MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2023, 19(12), em2371
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/13865

Exploring challenges and strategies in teaching hypothesis testing to engineering students from the perspective of educators

María Lidia Retamal Pérez [1], Hugo Alvarado Martínez [1*], Rosamel Sáez Espinoza [1], Jesús Guadalupe Lugo-Armenta [2]

[1] Universidad Católica de la Santísima Concepción, Concepción, CHILE
[2] Universidad de Los Lagos, Osorno, CHILE

Received 14 December 2022 ▪ Accepted 01 October 2023

Abstract

The study of hypothesis tests is essential in the training of engineers and although in recent years students' knowledge has been investigated, it is necessary to develop studies from the perspective of professors who teach this topic in engineering. For this reason, in this paper we present, from the teachers' perspective, the difficulties and errors exhibited by engineering students, the teaching methodologies and the resources available in relation to hypothesis testing. For this, a conceptual framework with ideas on statistical reasoning for inference was used. In this qualitative research, 19 teachers participated, to whom two questionnaire-type instruments were applied. Among the main results we have that, students have great difficulties with the approach of statistical hypotheses and the study of parameters in two populations, even when teachers attribute a high degree of importance to hypothesis tests in the subject of statistics for engineers.

Keywords: hypothesis tests, engineers training, teaching statistics

INTRODUCTION

In the engineering faculties, there is an important number of subjects considered as critical in basic science, whose passing percentages are not as expected, which brings effects on the retention rate of the first years and the timely graduation rate of the students. In particular, the statistics curricular activity program, which, in spite of involving experiences with the implementation of active methodologies, where students positively value innovations in statistics, increasing motivation, autonomy and application of strategies in data analysis, have declared feeling weaker in the soft skills of time management and computer resources, as well as in information search and hypothesis formulation (Alvarado et al., 2018). According to Ferrés (2017), practical investigative work begins with the process of asking questions and formulating hypotheses, finding out the answers through data analysis and examining the relationship between various variables. Especially, in order to guide research work, authors promote, as a starting point, the identification of research questions as the axis of scientific initiation projects (Caamaño, 2012; Ferrés et al., 2015). The formulation of hypotheses becomes essential in statistical inference and applications in engineering science, however, their implementation and appropriation have been partial, given the curricular and social situation. A pending task is to strengthen the appropriation and correct application of statistical knowledge, moving from data analysis to statistical inference; being the omnipresence of statistical hypothesis testing in the professional work of engineers one of the most relevant issues.

The teaching experience with students at the school and university system, future scientists and engineers, leads us to reflect and to differentiate the work, reasoning and development in the curricular activity programs of statistics and mathematics, which students do not usually distinguish. In connection with our statistical object of study, we conceive the hypothesis as a statement in which something is affirmed or denied from which consequences will emerge. This conception differs on the one hand, from the definition that is only to specify the meaning of what it is about and, on the other hand, from the axiom that must necessarily be admitted, contrary to what happens with the hypotheses (Muñoz & Velarde, 2000).

✉ lretamal@ucsc.cl ✉ alvaradomartinez@ucsc.cl **(*Correspondence)** ✉ rsaez@ucsc.cl ✉ jesus.lugo@ulagos.cl

Contribution to the literature

- This paper aims at characterize the difficulties and errors presented by engineering students, the teaching methodology and the resources available in the professional training of engineers.
- The paper provides guidance to bear in mind when developing strategies for teaching hypothesis testing with university students.
- The paper identifies the connections with other statistical ideas and the assessment of hypothesis tests from the teachers' perspective.

State statistical hypotheses is usually difficult for students since it comprises a system of concepts and properties that makes their study a complex one. Various investigations have documented these difficulties and pointed out the importance of introducing statistical hypotheses in the form of a question and evolving to the approach of hypotheses with statistical language (Lugo-Armenta & Pino-Fan, 2022; Pfannkuch & Wild, 2004; Stohl et al., 2010). Similarly, Inzunza and Jiménez (2013) mention the following emerging elements related to hypothesis testing: Population, sample, statistic test, sample distribution of the statistic test, level of significance, null hypothesis, alternative hypothesis, p-value, rejection region and non-rejection region. Inferring the value that the population parameter can take implies applying the procedures of point estimation, estimation by confidence intervals and hypothesis tests. In turn, such parameter is associated to one statistic that is obtained by some estimation method and evaluated in a random sample, which will be the necessary evidence to refute or not the null hypothesis. Therefore, we state that a hypothesis is a statement to be verified regarding the characteristics of one or more populations. There are usually two types of statements: Those that are related to the parameters and those that are related to the distribution of a random variable. In this research we focus on analyzing the former, considering the different parameters of one and two populations.

Research highlights the difficulties of university students in statistical hypothesis testing due to its complexity and the little amount of time spent in its teaching (e.g., Batanero et al., 2012; Figueroa & Baccelli, 2019; Inzunza & Jiménez, 2013; López-Martín et al., 2019; Vera, 2019). Such investigations present different approaches, namely, they illustrate the way of approaching the hypothesis test, meanings of the hypothesis contrast, associated epistemic configurations and semiotic conflicts, orientations and levels of study, difficulties encountered and the scope and implications. Thus, we consider that there is a lack of studies on teaching practice and reflection of teachers who teach and use statistics. In consequence, we formulate these questions: What are the difficulties and errors that teachers of future engineers show regarding hypothesis testing? Is the formulation of hypothesis tests one of the critical components considered by teachers in the inference? Considering this, in this study we intend to characterize the difficulties and errors presented by engineering students, the teaching methodology and the resources available in the professional training of engineers, the connections with other statistical ideas and the evaluation of hypothesis tests from the perspective of teachers who teach and use statistics (engineers).

CONCEPTUAL FRAMEWORK

We focus on promoting and developing statistical sense in the classroom, which is understood as the union of statistical culture and statistical reasoning (Batanero et al., 2013). On the one hand, statistical literacy implies understanding the fundamental statistical ideas, namely data, graphs, variation, distribution, association and correlation, probability, sampling and inference (Burrill & Biehler, 2011). The importance of these notions lies in the fact that they are the ones that we need in most of the problems, where we need to make use of statistics. On the other hand, statistical reasoning allows decisions to be made, predictions to be made from a set of data considering uncertainty, and it is associated with data analysis skills.

Consequently, students must correctly apply fundamental statistical ideas; one of them being statistical inference, which parameter hypothesis tests arise from. We conceive, by hypothesis testing, the procedure that allows verifying an assertion regarding the value of a population parameter from data obtained in a sample. In addition, in recent years in statistical education, specialists have reported teaching and research experiences on *statistical reasoning for inference*; for example, studying levels of statistical reasoning on hypothesis tests with mathematics students (Inzunza & Jiménez, 2013) and the impact of learning environments with the use of technologies (Cañadas et al., 2012, 2019). Other investigations refer to the incorrect use of statistical hypothesis tests, erroneous interpretations of the results, confusion between null and alternative hypotheses (Vallecillos, 1999), student confusion between statistical and research hypotheses (Chow, 1996), conflict in the logic of the process of Fisher's significance tests and Neyman-Pearson's hypothesis testing (Vallecillos, 1999), confusion between bilateral and unilateral tests, and the regions of acceptance and rejection (Vera, 2017), difficulties in understanding the level of significance α and p-value (López-Martín et al., 2019).

Furthermore, proposals for levels of inferential reasoning on hypothesis testing have emerged, where it is proposed to solve statistical inference problems initially with aspects of informal inference and progressively move towards formal notions and methods of statistical inference (Lugo-Armenta & Pino-Fan, 2021b). The acquisition of statistical concepts must be gradual for its correct application in problem solving. Therefore, a weak understanding of elementary statistical elements leads to serious appropriation difficulties in the sequential chain of contents in random variables, probability distributions and statistical inference. Various investigations have been carried out on the difficulties and errors of students in: Sample representativeness and sample variability (Kahneman et al., 1982; Well et al., 1990); levels of reasoning on inference (Moreno & Vallecillos, 2001); intuitive understanding of the concept of sampling distribution (Díaz & de la Fuente, 2004); understanding of the central limit theorem (Alvarado & Batanero, 2007; Alvarado & Retamal, 2010, 2012; delMas et al., 2004; Lipson, 1997); articulation of the sample distribution and confidence intervals (Alvarado & Segura, 2012; Alvarado et al., 2018; Retamal et al., 2007); and hypothesis tests (Figueroa & Baccelli, 2019; Vera, 2019).

Similarly, Inzunza and Islas (2019), in their study on learning trajectories to develop reasoning in samples, variability and sampling distributions, conclude that an informal approach based on the simulation of sampling with technology contributes to developing correct reasoning regarding sampling distributions and building the basis for understanding statistical inference. In addition, Vera (2019), in research on hypothesis testing, proposes that an element of meaning for statistical inference is the interpretation of results. Moreover, Figueroa and Baccelli (2019) analyze the productions of two problems on hypothesis tests by statistics students of engineering careers, discovering that a group of students fail to identify the appropriate hypotheses according to the parameter and do not discriminate the null hypothesis and alternative.

Therefore, although we are interested in the statistical sense, we are particularly interested in investigating the component of statistical reasoning focused on inference. Statistical inference is one of the classic topics in statistics courses oriented to different university careers. We are especially interested in analyzing the emergence of statistical hypothesis tests in the context of engineering.

METHODS

The procedure to carry out the qualitative research considered, first, a comparative exploratory study of content analysis through a sample of university statistics textbooks, and then an analysis of the difficulties in teaching the hypothesis testing from the institutional perspective.

For the content analysis of the hypothesis tests, 13 texts with chapters directly related to the theme of the hypothesis tests were selected; considering six books from the engineering area, two introductory books to engineering, three books from the business and economics area, and two books on statistics education.

The analysis process was inductive, continuously reviewing the chapters related to the subject oriented to engineering, which gave an approximate vision of what students receive in university education, and allowing the preparation of items 3, 6, 8, 12, and 14 (**Appendix A**) and items 8 and 9 (**Appendix B**).

In this way, we follow the guidelines of researchers in didactics of mathematics, who point out the importance of this didactic resource and of analyzing the meanings of the mathematical objects contained in them, through the characterization of the elements of meanings: Problem situations, language, procedure, concepts, propositions and arguments (Alvarado & Batanero, 2008; Alvarado & Segura, 2012; Carrera et al., 2021; Pino-Fan et al., 2019).

In parallel, items from hypothesis tests applied in related research were analyzed, which together with content analysis allowed determining an institutional reference meaning of the topic. The investigations are varied with different approaches; They give an account of the way of approaching the hypothesis test, meanings of the contrast of hypotheses, orientations and levels of study, difficulties encountered and the scope and implications. This allowed us to elaborate items 9, 10, 11, and 13 (**Appendix A**) and item 7 (**Appendix B**).

Next, a questionnaire guideline for teachers who teach and use statistics was developed. The constructed items, based on content analysis and related research, were consulted by experts, two from statistics and two from statistics education. We opted to use two questionnaires as an instrument of inquiry due to the research context, framed in the health contingency by COVID-19, the limitation of online time and giving teachers adequate time to respond thoughtfully to the proposed items of the questionnaire; so that it reports information on the identification of critical components in the reasoning process, associated concepts and calculations related to hypothesis testing.

In order to select the participants, non-probabilistic sampling by volunteers was used (Vieytes, 2005) thus, the sample of participants is made up of 19 academics. Ten of these academics teach statistics in the first courses of the university program; and nine are engineering teachers, who apply statistics in various situations problems of the specialties. The sample was intentional aiming to obtain several different responses from the teachers, through variables such as years of teaching and research experience, type of institution, where they

Table 1. Errors in the hypothesis tests declared by the teachers & the engineers

Hypothesis tests	Teachers	Engineers
Errors in the selection or approach	8	4
Conceptual errors	4	7
Procedural errors	1	3
Interpretation errors of the result of mathematical calculations	5	6

usually teach and use of means and resources for teaching. The teachers work in five different educational institutions; three are from South Central Chile, one from Argentina and one from Costa Rica. Statistics professors have at least 10 years of teaching experience, whereas engineers have at least 12 years of engineering teaching. The nine engineers teach in *civil engineering science* training programs, and whose preferred dedication is teaching and research.

The two questionnaire-type instruments were applied in one month to teachers who teach statistics and engineering teachers who use statistics, particularly hypothesis tests. Although there are similar items in the two questionnaires, the main differences are the conceptual emphasis on hypothesis testing in the first one and the need to formulate hypotheses in engineering science research projects in the second questionnaire.

In relation to the professors who teach statistics, ten responses were received to the 15-item questionnaire (see **Appendix A**), characterized in

(1) statistics curriculum in engineering careers,

(2) reference textbooks on the subject,

(3) difficulties in learning hypothesis tests,

(4) use of methodology in statistical inference,

(5) connections of hypothesis tests with other important units of statistics,

(6) use of statistical programs as a learning resource,

(7) importance and suggestions for appropriation, and

(8) application of hypothesis tests in engineering sciences.

Regarding the engineering professors, nine answers were received to the 10-item questionnaire (see **Appendix B**), which were structured in

(1) applications of hypothesis tests in engineering science,

(2) difficulties in learning hypothesis tests in engineering science,

(3) use of methodology in statistical inference,

(4) use of statistical programs as a learning resource,

(5) the importance and suggestions for appropriation, and

(6) application of tests of hypothesis in engineering science.

In this way, we were able to identify which the possible critical components of hypothesis tests revealed by teachers who teach and use statistics in engineering are; obtaining a broader vision to provide improvement guidelines in the study processes of hypothesis tests, and the improvement of the teaching of statistics in engineering schools. In the next section we will highlight some of the answers that the participants gave us through the various items of the questionnaires.

RESULTS & DISCUSSION

The main purpose of this study has been to identify the critical components in the understanding of hypothesis tests, taught by university professors who teach and use statistics in engineering careers, by characterizing the difficulties and errors that the engineering students present, the teaching methodology, the available resources, connections with other statistical ideas and assessment of the topic. The items of the questionnaires have been numbered, as follows, for teachers who teach statistics TIn° (**Appendix A**) and for engineers who use statistics EIn° (**Appendix B**).

Then, the errors and difficulties declared by the teachers in the application of the students' hypothesis tests are analyzed, followed by teachers' methodology, the technological resources they use, the relationship or connection with other statistical ideas and the importance that they confer to hypothesis testing.

Errors & Difficulties in Hypothesis Tests

In **Table 1**, we can see the responses of ten statistics teachers and nine engineers regarding learning errors on hypothesis testing. The item was adapted from López-Martín et al. (2019).

TI9–EI7. Where are the biggest errors observed in hypothesis tests by students?

10 teachers point out that the greatest error made by students is observed in the selection or formulation of hypotheses (eight teachers), the same error that is highlighted in the study by Figueroa and Baccelli (2019). It is followed by the interpretation of results (five) and misconceptions (four). This result does not match López-Martín et al. (2019) findings with mathematics teachers in training, where they found that the procedural errors and the interpretation of the results exceeded the conceptual errors and the hypothesis testing approach. The nine engineers who use statistics highlighted as major errors of the students: The conceptual (seven engineers) and the interpretation of results (six). The engineers did not highlight the procedural errors,

Table 2. Answers about the errors in the hypothesis tests declared by the teachers

Teachers' responses to item TI9
T: "Errors arise when the hypotheses are incorrectly stated because, although they identify the parameter, consistency is not maintained when identifying a bilateral or unilateral test with the proposition of the alternative hypothesis and the rejection zone of the null hypothesis."
T3: " Many times, the students do not write the conclusion of the results obtained."
T4: "It is difficult to define the hypotheses correctly. Sometimes, they conclude correctly but misinterpret the information that the problem gives them."
T9: "Errors in the selection or approach because of the lack of understanding of the text."

Table 3. Answers about the errors in the hypothesis tests declared by the engineers

Engineers' responses to item EI7
E1: "It is a topic that is very difficult for them, I'd answer "all of the above", if that option were available: they pose the hypotheses backwards, they present procedural errors (for example, when to use α or α/2), it is very difficult for them to draw correct conclusions based on the available information… but without a doubt the biggest errors are conceptual ones."
E2: "Errors in the selection or approach/Conceptual. I have noticed that the students are good at performing procedural tasks and quickly grasp the idea of interpreting the test result, but they have great difficulty in making the decision to take the test, they only take it when the academic explicitly requests it. In this way, it is evident that there is no clarity in the concepts or the benefits of performing the hypothesis test."
E9: "There are conceptual weaknesses and weaknesses in the approach; therefore, the analysis and interpretation of the results extracted from the data processing in the software are not correct."

understanding that as users of the statistics they are supported by computer resources for which the interpretation of the output result of the software becomes a fundamental part of the statistical work. Indeed, the interpretation of a hypothesis is a fundamental action in the decision-making process itself. However, we sometimes neglect the meaning of failing to reject a null hypothesis, which is often misunderstood as accepting the null hypothesis. More precisely,

> "A hypothesis consists of a simple or complex internally consistent conditional statement with an antecedent that is what is actually hypothetically proposed and a consequent that can be related to verified or verifiable facts. The falsity of the consequent implies that of the antecedent and invalidates the hypothesis, while the verification of the consequent does not allow establishing the necessary truth of the antecedent and, therefore, does not modify its hypothetical character" (Muñoz & Velarde, 2000).

The teachers have highlighted errors due to the lack of reading comprehension and interpretation in hypothesis tests, not writing the conclusions in a proper way. Some examples of responses are shown in **Table 2**.

The arguments of the engineers who use statistics refer to conceptual errors and the difficulty in making decisions when applying the test. In **Table 3**, we can see the responses of engineers 1, 2, and 9.

Based on the responses of teachers who teach and use statistics about student errors, we can say that in order to improve the teaching of hypothesis tests, the approach of hypotheses in context and the results found here, emphasizing on their interpretation should be considered for the decision-making process (Lugo-Armenta & Pino-Fan, 2021a, 2022). In engineering science, the interpretations of the results are important for decision making and it is clear that it is necessary to reinforce the conceptual foundations of hypothesis tests, and that there must be evident connections between the statistics course and the specialty courses, as they do when it comes to the requirements form the first course. It should be noted that one of the engineers who participated in this study indicates that errors occur in all the alternatives in **Table 1**. In relation to this assertion of the teacher who uses statistics (engineer), Inzunsa and Jiménez (2013) point out that university students have shown difficulties in all the stages of the process of a hypothesis test, from approaching the problem to the interpretation of the results.

Next, we will delve into the errors through the difficulties of understanding students' hypothesis tests for which three items are presented that were consulted only to teachers who teach statistics.

TI5. You believe that hypothesis testing is a complex concept even for students with a background in mathematics and probability.

Regarding this item, seven out of ten teachers indicated that it is a complex issue to address due to the number of related concepts, which coincides with the study by Inzunza and Jiménez (2013), where it is emphasized that hypothesis tests are even difficult to students with previous training in mathematical statistics and foundations of probability theory. In **Table 4**, we can see some of the teachers' responses.

Table 4. Teachers' responses on the complexity of the hypothesis tests

Teachers' responses to item TI5
T1: "There are a number of concepts that the student must recognize and relate later, with their appropriate notation, such as the concepts associated with a population and its parameters, or a sample and its statistics, or the sampling distribution of a statistic with the distribution of relative frequencies of its associated statistical variable, the difference between a parameter and a point estimate of it, the difference between a statistic and an estimate."
T2: "They usually have problems to state the statistical hypotheses, to interpret the p-value, or even to select the appropriate hypothesis test for the type of data and problem that we want to solve."
T7: "It is more difficult to establish the hypotheses due to the lack of understanding of logic, which is the basis for understanding the conjectures."

Table 5. Teachers' responses about the greatest difficulties in the hypothesis tests

Teachers' responses to item TI10
TI5: "In general, in the different hypothesis testing procedures, the difficulty persists in the adequate selection of the test statistic on which the decision will be made."
TI3: "I think the two-parameter hypothesis tests are the most complicated, mainly in terms of calculating the degrees of freedom associated with the t-test of two population means."

Table 6. Assignment of the difficulty levels in the hypothesis tests declared by the teachers

Procedure of a HT	1	2	3	%	4	5	%	6	7	%	Total
Definition of null and alternative hypotheses			2	20		1	0	6	1	80	10
Significance level and its interpretation		4	2	60	1	1	10	2		30	10
Identify and compute the test statistic		2	1	30		4	0	1	2	70	10
Set the critical region		2	3	50	3	2	30			20	10
p-value calculation		1	1	20	2	3	20	1	2	60	10
Use classical rejection region method		2	2	40	5		60			0	10
Use method via probability value			2	20	2	3	30	1	1	50	10
Finish and answer the problem		1	2	30	2	1	20	3	1	50	10

The answer given by teacher 7 is related to the need for a logical understanding of hypothesis tests, which was previously pointed out by Vallecillos (1997). Recently developed studies sought to promote the logic of hypothesis testing in order to promote inferential reasoning (Harradine et al., 2011; Liu & Thompson, 2009; Lugo-Armenta & Pino-Fan, 2021b). The following item was applied only in the teachers' questionnaire and deals with students' difficulties in the study of the hypothesis tests.

TI10. Where are the greatest students' difficulties in hypothesis tests?

a. HT of means
b. HT of variances
c. HT of proportions
d. HT of parameters in a population
e. HT of parameters in two populations
f. Others

According to the teachers who participated in this study, the greatest difficulties are observed when performing parameter hypothesis tests in two populations (eight teachers) and variance hypothesis testing in one population (three teachers), mean hypothesis testing in one population (three teachers). Figueroa and Baccelli (2019) determine as a difficulty of the erroneous conceptions in the formulation of hypotheses, the work in one and two populations with the parameter of variances. Next, two arguments from teachers are presented, whose difficulty may be due to the fact that the assumptions that are necessary to apply a particular hypothesis testing procedure are not studied (**Table 5**).

Item TI13 aims that teachers to assess the degree of difficulty of certain procedures that they have observed in their teaching experience. It should be noted that the options that have been provided for the assessment have been taken from various statistical education studies (e.g., López-Martín et al., 2019; Lugo-Armenta & Pino-Fan, 2022; Vera, 2017).

TI13. Assign a score from one to seven on the degree of difficulty that students face in the application of hypothesis tests: One indicates no difficulty and seven high difficulties. Discuss the two highest scores assigned (of highest difficulty) of what you consider most difficult for students to understand (**Table 6**).

Responses were grouped into three levels; from 1 to 3, 4 and 5 to 7. **Figure 1** considers the total amount of the values 5, 6, and 7 representing the levels of highest difficulty measured in percentage.

Teachers rated the definition of the null and alternative hypotheses (80%), identifying and calculating the test statistic (70%), calculating the p-value (60%), using the method via the probability value (50%)

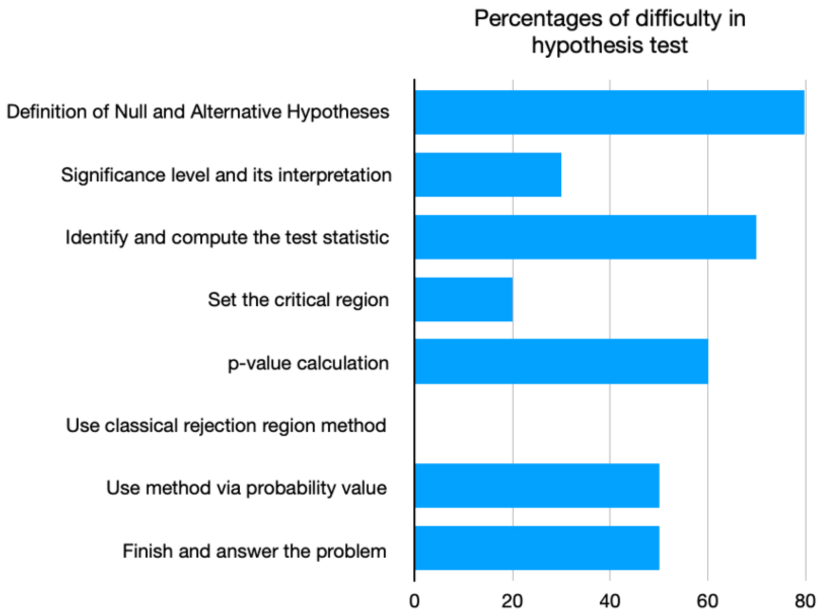

Figure 1. Levels of the highest difficulty expressed as a percentage (sum of values 5, 6, and 7) (Source: Authors' own elaboration)

Table 7. Teachers' responses on the degree of difficulty of the procedures

Teachers' responses to item TI13
T3: "I think that students get quite confused in the calculation of the p-value, since they have difficulties in understanding its definition and how it is calculated in the different types of hypotheses."
T4: "They find it difficult to define the hypotheses correctly. In working with means, they are often confused in choosing the appropriate distribution for calculating the test statistic."
T6: "I have seen that students have problems in stating the hypotheses, they get confused between statistics and parameters, they put=in alternative hypothesis, or they do not understand what the null hypothesis should be according to what was stated."
T9: "Define the null and alternative hypotheses, due to the lack of reading comprehension."

and conclude and answer the problem (50%). The arguments on this item issued by the teachers are related to the hypothesis approach and the calculation of the probability value. We can see some examples in **Table 7**.

On the other hand, teachers considered procedures related to the use of the classic method of the region of rejection and the level of significance as less difficult, because prior to hypothesis tests, confidence intervals are usually studied:

> "I think they can, indeed, indicate the level of significance in a good way, because they understand the level of confidence and therefore its complement, but it is not very clear to them what it means in the context of HT" (T6).

We consider it important that teachers spend time working on each of the stages of hypothesis testing that lead to decision making. Furthermore, research suggests that hypothesis tests should be studied from a progressive perspective, from the informal to the formal (Lugo-Armenta & Pino-Fan, 2021b). In addition, López-Martín et al. (2019) point out that the teaching of the subject should be supported by working on intuition through simulation in sample distributions.

Teaching Methodology for Hypothesis Tests

We present two items from the questionnaires of the current study, which are related to the Neyman-Pearson's approach (comparing the value of the statistic and the critical value) and Fisher's approach (obtaining and interpreting the p-value), with the purpose of delving into which of them is considered more relevant for teaching, in the case of teachers, or more useful, in the case of engineers.

TI12–EI8. According to your experience in the process of studying hypothesis tests, is teaching (utility) more pertinent through the classical method (comparing the value of the statistic and the critical value) and/or the p-value method?

Based on the teachers' experience in the process of studying hypothesis tests, most of them indicated that they use both methods (eight teachers), with the Neyman-Pearson approach being the preferred one due to the procedure followed by the students, and the Fisher's approach because of the scope of hypothesis tests in research and its use with statistical software. For their part, engineers mostly use Fisher's approach; p-value method, (eight engineers), indicating that it is more associated with the development of hypotheses

Table 8. Teachers' responses on the classical method Vs p-value method

Teachers' responses to item TI12
T2: "I believe that both methods should be taught. However, I understand that for some students it is easier to compare the statistic value and the critical value, it seems to make more sense than comparing the p-value and the level of significance."
T3: "Most statistical software provides the p-value, so more emphasis should be placed on this concept, but without ceasing to teach the classical method."
T4: "It is a process that teachers must work on. The way of working with formulas and the classic method is still very ingrained."
T6: "I think that both are pertinent, although much times research is supported with the p-value and that is why it may be more relevant."

Table 9. Engineers' responses on the classical method Vs p-value method

Engineers' responses to item EI8
E9: "In simulation, the classical method is widely used (confidence intervals). However, in regression models and forecasts, the p-value is used much more."
E5: "Conceptually, I prefer the first one, but for practical purposes and with the use of software, the p-value is better, but I do understand its reason for being."
E8: "Both methods, depending on the case. In general, students present more problems in the application of p-value tests."

Table 10. Teachers' responses regarding statistical software in teaching

Teachers' responses to item TI4
T1: "I use GeoGebra because it is free and friendly, since it requires very little information for its use, the online help allows self-learning, and the students can check their results."
T2: "I use Minitab because it allows us to carry out the hypothesis tests in a simple way (without so many calculations), it allows us to visualize the distribution graphs, the rejection zones, histograms or box-and-whisker graphs. In addition, the statistic value and the p-value provide more analysis options."
T5: "Fathom is practical, effective and easy-to-use for the student who does not know how to program, it allows the student to use statistical inference. In general, it offers many options in terms of simulation or mathematical modeling for descriptive and inferential data analysis."

than with the use of software. In **Table 8**, there are examples of teachers' responses to this item.

In the case of the engineers, there were statements, as shown in **Table 9**.

As we can see in **Table 8** and **Table 9**, both teachers and engineers promote both methods, although the research related to these two approaches has presented scientific disagreements, highlighting that difficulties in their application still persist. Thus, Wong et al. (2014) highlight the presence of erroneous conclusions when

> "using the p-value in which the data is analyzed without looking at the data: it only decides yes/no, according to the p-value."

Resources for Teaching Hypothesis Tests

Both teachers and engineers were consulted about the technologies used in their teaching performance. Below are the items and some answers provided by the study participants.

TI4. Do you use statistical programs in the teaching of hypothesis tests? Which ones and what benefits do they present?

In this regard, teachers answered that they consider that using software allows attention to be focused beyond calculations, for example, in the interpretation of results. Out of the 10 teachers, five of them are mainly inclined to the R Studio software because it allows versatility in handling, and the other five teachers preferred Excel. While Excel is not a statistical software, it does have statistical functions and a data analysis plug-in and is widely applicable in business. R Studio is a free distribution software that is quite comfortable to work with. In the case of hypotheses, the p-value is used, whereas for the hypothesis of the mean, it uses t-student, regardless of the sample size, just like SAS. Besides, there are teachers who use GeoGebra (two), Minitab (one), Fathom (one) and SPSS (one). Some comments regarding the use of GeoGebra, Minitab, and Fathom are presented in **Table 10**.

EI6. Do you use statistical software in the application of hypothesis tests in engineering science? Which ones and what benefits do they present?

In engineering science, we observe that there is no predominant software for engineers, they use tools such as Excel, R Studio, SPSS, MATLAB, Statgraphics, and

Table 11. Responses of the engineers on statistical programs in education

Engineers' responses to item EI6
E2: "Python and R. Ease of programming and a wide range of packages/libraries that allow the application of hypothesis tests."
E: "Statgraphics software, better when in Spanish, because it is very friendly and easy for students to use."

Python. Teachers who use statistics commented on the following (see **Table 11**).

Teachers and engineers use various technological resources in the classroom, depending on the conditions of laboratory equipment and software licenses of the institution. Cañadas et al. (2019) evaluated R and SPSS software in terms of utility, ease of use, ease of learning, and satisfaction with 168 first-year university students. They concluded that both are effective and present good results in the application of hypothesis tests in the comparison of means, although they emphasize that SPSS has shown better attitudes in various fields.

Connections Between Hypothesis Tests & Other Statistical Ideas

Firstly, the curricular plans of the engineering careers are striking, eight out of ten teachers teach classes on classic contents of university statistics offered in a single course (TI1). Furthermore, they consider that the course program considers little time for the teaching of the hypothesis testing unit (TI7). When asked about which textbooks they use or recommend in their subjects for the study of statistical inference, in particular hypothesis testing (TI3), teachers mentioned 17 books; the most cited being: Anderson (2016), Devore (2016), Lind (2012), Montgomery and Runger (2004), Navidi (2006), and Walpole et al. (2012).

Secondly, three items from the questionnaires are presented with which it is intended to inquire about the hypothesis tests and their link with the sequence of previous contents, link of descriptive statistics with inferential statistics and preference of teachers in the use of confidence intervals and/or hypothesis tests.

TI6. In the statistical inference unit, what is the sequence of contents that you do to get to the hypothesis tests unit? List the contents from highest to lowest emphasis given.

In their comments on this item, teachers have considered whether the curriculum consists of one or two statistics courses and the specialty of engineering. For example, a teacher taking a statistics course noted the following:

> "I list the order in which we work on the topics: Descriptive statistics, probability, probability distributions, sampling distributions, point and confidence interval estimation, hypothesis tests" (T4).

Another example of an interesting response belongs to teacher 10, who introduces the intuitive idea of hypothesis tests:

> "Before you see sampling distributions. Then, the intuitive concept of hypothesis tests (several intuitive examples are given and the idea of how to decide in real life and how to decide with hypothesis tests is contrasted); then, the theory of hypothesis tests for any statistic is developed, then this theory is applied to see hypotheses tests of a mean number, of a proportion …" (T10).

Teachers rely on classic statistics textbooks for design and classroom learning of concepts and properties; carrying out a text content analysis would show us the complexity of teaching fundamental statistical ideas, specifically about hypothesis test, and the richness of its varied fields of problems and representations (Alvarado & Batanero, 2008).

The presentation and emphasis of hypothesis testing concepts and procedures in textbooks is varied depending on the orientation of the context. It is common to see in textbooks the sequence of statistics content mentioned in the response by teacher 4 (e.g., Devore, 2016; Lind et al., 2019). The study by Lind et al. (2019), which is oriented to the area of business and economics, mentions types of errors I and II, and it shows six steps to carry out a statistical hypothesis test, and also presents exercises with applications using Minitab software and Excel. Besides, Devore (2016), from the area of engineering and science, presents a mathematical level of little difficulty for the learning of students in engineering careers. In the hypothesis test chapter, basic concepts and terminologies are discussed, and then decision-making procedures are developed, using Minitab software to analyze the data. On the other hand, we highlight the study by Milton and Arnold (2004), where a sequence of chapters is presented, starting with probabilities and, descriptive statistics (chapter 6), estimation (chapter 7) and inference linking confidence intervals and hypothesis tests (chapter 8, chapter 9, and chapter 10). Below, we ask the teachers about the importance of this relation.

TI11. Do you think it is feasible to connect descriptive statistics with inferential statistics?

Regarding this item, the teachers indicated that they agreed on the methodological relevance of communicating in the classroom the articulation from the description of the data to the justification of

Table 12. Teachers' responses on the complexity of the hypothesis tests

Teachers' responses to item TI11
T1: "It is essential. The link between a statistical variable and its associated random variable contributes to the understanding of inferential statistics by linking empirical sampling distributions with theoretical sampling distributions."
T2: "Yes, I think it is possible to connect descriptive statistics and inferential statistics. I think that ideas of inference can be promoted from the beginning of the probability and statistics course."
T10: "Given the insufficiency of descriptive statistics to understand certain problems, we endow it with probability, we create a fictitious random situation, this makes the concept of statistical variable (data set) look like a random variable, and so all concepts of the descriptive are redefined to create the inferential."

Table 13. Teachers' responses on confidence intervals and hypothesis tests

Teachers' responses to item TI14
T1: "There is a statement about the value of the population mean. The sample data coincide with the hypothesis that current concentrations exceed this mean."
T2: "I think they might choose to use a hypothesis test since we have the mean set by the federal ozone standard along with a hypothesis: Their current concentrations in air currents are thought to exceed that level."
T5: "Hypothesis tests, since the wording of the problem indicates "In order to verify it", this encourages the student to test a hypothesis."
T6: "I think they would use hypothesis test for a normally distributed mean, most of the time without checking assumptions."

Table 14. Engineers' responses on confidence intervals and hypothesis tests

Engineers' responses to item EI9
E2: "In quality management, we have used confidence intervals more, also as a first idea to determine the range of values where the true population parameter is contained."
E5: "Hypothesis tests are suggested in the area of econometrics or to diagnose multiple linear regression models."

hypotheses based on the sample evidence. **Table 12** shows some examples of teachers' responses.

We consider that the teaching of statistics can take on the challenge of introducing the notions of inference from an informal perspective when teaching descriptive statistics and probability, these notions will be progressively formalized as the study of the fundamental notions for statistical inference and the connections between them are deepened (Lugo-Armenta & Pino-Fan, 2021b).

Teachers were also asked: Which method would students use, confidence intervals or hypothesis test in an engineering context? (TI14 **Appendix A**). The following context was given:

> Ozone is a component of smog, which can cause damage to sensitive plants, even if its concentrations are low. A federal ozone standard of 0.12 ppm average was established. It is thought that its current concentrations in the air exceed this level. In order to verify this, air samples are obtained at 30 monitoring stations in that region, obtaining a sample mean of 0.135 ppm and a sample standard deviation of 0.03 ppm (Milton & Arnold, 2004, p. 299).

Three of the ten teachers indicated that most of the students would apply confidence intervals,

> "because it is easier for them to conclude" (T9).

Seven teachers indicated that the students would apply hypothesis tests based mainly on the wording of the problem.

Table 13 shows examples of the responses provided by teachers to this item.

Confidence intervals and hypothesis tests are commonly used in inference to study the characteristics of a population. Most teachers emphasize the importance of the context of the problem situation to use the most relevant inferential procedure to analyze a set of data. In this case, the hypothesis tests were better appreciated in the contamination situation and where a conjecture is raised.

The previous item, regarding the choice of a method, was also consulted to the engineers, this time without mentioning the previous context but rather considering the applications in the engineering area.

EI9. In engineering science applications, what do you recommend to your students, to use confidence intervals or hypothesis tests?

We observe that, in general, engineers use both confidence intervals and hypothesis tests depending on the objective of the study. Three of them indicated hypothesis tests only. **Table 14** shows examples from engineers' responses.

Table 15. Teachers' responses on the importance of hypothesis tests

Teachers' responses to item TI8
T1: "It is very important, it integrates descriptive statistics and inference, it helps in making decisions on inference problems."
T2: "Making inferences is one of the key aspects of statistics and one that will be most useful to students in their professional lives."

Importance Given to Hypothesis Tests in Basic Science & Engineering Science

Following this further, we ask teachers and engineers for their opinion on the value they attribute to hypothesis tests in their areas of study. Firstly, some responses from teachers are presented and secondly from engineers.

TI8–EI5. What degree of importance do you attribute to the topic of hypothesis tests?

Teachers attribute a high degree of importance to the topic of hypothesis tests as part of their courses, both for the connection between thematic units and for the professional field. They consider it a main topic of any statistics course that is not given the importance it deserves. Generally, little time is available at the end of the teaching semester, here hypothesis tests, the last learning outcome, is taught. This learning outcome is given considerable weight in the final grade of the subject (T6) and therefore can have its scope in academic performance. Other examples of teachers' responses are found in **Table 15**.

Various investigations in the field of engineering have used hypothesis testing as an instrument to verify whether the results obtained from the application of a methodology are significant, aiming to improve the process. Namely, Ortiz et al. (2010) gave great importance to hypothesis tests to validate that the results obtained from the application of autonomous maintenance produced a favorable change in company that manufactures gardening and construction products. Similarly, Wong et al. (2014) applied the t-student test to measure performance in the supply process of a beverage bottling industry, where they established that the average number of inventory days after the implementation of material needs planning is better than the previous process; that is to say, a significant reduction in the average of the inventories was evidenced. Another interesting study is the problem situation of air temperature estimation in the high mountains of Toluca through a terrain elevation model (Soto & Delgado, 2020). The tests that were carried out were the analysis of variance between groups, the t-student test for two samples and the test of similarity or difference of group variances; concluding that the estimated values are statistically similar to the values observed in eight weather stations. As a result, this model can be replicated in other areas.

In engineering, we often work with data samples or simulated data, where possible conditions or factors can be identified to study, modify and make decisions. According to engineers' knowledge and experience in the use of statistics, they indicate as essential the use of hypothesis tests in solving problems in engineering sciences. Specifically, they have indicated as degrees of very important, important, fundamental, high, great relevance; in matters of validation of technological developments in real productive environments (engineer 8), as well as for the practice in civil engineering, hypothesis tests arise in the specialties of hydraulics and hydrology (T4), econometrics, design of experiments, simulation, marketing (T5). Similarly, engineer 9 mentions highlighting the elaboration of hypotheses, preferably using objectives:

> "The formulation of hypotheses is essential in the training of engineers, as well as the setting of objectives and research questions. However, engineering generally works with the formulation of objectives when there is an identified and described problem or opportunity" (T9).

The previous applications have been developed in theses and research oriented to careers in the specialties of civil engineering, industrial civil, computer science, electrical engineering and geological, on simulation issues, regression models, forecasts, use of chi-square tests of independence and goodness of fit.

Specifically, 89% of nine engineers have applied hypothesis tests in undergraduate theses (item EI3) and 67% have participated in a research project, where they had to formulate a hypothesis and then verify its validity (item EI4). The contexts of said projects are mentioned in **Table 16**.

As evidenced in teachers and engineers' the responses, there is consensus that hypothesis tests are appreciated as the most popular tool for evaluating scientific hypotheses and making decisions. Rodríguez (2016) attributes statistics as essential for the work performed by an engineer, highlighting the specialties of information technology and industrial processes.

CONCLUSIONS

In this study we characterize the difficulties and errors presented by engineering students, the teaching methodology and the resources available in the professional training of engineers, the connections with other statistical ideas and the assessment of hypothesis tests from the teachers' perspective, who teach and use statistics (engineers). It has become clear, from the

Table 16. Responses of the engineers on the contexts of the projects

Engineers' responses to item EI4
E1: "Yes. Basically, to compare heuristic methods or to adjust their parameters. Since simulated (random) data is used in the validation stage, the comparison is made with hypothesis tests. Also, in fitting regression models in chemical processes, to see the impact of various parameters on the process, the null hypothesis is that these parameters have no impact on the process."
E3: "Yes. Various problems: Comparison of growth rates, indexes (ecological, biological, & pollutants), temporary abundances, temporary biomasses, presence of temporary organisms, their relationship with environmental conditions (metals & environmental variables), and identification of factors that explain variability."
E8: "Yes, I have participated in research projects where it was necessary to formulate a hypothesis and then verify its validity. The main difficulty has arisen when establishing and demonstrating the sequence "If … Then … Because" or "If ... Then ... Because of ...", especially when there is a certain degree of uncertainty in the analysis of causality due to: (a) small datasets, (b) datasets with a low level of refinement, (b) datasets with imbalance problems, and (d) when there are unobservable variables that can have an important impact on the predictions under different a priori analyses."

responses of the teachers, that there are, indeed, errors in the learning of hypothesis tests, highlighting as one of the greatest difficulties the elaboration of hypotheses as a critical component and its scope for decision-making in problems typical of the engineer.

We agree that hypothesis testing is a topic of statistics with various representations, procedures and arguments, which makes it difficult to appropriate the meaning and effectiveness of its applications. Teachers who teach and use statistics have discovered that their students present a variety of difficulties and errors in hypothesis tests. Teachers indicate that the greatest difficulties in hypothesis tests, mainly referring to the study of parameters in two populations, are linked to the difference in means and ratio of variances (item TI10). When asked about which textbooks are used or recommended, the answers were varied when it comes to courses, where statistical inference is studied, especially in hypothesis tests. We think that competence of text analysis must be present in teachers' professional practice, even more so when in our topic of study, we obtained many concepts that were related to each other, which suggests adequate further review of books that are oriented to certain engineering careers.

We have considered the research of teaching experiences and difficulties in the subject and professional training in statistics from the perspective of teachers in relation to hypothesis tests, to promote preliminary guidelines to consider in the design of understanding of the unit. In this work, the teachers (item ID15) and engineers have proposed some guidelines for the correct application of hypothesis tests, for example, giving spaces for reflection and interpretation, the use of real contexts and a gradual approach, articulating informal inferential reasoning through descriptive statistics with formal reasoning through inferential statistics.

We support the idea of keeping in mind the dedication of time and students' effort to the elaboration of hypotheses, strongly accompanied by the teachers, guiding their argumentation through the transition from descriptive statistics to inferential statistics by various means such as the use of technological learning resources. The work of statistics by projects or the methodology based on the creation of problems located in engineering are actions that could be considered for the development of skills and abilities of the future engineer. Similarly, the articulation of the bridge of descriptive statistics, sampling distributions, confidence intervals and hypothesis tests should be promoted.

Statistics in the university curriculum is recognized for contributing to decision-making based on information, the development of critical thinking and as a base knowledge of all disciplines. Studies support the importance of statistics in engineering (Behar et al., 2002; Rodríguez, 2016). However, it is a thematic area considered difficult by the students, who in turn usually receive decontextualized teaching that makes little sense to them. One of the challenges for teachers is to start the methodological work of elaborating statistical hypotheses, data analysis competence and its correct communication in the first levels, advancing towards the promotion of statistical sense in the school curriculum, which involves developing statistical reasoning for the proper application of fundamental ideas, such as hypothesis testing, in solving statistical situations. Thus, based on what has been analyzed and discussed throughout this article, we believe that this study provides guidance to bear in mind when developing strategies for teaching hypothesis testing with university students, especially engineering students. We also recognize that it is desirable to continue this research, increasing the sample of participants and differentiating teachers who teach and apply hypothesis tests at different higher levels.

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: This study was developed within the framework of research projects: DIREG 06-2021, Dirección de Investigación, Ingeniería 2030 ING222010004 e InEs de Género INGE220011, FAD 2021-24, Universidad Católica de la Santísima Concepción, Chile and Fondecyt 1200005, funded by Agencia Nacional de Investigación y Desarrollo (ANID) of Chile.

Ethical statement: The authors stated that the study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Universidad de Los Lagos (Folio CEC-ULagos H012/2020 and date of approval 7 July 2020). Written informed consents were obtained from the participants.

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Alvarado, H., & Batanero, C. (2007). Dificultades de comprensión de la aproximación normal a la distribución binomial [Difficulties in understanding the normal approximation to the binomial distribution]. *Números. Revista de Didáctica de las Matemáticas* [*Numbers. Mathematics Didactics Magazine*], *67*, 1-7.

Alvarado, H., & Batanero, C. (2008). Significado del teorema central del límite en textos universitarios de probabilidad y estadística [Meaning of the central limit theorem in university texts on probability and statistics]. *Estudios Pedagógicos* [*Pedagogical Studies*], *34*(2), 7-28. https://doi.org/10.4067/S0718-07052008000200001

Alvarado, H., & Retamal, L. (2010). La aproximación binomial por la normal: Una experiencia de reflexión sobre la práctica [The binomial approximation by the normal one: An experience of reflection on practice]. *Paradigma* [*Paradigm*], *31*(2), 89-108.

Alvarado, H., & Retamal, L. (2012). Dificultades de comprensión del teorema central del límite en estudiantes universitarios [Difficulties in understanding the central limit theorem in university students]. *Educación Matemática* [*Mathematics Education*], *24*(3), 119-130.

Alvarado, H., & Segura, N. (2012). Significado de las distribuciones muestrales en textos universitarios de estadística [Meaning of sampling distributions in university statistics texts]. *Revista Electrónica de Investigación en Educación en Ciencias* [*Electronic Journal of Research in Science Education*], *7*(2), 54-71. https://doi.org/10.54343/reiec.v7i2.117

Alvarado, H., Galindo, M., & Retamal, L. (2018). Evaluación del aprendizaje de la estadística orientada a proyectos en estudiantes de ingeniería [Evaluation of project-oriented statistics learning in engineering students]. *Revista Educación Matemática* [*Mathematics Education Magazine*], *30*(3), 151-183. https://doi.org/10.24844/EM3003.07

Batanero, C., Díaz, C., Contreras, J., & Roa, R. (2013). El sentido estadístico y su desarrollo [The statistical sense and its development]. *Números. Revista de Didáctica de las Matemáticas* [*Numbers. Mathematics Didactics Magazine*], *83*, 7-18.

Batanero, C., Vera, O., & Díaz, C. (2012). Dificultades de estudiantes de psicología en la comprensión del contraste de hipótesis [Difficulties of psychology students in understanding hypothesis contrast]. *Números. Revista de Didáctica de las Matemáticas* [*Numbers. Mathematics Didactics Magazine*], *80*, 91-101.

Behar, R., Klinger, R., Olaya, J., Andrade, M. Mesa, E., Conde, G., Delgado, J., Arbeláez, D., Trujillo, P., Solano, H., & Díaz, R. (2002). El rol de la estadística en el trabajo del ingeniero [The role of statistics in the engineer's work]. *Ingeniería y Competitividad* [*Engineering and Competitiveness*], *4*(1), 1-17. https://doi.org/10.25100/iyc.v4i1.2319

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education–A joint ICMI/IASE study* (pp. 57- 69). Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Caamaño, A. (2012). ¿Cómo introducir la indagación en el aula? Alambique [How to introduce inquiry in the classroom? Still]. *Didáctica de Las Ciencias*, *70*, 83-91.

Cañadas, C., Batanero, C., Díaz, C., & Roa, R. (2012). Psychology students' understanding of the Chi-squared test. *Statistique et Enseignement* [*Statistics and Teaching*], *3*(1), 3-18.

Cañadas, G., Molina-Portillo, E., Contreras, J. M., & Álvarez-Arroyo, R. (2019). Las tecnologías en el aula para la enseñanza del contraste de hipótesis [Technologies in the classroom for teaching hypothesis contrast]. *Revista Educación Matemática* [*Mathematics Education Magazine*], *31*(2), 195-211. https://doi.org/10.24844/EM3102.08

Carrera, P. B., Pino-Fan, L. R., Alvarado, H., & Lugo-Armenta, J. G. (2021). Practices of the random variable proposed in the Chilean mathematics curriculum of secondary education. *Mathematics*, *9*(19), 2435. https://doi.org/10.3390/math9192435

Chow, L. S. (1996). *Statistical significance: Rationale, validity and utility*. SAGE.

delMas, R., Garfield, J., & Chance, B. (2004). *Using assessment to study the development of students' reasoning about sampling distributions* [Paper presentation]. Annual Meeting of American Educational Research Association.

Devore, J. (2016). *Probabilidad y estadística para ingeniería y ciencias* [*Probability and statistics for engineering and science*]. Cengage Learning Publishers.

Díaz, C., & de la Fuente, I. (2004). Controversias en el uso de la inferencia en la investigación experimental [Controversies in the use of inference in experimental research]. *Metodología de las Ciencias del Comportamiento* [*Behavioral Sciences Methodology*], *2004*, 161-167.

Ferrés, C. (2017). El reto de plantear preguntas científicas investigables [The challenge of raising researchable scientific questions]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias* [*Eureka Magazine on Science Teaching and Dissemination*], *14*(2), 410-426. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2017.v14.i2.09

Ferrés, C., Marbà, A., & Sanmartí, N. (2015). Trabajos de indagación de los alumnos: Instrumentos de evaluación e identificación de dificultades [Student inquiry work: Evaluation instruments and identification of difficulties]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias* [*Eureka Magazine on Science Teaching and Dissemination*], *12*(1), 22-37. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2015.v12.i1.03

Figueroa, S., & Baccelli, S. (2019). Significados personales sobre los contrastes de hipótesis en estudiantes de ingeniería [Personal meanings about hypothesis contrasts in engineering students]. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística* [*Proceedings of the 3rd Virtual International Congress on Statistics Education*].

Inzunza, S., & Islas, E. (2019). Diseño y evaluación de una trayectoria hipotética de aprendizaje para intervalos de confianza basada en simulación y datos reales [Design and evaluation of a hypothetical learning trajectory for confidence intervals based on simulation and real data]. *Bolema: Boletim de Educação Matemática* [*Bulletin: Mathematics Education Bulletin*], *33*(63), 1-26. https://doi.org/10.1590/1980-4415v33n63a01

Inzunza, S., & Jiménez, J. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis [Characterization of the statistical reasoning of university students about hypothesis testing]. *Revista Latinoamericana de Investigación en Matemática Educativa* [*Latin American Journal of Research in Educational Mathematics*], *16*(2), 179-211. https://doi.org/10.12802/relime.13.1622

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809477

Lind, D., Marchal, W., & Wathen, S. (2019). *Estadística aplicada a los negocios y la economía* [*Statistics applied to business and economy*]. McGraw Hill.

López-Martín, M., Batanero, C., & Gea, M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en inferencia estadística? [Do future teachers know their students' errors in statistical inference?] *Bolema: Boletim de Educação Matemática* [*Bulletin: Mathematics Education Bulletin*], *33*(64), 672-693. https://doi.org/10.1590/1980-4415v33n64a11

Lugo-Armenta, J. G., & Pino-Fan, L. R. (2021a). Inferential reasoning of secondary school mathematics teachers on the Chi-square statistic. *Mathematics*, *9*(19), 2416. https://doi.org/10.3390/math9192416

Lugo-Armenta, J. G., & Pino-Fan, L. R. (2021b). Niveles de razonamiento inferencial para el estadístico t-student. *Bolema: Boletim de Educação Matemática* [*Bulletin: Mathematics Education Bulletin*], *35*, 1776-1802. https://doi.org/10.1590/1980-4415v35n71a25

Lugo-Armenta, J. G., & Pino-Fan, L. R. (2022). Razonamiento inferencial de docentes de matemáticas de enseñanza media sobre el estadístico t-student [Inferential reasoning of high school mathematics teachers on the t-student statistic]. *Uniciencia* [*Unscience*], *36*(1), 1-29. https://doi.org/10.15359/ru.36-1.25

Milton, S., & Arnold, J. (2004). *Probabilidad y estadística con aplicaciones para ingeniería y ciencias computacionales* [*Probability and statistics with applications for engineering and computer science*]. McGraw Hill.

Moreno, A., & Vallecillos, A. (2001). Exploratory study on inferential' concepts learning in secondary level in Spain. In M. van der Heuvel (Ed.), *Proceedings of the 25th Conference of the International Group of the Psychology of Mathematics Education*. Freudenthal Institute and Utrecht University.

Muñoz, J., & Velarde, J. (2000). *Compendio de epistemología* [*Compendium of epistemology*]. Trotta Publishing.

Ortiz, F., Viclhis, J., Báez, O., & Alvarado, A. (2010). Aplicación de pruebas de hipótesis para validar los tiempos de paro después de aplicar mantenimiento autónomo [Application of hypothesis tests to validate downtime after applying autonomous maintenance]. *Unidad Profesional Interdisciplinaria de Ingeniería, Ciencias Sociales y Administrativas UPIICSA* [*Interdisciplinary Professional Unit of Engineering, Social and Administrative Sciences UPIICSA*], *18*(7), 14-21.

Pfannkuch, M., & Wild, C. (2004). Towards and understanding of statistical thinking. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Kluwer Academic Publisher. https://doi.org/10.1007/1-4020-2278-6_2

Pino-Fan, L., Parra-Urrea, Y., & Castro, W. F. (2019). Significados de la función pretendidos por el currículo de matemáticas Chileno [Function meanings intended by the Chilean mathematics curriculum]. *Revista Internacional de Investigación en Educación* [*International Journal of Research in Education*], *11*(23), 201-220. https://doi.org/10.11144/Javeriana.m11-23.sfpc

Retamal, L., Alvarado, H., & Rebolledo, R. (2007). Comprensión de las distribuciones muestrales en un curso de estadística para ingenieros [Understanding sampling distributions in a statistics course for engineers]. *Ingeniare* [*Engineer*], *15*, 6-17. https://doi.org/10.4067/S0718-33052007000100002

Rodríguez, D. (2016). Importancia de la estadística en ingeniería [Importance of statistics in engineering]. *Revista Iberoamericana de Producción Académica y Gestión Educativa* [*Ibero-American Journal of Academic Production and Educational Management*], *3*(6), 17.

Soto, V., & Delgado, H. (2020). Estimación de la temperatura del aire en la alta montaña mexicana mediante un modelo de elevación del terreno: Caso del volcán Nevado de Toluca (México) [Estimation of air temperature in the Mexican high mountains using a terrain elevation model: Case of the Nevado de Toluca volcano (Mexico)]. *Ería, Revista Cuatrimestral de Geografía* [*Ería, Quarterly Magazine of Geography*], *XL*(2), 167-182. https://doi.org/10.17811/er.2.2020.167-182

Stohl Lee, H., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal, Auckland, 9*(1), 68-96. https://doi.org/10.52041/serj.v9i1.388

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. In *Proceedings of the 52nd Session of the International Statistical Institute* (pp. 201-204).

Vallecillos, A., & Batanero, C. (1997). Aprendizaje y enseñanza del contraste de hipótesis: Concepciones y errores [Learning and teaching hypothesis testing: Conceptions and errors]. *Enseñanza de las Ciencias* [*Science Teaching*], *15*(2), 189-197. https://doi.org/10.5565/rev/ensciencias.4175

Vera, O. (2017). Análisis de varianza elemental versus contraste de hipótesis: Comprensión de las hipótesis estadísticas mediante la identificación y análisis de conflictos semióticos comparados [Elemental variance analysis versus hypothesis contrast: Understanding statistical hypotheses through the identification and analysis of comparative semiotic conflicts]. In J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M. Gea, B. Giacomone, & M. López-Martín (Eds.), *Actas del Segundo Congreso International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos* [*Proceedings of the 2nd International Virtual Congress on the Ontosemiotic Approach to Mathematical Knowledge and Instruction*].

Vera, O. (2019). Un elemento de significado en el contraste de hipótesis: La interpretación de los resultados [An element of meaning in the hypothesis contrast: The interpretation of the results]. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística* [*Proceedings of the 3rd Virtual International Congress on Statistics Education*].

Vieytes, R. (2005). *Metodología de la investigación en organizaciones, mercado y sociedad* [*Research methodology in organizations, market and society*]. Sciences Publishing House.

Well, A., Pollatsek, A., & Boyce, S. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Processes, 47*, 289-312. https://doi.org/10.1016/0749-5978(90)90040-G

Wong, J., Ruiz, C., Lerma, D., Lugo, F., & Delgado, X. (2014). Caso de estudio de utilización de prueba de hipótesis en la medición del desempeño en el proceso de suministro de una industria embotelladora [Case study of the use of hypothesis testing in measuring performance in the supply process of a bottling industry]. In M. Ramos, & V. Aguilera (Eds.), *Ciencias de la ingeniería y tecnología, handbook T-V ECORFAN* [*Engineering sciences and technology, handbook T-V ECORFAN*] (pp. 1-19).

APPENDIX A

Questionnaire Applied to Teachers Who Teach Statistics

Exploratory study based on the experience of teachers about the difficulties of understanding hypothesis tests in engineering careers.

Item 1. Does the curriculum, where you teach establish one or two statistics courses with the classic content oriented to engineering careers?

Item 2. Your experience teaching hypothesis tests has been with students specializing in: …

Item 3. What textbooks do you use or recommend in your courses for the study of statistical inference, in particular hypothesis tests?

Item 4. Do you use statistical software in the teaching of hypothesis tests? Which ones and what benefits do they present?

Item 5. Do you consider that hypothesis tests are a complex concept even for students who have a background in mathematics and probability?

Item 6. In the statistical inference unit, what is the sequence of contents that you follow to get to the hypothesis tests unit? Write them from highest to lowest emphasis.

Item 7. How much time do you have for teaching hypothesis tests? Do you consider that the time is enough to develop the unit?

Item 8. What degree of importance do you attribute to the topic of hypothesis tests in the course?

Item 9. Where are the greatest errors observed in hypothesis tests by students? Comment

a. Errors in the selection or approach
b. Conceptual errors
c. Procedural errors
d. Interpretation errors of the result of mathematical calculations

Item 10. Where are the students' greatest difficulties in hypothesis tests? Comment

a. HT of means
b. HT of variances
c. HT of proportions
d. HT of parameters in a population
e. HT of parameters in two populations
f. Others

Item 11. Do you see it feasible to connect descriptive statistics with inferential statistics?

Item 12. According to your experience in the process of studying hypothesis tests, is teaching by using the classical method (comparing the value of the statistic and the critical value) and/or the p-value method more relevant?

Item 13. Rate the degree of difficulty of the students in the application of hypothesis tests: 1 indicates no difficulty and 7 great difficulties: 1 2 3 4 5 6 7

a. Define the null and alternative hypotheses
b. Indicate level of significance and its interpretation
c. Identify and compute the test statistic
d. Set the critical region
e. Calculate p-value
f. Use classical rejection region method
g. Use method via probability value
h. Finish and answer the problem
i. Discuss the two highest scores assigned (of highest difficulty) of what you consider most difficult for students to understand.

Item 14. Ozone is a component of smog, which can cause damage to sensitive plants, even if its concentrations are low. A federal ozone standard of 0.12 ppm average was established. It is thought that its current concentrations in the air exceed this level. In order to verify this, air samples are obtained at 30 monitoring stations in that region, obtaining a sample mean of 0.135 ppm and a sample standard deviation of 0.03 ppm

As a verification methodology, which method do you think most students would use, confidence intervals or hypothesis tests? Argue without resolving it.

Item 15. Based on your teaching experience, what guidelines do you suggest for the understanding and application of students in hypothesis tests?

APPENDIX B

Questionnaire Applied to Engineering Teachers Who Use Statistics

Exploratory study through the experience of teachers about the difficulties of understanding hypothesis tests in engineering careers.

Item 1. In the courses you have taught, what areas of the specialty have required the application of hypothesis tests?

Item 2. In relation to the previous question, indicate the engineering careers that have applied hypothesis tests.

Item 3. Have you required application of hypothesis tests in the undergraduate theses you have participated in?

Item 4. Have you participated in any research project, where you had to formulate a hypothesis and then verify its validity? If possible, can you comment on what was discussed and what difficulty you encountered?

Item 5. According to your experience and knowledge, what degree of importance do you attribute to statistical inference, in particular hypothesis tests, in engineering science?

Item 6. Do you use statistical software in the application of hypothesis tests in engineering science? Which ones and what benefits do they present?

Item 7. In applications of hypothesis tests in engineering science, where are the greatest students' errors observed? Comment

a. Errors in the selection or approach

b. Conceptual errors

c. Procedural errors

d. Interpretation errors of the result of mathematical calculations

Item 8. Based on your experience in hypothesis test applications, do you use the classical method (compare statistical value and critical value) and/or the p-value method?

Item 9. In applications in engineering science, what do you recommend your students use, confidence intervals or hypothesis tests?

Item 10. Based on your teaching experience, what guidelines do you suggest for students' understanding and application of hypothesis tests in engineering science?

https://www.ejmste.com