of the students assumed that CI is a sort of truncated range of population values, confirming Fidler's results (2005). This is corroborated in question in the milking cows problem, (7). The confidence level is assumed by them to be a percentage of population values that are contained in the confidence interval. In this sense, it is quite compelling that when the question is asked in terms of the majority (2) instead of a 95%, the percentage of experts that respond affirmatively goes beyond half while that of students reaches almost 85%.

This seems to be the manifestation of a bad conception that present the students with respect to the sample distributions in the sense that they confuse the distribution of the sample with the mean sample distribution. (Saldaña, 2004; Inzunsa, 2006). It could be that by not perceiving the sample mean variability (perhaps because only one value for this mean is available), but perceiving the variability of population (several population values are available, as many as the size of the collected sample), they end up linking confidence intervals to population values.

In accordance with the items (1) and (7), 32% of experts do not accept that the sample mean is contained in the confidence interval (6). In contrast, the percentage of students that does not accept this fact is 30%.

In order to explain the fact that a third of the experts does not consider that the mean sample belongs at the confidence interval, it would be possible to think that they considered the mean sample like a random variable and, therefore, they assume the confidence interval centered in the population mean which is also the mean of the distribution of the sample means.

The percentages of erroneous answers to the items (3), (5), (8) and in less acute form to the item (9), demonstrate that the experts and the students do not interpret in correct form the level of confidence associated at the interval. It would be possible to say that they assume this level like a certainty measurement that is applicable indifferently to all affirmation that makes reference to values of averages without mattering if they are sample o population mean. In item 6, where any reference to the confidence level is omitted and the question is given in a general form, close to a third of the experts and students do not admit that the mean sample is part of the interval constructed from it. Perhaps is the same again: the mean sample is assumed to be a random variable.

The answers to the fourth item evidences the lack of frequency interpretation of the confidence level or of the probability in the studied populations. More than half of the experts and 65% of the students deny that in the long run, if the sample were remade many times, such intervals, in a percentage equal to the confidence level would include the population parameter μ, allowing thus the existence of some intervals that do not include it.

After looking at the relationships between level of confidence, sample size and population variability, which ultimately examines the understanding of CIs’ construction process, a series of facts came out. Approximately 40% of experts and almost half the students are not clear on the fact that the interval width increases when the confidence level is raised (10). Formulating the question in descending terms (13) the percentage of wrong answers goes up to almost a 50% among the experts whereas it still keeps close to 50% among the students, which shows a low consistency in the answers of the former.

At the same time, 17% of the experts and up to half of the students do not understand that the relation between the interval width and sample size is inverse, namely, increasing the sample decreases the size of the interval, (11). These results are considerably better than those reported by Fidler (2005), who found that only 16% (of 180 students) answered that increasing the sample size diminishes the interval size.

The effect of population variability on the CIs shows the best results: around 90% of the experts and 60% of the students identified that when the variability in the population increases the interval size also increases, (12) and (14).

4. Analysis and Conclusions

Although we knew beforehand the misconceptions common among students about intervals of confidence, this study points out to a reality that must be addressed with priority: their misconceptions are often born of their own teachers' misconceptions.

These results allow us to note that both experts and students have misconceptions both at the conceptual level of confidence intervals as their relations with the sample size and level of confidence.

It is assumed that the confidence intervals are descriptive statistics in the sense that that contain possible values of the average sample or simply population values, and not inferential statistical that estimate the value of a population parameter. The confidence level is interpreted as a measure of certainty various assertions about confidence intervals without any reference frequency.

Similarly, a high percentage of experts and students assume that the size of the interval is directly related to the confidence level. Although in a less degree for experts, both samples assume that the size of the intervals is directly proportional to the size of the sample.

To explain the fact that the experts, rather than the student, who do not always consider the confidence interval contains the sample mean, we present a theoretical analysis of how to get the confidence intervals.

Let X be a random variable with distribution N (μ, σ) and let (X_1, X_2, X_3, .. , X_n) be a random sample of the population. For simplicity, suppose that σ is known. Thus,

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (1)$$

From (1) and by a standardization process is obtained the expression

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad (2)$$

From (2) and assuming a confidence of 1-α is obtained the following expression

$$P\left\{-Z_{1-\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < Z_{1-\frac{\alpha}{2}}\right\} = 1 - \alpha \quad (3)$$

Where $\Phi\left(Z_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$ being $\Phi(.)$ the cumulative distribution function of probability for standard normal distribution.

From (3), solved properly are obtained two separate intervals for $\bar{X}_n$ and μ:

$$P\left\{\mu - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \quad (4)$$

$$P\left\{\bar{X}_n - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \quad (5)$$

The procedure leads to two intervals that have the same level of confidence: one centered in μ, in (4), which refers to the probability of finding values $\bar{X}_n$ in this interval (at some distance from the parameter μ) and the other centered in $\bar{X}_n$, in (5), which refers to the probability that the parameter μ is in this interval (at some distance of the value $\bar{X}_n$). The first interval, denoted as $\mu \pm \rho$, evidences that the meaning of the margin of error ρ, because when is said that $\bar{X}_n$ belongs to this interval it is said that the maximum error, when it is used as an estimator of μ, is ρ. When μ is known, this is the kind of interval that refers to possible values of the sample mean and that is generated when sampling distributions are studied. The second interval, denoted as $\bar{X}_n \pm \rho$, is the interval with a confidence of 100(1-α)% which contains the parameter μ, that means, it is the confidence interval. While the two intervals are obtained from (3), they are by no means equal or have not the same meaning. Perhaps the misinterpretations that associate the confidence intervals with the sample mean are consequence of confusing these two intervals.

Additional to this mathematical explication, we would have to mention, also, the confusions that can arise in a process based on the behavior of many samples when only one sample is available. This situation can be the reason to interpret the probability value as a measure of certainty away from any frequency interpretation. In fact, the frequency and subjective or epistemological interpretations of probability that appear to be incompatible constitute an essential part of prediction and inference processes where a probability measure is necessary. This duality which, according to Hacking (1995) hindered the earlier development of probability theory, is reflected on Konold's *outcome approach* (1989), who found that people, upon being asked to assign a probability value to an event, do so by thinking about the degree of certainty they have of outcome on the next rerun of the experiment which defines the random event. This interpretation stands in the way for the adoption of a frequency approach to probability. We believe that this approach deserves a more profound study, for we are convinced that it is related to wrong conceptions about sampling distributions, and, in particular, intervals of confidence.

In summary, one could say that the full understanding of the confidence intervals requires the integration of the particular and general simultaneously. The particular is given by the sample and the sample mean, and the general by the distribution of all possible samples. The sample provides the center of the interval and assists in determining its size, and the distribution, given the confidence level, provides as missing to determine the size of the interval.

5. References

1. Bower, D.M. (2003). Some Misconceptions about Confidence Intervals. *Six Sigma Forum-American Society for Quality,* July.

2. Cumming, G., & Williams, J., & Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. Understanding Statistics, 3, 299 – 311.

3. Fidler, F. (2005). From Statistical Significance to Effect Estimation. Ph.D. Thesis, unpublished. Universidad La Trobe, Melbourne, Australia.

4. Glasersfeld, E. (1995). Radical constructivism: A way of knowing and learning. London: Falmer Press.

5. Hacking, I. (1995). *El surgimiento de la probabilidad*. Editorial Gedisa, España. [Translated of the original in english: *Emergence of Probability*, 1975, Cambridge University Press].

6. Insunza, S. (2006). Significados que estudiantes universitarios atribuyen a las distribuciones muestrales en un ambiente de simulación computacional y estadística dinámica. Ph.D. Thesis, unpublished. Centro de Investigaciones y de Estudios Avanzados, CINVESTAV-IPN, México.

7. Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6(1*), 59-98.

8. Saldaña, L. (2004). "Is this sample unusual?" An investigation of students exploring connections between sampling distributions and statistical inference. PhD Thesis, unpublished. Vanderbilt University, USA.