http://doi.org/10.14393/ER-v31e2024-03

Future Mathematics and Statistics teachers' understanding of confidence intervals about the arithmetic average[1]

Nicolás Sánchez Acevedo[2]

ABSTRACT

Confidence intervals are an important aspect of statistical inference. Currently this topic has been included into the curriculum of Mathematics teacher training and secondary school study programs in Chile. According to this subject, the understanding that future Mathematics and Statistics teachers express about confidence intervals on the arithmetic average will be explored. A qualitative and exploratory-descriptive methodology has been adopted. To do this, the responses of 11 future Mathematics and Statistics teachers, who were taking the statistical inference course, will be analyzed based on an open-response questionnaire with two activities on confidence intervals for the arithmetic average and the differences of those averages. The results show confusion in the calculation of confidence intervals for the average, as the deterministic interpretations and errors in the determination of Z or T quantiles. Despite this, the vast majority of future teachers are aware of the assumptions underlying the construction of these confidence intervals. These results are relevant, both for future teachers and for mathematics teacher trainers, given the few studies on this topic in the teachers' context.

KEYWORDS: Confidence intervals; Future Mathematics teachers; Statistics; Qualitative.

[1] English version by Angel Aliaga Sepúlveda. *E-mail*: mr.aaliaga@gmail.com.

[2] Dr(c) en Ciencias, área de Matemática Educativa (CICATA-IPN, México). Universidad Alberto Hurtado/Universidad Central de Chile, Santiago, Chile. Orcid: https://orcid.org/0000-0002-0665-6102. *E-mail*: nsanchez@uahurtado.cl.

http://doi.org/10.14393/ER-v31e2024-03

Comprensión que muestran futuros profesores de Matemática y estadística de los intervalos de confianza sobre la media

RESUMEN

Los intervalos de confianza son un aspecto importante en la inferencia estadística. Actualmente este tema se ha incluido en el curriculum de la formación de profesores de Matemática y los programas de estudio de secundaria en Chile. De acuerdo con esto, se explora en la comprensión que manifiestan futuros profesores de Matemática y Estadística sobre los intervalos de confianza sobre la media. Se ha adoptado una metodología cualitativa y de tipo exploratoria-descriptiva. Para ello se analizan las respuestas de 11 futuros profesores de Matemática y Estadística, que llevaban el curso de inferencia estadística, a partir de un cuestionario de respuesta abierta con dos actividades sobre intervalos de confianza para la media y diferencia de medias. Los resultados evidencian confusión en el cálculo de intervalos de confianza para la media, interpretaciones deterministas y errores en la determinación de cuantiles Z o T. A pesar de ello, la gran mayoría de los futuros profesores es consciente de los supuestos que subyacen a la construcción de estos intervalos de confianza. Estos resultados son relevantes, tanto para los futuros profesores, como formadores de profesores de Matemáticas, dado los pocos estudios sobre este tema en el contexto del profesorado.

PALABRAS CLAVE: Intervalos de confianza; Futuros profesores de matemática; Estadísticas; Cualitativo.

Compreensão de futuros professores de Matemática e Estatística em relação aos intervalos de confiança para a média

RESUMO

Os intervalos de confiança, assim como os testes de hipóteses, desempenham um papel importante na inferência estatística. Atualmente, este tema foi incorporado no currículo da formação de professores de Matemática e nos programas de estudo do ensino secundário no Chile. Nesse sentido, investigamos a compreensão demonstrada pelos futuros professores de Matemática e Estatística em

http://doi.org/10.14393/ER-v31e2024-03

relação aos intervalos de confiança para a média. Adotamos uma metodologia qualitativa de natureza exploratória-descritiva. Para isso, analisamos as respostas de 11 futuros professores de Matemática e Estatística, que estavam cursando um curso de inferência estatística, por meio de um questionário de resposta aberta com duas atividades relacionadas aos intervalos de confiança para a média e diferenças de médias. Os resultados revelam confusão no cálculo dos intervalos de confiança para a média, interpretações determinísticas e erros na determinação dos quantis Z ou T. No entanto, a grande maioria dos futuros professores está ciente das premissas subjacentes à construção desses intervalos de confiança. Esses resultados são relevantes tanto para os futuros professores quanto para os formadores de professores de Matemática, dado o escasso número de estudos sobre esse tema no contexto do ensino.

PALAVRAS-CHAVE: Intervalos de confiança; Futuros professores de matemática; Estatísticas; Qualitativo.

* * *

Introduction

The learning and teaching of statistics have gained significant importance due to their role in fostering critical citizenship (MINEDUC, 2015; NCTM, 2000). The growing significance of statistics in society, driven by the massive amount of available information, has led various researchers to engage in research across diverse professional application areas, including basic sciences, engineering, secondary education, pre-service and in-service teachers. Their goal is to delve into the difficulties and comprehension of various statistical concepts and procedures. The intention behind these efforts is to gather evidence of these challenges and, in doing so, propose alternative ways to construct teaching and learning situations that promote a deeper conceptual understanding in this field of knowledge (Shaughnessy, 2007) and develop statistical thinking.

As a part of statistical thinking, inference plays a pivotal role in understanding population characteristics. In this context, confidence intervals serve as a fundamental mechanism for estimating population characteristics under investigation and complement hypothesis testing (Roldán, Batanero; Álvarez-Arroyo, 2020). Both confidence intervals and hypothesis testing are integral components of the statistical mindset as a methodological discipline.

Today, the treatment of confidence intervals is promoted at the professional level to inform and comprehend various phenomena (Coulson, Healey, Fidel; Cumming, 2010; Yaremko, Harari, Harrison; Lynn, 2013) even more so than hypothesis testing.

Given the significance of confidence intervals [and hypothesis testing] in the development of statistical literacy (Gal, 2002), these topics have been incorporated into the teaching guidelines of various curricula at both the school and university levels (e.g., Franklin et al., 2007; MEC, 2015; MINEDUC, 2021a; MINEDUC, 2021b; NCTM, 2000).

In Chile, the inclusion of the Statistics and Probability unit (since 2009) in the mathematics curriculum (MINEDUC, 2012; MINEDUC, 2021a; MINEDUC, 2021b) encompasses content ranging from descriptive statistics in the early school years to the concept of sampling distribution, estimation (point and interval estimation), and hypothesis testing for the population mean, assuming known variance (MINEDUC, 2021b). The integration of these topics into the curriculum for student learning prompted the need for curriculum reforms in the training of future mathematics teachers in Chile, ensuring that their education aligns with the needs of society. As a result, most university programs for teacher preparation offer at least two statistics courses, with some delving deeper into the field, including up to four disciplinary courses[3], which may include courses on the teaching of statistics and probability to strengthen the aspects of statistics and its instruction.

[3] In Chile, among the currently active teacher education programs for mathematics teachers in higher education institutions, only one of them offers a disciplinary specialization in the field of statistics. This teacher education program, in addition to including courses in descriptive statistics, probability, and statistical inference, features a course titled "Statistical Models," which covers classical linear models (OLS and RLM) and some non-linear models for the context of teaching and learning.

The importance of promoting learning about statistics in general, and confidence intervals in particular, has been recognized in the Chilean secondary education curriculum (MINEDUC, 2021a) in the 3rd (16 to 17 years) and 4th-year (17 to 18 years) study program. It declares: (i) "that students understand the use of inferential statistics and how to estimate population parameters from sample statistics" (MINEDUC, 2021a, p. 159) and (ii) "that students can calculate confidence intervals to make inferences about the mean of a population in different contexts [...] according to the required level of confidence" (MINEDUC, 2021a, p. 165).

To promote student learning in accordance with the stated learning objectives, the Pedagogical and Disciplinary Standards for Mathematics Teacher Education[4] (MINEDUC, 2021b) propose that future mathematics teachers must be able to "facilitate classroom discussions to monitor various levels of reasoning and the difficulties their students encounter when interpreting confidence intervals in statistical inference problems" (MINEDUC, 2021b, p. 89).

Both the 3rd and 4th-year study program (MINEDUC, 2021a) and the Pedagogical and Disciplinary Standards for Mathematics Teacher Education[5] (MINEDUC, 2021b) demonstrate coherence between what mathematics teachers should know and what they should promote in students to understand these contents. Notably, the Pedagogical and Disciplinary Standards for Mathematics Teacher Education (MINEDUC, 2021b) lack explicit elucidation of the assumptions that must be met in constructing confidence intervals for population parameters μ and p.

Since confidence intervals are a relatively new topic in the Chilean curriculum, there is a demand to explore the understanding of

[4] In Chile, teacher education in various disciplines is structured according to the Initial Teacher Training Standards (https://estandaresdocentes.mineduc.cl/), which "aims to strengthen the training of teachers and ensure that educators are better prepared to meet the significant challenges presented by this dynamic and globalized knowledge society" (MINEDUC, 2021, p. 3).

[5] In Chile, teacher education in various disciplines is regulated by the Initial Teacher Training Standards (https://estandaresdocentes.mineduc.cl/), which have the purpose of strengthening the training of educators and ensuring that teachers are better prepared to confront the significant challenges posed by this dynamic and globalized knowledge society (MINEDUC, 2021, p. 3).

future mathematics teachers in instructional contexts. Existing research in this field is scarce, both in teacher education and in-service teaching, as much of the studies on confidence intervals have focused on engineering students or professionals from other areas (as described in the theoretical foundation section).

On this regards, the objective of this exploratory research is to assess the comprehension demonstrated by future mathematics and statistics teachers related confidence intervals.

Theoretical foundations

Currently, research in statistical education has significantly increased. However, studies related to the understanding of confidence intervals are scarce. Those researchers have been focused on professional investigators, university students, and other kind of professionals.

One study that involved educational professionals was conducted by Behar (2001), who explored the understanding of confidence intervals in 47 statistical professionals who were also teachers and 297 engineering students. They were given a questionnaire on confidence intervals and hypothesis testing. From the responses, it was found that 29% of the statistical professionals and 59% of the students considered the confidence level to represent the percentage within which population data falls rather than the parameter. Some errors included giving Bayesian interpretations, establishing direct proportional relationships between the interval width and sample size, and not recognizing the utility of confidence intervals in making decisions about hypotheses, as they did not view a confidence interval as a range of possible values for the true population parameter.

In the context of research with researchers, Cumming, William, and Fidler (2004) conducted a study involving 134 researchers working with confidence intervals. They were provided with a 95% confidence interval to estimate the population average and were asked to provide plausible values for the mean in nine experiments. Among these researchers, 78% believed that the confidence interval would contain the true parameter 95% of the time, revealing misconceptions regarding the constancy of confidence interval limits.

Due to the misuse and misinterpretation of hypothesis testing, Hoekstra, Morey, Rouder, and Wagenmakers (2014) asked 120 researchers and 442 psychology students to assess the accuracy of six statements related to the interpretation of confidence intervals, all of which were false. However, both researchers and students, on average, endorsed three or more false statements, indicating an inadequate understanding of confidence intervals. Surprisingly, a few researchers did not exhibit a better understanding than the students, even though those students had not taken courses in statistical inference. These results suggest that many researchers lack a correct interpretation of confidence intervals, which is a serious problem when making data-based decisions.

Rashidah, Razak, Baharun, and Arul (2018) conducted a research on confidence intervals but within the context of estimating parameters in a simple linear regression model. The study involved 197 second-year students in statistics and actuarial science, who were given a questionnaire to calculate and interpret confidence intervals for the slope parameterβ_0. Only 48% of the students correctly calculated the confidence interval for the slope parameterβ_1, and 68.5% of the students provided incorrect interpretations of the estimated slope. Errors included inadequate identification of degrees of freedom and improper evaluation of the standard error of the slope.

Within the context of research involving students at various levels, Yañez and Behar (2000) conducted a study to explore potential explanations for the misconceptions held by students and teachers regarding confidence intervals and levels of confidence. The study used epistemological analysis and clinical interviews. The results showed that misconceptions were primarily due to difficulties in the statistical-algebraic treatment of confidence interval construction and an incorrect extrapolation of the probability of the confidence level to the true parameter being within the given confidence interval.

Batanero and Olivo (2007) assessed the understanding of confidence intervals among 48 engineering students. Data were collected through a multiple-choice questionnaire. The results indicated a high conceptual understanding of confidence intervals but revealed procedural and conceptual errors, such as confusing the sampling distribution, failing to determine the critical value, and mistaking the statistic for the parameter.

Olivo, Batanero, and Díaz (2008) conducted research to evaluate the comprehension of university students majoring in engineering regarding confidence intervals. A questionnaire with 12 items and two open-ended situations was administered to 252 students with prior exposure to confidence intervals. The results showed a lack of understanding regarding the randomness of the interval endpoints, an inability to attribute interval width to various factors, incorrect selection of the appropriate sampling distribution based on the provided information, and errors in determining critical values.

Henriques (2016) studied the difficulties in understanding confidence intervals in a sample of 33 second-year naval school students through a written test as part of an introductory statistics course. The test included four multiple-choice questions and five open-ended questions related to confidence intervals. While the results indicated that students understood the concept of confidence intervals, they also revealed some procedural and conceptual difficulties. These included difficulties with defining a confidence interval, interpreting the descriptive meaning of the confidence interval, selecting the appropriate sampling distribution (particularly for variance), and determining critical values.

Crooks, Bartel, and Alibali (2019) conducted a study with undergraduate and postgraduate students in sociology to assess their conceptual understanding of confidence intervals. The results revealed conceptual errors in the treatment of confidence intervals, a lack of connection between confidence intervals and the estimation of the sample mean, and a disconnect between confidence intervals and hypothesis testing. These findings suggest a need for a deeper exploration of basic elements of confidence intervals.

Roldán et al. (2020) conducted research to evaluate the understanding of confidence intervals among high school students through a questionnaire that included both multiple-choice and open-response items. The results demonstrated limited comprehension of confidence intervals for parameter estimation, with no more than 40% of the evaluated students (58) successfully completing the activities.

Based on the reported research and others (e.g., Andrade; Fernández, 2016; Reaburn, 2014), there is a scarcity, if not a complete absence, of research on the understanding of confidence intervals in the context of future mathematics teachers. This presents an opportunity for exploration within this context.

An Approach to the Concept of Confidence Interval

One of the relevant aspects in statistics is to transform data from a sample into useful information for making inferences about population characteristics and decisions. Considering that populations are characterized by descriptive measures known as parameters, statistical research seeks to calculate the value of these parameters (Wackerly; Mendenhall; Scheaffer, 2010).

Confidence intervals are not an easily comprehensible concept. In the case of frequentist interpretation, it is established that the confidence interval for the $(100-\alpha)\%$ allows characterizing a process of generating intervals in which, through the repetition of the process, the $(100-\alpha)\%$ of the generated intervals contain the true unknown parameter. [...] The typical treatment of confidence intervals often begins with a sampling distribution of a statistic, such as the sample mean (Engel, 2010).

In general, in the case of classical [parametric] inference, unbiased estimators are assumed, for example, for μ, p, $\mu_1-\mu_2$, p_1-p_2 where it is assumed that, for large samples, point estimators have approximately normal sampling distributions with specific standard errors[6], and furthermore, independence of samples is assumed. That is, for large samples $Z=\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}$, approximately follows a standard normal distribution, serving as a pivotal quantity for the confidence interval of the target parameter θ.

For the case of this research, we extensively present, considering the assumptions, the confidence intervals for the mean (with an unknown σ), and for the difference of means in the case of unknown and unequal variances. Both interval estimations are described as follows in Wackerly et al. (2010, pp. 425-428):

- For a confidence interval of μ when is unknown, we assume a random sample $X_1, X_2, \ldots, X_n$ from a normal population with mean $\bar{X}$and estimated sample varianceS^2. The confidence interval for the population meansμ, when $V(Y_i)=\sigma^2$ is unknown and the sample size is small, involves using the pivotal quantity $T=\frac{X-\mu}{S/\sqrt{n}}$, which follows a T-distribution with $(n-1)$ degrees of freedom. By choosing appropriate values for $t\alpha/2$ and $-t\alpha/2$, one can construct the interval and$P\left(-t\alpha/_2 \leq T \leq t\alpha/_2\right)=1-\alpha$, after some algebraic manipulation,

[6] Ver Wackerly et al. (2010, p. 397)

obtain the confidence interval for μ with $\bar{X} \pm t_{\alpha/2,\ n-1} \cdot \left(\frac{S}{\sqrt{n}}\right)$. It is worth noting that an alternative approach to interval estimation for μ involves using the Central Limit Theorem (CLT), which states that if we have a sequence $X_1, X_2, \ldots, X_n$ of random variables originating from a sample of a distribution with mean μ and variance σ^2, for sufficiently large n ($n \geq 30$), the sampling distribution of $\bar{X}$ is approximately normal with mean $\mu_{\bar{X}} = \mu$and $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ (Devore, 2012, p. 226). In this work, the CLT response is not considered, as it is not part of the objectives of the inference course for future mathematics teachers.

- In Walpole et al. (2012, pp. 289-290), the confidence interval for the difference of means$\mu_1 - \mu_2$ in the case of having unknown and unequal variances is described as follows: If $\bar{X}_1$ and s_1^2 and, $\bar{X}_2$and s_2^2 are the means and variances of independent random samples of sizesn_1 and n_2, respectively, taken from approximately normal populations with unknown and different variances, an approximate confidence interval for $100(1-\alpha)\%$is given by $\bar{x}_1 - \bar{x}_2 \pm t_{v,\ \alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. The degrees of freedom $v$are approximated[7].

Methodology

According to the objective set on this research, this study is based on a qualitative methodology (STAKE, 2007) of an exploratory-descriptive nature (Pérez-Serrano, 1994; Hernández; Fernández; Baptista, 2010). Our interest lies in delving into the detailed understanding displayed by a group

[7] $v = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left[\left(s_1^2/n_1\right)^2/(n_1-1)\right] + \left[\left(s_2^2/n_2\right)^2/(n_2-1)\right]}$ que son los grados de libertad estimados.

http://doi.org/10.14393/ER-v31e2024-03

of prospective mathematics and statistics teachers from a Chilean university regarding confidence intervals for the mean and the difference in means. To analyse the responses of the prospective teachers, we employ content analysis (Noguero, 2002), following these steps: (I) selection of the activities and respective prompts to be analysed, as each response can be broken down into components related to the comprehension of confidence intervals; (II) based on the elements of confidence interval comprehension discussed in the theoretical framework, an analysis of the responses of the prospective teachers is conducted, categorizing them as achieved, partially achieved, not achieved, or non-responsive; and (III) a descriptive analysis of the data is performed, utilizing tables to summarize and systematize the information obtained.

Prospective Teachers and the Training Context

The population (N=11) of teachers was purposefully selected and accessed due to the ease of reaching students enrolled in the Statistical Inference course and their willingness to collaborate in the research. The prospective teachers, consisting of 3 females (27.3%) and 8 males (72.7%), were in their 8th semester of the Pedagogy in Mathematics and Statistics program at a private institution of higher education that specializes in teacher training in Chile. At the time of responding to the open-response questionnaire (as part of a process evaluation), the students had completed courses in Descriptive Statistics and Probability. The Statistical Inference course serves as a prerequisite for the final course in the statistics track, called Statistical Models. As previously mentioned, this program is the only teacher training program for mathematics in Chile that delves into the field of statistics and its pedagogy.

The Statistical Inference course (where data was collected) consists of three major units of instruction: (I) the concept of estimation and methods of point estimation, covering parameter estimation methods (maximum likelihood and moments) and the properties of unbiasedness, efficiency, and consistency, as well as bias and mean square error; (II) interval estimation, which involves constructing confidence intervals for classic cases, addressing both procedural and conceptual aspects of confidence intervals for the mean, difference of means, proportion, difference of proportions, and variance; and finally, (III) hypothesis testing, encompassing conceptual aspects, types of errors (Type I and Type II), and classic hypothesis tests.

Activities on Confidence Intervals

As part of a process evaluation, a questionnaire containing two open-response activities on estimation through confidence intervals for the mean (μ) and the difference in means ($\mu_1 - \mu_2$), both in cases with unknown variances, was administered to the 11 students (the study population).

The questionnaire items were accompanied by prompts designed to assess the students' problem-solving strategies, argumentation skills, and their knowledge of the assumptions that underlie interval construction. All of this was done in consideration of the course's learning outcomes and the Initial Teacher Training Standards for Mathematics (MINEDUC, 2021b).

Following the activity selection method by Olivo et al. (2008), the two chosen activities align with the objectives of the statistical inference course. Activity 1, adapted from Wackerly et al. (2010, p. 412), assesses procedural knowledge in constructing and estimating confidence intervals for the mean of a normally distributed population with an unknown variance, as well as the conceptual understanding regarding interpretation and assumptions. Activity 2, adapted from Wackerly et al. (2010, p. 440), evaluates understanding of the assumptions necessary for calculating confidence intervals for differences in means and assesses both procedural and conceptual knowledge for the constructed confidence interval.

Activity 1: The purchase times of n = 64 randomly selected customers at a local supermarket were recorded. The sample mean and variance for the 64 purchase times were 33 minutes and 256 minutes, respectively. (a) Estimate μ, the true average purchase time per customer, with a confidence level of α. (b) Interpret the confidence interval in the context of the situation, and (c) What assumptions should be made for constructing the confidence interval?

Expected responses.

(a) Assuming a 90% confidence level and considering that the variance is unknown, what we have is an estimation of the population variance. Considering a $n = 64$, you obtain a $\bar{x} = 33\ min.$ and a $s^2 = 256\ min$. Using $\bar{X} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ under a T-Student distribution with $n - 1\ gl = 63\ gl$ degrees of freedom, the interval estimate for μ is $33 \pm 1{,}67 \cdot \left(\frac{16}{8}\right)$, which is $29{,}66 < \mu < 36{,}34$. **(b)** A suitable interpretation would be: It is expected that with 95% confidence, the true population parameter for the mean is between 29.66 minutes and 36.34 minutes, implying that, in the sampling, for example, 10% of the samples of size $n = 64$ will not contain the true average time μ of the population. **(c)** It is expected that future teachers (discussing aspects in accordance with the assumptions to be considered in class) reflect on the fact that the use of this confidence interval necessitates an estimation of the variance due to the characteristics of the situation presented. Additionally, they should consider that the random variable samples X_i come from normal distributions, or alternatively, the sample size is sufficiently large, and each of them is independent and identically distributed. In the latter case, the construction of the interval can be based on the Central Limit Theorem (CLT).

Activity 2: Two methods for teaching reading were applied to two randomly selected groups of elementary school children and were compared based on a reading comprehension exam at the end of the teaching period. The sample means and variances calculated from the exam scores are shown in the following table.

Estimator	Method 1	Method 2
Number of children in the group	11	14
$\bar{x}$	64	69
s^2	52	71

You are asked to (a) What assumptions are necessary? (b) Find a 95% confidence interval for $(\mu_1 - \mu_2)$ and interpret the interval in the context of the situation presented.

Expected responses.

(a) In the case of this prompt, students are expected to recognize that estimating the confidence interval necessitates knowing that the samples from the two populations are random and independent. Furthermore, they should acknowledge that these samples come from distributions that are approximately normal, where the equality of variances cannot be assumed.**(b)** Based on the assumptions in (a), the expression $\bar{x}_1 - \bar{x}_2 \pm t_{\nu,\ \alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ is used to estimate the true difference in means. The estimation of degrees of freedom is calculated for the case of unequal variances, $\nu \approx 22{,}79 = 23$ and with that, we have $(64 - 69) \pm 2{,}069 \cdot \sqrt{\frac{52}{11} + \frac{71}{14}}$. Solving for it yields $-11{,}476 < \mu_1 - \mu_2 < 1{,}477$. With this interval, a possible interpretation would be: It is expected that with 95% confidence, the true difference in means is between -11.476 and 1.477, or that the true difference in scores on the reading comprehension exam falls within these values. Some students may also notice the connection to hypothesis testing by mentioning that the performance of one group appears to be better than that of the other group.

http://doi.org/10.14393/ER-v31e2024-03

Results

In Table 1, we present the achievement percentages derived from the responses of the prospective teachers regarding prompts 1a, 1b, and 1c of Activity 1. Student achievement levels were categorized as achieved, partially achieved, not achieved, and no response based on their ability to: estimate the parameter μ, interpret the confidence interval, and understand the assumptions for construction.

TABLE 1: Percentages of achievement of activity 1.

Consignment by activity	Accomplished	Partially achieved	Unachieved	Does not respond
1ª	*0%*	81,82%	18,18%	0%
1b	*0%*	81,82%	18,18%	0%
1c	72,73%	18,18%	9,09%	0%

Source: Prepared by the authors.

The data in Table 1 reveals that all the teachers responded to the given tasks (1a, 1b, and 1c). Additionally, it can be observed that 72.73% of the future teachers are familiar with the assumptions required for constructing confidence intervals for the mean μ, which is a relevant aspect in the context of future mathematics teachers. Similarly, 18.18% are unable to construct the confidence interval and fail to provide a proper interpretation of it.

In Table 2, the percentage of achievement for tasks 2a and 2b in activity 2 is displayed, based on the predefined response levels (achieved, partially achieved, not achieved, and no response) for the necessary assumptions for constructing the confidence interval of the difference $(\mu_1 - \mu_2)$ and the procedure (proc.) and interpretation (int.) of the confidence interval.

TABLE 2: Percentages of achievement of activity 2.

Consignment by activity	Accomplished	Partially achieved	Unachieved	Does not respond
2ª	72,73%	0%	27,27%	0%
2b(proc.)	*45,45%*	27,27%	18,18%	9,1%
2b(int.)	*18,18%*	*45,45%*	18,18%	18,18%

Source: Prepared by the authors.

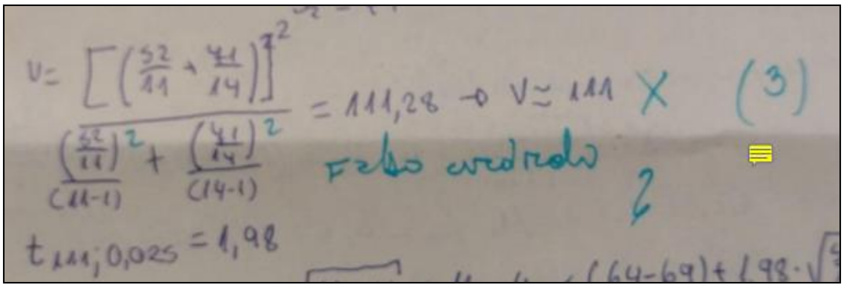
Regarding this second item, it's worth noting that 45.45% manage to determine the confidence interval for the difference of means, but only 18.18% interpret it correctly, with another 45.45% interpreting it partially.

Some of the correct procedures and errors made by future mathematics and statistics teachers were related to procedural aspects, for instance, assuming known variances when it wasn't explicitly stated, calculating the estimation of degrees of freedom for the interval of mean differences, procedural mistakes in constructing the confidence intervals, looking up percentiles in the standard normal or t-Student tables. As for conceptual aspects, they confused sample variance with population variance, used the confidence interval for known σ^2, interpreted the confidence interval deterministically, assumed without justification that variances are known for the confidence interval of the difference in means, and confused the confidence level with probability.

We present some examples of responses from future teachers regarding the calculation and interpretation of confidence intervals for the population mean and the difference in means.

One of the common errors made by the future teachers was the estimation of degrees of freedom in the confidence interval for the difference in means, particularly in the incorrect use of the formula $\nu = \frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\left[{\left(s_1^2/n_1\right)^2}/{(n_1-1)}\right]+\left[{\left(s_2^2/n_2\right)^2}/{(n_2-1)}\right]}$, as seen in Figure 1.

FIGURE 1: Error in the calculation of the estimation of the degrees of freedom

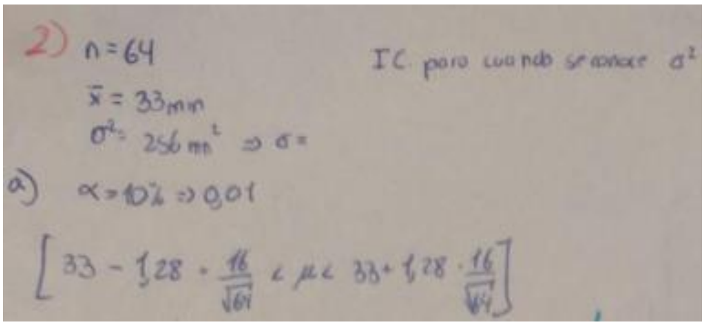


Source: Student response 6.

In this case, we can see that the procedural error stems from not considering the squares in the terms $(s_1^2/n_1)^2$ and $(s_2^2/n_2)^2$, which appear in the denominator of the formula, possibly because they thought they were already dealing with variance based on the data provided in the activity. This error leads the future teachers to obtain 111 degrees of freedom, and consequently, they fail to determine the correct quantile in the table of the T-Student distribution with 95% confidence.

Another error made by the future teachers is the determination of the critical value in constructing the confidence interval for the mean when the population variance is unknown (Figure 2).

FIGURE 2: Error in quantile identification



Source: Student response 1.

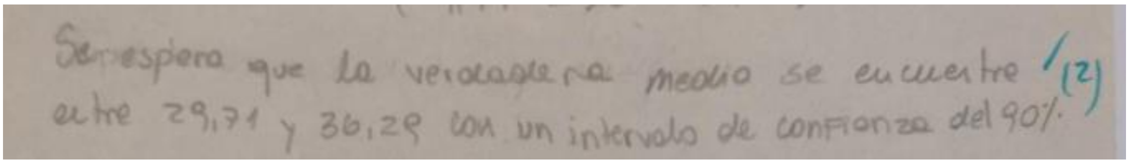
In this case, Student 1 assumes that the variance is known. Based on this, they look for the quantile in the standard normal distribution table but mistakenly overlook that the requested confidence interval is two-sided,

http://doi.org/10.14393/ER-v31e2024-03

which means the Z_α quantile needs to be split into each of the two tails, covering the α area, resulting in lower and upper quantiles $Z_{\alpha/2} = 1{,}64$.

Regarding the interpretation of the confidence intervals, as shown in Table 1 and Table 2, there were no future teachers who interpreted the constructed intervals correctly. In the case of the interpretation of the confidence interval for the mean, there is a partially achieved response (Figure 3), where the teacher assumed a known variance, which was repeated by the other future teachers.

FIGURE 3: Partially achieved interpretation of the CI for the mean


Se espera que la verdadera media se encuentre entre 29,71 y 36,29 con un intervalo de confianza del 90%. /(2)

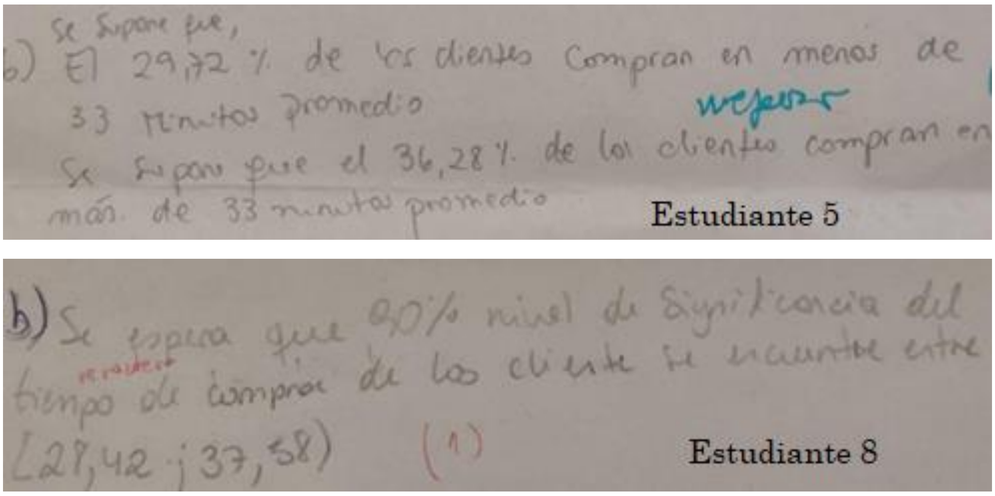
Source: Student response 2.

In this case, we can observe that the student provides a moderately correct interpretation, as she recognizes the possibility that the confidence interval, based on a sample, contains the true population mean. However, it's not clear what she means when mentioning "with a 90% confidence interval," as she may be confusing the confidence level with the area corresponding to the tails. This confusion may arise from the assumption that the student teacher is making that the distribution is normal.

Similarly, we find cases where future teachers provide a deterministic interpretation of the confidence intervals, both for the estimation of the mean μ and for the difference in means $\mu_1 - \mu_2$ (Figure 4).

http://doi.org/10.14393/ER-v31e2024-03

FIGURE 4: Deterministic interpretation of the confidence interval in both cases

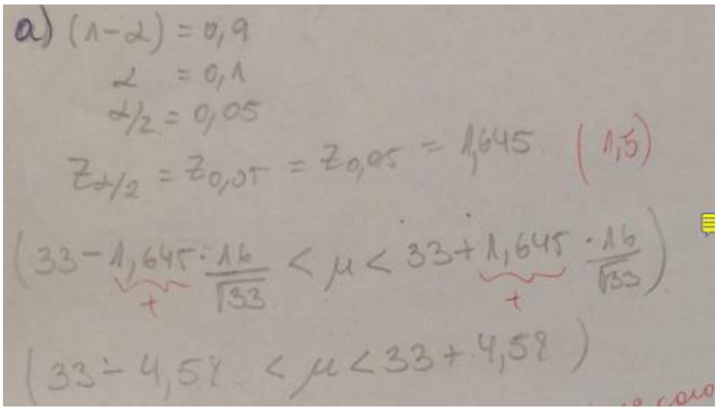


Source: Student response 5 and 8.

It can be observed that the future teacher (Student 5) interprets the confidence interval in terms of the sample mean, not the population mean. Furthermore, she interprets the confidence interval in terms of percentages for both the lower and upper bounds. She considers the regions that do not include the true population mean, namely the sample mean values less than $-Z\alpha/_2$ and those greater than $Z\alpha/_2$, without realizing that the estimation of the true population mean is within the region defined by $1-\alpha$. In the case of Student 8, while he includes the phrase "it is expected," he mistakenly believes that the 90% represents the significance level. Additionally, he does not specify that the time should refer to the true population average time and does not consider the randomness of the sampling intervals since he asserts that this average time will be (deterministically) between 28.42 and 37.58.

Another confusion observed (Student 8) is that in constructing the confidence interval for the mean with an unknown variance, he replaces $n = 64$ in the formula $\bar{X} \pm t\alpha/_2 \left(\frac{S}{\sqrt{n}}\right)$ with $n = 33$ (Figure 5).

FIGURE 5: Procedure error when confusing the sample size value with the mean value

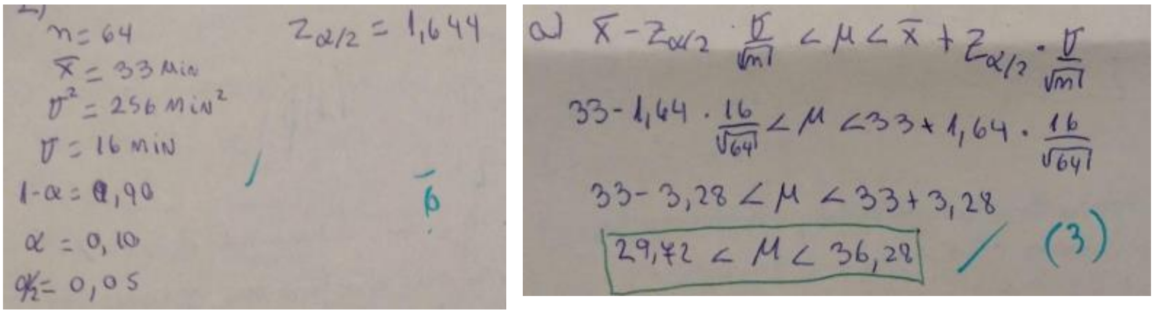


Source: Student response 8.

This error in replacing $n = 33$ in the confidence interval formula could be due to a slip-up, as the student correctly records the data provided in the activity statement. For instance, he accurately locates the value of the quantile Z (even though he assumes known variance), which might introduce more complexity, and the use of the normal distribution could be justified through the Central Limit Theorem.

An interesting aspect in the procedures of the future teachers is that 10 of them assume that the variance is known, $\sigma^2 = 256\ min.$, even though the mean $\bar{x} = 33\ min.$ and the variance $s^2 = 256\ min.$ are explicitly given for a sample of size $n = 64$ (Figure 6).

FIGURE 6: Procedure where it is assumed that the variance is known

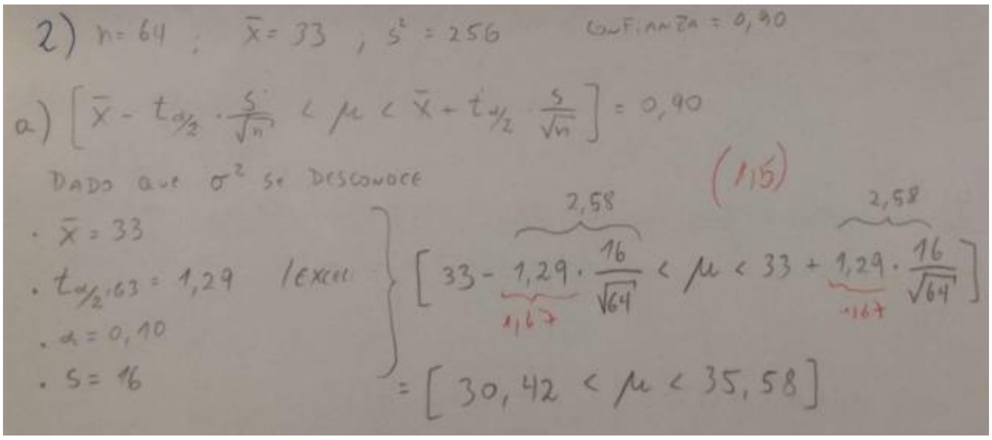


Source: Student response 6.

Although this consideration can be viewed as a conceptual error on the part of the students for not considering the estimation of the variance and, consequently, the use of the T-Student distribution, the procedures identified, when considering the normal distribution, are correct, representing an 82.82% of responses that are partially achieved in addressing item 1a.

On the contrary, a different case occurs with a future teacher who does not assume normality, meaning he identifies that $s^2 = 256$ is an estimate of the population variance (Figure 7).

FIGURE 7: Correct procedure with unknown variance, but does not determine the associated percentile

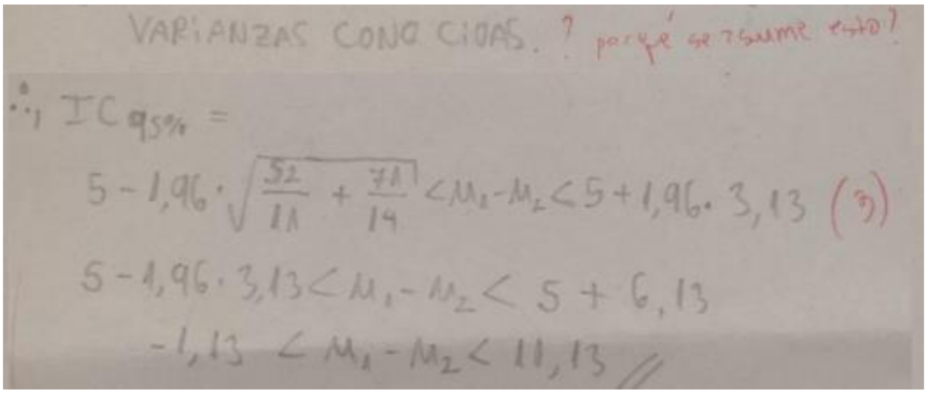


Source: Student response 6.

This future teacher, while correctly interpreting the context of the problem by identifying that $s^2 = 256$is an estimate of the population variance (Figure 7) and using the t-Student distribution, does not consider that the confidence interval is two-sided. He places the value of the percentile as $t_{\alpha,n-1} = t_{0,1,63} = 1{,}29$ instead of $t_{\alpha/2;\, n-1} = t_{0{,}05;\ 63} = 1{,}669$.

Finally, we encounter the case of the confidence interval for the difference in means with unknown and unequal variances. One student assumes, without further assumption, unknown but equal variances, which implies using an estimate of the variance S_p^2 under the T-Student distribution, where the degrees of freedom are determined based on the expression $n_1 + n_2 - 2$ (where $n_1$andn_2 are the sizes of each independent sample) from the two independent samples (Figure 8).

FIGURE 8: Construction of the confidence interval assuming known variances



Source: Student response 11.

It can be observed that this future teacher, firstly, incorrectly calculates the difference in means $\mu_1 - \mu_2 = 5$ when the correct difference is $\mu_1 - \mu_2 = -5$. Secondly, he only uses the expression $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ in the denominator of the pivotal quantity and does not calculate the weighted estimated variance for S_p^2, which indicates a difficulty in understanding phenomena where certain initial conditions can or cannot be assumed for the estimation, in this case, through intervals, to be more robust to the true population parameter.

Discussion and Conclusions

Concerning the objective of this research, which is to "assess the understanding of prospective Mathematics and Statistics teachers regarding confidence intervals," this study provides exploratory insights into how prospective Mathematics and Statistics teachers comprehend confidence intervals in terms of their construction, the use of definitions, underlying assumptions, and properties. While various studies have contributed to this topic in different professional contexts (e.g., Andrade; Fernández, 2016; Henriques, 2016; Reaburn, 2014; Roldán et al., 2020), this work contributes to the training of Mathematics teachers who will teach statistics. This is particularly relevant on the Chilean context, where, it is a productive area for research. The results reported in this research align with the new Initial Teacher Training Standards (MINEDUC, 2021b), which demand that future Mathematics teachers are capable of teaching statistics and probability, with confidence intervals being a topic covered in 11th and 12th grade (ages 16-18), justifying the assessment of prospective teachers' understanding.

As part of the findings, it is evident that prospective Mathematics and Statistics teachers face various challenges in their understanding of confidence intervals, both conceptually and procedurally. In general, they make errors in estimating degrees of freedom, overlook assumptions in the construction of confidence intervals – results like those reported in other studies (Behar, 2001; Coulson et al., 2010; Henriques, 2016) but in different disciplinary areas unrelated to teacher training.

The major prominent difficulty observed in prospective teachers is the misconception that the population variance formula is equivalent to the sample variance, even when the activities explicitly provide the sample size. This situation leads them to use the normal distribution rather than the T-Student distribution (which could be justified under the CLT).

Additionally, some prospective Mathematics and Statistics teachers exhibit deterministic interpretations, as reported in Roldán et al. (2020). These interpretations may be related to deterministic rather than statistical thinking. According to this, the prospective teachers do not consider that $(1-\alpha)\%$ determines the probability that the CI will contain the true population parameter (Cumming, William, Fidler; 2004). Furthermore, interpretations of confidence intervals as percentages emerge, meaning that the type of variable under study is not considered (Activity 1), and it is assumed that the upper and lower limits are percentages within which the true parameter must fall.

Procedural errors are consistent with those reported by Olivo, Batanero, and Díaz (2008), Rashidah et al. (2018), and Roldán et al. (2020). These errors include incorrect calculation of the quantile derived from the standard normal or T-Student table, incorrect estimation of degrees of freedom in the calculation of confidence intervals for the difference in means with unknown and unequal variances, failure to recognize that confidence intervals are two-tailed, miscalculations in the formula for confidence intervals for the mean (μ) or the difference in means $\mu_1 - \mu_2$, confusion between sample size and the value of the sample mean when substituting in the formula, and the omission of relevant relationships among elements that are part of confidence intervals. These findings align with previous research in areas unrelated to teacher training, such as Andrade; Fernández (2016), Crooks et al. (2019), and Roldán et al. (2020).

Both procedural and conceptual errors in the construction of confidence intervals and the resulting interpretations raise the need to review both the curriculum (MINEDUC, 2021a) and the Initial Teacher Training Standards (MINEDUC, 2021b). This is because these prospective Mathematics teachers make errors and have interpretations like those in other knowledge areas. However, the challenge lies in the fact that these "biased" understandings will later be taught in high school classrooms, likely perpetuating the same errors and modes of interpretation.

This initial exploratory approach, based on the findings, highlights the importance of focusing on Mathematics [and Statistics] teacher training, especially in a context where, as reported in various curriculum studies (e.g., Bakker; Derry, 2011; Del Pino; Estrella, 2012; Sánchez; Ruiz, 2022), statistics education is fragmented and emphasizes mechanistic work, often neglecting the promotion of underlying conceptual aspects of the discipline.

Recommendations stemming from this type of work can be directed in two main directions: (1) a shift toward statistics education that considers innovative research in the light of teaching and learning curriculum proposals, such as the GAISE project (Franklin et al., 2007). This could promote a culture of statistical reasoning (Garfield; Ben-zvi, 2009). (2) Attention should be directed towards teacher educators, as it is worth considering that one of the factors contributing to mechanistic and fragmented teaching (as reported) may be the way in which these teachers were educated themselves – by the educators who teach statistics in university courses for teacher training. This is not a criticism considering the results of this research, but it aims to stimulate reflection among teacher educators, especially those responsible for training future Mathematics teachers, who will teach statistics and must be able to meet the current demands of the education system and society.

References

ANDRADE, L.; FERNÁNDEZ, F. Interpretation of Confidence Interval Facing the Conflict. *Universal Journal of Educational Research*, v. 4, p, 2687–2700, 2016. DOI: http://dx.doi.org/10.13189/ujer.2016.041201.

BAKKER, A.; DERRY, J. Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, v. 13, n. 1–2, p. 5–26, 2011. DOI: http://dx.doi.org/10.1080/10986065.2011.538293.

BEHAR, R. *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*. Trabajo de Tesis Doctoral. Universidad Politécnica de Cataluña, 2011.

COULSON, M., HEALEY, M.; FIDLER, F.; CUMMING, G. Confidence intervals permit, but do not guarantee, better inference than statistical significance testing. *Frontiers in Psychology*, v. 1, n. 26, p. 1-9, 2010. DOI: https://doi.org/10.3389/fpsyg.2010.00026.

CROOKS, N. M.; BARTEL, A. N.; ALIBALI, M. W. Conceptual knowledge of confidence intervals in psychology undergraduate and graduate students. *Statistics Education Research Journal*, v. 18, n. 1, p. 46–62, 2019. DOI: https://doi.org/10.52041/serj.v18i1.149.

CUMMING, G.; WILLIAMS, J.; FIDLER, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics,* v. 18, n. 3, p. 299-311, 2004.

DEL PINO, G.; ESTRELLA, S. Educación estadística: relaciones con la matemática. *Pensamiento Educativo, Revista De Investigación Latinoamericana (PEL)*, v. 49, n. 1, p. 53–64, 2012. DOI: https://doi.org/10.7764/PEL.49.1.2012.5.

ENGEL, J. On teaching bootstrap confidence intervals. In C. READING (Ed.), *Proceedings of the eighth International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute. 2010. Disponible en: http://www.stat.auckland.ac.nz/~iase/publications.php.

FRANKLIN, C.; KADER, G.; MEWBORN, D.; MORENO, J.; PECK, R.; PERRY, M.; OTROS. *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association, 2007.

GAL, I. Adults' Statistical literacy: meanings, components, responsibilities. *International Statistical Review*, Nederlandn, n. 70, p. 1-25, 2002.

GARFIELD, J. B.; BEN-ZVI, D. Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics,* v. 31, n. 3, p. 72–77, 2009.

HENRIQUES, A. Students' difficulties in understanding of confidence intervals. En D. BEN-ZVI.; K. MAKAR (Eds.), *The Teaching and Learning of Statistics* (pp. 129-138). Cham, Switzerland: Springer, 2016.

HERNÁNDEZ, R.; FERNÁNDEZ, C.; BAPTISTA, P. Metodología de la Investigación. México: McGraw Hill, 2010.

HOEKSTRA, R.; MOREY, R. D.; ROUDER, J. N.; WAGENMAKERS, E. J. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, v. 21, p. 1157–1164, 2014.

MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE, MECD. *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.* Madrid: Autor, 2015.

MINEDUC. *Bases curriculares 2012. Matemática, educación básica*. Santiago: Ministerio de Educación, 2012.

MINEDUC. *Programa de Estudio Probabilidades y Estadística Descriptiva e Inferencial 3° y 4° medio – Formación Diferenciada*. Unidad de Currículum y Evaluación Ministerio de Educación, 2021a.

MINEDUC. *Estándares para la formación inicial docente de Matemáticas*. Unidad de Currículum y Evaluación Ministerio de Educación, 2021b.

NCTM. *Principles and standards for school mathematics*. Reston, VA: NCTM, 2000.

OLIVO, E.; BATANERO, C. Un estudio exploratorio de dificultades de comprensión del intervalo de confianza. *Union*, v. 12, p. 37–51, 2007.

OLIVO, E.; BATANERO, C., DÍAZ, C. Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación Matemática*, v. 20, n. 3, p. 5-32, 2008.

PÉREZ-SERRANO, G. Investigación cualitativa. Retos e interrogantes. Madrid: La Muralla, 1994.

RASHIDAH, N.; RAZAK, F. A.; BAHARUN, N.; ARUL, E. S. G. Investigating Students' Difficulties in Understanding Confidence Intervals in Linear Regression Models. *International Journal of Engineering & Technology*, v. 7, n. 4.33, p. 60-64, 2018. DOI: http://dx.doi.org/10.14419/ijet.v7i4.33.23485.

REABURN, R. Students' understanding of confidence intervals. In K. MAKAR, B. DE SOUSA.; R. GOULD (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: IASE, 2014.

ROLDÁN, A. F.; BATANERO, C.; ÁLVAREZ-ARROYO, R. Comprensión del intervalo de confianza por estudiantes de Bachillerato *Avances De Investigación En Educación Matemática*, v. 18, p. 103–117, 2020. DOI: https://doi.org/10.35763/aiem.v0i18.284.

SÁNCHEZ, N.; RUIZ, B. Análisis de las actividades que proponen dos libros de texto de educación primaria. Un acercamiento comparativo desde la perspectiva de la inferencia informal. *REXE- Revista De Estudios Y Experiencias En Educación*, v. 21, n. 46, p. 76–101, 2022. DOI: https://doi.org/10.21703/0718-5162.v21.n46.2022.004.

SHAUGHNESSY, J. M. Research on statistics learning and reasoning. In F. K. LESTER JR. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Greenwich, CT: Information Age/NCTM, 2007.

STAKE, R. *Investigación con estudio de casos*. 4. Ed. Madrid: Morata, S.L, 2007.

PÉREZ-SERRANO, G. *Investigación cualitativa: retos e interrogantes*. Madrid: La Muralla, 1994.

VÁSQUEZ, C.; Y ALSINA, Á. Enseñanza de la Probabilidad en educación primaria. Un desafío para la formación inicial y continua del profesorado. *Números. Revista de Didáctica de las Matemáticas*, v. 85, p. 5-23, 2014. Disponible en: http://www.sinewton.org/numeros/numeros/85/Articulos_01.pdf.

http://doi.org/10.14393/ER-v31e2024-03

YÁÑEZ, G.; BEHAR, R. Interpretaciones erradas del nivel de confianza en los intervalos de confianza y algunas explicaciones plausibles. En M. J. González, M. T. González y J. Murillo (Eds.). *Investigación en Educación Matemática XIII, Comunicaciones de los grupos de investigación*. XIII Simposio de la SEIEM. Santander: SEIEM. 2009. Disponible en: https://seiem.es/docs/comunicaciones/GruposXIII/depc/Yanez_Behar_R.pdf.

YAREMKO, R. M.; HARARI, H.; HARRISON, R. C.; LYNN, E. *Handbook of research and quantitative methods in psychology: For students and professionals*. Hilldale, NJ: Erlbaum, 2013.

WACKERLY, D. D.; MENDENHALL III, W.; SCHEAFFER, R. L. *estadística matemática con aplicaciones*. México, D.F.: Cengage Learning Editores, 2010.

WALPOLE, R.E.; MYERS, R.H.; MYERS, S.L.; YE, K. *probabilidad y estadística para ingeniería y ciencias*. 9° edición. Pearson Educación México, 2012.

Received in June 2023.

Approved in November 2023.